

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΕΚΤΙΜΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΔΙΣΔΙΑΣΤΑΤΗΣ ΚΑΙ  
ΤΡΙΣΔΙΑΣΤΑΤΗΣ ΔΟΜΗΣ ΚΑΙ ΚΙΝΗΣΗΣ  
ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ ΑΚΟΛΟΥΘΙΕΣ ΕΙΚΟΝΩΝ

Ιωάννης Σ. Ξηρουχάκης

Ηλ/γος Μηχ/κός και Μηχ/κός Η/Υ Ε.Μ.Π.

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2001

## Σημείωμα

Ο υπογράφων επιθυμεί να ευχαριστήσει τον κ. Αναστάσιο Ντελόπουλο, Διδάκτορα Ερευνητή ΕΠ.Ι.Σ.Ε.Υ., χωρίς τη συμβολή του οποίου το παρόν θα ήταν αρκετά φτωχότερο. Επίσης, τους κκ. Αθανάσιο Δροσόπουλο, Γαβριήλ Τσεχπενάκη και Γεώργιο Βότση, Υποψήφιους Διδάκτορες Ε.Μ.Π., και τον κ. Ιωάννη Αβρίθη, Διδάκτορα Ε.Μ.Π., για τη συμβολή τους στις επιστημονικές εργασίες που δημοσιεύθηκαν.

Ο υπογράφων επιθυμεί να εκφράσει τις θερμές του ευχαριστίες στον κ. Στέφανο Κόλλια, Καθηγητή Ε.Μ.Π., για την κατανόηση και υποστήριξή του από τον Σεπτέμβριο του 1996, οπότε και ο υπογράφων εντάχθηκε στο δυναμικό του Εργαστηρίου Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων Ε.Μ.Π. (ΨΕΕΒΠ) για την εκπόνηση της πτυχιακής του εργασίας, καθώς και τα μέλη της τριμελούς επιτροπής Ανδρέα Σταφυλοπάτη και Παναγιώτη Τσανάκα, Καθηγητές Ε.Μ.Π.. Επίσης, να ευχαριστήσει το σύνολο του ερευνητικού προσωπικού του εργαστηρίου ΨΕΕΒΠ για τη δημιουργική τετράχρονη συμβίωση.

Ο υπογράφων οφείλει να εκφράσει τις ευχαριστίες του στους γονείς του για την ηθική και υλική τους υποστήριξη καθ'όλη τη διάρκεια των σπουδών του. Τέλος, να ευχαριστήσει το σύνολο των μελών της φοιτητικής ομάδας Ε.Μ.Π. 'NYΓMA' για τη συνεχή τους συμπαράσταση και τα βαθιά αμοιβαία αισθήματα φιλίας.

Το παρόν αφιερώνεται στην Τόνια για την υποστήριξη και υπομονή της καθ'όλη τη διάρκεια της συγγραφής του.

Ιωάννης Σ. Ξηρουχάκης

## Περιεχόμενα

σελίδες

Σημείωμα . . . . .	ii
Πρόλογος . . . . .	vii
Σχήματα . . . . .	xii
Πίνακες . . . . .	xviii
1. Εισαγωγή . . . . .	1
I Εκτίμηση Χαρακτηριστικών Δισδιάστατης Δομής και Κίνησης . . . .	7
2. Εκτίμηση Χαρακτηριστικών στις Δύο Διαστάσεις . . . . .	8
3. Εξαγωγή και Μοντελοποίηση της Χωρικής Πληροφορίας Χρώματος Εικό- νων . . . . .	16
3.1 Εισαγωγή / Αναδρομή . . . . .	16
3.2 Το χρώμα ως χαρακτηριστικό της εικόνας . . . . .	20
3.3 Εξαγωγή χρωματικών αντικειμένων . . . . .	22
3.4 Αναπαράσταση με χρήση γράφου γειτνίασης . . . . .	25
3.5 Ελάττωση του γράφου γειτνίασης . . . . .	27
3.6 Ταυτοποίηση χαρακτηριστικών με χρήση μετρικών ομοιότητας . . . . .	31
3.6.1 Η μετρική ‘κάλυψη’ . . . . .	32
3.6.2 Η μετρική ‘συνδεσιμότητα’ . . . . .	35
3.6.3 Η συνολική μετρική ομοιότητας . . . . .	37
3.7 Πειράματα και συζήτηση . . . . .	38
3.8 Συμπεράσματα και επεκτάσεις . . . . .	40

<b>4. Εντοπισμός και Εξαγωγή Κύριων Κινούμενων Αντικειμένων σε Ακολουθίες Εικόνων</b>	<b>42</b>
4.1 Εισαγωγή / Αναδρομή	42
4.2 Τοποθέτηση του προβλήματος	46
4.3 Η μονάδα κανονικοποίησης φόντου	47
4.4 Η μονάδα ενημέρωσης φόντου	49
4.5 Εξαγωγή κινούμενων ακμών	51
4.6 Ταξινόμηση κινούμενων ακμών	52
4.7 Εξαγωγή κινούμενων αντικειμένων	53
4.8 Μία πολυδιακριτική υλοποίηση	55
4.9 Προσαρμογή παραμέτρων	56
4.10 Ο αλγόριθμος σε βήματα	58
4.11 Πειράματα και συζήτηση	60
4.12 Συμπεράσματα και επεκτάσεις	66
<b>5. Κανονικοποίηση Καμπυλών για Μοντελοποίηση Αντικειμένων με βάση το Σχήμα</b>	<b>68</b>
5.1 Εισαγωγή / Αναδρομή	68
5.2 Περιγραφή του προβλήματος	72
5.3 Αναπαράσταση με χρήση B-splines	74
5.3.1 Μοντελοποίηση καμπύλης	74
5.3.2 Αναδιάταξη των σημείων σύνδεσης	77
5.4 Ορθογωνοποίηση καμπύλης	78
5.4.1 Η διαδικασία ορθογωνοποίησης	79
5.4.2 Αναλλοίωτο σε μετατόπιση και μεγέθυνση	82
5.5 Κανονικοποίηση σημείου έναρξης και περιστροφής	84
5.5.1 Κανονικοποίηση σημείου έναρξης	85
5.5.2 Κανονικοποίηση περιστροφής και κατοπτρισμού	88
5.6 Πειράματα και συζήτηση	91
5.7 Συμπεράσματα και επεκτάσεις	100

<b>II</b>	<b>Εκτίμηση Χαρακτηριστικών Τρισδιάστατης Δομής και Κίνησης . . .</b>	<b>103</b>
<b>6.</b>	<b>Εκτίμηση Χαρακτηριστικών στις Τρεις Διαστάσεις . . . . .</b>	<b>104</b>
<b>7.</b>	<b>Υπολογισμός Τρισδιάστατης Δομής και Κίνησης από Προβολές Τρισδιάστατων Στερεών Αντικειμένων . . . . .</b>	<b>109</b>
7.1	Εισαγωγή / Αναδρομή . . . . .	109
7.2	Ο πίνακας περιστροφής επιπέδου . . . . .	114
7.2.1	Το μοντέλο παράλληλης προβολής . . . . .	114
7.2.2	Προκαταρκτικοί ορισμοί . . . . .	114
7.2.3	Υπολογισμός των πινάκων περιστροφής επιπέδου βάσει των διανυσμάτων κίνησης . . . . .	116
7.3	Υπολογισμός των παραμέτρων δομής και κίνησης . . . . .	117
7.4	Υπολογισμός παρουσία θορύβου . . . . .	123
7.5	Συζήτηση και σύγκριση με άλλες μεθόδους . . . . .	134
7.6	Πειράματα και συζήτηση . . . . .	137
7.6.1	Πεδία κίνησης χωρίς σφάλματα / θόρυβο . . . . .	137
7.6.2	Θορυβώδη διανύσματα κίνησης . . . . .	139
7.6.3	Θορυβώδεις αντιστοιχίσεις σημείων . . . . .	145
7.6.4	Σύγκριση με υπάρχουσες μεθόδους . . . . .	146
7.7	Συμπεράσματα και επεκτάσεις . . . . .	149
<b>8.</b>	<b>Υπολογισμός Τρισδιάστατης Κίνησης της Κάμερας σε Εικονικά Περιβάλλοντα Καταγραφής . . . . .</b>	<b>151</b>
8.1	Εισαγωγή / Αναδρομή . . . . .	151
8.2	Περιγραφή του συστήματος . . . . .	155
8.2.1	Σχεδιασμός του συστήματος οπτικής παρακολούθησης . . . . .	155
8.2.2	Σχεδιασμός του κυανού πετάσματος . . . . .	157
8.2.3	Εντοπισμός των δισδιάστατων χαρακτηριστικών . . . . .	158
8.2.4	Υπολογισμός των παραμέτρων τρισδιάστατης κίνησης της κάμερας . . .	158
8.3	Κατασκευή του δυαδικού πίνακα του πετάσματος . . . . .	161

8.3.1	Ένα ειδικό δισδιάστατο δυαδικό πεδίο . . . . .	162
8.3.2	Ο αλγόριθμος κατασκευής . . . . .	164
8.4	Υπολογισμός της θέσης και του προσανατολισμού της κάμερας . . . . .	166
8.5	Ασυμπτωτική ανάλυση σφάλματος . . . . .	175
8.6	Πειράματα και συζήτηση . . . . .	177
8.6.1	Πειράματα με εικονική κάμερα . . . . .	177
8.6.2	Πειράματα με πραγματική κάμερα . . . . .	183
8.7	Συμπεράσματα και επεκτάσεις . . . . .	184
9.	Συμπεράσματα και Επεκτάσεις . . . . .	187
	Βιβλιογραφία . . . . .	191

## Πρόλογος

Το παρόν αποτελεί την διδακτορική διατριβή του Ιωάννη Σ. Ξηρουχάκη, Ηλεκτρολόγου Μηχανικού και Μηχανικού Ηλεκτρονικών Υπολογιστών Ε.Μ.Π., η οποία εκπονήθηκε στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων Ε.Μ.Π. (ΨΕΕΒΠ). Το περιεχόμενό του βασίζεται στην έρευνα που διεξήγαγε στα πλαίσια της διδακτορικής του διατριβής στις επιστημονικές περιοχές της Όρασης Υπολογιστών και της Κατανόησης Περιεχομένων Βίντεο κατά τη χρονική περίοδο Σεπτέμβριος 1997 - Ιούνιος 2001.

Οι περιοχές της Όρασης Υπολογιστών (Computer Vision) και της Κατανόησης Περιεχομένων Εικόνας και Βίντεο (Image and Video Understanding) αποτελούν επιστημονικά αντικείμενα με μεγάλο ενδιαφέρον, αφού επιχειρούν να προσεγγίσουν δυσεπίλυτα προβλήματα, που σχετίζονται εν πολλοίς με τη δυνατότητα των υπολογιστικών μηχανών (π.χ. ρομπότ) να βλέπουν, να επεξεργάζονται, να κατανοούν και να αντιλαμβάνονται το περιβάλλον τους. Οι δυνατότητες της αντίληψης (perception) και της κατανόησης (understanding) των υπολογιστικών μηχανών απασχολούν κατά ένα αρκετά μεγάλο ποσοστό την επιστημονική κοινότητα των υπολογιστών, ενώ στο άμεσο μέλλον αναμένεται να αποτελέσουν την κυριότερη ίσως κατεύθυνση της επιστημονικής έρευνας. Ουσιαστικά, η αντίληψη των υπολογιστικών μηχανών αντιπροσωπεύει τη δυνατότητά τους να εντοπίζουν και να αναγνωρίζουν συγκεκριμένα χαρακτηριστικά των εικόνων (π.χ. χρώμα, σχήμα, κίνηση κ.λπ.), τα οποία με τη σειρά τους επεξεργάζονται από μόνα τους (π.χ. υπολογισμός τρισδιάστατης κίνησης από τα δεδομένα κίνησης της εικόνας) ή σε σύγκριση με ήδη γνωστά δεδομένα (π.χ. σύγκριση σχήματος αντικειμένου της εικόνας με υπάρχοντα πρότυπα). Σ' αυτά τα πλαίσια, η κατανόηση αναφέρεται στην αξιοποίηση της ωφέλιμης πληροφορίας που μπορεί να εξαχθεί. Ένα υπολογιστικό σύστημα που έχει, υπό αυτή την έννοια, αντίληψη και μπορεί να κατανοεί το περιβάλλον του, έχει αυτόματα και τη δυνατότητα αποφάσεων και αντιδράσεων στα εξωτερικά ερεθίσματα.

Στο πλαίσιο του παρόντος, προτείνεται ουσιαστικά ένας αριθμός από πρωτότυπες μεθόδους προς την κατεύθυνση του υπολογισμού χαρακτηριστικών δομής και κίνησης από εικόνες ή ακολουθίες εικόνων. Κατά περίπτωση και ανάλογα με την εφαρμογή, τα χαρακτηριστικά αυτά δημιουργούν αντίληψη στο υπολογιστικό σύστημα, την οποία μπορεί να εκμεταλλευτεί για να εκτελέσει διάφορες διεργασίες όπως η ταξινόμηση, η αναζήτηση ή ακόμη και η τρισδιάστατη μοντελοποίηση αντικειμένων.

Όπως μπορεί κανείς να αντιληφθεί, τα χαρακτηριστικά δομής και κίνησης που μπορούν να εξαχθούν από τις εικόνες και τις ακολουθίες τους είναι ποικίλα - περισσότερο δε αυτό ισχύει

για τα χαρακτηριστικά δομής αφού στην έννοια ‘δομή’ μπορούν να δοθούν διάφορες σημασίες. Στο πλαίσιο της διατριβής, αποφεύχθηκε η βιβλιογραφική κάλυψη κάθε δυνατής αντιμετώπισης των χαρακτηριστικών δομής / σχήματος ή κίνησης. Αντίθετα, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στην πλούσια βιβλιογραφία, η οποία αναλύεται εν τάχει στις εισαγωγικές υποενότητες των κεφαλαίων, ενώ στο κύριο σώμα της διατριβής δίνεται έμφαση στις πρωτότυπες μεθόδους που αναπτύχθηκαν.

Στο Κεφάλαιο 1, αντί εισαγωγής τοποθετείται το πρόβλημα της εξαγωγής χαρακτηριστικών, η σημασία του και η ευρεία εφαρμογή του. Αναφέρονται οι επιστημονικές περιοχές που το θέμα αυτό αγγίζει, καθώς και το πού βρίσκεται η επιστημονική πρόοδος αλλά και το μελλοντικό επιστημονικό ενδιαφέρον στο αντικείμενο. Στη συνέχεια, η διατριβή χωρίζεται πρακτικά σε δύο μέρη, με κριτήριο τη διάσταση των χαρακτηριστικών που θέλουμε να εξάγουμε. Ο διαχωρισμός αυτός δεν είναι αναγκαίος, χρησιμοποιήθηκε όμως για να γίνει φανερή η διάκριση στην πολυπλοκότητα και στη μεθοδολογία των δισδιάστατων / τρισδιάστατων προβλημάτων. Το γενικότερο πλαίσιο της έρευνας για εξαγωγή δισδιάστατων και τρισδιάστατων χαρακτηριστικών δομής και κίνησης περιγράφεται στα Κεφάλαια 2 και 6.

Το Κεφάλαιο 3 ασχολείται με το πρόβλημα της εξαγωγής τοπολογικής πληροφορίας χρώματος από εικόνες. Η ‘κατανομή χρώματος’ είναι το πρώτο σχετικά υψηλού επιπέδου χαρακτηριστικό δομής που μπορεί να λάβει κανείς από μία εικόνα. Επειδή ακριβώς μοντελοποιεί το χρώμα και την κατανομή του επάνω στην εικόνα στα πλαίσια χρωματικών περιοχών, αποτελεί χρήσιμο χαρακτηριστικό για τη σύγκριση εικόνων (ιδιαίτερα σε εφαρμογές αναζήτησης, ανάκτησης και ταξινόμησης εικόνων). Το κεφάλαιο βασίζεται στην εργασία [189]\* των Ξηρουχάκη και Ντελόπουλου και αποτελεί μία πρωτότυπη μέθοδο εξαγωγής, μοντελοποίησης και σύγκρισης του χαρακτηριστικού της κατανομής χρώματος. Ο αλγόριθμος αυτός περιγράφηκε σε αδρές γραμμές νωρίτερα στην αναφορά [187] των ιδίων, γενικότερα όμως βασίζεται στην εργασία [185] των Ξηρουχάκη, Τυράκη και Ντελόπουλου.

Το Κεφάλαιο 4 βασίζεται στο θέμα του εντοπισμού και εξαγωγής κινούμενων αντικειμένων σε ακολουθίες εικόνων. Στο κεφάλαιο αυτό γίνεται προσπάθεια κάλυψης όλων των δυνατών περιπτώσεων καταγραφής, με στατική / κινούμενη κάμερα σε εσωτερικό / εξωτερικό χώρο κ.λπ., βάσει ενός πρωτότυπου παραμετροποιημένου συστήματος που προτείνεται. Ουσιαστικά πραγματοποιείται εξαγωγή της δισδιάστατης κίνησης των κύριων (μεϊζονων) κινούμενων αντικειμένων στις ακολουθίες εικόνων, ενώ παράλληλα η δομή τους (το σχήμα του περιγράμματός τους) μοντελοποιείται με κλειστά πολύγωνα. Το κεφάλαιο βασίζεται στην εργασία [191] των Ξηρουχάκη, Τσεχπενάκη και Ντελόπουλου. Μέρη του αλγορίθμου αυτού εμφανίστηκαν και

---

\*Οι βιβλιογραφικές αναφορές σημειώνονται με τον αντίστοιχο αριθμό εντός αγκυλών.



στη σχετική αναφορά [166] των ιδίων, ενώ κάποιες πρώτες ιδέες δόθηκαν στην εργασία [194] των Ξηρουχάκη, Μαθιουδάκη και Ντελόπουλου.

Στο Κεφάλαιο 5 εξετάζουμε τη μοντελοποίηση των αντικειμένων με βάση το σχήμα, το οποίο αναπαριστούμε με το κανονικοποιημένο περίγραμμά του. Πρακτικά, με χρήση ακριβούς κατάτμησης κίνησης ή/και χρώματος εξάγουμε τα περιγράμματα αντικειμένων με σημασιολογικό περιεχόμενο από ακολουθίες εικόνων, τα μοντελοποιούμε με κλειστές B-splines και στη συνέχεια προτείνουμε μία τεχνική κανονικοποίησής τους, ώστε να είναι διαθέσιμα προς σύγκριση ανεξάρτητα των μετασχηματισμών όψεως (affine transformations) ή των μικρών αλλοιώσεων (deformations) που έχουν υποστεί. Το κεφάλαιο βασίζεται στην πρωτότυπη εργασία [9] των Αβρίθη, Ξηρουχάκη και Κόλλια. Μέρη της μεθόδου και συναφής μεθοδολογία παρουσιάστηκαν νωρίτερα στις αναφορές [8, 7] των ιδίων, ενώ η σημασία τέτοιων μεθόδων και η κατάλληλη εφαρμογή της μοντελοποίησης καμπυλών με χρήση B-splines δόθηκε αρχικά από τους Ξηρουχάκη, Αβρίθη και Κόλλια στην [195].

Το Κεφάλαιο 7 αντιμετωπίζει το πρόβλημα του υπολογισμού της τρισδιάστατης δομής και κίνησης αντικειμένων από τις δισδιάστατες προβολές τους. Ουσιαστικά, δεδομένων των ορθογραφικών προβολών ενός στερεού αντικειμένου και των διανυσμάτων προβεβλημένης κίνησης που υπολογίζονται για τα σημεία των προβολών, προτείνεται ένας πρωτότυπος αλγόριθμος ανακατασκευής της τρισδιάστατης δομής του και υπολογισμού της κίνησής του στον τρισδιάστατο χώρο για κάθε μετάβαση. Το κεφάλαιο βασίζεται στην πρωτότυπη εργασία [188] των Ξηρουχάκη και Ντελόπουλου, η οποία δανείζεται τις βασικές ιδέες από παλαιότερη εργασία [25] των ιδίων. Εντωμεταξύ, οι αλγοριθμικές λεπτομέρειες της προσέγγισης δίνονται στην αναφορά [190] των Ξηρουχάκη, Τσεχπενάκη και Ντελόπουλου, ενώ μία πρώτη μελέτη της επίδοσης του αλγορίθμου σε πεδία κίνησης που λήφθηκαν από γνωστές μεθόδους υπολογισμού δισδιάστατης κίνησης σε πραγματικές εικόνες δίνεται στην αναφορά [165] των ιδίων. Μία ενδιαφέρουσα προσέγγιση για τη βελτίωση των αποτελεσμάτων ανακατασκευής με το χρόνο δίνεται στην εργασία των Μπριασούλη, Ξηρουχάκη και Ντελόπουλου [19], ενώ σύγκριση της επίδοσης του αλγορίθμου με άλλες μεθόδους στην περιοχή πραγματοποιείται στην εργασία [186] των Ξηρουχάκη και Ντελόπουλου. Στο κεφάλαιο αυτό, για λόγους ισοδυναμίας με τα υπόλοιπα κεφάλαια αλλά και για λόγους έμφασης στις θεωρητικές ιδιότητες της βασικής μεθόδου, βασιστήκαμε μόνο στις αναφορές [188, 190, 186] από τις παραπάνω.

Στο Κεφάλαιο 8 παρουσιάζουμε μία αποδοτική λύση στο πρόβλημα του υπολογισμού της τρισδιάστατης κίνησης της κάμερας, με απευθείας εφαρμογή σε εικονικά περιβάλλοντα / εικονικά στούντιο. Πρακτικά, προτείνεται μία πλήρης μεθοδολογία κατασκευής του κυανού

πέτασματος ενός στούντιο με την εισαγωγή ενός σετ χαρακτηριστικών στο πέτασμα για λόγους υπολογισμού κίνησης. Παράλληλα, προτείνεται μία πρωτότυπη τεχνική υπολογισμού της τρισδιάστατης κίνησης από αντιστοιχίσεις γραμμών. Το κεφάλαιο βασίζεται στην εργασία [184] των Ξηρουχάκη, Δροσόπουλου και Ντελόπουλου. Οι βασικές ιδέες για την επίλυση του προβλήματος και οι πρώτες επιτυχείς μέθοδοι επίλυσής του με υποδεέστερη όμως επίδοση, παρουσιάστηκαν στις εργασίες [31, 182] των ιδίων. Ο αλγόριθμος υπολογισμού της τρισδιάστατης κίνησης από αντιστοιχίσεις γραμμών δίνεται στην εργασία [183], ενώ η αντιμετώπιση περιπτώσεων άμβλυνσης (blurring) λόγω κίνησης ή ανεστίασης και των περιπτώσεων εκφυλισμού της προοπτικής προβολής αντιμετωπίζονται στην [32] των ιδίων. Από τις παραπάνω αναφορές, το παρόν κεφάλαιο βασίζεται για λόγους έκτασης και έμφασης μόνο στις [184, 183].

Στο πλαίσιο της έρευνας για την εκπόνηση της διδακτορικής διατριβής πραγματοποιήθηκαν και άλλες συναφείς εργασίες, των οποίων η ανάλυση στο κυρίως σώμα του παρόντος δεν κρίθηκε σκόπιμη. Οι εργασίες αυτές αποτελούν κυρίως εφαρμογές των θεωρητικών αποτελεσμάτων της έρευνας ή/και πρότυπα συστήματα που σχεδιάστηκαν στα πλαίσια ερευνητικών προγραμμάτων.

Οι εργασίες των Ξηρουχάκη, Βότση και Ντελόπουλου [192] και Βότση, Ξηρουχάκη, Κόλλια και Σταφυλοπάτη [170] αφορούν σε τρισδιάστατη ανακατασκευή προσώπων από τις προβολές τους. Η έρευνα στην περιοχή αυτή έδειξε ότι τέτοια εγχειρήματα μπορεί να είναι επιτυχή μόνο κάτω από αυστηρούς περιορισμούς ακρίβειας.

Οι εργασίες [193, 169] των Ξηρουχάκη, Βότση, Καρπούζη και Κόλλια προτείνουν ένα σύστημα αρχειοθέτησης ψηφιακών εικόνων όπου η εφαρμογή τεχνικών αναζήτησης με βάση το περιεχόμενο και τεχνικών ευφυών πρακτόρων (intelligent agents) υποβοηθά στην ταχεία αναζήτηση σε βάσεις με μεγάλο όγκο δεδομένων. Η εργασία [73] των Καρπούζη, Βότση, Ξηρουχάκη, Στάμου και Κόλλια εισάγει τεχνικές ασαφούς λογικής στην εκπαίδευση του συστήματος ως προς τις προτιμήσεις των χρηστών.

Η εργασία [26] των Ντελόπουλου, Ραγκούση, Ξηρουχάκη και Δροσόπουλου αφορά στην ανάπτυξη ενός συστήματος αναγνώρισης κομματιών ήχου σε πραγματικό χρόνο. Η εργασία αυτή, εάν και δεν φαίνεται σχετική με το αντικείμενο της διατριβής, παρουσιάζει ενδιαφέρον στο πλαίσιο αυτής ως προς την εξαγωγή χαρακτηριστικών από πληροφορία ήχου.

Η εργασία [168] των Βότση, Δροσόπουλου, Ακρίβα, Τζουβάρα και Ξηρουχάκη περιγράφει ένα σύστημα ψηφιακής αρχειοθέτησης και δεικτοδότησης υλικού στα πλαίσια του προτύπου MPEG-7.

Όπως γίνεται φανερό από τα παραπάνω, μέρος των αποτελεσμάτων από τις εργασίες που παρουσιάζονται εδώ ανήκει στους κκ. Δροσόπουλο, Τσεχπενάκη και Αβρίθη, άρα και θα εμφανίζει κάποια μικρή επικάλυψη με τις διδακτορικές τους διατριβές. Στο παρόν πραγματοποιήθηκε γενικά προσπάθεια έμφασης μόνο στα σημεία που αφορούν το αντικείμενο της εν λόγω διατριβής. Βεβαίως, αυτό δεν ήταν δυνατό να πραγματοποιηθεί σε σημεία όπου η απάλειψη τμημάτων θα λειτουργούσε ανασταλτικά ως προς την κατανόηση και την ενότητα των κεφαλαίων.

Πρέπει να τονιστεί ότι η διδακτορική διατριβή επελέγη και χρηματοδοτήθηκε από τη Γενική Γραμματεία Έρευνας και Τεχνολογίας σε συνεργασία με την INTRAKOM στα πλαίσια του έργου ΥΠΕΡ-97 'Αποτελεσματική Αναζήτηση και Ανάκληση Εικόνων με βάση το Περιεχόμενο σε Σύγχρονα Συστήματα Πολυμέσων', 1998-2001.

Τα κεφάλαια της διατριβής έχουν γραφεί με τέτοιο τρόπο ώστε να είναι αυτοτελείς ενότητες και να μπορούν έτσι να διαβαστούν ανεξάρτητα. Για το λόγο αυτό, μικρές επικαλύψεις μπορούν να βρεθούν στις εισαγωγικές τους ενότητες.

Η συγγραφή του παρόντος πραγματοποιήθηκε αποκλειστικά με χρήση της έκδοσης 2.09 του L<sup>A</sup>T<sub>E</sub>X [79].

## Σχήματα

αρ. σχήματος

σελίδα

- 3.1 Το προτεινόμενο σύστημα εξαγωγής χρώματος σε μπλοκ-διάγραμμα. . . . . 20
- 3.2 Παράδειγμα εξαγωγής και μείωσης του γράφου γειτνίασης: (α) τεχνητά παρηγμένη εικόνα με (γ) τη χρωματική κατάταξή της και (ε) το γράφο γειτνίασης αντίστοιχα, (β) η εικόνα και (δ) η αντίστοιχη χρωματική κατάταξή της που προήλθαν από τον μειωμένο γράφο γειτνίασης (στ). . . . . 27
- 3.3 Δύο παραδείγματα (α,β) που επιδεικνύουν τη σημασία του χαρακτηριστικού όριο σύνδεσης: η περιοχή  $r_2$  θα ανατεθεί στην  $r_1$  αφού μοιράζονται το μεγαλύτερο όριο σύνδεσης. . . . . 30
- 3.4 Τρία παραδείγματα που επιδεικνύουν τη σημασία της μετρικής κάλυψη: η εικόνα εισόδου (α) παρουσιάζει 100% κάλυψη στην (β), 100% στην (γ) ως υποσύνολό της, 85% στην (δ) ως υπερσύνολό της. . . . . 33
- 3.5 Παράδειγμα που επιδεικνύει περιορισμούς στη μετρική κάλυψη: (α,β) εικόνες που ενώ διαφέρουν, παρουσιάζουν πλήρη σύμπτωση στα πλαίσια της κάλυψης. Παράδειγμα που εξηγεί τον ορισμό της μετρικής συνδεσιμότητα: σχεδόν πανομοιότυπες εικόνες (γ,δ) με διαφορετικούς γράφους γειτνίασης (ε,στ). . . . . 35
- 3.6 Πειραματικά αποτελέσματα με μοντέλο μία σκηνή με πολλά αντικείμενα, (α-ε) διαφορετικές λήψεις της ίδιας σκηνής και (στ) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της εικόνας (δ). . . . . 36
- 3.7 Πειραματικά αποτελέσματα με μοντέλο ένα πολύχρωμο πακέτο: (α-ε) όψεις του πακέτου, (στ) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της εικόνας (α). . . . . 37
- 3.8 20 ενδεικτικές φωτογραφίες από τη βάση δεδομένων με 1393 εικόνες από το διάστημα, όπου πραγματοποιήθηκαν πειραματικές αναζητήσεις εικόνων με βάση το περιεχόμενο. . . . . 38
- 3.9 Αποτελέσματα αναζήτησης βάσει της προτεινόμενης μεθόδου, (α) εικόνα εισόδου, (β-θ) καλύτεροι αντιπρόσωποι σε φθίνουσα σειρά ομοιότητας με την εικόνα εισόδου, (ι) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της (α). . 39

- 3.10 Αποτελέσματα αναζήτησης βάσει της προτεινόμενης μεθόδου, (α) εικόνα εισόδου, (β-θ) καλύτεροι αντιπρόσωποι σε φθίνουσα σειρά ομοιότητας με την εικόνα εισόδου, (ι) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της (α). . . . . 41
- 4.1 Εντοπισμός κινούμενων αντικειμένων για ένα πολύπλοκο παραμορφώσιμο αντικείμενο (ανθρώπινο σώμα) σε μία ακολουθία εσωτερικού χώρου που ελήφθη με στατική κάμερα, υπό μεταβαλλόμενες συνθήκες φωτισμού και μερική απόκρυψη, (α,δ,ζ) τρία από τα διαθέσιμα καρέ, (β,ε,η) αποτελέσματα ενημέρωσης φόντου, (γ,στ,θ) κινούμενες ακμές και οριακό πολύγωνο που λαμβάνεται. . . . . 57
- 4.2 Εντοπισμός κινούμενων αντικειμένων σε μία ακολουθία εξωτερικού χώρου που ελήφθη με μία ελαφρώς περιστρεφόμενη κάμερα χειρός παρουσία σημαντικού θορύβου (βλάστηση), (α,δ,ζ) τρία μη διαδοχικά καρέ, (β,ε,η) τα αντίστοιχα καρέ κανονικοποιημένα ως προς την κίνηση ως προς καρέ αναφοράς, (γ,στ,θ) αποτελέσματα εντοπισμού ακμών. . . . . 59
- 4.3 Αποτελέσματα που επιδεικνύουν την ενημέρωση του φόντου και την εξαγωγή των κινούμενων ακμών για τα καρέ που εμφανίζονται στο Σχήμα 4.2 (η ενημέρωση του φόντου βάσει μόνο των κινούμενων ακμών σε μία ανάλυση φαίνεται ανεπαρκής), (α,δ,ζ) το ενημερωμένο φόντο, (β,ε,η) οι εξαγόμενες κινούμενες ακμές, (γ,στ,θ) οι κινούμενες ακμές μετά από συνέλιξη με ένα φίλτρο εξασθένισης. . . . . 60
- 4.4 Βελτίωση των αποτελεσμάτων συνδυάζοντας τις κινούμενες ακμές σε δύο αναλύσεις για την ακολουθία των Σχημάτων 4.2, 4.3, (α,δ,ζ) οι κινούμενες ακμές σε χαμηλότερη ανάλυση, (β,ε,η) οι κινούμενες ακμές που λαμβάνονται συνδυάζοντας τις δύο αναλύσεις, (γ,στ,θ) οι κινούμενες ακμές μετά από συνέλιξη με φίλτρο εξασθένισης. . . . . 62
- 4.5 Βελτίωση των αποτελεσμάτων χρησιμοποιώντας και τα δύο κριτήρια ενημέρωσης του φόντου (επιμονή ακμών και αλλαγή κατεύθυνσης της φωτεινότητας), (α,δ,ζ) τρία μη διαδοχικά καρέ, (β,ε,η) κινούμενες ακμές χρησιμοποιώντας το κριτήριο της επιμονής ακμών για την ενημέρωση του φόντου, (γ,στ,θ) κινούμενες ακμές χρησιμοποιώντας και τα δύο κριτήρια. . . . . 63

4.6	Ταξινόμηση κινούμενων ακμών και εξαγωγή των οριακών πολυγώνων για μία ακολουθία που περιέχει περισσότερα από ένα κινούμενα αντικείμενα, εκμεταλλεύόμενοι όλα τα βήματα της προτεινόμενης μεθοδολογίας, (α,δ,ζ) τρία μη διαδοχικά καρέ, (β,ε,η) κανονικοποιημένα ως προς την κίνηση καρέ ως προς διαφορετική αναφορά, (γ,στ,θ) οριακά πολύγωνα των κινούμενων αντικειμένων που εντοπίστηκαν. . . . .	64
4.7	Ενδεικτικά αποτελέσματα που λαμβάνονται χρησιμοποιώντας την προτεινόμενη μέθοδο, (αριστερά) καρέ μετά από κανονικοποίηση κίνησης, (μέση) εξαγόμενες ακμές, (δεξιά) κινούμενες ακμές και οριακά πολύγωνα των κινούμενων αντικειμένων που εντοπίστηκαν. . . . .	65
5.1	Ενδεικτικά αποτελέσματα κατάτμησης χρώματος για μία φωτογραφία που περιέχει ένα αυτοκίνητο, (α) αρχική εικόνα, (β) κατετμημένη εικόνα (πρώτο επίπεδο), (γ) κατετμημένη εικόνα (τελικό αποτέλεσμα) και (δ) εξαγόμενο περίγραμμα αντικειμένου. . . . .	88
5.2	Ενδιάμεσα και τελικά αποτελέσματα της προτεινόμενης διαδικασίας ορθογωνοποίησης στα δείγματα του περιγράμματος ενός ψαριού (χωρίς χρήση τεχνικής ανα-δειγματοληψίας), (α) αρχική καμπύλη και ζεύγος τυχαίων μετασχηματισμών affine, (β) καμπύλες μετά την κανονικοποίηση μετατόπισης, (γ) καμπύλες μετά την κανονικοποίηση μεγέθυνσης και (δ) τελικές καμπύλες. . . . .	92
5.3	Αποτελέσματα από την προτεινόμενη τεχνική ορθογωνοποίησης στα δείγματα του περιγράμματος ενός ψαριού μετά από μη-ομοιόμορφη δειγματοληψία στις καμπύλες εισόδου, (α) αρχική καμπύλη και ζεύγος τυχαίων μετασχηματισμών affine, (β,γ,δ) κακή εφαρμογή των καμπυλών μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής. . . . .	93
5.4	Αποτελέσματα από την προτεινόμενη τεχνική ορθογωνοποίησης στα δείγματα του περιγράμματος ενός ψαριού μετά από διαφορετική μη-ομοιόμορφη δειγματοληψία στην ίδια ακριβώς καμπύλη, (α) τρία επικαλυπτόμενα στιγμιότυπα της αρχικής καμπύλης, (β,γ,δ) κακό ταίριασμα των καμπυλών μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής. . . . .	94

5.5	Βελτίωση των αποτελεσμάτων ορθογωνοποίησης μετά από χρήση B-spline μοντελοποίησης και αναδιάταξη των σημείων σύνδεσης στις μη ομοιόμορφα δειγματοληπτημένες καμπύλες του Σχήματος 5.3, (α) αρχική καμπύλη και ζεύγος τυχαίων μετασχηματισμών affine, (β,γ,δ) καμπύλες μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής αντίστοιχα. . . . .	95
5.6	Άμεση ευθυγράμμιση καμπυλών μέσω της διαδικασίας κανονικοποίησης για νοηματικά όμοια (χωρικά όμοια) περιγράμματα αντικειμένων, (α) αρχικές καμπύλες για τρία διαφορετικά αεροπλάνα, (β,γ,δ) καμπύλες μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής αντίστοιχα. . . . .	96
5.7	Ευθυγράμμιση καμπυλών μέσω της διαδικασίας κανονικοποίησης για ουσιαστικά διαφορετικά περιγράμματα αντικειμένων, (α) αρχικές καμπύλες για ένα σφυρί, ένα αυτοκίνητο και ένα αεροπλάνο, (β,γ,δ) καμπύλες μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής αντίστοιχα. . . . .	97
5.8	Αποτελέσματα κανονικοποίησης παρουσία τυχαίου θορύβου στα δείγματα των αρχικών καμπυλών, (α,β) αρχικές και τελικές καμπύλες για τρία στιγμιότυπα της ίδιας καμπύλης, (γ,δ) αρχικές και τελικές καμπύλες για τρία όμοια νοηματικά αντικείμενα (αεροπλάνα). . . . .	98
5.9	Ενδεικτικές αποστάσεις περιγραμμάτων για σκοπούς ταξινόμησης. . . . .	99
5.10	Αποτελέσματα ανάκτησης εικόνων (αναζήτηση-βάσει-παραδείγματος) βάσει ομοιότητας περιγράμματος για μία φωτογραφία που περιέχει ένα αεροπλάνο, (α) φωτογραφία εισόδου, (β) εξαγόμενο περίγραμμα (γ) φωτογραφίες κατά σειρά ανάκτησης (προς τα δεξιά και προς τα κάτω). . . . .	100
5.11	Αποτελέσματα ανάκτησης εικόνων (αναζήτηση-βάσει-παραδείγματος) βάσει ομοιότητας περιγράμματος για μία φωτογραφία που περιέχει ένα αυτοκίνητο, (α) φωτογραφία εισόδου, (β) εξαγόμενο περίγραμμα (γ) φωτογραφίες κατά σειρά ανάκτησης (προς τα δεξιά και προς τα κάτω). . . . .	101
7.1	Κίνηση επίπεδης επιφάνειας στον 3Δ χώρο. . . . .	115
7.2	Αντικείμενο που αποτελείται από δύο επίπεδες επιφάνειες (α) σκηνή 1, (β) σκηνή 2, (γ) σκηνή 3, (δ) πεδίο κίνησης (καρέ 1-2), (ε) πεδίο κίνησης (καρέ 1-3), (στ) διαφορικό πεδίο κίνησης (καρέ 1-2). . . . .	138
7.3	Λεία τρισδιάστατη επιφάνεια (α) σκηνή 1, (β) πεδίο κίνησης (καρέ 1-2), (γ) πεδίο κίνησης (καρέ 1-3), (δ) διαφορικό πεδίο κίνησης (καρέ 1-3). . . . .	139

7.4	Υπολογισμός και ανακατασκευή παρουσία θορύβου για το αντικείμενο που αποτελείται από δύο επίπεδες επιφάνειες (α) 10db θορυβώδες πεδίο κίνησης (καρέ 1-2), (β) 10db θορυβώδες πεδίο κίνησης (καρέ 1-3), (γ) αντίστοιχο θορυβώδες διαφορικό πεδίο κίνησης (καρέ 1-2), (δ) ανακατασκευασμένη σκηνή 1. . . . .	140
7.5	Υπολογισμός και ανακατασκευή παρουσία θορύβου για τη λεία επιφάνεια, (α) 10db θορυβώδες πεδίο κίνησης (καρέ 1-2), (β) 10db θορυβώδες πεδίο κίνησης (καρέ 1-3), (γ) αντίστοιχο θορυβώδες διαφορικό πεδίο κίνησης (καρέ 1-3), (δ) ανακατασκευασμένη σκηνή 1. . . . .	142
7.6	Υπολογισμός παραμέτρων περιστροφής για το αντικείμενο που αποτελείται από δύο επίπεδες επιφάνειες, (α,β,γ) μετρήσεις των $\phi_R$ , $\phi_S$ , $w$ ως προς τον αριθμό των δειγμάτων, (δ,ε,στ) μετρήσεις των $\phi_R$ , $\phi_S$ , $w$ ως προς το βήμα $x$ . . . . .	144
7.7	Υπολογισμός παραμέτρων κίνησης ενός οχήματος διαθέσιμου σε 3 καρέ χρησιμοποιώντας ανάθεση των χαρακτηριστικών σημείων με το χέρι (α) χαρακτηριστικά σημεία (καρέ 1), (β) χαρακτηριστικά σημεία (καρέ 2), (γ) χαρακτηριστικά σημεία (καρέ 3), (δ) αντιστοιχίσεις σημείων (καρέ 1-2). . . . .	145
7.8	Μετρήσεις για διαφορετικό επίπεδο SNR στα διαφορικά πεδία κίνησης, για το αντικείμενο με τις δύο επίπεδες επιφάνειες, (α,γ,ε) και (β,δ,στ) $\phi_R$ και $\phi_S$ χρησιμοποιώντας την factorization, την epipolar και την προτεινόμενη μέθοδο αντίστοιχα. . . . .	147
7.9	Μετρήσεις για διαφορετικό επίπεδο SNR στα διαφορικά πεδία κίνησης, για τη λεία επιφάνεια, (α,γ,ε) και (β,δ,στ) $\phi_R$ και $\phi_S$ χρησιμοποιώντας την factorization, την epipolar και την προτεινόμενη μέθοδο αντίστοιχα. . . . .	148
8.1	Το προτεινόμενο σύστημα υπολογισμού κίνησης σε μπλοκ-διάγραμμα. . . . .	161
8.2	Η γεωμετρία του προβλήματος. Το σύστημα συντεταγμένων κινείται με το επίπεδο του CCD (camera-based). . . . .	167
8.3	Παράδειγμα κυανού πετάσματος δύο χρωματικών τόνων (α) και καταγραφόμενο από την κάμερα τμήμα του πετάσματος (β). . . . .	179
8.4	Όλα τα βήματα της προτεινόμενης μεθοδολογίας για τον υπολογισμό της κίνησης της κάμερας στο Σχήμα 8.3(β), (α) το καταγραφόμενο καρέ, (β) οι ακμές που εντοπίστηκαν στο καρέ, (γ) οι εξαγόμενες γραμμές αναφοράς, (δ) το καρέ όπως αποδίδεται από τις υπολογιζόμενες παραμέτρους κίνησης. . . . .	180



8.5	(α) Καρέ που καταγράφεται με απώστες κάποιες από τις γραμμές αναφοράς και (β) οι γραμμές αναφοράς που εξάγονται. . . . .	182
8.6	Εκτιμήσεις (α) της γωνίας περιστροφής, (β) του μήκους εστίασης, (γ) της πρώ- της συνιστώσας του άξονα περιστροφής και (δ) της πρώτης συνιστώσας του διανύσματος μετατόπισης (απλές γραμμές) συναρτήσει του χρόνου, για μία τυ- χαία ακολουθία. . . . .	184
8.7	Ένα φυσικό παράδειγμα: (α) το σύστημα-μινιατούρα ενός εικονικού στούντιο, (β) ένα καρέ που καταγράφεται από μία πραγματική κάμερα, (γ) οι ακμές που εξάγονται (δ) και το καρέ που παράγεται χρησιμοποιώντας τις παραμέτρους της κάμερας που υπολογίστηκαν. . . . .	185

## Πίνακες

<u>αρ. πίνακα</u>	<u>σελίδα</u>
3.1 Η στρατηγική της πλημμύρας . . . . .	22
3.2 Εξαγωγή χαρακτηριστικών και κατασκευή του γράφου γειτνίασης . . . . .	24
3.3 Συμβολισμός των χαρακτηριστικών που συμπεριλαμβάνονται στο γράφο γειτνίασης . . . . .	25
3.4 Μείωση του γράφου γειτνίασης . . . . .	29
3.5 Η διαδικασία σύγκρισης διανυσμάτων χαρακτηριστικών . . . . .	34
3.6 Η βελτιωμένη εκδοχή της διαδικασίας σύγκρισης . . . . .	37
4.1 Ανοιχτές παράμετροι του συστήματος εντοπισμού κινούμενων αντικειμένων . . . . .	55
5.1 Πίνακας μετασχηματισμών καμπυλών και συναφών ποσοτήτων που περιλαμβάνει η προτεινόμενη διαδικασία κανονικοποίησης καμπύλης. . . . .	79
7.1 Διαδικασία υπολογισμού απουσία θορύβου. . . . .	122
7.2 Διαδικασία υπολογισμού παρουσία θορύβου. . . . .	135
7.3 Υπολογισμός με θορυβώδη πεδία κίνησης (βλ. και Σχήματα 7.4,7.5). . . . .	141
7.4 Μείωση της ποσότητας $\sigma_k^2$ με το βήμα $x$ . . . . .	142
7.5 Υπολογισμός με θορυβώδεις αντιστοιχίσεις σημείων (βλ. Σχήμα 7.7). . . . .	146
8.1 Κατασκευάζοντας το δυαδικό χάρτη του κυανού πετάσματος, άνω γραμμή: η ακολουθία των διαφορών των μετατοπίσεων, κάτω γραμμή: η ακολουθία των μετατοπίσεων. . . . .	165
8.2 Εντοπισμός των γραμμών αναφοράς που απουσιάζουν. Παράμετροι γραμμών που υπολογίζονται (ξεκινώντας από αριστερά) από τον μετασχηματισμό Hough, μετά από κανονικοποίηση μεγέθυνσης και με τις πραγματικές διαστάσεις του CCD, κανονικοποιημένες αποστάσεις γραμμών σε ζεύγη όπως υπολογίζονται από την Πρόταση 8.3. . . . .	183

## ΚΕΦΑΛΑΙΟ 1

### Εισαγωγή

Είναι σήμερα ίσως περισσότερο από ποτέ έντονο το συναίσθημα ότι η εποχή μας ακολουθεί κατά πόδας και με ταχύ ρυθμό τις επιταγές των σεναρίων των ταινιών επιστημονικής φαντασίας. Ποιος θα φανταζόταν μερικές μόλις δεκαετίες πριν ότι ο φούρνος θα μαγειρεύε σχεδόν αυτόβουλα το φαγητό, βάσει κάποιου έξυπνου προγράμματος ψησίματος και θα ειδοποιούσε με ένα ηχητικό σήμα όταν αυτό είναι έτοιμο; Ποιος θα πίστευε ότι ένα μικρό κουτάκι, που θα χωρούσε άνετα στην τσέπη του, θα του έδινε τη δυνατότητα να δέχεται τις κλήσεις του από την άλλη άκρη του κόσμου, όπου κι αν αυτός βρισκόταν; Ποιος θα πειθόταν ότι ένα καλώδιο θα ήταν αρκετό για να συνδέσει το σπίτι του με τεράστιες κατανεμημένες βιβλιοθήκες πληροφοριών σε οποιοδήποτε θέμα και γλώσσα, με χώρους ηλεκτρονικής ψυχαγωγίας όπου θα μπορούσε να γνωριστεί ή/και να ψυχαγωγηθεί ταυτόχρονα με χιλιάδες άλλους ανθρώπους από όλα τα μέρη του κόσμου, με εκατοντάδες κανάλια τα οποία θα του παρείχαν σχεδόν οποιοδήποτε ψυχαγωγικό / πληροφοριακό πρόγραμμα την ώρα που αυτός το επιθυμούσε; Και την ίδια στιγμή, ποιος θα μπορούσε να φανταστεί ότι όλα αυτά (και πολλά άλλα σε μία λίστα που μεγαλώνει κάθε μέρα) θα μπορούσε να τα απολαύσει σε ένα τόσο προσιτό κόστος;

Πέρα από τις αντιρρήσεις που μπορεί κανείς να προβάλλει (και που όντως προβάλλονται) σχετικά με το κοινωνικό όφελος και την ανθρωπιστική διάσταση της καλπάζουσας τεχνολογίας, είναι αναμενόμενο ότι η τεχνολογική ανάπτυξη θα συνεχίσει να μας εκπλήσσει κάθε μέρα με τις καινούριες κατακτήσεις της, που θα ικανοποιούν και τον πιο απαιτητικό λάτρη ταινιών επιστημονικής φαντασίας. Θα έχουμε έτσι στο κοντινό (;) μέλλον τη δυνατότητα να διαλέγουμε το φύλο και το χρώμα των ματιών των παιδιών μας, να τηλε-μεταφερόμαστε για ένα επαγγελματικό ραντεβού στον Άρη και να είμαστε πίσω πριν αυτά γυρίσουν από το σχολείο και αντί για σκύλο να τους αγοράζουμε ένα μικρό φίλο-ρομπότ, ο οποίος θα είναι πιο ψυχαγωγικός, πιο ευχάριστος, πιο πιστός και πιο αποτελεσματικός στην προστασία τους χάριν σε ένα εξελιγμένο πακέτο λογισμικού που θα του δίνει ‘αντίληψη’.

Συμβαίνει πολλές φορές να μας συναρπάζουν μηχανές που εμφανίζουν ιδιότητες όμοιες με αυτές του ανθρώπου όπως να βγάζουν ανθρώπινους ήχους, να περπατάνε στα δύο πόδια ή να κουρεύουν το γκαζόν στον κήπο. Οι πραγματικά συναρπαστικές μηχανές όμως δεν έχουν απαραίτητα ανθρωπόμορφο σχήμα και δεν παρουσιάζουν απαραίτητα απλές ανθρώπινες ιδιότητες και ασχολίες όπως οι παραπάνω, αλλά τη χαρακτηριστική ανθρώπινη λειτουργία της αντίληψης. Έτσι, για παράδειγμα έχουμε σήμερα συστήματα που αναγνωρίζουν την ανθρώπινη φωνή, που μπορούν να ξεχωρίσουν και να αναπαράγουν τις λέξεις. Κινητά συστήματα που

μπορούν να αντιληφθούν την παρουσία εμποδίων στο δρόμο τους και να διορθώσουν την κατεύθυνσή τους ώστε να τα αποφύγουν. Συστήματα επιτήρησης που μπορούν να εντοπίσουν και να αναγνωρίσουν 'στόχους' αλλά και να παρακολουθήσουν την κίνησή τους. Όπως δε είναι αναμενόμενο, όσο 'καλύτερη αντίληψη' του περιβάλλοντός της μπορεί να έχει μία υπολογιστική μηχανή, τόσο περισσότερο πολύπλοκες και θαυμαστές είναι οι λειτουργίες που μπορούμε να την προγραμματίσουμε να επιτελέσει.

Η αντίληψη (perception) προέρχεται κυρίως μέσω των ερεθισμάτων των αισθήσεων. Μία από τις αισθήσεις είναι και η όραση (vision), η οποία τεχνικώς επιτελείται σε ένα υπολογιστικό σύστημα μέσω ενός συστήματος καταγραφής, δηλαδή από μία ή περισσότερες κάμερες. Σε αυτά τα πλαίσια, ένας προσωπικός υπολογιστής συνδεδεμένος με μία απλή κάμερα μπορεί τεχνικώς 'να βλέπει'. Όταν αναφερόμαστε στην ικανότητα της όρασης στους ανθρώπινους οργανισμούς θεωρούμε δεδομένη τη δυνατότητα της αντίληψης αυτών που 'βλέπουν', αυτή όμως είναι και η δυσκολότερη εργασία που πρέπει να επιτελέσει ένα υπολογιστικό σύστημα. Δηλαδή, να καταφέρει να αντιληφθεί αυτά που 'βλέπει', να εξάγει κάποια χαρακτηριστικά τους, βάσει των οποίων θα μπορέσει να τα διακρίνει, να τα συγκρίνει, να τα αναγνωρίσει και να τα χαρακτηρίσει. Στην ιδανική περίπτωση να πάρει και μία απόφαση για το αν θα πρέπει να εκτελέσει κάποια ενέργεια και ποια είναι αυτή. Εάν λοιπόν, για παράδειγμα, επιθυμούμε να κατασκευάσουμε ένα υπολογιστικό σύστημα με τη δυνατότητα να διασχίζει το δρόμο, από πεζοδρόμιο σε πεζοδρόμιο, χωρίς να κινδυνεύει από τα επερχόμενα αυτοκίνητα, θα πρέπει τουλάχιστον να το 'διδάξουμε' να ξεχωρίζει τις σιλουέτες των επερχόμενων αυτοκινήτων και να αποφασίζει εάν κινδυνεύει από αυτές καθώς προχωράει, κρίνοντας από την κατεύθυνσή τους, την ταχύτητά τους και τη σχετική τους απόσταση από αυτό. Με δεδομένες μόνο τις εικόνες που λαμβάνει από μία κάμερα, κάθε ένα από τα παραπάνω αποτελεί ένα χωριστό και δυσεπίλυτο στη γενική περίπτωση πρόβλημα, τουλάχιστον συγκρινόμενο ως προς τη δυνατότητα του συστήματος να στέκεται στα δύο του πόδια.

Με προβλήματα όπως το παραπάνω ασχολείται η Όραση Υπολογιστών (Computer Vision), όπως και η συναφής περιοχή της Κατανόησης Περιεχομένων Εικόνων και Βίντεο (Image and Video Understanding). Κοντινές επιστημονικές περιοχές, με δυσδιάκριτα σύνορα μεταξύ αυτών και των παραπάνω, είναι αυτές της Ψηφιακής Επεξεργασίας Εικόνων και Βίντεο (Digital Image and Video Processing), της Τεχνητής Νοημοσύνης (Artificial Intelligence), των Συστημάτων Βάσεων Δεδομένων (Database Systems), των Συστημάτων Πολυμέσων (Multimedia Systems), της Ρομποτικής και των Αυτοματισμών (Robotics and Automation) και άλλες. Ο βαθμός συσχέτισης των παραπάνω περιοχών με την Όραση Υπολογιστών εξαρτάται βεβαίως

και από την κατηγορία εφαρμογών που εξετάζουμε, αφού το πεδίο εφαρμογών της Όρασης Υπολογιστών είναι ιδιαίτερα ευρύ.

Στη γενική περίπτωση θα λέγαμε ότι έχουμε στη διάθεσή μας μία εικόνα ή μία ακολουθία βίντεο η οποία φθάνει στη μονάδα επεξεργασίας μας μέσω ενός μηχανισμού καταγραφής (κάμερα) ή αναπαραγωγής (αρχείο εικόνας ή βίντεο). Τα δεδομένα εισόδου (raw data) προ της επεξεργασίας τους δεν έχουν καμία πρακτική χρησιμότητα, αναφορικά με την αντίληψη του συστήματος, πέρα από την αποθήκευση και την επίδειξή τους, αφού δεν είναι παρά οι συντεταγμένες των εικονοστοιχείων (pixels) των εικόνων στο χώρο των χρωμάτων. Βάσει μόνο της πληροφορίας αυτής, το σύστημα καλείται να επιτελέσει υψηλού επιπέδου εργασίες που σε κάποιο βαθμό να φανερώνουν την αντίληψή του. Η μεθοδολογία που ακολουθείται στη λύση τέτοιων προβλημάτων (ώστε να επιτελεστούν οι εργασίες αυτές στη γενική περίπτωση) εξαρτάται από ένα σύνολο παραμέτρων που έχουν να κάνουν με τα επιθυμητά αποτελέσματα εξόδου ως προς τη φύση τους, την πολυπλοκότητά τους, την ακρίβεια και την ταχύτητα με την οποία θέλουμε να λαμβάνονται και με ποικίλους άλλους παράγοντες. Πολύ διαφορετική μεθοδολογία θα ακολουθήσουμε για παράδειγμα όταν θέλουμε να μετρήσουμε το ποσοστό κόκκινου χρώματος μίας εικόνας, όταν θέλουμε να ανιχνεύσουμε τα κινούμενα αντικείμενα σε μία ακολουθία βίντεο ή όταν θέλουμε να εκτιμήσουμε την σχετική απόσταση των αντικειμένων από το υπολογιστικό σύστημα / την κάμερα.

Πρακτικά, ένα σύστημα ‘αντιλαμβάνεται’ ή ‘κατανοεί’ τα περιεχόμενα μίας τέτοιας πληροφορίας, όταν μπορεί να μετασχηματίσει την χρωματική πληροφορία των pixels σε κάποια άλλη μορφή πληροφορίας που να έχει περισσότερο σημασιολογικό περιεχόμενο (semantics). Για παράδειγμα, η οργάνωση μίας εικόνας σε χρωματικές περιοχές, η εξαγωγή των ακμών της εικόνας ή των περιγραμμάτων χαρακτηριστικών περιοχών της, ο υπολογισμός της κίνησης των pixels / σημείων από καρέ σε καρέ σε μία ακολουθία εικόνων αποτελούν κάποια πρώτα στοιχεία που μπορεί κάποιος να εξάγει και που περιέχουν πληροφορία που μπορεί να είναι περισσότερο χρήσιμη στη συνέχεια. Τα στοιχεία αυτά ονομάζονται χαρακτηριστικά (features) και μπορούν να είναι πολύ διαφορετικά, ακόμα και όταν αφορούν την ίδια μορφή πληροφορίας, για παράδειγμα οι ακμές μίας εικόνας, τα περιγράμματα των αντικειμένων της, ο σκελετός των αντικειμένων της, η σχετική θέση των αντικειμένων της επάνω στην εικόνα, το μέγεθός τους και η εκκεντρότητά τους μπορούν όλα να θεωρηθούν πληροφορία ‘σχήματος’. Επομένως, τα χαρακτηριστικά που μπορεί κανείς να εξάγει από μία εικόνα ή ακολουθία εικόνων είναι ποικίλα και η ποιότητά τους ή η επιτυχία του αλγορίθμου εξαγωγής τους εξαρτάται απευθείας από την εφαρμογή στην οποία θα χρησιμοποιηθούν. Παράλληλα, κάποιο χαρακτηριστικό μπορεί να είναι ικανά ‘πληροφοριακό’ ή ‘σημασιολογικό’ για μία εφαρμογή και λιγότερο (ή

εντελώς ακατάλληλο) για μία άλλη - για παράδειγμα η χρωματική κατανομή των pixels σε μία εικόνα μπορεί να είναι ενδιαφέρουσα πληροφορία για σκοπούς αναζήτησης εικόνων με βάση το χρωματικό περιεχόμενο αλλά μάλλον αδιάφορη για την τρισδιάστατη ανακατασκευή των αντικειμένων που αυτές περιέχουν.

Θα μπορούσαμε γενικά να πούμε ότι οι μέθοδοι εξαγωγής χαρακτηριστικών εργάζονται με παρεμφερή τρόπο. Αρχικά, ένα ή περισσότερα χαμηλού επιπέδου χαρακτηριστικά εξάγονται από την εικόνα ή ακολουθία, όπως το χρώμα (πρακτικά, ένα 'χαρακτηριστικό χρώματος' από τα πολλά που μπορούν να εξαχθούν), η υφή, το σχήμα, η κίνηση κ.λπ. με χρήση ενός κατάλληλου αλγορίθμου. Εν συνεχεία, από το χαρακτηριστικό αυτό και συχνά με χρήση κάποιας σωρρευμένης γνώσης, γενικής ή ειδικότερης σχετιζόμενης με την εφαρμογή, εξάγεται πληροφορία υψηλότερου επιπέδου, πρακτικά υπολογίζονται κάποια νέα χαρακτηριστικά με μεγαλύτερη σημασιολογική πληροφορία. Τα χαρακτηριστικά αυτά με τη σειρά τους μπορούν να χρησιμοποιηθούν για διάφορες λειτουργίες. Μπορούν για παράδειγμα να χρησιμοποιηθούν από μόνα τους χαρακτηρίζοντας την εικόνα ή ακολουθία ή ακόμη και περιοχές ή αντικείμενα αυτής, όπως για παράδειγμα το βάθος των αντικειμένων από το σύστημα καταγραφής ή το τρισδιάστατο μοντέλο ενός αντικειμένου του οποίου οι όψεις εμφανίζονται στην ακολουθία. Σε άλλες περιπτώσεις, μπορούν να χρησιμοποιηθούν σε σύγκριση με κάποια συγκεκριμένα αντίστοιχα δεδομένα για εφαρμογές αναγνώρισης, όπως για παράδειγμα τα χαρακτηριστικά που εξάγονται από πρόσωπα για αυτόματη ταυτοποίησή τους σε μία βάση δεδομένων προσώπων. Τέλος, μπορούν να χρησιμοποιηθούν σε εφαρμογές ταξινόμησης ή αναζήτησης ομοίων, όπως για παράδειγμα σε εφαρμογές αναζήτησης με βάση το περιεχόμενο.

Όπως γίνεται εύκολα κατανοητό, η 'διαδρομή' από την όραση σε επίπεδο καταγραφής εικόνων έως την αντίληψη σε επίπεδο κατανόησης και απόφασης είναι πολύ μεγάλη. Τα ολοκληρωμένα συστήματα που εκτελούν επιτυχημένα τέτοιες λειτουργίες και που προσομοιώνουν σε ικανοποιητικό βαθμό την ανθρώπινη αντίληψη είναι σήμερα λίγα σε αριθμό, εκτελούν ως επί το πλείστον πολύ συγκεκριμένες λειτουργίες και με αρκετά σημαντικούς περιορισμούς στη λειτουργία τους, ενώ ταυτόχρονα απαιτούν πολύπλοκους και συχνά χρονοβόρους αλγορίθμους. Το τελευταίο είναι αποτέλεσμα ενός αριθμού από ενδιάμεσα προβλήματα τα οποία πρέπει συχνά να αντιμετωπιστούν με όσο το δυνατόν γενικότερο τρόπο. Για παράδειγμα, συνήθη πολύπλοκα ενδιάμεσα προβλήματα είναι αυτά της προβολής και της κλίμακας. Το πρώτο αφορά το γεγονός ότι οι εικόνες που καταγράφει μία κάμερα είναι αποτέλεσμα της προβολής του τρισδιάστατου κόσμου σε ένα δισδιάστατο επίπεδο. Αυτό προξενεί έναν αριθμό από σημαντικές ασάφειες και απώλειες πληροφορίας, οι οποίες πρέπει να καταργηθούν είτε με έμμεσο υπολογισμό των ποσοτήτων που απουσιάζουν (π.χ. υπολογισμός βάθους με τρισδιάστατη ανακατασκευή) είτε να

αποφευχθούν ορίζοντας δισδιάστατα χαρακτηριστικά ως αποτέλεσμα προβολής που να είναι αρκούντως πληροφοριακά (π.χ. μοντελοποίηση γωνιών και ακμών σε προβολές προσώπων), είτε να μοντελοποιηθούν με χρήση κατάλληλων μετασχηματισμών που να περιγράφουν ικανοποιητικά το αποτέλεσμα της προβολής (π.χ. εκτίμηση προβολής καμπυλών με μετασχηματισμούς affine). Είναι βεβαίως αναμενόμενο ότι σε κάθε περίπτωση, η επιτυχία των παραπάνω εξαρτάται από την ικανοποίηση σημαντικών περιορισμών. Για παράδειγμα, δεν μπορούμε να εκτιμήσουμε το σχήμα μίας όψης του αντικειμένου που δεν είναι ποτέ ορατή από την κάμερα. Ή ακόμη, δεν μπορούμε να αναγνωρίσουμε ένα αντικείμενο που διαρκώς κρύβεται σε σημαντικό βαθμό από ένα άλλο. Το πρόβλημα της αλλαγής κλίμακας είναι επίσης πολύ σημαντικό σε εφαρμογές αναγνώρισης, ιδιαίτερα σε συνδυασμό με τη διακριτότητα του pixel που επιτρέπει ένα μέγιστο επίπεδο λεπτομέρειας ή και με την απόκρυψη μέρους του αντικειμένου από τα όρια της εικόνας.

Παρόλη την πολυπλοκότητα των προβλημάτων στην περιοχή της Όρασης Υπολογιστών, αλλά και άλλα σημαντικά θέματα όπως η συνήθως σημαντική εξάρτηση των αποτελεσμάτων από τις ανακρίβειες ή το θόρυβο, η επιστημονική κοινότητα των υπολογιστών έχει δώσει μεγάλη έμφαση προς την κατεύθυνση αυτή αναμένοντας να δρέψει τους καρπούς της προσπάθειας στο κοντινό μέλλον. Ήδη σήμερα, μπορούμε να βρούμε πρότυπα συστήματα να λειτουργούν με σημαντικούς περιορισμούς και περιορισμένη ακρίβεια, πολλά υποσχόμενα όμως στο μέλλον. Μπορεί κανείς λοιπόν σήμερα ίσως και να εντυπωσιαστεί από τις δυνατότητες συστημάτων εντοπισμού και αναγνώρισης προσώπων (face detection and recognition), αναζήτησης και ανάκτησης εικόνων με βάση το περιεχόμενο (content-based image query and retrieval), εντοπισμού εισβολέα σε παρακολουθούμενους χώρους (intruder / mobile object detection) ή ακόμη και συστημάτων εκτίμησης της τρισδιάστατης κίνησης (three-dimensional motion estimation) ή αναγνώρισης συναισθημάτων από τις χειρονομίες και την έκφραση (emotion recognition).

Επηρεασμένοι συχνά από τα σενάρια επιστημονικής φαντασίας, νομίζουμε ότι η τεχνητή νοημοσύνη θα κατακτηθεί όταν πια θα συζητάμε τα προβλήματά μας με τον καλό μας φίλο-ρομπότ ή τουλάχιστον όταν αυτός θα πηγαίνει να κάνει τα ψώνια και αισθανόμαστε άσχημα όταν βλέπουμε ότι ακόμη και μία αναζήτηση με βάση το περιεχόμενο δεν μπορεί να επιτύχει ο υπολογιστής μας με ακρίβεια. Ταυτόχρονα, αισθανόμαστε δέος μπροστά σε όρους όπως αναγνώριση προσώπων και συναισθημάτων ή αναγνώριση και παρακολούθηση στόχου και νομίζουμε ότι οι υπολογιστές-ρομπότ θα αρχίσουν να συμπεριφέρονται αυτόβουλα και κάποια στιγμή ίσως αποτελέσουμε και τον . . . υπό παρακολούθηση στόχο. Η αλήθεια βρίσκεται σήμερα μακριά και από τις δύο σκέψεις, αφού η υπολογιστική όραση και η τεχνητή νοημοσύνη έκαναν τα πρώτα σημαντικά τους βήματα τις τελευταίες δεκαετίες. Σήμερα θα λέγαμε ότι τουλάχιστον

η υπολογιστική όραση προσπαθεί να κατανοήσει τις αντιληπτικές λειτουργίες της όρασης του ανθρώπου και να τις προσομοιώσει ικανοποιητικά, αναφορικά μόνο ως προς ορισμένες σχετικά απλές εφαρμογές που ένα υπολογιστικό σύστημα θα μπορούσε να επιτελέσει, π.χ. αναγνώριση προτύπων κομματιών από ένα υπολογιστικό σύστημα σε μία γραμμή μεταφοράς εργοστασίου.

Ακριβώς όμως επειδή εφαρμογές όπως αυτές που αναφέρθηκαν βρίσκονται σε ευθεία αντιπαραβολή με τη συμπεριφορά ενός ανθρώπου μπροστά στο ίδιο πρόβλημα, η επίλυσή τους μπορεί να είναι πολύ δύσκολη αλλά ταυτόχρονα ενδιαφέρουσα γιατί συχνά αποτελεί προσπάθεια προσομοίωσης της ανθρώπινης συμπεριφοράς. Με άλλα λόγια, λύνοντας τέτοια προβλήματα, συχνά προσπαθούμε να προσομοιώσουμε και άρα να κατανοήσουμε την ανθρώπινη αντίληψη, ιδανικά να κατασκευάσουμε συστήματα που θα αντιλαμβάνονται και θα κατανοούν όπως οι άνθρωποι, γενικά όμως με πολύ μεγαλύτερες δυνατότητες αποθήκευσης και ταχύτητα επεξεργασίας. Στον μακρύ δρόμο μέχρι την επίτευξη του παραπάνω θαυμαστού και λίγο φανταστικού στόχου, η υπολογιστική όραση αναμένεται να μας εφοδιάσει με ενδιαφέρουσες τεχνολογίες και υπηρεσίες στην εκπαίδευση, στην εργασία, στη διασκέδαση και στους περισσότερους υπόλοιπους τομείς της καθημερινής ζωής.



## Μέρος Ι

Εκτίμηση Χαρακτηριστικών Δισδιάστατης Δομής και Κίνησης

## ΚΕΦΑΛΑΙΟ 2

### Εκτίμηση Χαρακτηριστικών στις Δύο Διαστάσεις

Με δεδομένο το γεγονός ότι μία εικόνα αποτελεί προβολή του τρισδιάστατου κόσμου σε ένα δισδιάστατο επίπεδο, είναι λογικό η εξαγωγή δισδιάστατων χαρακτηριστικών (features) από αυτήν να είναι πιο ‘προφανής’ αλλά και πιο απλή. Στην πραγματικότητα, από μία δισδιάστατη εικόνα μπορούμε να εξάγουμε μόνο δισδιάστατα χαρακτηριστικά, όπως είναι λογικό, ενώ η εξαγωγή τρισδιάστατων χαρακτηριστικών μπορεί να πραγματοποιηθεί εμμέσως (π.χ. από πρακτική γνώση του πώς προβάλλονται οι τρισδιάστατες επιφάνειες στο επίπεδο) αλλά και από συνδυασμένη επεξεργασία περισσότερων από μίας εικόνες (π.χ. μεταξύ διαδοχικών καρρέ σε μία ακολουθία βίντεο).

Το πιο απλό πρόβλημα ‘όρασης’ είναι ίσως αυτό της ταυτοποίησης μίας εικόνας μέσα σε ένα σύνολο εικόνων. Με δεδομένο λοιπόν ένα σετ από εικόνες σε μία βάση δεδομένων έστω ότι επαναισθάνουμε μία από τις εικόνες του σετ στο σύστημα και ζητάμε από αυτό να μας πει εάν η εικόνα υπάρχει ήδη στη βάση δεδομένων και με ποια είναι όμοια. Το πρόβλημα αυτό μπορεί να λυθεί από το σύστημα με απλό και σχετικά γρήγορο τρόπο: οι χρωματικές τιμές των pixels της εικόνας μας συγκρίνονται μία προς μία με τις χρωματικές τιμές των pixels των εικόνων της βάσης δεδομένων. Σε αυτή την περίπτωση δεν υπάρχει κανένα περιθώριο λάθους, οπότε π.χ. τα αντικείμενα, τα πρόσωπα και το περιβάλλον τους ταυτοποιούνται επιτυχώς. Οριακά και τότε πραγματοποιείται εξαγωγή δισδιάστατων χαρακτηριστικών από τις εικόνες, με τα χαρακτηριστικά να είναι η θέση και το χρώμα των pixels της κάθε εικόνας. Το διάνυσμα χαρακτηριστικών (feature vector) είναι τότε το σετ που αποτελείται από τις θέσεις και τις χρωματικές τιμές όλων των pixels της εικόνας.

Συνεχίζοντας, παρατηρούμε ότι ένα τέτοιο σύστημα θα αποτύχει δραματικά όταν στην εικόνα που υποβάλλεται έχουμε π.χ., ως προς τις εικόνες που ομοιάζουν με αυτή, μετακίνηση κάποιου αντικειμένου, στροφή κάποιου προσώπου, αλλαγή συνθηκών φωτισμού, πολύ δε περισσότερο όταν εμφανιστούν νέα αντικείμενα στη σκηνή ή όταν αλλάξει το σημείο παρατήρησης. Οι σχέσεις αυτές μας οδηγούν ασφαλώς στο συμπέρασμα ότι οι συγκρίσεις σε επίπεδο pixel των καταγραφόμενων εικόνων είναι ουσιαστικά χωρίς νόημα. Το τελευταίο έγινε κατανοητό πολύ νωρίς, ακόμη και σε εφαρμογές αναζήτησης με βάση το περιεχόμενο [50], αν και αυτές ενδιαφέρονται περισσότερο για ταχύτητα παρά για ακρίβεια. Ένα σύστημα που αντιλαμβάνεται το περιβάλλον του θα πρέπει να μπορεί να εξάγει πληροφορία πέραν των χρωματικών συντεταγμένων των pixels των εικόνων που δέχεται. Αυτή η υψηλότερη σημασιολογικά πληροφορία

(semantics) θα καταγραφεί από το σύστημα ξανά στη μορφή κάποιων χαρακτηριστικών (features). Όπως ήδη αναφέραμε, η καταλληλότητα των εξαγόμενων χαρακτηριστικών, καθώς και του τρόπου με τον οποίο εξάγονται, εξαρτάται από ποικίλους παράγοντες και τελικώς από την εφαρμογή.

Όσον αφορά στα δισδιάστατα χαρακτηριστικά, μπορούμε να βρούμε ένα σύνολο από χαρακτηριστικά υψηλότερου επιπέδου από την πληροφορία των pixels, αλλά με σχετικά χαμηλό σημασιολογικό περιεχόμενο, τα οποία ταυτόχρονα να έχουν κάποιο νόημα στην εξαγωγή τους αναφορικά και με την ανθρώπινη αντίληψη. Στη συνέχεια αναφέρουμε μερικά από αυτά.

**Χρώμα:** Το χρώμα αποτελεί ίσως την πιο άμεσα εξαγόμενη πληροφορία από μία εικόνα. Όπως όλα τα χαρακτηριστικά, έτσι και το χρώμα μπορεί να μοντελοποιηθεί με διάφορους τρόπους, για παράδειγμα ακόμη και μία ένδειξη του τύπου 'υπάρχει / δεν υπάρχει κόκκινο χρώμα στην εικόνα' θα αποτελούσε ένα χαρακτηριστικό χρώματος. Βεβαίως, στη γενικότερη περίπτωση προσπαθούμε τα χαρακτηριστικά που εξάγουμε να είναι αρκούντως πληροφοριακά. Έτσι, το χρώμα μοντελοποιείται συνήθως από έναν μικρό αριθμό από κύρια χρώματα [50] ή μέσω του χρωματικού ιστογράμματος [152]. Προς την κατεύθυνση της εισαγωγής χωρικής πληροφορίας στο χαρακτηριστικό χρώματος, προτείνονται διάφορες μοντελοποιήσεις στη βιβλιογραφία όπως αυτές που χρησιμοποιούν τις χρωματικές ροπές [110] και την χωρική συσχέτιση [59]. Το χρώμα συνδυάζεται συχνά με τη χωρική κατανομή του στο χαρακτηριστικό της κατανομής χρώματος που περιλαμβάνει τοπολογική πληροφορία για την κατανομή χρώματος στην εικόνα [72, 109, 185].

**Υφή:** Όμοια με το χρώμα αντιμετωπίζεται και η υφή (texture) της εικόνας. Διάφορα μοντέλα έχουν προταθεί για τη μοντελοποίηση της υφής [68, 89, 120]. Όπως ακριβώς και το χρώμα, η υφή συνδυάζεται με τη χωρική κατανομή της στο χαρακτηριστικό της κατανομής υφής [95].

**Σχήμα:** Το σχήμα μπορεί γενικά να θεωρηθεί ως υψηλότερου επιπέδου χαρακτηριστικό. Βασικά, μοντελοποιείται και αυτό σε χαμηλό επίπεδο βάσει των ιδιοτήτων σχήματος τμημάτων της εικόνας όπως είναι, οι ακμές, το εμβαδό, η θέση, η επιμήκυνση και η εκκεντρότητα [50, 81]. Το περίγραμμα ή ο σκελετός σημασιολογικών αντικειμένων της εικόνας εξάγονται, μοντελοποιούνται και χρησιμοποιούνται σε υψηλότερο επίπεδο [117, 58]. Ιδιαίτερο ενδιαφέρον παρουσιάζει στην περίπτωση αυτή η αντιμετώπιση των μετασχηματισμών όψεως [62, 195, 9] και οι παραμορφώσεις [69] στις οποίες υπόκειται ένα αντικείμενο που εμφανίζεται στην εικόνα.

**Δομή:** Το χαρακτηριστικό της δομής ουσιαστικά επικαλύπτεται σε μεγάλο βαθμό από το χαρακτηριστικό του σχήματος υψηλότερου επιπέδου (π.χ. σκελετός, περίγραμμα) και από

τα χαρακτηριστικά κατανομής χρώματος και κατανομής υψής. Άλλα χαρακτηριστικά δομής προτείνονται στη βιβλιογραφία για να περιγράψουν τη δομική πληροφορία της εικόνας όπως αυτά της σύνθεσης [51] και της μορφής [116]. Ενδιαφέροντα χαρακτηριστικά δομής προκύπτουν και από μορφολογικές μεθόδους (σχετιζόμενα εν πολλοίς και με χαρακτηριστικά σχήματος).

*Κίνηση:* Η κίνηση αποτελεί ένα βασικό χαρακτηριστικό, το οποίο εξάγεται από τουλάχιστον δύο (συνήθως διαδοχικά) καρέ μίας ακολουθίας βίντεο. Η βιβλιογραφία βρίθει σε μεθόδους υπολογισμού κίνησης [155], αφού η εξαγωγή του χαρακτηριστικού αυτού οδηγεί σε έναν μεγάλο αριθμό από ενδιαφέρουσες εφαρμογές με την κατάτμηση κίνησης [10] και τον υπολογισμό τρισδιάστατης κίνησης [188] να είναι μόνο δύο από αυτές. Το χαρακτηριστικό δισδιάστατης κίνησης μπορεί να περιγραφεί με διάφορους τρόπους όπως για παράδειγμα το διάνυσμα μετατόπισης και τη στροφή των αντικειμένων μίας εικόνας πάνω στο επίπεδο, το σύνολο όμως των τρόπων αυτών προκύπτουν απευθείας από τα διανύσματα κίνησης των pixels ή των blocks της εικόνας, τα οποία λαμβάνονται μέσω των μεθόδων υπολογισμού κίνησης.

*Πρότυπα:* Είναι σύνηθες σε συγκεκριμένες κατηγορίες εφαρμογών αντί των παραπάνω χαρακτηριστικών να εξάγουμε πρότυπα (π.χ. τμήματα εικόνων) τα οποία εν συνεχεία να χρησιμοποιούμε αυτούσια, μετασχηματισμένα ή παραμορφωμένα [69] σε διαδικασίες εντοπισμού, αναγνώρισης κ.λπ. Διαδεδομένη για παράδειγμα είναι η εξαγωγή προτύπων από εικόνες με ανθρώπινα πρόσωπα [170].

*Μετασχηματισμοί:* Ενδιαφέροντα χαρακτηριστικά για εφαρμογές υπολογιστικής όρασης προκύπτουν και από μετασχηματισμό των εικόνων (ή περιοχών τους) από το πεδίο του χώρου σε άλλα πεδία μέσω κατάλληλων μετασχηματισμών, όπως για παράδειγμα ο μετασχηματισμός στο πεδίο της συχνότητας μέσω του μετασχηματισμού Fourier.

*Άλλα χαρακτηριστικά:* Πέραν των χαρακτηριστικών που αναφέραμε παραπάνω, υπάρχει μία πλειάδα χαρακτηριστικών που εξάγονται, μοντελοποιούνται και χρησιμοποιούνται σε εφαρμογές υπολογιστικής όρασης, για σκοπούς αναγνώρισης ή ανάκτησης ή και για εξαγωγή στη συνέχεια άλλων χαρακτηριστικών, συνήθως υψηλότερου επιπέδου. Ως παράδειγμα θα μπορούσαμε να αναφέρουμε την εξαγωγή της δισδιάστατης πληροφορίας εστίασης ή σκίασης από εικόνες για τον υπολογισμό του βάθους των αντικειμένων που περιέχουν [199, 4, 78].

Ενδιαφέρον έχει να δούμε μερικές από τις εφαρμογές όπου δισδιάστατα χαρακτηριστικά διαδραματίζουν σημαντικό ρόλο, αναφορικά πάντα με την υπολογιστική όραση και τη λειτουργία της αντίληψης των υπολογιστικών συστημάτων που προέρχεται από την όραση. Όπως είδαμε μέχρι τώρα, σε γενικές γραμμές τέτοιες εφαρμογές συγκεντρώνονται στον εντοπισμό,

στη χωροθέτηση, στην αναγνώριση, στην ταυτοποίηση, στην ταξινόμηση και στην ανάκτηση εικόνων, περιοχών τους και αντικειμένων τους.

Σημαντική έμφαση έχει δοθεί περισσότερο την τελευταία δεκαετία στο θέμα της αναζήτησης και ανάκτησης εικόνων και βίντεο με βάση το περιεχόμενο (content-based image and video query and retrieval). Συνήθη χαρακτηριστικά που χρησιμοποιούνται σε τέτοιες εφαρμογές είναι το χρώμα, η υφή και τα χαρακτηριστικά σχήματος χαμηλού επιπέδου. Ακριβώς επειδή τα αντίστοιχα συστήματα λειτουργούν συνήθως σε πραγματικό χρόνο, τα διανύσματα χαρακτηριστικών κρατώνται μικρά, ώστε τα πολύπλοκα χαρακτηριστικά σχήματος και δομής να αποφεύγονται. Ενδιαφέρον παρουσιάζεται στην ύπαρξη προσώπων ή κινούμενων αντικειμένων στις εικόνες και τις ακολουθίες τους, φθάνει η εξαγωγή τους να μη δημιουργεί μεγάλο υπολογιστικό φόρτο. Στην περίπτωση των ακολουθιών εικόνων, χαρακτηριστικά κίνησης εξάγονται επίσης από τα διανύσματα κίνησης, κυρίως όμως σε επίπεδο κινούμενων χωρίων ή αντικειμένων. Στη συνήθη περίπτωση οι ακολουθίες εικόνων αντιπροσωπεύονται από τα χαρακτηριστικά των *χαρακτηριστικών καρτέ*, τα οποία με τη σειρά τους επιλέγονται από τις *σκηνές* κάθε ακολουθίας. Η ανακρίβεια που εισάγεται από τη χρήση χαμηλού επιπέδου χαρακτηριστικών προς όφελος της ταχύτητας, καλύπτεται συχνά και από ένα σύστημα μάθησης ή ανατροφοδότησης (relevance feedback), ακολουθώντας την καθοδήγηση του χρήστη.

Ο *εντοπισμός κύριων κινούμενων αντικειμένων* (main mobile object detection) αποτελεί μία διαδικασία με ευρεία εφαρμογή, για παράδειγμα για σκοπούς παρακολούθησης εισβολέων ή στόχων (intrusion detection, target tracking) ή ακόμη και για χωροθέτηση ή εξαγωγή κινούμενων αντικειμένων (mobile object localization / extraction). Σε τέτοιες εφαρμογές χρησιμοποιούνται συνήθως χαρακτηριστικά κίνησης ή και συνδυασμός τους με χαρακτηριστικά σχήματος (π.χ. κίνηση περιοχών ή ακμών). Όταν είναι επιθυμητή η ακριβής εξαγωγή των αντικειμένων, χαρακτηριστικά κίνησης ή/και χρώματος χρησιμοποιούνται για *κατάτμηση κίνησης / χρώματος* (motion / color segmentation).

Σε εφαρμογές αναγνώρισης αντικειμένων (object recognition), τα χαρακτηριστικά σχήματος είναι περισσότερο κατάλληλα, ειδικότερα σε υψηλό επίπεδο όπως αυτά του περιγράμματος. Βεβαίως, όταν τα υπό αναγνώριση αντικείμενα έχουν ιδιότητες που τα διακρίνουν στο πεδίο του χρώματος, της συχνότητας ή της υφής, δεν αποκλείεται και η χρήση των χαρακτηριστικών αυτών, αν και συνήθως η πληροφορία που προσφέρουν είναι φτωχή για εφαρμογές αναγνώρισης. Σε συγκεκριμένες περιπτώσεις, όπως σε αυτή της αναγνώρισης ανθρώπινων προσώπων (face recognition), η εξαγωγή προτύπων κρίνεται συχνά δόκιμη.

Όπως μπορεί να συνάγει κανείς από τη μελέτη των χαρακτηριστικών και των εφαρμογών τους, στην προσπάθεια εξαγωγής πληροφορίας που έχει σημασιολογικό περιεχόμενο ώστε να προσομοιώνει καλύτερα την ανθρώπινη όραση, τα χαρακτηριστικά της δομής και της κίνησης παρουσιάζουν αυξημένο ενδιαφέρον. Ενδεικτικά για τις εφαρμογές που αναφέρθηκαν, η κατανομή χρώματος, η περιοχή κάλυψης ή το περίγραμμα και η κίνηση εμφανίζονται περισσότερο ‘πληροφοριακά’ από άλλα χαρακτηριστικά. Ακριβώς επειδή πρόκειται για χαρακτηριστικά σχετικά υψηλού επιπέδου η εξαγωγή τους δεν είναι όσο ακριβής ή όσο ταχεία είναι για παράδειγμα η εξαγωγή του χρωματικού ιστογράμματος. Λόγω λοιπόν της σημασίας τους αλλά και της δυσκολίας στην εξαγωγή και μοντελοποίησή τους, προσελκύουν την προσοχή μεγάλου μέρους της επιστημονικής κοινότητας, ώστε να αναμένονται σημαντικά αντίστοιχα αποτελέσματα στο άμεσο μέλλον.

Ας υποθέσουμε ότι έχουμε στη διάθεσή μας ένα υπολογιστικό σύστημα με μία στοιχειώδη δυνατότητα όρασης. Σύμφωνα με όσα περιγράψαμε, αυτό δεν είναι τίποτα άλλο παρά ένας απλός υπολογιστής, δηλαδή βασικά ένας επεξεργαστής και μία μνήμη αποθήκευσης, συνοδευόμενος από ένα σύστημα καταγραφής, δηλαδή μία κάμερα. Τα δεδομένα καταγραφής φθάνουν από την κάμερα στη μνήμη αποθήκευσης κι εκεί επεξεργάζονται από έναν κατάλληλο αλγόριθμο εξαγωγής χαρακτηριστικών με την ισχύ που παρέχει στο σύστημά μας ο επεξεργαστής. Είναι βεβαίως αναμενόμενο ότι τα δεδομένα καταγραφής είναι τόσο μεγάλου όγκου που πρέπει να παρεμβάλουμε μία μεγαλύτερη μονάδα αποθήκευσης (π.χ. ένα σκληρό δίσκο) ώστε να τα φιλοξενήσει. Πρακτικά λοιπόν, δεν χρειάζεται κανείς τίποτα περισσότερο από έναν προσωπικό υπολογιστή για να κατασκευάσει ένα σύστημα που ‘βλέπει’.

Οι εφαρμογές υπολογιστικής όρασης απαιτούν συνήθως επεξεργασία σε πραγματικό χρόνο. Αυτό σημαίνει ενδεικτικά ότι το υπολογιστικό μας σύστημα πρέπει να λαμβάνει μία interlaced ακολουθία εικόνων με ρυθμό 25 καρέ/δευτερόλεπτο (στην Ευρώπη) με μέγεθος καρέ  $576 \times 720$  εικονοστοιχεία (pixels) για απεικόνιση PAL/SECAM. Στη συνέχεια πρέπει να επεξεργάζεται αυτά τα στοιχεία με έναν πολύπλοκο αλγόριθμο, ο οποίος πιθανότατα χρησιμοποιεί δεδομένα καταγραφής προηγούμενων χρονικών στιγμών ή/και δεδομένα αποθηκευμένα στη μνήμη, στο δίσκο ή σε μία τοπική ή απομακρυσμένη βάση δεδομένων και μπορεί να παίρνει και μία απόφαση και να εκτελεί μία ενέργεια από ένα σύνολο προκαθορισμένων ενεργειών. Στη γενική περίπτωση συστημάτων πραγματικού χρόνου, οι παραπάνω πράξεις πρέπει να διαρκούν κατά μέσο όρο ένα εικοστό πέμπτο του δευτερολέπτου. Ένας μέσος προσωπικός υπολογιστής σήμερα μπορεί τουλάχιστον να δέχεται τα δεδομένα από ένα σύστημα καταγραφής και να τα απεικονίζει σε πραγματικό χρόνο, μετά και από μία απλή επεξεργασία τους σε πραγματικό χρόνο (για παράδειγμα, μία διεργασία αποκωδικοποίησης). Σημαντικό όμως βάρος επεξεργα-

σίας προστίθεται από τους πολύπλοκους αλγόριθμους εξαγωγής χαρακτηριστικών, όπως επίσης και τους αλγόριθμους σύγκρισης, ταυτοποίησης και αναγνώρισης αυτών. Μιλάμε τότε για το υπολογιστικό κόστος (computational cost) των αλγορίθμων που χρησιμοποιούμε. Όταν το υπολογιστικό κόστος είναι μεγάλο, αυτό υποδεικνύει ότι ένας μέσος προσωπικός υπολογιστής δεν έχει τη δυνατότητα να εκτελεί τις διεργασίες σε πραγματικό χρόνο.

Το υπολογιστικό κόστος έρχεται σε ευθεία αντιπαράβολή με την επιθυμητή ακρίβεια (accuracy). Όπως γίνεται φανερό, η τελευταία έχει να κάνει με την επιτυχία που αναμένουμε το σύστημά μας να έχει στις εργασίες που εκτελεί. Μεγάλη ακρίβεια απαιτεί συνήθως περισσότερο πολύπλοκους υπολογισμούς, άρα και έχει μεγαλύτερο υπολογιστικό κόστος. Στις εφαρμογές υπολογιστικής όρασης και ιδιαίτερα στις εφαρμογές που απαιτούν εξαγωγή χαρακτηριστικών μιλάμε για συμβιβασμό (trade-off) ακρίβειας και υπολογιστικού κόστους. Ένα σύστημα μεγάλης υπολογιστικής ακρίβειας έχει κατά κανόνα πολύ μεγάλες απαιτήσεις σε ισχύ επεξεργασίας και σε μνήμη ή ακόμη και σε ταχείς μονάδες αποθήκευσης. Στην περίπτωση που τα τελευταία είναι δυνατά από την υπάρχουσα τεχνολογία, συμβαίνει να έχουν ιδιαίτερα υψηλό κόστος.

Το παραπάνω ζήτημα έρχεται να εξηγήσει με απλό τρόπο το σκεπτικό των αλγορίθμων εξαγωγής χαρακτηριστικών, που θα παρουσιαστούν στα επόμενα κεφάλαια. Μπορεί βέβαια να ισχύει ο γενικός κανόνας της επίτευξης της μέγιστης δυνατής ακρίβειας στον ελάχιστο δυνατό χρόνο, πώς όμως θα καθορίσουμε την καταλληλότητα μίας προτεινόμενης τεχνικής και πώς θα την συγκρίνουμε με παρεμφερείς τεχνικές; Η απάντηση έρχεται συνήθως από την ίδια την εφαρμογή. Η εφαρμογή είναι αυτή που κατά βάση θα υποδείξει εάν ένα εξαγόμενο χαρακτηριστικό είναι κατάλληλο και ποιος είναι ο καταλληλότερος τρόπος εξαγωγής του, μοντελοποίησής του και χρήσης του. Για να γίνει αυτό κατανοητό μπορούμε να εξετάσουμε ένα απλό παράδειγμα αναγνώρισης προσώπου σε μία βάση δεδομένων προσώπων. Έστω ότι θέλουμε να αναπτύξουμε μία εφαρμογή εντοπισμού και αναγνώρισης προσώπων σε μία βάση δεδομένων με ακολουθίες εικόνων τηλεοπτικών προγραμμάτων. Στα τηλεοπτικά προγράμματα δεν μπορούν να τεθούν οποιοδήποτε περιορισμοί, οπότε είμαστε αναγκασμένοι να ψάξουμε για πρόσωπα σε οποιαδήποτε θέση και προσανατολισμό μέσα στην εικόνα. Το υπολογιστικό κόστος είναι εκ των πραγμάτων πολύ μεγάλο σε αυτήν την περίπτωση, όμως είναι απίθανο να θέλουμε να μοντελοποιήσουμε την πληροφορία αυτή σε πραγματικό χρόνο. Έτσι, αρκεί να θέσουμε το υπολογιστικό μας σύστημα σε λειτουργία και να αναμένουμε την ολοκλήρωση του εντοπισμού προσώπων μετά από ημέρες. Ταυτόχρονα, μπορούμε να απαιτήσουμε η διεργασία αυτή να πραγματοποιηθεί με πολύ μεγάλη ακρίβεια και το σύστημα να αποθηκεύσει τα αποτελέσματα σε μία νέα μορφή δεδομένων (metadata) η οποία να είναι ταχέως προσπελάσιμη

για μελλοντικές αναζητήσεις. Σαφώς δε, το σύνολο των προσώπων που ενδιαφερόμαστε να αναγνωρίσουμε σε τέτοιες ακολουθίες είναι σχετικά λίγα (π.χ. πολιτικοί, καλλιτέχνες κ.λπ.). Έτσι, όταν η διαδικασία (ή μέρος της) ολοκληρωθεί, θα είμαστε σε θέση να αναζητούμε το πρόσωπο X (από το συγκεκριμένο σύνολο) στη βάση των εικόνων και το σύστημα να το αναζητεί σε πραγματικό χρόνο 'βλέποντας' μόνο τα metadata. Παράλληλα, όταν θέλουμε να μεγάλωσουμε το σετ των υπό αναζήτηση προσώπων και να προσθέσουμε ένα πρόσωπο Y, μπορούμε σε μη-πραγματικό χρόνο να ζητήσουμε από το σύστημα να μας το αναγνωρίσει στα πρόσωπα που έχουν εντοπιστεί και να ενημερώσει τα metadata για μελλοντικές αναζητήσεις. Σε μία θεωρητικά όμοια εφαρμογή αναγνώρισης προσώπων, όπου θέλουμε να αναζητήσουμε ένα πρόσωπο σε μία βάση υπόπτων για κακουργηματικές πράξεις, ο παραπάνω αλγόριθμος θα ήταν πλήρως ακατάλληλος, αφού θα θέλαμε η αναζήτηση να πραγματοποιείται άμεσα (και όχι μετά από ημέρες/μήνες), ενώ κατά την εισαγωγή του υπόπτου σε περιορισμένο αριθμό φωτογραφιών θα μπορούσαμε να υποβάλλουμε συγκεκριμένους περιορισμούς για τη θέση, τον προσανατολισμό και το μέγεθος του προσώπου. Σε αυτήν την περίπτωση λοιπόν, ένας γρήγορος αλγόριθμος αναγνώρισης προσώπων που βρίσκονται σε συγκεκριμένο μέγεθος και θέση στην εικόνα και έχουν κάποια χαρακτηριστικά που ομοιάζουν στο διαθέσιμο προσχέδιο (παράδειγμα) θα ήταν καταλληλότερος. Παράλληλα, θα απαιτούσαμε από το σύστημα να μας παράσχει ένα σύνολο από όμοια πρόσωπα, παρά να αναγνωρίσει ακριβώς το πρόσωπο στη βάση, ώστε να διεξάγουμε έναν δεύτερο οπτικό πλέον έλεγχο. Στα δύο αυτά λοιπόν θεωρητικώς όμοια παραδείγματα οι αλγόριθμοι εντοπισμού, εξαγωγής, αναγνώρισης και απόφασης, αλλά και τα εξαγόμενα metadata διαφέρουν δραματικά.

Αυτό βεβαίως δε σημαίνει ότι δεν υπάρχει μέτρο σύγκρισης των αλγορίθμων υπολογιστικής όρασης. Μπορεί κανείς να δει ότι δύο αλγόριθμοι με όμοια δεδομένα εισόδου και εξόδου μπορούν να συγκριθούν ευθέως ως προς την ακρίβεια και το υπολογιστικό τους κόστος. Στην περίπτωση που τα αποτελέσματα εξαρτώνται ευθέως από την εφαρμογή, έχουν οριστεί σε πολλές περιπτώσεις εφαρμογών πρότυπες ακολουθίες εικόνων όπου οι ερευνητές μπορούν να ελέγχουν την επιτυχία των τεχνικών που προτείνουν ως προς την ταχύτητα εκτέλεσης και την επιτευχθείσα ακρίβεια, σε αντιπαράβολή με άλλες τεχνικές στην ίδια περιοχή. Από την άλλη μεριά, εξετάζοντας μία προτεινόμενη τεχνική θα πρέπει κανείς να έχει υπόψη του τις εφαρμογές όπου αυτή θα εφαρμοστεί αλλά και την αντιπαράθεση ακρίβειας - υπολογιστικού κόστους. Ουσιαστικά, στην υπολογιστική όραση κρίνουμε την 'επιτυχία στην εξισορρόπηση των δύο αυτών παραγόντων ως προς την οικογένεια των εφαρμογών όπου η τεχνική εφαρμόζεται'.

Στα επόμενα κεφάλαια του πρώτου μέρους παρουσιάζουμε πρωτότυπες μεθόδους εξαγωγής χαρακτηριστικών δομής και κίνησης στις δύο διαστάσεις. Σύμφωνα με όσα αναφέραμε παρα-



πάνω, για κάθε χαρακτηριστικό που προτείνεται, εξάγεται, μοντελοποιείται και χρησιμοποιείται, δίνουμε ενδεικτικά τις εφαρμογές όπου η χρήση του θα ήταν κατάλληλη (ή ακατάλληλη), σχολιάζουμε και παραπέμπουμε σε συγγενείς κατά οποιονδήποτε τρόπο μεθόδους της βιβλιογραφίας, μελετούμε την επίδοση της προτεινόμενης μεθόδου σε πειραματικές ακολουθίες και σχολιάζουμε την επίδοση σε ακρίβεια και υπολογιστικό κόστος. Στα συμπεράσματα κάθε κεφαλαίου αναφερόμαστε ενδεικτικά σε μελλοντικές επεκτάσεις και σχολιάζουμε τη συνεισφορά της εργασίας μας στο γενικότερο αντικείμενο, αφού όπως είναι φανερό βρισκόμαστε σήμερα μακριά από την κατασκευή ενός υπολογιστικού συστήματος όρασης γενικού σκοπού.

## ΚΕΦΑΛΑΙΟ 3

### Εξαγωγή και Μοντελοποίηση της Χωρικής Πληροφορίας Χρώματος Εικόνων

#### 3.1 Εισαγωγή / Αναδρομή

Την τελευταία δεκαετία, στο πλαίσιο της επιτυχούς και αποδοτικής συλλογής πληροφορίας, του χειρισμού και της διανομής αυτής, έχει εμφανιστεί έντονα η ανάγκη για κατάλληλη οργάνωση και διάθεση του οπτικοακουστικού υλικού. Οι κατανεμημένες ψηφιακές βιβλιοθήκες αναμένονται να διευκολύνουν τη διαχείριση και ανταλλαγή της οπτικοακουστικής πληροφορίας, η οποία παραδοσιακά δεν ήταν διαθέσιμη εκτός των οργανισμών όπου παραγόταν και χρησιμοποιούταν. Ταυτόχρονα, ένας αριθμός από βασικές απαιτήσεις έχουν παρουσιαστεί για τον αποδοτικό χειρισμό δεδομένων τέτοιου μεγέθους, με πιο σημαντικές τη δεικτοδότηση, την αναζήτηση και την ανάκτηση της πληροφορίας. Σ' αυτό το πλαίσιο, η ιδέα της ανάκτησης του οπτικοακουστικού υλικού με βάση το περιεχόμενο έχει συγκεντρώσει την προσοχή, υποστηριζόμενη κυρίως από την αντίστοιχη πρόοδο στα συστήματα βάσεων δεδομένων, στην επεξεργασία εικόνων και στις τεχνολογίες αναγνώρισης προτύπων.

Συστήματα με δυνατότητες αναζήτησης και ανάκτησης εικόνων με βάση το περιεχόμενο έχουν περιγραφεί στη βιβλιογραφία κατά την τελευταία δεκαετία με τα Virage [51], QBIC [37], Photobook [115, 116], VisualSEEk [145, 146], C-BIRD [81] και Chabot [102] να είναι μερικά από τα πιο γνωστά μεταξύ άλλων. Στην πραγματικότητα, το ενδιαφέρον σ' αυτό το πεδίο αυξάνει με τόσο ταχείς ρυθμούς, ώστε η βιβλιογραφία να βρίθει από σχετικές προσεγγίσεις, οι οποίες εκτείνονται από απλές προσθήκες-επεκτάσεις σε υπάρχουσες τεχνικές έως πολύπλοκα ολοκληρωμένα συστήματα. Όσον αφορά τα τελευταία, τα Netra [86], MARS [131], SaFe [147], PISARO [137], Enigma [43], Webseer [38], ImageRover [136], PicToSeek [44], Color-WISE [139] και MARCO [134] έχουν σχεδιασθεί και υλοποιηθεί. Ταυτόχρονα, νέα συστήματα αναπτύσσονται (ή υπάρχοντα συστήματα 'αναβαθμίζονται') ώστε να συμπεριλάβουν πληροφορία βίντεο και ήχου, όπως για παράδειγμα τα VideoQ [20] και DiVAN [158] αντίστοιχα. Ένα υποσύνολο των παραπάνω συστημάτων, π.χ. τα [51, 37, 158], έχουν υλοποιηθεί και βρίσκονται επί του παρόντος στο στάδιο της εκτίμησης ή ακόμη και της εκμετάλλευσης ως εμπορικά προϊόντα.

Τα συστήματα αναζήτησης εικόνων με βάση το περιεχόμενο βασίζονται στη μοντελοποίηση της πληροφορίας περιεχομένου των εικόνων στο πλαίσιο κάποιων προαποφασισμένων χαρακτηριστικών. Ένας αριθμός από τέτοια χαρακτηριστικά έχουν προταθεί είτε ως ειδικά μέρη ολοκληρωμένων συστημάτων είτε ως ανεξάρτητες προσεγγίσεις. Σχεδόν όλα τα χαρακτηριστικά

που έχουν προταθεί στη βιβλιογραφία, μαζί με τις τεχνικές μοντελοποίησης που τα συνοδεύουν, μπορούν να κατηγοριοποιηθούν σε έναν περιορισμένο αριθμό από οπτικά κατανοητές και διαχωρίσιμες ιδιότητες της εικόνας, όπως το χρώμα, η υφή και το σχήμα. Αρκετές στρατηγικές μοντελοποίησης αυτών των χαρακτηριστικών της εικόνας έχουν υιοθετηθεί, διαφέροντας κυρίως στην κάλυψη (support) των χαρακτηριστικών (τοπική ή ολική) και στη μέθοδο ανάλυσής τους. Η μοντελοποίηση των χαρακτηριστικών επηρεάζει (και επηρεάζεται από) τις διαδικασίες εξαγωγής και σύγκρισης των χαρακτηριστικών αυτών, οι οποίες με τη σειρά τους προσπαθούν να προσομοιώσουν τον τρόπο με τον οποίο το ανθρώπινο μάτι προσλαμβάνει και συγκρίνει τις εικόνες βάσει των χαρακτηριστικών τους. Τέλος, η μέθοδος μοντελοποίησης των χαρακτηριστικών μπορεί να εξαρτάται από την συγκεκριμένη κάθε φορά εφαρμογή (application-specific), όταν τα περιεχόμενα της βάσης δεδομένων είναι γενικά γνωστά. Για παράδειγμα, ποιος θα μοντελοποιούσε το χαρακτηριστικό χρώματος με ένα ή δύο κύρια χρώματα (dominant colors) εάν η βάση περιείχε μόνο φωτογραφίες από το βυθό ή από το λυκόφως;

Το *χρώμα* (color) μοντελοποιείται συνήθως από έναν μικρό αριθμό από κύρια (dominant) χρώματα ή ακόμη μέσω του χρωματικού ιστογράμματος (color histogram) [50]. Τέτοιες προσεγγίσεις μοντελοποιούν το χαρακτηριστικό *συνολικό χρώμα* (global color). Όπως όμως θα αναλύσουμε στη συνέχεια, το *χαρακτηριστικό κατανομή χρώματος* (color composition), που περιλαμβάνει τοπολογική πληροφορία για την κατανομή του χρώματος στην εικόνα, μπορεί να δώσει καλύτερα αποτελέσματα σύγκρισης και ανάκτησης [72, 109, 185]. Η *υφή* (texture) ως ένα συνολικό χαρακτηριστικό μοντελοποιείται μέσω στατιστικών και δομικών μοντέλων [68], ενώ διάφορες μέθοδοι έχουν ακόμη προταθεί, περιλαμβάνοντας πολυδιακριτικές (multiresolution) προσεγγίσεις [89] και ανάλυση σε ορθογώνιες συνιστώσες [120, 115]. Όμοια με την κατανομή χρώματος, το *χαρακτηριστικό κατανομή υφής* (texture composition) περιλαμβάνει την πληροφορία υφής για οπτικά διαχωρίσιμα τμήματα της εικόνας [95]. Το *σχήμα* (shape) μοντελοποιείται συνήθως βάσει των ιδιοτήτων σχήματος τμημάτων της εικόνας όπως είναι το εμβαδό, η επιμήκυνση και η εκκεντρότητα [50] ή ακόμη και ο προσανατολισμός των ακμών [81]. Η πληροφορία σχήματος σε μία εικόνα θεωρείται συχνά ως σημασιολογική (semantic) πληροφορία υψηλότερου επιπέδου. Σ' αυτό το πλαίσιο, το περίγραμμα (contour) ή ο σκελετός (skeleton) σημασιολογικών αντικειμένων της εικόνας εξάγονται, μοντελοποιούνται και χρησιμοποιούνται στη συνέχεια για σύγκριση και ανάκτηση, ενώ μετασχηματισμοί [62, 195, 9] και παραμορφώσεις [69] των αντικειμένων λαμβάνονται επίσης υπόψιν. Ένας αριθμός από πρόσθετα χαρακτηριστικά έχουν προταθεί στη βιβλιογραφία για να υποβοηθήσουν / βελτιώσουν την ανάκτηση με βάση το περιεχόμενο, περιλαμβάνοντας για παράδειγμα εκδοχές του μετασχηματισμού Karhunen-Loeve [116] ή ανάλυση σε wavelets.

Τα πρότυπα συστήματα αναζήτησης εικόνων έχουν επεκταθεί ώστε να συμπεριλάβουν έναν αριθμό από πρόσθετες δυνατότητες. Αλγόριθμοι εντοπισμού και αναγνώρισης προσώπων (face detection and recognition) έχουν ήδη συμπεριληφθεί σε αρκετά συστήματα όπως η eigenface image database [114] συγγενής με το σύστημα Photobook. Η σημασιολογική πληροφορία στα πλαίσια των περιεχόμενων αντικειμένων και του σχήματός τους [9, 49] καθώς και των μεταξύ τους χωροχρονικών συσχετισμών χρησιμοποιείται επίσης, ειδικότερα δε στα πλαίσια της δεικτοδότησης βίντεο (video indexing) [128, 143]. Τα συστήματα αναζήτησης βίντεο με βάση το περιεχόμενο βασίζονται στον εντοπισμό σκηνών (shot detection) και στην εξαγωγή χαρακτηριστικών καρτέ (keyframes extraction) [10] παράλληλα με τις υπάρχουσες τεχνικές ανάκτησης στατικών εικόνων, ενώ η εξαγωγή οπτικά ανεξάρτητων αντικειμένων γίνεται δυνατή μέσω τεχνικών εντοπισμού κινούμενων αντικειμένων (mobile object detection), π.χ. [101, 194, 65]. Η επέκταση των συστημάτων ανάκτησης εικόνων στα συστήματα ανάκτησης βίντεο πραγματοποιείται σε συμφωνία με τις οδηγίες των προτύπων κωδικοποίησης βίντεο MPEG-4 και MPEG-7 για αντικειμενοστρεφή κωδικοποίηση και ανάκτηση πληροφορίας αντίστοιχα (βλ. για παράδειγμα την [133] και τις εκεί αναφορές). Η δεικτοδότηση του περιεχομένου βίντεο δίνει ένα σύνολο από πρόσθετα πληροφοριακά χαρακτηριστικά όπως είναι η *κίνηση* (motion), η οποία αντιστοιχεί στην προβεβλημένη κίνηση ενός τρισδιάστατου (3Δ) αντικειμένου στο επίπεδο της εικόνας. Αυτή η πληροφορία, πέρα από τη χρησιμοποίησή της ‘ως έχει’ για να χαρακτηρίσει οπτικό υλικό, μπορεί να χρησιμοποιηθεί και για να διευκολύνει την εξαγωγή τοπολογικής πληροφορίας ή ακόμη και την εκτίμηση 3Δ χαρακτηριστικών όπως η 3Δ κίνηση και δομή [188]. Τα συστήματα αναζήτησης με βάση το περιεχόμενο έχουν επίσης χρησιμοποιηθεί σε συνδυασμό με έξυπνους πράκτορες (intelligent agents) [193, 73] και εξελιγμένες τεχνικές αποθήκευσης (storage) για αποδοτική αναζήτηση και οργάνωση των δεδομένων.

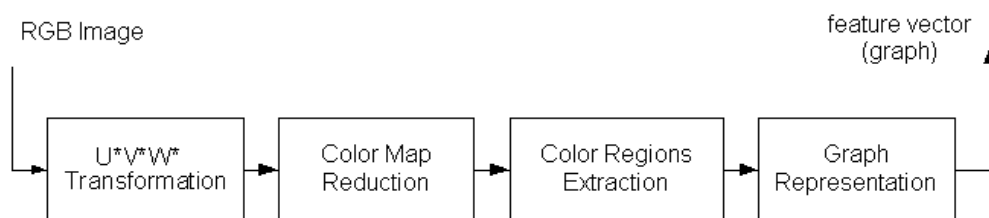
Παρόλη την αυξανόμενη προσπάθεια στην περιοχή, η καθημερινή πρακτική δείχνει ότι τα συστήματα αναζήτησης με βάση το περιεχόμενο δεν έχουν ακόμη αγγίξει τις προσδοκίες και συχνά θεωρούνται κατώτερα από συστήματα που υποστηρίζουν χαρακτηρισμό (annotation). Από την άλλη πλευρά, η αναγκαία επίπονη ανθρώπινη παρέμβαση και η αναπόφευκτη υποκειμενικότητα του χαρακτηριστή κάνουν τέτοια συστήματα ακατάλληλα στις περισσότερες περιπτώσεις. Οι τεχνικές που βασίζονται στην τοπολογική ανάλυση των εικόνων (region-based) είναι πιθανότατα οι περισσότερα υποσχόμενες για αυτόματη δεικτοδότηση βίντεο, αφού περιλαμβάνουν χαρακτηριστικά με τοπολογική πληροφορία, τα οποία περιγράφουν καλύτερα τα οπτικά δεδομένα, ενώ ταυτόχρονα υποστηρίζουν και μετρήσεις τοπικής ομοιότητας (μεταξύ τμημάτων των εικόνων). Προσεγγίσεις που περιλαμβάνουν την εξαγωγή τοπικής πληροφορίας δίνουν τη δυνατότητα για αποδοτικές συγκρίσεις μεταξύ τμημάτων διαφορετικού μεγέθους και

θέσης πάνω στην εικόνα, ή ακόμη και για μετασχηματισμένες και παραμορφωμένες εκδοχές του ίδιου αντικειμένου. Το τελευταίο παρουσιάζει μεγάλο ενδιαφέρον αφού, για παράδειγμα, μία σημαία ως αντικείμενο μπορεί να υπάρχει σε αρκετές φωτογραφίες, όμως πολύ σπάνια μπορεί να βρεθεί στην ίδια θέση, προσανατολισμό ή μέγεθος.

Σ' αυτό το πλαίσιο, το παρόν κεφάλαιο ασχολείται με το χαρακτηριστικό της κατανομής χρώματος και προτείνει ένα νέο σετ μεθόδων για την εξαγωγή του, τη μοντελοποίησή του και τη σύγκριση εικόνων ως προς αυτό. Η προτεινόμενη μέθοδος βασίζεται στην εύρεση κύριων χρωματικών αντικειμένων σε δεδομένες εικόνες και στη σύγκριση των τελευταίων βάσει των χρωματικών αντικειμένων που περιέχονται σε αυτές και της διασυνδεσιμότητάς τους. Η προσέγγιση αυτή φαίνεται να είναι αποδοτικότερη από τους παραδοσιακούς αλγορίθμους εξαγωγής της κατανομής χρώματος, εφόσον δίνει τη δυνατότητα εντοπισμού των χρωματικών αντικειμένων υπό διαφορετική κλίμακα και θέση στην εικόνα. Η μέθοδος δίνει έμφαση στην αποδοτική κατάτμηση της εικόνας (image segmentation) σε δεσπόζουσες χρωματικές περιοχές και στη μοντελοποίηση της εξαγόμενης πληροφορίας (μέγεθος περιοχής, χρώμα περιοχής και τρόπος σύνδεσης περιοχών) στα πλαίσια ενός απλού γράφου. Οι κόμβοι του γράφου αντιπροσωπεύουν τις εξαγόμενες περιοχές της εικόνας, ενώ οι ακμές του γράφου τη διασυνδεσιμότητα μεταξύ των αντίστοιχων κόμβων (περιοχών). Ως ένα ποσοτικό χαρακτηριστικό διασυνδεσιμότητας χρησιμοποιούμε εδώ το μήκος του κοινού ορίου. Η προτεινόμενη προσέγγιση έχει το άμεσο πλεονέκτημα ότι οδηγίες / μέθοδοι για αποδοτική κατασκευή και πρόσβαση στο γράφο δίνονται εκτεταμένα στη βιβλιογραφία. Παράλληλα, μοντελοποιώντας την κατανομή χρώματος μέσω ευέλικτων γράφων, επιτυγχάνουμε την απουσία περιορισμών θέσης και κλίμακας για τα περιεχόμενα αντικείμενα και την εξάρτησή της μόνο από το μέγεθος, το χρώμα και τη διασυνδεσιμότητα των αντικειμένων.

Η συγκεκριμένη μέθοδος έχει περιγραφεί αναλυτικά στην εργασία [189], ενώ βασίζεται σε παλαιότερες ιδέες, όπως αυτές δίνονται στις εργασίες [187, 185].

Η Ενότητα 3.2 πραγματοποιεί μία ανασκόπηση στη μελέτη των χαρακτηριστικών χρώματος και κατανομής χρώματος και υπογραμμίζει την ανωτερότητα του δεύτερου. Στην Ενότητα 3.3, δίνεται μία απλή και ταχεία μέθοδος εξαγωγής των χρωματικών αντικειμένων, το βήμα αυτό όμως μπορεί να αντικατασταθεί από οποιονδήποτε αλγόριθμο κατάτμησης χρώματος (color segmentation). Η μοντελοποίηση του χαρακτηριστικού κατανομής χρώματος βάσει ενός απλού γράφου εισάγεται στην Ενότητα 3.5, ενώ μία συνάρτηση κόστους προτείνεται στην Ενότητα 3.6 για τη σύγκριση εικόνων επί του χαρακτηριστικού αυτού. Τα αλγοριθμικά βήματα που περιλαμβάνονται σε κάθε ενότητα δίνονται σε μορφή πίνακα απλών βημάτων. Πειράματα που



**Σχήμα 3.1** Το προτεινόμενο σύστημα εξαγωγής χρώματος σε μπλοκ-διάγραμμα.

επιδεικνύουν την επίδοση του προτεινόμενου αλγορίθμου έχουν συμπεριληφθεί στην Ενότητα 3.7, ενώ τα σχετικά γενικότερα συμπεράσματα δίνονται στην Ενότητα 3.8.

### 3.2 Το χρώμα ως χαρακτηριστικό της εικόνας

Το χρώμα μίας εικόνας είναι ίσως το απλούστερο και το πιο πληροφοριακό χαρακτηριστικό που μπορεί κανείς να εξάγει στη γενική περίπτωση. Χωρίς να δώσει κανείς ιδιαίτερη προσοχή σε μία φωτογραφία, μπορεί σίγουρα να αποφανθεί εάν η συγκεκριμένη φωτογραφία είναι όμοια με κάποια άλλη ως προς το χρώμα. Ίσως αυτός είναι και ο λόγος που το χρωματικό ιστόγραμμα ήταν μία από τις πρώτες μετρικές που χρησιμοποιήθηκαν ευρέως όταν πρωτοεμφανίστηκε η ιδέα της δεικτοδότησης με βάση το περιεχόμενο.

Η προσέγγιση των Swain και Ballard [152] είναι ίσως η περισσότερο δημοφιλής για δεικτοδότηση χρώματος στα πλαίσια του χρωματικού ιστογράμματος. Η συγκεκριμένη μέθοδος, όπως και κάθε μέθοδος που βασίζεται στο ιστόγραμμα, συγκρίνει τις εικόνες βάσει των χρωμάτων τους λαμβάνοντας υπόψη το εμβαδό που καλύπτεται από τον κάθε χρωματικό τόνο. Μικρές αλλαγές στον προσανατολισμό και στη γωνία παρατήρησης δεν επηρεάζουν σημαντικά την ανάκτηση - αλλαγές όμως στις συνθήκες φωτισμού πρέπει να αντιμετωπιστούν χωριστά όπως στην εργασία των Funt και Finlayson [40]. Με τη σειρά τους οι Ennesser και Medioni στην [34] προτείνουν μία διαφορετική προσέγγιση στην τοπική σύγκριση των ιστογραμμάτων. Μία μέθοδος για βελτιωμένα αποτελέσματα υπό συνθήκες μεταβαλλόμενου σημείου παρατήρησης και φωτισμού δίνεται επίσης από τους Gevers και Smeulders στην [45]. Γενικά, τέτοιες μέθοδοι, που βασίζονται στην πληροφορία ιστογράμματος, χρησιμοποιούνται ευρέως στη δεικτοδότηση και ανάκτηση οπτικής πληροφορίας [37, 102].

Η ανάγκη ενσωμάτωσης χωρικής πληροφορίας χρώματος στην αναζήτηση εικόνων έχει μολτατά προκύψει. Ένα σύνολο μεθόδων έχουν προταθεί στη βιβλιογραφία στην κατεύθυνση της αναβάθμισης του ιστογράμματος, ούτως ώστε η χωρική ομοιογένεια των pixels να συμπεριλαμβάνεται σε αυτό. Το σύνολο αυτό των μεθόδων περιλαμβάνει την προσέγγιση του διανύσματος χρωματικής ομοιογένειας των Pass και Zabih [111], όπου ένα pixel είναι

‘ομοιογενές’ όταν είναι μέρος μίας περιοχής όμοιου χρώματος. Όμοια, οι Huang *et al.* [59] προτείνουν τα χρωματικά correlograms για να εκφράσουν τη χωρική συσχέτιση ζευγών από χρώματα ως προς την απόστασή τους. Οι προσεγγίσεις που αφορούν στα μείζονα χρώματα (dominant colors) προσπαθούν επίσης να βελτιώσουν τα κριτήρια που βασίζονται στο ιστόγραμμα, ιδιαίτερα όταν περιλαμβάνουν και χωρική πληροφορία όπως στη δουλειά των Hsu *et al.* [56]. Με έναν τρόπο όμοιο με τις μεθόδους ιστογράμματος, οι Paschos και Radev [110] αποθηκεύουν αντιπροσωπευτικές τιμές των χρωματικών ροπών (chromaticity moments) με κάθε εικόνα για δεικτοδότηση χρώματος, ενώ οι Stricker και Dimai [151] διαιρούν τις εικόνες σε έναν ορισμένο αριθμό από επικαλυπτόμενες περιοχές και εξάγουν τις τρεις πρώτες χρωματικές ροπές για κάθε περιοχή.

Προς την κατεύθυνση του συνδυασμού της πληροφορίας χρώματος με τοπολογική πληροφορία, οι Gong *et al.* [48] εξάγουν ένα μικρό σύνολο από αντιπροσωπευτικά χρώματα και ένα αντίστοιχο σύνολο από περιοχές σε κάθε εικόνα. Όμοια, οι εικόνες κατατιμώνται σε τάξεις χρώματος και σχήματος, που με τη σειρά τους μοντελοποιούνται χωριστά, στην εργασία των Mehtre *et al.* [93]. Η χωρική ταξινόμηση χρώματος (spatial color clustering) αντιμετωπίζεται σε μία σχετική εργασία των Kankanhalli *et al.* [71]. Με τη σειρά τους, οι Tao και Grosky [154] εξάγουν χαρακτηριστικά σημεία βάσει μη-επικαλυπτόμενων παραλληλογράμμων και μοντελοποιούν την κατανομή του χρώματος επί τριγώνων τα οποία παράγονται από το εξηγμένο σύνολο σημείων.

Οι περισσότεροι ίσως σχετικές προσεγγίσεις με το παρόν κεφάλαιο είναι αυτές που προτείνονται στις αναφορές [91, 109]. Στην [91] οι Matas *et al.* εισάγουν τον *γράφο χρωματικής γειτνίασης* (color adjacency graph), ο οποίος εκμεταλλεύεται την ιδέα του γράφου γειτνίασης (adjacency graph), που με τη σειρά του χρησιμοποιείται ευρέως στη βιβλιογραφία σε διάφορα ερευνητικά αντικείμενα. Οι Park *et al.* [109] επεκτείνουν τον τελευταίο για να κατασκευάσουν τον *τροποποιημένο γράφο χρωματικής γειτνίασης* (modified color adjacency graph) περιλαμβάνοντας την πληροφορία γειτνίασης των pixels στο διάνυσμα χαρακτηριστικών (feature vector). Παράλληλα, ο *γράφος χωρικής διασποράς* (spatial variance graph), που περιέχει πληροφορία αυτοσυσχέτισης και ετεροσυσχέτισης των χρωματικών περιοχών, προτείνεται για βελτίωση της επίδοσης του αλγορίθμου. Η προσέγγιση των Corridoni *et al.* [22] είναι επίσης σε κάποια έκταση σχετική, αφού χρησιμοποιεί μία δομή δεδομένων σε ένα γράφο πολλών επιπέδων (multi-layered graph) για να αναπαραστήσει τη χωρική κατανομή του χρώματος.

Στην προτεινόμενη προσέγγιση, ο γράφος χρωματικής γειτνίασης επεκτείνεται ώστε να συμπεριλάβει έναν αριθμό από χαρακτηριστικά των περιοχών όπως το εμβαδό, το χρώμα και οι γείτονες. Η γειτνίαση των περιοχών χαρακτηρίζεται από τον τύπο της διασύνδεσης και από

**Πίνακας 3.1** Η στρατηγική της πλημμύρας

- 
1. Ξεκίνα με οποιοδήποτε pixel  $p_0$  (για παράδειγμα από την άνω-αριστερά γωνία της εικόνας).
  2. Ανέθεσε το  $p_0$  στην περιοχή  $r_0$ .
  3. Σημάδεψε το  $p_0$  ως ‘επεξεργασμένο’.
  4. Εάν όλοι οι γείτονες του  $p_0$  έχουν ήδη επεξεργαστεί, ψάξε για το επόμενο μη-επεξεργασμένο pixel  $p_1$  και ανέθεσέ το σε μία νέα χρωματική περιοχή  $r_1$ . Πήγαινε στο βήμα 3 για το pixel  $p_1$ .
  5. Αλλιώς, για κάθε μη-επεξεργασμένο pixel  $p_2$  στην eight-connected γειτονιά του  $p_0$ , υπολόγισε την τιμή μίας συνάρτησης ομοιότητας  $f_p(p_2, \dots)$  με δυαδική έξοδο.
  6. Εάν 1, ανέθεσε το  $p_2$  στην  $r_0$  και βάλε το  $p_2$  σε μία στοίβα προτεραιότητας  $\Sigma$ .
  7. Βγάλε το πρώτο στοιχείο  $p_3$  από τη  $\Sigma$  και πήγαινε στο βήμα 3 για το pixel  $p_3$ .
- 

το μήκος του κοινού ορίου. Αυτά τα χαρακτηριστικά, μαζί με τους αλγορίθμους που δίνονται σε αυτό το κεφάλαιο, οδηγούν σε βελτιωμένες διαδικασίες κατάτμησης, αναπαράστασης με γράφο και σύγκρισης γράφων, πάντα ως προς το ζητούμενο χαρακτηριστικό της εικόνας, οι οποίες με τη σειρά τους δείχνουν να προσομοιώνουν ικανοποιητικώς την ανθρώπινη αντίληψη στη σύγκριση εικόνων. Το προτεινόμενο σύστημα σε μπλοκ-διάγραμμα απεικονίζεται στο Σχήμα 3.1.

Μετά από έναν κατάλληλο μετασχηματισμό του χρωματικού χώρου (στο χώρο  $U^*V^*W^*$  ή άλλον), η χρωματική παλέτα μειώνεται σε μία αντίστοιχη ομοιόμορφη και οι δεσπόζουσες χρωματικές περιοχές εξάγονται σε δύο βήματα: με μία απλή στρατηγική ‘πλημμύρας’ (flood) και ένα δεύτερο βήμα που εκμεταλλεύεται την προτεινόμενη αναπαράσταση με γράφο. Ο τελικώς παραγόμενος γράφος (ισοδύναμα ‘το διάγραμμα χαρακτηριστικών’) γίνεται διαθέσιμος για συγκρίσεις βάσει των αλγορίθμων που επίσης προτείνονται.

### 3.3 Εξαγωγή χρωματικών αντικειμένων

Το πρόβλημα της σύγκρισης δύο εικόνων ως προς την χωρική κατανομή χρώματος σε αυτές περιλαμβάνει αυτόματα τεχνικές κατάτμησης χρώματος (color segmentation) και/ή τεχνικές χωρικής αναπαράστασης χρώματος, για παράδειγμα ανάλυση σε τετραδικά δένδρα (quad-tree decomposition). Πρακτικά, η αναπαράσταση χρώματος με χρήση τετραδικών δένδρων, ή άλλες όμοιες τεχνικές, έχει το πλεονέκτημα της ταχείας υλοποίησης τόσο στην κατασκευή του δένδρου όσο και στη διαδικασία σύγκρισης, βασιζόμενη στα αποτελέσματα της θεωρίας και των εφαρμογών των τετραδικών δένδρων. Μπορεί όμως να δει κανείς ότι ακόμη και μία πολύ μικρή αλλαγή στη θέση ή στον προσανατολισμό των χρωματικών αντικειμένων στη σκηνή οδηγεί σε απαράδεκτα αποτελέσματα σύγκρισης. Ίσως το πιο χαρακτηριστικό παράδειγμα είναι η



περίπτωση μιας κινούμενης κάμερας σε μία καταγραφή βίντεο, όπου μία αλλαγή στην καταγραφόμενη ακολουθία λόγω της κίνησης της κάμερας οδηγεί σε πλήρως διαφορετική χρωματική αναπαράσταση με τετραδικό δένδρο ακόμη και για διαδοχικά καρέ.

Με αυτό το σκεπτικό, οι τεχνικές κατάτμησης χρώματος είναι περισσότερο κατάλληλες ως προς την εξαγωγή του χαρακτηριστικού κατανομής χρώματος σε εικόνες. Στο παρόν κεφάλαιο, συνδυάζουμε μία στρατηγική πλημμύρας για κατάτμηση χρώματος σε επίπεδο pixel ως πρώτο βήμα, μαζί με ένα δεύτερο βήμα που εκμεταλλεύεται την πληροφορία που αποθηκεύεται στον γράφο γειτνίασης.

Η ευρέως χρησιμοποιούμενη στρατηγική της πλημμύρας περιλαμβάνει την επίσκεψη όλων των pixels της εικόνας με έναν επαναληπτικό τρόπο, ώστε να δίνεται προτεραιότητα στα γειτονικά (eight-connected) pixels των οποίων οι τιμές ικανοποιούν ένα κριτήριο. Μ' αυτή τη λογική, έστω  $p$  ένα pixel της εικόνας και  $c_p$  το  $3 \times 1$  διάνυσμα που περιέχει τις συντεταγμένες του στον  $3\Delta$  χώρο των χρωμάτων. Υποθέτοντας ότι το pixel  $p$  έχει ήδη ανατεθεί σε μία χρωματική περιοχή  $r$ , τότε κάθε pixel στην eight-connected περιοχή του είναι επίσης υποψήφιο για την  $r$ . Έτσι, όταν ένας τέτοιος γείτονας δεν ανήκει ήδη σε άλλη περιοχή και ικανοποιεί μία συγκεκριμένη συνθήκη, ανατίθεται στην  $r$  και του δίνεται προτεραιότητα για να δεχθεί επίσκεψη όσο το δυνατόν πιο γρήγορα. Η εν λόγω συνθήκη είναι απλώς μία δυαδική απόφαση μεταξύ της τιμής μίας συνάρτησης ομοιότητας  $f_p$  κι ενός κατωφλίου (threshold). Ο Πίνακας 3.1 δίνει περιληπτικά τα βήματα του αλγορίθμου.

Ακολουθώντας όμως την συγκεκριμένη διαδικασία μπορεί κάποιος να οδηγηθεί σε ανεπιθύμητα αποτελέσματα. Για παράδειγμα, στη συνάρτηση  $f_p$  υπολογίζεται είτε η απόσταση μεταξύ των χρωματικών τιμών γειτονικών pixels, είτε η απόσταση μεταξύ των χρωματικών τιμών του τρέχοντος pixel και της γειτονικής περιοχής. Στην πρώτη περίπτωση είναι πιθανό ότι η τελικώς λαμβανόμενη περιοχή θα καλύπτει περισσότερες από μία διαχωρίσιμες χρωματικές περιοχές, όταν ο αλγόριθμος ακολουθεί μονοπάτια σταδιακών χρωματικών μεταβάσεων. Στη δεύτερη περίπτωση, είναι αναγκαίο να επανεκτιμάται το χρώμα της λαμβανόμενης χρωματικής περιοχής μετά από κάθε ανάθεση pixel σε αυτή. Με αυτόν τον τρόπο, η προτεραιότητα με την οποία επισκεπτόμαστε γειτονικά pixels επηρεάζει σημαντικά τα τελικά αποτελέσματα του αλγορίθμου άρα και τις λαμβανόμενες περιοχές. Και στις δύο περιπτώσεις, όμοιες εικόνες μπορούν να δώσουν πολύ διαφορετικές χρωματικές περιοχές, το οποίο είναι πρακτικά απαράδεκτο. Για να ξεπεράσουμε αυτές τις δυσκολίες, είναι συχνά προτιμητέο να μειώσουμε αρκετά τη χρωματική παλέτα πριν από την κατάτμηση, αντί του να χρησιμοποιήσουμε όλα τα διαθέσιμα χρώματα στην εικόνα. Μετά από κατάλληλο κβαντισμό, η συνάρτηση ομοιότητας μπορεί να επιτρέψει

**Πίνακας 3.2** Εξαγωγή χαρακτηριστικών και κατασκευή του γράφου γειτνίασης

- 
1. Θεώρησε ότι το τρέχον pixel  $p_0$  έχει ανατεθεί στην περιοχή  $r_0$ .
  2. Εάν η περιοχή  $r_0$  κατασκευάστηκε μόλις, θέσε το χρώμα της και το εμβαδό της στο χρώμα και το εμβαδό του  $p_0$ . Αλλιώς, προσμέτρησε το εμβαδό του  $p_0$  στο συνολικό εμβαδό της  $r_0$ .
  3. Εάν η eight-connected περιοχή του  $p_0$  ξεπερνάει τα όρια της εικόνας, τότε θεώρησε ότι η  $r_0$  είναι συνδεδεμένη (και) με μία υποτιθέμενη χρωματική περιοχή  $r_o$ .
  4. Για κάθε pixel  $p_1$  που έχει ήδη ανατεθεί στην  $r_1$  και βρίσκεται μέσα στην eight-connected περιοχή του  $p_0$ , συνδέσε την  $r_0$  και την  $r_1$  και αύξησε κατά ένα τον αριθμό των γειτόνων τους, εάν δεν είναι ήδη συνδεδεμένες. Μέτρησε τα  $p_0$  και  $p_1$  στο σύνορο μεταξύ των  $r_0$  και  $r_1$ .
  5. Αφού ο γράφος κατασκευαστεί, εάν η περιοχή  $r_1$  συνδέεται μόνο με την  $r_0$ , ενημέρωσε τους τύπους διασύνδεσής τους σε περιέχεται και περιέχει αντίστοιχα.
- 

μόνο σε pixels του ιδίου χρώματος να συμμετέχουν στην ίδια περιοχή. Το χρώμα κάθε περιοχής είναι τότε ταυτόσημο με το χρώμα όλων των pixels της. Αυτή η στρατηγική μπορεί να οδηγήσει σε υπερ-κατάτμηση (oversegmentation), όμως μας διασφαλίζει από τα ανεπιθύμητα αποτελέσματα που αναφέρθηκαν. Η υπερ-κατάτμηση μπορεί να αντιμετωπιστεί εκ των υστέρων βάσει του γράφου γειτνίασης, όμως θα εξηγήσουμε στη συνέχεια.

Εάν έχουμε ήδη διαθέσιμο ένα συγκεκριμένο σύστημα ανάκτησης εικόνων που να χρησιμοποιεί μία μετρική ομοιότητας συνολικού χρώματος (που να βασίζεται για παράδειγμα σε χρωματικά ιστογράμματα), ο αριθμός των χρωματικών σταθμών μπορεί να τεθεί ίσος με αυτόν που χρησιμοποιείται στο κριτήριο συνολικού χρώματος. Η κατάτμηση της εικόνας μπορεί να πραγματοποιηθεί σε οποιονδήποτε 3Δ χρωματικό χώρο, όμως ο  $U^*V^*W^*$  (ή ακόμη ο  $L^*a^*b^*$ ) είναι περισσότερο κατάλληλος για τέτοιες εφαρμογές [68, 172]. Άλλοι χρωματικοί χώροι έχουν επίσης αναφερθεί να δίνουν ικανοποιητικά αποτελέσματα για όμοιους σκοπούς [154, 48]. Στην γενική περίπτωση,  $n_U$ ,  $n_V$  και  $n_W$  χρώματα χρησιμοποιούνται οδηγώντας σε συνολικά  $\mathcal{N} = n_U n_V n_W$  χρωματικά επίπεδα.

Εναλλακτικά, οποιαδήποτε τεχνική κατάτμησης χρώματος μπορεί να χρησιμοποιηθεί (βλ. για παράδειγμα [179, 124, 82]) σε βάρος του χρόνου εκτέλεσης ή της ακρίβειας, το οποίο εξαρτάται από την επιθυμητή λειτουργία του συγκεκριμένου συστήματος. Για παράδειγμα, στην περίπτωση ενός on-line συστήματος στο διαδίκτυο, όπου ο χρήστης έχει τη δυνατότητα να υποβάλει μία εικόνα και να πραγματοποιήσει αναζητήσεις με οπτικό παράδειγμα (querying by visual example), οι επαναληπτικές τεχνικές (ή ακόμη και απλούστερες αυτών) είναι προτιμότερες. Αντιθέτως, όταν δεν υπάρχει απαίτηση για εξαγωγή του διανύσματος χαρακτηριστικών σε σχεδόν πραγματικό χρόνο, μία περισσότερο ακριβής και δαπανηρή τεχνική θα μπορούσε να

**Πίνακας 3.3** Συμβολισμός των χαρακτηριστικών που συμπεριλαμβάνονται στο γράφο γειτνίασης

$G$	γράφος γειτνίασης
$G.v$	σύνολο κόμβων γράφου (=περιοχών)
$G.r[i] \equiv G.r_i$	κόμβος γράφου (=περιοχή) $i$
$G.r[i] \rightarrow c$	χρώμα περιοχής
$G.r[i] \rightarrow a$	εμβαδό περιοχής
$G.r[i] \rightarrow n$	πλήθος γειτόνων περιοχής
$G.c[i, j]$	ακμή του γράφου που συνδέει τους κόμβους $i, j$
$G.c[i, j] \rightarrow t$	είδος σύνδεσης
$G.c[i, j] \rightarrow b$	μήκος συνόρου

χρησιμοποιηθεί, όπως για παράδειγμα αλγόριθμοι που βασίζονται σε χρωματικές περιοχές που μεγαλώνουν ανεξάρτητα (independently growing color regions).

Η απλή στρατηγική της πλημμύρας που χρησιμοποιείται εδώ για κατάτμηση χρώματος έχει ακόμη ένα πλεονέκτημα στην υλοποίησή μας, πέρα από την ακριβή και σχετικά ταχεία κατάτμηση. Εφόσον όλα τα pixels που ανήκουν σε μία συγκεκριμένη περιοχή  $r_0$  θα τα επισκεφθούμε πριν ξεκινήσουμε να επεξεργαζόμαστε μία άλλη περιοχή  $r_1$ , είναι δυνατόν να κατασκευάσουμε το γράφο γειτνίασης των περιοχών με παράλληλο και απλό τρόπο. Την ίδια στιγμή, ένας αριθμός από *χαρακτηριστικά της περιοχής* (region attributes) εξάγονται παράλληλα, τα οποία με τη σειρά τους θα χρησιμοποιηθούν στους ελέγχους ομοιότητας μεταξύ διαφορετικών εικόνων.

### 3.4 Αναπαράσταση με χρήση γράφου γειτνίασης

Ο γράφος γειτνίασης, ως ένας απλός γράφος, είναι μία συλλογή από κόμβους και ακμές. Στην περίπτωση μας, οι κόμβοι του γράφου υποδηλώνουν τις χρωματικές περιοχές και οι ακμές συνδέουν κόμβους που αντιστοιχούν σε γειτονικές περιοχές. Η χρήση τέτοιων γράφων στη βιβλιογραφία είναι μάλλον συνήθης για μία μεγάλη ποικιλία από εφαρμογές. Η κατασκευή ενός απλού γράφου γειτνίασης είναι σ' αυτό το σενάριο αυτόματη, αφού αρκεί να προστεθεί ένας καινούριος κόμβος όταν αρχικοποιούμε μία νέα περιοχή  $r_1$  και μία ακμή που να συνδέει τις  $r_1$  και  $r_0$  όταν ένα pixel της  $r_0$  βρεθεί στη γειτονιά ενός pixel της  $r_1$ . Ο αλγόριθμος του Πίνακα 3.1 μπορεί να τροποποιηθεί σύμφωνα με τα παραπάνω χωρίς στην πραγματικότητα ουσιαστικό υπολογιστικό κόστος. Την ίδια στιγμή όμως, θα πρέπει να δει κανείς ότι ένας τέτοιος γράφος γειτνίασης δίνει φτωχή πληροφορία για σύγκριση εικόνων ως προς το χαρακτηριστικό κατανομής χρώματος. Γι' αυτόν το σκοπό, στην προτεινόμενη προσέγγιση οι παρακάτω ιδιότητες των χρωματικών περιοχών εξάγονται παράλληλα με τη διαδικασία κατάτμησης:

(1) *χρώμα περιοχής*: το χρώμα της κάθε εξαγόμενης περιοχής ως το χρώμα των pixels της,

(2) *εμβαδό περιοχής*: το εμβαδό που καλύπτεται από την περιοχή ως το ποσοστό των pixels που της ανήκουν προς το σύνολο των pixels της εικόνας,

(3) *γείτονες περιοχής*: ο αριθμός των περιοχών που είναι γειτονικές στη συγκεκριμένη περιοχή,

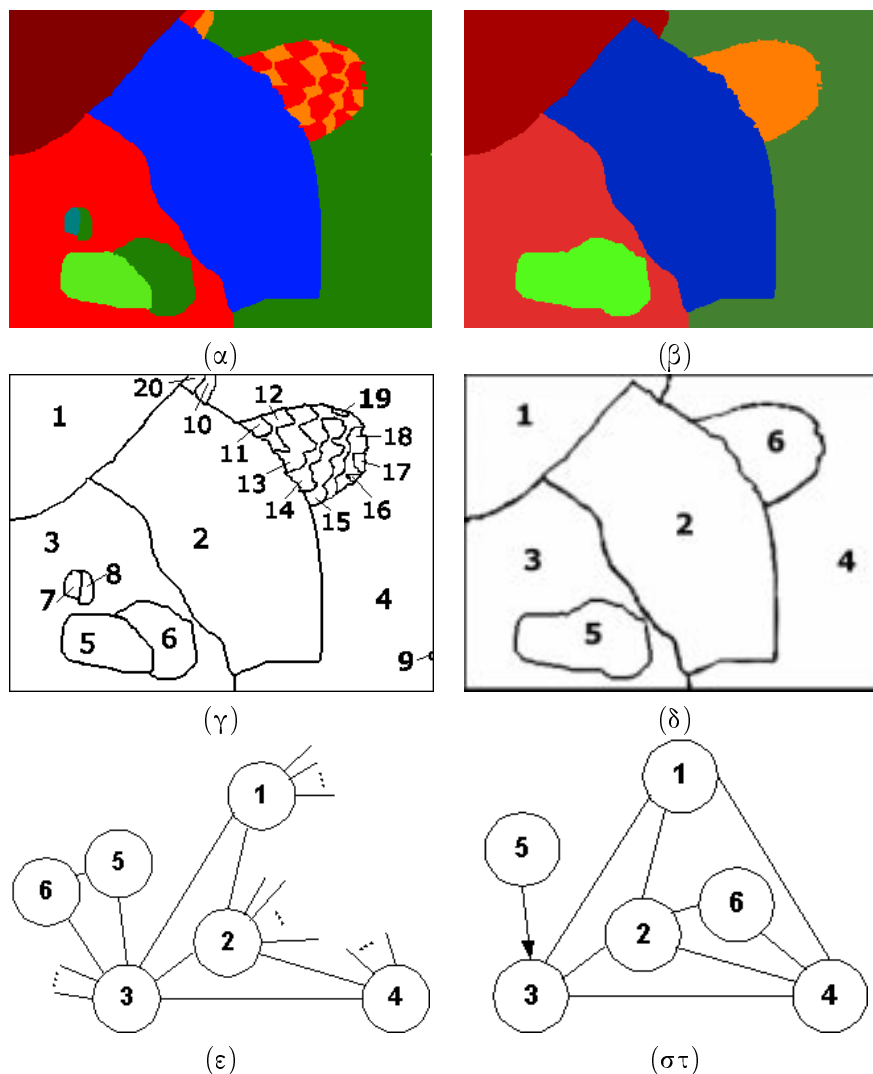
(4) *μήκος συνόρου*: ο αριθμός των pixels που βρίσκονται στο σύνορο μεταξύ δύο γειτονικών περιοχών,

(5) *είδος σύνδεσης*: το είδος της σύνδεσης μεταξύ δύο περιοχών  $r_0$  και  $r_1$  είναι ένα σύνολο από τέσσερις πιθανές τιμές - καμία όταν δεν υπάρχει σύνδεση, απλή όταν οι περιοχές συνδέονται απλώς, περιέχει όταν η  $r_0$  περιέχει την  $r_1$  και περιέχεται όταν η  $r_0$  περιέχεται στην  $r_1$ .

Η εξαγωγή των χαρακτηριστικών (1)-(4) μπορεί να ενσωματωθεί στη διαδικασία της πλημύρας χωρίς να αυξήσει σημαντικά το κόστος εκτέλεσης. Αντίθετα, για το χαρακτηριστικό (5), δεν μπορεί κανείς να αποφασίσει a priori τον τύπο σύνδεσης για όλες τις ακμές του γράφου. Για το λόγο αυτό, το χαρακτηριστικό (5) υπολογίζεται σε ένα επόμενο ταχύ βήμα διατρέχοντας το γράφο (και όχι την ίδια την εικόνα). Η κάθε περιοχή  $r_0$  που συνδέεται μόνο με μία άλλη περιοχή  $r_1$  είναι πολύ πιθανόν να περιέχεται στην  $r_1$ . Για να βοηθήσουμε στη λήψη της απόφασης αυτής, ο χώρος εκτός της εικόνας θεωρείται ως μία χρωματική περιοχή  $r_o$ . Όταν η  $r_0$  είναι γειτονική και με την  $r_1$  και με την  $r_o$  τότε απλώς συνδέεται με την  $r_1$ , ενώ, όταν είναι γειτονική μόνο με την  $r_1$  τότε περιέχεται σε αυτήν.

Ο αλγόριθμος εξαγωγής των χαρακτηριστικών δίνεται στον Πίνακα 3.2.

Σ' αυτό το σημείο, είναι σαφές ότι τόσο ο γράφος γειτνίασης όσο και τα παραπάνω χαρακτηριστικά των κόμβων και των ακμών του έχουν εξαχθεί. Έτσι, μπορεί κανείς να χρησιμοποιήσει μόνο τον εξαγόμενο γράφο για περαιτέρω ανάλυση και/ή σύγκριση. Σ' αυτό το πλαίσιο, ο γράφος καθεαυτός μπορεί να θεωρηθεί ως ένα διάνυσμα χαρακτηριστικών της συγκεκριμένης εικόνας ως προς το χαρακτηριστικό κατανομής χρώματος. Ταυτόχρονα, μπορεί κανείς να παρατηρήσει ότι ο λαμβανόμενος γράφος μπορεί να είναι ακατάλληλος για σύγκριση εικόνων, αφού οι αντίστοιχες εικόνες μπορεί να είναι υπερ-κατατμημένες λόγω χρωματικού θορύβου της εικόνας και/ή ασήμαντων (ελάχιστων) χρωματικών αντικειμένων. Όπως θα δούμε στην επόμενη ενότητα, η συγκεκριμένη αναπαράσταση του γράφου επιτρέπει στη δραματική και αποδοτική μείωσή του, υπό ορισμένους κανόνες, ώστε η υπερ-κατάτμηση να ξεπεραστεί



**Σχήμα 3.2** Παράδειγμα εξαγωγής και μείωσης του γράφου γειτνίασης: (α) τεχνητά παρηγμένη εικόνα με (γ) τη χρωματική κατάταξη της και (ε) το γράφο γειτνίασης αντίστοιχα, (β) η εικόνα και (δ) η αντίστοιχη χρωματική κατάταξη της που προήλθαν από τον μειωμένο γράφο γειτνίασης (στ).

και οι χρωματικές περιοχές να είναι όσο το δυνατόν πιο κοντά στις παρατηρούμενες με το μάτι.

### 3.5 Ελάττωση του γράφου γειτνίασης

Όπως περιγράψαμε στην προηγούμενη ενότητα, ο λαμβανόμενος γράφος αποτελείται τόσο από την πληροφορία γειτνίασης των περιοχών, όσο και από ποιοτική και ποσοτική πληροφορία για τα εξαγόμενα χαρακτηριστικά (1)-(5). Διαισθητικά θα λέγαμε ότι κάθε κόμβος ή ακμή του γράφου είναι πρακτικά μία δομή που περιέχει ένα συγκεκριμένο αριθμό στοιχείων. Έτσι, οι κόμβοι περιέχουν πληροφορία για τα *εμβάδο περιοχής*, *χρώμα περιοχής* και *γείτονες περιοχής*

ενώ οι ακμές για τα μήκος συνόρου και τύπο σύνδεσης. Όπως μπορεί να δει κανείς, περαιτέρω στοιχεία θα μπορούσαν να εξαχθούν και να χρησιμοποιηθούν στον προτεινόμενο γράφο χωρίς ουσιαστικά να αλλαχθεί κανένα από τα βασικά βήματα της μεθόδου. Ο Πίνακας 3.3 συνοψίζει τα χαρακτηριστικά που χρησιμοποιούνται και δίνει τον αντίστοιχο συμβολισμό.

Θεωρείστε για παράδειγμα την τεχνητά παρηγμένη εικόνα του Σχήματος 3.2(α). Εκτελώντας το βήμα της πλημμύρας στη διαδικασία κατάτμησης της Ενότητας 3.3, μπορεί κανείς να λάβει τα αποτελέσματα κατάτμησης του Σχήματος 3.2(γ), τα οποία αντιστοιχούν απευθείας στον μάλλον πολύπλοκο γράφο του Σχήματος 3.2(ε) (στο συγκεκριμένο σχήμα μόνο ένας ενδεικτικός υπογράφος έχει περιληφθεί). Πρακτικά, η λαμβανόμενη δομή γράφου μαζί με τα εξαγόμενα χαρακτηριστικά του Πίνακα 3.3 φυλάσσονται αποδοτικά σε μία συνδεδεμένη λίστα, ώστε η ίδια η εικόνα να μην είναι αναγκαία με κανέναν τρόπο στους υπολογισμούς.

Συγκρίνοντας το γράφο γειτνίασης με την πρωτότυπη εικόνα, μπορεί να σημειωθεί ότι, ακόμη και για την εικόνα αυτή όπου οι χρωματικές περιοχές είναι 'βολικά' διαχωρίσιμες, η πληροφορία που φυλάσσεται για σύγκριση με άλλες εικόνες είναι μάλλον περισσότερη από την απολύτως αναγκαία. Ως προς την κατανομή χρώματος, με μία πρώτη ματιά κανείς θα συγκρατούσε μία καφέ, μία μπλε, μία πράσινη και μία κόκκινη περιοχή, μαζί με μία αόριστη εντύπωση για τη διασυνδεσιμότητά τους - μάλλον λίγο πράσινο μέσα στην κόκκινη περιοχή και κάποιο πορτοκαλί πάνω στο πράσινο-μπλε σύνορο. Με άλλα λόγια, η συγκεκριμένη εικόνα είναι υπερ-κατατμημένη, γεγονός το οποίο μπορεί με τη σειρά του να οδηγήσει τόσο σε αυξημένη πολυπλοκότητα στη σύγκριση των εικόνων όσο και σε λανθασμένα αποτελέσματα συγκρίσεων λόγω μεγάλου επιπέδου λεπτομέρειας. Αυτά τα προβλήματα μεγεθύνονται όταν δουλεύουμε με φωτογραφίες του πραγματικού κόσμου όπου ο χρωματικός θόρυβος είναι μάλλον η συνήθης περίπτωση παρά η εξαίρεση.

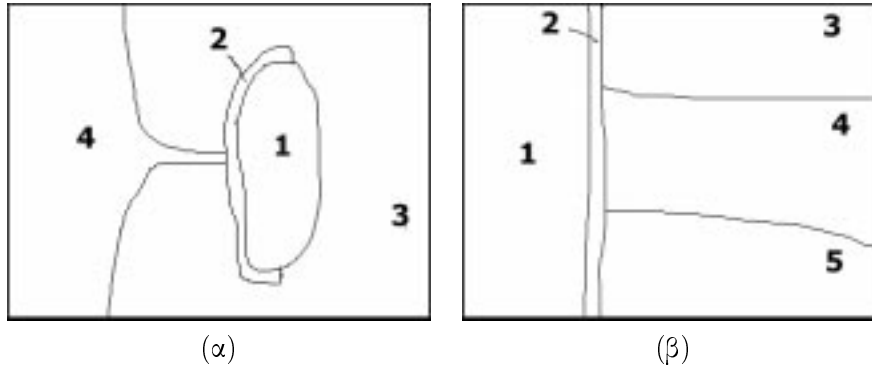
Με το παραπάνω σκεπτικό, μία τεχνική συνένωσης περιοχών (ισοδύναμα, μία τεχνική μείωσης του γράφου) εφαρμόζεται ως δεύτερο βήμα. Στην προσέγγισή μας, η επιλογή των περιοχών προς συνένωση βασίζεται στο σχετικό τους μέγεθος. Για παράδειγμα, μία περιοχή μπορεί να θεωρηθεί ως 'θόρυβος εικόνας' ή, καλύτερα, 'ασήμαντη λεπτομέρεια' όταν το μέγεθός της δεν ξεπερνάει ένα συγκεκριμένο κατώφλι  $\mathcal{A}$  σε σχέση με το συνολικό εμβαδό της εικόνας (για παράδειγμα  $\mathcal{A} \equiv 5\%$ ). Οι μικρές περιοχές εντοπίζονται και ανατίθενται σε όμοιες και μεγαλύτερες γειτονικές τους. Με αυτό το σκεπτικό, μία μετρική / συνάρτηση ομοιότητας  $f_r$  χρησιμοποιείται για να αποφανθούμε ποιος είναι ο καλύτερος υποψήφιος γείτονας. Στο σενάριό μας, όπως θα εξηγηθεί στη συνέχεια, η  $f_r$  βασίζεται στα εξηγμένα χαρακτηριστικά του χρώματος περιοχής και εμβαδού περιοχής. Η διαδικασία συνένωσης, όμως, θα πρέπει να λαμβάνει υπόψη της κάποιους παράγοντες ώστε η μείωση του γράφου να είναι επιτυχής:

Πίνακας 3.4 Μείωση του γράφου γειτνίασης

- 
1. Θεώρησε την τρέχουσα περιοχή  $r_0$  όπως λαμβάνεται από τον FIFO καταχωρητή  $R$ .
  2. Εάν είναι  $(G.r[0] \rightarrow a > \mathcal{A})$  πήγαινε στο βήμα 9.
  3. Πάρε τους γείτονες της περιοχής  $r_1 \cdots r_n$ .
  4. Εάν είναι  $(G.r[0] \rightarrow a > G.r[i] \rightarrow a)$  για κάθε γείτονα  $i$ , βάλε την  $r_0$  στην είσοδο του  $R$  και πήγαινε στο βήμα 9.
  5. Εάν η  $r_0$  έχει ένα μοναδικό γείτονα  $r_i$ , πήγαινε στο βήμα 8.
  6. Βρες τις χρωματικές αποστάσεις βάσει της  $f_r(G.r[0] \rightarrow c, G.r[i] \rightarrow c)$  για όλους τους γείτονες και διάλεξε ανάμεσά τους αυτούς που επιστρέφουν τη μέγιστη τιμή.
  7. Εάν δεν υπάρχει μία μοναδική τέτοια περιοχή  $i$ , διάλεξε το  $i$  ούτως ώστε  $(G.c[0, i] \rightarrow b > G.c[0, j] \rightarrow b)$  για όλους τους άλλους γείτονες  $r_j$ .
  8. Ανέθεσε την  $r_0$  στην  $r_i$ .
  9. Πάρε την επόμενη έξοδο του καταχωρητή και πήγαινε στο βήμα 2.
- 

- Η διαδικασία συνένωσης θα πρέπει να επιστρέφει τα ίδια αποτελέσματα ανεξάρτητα από τη διάταξη των χρωματικών περιοχών στο γράφο. Έτσι, οι μετασχηματισμένες υπό affine (λόγω π.χ. της κίνησης της κάμερας) συγγενείς εικόνες της διαθέσιμης εικόνας θα πρέπει να δίνουν όμοιο γράφο γειτνίασης μετά τη μείωση.
- Το χρώμα κάθε περιοχής ανανεώνεται συνεχώς ώστε τελικά να είναι ένας συνδυασμός από τα χρώματα των μικρότερων περιοχών που περιέχει.
- Το εμβαδό μίας συγκεκριμένης περιοχής μεγαλώνει όταν μικρότερες περιοχές γειτνιάζουν με αυτή.

Οι παραπάνω θεωρήσεις μπορούν να βρίσκονται σε ασυμφωνία μεταξύ τους. Έστω για παράδειγμα ότι η περιοχή  $r_5$  γειτνιάζει με τις  $r_3$  και  $r_6$ , με τις  $r_5, r_6$  να είναι ‘μικρές’ περιοχές, οι οποίες με τη σειρά τους όταν ενώνονται να σχηματίζουν μία ‘μεγάλη’ περιοχή. Ένας μη αποδοτικός αλγόριθμος, θα ένωνε την  $r_5$  με την  $r_3$  (όταν την επεξεργαζόταν), με αποτέλεσμα όλες οι περιοχές να ανατίθονταν στην  $r_3$ . Αυτό θα μπορούσε να συμβεί στο Σχήμα 3.2(α), εάν οι περιοχές  $r_5, r_6$  ανατίθονταν στην  $r_3$ , αντί να δημιουργήσουν μία χωριστή ‘μεγάλη’ περιοχή. Αυτό είναι ένα σύννηθες πρόβλημα αρκετών αλγορίθμων κατάτμησης χρώματος. Την ίδια στιγμή, η περιοχή  $r_{10}$  θα μπορούσε εσφαλμένα να ανατεθεί στην  $r_4$  αντί στην  $r_{20}$ , εάν δίνουμε προτεραιότητα στο εμβαδό κατά τη σύγκριση. Από την άλλη μεριά, μία μικρή περιοχή περικυκλωμένη από ακόμη μικρότερες θα μπορούσε να ενωθεί με κάποια από αυτές, προσπαθώντας απεγνωσμένα να μεγαλώσει και να επιβιώσει. Με τον τρόπο αυτό, το χρώμα της θα ανανεωνόταν συνεχώς, οδηγώντας πιθανότατα την προκύπτουσα περιοχή σε συνένωση με μία ακόμα μικρότερη και τελικώς οδηγώντας στην επιβίωση περιοχές οι οποίες θα έπρεπε να



**Σχήμα 3.3** Δύο παραδείγματα (α,β) που επιδεικνύουν τη σημασία του χαρακτηριστικού όριο σύνδεσης: η περιοχή  $r_2$  θα ανατεθεί στην  $r_1$  αφού μοιράζονται το μεγαλύτερο όριο σύνδεσης.

εξαφανιστούν. Αντίθετα, εάν αποφασίζαμε τη μη ανανέωση του χρώματος των λαμβανόμενων περιοχών από συνενώσεις, για την κόκκινη-πορτοκαλί περιοχή (σκακίερα)  $r_{11} - r_{19}$  του Σχήματος 3.2(γ) θα λαμβάναμε τελικώς μία μεγάλη πορτοκαλί ή κόκκινη περιοχή, βάσει της σειράς με την οποία επισκεπτόμαστε τις υποπεριοχές της.

Για να ξεπεράσουμε τις παραπάνω πιθανότητες, στην προτεινόμενη στρατηγική η διαδικασία συνένωσης περιοχών υπακούει σε μερικούς προ-αποφασισμένους κανόνες, όπου όλα τα εξηγμένα χαρακτηριστικά διαδραματίζουν σημαντικό ρόλο. Αρχικά, όλες οι περιοχές εισάγονται σε έναν καταχωρητή FIFO με τυχαία σειρά. Κάθε περιοχή  $r_1$  που λαμβάνεται από την έξοδο του καταχωρητή είναι υποψήφια για συνένωση όταν το εμβαδό της είναι μικρότερο από το κατώφλι  $\mathcal{A}$ . Ταυτόχρονα, η συνένωσή της μπορεί να πραγματοποιηθεί μόνο όταν υπάρχει ένας γείτονάς της  $r_0$  με  $G.r[0] \rightarrow a > G.r[1] \rightarrow a$ . Όταν η ‘μικρή’ περιοχή απλώς περιέχεται σε μία γειτονική της, ανατίθεται αυτομάτως στην περιέχουσα περιοχή. Όταν υπάρχουν περισσότερες από μία ‘μεγάλες’ γειτονικές περιοχές, επιλέγεται αυτή με το κοντινότερο χρώμα περιοχής. Ανάμεσα σε γείτονες με ίσες χρωματικές αποστάσεις, επιλέγουμε τον γείτονα με τον οποίο η περιοχή μοιράζεται το μακρύτερο μήκος συνόρου. Το τελευταίο μπορεί να δικαιολογηθεί διαισθητικά με επισκόπηση του Σχήματος 3.3. Υποθέτοντας ότι και στα δύο Σχήματα 3.3(α) και 3.3(β), έχουμε  $G.r[2] \rightarrow a < \mathcal{A}$  και  $f_r(G.r[2] \rightarrow c, G.r[1] \rightarrow c) = f_r(G.r[2] \rightarrow c, G.r[4] \rightarrow c)$ , η  $r_2$  ανατίθεται στην  $r_1$  εφόσον  $G.c[2,1] \rightarrow b > G.c[2,4] \rightarrow b$ . Η διαδικασία αυτή συνοψίζεται στον Πίνακα 3.4. Πρακτικά, με βάση τα παραπάνω, η συνάρτηση  $f_r$  μπορεί να επεκταθεί σε μία συνολική απόφαση του τύπου ‘σε ποιον γείτονα θα πρέπει η περιοχή  $r_i$  να ανατεθεί’ και να καλείται αντίστοιχα στη μορφή  $f_r(r_i, G)$ , αφού στην προτεινόμενη ανάλυση όλα τα χαρακτηριστικά των γειτόνων συμμετέχουν στη διαδικασία σύγκρισης.

Το Σχήμα 3.2(στ) απεικονίζει τη μειωμένη εκδοχή του γράφου γειτνίασης του Σχήματος 3.2(ε). Τα Σχήματα 3.2(β) και 3.2(δ) αντιστοιχούν στο μειωμένο γράφο γειτνίασης και θα



πρέπει να συγκριθούν απευθείας με τα 3.2(α) και 3.2(γ) αντίστοιχα. Ο σύνδεσμος με κατευθυντικότητα (βέλος) που συνδέει τις  $r_5$  και  $r_3$  υποδηλώνει ότι η  $r_3$  περιέχει την  $r_5$ .

### 3.6 Ταυτοποίηση χαρακτηριστικών με χρήση μετρικών ομοιότητας

Ακολουθώντας τα βήματα της μεθοδολογίας που αναπτύχθηκε στις προηγούμενες ενότητες, μία αναπαράσταση σε μορφή γράφου εξάγεται για κάθε εικόνα εισόδου. Πρακτικά, ο γράφος αυτός αντιπροσωπεύει το εξαγόμενο διάνυσμα χαρακτηριστικών (feature vector) που αντιστοιχεί στη συγκεκριμένη εικόνα (τουλάχιστον, το τμήμα του διανύσματος χαρακτηριστικών που αντιστοιχεί στο χαρακτηριστικό της κατανομής χρώματος). Η ανάκτηση εικόνων πραγματοποιείται τότε βάσει των αντίστοιχων διανυσμάτων χαρακτηριστικών χρησιμοποιώντας κάποια συνάρτηση ομοιότητας.

Η συγκεκριμένη συνάρτηση ομοιότητας που χρησιμοποιείται για τη σύγκριση των γράφων γειτνίασης μπορεί να διαφέρει εξαρτώμενη με τα επιθυμητά αποτελέσματα. Με άλλα λόγια, η προτεινόμενη αναπαράσταση του γράφου δεν υπαγορεύει την επιλογή της συνάρτησης. Σ' αυτό το κεφάλαιο, υλοποιήσαμε μία συνάρτηση ομοιότητας δίνοντας 'βαθμούς ομοιότητας' βάσει των ακόλουθων κριτηρίων:

- ομοιότητα χρώματος περιοχών
- ομοιότητα εμβαδού περιοχών
- ομοιότητα συνδεσιμότητας περιοχών (αριθμός και τύπος ακμών)

Μπορεί να δει κανείς ότι τα παραπάνω κριτήρια είναι σε αντιστοιχία με τα εξηγμένα χαρακτηριστικά των περιοχών, αφού οι μετρήσεις ομοιότητας βασίζονται στα αντίστοιχα δεδομένα χαρακτηριστικά.

Στην παρούσα προσέγγιση, εισάγονται βασικά δύο μετρικές, συγκεκριμένα η μετρική κάλυψη και η μετρική συνδεσιμότητα, οι οποίες αναλύονται παρακάτω. Πρακτικά, η κάλυψη αντιπροσωπεύει το ποσοστό του γράφου (του υπογράφου) μίας αποθηκευμένης εικόνας που 'συμπίπτει' με το γράφο (ή κάποιον υπογράφο) της εικόνας εισόδου. Ως συμπληρωματική μετρική, η συνδεσιμότητα ποσοτικοποιεί τη δομική ομοιότητα μεταξύ όμοιων γράφων.

### 3.6.1 Η μετρική ‘κάλυψη’

Συγκρίνοντας τους γράφους γειτνίασης  $G, H$  των Σχημάτων 3.4(α) και 3.4(β) αντίστοιχα, εάν  $G.r[i] \rightarrow c \equiv H.r[i] \rightarrow c$  για κάθε περιοχή  $r_i$ , κάποιος μπορεί να πει ότι οι αντίστοιχες εικόνες είναι ταυτόσημες μεταξύ τους, πέρα από κάποιες τοπικές παραμορφώσεις ή κάποιους μετασχηματισμούς λόγω της κίνησης της κάμερας ή των αντικειμένων. Αυτό είναι ισοδύναμο με την ακόλουθη δήλωση: ‘οι χρωματικές περιοχές καλύπτουν το ίδιο ποσοστό της εικόνας’ στα δύο σχήματα. Εφόσον τέτοιες περιπτώσεις συμβαίνουν συχνά ακόμη και σε διαδοχικά καρέ μίας ακολουθίας βίντεο, θα ήταν μάλλον υπερβολή να λάβουμε υπόψη οποιεσδήποτε τέτοιες υπάρχουσες διαφορές. Αν και η προσέγγισή μας επιτρέπει τέτοια αντιμετώπιση (μετρώντας για παράδειγμα τις διαφορές στα αντίστοιχα *εμβαδά περιοχής* ή προσθέτοντας χαρακτηριστικά επιμήκυνσης ή εκκεντρότητας), επιλέγουμε να επικεντρωθούμε σε μία μετρική κάλυψης χωρίς να συμπεριλάβουμε χαρακτηριστικά που σχετίζονται με το σχήμα και θα έβρισκαν αρκετά ανόμοιες τέτοιες εικόνες. Μ’ αυτή τη λογική, στο συγκεκριμένο παράδειγμα, έχουμε 100% σύμπτωση ως προς την κάλυψη. Όμοια, εάν η σκιασμένη περιοχή του Σχήματος 3.4(γ) αντιστοιχεί ως προς μία χρωματική αναλογία στο Σχήμα 3.4(α), τότε επίσης έχουμε 100% σύμπτωση ως προς την κάλυψη, αφού ο γράφος γειτνίασης του Σχήματος 3.4(α) αποτελεί έναν υπογράφο αυτού του Σχήματος 3.4(γ). Από την άλλη μεριά, παρατηρώντας τα Σχήματα 3.4(α) και 3.4(δ), με  $G.r[2] \rightarrow c$ ,  $G.r[3] \rightarrow c$ ,  $G.r[5] \rightarrow c$  να αντιστοιχούν ως προς το χρώμα στα  $H.r[2] \rightarrow c$ ,  $H.r[3] \rightarrow c$ ,  $H.r[1] \rightarrow c$  αντίστοιχα, μόνο ένας υπογράφος του  $G$  υπάρχει στο  $H$ , οπότε η κάλυψη του  $G$  στο  $H$  ισοδυναμεί με  $(G.r[2] \rightarrow a) + (G.r[3] \rightarrow a) + (G.r[5] \rightarrow a) \approx 60\%$ .

Γενικά, ορίζουμε την κάλυψη του  $G$  στο  $H$  ως

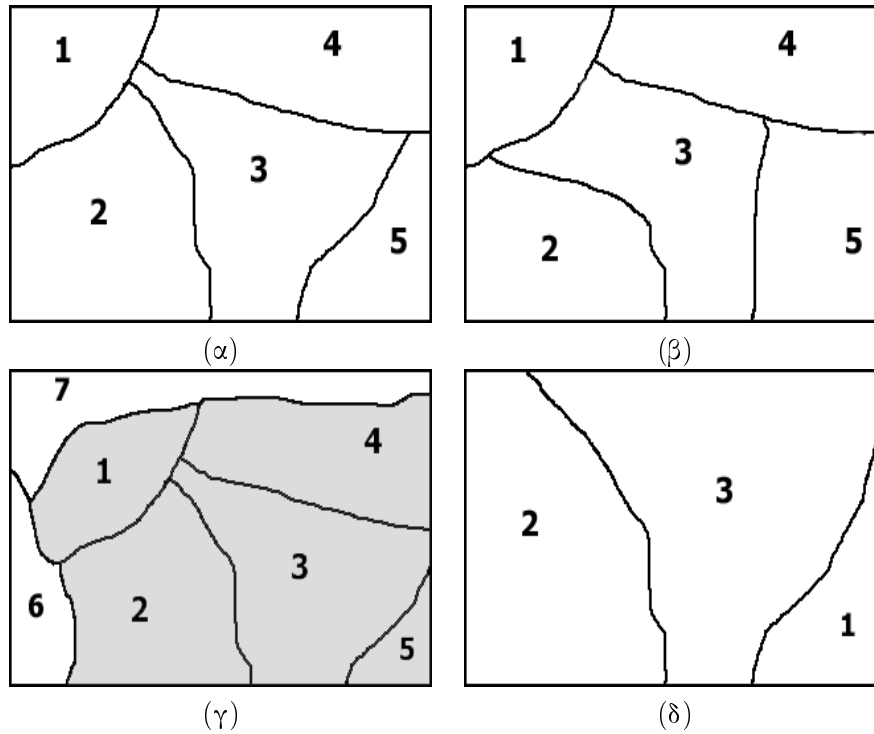
$$C(G, H) = \sum_{\forall i \text{ αντιστ.}} G.r[i] \rightarrow a \quad (3-1)$$

όπου  $G$  είναι ο γράφος γειτνίασης και  $r_i$  όλες οι περιοχές για τις οποίες κατάλληλες αντίστοιχες έχουν βρεθεί στο  $H$ .

Στην ίδια λογική, ας υποθέσουμε ότι το Σχήμα 3.4(α) συγκρινόταν με ένα άλλο, το οποίο ήταν διαφορετικό μόνο κατά το χρώμα της περιοχής  $r_2$  και ότι  $G.r[2] \rightarrow a \equiv 30\%$ . Σε αντιστοιχία με τα παραπάνω  $C(G, H) = 70\%$ . Όμως, ουσιαστικά η κάλυψη θα μπορούσε να επαναοριστεί ώστε να συμπεριλάβει την ομοιότητα ως προς χρώμα περιοχής ως

$$C(G, H) = \sum_i \Delta c_i \cdot G.r[i] \rightarrow a \quad (3-2)$$

όπου το  $\Delta c_i$  εισάγει τη διαφορά χρώματος μεταξύ της περιοχής  $G.r[i]$  και της αντίστοιχής της στον  $H$ , ως προς το λόγο της απόστασης των αντίστοιχων χρωματικών σταθμών  $\Delta b_i$  δια τη



**Σχήμα 3.4** Τρία παραδείγματα που επιδεικνύουν τη σημασία της μετρικής κάλυψη: η εικόνα εισόδου (α) παρουσιάζει 100% κάλυψη στην (β), 100% στην (γ) ως υποσύνολό της, 85% στην (δ) ως υπερσύνολό της.

μέγιστη δυνατή χρωματική απόσταση  $\Delta b_{\max}$ . Έτσι,  $\Delta c_i = \frac{\Delta b_{\max} - \Delta b_i}{\Delta b_{\max}}$ . Με αυτή τη λογική, όταν αντίστοιχες περιοχές είναι πολύ διαφορετικές ή όταν δεν μπορεί να βρεθεί αντίστοιχη για μία συγκεκριμένη περιοχή, το  $\Delta c_i$  πλησιάζει το μηδέν και το αντίστοιχο ζεύγος περιοχών δεν συνεισφέρει στο  $\mathcal{C}$ .

Πρακτικά, για να αποφύγουμε αναζητήσεις που καταναλώνουν πολύ χρόνο και την ίδια στιγμή να απορρίψουμε απευθείας ζεύγη από ανόμοιες περιοχές, το  $\Delta b_{\max}$  μπορεί να τεθεί αρκετά μικρότερο από τη μέγιστη δυνατή χρωματική απόσταση και ταυτόχρονα μόνο ζεύγη που δίνουν  $\Delta b_i \leq \Delta b_{\max}$  να γίνονται δεκτά. Με δεδομένο τον ορισμό της εξίσωσης (3-2) και υποθέτοντας ότι στο προηγούμενο παράδειγμα  $\Delta c_2 = 0.5$ , λαμβάνουμε  $\mathcal{C}(G, H) = 85\%$ .

Εκτός από τη μετρική ομοιότητας, που υλοποιήθηκε εδώ μέσω της μετρικής της κάλυψης, μία διαδικασία αναζήτησης πρέπει να οριστεί ώστε όλοι οι κόμβοι του γράφου να επεξεργαστούν και το καλύτερο (μεγαλύτερο) ποσοστό σύμπτωσης να εντοπιστεί. Σημαντικότερο δε είναι η διαδικασία που θα υιοθετηθεί να δίνει ταυτόσημα αποτελέσματα, ανεξάρτητα του κόμβου έναρξης. Όμως, αυτό μπορεί να πραγματοποιηθεί μόνο όταν η αναζήτηση αρχικοποιείται από κάθε κόμβο όμοιων περιοχών και επιλέγεται η καλύτερη σύμπτωση. Αυτό συμβαίνει βασικά λόγω του γεγονότος ότι όμοια πρότυπα μπορούν να υπάρχουν σε μία συγκεκριμένη εικόνα, ενώ ταυτόχρονα πολύπλοκοι γράφοι μπορούν να αποτελούνται από επικαλυπτόμενους

**Πίνακας 3.5** Η διαδικασία σύγκρισης διανυσμάτων χαρακτηριστικών

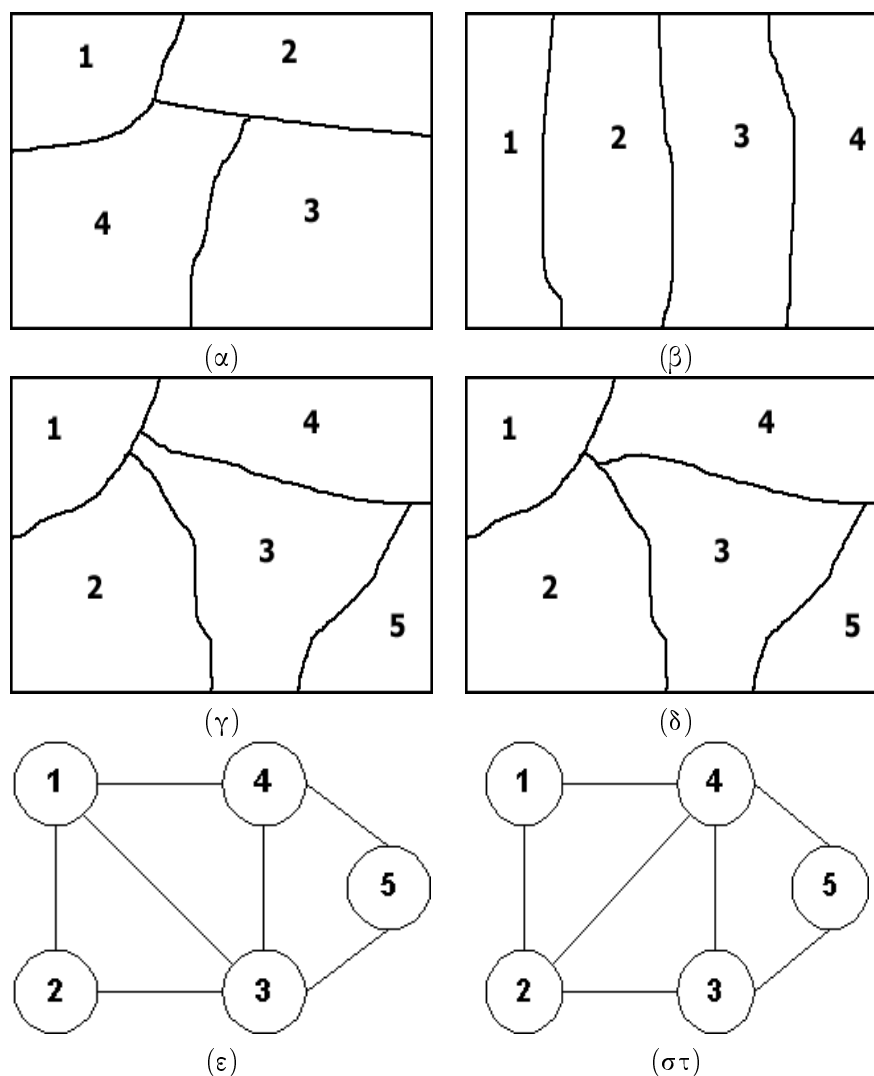
- 
1. Θεώρησε το τρέχον ζεύγος κόμβων  $(i, j)$  ( $r_i \in G, r_j \in H$ ) που έρχονται από την έξοδο του καταχωρητή FIFO  $R$  που περιέχει όλα τα δυνατά ζεύγη κόμβων. Δημιούργησε έναν άδειο καταχωρητή FIFO  $S$  για αυτό το ζευγάρι.
  2. Υπολόγισε την τιμή  $f_t(G.r[i], H.r[j])$ . Εάν  $f_t < 0$  πήγαινε στο βήμα 6. Αλλιώς, συμπεριέλαβε τη συνεισφορά του ζεύγους κόμβων στη μετρική της κάλυψης  $C_{ij}$  όπως στην εξίσωση (3-2).
  3. Για τα  $G.r[i], H.r[j]$  υπολόγισε τα καλύτερα ζεύγη γειτόνων και βάλε τα στον καταχωρητή  $S$ .
  4. Συνέχισε με το επόμενο ζεύγος στον  $S$  από το βήμα 2 έως ότου ο  $S$  είναι άδειος.
  5. Όταν ο  $S$  αδειάσει, το  $C_{ij}$  είναι η συνολική μέτρηση κάλυψης ξεκινώντας από το συγκεκριμένο ζεύγος κόμβων  $(i, j)$ . Συνέκρινε το  $C_{ij}$  με το καλύτερο  $C$  που επιτεύχθηκε μέχρι τώρα.
  6. Πήγαινε στο βήμα 1 για το επόμενο ζευγάρι κόμβων εάν ο  $R$  δεν είναι άδειος.
- 

γράφους ιδίου μεγέθους. Μία συνάρτηση ομοιότητας ορίζεται ούτως ώστε να συγκρίνουμε περιοχές που ανήκουν σε διαφορετικούς γράφους. Μπορεί κανείς να δει ότι τέτοιες συγκρίσεις πραγματοποιούνται ενώ ακολουθούμε πάνω στο γράφο ένα μονοπάτι, δηλ. επισκεπτόμενοι μία σειρά από κόμβους χωρίς επαναλήψεις.

Στην υλοποίησή μας, δύο περιοχές θεωρούνται ανόμοιες ως προς το χρώμα όταν οι αντίστοιχες χρωματικές στάθμες διαφέρουν περισσότερο από μία απόσταση κατωφλίου  $\Delta b_{\max}$ . Στην περίπτωση αυτή, η αναζήτηση στο συγκεκριμένο μονοπάτι για την καλύτερη σύμπτωση σταματάει μόνο όταν η διαφορά στο εμβαδό ξεπερνάει ταυτόχρονα ένα χαμηλό κατώφλι  $\Delta a_{\max}$ . Η συνάρτηση  $f_t$  χρησιμοποιείται όταν ελέγχουμε για τα καλύτερα γειτονικά ζεύγη μίας συγκεκριμένης περιοχής, για να αποφασίσουμε πώς θα συνεχιστεί η διαδικασία σύγκρισης. Όμως, θα πρέπει να παρατηρηθεί ότι δύο γείτονες των αρχικών περιοχών είναι ανόμοιοι εξαρχής όταν οι τύποι σύνδεσης μεταξύ των αρχικών περιοχών και των γειτόνων τους διαφέρουν.

Η διαδικασία σύγκρισης συνοψίζεται στον Πίνακα 3.5.

Η μετρική κάλυψης όπως εισάγεται εδώ μπορεί να είναι αρκετή ως μετρική ομοιότητας. Ας παρατηρήσουμε όμως τα Σχήματα 3.5(α) και 3.5(β) όπου υποθέτουμε  $G.r[i] \rightarrow c \equiv H.r[i] \rightarrow c$  για κάθε περιοχή  $r_i$  των αντίστοιχων γράφων γειτνίασης  $G, H$ . Η διαδικασία σύγκρισης του Πίνακα 3.5 θα επέστρεφε μία  $C = 100\%$  σύμπτωση στο μονοπάτι  $r_1 \rightarrow r_2 \rightarrow r_3 \rightarrow r_4$ , αν και η κατανομή των όμοιων χρωματικών περιοχών επάνω στην εικόνα είναι διαφορετική.

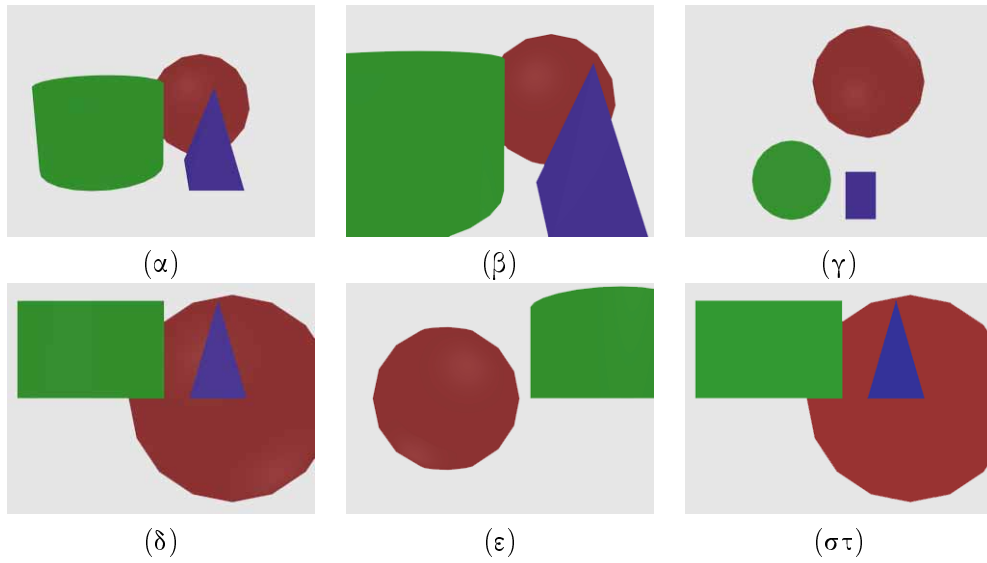


**Σχήμα 3.5** Παράδειγμα που επιδεικνύει περιορισμούς στη μετρική κάλυψη: (α,β) εικόνες που ενώ διαφέρουν, παρουσιάζουν πλήρη σύμπτωση στα πλαίσια της κάλυψης. Παράδειγμα που εξηγεί τον ορισμό της μετρικής συνδεσιμότητας: σχεδόν πανομοιότυπες εικόνες (γ,δ) με διαφορετικούς γράφους γειτνίασης (ε,στ).

### 3.6.2 Η μετρική ‘συνδεσιμότητα’

Για να εκμεταλλευτούμε την πληροφορία συνδεσιμότητας (connectivity) του γράφου παράλληλα με τα χαρακτηριστικά του τύπου σύνδεσης και του μήκους συνόρου, εφαρμόζεται ένα δεύτερο βήμα επεξεργασίας. Στην πραγματικότητα, το πρώτο βήμα της διαδικασίας σύγκρισης (Πίνακας 3.5) δεν δίνει μόνο την καλύτερη δυνατή μέτρηση της κάλυψης, αλλά και το μονοπάτι πάνω στο γράφο κατά το οποίο λαμβάνεται η μέτρηση αυτή. Έτσι, μπορεί κάποιος να εξάγει τους μεγαλύτερους υπογράφους  $G_M \subseteq G$ ,  $H_M \subseteq H$  κατά το μονοπάτι αυτό και να τους συγκρίνει ως προς τη συνδεσιμότητά τους.

Στο προηγούμενο παράδειγμα, λοιπόν, θα πρέπει να επιβληθεί κάποιο είδος ‘τιμωρίας’



**Σχήμα 3.6** Πειραματικά αποτελέσματα με μοντέλο μία σκηνή με πολλά αντικείμενα, (α-ε) διαφορετικές λήψεις της ίδιας σκηνής και (στ) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της εικόνας (δ).

όταν συγκρίνουμε τους γράφους γειτνίασης των Σχημάτων 3.5(α) και 3.5(β), εφόσον οι  $G.r[1], G.r[4]$  είναι συνδεδεμένες ενώ οι  $H.r[1], H.r[4]$  όχι. Η μετρική *συνδεσιμότητα* μπορεί τότε να οριστεί ως ο λόγος των κοινών ακμών  $e_{κοινές}$  των υπογράφων προς τις συνολικές ακμές  $e_{συνολικές}$

$$\mathcal{L}(G, H) \equiv \mathcal{L}(G_M, H_M) = \frac{e_{κοινές}}{e_{συνολικές}} \quad (3-3)$$

Ο τύπος σύνδεσης και οι γείτονες περιοχής παίζουν σημαντικό ρόλο σ'αυτά τα πλαίσια, αφού μόνο συνδέσεις ιδίου τύπου μπορούν να είναι συγκρίσιμες και ο αριθμός των γειτόνων κατά περιοχή μπορεί να δώσει μία γενική ιδέα της ισοδυναμίας συνδεσιμότητας (ή ακόμη και να επιταχύνει τη διαδικασία σύγκρισης).

Η χρήση του *μήκους συνόρου* ελαχιστοποιεί την πιθανότητα κακής εφαρμογής από διαφορές στη συνδεσιμότητα λόγω μικρών διαφοροποιήσεων στα σύνορα των περιοχών. Ας θεωρήσουμε για παράδειγμα τις εικόνες των Σχημάτων 3.5(γ) και 3.5(δ) και τους αντίστοιχους γράφους γειτνίασης των Σχημάτων 3.5(ε) και 3.5(στ). Αν και οι εικόνες είναι σχεδόν πανομοιότυπες, η μετρική *συνδεσιμότητα* δίνει  $\mathcal{L} = \frac{12}{14} \approx 85\%$ . Αυτό οφείλεται στη 'μικρή' διαφοροποίηση στη συνδεσιμότητα μεταξύ των  $r_1 - r_3$  και  $r_2 - r_4$ , η οποία πρακτικά μπορεί να οφείλεται σε συνήθεις αστοχίες της διαδικασίας κατάτμησης. Χρησιμοποιώντας το χαρακτηριστικό του *μήκους συνόρου*, τέτοιες περιπτώσεις μπορούν να εντοπιστούν και να αποφευχθούν. Στο συγκεκριμένο παράδειγμα, τα  $G.c[1, 3] \rightarrow b$  και  $H.c[2, 4] \rightarrow b$  είναι τόσο μικρά που μπορούν να αμεληθούν κατά τη σύγκριση. Εισάγοντας ένα κατάλληλα μικρό κατώφλι  $B$  για να χαρακτηρίσουμε το βάρος του *μήκους συνόρου*, λαμβάνουμε  $\mathcal{L} \equiv 100\%$ .

Πίνακας 3.6 Η βελτιωμένη εκδοχή της διαδικασίας σύγκρισης

- 
1. Πάρε τους υπογράφους  $G_M \subseteq G$ ,  $H_M \subseteq H$  κατά το βέλτιστο μονοπάτι των γράφων γειτνίασης.
  2. Για ένα ζεύγος κόμβων  $(i, j)$  ( $r_i \in G_M$ ,  $r_j \in H_M$ ) υπολόγισε τις ακμές που αναχωρούν από τους κόμβους και αύξησε τους μετρητές *εσυνολικές* για κάθε ακμή και *εκοινές* όταν η ακμή ανήκει στο βέλτιστο μονοπάτι.
  3. Για ακμές (που συνδέουν τους κόμβους  $i, j$ ) που δεν έχουν αντίστοιχές τους μείωσε τις *εσυνολικές* όταν  $(G.c[i][j] \rightarrow b < B)$ .
  4. Πήγαινε στο βήμα 2 για το επόμενο ζεύγος κόμβων.
- 



**Σχήμα 3.7** Πειραματικά αποτελέσματα με μοντέλο ένα πολύχρωμο πακέτο: (α-ε) όψεις του πακέτου, (στ) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της εικόνας (α).

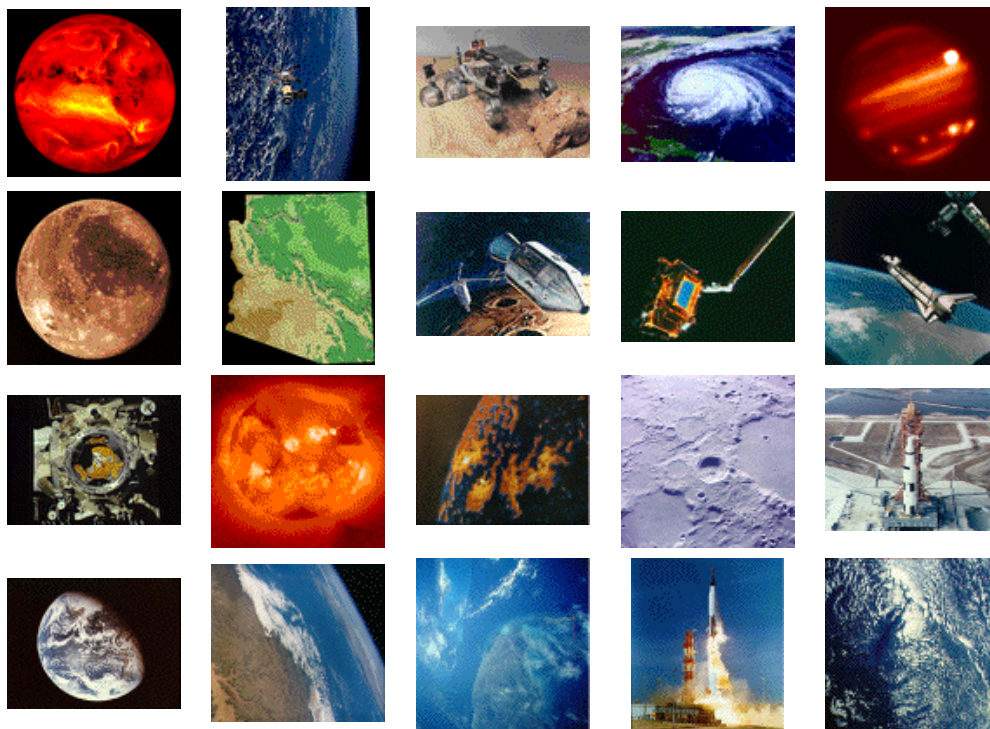
Ο Πίνακας 3.6 συνοψίζει το δεύτερο βήμα της διαδικασίας σύγκρισης.

### 3.6.3 Η συνολική μετρική ομοιότητας

Με δεδομένη την παραπάνω μεθοδολογία και τα ενδεικτικά αποτελέσματα, οι μετρικές *κάλυψη* και *συνδεσιμότητα* δείχνουν να συμπεριφέρονται καλώς στη γενική περίπτωση. Στη συγκεκριμένη υλοποίηση, αποφασίζουμε ότι η μετρική *ομοιότητας*  $\mathcal{R}$  δίνεται από το γινόμενο των δύο μετρικών

$$\mathcal{R} = \mathcal{C} \cdot \mathcal{L} \quad (3-4)$$

Κάποιος θα μπορούσε να χρησιμοποιήσει οποιαδήποτε άλλη συνάρτηση, που θα έδινε το  $\mathcal{R}$  συναρτήσει των  $\mathcal{C}$  και  $\mathcal{L}$ , ακόμη και να επιλέξει  $\mathcal{R} \equiv \mathcal{C}$ . Θα πρέπει όμως να σημειωθεί ότι χρησιμοποιώντας μόνο το  $\mathcal{L}$ , θα μπορούσε να χάσει την περίπτωση αντικειμένων που



**Σχήμα 3.8** 20 ενδεικτικές φωτογραφίες από τη βάση δεδομένων με 1393 εικόνες από το διάστημα, όπου πραγματοποιήθηκαν πειραματικές αναζητήσεις εικόνων με βάση το περιεχόμενο.

περιέχονται και στις δύο υπό σύγκριση εικόνες, δηλαδή μόνο όταν συγκεκριμένοι υπογράφοι των γράφων γειτνιάζουν ομοιάζουν.

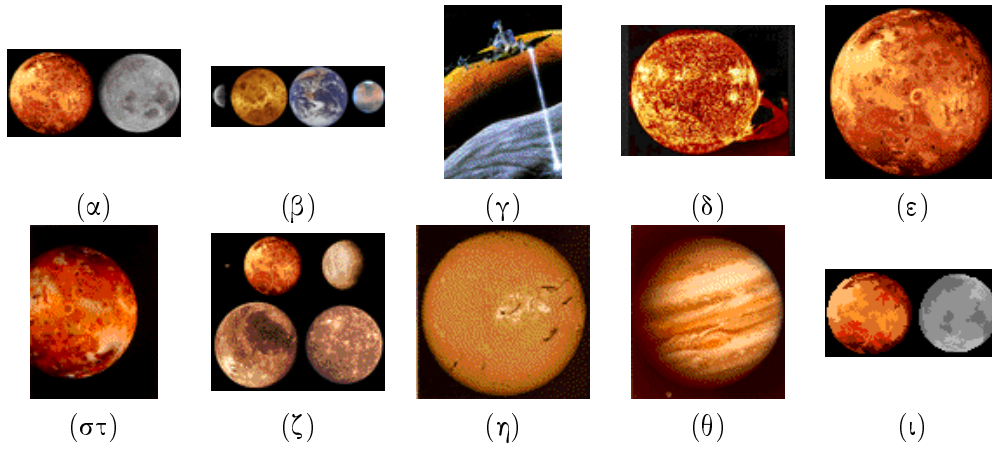
### 3.7 Πειράματα και συζήτηση

Η προτεινόμενη προσέγγιση ελέγχθηκε πειραματικά σε ένα μεγάλο αριθμό από απλές και σύνθετες εικόνες δίνοντας ικανοποιητικά αποτελέσματα. Ιδιαίτερη προσοχή δόθηκε στο εάν:

- (1) εικόνες με όμοια χρωματικά ιστογράμματα δύνανται να διακριθούν βάσει των διαφορών τους στην κατανομή χρώματος χρησιμοποιώντας τον προτεινόμενο αλγόριθμο,
- (2) εικόνες που αντιστοιχούν σε καρέ της ίδιας ακολουθίας βίντεο ταξινομούνται ως ιδιαίτερα όμοιες ακόμη και όταν τα αντικείμενα που περιέχουν υποβάλλονται σε σημαντική μετατόπιση, περιστροφή και μεγέθυνση.

Το Σχήμα 3.6 επιδεικνύει την επίδοση της μεθόδου σε εικόνες που λήφθηκαν με απόδοση (rendering) μίας συνθετικής σκηνής με περισσότερα από ένα αντικείμενα. Τα συγκεκριμένα χρωματικά αντικείμενα επιλέγησαν να είναι σχετικά απλά, ώστε τα λαμβανόμενα αριθμητικά αποτελέσματα να μπορούν εύκολα να εξηγηθούν ως προς τις οδηγίες που δίνονται στις προηγούμενες ενότητες. Τα Σχήματα 3.6(α-ε) περιέχουν τα αποδιδόμενα καρέ της στατικής σκηνής,





**Σχήμα 3.9** Αποτελέσματα αναζήτησης βάσει της προτεινόμενης μεθόδου, (α) εικόνα εισόδου, (β-θ) καλύτεροι αντιπρόσωποι σε φθίνουσα σειρά ομοιότητας με την εικόνα εισόδου, (ι) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της (α).

όπως αυτά λαμβάνονται από διαφορετικά σημεία παρατήρησης. Ενδεικτικά, το Σχήμα 3.6(στ) περιέχει την εικόνα που αντιστοιχεί στον μειωμένο γράφο γειτνίασης της εικόνας του Σχήματος 3.6(δ). Εάν το  $\mathcal{R}_{kl}$  αποτελεί την τιμή της συνολικής μετρικής ομοιότητας  $\mathcal{R}$  όταν το Σχήμα 3.6(k) είναι η εικόνα εισόδου που συγκρίνεται με την εικόνα του Σχήματος 3.6(l) στη βάση δεδομένων, λαμβάνουμε  $\mathcal{R}_{\alpha\beta} = 100\%$ ,  $\mathcal{R}_{\alpha\gamma} = 60\%$ ,  $\mathcal{R}_{\alpha\delta} = 60\%$ ,  $\mathcal{R}_{\alpha\epsilon} = 51\%$ . Οι μειωμένες τιμές των  $\mathcal{R}_{\alpha\gamma}$ ,  $\mathcal{R}_{\alpha\delta}$ ,  $\mathcal{R}_{\alpha\epsilon}$ , σε σχέση με το  $\mathcal{R}_{\alpha\beta}$ , οφείλονται κυρίως στην απουσία σύνδεσης ανάμεσα στα χρωματικά αντικείμενα στο Σχήμα 3.6(γ), το διαφορετικό τύπο σύνδεσης μεταξύ των αντικειμένων στο Σχήμα 3.6(δ) και την απουσία του μπλε αντικειμένου στο Σχήμα 3.6(ε). Το συγκεκριμένο παράδειγμα υποδηλώνει επίσης ότι η προτεινόμενη μέθοδος αντιμετωπίζει με επιτυχία την περίπτωση των σκηνών με πολλά αντικείμενα που παρακολουθείται από διαφορετικά σημεία του χώρου.

Στα Σχήματα 3.7(α-ε) απεικονίζονται καρέ μίας ακολουθίας που περιέχει ένα περιστρεφόμενο, σχετικά πολύπολοχο, χρωματιστό πακέτο. Ενδεικτικά, το Σχήμα 3.7(στ) περιέχει την εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της εικόνας του Σχήματος 3.7(α). Ο προτεινόμενος αλγόριθμος επιστρέφει σε αυτήν την περίπτωση, όπως αναμένεται, υψηλές τιμές για το συνολικό κριτήριο ομοιότητας  $\mathcal{R}$ . Ακολουθώντας το συμβολισμό της προηγούμενης παραγράφου, λαμβάνουμε  $\mathcal{R}_{\alpha\epsilon} = 99\%$ ,  $\mathcal{R}_{\alpha\beta} = 90\%$ ,  $\mathcal{R}_{\alpha\delta} = 87\%$ ,  $\mathcal{R}_{\alpha\gamma} = 77\%$ . Με δεδομένες τις αλλαγές στο φωτισμό, το μεγάλο ποσό μετασχηματισμών και/ή αποκρύψεων, τα αποτελέσματα που λαμβάνονται μπορούν να θεωρηθούν ιδιαίτερα ικανοποιητικά.

Η προτεινόμενη προσέγγιση ελέγχθηκε σε αρκετές περιπτώσεις επιστρέφοντας ικανοποιητικά αποτελέσματα, ιδιαίτερα στην περίπτωση που το σύνολο των εικόνων περιέχει εικόνες που είναι μερικώς όμοιες, που αποτελούνται από όμοια χρωματικά αντικείμενα τα οποία υπόκεινται

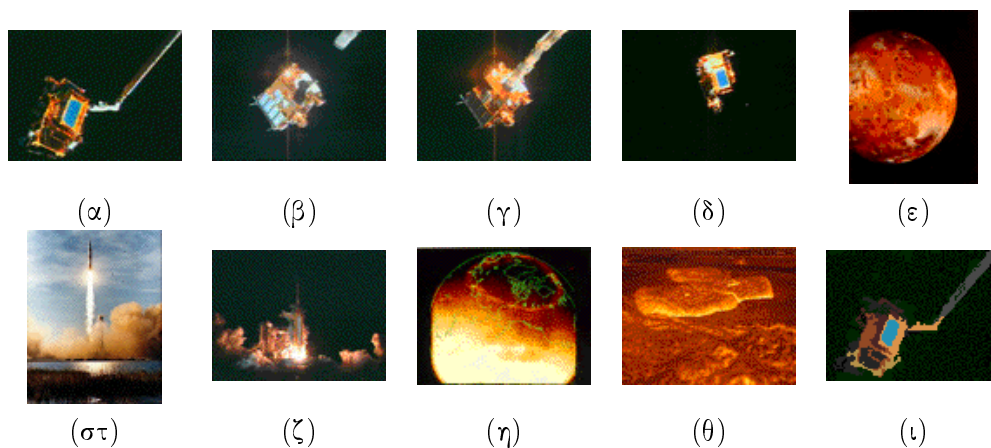
σε μετασχηματισμούς και/ή παραμορφώσεις ή αποτελούν καρέ της ίδιας ακολουθίας εικόνων. Το Σχήμα 3.8 περιέχει 20 ενδεικτικούς αντιπροσώπους από μία βάση δεδομένων 1393 εικόνων, που βρίσκεται διαθέσιμη στην ηλεκτρονική διεύθυνση <ftp://ftp.eecs.umich.edu/groups/ai/derwick/essbthm.zip> (η συγκεκριμένη βάση δεδομένων έχει επίσης χρησιμοποιηθεί και από άλλους ερευνητές για όμοιους σκοπούς, π.χ. στην [154]).

Το Σχήμα 3.9 απεικονίζει αποτελέσματα αναζήτησης που λήφθηκαν χρησιμοποιώντας την προτεινόμενη προσέγγιση στην προαναφερθείσα βάση δεδομένων εικόνων. Στα Σχήματα 3.9(α,ι) η εικόνα εισόδου και η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης έχουν περιληφθεί αντίστοιχα. Τα Σχήματα 3.9(β-θ) περιέχουν τους καλύτερους αντιπρόσωπους σε φθίνουσα σειρά ομοιότητας - εκτός από την πρωτότυπη εικόνα η οποία δίνει 100% σύμπτωση με τον εαυτό της. Για χάρη της πληρότητας, για τις φωτογραφίες που απεικονίζονται, η μεγαλύτερη σύμπτωση ήταν 0.73 ενώ η μικρότερη 0.54. Στα λαμβανόμενα αποτελέσματα μπορούμε να διακρίνουμε τους τόνους του πορτοκαλί, γκρι και μαύρου, οι οποίοι κατανέμονται τοπικά σε όλες τις φωτογραφίες. Ταυτόχρονα, οι περισσότερες εικόνες που λαμβάνονται περιέχουν το 'φλεγόμενο' ουράνιο σώμα. Πρέπει επίσης να σημειωθεί εδώ ότι χαρακτηριστικά σχήματος δεν περιλαμβάνονται στον αλγόριθμο, οπότε η παρουσία κυκλικών αντικειμένων στις περισσότερες εικόνες οφείλεται μόνο στα περιεχόμενα της συγκεκριμένης βάσης δεδομένων.

Το Σχήμα 3.10 απεικονίζει ένα ακόμη παράδειγμα επί των δεδομένων της ίδιας βάσης. Το ενδιαφέρον σημείο του παραδείγματος αυτού είναι ότι όλες οι εικόνες στη βάση που περιέχουν το αντικείμενο του Σχήματος 3.10(α), δηλαδή οι εικόνες στα Σχήματα 3.10(β-δ), λαμβάνονται με ιδιαίτερα υψηλό βαθμό ομοιότητας. Στην πραγματικότητα, οι υπόλοιπες εικόνες που λαμβάνονται ως όμοιες απλώς περιέχουν μεγάλες περιοχές μαύρου, πορτοκαλί ή γαλάζιου που αντιστοιχίζονται σε περιοχές ιδίου χρώματος της εικόνας εισόδου. Για χάρη της πληρότητας, για τις φωτογραφίες που απεικονίζονται, η μεγαλύτερη σύμπτωση ήταν 0.72 ενώ η μικρότερη 0.61.

### 3.8 Συμπεράσματα και επεκτάσεις

Στο κεφάλαιο αυτό παρουσιάσαμε ένα πρωτότυπο σύστημα για ανάκτηση εικόνων με βάση το περιεχόμενο, βασιζόμενο στο χαρακτηριστικό της κατανομής χρώματος. Η προσέγγιση αυτή διαφέρει ουσιαστικά από τις παραδοσιακές τεχνικές ανάλυσης σε τετραδικά δένδρα αλλά και τις παραδοσιακές τεχνικές γράφων γειτνίασης, καταλήγοντας σε μία πιο γενική και αποδοτική αναπαράσταση του χρωματικού περιεχομένου. Η αποδοτικότητα του αλγορίθμου



**Σχήμα 3.10** Αποτελέσματα αναζήτησης βάσει της προτεινόμενης μεθόδου, (α) εικόνα εισόδου, (β-θ) καλύτεροι αντιπρόσωποι σε φθίνουσα σειρά ομοιότητας με την εικόνα εισόδου, (ι) η εικόνα που αντιστοιχεί στο μειωμένο γράφο γειτνίασης της (α).

ελέγχθηκε σε έναν μεγάλο αριθμό από φυσικά παραδείγματα δίνοντας ικανοποιητικά αποτελέσματα.

Η προτεινόμενη τεχνική προσφέρει τη βάση για ένα σύνολο επεκτάσεων, οι σημαντικότερες από τις οποίες είναι:

- (1) η χρήση κάποιας κατάλληλης πολυδιακριτικής υλοποίησης σε δύο-τρία επίπεδα, ώστε η χρήση κατωφλίων να μην είναι κρίσιμη,
- (2) η ενσωμάτωση στα κριτήρια συγκρίσεων χαρακτηριστικών επιπλέον του χρώματος (π.χ. υψής) προς την κατεύθυνση ορισμού ενός γενικότερου χαρακτηριστικού σύνθεσης / κατανομής πληροφορίας,
- (3) η χρήση κατάτμησης κίνησης, από κοινού με τη χρωματική κατάτμηση, στη λογική των κινούμενων αντικειμένων, όταν πρόκειται για πληροφορία βίντεο.

## ΚΕΦΑΛΑΙΟ 4

### Εντοπισμός και Εξαγωγή Κύριων Κινούμενων Αντικειμένων σε Ακολουθίες Εικόνων

#### 4.1 Εισαγωγή / Αναδρομή

Ο εντοπισμός και η εξαγωγή κινούμενων αντικειμένων από ακολουθίες βίντεο αποτελεί την τελευταία δεκαετία μία από τις σημαντικότερες διαδικασίες της κατανόησης περιεχομένων βίντεο, η οποία είναι και άμεσα σχετιζόμενη με τη θεματολογία της ανάλυσης προτύπων, της υπολογιστικής όρασης και της τεχνητής νοημοσύνης. Ο εντοπισμός των κύριων κινούμενων αντικειμένων, ή εισβολέων, σε μία ακολουθία εικόνων μπορεί να αποτελέσει μέσο απόφασης για συστήματα συναγερμού πραγματικού χρόνου σε εφαρμογές παρακολούθησης ή ακόμη να βοηθήσει στη μέτρηση της ροής της κυκλοφορίας οχημάτων ή πεζών. Σε εφαρμογές στην περιοχή της βιολογίας, τεχνικές εντοπισμού κινούμενων αντικειμένων, υποστηριζόμενες από τεχνικές αναγνώρισης προτύπων, μπορούν να χρησιμοποιηθούν για καταμέτρηση μικροοργανισμών ή πληθυσμών υπό εξαφάνιση. Η χωροθέτηση των κινούμενων αντικειμένων στο επίπεδο προβολής (καρέ) μπορεί να χρησιμοποιηθεί ως ένα ενδιαφέρον χαρακτηριστικό για σκοπούς αναζήτησης εικόνων και βίντεο με βάση το περιεχόμενο, πιθανώς μαζί με κάποια επιπλέον χαρακτηριστικά σχήματος. Η εξαγωγή κινούμενων αντικειμένων μπορεί να χρησιμοποιηθεί, κατά περίπτωση και λαμβάνοντας υπόψη το επίπεδο ακρίβειας που επιτυγχάνεται, σε εφαρμογές αντικειμενοστρεφούς κωδικοποίησης (object-based coding) [133], ταξινόμησης αντικειμένων (object classification) [9], ανάλυσης ανθρώπινης κίνησης (human motion analysis) [126], ή ακόμη και σε εφαρμογές τρισδιάστατης ανακατασκευής (three-dimensional reconstruction) [188]. Επιπλέον του ευρείου πεδίου εφαρμογών τους, ο επιτυχής εντοπισμός και η εξαγωγή κινούμενων αντικειμένων έχουν συγκεντρώσει την προσοχή κατά την τελευταία δεκαετία, στα πλαίσια των προτύπων κωδικοποίησης MPEG-4 και MPEG-7, για αντικειμενοστρεφή κωδικοποίηση και αναζήτηση με βάση το περιεχόμενο αντίστοιχα [76].

Ένας μεγάλος αριθμός από αλγορίθμους εντοπισμού, χωροθέτησης και εξαγωγής κινούμενων αντικειμένων έχει προταθεί στη βιβλιογραφία, με έμφαση είτε στην ακρίβεια είτε στο χρόνο εκτέλεσης, ή ακόμη στις ανάγκες συγκεκριμένων ειδικών εφαρμογών. Το τελευταίο έχει ως αποτέλεσμα έναν αριθμό αλγορίθμων, οι οποίοι δεν είναι απευθείας συγκρίσιμοι, αντιμετωπίζουν όμως σημαντικές εφαρμογές, όπως η ανάλυση της περιοδικής κίνησης που είναι σημαντικό κριτήριο αναγνώρισης αντικειμένων και ενεργειών [24], η παρακολούθηση αρθρωτών δομών [126] και ο συνδυασμός ανάλυσης και παρακολούθησης σχήματος για τον εντοπισμό ανθρώπων και τη δημιουργία μοντέλων της εμφάνισής τους [52]. Μία ενδιαφέρουσα υποπερίπτωση

εντοπισμού και παρακολούθησης κινούμενων αντικειμένων εμφανίζεται όταν η τρισδιάστατη δομή των αντικειμένων που μας ενδιαφέρουν είναι γνωστή, η οποία αντιμετωπίζεται μέσω μεθόδων που βασίζονται σε μοντέλα [42, 85].

Τα κινούμενα αντικείμενα εντοπίζονται συνήθως βάσει ‘κινούμενων’ χαρακτηριστικών τους, όπως οι αλλαγές φωτεινότητας [144, 178, 57, 197], οι κινούμενες ακμές και γραμμές [23, 28] ή οι μετακινούμενες χρωματικές περιοχές [92, 36]. Στην ίδια κατεύθυνση, αντί της αντιστοίχισης χαρακτηριστικών, πολλές προσεγγίσεις αναζητούν ακριβή μοντέλα δισδιάστατης κίνησης για την κίνηση του φόντου και χρησιμοποιούν μία συνάρτηση ελέγχου κάποιου σχετικού περιορισμού, ως ένδειξη της ανεξάρτητης κίνησης του σώματος από το φόντο [121, 97]. Με όμοιο τρόπο, αρκετές μέθοδοι θεωρούν την τρισδιάστατη κίνηση της κάμερας γνωστή και εξάγουν περιορισμούς για τις παραμέτρους της δισδιάστατης κίνησης των αντικειμένων [157, 11]. Το πλεονέκτημα της ‘κανονικοποίησης’ της κίνησης (motion compensation) μεταξύ διαδοχικών καρέ, και εν συνεχεία της αντιστοίχισης χαρακτηριστικών ή του εντοπισμού τοπικών αλλαγών / διαφορών, είναι ότι η δεσπόζουσα κίνηση του φόντου (background dominant motion) εξάγεται με καλύτερη ακρίβεια συγκρινόμενη με τα δεδομένα παρακολούθησης χαρακτηριστικών (feature tracking data), λόγω της ποσότητας πληροφορίας που είναι διαθέσιμη σε κάθε περίπτωση. Από την άλλη πλευρά, η ποσότητα υπολογισμών που απαιτούνται για την ανάπτυξη σχετικά ακριβών δισδιάστατων μοντέλων κίνησης δεν δικαιολογείται πάντα από την επιτυγχανόμενη ακρίβεια, περισσότερο όταν η προβεβλημένη κίνηση είναι μικρή.

Μεταξύ μεθόδων που χρησιμοποιούν μοντέλα για να χωροθετήσουν τα κινούμενα αντικείμενα, τα παραμορφώσιμα μοντέλα (deformable models) θεωρούνται ανώτερα, εξαιτίας της ευκαμψότητάς τους και της δυνατότητάς τους να παρακολουθούν την κίνηση και στερεών και παραμορφώσιμων σωμάτων [118, 14, 74]. Ένα άμεσο πλεονέκτημα τέτοιων μεθόδων είναι ότι η δυναμική των εικόνων που μεταβάλλονται με το χρόνο εισάγεται με απλό τρόπο σε ένα μοντέλο δυνάμεων. Ενδιαφέρουσες είναι στην ίδια κατεύθυνση μέθοδοι που χρησιμοποιούν διάδοση καμπυλών (curve propagation), όπως οι μέθοδοι που βασίζονται σε level sets [107, 105]. Παραμορφώσιμα μοντέλα που βασίζονται σε πρωτότυπα (prototype-based) μπορούν να χρησιμοποιηθούν επίσης για να εντοπίσουν μεταβολές του σχήματος των κινούμενων αντικειμένων μεταξύ καρέ [201]. Το κύριο μειονέκτημα τέτοιων μεθόδων είναι ότι χρειάζονται μία καλή αρχική εκτίμηση του σχήματος του αντικειμένου, ή ακόμη πρότερη γνώση ενός προτύπου. Ταυτόχρονα, αναμένονται να αποτυγχάνουν όταν η μετατόπιση του αντικειμένου είναι σχετικά μεγάλη [97].

Μία από τις σημαντικότερες δυσκολίες στα προβλήματα εντοπισμού κινούμενων αντικειμένων είναι η απόρριψη των ελάσσονων ακμών, που εμφανίζονται λόγω ‘αποσπασματικών’

κινήσεων, όπως η κίνηση της βλάστησης στον άνεμο ή οι κατοπτρισμοί σε υγρά. Τέτοιες κινήσεις μπορούν να οδηγήσουν σε λανθασμένους συναγερμούς σε συστήματα παρακολούθησης χώρων για παράδειγμα, αλλά και σε μη ικανοποιητικά αποτελέσματα χωροθέτησης αντικειμένων παρουσία σημαντικού περιβαλλοντικού 'θορύβου' (temporal clutter). Οι συνήθειες προσεγγίσεις σε αυτό το πρόβλημα περιλαμβάνουν χωροχρονικό φιλτράρισμα [178] ή επιβολή περιορισμών για συμβατότητα στην κατεύθυνση του αντικειμένου με το χρόνο [180]. Παρόλ' αυτά μπορεί κανείς να δει ότι και οι δύο προσεγγίσεις δημιουργούν ανεπιθύμητες υποθέσεις για τη συμπεριφορά των κινούμενων αντικειμένων.

Μία ενδιαφέρουσα λύση στα περισσότερα συνήθη προβλήματα του εντοπισμού και χωροθέτησης κινούμενων αντικειμένων έρχεται από την περίπτωση της στατικής κάμερας, όπου οι προτεινόμενες προσεγγίσεις μπορούν να διαιρεθούν σε δύο κυρίως κατηγορίες: την βασιζόμενη σε χαρακτηριστικά (feature-based) που εξαρτάται από την εξαγωγή γενικών χαρακτηριστικών των ακολουθιών [108] και την βασιζόμενη σε pixels (pixel-based) που εξετάζει τις διαφορές μεταξύ διαδοχικών καρέ χρησιμοποιώντας pixels ως χαρακτηριστικά εισόδου. Μεταξύ των μεθόδων που βασίζονται σε pixels, μέθοδοι που βασίζονται στη φωτεινότητα (luminance-based) [54] και στις ακμές (edge-based) [87, 88, 100] έχουν προταθεί, με τις τελευταίες να αναφέρονται ως περισσότερο εύρωστες. Ειδικότερα, οι τεχνικές ενημέρωσης φόντου (background updating) [88] παρέχουν ένα βολικό τρόπο εντοπισμού κινούμενων χαρακτηριστικών χωρίς χρήση μεθόδων κατάτμησης κίνησης (motion-based segmentation), οι οποίες όχι μόνο παράγουν σημαντικό υπολογιστικό φορτίο, αλλά υποφέρουν και από τις συνήθεις αστοχίες της εκτίμησης κίνησης (motion estimation).

Μία σχετικά θεμελιώδης θεώρηση στην χωροθέτηση κινούμενων αντικειμένων είναι ο ανταγωνισμός μεταξύ ακρίβειας και αποδοτικότητας σε χρόνο. Για παράδειγμα, το πρότυπο MPEG-4 απαιτεί μία σχετικά ακριβή μάσκα που να περιέχει μόνο το κινούμενο αντικείμενο, ενώ για το πρότυπο MPEG-7 αρκεί να ανακτήσουμε το οριακό πολύγωνο (bounding polygon) ή ακόμη και το οριακό παραλληλόγραμμο. Στη γενική περίπτωση, πέρα από τα πρότυπα κωδικοποίησης, διαφορετικά επίπεδα ακρίβειας / ταχύτητας είναι αποδεκτά από διαφορετικές εφαρμογές - διαφορετικά επίπεδα λεπτομέρειας χρειάζονται για παράδειγμα για αναζήτηση με βάση το περιεχόμενο στα πλαίσια χαμηλού [187] ή υψηλού [195, 9] επιπέδου χαρακτηριστικών. Ο ανταγωνισμός ακρίβειας και αποδοτικότητας σε χρόνο δεν είναι τόσο σημαντικός στην περίπτωση της στατικής κάμερας όσο είναι στην περίπτωση που χρησιμοποιούμε κινούμενη κάμερα, απαιτώντας λειτουργίες μεγάλου υπολογιστικού κόστους όπως η εκτίμηση και κατάτμηση κίνησης. Εκμεταλλευόμενη αυτό το γεγονός, η μέθοδος που προτάθηκε στην [194] επέκτεινε έναν αλγόριθμο για στατική κάμερα [88] στην περίπτωση της κινούμενης κάμερας.

Στο παρόν κεφάλαιο, προτείνουμε έναν απλό αλγόριθμο για εντοπισμό και εξαγωγή κινούμενων αντικειμένων από ακολουθίες βίντεο, στη λογική του MPEG-7. Ο προτεινόμενος αλγόριθμος εκμεταλλεύεται τα πλεονεκτήματα των τεχνικών ενημέρωσης φόντου για να εξάγει τις κινούμενες ακμές που υπάρχουν στο τρέχον καρέ. Όπως μπορεί να επαληθευθεί πειραματικώς, μία απλή τεχνική ταιριάσματος μπλοκ με αναλυτική αναζήτηση (exhaustive-search block-matching) που εφαρμόζεται σε ένα μικρό υποσύνολο των pixels του τρέχοντος καρέ, εξάγει ικανοποιητικά την κίνηση του φόντου για σκοπούς κανονικοποίησης κίνησης, ως προς την καταλληλότητά της για το βήμα ενημέρωσης φόντου που ακολουθεί. Με αυτόν τον τρόπο, με μία ιδιαίτερα ταχεία υλοποίηση, τα προβλήματα των σταδιακών ή απότομων αλλαγών φωτισμού και της παρουσίας σημαντικού περιβαλλοντικού θορύβου αντιμετωπίζονται μέσω της τεχνικής ενημέρωσης φόντου και των προσθηκών / τροποποιήσεων που προτείνονται.

Η κύρια έμφαση της προτεινόμενης μεθόδου δίνεται στη διατήρηση χαμηλού υπολογιστικού κόστους, ώστε να μπορεί να χρησιμοποιηθεί αποδοτικά για συστήματα παρακολούθησης πραγματικού χρόνου και για συστήματα αναζήτησης βίντεο με βάση το περιεχόμενο. Αυτό επιτυγχάνεται μέσω της χρησιμοποίησης ενός απλού υποσυστήματος κανονικοποίησης κίνησης για διαδοχικά καρέ της ακολουθίας εικόνων και της δυαδικοποίησης όλων των λειτουργιών επί των ακμών που προκύπτουν. Αν και η προτεινόμενη μέθοδος δεν επιμένει στην ακρίβεια υπολογισμού των περιγραμμάτων των αντικειμένων, λόγω κυρίως της μεθόδου κανονικοποίησης κίνησης, φαίνεται να εξάγει τα κινούμενα αντικείμενα με ακρίβεια συγκρίσιμη με τις πιο πολύπλοκες μεθόδους παρακολούθησης περιγράμματος. Το τελευταίο είναι δυνατό κυρίως χάρη στο γεγονός ότι αυτή βασίζεται στην πληροφορία ακμών που λαμβάνεται από περισσότερα από δύο διαδοχικά καρέ, κρατώντας ταυτόχρονα χαμηλή ευαισθησία στις δευτερεύουσες ακμές. Το ιδιαίτερα χαμηλό υπολογιστικό κόστος της μεθόδου μας επιτρέπει επίσης να χρησιμοποιήσουμε τις ακμές που λαμβάνονται (ή το οριακό πολύγωνο) ως μία κατάλληλη αρχική εκτίμηση για μεθόδους που βασίζονται σε ενεργά περιγράμματα (active contours), όπως είναι για παράδειγμα τα snakes. Με αυτό το σκεπτικό, αντιμετωπίζονται τα προβλήματα της ακριβούς αρχικοποίησης και των περιορισμών στην ταχύτητα των αντικειμένων που εμφανίζουν οι τελευταίες. Θα πρέπει τέλος να σημειωθεί ότι ο προτεινόμενος αλγόριθμος δεν θέτει οποιουσδήποτε περιορισμούς στο μέγεθος ή στη συμπεριφορά των κινούμενων αντικειμένων.

Η προτεινόμενη μέθοδος έχει παρουσιαστεί στην εργασία [191], ενώ κάποιες πρώτες ιδέες αναφέρθηκαν στις [166, 194].

## 4.2 Τοποθέτηση του προβλήματος

Ο εντοπισμός και η χωροθέτηση κινούμενων αντικειμένων βασίζεται στον υπολογισμό των συνόρων των αντικειμένων ή ακόμη και του περιγράμματός τους από ακολουθίες εικόνων που καταγράφονται από κινούμενες ή στατικές κάμερες.

Στην περίπτωση της στατικής κάμερας, η σκηνή συμπεριλαμβανομένου του φόντου καταγράφεται βολικά από ένα συγκεκριμένο σταθερό σημείο παρατήρησης, ώστε όλη η δράση στην καταγραφόμενη ακολουθία να οφείλεται στην κίνηση κινούμενων αντικειμένων. Σ' αυτά τα πλαίσια, σχεδόν οποιαδήποτε αλλαγή στο πεδίο καταγραφής της κάμερας θα έπρεπε να υποδεικνύει την παρουσία ενός κινούμενου αντικειμένου που θα μπορούσε να εντοπιστεί και να χωροθετηθεί στην ακολουθία. Ταυτόχρονα όμως, υπάρχουν περιπτώσεις όπου, ακόμη κι αν υπάρχουν αλλαγές στην καταγραφόμενη ακολουθία, δεν υπάρχουν κινούμενα αντικείμενα ενδιαφέροντος προς παρακολούθηση - δηλαδή, όταν σταδιακές ή απότομες αλλαγές συμβαίνουν στις συνθήκες φωτισμού, όταν υπάρχουν μικρές τυχαίες αλλαγές στο φόντο (π.χ. όταν ο άνεμος κινεί φυτά ή κουρτίνες) ή όταν η κάμερα υποφέρει από μικρές δονήσεις. Βάσει των παραπάνω, το πρόβλημα του εντοπισμού κινούμενων αντικειμένων μετονομάζεται σε πρόβλημα εντοπισμού των 'κύριων' κινούμενων αντικειμένων, όπου αλλαγές όμοιες με αυτές που αναφέρθηκαν παραπάνω θεωρούνται ασήμαντες. Με άλλα λόγια, ο κύριος στόχος ενός εύρωστου αλγορίθμου εντοπισμού κύριων κινούμενων αντικειμένων για στατικές κάμερες είναι να διατηρεί υψηλή ευαισθησία στην παρουσία ενός σημαντικού κινούμενου αντικειμένου στην παρατηρούμενη σκηνή, ενώ να μειώνει την ευαισθησία του στις αλλαγές φωτεινότητας ή στις 'ασήμαντες' αλλαγές. Θα πρέπει να δοθεί προσοχή εδώ στο γεγονός ότι σε αντίθεση με τις μικρές χρονικές διαφοροποιήσεις λόγω 'θορύβου', μία αλλαγή στο φωτισμό μπορεί να οδηγήσει στη λανθασμένη θεώρηση μεγάλων περιοχών της παρατηρούμενης σκηνής ως κινούμενα αντικείμενα. Τέτοιες διαφοροποιήσεις είναι ιδιαίτερα σημαντικές στην περίπτωση καταγραφών εξωτερικού χώρου.

Στην περίπτωση κινούμενης κάμερας, πέρα από τις προαναφερθείσες δυσκολίες, θα πρέπει να αντιμετωπίσουμε και την κίνηση του φόντου λόγω της κίνησης της κάμερας. Εφόσον, για κινούμενες κάμερες, πραγματοποιούνται υπολογισμοί μεταξύ διαφορετικών καρέ για να εντοπίσουμε τα κινούμενα αντικείμενα, η πολυπλοκότητα του προβλήματος είναι ανάλογη με την πολυπλοκότητα της κίνησης της κάμερας, με το σκεπτικό ότι απότομη ή ισχυρά περιστροφική κίνηση αλλάζει σημαντικά το καταγραφόμενο φόντο. Στην περίπτωση που περισσότερα από ένα κινούμενα αντικείμενα υπάρχουν στη σκηνή, το σύστημα θα πρέπει να υπολογίσει τη σχετική ταχύτητα μεταξύ κάθε αντικειμένου και του φόντου (ή ακόμη να συμπεριλάβει κάποια



χωροχρονικά κριτήρια), ώστε να τα διακρίνει μεταξύ τους. Μία τέτοια διαδικασία γίνεται σημαντικά δύσκολη όταν τα αντικείμενα κινούνται σε μικρές αποστάσεις μεταξύ τους με όμοιες ταχύτητες, έχουν όμοιες χρωματικές ιδιότητες και ιδιότητες ακμών ή ακόμη επικαλύπτονται μερικώς.

Τα περισσότερα από τα συστήματα που προτείνονται στη βιβλιογραφία δείχνουν να επιβάλλουν ανεπιθύμητους περιορισμούς στην καταγραφόμενη ακολουθία ή λειτουργούν καλά μόνο κάτω από τέτοιους περιορισμούς. Η ανεπάρκεια τέτοιων μεθόδων εμφανίζεται παρουσία περισσότερων του ενός κινούμενων αντικειμένων, ειδικότερα όταν επικαλύπτονται μεταξύ τους, ή παρουσία σημαντικών ποσοτήτων θορύβου, ειδικότερα στην περιοχή των κινούμενων αντικειμένων (π.χ. όταν κινούμενα φυτά επικαλύπτουν μερικώς τα όρια του κινούμενου αντικειμένου). Επιπλέον, η υπόθεση ότι το όριο του αντικειμένου, ή ακόμη και το ίδιο το αντικείμενο, μπορεί να διακριθεί καλώς από το περιβάλλον του χρησιμοποιείται συχνά, δεν αποτελεί όμως τη συνήθη περίπτωση σε πραγματικές ακολουθίες. Τέλος, ακόμη και προσεγγίσεις που χρησιμοποιούν πολύπλοκα παραμετρικά μοντέλα κίνησης αποτυγχάνουν στην περίπτωση ισχυρής περιστροφικής κίνησης της κάμερας. Σε πολλές δε περιπτώσεις, η υπολογιστική πολυπλοκότητα των αλγορίθμων δεν δικαιολογείται από τα αποτελέσματα που επιτυγχάνονται.

Όπως θα δούμε στη συνέχεια του κεφαλαίου, η προτεινόμενη μέθοδος αντιμετωπίζει επιτυχώς όλα σχεδόν από τα παραπάνω προβλήματα και μπορεί να χρησιμοποιηθεί είτε ως ένα αποδοτικό εργαλείο για εντοπισμό αντικειμένων, όταν το ακριβές περίγραμμα των αντικειμένων δεν είναι αναγκαίο, είτε ως ένα ικανοποιητικό πρώτο βήμα για πολύπλοκους αλγορίθμους που απαιτούν την αρχικοποίηση του ενεργού περιγράμματος.

### 4.3 Η μονάδα κανονικοποίησης φόντου

Στα προβλήματα εντοπισμού κινούμενων αντικειμένων, ιδιαίτερα όταν οι ακολουθίες λαμβάνονται από κινούμενη κάμερα, η κίνηση του φόντου αποβάλλεται χρησιμοποιώντας μία μονάδα κανονικοποίησης κίνησης. Η διαδικασία της κανονικοποίησης κίνησης γενικά περιλαμβάνει τον υπολογισμό της κίνησης της κάμερας μέσω της εμφανιζόμενης κίνησης του φόντου, η οποία θα πρέπει στην πραγματικότητα να διαχωριστεί από την κίνηση των κινούμενων αντικειμένων που υπάρχουν στην ακολουθία.

Οι παραμετρικές μέθοδοι υπολογισμού κίνησης, ενώ βελτιώνουν ελαφρώς το λαμβανόμενο πεδίο διανυσμάτων κίνησης (motion field), αυξάνουν δραματικά το υπολογιστικό κόστος. Συνήθως σημειώνεται ότι μεταξύ διαδοχικών καρέ τα διανύσματα κίνησης που λαμβάνονται είναι

τόσο μικρά που θεωρούνται ακατάλληλα για την ανάπτυξη παραμετρικών μοντέλων, λαμβάνοντας επίσης υπόψη τις συνήθεις αστοχίες των τεχνικών εκτίμησης της διαστάσεως κίνησης. Επιπρόσθετα, οι τεχνικές κατάτμησης κίνησης που χρησιμοποιούνται για να διακρίνουμε τμήματα φόντου και κινούμενων αντικειμένων συχνά παράγουν απαράδεκτο κόστος για εφαρμογές πραγματικού χρόνου.

Στο παρόν κεφάλαιο, χρησιμοποιούμε την παραδοχή ότι για διαδοχικά (ή σχεδόν διαδοχικά) καρέ σε μία ακολουθία η κίνηση του φόντου είναι κίνηση μετατόπισης. Με αυτή τη λογική, μία απλή μη-παραμετρική μη-επαναληπτική τεχνική ταιριάσματος μπλοκ θα ήταν αρκετή. Μπορούμε όμως ακόμη, υπό την υπόθεση ότι η κίνηση του φόντου είναι κίνηση μετατόπισης και ότι τα κύρια κινούμενα αντικείμενα καλύπτουν μόνο ένα μέρος του φόντου, να αποφύγουμε τον υπολογισμό πυκνών πεδίων κίνησης και να κερδίσουμε περισσότερο υπολογιστικό χρόνο. Έτσι, μία τεχνική υπολογισμού κίνησης αναλυτικής αναζήτησης ταιριάσματος μπλοκ εκτελείται για ένα υποσύνολο των pixels του κάθε καρέ. Αν και είναι αλήθεια ότι τα κινούμενα αντικείμενα που μας ενδιαφέρουν καλύπτουν κυρίως το κέντρο της σκηνής / καρέ, δεν μπορεί κανείς να εισάγει κάποιον γενικό περιορισμό στη θέση και στο μέγεθός τους. Τα pixels για τα οποία αναζητείται ταιρίασμα μπλοκ επιλέγονται σε ένα ομοιογενές παραλληλόγραμμο πλέγμα, του οποίου η πυκνότητα αφήνεται ως ανοιχτή παράμετρος. Γενικά όμως, όπως δείχνουν τα πειράματά μας, αρκεί να χρησιμοποιήσουμε έναν ικανοποιητικό αριθμό από διανύσματα κίνησης, π.χ. 1%. Η τεχνική λογαριθμικής αναζήτησης σε τρία βήματα που χρησιμοποιείται στην [194], αν και ταχύτερη, αποδείχθηκε επισφαλής για συγκεκριμένα πειράματα.

Έστω ότι τα  $N \times M$  υποδηλώνουν τις διαστάσεις του καρέ και  $p_N, p_M$  το ποσοστό των pixels σε κάθε στήλη και γραμμή της εικόνας αντίστοιχα για το οποίο πραγματοποιείται εκτίμηση κίνησης. Εφόσον τα κύρια κινούμενα αντικείμενα καλύπτουν μόνο ένα μέρος σε κάθε καρέ, αναμένουμε ότι τα περισσότερα από τα λαμβανόμενα  $(p_N N) \times (p_M M)$  διανύσματα κίνησης ανήκουν στο φόντο και έτσι μοιράζονται όμοιες τιμές. Η κίνηση του φόντου επιλέγεται ως το συχνότερο διάνυσμα κίνησης του συνόλου  $\mathbf{U}$  των διανυσμάτων κίνησης, ή

$$\delta(f) = m(\mathbf{U}(f)) , \quad (4-1)$$

όπου το  $m()$  υποδηλώνει τη συνάρτηση 'πλειοψηφίας' (η οποία ισούται με μία πράξη διαστάσεως του ιστογράμματος με προαποφασισμένες στάθμες), το  $f$  υποδηλώνει το δείκτη του τρέχοντος καρέ και το  $\delta(f)$  την κίνηση του φόντου από το καρέ  $f-1$  στο καρέ  $f$ .

Με αυτή τη λογική, με δεδομένο ένα ζεύγος από διαδοχικά καρέ, λαμβάνουμε το  $\delta(f)$ . Χρησιμοποιώντας ένα σύνολο από τέτοιες διαδοχικές διαδικασίες υπολογισμού κίνησης, η κίνηση του φόντου λαμβάνεται για οποιαδήποτε  $\mathbf{I}(f_1)$  και  $\mathbf{I}(f_2)$  στις χρονικές στιγμές  $f_1$  και

$f_2$ . Έστω ότι το  $\mathbf{d}(f_1, f_2)$  υποδηλώνει την αντίστοιχη κίνηση του φόντου που λαμβάνεται με πρόσθεση των διαδοχικών  $\delta(f)$ . Τότε, το τρέχον καρέ  $\mathbf{I}(f)$  κανονικοποιείται ως προς κάποιο καρέ αναφοράς  $\mathbf{I}(f_r)$  με το διάνυσμα  $\mathbf{d}(f_r, f) \equiv [d_i \ d_j]$  ως,

$$\mathbf{I}^c(i, j, f) = \mathbf{I}(i + d_i, j + d_j, f_r) , \quad (4-2)$$

όπου το  $\mathbf{I}(i, j, f)$  γενικά υποδηλώνει την τιμή φωτεινότητας ενός pixel με χωρικές συντεταγμένες  $(i, j)$  στο καρέ  $f$  και ο εκθέτης  $(c)$  υποδηλώνει το κανονικοποιημένο όσον αφορά την κίνηση του φόντου καρέ (compensated). Η μέθοδος αυτή δεν είναι νέα στη βιβλιογραφία. Όμοιες προσεγγίσεις υπάρχουν για ‘σταθεροποίηση’ των καρέ μίας ακολουθίας, με κάποιες από αυτές να πραγματοποιούν ασαφή ταξινόμηση του τρεμοπαιξίματος των εικόνων [33].

#### 4.4 Η μονάδα ενημέρωσης φόντου

Η βασική ιδέα των τεχνικών ενημέρωσης φόντου είναι η αργή εισαγωγή των περιοχών μίας παρατηρούμενης σκηνής σε μία εικόνα αναφοράς που αντιπροσωπεύει το φόντο (reference background). Έτσι, μία περιοχή της παρατηρούμενης σκηνής που άλλαξε αλλά στη συνέχεια παραμένει απaráλλαχτη για ένα συγκεκριμένο χρονικό διάστημα αναγνωρίζεται ως ένα καινούριο στατικό χαρακτηριστικό της σκηνής και επομένως εισάγεται στο φόντο αναφοράς. Διαφορετικά, η παρατηρούμενη αλλαγή θεωρείται είτε κινούμενο αντικείμενο είτε θόρυβος. Στη συγκεκριμένη μονάδα ενημέρωσης φόντου επεκτείνουμε τη μέθοδο των Makarov *et al* [88] για στατικές κάμερες, η οποία βασίζεται στην πληροφορία ακμών και φαίνεται να είναι ιδιαίτερα εύρωστη τόσο σε ακολουθίες εσωτερικού όσο και εξωτερικού χώρου. Το κύριο πλεονέκτημα της μεθόδου είναι η δυνατότητά της να διακρίνει και να απορρίπτει αλλαγές, που συμβαίνουν λόγω της ύπαρξης υπερβολικού περιβαλλοντικού θορύβου στην παρατηρούμενη σκηνή, ακόμη και λόγω αντικειμένων (π.χ. δένδρα και φυτά) των οποίων η μικρή αλλά υπαρκτή κίνηση φαίνεται να επηρεάζει δυσμενώς την επίδοση άλλων μεθόδων.

Οι τεχνικές ενημέρωσης φόντου αποδίδουν ικανοποιητικά, ιδιαίτερα στην περίπτωση της στατικής κάμερας. Αυτό οφείλεται κυρίως στο γεγονός ότι αποτελούν έναν απλό και ταχύ τρόπο διαχωρισμού του φόντου από τις κινούμενες ακμές που οφείλονται στη δράση του προσκηνίου. Ένα άμεσο πλεονέκτημα των τεχνικών αυτών είναι ότι συνδυάζουν πληροφορία δομής από αρκετά διαδοχικά καρέ δίνοντας περισσότερο ευσταθή αποτελέσματα, συγκρινόμενες με μεθόδους εκτίμησης κίνησης που λαμβάνουν υπόψη τις διαφορές μεταξύ μόνο δύο διαδοχικών καρέ. Η πληροφορία που εξάγεται είναι δυαδική, δίνοντας δυνατότητα ταχύτατων δυαδικών υπολογισμών οι οποίοι μπορούν να υλοποιηθούν αποδοτικά και χρειάζονται ένα σημαντικό χαμηλό μέγεθος μνήμης. Ένα ακόμη άμεσο πλεονέκτημα έρχεται από τη χρήση των

κινούμενων / μεταβατικών ακμών (transient edges), υπό το πρίσμα του ότι τα κινούμενα αντικείμενα παράγουν ισχυρές ακμές στο κινούμενο περίγραμμά τους, το οποίο μπορεί να διακριθεί με το χρόνο, ακόμη και όταν το εσωτερικό του αντικειμένου δεν διακρίνεται καθαρά από το φόντο. Μπορεί να ειπωθεί επίσης ότι ένα κινούμενο περίγραμμα μπορεί να 'παρακολουθηθεί' ευκολότερα σε σχέση με μία κινούμενη περιοχή η οποία μπορεί να διαφέρει σημαντικά με το χρόνο λόγω κίνησης ή/και αλλαγών στο φωτισμό.

Για να ενημερωθεί το φόντο αναφοράς χρησιμοποιείται ένας μετρητής καρέ, του οποίου η μεταβολή εξαρτάται από το κριτήριο

$$\mathbf{C}^e(i, j, f) = \mathbf{C}^e(i, j, f - 1) + \mathbf{E}(i, j, f) , \quad (4-3)$$

όπου ο  $\mathbf{E}(i, j, f)$  είναι η τιμή του χάρτη ακμών στο pixel  $(i, j)$  τη χρονική στιγμή  $f$  που αντιστοιχεί στο pixel του καρέ  $\mathbf{I}(i, j, f)$  και  $\mathbf{C}^e(i, j, f)$  είναι η αντίστοιχη τιμή του μετρητή καρέ. Κάθε ακμή που εμφανίζεται στο pixel  $(i, j)$  στο τρέχον καρέ εισόδου  $f$  αυξάνει την τιμή στη θέση  $(i, j)$  του μετρητή καρέ, ενώ απουσία ακμής στην ίδια θέση οδηγεί σε μία αντίστοιχη μείωση. Έτσι, οι στατικές ακμές παράγουν υψηλές τιμές, ενώ οι μεταβατικές ακμές οδηγούν σε χαμηλότερες τιμές του μετρητή καρέ. Το φόντο αναφοράς λαμβάνεται τότε με κατωφλίωση του μετρητή,

$$\mathbf{B}(i, j, f) = \begin{cases} 1, & \text{εάν } \mathbf{C}^e(i, j, f) > \tau_a \\ 0, & \text{εάν } \mathbf{C}^e(i, j, f) \leq \tau_a \end{cases} , \quad (4-4)$$

όπου το κατώφλι  $\tau_a$  είναι ο χρόνος προσαρμογής (adaptation time) ή η καθυστέρηση εισαγωγής (insertion delay) της μονάδας ενημέρωσης φόντου.

Ουσιαστικά, η εξίσωση (4-3) υποδεικνύει ότι τα pixels σε ακμές που επιμένουν στις θέσεις τους ανατίθενται στο φόντο, αφού ο αντίστοιχος μετρητής διατηρεί μία υψηλή τιμή για ένα συγκεκριμένο χρονικό διάστημα. Η χρήση του συγκεκριμένου μετρητή οδηγεί σε μία ομαλή διαδικασία προσαρμογής του φόντου, αφού η απουσία της ακμής σε κάποιο ενδιαμέσο καρέ δεν υποδεικνύει την εξαφάνισή της αλλά μάλλον αποτυχία στη διαδικασία εξαγωγής των ακμών, για παράδειγμα λόγω των συνθηκών φωτισμού ή της διακριτότητας του pixel. Όμοια, μία νέα ακμή ανατίθεται στο φόντο μόνο όταν επιμένει στο χρόνο. Εφόσον οι παραπάνω διαδικασίες είναι ανταγωνιστικές, είναι η παράμετρος  $\tau_a$  που αποτιμά την επιμονή των ακμών. Ιδιαίτερα χαμηλές τιμές του  $\tau_a$  μπορούν να εξαφανίσουν κινούμενα pixels στο φόντο, ενώ ιδιαίτερα υψηλές τιμές μπορούν να εισάγουν στατικές ακμές στο σύνολο των μεταβατικών ακμών.

Αν και το κριτήριο της εξίσωσης (4-3) μπορεί να είναι κατάλληλο για στατικές κάμερες, η ενημέρωση φόντου στην περίπτωση της κινούμενης κάμερας μπορεί να αποτύχει λόγω κακής εφαρμογής των κανονικοποιημένων σε κίνηση καρέ, οδηγώντας σε ένα αυξημένο σύνολο

από μεταβατικές ακμές. Η ενημέρωση φόντου μπορεί να βελτιωθεί επιβάλλοντας περισσότερα κριτήρια πριν να αποφασίσουμε εάν μία περιοχή είναι μεταβατική ή στατική. Είναι φανερό ότι ενδιαφερόμαστε περισσότερο για μεταβατικές ακμές που ανήκουν στο περίγραμμα του αντικειμένου παρά στο εσωτερικό του. Ταυτόχρονα, το περίγραμμα του κινούμενου αντικειμένου καλύπτει / αποκαλύπτει μέρη του φόντου μεταξύ χρονικών στιγμών. Σε αυτά τα πλαίσια, οι αλλαγές στη φωτεινότητα μπορούν να υποδείξουν με αρκετή ασφάλεια τη θέση του κινούμενου περιγράμματος, οπότε μπορούν να χρησιμοποιηθούν για τον ορισμό ενός επιπλέον κριτηρίου για την ενημέρωση φόντου.

Έστω ότι το  $\nabla \mathbf{I}(i, j, f)$  συμβολίζει την κλίση της φωτεινότητας σε ένα pixel μίας εικόνας και το  $\mathbf{G}(i, j, f)$  δίνει την κατεύθυνσή της,

$$\mathbf{G}(i, j, f) = \angle \nabla \mathbf{I}(i, j, f) . \quad (4-5)$$

Χρησιμοποιώντας έναν όμοιο μετρητή  $\mathbf{C}^g(i, j, f)$  για να αποφασίσουμε εάν ένα pixel θα έπρεπε να εισαχθεί στο φόντο, θα αυξάναμε τον μετρητή όταν  $\mathbf{G}(i, j, f) \approx \mathbf{G}(i, j, f - 1)$ . Οπότε,

$$\mathbf{C}^g(i, j, f) = \mathbf{C}^g(i, j, f - 1) + (\Delta \mathbf{G}(i, j, f, f - 1) < t_g) , \quad (4-6)$$

όπου  $t_g$  είναι μία μικρή γωνία. Τότε το φόντο ενημερώνεται βάσει της

$$\mathbf{B}(i, j, f) = \begin{cases} 1, & \text{εάν } \mathbf{C}^g(i, j, f) > \tau_b \\ 0, & \text{εάν } \mathbf{C}^g(i, j, f) \leq \tau_b \end{cases} , \quad (4-7)$$

όπου το κατώφλι  $\tau_b$  είναι ο αντίστοιχος χρόνος προσαρμογής.

Όπως επιβεβαιώθηκε πειραματικά, οι δύο συνθήκες δίνουν καλύτερα αποτελέσματα μαζί από ότι η καθεμία χωριστά. Οι δύο συνθήκες (4-4) και (4-7) μπορούν να συνδυαστούν απλά σε μία μονάδα απόφασης - εάν επιπλέον επιλέξουμε  $\tau_a \equiv \tau_b$ , ένας μετρητής μπορεί να ενημερώνεται και για τα δύο κριτήρια.

#### 4.5 Εξαγωγή κινούμενων ακμών

Εφόσον το τρέχον καρέ και το προκύπτον φόντο αναφοράς είναι δυαδικές εικόνες ακμών, η εξαγωγή των μεταβατικών ακμών στο pixel  $(i, j)$  πραγματοποιείται μέσω ενός λογικού τελεστή που απομονώνει τις ακμές εισόδου που είναι απούσες στο ενημερωμένο φόντο αναφοράς:

$$\mathbf{T}(i, j, f) = \mathbf{E}(i, j, f) \wedge \overline{\mathbf{B}(i, j, f)} , \quad (4-8)$$

όπου τα  $(\wedge)$  και  $(\overline{\phantom{x}})$  υποδηλώνουν τους λογικούς τελεστές AND και NOT αντίστοιχα. Η εξίσωση (4-8) απλώς υποδεικνύει ότι οι μεταβατικές ακμές λαμβάνονται ως οι ακμές που

εξάγονται στο τρέχον καρέ, μετά από αφαίρεση των στατικών ακμών του φόντου την ίδια χρονική στιγμή.

Η χρήση των λογικών τελεστών AND/NOT οδηγεί σε μία δυαδική εικόνα, που αποτελείται από τις μεταβατικές ακμές του κινούμενου αντικειμένου (εάν υπάρχει) και από υπολείμματα ακμών φόντου, που οφείλονται στον περιβαλλοντικό θόρυβο. Στην [194] ο θόρυβος αντιμετωπίζεται με φιλτράρισμα median, αφού οι ακμές των κινούμενων αντικειμένων είναι χωρικά συμπαγείς, ενώ τα υπολείμματα του φόντου είναι απλωμένα τυχαία σε εκτεταμένες περιοχές των καρέ εισόδου. Όπως όμως αναφέρεται και στην [166], μπορεί κανείς να δει ότι το φιλτράρισμα median σε δυαδικές εικόνες είναι ισότιμο με δυαδικό φιλτράρισμα πλειοψηφίας, το οποίο οδηγεί σε ανεπαρκή αποτελέσματα για υπερβολικό περιβαλλοντικό θόρυβο. Με άλλα λόγια, ένα τέτοιο φιλτράρισμα δεν λαμβάνει υπόψη τις ιδιότητες συνδεσιμότητας των ακμών. Από την άλλη μεριά, αντίθετα με τον περιβαλλοντικό θόρυβο, τα περιγράμματα των κινούμενων αντικειμένων αποτελούνται από πλατιές μεταβατικές γραμμές. Σ' αυτά τα πλαίσια, η συνέλιξη με ένα φίλτρο εξασθένησης είναι περισσότερο κατάλληλη.

#### 4.6 Ταξινόμηση κινούμενων ακμών

Σύμφωνα με τις προηγούμενες ενότητες, μπορεί κανείς να καταλήξει σε ομάδες από pixels σε μεταβατικές ακμές, οι οποίες ορίζουν εμμέσως τη θέση και το μέγεθος του κινούμενου αντικειμένου [194]. Όπως μπορεί να επιβεβαιώσει κανείς πειραματικά, τα pixels που λαμβάνονται στο σημείο αυτό υποδεικνύουν στις περισσότερες περιπτώσεις τα κινούμενα αντικείμενα πάνω στο καρέ. Με αυτό το σκεπτικό, η απλούστερη και ταχύτερη επιλογή θα ήταν να χωροθετήσουμε τα κινούμενα αντικείμενα περιορίζοντας τη χωρική διασπορά των pixels που ανήκουν σε κάθε ένα από αυτά. Μία τέτοια στρατηγική θα έδινε ικανοποιητικά αποτελέσματα στην περίπτωση που τα κινούμενα αντικείμενα στη σκηνή είναι σχετικά μακριά μεταξύ τους.

Στη γενική όμως περίπτωση, περισσότερα από ένα κύρια κινούμενα αντικείμενα μπορούν να είναι παρόντα στη σκηνή, σχετικά κοντά μεταξύ τους, ή ακόμη επικαλυπτόμενα μεταξύ τους, ούτως ώστε να είναι αναγκαία η πληροφορία κίνησης παράλληλα με τη χωρική κατανομή των μεταβατικών pixels. Ταυτόχρονα, είναι ενδιαφέρον να υπολογίσουμε την κατεύθυνση της κίνησης για κάθε κινούμενο αντικείμενο. Μία τέτοια διαδικασία είναι επιβεβλημένη στην αυτόματη παρακολούθηση όπου κάμερες παρακολουθούν τον εισβολέα και σε συστήματα αναζήτησης με βάση το περιεχόμενο που επιτρέπουν ερωτήσεις όπως 'ένα κόκκινο αντικείμενο που κινείται νοτιοανατολικά'.

Για το σκοπό αυτό, μία μονάδα εκτίμησης κίνησης με ταίριασμα μπλοκ υπεισέρχεται στο σύστημα. Θα πρέπει ωστόσο να παρατηρηθεί ότι η πολυπλοκότητα σε χρόνο που απαιτεί αυτή η διαδικασία είναι αμελητέα, αφού η εκτίμηση κίνησης επιτελείται μόνο για pixels που ανήκουν σε μεταβατικές ακμές. Επιπλέον, η διαδικασία ταίριασματος μπλοκ μπορεί να εφαρμοστεί στις δυαδικές εικόνες, απλοποιώντας όλους τους υπολογισμούς σε λογικές πράξεις. Στην υλοποίησή μας, οριακά διαλέγουμε να μετρήσουμε τα pixels σε ακμές που κάθε μπλοκ περιλαμβάνει για να επιλέξουμε το καλύτερο ταίριασμα, αντί να υπολογίσουμε για παράδειγμα τη μέση απόλυτη διαφορά.

Σε μία εναλλακτική υλοποίηση, καλύτερα αποτελέσματα λαμβάνονται με εκτίμηση κίνησης μεταξύ των πρωτότυπων κανονικοποιημένων σε κίνηση καρέ, αντί των δυαδικών αντίστοιχών τους. Ακόμη και σε αυτήν την περίπτωση, το υπολογιστικό κόστος δεν είναι μεγάλο, αφού πάλι το υποσύνολο των pixels που χρησιμοποιούμε δεν είναι με κανέναν τρόπο συγκρίσιμο με το σύνολο των pixels στο καρέ. Είναι πολύ ενδιαφέρον να παρατηρήσουμε ότι σε αυτή την εναλλακτική τα αποτελέσματα της εκτίμησης κίνησης είναι όσο το δυνατόν ακριβέστερα, αφού τα κέντρα των μπλοκ βρίσκονται σε μεταβατικά pixels, δηλαδή pixels που βρίσκονται πάνω σε ακμές των εικόνων. Με αυτόν τον τρόπο, η εκτίμηση κίνησης πραγματοποιείται στο περισσότερο πληροφοριακό κομμάτι της εικόνας, δηλαδή στο σύνορο μεταξύ κινούμενων αντικειμένων και φόντου, μακριά από μεγάλες ομαλές περιοχές.

Έχοντας διαθέσιμη την πληροφορία κίνησης για τα pixels των ακμών, μία τεχνική ταξινόμησης κίνησης (π.χ. self-organizing map) αποφασίζει για την ταξινόμησή τους σε κινούμενα αντικείμενα απορρίπτοντας την ίδια στιγμή απομακρυσμένα σημεία. Στην υλοποίησή μας, χρησιμοποιούμε ένα διάνυσμα χαρακτηριστικών  $\mathbf{v}$  τάξεως τέσσερα,

$$\mathbf{v}_k = \begin{bmatrix} i & j & v_i & v_j \end{bmatrix}, \quad (4-9)$$

όπου το  $(i, j)$  αντιπροσωπεύει το διάνυσμα θέσεως του  $k$ -οστού μεταβατικού pixel και το  $(v_i, v_j)$  το αντίστοιχο υπολογιζόμενο διάνυσμα κίνησης. Μία άλλη, όχι τόσο προφανής, ιδιότητα της διαδικασίας αυτής είναι ότι επιτρέπει την απόρριψη των pixels ακμών που δεν υπακούουν σε κάποια κριτήρια, όταν για παράδειγμα δημιουργούν ένα μικρό ασήμαντο αντικείμενο.

#### 4.7 Εξαγωγή κινούμενων αντικειμένων

Για κάθε κινούμενο αντικείμενο εξάγεται τελικά το οριακό του πολύγωνο. Εφόσον περισσότερα από ένα πολύγωνα μπορούν να περιέχουν το κάθε κινούμενο αντικείμενο (ή σύνολο

από σημεία), υπολογίζουμε το κυρτό πολύγωνο που περιέχει το σύνολο σημείων (convex hull). Ενδεικτικά, χρησιμοποιούμε τη μέθοδο Graham Scan, η οποία είναι μικρού υπολογιστικού κόστους, ειδικά όταν οι συντεταγμένες των σημείων είναι ακέραιοι αριθμοί όπως στην περίπτωση μας (βλ. και την [138] για μία υλοποίηση).

Για να αποφασίσουμε εάν το τρέχον καρέ εισόδου περιέχει κινούμενα αντικείμενα, το άθροισμα των μεταβατικών pixels ανά καρέ μπορεί να χρησιμοποιηθεί ως κριτήριο, όπως στις [88, 194],

$$p(f) = \sum_i \sum_j \mathbf{T}(i, j, f) . \quad (4-10)$$

Μπορεί όμως να σημειωθεί ότι οι εξαγόμενες ομάδες σημείων δεν αποτελούν ένα ασφαλές κριτήριο, στο οποίο να μπορούμε να βασιστούμε για να αποφασίσουμε εάν ένα κινούμενο αντικείμενο είναι παρόν στη σκηνή. Αυτό συμβαίνει κυρίως λόγω του ότι ο αριθμός των pixels που ανιχνεύονται στο κινούμενο περίγραμμα δεν είναι πάντα ενδεικτικός του μεγέθους του αντικειμένου. Αφού καταφέρουμε στην προτεινόμενη υλοποίηση να υπολογίσουμε το οριακό πολύγωνο  $P_m$  για κάθε κινούμενο αντικείμενο  $m$ , η περίμετρος ή το εμβαδό του κάθε εξαγόμενου αντικειμένου θα μπορούσε να χρησιμοποιηθεί εναλλακτικά, ώστε να αποφασίσουμε εάν είναι αρκετά μεγάλο για να υποδεικνύει ένα κύριο κινούμενο αντικείμενο. Έτσι, η εξίσωση (4-10) ξαναγράφεται ως

$$p(f) = \max_f (a(P_m)) , \quad (4-11)$$

όπου  $a()$  είναι μία συνάρτηση που υπολογίζει το εμβαδό που περιλαμβάνεται σε ένα κυρτό πολύγωνο και  $\max_f$  είναι η μέγιστη τιμή μεταξύ των υπολογιζόμενων εμβαδών στο συγκεκριμένο καρέ  $f$ . Η διαδικασία αυτή μπορεί να είναι ευαίσθητη στον θόρυβο στην περίπτωση που η ταξινόμηση των μεταβατικών ακμών αποτύχει, ειδικά δε όταν τα απομακρυσμένα σημεία δεν απορριφθούν κατάλληλα.

Κατά τη διάρκεια του προαναφερθέντος χρόνου προσαρμογής  $\tau_a$ , η παρουσία ενός κινούμενου αντικειμένου στην παρατηρούμενη σκηνή δεν μπορεί να διακριθεί από τις μεταβολές στο φωτισμό. Το μονοδιάστατο σήμα  $p(f)$  χαρακτηρίζεται τότε από κορυφές μικρής διάρκειας όταν συμβαίνουν απότομες αλλαγές στο φωτισμό. Η μονάδα απόφασης για την παρουσία ή όχι ενός κινούμενου αντικειμένου στην παρατηρούμενη ακολουθία μπορεί να είναι εύρωστη σε απότομες μεταβολές φωτισμού, εάν οι εν λόγω κορυφές μικρής διάρκειας καταπιέζονταν με χρήση ενός χρονικού φίλτρου ελαχίστου (temporal min-filter) όπως στην [88], το οποίο δίνει το ελάχιστο δείγμα ενός σήματος εισόδου κατά τη διάρκεια ενός συγκεκριμένου χρονικού διαστήματος,

$$p_{min} = \min(p(t - \tau)), \text{ όπου } 0 \leq \tau \leq \tau_l . \quad (4-12)$$



**Πίνακας 4.1** Ανοιχτές παράμετροι του συστήματος εντοπισμού κινούμενων αντικειμένων

παράμετρος	περιγραφή	περιορισμοί-εξαρτήσεις
$r$	ρυθμός καρέ	σχετική ταχύτητα κάμερας
$a$	περιοχή αναζήτησης εκτίμησης κίνησης	σχετική ταχύτητα κάμερας
$\tau_u$	χρόνος ενημέρωσης καρέ αναφοράς	σχετική ταχύτητα κάμερας
$\tau_a$	χρόνος προσαρμογής φόντου	μεταβολές φωτισμού
$\tau_l$	μήκος φίλτρου ελαχίστου	μεταβολές φωτισμού
$t_d$	κατώφλι απόφασης εντοπισμού	εμβαδό οριακού πολυγώνου

Η χρήση του μετρητή καρέ  $\mathbf{C}$  για ενημέρωση φόντου σε συνδυασμό με το χρονικό φίλτρο ελαχίστου εξασφαλίζει ότι η μέθοδος εντοπισμού κινούμενων αντικειμένων είναι εύρωστη και σε σταδιακές και σε απότομες μεταβολές στον φωτισμό. Τελικά, η απόφαση  $D$  σχετικά με την παρουσία ή όχι ενός κινούμενου αντικειμένου στην παρατηρούμενη σκηνή λαμβάνεται συγκρίνοντας την έξοδο του χρονικού φίλτρου ελαχίστου με ένα κατώφλι  $t_d$ ,

$$D = \begin{cases} \text{true}, & \text{όταν } p_{\min} > t_d \\ \text{false}, & \text{όταν } p_{\min} \leq t_d \end{cases} \quad (4-13)$$

#### 4.8 Μία πολυδιακριτική υλοποίηση

Η κύρια πηγή ανακριβειών στο προτεινόμενο σύστημα είναι η παρουσία θορύβου, δηλαδή μεταβατικά pixels τα οποία δεν ανήκουν σε κινούμενα αντικείμενα, όμως παραμένουν στο σύνολο μεταβατικών ακμών δημιουργώντας προβλήματα στην εξαγωγή των οριακών πολυγώνων των κινούμενων αντικειμένων. Τέτοια ‘επίμονα’ pixels μπορούν να οφείλονται (α) σε ισχυρή περιστροφική κίνηση της κάμερας που δυσκολεύει τοπικά την τεχνική ενημέρωσης φόντου ή (β) σε σημαντικό θόρυβο περιβάλλοντος που επιτρέπει την επιβίωση τυχαίων pixels. Έως αυτό το σημείο, εκτός των κριτηρίων που επιβάλλονται στην τεχνική ενημέρωσης φόντου, δύο ακόμη μέτρα λαμβάνονται για να εξαφανίσουμε τέτοιες ανεπιθύμητες ακμές, συγκεκριμένα η χρήση του φίλτρου εξασθένισης στην Ενότητα 4.5 και η απόρριψη των απομακρυσμένων pixels κατά την ταξινόμηση στην Ενότητα 4.6.

Παρόλα τα παραπάνω, μπορεί να επαληθευθεί ότι μερικές ανεπιθύμητες ακμές παραμένουν στο σύνολο των μεταβατικών pixels. Σε όλες σχεδόν τις περιπτώσεις, αυτά μπορούν να απομακρυνθούν εάν ακολουθήσουμε μία πολυδιακριτική προσέγγιση. Η τελευταία δικαιολογείται, θεωρώντας ότι εάν υποδειγματοληπτήσουμε την εικόνα κατά έναν παράγοντα δύο, οι ακμές των κινούμενων αντικειμένων θα επιμείνουν, ενώ τα στατικά pixels θα εξαφανιστούν, αφού η

εμφάνισή τους δεν οφείλεται σε κάποιο συστηματικό λόγο. Στην πραγματικότητα, αυτό που συμβαίνει είναι ότι στις δύο αναλύσεις επιμένουν οι μεταβατικές ακμές στα ίδια σημεία, ενώ οι ανεπιθύμητες ακμές δεν εμφανίζονται στις ίδιες θέσεις λόγω της τυχειότητάς τους.

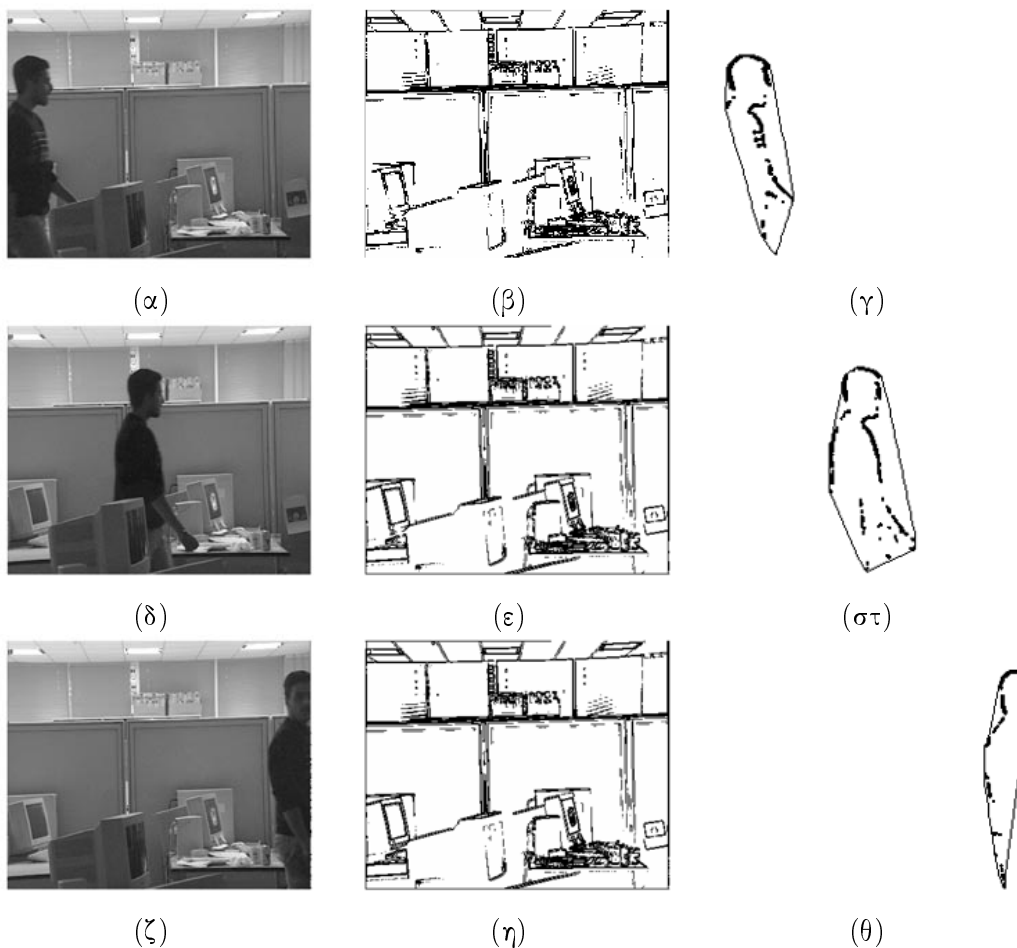
Πρακτικά, η πολυδιακριτική τεχνική εφαρμόζεται ως ακολούθως. Μαζί με κάθε καρέ εισόδου λαμβάνουμε το αντίστοιχό του με υποδειγματοληψία κατά δύο. Το τελευταίο υποβάλλεται σε όλα τα βήματα του αλγορίθμου έως ότου λάβουμε τις αντίστοιχες ομάδες σημείων, με τη διαφορά ότι η εκτίμηση κίνησης για το φόντο πραγματοποιείται μόνο στην πρωτότυπη ακολουθία και ανάλυση. Αφού οι μεταβατικές ακμές και στο τρέχον καταγραφόμενο καρέ και στο υποδειγματοληπτημένο αντίστοιχό του έχουν εξαχθεί, ο χάρτης ακμών του υποδειγματοληπτημένου καρέ επαναφέρεται στην αρχική ανάλυση και συγκρίνεται απευθείας με το χάρτη ακμών του πρωτότυπου καρέ. Αφού οι ελάχιστονες ακμές εμφανίζονται ως θόρυβος και στα δύο καρέ, μόνο οι μεταβατικές ακμές επιβιώνουν στο συνδυασμένο χάρτη ακμών. Με αυτόν τον τρόπο, η απόδοση της τεχνικής ταξινόμησης βελτιώνεται σημαντικά.

Όπως επιβεβαιώνεται πειραματικά, μία υποδειγματοληπτημένη κατά δύο εκδοχή της πρωτότυπης ακολουθίας είναι αρκετή για να λάβουμε ικανοποιητικά αποτελέσματα σε όλες τις περιπτώσεις.

#### 4.9 Προσαρμογή παραμέτρων

Το πρόβλημα του εντοπισμού και χωροθέτησης κινούμενων αντικειμένων είναι τόσο γενικό και χρησιμοποιείται σε τόσο μεγάλη ποικιλία εφαρμογών που είναι σχεδόν αδύνατο για έναν συγκεκριμένο αλγόριθμο να καλύψει όλες τις περιπτώσεις με επιτυχία. Για παράδειγμα, η περίπτωση μίας ακολουθίας που λαμβάνεται με μία στατική κάμερα για σκοπούς παρακολούθησης εσωτερικών χώρων διαφέρει αισθητά από αυτήν της κινούμενης κάμερας χειρός σε αγώνες αυτοκινήτων για σκοπούς δεικτοδότησης βίντεο. Ανάμεσα στα πλεονεκτήματα της προτεινόμενης προσέγγισης, μπορούμε να προσμετρήσουμε ότι η ρύθμιση των ανοιχτών παραμέτρων βελτιώνει τα λαμβανόμενα αποτελέσματα με τη λογική που περιγράφεται παρακάτω (βλ. και Πίνακα 4.1).

Ο ρυθμός των καρέ (frame-rate)  $r$  που δίνονται ως είσοδος στον αλγόριθμο μπορεί να είναι διαφορετικός (μικρότερος) από τον ρυθμό καρέ καταγραφής. Για παράδειγμα,  $r = 5$  καρέ/δευτερόλεπτο θα ήταν μία λογική επιλογή για μία ακολουθία εσωτερικού χώρου που παράγεται από μία στατική κάμερα. Η παράμετρος  $r$  γενικά εξαρτάται από τον αναμενόμενο ρυθμό αλλαγών στο πεδίο καταγραφής της κάμερας. Ειδικότερα, για καταγραφή με μία κινούμενη κάμερα, η σχετική ταχύτητα μεταξύ του φόντου και των κινούμενων αντικειμένων



**Σχήμα 4.1** Εντοπισμός κινούμενων αντικειμένων για ένα πολύπλοκο παραμορφώσιμο αντικείμενο (ανθρώπινο σώμα) σε μία ακολουθία εσωτερικού χώρου που ελήφθη με στατική κάμερα, υπό μεταβαλλόμενες συνθήκες φωτισμού και μερική απόκρυψη, (α,δ,ζ) τρία από τα διαθέσιμα καρέ, (β,ε,η) αποτελέσματα ενημέρωσης φόντου, (γ,στ,θ) κινούμενες ακμές και οριακό πολύγωνο που λαμβάνεται.

παίζει σημαντικό ρόλο. Όταν, για παράδειγμα, τα κινούμενα αντικείμενα αναμένονται να κινούνται ταχέως (π.χ. σε ακολουθίες αγώνων αυτοκινήτων), το  $r$  θα πρέπει να τεθεί σχετικά μεγάλο, ώστε να εντοπιστούν τα μεταβατικά περιγράμματα. Αντιθέτως, όταν το  $r$  είναι πολύ μεγάλο, υπάρχει περίπτωση κινούμενα αντικείμενα που κινούνται αργά να εξαφανιστούν στο φόντο. Γενικά, το τελευταίο δεν είναι τόσο πιθανό για πραγματικούς ρυθμούς καρέ 25 καρέ/δευτερόλεπτα σε φυσικές ακολουθίες.

Ο χρόνος προσαρμογής του φόντου (background adaptation time)  $\tau_a$  εξαρτάται από την αναμενόμενη διάρκεια ανεπιθύμητων απότομων αλλαγών φωτεινότητας. Για μία ακολουθία εξωτερικού χώρου τέτοιες αλλαγές θα μπορούσε για παράδειγμα να προξενήσει μία αστραπή, ενώ για καταγραφές εσωτερικού χώρου όμοια αποτελέσματα θα είχαμε τροποποιώντας την ένταση ή τη θέση των πηγών φωτός. Εάν αναμένουμε ότι τέτοια φαινόμενα θα διαρκέσουν

το πολύ  $t$  δευτερόλεπτα, ο χρόνος προσαρμογής δεν μπορεί να είναι μικρότερος από  $\tau_a = t r$  καρέ.

Το κατώφλι απόφασης εντοπισμού (detection decision threshold)  $t_d$  εξαρτάται από την περίμετρο των αναμενόμενων κινούμενων αντικειμένων σε pixels. Στην περίπτωση που αποφασίσουμε να χρησιμοποιήσουμε ένα κριτήριο που να βασίζεται στο εμβαδό που καλύπτεται από κάθε κινούμενο αντικείμενο, το  $t_d$  εξαρτάται τότε από το εμβαδό που καλύπτει ο εισβολέας. Στην πραγματικότητα, αρκεί το  $t_d$  να είναι αρκετά μεγαλύτερο από την περίμετρο ή το εμβαδό μικρών αντικειμένων που εντοπίζονται λόγω περιβαλλοντικού θορύβου, το οποίο μπορεί με τη σειρά του να αποφασιστεί δεδομένων των συνθηκών καταγραφής. Θα πρέπει όμως να παρατηρηθεί ότι η τιμή του  $t_d$  μπορεί πρακτικά να επιλεγεί από ένα μεγάλο εύρος τιμών, χωρίς να επηρεάζει σημαντικά την επίδοση του συστήματος. Το τελευταίο ισχύει χάριν του γεγονότος ότι εξάγονται τα οριακά πολύγωνα των αντικειμένων αντί των αντίστοιχων ομάδων από pixels [88].

Το μήκος του φίλτρου ελαχίστου (min-filter length)  $\tau_l$  εξαρτάται, όμοια με το χρόνο προσαρμογής  $\tau_a$ , από τις συνθήκες φωτισμού. Πρακτικά, το  $\tau_l$  θα πρέπει να επιλεγεί μεγαλύτερο από το  $\tau_a$ , ώστε το φόντο να έχει ήδη ενημερωθεί. Στην υλοποίησή μας επιλέγουμε  $\tau_a < \tau_l < 2\tau_a$ .

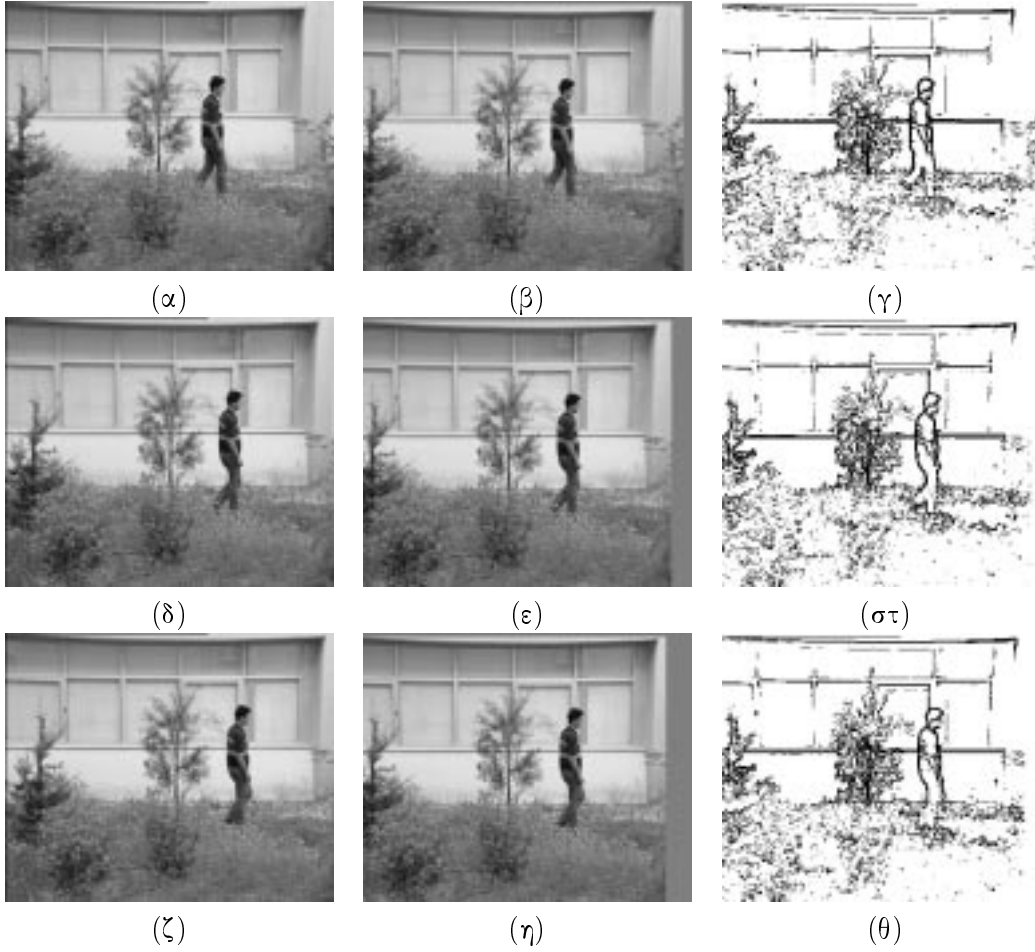
Ο χρόνος ενημέρωσης  $\tau_u$  του καρέ αναφοράς εξαρτάται ισχυρά από το σενάριο. Λαμβάνεται όσο το δυνατόν μικρότερος ώστε να αποφύγουμε τις αλλαγές σκηνών. Αντίθετα, πρέπει να είναι προφανώς αρκετά μεγαλύτερος από το  $\tau_l$ . Μπορούμε να δούμε ότι η παράμετρος  $\tau_u$  μπορεί να επιλεγεί πολύ μεγάλη για στατικές (ή σχεδόν στατικές) κάμερες.

#### 4.10 Ο αλγόριθμος σε βήματα

Βασιζόμενοι στις θεωρητικές και πρακτικές οδηγίες των προηγούμενων παραγράφων, ο προτεινόμενος αλγόριθμος ακολουθεί εν συντομία στη μορφή ακολουθιακών βημάτων.

1. Για το πρώτο καρέ  $\mathbf{I}_1(0)$ , σε πλήρες μέγεθος και στη χρονική στιγμή 0, εξήγαγε ένα (ή περισσότερα) υποδειγματοληπτημένα καρέ  $\mathbf{I}_{1/2}(0)$  και τους αντίστοιχους χάρτες ακμών  $\mathbf{E}_1(0)$ ,  $\mathbf{E}_{1/2}(0)$ . Θεώρησε το καρέ  $\mathbf{I}_1(0)$  ως το καρέ αναφοράς  $\mathbf{I}_1^r$ . Αρχικοποίησε τους μετρητές ακμών και κλίσης φωτεινότητας  $\mathbf{C}_1^e$ ,  $\mathbf{C}_{1/2}^e$  και  $\mathbf{C}_1^g$ ,  $\mathbf{C}_{1/2}^g$ .

2. Για κάθε καρέ  $f$  που δίνεται ως είσοδος στον αλγόριθμο:



**Σχήμα 4.2** Εντοπισμός κινούμενων αντικειμένων σε μία ακολουθία εξωτερικού χώρου που ελήφθη με μία ελαφρώς περιστρεφόμενη κάμερα χειρός παρουσία σημαντικού θορύβου (βλάστηση), (α,δ,ζ) τρία μη διαδοχικά καρέ, (β,ε,η) τα αντίστοιχα καρέ κανονικοποιημένα ως προς την κίνηση ως προς καρέ αναφοράς, (γ,στ,θ) αποτελέσματα εντοπισμού ακμών.

(α) υπολόγισε την κίνηση μετατόπισης του φόντου μεταξύ των  $I_1(f)$  και  $I_1(f - 1)$  και εξήγαγε τα κανονικοποιημένα σε κίνηση καρέ  $I_1^c(f)$ ,  $I_{1/2}^c(f)$  ως προς αναφορά  $I_1^r$  όπως φαίνεται στην Ενότητα 4.3,

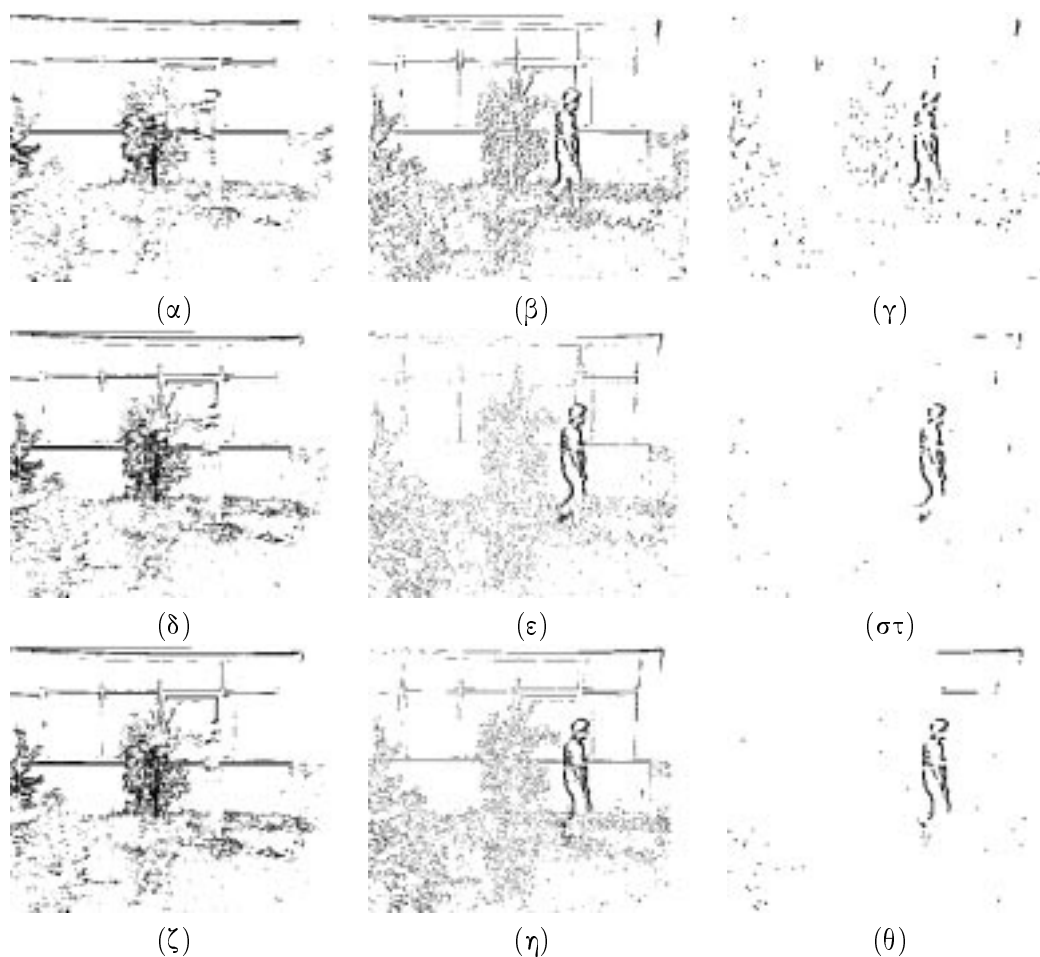
(β) παρήγαγε τα  $E_1(f)$ ,  $E_{1/2}(f)$  και/ή  $G_1(f)$ ,  $G_{1/2}(f)$  από τα  $I_1^c(f)$  και  $I_{1/2}^c(f)$ ,

(γ) ενημέρωσε τους  $C_1^e$ ,  $C_{1/2}^e$  και  $C_1^g$ ,  $C_{1/2}^g$  όπως υποδεικνύεται στις εξισώσεις (4-3) και (4-6) της Ενότητας 4.4,

(δ) υπολόγισε το φόντο  $B_1$ ,  $B_{1/2}$  και τα μεταβατικά pixels  $T_1$ ,  $T_{1/2}$  όπως υποδεικνύεται στις εξισώσεις (4-4), (4-7) και (4-8) των Ενότητων 4.4, 4.5,

(ε) εξήγαγε τα μεταβατικά pixels  $T$  από τα  $T_1$ ,  $T_{1/2}$  όπως εξηγεί η Ενότητα 4.8,

(στ) χώρισε τα μεταβατικά pixels σε ομάδες σημείων που να αντιστοιχούν σε κινούμενα αντικείμενα ακολουθώντας τη στρατηγική της Ενότητας 4.6,



**Σχήμα 4.3** Αποτελέσματα που επιδεικνύουν την ενημέρωση του φόντου και την εξαγωγή των κινούμενων ακμών για τα καρέ που εμφανίζονται στο Σχήμα 4.2 (η ενημέρωση του φόντου βάσει μόνο των κινούμενων ακμών σε μία ανάλυση φαίνεται ανεπαρκής), (α,δ,ζ) το ενημερωμένο φόντο, (β,ε,η) οι εξαγόμενες κινούμενες ακμές, (γ,στ,θ) οι κινούμενες ακμές μετά από συνέλιξη με ένα φίλτρο εξασθένισης.

(ζ) εξήγαγε τα οριακά πολύγωνα των κινούμενων περιοχών, υπολόγισε την τιμή του φίλτρου ελαχίστου και αποφάσισε για την παρουσία κύριων κινούμενων αντικειμένων όπως στις εξισώσεις (4-12) και (4-13) της Ενότητας 4.7,

(η) ενημέρωσε το καρέ αναφοράς  $I_1^r$ , εάν χρειάζεται σύμφωνα με την αντίστοιχη παράμετρο και αρχικοποίησε τους μετρητές.

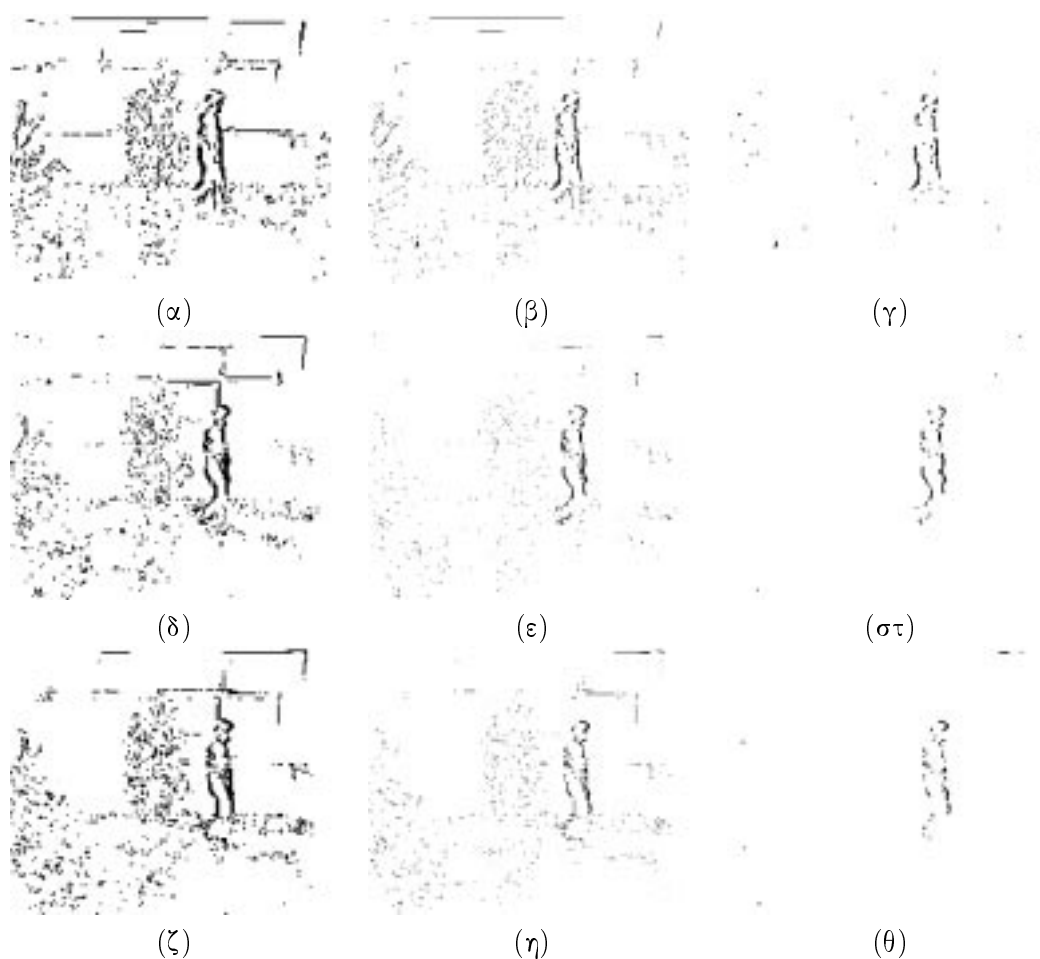
#### 4.11 Πειράματα και συζήτηση

Η επίδοση του προτεινόμενου αλγορίθμου δοκιμάστηκε σε ένα μεγάλο αριθμό από φυσικές ακολουθίες που λήφθηκαν μέσω μίας κάμερας χειρός και σε ακολουθίες δοκιμών διαθέσιμες από άλλους ερευνητές στο διαδίκτυο. Τα καρέ που δοκιμάστηκαν ήταν αρχικά διαθέσιμα σε μορφή PAL interlaced και στη συνέχεια υποδειγματοληπτήθηκαν κατάλληλα ώστε να λάβουμε

τα αντίστοιχα non-interlaced καρέ. Στα επόμενα, συμπεριλαμβάνουμε μερικά ενδεικτικά παραδείγματα που επιδεικνύουν την επίδοση της μεθόδου σε διαφορετικές περιπτώσεις και την αποδοτικότητα των χαρακτηριστικών επιλογών που έγιναν. Τα καρέ που εμφανίζονται στα σχήματα αυτής της ενότητας δεν είναι διαδοχικά, ώστε να επιδειχθεί ο εντοπισμός των κινούμενων αντικειμένων σε διαφορετικές χρονικές στιγμές.

Στα Σχήματα 4.1(α,δ,ζ) περιλαμβάνονται τρία καρέ μίας ακολουθίας εσωτερικού χώρου που καταγράφηκαν με μία στατική κάμερα. Ακολουθώντας τα βήματα του αλγορίθμου, όπως δίνεται στην Ενότητα 4.10, λαμβάνουμε το ενημερωμένο φόντο και τα κινούμενα αντικείμενα των Σχημάτων 4.1(β,ε,η) και 4.1(γ,στ,θ) αντίστοιχα. Στην περίπτωση αυτή, ο αλγόριθμος αποδείχθηκε εύρωστος παρά τις μεταβλητές συνθήκες φωτισμού που επιβλήθησαν κατά την καταγραφή αλλάζοντας τις πηγές φωτός. Λόγω της φύσης της μεθόδου, η μερική επικάλυψη του κινούμενου αντικειμένου δεν επηρεάζει τα λαμβανόμενα αποτελέσματα. Όπως μπορεί να δει κανείς, τα αποτελέσματα δεν θα διέφεραν σημαντικά στην περίπτωση αυτή, ακόμη κι αν παραλείπαμε τις μονάδες κανονικοποίησης κίνησης ως προς το φόντο και ταξινόμησης των μεταβατικών ακμών.

Τα Σχήματα 4.2 και 4.3 απεικονίζουν τα αποτελέσματα που λαμβάνονται για μία ακολουθία εξωτερικού χώρου που καταγράφηκε από μία κινούμενη κάμερα χειρός (με μικρή περιστροφική κίνηση και κραδασμούς). Στη συγκεκριμένη ακολουθία, η κύρια πηγή προβλημάτων είναι η παρουσία σημαντικού ποσού περιβαλλοντικού θορύβου, λόγω της κινούμενης βλάστησης η οποία καλύπτει σχεδόν το 50% της επιφάνειας του καρέ. Τα Σχήματα 4.2(α,δ,ζ) απεικονίζουν τρία μη διαδοχικά καρέ της ακολουθίας και τα Σχήματα 4.2(β,ε,η) τα αντίστοιχα κανονικοποιημένα σε κίνηση καρέ σε σχέση με το ίδιο καρέ αναφοράς, υποθέτοντας απλή μετατόπιση μεταξύ διαδοχικών καρέ όπως στην Ενότητα 4.3. Τα Σχήματα 4.2(γ,στ,θ) απεικονίζουν τότε τα αποτελέσματα εντοπισμού ακμών στα κανονικοποιημένα καρέ. Στο Σχήμα 4.3 περιλαμβάνονται μερικά ενδεικτικά αποτελέσματα εξαγωγής των αντίστοιχων κινούμενων ακμών. Στο πείραμα αυτό χρησιμοποιήσαμε μόνο το κριτήριο της επιμονής των ακμών για την ενημέρωση φόντου (βλ. Ενότητα 4.4) και επεξεργασία μόνο στην αρχική ανάλυση χωρίς να εκμεταλλευτούμε την Ενότητα 4.8. Ειδικότερα, τα Σχήματα 4.3(α,δ,ζ) περιέχουν το ενημερωμένο φόντο που λαμβάνουμε βασιζόμενοι στο κριτήριο της επιμονής των ακμών για τα καρέ των Σχημάτων 4.2(α,δ,ζ), τα Σχήματα 4.3(β,ε,η) περιέχουν τις μεταβατικές ακμές που εξάγονται και τα Σχήματα 4.3(γ,στ,θ) περιέχουν τις τελικές μεταβατικές ακμές μετά από συνέλιξη με έναν πυρήνα εξασθένισης ώστε να μειώσουμε το 'θόρυβο'. Μερικά ενδιαφέροντα σχόλια μπορούν να γίνουν σε αυτό το σημείο. Αρχικά, με απλή παρατήρηση των Σχημάτων 4.3(γ,στ,θ), μπορεί κανείς να πει ότι η ποσότητα του 'θορύβου' (περιβαλλοντικός θόρυβος ή κίνηση ασήμαντων

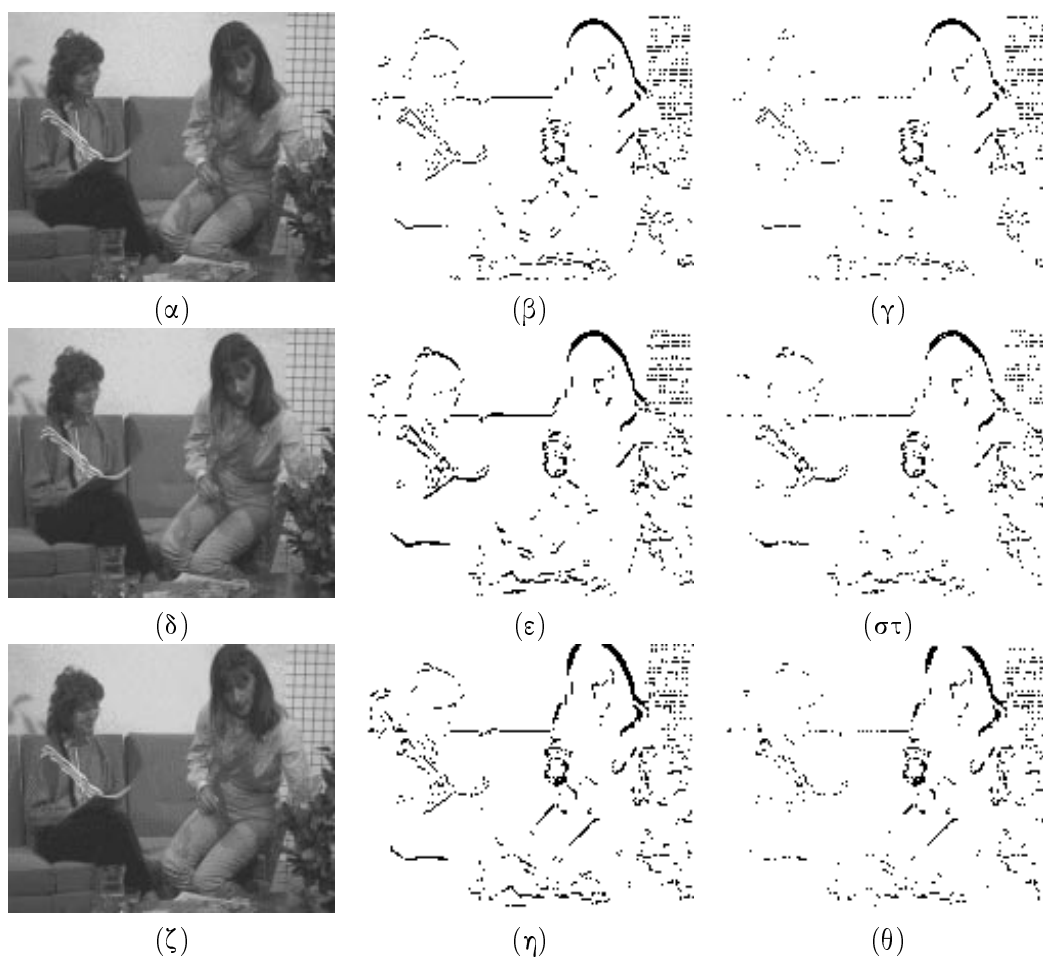


**Σχήμα 4.4** Βελτίωση των αποτελεσμάτων συνδυάζοντας τις κινούμενες ακμές σε δύο αναλύσεις για την ακολουθία των Σχημάτων 4.2, 4.3, (α,δ,ζ) οι κινούμενες ακμές σε χαμηλότερη ανάλυση, (β,ε,η) οι κινούμενες ακμές που λαμβάνονται συνδυάζοντας τις δύο αναλύσεις, (γ,στ,θ) οι κινούμενες ακμές μετά από συνέλιξη με φίλτρο εξασθένισης.

αντικειμένων) μειώνεται με το χρόνο. Αυτό είναι άμεση συνέπεια της τεχνικής ενημέρωσης φόντου που χρησιμοποιούμε, αφού όπως ο χρόνος περνάει οι μη επίμονες ακμές ανατίθενται στο φόντο. Επιπρόσθετα, μπορούμε να παρατηρήσουμε ότι το ποσό του θορύβου που υπεισέρχεται στις μεταβατικές ακμές και που απεικονίζεται στα Σχήματα 4.3(β,ε,η) δεν είναι αποδεκτό, υπό την έννοια ότι πυκνές ομάδες από μεταβατικά pixels μπορούν να οδηγήσουν σε λανθασμένους συναγερμούς. Την ίδια στιγμή, λόγω κυρίως της ελαφράς περιστροφικής κίνησης της κάμερας, μέρη από τις ευθείες ακμές του φόντου 'επιβιώνουν' στην ενημέρωση φόντου. Αστοχίες στην τεχνική εκτίμησης κίνησης σε επίπεδο pixel μπορούν να οδηγήσουν επίσης σε τέτοια λάθη.

Στο Σχήμα 4.4 επιδεικνύουμε τη δραματική βελτίωση των αποτελεσμάτων, όταν οι μεταβατικές ακμές  $T_1$  του πρωτοτύπου συνδυάζονται με τις μεταβατικές ακμές  $T_{1/2}$  σε μία χαμηλότερη ανάλυση, που λαμβάνονται σύμφωνα με την Ενότητα 4.10. Τα Σχήματα 4.4(α,δ,ζ) απεικονίζουν τις μεταβατικές ακμές που λαμβάνονται στη χαμηλότερη ανάλυση και αντιστοι-

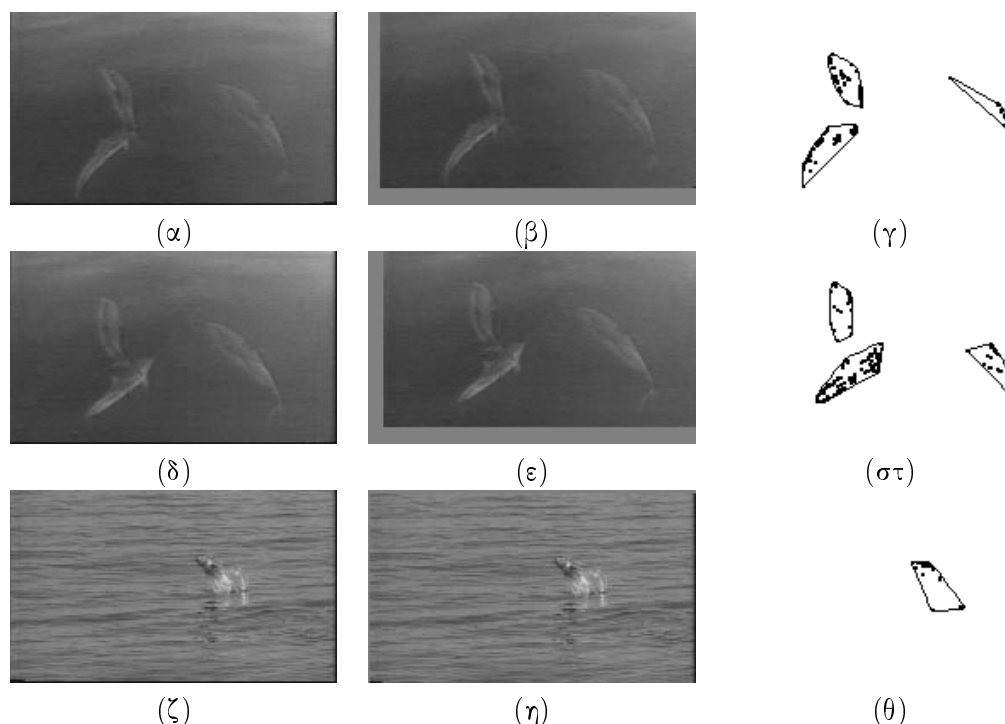




**Σχήμα 4.5** Βελτίωση των αποτελεσμάτων χρησιμοποιώντας και τα δύο κριτήρια ενημέρωσης του φόντου (επιμονή ακμών και αλλαγή κατεύθυνσης της φωτεινότητας), (α,δ,ζ) τρία μη διαδοχικά καρέ, (β,ε,η) κινούμενες ακμές χρησιμοποιώντας το κριτήριο της επιμονής ακμών για την ενημέρωση του φόντου, (γ,στ,θ) κινούμενες ακμές χρησιμοποιώντας και τα δύο κριτήρια.

χούν ακριβώς στην αρχική ανάλυση των Σχημάτων 4.3(β,ε,η). Συνδυάζοντας αυτούς τους χάρτες ακμών στα Σχήματα 4.4(β,ε,η), μπορεί κανείς να παρατηρήσει ότι ο περιβαλλοντικός θόρυβος που παράγεται από τη βλάστηση μειώνεται σε πολύ μεγάλο βαθμό και οι λανθασμένες ‘επίμονες’ ακμές του περιβάλλοντος σχεδόν εξαφανίζονται. Την ίδια στιγμή, οι ακμές στο όριο του κινούμενου αντικειμένου δεν επηρεάζονται σημαντικά. Με αυτή τη λογική, μετά από εξασθένιση ακμών, τα Σχήματα 4.4(γ,στ,θ) περιέχουν σχεδόν αποκλειστικά ακμές που βρίσκονται στο περίγραμμα του κινούμενου αντικειμένου.

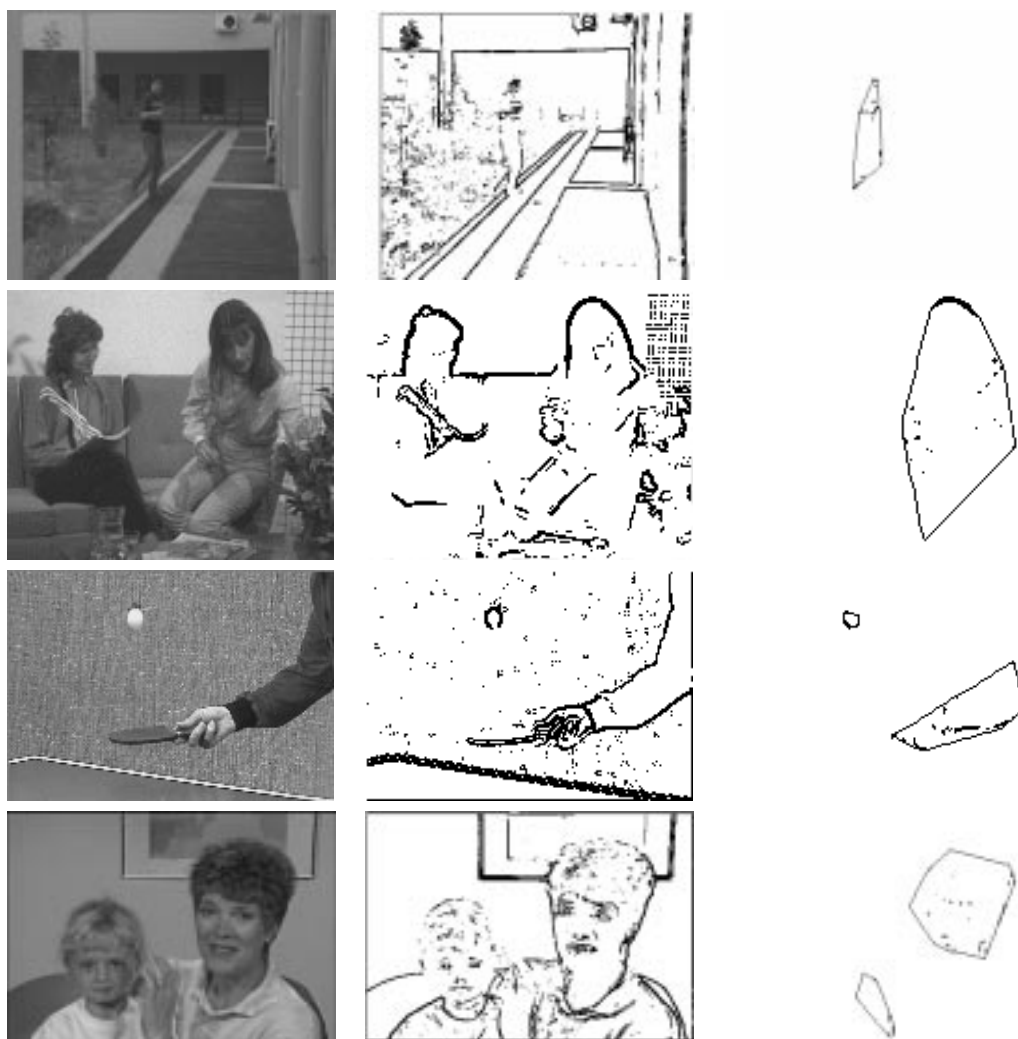
Στο Σχήμα 4.5 έχουμε συμπεριλάβει ένα χαρακτηριστικό παράδειγμα όπου η κίνηση της κάμερας είναι μικρή και υποφέρει από κραδασμούς που οδηγούν σε φτωχά αποτελέσματα ευθυγράμμισης του φόντου στην κανονικοποίηση κίνησης και συνακόλουθα στην ενημέρωση φόντου. Στην περίπτωση αυτή, αναμένεται ότι μερικές από τις δεσπόζουσες ακμές του φόντου θα επιβιώσουν στην ενημέρωση φόντου και θα εμφανιστούν ως μεταβατικές. Αυτό μπορεί να



**Σχήμα 4.6** Ταξινόμηση κινούμενων ακμών και εξαγωγή των οριακών πολυγώνων για μία ακολουθία που περιέχει περισσότερα από ένα κινούμενα αντικείμενα, εκμεταλλευόμενοι όλα τα βήματα της προτεινόμενης μεθοδολογίας, (α,δ,ζ) τρία μη διαδοχικά καρέ, (β,ε,η) κανονικοποιημένα ως προς την κίνηση καρέ ως προς διαφορετική αναφορά, (γ,στ,θ) οριακά πολύγωνα των κινούμενων αντικειμένων που εντοπίστηκαν.

επαληθευθεί στις μεταβατικές ακμές των Σχημάτων 4.5(β,ε,η) που αντιστοιχούν στα καρέ των Σχημάτων 4.5(α,δ,ζ). Στη μονάδα ενημέρωσης φόντου, χρησιμοποιήσαμε ξανά το κριτήριο της επιμονής των ακμών. Η συνέλιξη με ένα φίλτρο εξασθένισης στους χάρτες αυτούς θα έδινε μερικές μεταβατικές ακμές στο όριο του καναπέ και του ανθρώπινου σώματος που βρίσκεται στα αριστερά, πέρα από τις (σωστές) μεταβατικές γραμμές στο κυρίως κινούμενο αντικείμενο στα δεξιά. Στα Σχήματα 4.5(γ,στ,θ) μπορεί κανείς να παρατηρήσει πώς η χρήση του κριτηρίου της αλλαγής της κατεύθυνσης της κλίσης της φωτεινότητας της Ενότητας 4.4, από κοινού με το κριτήριο της επιμονής των ακμών, μπορεί να οδηγήσει στην αναγνώριση των μη μεταβατικών ακμών. Στα Σχήματα 4.5(γ,στ,θ) μόνο οι μεταβατικές ακμές του κύριου κινούμενου αντικειμένου επιμένουν, εφόσον γύρω από τις ακμές του καναπέ και του ανθρώπινου σώματος στα αριστερά η κλίση της φωτεινότητας δεν αλλάζει σημαντικά με το χρόνο.

Το Σχήμα 4.6 επιδεικνύει μία περίπτωση όπου περισσότερα από ένα κινούμενα αντικείμενα υπάρχουν στην ακολουθία, συγκεκριμένα μερικά κινούμενα δελφίνια σε μία σειρά ντοκιμανταίρ. Στην περίπτωση αυτή, η τεχνική ταξινόμησης εξάγει επιτυχώς τα σώματα των δελφινιών, δεδομένων των διανυσμάτων θέσης και κίνησης των μεταβατικών ακμών σύμφωνα με τις Ενότητες 4.6 και 4.7. Αν και τα κινούμενα αντικείμενα μπορούν να επικαλύπτονται



**Σχήμα 4.7** Ενδεικτικά αποτελέσματα που λαμβάνονται χρησιμοποιώντας την προτεινόμενη μέθοδο, (αριστερά) καρέ μετά από κανονικοποίηση κίνησης, (μέση) εξαγόμενες ακμές, (δεξιά) κινούμενες ακμές και οριακά πολύγωνα των κινούμενων αντικειμένων που εντοπίστηκαν.

μεταξύ τους, ο αλγόριθμος εξάγει επιτυχώς τις διαχωρίσιμες σιλουέτες - σε μία τέτοια περίπτωση όμως οι μεταβατικές ακμές κοντά στα όρια μπορούν να αποδοθούν σε οποιοδήποτε από τα κινούμενα αντικείμενα βάσει των ταχυτήτων τους. Οι αλλαγές σκηνής αντιμετωπίζονται από την επαναρχικοποίηση του καρέ αναφοράς και της επακόλουθης τεχνικής ενημέρωσης φόντου, κάθε τεχνική εντοπισμού σκηνών (scene detection) θα ήταν όμως χρήσιμη σ' αυτήν την περίπτωση.

Μερικά ενδεικτικά τελικά αποτελέσματα στην επίδοση του αλγορίθμου, όπως αυτός δίνεται στην Ενότητα 4.10, έχουν συμπεριληφθεί στο Σχήμα 4.7. Είναι ενδιαφέρον να παρατηρήσει κανείς, όπως είναι και αναμενόμενο, ότι ο αλγόριθμος εντοπίζει τα κινούμενα μέρη μεγάλων αντικειμένων που κινούνται ανεξάρτητα ως ανεξάρτητα κινούμενα αντικείμενα, αφού δεν εκμεταλλεύεται την πληροφορία διασυνδεσιμότητάς τους.

Ο προτεινόμενος αλγόριθμος έχει ελεγχθεί σε αντιπαραβολή με δημοφιλείς αλγορίθμους στην περιοχή του εντοπισμού και χωροθέτησης κινούμενων αντικειμένων, επιδεικνύοντας κατά κανόνα καλύτερη επίδοση σε ταχύτητα. Συγκρινόμενος με αλγορίθμους που εφαρμόζουν προχωρημένες τεχνικές εκτίμησης κίνησης (όπως πολυδιακριτικές ή επαναληπτικές πιθανοτικές τεχνικές) για να λάβουν πυκνά πεδία κίνησης για κάθε μετάβαση, η προτεινόμενη μέθοδος βασίζεται σε απλό ταίριασμα μπλοκ και σε εντοπισμό ακμών παρουσιάζοντας έτσι πολύ χαμηλό υπολογιστικό κόστος. Επιπρόσθετα, σε αντίθεση με τις τεχνικές κατάτμησης κίνησης και ταξινόμησης διανυσμάτων κίνησης, η προτεινόμενη μέθοδος βασίζεται μόνο σε ενημέρωση φόντου, η οποία με τη σειρά της υπολογίζει απλώς τις διαφορές των pixels ακμών, μία δυαδική πράξη με ασήμαντο κόστος. Με αυτόν τον τρόπο, η μέθοδος που παρουσιάστηκε επιβεβαιώνεται ότι είναι σημαντικά ταχύτερη. Συγκρινόμενη με τεχνικές που βασίζονται σε παρακολούθηση κίνησης ή παρακολούθηση περιγράμματος για να αποφύγουν τον υπολογισμό πυκνών πεδίων διανυσμάτων κίνησης για όλες τις μεταβάσεις, η μέθοδος αποδεικνύεται ανώτερη στην περίπτωση σημαντικής παραμόρφωσης του κινούμενου αντικειμένου ή μερικής επικάλυψης, αφού η ενημέρωση του φόντου πραγματοποιείται σε ένα μικρό παράθυρο που κινείται στο χρόνο. Εξετάζοντας τη θεωρητική θεμελίωση του αλγορίθμου, θα περίμενε κανείς ο αλγόριθμος να αποτυγχάνει όταν ισχυρές περιστροφές επιβάλλονται στην κινούμενη κάμερα. Έχει δειχθεί όμως πειραματικά ότι ο αλγόριθμος αποτυγχάνει μόνο στην περίπτωση μεγάλων περιστροφών, όπου ακόμη και αλγόριθμοι που περιλαμβάνουν παραμετρικά μοντέλα κίνησης φαίνονται να αποτυγχάνουν.

Από την άλλη πλευρά, η προτεινόμενη μέθοδος δεν είναι κατάλληλη όταν αναζητάμε ακριβή περιγράμματα αντικειμένων, αντί για το οριακό πολύγωνο ή παραλληλόγραμμο. Στην περίπτωση αυτή, ο αλγόριθμος που παρουσιάζεται μπορεί να χρησιμοποιηθεί για να λάβουμε μία άριστη αρχική εκτίμηση για τεχνικές που χρησιμοποιούν ενεργά περιγράμματα και μοντέλα διάδοσης περιγράμματος.

#### 4.12 Συμπεράσματα και επεκτάσεις

Στο κεφάλαιο αυτό προτείνεται ένας ταχύς αλγόριθμος εντοπισμού και χωροθέτησης κινούμενων αντικειμένων. Η συγκεκριμένη μέθοδος βασίζεται σε ιδέες ενημέρωσης φόντου οι οποίες, παράλληλα με μία απλή τεχνική κανονικοποίησης κίνησης, οδηγούν στην επιτυχή εξαγωγή του οριακού πολυγώνου των κύριων κινούμενων αντικειμένων.

Ο αλγόριθμος αποτελεί μία ικανοποιητική εναλλακτική σε τεχνικές που χρησιμοποιούν κατάτμηση βασισμένη σε κίνηση, ιδιαίτερα σε εφαρμογές όπου το υπολογιστικό κόστος είναι

κρίσιμο και δεν απαιτείται ακρίβεια σε επίπεδο pixel. Ταυτόχρονα, τα αποτελέσματα που λαμβάνονται μπορούν να χρησιμοποιηθούν ως άριστη αρχικοποίηση για περίπλοκες τεχνικές παρακολούθησης περιγράμματος, αντιμετωπίζοντας τα συνήθη προβλήματα των τελευταίων. Κατάλληλη ρύθμιση των παραμέτρων που αφήνονται ανοιχτές στην προτεινόμενη προσέγγιση επιτρέπουν την εφαρμογή του αλγορίθμου σε σχεδόν οποιαδήποτε φυσική ακολουθία.

Η συγκεκριμένη τεχνική, όπως αναλύθηκε σε αυτό το κεφάλαιο, ανοίγει τα παρακάτω άμεσα ζητήματα επεκτάσεων:

- (1) τη μελέτη της εισαγωγής της ως ένα υποσύστημα αρχικοποίησης σε μεθόδους παρακολούθησης περιγράμματος,
- (2) τη μοντελοποίηση της πληροφορίας των εξαγόμενων πολυγώνων για σκοπούς σύγκρισης και ταξινόμησης στα πλαίσια του MPEG-7,
- (3) τη βελτίωση των λαμβανόμενων αποτελεσμάτων εισάγοντας περιορισμούς 'διάδοσης' με το χρόνο στα οριακά πολύγωνα.

## ΚΕΦΑΛΑΙΟ 5

### Κανονικοποίηση Καμπυλών για Μοντελοποίηση Αντικειμένων με βάση το Σχήμα

#### 5.1 Εισαγωγή / Αναδρομή

Όπως αναλύσαμε και στο Κεφάλαιο 3, λόγω του αυξανόμενου ενδιαφέροντος σε εφαρμογές πολυμέσων τα τελευταία χρόνια ισχυροποιείται η απαίτηση για αποδοτική αποθήκευση, χειρισμό και αναζήτηση υλικού σε βάσεις δεδομένων πολυμέσων. Το τελευταίο έχει μάλιστα αποκτήσει μεγάλη σημασία μετά τις οδηγίες των πρόσφατων προτύπων κωδικοποίησης MPEG-4 και MPEG-7 της ομάδας MPEG (Moving Pictures Expert Group) [66, 67, 142]. Οι δυνατότητες υποβολής ερωτήσεων, δεικτοδότησης και ανάκτησης οπτικοακουστικού υλικού με βάση το περιεχόμενο είναι μεγάλης σημασίας στην αναζήτηση σε βάσεις δεδομένων ψηφιακής εικόνας και βίντεο, λόγω του μεγάλου όγκου πληροφορίας που αυτές περιέχουν [185, 5, 29]. Αρκετά πρότυπα συστήματα έχουν υλοποιηθεί δίνοντας δυνατότητες αναζήτησης με βάση το περιεχόμενο, περιλαμβάνοντας για παράδειγμα τα VIRAGE [51], QBIC [37], Photobook [116], VisualSEEk [145], Netra [86], MARS [131], VideoQ [20] και πολλά άλλα.

Η πληροφορία περιεχομένου σε συστήματα αναζήτησης με βάση το περιεχόμενο συνήθως μοντελοποιείται βάσει χαρακτηριστικών σχετικά χαμηλού επιπέδου, όπως η κατανομή χρώματος και υψής [185, 6], τα πεδία διανυσμάτων κίνησης και οι χάρτες βάθους (depth maps) [30], ή/και χαρακτηριστικά σχήματος [195]. Χαρακτηριστικά υψηλότερου επιπέδου όπως ‘σημαντικά’ αντικείμενα μπορούν να ληφθούν με κατάλληλη επεξεργασία χαμηλού επιπέδου χαρακτηριστικών, περισσότερο στα πλαίσια συγκεκριμένων εφαρμογών. Όπως αναφέρεται στην [117], εάν η κύρια πληροφορία περιγραφής ή ταξινόμησης ενός αντικειμένου μπορεί να βρεθεί στο περίγραμμά του, είναι φυσικό να διατηρηθεί μόνο το περίγραμμα για περαιτέρω επεξεργασία. Τέτοιες περιπτώσεις μπορούν να εμφανιστούν για παράδειγμα στην ταξινόμηση σιλουετών αεροπλάνων ή δορυφόρων, στην αναγνώριση χαρακτήρων και στην αυτόματη επεξεργασία κειμένου [70]. Γενικά, όμως, η μελέτη του σχήματος με στόχο την ταξινόμηση, αναγνώριση και ανάκτηση τυχαίων αντικειμένων, από μόνη της ή σε συνδυασμό με άλλα χαρακτηριστικά του αντικειμένου, είναι ένα ενεργό πεδίο επιστημονικής έρευνας [80, 153, 198].

Υπάρχουν δύο κυρίως λόγοι για το αυξανόμενο ενδιαφέρον στην ανάλυση σχήματος. Πρώτον, το σχήμα ενός αντικειμένου μπορεί να παρέχει σημαντικό εργαλείο για την ανάκτηση οπτικής πληροφορίας (εικόνας), μέσω ενός μηχανισμού αναζήτησης-με-σκίτσο (query-by-sketch) [13], όπου πρότυπα σχεδιασμένα από το χρήστη χρησιμοποιούνται για σύγκριση με τα σχήματα

των αντικειμένων που περιέχονται στη βάση δεδομένων. Η διαδικασία ταυτοποίησης εφαρμόζεται μεταξύ του σκίτσου του χρήστη και των σχημάτων της βάσης. Δεύτερον, λειτουργίες αναζήτησης με βάση το περιεχόμενο ενσωματώνονται στα νέα πρότυπα κωδικοποίησης πληροφορίας πολυμέσων (βλ. ενδεικτικά [67]). Σ' αυτή την κατεύθυνση, τεχνικές κωδικοποίησης δεύτερης γενιάς έχουν προταθεί [160], όπου η κωδικοποίηση βίντεο βασίζεται σε κατάτμηση εικόνας (segmentation) και επιτρέπει το χειρισμό αντικειμένων με βάση το περιεχόμενο [133]. Με τον τρόπο αυτό, η πληροφορία σχήματος ενσωματώνεται στα επίπεδα αντικειμένων βίντεο (VOPs - video object planes) στη μορφή δυαδικών ακολουθιών εικόνων και μπορεί να χρησιμοποιηθεί για πρόβλεψη κατάτμησης εικόνας με σημαντικές σχετικές εφαρμογές (partition interpolation ή extrapolation) [90].

Αρκετές μέθοδοι έχουν προταθεί στη βιβλιογραφία για ανάλυση σχήματος, για μοντελοποίηση και αναπαράστασή του, βασιζόμενες σε κωδικοποίηση αλυσίδας (chain coding) [39], πολυγωνική προσέγγιση (polygonal approximation) [112], μετασχηματισμό διάμεσου άξονα / σκελετού (medial axis transform / skeleton) [17], περιγραφείς Fourier (Fourier descriptors) [117], ροπές καμπυλών (curve moments) [58], B-splines [21], χώρους καμπυλότητας (curvature scale spaces) [98], σημεία ενδιαφέροντος (interest points) [171], μετασχηματισμό συνημιτόνου (sinusoidal transform) [123], περιγραφείς Legendre (Legendre descriptors) και ροπές Zernike (Zernike moments) [75]. Βασικά, οι περισσότερες προσεγγίσεις εκμεταλλεύονται γεωμετρικά χαρακτηριστικά των καμπυλών, είτε ολικά (όπως ροπές, μήκη, κύριους άξονες, επιμήκυνση ή κλειστότητα) είτε τοπικά (όπως σημεία ενδιαφέροντος, μετρήσεις καμπυλότητας, έμμεσα πολυώνυμα) ώστε να επιτύχουν ταυτοποίηση σχήματος, αναγνώριση ή ταξινόμηση. Όποια κι αν είναι η εφαρμογή, όλες οι μέθοδοι ανάλυσης σχήματος μοιράζονται ένα βασικό κοινό πρόβλημα: οι προβολές ενός αντικειμένου μπορούν να διαφέρουν δραματικά όταν το σημείο παρακολούθησης (viewpoint) αλλάζει, λόγω και του προοπτικού μετασχηματισμού. Οι περισσότερες εργασίες προσεγγίζουν την αλλαγή του σημείου παρακολούθησης με έναν μετασχηματισμό affine, ο οποίος είναι μία αρκετά καλή προσέγγιση όταν το αντικείμενο είναι μακριά από την κάμερα, οπότε και μία μικρή αλλαγή λόγω προοπτικής προβολής μπορεί να ιδωθεί ως κάποιου είδους παραμόρφωση. Με στόχο να αποφευχθεί η αποθήκευση πολλών διαφορετικών πρότυπων σχημάτων που να διαφέρουν μεταξύ τους κατά έναν μετασχηματισμό affine - περιέχοντα παράγοντες περιστροφής (rotation), μετακίνησης (translation) και μεγέθυνσης (skew / scaling) - και η ταυτοποίηση του μόλις εξηγμένου σχήματος με όλα τα πρότυπα αυτά, θα πρέπει να οριστούν ποσότητες αμετάβλητες ως προς τον μετασχηματισμό affine ή αμετάβλητες-σε-affine (affine invariants), δηλαδή ιδιότητες που θα παραμένουν σταθερές υπό τυχαίους affine μετασχηματισμούς.

Μία ιδιότητα που έχουν κοινή οι περισσότερες αμετάβλητες-σε-affine μέθοδοι στην βιβλιογραφία είναι ότι οι αμετάβλητες ποσότητες χρησιμοποιούνται κατά τη διαδικασία ταυτοποίησης, αναγνώρισης ή υπολογισμού των κριτηρίων ομοιότητας. Για παράδειγμα, μία μετρική ομοιότητας αμετάβλητη σε περιστροφή, μετακίνηση και μεγέθυνση βασισμένη σε συναρτήσεις στροφής (turning functions) για σύγκριση πολυγωνικών σχημάτων έχει προταθεί στο [3] και μία μετρική βασισμένη σε τροποποιημένους περιγραφείς Fourier παρουσιάστηκε στο [130]. Για τον ίδιο σκοπό, αμετάβλητες ροπών χρησιμοποιούνται στο [12]. Η αναγνώριση αντικειμένων με χρήση κανονικοποιημένων περιγραφών Fourier και νευρωνικά δίκτυα έχει περιγραφεί στο [173], ενώ γενετικοί αλγόριθμοι για αμετάβλητη-σε-affine αναγνώριση σχήματος έχουν προταθεί στο [164]. Στις τεχνικές που βασίζονται σε τοπικά χαρακτηριστικά των καμπυλών περιλαμβάνονται τοπικές αμετάβλητες στην κλίμακα του γκρι βασισμένες σε αυτόματα εντοπιζόμενα σημεία ενδιαφέροντος για ανάκτηση εικόνων [135], αμετάβλητες-σε-affine βασισμένες σε κυρτούς φλοιούς (convex hulls) για εγγραφή εικόνων (image registration) [198], καθώς και αμετάβλητες σε τοπικές παραμορφώσεις για αναγνώριση καμπυλών με έμμεσα πολυώνυμα (implicit polynomials) [127]. Μία άλλη προσέγγιση είναι η ταυτοποίηση δύο δεδομένων καμπυλών με βέλτιστο υπολογισμό των παραμέτρων affine που μεγιστοποιούν την ομοιότητα μεταξύ τους. Η βελτιστοποίηση αυτή βασίζεται για παράδειγμα σε ροπές καμπυλών [62] ή περιγραφείς Fourier [117].

Το κύριο μειονέκτημα της πρώτης προσέγγισης - χρήση των αμετάβλητων στην διαδικασία ταυτοποίησης ή αναγνώρισης - είναι ότι στις περισσότερες περιπτώσεις κάποια πληροφορία για την αρχική καμπύλη χάνεται. Εντωμεταξύ, η δεύτερη προσέγγιση - υπολογισμός των παραμέτρων affine μεταξύ δύο εκδοχών ενός αντικειμένου - απαιτεί εκ των προτέρων γνώση και των δύο εκδοχών, οπότε μπορεί να χρησιμοποιηθεί μόνο για ταυτοποίηση ενός συγκεκριμένου ζεύγους καμπυλών και όχι για παράδειγμα για αναγνώριση με χρήση ενός νευρωνικού δικτύου ή άλλου μέσου ευθείας ταξινόμησης. Επίσης, συνήθως απαιτεί πολύ μεγάλο υπολογιστικό κόστος. Λαμβάνοντας υπόψη όλα τα παραπάνω, η μέθοδος της κανονικοποίησης (normalization) έχει πρόσφατα προταθεί ως μία εναλλακτική στο πρόβλημα αυτό. Μια εικόνα ή καμπύλη μπορεί να κανονικοποιηθεί σε μία 'σταθερή' θέση / εκδοχή, η οποία ορίζεται με τέτοιο τρόπο ώστε όλοι οι μετασχηματισμοί affine του ιδίου αντικειμένου να κανονικοποιούνται στην ίδια εκδοχή. Εκτός από τις παραμέτρους του affine μετασχηματισμού, στις οποίες η κανονικοποίηση είναι αμετάβλητη, καμία άλλη πληροφορία δεν χάνεται. Η διαδικασία κανονικοποίησης αποτελείται στην πραγματικότητα από έναν affine (γραμμικό) μετασχηματισμό και έτσι το σχήμα της αρχικής καμπύλης παραμένει στην πραγματικότητα αμετάβλητο. Μία γενική διαδι-



κασία κανονικοποίησης για τον υπολογισμό αμετάβλητων ποσοτήτων δίνεται στο [129], ενώ η κανονικοποίηση εικόνων αντιμετωπίζεται στο [141].

Ταυτόχρονα, προσεγγίσεις έχουν προταθεί στη βιβλιογραφία για ταυτοποίηση καμπυλών από τυχαίες παραμορφώσεις, βασισμένες σε παραμορφώσιμα πρότυπα (deformable templates) [13]. Τα παραμορφώσιμα πρότυπα λαμβάνονται πραγματοποιώντας παραμετρικούς μετασχηματισμούς στην αρχική καμπύλη ενώ η παραμορφωσιμότητα των προτύπων εξασφαλίζεται με πιθανοτικό τρόπο [69]. Τα μοντέλα ενεργών περιγραμμάτων (active contour models / snakes) είναι επίσης κατάλληλα γι' αυτό το σκοπό [77]. Αν και τα παραμορφώσιμα πρότυπα αντιμετωπίζουν με επιτυχία το θόρυβο στις εικόνες και τις τοπικές παραμορφώσεις των καμπυλών λόγω τοπικών ανομοιοτήτων ή ακόμα και μερικής απόκρυψης (occlusion), οι τοπικές παραμορφώσεις μπορούν συχνά να 'μπερδευτούν' με συνολικές αλλαγές σχήματος λόγω κίνησης του στερεού ή μετασχηματισμών affine. Μ' αυτό το σκεπτικό, η επίδοση των αλγορίθμων αυτών είναι περιορισμένη. Παρόλ' αυτά, βελτιωμένα αποτελέσματα μπορούν να ληφθούν, όταν χρησιμοποιούνται ως δεύτερο βήμα, μετά την κανονικοποίηση.

Στο κεφάλαιο αυτό, παρουσιάζεται μία νέα μέθοδος κανονικοποίησης δισδιάστατων ( $2\Delta$ ) καμπυλών στην περίπτωση της προσέγγισης με affine μετασχηματισμούς, δίνοντας τη δυνατότητα εξαγωγής μίας αμετάβλητης-σε-affine αναπαράστασης καμπυλών χωρίς πραγματικά να χάνεται καμία πληροφορία σχήματος της αρχικής καμπύλης. Στην περίπτωση κλειστών καμπυλών, η αναπαράσταση είναι επίσης αμετάβλητη στο σημείο έναρξης (starting point). Ειδικότερα, μία  $2\Delta$  κλειστή καμπύλη που αναπαριστά το περίγραμμα εντός αντικειμένου αρχικά μοντελοποιείται με μία κυβική B-spline, ούτως ώστε το σχήμα απλοποιείται και ο θόρυβος κατάτμησης μειώνεται, ενώ ομοιόμορφη δειγματοληψία της καμπύλης βάσει του μήκους τόξου λαμβάνεται επαναυπολογίζοντας τα σημεία δεσμών (knot points) της B-spline. Εν συνεχεία, η δειγματοληπτημένη καμπύλη κανονικοποιείται σε διαδοχικά βήματα, ώστε να εξαφανίσουμε μετασχηματισμούς μετακίνησης, μεγέθυνσης, σημείου έναρξης, περιστροφής και ανάκλασης (reflection). Η κανονικοποίηση βασίζεται σε έναν συνδυασμό χαρακτηριστικών της καμπύλης που περιλαμβάνει χωρικές ροπές και περιγραφείς Fourier. Όλα τα χαρακτηριστικά υπολογίζονται από όλα τα διαθέσιμα σημεία επί της καμπύλης όποτε δεν χρησιμοποιείται τοπική πληροφορία. Το υπολογιστικό κόστος είναι αμελητέο, ούτως ώστε η μέθοδος να μπορεί εύκολα να συμπεριληφθεί σε ένα πραγματικού χρόνου σύστημα ανάκτησης εικόνων ή κωδικοποίησης βίντεο. Αποδεικνύεται ότι κάθε κανονικοποιημένη καμπύλη αντιστοιχεί μοναδικά σε ένα σύνολο καμπυλών που σχετίζονται μέσω ενός τυχαίου affine μετασχηματισμού. Επίσης, η κανονικοποιημένη αναπαράσταση μαζί με τις υπολογιζόμενες παραμέτρους affine, που σχετίζουν την πρότυπη καμπύλη με την κανονικοποιημένη εκδοχή της, περιγράφουν ακριβώς την

πρότυπη καμπύλη, αφού αυτή μπορεί να ανακατασκευαστεί από αυτά τα δεδομένα ακριβώς. Βασιζόμενοι στις παραπάνω ιδιότητες πληρότητας και μοναδικότητας, η ανάλυση καμπυλών στην κανονικοποιημένη εκδοχή τους και στις σχετικές affine παραμέτρους είναι δυνατή. Έτσι, η προτεινόμενη μέθοδος μπορεί να εφαρμοστεί ως ένα βήμα προ-επεξεργασίας σε κάθε τεχνική αναπαράστασης σχήματος, ταξινόμησης, αναγνώρισης ή ανάκτησης, εφόσον αποδοτικά διαχωρίζει το πρόβλημα της αμετάβλητης σε affine περιγραφής από την εξαγωγή χαρακτηριστικών και ταυτοποίηση προτύπων. Αρκετές δημοφιλείς μετρικές ομοιότητας καμπυλών χρησιμοποιήθηκαν για να επιδείξουν την ικανότητα της προτεινόμενης αναπαράστασης να διατηρεί όλη την πληροφορία της καμπύλης πέραν των τυχαίων μετασχηματισμών affine, στους οποίους είναι αμετάβλητη. Σε όλες τις περιπτώσεις, δείχνεται πειραματικά ότι η μέθοδος είναι σημαντικά εύρωστη στο θόρυβο και στις παραμορφώσεις σχήματος.

Η προτεινόμενη τεχνική δόθηκε στην εργασία [8], ενώ βασίζεται σε παλαιότερες σχετικές εργασίες στην κανονικοποίηση περιγραμμάτων αντικειμένων [9, 7] και στη μοντελοποίηση καμπυλών [195]. Το κεφάλαιο έχει οργανωθεί ως εξής: Η Ενότητα 5.2 εισάγει τις απαραίτητες λεπτομέρειες / ιδιότητες του περιβάλλοντος όπου η προτεινόμενη μέθοδος μπορεί να εφαρμοστεί, τις παραδοχές που έγιναν και γνωστούς περιορισμούς. Η Ενότητα 5.3 παρουσιάζει τη διαδικασία μοντελοποίησης καμπύλης με B-splines, καθώς και την εξαγωγή ομοιόμορφων δειγμάτων βάσει των σημείων δεσμών. Η Ενότητα 5.4 περιγράφει τη διαδικασία κανονικοποίησης καμπύλης που αναπτύχθηκε για την εξαφάνιση των παραμέτρων μετακίνησης και μεγέθυνσης, ενώ η υπόλοιπη διαδικασία κανονικοποίησης ως προς το σημείο έναρξης, την περιστροφή και την ανάκλαση δίνεται στην Ενότητα 5.5. Τέλος, πειράματα σε αρκετές φυσικές και συνθετικές ακολουθίες εικόνων και βίντεο δίνονται στην Ενότητα 5.6 για την αποτίμηση της επίδοσης, της απόδοσης και της ευρωστίας της προτεινόμενης διαδικασίας κανονικοποίησης. Τα σχετικά συμπεράσματα αλλά και τα στενά συνδεδεμένα ανοιχτά ζητήματα έρευνας αναλύονται στην Ενότητα 5.7.

## 5.2 Περιγραφή του προβλήματος

Στη συνέχεια, υποθέτουμε ότι το σχήμα του περιγράμματος ενός αντικειμένου είναι διαθέσιμο και αναπαριστάται από ένα σύνολο διατεταγμένων σημείων, φτιάχνοντας μία επίπεδη και κλειστή 2D επιφάνεια. Το σύνολο των δειγματοληπτούμενων σημείων λαμβάνεται από τα δεδομένα της εικόνας με αυτόματη ή χειροκίνητη κατάτμησή της (image segmentation). Στην πράξη, οποιοσδήποτε αλγόριθμος κατάτμησης μπορεί να εφαρμοστεί, βασιζόμενος για παράδειγμα στην ομοιογένεια χρώματος ή κίνησης, σε εξαγωγή ακμών ή σε μορφολογικά

εργαλεία [132]. Μία πολυδιακριτική (multiresolution) υλοποίηση του αλγορίθμου Recursive Shortest Spanning Tree (RSST), ο M-RSST αλγόριθμος [10], χρησιμοποιήθηκε για κατάτμηση χρώματος στα πειράματα αυτού του κεφαλαίου, όπως φαίνεται και στην Ενότητα 5.6. Στην περίπτωση ακολουθιών βίντεο, κάθε ακολουθία πρώτα αναλύεται σε σκηνές (video shots) οι οποίες αντιστοιχούν σε συνεχή λειτουργία της κάμερας. Στη συνέχεια, χαρακτηριστικά καρέ (key-frames) εξάγονται από τις σκηνές και η κατάτμηση εφαρμόζεται στα καρέ αυτά του βίντεο. Στην περίπτωση αυτή, η κατάτμηση διευκολύνεται με την εκμετάλλευση της πληροφορίας κίνησης. Συγκεκριμένα, 2Δ παραμετρικά μοντέλα κίνησης χρησιμοποιήθηκαν για κατάτμηση κίνησης (motion segmentation) [101, 155], καθώς και τεχνικές εξαγωγής κύριων κινούμενων αντικειμένων (main mobile object detection) [194]. Αν και η κανονικοποίηση ως προς το σημείο έναρξης, που παρουσιάζεται στην Ενότητα 5.5, εφαρμόζεται μόνο σε κλειστές καμπύλες, τα υπόλοιπα βήματα κανονικοποίησης ισχύουν και για ανοικτές καμπύλες.

Θεωρείται επίσης ότι οι καμπύλες αντιστοιχούν σε μη επικαλυπτόμενα περιγράμματα αντικειμένων τα οποία είναι εξολοκλήρου γνωστά. Το τελευταίο μπορεί να οδηγήσει σε προβλήματα που σχετίζονται με επικαλύψεις αντικειμένων (object occlusion), τα οποία μπορούν να ξεπεραστούν μόνο μέσω τοπικών αμετάβλητων (local invariants), ενώ στην προτεινόμενη μέθοδο μόνο ολικά (global) χαρακτηριστικά χρησιμοποιούνται. Δυστυχώς, δεν υπάρχει γνωστή από τη βιβλιογραφία μέθοδος, η οποία να διατηρεί όλη την πληροφορία της καμπύλης, ενώ ταυτόχρονα να αντιμετωπίζει κατάλληλα και τα προβλήματα επικάλυψης, αν και σημαντικές προσπάθειες έχουν καταβληθεί - βλ. για παράδειγμα την [62]. Στην ειδική περίπτωση που η επικάλυψη είναι αρκετά 'μικρή', μπορεί να θεωρηθεί ως μία μικρή αλλοίωση του σχήματος (deformation) και να αντιμετωπιστεί στο στάδιο της σύγκρισης καμπυλών, αμέσως μετά την κανονικοποίηση.

Κάθε καμπύλη εισόδου υποτίθεται ότι υποβάλλεται σε δύο τύπους μετασχηματισμών: μετασχηματισμούς παραμέτρων και συντεταγμένων. Οι μετασχηματισμοί παραμέτρων οφείλονται στο γεγονός ότι οι καμπύλες λαμβάνονται μέσω κατάτμησης εικόνας, οπότε ενέχεται διακριτοποίηση συνεχών αντικειμένων, με αναπόφευκτο αποτέλεσμα κάποιο θόρυβο κατάτμησης, καθώς και μη ομοιόμορφη δειγματοληψία ως προς μήκος τόξου. Επίσης, εφόσον τα περιγράμματα αντικειμένων αντιστοιχούν σε κλειστές καμπύλες, ένα ιδεατό σημείο που κινείται πάνω στο περίγραμμα δημιουργεί συντεταγμένες καμπύλης, οι οποίες είναι περιοδικές συναρτήσεις του μήκους τόξου και ένα τυχαίο σημείο έναρξης μπορεί να επιλεγεί για την περιγραφή μίας περιόδου. Επιπρόσθετα, ο προσανατολισμός της καμπύλης μπορεί να είναι είτε σύμφωνος είτε αντίθετος με τη φορά του ρολογιού. Από την άλλη μεριά, οι μετασχηματισμοί συντεταγμένων οφείλονται στο γεγονός ότι οι εικόνες παράγονται από προβολή τρισδιάστατων (3Δ) αντικει-

μένων στο  $2\Delta$  επίπεδο, οδηγώντας στον μη γραμμικό προοπτικό μετασχηματισμό (perspective transformation). Υποθέτοντας ότι ένα αντικείμενο είναι αρκετά μακριά από την κάμερα, ο τελευταίος μπορεί να προσεγγιστεί από έναν γραμμικό μετασχηματισμό affine. Το πρόβλημα είναι επομένως να κανονικοποιηθεί μία καμπύλη και να εξαχθεί μία αναπαράστασή της η οποία να είναι αμετάβλητη και σε παραμετρικούς μετασχηματισμούς και σε μετασχηματισμούς συντεταγμένων, ταυτόχρονα όμως να διατηρεί όλη την πληροφορία της αρχικής καμπύλης.

Οι μετασχηματισμοί παραμέτρων, λόγω της ανομοιόμορφης δειγματοληψίας και του θορύβου κατάτμησης, αντιμετωπίζονται μέσω μοντελοποίησης των καμπυλών με B-splines, όπως περιγράφεται στην επόμενη ενότητα. Μέρος της πληροφορίας χάνεται στην πραγματικότητα σ' αυτήν την διαδικασία, αφού αυτό είναι απαραίτητο για την απόρριψη του θορύβου και την απλοποίηση της καμπύλης, αντίθετα όμως τα επόμενα βήματα κανονικοποίησης είναι απολύτως αναστρέψιμα. Οι μετασχηματισμοί συντεταγμένων αναλύονται σε μετασχηματισμούς μετακίνησης (translation), μεγέθυνσης (skew, scaling), περιστροφής (rotation) και κατοπτρισμού (reflection). Μία διαδικασία ορθογωνοποίησης της καμπύλης προτείνεται για την απόρριψη της μετακίνησης και της μεγέθυνσης βάσει των ροπών της (curve moments). Στη συνέχεια, ακολουθεί μία διαδικασία κανονικοποίησης, με χρήση περιγραφέων Fourier (Fourier descriptors) για την κανονικοποίηση του σημείου έναρξης, της περιστροφής, του κατοπτρισμού και του προσανατολισμού.

### 5.3 Αναπαράσταση με χρήση B-splines

#### 5.3.1 Μοντελοποίηση καμπύλης

Οι B-splines έχουν χρησιμοποιηθεί ευρέως στην ανάλυση και μοντελοποίηση σχήματος, εφόσον έχουν έναν αριθμό από σημαντικές ιδιότητες όπως η ομαλότητα και η συνέχεια, η τοπική ελεγχσιμότητα και το αναλλοίωτο σε μετασχηματισμούς affine [21]. Σ' αυτό το κεφάλαιο, οι B-splines χρησιμοποιούνται με στόχο τη λήψη μίας συνεχούς και ομαλής αναπαράστασης καμπυλών, οι οποίες είναι αρχικά διαθέσιμες ως σύνολα σημείων, τα οποία με τη σειρά τους έχουν γενικά ληφθεί με μη ομοιόμορφη δειγματοληψία. Στην περίπτωση μας τέτοια σύνολα δεδομένων δίνονται από μία μονάδα κατάτμησης κίνησης ή/και χρώματος. Στη συνέχεια της ενότητας αυτής θα αναφερθούμε σύντομα στη γνωστή θεωρία προσέγγισης συνόλων σημείων με καμπύλες B-splines.

Οι κυβικές B-splines είναι σύνθετες καμπύλες που αποτελούνται γενικά από έναν μεγάλο αριθμό από τμήματα καμπύλης με  $C^2$  συνέχεια στα σημεία των συνδέσεων / τομών. Γενικά μία B-spline  $k$ -τάξης είναι  $C^{k-1}$  συνεχής. Στη συνέχεια θα αναφερόμαστε στην περίπτωση

των κυβικών B-splines. Υποθέτοντας ότι μία κλειστή κυβική B-spline αποτελείται από  $n + 1$  συνδεδεμένα τμήματα καμπύλης  $\mathbf{r}_i$  με  $i = 0, 1, \dots, n$  και  $\mathbf{r}_i(u) = (x_i(u), y_i(u))$ , κάθε ένα από τα τμήματα αυτά είναι ένας γραμμικός συνδυασμός τεσσάρων κυβικών πολυωνύμων, γνωστών και ως συναρτήσεις βάσης (basis functions), στην παράμετρο  $u$  στο  $[0, 1]$ :

$$\mathbf{r}_i(u) = \mathbf{C}_{i-1} Q_0(u) + \mathbf{C}_i Q_1(u) + \mathbf{C}_{i+1} Q_2(u) + \mathbf{C}_{i+2} Q_3(u) , \quad (5-1)$$

για  $i = 0, 1, \dots, n$  όπου

$$Q_k(u) = a_{k0} u^3 + a_{k1} u^2 + a_{k2} u + a_{k3} , \quad (5-2)$$

για  $k = 0, 1, 2, 3$ . Μπορεί να δει κανείς ότι ολόκληρη η B-spline  $\mathbf{r}$ , αποτελούμενη από  $n + 1$  συνδεδεμένα τμήματα καμπύλης  $\mathbf{r}_i$ , χαρακτηρίζεται από  $n + 1$  παραμέτρους, δηλαδή τα σημεία ελέγχου  $\mathbf{C}_i$  (control points).

Ούτως ώστε να ικανοποιηθούν οι  $C^0$ ,  $C^1$  και  $C^2$  περιορισμοί συνέχειας στα σημεία σύνδεσης μεταξύ τμημάτων, 15 εξισώσεις με τα άγνωστα  $a_k$  μπορούν να γραφούν (για παράδειγμα  $\mathbf{r}_i(1) = \mathbf{r}_{i+1}(0)$  για  $C^0$  συνέχεια ή συνέχεια θέσεως). Ένας επιπλέον περιορισμός, συγκεκριμένα η ιδιότητα του αναλλοίωτου στους μετασχηματισμούς συντεταγμένων  $\sum_{k=0}^3 Q_k(u) = 1$  με  $u$  στο  $[0, 1]$ , επιτρέπει τέλος τον υπολογισμό και των 16 άγνωστων παραμέτρων  $a_k$ . Με τον τρόπο αυτόν, ορίζονται οι συναρτήσεις βάσεως  $Q_k(u)$ .

Αφού βρεθούν οι συναρτήσεις βάσης, η παραμετροποίηση του συνόλου της καμπύλης είναι απαραίτητη ώστε να περιγραφεί η B-spline. Θεωρούμε τότε τη μεταβλητή  $u'$  ορισμένη στο διάστημα  $[0, n + 1]$ . Τότε για το  $i$ -οστό τμήμα,  $u' = u + i$  με  $u$  στο  $[0, 1]$ . Η καμπύλη B-spline δίνεται τότε βάσει των τμημάτων της από την

$$\mathbf{r}(u') = \sum_{i=0}^n \mathbf{r}_i(u) = \sum_{i=0}^n \mathbf{r}_i(u' - i) , \quad (5-3)$$

όπου το  $\mathbf{r}_i(u' - i)$  είναι μη μηδενικό μόνο για  $u$  στο  $[0, 1]$  ή αντίστοιχα  $u'$  στο  $[i, i + 1]$ . Χρησιμοποιώντας την εξίσωση (5-1), η εξίσωση (5-3) μπορεί να γραφεί με 'βολικότερο' τρόπο ως

$$\mathbf{r}(u') = \sum_{i=0}^{n+3} \mathbf{C}_i N_i(u') , \quad (5-4)$$

όπου τα  $\mathbf{C}_i$  ορίζονται για  $i = 0, 1, \dots, n$  και  $\mathbf{C}_{-1} = \mathbf{C}_n$ ,  $\mathbf{C}_{n+1} = \mathbf{C}_0$ ,  $\mathbf{C}_{n+2} = \mathbf{C}_1$ ,  $\mathbf{C}_{n+3} = \mathbf{C}_2$ . Με  $N_i(u')$  στην παραπάνω εξίσωση συμβολίζουμε τις συναρτήσεις εναρμόνισης ή συνδυασμού (blending functions) [174]:

$$N_i(u') = \begin{cases} Q_3(u' - i + 3), & i - 3 \leq u' < i - 2 \\ Q_2(u' - i + 2), & i - 2 \leq u' < i - 1 \\ Q_1(u' - i + 1), & i - 1 \leq u' < i \\ Q_0(u' - i), & i \leq u' < i + 1 \\ 0, & \text{αλλιώς} \end{cases} , \quad (5-5)$$

Παράλληλα με τα σημεία ελέγχου (control points), τα σημεία σύνδεσης (knot points) ορίζονται ως τα σημεία τομής  $\mathbf{p}_i$  μεταξύ των τμημάτων της καμπύλης. Γενικά,  $\mathbf{p}_i = \mathbf{r}_i(0) = \mathbf{r}_{i-1}(1)$ . Με δεδομένα τα σημεία ελέγχου, τα σημεία σύνδεσης μπορούν να υπολογιστούν μοναδικά, θεωρώντας ομοιόμορφη κατανομή των δεσμών, ως

$$\mathbf{p}_i = \frac{1}{6}\mathbf{C}_{i-1} + \frac{2}{3}\mathbf{C}_i + \frac{1}{6}\mathbf{C}_{i+1}, \quad (5-6)$$

για  $i = 0, 1, \dots, n$ . Βάσει της παραπάνω θεωρητικής ανάλυσης, μπορεί να παρατηρηθεί ότι διαφορετικά ζεύγη από σημεία ελέγχου και σύνδεσης μπορούν να περιγράφουν την ίδια B-spline καμπύλη.

Με δεδομένο ένα ‘πυκνό’ σύνολο από  $m$  σημεία  $\mathbf{s}_j$ , με  $j = 0, 1, \dots, m-1$  που ανήκουν σε μία άγνωστη καμπύλη, έχοντας ως στόχο να προσαρμόσουμε μία κατάλληλη B-spline, τα σημεία ελέγχου  $\mathbf{C}_i$  πρέπει να υπολογιστούν. Η προσέγγιση που ακολουθείται σ’ αυτό το κεφάλαιο εντοπίζει μία προσεγγιστική B-spline, τέτοια ώστε το σφάλμα μεταξύ των παρατηρούμενων δεδομένων και της εξαχθείσας B-spline να είναι ελάχιστο. Έτσι, η ποσότητα

$$d^2 = \sum_{j=0}^{m-1} \|\mathbf{s}_j - \mathbf{r}(u'_j)\|^2, \quad (5-7)$$

όπου  $u'_j$  κατάλληλες παραμετρικές τιμές του  $u'$ , πρέπει να ελαχιστοποιηθεί. Εάν κατάλληλες παραμετρικές τιμές του  $u'$  σημειώνονταν στην καμπύλη, τότε η MMSE λύση για τα σημεία ελέγχου θα δινόταν σε μορφή πίνακα από τη σχέση

$$\mathbf{C}_f = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{f}, \quad (5-8)$$

όπου τα  $\mathbf{f}$  και  $\mathbf{C}_f$  είναι μεγέθους  $m \times 2$  και  $(n+1) \times 2$  αντίστοιχα, περιέχοντας τα δεδομένα σημεία  $\mathbf{s}_j$  και τα σημεία ελέγχου  $\mathbf{C}_i$  αντίστοιχα. Ο  $m \times (n+1)$  πίνακας  $\mathbf{P}$  περιέχει κατάλληλες τιμές για τις συναρτήσεις εναρμόνισης, υπολογισμένες στα σημεία  $\mathbf{r}(u'_j)$ ,

$$\mathbf{P} = \begin{bmatrix} N_0(u'_0) + N_{n+1}(u'_0) & N_0(u'_1) + N_{n+1}(u'_1) & \cdots & N_0(u'_{m-1}) + N_{n+1}(u'_{m-1}) \\ N_1(u'_0) + N_{n+2}(u'_0) & N_1(u'_1) + N_{n+2}(u'_1) & \cdots & N_1(u'_{m-1}) + N_{n+2}(u'_{m-1}) \\ N_2(u'_0) + N_{n+3}(u'_0) & N_2(u'_1) + N_{n+3}(u'_1) & \cdots & N_2(u'_{m-1}) + N_{n+3}(u'_{m-1}) \\ N_3(u'_0) & N_3(u'_1) & \cdots & N_3(u'_{m-1}) \\ \vdots & \vdots & \cdots & \vdots \\ N_n(u'_0) & N_n(u'_1) & \cdots & N_n(u'_{m-1}) \end{bmatrix}^T. \quad (5-9)$$

**Απόδειξη:** Για οποιοδήποτε σημείο  $\mathbf{s}_j$  της κυβικής B-spline  $\mathbf{r}(u')$  ισχύει  $\mathbf{r}(u'_j) = \mathbf{s}_j$  ή

$$\begin{aligned} \mathbf{s}_j &= \sum_{i=0}^{n+3} \mathbf{C}_i N_i(u'_j) = \mathbf{C}_0(N_0(u'_j) + N_{n+1}(u'_j)) + \mathbf{C}_1(N_1(u'_j) + N_{n+2}(u'_j)) + \\ &+ \mathbf{C}_2(N_2(u'_j) + N_{n+3}(u'_j)) + \sum_{i=3}^n \mathbf{C}_i N_i(u'_j) \end{aligned}$$

Ισοδύναμα σε μορφή πίνακα

$$\begin{bmatrix} N_0(u'_j) + N_{n+1}(u'_j) & N_1(u'_j) + N_{n+2}(u'_j) & N_2(u'_j) + N_{n+3}(u'_j) & N_3(u'_j) & \cdots \\ \cdots & N_n(u'_j) \end{bmatrix} \mathbf{C}_f = \mathbf{s}_j$$

Γράφοντας  $m$  τέτοιες εξισώσεις, για όλα τα διαθέσιμα σημεία στην καμπύλη λαμβάνουμε

$$\mathbf{P} \mathbf{C}_f = \mathbf{f}$$

όπου ο  $\mathbf{P}$  δίνεται από την εξίσωση (5-9). Η εξίσωση (5-8) δίνει τότε την MMSE λύση για τα σημεία ελέγχου. *O.E.Δ.*

Για την εκχώρηση των παραμετρικών τιμών του  $u'$ , χρησιμοποιείται η μέθοδος του μήκους χορδής (chord length μέθοδος ή CL μέθοδος). Συγκεκριμένα, για  $u'_1 = 0$  και  $u'_{\max} = n - 2$ , το  $u'_j$  το σχετιζόμενο με το δεδομένο σημείο  $\mathbf{s}_j$  υπολογίζεται από τη σχέση

$$u'_j = u'_{j-1} + u'_{\max} \frac{\|\mathbf{s}_j - \mathbf{s}_{j-1}\|}{\sum_{k=2}^m \|\mathbf{s}_k - \mathbf{s}_{k-1}\|} \quad (5-10)$$

όπου  $j = 0, 1, \dots, m - 1$ .

Η μέθοδος CL βασίζεται στο γεγονός ότι το μήκος της χορδής μεταξύ δύο σημείων είναι μία πολύ ακριβής προσέγγιση του μήκους του τόξου της καμπύλης, υπό την υπόθεση ότι ένα μόριο κινείται επάνω στην καμπύλη με σταθερή ταχύτητα. Είναι αρκετά εύρωστη σε ομοιόμορφα κατανεμημένο θόρυβο, αλλά υποφέρει από ανομοιόμορφα κατανεμημένο θόρυβο και ανομοιόμορφη δειγματοληψία. Εναλλακτικά, η μέθοδος αντίστροφου μήκους χορδής (inverse chord length ή ICL) μπορεί να χρησιμοποιηθεί με καλύτερα αποτελέσματα στη μη ομοιόμορφη περίπτωση όπως αναφέρεται στο [62].

### 5.3.2 Αναδιάταξη των σημείων σύνδεσης

Ας υποθέσουμε ότι ένα σύνολο από διαφορετικές καμπύλες (σύνολα σημείων για παράδειγμα) είναι διαθέσιμες σε μία βάση δεδομένων. Αφού έχουμε μοντελοποιήσει τα σύνολα αυτά σημείων με κλειστές κυβικές B-splines σύμφωνα με την προηγούμενη παράγραφο, μπορούμε να δούμε ότι τα σημεία ελέγχου τους δεν μπορούν να χρησιμοποιηθούν ως κριτήριο για την μεταξύ τους ομοιότητα, αφού γενικά διαφορετικά σύνολα σημείων ελέγχου μπορούν να περιγράφουν την ίδια καμπύλη. Γι' αυτό το λόγο, είναι βολικό για κάθε καμπύλη να εξάγουμε τα σημεία σύνδεσης  $\mathbf{p}_i$ ,  $i = 0, 1, \dots, n$ , χρησιμοποιώντας τα σημεία ελέγχου που έχουμε ήδη

εξάγει. Για κλειστές κυβικές B-splines αυτό πραγματοποιείται από την εξίσωση (5-6), ή σε μητρική μορφή

$$\mathbf{p}_f = \mathbf{A} \mathbf{C}_f \quad (5-11)$$

όπου  $\mathbf{p}_f$  είναι ο  $(n+1) \times 2$  πίνακας που περιέχει τα σημεία σύνδεσης και  $\mathbf{A}$  είναι ο  $(n+1) \times (n+1)$  κυκλικός πίνακας

$$\mathbf{A} = \begin{bmatrix} \frac{2}{3} & \frac{1}{6} & 0 & 0 & \cdots & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{6} & 0 & 0 & 0 & \cdots & \frac{1}{6} & \frac{2}{3} \end{bmatrix}. \quad (5-12)$$

Θα πρέπει να τονιστεί εδώ ότι τα σημεία σύνδεσης ανήκουν στην εξαγόμενη B-spline (δηλαδή η B-spline περνάει από τα σημεία αυτά). Όμως, μπορούμε να δούμε ότι για οποιεσδήποτε δύο καμπύλες, δεν είναι δεδομένο ότι τα εξαγόμενα σημεία σύνδεσής τους αντιστοιχούν ένα-προς-ένα, ακόμη κι αν είναι ίσα σε αριθμό. Γι' αυτό το λόγο θα πρέπει τα σημεία σύνδεσης να αναδιαταχθούν πάνω στις εξαγόμενες καμπύλες [21]. Συγκεκριμένα, τοποθετούμε  $l$  σημεία σύνδεσης σε ίσες αποστάσεις ως προς  $u'$  σε κάθε καμπύλη, με το πρώτο σημείο σε μία τυχαία θέση επάνω στην spline. Όταν το σωστό σημείο έναρξης (starting point) υπολογιστεί από τη διαδικασία κανονικοποίησης, πραγματοποιείται νέα αναδιάταξη, όπως θα δούμε και στη συνέχεια. Το τελευταίο μας εξασφαλίζει ότι για κάθε καμπύλη αποθηκευμένη στο σύστημά μας τα αναδιατεταγμένα σημεία σύνδεσης βρίσκονται σε αντιστοιχία.

Ένας ταξινομητής (classifier) των καμπυλών βάσει των αναδιατεταγμένων σημείων σύνδεσης θα μπορούσε να ελαχιστοποιεί κάποια μετρική όπως η

$$d^2 = \sum_{i=1}^l \|\mathbf{p}_i^{(a)} - \mathbf{p}_i^{(b)}\|^2, \quad (5-13)$$

όπου με  $(a)$  και  $(b)$  σημειώνονται οι  $a$ -οστή και  $b$ -οστή B-spline καμπύλες υποκείμενες σε σύγκριση. Άλλες μετρικές μπορούν επίσης να χρησιμοποιηθούν.

## 5.4 Ορθογωνοποίηση καμπύλης

Έως αυτό το σημείο, με τη χρήση των B-splines για την αναπαράσταση δισδιάστατων (επίπεδων) κλειστών καμπυλών έχουμε κατορθώσει να απλοποιήσουμε το σχήμα της καμπύλης, να μειώσουμε το θόρυβο κατάτμησης και να πάρουμε ομοιόμορφη δειγματοληψία σημείων πάνω στην καμπύλη ως προς τα μήκη τόξου. Μία μέθοδος ορθογωνοποίησης της καμπύλης προτείνεται σε αυτή την ενότητα, ως ένα πρώτο βήμα προς την κανονικοποίησή της. Η διαδικασία αυτή κανονικοποιεί την καμπύλη ως προς πιθανή μετατόπιση και μεγέθυνση και ουσιαστικά μειώνει



**Πίνακας 5.1** Πίνακας μετασχηματισμών καμπυλών και συναφών ποσοτήτων που περιλαμβάνει η προτεινόμενη διαδικασία κανονικοποίησης καμπύλης.

$n_a(.)$	μετασχηματισμός ‘ορθογωνοποίησης’ καμπύλης, ορισμός (5-19)
$N(.)$	πίνακας ‘ορθογωνοποίησης’ καμπύλης, ορισμός (5-19)
$S_m(.)$	συνάρτηση κυκλικής μετατόπισης σημείου έναρξης καμπύλης κατά $m$ δείγματα, ορισμός (5-28)
$n_p(.)$	μετασχηματισμός κυκλικής μετατόπισης καμπύλης, ορισμός (5-31) και (5-32)
$p(.)$	βασική κυκλική μετατόπιση καμπύλης, ορισμός (5-30)
$n_r(.)$	μετασχηματισμός κανονικοποίησης περιστροφής και κατοπτρισμού καμπύλης, ορισμός (5-39)
$r(.)$	γωνία στροφής καμπύλης, ορισμός (5-38)

το μετασχηματισμό affine της καμπύλης ως προς ένα πρότυπο σε έναν ορθογώνιο μετασχηματισμό, δηλαδή ένα μετασχηματισμό που περιλαμβάνει μόνο περιστροφή ή/και κατοπτρισμό. Στη συνέχεια του κεφαλαίου ορίζονται και χρησιμοποιούνται αρκετοί μετασχηματισμοί καμπυλών. Για ευκολία στην ανάγνωση, οι μετασχηματισμοί δίνονται περιληπτικά στον Πίνακα 5.1.

#### 5.4.1 Η διαδικασία ορθογωνοποίησης

Έστω  $\mathbf{s}_i = [x_i \ y_i]^T$ ,  $i = 0, 1, \dots, N-1$ , ένα σύνολο από  $N$  σημεία καμπύλης λαμβανόμενα μέσω της μοντελοποίησης καμπύλης με B-spline. Ένας πίνακας  $2 \times N$  της μορφής  $\mathbf{s} = [\mathbf{s}_0 \ \mathbf{s}_1 \ \dots \ \mathbf{s}_{N-1}]$  θα χρησιμοποιείται στη συνέχεια για την αναπαράσταση της καμπύλης, ώστε να επιτρέπει την απλούστερη αναπαράσταση μετασχηματισμών που περιλαμβάνουν πολλαπλασιασμούς με  $2 \times 2$  πίνακες. Όμοια, οι οριζόντιες και κατακόρυφες συντεταγμένες των σημείων θα αναπαρίστανται από τα  $1 \times N$  διανύσματα  $\mathbf{x} = [x_0 \ x_1 \ \dots \ x_{N-1}]$  και  $\mathbf{y} = [y_0 \ y_1 \ \dots \ y_{N-1}]$  αντίστοιχα ώστε  $\mathbf{s} = [\mathbf{x}^T \ \mathbf{y}^T]$ . Για κάθε καμπύλη  $\mathbf{s}$  οι  $(p, q)$ -τάξεως ροπές

$$m_{pq}(\mathbf{s}) = \frac{1}{N} \sum_{i=0}^{N-1} x_i^p y_i^q, \quad (5-14)$$

μέχρι τάξεως 2 χρησιμοποιούνται για την κατασκευή της αντίστοιχης κανονικοποιημένης καμπύλης  $n_a(\mathbf{s})$ . Χωρίς βλάβη της γενικότητας, υποτίθεται στη συνέχεια ότι η καμπύλη  $\mathbf{s}$  δεν είναι ένα γραμμικό τμήμα, οπότε πάντοτε  $m_{20}(\mathbf{s}) \neq 0$  και  $m_{02}(\mathbf{s}) \neq 0$ . Σ’αυτήν την περίπτωση τα βήματα της κανονικοποίησης που περιλαμβάνουν διαίρεση με τις παραπάνω ποσότητες πρέπει να παραλειφθούν.

Η διαδικασία ορθογωνοποίησης αποτελείται από ένα σύνολο γραμμικών πράξεων (μετατόπιση, μεγέθυνση και περιστροφή) που δεν εξαρτώνται από το επιλεγμένο σημείο έναρξης της κλειστής καμπύλης (ή, γενικότερα από τη σειρά των σημείων πάνω στην καμπύλη). Για

απλότητα, στην παρακάτω ανάλυση, πρόσθεση ή αφαίρεση μιας βαθμωτής ποσότητας από ένα διάνυσμα (ή από ένα διάνυσμα-γραμμή ενός πίνακα) υποδηλώνει πρόσθεση ή αφαίρεση της ποσότητας από όλα του τα στοιχεία. Σε τέτοιες περιπτώσεις ο πίνακας κατάλληλων διαστάσεων που παράγεται από τη βαθμωτή ποσότητα  $a$ , συμβολίζεται με  $[a]$ .

(α) Το κέντρο βάρους της καμπύλης κανονικοποιείται ώστε να συμπίπτει με την αρχή των αξόνων:

$$\mathbf{s}_1 = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{x} - [\mu_x] \\ \mathbf{y} - [\mu_y] \end{bmatrix}, \quad (5-15)$$

όπου  $\mu_x = m_{10}(\mathbf{s})$  και  $\mu_y = m_{01}(\mathbf{s})$ .

(β) Η καμπύλη μεγεθύνεται οριζόντια και κατακόρυφα ώστε οι δεύτερης τάξης ροπές της να γίνουν ίσες με τη μονάδα:

$$\mathbf{s}_2 = \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \sigma_x \mathbf{x}_1 \\ \sigma_y \mathbf{y}_1 \end{bmatrix}, \quad (5-16)$$

όπου  $\sigma_x = \frac{1}{\sqrt{m_{20}(\mathbf{s}_1)}}$  και  $\sigma_y = \frac{1}{\sqrt{m_{02}(\mathbf{s}_1)}}$ .

(γ) Η καμπύλη περιστρέφεται κατά  $\theta_0 = \frac{\pi}{4}$  αντίθετα με τη φορά του ρολογιού:

$$\mathbf{s}_3 = \mathbf{R}_{\frac{\pi}{4}} \mathbf{s}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{x}_2 - \mathbf{y}_2 \\ \mathbf{x}_2 + \mathbf{y}_2 \end{bmatrix}, \quad (5-17)$$

όπου με  $\mathbf{R}_\theta$  δηλώνεται ο  $2 \times 2$  πίνακας αντιστοιχών σε αντισωρολογιακή περιστροφή κατά  $\theta$  rad.

(δ) Τέλος, η καμπύλη μεγεθύνεται ξανά όπως στο βήμα (β):

$$\mathbf{s}_4 = \begin{bmatrix} \mathbf{x}_4 \\ \mathbf{y}_4 \end{bmatrix} = \begin{bmatrix} \tau_x \mathbf{x}_3 \\ \tau_y \mathbf{y}_3 \end{bmatrix}, \quad (5-18)$$

όπου  $\tau_x = \frac{1}{\sqrt{m_{20}(\mathbf{s}_3)}}$  και  $\tau_y = \frac{1}{\sqrt{m_{02}(\mathbf{s}_3)}}$ .

Η κανονικοποιημένη καμπύλη  $n_a(\mathbf{s}) \equiv \mathbf{s}_4$  μπορεί επίσης να γραφεί ως

$$n_a(\mathbf{s}) = \mathbf{N}(\mathbf{s}) (\mathbf{s} - [\mu(\mathbf{s})]) = \frac{1}{\sqrt{2}} \begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix} \begin{bmatrix} \mathbf{x} - [\mu_x] \\ \mathbf{y} - [\mu_y] \end{bmatrix}, \quad (5-19)$$

όπου  $\mu(\mathbf{s}) = [\mu_x \ \mu_y]^T$  και ο  $\mathbf{N}(\mathbf{s})$  συμβολίζει τον  $2 \times 2$  πίνακα ορθογωνοποίησης της  $\mathbf{s}$ . Αν και η εξάρτηση των παραμέτρων  $\mu_x, \mu_y, \sigma_x, \sigma_y, \tau_x, \tau_y$  από την  $\mathbf{s}$  παραλείπεται εδώ για χάριν απλότητας, ο πίνακας ορθογωνοποίησης είναι μία συνάρτηση του  $\mathbf{s}$ . Μπορεί κανείς να δει ότι η καμπύλη  $\mathbf{s}_2$  του βήματος (β) ικανοποιεί  $m_{10}(\mathbf{s}_2) = m_{01}(\mathbf{s}_2) = 0$  και  $m_{20}(\mathbf{s}_2) = m_{02}(\mathbf{s}_2) = 1$ . Τα επιπλέον βήματα περιστροφής και μεγέθυνσης (γ) και (δ) είναι αναγκαία ώστε  $m_{11}(n_a(\mathbf{s})) = 0$ . Η παρακάτω πρόταση δίνει τις αναγκαίες και ικανές συνθήκες για την ορθογωνοποίηση καμπύλης.

**Πρόταση 5.1:** Για κάθε αρχική καμπύλη  $\mathbf{s}$ , η κανονικοποιημένη καμπύλη  $n_a(\mathbf{s})$  όπως προκύπτει από τις εξ. (5-15)-(5-18) έχει τις παρακάτω ιδιότητες

$$\begin{aligned} m_{10}(n_a(\mathbf{s})) &= m_{01}(n_a(\mathbf{s})) = m_{11}(n_a(\mathbf{s})) = 0 , \\ m_{20}(n_a(\mathbf{s})) &= m_{02}(n_a(\mathbf{s})) = 1 . \end{aligned} \quad (5-20)$$

Επίσης, οι παραπάνω συνθήκες μπορούν να ληφθούν μόνον όταν η γωνία περιστροφής  $\theta_0$  στο βήμα κανονικοποίησης ( $\gamma$ ) είναι ίση με  $k \frac{\pi}{2} + \frac{\pi}{4}$ , με  $k$  να ανήκει στο  $Z$ .

**Απόδειξη:** Ο περιορισμός ότι το  $\mathbf{s}$  δεν αποτελεί ευθύ γραμμικό τμήμα είναι αναγκαίος ώστε με κάθε μετατόπιση, περιστροφή ή μεγέθυνση, η  $\mathbf{s}$  να μην βρίσκεται εξολοκλήρου στον  $x$  ή στον  $y$  άξονα, ώστε  $m_{20}(\mathbf{s}) \neq 0$ ,  $m_{02}(\mathbf{s}) \neq 0$  και όλες οι ποσότητες των εξισώσεων (5-15)-(5-18) να μπορούν να οριστούν. Μπορούμε να δούμε αμέσως ότι  $m_{10}(\mathbf{s}_1) = m_{01}(\mathbf{s}_1) = 0$ . Αυτή η ιδιότητα παραμένει επίσης και στο  $n_a(\mathbf{s}) \equiv \mathbf{s}_4$ , αφού τα ενδιάμεσα βήματα κανονικοποίησης περιλαμβάνουν μόνο περιστροφές και μεγεθύνσεις. Είναι επίσης φανερό ότι  $m_{20}(\mathbf{s}_2) = m_{02}(\mathbf{s}_2) = m_{20}(\mathbf{s}_4) = m_{02}(\mathbf{s}_4) = 1$ . Παρατηρείται τότε ότι

$$m_{11}(\mathbf{s}_3) = \mathbf{x}_3 \mathbf{y}_3^T = \frac{1}{2}(\mathbf{x}_2 - \mathbf{y}_2)(\mathbf{x}_2 + \mathbf{y}_2)^T = \frac{1}{2}(m_{20}(\mathbf{s}_2) - m_{02}(\mathbf{s}_2)) = 0 \quad (5-21)$$

οπότε

$$m_{11}(\mathbf{s}_4) = \frac{m_{11}(\mathbf{s}_3)}{\sqrt{m_{20}(\mathbf{s}_3) m_{02}(\mathbf{s}_3)}} = 0 ,$$

επομένως η  $n_a(\mathbf{s})$  ικανοποιεί τις συνθήκες της εξ. (5-20).

Εάν η γωνία  $\theta_0$  αντικατασταθεί από μία τυχαία γωνία  $\theta$  στην εξ. (5-17), τότε

$$\mathbf{s}_3 = \mathbf{R}_\theta \mathbf{s}_2 = \begin{bmatrix} \mathbf{x}_2 \cos\theta - \mathbf{y}_2 \sin\theta \\ \mathbf{x}_2 \sin\theta + \mathbf{y}_2 \cos\theta \end{bmatrix}$$

και η εξίσωση (5-21) μπορεί να γραφεί ως

$$m_{11}(\mathbf{s}_3) = \mathbf{x}_3 \mathbf{y}_3^T = \sin\theta \cos\theta (\mathbf{x}_2 \mathbf{x}_2^T - \mathbf{y}_2 \mathbf{y}_2^T) + (\cos^2\theta - \sin^2\theta) \mathbf{x}_2 \mathbf{y}_2^T = \cos(2\theta) \mathbf{x}_2 \mathbf{y}_2^T ,$$

οπότε για να είναι το  $m_{11}(\mathbf{s}_3)$  πάντα ίσο με το μηδέν για κάθε αρχική καμπύλη, το  $\theta$  θα πρέπει να ικανοποιεί  $\cos(2\theta) = 0$  και επομένως να λαμβάνει τιμές  $k \frac{\pi}{2} + \frac{\pi}{4}$ , με  $k$  να ανήκει στο  $Z$ . *Ο.Ε.Δ.*

### 5.4.2 Αναλλοίωτο σε μετατόπιση και μεγέθυνση

Ας θεωρήσουμε δύο καμπύλες  $\mathbf{s}, \mathbf{s}'$  συσχετιζόμενες μέσω ενός μετασχηματισμού affine

$$\mathbf{s}' = \mathbf{A} \mathbf{s} + \mathbf{t} \Rightarrow \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \quad (5-22)$$

όπου ο πίνακας  $\mathbf{A}$  υποτίθεται ότι είναι βαθμού 2 (full rank). Διαφορετικά, στην περίπτωση που  $\det \mathbf{A} = 0$ , οποιοδήποτε σύνολο σημείων μετασχηματίζεται σε μία ευθεία γραμμή και η κανονικοποίηση είναι αδύνατη. Τότε,  $\mu(\mathbf{s}') = \mathbf{A}\mu(\mathbf{s}) + \mathbf{t}$  ώστε  $\mathbf{s}'_1 = \mathbf{s}' - [\mu(\mathbf{s}')] = \mathbf{A}(\mathbf{s} - [\mu(\mathbf{s})]) = \mathbf{A} \mathbf{s}_1$ . Έτσι, μετά το βήμα (α) για κανονικοποίηση της μετατόπισης, οι καμπύλες  $\mathbf{s}_1$  και  $\mathbf{s}'_1$  συνδέονται απλώς μέσω ενός πολλαπλασιασμού με έναν  $2 \times 2$  πίνακα. Σε αυτήν την περίπτωση, η σχέση μεταξύ των ροπών των καμπυλών διαμορφώνεται ως εξής

$$\begin{aligned} m_{20}(\mathbf{s}'_1) &= a^2 m_{20}(\mathbf{s}_1) + b^2 m_{02}(\mathbf{s}_1) + 2ab m_{11}(\mathbf{s}_1) \\ m_{02}(\mathbf{s}'_1) &= c^2 m_{20}(\mathbf{s}_1) + d^2 m_{02}(\mathbf{s}_1) + 2cd m_{11}(\mathbf{s}_1) \\ m_{11}(\mathbf{s}'_1) &= ac m_{20}(\mathbf{s}_1) + bd m_{02}(\mathbf{s}_1) + (ad + bc) m_{11}(\mathbf{s}_1) \end{aligned} \quad (5-23)$$

Παρατηρείται στη συνέχεια ότι εάν η  $\mathbf{s}_1$  κανονικοποιηθεί, δηλαδή ικανοποιεί την (5-20), οι παραπάνω εξισώσεις απλοποιούνται στις

$$\begin{aligned} m_{20}(\mathbf{s}'_1) &= a^2 + b^2 \\ m_{02}(\mathbf{s}'_1) &= c^2 + d^2 \\ m_{11}(\mathbf{s}'_1) &= ac + bd \end{aligned} \quad (5-24)$$

Αυτό σημαίνει ότι για παράδειγμα όταν μία κανονικοποιημένη καμπύλη υποβληθεί σε περιστροφή ή/και κατοπτρισμό, οπότε ο  $\mathbf{A}$  είναι ορθογώνιος, αυτή παραμένει κανονικοποιημένη. Αντιστρόφως, εάν δύο καμπύλες είναι κανονικοποιημένες, τότε ο  $\mathbf{A}$  που τις συνδέει θα πρέπει να είναι ορθογώνιος, εάν αυτές προέρχονται βέβαια από την ίδια αρχική καμπύλη. Με αυτό το σκεπτικό, η παρακάτω πρόταση δίνει τη σχέση δύο καμπυλών βάσει των αντίστοιχών τους κανονικοποιημένων.

**Πρόταση 5.2:** Εάν δύο καμπύλες  $\mathbf{s}, \mathbf{s}'$  σχετίζονται μέσω ενός μετασχηματισμού affine, υπάρχει ένας ορθογώνιος  $2 \times 2$  πίνακας  $\mathbf{Q}$  τέτοιος ώστε:

$$n_a(\mathbf{s}') = \mathbf{Q} n_a(\mathbf{s}). \quad (5-25)$$

Επιπλέον, η ίδια σχέση ισχύει μεταξύ της  $n_a(\mathbf{s})$  και μίας οποιασδήποτε κανονικοποιημένης καμπύλης  $\tilde{n}_a(\mathbf{s})$  που να έχει προκύψει από μία οποιαδήποτε διαδικασία κανονικοποίησης σύμφωνα με τις ιδιότητες της εξ. (5-20).

**Απόδειξη:** Από την εξίσωση (5-19) έχουμε  $n_a(\mathbf{s}) = \mathbf{N}(\mathbf{s})(\mathbf{s} - [\mu(\mathbf{s})]) = \mathbf{N}(\mathbf{s}) \mathbf{s}_1$  και  $n_a(\mathbf{s}') = \mathbf{N}(\mathbf{s}') \mathbf{s}'_1 = \mathbf{N}(\mathbf{s}') \mathbf{A} \mathbf{s}_1$ . Εφόσον οι  $\mathbf{s}$  και  $\mathbf{s}'$  δεν αντιστοιχούν σε ευθύγραμμα τμήματα,  $\det \mathbf{N}(\mathbf{s}) = \frac{\sigma_x \sigma_y \tau_x \tau_y}{\sqrt{2}} \neq 0$  (ομοίως και για το  $\mathbf{N}(\mathbf{s}')$ ), οπότε οι πίνακες κανονικοποίησης είναι ομαλοί και μπορούμε να ορίσουμε

$$\mathbf{Q} = \mathbf{N}(\mathbf{s}') \mathbf{A} \{\mathbf{N}(\mathbf{s})\}^{-1} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

ούτως ώστε  $n_a(\mathbf{s}') = \mathbf{Q} n_a(\mathbf{s})$ . Τώρα, αφού τόσο η  $n_a(\mathbf{s}')$  όσο και η  $n_a(\mathbf{s})$  είναι κανονικοποιημένες, έχουμε από την (5-23)

$$m_{20}(n_a(\mathbf{s}')) = q_{11}^2 + q_{12}^2 = 1$$

$$m_{02}(n_a(\mathbf{s}')) = q_{21}^2 + q_{22}^2 = 1$$

$$m_{11}(n_a(\mathbf{s}')) = q_{11}q_{21} + q_{12}q_{22} = 0$$

Έτσι  $\mathbf{Q} \mathbf{Q}^T = \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_2$  και ο  $\mathbf{Q}$  είναι ορθομοναδιαίος πίνακας, δηλαδή οι  $n_a(\mathbf{s}')$  και  $n_a(\mathbf{s})$  διαφέρουν μόνο κατά ένα μετασχηματισμό περιστροφής (εάν  $\det \mathbf{Q} = 1$ ), συν ένα μετασχηματισμό κατοπτρισμού όταν  $\det \mathbf{Q} = -1$ .

Έστω τώρα ότι  $\tilde{\mathbf{N}}(\mathbf{s})$  είναι μία άλλη διαδικασία κανονικοποίησης για το  $\mathbf{s}$  επιστρέφοντας μία κανονικοποιημένη καμπύλη  $\tilde{n}_a(\mathbf{s}) = \tilde{\mathbf{N}}(\mathbf{s}) \mathbf{s}_1$ . Σ'αυτήν την περίπτωση μπορούμε να ορίσουμε  $\tilde{\mathbf{Q}} = \tilde{\mathbf{N}}(\mathbf{s})\{\mathbf{N}(\mathbf{s})\}^{-1}$  ώστε  $\tilde{n}_a(\mathbf{s}) = \tilde{\mathbf{Q}} n_a(\mathbf{s})$ , και εφόσον τόσο η  $\tilde{n}_a(\mathbf{s})$  όσο και η  $n_a(\mathbf{s})$  είναι κανονικοποιημένες, μπορούμε να συμπεράνουμε ότι ο  $\tilde{\mathbf{Q}}$  είναι ορθομοναδιαίος. Έτσι όλες οι διαδικασίες κανονικοποίησης που βασίζονται στις ιδιότητες της εξ. (5-20) οδηγούν σε μία καμπύλη που σχετίζεται με την προτεινόμενη  $n_a(\mathbf{s})$  με μία απλή περιστροφή (ή/και κατοπτρισμό). *Ο.Ε.Δ.*

Με αυτόν τον τρόπο, η προτεινόμενη μέθοδος κανονικοποίησης 'εκφυλλίζει' τους μετασχηματισμούς affine σε ορθογώνιους μετασχηματισμούς που μπορούν να περιέχουν μόνο περιστροφή ή/και κατοπτρισμό (βάσει του προσήμου του  $\det \mathbf{Q}$ ), και επομένως οι κανονικοποιημένες καμπύλες είναι αναλλοίωτες σε μετατόπιση και μεγέθυνση. Είναι πολύ σημαντικό να τονιστεί εδώ ότι η κανονικοποίηση πραγματοποιείται χωρίς γνώση των παραμέτρων affine  $\mathbf{A}$  και  $\mathbf{t}$  και χωρίς ένα-προς-ένα αντιστοίχιση μεταξύ των καμπυλών  $\mathbf{s}$  και  $\mathbf{s}'$ , δηλαδή το  $\mathbf{s}$  κανονικοποιείται χωρίς γνώση του  $\mathbf{s}'$  και αντιστρόφως. Παρατηρήστε επίσης ότι οι παράμετροι του μετασχηματισμού  $\{\mu_x, \mu_y, \sigma_x, \sigma_y, \tau_x, \tau_y\}$  υπολογίζονται απευθείας από τις ροπές πρώτης και δεύτερης τάξεως των δεδομένων της καμπύλης. Επίσης, μπορούμε να δούμε ότι αυτό το σύνολο παραμέτρων μαζί με την  $n_a(\mathbf{s})$  περιέχουν όλη την αναγκαία πληροφορία για την αρχική

καμπύλη  $\mathbf{s}$  - αν και η  $n_a(\mathbf{s})$  ικανοποιεί πέντε περιορισμούς, βλ. (5-20), οι παράμετροι είναι έξι αφού οι  $\tau_x$  και  $\tau_y$  συνδέονται με την  $\frac{1}{\tau_x^2} + \frac{1}{\tau_y^2} = 1$ . Από τα παραπάνω, η μοναδικότητα και η πληρότητα του  $n_a(\mathbf{s})$  επαληθεύονται, αφού αυτό περιέχει όλη την πληροφορία του  $\mathbf{s}$ , εκτός από το μετασχηματισμό ως προς τον οποίο είναι αναλλοίωτο. Το παραπάνω μπορεί να μας οδηγήσει σε ασφαλείς και ακριβείς συγκρίσεις καμπυλών για λόγους ταξινόμησης, αναγνώρισης ή αναζήτησης με βάση το περιεχόμενο (και συγκεκριμένα το σχήμα). Εφόσον η διαδικασία της κανονικοποίησης δεν εξαρτάται από μηχανισμούς σύγκρισης, οποιαδήποτε διαδικασία ταξινόμησης μπορεί να χρησιμοποιηθεί, συμπεριλαμβανομένων για παράδειγμα των νευρωνικών δικτύων (neural networks). Τέλος, όπως φαίνεται και στις πειραματικές μετρήσεις, το  $n_a(\mathbf{s})$  είναι ανθεκτικό στο θόρυβο (λόγω κατάτμησης ή δειγματοληψίας) και στις τοπικές διαφορές όμοιων καμπυλών, αφού βασίζεται σε 'ολιστικές' ιδιότητες των καμπυλών, όπως οι ροπές τους.

### 5.5 Κανονικοποίηση σημείου έναρξης και περιστροφής

Έως αυτό το σημείο, κατορθώσαμε να απλοποιήσουμε τους μετασχηματισμούς affine σε ορθογώνιους μετασχηματισμούς, ώστε να αρκεί η χρήση ενός επιπλέον μετασχηματισμού αναλλοίωτου σε περιστροφή και κατοπτρισμό. Μία υποθετική συνολική διαδικασία κανονικοποίησης (ορθογωνοποίηση και κανονικοποίηση ως προς περιστροφή και κατοπτρισμό) θα ήταν αναλλοίωτη γενικά σε μετασχηματισμούς affine. Μπορεί όμως να σημειωθεί ότι η τελευταία διαδικασία κανονικοποίησης εμπεριέχει πράξεις που εξαρτώνται από το σημείο έναρξης της κλειστής καμπύλης, και γενικότερα από την σειρά διάταξης των σημείων επάνω στην καμπύλη. Για παράδειγμα, η κανονικοποίηση της περιστροφής εξαρτάται από το σημείο έναρξης και η κανονικοποίηση κατοπτρισμού εξαρτάται από την περιστροφή. Γι'αυτό το λόγο η κανονικοποίηση του σημείου έναρξης πραγματοποιείται πρώτη, ενώ ακολουθούν η κανονικοποίηση περιστροφής και κατοπτρισμού.

Στην παρακάτω ανάλυση, η κανονικοποίηση βασίζεται στον διακριτό μετασχηματισμό Fourier, σε αντίθεση με τις έως τώρα χρησιμοποιούμενες ροπές καμπυλών. Γι'αυτό το λόγο, η διανυσματική μιγαδική αναπαράσταση της καμπύλης είναι προτιμότερη από την αναπαράσταση με πίνακες. Οι οριζόντιες και κατακόρυφες συντεταγμένες των σημείων στην καμπύλη αναπαριστώνται από τα  $N \times 1$  διανύσματα  $\mathbf{x} = [x_0 \ x_1 \ \cdots \ x_{N-1}]^T$  και  $\mathbf{y} = [y_0 \ y_1 \ \cdots \ y_{N-1}]^T$  αντίστοιχα, ενώ μία καμπύλη αναπαρίσταται από το μιγαδικό  $N \times 1$  διάνυσμα  $\mathbf{z} = \mathbf{x} + j \mathbf{y} = [z_0 \ z_1 \ \cdots \ z_{N-1}]^T$  όπου το  $z_i = x_i + j y_i$ ,  $i = 0, 1, \dots, N-1$  αντιστοιχεί σε ένα σημείο της καμπύλης.

### 5.5.1 Κανονικοποίηση σημείου έναρξης

Η διαδικασία κανονικοποίησης για το σημείο έναρξης βασίζεται στον διακριτό μετασχηματισμό Fourier του μιγαδικού διανύσματος  $\mathbf{z}$  που αναπαριστά μία καμπύλη:

$$\mathbf{u} = \mathcal{F}(\mathbf{z}) = \mathbf{W} \mathbf{z} = \begin{bmatrix} w^0 & w^0 & \cdots & w^0 \\ w^0 & w^1 & \cdots & w^{N-1} \\ \vdots & \vdots & & \vdots \\ w^0 & w^{N-1} & \cdots & w^{(N-1)^2} \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{N-1} \end{bmatrix} \quad (5-26)$$

όπου  $w = e^{j \frac{2\pi}{N}}$ , ώστε  $w^{\lambda N} = 1$ , με  $\lambda$  στο  $Z$ . Ισοδύναμα,

$$u_k = \sum_{i=0}^{N-1} z_i w^{-k i}, \quad k = 0, 1, \dots, N-1 \quad (5-27)$$

Για κάθε στοιχείο του μετασχηματισμού Fourier θα χρησιμοποιήσουμε τη φάση του (phase ή primary argument) ως  $a_k = \text{Arg } u_k \equiv \theta$  με  $\theta$  στο  $[0, 2\pi)$  ώστε  $u_k = r e^{j\theta}$  (και με  $r$  στο  $R^+$ ). Το αντίστοιχο διάνυσμα των φάσεων είναι τότε  $\mathbf{a} = \text{Arg } \mathbf{u} = [a_0 \ a_1 \ \cdots \ a_{N-1}]$ . Θεωρώντας τώρα μία δεύτερη καμπύλη  $\mathbf{z}' = [z'_0 \ z'_1 \ \cdots \ z'_{N-1}]^T$  ακριβώς όμοια με την  $\mathbf{z}$  διαφέροντας μόνο στο σημείο έναρξης, για παράδειγμα κυκλικά μετατοπισμένη ως προς τη  $\mathbf{z}$  κατά  $m$  δείγματα (με  $m$  στο  $\{0, 1, \dots, N-1\}$ ),

$$\mathbf{z}' = S_m(\mathbf{z}) : z'_i = z_{(i+m) \bmod N}, \quad i = 0, 1, \dots, N-1. \quad (5-28)$$

Εάν τότε ορίσουμε ομοίως  $\mathbf{u}' = \mathcal{F}(\mathbf{z}')$  και  $\mathbf{a}' = \text{Arg } \mathbf{u}'$ , μπορεί ναδειχθεί ότι  $u'_k = w^{k m} u_k$ , ή

$$a'_k = (a_k + 2\pi \frac{km}{N}) \bmod 2\pi \quad (5-29)$$

οπότε μία κυκλική μετατόπιση μεταξύ των  $\mathbf{z}$  και  $\mathbf{z}'$  μειώνεται σε μία απλή σχέση μεταξύ των  $\mathbf{a}$  και  $\mathbf{a}'$ . Εάν οι δύο καμπύλες  $\mathbf{z}$  και  $\mathbf{z}'$  ήταν γνωστές, η απευθείας σύγκριση μεταξύ τους θα ήταν δυνατή, αφού μετά τον υπολογισμό του  $m$  της εξίσωσης (5-29) θα αρκούσε μία κυκλική μετατόπιση στο  $\mathbf{z}'$ . Αυτό αντίθετα δεν είναι δυνατό σε μία τεχνική κανονικοποίησης, αφού μία καμπύλη είναι γνωστή σε κάθε στιγμή (δεν είναι δηλαδή μία συγκριτική διαδικασία). Ακόμα όμως κι όταν οι δύο καμπύλες είναι γνωστές, το  $m$  μπορεί να υπολογιστεί με απόλυτη ακρίβεια μόνο όταν αντιστοιχούν ακριβώς με στην ίδια αρχική καμπύλη (πέρα από το σημείο έναρξης). Σε κάθε άλλη περίπτωση, οι φάσεις του Fourier είναι τόσο θορυβώδεις που η (5-29) δεν δίνει ουσιαστικά καμία χρήσιμη πληροφορία. Γι'αυτούς τους λόγους είναι χρήσιμο να ορίσουμε μία βασική κυκλική μετατόπιση για κάθε καμπύλη, βασιζόμενοι στη διαφορά της πρώτης και της τελευταίας φάσης Fourier:

$$p(\mathbf{z}) = [\frac{N}{4\pi}(a_1 - a_{N-1})] \bmod \frac{N}{2} \quad (5-30)$$

και στη συνέχεια να χρησιμοποιήσουμε την αντίθετη μετατόπιση για να κανονικοποιήσουμε την καμπύλη

$$n_p(\mathbf{z}) = S_{-p(\mathbf{z})}(\mathbf{z}) \quad (5-31)$$

Είναι εμφανές από την (5-29) ότι η διαφορά μεταξύ δύο διαδοχικών στοιχείων του  $\mathbf{a}'$  (για παράδειγμα  $a'_2 - a'_1$ ) σχετίζεται με την αντίστοιχη διαφορά  $(a_2 - a_1)$  μέσω μίας απλής πρόσθεσης  $\frac{2\pi m}{N}$  modulo  $2\pi$ . Έτσι, εάν  $a_2 - a_1 \equiv 0$ , μπορούμε άμεσα να υπολογίσουμε το  $m$  διαιρώντας τη διαφορά με  $\frac{2\pi}{N}$  και παίρνοντας modulo  $N$ . Για την επιλογή όμως της πρώτης και της τελευταίας φάσης, όμως, όπως εξηγείται παρακάτω, τα στοιχεία  $a_1$  και  $a_{N-1}$  δεν είναι διαδοχικά - διαφέρουν κατά δύο δείγματα αφού οι συντελεστές Fourier είναι περιοδικοί με περίοδο  $N$ . Μία ακόμα διαίρεση με το 2 είναι αναγκαία λοιπόν σε αυτήν την περίπτωση. Τα παραπάνω εξηγούν τον ορισμό του  $p(\mathbf{z})$  στην εξ. (5-30), όπου το ακέραιο μέρος είναι αναγκαίο για να πραγματοποιηθεί επιτυχώς η κυκλική μετατόπιση στην (5-31).

Μπορεί να δειχθεί ότι ακόμη και όταν το  $N$  είναι άρτιος, η παραπάνω κανονικοποίηση είναι αναλλοίωτη ως προς το σημείο έναρξης, εκτός από μία αβεβαιότητα στη βασική κυκλική μετατόπιση, που μπορεί να δημιουργήσει μία ακόμη μετατόπιση κατά  $\frac{N}{2}$ . Εάν το  $N$  δεν είναι άρτιος, μία διαφορά κατά ένα περισσότερο/λιγότερο δείγμα μεταξύ  $n_p(\mathbf{z})$  και  $n_p(\mathbf{z}')$  μπορεί να προκύψει.

**Πρόταση 5.3:** Εάν δύο καμπύλες  $\mathbf{z}, \mathbf{z}'$  σχετίζονται μέσω ενός μετασχηματισμού κυκλικής μετατόπισης, τότε

$$\begin{aligned} p(\mathbf{z}') &= (p(\mathbf{z}) + m) \bmod \frac{N}{2} \\ p(n_p(\mathbf{z}')) &= p(n_p(\mathbf{z})) = 0 \\ n_p(\mathbf{z}') &= \begin{cases} n_p(\mathbf{z}), & 0 \leq p(\mathbf{z}) + m < \frac{N}{2} \\ S_{\frac{N}{2}}(n_p(\mathbf{z})), & \frac{N}{2} \leq p(\mathbf{z}) + m < N \end{cases} \end{aligned} \quad (5-32)$$

**Απόδειξη:** Εάν  $\mathbf{z}' = S_m(\mathbf{z})$ , τότε η εξίσωση (5-29) ισχύει και μπορούμε να αντικαταστήσουμε  $a'_1 = (a_1 + 2\pi \frac{m}{N}) \bmod 2\pi$  και  $a'_{N-1} = (a_{N-1} - 2\pi \frac{m}{N}) \bmod 2\pi$  στην

$$p(\mathbf{z}') = [\frac{N}{4\pi}(a'_1 - a'_{N-1})] \bmod \frac{N}{2} . \quad (5-33)$$

Επιπρόσθετα, εάν το  $N$  είναι άρτιο, το  $\frac{N}{2}$  είναι ακέραιος και μπορούμε να αλλάξουμε σειρά στο ακέραιο μέρος και τον τελεστή modulo  $\frac{N}{2}$ . Αυτό δίνει:

$$p(\mathbf{z}') = ([\frac{N}{4\pi}(a_1 - a_{N-1})] \bmod \frac{N}{2} + m) \bmod \frac{N}{2} ,$$



το οποίο οδηγεί στο πρώτο σκέλος της εξίσωσης (5-32). Θεωρώντας τώρα ότι το  $n_p(\mathbf{z})$  λαμβάνεται από το  $\mathbf{z}$  μέσω μίας κυκλικής μετατόπισης του  $N - p(\mathbf{z})$  μπορούμε να χρησιμοποιήσουμε το πρώτο σκέλος της (5-32) για τα  $\mathbf{z}$  και  $n_p(\mathbf{z})$ :

$$p(n_p(\mathbf{z})) = p(S_{N-p(\mathbf{z})}(\mathbf{z})) = (p(\mathbf{z}) + (N - p(\mathbf{z}))) \bmod \frac{N}{2} = 0, \quad (5-34)$$

και με μία όμοια σκέψη για τα  $\mathbf{z}'$  και  $n_p(\mathbf{z}')$  οδηγούμαστε στο δεύτερο σκέλος της (5-32).

Στη συνέχεια, μπορούμε να δούμε ότι  $n_p(\mathbf{z}) = S_{-p(\mathbf{z})}(\mathbf{z})$  και  $n_p(\mathbf{z}') = S_{-p(\mathbf{z}')}(\mathbf{z}') = S_{m-p(\mathbf{z}')}(\mathbf{z})$  ώστε

$$n_p(\mathbf{z}') = S_{p(\mathbf{z})+m-p(\mathbf{z}')} (n_p(\mathbf{z})) .$$

Τότε όμως μπορούμε να αντικαταστήσουμε

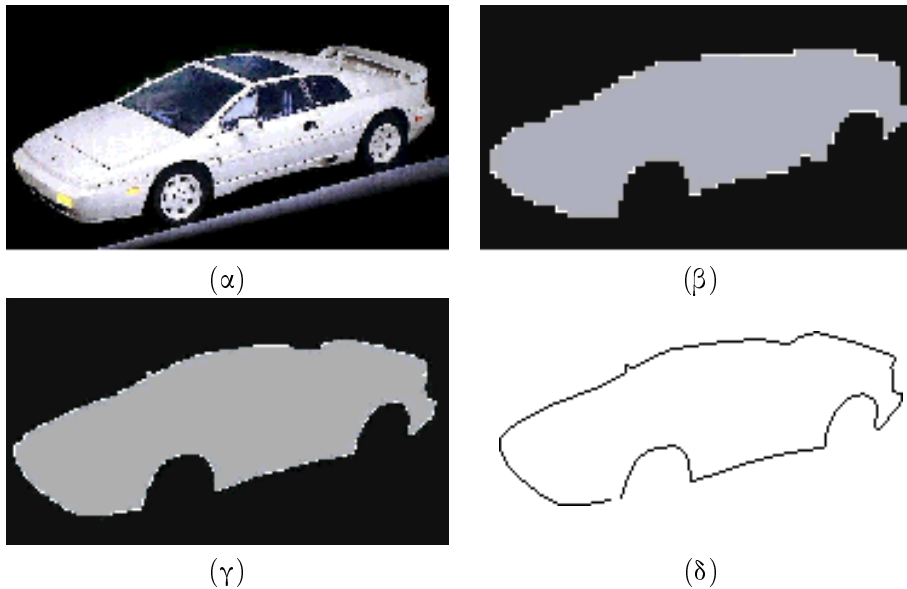
$$p(\mathbf{z}') = (p(\mathbf{z}) + m) \bmod \frac{N}{2} = \begin{cases} p(\mathbf{z}) + m, & 0 \leq p(\mathbf{z}) + m < \frac{N}{2} \\ p(\mathbf{z}) + m - \frac{N}{2}, & \frac{N}{2} \leq p(\mathbf{z}) + m < N \end{cases} \quad (5-35)$$

στην προηγούμενη σχέση οπότε προκύπτει το τρίτο σκέλος της (5-32).

Τέλος, αξίζει να προσεχθεί ότι όταν ο  $N$  δεν είναι άρτιος, αλλάζοντας σειρά στους ακέραιους μέρος και modulo  $\frac{N}{2}$  στην (5-33) μπορεί να οδηγήσει σε αβεβαιότητα ενός δείγματος πάνω ή κάτω στον υπολογισμό του  $p(\mathbf{z}')$  από το  $p(\mathbf{z})$ , πιθανότατα οδηγώντας σε μία αντίστοιχη αβεβαιότητα δείγματος μεταξύ  $n_p(\mathbf{z}')$  και  $n_p(\mathbf{z})$ . *Ο.Ε.Δ.*

Μπορεί να παρατηρηθεί ότι η προσθετική σχέση του πρώτου σκέλους της (5-32) οδηγεί σε κανονικοποίηση του σημείου έναρξης εφόσον απαιτούμε - μέσω της κυκλικής μετατόπισης της (5-31) - οι κανονικοποιημένες καμπύλες να υπακούουν στο δεύτερο σκέλος της (5-32), ακριβώς όπως απαιτούμε οι 'ορθογωνοποιημένες' καμπύλες να υπακούουν τις συνθήκες της εξ. (5-20). Παρατηρείται επίσης ότι η συνθήκη του δεύτερου σκέλους της (5-32) είναι ισοδύναμη με το να αναγκάσουμε τους συντελεστές Fourier  $u_1$  και  $u_{N-1}$  να έχουν ίσες φάσεις. Όπως θα δούμε παρακάτω, εάν αυτές οι φάσεις τεθούν ίσες με μηδέν, ούτως ώστε τα  $u_1$  και  $u_{N-1}$  να γίνουν πραγματικά και θετικά, τότε μπορεί να επιτευχθεί και κανονικοποίηση της περιστροφής. Παρόλ'αυτά, αφού το πρώτο σκέλος της (5-32) ισχύει για modulo  $\frac{N}{2}$  και όχι για modulo  $N$ , ακριβής κανονικοποίηση σημείου έναρξης μπορεί να επιτευχθεί μόνο όταν  $0 \leq p(\mathbf{z}) + m < \frac{N}{2}$ . Στην αντίθετη περίπτωση, η κανονικοποίηση οδηγεί σε μία επιπλέον μετατόπιση κατά  $\frac{N}{2}$  στο τρίτο σκέλος της (5-32). Αυτή η αβεβαιότητα οφείλεται στην επιλογή των συντελεστών  $u_1$  και  $u_{N-1}$  και μπορεί να απαλειφθεί μόνο μετά την κανονικοποίηση περιστροφής όπως θα φανεί παρακάτω.

Αν και η επιλογή δύο διαδοχικών συντελεστών όπως οι  $a_1$  και  $a_2$  θα οδηγούσε σε κανονικοποίηση ως προς το σημείο έναρξης χωρίς προβλήματα, η προτεινόμενη επιλογή της εξίσωσης (5-30) έχει δύο σημαντικά πλεονεκτήματα. Πρώτον, μπορεί να εντοπίσει κατοπτρικές συμμετρίες της καμπύλης (reflectional curve symmetries) και δεύτερον, όπως θα φανεί στις πειραματικές μετρήσεις, είναι ανθεκτικότερη στο θόρυβο και στις τοπικές διαφορές μεταξύ καμπυλών. Για παράδειγμα, η επιλογή των  $a_1$  και  $a_2$  για κανονικοποίηση οδηγεί σε τελείως διαφορετικά (σχεδόν τυχαία) σημεία έναρξης για δύο καμπύλες που διαφέρουν ελάχιστα. Η προτεινόμενη επιλογή των  $a_1$  και  $a_{N-1}$  έχει επίσης χρησιμοποιηθεί στην [117] για τον εντοπισμό του σημείου έναρξης σε καμπύλες που περιγράφουν γραμμικά πρότυπα, όπως τα όρια γραμμάτων του αλφάβητου. Η χρήση άλλων συντελεστών Fourier είναι επίσης δυνατή [90, 141]. Αυτόματος εντοπισμός των κατάλληλων συντελεστών βασίζεται συνήθως στο πλάτος τους πέραν από τη φάση τους και μπορεί να αντιμετωπίσει καλύτερα περιπτώσεις κατοπτρικών συμμετριών. Όμως, τέτοιες περιπτώσεις είναι μάλλον σπάνιες στον πραγματικό κόσμο και η επίδοση της προτεινόμενης διαδικασίας κανονικοποίησης είναι ικανοποιητική για πραγματικές εικόνες.



**Σχήμα 5.1** Ενδεικτικά αποτελέσματα κατάτμησης χρώματος για μία φωτογραφία που περιέχει ένα αυτοκίνητο, (α) αρχική εικόνα, (β) κατετμημένη εικόνα (πρώτο επίπεδο), (γ) κατετμημένη εικόνα (τελικό αποτέλεσμα) και (δ) εξαγόμενο περίγραμμα αντικειμένου.

### 5.5.2 Κανονικοποίηση περιστροφής και κατοπτρισμού

Ας υποθέσουμε ότι δύο αρχικές καμπύλες έχουν υποστεί 'ορθογωνοποίηση' και στη συνέχεια κανονικοποίηση ως προς το σημείο έναρξης και ότι οι ληφθείσες καμπύλες είναι οι  $s$  και  $s'$ . Σύμφωνα με την Πρόταση 5.2, οι κανονικοποιημένες καμπύλες θα ικανοποιούν την  $s' = Qs$ , όπου  $Q$  ένας ορθομοναδιαίος  $2 \times 2$  πίνακας. Τότε, μπορούμε με μοναδικό τρόπο να

αναλύσουμε τον  $\mathbf{Q}$  ως

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \quad (5-36)$$

όπου το  $\theta$  παίρνει τιμές στο  $[0, \pi)$  και  $|s_x| = |s_y| = 1$ , οπότε υπάρχει ένα-προς-ένα σχέση μεταξύ αυτών των παραμέτρων και των στοιχείων του  $\mathbf{Q}$ . Δηλαδή, το  $\mathbf{Q}$  αντιστοιχεί σε μία περιστροφή από 0 έως  $\pi$  rads ή/και σε μία οριζόντια ή κατακόρυφη αντανάκλαση. Για απλότητα, θα χρησιμοποιήσουμε ξανά την αναπαράσταση της καμπύλης βάσει του μιγαδικού διανύσματος  $\mathbf{z}$ ,  $\mathbf{z}'$  για δύο καμπύλες  $\mathbf{s}$ ,  $\mathbf{s}'$  και θα ξαναγράψουμε τη μεταξύ τους σχέση ως

$$\mathbf{z}' = (s_x \mathbf{x} + j s_y \mathbf{y}) e^{j\theta} \quad (5-37)$$

συμπεριλαμβάνοντας το μετασχηματισμό περιστροφής και κατοπτρισμού. Η παραπάνω σχέση θα μας οδηγήσει στην κανονικοποίηση διαδοχικά ως προς την περιστροφή και τον κατοπτρισμό σε δύο βήματα. Η διαδικασία ορίζεται για το  $\mathbf{z}$  και εφαρμόζεται με τον ίδιο ακριβώς τρόπο στο  $\mathbf{z}'$ . Όπως και στην περίπτωση του σημείου έναρξης, η κανονικοποίηση βασίζεται στην πρώτη  $a_1$  και την τελευταία  $a_{N-1}$  φάση του μετασχηματισμού Fourier του  $\mathbf{z}$ , όπου  $\mathbf{a} = \text{Arg}\mathcal{F}(\mathbf{z}) = [a_0 \ a_1 \ \cdots \ a_{N-1}]$ :

(α) Η περιστροφή της αρχικής καμπύλης  $\mathbf{z}$  κανονικοποιείται σύμφωνα με το μέσο όρο των  $a_1$  και  $a_{N-1}$ :

$$\begin{aligned} r(\mathbf{z}) &= \left(\frac{1}{2}(a_1 + a_{N-1})\right) \bmod \pi \\ \mathbf{z}_1 &= \mathbf{z} e^{-j r(\mathbf{z})} \end{aligned} \quad (5-38)$$

(β) Η οριζόντια και κατακόρυφη ανάκλαση κανονικοποιείται με χρήση των ροπών τρίτης τάξης του  $\mathbf{z}_1$ :

$$\begin{aligned} v(\mathbf{z}_1) &= v_x(\mathbf{z}_1) + j v_y(\mathbf{z}_1) = \text{sgn}\{m_{12}(\mathbf{z}_1)\} + j \text{sgn}\{m_{21}(\mathbf{z}_1)\} \\ n_r(\mathbf{z}) &= \mathbf{z}_2 = v_x(\mathbf{z}_1) \mathbf{x}_1 + j v_y(\mathbf{z}_1) \mathbf{y}_1 \end{aligned} \quad (5-39)$$

όπου το  $\text{sgn}$  υποδηλώνει τη συνάρτηση προσήμου. Αξίζει να προσεχθεί εδώ ότι θα ήταν επίσης αρκετό να ορίσουμε  $r(\mathbf{z}) = a_1$ . Ο προτεινόμενος όμως ορισμός οδηγεί στην εξαφάνιση του φαινομένου της εξ. (5-29) ως προς το σημείο έναρξης, ούτως ώστε η διαδικασία κανονικοποίησης της περιστροφής να είναι ανεξάρτητη του σημείου έναρξης, όπως εξηγείται και παρακάτω. Ομοίως προς τη διαδικασία ορθογωνοποίησης, η μοναδικότητα και η πληρότητα του  $n_r(\mathbf{z})$  μπορούν να αποδειχθούν, δηλαδή να δειχθεί ότι το  $n_r(\mathbf{z})$  είναι ανεξάρτητο των μετασχηματισμών περιστροφής και κατοπτρισμού και ότι συγκρατεί όλη την πληροφορία του  $\mathbf{s}$ , εκτός από τους μετασχηματισμούς στο οποίο είναι αναλλοίωτο.

**Πρόταση 5.4:** Εάν δύο καμπύλες  $\mathbf{s}$  και  $\mathbf{s}'$  σχετίζονται μέσω ενός μετασχηματισμού της μορφής  $\mathbf{s} = \mathbf{Q} \mathbf{s}'$ , όπου  $\mathbf{Q}$  είναι ένας ορθομοναδιαίος  $2 \times 2$  πίνακας (αναλυόμενος σε ένα πίνακα περιστροφής και ένα πίνακα ανάκλασης), τότε για τα αντίστοιχα μιγαδικά διανύσματα  $\mathbf{z}$  και  $\mathbf{z}'$  ισχύουν τα ακόλουθα:

$$\begin{aligned} r(\mathbf{z}') &= (\lambda r(\mathbf{z}) + \theta) \bmod \pi \\ n_r(\mathbf{z}') &= n_r(\mathbf{z}) \\ r(n_r(\mathbf{z}')) &= r(n_r(\mathbf{z})) = 0, \quad v_x(n_r(\mathbf{z}')) = v_y(n_r(\mathbf{z}')) = v_x(n_r(\mathbf{z})) = v_y(n_r(\mathbf{z})) = 1 \end{aligned} \quad (5-40)$$

όπου  $\lambda = s_x s_y = \pm 1$ .

**Απόδειξη:** Εάν  $\mathbf{s} = \mathbf{Q} \mathbf{s}'$  και ο  $\mathbf{Q}$  αναλύεται όπως στην εξίσωση (5-36), τότε η (5-37) ισχύει και μπορούμε να βρούμε σχέση μεταξύ των διακριτών μετασχηματισμών Fourier  $\mathbf{u} = \mathcal{F}(\mathbf{z})$  και  $\mathbf{u}' = \mathcal{F}(\mathbf{z}')$ :

$$u'_k = \sum_{i=0}^{N-1} z'_i w^{-k i} = s_x e^{j\theta} \sum_{i=0}^{N-1} (x_i + j \lambda y_i) w^{-k i} = \begin{cases} s_x e^{j\theta} u_k, & \lambda = 1 \\ s_x e^{j\theta} u_{N-k}^*, & \lambda = -1 \end{cases} \quad (5-41)$$

για  $k = 0, 1, \dots, N-1$  όπου  $\lambda = s_x s_y$ .

Εάν υπολογίσουμε τις φάσεις  $\mathbf{a} = \text{Arg } \mathbf{u}$  και  $\mathbf{a}' = \text{Arg } \mathbf{u}'$ , η εξ. (5-41) οδηγεί στα

$$\begin{aligned} a'_1 &= \begin{cases} a_1 + \theta + \text{Arg } s_x, & \lambda = 1 \\ -a_{N-1} + \theta + \text{Arg } s_x, & \lambda = -1 \end{cases} \\ a'_{N-1} &= \begin{cases} a_{N-1} + \theta + \text{Arg } s_x, & \lambda = 1 \\ -a_1 + \theta + \text{Arg } s_x, & \lambda = -1 \end{cases} \end{aligned} \quad (5-42)$$

Προσθέτοντας τις δύο προηγούμενες σχέσεις, λαμβάνουμε  $(a'_1 + a'_{N-1}) \bmod 2\pi = (\lambda(a_1 + a_{N-1}) + 2\theta) \bmod 2\pi$  αφού  $|s_x| = 1$ , οπότε  $(2 \text{Arg } s_x) \bmod 2\pi = 0$ . Από τον ορισμό του  $r(\mathbf{z})$  στο πρώτο σκέλος της (5-38), το πρώτο σκέλος της (5-40) προκύπτει αμέσως.

Τώρα, ο υπολογισμός της  $\mathbf{z}'_1$  από το δεύτερο σκέλος της (5-38) δίνει  $\mathbf{z}'_1 = \mathbf{z}' e^{-j r(\mathbf{z}')} = s_x (\mathbf{x} + j \lambda \mathbf{y}) e^{j(\theta - r(\mathbf{z}'))}$ . Εάν ορίσουμε

$$l(\mathbf{z}) = \begin{cases} 1, & \lambda r(\mathbf{z}) + \theta \in [0, \pi) \\ -1, & \lambda r(\mathbf{z}) + \theta \in [\pi, 2\pi) \end{cases} \quad (5-43)$$

τότε από το πρώτο σκέλος της (5-40) προκύπτει ότι

$$\mathbf{z}'_1 = s_x (\mathbf{x} + j \lambda \mathbf{y}) e^{j(\theta - ((\lambda r(\mathbf{z}) + \theta) \bmod \pi))} = s_x l(\mathbf{z}) (\mathbf{x} + j \lambda \mathbf{y}) e^{-j \lambda r(\mathbf{z})} = \begin{cases} s_x l(\mathbf{z}) \mathbf{z}_1, & \lambda = 1 \\ s_x l(\mathbf{z}) \mathbf{z}_1^*, & \lambda = -1 \end{cases} \quad (5-44)$$

ούτως ώστε  $\mathbf{x}'_1 = s_x l(\mathbf{z}) \mathbf{x}_1$  και  $\mathbf{y}'_1 = \lambda s_x l(\mathbf{z}) \mathbf{y}_1 = s_y l(\mathbf{z}) \mathbf{y}_1$ . Οπότε τα  $\mathbf{x}'_1$  και  $\mathbf{y}'_1$  διαφέρουν (ενδεχομένως) από τα  $\mathbf{x}_1$  και  $\mathbf{y}_1$  μόνο στα πρόσημά τους και ο υπολογισμός των

ροπών τρίτης τάξεως στο πρώτο σκέλος της (5-39) δίνει  $v_x(\mathbf{z}'_1) = s_x l(\mathbf{z}) v_x(\mathbf{z}_1)$  και  $v_y(\mathbf{z}'_1) = s_y l(\mathbf{z}) v_y(\mathbf{z}_1)$ . Τότε, η επόμενη κανονικοποίηση του δεύτερου σκέλους της (5-39) δίνει  $\mathbf{x}'_2 = (s_x l(\mathbf{z}))^2 v_x(\mathbf{z}_1) \mathbf{x}_1 = \mathbf{x}_2$  και ομοίως  $\mathbf{y}'_2 = \mathbf{y}_2$  και το δεύτερο μέλος της (5-40) αποδεικνύεται.

Τέλος οι συνθήκες του τρίτου σκέλους έπονται απευθείας από το πρώτο μέλος της εξίσωσης και την εξ. (5-39). Ο.Ε.Δ.

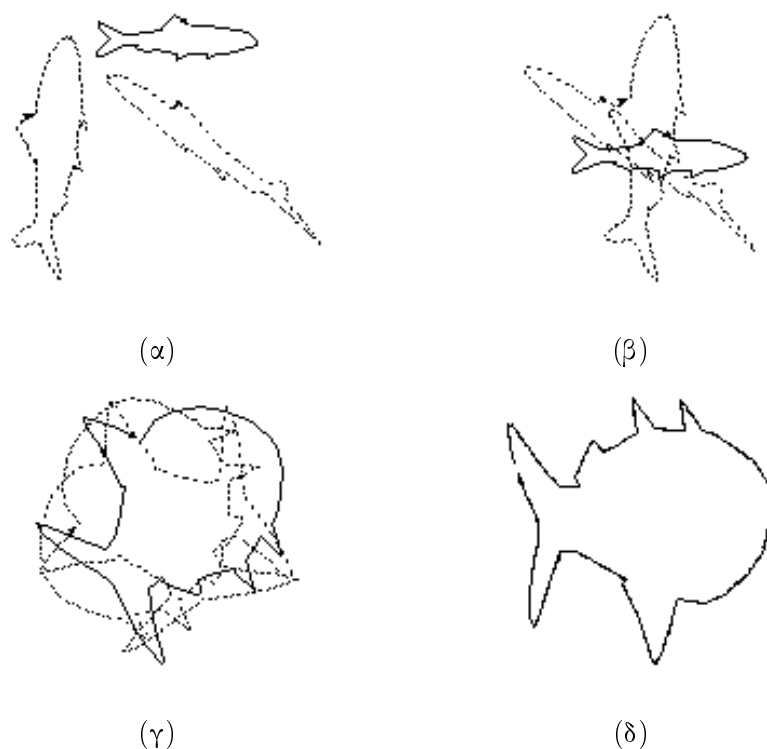
Μπορεί εδώ να σημειωθεί ότι όπως στην (5-20) και στο δεύτερο σκέλος της (5-32), οι μετασχηματισμοί περιστροφής και ανάκλασης καταργούνται ώστε οι κανονικοποιημένες καμπύλες να ικανοποιούν τις συνθήκες του τρίτου σκέλους της (5-40). Επιπλέον, όπως και στη διαδικασία της ορθογωνοποίησης, το σύνολο παραμέτρων  $\{r(\mathbf{z}), v_x(\mathbf{z}), v_y(\mathbf{z})\}$  μαζί με το  $n_r(\mathbf{z})$  περιέχει όλη την πληροφορία για την αρχική καμπύλη  $\mathbf{z}$ . Μπορεί επίσης να συναχθεί από την απόδειξη της Πρότασης 5.4 ότι η διαδικασία κανονικοποίησης είναι ανεξάρτητη από το σημείο έναρξης, δηλαδή  $n_r(S_m(\mathbf{z})) = S_m(n_r(\mathbf{z}))$ , για κάθε  $m \in \{0, 1, \dots, N-1\}$ , ακριβώς όπως  $n_a(S_m(\mathbf{z})) = S_m(n_a(\mathbf{z}))$ . Έτσι η κανονικοποίηση περιστροφής και ανάκλασης μπορεί να εφαρμοστεί ακόμα και πριν την κανονικοποίηση σημείου έναρξης. Σ'αυτήν την περίπτωση είναι σημαντικό να σημειωθεί ότι συνδυάζοντας τα αποτελέσματα των Προτάσεων 5.2 και 5.4, η ορθογωνοποίηση καμπύλης, ακολουθούμενη από κανονικοποίηση περιστροφής και ανάκλασης, δηλαδή  $n_r(n_a(\mathbf{z}))$  είναι αναλλοίωτη σε κάθε μετασχηματισμό *affine*. Η ευρρωστέα της μεθόδου στο θόρυβο και στις παραμορφώσεις της καμπύλης περιγράφεται στα πειράματα.

Δύο τελευταία βήματα είναι αναγκαία μετά την κανονικοποίηση περιστροφής και ανάκλασης. Πρώτον, ας θυμηθούμε ότι η κανονικοποίηση του σημείου έναρξης μας αφήνει με μία αβεβαιότητα  $\frac{N}{2}$ . Μπορούμε να δούμε ότι μετά την κανονικοποίηση της περιστροφής, το εντοπιζόμενο σημείο έναρξης  $(x_0, y_0)$  βρίσκεται πολύ κοντά στον  $x$ -άξονα ( $y_0 \approx 0$ ) και μακριά από τον  $y$ -άξονα σε κάθε πλευρά. Η αβεβαιότητα του σημείου εκκίνησης λύνεται τότε εφαρμόζοντας μία κυκλική μετατόπιση κατά  $\frac{N}{2}$  δείγματα όταν  $x_0 < 0$ , διαφορετικά η καμπύλη μένει ως έχει. Δεύτερον, η κατευθυντικότητα (διαγραφής) της καμπύλης κανονικοποιείται σε αντιστροφολογική.

Στο σημείο αυτό η διαδικασία κανονικοποίησης έχει πλέον ολοκληρωθεί.

## 5.6 Πειράματα και συζήτηση

Στην ενότητα αυτή, επιδεικνύεται η επίδοση του προτεινόμενου αλγορίθμου και συζητώνται κατάλληλες αλγοριθμικές λεπτομέρειες / οδηγίες. Αρχικά, δίνονται κάποια ενδεικτικά

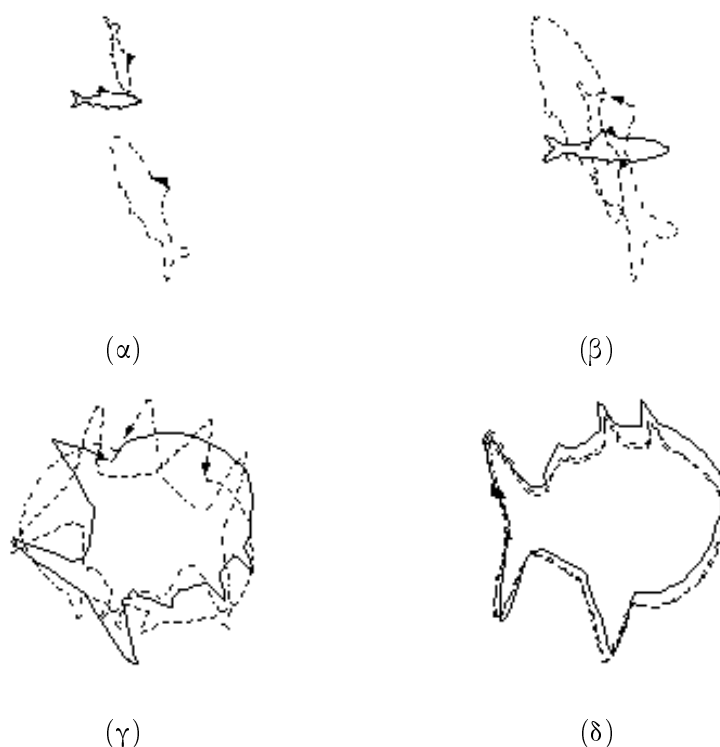


**Σχήμα 5.2** Ενδιάμεσα και τελικά αποτελέσματα της προτεινόμενης διαδικασίας ορθογωνοποίησης στα δείγματα του περιγράμματος ενός ψαριού (χωρίς χρήση τεχνικής αναδειγματοληψίας), (α) αρχική καμπύλη και ζεύγος τυχαίων μετασχηματισμών affine, (β) καμπύλες μετά την κανονικοποίηση μετατόπισης, (γ) καμπύλες μετά την κανονικοποίηση μεγέθυνσης και (δ) τελικές καμπύλες.

αποτελέσματα σε κατάτμηση χρώματος και κίνησης για την εξαγωγή των σημείων περιγράμματος των υπό εξέταση αντικειμένων. Στη συνέχεια, δίνονται αποτελέσματα κανονικοποίησης για όλα τα προτεινόμενα βήματα της διαδικασίας κανονικοποίησης, δηλαδή μετατόπισης, μεγέθυνσης, περιστροφής και σημείου έναρξης. Η επίδοση του αλγορίθμου εξετάζεται τότε για μετασχηματισμούς affine (α) του ίδιου αντικειμένου, (β) του ίδιου αντικειμένου λαμβανόμενου με διαφορετική δειγματοληψία, (γ) όμοιων αντικειμένων και (δ) ουσιαστικά διαφορετικών αντικειμένων. Παράλληλα, η θεωρία των κλειστών κυβικών B-splines της Ενότητας 5.3 υλοποιήθηκε για την αντιμετώπιση προβλημάτων προσαρμογής λόγω θορύβου ή/και ανομοιομορφίας δειγματοληψίας.

Όπως αναφέρθηκε στην Ενότητα 5.2, τα αποτελέσματα της διαδικασίας κατάτμησης χρώματος και κίνησης επηρεάζουν σημαντικά την επίδοση των τεχνικών κανονικοποίησης. Το τελευταίο ισχύει για κάθε τεχνική μοντελοποίησης περιγράμματος (contour modeling) ή σύγκρισης καμπυλών (curve matching) που βασίζεται σε πραγματικά δεδομένα. Για τις ανάγκες του κεφαλαίου, υπάρχοντες αλγόριθμοι κατάτμησης χρώματος και κίνησης της βιβλιογραφίας εξετάστηκαν, ώστε να ληφθούν τα καλύτερα δυνατά περιγράμματα από στατικές εικόνες και

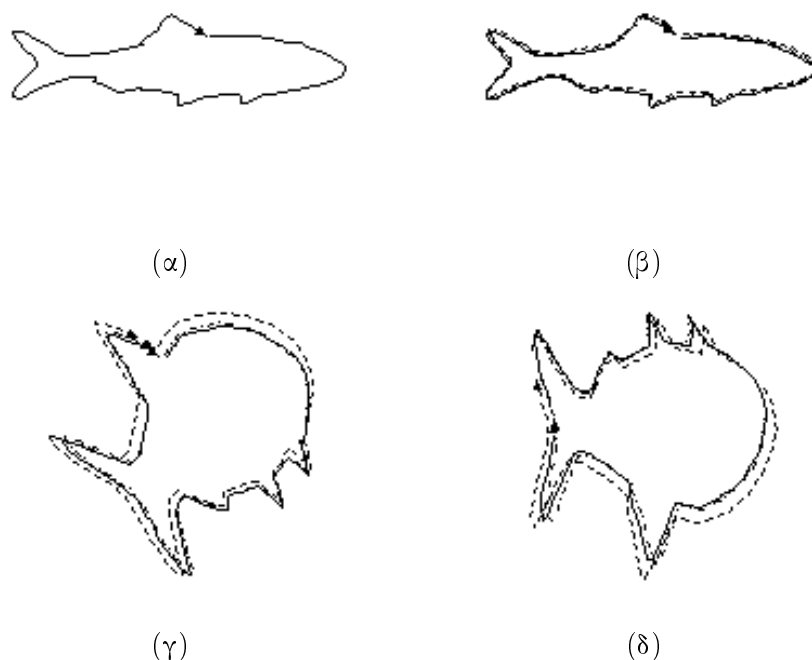
ακολουθίες εικόνων. Μ' αυτό το σκεπτικό, ο αλγόριθμος M-RSST φάνηκε να επιστρέφει αρκετά ικανοποιητικά αποτελέσματα για χρωματικά αντικείμενα σε εικόνες υψηλής ανάλυσης, ενώ η κατάτμηση κίνησης πραγματοποιήθηκε με βάση 2Δ παραμετρικά μοντέλα κίνησης. Μερικά ενδεικτικά αποτελέσματα κατάτμησης χρώματος για ένα ακίνητο αντικείμενο δίνονται στο Σχήμα 5.1. Στο Σχήμα 5.1(α) φαίνεται η αρχική εικόνα, ενώ στα Σχήματα 5.1(β) και 5.1(γ) εμφανίζονται τα αποτελέσματα του πρώτου και του τελευταίου βήματος αντίστοιχα της διαδικασίας κατάτμησης (M-RSST). Το εξαγόμενο περίγραμμα του αντικειμένου βάσει των λαμβανόμενων δειγμάτων φαίνεται στο Σχήμα 5.1(δ).



**Σχήμα 5.3** Αποτελέσματα από την προτεινόμενη τεχνική ορθογωνοποίησης στα δείγματα του περιγράμματος ενός ψαριού μετά από μη-ομοιόμορφη δειγματοληψία στις καμπύλες εισόδου, (α) αρχική καμπύλη και ζεύγος τυχαίων μετασχηματισμών affine, (β,γ,δ) κακή εφαρμογή των καμπυλών μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής.

Τα προκύπτοντα περιγράμματα δίνονται τότε ως είσοδος στη μέθοδο ορθογωνοποίησης. Θα πρέπει εδώ να αποσαφηνιστεί ότι ο προτεινόμενος αλγόριθμος χρησιμοποιείται ευθέως και ως έχει, χωρίς να χρειάζεται κάποιου είδους ανθρώπινη παρέμβαση, αφού οι όποιες ελεύθερες παράμετροι τίθενται μία φορά με την εκκίνηση του συστήματος. Αποτελέσματα από την επίδοση του αλγορίθμου με βάση ένα σύνολο σημείων που έχουν εξαχθεί πάνω στο περίγραμμα του αντικειμένου βλέπουμε στο Σχήμα 5.2. Ένα περίγραμμα αποτελούμενο από 100 σημεία / δείγματα, που έχει ληφθεί μέσω κατάτμησης χρώματος από μία φωτογραφία που περιέχει ένα ψάρι εμφανίζεται πάνω-δεξιά στο Σχήμα 5.2(α). Στο ίδιο σχήμα έχουν συμπεριληφθεί

ακόμη δύο άλλα σύνολα σημείων αποτελούμενα από 100 σημεία, έχοντας ληφθεί μέσω τυχαίων μετασχηματισμών affine πάνω στα σημεία του πρώτου περιγράμματος. Ακολουθώντας τη θεωρητική μεθοδολογία αυτού του κεφαλαίου, τα σύνολα σημείων υπεβλήθησαν σε κανονικοποίηση μετατόπισης, μεγέθυνσης, περιστροφής και σημείου έναρξης. Τα προκύπτοντα σύνολα δεδομένων εμφανίζονται στα Σχήματα 5.2(β), 5.2(γ) και 5.2(δ) μετά από κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής/έναρξης αντίστοιχα. Μπορεί κανείς να δει ότι οι τελικές καμπύλες εφαρμόζουν τέλεια όταν οι αρχικές καμπύλες είναι μετασχηματισμοί affine της ίδιας καμπύλης, επαληθεύοντας και τα αποτελέσματα των Ενοτήτων 5.4 και 5.5.

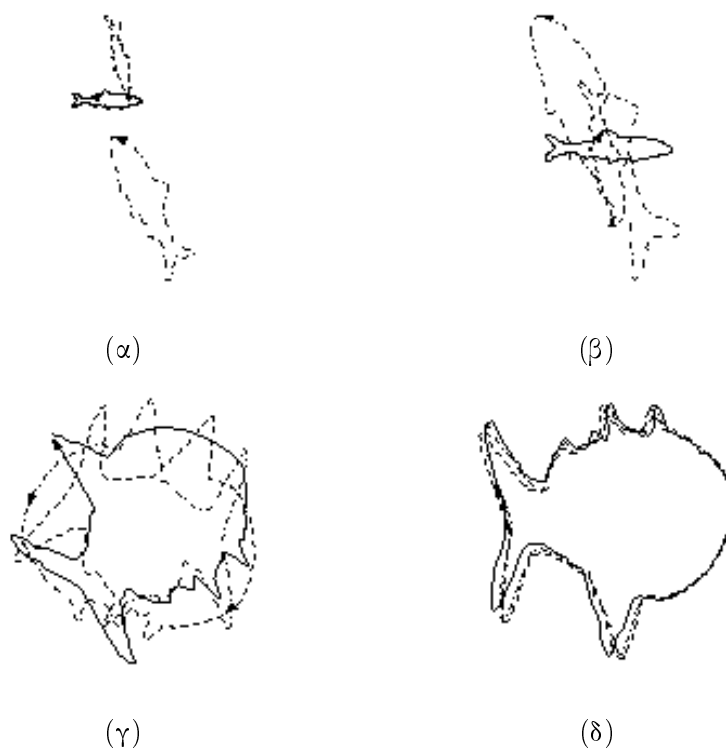


**Σχήμα 5.4** Αποτελέσματα από την προτεινόμενη τεχνική ορθογωνοποίησης στα δείγματα του περιγράμματος ενός ψαριού μετά από διαφορετική μη-ομοιόμορφη δειγματοληψία στην ίδια ακριβώς καμπύλη, (α) τρία επικαλυπτόμενα στιγμιότυπα της αρχικής καμπύλης, (β,γ,δ) κακό ταίριασμα των καμπυλών μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής.

Στην πραγματικότητα, τα προκύπτοντα (ορθογωνοποιημένα) σύνολα δεδομένων είναι πανομοιότυπα μόνο όταν προέρχονται από μετασχηματισμούς affine της ίδιας καμπύλης. Σύμφωνα με αυτή τη διαπίστωση, τα Σχήματα 5.3 και 5.4 επιδεικνύουν την κακή εφαρμογή ορθογωνοποιημένων καμπυλών προερχόμενων από μία συγκεκριμένη αρχική καμπύλη, έχοντας όμως υποστεί διαφορετική (και μη-ομοιόμορφη) δειγματοληψία. Στο Σχήμα 5.3(α), η αρχική καμπύλη μετασχηματίζεται με χρήση ενός ζεύγους τυχαίων μετασχηματισμών affine και οι προκύπτουσες καμπύλες ανα-δειγματοληπτούνται με μη-ομοιόμορφη δειγματοληψία. Τα αποτελέσματα του κάθε βήματος κανονικοποίησης για τις καμπύλες αυτές φαίνονται τότε στα Σχήματα



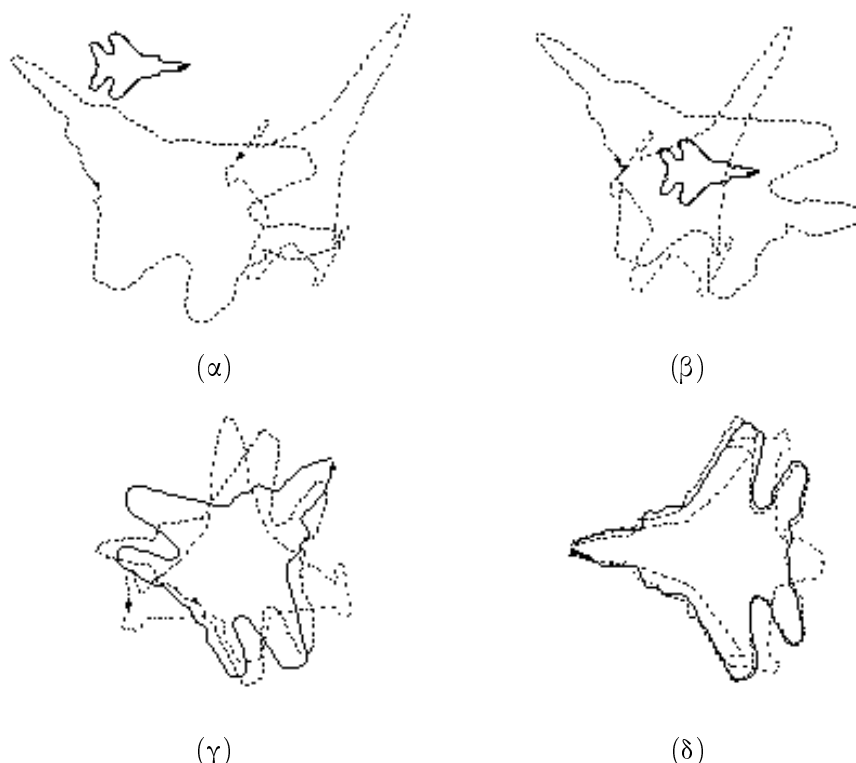
5.3(β)-(δ). Στα αντίστοιχα σχήματα του Σχήματος 5.4, η τελευταία διαδικασία επαναλαμβάνεται για τρία διαφορετικά δειγματοληπτημένα στιγμιότυπα της ίδιας ακριβώς καμπύλης. Και στις δύο περιπτώσεις, τα αποτελέσματα δεν είναι πολύ μακριά από τα επιθυμητά, ακόμα και για μεγάλες διαφορές στην διαδικασία της δειγματοληψίας. Όμως, οι παρατηρούμενες διαφοροποιήσεις μπορούν να είναι ιδιαίτερος ενοχλητικές όταν μεγάλη ακρίβεια απαιτείται στο ταίριασμα καμπυλών από μία συγκεκριμένη εφαρμογή.



**Σχήμα 5.5** Βελτίωση των αποτελεσμάτων ορθογωνοποίησης μετά από χρήση B-spline μοντελοποίησης και αναδιάταξη των σημείων σύνδεσης στις μη ομοιόμορφα δειγματοληπτημένες καμπύλες του Σχήματος 5.3, (α) αρχική καμπύλη και ζεύγος τυχαίων μετασχηματισμών affine, (β,γ,δ) καμπύλες μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής αντίστοιχα.

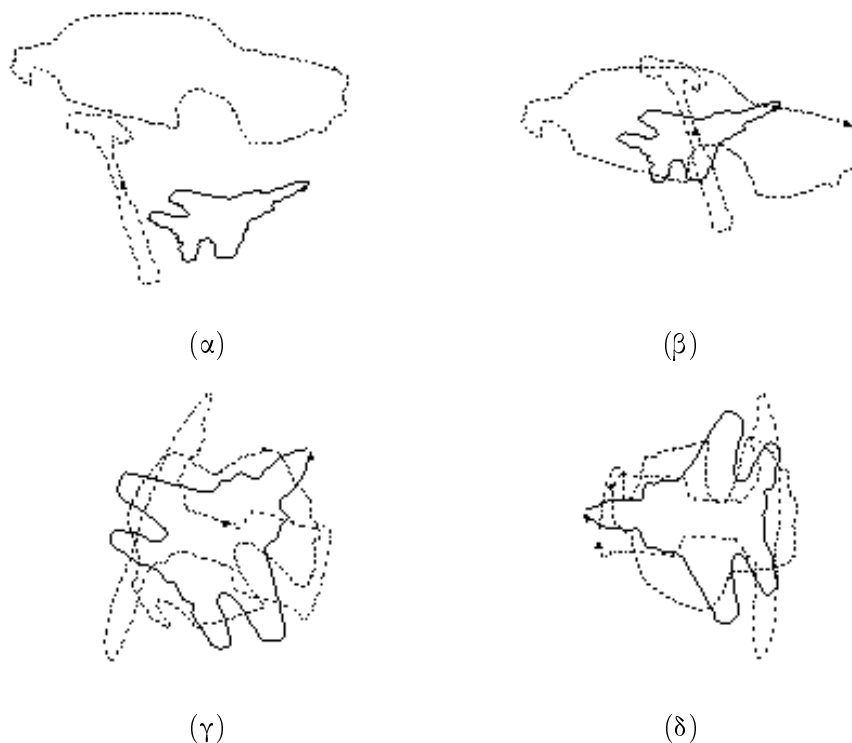
Τα λαμβανόμενα αποτελέσματα μπορούν να βελτιωθούν σημαντικά εφαρμόζοντας μία τεχνική ανα-δειγματοληψίας μαζί με την διαδικασία κανονικοποίησης. Κυβικές B-splines χρησιμοποιήθηκαν εδώ για το σκοπό αυτό, αφού το ταίριασμα των σημείων σύνδεσης και η ανατοποθέτησή τους φάνηκαν να δίνουν πολύ καλά αποτελέσματα όσον αφορά τη δειγματοληψία με ομοιόμορφο τρόπο (βλ. και Ενότητα 5.3). Το Σχήμα 5.5 επιδεικνύει τη βελτίωση στο ταίριασμα μεταξύ ορθογωνοποιημένων καμπυλών μετά από μοντελοποίηση με B-spline και ανατοποθέτηση των σημείων σύνδεσης δύο φορές: πριν από την κανονικοποίηση θέσης και πριν από την κανονικοποίηση περιστροφής / σημείου έναρξης. Ένα τελικό βήμα ανα-δειγματοληψίας είναι χρήσιμο επί των τελικών καμπυλών, ώστε τα σημεία σύνδεσης να αντιστοιχούν ακριβώς

και η διαδικασία σύγκρισης να δώσει καλύτερα αποτελέσματα. Οι παραστάσεις στο Σχήμα 5.5 είναι απευθείας συγκρίσιμες με τις αντίστοιχες του Σχήματος 5.3. Όταν αναφερόμαστε σε ορθογωνοποιημένες καμπύλες στη συνέχεια θα υποθέτουμε ότι η μοντελοποίηση με B-spline και η ανα-τοποθέτηση των σημείων σύνδεσης εφαρμόζεται δις, σύμφωνα και με τα όσα ειπώθηκαν παραπάνω.



**Σχήμα 5.6** Άμεση ευθυγράμμιση καμπυλών μέσω της διαδικασίας κανονικοποίησης για νοηματικά όμοια (χωρικά όμοια) περιγράμματα αντικειμένων, (α) αρχικές καμπύλες για τρία διαφορετικά αεροπλάνα, (β,γ,δ) καμπύλες μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής αντίστοιχα.

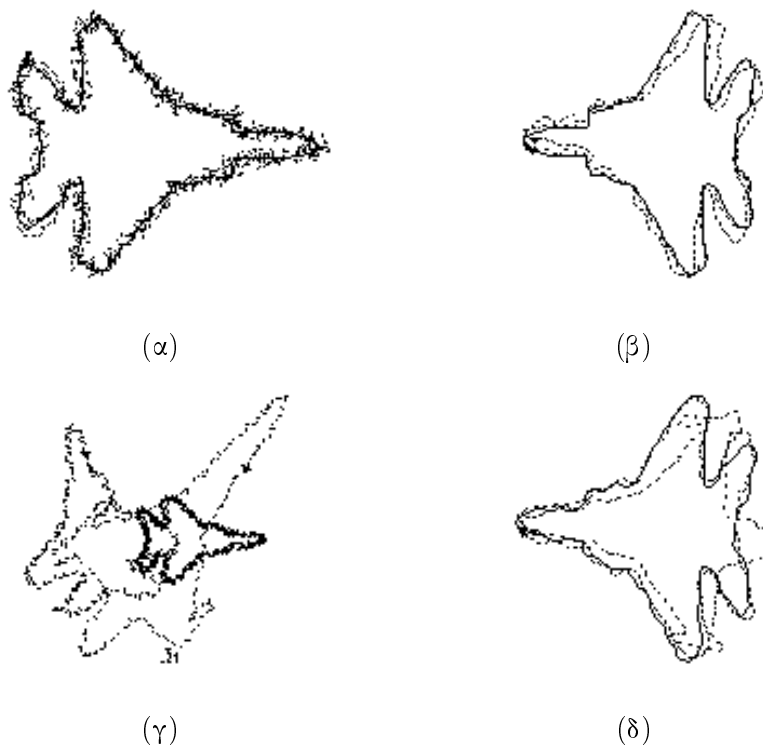
Ίσως η πιο σημαντική ιδιότητα της προτεινόμενης προσέγγισης είναι η δυνατότητά της να ευθυγραμμίζει καμπύλες που φαίνονται χωρικά ‘όμοιες’. Το τελευταίο δοκιμάστηκε σε μία σειρά από περιγράμματα αντικειμένων ανήκοντας σε διαφορετικές κλάσεις/κατηγορίες, όπως αεροπλάνα, αυτοκίνητα, ποτήρια, ψάρια και σφυριά. Το Σχήμα 5.6 περιέχει αποτελέσματα κανονικοποίησης για τρία διαφορετικά αεροπλάνα που λήφθηκαν μέσω κατάτμησης χρώματος, επομένως και ουσιαστικά με διαφορετική δειγματοληψία. Η προτεινόμενη μέθοδος, μαζί με την ανα-δειγματοληψία με κλειστές κυβικές B-splines αποδεικνύεται να ευθυγραμμίζει με μεγάλη επιτυχία χωρικά σχετικά περιγράμματα αντικειμένων. Τα αποτελέσματα της ορθογωνοποίησης του Σχήματος 5.6 μπορούν να συγκριθούν απευθείας με αυτά του Σχήματος 5.7, όπου τρεις καμπύλες από διαφορετικές κλάσεις αντικειμένων έχουν χρησιμοποιηθεί. Μπορεί κανείς να δει ότι οι τελικές καμπύλες του Σχήματος 5.6(δ) θα δώσουν μεγαλύτερη ομοιότητα με χρήση



**Σχήμα 5.7** Ευθυγράμμιση καμπυλών μέσω της διαδικασίας κανονικοποίησης για ουσιαστικά διαφορετικά περιγράμματα αντικειμένων, (α) αρχικές καμπύλες για ένα σφυρί, ένα αυτοκίνητο και ένα αεροπλάνο, (β,γ,δ) καμπύλες μετά την κανονικοποίηση μετατόπισης, μεγέθυνσης και περιστροφής αντίστοιχα.

οποιασδήποτε τεχνικής ταιριάσματος σε σχέση με αυτές του Σχήματος 5.7(δ).

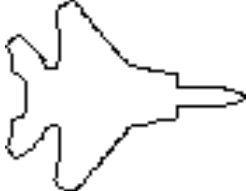









Η προτεινόμενη προσέγγιση φαίνεται να δίνει ικανοποιητικά αποτελέσματα ακόμα και παρουσία σημαντικού ποσοστού θορύβου στις καμπύλες εισόδου. Με σκοπό να προσομοιωθεί ο θόρυβος που εμπλέκεται στα περιγράμματα των αντικειμένων λόγω σφαλμάτων της διαδικασίας κατάκτησης χρώματος / κίνησης, τυχαίος θόρυβος σχετικά υψηλής διασποράς εισήχθη με τεχνητό τρόπο στα διαθέσιμα περιγράμματα. Η επίδοση του αλγορίθμου ήταν ικανοποιητική ακόμη και σε αυτήν την περίπτωση, έως ότου το περίγραμμα του αντικειμένου δεν ήταν σημαντικά παραμορφωμένο. Ενδεικτικά πειραματικά αποτελέσματα δίνονται στο Σχήμα 5.8. Στο Σχήμα 5.8(α), ένα περίγραμμα αεροπλάνου, συνοδευόμενο από δύο θορυβώδεις εκδοχές του, έχει συμπεριληφθεί, όπου η μία εκ των δύο εκδοχών έχει 'μολυνθεί' με διπλάσια ποσότητα θορύβου (σε επίπεδο τυπικής απόκλισης) σε σχέση με την άλλη. Στο Σχήμα 5.8(β) δίνονται τα αντίστοιχα αποτελέσματα κανονικοποίησης. Θα πρέπει να τονιστεί εδώ ότι ο τεχνητός θόρυβος εισήχθη στα δείγματα των αρχικών καμπυλών (λαμβάνόμενα με μη-ομοιόμορφη δειγματοληψία), πριν από τη μοντελοποίηση του περιγράμματος με B-spline. Όπως φαίνεται στο Σχήμα 5.8(β), η λιγότερο θορυβώδης καμπύλη δίνει καλύτερη προσαρμογή με την αντίστοιχη της



**Σχήμα 5.8** Αποτελέσματα κανονικοποίησης παρουσία τυχαίου θορύβου στα δείγματα των αρχικών καμπυλών, (α,β) αρχικές και τελικές καμπύλες για τρία στιγμιότυπα της ίδιας καμπύλης, (γ,δ) αρχικές και τελικές καμπύλες για τρία όμοια νοηματικά αντικείμενα (αεροπλάνα).

από ότι η περισσότερη θορυβώδης. Στα Σχήματα 5.8(γ) και 5.8(δ), αντίστοιχα αποτελέσματα δίνονται για τρία χωρικά όμοια αντικείμενα (αεροπλάνα).

Η ταξινόμηση αντικειμένων βάσει της πληροφορίας περιγράμματος έχει αντιμετωπιστεί από αρκετούς ερευνητές, όπως είδαμε και στην Ενότητα 5.1, επομένως δεν κρίνεται σκόπιμο να αναλυθεί εδώ σε μεγάλη έκταση. Αντιθέτως, μόνο μερικά ενδεικτικά αποτελέσματα ταξινόμησης έχουν συμπεριληφθεί στο Σχήμα 5.9. Η αρχική πρότυπη Καμπύλη 1, όπως και οι αποθηκευμένες Καμπύλες 2 υπέστησαν αρχικώς τη διαδικασία κανονικοποίησης. Στη συνέχεια, οι 'κανονικοποιημένες' εκδοχές τους συγκρίθηκαν μεταξύ τους ως προς τρία απλά κριτήρια, συγκεκριμένα την Ευκλείδεια απόσταση (α) των αντίστοιχων σημείων/δειγμάτων, (β) των υπολογιζόμενων περιγραφών Fourier (Fourier Descriptors - FD) και (γ) των τροποποιημένων περιγραφών Fourier (Modified Fourier Descriptors - MFD) [130]. Όπως κανείς θα ανέμενε εξ αρχής, αλλά και όπως επιβεβαίωσαν τα πειράματα, οι τρεις αυτές μετρικές αποτυγχάνουν στο στόχο τους για τις αρχικές (μη κανονικοποιημένες) καμπύλες. Αντιθέτως αποτελούν ικανοποιητικά μέτρα της ομοιότητας των κανονικοποιημένων καμπυλών (βλ. Σχήμα 5.9). Κάθε σύνολο από αποστάσεις στο σχήμα αυτό κανονικοποιήθηκε, ούτως ώστε σε κάθε περίπτωση οι τιμές μηδέν (0) και ένα (1) να υποδηλώνουν τέλεια εφαρμογή και καμία ομοιότητα αντίστοιχα.

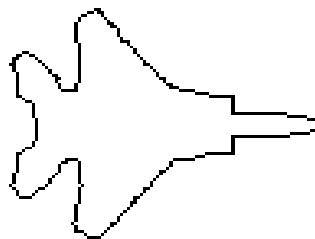
Καμπύλη 1	Καμπύλη 2	Σημεία	FD	MFD
		0.01	0.02	0.01
		0.19	0.12	0.11
		0.75	0.41	0.57
		0.89	0.62	0.65
		0.76	0.25	0.32

**Σχήμα 5.9** Ενδεικτικές αποστάσεις περιγραμμάτων για σκοπούς ταξινόμησης.

Τελικά, η επίδοση του προτεινόμενου αλγορίθμου ελέγχθηκε επιτυχώς για σκοπούς ‘αναζήτησης με βάση το περιεχόμενο βάσει ομοιότητας περιγράμματος’ σε μία μικρή βάση δεδομένων που περιείχε 50 ακίνητες εικόνες, οι οποίες με τη σειρά τους ανήκαν σε 5 οπτικά (και εννοιολογικά) διαφορετικές κλάσεις αντικειμένων - συγκεκριμένα αεροπλάνα, αυτοκίνητα, ψάρια, σφυριά και ποτήρια. Για καλύτερα αποτελέσματα, αυξημένη προσοχή δόθηκε ώστε τα περιγράμματα των πιο σημαντικών αντικειμένων να εξαχθούν επιτυχώς. Στο Σχήμα 5.10, σχετικά πειραματικά αποτελέσματα εμφανίζονται για μία εικόνα εισόδου που περιέχει ένα αεροπλάνο. Η εικόνα εισόδου εμφανίζεται στο Σχήμα 5.10(α), το εξαγόμενο περίγραμμα στο Σχήμα 5.10(β), ενώ οι λαμβανόμενες (υποτιθέμενες όμοιες) εικόνες στο Σχήμα 5.10(γ), με φθίνουσα σειρά ομοιότητας προς τα δεξιά και κάτω. Όμοια αποτελέσματα για μία εικόνα εισόδου που περιέχει ένα αυτοκίνητο δίνονται στο Σχήμα 5.11. Θα πρέπει να σημειωθεί για το Σχήμα 5.10 ότι η χρήση μόνο του κριτηρίου της ομοιότητας περιγράμματος για τη σύγκριση μεταξύ δύο εικό-



(α)



(β)



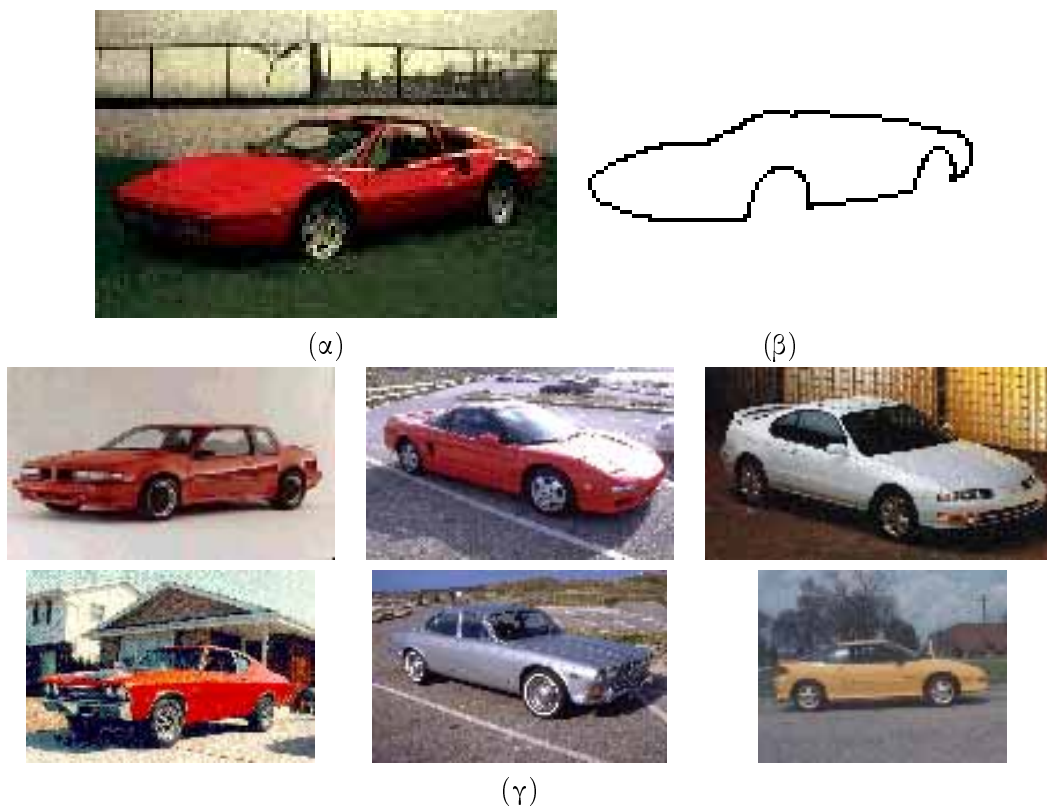
(γ)

**Σχήμα 5.10** Αποτελέσματα ανάκτησης εικόνων (αναζήτηση-βάσει-παραδείγματος) βάσει ομοιότητας περιγράμματος για μία φωτογραφία που περιέχει ένα αεροπλάνο, (α) φωτογραφία εισόδου, (β) εξαγόμενο περίγραμμα (γ) φωτογραφίες κατά σειρά ανάκτησης (προς τα δεξιά και προς τα κάτω).

νων / αντικειμένων μπορεί να οδηγήσει σε (εννοιολογικά) λάθη, αφού ένα ψάρι μοιάζει ίσως περισσότερο με ένα αεροπλάνο F-15 από όσο μοιάζει ένα αεροπλάνο Stealth με το F-15. Για τέτοιους λόγους, η προτεινόμενη μεθοδολογία θα είχε ακόμα καλύτερα αποτελέσματα ταξινόμησης και ανάκτησης, χρησιμοποιούμενη παράλληλα με άλλα κριτήρια ομοιότητας, όπως αυτά που εκμεταλλεύονται τα χαρακτηριστικά χρώματος και υφής, σε ένα ολοκληρωμένο σύστημα αναζήτησης εικόνων με βάση το περιεχόμενο.

## 5.7 Συμπεράσματα και επεκτάσεις

Χρησιμοποιώντας τη διαδικασία κανονικοποίησης καμπυλών που προτάθηκε σε αυτό το κεφάλαιο, είναι δυνατόν να λάβει κανείς μία αναπαράσταση καμπύλης αναλλοίωτη σε μετασχηματισμούς affine, χωρίς πραγματική απώλεια πληροφορίας της αρχικής καμπύλης. Η διαδικασία αυτή μπορεί να χρησιμοποιηθεί ως ένα βήμα προεπεξεργασίας σε κάθε τεχνική αναπαράστασης



**Σχήμα 5.11** Αποτελέσματα ανάκτησης εικόνων (αναζήτηση-βάσει-παραδείγματος) βάσει ομοιότητας περιγράμματος για μία φωτογραφία που περιέχει ένα αυτοκίνητο, (α) φωτογραφία εισόδου, (β) εξαγόμενο περίγραμμα (γ) φωτογραφίες κατά σειρά ανάκτησης (προς τα δεξιά και προς τα κάτω).

σχήματος (shape representation) ή ταξινόμησης (classification), αναγνώρισης (recognition) και ανάκτησης (retrieval) αντικειμένων με βάση το περίγραμμα, αφού καταφέρνει επιτυχώς να αποσυνδέσει το πρόβλημα του 'αναλλοίωτου σε μετασχηματισμούς affine' από την εξαγωγή χαρακτηριστικών (feature extraction) και το ταίριασμα προτύπων (pattern matching). Αυτό επιβεβαιώνεται χρησιμοποιώντας ένα πλήθος από γνωστές μεθόδους ταιριάσματος καμπυλών στην περιοχή της ανάκτησης με βάση το περιεχόμενο από βάσεις δεδομένων εικόνων και ακολουθιών εικόνων. Σε κάθε περίπτωση, η προτεινόμενη κανονικοποίηση δείχνει πειραματική ευρωστία στο θόρυβο και στις μικρές παραμορφώσεις του σχήματος. Επιπλέον, η τεχνική είναι εξαιρετικά αποδοτική και μπορεί να ενσωματωθεί σε οποιοδήποτε σύστημα πραγματικού χρόνου για αναζήτηση με βάση το περιεχόμενο ή ακόμη και για κωδικοποίηση βίντεο, αφού το υπολογιστικό κόστος είναι αμελητέο.

Βελτίωση αλλά και επέκταση της διαδικασίας κανονικοποίησης είναι δυνατή προς αρκετές κατευθύνσεις. Πρώτον, θα μπορούσε να τροποποιηθεί ώστε να χειριστεί σύνολα από συνδεδεμένες ή μη, ανοικτές ή κλειστές καμπύλες για σκοπούς οπτικής αναγνώρισης χαρακτήρων (optical character recognition) ή δεικτοδότησης βάσεων δεδομένων (database indexing) που

περιέχουν σκίτσα ή τεχνικά σχέδια. Η κανονικοποίηση δισδιάστατων δεδομένων εικόνας / αντικειμένων πέραν των περιγραμμάτων τους, ή ακόμη και η κανονικοποίηση τρισδιάστατων μοντέλων ή πολυδιάστατων δεδομένων είναι επίσης γενικά επιθυμητή, ενώ ένα θέμα με έντονο ερευνητικό ενδιαφέρον είναι η κανονικοποίηση ως προς προοπτικούς μετασχηματισμούς.



## Μέρος II

Εκτίμηση Χαρακτηριστικών Τρισδιάστατης Δομής και Κίνησης

## ΚΕΦΑΛΑΙΟ 6

### Εκτίμηση Χαρακτηριστικών στις Τρεις Διαστάσεις

Στο Κεφάλαιο 2 είδαμε τα πολλά δισδιάστατα χαρακτηριστικά που μπορούν να εξαχθούν από εικόνες και ακολουθίες τους και αναδείξαμε τη σημασία των χαρακτηριστικών δομής και κίνησης ανάμεσά τους. Αναφορικά με τα χαρακτηριστικά τριών διαστάσεων, μπορούμε να παρατηρήσουμε ότι η φιλοσοφία στη χρήση τους δεν διαφέρει σημαντικά από την αντίστοιχη στις δύο διαστάσεις.

Πρακτικά, δεν είναι δυνατόν να εξάγουμε τρισδιάστατα χαρακτηριστικά από εικόνες, χωρίς να βασιστούμε σε κάποια ήδη εξηγημένα χαρακτηριστικά δύο διαστάσεων. Αυτό γίνεται αμέσως φανερό, εάν αναλογιστούμε ότι η πληροφορία τριών διαστάσεων είναι ουσιαστικά απύσχα λόγω προβολής στις δισδιάστατες εικόνες και ουσιαστικά προσπαθούμε να την εξάγουμε εμμέσως βασιζόμενοι σε δισδιάστατη πληροφορία. Έτσι, για παράδειγμα, δεν μπορούμε να εκτιμήσουμε την τρισδιάστατη κίνηση σωμάτων που εμφανίζονται σε ακολουθίες εικόνων, εάν πρώτα δεν εκτιμήσουμε τη δισδιάστατη (προβλεπόμενη) κίνησή τους. Επίσης, δεν μπορούμε να υπολογίσουμε την τρισδιάστατη δομή αντικειμένων, εάν πρώτα δεν εκτιμήσουμε την προβλεπόμενη δομή τους.

Ουσιαστικά, αυτό που προσπαθούμε να επιτύχουμε κατά την εξαγωγή τρισδιάστατων χαρακτηριστικών είναι να εκτιμήσουμε την απωλεσθείσα λόγω προβολής τρίτη διάσταση. Αυτό που περιέχει η τρίτη διάσταση και χάνεται λόγω προβολής είναι κατά βάση το βάθος, το οποίο ουσιαστικά αποτελεί το βασικό χαρακτηριστικό (feature) που αναζητούμε στην τρίτη διάσταση. Με τον όρο βάθος συνήθως αναφερόμαστε στην απόσταση του κάθε pixel/σημείου από την κάμερα, ούτως ώστε σε μία αεροφωτογραφία κάτοψης μίας οροσειράς θα αναζητούσαμε τις ισοϋψείς ή το ανάγλυφο. Ως μοντελοποίηση του βάθους θα μπορούσαμε να έχουμε τις αποστάσεις των αντικειμένων που εμφανίζονται σε μία εικόνα από την κάμερα ή το εύρος σε βάθος που καταλαμβάνει ένα καταγραφόμενο αντικείμενο στο χώρο. Τα τελευταία δίνει περισσότερο την έννοια του *σχετικού βάθους*, το οποίο ως χαρακτηριστικό αντικατοπτρίζει λιγότερο το βάθος και περισσότερο την σχετική τοποθέτηση των αντικειμένων κατά την τρίτη διάσταση μέσα στη σκηνή. Όταν δε αναφερόμαστε στο σχετικό βάθος των σημείων που ανήκουν σε ένα σώμα, ουσιαστικά ανακατασκευάζουμε την τρισδιάστατη δομή του. Στην πραγματικότητα, όλα τα παραπάνω χαρακτηριστικά θα μπορούσαμε να πούμε ότι αντικατοπτρίζουν την τρισδιάστατη δομή της εικόνας, της σκηνής ή των αντικειμένων που περιέχει.

Ένα ακόμη ενδιαφέρον χαρακτηριστικό που μπορεί να εξαχθεί κατά την τρίτη διάσταση είναι η *κίνηση* των αντικειμένων στον τρισδιάστατο χώρο. Εφόσον γενικά αναφερόμαστε

σε στερεά ή ελαστικά αντικείμενα, η κίνηση μοντελοποιείται περισσότερο ως η ποσότητα κίνησης κατά τους βαθμούς ελευθερίας και λιγότερο ως η κίνηση κάθε σημείου ενός σώματος. Αυτό αναδεικνύει την υπεροχή του χαρακτηριστικού κίνησης τριών διαστάσεων ως προς το συγγενές του χαρακτηριστικό στις δύο διαστάσεις, αφού λίγες παράμετροι κίνησης μπορούν να περιγράψουν πολύ ικανοποιητικά την κίνηση και την προβολή του.

Από τα παραπάνω διαφαίνεται ότι μιλώντας για εξαγωγή χαρακτηριστικών τριών διαστάσεων αναφερόμαστε κυρίως σε εξαγωγή χαρακτηριστικών τρισδιάστατης δομής και κίνησης. Αφότου εξάγουμε τη δομή ή/και την κίνηση των αντικειμένων που εμφανίζονται, μπορούμε στη συνέχεια σε συνδυασμό με άλλα δισδιάστατα χαρακτηριστικά που εξάγουμε, να εκτιμήσουμε τα τελευταία και στον τρισδιάστατο χώρο. Για παράδειγμα, έχοντας το τρισδιάστατο χαρακτηριστικό της δομής και το δισδιάστατο χαρακτηριστικό της κατανομής χρώματος μπορούμε να εκτιμήσουμε το τρισδιάστατο χαρακτηριστικό της κατανομής χρώματος. Στην πραγματικότητα, το πρόβλημα υπολογισμού τρισδιάστατης δομής και κίνησης είναι τόσο πολύπλοκο στη γενική του περίπτωση, ώστε η μοντελοποίηση άλλων τρισδιάστατων χαρακτηριστικών που βασίζονται σε αυτό να θεωρείται δευτερεύουσα.

Η εξαγωγή τρισδιάστατων χαρακτηριστικών δομής και κίνησης συγκεντρώνει σήμερα το ενδιαφέρον της ερευνητικής κοινότητας της υπολογιστικής όρασης. Ακριβώς λόγω της πολυπλοκότητας του προβλήματος αλλά και των ζητημάτων ασάφειας και ευαισθησίας στο θόρυβο, το πρόβλημα έχει προσεγγιστεί με πολλούς διαφορετικούς τρόπους. Υπάρχει σήμερα μία μεγάλη κατηγορία προβλημάτων που είναι γνωστά και ως *δομή από X* (structure from X), των οποίων στόχος είναι η εξαγωγή του χαρακτηριστικού της δομής από οποιαδήποτε χρήσιμη πληροφορία (χαρακτηριστικά δύο διαστάσεων). Πρακτικά το X σήμερα αντιστοιχεί στα χαρακτηριστικά κίνησης, σκίασης, εστίασης, ανεστίασης κ.λπ. περιγράφοντας τις μεθόδους δομής από κίνηση (structure from motion - SFM) [188, 148, 159], δομής από σκίαση (structure from shading - SFS) [199], δομής από εστίαση (structure from focus - SFF) [4] ή ανεστίαση (structure from defocus - SFD) [78]. Η επιτυχία των μεθόδων αυτών εξαρτάται από πολλούς παράγοντες, σημαντικότερος εκ των οποίων είναι η επιτυχία με την οποία εξάγονται τα αντίστοιχα δισδιάστατα χαρακτηριστικά (κίνηση, σκίαση, εστίαση κ.λπ.). Θα έλεγε κανείς σήμερα ότι μεγαλύτερο ενδιαφέρον παρουσιάζουν οι μέθοδοι δομής από κίνηση, ακριβώς γιατί μπορούν να εκμεταλλευτούν τη μεγάλη πληροφορία δισδιάστατης κίνησης που εξάγεται από μία ακολουθία εικόνων αλλά και γιατί η τρισδιάστατη δομή και κίνηση συνδέονται με άμεσο τρόπο με τα αντίστοιχα δισδιάστατα χαρακτηριστικά.

Σε κάθε μία από τις παραπάνω κατηγορίες μεθόδων, ειδικότερα δε στις μεθόδους δομής από κίνηση που συγκεντρώνουν το ενδιαφέρον στις τελευταίες δύο δεκαετίες, εμφανίζεται με-

γάλη διαφοροποίηση μεθόδων, κυρίως γιατί οι επιλογές που πραγματοποιούνται σε αυτές (ως προς τα δεδομένα εισόδου, τον τρόπο αντιστοίχισής τους, το θεωρούμενο μοντέλο προβολής, τον τρόπο επεξεργασίας των δεδομένων) μπορούν να είναι ποικίλες, αλλά και κατάλληλες για διαφορετικού είδους εφαρμογές. Ας θεωρήσουμε ως ενδεικτικό παράδειγμα τις μεθόδους υπολογισμού τρισδιάστατης δομής και κίνησης. Ως προς τα δεδομένα εισόδου (input data), αυτά μπορούν να είναι αντιστοιχίσεις σημείων, αντιστοιχίσεις ακμών, αντιστοιχίσεις καμπυλών ή πυκνά πεδία διανυσμάτων κίνησης. Ως προς τον τρόπο αντιστοίχισης (feature correspondences), ανάλογα με τη διάσταση των χαρακτηριστικών που παρακολουθούμε, δισδιάστατα (2Δ) ή τρισδιάστατα (3Δ), διακρίνουμε 3Δ-σε-3Δ, 2Δ-σε-2Δ και 3Δ-σε-2Δ αντιστοιχίσεις. Το θεωρούμενο μοντέλο προβολής (projection model) μπορεί να είναι το σύνθετες προοπτικό ή το παράλληλο ή ακόμη και ενδιάμεσες προσεγγίσεις τους. Ο τρόπος επεξεργασίας είναι συνήθως μαζικός (batch processing) ή επαναληπτικός (recursive processing), ανάλογα με το εάν τα δεδομένα διατίθενται μαζικά ή σταδιακά με το χρόνο.

Το κλασικό πρόβλημα δομής από κίνηση αφορά στις 2Δ-σε-2Δ αντιστοιχίσεις, όπου επιθυμούμε να ανακατασκευάσουμε τη δομή ενός τρισδιάστατου αντικειμένου από τις προβολές του στα καρέ μίας ακολουθίας εικόνων (πρόβλημα SFM). Το πρόβλημα αυτό απαντάται συνήθως ως τρισδιάστατη ανακατασκευή από προβολές ή υπολογισμός δομής από κίνηση ή εκτίμηση τρισδιάστατης δομής και κίνησης (3D reconstruction, 3D motion and structure estimation). Η μοναδική διαθέσιμη πληροφορία σε αυτήν την περίπτωση είναι η πληροφορία δισδιάστατης κίνησης που λαμβάνεται από τις προβολές μέσω κάποιας τεχνικής υπολογισμού κίνησης (2D motion estimation). Στην περίπτωση που θέλουμε να ανακατασκευάσουμε μέρος της εικόνας που αντιστοιχεί σε ένα μοναδικό αντικείμενο, αναγκαίο είναι να το εξάγουμε από την εικόνα, για παράδειγμα με κάποια τεχνική κατάτμησης κίνησης (motion segmentation). Ανάλογα με το μοντέλο προβολής, τα χρησιμοποιούμενα χαρακτηριστικά και τη μέθοδο επεξεργασίας, διάφορες τεχνικές έχουν προταθεί [188, 162, 149, 99].

Οι 3Δ-σε-3Δ αντιστοιχίσεις αφορούν προβλήματα όπου γνωρίζουμε την τρισδιάστατη κίνηση του σώματος και θέλουμε να υπολογίσουμε τη δομή του, ή ακόμη γνωρίζουμε και τη δομή και την κίνηση κάποιου χώρου και επιθυμούμε να τοποθετηθούμε μέσα σε αυτόν. Ως παράδειγμα μπορούμε να δούμε εφαρμογές χωροθέτησης αντικειμένων με τεχνικές στέρεο (positioning of 3D objects using stereo) [16]. Αντίστοιχα, οι 3Δ-σε-2Δ αντιστοιχίσεις [83, 184] αφορούν εφαρμογές βαθμονόμησης κάμερας (camera calibration) όταν γνωρίζουμε τη θέση κάποιων χαρακτηριστικών του χώρου που καταγράφεται ή παθητικής καθοδήγησης οχημάτων (passive navigation) βασιζόμενοι σε σημάνσεις (landmarks).

Σχεδόν το σύνολο των εφαρμογών αναφέρονται σε στερεά (rigid) τρισδιάστατα αντικείμενα. Ενδιαφέρον παρουσιάζει επίσης η περίπτωση παραμορφώσιμων (deformable) τρισδιάστατων αντικειμένων και σημαντική προσπάθεια γίνεται προς την κατεύθυνση της ανακατασκευής τους μέσω κατάλληλης μοντελοποίησής τους [156, 96, 113]. Ουσιαστικά, το πρόβλημα της ανακατασκευής των στερεών αντικειμένων, πολύ περισσότερο δε των παραμορφώσιμων, παρουσιάζει σημαντική ασάφεια λόγω της απώλειας της τρίτης διάστασης. Η ασάφεια αυτή είναι πολύ σημαντικότερη και πολύ δυσκολότερα επιλύσιμη από ασάφειες / προβλήματα που εμφανίζονται κατά την εξαγωγή δισδιάστατων χαρακτηριστικών. Για παράδειγμα, ένα άπειρο επίπεδο που βρίσκεται μακριά από την κάμερα προβάλλεται ακριβώς με τον ίδιο τρόπο πάνω σε αυτήν. Επίσης, δεν μπορούμε να ξεχωρίσουμε την κίνηση κατά τη διεύθυνση του οπτικού άξονα ενός αντικειμένου που το πάχος του είναι μικρό σε σχέση με την απόστασή του από την κάμερα. Επίσης, οι εκτιμήσεις για τρισδιάστατα μεγέθη υποφέρουν από όλα τα προβλήματα στην ανάκτηση των δισδιάστατων χαρακτηριστικών, αφού προκύπτουν μέσω αυτών. Τέλος, επειδή ακριβώς έχουμε στη διάθεσή μας πληροφορία προβολής, τόσο μεγαλύτερη είναι η εισερχόμενη ασάφεια, όσο περισσότεροι είναι οι βαθμοί ελευθερίας των τρισδιάστατων σωμάτων που καταγράφονται. Γι'αυτόν ακριβώς το λόγο χρησιμοποιούμε όσο το δυνατόν περισσότερη πληροφορία, πρακτικά όσες περισσότερες αντιστοιχίσεις χαρακτηριστικών και όσα περισσότερα καρέ είναι δυνατόν. Στην οριακή περίπτωση, είναι αδύνατον να ανακατασκευάσουμε ένα σώμα που κινείται χωρίς περιορισμούς στερεότητας ή ελαστικότητας, π.χ. ένα ρευστό. Φανερό είναι δε ότι δεν μπορούμε να ανακτήσουμε πληροφορία που δεν υπάρχει στα διαθέσιμα δισδιάστατα χαρακτηριστικά, π.χ. τη μορφή μίας επιφάνειας που επικαλύπτεται από το ίδιο ή από άλλα σώματα.

Πέρα από τις ασάφειες, οι εκτιμήσεις τρισδιάστατων χαρακτηριστικών υποφέρουν και από το θόρυβο ή τα λάθη που υπεισέρχονται στις μετρήσεις δύο διαστάσεων. Για παράδειγμα, είναι σαφές ότι οι μέθοδοι υπολογισμού δισδιάστατης κίνησης υποφέρουν από σοβαρές ανακρίβειες. Οι ανακρίβειες αυτές είναι σε συνάρτηση με το υπολογιστικό κόστος που οι μέθοδοι αυτές απαιτούν για να προσεγγίσουν καλύτερα την πραγματική κίνηση, αλλά και σε προβλήματα επικαλύψεων, φωτισμού και της διακριτικότητας του pixel. Σε γενικές γραμμές, λόγω του ότι οι μέθοδοι ανακατασκευής βασίζονται σε διαφορικούς υπολογισμούς στα πεδία κίνησης, οι υπάρχουσες ανακρίβειες 'πολλαπλασιάζονται'.

Στα επόμενα κεφάλαια του δεύτερου μέρους παρουσιάζουμε πρωτότυπες μεθόδους εξαγωγής χαρακτηριστικών δομής και κίνησης στις τρεις διαστάσεις. Ουσιαστικά εξετάζουμε το πρόβλημα ανακατασκευής στερεών αντικειμένων από 2Δ-σε-2Δ αντιστοιχίσεις σημείων σε παράλληλη προβολή και τον υπολογισμό της κίνησης της κάμερας από 3Δ-σε-2Δ αντιστοι-

χίσεις γραμμών σε προοπτική προβολή. Στην ανάλυση, σχολιάζουμε και παραπέμπουμε σε συγγενείς κατά οποιονδήποτε τρόπο μεθόδους της βιβλιογραφίας, μελετούμε την επίδοση των προτεινόμενων μεθόδων σε πειραματικές ακολουθίες και σχολιάζουμε την επίδοση σε ακρίβεια και υπολογιστικό κόστος. Στα συμπεράσματα κάθε κεφαλαίου αναφερόμαστε ενδεικτικά σε μελλοντικές επεκτάσεις και σχολιάζουμε τη συνεισφορά της εργασίας μας στο γενικότερο αντικείμενο.

## ΚΕΦΑΛΑΙΟ 7

### Υπολογισμός Τρισδιάστατης Δομής και Κίνησης από Προβολές Τρισδιάστατων Στερεών Αντικειμένων

#### 7.1 Εισαγωγή / Αναδρομή

Ο υπολογισμός της τρισδιάστατης δομής και κίνησης (3D structure and motion) κινούμενων στερεών αντικειμένων που καταγράφονται σε διαφορετικές χρονικές στιγμές και/ή από πολλαπλές κάμερες, έχει απασχολήσει σημαντικά τις περιοχές της Όρασης Υπολογιστών και της Κατανόησης Περιεχομένων Βίντεο την τελευταία δεκαετία. Όπως προτείνουν οι Huang και Netravali [60], τρεις κυρίως τύποι σχετικών προβλημάτων μπορούν να τεθούν σε σχέση με τον τύπο της αντιστοίχισης των χαρακτηριστικών που παρακολουθούνται (feature correspondence). Η τρισδιάστατη-σε-τρειςδιάστατη (3Δ-σε-3Δ) αντιστοίχιση χαρακτηριστικών χρησιμοποιείται για τον υπολογισμό κίνησης ή τη χωροθέτηση (positioning) τρισδιάστατων αντικειμένων με τεχνικές στέρεο. Η δισδιάστατη-σε-τρειςδιάστατη (2Δ-σε-3Δ) αντιστοίχιση χαρακτηριστικών χρησιμοποιείται για αυτόματη πλοήγηση (passive navigation) οχημάτων με τρισδιάστατα σημεία αναφοράς (landmarks), αλλά και για διαμέτρηση (calibration) της κάμερας. Η δισδιάστατη-σε-δισδιάστατη (2Δ-σε-2Δ) αντιστοίχιση χαρακτηριστικών λαμβάνεται για τον υπολογισμό των παραμέτρων κίνησης της κάμερας (egomotion), για τον υπολογισμό των σχετικών παραμέτρων σε συστήματα από κάμερες που παρακολουθούν τα ίδια χαρακτηριστικά, ή για τον υπολογισμό της τρισδιάστατης δομής και κίνησης αντικειμένων που κινούνται σε σχέση με μία κάμερα. Στην τελευταία περίπτωση δίνεται ιδιαίτερη προσοχή σήμερα, λόγω και των πρόσφατων αντικειμενοστρεφών (object-based) προτύπων κωδικοποίησης MPEG-4 και MPEG-7.

Η εξαγωγή των παραμέτρων δομής και κίνησης ενός κινούμενου στερεού τρισδιάστατου (3Δ) αντικειμένου από μία δισδιάστατη (2Δ) ακολουθία εικόνων ονομάζεται συνήθως ως το πρόβλημα Δομής Από Κίνηση (Structure From Motion ή SFM), γιατί βασίζεται στην παρατηρούμενη 2Δ προβλεβλημένη κίνηση του σώματος για να υπολογίσει τη δομή του - στη συνέχεια θα αναφερόμαστε σε αυτό με τα αρχικά SFM. Το SFM πρόβλημα έχει απασχολήσει στο παρελθόν αρκετούς ερευνητές. Διάφορες προσεγγίσεις έχουν προταθεί στη διεθνή βιβλιογραφία, οι οποίες διαφέρουν στο θεωρούμενο μοντέλο προβολής (projection model), στα χαρακτηριστικά (features) και τις μετρήσεις εισόδου (input measurements) που χρησιμοποιούνται [1], καθώς και στη μέθοδο επεξεργασίας των δεδομένων (data processing) [148]. Όσον αφορά στα 2Δ χαρακτηριστικά, αντιστοιχίσεις ευθειών, καμπυλών και σημείων έχουν χρησιμοποιηθεί [35], με τις αντιστοιχίσεις σημείων να αποτελούν την πιο δημοφιλή περίπτωση. Τα δύο

συχνότερα χρησιμοποιούμενα μοντέλα προβολής είναι το προοπτικό (perspective) και το παράλληλο (orthographic), με το τελευταίο να αποτελεί μία εξαίρετη προσέγγιση του πραγματικού προοπτικού όταν το αντικείμενο βρίσκεται μακριά από την κάμερα [60]. Ακριβείς μαθηματικές λύσεις έχουν δοθεί και για τις δύο περιπτώσεις συμπεριλαμβάνοντας για παράδειγμα τις [84, 161, 163, 162, 150] για προοπτική προβολή και [61, 2, 159] για παράλληλη προβολή. Διαφορετικού τύπου προβολές βασισμένες σε προσεγγίσεις της προοπτικής και της παράλληλης προβολής έχουν επίσης προταθεί, όπως είναι η ασθενής προοπτική / αναλογική παράλληλη (weak perspective / scaled orthographic) και η ορθοπροοπτική (orthoperspective) προβολή (βλ. για παράδειγμα την εργασία των DeMenthon και Davis [27] και τις εκεί αναφορές).

Όσον αφορά την περίπτωση της παράλληλης προβολής, είναι εμφανές ότι το απόλυτο βάθος ενός αντικειμένου δεν μπορεί να υπολογιστεί, αφού η προβολή του στο επίπεδο της κάμερας δεν επηρεάζεται από οποιαδήποτε 3Δ μετακίνηση του αντικειμένου κάθετη στην επιφάνεια προβολής. Όπως υποδεικνύεται στη βιβλιογραφική μελέτη των Huang και Netravali [60] στο SFM πρόβλημα, η παράλληλη προβολή είναι μία καλή προσέγγιση της προοπτικής όταν το αντικείμενο είναι σχετικά μακριά από την κάμερα. Ακριβέστερα, ο Tekalp [155] συμπληρώνει ότι η παράλληλη προβολή δίνει αποδεκτά πειραματικά αποτελέσματα όταν η μέγιστη παρατηρούμενη σχετική διάσταση βάθους του αντικειμένου (μεταξύ δηλαδή των 3Δ σημείων της επιφανείας του) είναι μικρότερη από το 10% της απόστασης του κέντρου βάρους του από το επίπεδο προβολής. Στο κεφάλαιο αυτό, ασχολούμαστε κυρίως με το μοντέλο της παράλληλης προβολής, τα συμπεράσματά του όμως μπορούν να αναχθούν και στις περιπτώσεις προοπτικής προβολής και των προσεγγίσεων αυτών.

Ο Ullman στην κλασική του εργασία [167] αποδεικνύει ότι τέσσερις αντιστοιχίσεις σημείων (point correspondences) σε τρία καρέ (frames) είναι αρκετές για να δώσουν μοναδική λύση στο SFM πρόβλημα σε παράλληλη προβολή, με αβεβαιότητα καθρεπτισμού του σώματος ως προς το επίπεδο προβολής λόγω συμμετρίας. Το συμπέρασμα αυτό επιβεβαιώνεται σε όλους τους προτεινόμενους αλγόριθμους (για παράδειγμα [61, 25]). Στην πραγματικότητα, χρησιμοποιώντας τρία καρέ, ισοδύναμα δύο μεταβάσεις (transitions), είναι δυνατό να υπολογίσουμε το σχετικό βάθος των σημείων του σώματος, δηλαδή τη δομή του 3Δ αντικειμένου. Οι Huang και Lee [61] προτείνουν ένα γραμμικό αλγόριθμο για να εξάγουν τις παραμέτρους τριδιάστατης δομής και κίνησης ενός αντικειμένου χρησιμοποιώντας τέσσερις αντιστοιχίσεις σημείων σε τρία καρέ. Οι Aizawa *et al* [2] χρησιμοποιούν μικρές γωνίες Euler, δηλαδή μικρές γωνίες περιστροφής, και έναν επαναληπτικό αλγόριθμο δύο βημάτων για υπολογισμό της 3Δ κίνησης και δομής, όπου θεωρούν ότι μία αρχική εκτίμηση του βάθους των 3Δ σημείων είναι διαθέσιμη a-priori βάσει ενός τριδιάστατου μοντέλου. Αυτό είναι βολικό στην περίπτωση



που εξετάζουν, αφού θεωρούν μόνο την περίπτωση κινούμενων προσώπων σε εφαρμογές τηλεδιάσκεψης (videoconferencing), ενώ πράγματι γενικά 3Δ μοντέλα προσώπων για τέτοιους σκοπούς είναι σήμερα διαθέσιμα. Στην εργασία τους [18], οι Bozdagi *et al* ορίζουν ένα κριτήριο διόρθωσης σφάλματος και προτείνουν ένα προσθετικό βήμα αναζήτησης με βάση την παράγωγο στον παραπάνω αλγόριθμο. Η περισσότερο ακριβής μαθηματικά λύση από τις παραπάνω είναι πιθανότατα η [61]. Η μέθοδος αυτή δίνει ακριβείς εκτιμήσεις των παραμέτρων 3Δ περιστροφής των σωμάτων όταν απόλυτα ακριβείς μετρήσεις των διανυσμάτων 2Δ κίνησης (2D motion vectors) έχουν εξαχθεί. Εντούτοις, δεν χειρίζεται την περίπτωση διανυσμάτων κίνησης προσβεβλημένων από θόρυβο, οπότε στη ρεαλιστική αυτή περίπτωση ο αλγόριθμος αποτυγχάνει.

Πιο σύγχρονες προσεγγίσεις στο πρόβλημα SFM με παράλληλη προβολή περιλαμβάνουν τη μέθοδο factorization των Tomasi και Kanade [159] και τα ποικίλα παράγωγά της, τις epipolar μεθόδους των Shapiro *et al* [140], Ostuni και Dunn [106], Xu και Sugimoto [196], καθώς και τη μέθοδο των κινούμενων επιπέδων των Ξηρουχάκη και Ντελόπουλου [188, 25]. Οι epipolar μέθοδοι γενικά αναφέρονται σε εικόνες υπό ασθενή προοπτική προβολή (weak perspective), εντούτοις, μπορεί ναδειχθεί ότι το μοντέλο ασθενούς προοπτικής προβολής διαφέρει από αυτό της παράλληλης μόνο κατά ένα παράγοντα αναλογίας (scaling) μεταξύ διαφορετικών όψεων του αντικειμένου.

Η λύση που προτείνουν οι Tomasi και Kanade [159] στο SFM πρόβλημα βασίζεται στη θεώρηση της κάμερας ως κέντρο της αναπαράστασης (camera-based model), ενώ η μορφή της λύσης επιτρέπει τη χρησιμοποίηση τυχαίου (μεγάλου) αριθμού σημείων και μεταβάσεων για να επιτύχει καλύτερη εκτίμηση της λύσης παρουσία θορύβου. Η λύση βασίζεται στην ανάλυση (decomposition) του πίνακα που περιέχει τις μετρήσεις σε δύο άλλους πίνακες, που με τη σειρά τους περιέχουν τις παραμέτρους κίνησης και σχήματος αντίστοιχα. Η μέθοδος αυτή, η οποία έχει επικρατήσει με το όνομα factorization, επεκτείνεται στην περίπτωση της ασθενούς προοπτικής προβολής από τους Poelman και Kanade [122], ενώ οι Morita και Kanade [99] δίνουν μέσα από ένα λίγο διαφορετικό φορμαλισμό στη μέθοδο τη δυνατότητα ακολουθιακής επεξεργασίας (sequential processing) στο χρόνο. Όσον αφορά στις epipolar μεθόδους, οι Shapiro *et al* βασίζονται στις ιδιότητες των affine epipolar γραμμών (όπως ονομάζονται) και λύνουν την αντίστοιχη εξίσωση κίνησής τους. Σε ένα επόμενο βήμα, όλες οι άγνωστες 3Δ παράμετροι κίνησης υπολογίζονται (βλέπε [140] και τις εκεί αναφορές). Με τον ίδιο τρόπο, οι Xu και Sugimoto [196] λύνουν την epipolar εξίσωση, για να βρουν τις τρεις γωνίες περιστροφής (γωνίες Euler) σε ένα δεύτερο βήμα. Οι Ostuni και Dunn [106] επίσης χρησιμοποιούν την epipolar εξίσωση, αλλά με διαφορετική παραμετροποίηση του πίνακα περιστροφής. Το

κεφάλαιο αυτό θα ασχοληθεί κυρίως με τη μέθοδο των Xirouhakis και Delopoulos [188, 25], η οποία φαίνεται να υπερέχει από τις δύο παραπάνω κατηγορίες μεθόδων.

Οι περισσότερες εργασίες που ασχολούνται με το SFM πρόβλημα, υποθέτουν ότι το παρατηρούμενο αντικείμενο είναι απόλυτα στερεό (μη-παραμορφώσιμο). Αυτή η παραδοχή δεν μπορεί να ξεπεραστεί εύκολα, λόγω κυρίως της πολυπλοκότητας στον υπολογισμό της 3Δ δομής και κίνησης παραμορφώσιμων σωμάτων. Πρώιμες προσεγγίσεις παρουσία παραμόρφωσης των στόχων περιλαμβάνουν την μοντελοποίησή της μέσω ελεύθερων ταλαντώσεων ή συστημάτων εσωτερικών και εξωτερικών δυνάμεων [96, 113, 156]. Στην περίπτωση που περισσότερα του ενός αντικείμενα υπάρχουν στη σκηνή, εμφανίζεται το πρόβλημα της μεταξύ τους μερικής επικάλυψης (occlusion). Το τελευταίο επηρεάζει τη διαδικασία υπολογισμού της 2Δ κίνησης. Το θέμα της παρακολούθησης αντικειμένων παρουσία επικαλύψεων μπορεί να αντιμετωπιστεί με διάφορες μεθόδους και είναι έξω από τα πλαίσια αυτού του κεφαλαίου. Λύσεις παρουσία θορύβου / λαθών στα 2Δ δεδομένα εισόδου (διανύσματα 2Δ κίνησης) προτείνονται στις αναφορές [119, 176, 177, 175] για προοπτικές προβολές. Για παράλληλες προβολές, λύσεις παρουσία θορύβου προτείνονται στις αναφορές [159, 140, 106, 196, 25, 188].

Στο κεφάλαιο αυτό ασχολούμαστε με παράλληλες προβολές. Συγκεκριμένα αναλύουμε τη μέθοδο των Xirouhakis και Delopoulos [188] για το SFM πρόβλημα για τέσσερις ή περισσότερες αντιστοιχίσεις σημείων στερεών σωμάτων σε δύο μεταβάσεις. Μερικά από τα βασικά συμπεράσματα της έρευνας αυτής παρουσιάστηκαν πρώτα από τους Delopoulos και Xirouhakis στην [25].

Αρχικά, καθορίζουμε ένα σύνολο σχέσεων μεταξύ των παραμέτρων κίνησης, όπως αυτές ορίζονται από τους πίνακες περιστροφής και μετατόπισης, και το 2Δ πεδίο κίνησης (2D motion field). Εν συνεχεία, αποδεικνύουμε ότι τα στοιχεία του πίνακα περιστροφής (rotation matrix) μπορούν να υπολογιστούν μέσω των ιδιοδιανυσμάτων (eigenvectors) και των ιδιοτιμών (eigenvalues) κατάλληλα ορισμένων  $2 \times 2$  πινάκων - τα τελευταία είναι απλές εκφράσεις τεσσάρων διανυσμάτων κίνησης σε δύο μεταβάσεις (τρία καρέ). Τέλος, αποδεικνύουμε ότι οι προκύπτουσες εκφράσεις μπορούν να χρησιμοποιηθούν για τον υπολογισμό των εμπλεκόμενων πινάκων περιστροφής κατά τη λογική των ελάχιστων τετραγώνων (least squares), στην περίπτωση που είναι διαθέσιμα περισσότερα από τέσσερα διανύσματα κίνησης ανά μετάβαση. Ακριβής ασυμπτωτική ανάλυση της προτεινόμενης διαδικασίας υπολογισμού με ελάχιστα τετράγωνα υποδεικνύει ότι τα αποτελέσματα που προκύπτουν δεν επηρεάζονται από την παρουσία θορύβου στα 2Δ διανύσματα κίνησης.

Η Ενότητα 7.2 του κεφαλαίου δίνει τους απαραίτητους ορισμούς / συμβολισμούς και ορίζει τους πίνακες  $\mathbf{K}$  και  $\tilde{\mathbf{K}}$ , οι οποίοι αποτελούν τις κυρίως χρησιμοποιούμενες ενδιάμεσες ποσότητες για την έκφραση των παραμέτρων περιστροφής σε σχέση με τα 2Δ διανύσματα κίνησης. Αποδεικνύεται ότι οι πίνακες  $\mathbf{K}$  και  $\tilde{\mathbf{K}}$  εξάγονται απευθείας από τα τελευταία.

Η Ενότητα 7.3 παρέχει τον προτεινόμενο αλγόριθμο για τον υπολογισμό των πινάκων περιστροφής  $\mathbf{R}$  και  $\mathbf{S}$ , που αντιστοιχούν σε δύο διαδοχικές περιστροφές ενός 3Δ αντικειμένου βάσει ενός ικανού αριθμού από  $\mathbf{K}$  πίνακες. Τα τρία βήματα του αλγορίθμου δίνονται με τη μορφή Προτάσεων (7.3-7.5) και περιλαμβάνουν τον απλό υπολογισμό ιδιοτιμών και ιδιοδιανυσμάτων  $2 \times 2$  πινάκων. Η ανάλυση της Ενότητας 7.3 αναφέρεται σε πραγματικά (χωρίς θόρυβο) πεδία διανυσμάτων κίνησης. Οι οριακές συνθήκες / περιπτώσεις του SFM προβλήματος σε παράλληλη προβολή δεν εξετάζονται εδώ, αφού έχουν σε μεγάλο βαθμό εξεταστεί στη διεθνή βιβλιογραφία (βλ. για παράδειγμα [18, 196]).

Στην Ενότητα 7.4 δείχνουμε τον τρόπο με τον οποίο οι υπολογισμοί της Ενότητας 7.3 μπορούν να δώσουν συνεπείς εκτιμήτριες (estimators) των παραμέτρων περιστροφής. Οι Προτάσεις 7.6, 7.7 και 7.8 της ενότητας - ευθέως αντίστοιχες με τις 7.3, 7.4 και 7.5 - παρέχουν ισχυρά συνεπείς εκτιμήτριες, στη λογική των ελάχιστων τετραγώνων των πινάκων  $\mathbf{R}$  και  $\mathbf{S}$ . Η ασυμπτωτική συμπεριφορά των αποτελεσμάτων εξετάζεται σε σχέση με τον αριθμό των διανυσμάτων κίνησης που περιλαμβάνονται στους υπολογισμούς και την τοποθέτηση των διανυσμάτων θέσης. Οι προκύπτουσες εκφράσεις της διασποράς του θορύβου αποτελούν σημαντικό βοήθημα στην επιλογή των διανυσμάτων κίνησης που θα περιληφθούν στους υπολογισμούς. Οι Προτάσεις 7.6-7.8 βασίζονται στη χρήση των Λημμάτων 7.1 και 7.2 της ίδιας ενότητας. Κατάλληλες αποδείξεις συμπεριλαμβάνονται όπου κρίνεται απαραίτητο.

Στην Ενότητα 7.5 ο προτεινόμενος αλγόριθμος συγκρίνεται θεωρητικά με υπάρχουσες μεθόδους για παράλληλες προβολές. Η Ενότητα 7.6 του κεφαλαίου παρουσιάζει ένα σύνολο πειραμάτων, τα οποία επιβεβαιώνουν τα αποτελέσματα των Ενότητων 7.3 και 7.4. Στην τελευταία υποενότητα της Ενότητας 7.6, η επίδοση του αλγορίθμου εξετάζεται απέναντι σε δημοφιλείς αλγορίθμους για τον υπολογισμό της 3Δ δομής και κίνησης από παράλληλες προβολές [159, 196, 140]. Συμπεράσματα και ζητήματα συζήτησης, περιλαμβάνοντας θέματα μελλοντικής επέκτασης και έρευνας, δίνονται στην Ενότητα 7.7.

Οι αποδείξεις για τις οκτώ Προτάσεις και τα δύο Λήμματα παρουσιάζονται όσο το δυνατόν πιο συνοπτικά. Εντούτοις, οι αναγνώστες που επιθυμούν να δώσουν έμφαση στο αλγοριθμικό μέρος του κεφαλαίου μπορούν να τις παραλείψουν σε πρώτη ανάγνωση.

## 7.2 Ο πίνακας περιστροφής επιπέδου

### 7.2.1 Το μοντέλο παράλληλης προβολής

Κάθε κίνηση ενός στερεού αντικειμένου στον 3Δ χώρο είναι υπέρθεση μίας 3Δ περιστροφής (rotation) και μίας 3Δ μετατόπισης (translation). Επομένως, όταν ένα σημείο  $(x, y, z)$  του σώματος κινείται στο  $(x', y', z')$ , ισχύει:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{T} \quad (7-1)$$

όπου  $\mathbf{R}$  είναι ο  $3 \times 3$  πίνακας περιστροφής, ενώ  $\mathbf{T}$  το  $3 \times 1$  διάνυσμα μετατόπισης. Στη συνέχεια, όταν αναφερόμαστε στα στοιχεία τους, υποθέτουμε ότι

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^T.$$

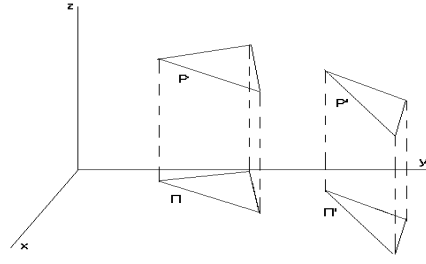
Τα στοιχεία του πίνακα περιστροφής  $r_{mn}$  υπακούουν στις συνθήκες ορθογωνιότητας, σύμφωνα με το γεγονός ότι ο  $\mathbf{R}$  είναι ορθομοναδιαίος, δηλαδή,  $\mathbf{R}^T \mathbf{R} = \mathbf{R} \mathbf{R}^T = \mathbf{I}$ .

### 7.2.2 Προκαταρκτικοί ορισμοί

Η παρακάτω ανάλυση, τόσο σε επίπεδο περιγραφής της 3Δ κίνησης και της 2Δ προβολής στερεού αντικειμένου, όσο και ως προς τον ορισμό των βασικών ποσοτήτων (πίνακες  $\mathbf{K}$  κ.λπ.), δόθηκε πρώτη φορά στην πρωτότυπη διπλωματική εργασία της αναφοράς [181]. Αρκετά σημαντικές διαφοροποιήσεις παρουσιάζονται εδώ, κυρίως ως προς τις αποδείξεις αλλά και τα απαραίτητα κατά περίπτωση σχόλια.

Έστω ένα σύνολο  $P$  περιέχον τρία σημεία, τα οποία με τη σειρά τους ορίζουν μία επίπεδη επιφάνεια στον 3Δ χώρο. Έστω επίσης ότι τα σημεία αυτά προβάλλονται σε μία τριπλέτα σημείων  $\Pi$  στο επίπεδο προβολής. Στη συνέχεια, τέτοιες επίπεδες επιφάνειες θα ονομάζονται *τρίγωνα*. Οι 3Δ χώροι θα καλούνται *σκηνές*, ενώ οι προβολές τους στο επίπεδο προβολής *καρέ*. Θεωρούμε ότι το τρίγωνο  $P$  'κινείται' σε ένα τρίγωνο  $P'$  στη δεύτερη σκηνή και προβάλλεται στο  $\Pi'$  στο δεύτερο καρέ (Σχήμα 7.1).

Έστω  $\boldsymbol{\nu}$ ,  $\boldsymbol{\mu}$  τα μοναδιαία διανύσματα στο επίπεδο που ορίζει το τρίγωνο  $P$  τα οποία κινούνται στα  $\boldsymbol{\nu}'$ ,  $\boldsymbol{\mu}'$  στο τρίγωνο  $P'$ . Έστω επίσης οι προβολές τους  $\mathbf{n}$ ,  $\mathbf{m}$  και  $\mathbf{n}'$ ,  $\mathbf{m}'$  αντίστοιχα. Γενικά, στο κεφάλαιο αυτό, θα αναφερόμαστε με Ελληνικούς χαρακτήρες στα 3Δ διανύσματα και με Λατινικούς στα 2Δ διανύσματα. Θεωρήστε ένα σημείο  $\boldsymbol{\rho} = (x, y, z)$  στο επίπεδο του  $P$  το οποίο κινείται στο  $\boldsymbol{\rho}' = (x', y', z')$  του  $P'$  και τις παράλληλες προβολές τους



**Σχήμα 7.1** Κίνηση επίπεδης επιφάνειας στον 3Δ χώρο.

$\mathbf{r} = (x, y)$ ,  $\mathbf{r}' = (x', y')$  αντίστοιχα. Η επιφάνεια προβολής (image plane) θεωρείται, χωρίς βλάβη της γενικότητας, ως το επίπεδο  $xy$  ( $z = 0$ ). Τότε, για κάποια  $a, b$ ,

$$\boldsymbol{\rho} = a\boldsymbol{\nu} + b\boldsymbol{\mu} + \boldsymbol{\gamma}, \quad \boldsymbol{\rho}' = a\boldsymbol{\nu}' + b\boldsymbol{\mu}' + \boldsymbol{\gamma}'$$

όπου το  $\boldsymbol{\gamma}$  αντιστοιχεί σε ένα τυχαίο σταθερό (δηλαδή, ανεξάρτητο του  $\boldsymbol{\rho}$ ) σημείο του  $P$  και  $\boldsymbol{\gamma}'$  είναι το αντίστοιχό του στο  $P'$ . Εάν τα διανύσματα  $\boldsymbol{\gamma}, \boldsymbol{\delta}$  προβάλλονται στα  $\mathbf{c}, \mathbf{d}$ , αντίστοιχα, ισχύει

$$\mathbf{r} = a\mathbf{n} + b\mathbf{m} + \mathbf{c}, \quad \mathbf{r}' = a\mathbf{n}' + b\mathbf{m}' + \mathbf{c}'. \quad (7-2)$$

Τότε, το διάνυσμα κίνησης  $\mathbf{v}$  που αντιστοιχεί στην προβαλλόμενη κίνηση του σημείου δίνεται από την

$$\mathbf{v} = \mathbf{r}' - \mathbf{r} = a(\mathbf{n}' - \mathbf{n}) + b(\mathbf{m}' - \mathbf{m}) + (\mathbf{c}' - \mathbf{c}) \quad (7-3)$$

Οι εξισώσεις (7-2) και (7-3) υποδεικνύουν ότι (α) όλα τα  $\mathbf{r}$  που ανήκουν σε ένα συγκεκριμένο τρίγωνο (επίπεδο) μπορούν να εκφραστούν ως ένας affine μετασχηματισμός των  $\mathbf{n}, \mathbf{m}$ , (β) όλα τα  $\mathbf{r}'$  ως ένας affine μετασχηματισμός των  $\mathbf{n}', \mathbf{m}'$  και (γ) όλα τα  $\mathbf{v} \triangleq \mathbf{r}' - \mathbf{r}$  μπορούν να εκφραστούν ως affine μετασχηματισμός των  $\mathbf{n}' - \mathbf{n}, \mathbf{m}' - \mathbf{m}$ .

Αξιοποιώντας την παραπάνω παρατήρηση, ορίζουμε τους παρακάτω  $2 \times 2$  πίνακες  $\mathbf{K}, \tilde{\mathbf{K}}$  ως

$$\mathbf{K} \triangleq [\mathbf{n}' \ \mathbf{m}'] [\mathbf{n} \ \mathbf{m}]^{-1} \quad (7-4)$$

$$\tilde{\mathbf{K}} \triangleq [\mathbf{n}' - \mathbf{n} \ \mathbf{m}' - \mathbf{m}] [\mathbf{n} \ \mathbf{m}]^{-1} = \mathbf{K} - \mathbf{I}. \quad (7-5)$$

Στη συνέχεια, θα δειχθεί ότι αυτοί οι πίνακες χαρακτηρίζουν την 3Δ περιστροφή ενός συγκεκριμένου συνόλου 3Δ σημείων. Πρέπει να παρατηρηθεί εδώ ότι γενικά ο  $[\mathbf{n} \ \mathbf{m}]$  είναι ομαλός, εφόσον τα τρία σημεία του  $\Pi$  είναι μη συνευθειακά, οπότε και μπορεί να αντιστραφεί. Ο πίνακας  $\mathbf{K}$  (ισοδύναμα ο  $\tilde{\mathbf{K}}$ ) καθορίζει τη δομή και την κίνηση ενός τριγώνου στον 3Δ χώρο. Αυτό τυπικά δίνεται στην ακόλουθη Πρόταση.

**Πρόταση 7.1:** Έστω ότι,

(α) το  $\xi = \nu \times \mu = [\xi_x \ \xi_y \ \xi_z]^T$  αντιπροσωπεύει το μοναδιαίο κάθετο διάνυσμα σε ένα τρίγωνο και τα  $p = -\xi_x/\xi_z$ ,  $q = -\xi_y/\xi_z$  περιέχουν την πληροφορία προσανατολισμού του επιπέδου,

(β) ο  $\mathbf{R}_{2 \times 3}$  είναι ο  $2 \times 3$  πίνακας που περιέχει τις δύο πρώτες γραμμές του  $3 \times 3$  πίνακα περιστροφής του τριγώνου ως προς τυχαίο άξονα.

Τότε, ο πίνακας  $\mathbf{K}$  μπορεί να εκφραστεί ως

$$\mathbf{K} = \mathbf{R}_{2 \times 3} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ p & q \end{bmatrix}. \quad (7-6)$$

**Απόδειξη:** Εφόσον τα  $\nu$  και  $\mu$  περιστρέφονται στα  $\nu'$  και  $\mu'$  αντίστοιχα,

$$\nu' = \mathbf{R} \nu, \quad \mu' = \mathbf{R} \mu \quad \text{και} \quad \mathbf{n}' = \mathbf{R}_{2 \times 3} \nu, \quad \mathbf{m}' = \mathbf{R}_{2 \times 3} \mu.$$

Τότε

$$\begin{aligned} \mathbf{K} &= [\mathbf{n}' \ \mathbf{m}'] [\mathbf{n} \ \mathbf{m}]^{-1} = \begin{bmatrix} \mathbf{R}_{2 \times 3} \nu & \mathbf{R}_{2 \times 3} \mu \end{bmatrix} [\mathbf{n} \ \mathbf{m}]^{-1} = \mathbf{R}_{2 \times 3} [\nu \ \mu] [\mathbf{n} \ \mathbf{m}]^{-1} \\ &= \mathbf{R}_{2 \times 3} \begin{bmatrix} [\mathbf{n} \ \mathbf{m}] [\mathbf{n} \ \mathbf{m}]^{-1} \\ [\nu_z \ \mu_z] [\mathbf{n} \ \mathbf{m}]^{-1} \end{bmatrix}. \end{aligned}$$

Εφόσον  $[\nu_z \ \mu_z] [\mathbf{n} \ \mathbf{m}]^{-1} = \frac{1}{\nu_z \mu_y - \nu_y \mu_z} [\nu_z \mu_y - \nu_y \mu_z \ \nu_x \mu_z - \nu_z \mu_x] = [-\xi_x/\xi_z \ -\xi_y/\xi_z] = [p \ q]$ , η εξίσωση (7-6) έχει αποδειχθεί. *Ο.Ε.Δ.*

### 7.2.3 Υπολογισμός των πινάκων περιστροφής επιπέδου βάσει των διανυσμάτων κίνησης

Από τον ορισμό του ο  $\mathbf{K}$ , ισοδύναμα ο  $\tilde{\mathbf{K}}$ , είναι έκφραση του 3Δ σχήματος και κίνησης των προβαλλόμενων αντικειμένων. Η επόμενη Πρόταση δείχνει ότι οι πίνακες αυτοί μπορούν να υπολογιστούν βάσει της διαθέσιμης πληροφορίας στον 2Δ χώρο προβολής.

**Πρόταση 7.2:** Οι πίνακες  $\mathbf{K}$ ,  $\tilde{\mathbf{K}}$  μπορούν να εκφραστούν βάσει των διανυσμάτων θέσης και κίνησης τριών σημείων ως

$$\Delta \mathbf{r}' = \mathbf{K} \Delta \mathbf{r} \quad (7-7)$$

$$\Delta \mathbf{v} = \tilde{\mathbf{K}} \Delta \mathbf{r} \quad (7-8)$$

όπου  $\Delta \mathbf{r} = [\mathbf{r}_3 - \mathbf{r}_1 \ \mathbf{r}_2 - \mathbf{r}_1]$ ,  $\Delta \mathbf{r}' = [\mathbf{r}'_3 - \mathbf{r}'_1 \ \mathbf{r}'_2 - \mathbf{r}'_1]$  και  $\Delta \mathbf{v} = [\mathbf{v}_3 - \mathbf{v}_1 \ \mathbf{v}_2 - \mathbf{v}_1]$ , με τα  $\mathbf{v}_i$ ,  $\mathbf{r}_i$  να είναι τα διανύσματα κίνησης και θέσης αντίστοιχα τριών σημείων στο πρώτο καρέ και τα  $\mathbf{r}'_i$  τα διανύσματα θέσης των αντίστοιχων σημείων στο δεύτερο καρέ. Μπορεί να παρατηρήσει κανείς ότι οι πίνακες  $\mathbf{K}$ ,  $\tilde{\mathbf{K}}$  είναι οι ίδιοι για οποιαδήποτε τρία σημεία στο ίδιο τρίγωνο (δηλαδή, στο ίδιο επίπεδο). Επιπρόσθετα, τα  $\mathbf{r}_i$  μπορούν να θεωρηθούν ότι ορίζουν το  $P$  ενώ τα  $\mathbf{r}'_i$  ότι ορίζουν το  $P'$ .

**Απόδειξη:** Οι εξισώσεις (7-7) και (7-8) μπορούν να αποδειχθούν κατευθείαν από τις (7-2) και (7-3) αντίστοιχα.

Τα τρία σημεία  $\mathbf{r}_i$  μπορούν να επιλεγούν στο ορθογώνιο πλέγμα του επιπέδου προβολής, ώστε  $\Delta \mathbf{r} = \mathbf{I}$ . Στην ειδική αυτή περίπτωση, η αντιστροφή του  $\Delta \mathbf{r}$  είναι δεδομένη και  $\tilde{\mathbf{K}} = \Delta \mathbf{v}$ ,  $\mathbf{K} = \Delta \mathbf{r}'$ .

Η Πρόταση 7.2 προσφέρει ένα τρόπο υπολογισμού των  $\mathbf{K}$ ,  $\tilde{\mathbf{K}}$  βάσει της πληροφορίας που γίνεται διαθέσιμη μέσω της διαδικασίας υπολογισμού κίνησης στο χώρο των 2Δ εικόνων. Εν συνεχεία, η Πρόταση 7.2 θα χρησιμοποιείται για τον υπολογισμό του  $\mathbf{K}$  στην πράξη.

Μία σημαντική ιδιότητα σχετική με τις επίπεδες επιφάνειες, διαφαίνεται από την εξίσωση (7-8): εάν  $\mathbf{r}_3 - \mathbf{r}_1 = \lambda(\mathbf{r}_2 - \mathbf{r}_1)$ , τότε  $\mathbf{v}_3 - \mathbf{v}_1 = \lambda(\mathbf{v}_2 - \mathbf{v}_1)$ . Ορίζοντας το διαφορικό διάνυσμα κίνησης  $\mathbf{v}_i - \mathbf{v}_j$  που αντιστοιχεί στα σημεία  $i, j$  της εικόνας, το παραπάνω μπορεί να ερμηνευθεί ως εξής: όλα τα διαφορικά διανύσματα κίνησης που αντιστοιχούν σε συνευθειακά σημεία της εικόνας είναι παράλληλα μεταξύ τους και το μέγεθός τους εξαρτάται μόνο από τις διαφορικές αποστάσεις των διανυσμάτων θέσης (της μορφής  $\mathbf{r}_i - \mathbf{r}_j$ ), αρκεί τα εν λόγω σημεία να αποτελούν προβολές σημείων που ανήκουν στο ίδιο 3Δ επίπεδο. Αυτή η ιδιότητα, η οποία δείχνεται και πειραματικά στην Ενότητα 7.6, μπορεί να χρησιμοποιηθεί για την κατάτμηση των εικόνων σε περιοχές που να αντιστοιχούν σε 3Δ ομοεπίπεδες περιοχές του στερεού αντικειμένου.

### 7.3 Υπολογισμός των παραμέτρων δομής και κίνησης

Τέσσερις πίνακες  $\mathbf{K}$  περιέχουν ικανή πληροφορία για τον υπολογισμό των παραμέτρων δομής και κίνησης ενός στερεού αντικειμένου. Συγκεκριμένα, εάν πίνακες  $\mathbf{K}_{aj}$ ,  $a = R, S$ ,  $j = 1, 2$  είναι διαθέσιμοι (όπου  $\mathbf{R}$ ,  $\mathbf{S}$  συμβολίζουν τους πίνακες περιστροφής για κίνηση από τη σκηνή 1 στη 2 και από τη σκηνή 1 στην 3 αντίστοιχα και όπου  $j = 1, 2$  συμβολίζουν δύο τρίγωνα) ακριβής ανάκτηση των  $\mathbf{R} = [r_{mn}]$  και  $\mathbf{S} = [s_{mn}]$  είναι δυνατή. Επιπρόσθετα, οι σχετικές 3Δ συντεταγμένες των προβαλλόμενων σημείων του στερεού μπορούν να υπολογιστούν. Στη συνέχεια, ο adjoint πίνακας  $\mathbf{L} = \text{adj}(\mathbf{K}) \equiv \text{adj}(\tilde{\mathbf{K}} + \mathbf{I})$  θα χρησιμοποιηθεί για

απλούστερες εξισώσεις. Σ' αυτήν την ενότητα, παραθέτουμε ένα σύνολο Προτάσεων οι οποίες επιτρέπουν τον υπολογισμό των πινάκων περιστροφής  $\mathbf{R}$  και  $\mathbf{S}$  βάσει των τεσσάρων πινάκων  $\mathbf{K}_{aj}$ . Στην πραγματικότητα, εφόσον οι  $\mathbf{R}$  και  $\mathbf{S}$  είναι ορθοκανονικοί, αρκεί να υπολογίσουμε μόνο ένα υποσύνολο των στοιχείων τους. Στην ανάλυση που ακολουθεί, τα στοιχεία  $r_{13}$ ,  $r_{23}$ ,  $r_{31}$ ,  $r_{32}$  και  $r_{33}$  και τα αντίστοιχά τους  $s_{13}$ ,  $s_{23}$ ,  $s_{31}$ ,  $s_{32}$  και  $s_{33}$  χρησιμοποιούνται.

Έστω λοιπόν  $\mathbf{L}_{aj} = \text{adj}(\mathbf{K}_{aj})$  ο adjoint πίνακας για κάθε έναν από τους πίνακες  $\mathbf{K}$ . Έστω επίσης ο πίνακας  $\mathbf{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ . Έστω ότι τα  $p_j$ ,  $q_j$ , όπου  $j = 1, 2$ , αντιπροσωπεύουν τις παραμέτρους δομής για κάθε τρίγωνο, σύμφωνα με την Πρόταση 7.1. Έπειτα από πράξεις, και χρησιμοποιώντας την ταυτότητα  $\mathbf{A}^T \equiv \mathbf{J} \text{adj}(\mathbf{A}) \mathbf{J}^T$  για  $2 \times 2$  πίνακες, η εξίσωση (7-6) δίνει,

$$\mathbf{L}_{Rj} = \mathbf{J} \begin{bmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{bmatrix} \mathbf{J}^T + \mathbf{J} \begin{bmatrix} p_j \\ q_j \end{bmatrix} \begin{bmatrix} r_{13} & r_{23} \end{bmatrix} \mathbf{J}^T \quad (7-9)$$

Βάσει της (7-9), οι Προτάσεις 7.3, 7.4 και 7.5 επιτρέπουν τον υπολογισμό των πινάκων περιστροφής  $\mathbf{R}$ ,  $\mathbf{S}$  με χρήση των  $\mathbf{L}_{Rj}$  και  $\mathbf{L}_{Sj}$ ,  $j = 1, 2$ . Ως πρώτο βήμα, η Πρόταση 7.3 επιτρέπει τον υπολογισμό των άγνωστων διανυσμάτων  $\mathbf{r}_1 \triangleq [r_{13} \ r_{23}]^T$  και  $\mathbf{r}_2 \triangleq [r_{31} \ r_{32}]^T$  με ένα πολλαπλασιαστικό παράγοντα αβεβαιότητας  $\rho$ , χρησιμοποιώντας μόνο την πληροφορία των διανυσμάτων κίνησης (ισοδύναμα των πινάκων  $\mathbf{L}_{Rj}$ ) για την περιστροφή  $\mathbf{R}$ . Όμοια, τα διανύσματα  $\mathbf{s}_1 \triangleq [s_{13} \ s_{23}]^T$  και  $\mathbf{s}_2 \triangleq [s_{31} \ s_{32}]^T$  υπολογίζονται χωριστά με αβεβαιότητα ενός πολλαπλασιαστικού παράγοντα  $\sigma$  για την περιστροφή  $\mathbf{S}$ . Ως δεύτερο βήμα, η Πρόταση 7.4 δίνει μία απλή έκφραση του λόγου  $w = \frac{\sigma}{\rho}$  των αβεβαιοτήτων  $\rho$  και  $\sigma$ . Αυτό αποτελεί ένα απαραίτητο ενδιάμεσο αποτέλεσμα. Ο λόγος  $w$  χρησιμοποιείται στη συνέχεια στην Πρόταση 7.5, ώστε να υπολογιστούν τα στοιχεία  $r_{33}$  και  $s_{33}$  λύνοντας ένα από  $2 \times 2$  γραμμικό σύστημα. Τότε,  $\rho^2 = 1 - (r_{33})^2$  και  $\sigma^2 = 1 - (s_{33})^2$ , οπότε και όλα τα στοιχεία των πινάκων περιστροφής υπολογίζονται.

**Πρόταση 7.3:** Έστω

$$\mathbf{r}_1 \triangleq [r_{13} \ r_{23}]^T, \quad \mathbf{r}_2 \triangleq [r_{31} \ r_{32}]^T, \quad (7-10)$$

$$\mathbf{s}_1 \triangleq [s_{13} \ s_{23}]^T, \quad \mathbf{s}_2 \triangleq [s_{31} \ s_{32}]^T,$$

$$\Delta \mathbf{L}_R \triangleq \mathbf{L}_{R1} - \mathbf{L}_{R2}, \quad \Delta \mathbf{L}_S \triangleq \mathbf{L}_{S1} - \mathbf{L}_{S2}. \quad (7-11)$$

Τότε:

(α) Ισχύει,

$$\|\mathbf{r}_1\|_2^2 = \|\mathbf{r}_2\|_2^2 = \rho^2 = 1 - (r_{33})^2 \quad \text{και} \quad \|\mathbf{s}_1\|_2^2 = \|\mathbf{s}_2\|_2^2 = \sigma^2 = 1 - (s_{33})^2 \quad (7-12)$$



και επίσης

$$\mathbf{L}_{Rj} \mathbf{r}_1 = -\mathbf{r}_2, \quad \mathbf{L}_{Sj} \mathbf{s}_1 = -\mathbf{s}_2, \quad j = 1, 2. \quad (7-13)$$

(β) Εάν  $\mathbf{c}_1, \mathbf{d}_1$  συμβολίζουν τα μοναδιαία ιδιοδιανύσματα των τάξης-1 θετικά ημιορισμένων πινάκων  $(\Delta \mathbf{L}_R)^T (\Delta \mathbf{L}_R)$ ,  $(\Delta \mathbf{L}_S)^T (\Delta \mathbf{L}_S)$  αντίστοιχα, που αντιστοιχούν στις μηδενικές τους ιδιοτιμές, τότε τα  $\mathbf{r}_1, \mathbf{s}_1$  μπορούν να υπολογιστούν με μία πολλαπλασιαστική αβεβαιότητα από τις

$$\mathbf{r}_1 = \rho \mathbf{c}_1, \quad \mathbf{s}_1 = \sigma \mathbf{d}_1. \quad (7-14)$$

(γ) Επίσης τα  $\mathbf{r}_2, \mathbf{s}_2$  μπορούν να υπολογιστούν με την ίδια πολλαπλασιαστική αβεβαιότητα ως

$$\mathbf{r}_2 = \rho \mathbf{c}_2, \quad \mathbf{s}_2 = \sigma \mathbf{d}_2, \quad (7-15)$$

όπου  $\mathbf{c}_2 \triangleq -\mathbf{L}_{Rj} \mathbf{c}_1$ ,  $\mathbf{d}_2 \triangleq -\mathbf{L}_{Sj} \mathbf{d}_1$  για  $j = 1, 2$  όπως μπορεί να παρατηρηθεί από τις εξισώσεις (7-13), (7-14) και (7-15).

**Απόδειξη:**

(α) Από τις ιδιότητες ορθοκανονικότητας του πίνακα περιστροφής:

$$\|\mathbf{r}_1\|_2^2 = r_{13}^2 + r_{23}^2 = 1 - r_{33}^2$$

Ομοίως αποδεικνύονται και οι υπόλοιποι τύποι. Επίσης από την εξίσωση (7-9), έχουμε

$$\mathbf{L}_{Rj} \mathbf{r}_1 = \begin{bmatrix} r_{22} & -r_{12} \\ -r_{21} & r_{11} \end{bmatrix} \begin{bmatrix} r_{13} \\ r_{23} \end{bmatrix} + \begin{bmatrix} -q_j \\ p_j \end{bmatrix} \begin{bmatrix} -r_{23} & r_{13} \end{bmatrix} \begin{bmatrix} r_{13} \\ r_{23} \end{bmatrix} = -\mathbf{r}_2.$$

$$(β) \text{ Είναι } \Delta \mathbf{L}_R \mathbf{r}_1 = \mathbf{L}_{R1} \mathbf{r}_1 - \mathbf{L}_{R2} \mathbf{r}_1 = \mathbf{0} \quad \text{και} \quad \Delta \mathbf{L}_R^T \Delta \mathbf{L}_R \mathbf{r}_1 = \mathbf{0}.$$

Φαίνεται τότε ότι το  $\mathbf{r}_1$  είναι το ιδιοδιάνυσμα του  $(\Delta \mathbf{L}_R^T)(\Delta \mathbf{L}_R)$  που αντιστοιχεί στη μηδενική του ιδιοτιμή. Εάν το  $\mathbf{c}_1$  είναι το αντίστοιχο μοναδιαίο ιδιοδιάνυσμα, τότε  $\mathbf{r}_1 = \rho \mathbf{c}_1$ . Το  $\mathbf{r}_1$  είναι επίσης ιδιοδιάνυσμα του  $\Delta \mathbf{L}_R$  και αντιστοιχεί επίσης στην μηδενική ιδιοτιμή του. Όμοια απόδειξη ισχύει και για το  $\mathbf{s}_1$ .

(γ) Έχουμε  $\mathbf{r}_2 = -\mathbf{L}_{Rj} \mathbf{r}_1 = -\rho \mathbf{L}_{Rj} \mathbf{c}_1$ . Τότε, για  $\mathbf{c}_2 = -\mathbf{L}_{Rj} \mathbf{c}_1$ , είναι  $\mathbf{r}_2 = \rho \mathbf{c}_2$ . Όμοια και για το  $\mathbf{s}_2$ . *Ο.Ε.Δ.*

Μετά από πράξεις, χρησιμοποιώντας τις ιδιότητες του πίνακα περιστροφής και τους ορισμούς της Πρότασης 7.3

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = -r_{33} \mathbf{c}_1 \mathbf{c}_2^T - (\mathbf{J} \mathbf{c}_1) (\mathbf{J} \mathbf{c}_2)^T, \quad (7-16)$$

οπότε η εξίσωση (7-9) μπορεί να γραφεί ως:

$$\mathbf{L}_{Rj} = -(\mathbf{J}\mathbf{c}_2)(\mathbf{J}\mathbf{c}_1)^T r_{33} - \mathbf{c}_2 \mathbf{c}_1^T + \rho \mathbf{J} \begin{bmatrix} p_j \\ q_j \end{bmatrix} (\mathbf{J}\mathbf{c}_1)^T. \quad (7-17)$$

Όμοιες εξισώσεις ισχύουν και για τα στοιχεία του πίνακα  $\mathbf{S}$ . Χρησιμοποιώντας την Πρόταση 7.3 και τις παραπάνω εξισώσεις, οι επόμενες Προτάσεις 7.4 και 7.5 επιτρέπουν τον υπολογισμό των  $r_{33}$ ,  $s_{33}$  και των αβεβαιοτήτων  $\rho$  και  $\sigma$ .

**Πρόταση 7.4:** Ο λόγος  $w^2 = (\frac{\sigma}{\rho})^2$  ισούται με  $\frac{\lambda_S}{\lambda_R}$  όπου  $\lambda_R$ ,  $\lambda_S$  είναι οι μη μηδενικές ιδιοτιμές των πινάκων  $(\Delta\mathbf{L}_R)^T(\Delta\mathbf{L}_R)$ ,  $(\Delta\mathbf{L}_S)^T(\Delta\mathbf{L}_S)$  αντίστοιχα. Έτσι το  $w = \sigma/\rho$  μπορεί να υπολογιστεί με μία αβεβαιότητα προσήμου. Μπορεί ναδειχθεί ότι η μία λύση απορρίπτεται αμέσως.

**Απόδειξη:**

Έστω  $\Delta\mathbf{L}_R = \mathbf{L}_{R1} - \mathbf{L}_{R2}$ ,  $\Delta p = p_1 - p_2$ ,  $\Delta q = q_1 - q_2$ . Χρησιμοποιώντας την (7-17) για  $j = 1, 2$  και πολλαπλασιάζοντας στη συνέχεια με  $\mathbf{J}\mathbf{c}_1$ , έχουμε:

$$\Delta\mathbf{L}_R \mathbf{J}\mathbf{c}_1 = \rho \begin{bmatrix} -\Delta q \\ \Delta p \end{bmatrix}, \quad \Delta\mathbf{L}_R = \rho \begin{bmatrix} -\Delta q \\ \Delta p \end{bmatrix} (\mathbf{J}\mathbf{c}_1)^T.$$

Οπότε

$$(\Delta\mathbf{L}_R)^T (\Delta\mathbf{L}_R) \mathbf{J}\mathbf{c}_1 = \rho^2 ((\Delta q)^2 + (\Delta p)^2) \mathbf{J}\mathbf{c}_1. \quad (7-18)$$

Φαίνεται τότε ότι το  $\mathbf{J}\mathbf{c}_1$  είναι το μοναδιαίο ιδιοδιάνυσμα του  $(\Delta\mathbf{L}_R)^T(\Delta\mathbf{L}_R)$  που αντιστοιχεί στη μη μηδενική ιδιοτιμή του

$$\lambda_R = \rho^2 ((\Delta q)^2 + (\Delta p)^2). \quad (7-19)$$

Ας θυμηθούμε ότι το  $\mathbf{c}_1$  είναι το μοναδιαίο ιδιοδιάνυσμα του  $(\Delta\mathbf{L}_R)^T(\Delta\mathbf{L}_R)$  που αντιστοιχεί στη μηδενική ιδιοτιμή του.

Όμοια,  $\mathbf{J}\mathbf{d}_1$  είναι το μοναδιαίο ιδιοδιάνυσμα του  $(\Delta\mathbf{L}_S)^T(\Delta\mathbf{L}_S)$  που αντιστοιχεί στη μηδενική ιδιοτιμή του  $\lambda_S = \sigma^2 ((\Delta q)^2 + (\Delta p)^2)$ .

Έτσι,  $w^2 = \frac{\sigma^2}{\rho^2} = \frac{\lambda_S}{\lambda_R}$ . Ο.Ε.Δ.

**Πρόταση 7.5:** Τα άγνωστα στοιχεία  $r_{33}$ ,  $s_{33}$  των πινάκων περιστροφής μπορούν να υπολογιστούν με μοναδικό τρόπο από την

$$\begin{bmatrix} w\mathbf{J}\mathbf{c}_2 & -\mathbf{J}\mathbf{d}_2 \end{bmatrix} \begin{bmatrix} r_{33} \\ s_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{Sj} & \mathbf{L}_{Rj} \end{bmatrix} \begin{bmatrix} \mathbf{J}\mathbf{d}_1 \\ -w\mathbf{J}\mathbf{c}_1 \end{bmatrix}. \quad (7-20)$$

**Απόδειξη:** Πολλαπλασιάζοντας την (7-17) με  $\mathbf{Jc}_1$ ,

$$\mathbf{L}_{Rj} \mathbf{Jc}_1 = -(\mathbf{Jc}_2) r_{33} + \rho \mathbf{J} \begin{bmatrix} p_j \\ q_j \end{bmatrix}.$$

Συνδυάζοντας την εξίσωση αυτή με μία αντίστοιχη για τον  $\mathbf{L}_{Sj}$

$$(\mathbf{L}_{Rj} \mathbf{Jc}_1 + \mathbf{Jc}_2 r_{33}) \sigma = (\mathbf{L}_{Sj} \mathbf{Jd}_1 + \mathbf{Jd}_2 s_{33}) \rho.$$

Γράφοντας την παραπάνω εξίσωση σε μορφή πινάκων λαμβάνουμε την εξίσωση (7-20).  
Ο.Ε.Δ.

Έχοντας διαθέσιμα τα  $r_{33}$ ,  $s_{33}$ , οι σταθερές  $\rho$  και  $\sigma$  υπολογίζονται από τις εξισώσεις (7-12). Στο σημείο αυτό, οι πίνακες περιστροφής  $\mathbf{R}$  και  $\mathbf{S}$  υπολογίζονται. Μπορεί να τονιστεί εδώ ότι τα  $\mathbf{r}_1$ ,  $\mathbf{r}_2$  δέχονται δύο πραγματικές λύσεις, που αντιστοιχούν σε κατοπτρισμό του στερεού αντικειμένου ως προς το επίπεδο προβολής [167]. Το τελευταίο μπορεί ναδειχθεί πολύ απλά, παρατηρώντας ότι δεν μπορούμε να διαχωρίσουμε μεταξύ των δύο παρακάτω εξισώσεων

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{T} \quad (7-21)$$

$$\begin{bmatrix} x' \\ y' \\ -z' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & -r_{13} \\ r_{21} & r_{22} & -r_{23} \\ -r_{31} & -r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ -z \end{bmatrix} + \mathbf{T} \quad (7-22)$$

όταν γνωρίζουμε μόνο τα διανύσματα κίνησης (δηλαδή τα  $x, x', y, y'$ ).

Η εξίσωση (7-22) δίνει ουσιαστικά την κίνηση του ειδώλου του αντικειμένου όταν αυτό καθρεπτίζεται επάνω στο επίπεδο προβολής. Διατυπώνεται συχνά ότι αυτή η λύση μπορεί εύκολα να απορριφθεί αφού δίνει αρνητικές  $z$ -συντεταγμένες, το οποίο όμως δεν είναι αλήθεια. Όπως είπαμε και νωρίτερα, μόνο το σχετικό βάθος (δηλαδή, η δομή) του αντικειμένου μπορεί να υπολογιστεί από διαδικασίες ανακατασκευής σε παράλληλη προβολή. Σύμφωνα με αυτόν τον περιορισμό, υπολογίζουμε το βάθος  $z_p$  του σημείου  $(x_p, y_p, z_p)$  ως  $\alpha + z_p$  με μία αβεβαιότητα  $\alpha$ . Επομένως, το πρόσημο των απόλυτων συντεταγμένων βάθους εξαρτάται από την αυθαίρετη επιλογή του  $\alpha$  και δεν είναι κατ'ανάγκη θετικό ή αρνητικό σε κάθε περίπτωση. Στην πραγματικότητα, ο χρήστης μπορεί να διαχωρίσει τις δύο λύσεις (7-21) και (7-22), βασιζόμενος σε ευρεστικούς περιορισμούς επί του προσανατολισμού γνωστών αντικειμένων.

Αφού υπολογιστούν τα  $\mathbf{R}$  και  $\mathbf{S}$ , ο υπολογισμός των διανυσμάτων μετατόπισης (translation) και των παραμέτρων δομής είναι άμεσος. Για να απαλλαγούμε από τις αβεβαιότητες

**Πίνακας 7.1** Διαδικασία υπολογισμού απουσία θορύβου.**Για κάθε μετάβαση:**

- (i) παράγουμε τουλάχιστον δύο  $\mathbf{L}_{Rj}$  (ισοδύναμα  $\mathbf{K}_{Rj}$ ) πίνακες βάσει τουλάχιστον δύο τριπλετών σημείων χρησιμοποιώντας τις εξισώσεις (7-7) ή (7-8),
- (ii) παράγουμε τουλάχιστον ένα  $\Delta\mathbf{L}_{Rk}$  πίνακα από τους  $\mathbf{L}_{Rj}$  χρησιμοποιώντας την εξίσωση (7-11),
- (iii) υπολογίζουμε το μοναδιαίο ιδιοδιάνυσμα  $\mathbf{c}_1$  που αντιστοιχεί στη μηδενική ιδιοτιμή του  $\Delta\mathbf{L}_{Rk}^T \Delta\mathbf{L}_{Rk}$ , καθώς και τη μη μηδενική του ιδιοτιμή  $\lambda_R$  από την Πρόταση 7.3.

**Για κάθε ζεύγος μεταβάσεων:**

- (i) υπολογίζουμε το  $w^2$  ως το λόγο των μη μηδενικών ιδιοτιμών από την Πρόταση 7.4,
- (ii) υπολογίζουμε το  $r_{33}$  και για τους δύο πίνακες περιστροφής από την (7-20) σύμφωνα με την Πρόταση 7.5.

**Για κάθε μετάβαση:**

- (i) υπολογίζουμε όλα τα στοιχεία του πίνακα περιστροφής από τις (7-12), (7-14), (7-15) και (7-16),
- (ii) υπολογίζουμε τα διανύσματα μετατόπισης υπό τις υποθέσεις για το γεωμετρικό κέντρο από την εξίσωση (7-24),
- (iii) υπολογίζουμε το σχετικό βάθος από την εξίσωση (7-1).

του μοντέλου παράλληλης προβολής στο απόλυτο βάθος σημείων ( $z$ -συντεταγμένη) και στη συνιστώσα μετατόπισης  $t_3$  κατά τον άξονα- $z$ , είναι αναγκαίο να τεθούν, χωρίς βλάβη της γενικότητας, οι παρακάτω περιορισμοί (ή άλλοι ισοδύναμοι). Θεωρούμε ότι το γεωμετρικό κέντρο του σώματος στην πρώτη σκηνή  $(x_m, y_m, z_m) = (\frac{1}{P} \sum_{p=1}^P x_p, \frac{1}{P} \sum_{p=1}^P y_p, \frac{1}{P} \sum_{p=1}^P z_p)$  βρίσκεται πάνω στο επίπεδο προβολής  $z$ , δηλαδή  $z_m = 0$ . Τότε από την εξίσωση (7-1), λαμβάνουμε για τις δύο πρώτες συνιστώσες της μετατόπισης

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} x'_m \\ y'_m \end{bmatrix} - \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} x_m \\ y_m \end{bmatrix}. \quad (7-23)$$

Με όμοια συλλογιστική, υπολογίζεται και η τρίτη συνιστώσα της μετατόπισης κρατώντας το γεωμετρικό κέντρο στο επίπεδο προβολής και στην επόμενη σκηνή, δηλαδή  $z'_m = 0$ . Τότε

$$\mathbf{T} \equiv \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} x'_m \\ y'_m \\ 0 \end{bmatrix} - \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \\ r_{31} & r_{32} \end{bmatrix} \begin{bmatrix} x_m \\ y_m \end{bmatrix}. \quad (7-24)$$

Όπως γίνεται φανερό από τα παραπάνω, η μετατόπιση  $\mathbf{T}$  υπολογίζεται τότε για κάθε μετάβαση από την εξίσωση (7-24), κρατώντας διαρκώς τη συνιστώσα βάθους του γεωμετρικού κέντρου του σώματος στο επίπεδο προβολής.

Τέλος, έχοντας υπολογίσει τις παραμέτρους 3Δ κίνησης του σώματος, η πληροφορία σχετικού βάθους για όλα τα διαθέσιμα σημεία δίνεται από την εξίσωση (7-1).

Ολόκληρη η διαδικασία υπολογισμού περιγράφεται σε βήματα στον Πίνακα 7.1.

## 7.4 Υπολογισμός παρουσία θορύβου

Στην πράξη οι τεχνικές υπολογισμού κίνησης μπορούν να εισάγουν σφάλματα (θόρυβο) στους πίνακες  $\mathbf{K}_{Rj}$ ,  $\mathbf{K}_{Sj}$ . Τα σφάλματα αυτά οφείλονται είτε σε περιορισμούς της διαδικασίας υπολογισμού είτε στην ανάλυση του πλέγματος πάνω στο επίπεδο προβολής (ακρίβεια pixel). Στην ενότητα αυτή, εισάγουμε μία τροποποιημένη εκδοχή της διαδικασίας που περιγράφεται στις Προτάσεις 7.3, 7.4 και 7.5, η οποία χειρίζεται θορυβώδη πεδία διανυσμάτων κίνησης.

Η παρακάτω ανάλυση βασίζεται σε δύο παραδοχές:

- ένας μεγάλος αριθμός ( $N$ ) πινάκων  $\hat{\mathbf{L}}_{Rj}$ ,  $\hat{\mathbf{L}}_{Sj}$  ( $j = 1 \dots N$ ) προσβεβλημένων από σφάλμα είναι διαθέσιμος, και
- το σφάλμα μοντελοποιείται σαν ανεξάρτητος ομοιόμορφα κατανεμημένος προσθετικός θόρυβος (i.i.d. additive noise) τέτοιος ώστε  $\hat{\mathbf{L}}_{Rj} = \mathbf{L}_{Rj} + \mathbf{L}_{Rj}^e$  και  $\hat{\mathbf{L}}_{Sj} = \mathbf{L}_{Sj} + \mathbf{L}_{Sj}^e$  όπου οι δεύτεροι όροι στα δεξιά αντιπροσωπεύουν τυχαίους  $2 \times 2$  πίνακες.

Στη συνέχεια, θα χρησιμοποιήσουμε εκθέτες ( $\hat{\phantom{x}}$ ) και ( $^e$ ) για να συμβολίσουμε θορυβώδεις πίνακες και συνιστώσες θορύβου αντίστοιχα. Πριν προχωρήσουμε στην τεκμηρίωση κατάλληλων εκτιμητριών των  $\mathbf{R}$  και  $\mathbf{S}$  παρουσία θορύβου, θα πρέπει να εξετάσουμε πώς τα λάθη στο πεδίο κίνησης επηρεάζουν τους πίνακες  $\mathbf{K}$ ,  $\tilde{\mathbf{K}}$  και  $\mathbf{L}$ . Το επόμενο Λήμμα εμπεριέχει την επήρεια αυτή.

Μια παρατήρηση θα πρέπει να γίνει σ' αυτό το σημείο για να γίνει σαφής η παρακάτω ανάλυση. Στα επόμενα, ο δείκτης  $j = 1 \dots N$  θα αναφέρεται σε πίνακες  $\mathbf{L}$ , ενώ ο δείκτης  $k = 1 \dots M$  σε πίνακες  $\Delta \mathbf{L}$ . Τα  $M$ ,  $N$  αντιπροσωπεύουν τον αριθμό των πινάκων  $\Delta \mathbf{L}_k$  και  $\mathbf{L}_j$  που περιλαμβάνονται στους υπολογισμούς. Έστω επίσης ότι ο αριθμός  $P$  συμβολίζει τον αριθμό των σημείων (και αντίστοιχων διανυσμάτων κίνησης) που χρησιμοποιούνται. Η σχέση μεταξύ των  $M$  και  $N$  διαφοροποιείται σύμφωνα με τη στρατηγική παραγωγής των πινάκων  $\Delta \mathbf{L}_k$  από τους πίνακες  $\mathbf{L}_j$ . Όπως θα φανεί και στη συνέχεια, στη γενική περίπτωση  $M = M(N) \cong N/2$  και  $M(N) \rightarrow \infty$  όπως  $N \rightarrow \infty$ . Η σχέση μεταξύ  $N$  και  $P$  εξαρτάται από τη στρατηγική παραγωγής πινάκων  $\mathbf{L}_j$  από τα διαθέσιμα σημεία και, όπως θα δούμε, γενικά  $N = N(P) \cong P/3$ .

**Λήμμα 7.1:** Έστω ότι  $\mathbf{r}_1^j, \mathbf{r}_2^j, \mathbf{r}_3^j$  είναι μία συλλογή από τριπλέτες διανυσμάτων θέσης ( $j = 1 \dots N$ ) με αντίστοιχα διανύσματα κίνησης  $\mathbf{v}_1^j, \mathbf{v}_2^j, \mathbf{v}_3^j$  ( $j = 1 \dots N$ ). Έστω επίσης ότι  $\hat{\mathbf{v}}_1^j = \mathbf{v}_1^j + \mathbf{e}_1^j$ ,  $\hat{\mathbf{v}}_2^j = \mathbf{v}_2^j + \mathbf{e}_2^j$ ,  $\hat{\mathbf{v}}_3^j = \mathbf{v}_3^j + \mathbf{e}_3^j$  είναι τα μετρώμενα θορυβώδη

διανύσματα κίνησης. Ας υποθέσουμε ότι οι  $2 \times 1$  προσθετικές συνιστώσες θορύβου  $\mathbf{e}_1^j$ ,  $\mathbf{e}_2^j$ ,  $\mathbf{e}_3^j$  είναι μηδενικού μέσου και αμοιβαίως ασυσχέτιστες με τη λογική ότι

$$E\{\mathbf{e}_m^i (\mathbf{e}_n^j)^T\} = \delta(m-n)\delta(i-j) \sigma_e^2 \mathbf{I} , \quad (7-25)$$

όπου  $\sigma_e^2$  είναι η διασπορά του θορύβου σε κάθε συνιστώσα του διανύσματος κίνησης και η διασπορά του θορύβου είναι ίδια και προς τις δύο κατευθύνσεις,  $\sigma_e^2 = \sigma_{ex}^2 = \sigma_{ey}^2$ .

Για  $\hat{\mathbf{K}}_j \triangleq [\hat{\mathbf{v}}_3^j - \hat{\mathbf{v}}_1^j \quad \hat{\mathbf{v}}_2^j - \hat{\mathbf{v}}_1^j] [\mathbf{r}_3^j - \mathbf{r}_1^j \quad \mathbf{r}_2^j - \mathbf{r}_1^j]^{-1}$ ,  $\hat{\mathbf{L}}_j = \text{adj}(\hat{\mathbf{K}}_j + \mathbf{I})$  και  $\Delta \hat{\mathbf{L}}_k = \hat{\mathbf{L}}_j - \hat{\mathbf{L}}_i$ ,  $\Delta \mathbf{L}_k = \mathbf{L}_j - \mathbf{L}_i$  για κάποια  $i, j = 1 \cdots N$  και  $k = 1 \cdots M$ , τέτοια ώστε

(ι) κάθε  $\hat{\mathbf{K}}_j$  παράγεται από διαφορετικά σημεία αναφοράς και διανύσματα κίνησης,

(ιι) κάθε  $\Delta \hat{\mathbf{L}}_k$  παράγεται βάσει διαφορετικών πινάκων  $\hat{\mathbf{L}}_j$ ,

οι επόμενες στατιστικές εκφράσεις του μέσου όρου και της διασποράς ισχύουν:

$$E\{\hat{\mathbf{L}}_j\} = \mathbf{L}_j , \quad E\{\Delta \hat{\mathbf{L}}_k\} = \Delta \mathbf{L}_k , \quad (7-26)$$

$$E\{\hat{\mathbf{L}}_j^T \hat{\mathbf{L}}_i\} = \mathbf{L}_j^T \mathbf{L}_i + \delta(j-i) \pi_j^2 \mathbf{I} , \quad (7-27)$$

$$E\{\Delta \hat{\mathbf{L}}_k^T \Delta \hat{\mathbf{L}}_l\} = \Delta \mathbf{L}_k^T \Delta \mathbf{L}_l + \delta(k-l) \sigma_k^2 \mathbf{I} , \quad (7-28)$$

όπου

$$\pi_j^2 \triangleq \sigma_e^2 \frac{\|\mathbf{r}_3^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_2^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_3^j - \mathbf{r}_2^j\|^2}{|\det[\mathbf{r}_3^j - \mathbf{r}_1^j \quad \mathbf{r}_2^j - \mathbf{r}_1^j]|^2} \quad (7-29)$$

και

$$\sigma_k^2 \triangleq \pi_j^2 + \pi_i^2 \quad (7-30)$$

**Απόδειξη:** Για ευκολία στην αναπαράσταση και την κατανόηση, στην απόδειξη παραλείπουμε τους δείκτες  $R(S)$  από πίνακες και διανύσματα.

(α) Από τον ορισμό των πινάκων  $\mathbf{L}$  και  $\hat{\mathbf{L}}$  λαμβάνουμε  $\hat{\mathbf{L}}_j = \mathbf{L}_j + \mathbf{L}_j^e$  όπου

$$\mathbf{L}_j^e = \mathbf{J} ([\mathbf{e}_3^j - \mathbf{e}_1^j \quad \mathbf{e}_2^j - \mathbf{e}_1^j][\mathbf{r}_3^j - \mathbf{r}_1^j \quad \mathbf{r}_2^j - \mathbf{r}_1^j]^{-1})^T \mathbf{J}^T .$$

Εφόσον οι συνιστώσες προσθετικού θορύβου  $\mathbf{e}_1^j$ ,  $\mathbf{e}_2^j$ ,  $\mathbf{e}_3^j$  είναι μηδενικού μέσου λαμβάνουμε  $E\{\hat{\mathbf{L}}_j\} = \mathbf{L}_j$  για όλα τα  $j$ . Με βάση το παραπάνω αποτέλεσμα εάν  $\Delta \hat{\mathbf{L}}_k = \hat{\mathbf{L}}_j - \hat{\mathbf{L}}_i$  και  $\Delta \mathbf{L}_k = \mathbf{L}_j - \mathbf{L}_i$  καταλήγουμε ότι  $E\{\Delta \hat{\mathbf{L}}_k\} = \Delta \mathbf{L}_k$  το οποίο ολοκληρώνει την απόδειξη του πρώτου μέρους του Λήμματος.

(β) Για δύο θορυβώδεις  $\mathbf{L}$ -πίνακες  $\hat{\mathbf{L}}_j$  και  $\hat{\mathbf{L}}_i$ , έχουμε

$$E\{\hat{\mathbf{L}}_j^T \hat{\mathbf{L}}_i\} = \mathbf{L}_j^T \mathbf{L}_i + \mathbf{A}(j, i) ,$$

όπου  $\mathbf{A}(j, i) = E\{(\mathbf{L}_j^e)^T \mathbf{L}_i^e\}$  αφού όλοι οι άλλοι όροι χάνονται λόγω της ιδιότητας μηδενικού μέσου του  $\mathbf{e}$ . Στην έκφραση του  $\mathbf{A}$  ο δείκτης  $e$  των πινάκων  $\mathbf{L}$  συμβολίζει ότι αυτοί οι πίνακες αντιστοιχούν μόνο σε προσθετικές συνιστώσες των διανυσμάτων κίνησης. Υπό την υπόθεση ότι όλες οι προσθετικές συνιστώσες θορύβου των παρατηρούμενων διανυσμάτων κίνησης είναι αμοιβαίως ασυσχέτιστες και εφόσον κάθε πίνακας  $\hat{\mathbf{L}}_j$  ορίζεται βάσει τριών διαφορετικών διανυσμάτων κίνησης (δηλαδή, κανένα διάνυσμα κίνησης δεν χρησιμοποιείται για την παραγωγή περισσότερων του ενός πίνακα  $\hat{\mathbf{L}}_j$ ) καταλήγουμε ότι  $\mathbf{A}(j, i) = \mathbf{0}$  εάν  $j \neq i$ . Από την άλλη πλευρά, εάν  $j = i$ ,

$$\mathbf{A}(j, j) = E\{(\mathbf{L}_j^e)^T \mathbf{L}_j^e\} = \mathbf{J}^T E\{\tilde{\mathbf{K}}_j^e (\tilde{\mathbf{K}}_j^e)^T\} \mathbf{J}$$

όπου

$$\tilde{\mathbf{K}}_j^e \triangleq [\mathbf{e}_3^j - \mathbf{e}_1^j \quad \mathbf{e}_2^j - \mathbf{e}_1^j][\mathbf{r}_3^j - \mathbf{r}_1^j \quad \mathbf{r}_2^j - \mathbf{r}_1^j]^{-1}.$$

Βάσει αυτών των εκφράσεων και μετά από μερικές αλγεβρικές πράξεις που χρησιμοποιούν την (7-25) καταλήγουμε στην ακόλουθη έκφραση του υπολειπόμενου όρου  $\mathbf{A}(j, j)$

$$\mathbf{A}(j, j) = \sigma_e^2 \frac{\|\mathbf{r}_3^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_2^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_3^j - \mathbf{r}_2^j\|^2}{|\det[\mathbf{r}_3^j - \mathbf{r}_1^j \quad \mathbf{r}_2^j - \mathbf{r}_1^j]|^2} \mathbf{I} \equiv \pi_j^2 \mathbf{I}. \quad (7-31)$$

Συγκεκριμένα, έστω  $\Delta \mathbf{r} = [\delta_1 \delta_2]$ ,  $\Delta \mathbf{e} = [\epsilon_1 \epsilon_2] = \begin{bmatrix} \epsilon_{1x} & \epsilon_{2x} \\ \epsilon_{1y} & \epsilon_{2y} \end{bmatrix}$ , όπου οι δείκτες  $i, j$  παραλείπονται για απλότητα αφού  $i = j$ . Ισχύει,

$$\tilde{\mathbf{K}}^e (\tilde{\mathbf{K}}^e)^T = \Delta \mathbf{e} \Delta \mathbf{r}^{-1} \Delta \mathbf{r}^{-T} \Delta \mathbf{e}^T = \Delta \mathbf{e} (\Delta \mathbf{r}^T \Delta \mathbf{r})^{-1} \Delta \mathbf{e}^T$$

και

$$\Delta \mathbf{r}^T \Delta \mathbf{r} = \begin{bmatrix} \delta_1^T \delta_1 & \delta_1^T \delta_2 \\ \delta_2^T \delta_1 & \delta_2^T \delta_2 \end{bmatrix}, \quad \det(\Delta \mathbf{r}^T \Delta \mathbf{r}) = (\det(\Delta \mathbf{r}))^2 = (\det[\delta_1 \delta_2])^2,$$

$$(\Delta \mathbf{r}^T \Delta \mathbf{r})^{-1} = \frac{1}{(\det[\delta_1 \delta_2])^2} \begin{bmatrix} \delta_2^T \delta_2 & -\delta_1^T \delta_2 \\ -\delta_2^T \delta_1 & \delta_1^T \delta_1 \end{bmatrix}.$$

Ας θυμηθούμε ότι για τις συνιστώσες συνιστώσες θορύβου στα διανύσματα κίνησης ισχύει  $E\{e_{pm} e_{qn}\} = \delta(p - q) \delta(m - n) \sigma_e^2$ . Χρησιμοποιώντας την εξίσωση αυτή και μετά τον υπολογισμό των προσδοκητών τιμών για τα διαφορικά διανύσματα κίνησης:

$$\begin{aligned} E\{\tilde{\mathbf{K}}^e (\tilde{\mathbf{K}}^e)^T\} &= \frac{1}{(\det[\delta_1 \delta_2])^2} \begin{bmatrix} \delta_2^T \delta_2 + \delta_1^T \delta_1 - \delta_1^T \delta_2 & 0 \\ 0 & \delta_2^T \delta_2 + \delta_1^T \delta_1 - \delta_1^T \delta_2 \end{bmatrix} 2\sigma_e^2 \\ &= \frac{\delta_2^T \delta_2 + \delta_1^T \delta_1 - \delta_1^T \delta_2}{(\det[\delta_1 \delta_2])^2} 2\sigma_e^2 \mathbf{I} = \frac{\|\delta_2\|^2 + \|\delta_1\|^2 + \|\delta_1 - \delta_2\|^2}{(\det[\delta_1 \delta_2])^2} \sigma_e^2 \mathbf{I} \end{aligned}$$

το οποίο καταλήγει στην εξίσωση (7-31) και ολοκληρώνει το δεύτερο μέρος του Λήμματος 7.1 (εξίσωση (7-27)).

(γ) Χρησιμοποιώντας δύο θορυβώδεις  $\Delta\hat{\mathbf{L}}$ -πίνακες  $\Delta\hat{\mathbf{L}}_k$  και  $\Delta\hat{\mathbf{L}}_l$ , έχουμε ότι,

$$E\{\Delta\hat{\mathbf{L}}_k^T \Delta\hat{\mathbf{L}}_l\} = \Delta\mathbf{L}_k^T \Delta\mathbf{L}_l + \mathbf{B}(k, l),$$

όπου  $\mathbf{B}(k, l) = E\{(\Delta\mathbf{L}_k^e)^T \Delta\mathbf{L}_l^e\}$  εφόσον όλοι οι υπόλοιποι όροι χάνονται λόγω της ιδιότητας του μηδενικού μέσου όρου του  $\mathbf{e}$ . Όμοια με το (β), καταλήγουμε ότι  $\mathbf{B}(k, l) = \mathbf{0}$  εάν  $k \neq l$ . Αντίθετα, όταν  $k = l$ ,

$$\mathbf{B}(k, k) = E\{[\mathbf{L}_j^e - \mathbf{L}_i^e]^T [\mathbf{L}_j^e - \mathbf{L}_i^e]\} = \mathbf{J}^T E\{\tilde{\mathbf{K}}_j^e (\tilde{\mathbf{K}}_j^e)^T\} \mathbf{J} + \mathbf{J}^T E\{\tilde{\mathbf{K}}_i^e (\tilde{\mathbf{K}}_i^e)^T\} \mathbf{J}.$$

Χρησιμοποιώντας το μέρος (β) της απόδειξης, καταλήγουμε ότι:

$$\mathbf{B}(k, k) = \mathbf{J}^T \pi_j^2 \mathbf{J} + \mathbf{J}^T \pi_i^2 \mathbf{J} = (\pi_j^2 + \pi_i^2) \mathbf{I}.$$

Αυτό ολοκληρώνει την απόδειξη του τρίτου μέρους του Λήμματος 7.1 (εξίσωση (7-28)).  
Ο.Ε.Δ.

Είναι χρήσιμα μερικά σχόλια σε αυτό το σημείο:

[C1] Το Λήμμα 7.1 υποδεικνύει ότι ο πίνακας συσχέτισης των  $\Delta\hat{\mathbf{L}}_k, \Delta\hat{\mathbf{L}}_l$  ( $\hat{\mathbf{L}}_j, \hat{\mathbf{L}}_i$ ) είναι ίσος με το γινόμενο  $\Delta\mathbf{L}_k^T \Delta\mathbf{L}_l$  ( $\hat{\mathbf{L}}_j^T \hat{\mathbf{L}}_i$ ) συν το γινόμενο του μοναδιαίου πίνακα  $\mathbf{I}$  με μία βαθμωτή ποσότητα αρκεί κάποια, καθοριζόμενη από το χρήστη, στρατηγική να ακολουθείται κατά την επιλογή των σημείων που χρησιμοποιούνται στην παραγωγή των πινάκων  $\tilde{\mathbf{K}}$  και στην επιλογή των πινάκων  $\mathbf{L} = \mathbf{J}\tilde{\mathbf{K}}^T \mathbf{J}^T + \mathbf{I}$  στην παραγωγή των  $\Delta\mathbf{L}$ . Συγκεκριμένα, (α) κάθε σημείο πρέπει να χρησιμοποιείται στην παραγωγή μόνο ενός  $\tilde{\mathbf{K}}$ , (β) κάθε  $\tilde{\mathbf{K}}$  ή ισοδύναμα  $\mathbf{L}$  πρέπει να χρησιμοποιείται στην παραγωγή μόνο ενός  $\Delta\mathbf{L}$ .

[C2] Υπό τις προϋποθέσεις του [C1] η επίδραση του θορύβου στα διανύσματα κίνησης συγκεντρώνεται στη βαθμωτή ποσότητα 'διασποράς', παράμετρο  $\sigma_k^2$  (ισοδύναμα  $\pi_j^2$ ). Με τη σειρά της, η ποσότητα αυτή εξαρτάται από τη διασπορά του θορύβου  $\sigma_e^2$  και τη γεωμετρία των εμπλεκόμενων τριπλετών διανυσμάτων θέσης. Αν και το  $\sigma_e^2$  δεν μπορεί να ελεγχθεί από το χρήστη, αυτός (αυτή) μπορούν να ελαττώσουν την ποσότητα  $\sigma_k^2$  (ισοδύναμα  $\pi_j^2$ ) διαλέγοντας τις τριπλέτες διανυσμάτων θέσης  $\{\mathbf{r}_1^j, \mathbf{r}_2^j, \mathbf{r}_3^j\}$  με μικρό παράγοντα  $\frac{\|\mathbf{r}_3^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_2^j - \mathbf{r}_1^j\|^2 + \|\mathbf{r}_3^j - \mathbf{r}_2^j\|^2}{|\det[\mathbf{r}_3^j - \mathbf{r}_1^j \quad \mathbf{r}_2^j - \mathbf{r}_1^j]|^2}$ . Είναι χρήσιμο να παρατηρήσουμε ότι ο αριθμητής της έκφρασης αυτής είναι απλώς το άθροισμα των τετραγώνων των πλευρών του τριγώνου που σχηματίζουν τα  $2\Delta$  διανύσματα  $\mathbf{r}_3^j - \mathbf{r}_1^j, \mathbf{r}_3^j - \mathbf{r}_2^j, \mathbf{r}_2^j - \mathbf{r}_1^j$ , ενώ ο παρονομαστής είναι το εμβαδό του παραλληλογράμμου που ορίζουν



στο τετράγωνο.

**Λήμμα 7.2:** Έστω ότι οι συνθήκες του Λήμματος 7.1 ισχύουν για μία συλλογή πινάκων  $\hat{\mathbf{L}}_j$   $j = 1 \cdots N$  και  $\Delta \hat{\mathbf{L}}_k$   $k = 1 \cdots M$ . Ορίζουμε τους πίνακες

$$\hat{\mathbf{Z}}^{(N)} \triangleq \frac{1}{N} \sum_{j=1}^N \hat{\mathbf{L}}_j, \quad \mathbf{Z}^{(N)} \triangleq \frac{1}{N} \sum_{j=1}^N \mathbf{L}_j \quad (7-32)$$

$$\hat{\mathbf{X}}^{(N)} \triangleq \frac{1}{N} \sum_{j=1}^N \hat{\mathbf{L}}_j^T \hat{\mathbf{L}}_j, \quad \mathbf{X}^{(N)} \triangleq \frac{1}{N} \sum_{j=1}^N \mathbf{L}_j^T \mathbf{L}_j, \quad (7-33)$$

$$\hat{\mathbf{Y}}^{(M)} \triangleq \frac{1}{M} \sum_{k=1}^M \Delta \hat{\mathbf{L}}_k^T \Delta \hat{\mathbf{L}}_k, \quad \mathbf{Y}^{(M)} \triangleq \frac{1}{M} \sum_{k=1}^M \Delta \mathbf{L}_k^T \Delta \mathbf{L}_k. \quad (7-34)$$

Υποθέτουμε επίσης ότι τα  $\sigma_k^2, \pi_j^2$  αποτελούν φραγμένες ακολουθίες ως προς τα  $k, j$  αντίστοιχα και ότι το  $\mathbf{Y}^{(M)}$  αποτελεί φραγμένη ακολουθία ως προς  $M$ . Τότε,

$$E\{\hat{\mathbf{X}}^{(N)}\} = \mathbf{X}^{(N)} + \frac{\sum_{j=1}^N \pi_j^2}{N} \mathbf{I}, \quad (7-35)$$

$$E\{\hat{\mathbf{Y}}^{(M)}\} = \mathbf{Y}^{(M)} + \frac{\sum_{k=1}^M \sigma_k^2}{M} \mathbf{I},$$

$$E\{\hat{\mathbf{Z}}^{(N)}\} = \mathbf{Z}^{(N)}.$$

Ενώ,

$$\hat{\mathbf{X}}^{(N)} \rightarrow E\{\hat{\mathbf{X}}^{(N)}\}, \quad \hat{\mathbf{Y}}^{(M)} \rightarrow E\{\hat{\mathbf{Y}}^{(M)}\}, \quad \hat{\mathbf{Z}}^{(N)} \rightarrow E\{\hat{\mathbf{Z}}^{(N)}\} \quad (7-36)$$

στην MSE λογική, όταν  $M, N \rightarrow \infty$ .

**Απόδειξη:**

(α) Το πρώτο μέρος του Λήμματος 7.2 (εξίσωση (7-35)) είναι ευθύ αποτέλεσμα του Λήμματος 7.1 (εξισώσεις (7-26)-(7-30)). Για παράδειγμα,

$$E\{\hat{\mathbf{X}}^{(N)}\} = \frac{1}{N} \sum_{j=1}^N E\{\hat{\mathbf{L}}_j^T \hat{\mathbf{L}}_j\} = \frac{1}{N} \sum_{j=1}^N \mathbf{L}_j^T \mathbf{L}_j + \frac{\sum_{j=1}^N \pi_j^2}{N} \mathbf{I}.$$

(β) Και οι τρεις τύποι της εξίσωσης (7-36) αποδεικνύονται με τον ίδιο τρόπο. Εδώ, θα αποδείξουμε ενδεικτικά τον τύπο που αφορά το  $\hat{\mathbf{X}}^{(N)}$ .

Μπορεί να αποδειχθεί ότι  $\frac{1}{N} \sum_{j=1}^N (\hat{\mathbf{L}}_j^T \hat{\mathbf{L}}_j - E\{\hat{\mathbf{L}}_j^T \hat{\mathbf{L}}_j\}) \rightarrow \mathbf{0}$ , ή:

$$\frac{1}{N} \sum_{j=1}^N (\mathbf{L}_j^T \mathbf{L}_j^e + \mathbf{L}_j^e^T \mathbf{L}_j + \mathbf{L}_j^e^T \mathbf{L}_j^e - \pi_j^2 \mathbf{I}) \rightarrow \mathbf{0} \quad (7-37)$$

Για να αποδειχθεί αυτή η εξίσωση, θα δείξουμε ότι κάθε ένας από τους προσθετέους είναι φραγμένος με χρήση της συνάρτησης *trace*. Γι' αυτό το λόγο, οι ιδιότητες της συνάρτησης *trace* για συμμετρικούς πίνακες θα χρησιμοποιηθούν όπου είναι απαραίτητο.

(β.1) Για τον όρο  $\frac{1}{N} \sum_{j=1}^N \mathbf{L}_j^T \mathbf{L}_j^e$ , έχουμε:

$$E\{ \text{tr}(\mathbf{L}_j \mathbf{L}_j^T \mathbf{L}_j^e \mathbf{L}_j^{eT}) \} = \text{tr}(\mathbf{L}_j \mathbf{L}_j^T E\{ \mathbf{L}_j^e \mathbf{L}_j^{eT} \}) \leq \frac{1}{2} ( (\text{tr}(\mathbf{L}_j \mathbf{L}_j^T))^2 + (\text{tr}(E\{ \mathbf{L}_j^e \mathbf{L}_j^{eT} \}))^2 )$$

Μετά από μερικές πράξεις

$$\text{tr}(\mathbf{L}_j \mathbf{L}_j^T) \leq 2 \text{tr}(\hat{\mathbf{K}}_j^T \hat{\mathbf{K}}_j) + 4 \leq 4 \text{tr}(\mathbf{K}_j^T \mathbf{K}_j) + 12 .$$

$$\text{Εφόσον, } \mathbf{K}_j = \mathbf{R}_{2 \times 3} \begin{bmatrix} 1 & 0 & p_j \\ 0 & 1 & q_j \end{bmatrix}^T, \text{ είναι}$$

$$\mathbf{R} \begin{bmatrix} 1 & 0 & p_j \\ 0 & 1 & q_j \end{bmatrix}^T = \begin{bmatrix} \mathbf{K} & \\ r_{13} + r_{33}p & r_{23} + r_{33}q \end{bmatrix} ,$$

και  $\|\mathbf{R} [1 \ 0 \ p_j]^T\| = \|[1 \ 0 \ p_j]^T\|$ ,  $\|\mathbf{R} [0 \ 1 \ q_j]^T\| = \|[0 \ 1 \ q_j]^T\|$ . Συνδυάζοντας τις παραπάνω εξισώσεις,

$$\text{tr}(\mathbf{K}_j^T \mathbf{K}_j) = \left\| \begin{bmatrix} 1 \\ 0 \\ p_j \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 0 \\ 1 \\ q_j \end{bmatrix} \right\|^2 - (r_{13} + r_{33}p_j)^2 - (r_{23} + r_{33}q_j)^2 \leq 2 + p_j^2 + q_j^2$$

Μπορούμε να δούμε ότι  $p_j^2 + q_j^2 = \frac{1}{\xi_{zj}^2} - 1$ , οπότε τελικά  $\text{tr}(\mathbf{L}_j \mathbf{L}_j^T) \leq \frac{4}{\xi_{zj}^2} + 16$ .

Μπορεί να τονιστεί εδώ ότι το  $\text{tr}(\mathbf{L}_j \mathbf{L}_j^T)$  είναι φραγμένο, αφού το  $\xi_{zj}$  είναι μηδέν μόνο όταν η επίπεδη επιφάνεια που ορίζεται από την  $j$  τριπλέτα σημείων είναι κάθετη στο επίπεδο προβολής. Μπορεί κανείς να δει ότι τέτοιες τριπλέτες σημείων δεν μπορούν να επιλεγούν στο επίπεδο προβολής αφού προβάλλονται πάνω σε μία ευθεία γραμμή. Γενικά, είναι στην ευχέρεια του χρήστη να διαλέξει τέτοιες τριπλέτες σημείων ώστε το  $|\xi_{zj}|$  να είναι αρκετά μεγαλύτερο από το μηδέν.

Παράλληλα,  $\text{tr}(E\{ \mathbf{L}_j^e \mathbf{L}_j^{eT} \}) = \text{tr}(E\{ \mathbf{L}_j^T \mathbf{L}_j^e \}) = 2\pi_j^2$  και το  $\text{tr}(E\{ \mathbf{L}_j^e \mathbf{L}_j^{eT} \})$  είναι φραγμένο, αφού η ορίζουσα στο  $\pi_j^2$  είναι μηδέν μόνο όταν τα εμπλεκόμενα διαφορικά διανύσματα θέσης είναι συνευθειακά. Επιπρόσθετα, η ορίζουσα αυτή μπορεί να επιλεγεί λογικά μικρή βάσει της επιλογής των τριπλετών σημείων. Έτσι, τελικά,  $\frac{1}{N} \sum_{j=1}^N \mathbf{L}_j^T \mathbf{L}_j^e \rightarrow \mathbf{0}$ .

(β.2) Όμοια με το (β.1) αποδεικνύεται ότι  $\frac{1}{N} \sum_{j=1}^N \mathbf{L}_j^T \mathbf{L}_j \rightarrow \mathbf{0}$ .

(β.3) Η απόδειξη της (7-37) ολοκληρώνεται τελικά αποδεικνύοντας ότι  $\frac{1}{N} \sum_{j=1}^N (\mathbf{L}_j^{eT} \mathbf{L}_j^e - \pi_j^2 \mathbf{I}) \rightarrow \mathbf{0}$  ή  $\frac{1}{N} \sum_{j=1}^N (\mathbf{L}_j^{eT} \mathbf{L}_j^e - E\{\mathbf{L}_j^{eT} \mathbf{L}_j^e\}) \rightarrow \mathbf{0}$ . Μπορεί να δειχθεί μετά από πράξεις ότι αρκεί να δείξουμε ότι

$$\frac{1}{N} \sum_{j=1}^N (E\{\text{tr}(\mathbf{L}_j^e \mathbf{L}_j^{eT} \mathbf{L}_j^e \mathbf{L}_j^{eT})\} - \text{tr}(E\{\mathbf{L}_j^{eT} \mathbf{L}_j^e\} E\{\mathbf{L}_j^e \mathbf{L}_j^{eT}\})) \text{ φραγμένο.} \quad (7-38)$$

Ο όρος  $\text{tr}(E\{\mathbf{L}_j^{eT} \mathbf{L}_j^e\} E\{\mathbf{L}_j^e \mathbf{L}_j^{eT}\}) \leq (\text{tr}(E\{\mathbf{L}_j^{eT} \mathbf{L}_j^e\}))^2$  έχει δειχθεί ήδη ότι είναι φραγμένος. Για τον όρο  $E\{\text{tr}(\mathbf{L}_j^e \mathbf{L}_j^{eT} \mathbf{L}_j^e \mathbf{L}_j^{eT})\} \leq E\{(\text{tr}(\mathbf{L}_j^e \mathbf{L}_j^{eT}))^2\}$ , χρησιμοποιώντας τις αλγεβρικές πράξεις της απόδειξης του Λήμματος 7.1, είναι:

$$\text{tr}(\mathbf{L}_j^e \mathbf{L}_j^{eT}) = \frac{1}{(\det[\boldsymbol{\delta}_1 \boldsymbol{\delta}_2])^2} \text{tr} \left( \begin{bmatrix} \|\boldsymbol{\delta}_1^j\|^2 & -\langle \boldsymbol{\delta}_1^j, \boldsymbol{\delta}_2^j \rangle \\ -\langle \boldsymbol{\delta}_1^j, \boldsymbol{\delta}_2^j \rangle & \|\boldsymbol{\delta}_2^j\|^2 \end{bmatrix} \begin{bmatrix} \|\boldsymbol{\epsilon}_2^j\|^2 & \langle \boldsymbol{\epsilon}_1^j, \boldsymbol{\epsilon}_2^j \rangle \\ \langle \boldsymbol{\epsilon}_1^j, \boldsymbol{\epsilon}_2^j \rangle & \|\boldsymbol{\epsilon}_1^j\|^2 \end{bmatrix} \right)$$

και

$$\begin{aligned} \text{tr}(\mathbf{L}_j^e \mathbf{L}_j^{eT}) &\leq \frac{1}{2(\det[\boldsymbol{\delta}_1 \boldsymbol{\delta}_2])^2} ((\text{tr} \left( \begin{bmatrix} \|\boldsymbol{\delta}_1^j\|^2 & -\langle \boldsymbol{\delta}_1^j, \boldsymbol{\delta}_2^j \rangle \\ -\langle \boldsymbol{\delta}_1^j, \boldsymbol{\delta}_2^j \rangle & \|\boldsymbol{\delta}_2^j\|^2 \end{bmatrix} \right))^2 \\ &+ (\text{tr} \left( \begin{bmatrix} \|\boldsymbol{\epsilon}_2^j\|^2 & \langle \boldsymbol{\epsilon}_1^j, \boldsymbol{\epsilon}_2^j \rangle \\ \langle \boldsymbol{\epsilon}_1^j, \boldsymbol{\epsilon}_2^j \rangle & \|\boldsymbol{\epsilon}_1^j\|^2 \end{bmatrix} \right))^2) \end{aligned}$$

Όσον αφορά το τελευταίο, εφόσον η ορίζουσα μπορεί να διατηρηθεί λογικά μικρή από το χρήστη όπως αναφέρθηκε προηγουμένως και οι προσδοκητές τιμές για τις αμιγώς στοχαστικές ποσότητες δεν επηρεάζονται από το  $j$ , συμπεραίνουμε ότι το  $\frac{1}{N} \sum_{j=1}^N E\{(\text{tr}(\mathbf{L}_j^e \mathbf{L}_j^{eT}))^2\}$  είναι επίσης φραγμένο.

Οπότε, η εξίσωση (7-38) ισχύει και η εξίσωση (7-37) αποδεικνύεται. Με αυτόν τον τρόπο ολοκληρώνεται η απόδειξη της εξίσωσης (7-36) του Λήμματος 7.2 για το  $\hat{\mathbf{X}}^{(N)}$ . Με όμοιο τρόπο, η εξίσωση (7-36) αποδεικνύεται και για τα  $\hat{\mathbf{Z}}^{(N)}, \hat{\mathbf{Y}}^{(M)}$ . Ο.Ε.Δ.

Βάσει των προηγούμενων δύο Λημμάτων, οι επόμενες Προτάσεις 7.6, 7.7 και 7.8 τεκμηριώνουν μία διαδικασία η οποία επιτρέπει τον υπολογισμό των πινάκων περιστροφής  $\mathbf{R}$  και  $\mathbf{S}$  από θορυβώδη διανύσματα κίνησης. Οι Προτάσεις 7.6, 7.7 και 7.8 επεκτείνουν (στην πραγματικότητα αντικαθιστούν) τις Προτάσεις 7.3, 7.4 και 7.5 σε θορυβώδη περιβάλλοντα και υποδεικνύουν πώς η χρήση όσο το δυνατόν περισσότερων διανυσμάτων θέσης και κίνησης, διαθέσιμων από τρεις σκηνές, μπορεί να δώσει ασυμπτωτικά ακριβείς μετρήσεις των  $\mathbf{R}$  και  $\mathbf{S}$ .

Για την τεκμηρίωση των Προτάσεων 7.6, 7.7 και 7.8 θα υιοθετηθεί ο παρακάτω συμβολισμός. Έστω τα  $\mathbf{r}_1, \mathbf{r}_2, \mathbf{c}_1, \mathbf{c}_2, \mathbf{s}_1, \mathbf{s}_2, \mathbf{d}_1, \mathbf{d}_2, \rho, \sigma, w$  όπως ορίζονται στις Προτάσεις 7.3, 7.4 και 7.5. Έστω επίσης τα  $\hat{\mathbf{Y}}_R^{(M)}, \mathbf{Y}_R^{(M)}, \hat{\mathbf{Z}}_R^{(N)}, \mathbf{Z}_R^{(N)}$  και  $\hat{\mathbf{Y}}_S^{(M)}, \mathbf{Y}_S^{(M)}, \hat{\mathbf{Z}}_S^{(N)}, \mathbf{Z}_S^{(N)}$  όπως ορίζονται

στο Λήμμα 7.2, όπου οι δείκτες  $R$  και  $S$  συμβολίζουν ότι τα αντίστοιχα πεδία κίνησης αντιστοιχούν σε περιστροφές που περιγράφονται από πίνακες  $\mathbf{R}$  και  $\mathbf{S}$  αντίστοιχα. Υποθέτουμε ότι τα εμπλεκόμενα πεδία κίνησης και ο θόρυβος στις μετρήσεις τους ικανοποιούν τις συνθήκες των Λημμάτων 1 και 2. Έστω επιπρόσθετα ότι τα  $\lambda_{Rk}, \lambda_{Sk}$  συμβολίζουν τις ιδιοτιμές των τάξης-1 πινάκων  $(\Delta \mathbf{L}_{Rk})^T \Delta \mathbf{L}_{Rk}, (\Delta \mathbf{L}_{Sk})^T \Delta \mathbf{L}_{Sk}$  αντίστοιχα.

### Πρόταση 7.6:

(α) Οι ιδιοτιμές της προσδοκητής τιμής  $E\{\hat{\mathbf{Y}}_R^{(M)}\}$  είναι ίσες με  $\lambda_{Rmax} = \lambda_R + \sigma_R^2$  και  $\lambda_{Rmin} = \sigma_R^2$  και αντιστοιχούν στα μοναδιαία ιδιοδιανύσματα  $\mathbf{Jc}_1$  και  $\mathbf{c}_1$  αντίστοιχα, όπου  $\lambda_R = \frac{1}{M} \sum_{k=1}^M \lambda_{Rk}$  είναι η μη μηδενική ιδιοτιμή του πίνακα  $\mathbf{Y}_R^{(M)}$  και  $\sigma_R^2 = \frac{\sum_{k=1}^M \sigma_{Rk}^2}{M}$ .

(β) Όμοιες εκφράσεις ισχύουν για τον πίνακα  $\mathbf{S}$  για κατάλληλα ορισμένους  $\hat{\mathbf{Y}}_S^{(M)}, \mathbf{Y}_S^{(M)}$ .

(γ) Η εξίσωση (7-13) ισχύει εάν αντικαταστήσουμε τα  $\mathbf{L}_{Rj}, \mathbf{L}_{Sj}$  με  $E\{\mathbf{Z}_R^{(N)}\}$  και  $E\{\mathbf{Z}_S^{(N)}\}$  αντίστοιχα.

(δ) Όπως  $M \rightarrow \infty$  ισχυρά συνεπείς εκτιμήσεις  $\hat{\mathbf{c}}_1^{(M)}, \hat{\mathbf{c}}_2^{(M)}, \hat{\mathbf{d}}_1^{(M)}, \hat{\mathbf{d}}_2^{(M)}$  των  $\mathbf{c}_1, \mathbf{c}_2, \mathbf{d}_1, \mathbf{d}_2$  λαμβάνονται αντιστοίχως εάν στα (α), (β) και (γ) αντικαταστήσουμε το  $E\{\hat{\mathbf{Y}}_R^{(M)}\}$  με  $\hat{\mathbf{Y}}_R^{(M)}$ , το  $E\{\hat{\mathbf{Y}}_S^{(M)}\}$  με  $\hat{\mathbf{Y}}_S^{(M)}$ , το  $E\{\hat{\mathbf{Z}}_R^{(N)}\}$  με  $\hat{\mathbf{Z}}_R^{(N)}$  και το  $E\{\hat{\mathbf{Z}}_S^{(N)}\}$  με  $\hat{\mathbf{Z}}_S^{(N)}$ .

### Απόδειξη:

(α) Στην Πρόταση 7.3, δείξαμε ότι τα μοναδιαία ιδιοδιανύσματα του πίνακα  $(\Delta \mathbf{L}_R)^T (\Delta \mathbf{L}_R)$  είναι  $\mathbf{c}_1$  και  $\mathbf{Jc}_1$  και αντιστοιχούν σε ιδιοτιμές 0 και  $\lambda_R$  αντίστοιχα. Ο πίνακας  $\mathbf{Y}_R^{(M)}$ , ο οποίος είναι άθροισμα από πίνακες  $(\Delta \mathbf{L}_R)^T (\Delta \mathbf{L}_R)$ , έχει τα ίδια ιδιοδιανύσματα, που αντιστοιχούν τώρα σε ιδιοτιμές 0 και  $\lambda_R = \frac{1}{M} \sum_{k=1}^M \lambda_{Rk}$ , όπου  $\lambda_{Rk}$  είναι η μη μηδενική ιδιοτιμή του πίνακα  $(\Delta \mathbf{L}_{Rk})^T (\Delta \mathbf{L}_{Rk})$ , σύμφωνα με την εξίσωση (7-19). Αυτή η ιδιότητα του  $\mathbf{Y}_R^{(M)}$ , σε σχέση και με τις εξισώσεις (7-35) του Λήμματος 7.2, τεκμηριώνει την απόδειξη. Είναι σημαντικό να παρατηρήσει κανείς ότι  $\lambda_{Rmax} > \lambda_{Rmin}$  αφού  $\lambda_R > 0$  και ο  $\mathbf{Y}_R^{(M)}$  είναι θετικά ημιορισμένος.

(β) Όμοιες εκφράσεις ισχύουν για τον πίνακα  $\mathbf{S}$  χρησιμοποιώντας την ίδια απόδειξη και αντικαθιστώντας τον δείκτη  $R$  με  $S$ .

(γ) Είναι άμεση συνέπεια του Λήμματος 7.2 (εξισώσεις (7-32) και (7-35)).

(δ) Είναι άμεση συνέπεια των ιδιοτήτων ασυμπτωτικής σύγκλισης των  $\hat{\mathbf{Y}}_R^{(M)}, \hat{\mathbf{Y}}_S^{(M)}, \hat{\mathbf{Z}}_R^{(N)}$  και  $\hat{\mathbf{Z}}_S^{(N)}$  όπως παρουσιάζονται στο Λήμμα 7.2, εξισώσεις (7-36). Ο.Ε.Δ.

Αξιίζει να παρατηρηθεί εδώ ότι ο υπολογισμός του  $\mathbf{c}_1$  με την παραπάνω μέθοδο είναι η λύση με ελάχιστα τετράγωνα του υπερ-ορισμένου συστήματος:

$$\left[ \begin{array}{ccc} (\Delta \hat{\mathbf{L}}_{R1}^T \Delta \hat{\mathbf{L}}_{R1})^T & \cdots & (\Delta \hat{\mathbf{L}}_{RM}^T \Delta \hat{\mathbf{L}}_{RM})^T \end{array} \right]^T \mathbf{c}_1 = \mathbf{0}$$

υπό τον περιορισμό  $\|\mathbf{c}_1\|^2 = \mathbf{c}_1^T \mathbf{c}_1 = 1$  (βλ. για παράδειγμα και στο [104], σελ. 151-155).

**Πρόταση 7.7:** Εάν  $\lambda_{Rmax}$  ( $\lambda_{Smax}$ ) και οι  $\lambda_{Rmin}$  ( $\lambda_{Smin}$ ) ορίζονται όπως στην Πρόταση 7.6, τότε:

(α) Το  $w$  δίνεται τώρα από την

$$w^2 = \frac{\lambda_{Smax} - \lambda_{Smin}}{\lambda_{Rmax} - \lambda_{Rmin}} \quad (7-39)$$

Το  $w$  υπακούει στην εξίσωση (7-39) και στην περίπτωση απουσίας θορύβου, όπου  $\lambda_{Rmin} = \lambda_{Smin} = 0$ .

(β) Όπως  $M \rightarrow \infty$  ισχυρά συνεπείς εκτιμήσεις  $\hat{w}^{(M)}$  του  $w$  λαμβάνονται, εάν στη διαδικασία υπολογισμού των ιδιοτιμών αντικαταστήσουμε το  $E\{\hat{\mathbf{Y}}_R^{(M)}\}$  με  $\hat{\mathbf{Y}}_R^{(M)}$  και το  $E\{\hat{\mathbf{Y}}_S^{(M)}\}$  με  $\hat{\mathbf{Y}}_S^{(M)}$ .

**Απόδειξη:** (α) Είναι

$$w^2 = \frac{\lambda_{Smax} - \lambda_{Smin}}{\lambda_{Rmax} - \lambda_{Rmin}} = \frac{\lambda_S + \sigma_S^2 - \sigma_S^2}{\lambda_R + \sigma_R^2 - \sigma_R^2} = \frac{\lambda_S}{\lambda_R} = \frac{\sigma^2 \frac{1}{M} \sum_{k=1}^M ((\Delta q_k)^2 + (\Delta p_k)^2)}{\rho^2 \frac{1}{M} \sum_{k=1}^M ((\Delta q_k)^2 + (\Delta p_k)^2)} = \frac{\sigma^2}{\rho^2}$$

(β) Είναι άμεση συνέπεια της ασυμπτωτικής σύγκλισης των  $\hat{\mathbf{Y}}_R^{(M)}$ ,  $\hat{\mathbf{Y}}_S^{(M)}$ , στις προσδοκητές τιμές τους του Λήμματος 7.2, εξισώσεις (7-36). *Ο.Ε.Δ.*

Η επόμενη Πρόταση αντικαθιστά την Πρόταση 7.5 για τον υπολογισμό των άγνωστων στοιχείων  $r_{33}$ ,  $s_{33}$  των πινάκων περιστροφής. Μερικοί ορισμοί είναι αναγκαίοι σ' αυτό το σημείο. Η εξίσωση (7-20) μπορεί να ξαναγραφεί στη μορφή:

$$-\left[ \begin{array}{cc} w\mathbf{J}\mathbf{L}_{Rj}\mathbf{c}_1 & -\mathbf{J}\mathbf{L}_{Sj}\mathbf{d}_1 \end{array} \right] \left[ \begin{array}{c} r_{33} \\ s_{33} \end{array} \right] = \left[ \begin{array}{cc} \mathbf{L}_{Sj} & \mathbf{L}_{Rj} \end{array} \right] \left[ \begin{array}{c} \mathbf{J}\mathbf{d}_1 \\ -w\mathbf{J}\mathbf{c}_1 \end{array} \right]$$

Εάν το  $\theta = [r_{33} \ s_{33}]^T$  είναι ένα διάνυσμα περιέχον τους αγνώστους, μπορούμε να γράψουμε την προηγούμενη εξίσωση ως  $\mathbf{A}_j \theta = \mathbf{b}_j$ . Χρησιμοποιώντας  $N$  τριπλέτες σημείων,

$$\left[ \begin{array}{c} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_N \end{array} \right] \theta = \left[ \begin{array}{c} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{array} \right], \quad \text{ή} \quad \mathbf{A}\theta = \mathbf{b}$$

Στην περίπτωση με θορυβώδεις μετρήσεις, αντί των πινάκων  $\mathbf{A}$ ,  $\mathbf{b}$ , λαμβάνουμε τους πίνακες  $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{A}^e$  και  $\hat{\mathbf{b}} = \mathbf{b} + \mathbf{b}^e$  αντίστοιχα. Δημιουργούμε τότε το υπερ-ορισμένο σύστημα εξισώσεων:

$$\mathbf{A}\theta = \hat{\mathbf{b}} \quad (7-40)$$

το οποίο λύνεται με ελάχιστα τετράγωνα. Δυστυχώς ο πίνακας  $\mathbf{A}$  στην εξίσωση (7-40) δεν είναι τότε διαθέσιμος. Οπότε η λύση ελαχίστων τετραγώνων  $\hat{\theta}_{LS}$  της (7-40) δεν υπολογίζεται άμεσα. Η επόμενη Πρόταση προτείνει μία εναλλακτική εκτιμήτρια  $\tilde{\theta}_{LS}^{(N)}$  που φαίνεται να συγκλίνει ασυμπτωτικά στο  $\theta_{LS}^{(N)}$ . Η εναλλακτική έκφραση εκφράζεται βάσει των παρατηρούμενων διανυσμάτων κίνησης.

### Πρόταση 7.8:

(α) Η λύση ελαχίστων τετραγώνων  $\hat{\theta}_{LS}^{(N)} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \hat{\mathbf{b}}$  της (7-40) μπορεί να προσεγγιστεί ασυμπτωτικά από την εκτιμήτρια

$$\tilde{\theta}_{LS}^{(N)} = \frac{1}{N} \left( \frac{1}{N} \hat{\mathbf{A}}^T \hat{\mathbf{A}} - \begin{bmatrix} w^2 \pi_R^2 & 0 \\ 0 & \pi_S^2 \end{bmatrix} \right)^{-1} \hat{\mathbf{A}}^T \hat{\mathbf{b}} \quad (7-41)$$

όπου  $\pi_R^2 = \frac{\sum_{j=1}^N \pi_{Rj}^2}{N}$ ,  $\pi_S^2 = \frac{\sum_{j=1}^N \pi_{Sj}^2}{N}$ . Όπως  $N \rightarrow \infty$  ισχυρά συνεπείς μετρήσεις  $\hat{r}_{33}^{(N)}$ ,  $\hat{s}_{33}^{(N)}$  των  $r_{33}$  και  $s_{33}$  λαμβάνονται αντίστοιχα.

(β) Υπό την υπόθεση ότι πίνακες  $\Delta \mathbf{L}_k$  υπολογίζονται από διαφορετικούς πίνακες  $\mathbf{L}$ , έχουμε  $\pi_R^2 = \frac{M}{N} \sigma_R^2$  και  $\pi_S^2 = \frac{M}{N} \sigma_S^2$ .

### Απόδειξη:

(α) Η απόδειξη στηρίζεται στην υπόθεση ότι οι πραγματικές τιμές των  $\mathbf{c}_1$ ,  $\mathbf{d}_1$  και  $w$  χρησιμοποιούνται στις εκφράσεις των  $\mathbf{A}_j$  ( $j = 1 \cdots N$ ). Τα ίδια αποτελέσματα σύγκλισης μπορούν να τεκμηριωθούν ακόμη κι αν συνεπείς μετρήσεις αυτών χρησιμοποιούνται, οδηγώντας όμως σε πολύπλοκη μαθηματική ανάλυση. Υπό την υπόθεση αυτή, η σύγκλιση τεκμηριώνεται αποδεικνύοντας ότι (ι)  $\frac{1}{N} \hat{\mathbf{A}}^T \hat{\mathbf{A}} - \begin{bmatrix} w^2 \pi_R^2 & 0 \\ 0 & \pi_S^2 \end{bmatrix} \rightarrow \frac{1}{N} \mathbf{A}^T \mathbf{A}$  και (ii)  $\frac{1}{N} \hat{\mathbf{A}}^T \hat{\mathbf{b}} \rightarrow \frac{1}{N} \mathbf{A}^T \mathbf{b}$ .

(i) Από τον ορισμό του πίνακα  $\hat{\mathbf{A}}$  λαμβάνουμε

$$\frac{1}{N} \hat{\mathbf{A}}^T \hat{\mathbf{A}} = \begin{bmatrix} w^2 \mathbf{c}_1^T (\frac{1}{N} \sum_{j=1}^N \hat{\mathbf{L}}_{Rj}^T \hat{\mathbf{L}}_{Rj}) \mathbf{c}_1 & -w \mathbf{c}_1^T (\frac{1}{N} \sum_{j=1}^N \hat{\mathbf{L}}_{Rj}^T \hat{\mathbf{L}}_{Sj}) \mathbf{d}_1 \\ -w \mathbf{d}_1^T (\frac{1}{N} \sum_{j=1}^N \hat{\mathbf{L}}_{Sj}^T \hat{\mathbf{L}}_{Rj}) \mathbf{c}_1 & \mathbf{d}_1^T (\frac{1}{N} \sum_{j=1}^N \hat{\mathbf{L}}_{Sj}^T \hat{\mathbf{L}}_{Sj}) \mathbf{d}_1 \end{bmatrix}.$$

Χρησιμοποιώντας την εξίσωση (7-35) για τον  $\hat{\mathbf{X}}^{(N)}$ , συμπεραίνουμε ότι  $E\{ \frac{1}{N} \hat{\mathbf{A}}^T \hat{\mathbf{A}} \} = \frac{1}{N} \mathbf{A}^T \mathbf{A} + \begin{bmatrix} w^2 \pi_R^2 & 0 \\ 0 & \pi_S^2 \end{bmatrix}$ .

Επιπρόσθετα, θαδειχτεί ότι η ποσότητα  $\frac{1}{N}\hat{\mathbf{A}}^T\hat{\mathbf{A}}$  συγκλίνει στην προσδοκητή της τιμή. Για τους αντι-διαγώνιους όρους, προκύπτουν οι παρακάτω σχέσεις, οι οποίες και αποδεικνύονται χρησιμοποιώντας τη μεθοδολογία που ακολουθείται στην απόδειξη του Λήμματος 7.2:

$$\frac{1}{N}\sum_{j=1}^N\hat{\mathbf{L}}_{Rj}^T\hat{\mathbf{L}}_{Sj}\rightarrow E\left\{\frac{1}{N}\sum_{j=1}^N\hat{\mathbf{L}}_{Rj}^T\hat{\mathbf{L}}_{Sj}\right\}, \quad \frac{1}{N}\sum_{j=1}^N\hat{\mathbf{L}}_{Sj}^T\hat{\mathbf{L}}_{Rj}\rightarrow E\left\{\left(\frac{1}{N}\sum_{j=1}^N\hat{\mathbf{L}}_{Sj}^T\hat{\mathbf{L}}_{Rj}\right)\right\}.$$

Από τις οριακές αυτές σχέσεις των αντι-διαγώνιων όρων και τις οριακές σχέσεις του Λήμματος 7.2 (την εξίσωση (7-36) για τον  $\hat{\mathbf{X}}^{(N)}$ ) για τους διαγώνιους όρους, καταλήγουμε στην επιθυμητή τελική οριακή σχέση.

(ii) Εφόσον  $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{A}^e$ , για να αποδείξουμε τη δεύτερη οριακή σχέση, αρκεί να δείξουμε ότι  $\frac{1}{N}(\mathbf{A}^e)^T\hat{\mathbf{b}} \rightarrow \mathbf{0}$ . Με αυτό το στόχο, αρχικά παρατηρούμε ότι η μέση τιμή της έκφρασης αυτής είναι μηδενική,

$$E\left\{\frac{1}{N}(\mathbf{A}^e)^T\hat{\mathbf{b}}\right\} = E\left\{\frac{1}{N}(\mathbf{A}^e)^T\mathbf{b}^e\right\} =$$

$$\begin{bmatrix} -w\mathbf{c}_1^T E\left\{\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Rj}^e)^T\mathbf{L}_{Sj}^e\mathbf{J}\right\}\mathbf{d}_1 + w^2\mathbf{c}_1^T E\left\{\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Rj}^e)^T\mathbf{L}_{Rj}^e\mathbf{J}\right\}\mathbf{c}_1 \\ \mathbf{d}_1^T E\left\{\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Sj}^e)^T\mathbf{L}_{Sj}^e\mathbf{J}\right\}\mathbf{d}_1 - w\mathbf{d}_1^T E\left\{\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Sj}^e)^T\mathbf{L}_{Rj}^e\mathbf{J}\right\}\mathbf{c}_1 \end{bmatrix} = \mathbf{0} \quad (7-42)$$

Οι άνω-αριστερά και κάτω-δεξιά όροι της (7-42) εξαφανίζονται, εφόσον οι όροι σφάλματος που αντιστοιχούν σε διαδοχικές περιστροφές είναι αμοιβαίως ασυσχέτιστοι. Στη συνέχεια παρατηρούμε ότι ο πάνω-δεξιά όρος του πίνακα είναι επίσης μηδέν, αφού

$$E\{(\mathbf{J}\mathbf{L}_{Rj}^e)^T\mathbf{L}_{Rj}^e\mathbf{J}\} = E\{\mathbf{J}^T\tilde{\mathbf{K}}_{Rj}^e\mathbf{J}(\tilde{\mathbf{K}}_{Rj}^e)^T\} = E\{\det(\tilde{\mathbf{K}}_{Rj}^e)\}\mathbf{I} = \mathbf{0} \quad (7-43)$$

Στην πρώτη ισότητα χρησιμοποιήσαμε τον ορισμό των πινάκων  $\mathbf{L}_{Rj}$  ενώ το τελευταίο βήμα είναι επόμενο της υπόθεσης μη συσχέτισης για τους όρους σφάλματος (εξ. (7-25)). Όμοια επιχειρήματα μπορούν να χρησιμοποιηθούν για να αποδείξουμε ότι ο κάτω-αριστερά όρος της (7-42) είναι επίσης μηδέν.

Το επόμενο βήμα είναι να δείξουμε ότι η ισχύς του  $\frac{1}{N}(\mathbf{A}^e)^T\hat{\mathbf{b}}$  τείνει στο μηδέν. Χρησιμοποιώντας τον ορισμό των  $\mathbf{A}$  και  $\mathbf{b}$  λαμβάνουμε

$$\begin{aligned} \frac{1}{N}(\mathbf{A}^e)^T\hat{\mathbf{b}} &= \frac{1}{N}\sum_{j=1}^N(\mathbf{A}_j^e)^T\hat{\mathbf{b}}_j = \\ &\begin{bmatrix} -w\mathbf{c}_1^T\left(\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Rj}^e)^T\hat{\mathbf{L}}_{Sj}\mathbf{J}\right)\mathbf{d}_1 + w^2\mathbf{c}_1^T\left(\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Rj}^e)^T\hat{\mathbf{L}}_{Rj}\mathbf{J}\right)\mathbf{c}_1 \\ \mathbf{d}_1^T\left(\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Sj}^e)^T\hat{\mathbf{L}}_{Sj}\mathbf{J}\right)\mathbf{d}_1 - w\mathbf{d}_1^T\left(\frac{1}{N}\sum_{j=1}^N(\mathbf{J}\mathbf{L}_{Sj}^e)^T\hat{\mathbf{L}}_{Rj}\mathbf{J}\right)\mathbf{c}_1 \end{bmatrix} \end{aligned} \quad (7-44)$$

Αρκεί να δείξουμε ότι η ισχύς των κανονικοποιημένων αθροισμάτων στις παρενθέσεις τείνει στο μηδέν όπως το  $N \rightarrow \infty$ . Διαλέγουμε να σχηματίσουμε την απόδειξη για τον δεύτερο

όρο της πρώτης συνιστώσας της (7-44). Η απόδειξη για τα υπόλοιπα τρία αθροίσματα εξάγεται με τον ίδιο ακριβώς τρόπο. Ως εκτίμηση της ισχύος του αθροίσματος αυτού χρησιμοποιούμε το ίχνος (*trace*) του πίνακα συμμεταβλητότητας, δηλαδή

$$\text{tr}(E\{\{\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{J}^T \hat{\mathbf{L}}_{Ri}^T \mathbf{J} \mathbf{L}_{Ri}^e (\mathbf{L}_{Rj}^e)^T \mathbf{J}^T \hat{\mathbf{L}}_{Rj} \mathbf{J}\}\}) = \frac{1}{N^2} \sum_{j=1}^N \text{tr}(E\{\hat{\mathbf{L}}_{Rj} \hat{\mathbf{L}}_{Rj}^T \mathbf{J} \mathbf{L}_{Rj}^e (\mathbf{L}_{Rj}^e)^T \mathbf{J}^T\}) \quad (7-45)$$

όπου χρησιμοποιήσαμε το γεγονός ότι τα  $\mathbf{J}^T \hat{\mathbf{L}}_{Ri}^T \mathbf{J} \mathbf{L}_{Ri}^e$  και  $(\mathbf{L}_{Rj}^e)^T \mathbf{J}^T \hat{\mathbf{L}}_{Rj} \mathbf{J}$  είναι αμοιβαία ασυσχέτιστα για  $i \neq j$  και οι μέσες τους τιμές μηδενικές (7-43). Επίσης εκμεταλλευτήκαμε τις ιδιότητες του τελεστή *trace*.

Περαιτέρω διεύρυνση της (7-45) οδηγεί σε ένα άθροισμα γινομένων που περικλείει αμιγώς ντετερμινιστικές και αμιγώς στοχαστικές εκφράσεις των πινάκων  $\mathbf{L}$ . Το κοινό χαρακτηριστικό αυτών των αθροισμάτων είναι ότι είναι τάξεως  $o(N)$ , το οποίο με τη σειρά του υποδηλώνει ότι πολλαπλασιασμός με  $\frac{1}{N^2}$  κάνει τη συνολική παράσταση να μηδενίζεται όταν  $N \rightarrow \infty$ . Η ακριβής απόδειξη αυτού του ισχυρισμού παραλείπεται εδώ χάριν απλότητας. Από την άλλη πλευρά, όμοιοι ισχυρισμοί χρησιμοποιούνται και στην απόδειξη του Λήμματος 7.2.

(β) Από τον ορισμό τους,  $\sigma_R^2 = \frac{\sum_{k=1}^M \sigma_{Rk}^2}{M}$  και  $\pi_R^2 = \frac{\sum_{j=1}^N \pi_{Rj}^2}{N}$ . Χρησιμοποιώντας την εξίσωση (7-30) λαμβάνουμε  $\sigma_R^2 = \frac{\sum_{k=1, k \sim (i,j)}^M (\pi_{Rj}^2 + \pi_{Ri}^2)}{M} = \frac{\sum_{j=1}^N \pi_{Rj}^2}{M}$ , εφόσον οι πίνακες  $\Delta \mathbf{L}_k$  κατασκευάζονται βάσει διαφορετικών πινάκων. Έτσι,  $N \pi_R^2 = M \sigma_R^2$  ή  $\pi_R^2 = \frac{M}{N} \sigma_R^2$ . Ομοίως,  $\pi_S^2 = \frac{M}{N} \sigma_S^2$ . Ο.Ε.Δ.

Ο Πίνακας 7.2 συνοψίζει τα βήματα του αλγορίθμου.

## 7.5 Συζήτηση και σύγκριση με άλλες μεθόδους

Στην ενότητα αυτή, η προτεινόμενη μέθοδος συγκρίνεται με δύο δημοφιλείς μεθόδους της βιβλιογραφίας για την εξαγωγή της  $3\Delta$  δομής και κίνησης από ακολουθίες εικόνων από παράλληλη προβολή - συγκεκριμένα, τις epipolar [61, 140, 106, 196] και τις factorization [159, 122, 99] μεθόδους. Μπορεί ναδειχθεί ότι η προτεινόμενη μέθοδος είναι, κατά κάποιον τρόπο, σχετιζόμενη και με τις δύο παραπάνω.

Στην πραγματικότητα, ο πίνακας  $\mathbf{L}$  (ισοδύναμα  $\mathbf{K}$  ή  $\tilde{\mathbf{K}}$ ) εμπεριέχει την epipolar εξίσωση για μία τριπλέτα αντιστοιχίσεων σημείων κατά μία μετάβαση. Θαδειχθεί ότι η κατασκευή του πίνακα  $\mathbf{K}$  στην εξίσωση (7-7) όπως και η epipolar εξίσωση για τρία σημεία οδηγούν στην ταυτότητα της εξίσωσης (7-13) για τον πίνακα  $\mathbf{L}$ . Για κάθε ζεύγος αντιστοιχίσεων σημείων



**Πίνακας 7.2** Διαδικασία υπολογισμού παρουσία θορύβου.**Για κάθε μετάβαση:**

- (i) μοιράζουμε τις διαθέσιμες αντιστοιχίσεις σημείων σε τριπλέτες (σύμφωνα με την Υποενότητα 7.6.2), και υπολογίζουμε τους πίνακες  $\hat{\mathbf{L}}_j$  από τις (7-7)-(7-8).
- (ii) ταξινομούμε τους πίνακες  $\hat{\mathbf{L}}_j$  σε ζεύγη (σύμφωνα με την Υποενότητα 7.6.2), και υπολογίζουμε τους πίνακες  $\Delta\hat{\mathbf{L}}_k$  όμοια με την εξίσωση (7-11).
- (iii) κατασκευάζουμε τον πίνακα  $\mathbf{Y}$  από την εξίσωση (7-34) και υπολογίζουμε ιδιοτιμές και ιδιοδιανύσματα  $\lambda_{max}$ ,  $\lambda_{min}$ ,  $\mathbf{c}_1$  σύμφωνα με την Πρόταση 7.6.

**Για κάθε ζεύγος μεταβάσεων:**

- (i) υπολογίζουμε το  $w^2$  από την εξίσωση (7-39) της Πρότασης 7.7.
- (ii) υπολογίζουμε το  $r_{33}$  για τους δύο πίνακες περιστροφής από την (7-41) της Πρότασης 7.8.

**Για κάθε μετάβαση:**

- (i) υπολογίζουμε τα υπόλοιπα στοιχεία του πίνακα περιστροφής, τα διανύσματα μετατόπισης και το σχετικό βάθος όπως και στην περίπτωση απουσίας θορύβου (Πίνακας 7.1).

( $\mathbf{p}_1 \rightarrow \mathbf{p}'_1$  και  $\mathbf{p}_2 \rightarrow \mathbf{p}'_2$ ), η epipolar εξίσωση, όπως εισάγεται στο [61], μπορεί να γραφεί με βάση το φορμαλισμό που υιοθετήθηκε, ως

$$(\mathbf{p}'_2 - \mathbf{p}'_1)^T \mathbf{J} \mathbf{r}_1 = -(\mathbf{p}_2 - \mathbf{p}_1)^T \mathbf{J} \mathbf{r}_2, \quad (7-46)$$

και εισάγοντας μία τρίτη αντιστοιχίση ( $\mathbf{p}_3 \rightarrow \mathbf{p}'_3$ ),

$$\begin{bmatrix} \mathbf{p}'_3 - \mathbf{p}'_1 & \mathbf{p}'_2 - \mathbf{p}'_1 \end{bmatrix}^T \mathbf{J} \mathbf{r}_1 = \begin{bmatrix} \mathbf{p}_3 - \mathbf{p}_1 & \mathbf{p}_2 - \mathbf{p}_1 \end{bmatrix}^T \mathbf{J} \mathbf{r}_2.$$

Με αντιστροφή και πράξεις, η εξίσωση (7-13) λαμβάνεται για τον πίνακα  $\mathbf{L}$ .

Σε σχέση με τις epipolar μεθόδους [140, 106, 196], μπορεί ακόμη ναδειχθεί ότι η προτεινόμενη μέθοδος δουλεύει ακόμη και για το μοντέλο της ασθενούς προοπτικής προβολής (ή αναλογικής παράλληλης). Σ'αυτήν την περίπτωση, για κάθε μετάβαση, συμπεριλαμβάνεται ένας ενιαίος παράγοντας μεγέθυνσης  $f$ , επιτρέποντας αντιστοιχίσεις χαρακτηριστικών του τύπου  $\mathbf{p}_1 \rightarrow f\mathbf{p}'_1$ . Μπορεί ναδειχθεί, μετά από πράξεις, ότι οι άγνωστοι πολλαπλασιαστικοί παράγοντες και για τις δύο μεταβάσεις εξαφανίζονται στις τελικές σχέσεις, οι οποίες δίνουν ακριβώς τις παραμέτρους περιστροφής. Με αυτή τη συλλογιστική, η προτεινόμενη μεθοδολογία εξάγει την 3Δ δομή και κίνηση και για παράλληλη και για αναλογική παράλληλη προβολή (όμοια με την epipolar μέθοδο [196]).

Όσον αφορά στην προτεινόμενη μέθοδο, σε σχέση με τη μέθοδο factorization, μπορεί ναδειχθεί ότι όμοιες εκφράσεις χρησιμοποιούνται και στις δύο περιπτώσεις για τον υπολογισμό της 3Δ κίνησης. Στην αναφορά [159], ένας πίνακας  $R$  ορίζεται να περιέχει τους οριζόντιους

και κάθετους άξονες αναφοράς της κάμερας μέσα στην ακολουθία ('the horizontal and vertical camera reference axes throughout the stream', σελίδα 139 της [159]). Στη συνέχεια, οι άξονες αναφοράς της κάμερας  $\mathbf{i}_0$  και  $\mathbf{j}_0$  αρχικοποιούνται ούτως ώστε να συμβαδίζουν με το παγκόσμιο σύστημα αναφοράς (για παράδειγμα  $\mathbf{i}_0 = [100]^T$ ,  $\mathbf{j}_0 = [010]^T$ ). Από την άλλη πλευρά, θεωρώντας τον  $3 \times 3$  πίνακα περιστροφής  $\mathbf{R}$ , όταν το αντικείμενο περιστρέφεται σύμφωνα με τον  $\mathbf{R}$  ισοδύναμα η κάμερα περιστρέφεται σύμφωνα με τον  $\mathbf{R}^T$ . Με αυτή τη συλλογιστική ο πίνακας,  $R$ , περιλαμβάνει για κάθε μετάβαση τις δύο πρώτες γραμμές του πίνακα περιστροφής για κάθε μετάβαση. Το τελευταίο σχετίζεται ευθέως με τον πίνακα  $\mathbf{K}$ , όπως εισάγεται από την εξίσωση (7-6) του παρόντος κεφαλαίου, δηλαδή ο πίνακας  $\mathbf{K}$  μπορεί να αναλυθεί ('factorized') στον πίνακα που περιέχει τις παραμέτρους περιστροφής (στον  $\mathbf{R}_{2 \times 3}$  που με τη σειρά του περιέχει τις 2 πρώτες γραμμές του πίνακα περιστροφής) και στον πίνακα που περιέχει τις παραμέτρους δομής  $(p, q)$ . Όπως και στη μέθοδο factorization, ένας σύνθετος (block) πίνακας μπορεί όμοια να ληφθεί επανυψώνοντας τον πίνακα  $\mathbf{K}$ , δηλαδή αυξάνοντας τις στήλες του και τις γραμμές του με νέους πίνακες  $\mathbf{K}$  για περισσότερες τριπλέτες και περισσότερες μεταβάσεις αντίστοιχα. Η στρατηγική που ακολουθείται στη μέθοδο factorization παρουσιάζεται βολική ως προς την εισαγωγή περισσότερων των τριών καρέ στον υπολογισμό της τρισδιάστατης δομής, οδηγώντας όμως σε έναν υπερβολικά μεγάλο πίνακα μετρήσεων (measurements matrix). Από την άλλη πλευρά, η επιλογή κατάλληλων τριπλετών σημείων είναι με τη μέθοδο factorization αδύνατη.

Γενικά, μπορεί κανείς να δει ότι οι τρεις μέθοδοι (η factorization, η epipolar και η προτεινόμενη) οδηγούν στη λύση ενός υπερ-ορισμένου ομογενούς συστήματος για κάποιες παραμέτρους περιστροφής σε ένα πρώτο βήμα. Επιπρόσθετα, όλες οι μέθοδοι εκμεταλλεύονται τις ιδιότητες της ανάλυσης σε ιδιοτιμές ή ιδιάζουσες τιμές (eigen-decomposition ή singular-value-decomposition) για να εξαφανίσουν την επιρροή της διασποράς του i.i.d. θορύβου στα διανύσματα κίνησης. Η προτεινόμενη μέθοδος είναι ανώτερη σ' αυτό το σημείο, αφού καταφέρνει να αυξήσει το  $SNR$  στο διαφορικό πεδίο κίνησης με κατάλληλη επιλογή τριπλετών σημείων και ζευγών τριπλετών (βλ. Υποενότητα 7.6.2). Τέλος, ακόμη ένα σημείο που τονίζεται στο κεφάλαιο αυτό είναι ότι η επιτυχία των αλγορίθμων στην ερευνητική αυτή περιοχή στηρίζεται σε μεγάλο βαθμό στη σύγκλιση. Γι' αυτό το σκοπό ολόκληρα πεδία κίνησης θα πρέπει να χρησιμοποιούνται αντί μεμονωμένων αντιστοιχίσεων χαρακτηριστικών.

Σχετικά και με τα παραπάνω, η Υποενότητα 7.6.4, στα επόμενα, αναδεικνύει τις ιδιότητες αυτές της προτεινόμενης μεθόδου βάσει συγκριτικών πειραμάτων.

## 7.6 Πειράματα και συζήτηση

Τα πειραματικά αποτελέσματα που παρουσιάζονται σε αυτήν την ενότητα αναδεικνύουν την ικανότητα της προτεινόμενης μεθόδου να ανακτήσει την πληροφορία δομής και κίνησης, ακόμη και παρουσία θορύβου στα πεδία διανυσμάτων κίνησης. Η ενότητα χωρίζεται σε τρεις υποενότητες.

Στην πρώτη υποενότητα, επιδεικνύουμε την προτεινόμενη διαδικασία υπολογισμών στην περίπτωση των απαλλαγμένων από θόρυβο πεδίων διανυσμάτων κίνησης. Δύο 3Δ δομές κατασκευασμένες μέσω H/Y και οι περιστραμμένες εκδοχές τους χρησιμοποιούνται ώστε να παραχθούν πεδία κίνησης, τα οποία εν συνεχεία τροφοδοτούνται στον αλγόριθμο.

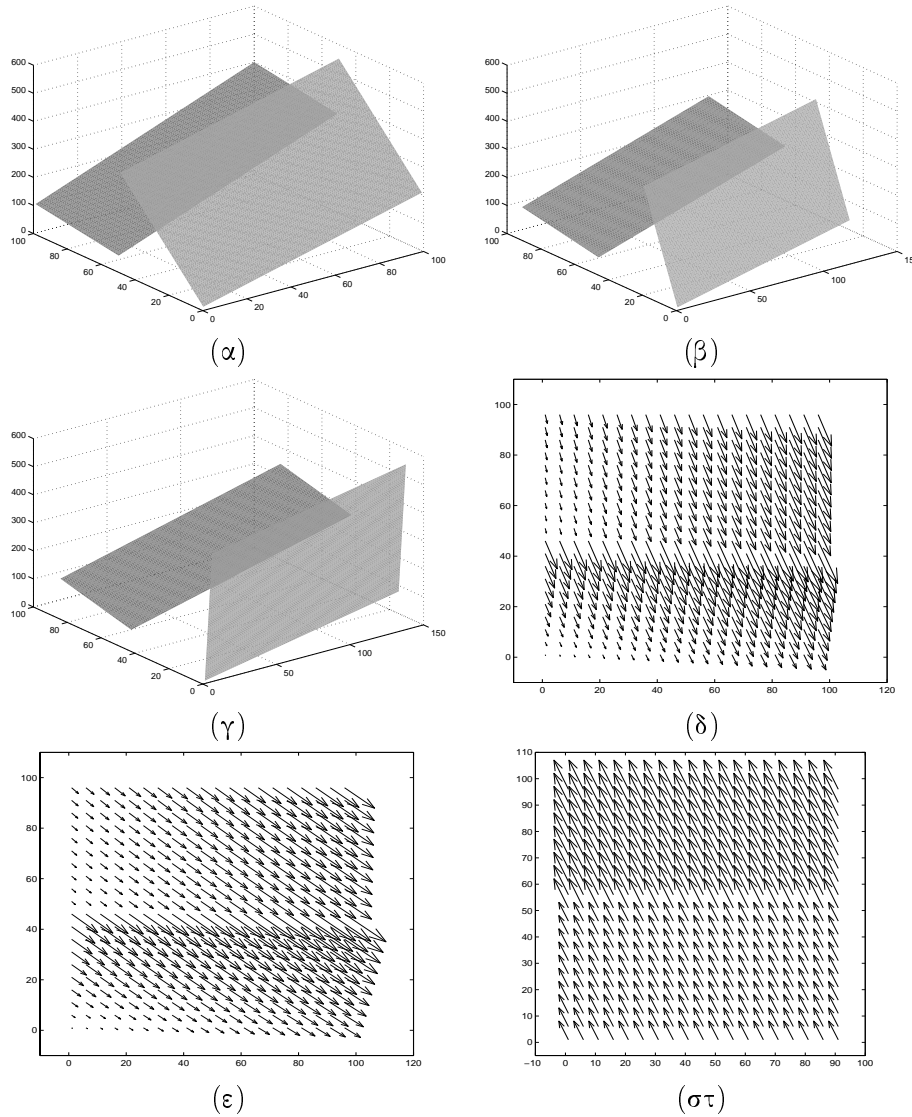
Η δεύτερη υποενότητα περιέχει πειραματικά αποτελέσματα τα οποία λαμβάνονται έπειτα από εισαγωγή στα πεδία κίνησης της προηγούμενης υποενότητας ψευδοτυχαίου προσθετικού θορύβου διαφόρων επιπέδων. Η σύγκλιση του προτεινόμενου αλγορίθμου στις πραγματικές παραμέτρους 3Δ κίνησης (γωνίες και άξονες περιστροφής) επαληθεύεται. Η μέση τιμή και η διασπορά των μετρήσεων που λαμβάνονται υπολογίζονται μέσω πειραμάτων Monte Carlo για διαφορετικό αριθμό διαθέσιμων διανυσμάτων θέσης και κίνησης και για διαφορετικό μέγεθος τριγώνων που χρησιμοποιούνται στην κατασκευή των πινάκων  $\mathbf{K}_j$ .

Η τρίτη υποενότητα εξετάζει την επίδοση του αλγορίθμου όταν αυτός εφαρμόζεται σε περιορισμένο αριθμό από διανύσματα κίνησης τα οποία προσδιορίζονται με το χέρι από μία ακολουθία εικόνων χαμηλής ανάλυσης.

### 7.6.1 Πεδία κίνησης χωρίς σφάλματα / θόρυβο

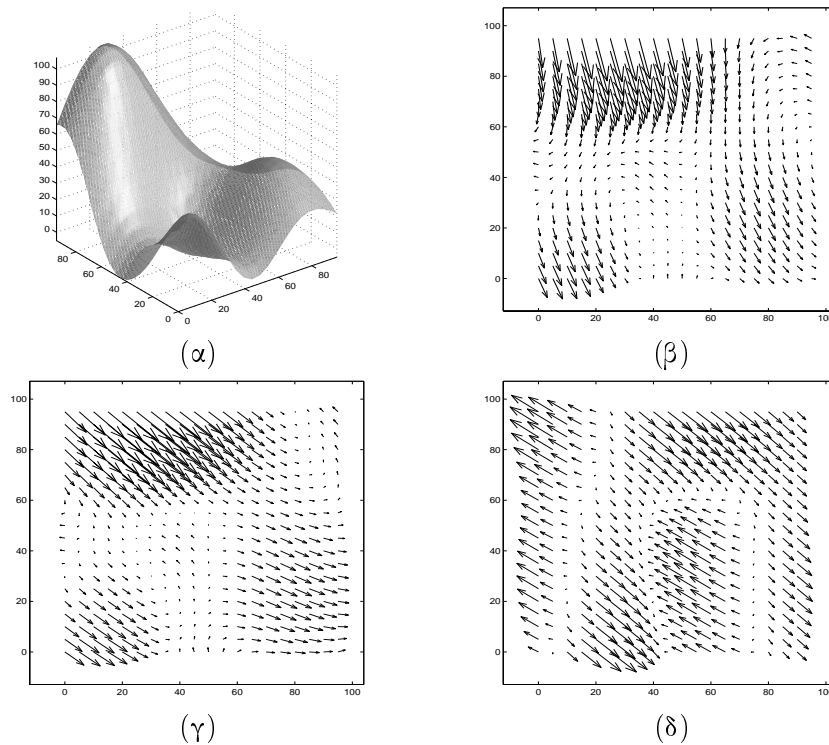
Σ' αυτήν την υποενότητα η επίδοση του αλγορίθμου εξετάζεται σε μη θορυβώδη περιβάλλοντα. Δύο παραδείγματα επιδεικνύονται. Στο (α) το διαθέσιμο αντικείμενο αποτελείται από δύο πολύγωνα, ενώ στο (β) το αντικείμενο είναι μία λεία επιφάνεια.

(α) Τα Σχήματα 7.2(α), 7.2(β) και 7.2(γ) παρουσιάζουν τρεις 3Δ σκηνές που περιέχουν ένα αντικείμενο αποτελούμενο από δύο επίπεδες επιφάνειες. Θεωρώντας ότι η θέση αναφοράς του αντικειμένου είναι αυτή του Σχήματος 7.2(α), το αντικείμενο περιστρέφει κατά  $\phi_R = 4^\circ$  και  $\phi_S = 7^\circ$  γύρω από άξονες  $\mathbf{u}_R = [.9129 \ .3651 \ .1826]^T$  και  $\mathbf{u}_S = [.6172 \ .7715 \ .1543]^T$ . Τα σχήματα 7.2(δ), 7.2(ε) παρουσιάζουν τα πεδία κίνησης που αντιστοιχούν σε κίνηση από το καρέ 1 στο 2 και από το καρέ 1 στο 3 αντίστοιχα. Στη συνέχεια θα χρησιμοποιήσουμε το συμβολισμό *κίνηση καρέ 1-2* και *κίνηση καρέ 1-3* αντίστοιχα. Στην πραγματικότητα, μόνο ένα ενδεικτικό υποσύνολο των διανυσμάτων κίνησης εμφανίζεται στα παραπάνω σχήματα.



**Σχήμα 7.2** Αντικείμενο που αποτελείται από δύο επίπεδες επιφάνειες (α) σκηνή 1, (β) σκηνή 2, (γ) σκηνή 3, (δ) πεδίο κίνησης (καρέ 1-2), (ε) πεδίο κίνησης (καρέ 1-3), (στ) διαφορικό πεδίο κίνησης (καρέ 1-2).

Το Σχήμα 7.2(στ) εμφανίζει το διαφορικό πεδίο κίνησης, το αντίστοιχο στην κίνηση καρέ 1-2. Αυτό το σχήμα περιέχει τα διαφορικά διανύσματα κίνησης σε μία συγκεκριμένη κατεύθυνση - σ' αυτό το παράδειγμα κάθε διαφορικό διάνυσμα κίνησης υπολογίζεται ως η διαφορά  $\Delta \mathbf{v}(i, j) = \mathbf{v}(i, j) - \mathbf{v}(i + 1, j)$ , όπου  $(i, j)$  οι συντεταγμένες ενός pixel της εικόνας. Το παράδειγμα επιβεβαιώνει την προαναφερθείσα ιδιότητα μεταξύ των διαφορικών διανυσμάτων κίνησης και θέσης στην περίπτωση των επίπεδων επιφανειών. Στην πραγματικότητα, όπως έχει ήδη δείχθει θεωρητικά, ο αλγόριθμος χρησιμοποιεί πίνακες  $\mathbf{L}_j$ , δηλαδή, ζεύγη από διαφορικά διανύσματα κίνησης σε διαφορετικές διευθύνσεις. Επιπρόσθετα, οι πίνακες  $\Delta \mathbf{L}_k$  θα ήταν καλύτερο εάν κατασκευάζονταν από ζεύγη πινάκων  $\mathbf{L}_j$  που να ανήκουν σε διαφορετικές επίπεδες επιφάνειες. Διαφορετικά,  $\Delta \mathbf{L}_k = \mathbf{0}$  και τέτοια ζεύγη  $\mathbf{L}_j$  πινάκων δεν συνεισφέρουν πρόσθετη



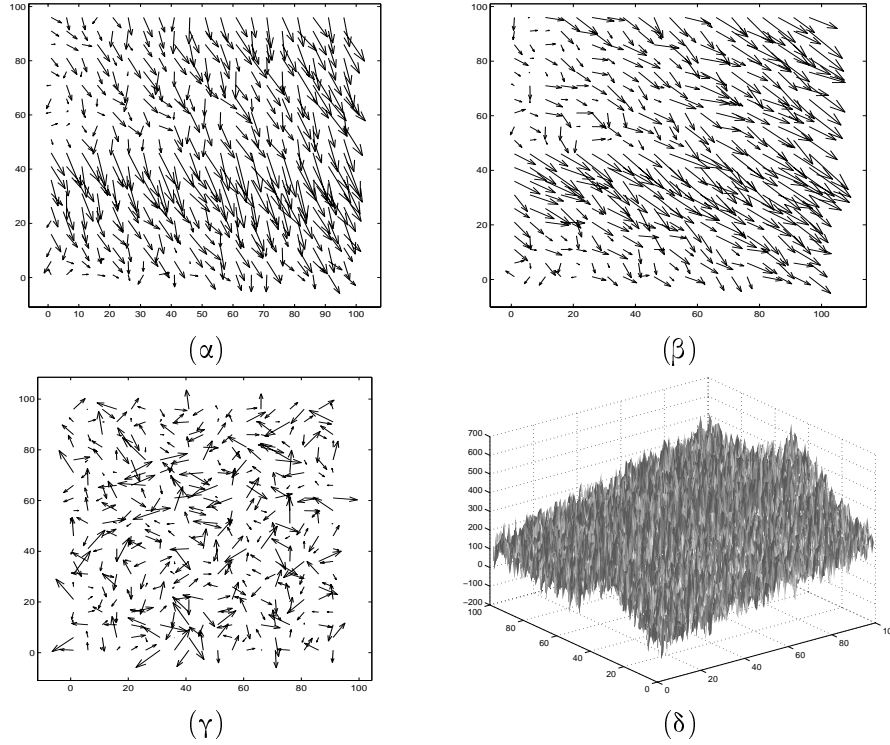
**Σχήμα 7.3** Λεία τρισδιάστατη επιφάνεια (α) σκηνή 1, (β) πεδίο κίνησης (καρέ 1-2), (γ) πεδίο κίνησης (καρέ 1-3), (δ) διαφορικό πεδίο κίνησης (καρέ 1-3).

πληροφορία. Σε συμφωνία με τα θεωρητικά αποτελέσματα των Ενοτήτων 7.3 και 7.4, σ' αυτό το παράδειγμα ο αλγόριθμος επιστρέφει ακριβείς εκτιμήσεις. Οι παράμετροι κίνησης υπολογίζονται να είναι πανομοιότυπες με αυτές που χρησιμοποιήθηκαν για την παραγωγή των τριών σκηνών και το ανακατασκευασμένο αντικείμενο είναι πανομοιότυπο με αυτό που εμφανίζεται στα Σχήματα 7.2(α), 7.2(β) και 7.2(γ). Γι' αυτό το λόγο τα σχετικά σχήματα παραλείπονται.

(β) Στο Σχήμα 7.3(α) εμφανίζεται μία λεία επιφάνεια, το δεύτερο 3Δ αντικείμενο που εξετάζεται, το οποίο περιστράφη κατά  $\phi_R = 4^\circ$  και  $\phi_S = 7^\circ$  γύρω από άξονες  $\mathbf{u}_R = [.9129 \ .3651 \ .1826]^T$  και  $\mathbf{u}_S = [.6172 \ .7715 \ .1543]^T$ . Στα Σχήματα 7.3(β) και 7.3(γ) εμφανίζονται τα απαλλαγμένα από θόρυβο διανύσματα κίνησης, τα αντίστοιχα με τις κινήσεις καρέ 1-2 και καρέ 1-3 αντίστοιχα. Στο Σχήμα 7.3(δ), εμφανίζεται το διαφορικό πεδίο κίνησης, το αντίστοιχο σε κίνηση καρέ 1-3. Ο αλγόριθμος επιστρέφει ακριβείς μετρήσεις και το ανακατασκευασμένο αντικείμενο είναι πανομοιότυπο με αυτό που εμφανίζεται στο Σχήμα 7.3(α).

### 7.6.2 Θορυβώδη διανύσματα κίνησης

Σ' αυτήν την υποενότητα, τα πεδία κίνησης τα παραγόμενα στην προηγούμενη υποενότητα είναι προσβεβλημένα από i.i.d. θόρυβο. Οι περιπτώσεις (α) και (β) της υποενότητας



**Σχήμα 7.4** Υπολογισμός και ανακατασκευή παρουσία θορύβου για το αντικείμενο που αποτελείται από δύο επίπεδες επιφάνειες (α) 10db θορυβώδες πεδίο κίνησης (καρέ 1-2), (β) 10db θορυβώδες πεδίο κίνησης (καρέ 1-3), (γ) αντίστοιχο θορυβώδες διαφορικό πεδίο κίνησης (καρέ 1-2), (δ) ανακατασκευασμένη σκηνή 1.

αντιστοιχούν στις αντίστοιχες της Υποενότητας 7.6.1.

Στα επόμενα πειράματα, στα πεδία κίνησης προστέθηκε 10db ομοιόμορφα κατανεμημένος τυχαίος θορύβος. Το επίπεδο SNR ορίζεται ως  $SNR_v = 10\log_{10}(\frac{\sigma_v^2}{\sigma_e^2})$ , όπου  $\sigma_v^2 \triangleq \frac{1}{P} \sum_{i=1}^P \mathbf{v}_i^T \mathbf{v}_i$  και  $\sigma_e^2 \triangleq E\{\mathbf{e}_i^T \mathbf{e}_i\}$ . Στην πραγματικότητα, η επίδοση του προτεινόμενου αλγορίθμου επηρεάζεται απευθείας από το επίπεδο SNR του διαφορικού πεδίου κίνησης και όχι από το πραγματικό πεδίο κίνησης. Αυτό υποδεικνύει ότι  $SNR_{dv} = 10\log_{10}(\frac{\sigma_{dv}^2}{\sigma_{de}^2})$  όπου  $\sigma_{dv}^2 \triangleq \frac{1}{P} \sum_{i \neq j} (\mathbf{v}_i - \mathbf{v}_j)^T (\mathbf{v}_i - \mathbf{v}_j)$  και  $\sigma_{de}^2 \triangleq E\{(\mathbf{e}_i - \mathbf{e}_j)^T (\mathbf{e}_i - \mathbf{e}_j)\} = 2\sigma_e^2$ , το οποίο αντιστοιχεί σε ένα διαφορικό πεδίο κίνησης αναφοράς ληφθέν για μέγεθος βήματος  $x=1$  (για παράδειγμα  $\Delta \mathbf{v}(n, m) = \mathbf{v}(n, m) - \mathbf{v}(n+1, m)$ , όπου  $(n, m)$  οι συντεταγμένες ενός pixel στην εικόνα). Όπως θα φανεί και στη συνέχεια, βελτιωμένο επίπεδο SNR επιτυγχάνεται για τις ίδιες μετρήσεις του διαφορικού πεδίου κίνησης αυξάνοντας το μέγεθος βήματος  $x$ .

(α) Θεωρούμε ξανά το αντικείμενο που αποτελείται από τις δύο επίπεδες επιφάνειες του Σχήματος 7.2(α). Υποθέτουμε ότι το αντικείμενο περιστράφη αρχικά σύμφωνα με τις ίδιες γωνίες και άξονες. Τα (γνωστά) προβεβλημένα πεδία κίνησης προσεβλήθησαν από 10db i.i.d. θόρυβο. Στα Σχήματα 7.4(α), 7.4(β) εμφανίζονται τα θορυβώδη πεδία κίνησης (τα σχήματα αυτά πρέπει να συγκριθούν απευθείας με τα 7.2(β), 7.2(γ)). Το θορυβώδες διαφορικό πεδίο

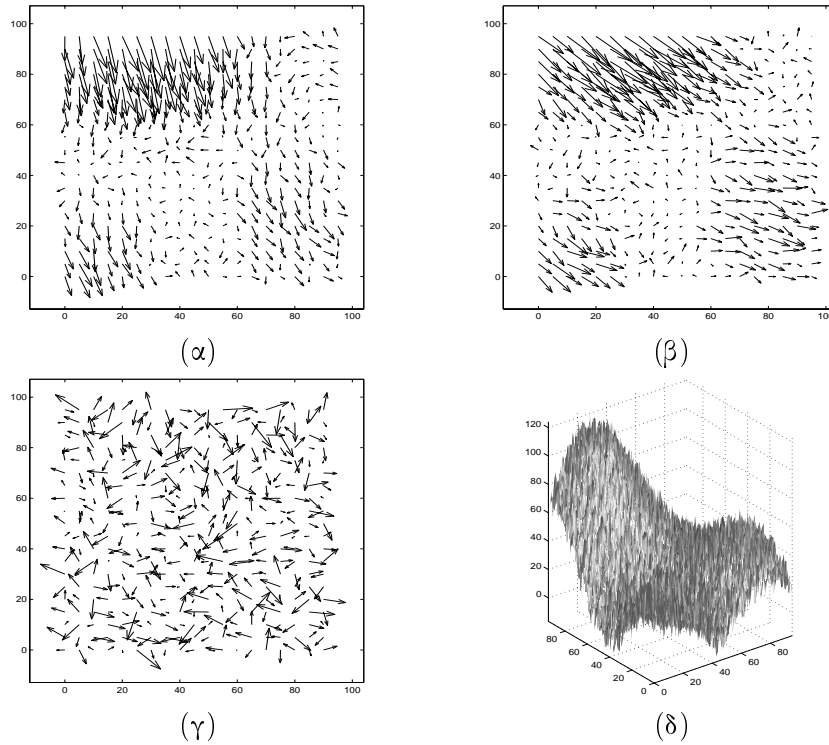
κίνησης που αντιστοιχεί στην κίνηση καρέ 1-2 εμφανίζεται στο 7.4(γ) και πρέπει να συγκριθεί με το 7.2(δ). Σε συμφωνία με τη θεωρητική μας ανάλυση, για την παραγωγή πινάκων  $\mathbf{L}_j$  από τριπλέτες σημείων, λαμβάνουμε υπόψη (ι) ότι κάθε σημείο πρέπει να συνεισφέρει στην παραγωγή μόνο ενός  $\mathbf{L}_j$ , (ii) όσο μεγαλύτερη είναι η επιφάνεια του τριγώνου που ορίζεται από μία τριπλέτα σημείων, τόσο μειώνεται το σφάλμα που σχετίζεται με τον αντίστοιχο  $\mathbf{L}_j$ . Σ' αυτό το παράδειγμα, χρησιμοποιήσαμε ορθογώνια ισοσκελή τρίγωνα με μήκος πλευράς  $x = 30$  pixels. Επιπρόσθετα, ανακαλούμε ότι τα ζεύγη  $\mathbf{L}_j$  πινάκων που ανήκουν στο ίδιο επίπεδο οδηγούν σε  $\Delta \mathbf{L}_k = \mathbf{0}$  και έτσι δε συνεισφέρουν πρόσθετη πληροφορία. Απουσία θορύβου, το παραπάνω δεν αποτελεί σημαντικό πρόβλημα. Αντιθέτως, παρουσία θορύβου, τέτοια ζεύγη συνεισφέρουν στους παράγοντες θορύβου χωρίς να παρέχουν χρήσιμη πληροφορία, οπότε θα πρέπει να αποφεύγονται. Ανακαλούμε επίσης ότι κάθε πίνακας  $\Delta \mathbf{L}_k$  πρέπει να παράγεται βάσει διαφορετικών πινάκων  $\mathbf{L}_j$ . Οι υπολογιζόμενοι άξονες και γωνίες περιστροφής παρουσιάζονται στον Πίνακα 7.3 (3η στήλη). Το ανακατασκευασμένο αντικείμενο εμφανίζεται στο Σχήμα 7.4(δ) το οποίο πρέπει να συγκριθεί απευθείας με το Σχήμα 7.2(α).

Το ανακατασκευασμένο αντικείμενο φαίνεται σημαντικά παραμορφωμένο σε σχέση με το πρωτότυπο. Αυτό οφείλεται στο θόρυβο που περιλαμβάνεται στα διανύσματα κίνησης. Διάφορες τεχνικές μπορούν να υιοθετηθούν για τη λείανση (smoothing) της ανακατασκευασμένης επιφάνειας. Περιορισμοί (constraints) και/ή πρότερη γνώση της τοπικής μεταβλητότητας της 3Δ επιφάνειας μπορούν να είναι χρήσιμα γι' αυτό το σκοπό.

(β) Η λεία επιφάνεια του Σχήματος 7.3(α) χρησιμοποιήθηκε εδώ και περιστράφη με την ίδια γωνία και γύρω από τον ίδιο άξονα με το προηγούμενο παράδειγμα. Τα (γνωστά) προβλεπόμενα πεδία κίνησης προσεβλήθησαν από 10db i.i.d. θόρυβο. Στα Σχήματα 7.5(α), 7.5(β) εμφανίζονται τα θορυβώδη πεδία κίνησης (τα σχήματα αυτά πρέπει να συγκριθούν με τα 7.3(β), 7.3(γ)). Το θορυβώδες πεδίο κίνησης που αντιστοιχεί στην κίνηση καρέ 1-3 εμφανίζεται στο 7.5(γ) και πρέπει να συγκριθεί με το 7.3(δ). Στην παραγωγή των πινάκων  $\mathbf{L}_j$  και  $\Delta \mathbf{L}_k$ , λαμβάνουμε υπόψη ότι αναφέρθηκε στο μέρος (α). Στο παράδειγμα αυτό, χρησιμοποιήσαμε πάλι ορθογώνια ισοσκελή τρίγωνα με μήκος πλευράς  $x = 20$  pixels. Η επιλογή των ζευγών πινά-

**Πίνακας 7.3** Υπολογισμός με θορυβώδη πεδία κίνησης (βλ. και Σχήματα 7.4,7.5).

	πραγματικές τιμές	2 επίπεδα	λεία επιφάνεια
$w$	1.7557	1.7768	1.7450
$\phi_R$	4°	4.2°	3.7°
$\phi_S$	7°	7.4°	6.96°
$\mathbf{u}_R$	$[\ .9129 \ .3651 \ .1826 ]^T$	$[\ .9157 \ .3645 \ .1691 ]^T$	$[\ .9179 \ .3351 \ .2125 ]^T$
$\mathbf{u}_S$	$[\ .6172 \ .7715 \ .1543 ]^T$	$[\ .6162 \ .7765 \ .1319 ]^T$	$[\ .6382 \ .7557 \ .1471 ]^T$



**Σχήμα 7.5** Υπολογισμός και ανακατασκευή παρουσία θορύβου για τη λεία επιφάνεια, (α) 10db θορυβώδες πεδίο κίνησης (καρέ 1-2), (β) 10db θορυβώδες πεδίο κίνησης (καρέ 1-3), (γ) αντίστοιχο θορυβώδες διαφορικό πεδίο κίνησης (καρέ 1-3), (δ) ανακατασκευασμένη σκηνή 1.

κων  $\mathbf{L}_j$  εδώ είναι αρκετά απλούστερη διαδικασία, αφού για μία λεία επιφάνεια, δύο τριπλέτες σημείων σπανίως ανήκουν στην ίδια επιφάνεια.

Οι υπολογισθείσες γωνίες και άξονες περιστροφής εμφανίζονται στον Πίνακα 7.3 (4η στήλη). Το ανακατασκευασμένο αντικείμενο παρουσιάζεται στο Σχήμα 7.5(δ) το οποίο θα πρέπει να συγκριθεί με το 7.3(α).

**Οδηγίες για τις επιλογές από το χρήστη.** Στην παράγραφο αυτή, τονίζονται μερικά σημαντικά ζητήματα στην επίδοση του αλγορίθμου, ώστε να χρησιμοποιηθούν ως οδηγίες για την επιλογή των τριπλετών σημείων και των πινάκων  $\mathbf{L}_j$  και  $\Delta\mathbf{L}_k$ .

(1) Όπως προτείνεται στην Ενότητα 7.4, είναι αναγκαίο οι πίνακες  $\mathbf{L}_j$  να παράγονται από διαφορετικές τριπλέτες σημείων. Επιπρόσθετα, μπορεί να δει κανείς ότι οι τριπλέτες σημείων πρέπει να επιλεγούν έτσι ώστε οι αντίστοιχες παράμετροι  $\pi_j^2$  να ελαχιστοποιούνται. Με αυ-

**Πίνακας 7.4** Μείωση της ποσότητας  $\sigma_k^2$  με το βήμα  $x$ .

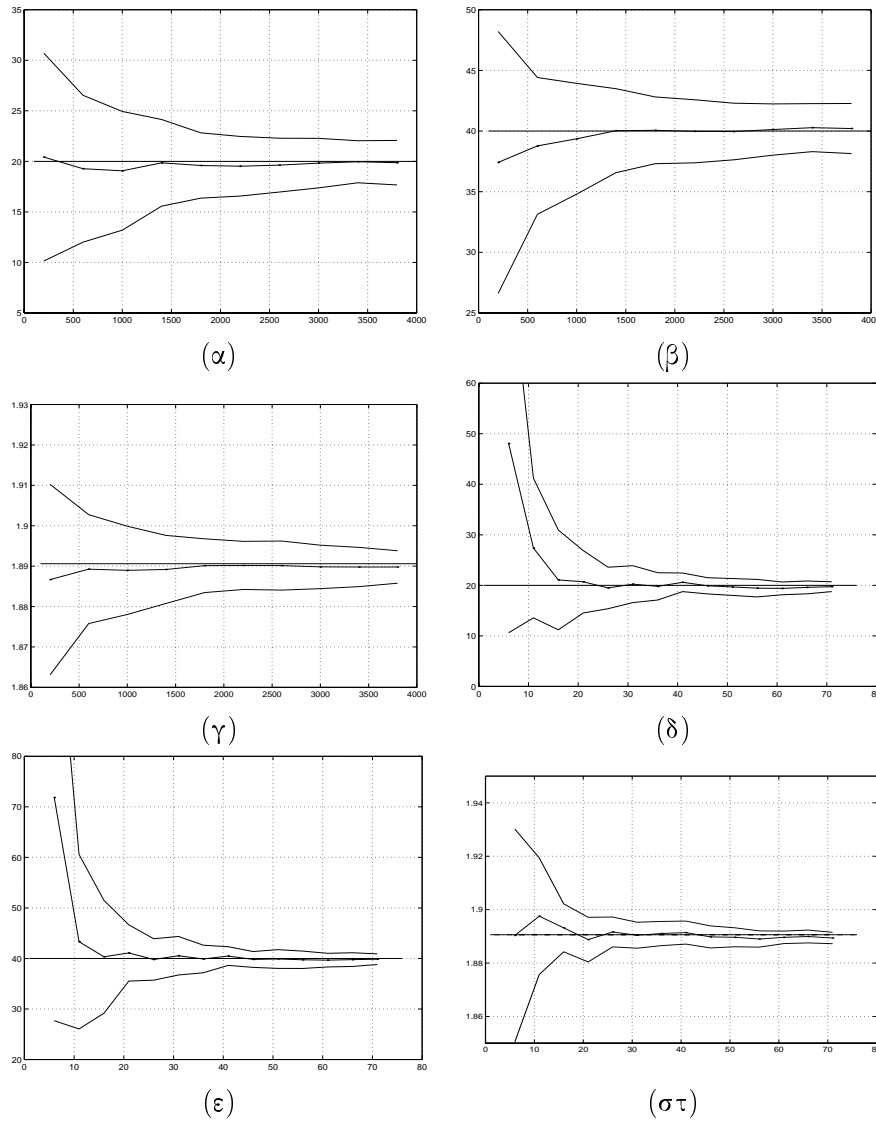
$x$	1	2	3	4	5	10	20	50	100
$\sigma_k^2/\sigma_e^2$	8	2	.8889	.5	.32	.08	.02	.0032	.0008



τόν τον τρόπο, όπως φαίνεται στις αποδείξεις των Λημμάτων 1 και 2, για περιορισμένο αριθμό σημείων  $P$ , οι όροι σφάλματος που περιλαμβάνονται στις ενδιάμεσες ποσότητες ελαχιστοποιούνται και η σύγκλιση των εκτιμητριών διασφαλίζεται. Ειδικότερα, μπορεί ναδειχθεί ότι τόσο οι διαγώνιοι όσο και αντιδιαγώνιοι παράγοντες σφάλματος του πίνακα  $\hat{\mathbf{Y}}^{(M)}$  εξαφανίζονται για αυξανόμενα  $\pi_j^2$ . Ελαχιστοποίηση του  $\pi_j^2$  (και συνεπώς του  $\sigma_k^2$ ) για δεδομένο  $\sigma_e^2$  επιγχάνεται μεγιστοποιώντας τα διαφορικά διανύσματα θέσης και έτσι αυξάνοντας το εμβαδό των τριγώνων που ορίζονται από τις τριπλέτες σημείων που χρησιμοποιούνται. Υποθέτοντας ότι οι τριπλέτες σημείων επιλέγονται σε ορθογώνια τρίγωνα και ότι όλα τα διαφορικά διανύσματα θέσης είναι ίσα με ένα βήμα  $x$ , ο Πίνακας 7.4 επιδεικνύει την ελάττωση του  $\sigma_k$  για αυξανόμενο βήμα  $x$ . Είναι ενδιαφέρον να παρατηρήσει κανείς ότι όλοι οι παράγοντες σφάλματος στον  $\hat{\mathbf{Y}}^{(M)}$  πολλαπλασιάζονται με αυτόν τον παράγοντα. Σ' αυτό το σημείο, η ερώτηση είναι εάν τα διαφορικά διανύσματα κίνησης διατηρούν το μέγεθός τους, δηλαδή εάν η απαλλαγμένη από σφάλματα συνιστώσα  $\mathbf{Y}^{(M)}$  του  $\hat{\mathbf{Y}}^{(M)}$  διατηρεί την ισχύ της. Στην περίπτωση των επίπεδων επιφανειών, μπορεί ναδειχθεί ότι για τριπλέτες σημείων στο ίδιο επίπεδο, αύξηση στα διαφορικά διανύσματα θέσεως οδηγεί σε ανάλογη αύξηση τα διαφορικά διανύσματα κίνησης, δηλαδή οι πίνακες  $\mathbf{L}_j$  και συνεπώς η συνιστώσα  $\mathbf{Y}^{(M)}$  διατηρούν την ισχύ τους. Με αυτόν τον τρόπο, το SNR στον πίνακα  $\hat{\mathbf{Y}}^{(M)}$  αυξάνεται δραματικά. Στην γενικότερη περίπτωση, μπορούμε να δούμε ότι για αυξανόμενα διαφορικά διανύσματα θέσεως, δεν μπορεί να προκαθοριστεί εάν τα διαφορικά διανύσματα θέσεως κερδίζουν σε ισχύ ή όχι. Με άλλα λόγια, δεν μπορεί γενικά να διατυπωθεί ότι η προτεινόμενη στρατηγική αυξάνει/μειώνει το SNR στον πίνακα  $\hat{\mathbf{Y}}^{(M)}$ .

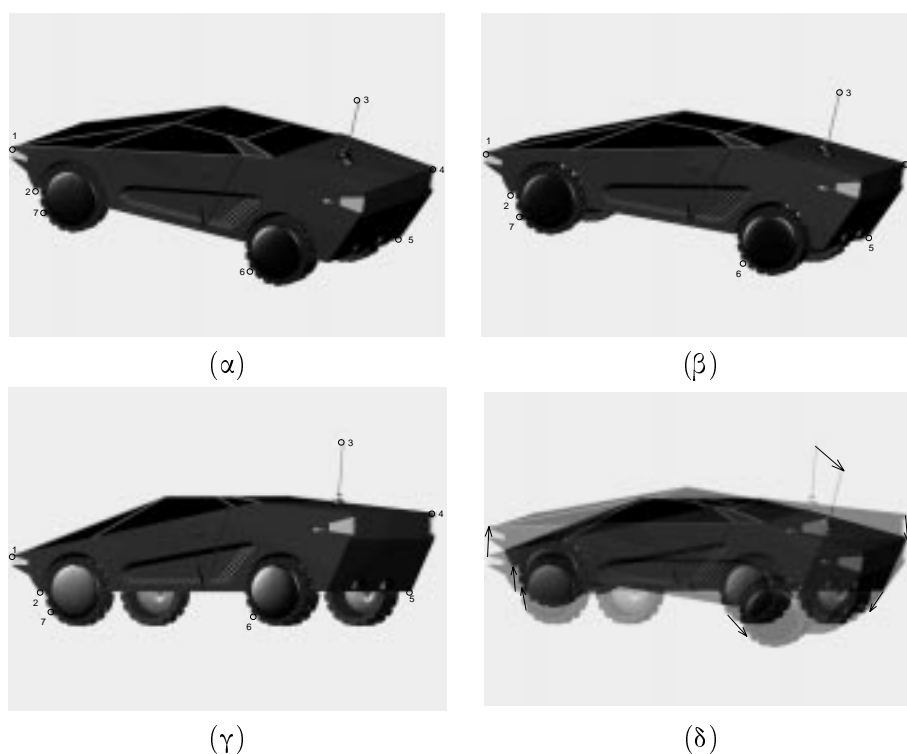
Στο Σχήμα 7.6, η αποδοτικότητα των παραπάνω ισχυρισμών παρουσιάζεται για το αντικείμενο που αποτελείται από 2 επίπεδες επιφάνειες, το οποίο περιστρέφεται κατά γωνίες  $\phi_R = 20^\circ$  και  $\phi_S = 40^\circ$  με άξονες  $\mathbf{u}_R = [.9129 \ .3651 \ .1826]^T$  και  $\mathbf{u}_S = [.6172 \ .7715 \ .1543]^T$ . Στα Σχήματα 7.6(δ), 7.6(ε) δίνονται μετρήσεις της μέσης τιμής (συνεχής γραμμή) και της τυπικής απόκλισης (διακεκομμένη γραμμή) των γωνιών περιστροφής  $\phi_R, \phi_S$  για 50 Monte Carlo πειράματα ως προς αυξανόμενα διαφορικά διανύσματα θέσης. Στο Σχήμα 7.6(στ) δίνονται μετρήσεις του ενδιάμεσου άγνωστου  $w$  στην περίπτωση αυτή.

(2) Για ελαττούμενο  $\pi_j^2$ , βλέπουμε από την εξίσωση (7-30) ότι το  $\sigma_k^2$  ελαττώνεται και έτσι η μέθοδος που περιγράφεται στο (1) αυξάνει το SNR στους πίνακες  $\Delta\mathbf{L}_k$ , δεδομένου ότι παράγονται βάσει διαφορετικών πινάκων  $\mathbf{L}_j$ . Παρόλ'αυτά, όπως αναφέρθηκε παραπάνω, είναι αναγκαίο οι πίνακες  $\Delta\mathbf{L}_k$  να ορίζονται από διαφορετικούς πίνακες  $\mathbf{L}_j$  που να αντιστοιχούν και σε διαφορετικά επίπεδα, αλλιώς οι όροι τους με ωφέλιμο σήμα είναι μηδενικοί. Αυτό δεν θέτει σημαντικό πρόβλημα στην περίπτωση των τυχαίων επιφανειών, όμως πρέπει να ληφθεί υπόψη για επίπεδα αντικείμενα.



**Σχήμα 7.6** Υπολογισμός παραμέτρων περιστροφής για το αντικείμενο που αποτελείται από δύο επίπεδες επιφάνειες,  $(\alpha, \beta, \gamma)$  μετρήσεις των  $\phi_R, \phi_S, w$  ως προς τον αριθμό των δειγμάτων,  $(\delta, \epsilon, \sigma\tau)$  μετρήσεις των  $\phi_R, \phi_S, w$  ως προς το βήμα  $x$ .

(3) Όπως φαίνεται και στην απόδειξη του Λήμματος 7.2, οι χρησιμοποιούμενες εκτιμήτριες συγκλίνουν στις πραγματικές τους τιμές για υψηλό αριθμό σημείων  $P$ . Αν και κατάλληλη επιλογή των τριπλετών σε συμφωνία με το (1) βελτιώνει το ρυθμό σύγκλισης, είναι ακόμη αναγκαίο να χρησιμοποιούμε το μέγιστο δυνατό αριθμό από πίνακες  $\mathbf{L}_j$ . Στα Σχήματα 7.6(α), 7.6(β) δίνονται μετρήσεις των γωνιών  $\phi_R, \phi_S$  για αυξανόμενο αριθμό χρησιμοποιούμενων πινάκων. Στο Σχήμα 7.6(γ) μετρήσεις του ενδιάμεσου αγνώστου δίνονται επίσης. Για το παράδειγμα αυτό, οι τριπλέτες σημείων διαλέγονται σε ισοσκελή ορθογώνια τρίγωνα πλευράς  $x = 30$ .



**Σχήμα 7.7** Υπολογισμός παραμέτρων κίνησης ενός οχήματος διαθέσιμου σε 3 καρέ χρησιμοποιώντας ανάθεση των χαρακτηριστικών σημείων με το χέρι (α) χαρακτηριστικά σημεία (καρέ 1), (β) χαρακτηριστικά σημεία (καρέ 2), (γ) χαρακτηριστικά σημεία (καρέ 3), (δ) αντιστοιχίσεις σημείων (καρέ 1-2).

### 7.6.3 Θορυβώδεις αντιστοιχίσεις σημείων

Σε αυτήν την υποενότητα, δεδομένων τριών καρέ μίας συνθετικής ακολουθίας βίντεο, ένας αριθμός από αντιστοιχίσεις σημείων λαμβάνεται με ευρεστικό τρόπο (ή ακόμη σημειώνονται και με το χέρι). Αυτές οι αντιστοιχίσεις σημείων τροφοδοτούνται στον προτεινόμενο αλγόριθμο και οι υπολογιζόμενες παράμετροι κίνησης συγκρίνονται με τις πραγματικές. Σημαντικό σφάλμα εισάγεται στις μετρήσεις λόγω της χαμηλής ανάλυσης των καρέ που χρησιμοποιούνται. Επιπρόσθετα, εφόσον χρησιμοποιείται μικρός αριθμός διανυσμάτων κίνησης, η ασυμπτωτική συμπεριφορά της προτεινόμενης διαδικασίας δεν επιτυγχάνεται. Παρόλ'αυτά, ο αλγόριθμος εμφάνισε καλή συμπεριφορά και σε αυτήν την περίπτωση.

Στα Σχήματα 7.7(α), 7.7(β) και 7.7(γ) εμφανίζονται τα τρία εμπλεκόμενα καρέ. Στα σχήματα αυτά εμφανίζονται επίσης τα χαρακτηριστικά σημεία που χρησιμοποιήθηκαν. Τα καρέ παράχθηκαν συνθετικά με κίνηση και απόδοση (rendering) του 3Δ συνθετικού μοντέλου "Future Car" χρησιμοποιώντας ένα εμπορικό λογισμικό πακέτο. Στο Σχήμα 7.7(δ) εμφανίζονται οι αντιστοιχίσεις σημείων για κίνηση καρέ 1-2. Βάσει των διαθέσιμων αντιστοιχίσεων σημείων, οι παράμετροι κίνησης υπολογίζονται στη συνέχεια. Οι πίνακες  $\mathbf{L}$  παράγονται σύμφωνα με

τους παραπάνω ισχυρισμούς, δηλαδή οι αντιστοιχίσεις σημείων για κάθε  $\mathbf{L}_j$  επιλέγονται σχετικώς μακριά μεταξύ τους. Πρόσθετη φροντίδα λαμβάνεται ώστε οι παραγόμενοι πίνακες  $\mathbf{L}$  να αντιστοιχούν σε διαφορετικά 3Δ επίπεδα.

Ο Πίνακας 7.5 περιέχει τις πραγματικές και υπολογιζόμενες παραμέτρους κίνησης.

**Πίνακας 7.5** Υπολογισμός με θορυβώδεις αντιστοιχίσεις σημείων (βλ. Σχήμα 7.7).

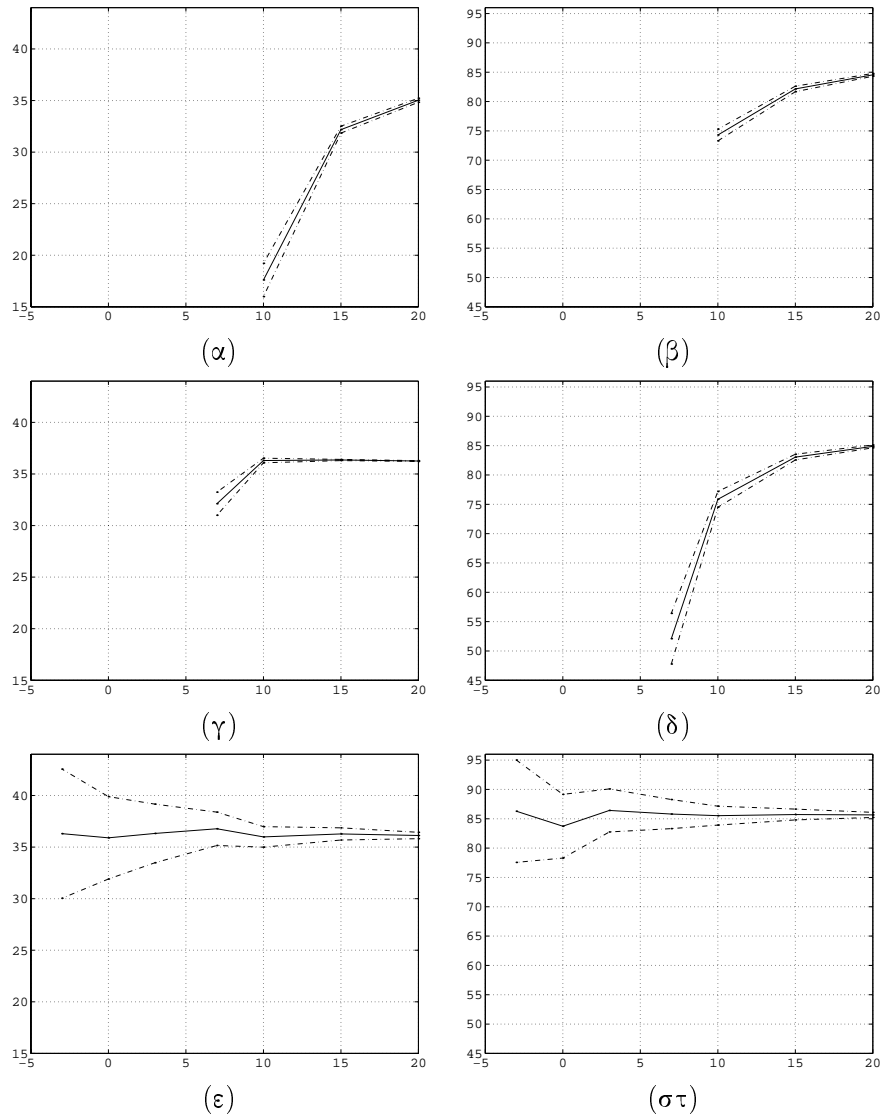
	πραγματικές τιμές	υπολογιζόμενες τιμές
$\phi_R$	$18.3^\circ$	$18.1^\circ$
$\phi_S$	$16.5^\circ$	$15.9^\circ$
$\mathbf{u}_R$	$[\ .1991 \ .5775 \ .7918 ]^T$	$[\ .1961 \ .5883 \ .7845 ]^T$
$\mathbf{u}_S$	$[\ .5005 \ .8125 \ .2990 ]^T$	$[\ .4770 \ .8296 \ .2930 ]^T$

#### 7.6.4 Σύγκριση με υπάρχουσες μεθόδους

Σ'αυτήν την τελευταία υποενότητα, ο προτεινόμενος αλγόριθμος αντιπαραβάλλεται σε υπάρχουσες προσεγγίσεις για την εκτίμηση της 3Δ δομής και κίνησης στην παράλληλη προβολή. Γι'αυτό το σκοπό, μία μέθοδος factorization και μία μέθοδος που χρησιμοποιεί τον epipolar περιορισμό υλοποιήθηκαν και εξετάστηκαν ενάντια στην προτεινόμενη προσέγγιση. Η μέθοδος factorization υλοποιήθηκε σύμφωνα με τον αλγόριθμο των Tomasi και Kanade όπως περιγράφεται στο [159]. Για την epipolar μέθοδο, ο αλγόριθμος που προτείνεται στην [196] χρησιμοποιήθηκε. Ο τελευταίος χρησιμοποιεί ως πρώτο βήμα τη λύση της epipolar εξίσωσης όπως δίνεται από τους Shapiro *et al.* στο [140]. Και οι δύο αλγόριθμοι υλοποιήθηκαν αυστηρά σύμφωνα με τις οδηγίες των αντίστοιχων συγγραφέων.

Η προτεινόμενη προσέγγιση έδειξε ανώτερη σε σχεδόν όλα τα πειράματα που διεξήχθησαν. Στα Σχήματα 7.8 και 7.9, μετρήσεις των γωνιών περιστροφής  $\phi_R$  και  $\phi_S$  για διάφορα επίπεδα SNR εμφανίζονται, για το αντικείμενο που αποτελείται από δύο επίπεδες επιφάνειες και για τη λεία επιφάνεια της Υποενότητας 7.6.1. Κάθε αντικείμενο περιστράφηκε για να παράγει δύο σύνολα από πεδία κίνησης, τα οποία στη συνέχεια προσβλήθηκαν από i.i.d. θόρυβο, όμοια με την Υποενότητα 7.6.2. Το επίπεδο SNR καθορίστηκε σε συμφωνία με τα διαφορετικά πεδία κίνησης, όπως εξηγείται στην ίδια υποενότητα.

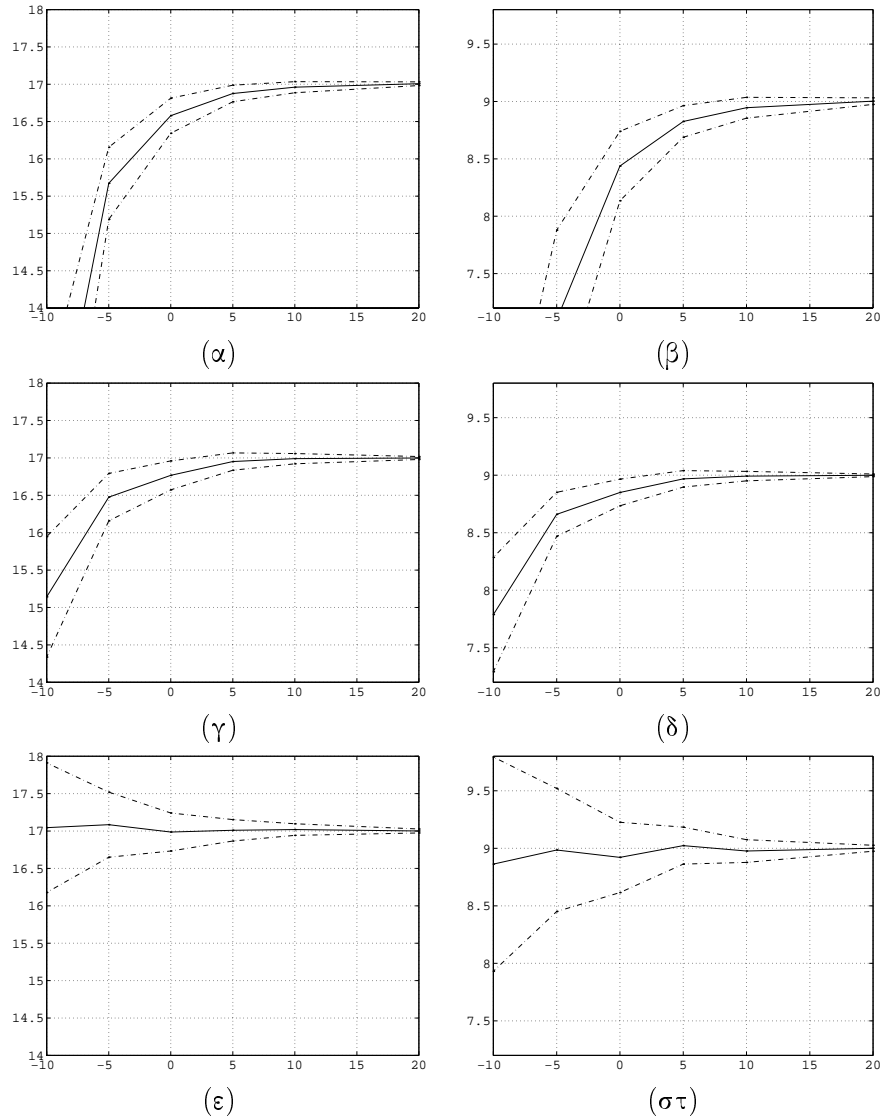
Στα Σχήματα 7.8(α,β), 7.8(γ,δ) και 7.8(ε,στ), μετρήσεις της μέσης τιμής (συνεχής γραμμή) και της τυπικής απόκλισης (διακεκομμένη γραμμή) των γωνιών περιστροφής  $\phi_R$ ,  $\phi_S$  εμφανίζονται για το αντικείμενο των δύο επιπέδων, για τη factorization, την epipolar και την προτεινόμενη μέθοδο. Οι μετρήσεις λήφθηκαν μέσω 50 πειραμάτων Monte Carlo για κάθε επίπεδο SNR. Σε κάθε πείραμα, το ίδιο σύνολο από θορυβώδεις αντιστοιχίσεις σημείων τροφοδοτήθηκε και στους τρεις αλγόριθμους. Για το παράδειγμά μας, αρκετές διαφορετικές γωνίες και άξονες



**Σχήμα 7.8** Μετρήσεις για διαφορετικό επίπεδο SNR στα διαφορεικά πεδία κίνησης, για το αντικείμενο με τις δύο επίπεδες επιφάνειες, (α,γ,ε) και (β,δ,στ)  $\phi_R$  και  $\phi_S$  χρησιμοποιώντας την factorization, την epipolar και την προτεινόμενη μέθοδο αντίστοιχα.

εξετάστηκαν. Οι πραγματικές γωνίες περιστροφής για τα Σχήματα 7.8(α,γ,ε) και 7.8(β,δ,στ) ήταν  $\phi_R = 36.19^\circ$  και  $\phi_S = 85.64^\circ$  αντιστοιχώντας σε άξονες  $\mathbf{u}_R = [.7082 \ .3963 \ .5843]^T$  και  $\mathbf{u}_S = [.7713 \ -.6358 \ -.0291]^T$ .

Στα Σχήματα 7.9(α,β), 7.9(γ,δ) και 7.9(ε,στ), μετρήσεις της μέσης τιμής (συνεχής γραμμή) και της τυπικής απόκλισης (διακεκομμένη γραμμή) των γωνιών περιστροφής  $\phi_R$ ,  $\phi_S$  εμφανίζονται για τη λεία επιφάνεια χρησιμοποιώντας την factorization, την epipolar και την προτεινόμενη μέθοδο αντίστοιχα. Οι πραγματικές γωνίες περιστροφής στα Σχήματα 7.9(α,γ,ε) και 7.9(β,δ,στ) ήταν  $\phi_R = 17^\circ$  και  $\phi_S = 9^\circ$  αντιστοιχώντας σε άξονες  $\mathbf{u}_R = [.3518 \ -.6697 \ .6540]^T$  και  $\mathbf{u}_S = [-.7208 \ -.3716 \ -.5850]^T$ .



**Σχήμα 7.9** Μετρήσεις για διαφορετικό επίπεδο SNR στα διαφορεικά πεδία κίνησης, για τη λεία επιφάνεια, (α,γ,ε) και (β,δ,στ)  $\phi_R$  και  $\phi_S$  χρησιμοποιώντας την factorization, την epipolar και την προτεινόμενη μέθοδο αντίστοιχα.

Όπως μπορεί να δειχθεί και στα δύο Σχήματα 7.8 και 7.9, η προτεινόμενη προσέγγιση είναι ανώτερη και από την factorization και από την epipolar μέθοδο. Περισσότερο, στην περίπτωση του αντικειμένου που αποτελείται από επίπεδες επιφάνειες, οι άλλες προσεγγίσεις αποτυγχάνουν όσο το επίπεδο SNR προσεγγίζει τα 10dB (απουσία μετρήσεων μέσης τιμής και απόκλισης υποδεικνύει παρουσία σημαντικών φανταστικών μερών στις υπολογιζόμενες γωνίες).

Μπορεί παρόλ'αυτά να σημειωθεί ότι και η factorization και η epipolar μέθοδος επιδεικνύουν μικρότερες μετρήσεις τυπικής απόκλισης συγκρινόμενες με την προτεινόμενη μέθοδο για υψηλότερα επίπεδα SNR (περισσότερο από 5dB στην περίπτωση της λείας επιφάνειας). Αυτό οφείλεται πιθανότατα στη στρατηγική που χρησιμοποιείται για την παραγωγή των τριπλε-

τών σημείων και των ζευγών τριπλετών. Στα συγκεκριμένα πειράματα ορθογώνια ισοσκελή τρίγωνα χρησιμοποιήθηκαν, ενώ τα ζεύγη τριπλετών κατασκευάστηκαν βάσει της μέγιστης ισχύος στους πίνακες  $\Delta \mathbf{L}_k$ . Μπορεί να επαληθευτεί ότι βελτιωμένες μετρήσεις τυπικής απόκλισης μπορούν να ληφθούν κατασκευάζοντας περισσότερες από  $N = N(P) \cong P/3$  triplets και επιλέγοντας ζεύγη τριπλετών που να μην περιέχουν το ίδιο σημείο δύο φορές.

## 7.7 Συμπεράσματα και επεκτάσεις

Στο κεφάλαιο αυτό παρουσιάσαμε μία νέα μέθοδο για τον υπολογισμό των  $3 \times 3$  πινάκων περιστροφής που χαρακτηρίζουν την κίνηση ενός 3Δ στερεού αντικειμένου από τις παράλληλες 2Δ προβολές του. Οι ληφθείσες εκφράσεις είναι απλές στον υπολογισμό τους μέσω της ανάλυσης σε ιδιοτιμές συμμετρικών  $2 \times 2$  πινάκων οι οποίοι περιέχουν τα αντίστοιχα διανύσματα κίνησης. Όταν τα διαθέσιμα διανύσματα κίνησης είναι απαλλαγμένα από θόρυβο, οι υπολογισμοί δίνουν τις ακριβείς τιμές για τις παραμέτρους κίνησης και σχήματος του αντικειμένου, ακόμη κι αν μόνο τέσσερα διανύσματα κίνησης ανά μετάβαση είναι διαθέσιμα - αυτό είναι συνεπές και με τα θεωρητικά αποτελέσματα του [61]. Οι Προτάσεις 7.3, 7.4 και 7.5 της Ενότητας 7.3 παρέχουν τις απαραίτητες αλγοριθμικές λεπτομέρειες. Πιθανότατα, η μεγαλύτερη συνεισφορά του κεφαλαίου είναι ότι μία μικρή τροποποίηση των Προτάσεων αυτών επιτρέπει να συμπεριληφθεί ένας τυχαία μεγάλος αριθμός από θορυβώδη διανύσματα κίνησης στους υπολογισμούς. Οι Προτάσεις 7.6, 7.7 και 7.8 τεκμηριώνουν ισχυρή σύγκλιση των τροποποιημένων εκφράσεων στις πραγματικές παραμέτρους περιστροφής, ακόμη και παρουσία θορύβου στα πεδία κίνησης. Η ασυμπτωτική συμπεριφορά των εκτιμήσεων παρουσία θορύβου εξετάστηκε με θεωρητικό και πειραματικό τρόπο. Πιο σημαντικό ακόμη είναι το γεγονός ότι η θεωρητική ανάλυση και τα πειράματα υποδεικνύουν ότι η ποιότητα των μετρήσεων μπορεί να βελτιωθεί εάν μία συγκεκριμένη στρατηγική υιοθετηθεί στη επιλογή των διανυσμάτων κίνησης που χρησιμοποιούνται στους εμπλεκόμενους πίνακες  $\hat{\mathbf{K}}_j$ . Τα αποτελέσματα που παρουσιάζονται στο κεφάλαιο αυτό δίνουν τη βάση για περαιτέρω βελτιώσεις κατά τους ακόλουθους τρόπους:

(α) Βελτίωση των εκτιμήσεων των  $\mathbf{R}$  και  $\mathbf{S}$  χρησιμοποιώντας έναν κατάλληλο μηχανισμό βαρών στις εμπλεκόμενες διαδικασίες ελαχίστων τετραγώνων. Τα βάρη μπορούν φυσικά να αντικατοπτρίζουν τη σχετική εμπιστοσύνη που έχουμε για τα διαθέσιμα διανύσματα κίνησης. Κάποια πρώτα αποτελέσματα σ' αυτό το θέμα μπορούν να βρεθούν στο [190].

(β) Βελτίωση του ανακατασκευασμένου 3Δ σχήματος (wireframe) το οποίο τείνει να είναι αρκετά παραμορφωμένο όταν ο θόρυβος στις μετρήσεις είναι χαμηλού SNR. Μία μέθοδος

λείανσης, πιθανότατα βασισμένη σε μερική a priori γνώση του αντικειμένου, μπορεί να είναι χρήσιμη σε αυτή την κατεύθυνση.

(γ) Επαναυπολογισμός των διανυσμάτων κίνησης χρησιμοποιώντας τον προτεινόμενο αλγόριθμο ως περιορισμό επιβαλλόμενο από τη στερεά φύση του θεωρούμενου  $3\Delta$  αντικειμένου.

(δ) Βελτίωση των μετρήσεων σχήματος (βάθους) μπορεί επίσης να επιτευχθεί μέσω χρονικών (σε επίπεδο καρτέ) επαναλήψεων όπως αναφέρεται για παράδειγμα στα [64, 19].



## ΚΕΦΑΛΑΙΟ 8

### Υπολογισμός Τρισδιάστατης Κίνησης της Κάμερας σε Εικονικά Περιβάλλοντα Καταγραφής

#### 8.1 Εισαγωγή / Αναδρομή

Η ταχεία ανάπτυξη της επιστήμης των υπολογιστών έχει θέλξει τα τελευταία χρόνια ένα μεγάλο μέρος της βιομηχανίας της ψυχαγωγίας (συμπεριλαμβανομένης της Τηλεόρασης και του Κινηματογράφου). Η πρόσφατη πρόοδος στους τομείς της επεξεργασίας εικόνων, της τεχνολογίας βίντεο και των γραφικών με υπολογιστές, όπως επίσης και η αυξανόμενη υπολογιστική ισχύς που παρέχεται από τα υπολογιστικά συστήματα, είναι αντικείμενο ευρείας εκμετάλλευσης για την εμπορική παραγωγή υλικού βίντεο. Τα virtual (εικονικά) studios χρησιμοποιούνται στην βιομηχανία βίντεο για ποικίλες παραγωγές, με αποτέλεσμα ποικίλα ενδιαφέροντα οπτικά εφέ, η πλέον συνήθης από τις οποίες είναι το δελτίο καιρού. Οι ακολουθίες βίντεο που παράγονται από τέτοια συστήματα είναι βασικά συνθέσεις φυσικών και συνθετικών ακολουθιών, οι οποίες είτε είναι ήδη αποθηκευμένες σε ψηφιακά μέσα είτε παράγονται σε πραγματικό χρόνο. Στη συνήθη περίπτωση, η ‘μικτή’ ακολουθία (ως αποτέλεσμα της σύνθεσης), η οποία αναφέρεται και ως virtual set (εικονικό σύνολο), αποτελείται από ένα φυσικό βίντεο πραγματικού χρόνου και μία αποθηκευμένη ακολουθία συνθετικής ή φυσικής οπτικής πληροφορίας [46]. Σ’ αυτό το πλαίσιο, αυξανόμενο ενδιαφέρον δίνεται σε τέτοια συστήματα, ιδιαίτερα αναφορικά με τις οδηγίες του προτύπου MPEG-4 σχετικά με την αντικειμενοστραφή και συνθετική-φυσική-υβριδική κωδικοποίηση.

Ένα σύστημα virtual studio αποτελείται βασικά από τρία μέρη, συγκεκριμένα τα υποσυστήματα παρακολούθησης της κάμερας (camera tracking module), απόδοσης (rendering module) και σύνθεσης (compositing module). Το τελευταίο παραδοσιακά περιλαμβάνει ένα κυανό πέτασμα (blue screen) στο φόντο, εμπροσθεν του οποίου εκτυλίσσεται η δράση του προσκηνίου, και μία τεχνική κατάτμησης χρώματος (chromakeying) για το διαχωρισμό του φόντου από το προσκηνίο. Η δράση που εκτυλίσσεται στο προσκηνίο συντίθεται τότε με τη διαθέσιμη ακολουθία φόντου (η οποία είναι συνήθως προ-αποθηκευμένη). Με αυτόν τον τρόπο, ο μετεωρολόγος μπορεί να εμφανίζεται στην τηλεόραση σαν να ευρίσκετο μπροστά από έναν μετεωρολογικό / γεωγραφικό χάρτη, ενώ στην πραγματικότητα στέκεται μπροστά από ένα κυανό παραπέτασμα. Πρακτικά, το υποσύστημα σύνθεσης, χρησιμοποιώντας ένα κατατμητή χρώματος (chromakeyer), αντικαθιστά τις περιοχές του μπλε χρώματος (το χρώμα-κλειδί μπορεί γενικά να είναι οποιοδήποτε, όμως η χρήση του μπλε είναι πλέον ενδεδειγμένη) στο βίντεο

πραγματικού χρόνου με το προ-αποθηκευμένο φόντο. Οι παραδοσιακές τεχνικές chromakeying έχουν διαφοροποιηθεί και επεκταθεί, ώστε να μειωθούν οι ατέλειες και να βελτιωθούν τα αποτελέσματα σύνθεσης (βλ. για παράδειγμα τα [53, 63]). Το υποσύστημα απόδοσης είναι υπεύθυνο για τη συμβατότητα του καταγεγραμμένου προσκηνίου με το φόντο. Το τελευταίο είναι κυρίως διαθέσιμο είτε ως ανεξάρτητη ακολουθία είτε ως ένας εικονικός κόσμος / μοντέλο. Στην πρώτη περίπτωση, η ακολουθία που αντιστοιχεί στο φόντο πρέπει να μετασχηματιστεί ώστε να είναι συμβατή με αυτή του προσκηνίου, ενώ στη δεύτερη περίπτωση, το εικονικό φόντο αποδίδεται μέσω μίας επίσης εικονικής κάμερας η κίνηση της οποίας συμφωνεί με την κίνηση της κάμερας του πραγματικού κόσμου. Το υποσύστημα παρακολούθησης της κάμερας είναι ουσιαστικά το πιο σημαντικό μέρος των συστημάτων virtual studio. Ο αποδοτικός υπολογισμός της κίνησης της κάμερας φροντίζει ώστε το βίντεο πραγματικού χρόνου και το διαθέσιμο φόντο να είναι οπτικά συμβατά. Μέχρι περίπου το 1990, το υποσύστημα αυτό απουσίαζε από τα virtual studios, περιορίζοντας τις δυνατότητες των συστημάτων σε εφαρμογές όπως τα τηλεοπτικά δελτία καιρού. Από τότε, ένας αριθμός τεχνικών υπολογισμού της κίνησης της κάμερας έχουν προταθεί για την επέκταση των δυνατοτήτων των παραδοσιακών virtual studios.

Τα συστήματα παρακολούθησης της κάμερας που χρησιμοποιούνται στα virtual studios ταξινομούνται γενικά σε δύο ευρείες κατηγορίες: τα ηλεκτρομηχανικά και τα οπτικά (ή ακόμη και σε συνδυασμό των δύο). Αρκετά συστήματα virtual studios έχουν αναπτυχθεί, ως πρότυπα ή ως εμπορικά προϊόντα, περιλαμβάνοντας τα ηλεκτρομηχανικά Elset, 3DK και Synthevision και τα οπτικά Cyberset και Mindset μεταξύ άλλων (βλ. και το [46] για λεπτομέρειες). Η ηλεκτρομηχανική παρακολούθηση έχει υιοθετηθεί ευρέως, αφού μπορεί να είναι υψηλής ακριβείας. Σε τέτοιες προσεγγίσεις, σερβομηχανισμοί ελέγχου χρησιμοποιούνται για τον έλεγχο της κάμερας όταν η τρισδιάστατη (3Δ) κίνησή της είναι προκαθορισμένη (ενεργά συστήματα / active systems). Όταν η κίνηση είναι άγνωστη, κατάλληλοι αισθητήρες (sensors) τοποθετούνται στο σκελετό της κάμερας, ώστε να ανιχνεύσουν την κίνησή της (παθητικά συστήματα / passive systems). Για τα ηλεκτρομηχανικά συστήματα, αντίθετα, αναφέρεται ότι απαιτούνται εκτενείς και χρονοβόρες διαδικασίες ρυθμίσεων (calibration), ενώ ταυτόχρονα οι αισθητήρες υποφέρουν από συνήθεις τυχαίες δονήσεις με αποτέλεσμα δυσάρεστα οπτικά αποτελέσματα. Επιπρόσθετα, ο ειδικός εξοπλισμός (π.χ. αισθητήρες) που απαιτείται μπορεί να είναι πολύ ακριβός, ιδιαίτερα όταν αυξάνονται οι επιθυμητοί βαθμοί ελευθερίας στην κίνηση της κάμερας. Τα οπτικά συστήματα βασίζονται σε τεχνικές επεξεργασίας εικόνας ώστε να εξάγουν την κίνηση της κάμερας βάσει μόνο των νέων καρέ που καταγράφονται. Γι' αυτό το λόγο, το 'μονόχρωμο' κυανό παραπέτασμα επεκτείνεται ώστε να περιλάβει κατάλληλα χαρακτηριστικά αναφοράς για

παρακολούθηση στις δύο διαστάσεις, όπως σημεία ή ευθείες γραμμές. Αν και, με αυτόν τον τρόπο, τα προβλήματα των χρονοβόρων ρυθμίσεων και των δονήσεων της κάμερας ξεπερνώνται, μπορεί κανείς να δει ότι τα οπτικά συστήματα αποτυγχάνουν όταν τα χαρακτηριστικά αναφοράς είναι εκτός εστίασης, αποκρύπτονται μερικώς από το προσκήνιο ή, ακόμη, είναι εκτός πλάνου. Αναφέρεται επίσης ότι οι χαρακτηριστικές ενδείξεις (markers) στο πέτασμα προξενούν προβλήματα σύνθεσης, αφού πρέπει να είναι διακριτές από το κυανό φόντο [46].

Στο παρόν κεφάλαιο, εξετάζεται το πρόβλημα της κατασκευής ενός κατάλληλου πετάσματος που να επιτρέπει ταυτόχρονα την άμεση αναγνώριση του οπτικού πεδίου της κάμερας και τον υπολογισμό της 3Δ κίνησής της. Η συγκεκριμένη μέθοδος προτάθηκε στην εργασία [184], ενώ παλαιότερες προς την ίδια κατεύθυνση αλλά ριζικά διαφορετικές προσεγγίσεις δόθηκαν στις εργασίες [31, 182].

Όπως θα φανεί και στα επόμενα, το προτεινόμενο σύστημα παρακολούθησης αντιμετωπίζει επιτυχώς τα προαναφερθέντα συνήθη προβλήματα, οδηγώντας σε υπολογισμό της 3Δ κίνησης της κάμερας με μεγάλη ακρίβεια, χωρίς να επηρεάζεται από απότομες αλλαγές στο οπτικό πεδίο της κάμερας ή να δημιουργεί προβλήματα σύνθεσης. Οι κύριες ιδέες στην κατασκευή του κυανού πετάσματος, όπως φαίνεται στην Ενότητα 8.3, βασίζονται στη χρήση δύο χοντινών σταθμών μπλε χρώματος, αντί του παραδοσιακού μονοχρωματικού πετάσματος. Τα δύο επίπεδα του μπλε κατανέμονται κατάλληλα στο πέτασμα του φόντου, ώστε να επιτρέπουν αναγνώριση του οπτικού πεδίου της κάμερας. Όμοιες προσεγγίσεις έχουν ακολουθηθεί στις [15] και [103], στο βαθμό που το μονοχρωματικό πέτασμα έχει επεκταθεί ώστε να περιέχει περισσότερα επίπεδα του μπλε. Στην [15], ένα μη ομοιόμορφο μπλε φόντο χρησιμοποιείται για να διευκολύνει τον δισδιάστατο (2Δ) υπολογισμό κίνησης με χρήση της μεθόδου οπτικής ροής (optical flow) της [55], η οποία αναφέρεται ως αρκετά εύρωστη ακόμη και για αντικείμενα εκτός εστίασης. Παρόλ'αυτά, η χρήση μίας τέτοιας προσέγγισης φαίνεται να υποφέρει από τις συνήθεις αστοχίες των μεθόδων 2Δ υπολογισμού κίνησης και δυσκολεύει την επιλογή των μπλε σκιών (blue shadows) στην προσπάθεια εξαγωγής σκιών. Στο [103], το αντίστοιχο εμπορικό προϊόν χρησιμοποιεί ένα κυανό πέτασμα δύο τόνων και μεθόδους αναγνώρισης προτύπων για να εξάγει το οπτικό πεδίο της κάμερας. Όμως, σύμφωνα με το [103], το κυανό πέτασμα λαμβάνεται βάσει 'δοκιμής και σφάλματος' ('trial and error') και η χρησιμοποιούμενη μέθοδος αναγνώρισης προτύπου, αν και δεν αναλύεται, φαίνεται να απαιτεί σημαντικό υπολογιστικό φόρτο.

Στο κεφάλαιο αυτό, το προτεινόμενο πέτασμα δύο τόνων κατασκευάζεται βάσει της καλά ορισμένης θεωρίας της αλγεβρικής κωδικοποίησης (algebraic coding), ενώ βασίζεται σε αρχέτυπα πολυώνυμα (primitive polynomials) και δυαδικούς κώδικες μέγιστου μήκους (binary

maximal length codes) [47] του μονοδιάστατου χώρου. Η μονοδιάστατη (1Δ) περίπτωση επεκτείνεται κατάλληλα στη 2Δ περίπτωση (ουσιαστικά, στην περίπτωση των 2Δ επίπεδων επιφανειών) για την κατασκευή ενός πετάσματος κατάλληλου μεγέθους. Συγκεκριμένα, η προτεινόμενη μεθοδολογία περιλαμβάνει τον κατακερματισμό του πετάσματος σε ισομεγέθη παραλληλόγραμμα και τον χρωματισμό τους με σκούρο ή ανοιχτό μπλε. Με αυτόν τον τρόπο, κάθε περιοχή του πετάσματος που υπερβαίνει ένα καθορισμένο μέγεθος (για παράδειγμα, το ελάχιστο οπτικό πεδίο της κάμερας) μπορεί να ταυτοποιηθεί μοναδικά πάνω σ'ολόκληρο το πέτασμα. Ταυτόχρονα, τα σημεία και οι γραμμές αναφοράς μπορούν να εξαχθούν εμμέσως, ως οι τομές των παραλληλογράμμων. Μερικές πρώτες ιδέες για την κατασκευή του πετάσματος βάσει μέγιστων δυαδικών ακολουθιών παρουσιάστηκαν στα [31, 182].

Ιδιαίτερη προσοχή δίνεται στην εξαγωγή των 2Δ χαρακτηριστικών του φόντου και στην μέθοδο εξαγωγής της 3Δ κίνησης της κάμερας. Στο [31], ο υπολογισμός της κίνησης της κάμερας γίνεται βάσει χαρακτηριστικών σημείων που εξάγονται από τα κοινά όρια των παραλληλογράμμων, τα οποία στη συνέχεια τροφοδοτούνται σε κατάλληλο αλγόριθμο υπολογισμού 3Δ κίνησης [163, 176, 140, 25] ανάλογα και με το μοντέλο προβολής [155]. Στο [182], ο υπολογισμός 3Δ κίνησης πραγματοποιείται βάσει των 2Δ χαρακτηριστικών γραμμών (οι οποίες πάλι εξάγονται από το κατατμημένο πέτασμα) λύνοντας ένα σύστημα από γραμμικές ομογενείς εξισώσεις με αγνώστους τις παραμέτρους περιστροφής. Σε σύγκριση και με τις δύο προσεγγίσεις, η [184] παρουσιάζεται ανώτερη ως προς την ακρίβεια που επιτυγχάνεται στην εξαγωγή των 2Δ χαρακτηριστικών. Στην προσέγγιση αυτή προτείνεται ένας νέος αλγόριθμος για τον υπολογισμό 3Δ κίνησης από 3Δ-σε-2Δ αντιστοιχίσεις γραμμών. Για παρόμοιες εργασίες, άλλοι αλγόριθμοι έχουν προταθεί στη βιβλιογραφία περιλαμβάνοντας τους [125, 200, 149, 83] μεταξύ άλλων, εμβαθύνοντας κυρίως στην 2Δ-σε-2Δ περίπτωση (για μία εκτενέστερη ανάλυση στον 3Δ υπολογισμό δομής και κίνησης από αντιστοιχίσεις χαρακτηριστικών βλέπε [60, 35]). Ο προτεινόμενος αλγόριθμος ασχολείται με την περίπτωση των 3Δ-σε-2Δ αντιστοιχίσεων, για το πλέγμα γραμμών που εξάγεται έμμεσα από το προτεινόμενο κυανό πέτασμα. Μία εύρωστη και κομψή λύση παρουσιάζεται βασισμένη στην ανάλυση σε ειδικές τιμές (singular value decomposition) ενός απλού πίνακα.

Η μέθοδος κατασκευής του πετάσματος και ο αλγόριθμος υπολογισμού της 3Δ κίνησης συνδυάζονται αποδοτικά δίνοντας βελτιωμένες μετρήσεις της 3Δ κίνησης της κάμερας. Στην Ενότητα 8.2 ο αλγόριθμος δίνεται στη μορφή ακολουθιακών βημάτων και όλα τα σχετικά ζητήματα / προβλήματα τονίζονται και σχολιάζονται. Στην Ενότητα 8.3 δίνεται η προτεινόμενη μεθοδολογία κατασκευής του πετάσματος, συνοδευόμενη από τις απαραίτητες επεκτάσεις της θεωρίας των 1Δ μέγιστων δυαδικών ακολουθιών. Η Ενότητα 8.4 στη συνέχεια εμπεριέχει

τις θεωρητικές βάσεις για τον υπολογισμό των παραμέτρων της 3Δ περιστροφής και μετατόπισης της κάμερας βάσει των 3Δ-σε-2Δ αντιστοιχίσεων γραμμών και υποδεικνύει πώς τα συνήθη προβλήματα απόκρυψης ή απουσίας χαρακτηριστικών ξεπερνώνται, ενώ η Ενότητα 8.5 εισάγει την αντίστοιχη ασυμπτωτική ανάλυση θορύβου. Η Ενότητα 8.6 δίνει στον ενδιαφερόμενο αναγνώστη πρακτικές αλγοριθμικές λεπτομέρειες και οδηγίες υλοποίησης, ενώ κατάλληλα πειραματικά αποτελέσματα περιλαμβάνονται για να επιδείξουν την ευρωστέια της μεθόδου. Τέλος, η Ενότητα 8.7 επαναλαμβάνει τις θεωρητικές και πρακτικές καινοτομίες του κεφαλαίου. Σύντομες αποδείξεις όλων των θεωρητικών Προτάσεων δίνονται μέσα στο κείμενο.

## 8.2 Περιγραφή του συστήματος

Σ'αυτή την ενότητα, ο προτεινόμενος αλγόριθμος παρουσιάζεται στη μορφή ακολουθιακών διαδικασιών. Επιπρόσθετα, τα προβλήματα των συστημάτων οπτικής παρακολούθησης υπογραμμίζονται και αναλύονται σε σχέση με την προτεινόμενη προσέγγιση.

### 8.2.1 Σχεδιασμός του συστήματος οπτικής παρακολούθησης

Τα συστήματα οπτικής παρακολούθησης στα virtual studios εκμεταλλεύονται το γεγονός ότι κάθε καρέ που καταγράφεται περιέχει ήδη μία μεγάλη ποσότητα πληροφορίας για το φόντο, η οποία θα μπορούσε να χρησιμοποιηθεί και για παρακολούθηση της κίνησης, πέρα από το διαχωρισμό του φόντου. Έτσι, κατάλληλα χαρακτηριστικά (features), τα οποία μπορούν να εντοπιστούν και να γίνουν αντικείμενο παρακολούθησης με την πάροδο του χρόνου ενσωματώνονται στο κυανό πέτασμα.

Για σκοπούς παρακολούθησης της κίνησης της κάμερας, τα χρησιμοποιούμενα χαρακτηριστικά θα πρέπει να είναι αρκετά σε πλήθος, διαχωρίσιμα από το ομοιόμορφο κυανό φόντο αλλά και μεταξύ τους κατά κάποιον τρόπο, για παράδειγμα διαχωρίσιμα σε χρωματική τονικότητα ή χωρική διάταξη. Μπορεί να δει κανείς, ότι γνώση της πραγματικής ή ακόμη και της σχετικής θέσης τους στο φόντο δεν είναι αρκετή, αφού το οπτικό πεδίο της κάμερας είναι άγνωστο με το χρόνο και η απόσταση των χαρακτηριστικών καθορίζεται από το μοντέλο προβολής και την θέση της κάμερας στον 3Δ χώρο. Ταυτόχρονα, η ποσότητα της πληροφορίας που εισάγεται στο σύστημα για λόγους παρακολούθησης, δεν θα πρέπει να προξενεί δυσκολίες σύνθεσης (compositing). Γενικά τα παρακάτω ζητήματα θα πρέπει να ληφθούν υπόψη:

1. ο συνολικός αριθμός των σταθμών μπλε χρώματος θα πρέπει να μείνει ουσιαστικά μικρός, ώστε να επιτρέπει τον αποδοτικό διαχωρισμό χρώματος (chromakeying), αλλά ακόμη και ικανοποιητική εξαγωγή σκιών (shadows), όταν το τελευταίο είναι επιθυμητό,

2. τα χαρακτηριστικά που χρησιμοποιούνται θα πρέπει να είναι διακριτά, ώστε να επιτρέπουν τη χωροθέτηση (localization) του οπτικού πεδίου της κάμερας,

3. το σύνολο των χαρακτηριστικών θα πρέπει να περιέχει ικανή πληροφορία τοπικά, ακόμη κι όταν ένα σημαντικό μέρος του φόντου καλύπτεται από τη δράση στο προσκήνιο σε κάθε καρέ που καταγράφεται,

4. η τεχνική παρακολούθησης θα πρέπει να επιτρέπει απότομες κινήσεις της κάμερας σε όλες τις κατευθύνσεις (όλους τους βαθμούς ελευθερίας).

Γενικά, μπορεί κανείς να δει ότι τα περισσότερα από τα υπάρχοντα συστήματα δεν ικανοποιούν το σύνολο των παραπάνω απαιτήσεων. Για παράδειγμα, τα συστήματα που βασίζονται σε υπολογισμό της οπτικής ροής (optical flow) εισάγουν δυσκολίες στον διαχωρισμό χρώματος, λόγω του μεγάλου αριθμού των σταθμών του μπλε χρώματος που απαιτούν, ενώ τα συστήματα που βασίζονται σε σημάνσεις (markers) υποφέρουν από τοπική απόκρυψη (partial occlusion) ή απουσία των σημάνσεων στο οπτικό πεδίο. Επιπρόσθετα, οι περισσότερες προσεγγίσεις υποθέτουν μικρή και ‘προσεκτική’ κίνηση της κάμερας ή θέτουν περιορισμούς στους βαθμούς ελευθερίας που επιτρέπονται. Τέλος, μία ακόμη απαίτηση πρέπει να τεθεί, συγκεκριμένα

5. η τεχνική παρακολούθησης θα πρέπει να επιτρέπει την πλήρη αλλαγή του οπτικού πεδίου της κάμερας.

Η τελευταία απαίτηση προκύπτει θεωρώντας το επόμενο σενάριο. Ας υποθέσουμε ότι μόνο μία κάμερα είναι διαθέσιμη για να καλύψει μία σκηνή όπου οι δύο ήρωες συζητούν, οφείλοντας να κινείται με ταχύτητα μεταξύ τους. Ισοδύναμα, θεωρείστε την περίπτωση όπου για μία πολύπλοκη σκηνή περισσότερες από μία κάμερες χρησιμοποιούνται, ο αριθμός όμως των συστημάτων επεξεργασίας (για παράδειγμα, ο αριθμός των διαχωριστών χρώματος) είναι μικρότερος από το πλήθος των μονάδων καταγραφής. Σε αυτές τις περιπτώσεις, όλη η επεξεργασία θα πρέπει να εκτελεστεί σε μη πραγματικό χρόνο για όλες τις κάμερες και οι σύνθετες (composite) ακολουθίες θα πρέπει στη συνέχεια να οδηγηθούν στη διαδικασία του μοντάζ. Στην περίπτωση που αυτό το κόστος σε χρόνο δεν είναι αποδεκτό (για παράδειγμα, όταν πρόκειται για μεταδόσεις πραγματικού χρόνου), ένας διαχωριστής χρώματος θα πρέπει να επεξεργαστεί αντί για την ακολουθία βίντεο που έρχεται από την κάμερα, την ακολουθία βίντεο που έρχεται από το μίκτη, δηλαδή μία ακολουθία που παράγεται από διαδοχικές μεταβάσεις από κάμερα σε κάμερα. Σύμφωνα με τα παραπάνω, μπορεί κανείς να δει ότι μεταξύ μεταβάσεων, το οπτικό πεδίο της κάμερας, όπως αυτό παρουσιάζεται στην ακολουθία που φθάνει στο διαχωριστή χρώματος, αλλάζει δραματικά, ώστε να μην υπάρχουν ορατά χαρακτηριστικά που να μπορούν να

ακολουθηθούν στο χρόνο. Η καλύτερη ίσως λύση σε αυτές τις περιπτώσεις είναι το σύστημα να υπολογίζει την 3Δ θέση της κάμερας βάσει ενός καρέ (και ως προς κάποια αναφορά), παρά την 3Δ κίνηση της κάμερας βάσει δύο ή περισσότερων καρέ.

Όπως θα δειχθεί στα επόμενα, το προτεινόμενο σύστημα ικανοποιεί όλες τις παραπάνω πέντε γενικές απαιτήσεις.

### 8.2.2 Σχεδιασμός του κυανού πετάσματος

Στην προσέγγισή μας, το κυανό πέτασμα κατασκευάζεται χρησιμοποιώντας δύο σχετικά κοντινούς τόνους του μπλε. Με αυτόν τον τρόπο αποφεύγονται πιθανά προβλήματα διαχωρισμού χρώματος, αφού τα επίπεδα του μπλε μπορούν να επιλεγούν αρκετά κοντά μεταξύ τους στο χώρο των χρωμάτων. Ικανοποιείται έτσι η απαίτηση (1).

Αφότου διαιρεθεί ο τοίχος, σε σχήμα παραλληλογράμμου, σε  $N \times M$  ισομεγέθη παραλληλόγραμμα, το κάθε ένα από αυτά βάφεται με σκούρο ή ανοιχτό μπλε. Έτσι, λαμβάνεται ένας δυαδικός πίνακας διαστάσεων  $N \times M$ , ο οποίος στη συνέχεια θα καλείται ‘δυαδικός χάρτης του κυανού πετάσματος’. Σε κάθε καρέ, η κάμερα καταγράφει ένα μικρό μέρος του κυανού πετάσματος, το οποίο με τη σειρά του αντιστοιχεί ακριβώς σε ένα μικρό μέρος του δυαδικού χάρτη. Υποθέτοντας ότι ένας τυχαίος υποπίνακας μεγέθους τουλάχιστον  $n \times m$  είναι πάντα ορατός από την κάμερα, αρκεί κάθε τέτοιος υποπίνακας να εμφανίζεται μόνο μία φορά σε ολόκληρο το κυανό πέτασμα (δηλ. σε ολόκληρο το δυαδικό χάρτη). Σύμφωνα με τα παραπάνω, αφότου ένας συγκεκριμένος υποπίνακας εξάγεται από το ορατό μέρος του πετάσματος, η θέση του οπτικού πεδίου της κάμερας επάνω στο πέτασμα μπορεί τότε να εκτιμηθεί.

Ένα τέτοιο κυανό πέτασμα μπορεί να κατασκευαστεί, επεκτείνοντας την βασική θεωρία των 1Δ ψευδοτυχαίων δυαδικών ακολουθιών μέγιστου μήκους (maximal length 1D pseudo-random binary sequences), οι οποίες παρουσιάζουν ένα σύνολο από χρήσιμες ιδιότητες. Για παράδειγμα, είναι γνωστό ότι σε μία δυαδική ακολουθία μέγιστου μήκους με συνολικό μήκος  $2^n - 1$ , κάθε υποακολουθία ή κατάσταση (state) μήκους  $n$  μπορεί να βρεθεί μόνο μία φορά. Τέτοιες ιδιότητες των ‘μέγιστων ακολουθιών’ (καλούνται συνήθως και ως maximal sequences εν συντομία) του 1Δ χώρου μπορούν να επεκταθούν και στον 2Δ χώρο, ακολουθώντας την Πρόταση 8.1 της Ενότητας 8.3.

Μπορεί τώρα να τονισθεί ότι ακόμα δύο από τις απαιτήσεις μας ικανοποιούνται αυτόματα. Υποθέτοντας ότι οποιοσδήποτε τυχαίος υποπίνακας του δυαδικού χάρτη, που είναι ορατός από την κάμερα, μπορεί να εντοπιστεί και εν συνεχεία να αντιστοιχηθεί μοναδικά επάνω στο δυαδικό χάρτη, τότε η χωροθέτηση του οπτικού πεδίου της κάμερας είναι δυνατή (βλ. απαίτηση

(2)). Επιπρόσθετα, θέτοντας επιτυχώς τον μικρότερο υποπίνακα που μπορεί να είναι ορατός από την κάμερα (συμπεριλαμβάνοντας την περίπτωση της μέγιστης δυνατής επικάλυψης του φόντου), είναι βέβαιο ότι το σύνολο των χαρακτηριστικών (features) περιέχει πάντα ικανή πληροφορία για τον υπολογισμό της κίνησης (βλ. απαίτηση (3)).

### 8.2.3 Εντοπισμός των δισδιάστατων χαρακτηριστικών

Τα χαρακτηριστικά που χρησιμοποιούνται για τον υπολογισμό της κίνησης της κάμερας είναι πρακτικά οι γραμμές αναφοράς (grid lines), οι οποίες λαμβάνονται εμμέσως από την κατασκευή του δυαδικού χάρτη. Έτσι, ένα επιπλέον πλεονέκτημα της προτεινόμενης μεθόδου είναι ότι τα χαρακτηριστικά είναι διαχωρίσιμα χωρίς να είναι αναγκαία η αύξηση της χρωματικής περιοχής του χαρακτηριστικού χρώματος και με τρόπο ακριβή και καλώς ορισμένο, σε πλήρη αντίθεση με τις προσεγγίσεις που χρησιμοποιούν σημάνσεις (markers).

Ας υποθέσουμε για το καρέ που καταγράφεται την τρέχουσα χρονική στιγμή, ότι το ορατό από την κάμερα τμήμα του πετάσματος έχει ληφθεί με οποιαδήποτε τεχνική διαχωρισμού χρώματος. Τότε, τα περιγράμματα των τετραπλεύρων του πετάσματος εξάγονται χρησιμοποιώντας μία τεχνική εντοπισμού γραμμών (edge detection) με ένα συνήθη τελεστή κλίσης, π.χ. Sobel. Στην πραγματικότητα, τα περιγράμματα των τετραπλεύρων εξάγονται μόνο για γειτονικά τετράπλευρα διαφορετικού χρωματικού τόνου. Η ακόλουθη εφαρμογή του μετασχηματισμού Hough επιτρέπει τον εντοπισμό των γραμμών, που πρακτικά αποτελούν τις μετασχηματισμένες γραμμές αναφοράς του κυανού πετάσματος. Μία απλή μέθοδος ελαχίστων τετραγώνων χρησιμοποιείται τότε για την εξαγωγή των ακριβέστερων δυνατών παραμέτρων των γραμμών. Εφόσον στο ορατό μέρος του πετάσματος μπορούν να περιέχονται ομάδες από γειτονικά τετράπλευρα ιδίου χρωματικού τόνου, είναι πιθανόν κάποιες από τις γραμμές αναφοράς να μη δύνανται να εξαχθούν. Σε μία τέτοια περίπτωση, ο υποπίνακας που αντιστοιχεί στο ορατό μέρος του δυαδικού χάρτη δεν μπορεί ακόμη να εξαχθεί.

### 8.2.4 Υπολογισμός των παραμέτρων τρισδιάστατης κίνησης της κάμερας

Αν και υπάρχουν περιπτώσεις που το ορατό μέρος του δυαδικού χάρτη δεν μπορεί ακόμη να εξαχθεί, η περιστροφή στην οποία υπεβλήθη η κάμερα μπορεί να υπολογιστεί άμεσα. Χρησιμοποιώντας το σύνολο των γραμμών από την έξοδο του Hough μετασχηματισμού, οι καλύτερες 2Δ γραμμές επιλέγονται βάσει της ακρίβειας με την οποία εξήχθησαν από την διαδικασία εντοπισμού. Το λαμβανόμενο σύνολο γραμμών διαιρείται τότε σε δύο υποσύνολα, βάσει των κλίσεών τους, ώστε τα δύο υποσύνολα να αντιστοιχούν στις οριζόντιες και κατακόρυφες γραμμές αναφοράς του κυανού πετάσματος στον πραγματικό χώρο. Για κάθε γραμμή, οι εξαγόμενες



παράμετροι σε μονάδες pixel κανονικοποιούνται σε μετρήσεις πραγματικού χώρου χρησιμοποιώντας το λόγο των διαστάσεων του πραγματικού CCD ως προς τις διαστάσεις του καρέ σε pixels.

Εφόσον η πραγματική θέση των γραμμών αναφοράς στο κυανό πέτασμα είναι γνωστή εκ κατασκευής, μπορούμε να λύσουμε ένα πρόβλημα υπολογισμού 3Δ κίνησης από προβολές, βάσει των 3Δ-σε-2Δ αντιστοιχίσεων των γραμμών. Με αυτό το σκοπό, μία εικονική 3Δ σκηνή αναφοράς κατασκευάζεται βάσει

(α) των οριζόντιων και κατακόρυφων μηκών των μπλοκ (που δίνουν τη θέση των γραμμών αναφοράς) και

(β) την απόσταση του κυανού πετάσματος από την κάμερα.

Η τελευταία μπορεί να επιλεγεί τυχαία ως η αρχική θέση του πετάσματος στην εικονική σκηνή αναφοράς.

Όπως μπορεί να δει κανείς στην Ενότητα 8.4, ο πίνακας περιστροφής υπολογίζεται τότε αποδοτικά από την ανάλυση σε ειδικές τιμές (singular value decomposition) ενός απλού πίνακα. Πρακτικά, η Πρόταση 8.2 της Ενότητας 8.4 δείχνει ότι η γνώση της ακριβούς θέσης των γραμμών στον 3Δ χώρο στη σκηνή αναφοράς δεν είναι αναγκαία για τον υπολογισμό της περιστροφής - αρκεί απλώς οι οριζόντιες γραμμές να είναι κάθετες ως προς τις κατακόρυφες.

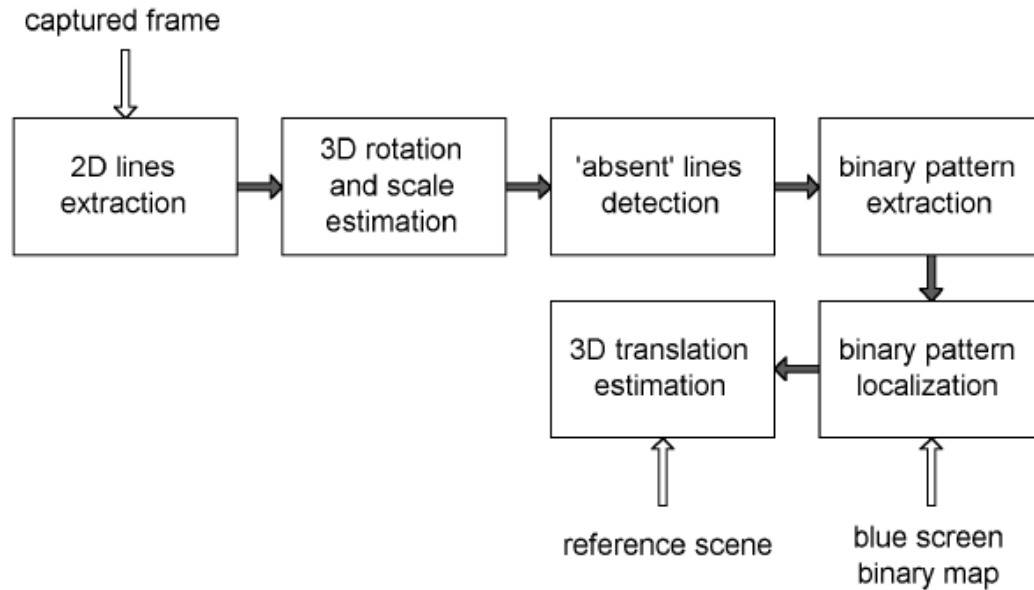
Ο υπολογισμός της 3Δ μετατόπισης, αντίθετα με τον υπολογισμό της 3Δ περιστροφής, απαιτεί τη γνώση των 3Δ παραμέτρων των γραμμών στη σκηνή αναφοράς αλλά και την ακριβή αντιστοίχιση ανάμεσα στις 3Δ γραμμές αναφοράς και τις 2Δ γραμμές που εντοπίζονται στο τρέχον καρέ. Σύμφωνα με τα παραπάνω, το τελευταίο καθίσταται δυνατό αναγνωρίζοντας σε ποιο τμήμα του πραγματικού κυανού πετάσματος αντιστοιχεί το καταγραφόμενο από την κάμερα τμήμα. Από την άλλη πλευρά, δεν είναι δυνατόν να εντοπίσουμε το ακριβές δυαδικό πρότυπο, αφού κάποιες από τις μετασχηματισμένες γραμμές αναφοράς μπορούν να απουσιάζουν από το καταγραφόμενο καρέ. Αυτές οι δύο 'δυσκολίες' λύνονται παράλληλα. Η Πρόταση 8.3 της Ενότητας 8.4 δείχνει πώς, για οποιεσδήποτε δύο γραμμές, η σχετική τους απόσταση στη σκηνή αναφοράς μπορεί να υπολογιστεί από τις εξαγόμενες 2Δ γραμμές. Με αυτόν τον τρόπο, οι γραμμές που απουσιάζουν μπορούν να εντοπιστούν και το ορατό δυαδικό πρότυπο να εξαχθεί. Η αντιστοίχιση μεταξύ των 2Δ και 3Δ γραμμών είναι τότε δυνατή μετά τη χωροθέτηση του τελευταίου επάνω στο δυαδικό χάρτη. Το διάνυσμα 3Δ μετατόπισης υπολογίζεται τελικώς από τον πίνακα περιστροφής, από τις γραμμές των οποίων οι αντιστοιχίσεις έχουν υπολογιστεί και από το βάθος του πετάσματος που τέθηκε στη σκηνή αναφοράς (βλ. Πρόταση 8.4 στην Ενότητα 8.4).

Μία από τις κύριες συνεισφορές του παρόντος κεφαλαίου, είναι ότι η προτεινόμενη λύση επιτρέπει επίσης τον υπολογισμό του άγνωστου εστιακού μήκους (focal length)  $f$  σε κάθε καταγραφόμενο καρέ. Η Πρόταση 8.5 στην Ενότητα 8.4 παρουσιάζει έναν κατάλληλο τρόπο υπολογισμού του εστιακού μήκους, βάσει των παραμέτρων των γραμμών που εξάγονται από το τρέχον καταγραφόμενο καρέ. Πρακτικά, η Πρόταση 8.5 συμπληρώνει τις Προτάσεις 8.2, 8.3 και 8.4, οδηγώντας στον υπολογισμό όλων των αγνώστων παραμέτρων κίνησης, δηλ. 3Δ περιστροφής, μετατόπισης και μεγέθυνσης.

Σύμφωνα με τα παραπάνω, η προτεινόμενη προσέγγιση αντιμετωπίζει με επιτυχία και τις δύο τελευταίες από τις απαιτήσεις, δηλαδή την ελεύθερη κίνηση της κάμερας (βλ. απαίτηση (4)) και την αποδοτική εκτίμηση κίνησης ακόμη και για ολική αλλαγή στο οπτικό πεδίο της κάμερας (βλ. απαίτηση (5)).

Εκ των πραγμάτων, ο προτεινόμενος αλγόριθμος υπολογίζει μάλλον την 3Δ θέση και τον προσανατολισμό της κάμερας συναρτήσει του χρόνου, παρά την 3Δ κίνηση της κάμερας. Το τελευταίο γίνεται δυνατό κύριως χάριν στο γεγονός ότι δεν γίνεται χρήση καμίας τεχνικής υπολογισμού κίνησης. Αντιθέτως, τα χρησιμοποιούμενα χαρακτηριστικά εντοπίζονται βάσει της μοναδικής τους εμφάνισης στο κυανό πέτασμα κι έτσι δεν υπάρχει ανάγκη παρακολούθησής τους και 2Δ υπολογισμού κίνησης. Ταυτόχρονα, μπορεί να δει κανείς ότι δεν είναι αναγκαίες επίπονες διαδικασίες ρυθμίσεων (calibration) αφού η κατασκευή της σκηνής αναφοράς είναι εικονική. Για να γίνει το τελευταίο σημείο περισσότερο κατανοητό, θεωρήστε την ενδιαφέρουσα περίπτωση όπου μία πραγματική κάμερα καταγράφει τη δράση στο προσκήνιο έναντι του προτεινόμενου κυανού πετάσματος, ενώ μία εικονική κάμερα αποδίδει ταυτόχρονα ένα εικονικό μοντέλο / κόσμο. Σε αυτήν την περίπτωση, για να λάβουμε συμβατά αποτελέσματα σύνθεσης (compositing), η εικονική κάμερα θα πρέπει να ακολουθεί την κίνηση της πραγματικής κάμερας (στα ηλεκτρομηχανικά συστήματα το τελευταίο ονομάζεται camera ganging). Η μοναδική αναγκαία ρύθμιση είναι, σε κάθε περίπτωση, οι υποτιθέμενες αποστάσεις των γραμμών στην εικονική σκηνή αναφοράς να συμπίπτουν με τις αποστάσεις των γραμμών στον πραγματικό χώρο. Αυτό προϋποθέτει ότι το κυανό πέτασμα κατασκευάζεται με ακρίβεια, ενώ καμιά ρύθμιση δεν είναι αναγκαία στη συνέχεια σε επίπεδο καταγραφής. Τέλος, το προτεινόμενο σύστημα, ως ένα οπτικό σύστημα παρακολούθησης (optical tracking system), αλλά παράλληλα και λόγω του ότι δεν θέτει περιορισμούς στους βαθμούς ελευθερίας της κάμερας, επιτρέπει την καταγραφή βίντεο με χρήση ακόμη και μίας κάμερας χειρός.

Στο Σχήμα 8.1 το προτεινόμενο σύστημα δίνεται σε μπλοκ-διάγραμμα.



Σχήμα 8.1 Το προτεινόμενο σύστημα υπολογισμού κίνησης σε μπλοκ-διάγραμμα.

### 8.3 Κατασκευή του δυαδικού πίνακα του πετάσματος

Σε ένα οπτικό σύστημα υπολογισμού της κίνησης της κάμερας, κύριος σκοπός του συστήματος είναι η εξαγωγή όσο το δυνατόν περισσότερης ωφέλιμης πληροφορίας από την καταγραφόμενη ακολουθία. Στη γενική περίπτωση, η καταγραφόμενη πληροφορία αποτελείται από τα αντικείμενα του προσκηνίου και το κυανό πέτασμα στο φόντο. Εφόσον υπάρχει ήδη μία μεγάλη ποσότητα πληροφορίας στο φόντο, για να χρησιμοποιηθεί μόνο για σκοπούς διαχωρισμού χρώματος, ο στόχος μετατίθεται στην εισαγωγή πρόσθετης πληροφορίας πάνω στο πέτασμα, η οποία να χρησιμοποιηθεί αργότερα στον υπολογισμό κίνησης της κάμερας. Υπάρχουν δύο κύριες τεχνικές για αυτόν το σκοπό: η τοποθέτηση σημάνσεων (markers) στο πέτασμα και η χρήση περισσότερων του ενός τόνων του μπλε για την κατασκευή του.

Το παρόν κεφάλαιο προτείνει την κατασκευή ενός κυανού πετάσματος αποτελούμενου από ισομεγέθη παραλληλόγραμμα, βαμμένα με έναν από δύο χρησιμοποιούμενους τόνους του μπλε. Η τεχνική αυτή δεν διαφοροποιεί σημαντικά την υλοποίηση ή την πολυπλοκότητα του αλγορίθμου διαχωρισμού χρώματος. Ταυτόχρονα, προσφέρει έναν απλό και αποδοτικό τρόπο εισαγωγής δυαδικής πληροφορίας στο κυανό πέτασμα.

Στα επόμενα, τα σύμβολα 1 και 0 θα υπονοούν τον ανοιχτό και τον σκούρο τόνο του μπλε αντίστοιχα. Ένα πέτασμα, διαιρεμένο σε  $N \times M$  παραλληλόγραμμα, μπορεί να οριστεί μαθηματικά από

- τον αντίστοιχο  $N \times M$  δυαδικό χάρτη  $\mathbf{B}$ , περιέχοντα 1 και 0 στις κατάλληλες θέσεις και
- τις διαστάσεις του στον πραγματικό χώρο.

Σε κάθε καρέ, η κάμερα καταγράφει ένα μικρό τμήμα του τοίχου, το οποίο αντιστοιχεί σε ένα τμήμα  $\Pi$  του πίνακα  $\mathbf{B}$ . Το  $\Pi$  είναι ουσιαστικά μία συλλογή στοιχείων του  $\mathbf{B}$  και δεν αντιστοιχεί γενικά σε κάποιον παραλληλόγραμμο υποπίνακα του  $\mathbf{B}$ . Έστω ότι ο  $\mathbf{S}$  υποδηλώνει τον μέγιστο υποπίνακα του  $\mathbf{B}$ , του οποίου τα στοιχεία (τετράπλευρα) ανήκουν όλα στο ορατό τμήμα του κυανού πετάσματος  $\Pi$ . Μπορεί να δει κανείς ότι, εάν και μόνο εάν ο  $\mathbf{S}$ , ή οι υποπίνακές του που ξεπερνούν ένα προαποφασισμένο μέγεθος, εμφανίζεται μόνο μία φορά μέσα στον  $\mathbf{B}$ , τότε το οπτικό πεδίο της κάμερας ως προς το πέτασμα μπορεί να αναγνωριστεί μοναδικά. Αυτό μπορεί να καταστεί δυνατό μόνο με χρήση μίας κατάλληλης μεθοδολογίας στην τοποθέτηση των 1 και 0 στον πίνακα  $\mathbf{B}$ . Θαδειχθεί ότι ένας τρόπος εξασφάλισης μοναδικότητας σε όλους τους υποπίνακες του  $\mathbf{B}$  που ισούνται ή ξεπερνούν ένα συγκεκριμένο μέγεθος  $n \times m$ , δίνεται μέσα από τη χρήση των τεχνικών αλγεβρικής κωδικοποίησης (algebraic coding techniques).

Στην επόμενη υποενότητα δίνονται θεωρητικές οδηγίες για την κατασκευή ενός τέτοιου πετάσματος, ενώ στην Υποενότητα 8.3.2 παρουσιάζεται ο αντίστοιχος αλγόριθμος.

### 8.3.1 Ένα ειδικό δισδιάστατο δυαδικό πεδίο

Γενικά, η δομή των πεδίων Galois (Galois fields) και οι ιδιότητες των αρχέτυπων πολυωνύμων (primitive polynomials) αποτελούν ένα αποδοτικό εργαλείο για την κατασκευή οποιουδήποτε κυκλικού κώδικα μέγιστου μήκους. Υποθέστε ότι  $\mathbf{h}$  είναι το μικρότερο πολυώνυμο ενός αρχέτυπου στοιχείου στο  $\text{GF}(p^n)$ , όπου  $p$  ένας πρώτος αριθμός. Ένας κυκλικός κώδικας  $\mathbf{c}$  με μήκος περιόδου  $p^n - 1$ , για τον οποίο το  $\mathbf{h}$  είναι το πολυώνυμο ελέγχου (check polynomial) δηλαδή αρχέτυπο πολυώνυμο βαθμού  $n$ , είναι γνωστός ως κώδικας μέγιστου μήκους (maximal length code) [41]. Στα επόμενα, όταν αναφερόμαστε στην ακολουθία / κώδικα μέγιστου μήκους, θα εννοούμε οποιαδήποτε περίοδο του τελευταίου.

Οι κωδικές λέξεις (codewords) ενός τέτοιου κώδικα μπορούν να παραχθούν από ένα κύκλωμα καταχωρητή ολίσθησης (shift register). Εάν  $R(\mathbf{h})$  είναι ο αντίστοιχος απλός καταχωρητής ολίσθησης με  $n$  θέσεις, τότε θα υπάρχει σύνδεση ανατροφοδότησης στην θέση  $i$ , εάν και μόνον εάν  $h_{i+1} = 1$ . Ως αποτέλεσμα, ο καταχωρητής περνάει από όλες τις  $p^n - 1$  καταστάσεις, που παράγονται από τις  $n$  θέσεις ψηφίων με χρήση  $p$  διαχωρίσιμων ψηφίων (και έχοντας εξαιρέσει τη μηδενική κατάσταση), οπότε και η ακολουθία εξόδου  $\mathbf{c}$  έχει περίοδο μήκους  $p^n - 1$  [47].

Για  $p \equiv 2$ , η παραπάνω θεωρία δίνει μία μέγιστη ακολουθία στο  $\text{GF}(2^n)$  για κάθε αριθμό  $n$ , και έτσι διασφαλίζει τη μοναδικότητα κάθε δυαδικού προτύπου μεγέθους  $n \times 1$  σε μία ακολουθία μεγέθους  $(2^n - 1) \times 1$ . Έτσι, διασφαλίζεται ότι κάθε  $n \times 1$  πρότυπο είναι μοναδικό σε μία τέτοια στήλη του δυαδικού χάρτη (του πετάσματος) με μέγεθος  $(2^n - 1) \times 1$ . Το επόμενο βήμα είναι η επέκταση αυτής της 1Δ ιδιότητας και στις δύο διαστάσεις του κυανού πετάσματος, ώστε να έχουμε μοναδικά  $n \times m$  πρότυπα στο συνολικό  $N \times M$  πέτασμα. Μία πρώτη προσέγγιση δόθηκε στις αναφορές [31, 182], πολλαπλασιάζοντας δύο κώδικες μέγιστου μήκους, που λαμβάνονταν από δύο αρχέτυπα πολυώνυμα τάξεως  $n$  και  $m$ . Μία βελτιωμένη λύση δίνεται στο κεφάλαιο αυτό, εκμεταλλευόμενη την ιδιότητα της κυκλικότητας των ακολουθιών μέγιστου μήκους, δηλ. ότι όλες οι πιθανές ολισθήσεις ενός κώδικα  $\mathbf{c}$  είναι μέγιστες ακολουθίες που αποτελούνται από τις καταστάσεις του ίδιου καταχωρητή ολίσθησης.

Σ' αυτό το πνεύμα, η επόμενη πρόταση παρουσιάζει έναν κατάλληλο τρόπο κατασκευής ενός 2Δ δυαδικού χάρτη  $\mathbf{B}$ , όπου κάθε  $n \times m$  (ή μεγαλύτερος) υποπίνακας  $\mathbf{S}$  είναι μοναδικός.

**Πρόταση 8.1:** Έστω  $\mathbf{c}$  ένας κώδικας μέγιστου μήκους με μήκος  $2^n - 1$ , και  $\mathbf{c}_r, \mathbf{c}_s$  δύο ολισθημένες εκδοχές του  $\mathbf{c}$ , όπου  $0 \leq s, r \leq 2^n - 2$ . Έστω επίσης,  $\mathbf{B} \equiv [b_{ij}]$  και έστω  $\mathbf{b}_j$  η  $j$ -οστή στήλη του  $\mathbf{B}$ . Για ένα ζεύγος από γειτονικές στήλες του  $\mathbf{B}$ , έστω  $\mathbf{b}_j \equiv \mathbf{c}_r$  και  $\mathbf{b}_{j+1} \equiv \mathbf{c}_s$ , και το  $d_j \equiv s - r$  να υποδηλώνει την διαφορά ολίσθησης για το ζεύγος στηλών  $(j, j + 1)$ .

Όμοια, για  $m$  γειτονικές στήλες έστω  $\mathbf{d}_j \triangleq [d_j \ d_{j+1} \ \cdots \ d_{j+m-2}]^T$  το διάνυσμα των διαφορών ολίσθησης. Τότε,

(α) δύο  $n \times m$  υποπίνακες του  $\mathbf{B}$  με τα πρώτα τους στοιχεία στα  $b_{i_1 j_1}, b_{i_2 j_2}$  είναι διαφορετικοί μεταξύ τους, εάν  $\mathbf{d}_{j_1} \neq \mathbf{d}_{j_2}$ ,

(β)  $\mathbf{d}_j$  ( $m - 1$ )-άδες μπορούν να τεθούν σε σειρά βάσει μίας μέγιστης ακολουθίας που λαμβάνεται από οποιοδήποτε πολυώνυμο ελέγχου στο  $\text{GF}((2^n - 1)^{m-1})$ .

**Απόδειξη:**

(α) Έστω  $\mathbf{S}_1$  και  $\mathbf{S}_2$  δύο  $n \times m$  υποπίνακες του  $\mathbf{B}$  με τα πρώτα τους στοιχεία να συμπίπτουν με τα  $b_{i_1 j_1}$  και  $b_{i_2 j_2}$  αντίστοιχα. Μπορούμε να χειριστούμε χωριστά τις ακόλουθες περιπτώσεις:

(α<sub>1</sub>) Εάν  $j_1 = j_2 \equiv j$ , τότε οι πρώτες στήλες των  $\mathbf{S}_1$  και  $\mathbf{S}_2$  βρίσκονται στην  $j$ -οστή στήλη του  $\mathbf{B}$ . Εφόσον η τελευταία είναι μία ολισθημένη εκδοχή της ακολουθίας μέγιστου μήκους  $\mathbf{c}$ , τότε τουλάχιστον οι πρώτες στήλες των  $\mathbf{S}_1$  και  $\mathbf{S}_2$  διαφέρουν για  $i_1 \neq i_2$ . Οπότε,  $\mathbf{S}_1 \neq \mathbf{S}_2$ .

(α<sub>2</sub>) Για  $j_1 \neq j_2$  και  $\mathbf{d}_{j_1} \neq \mathbf{d}_{j_2}$ , οι πρώτες στήλες των  $\mathbf{S}_1$  και  $\mathbf{S}_2$  μπορούν να είναι πανομοιότυπες. Όμως, εάν για παράδειγμα τα  $\mathbf{d}_{j_1}$  και  $\mathbf{d}_{j_2}$  διαφέρουν στο  $k$ -οστό τους στοιχείο,

τότε οι  $k + 1$ -οστές στήλες των  $\mathbf{S}_1$  και  $\mathbf{S}_2$  θα διαφέρουν, αφού θα αντιστοιχούν στο ίδιο  $n \times 1$  τμήμα διαφορετικών ολισθήσεων του  $\mathbf{c}$ , δηλ. σε διαφορετικές καταστάσεις στο  $\mathbf{c}$ . Οπότε, πάλι  $\mathbf{S}_1 \neq \mathbf{S}_2$ .

Μπορεί να δει κανείς ότι εάν η απαίτηση  $\mathbf{d}_{j_1} \neq \mathbf{d}_{j_2}$  αποσυρθεί, τότε θα μπορούσαν γενικά να υπάρχουν τέτοια  $i_1, i_2$  για  $j_1 \neq j_2$ , ώστε  $\mathbf{S}_1 = \mathbf{S}_2$ .

(β) Από το (α), συνάγουμε ότι ένας τρόπος να διασφαλίσουμε ότι κάθε  $n \times m$  πίνακας  $\mathbf{S}$  είναι μοναδικός στο  $\mathbf{B}$ , είναι όταν το (α) ισχύει για οποιουδήποτε δύο  $n \times m$  υποπίνακες στο  $\mathbf{B}$ . Ακόμη, μπορεί να σημειωθεί ότι το  $\mathbf{d}_{j+1}$  είναι εκ κατασκευής μία ολισθημένη εκδοχή του  $\mathbf{d}_j$ . Επιπρόσθετα, με στόχο να μεγιστοποιήσουμε τον αριθμό των στηλών του  $\mathbf{B}$ , όλες οι πιθανές ‘καταστάσεις’ του  $\mathbf{d}_j$  είναι αναγκαίες. Μία μέγιστη ακολουθία που περιέχει όλες τις πιθανές  $(m - 1)$ -άδες από  $\mathbf{d}_j$  λαμβάνεται από οποιοδήποτε πολυώνυμο ελέγχου στο  $\text{GF}((2^n - 1)^{m-1})$ . Ο.Ε.Δ.

Η Πρόταση 8.1 πρακτικά δηλώνει ότι ο  $\mathbf{B}$  κατασκευάζεται βάσει ολισθημένων εκδοχών του  $\mathbf{c}$  τοποθετημένων τη μία δίπλα στην άλλη. Η κατάλληλη σειρά τοποθέτησής τους αποφασίζεται από τη μέγιστη ακολουθία όλων των δυνατών διαφορών ολίσθησης. Αφού η ‘ακολουθία διαφορών ολίσθησης’ ληφθεί, η ‘ακολουθία ολισθήσεων’ λαμβάνεται θέτοντας την πρώτη ολίσθηση ίση για παράδειγμα με 0, δηλ.  $\mathbf{b}_1 \equiv \mathbf{c}_0$ . Με αυτόν τον τρόπο, λαμβάνονται όλες οι στήλες του και πρακτικά ο  $\mathbf{B}$ .

### 8.3.2 Ο αλγόριθμος κατασκευής

Η κατασκευή του δυαδικού χάρτη του κυανού πετάσματος μπορεί να δοθεί περιληπτικά σε μερικά αλγοριθμικά βήματα. Αυτό πρακτικά δίνει σε κάποιον τη δυνατότητα να κατασκευάσει ένα κυανό πέτασμα με σχεδόν οποιαδήποτε επιθυμητή διάσταση.

- Αποφάσισε το μέγεθος  $n \times m$  του μικρότερου υποπίνακα του  $\mathbf{B}$  που είναι ορατός από την κάμερα.
- Βρες ένα αρχέτυπο πολυώνυμο βαθμού  $n$  στο  $\text{GF}(2^n)$  και υπολόγισε τον αντίστοιχο κώδικα μέγιστου μήκους  $\mathbf{c}$ .
- Θέσε μία τυχαία περίοδο του  $\mathbf{c}$  ( $\mathbf{c}_0$ ) ως πρώτη στήλη στον πίνακα  $\mathbf{B}$ .
- Βρες ένα αρχέτυπο πολυώνυμο βαθμού  $2^n - 1$  στο  $\text{GF}((2^n - 1)^{m-1})$  και υπολόγισε τον αντίστοιχο κώδικα μέγιστου μήκους. Η ακολουθία των διαφορών ολίσθησης έχει ληφθεί.

$d_j$		0	1	1	2	3	1	...
$s$	1	1	2	3	5	1	2	...

**Πίνακας 8.1** Κατασκευάζοντας το δυαδικό χάρτη του κυανού πετάσματος, άνω γραμμή: η ακολουθία των διαφορών των μετατοπίσεων, κάτω γραμμή: η ακολουθία των μετατοπίσεων.

- Αρχίζοντας από την πρώτη στήλη του  $\mathbf{B}$ , ακολουθιακά υπολόγισε όλες τις στήλες ώστε να είναι ολισθημένες εκδοχές του  $\mathbf{c}_0$ , ως  $\mathbf{b}_{j+1} = \mathbf{c}_{r+d_j}$  όταν  $\mathbf{b}_j \equiv \mathbf{c}_r$ . Η ακολουθία των ολισθήσεων και ολόκληρος ο δυαδικός χάρτης έχουν πλέον ληφθεί.

Με στόχο να αποσαφηνίσουμε την προτεινόμενη μεθοδολογία, ο Πίνακας 8.1 περιέχει ένα τμήμα από τη λαμβανόμενη ακολουθία με τις διαφορές ολίσθησης και το αντίστοιχο τμήμα των εξαγόμενων ολισθήσεων, για ορατό τμήμα μεγέθους  $3 \times 3$ . Πρέπει να σημειωθεί ότι σε αυτό το παράδειγμα, στη στήλη των ολισθήσεων ως εναρκτήριο ψηφίο χρησιμοποιήθηκε το '1' ( $\mathbf{c}_1$ ) και όλες οι πράξεις πραγματοποιήθηκαν με αριθμητική modulo-7 (γενικά modulo- $(2^n - 1)$ ).

Είναι ενδιαφέρον να υπολογίσει κανείς τις τελικές διαστάσεις του δυαδικού πίνακα, που παράγεται ακολουθώντας τα βήματα της προτεινόμενης μεθοδολογίας, δηλ. δεδομένου ότι κάθε  $n \times m$  πρότυπο πρέπει να είναι μοναδικό στον  $\mathbf{B}$ , να βρούμε τα μέγιστα επιτρεπόμενα μήκη  $N$  και  $M$ . Αφού κάθε στήλη αντιστοιχεί σε  $2^n - 1$  καταστάσεις, περιέχει  $2^n - 1 + (n - 1) = 2^n + n - 2$  τετράπλευρα, αφού καταργήσουμε την κυκλική ιδιότητα του  $\mathbf{c}_s$  σε κάθε στήλη. Με αυτόν τον τρόπο, ορίζουμε το  $N$  ως  $N \triangleq 2^n + n - 2$ . Όμοια και σύμφωνα με τα παραπάνω,  $(2^n - 1)^{m-1} - 1$  καταστάσεις λαμβάνονται από τις διαφορές ολίσθησης, ενώ καταργώντας την κυκλική ιδιότητα, οι διαφορές ολίσθησης αυξάνονται στις  $(2^n - 1)^{m-1} + m - 2$ . Το  $M$  δίνεται από τις ολισθήσεις, οπότε ορίζεται ως  $M \triangleq (2^n - 1)^{m-1} + m - 1$ . Μπορεί να παρατηρηθεί εδώ ότι η παραπάνω μεθοδολογία θα μπορούσε να χρησιμοποιηθεί εναλλακτικά για να παράγει έναν πίνακα  $M \times N$ , θέτοντας την πρώτη γραμμή του  $\mathbf{B}$  και έπειτα υπολογίζοντας την ακολουθία των διαφορών ολίσθησης για τις γραμμές του ίδιου πίνακα.

Η παραπάνω θεωρία προσφέρει μία καλώς-ορισμένη και αποδοτική μέθοδο κατασκευής του δυαδικού χάρτη ενός κυανού πετάσματος, που να περιέχει μοναδικά πρότυπα προκαθορισμένου μεγέθους. Μερικά επιπλέον πλεονεκτήματα της μεθοδολογίας εξετάζονται ώστε να παραχθούν λύσεις με δυνατότητες διόρθωσης σφάλματος (error correction), στην περίπτωση που το χρώμα ενός παραλληλογράμμου του πετάσματος κρίνεται εσφαλμένα, λόγω πρακτικών προβλημάτων πραγματικού χρόνου όπως ο φωτισμός, οι σκιές ή η μερική επικάλυψη. Μπορεί να δει κανείς ότι η προτεινόμενη μεθοδολογία παράγει  $M \gg N$ . Με αυτό το σκεπτικό, ένα υποσύνολο των διαθέσιμων, για παράδειγμα, στηλών είναι αρκετό να χρησιμοποιηθεί. Κατάλληλη επιλογή του υποσυνόλου αυτού μπορεί να επιτρέψει διαδικασίες εντοπισμού και διόρθωσης σφάλματος (error detection and correction).

## 8.4 Υπολογισμός της θέσης και του προσανατολισμού της κάμερας

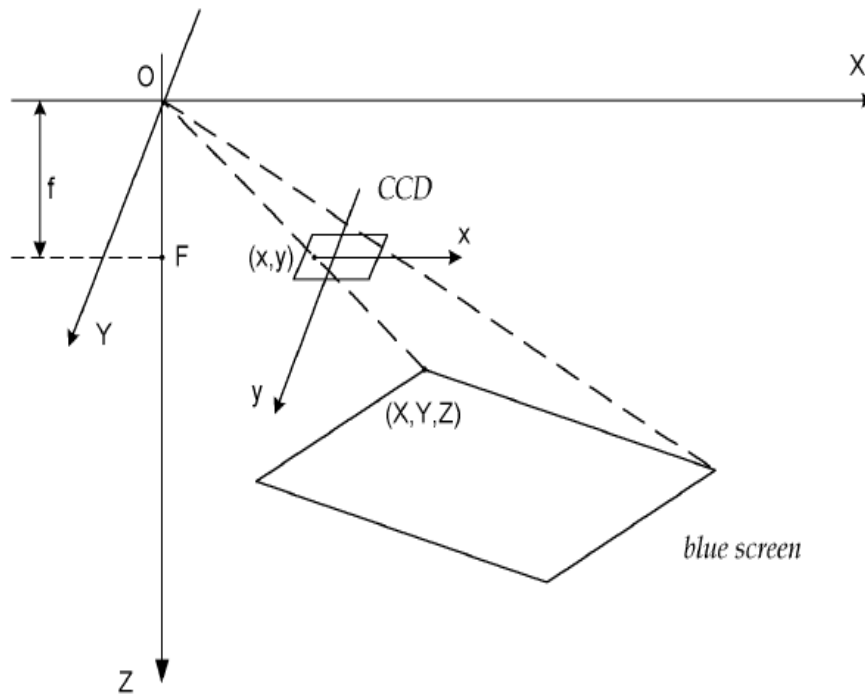
Σύμφωνα με το σενάριο που περιγράψαμε στην Ενότητα 8.2, η άγνωστη κίνηση της κάμερας για κάθε μετάβαση πρέπει να υπολογιστεί βάσει της γνωστής 3Δ δομής της σκηνής αναφοράς και της προβεβλημένης 2Δ δομής στο τρέχον καρέ. Αυτό αποτελεί ένα πρόβλημα υπολογισμού 3Δ κίνησης με βάση 3Δ-σε-2Δ αντιστοιχίσεις χαρακτηριστικών. Στον προτεινόμενο αλγόριθμο, τα χαρακτηριστικά επιλέγονται να είναι ευθείες γραμμές, δηλ. οι γραμμές αναφοράς που γίνονται εμμέσως διαθέσιμες από την κατασκευή του δυαδικού χάρτη πάνω στο κυανό πέτασμα.

Όπως εξηγείται στην Ενότητα 8.2, η μερική επικάλυψη ή ολική αλλαγή των χαρακτηριστικών εντός του οπτικού πεδίου της κάμερας μεταξύ μεταβάσεων καθιστά ακατάλληλη οποιαδήποτε τεχνική υπολογισμού κίνησης, παρακολούθησης χαρακτηριστικών ή τοπικής συσχέτισης. Αντιθέτως, θεωρώντας την 3Δ θέση της κάμερας ως προς το επίπεδο του πετάσματος αντί της 3Δ κίνησης της κάμερας, ο υπολογισμός των παραμέτρων 3Δ κίνησης της κάμερας μπορεί να πραγματοποιηθεί αποδοτικά. Για το μοντέλο προοπτικής προβολής, δύο όψεις περιέχουν ικανή πληροφορία για δεδομένες 3Δ-σε-2Δ αντιστοιχίσεις χαρακτηριστικών [83], ενώ χρησιμοποιώντας 2Δ-σε-2Δ αντιστοιχίσεις γραμμών τρεις όψεις είναι αναγκαίες [149]. Αυτό υποδηλώνει ότι η επιλογή των γραμμών ως χαρακτηριστικών για ‘παρακολούθηση’ μεταξύ δύο όψεων απαιτεί γνώση των 3Δ παραμέτρων των γραμμών τουλάχιστον σε μία από τις δύο διαθέσιμες όψεις. Με αυτό το σκεπτικό, έχοντας τις 2Δ παραμέτρους των γραμμών από το τρέχον καρέ, η 3Δ θέση της κάμερας μπορεί να υπολογιστεί θεωρώντας μία τυχαία 3Δ σκηνή αναφοράς όπου οι 3Δ παράμετροι των γραμμών, για τις ίδιες γραμμές, είναι γνωστές.

Ακόμη και σε αυτήν την περίπτωση, αναφέρεται στην [83] ότι γενικά οκτώ ή περισσότερες αντιστοιχίσεις γραμμών είναι απαραίτητες ώστε να εκτιμηθεί ο πίνακας περιστροφής με χρήση ενός γραμμικού αλγορίθμου, ενώ τρεις ή περισσότερες αντιστοιχίσεις είναι απαραίτητες χρησιμοποιώντας μία μη-γραμμική προσέγγιση. Από την άλλη πλευρά, θαδειχθεί ότι τοποθετώντας κατάλληλα τις γραμμές στη σκηνή αναφοράς (στο πέτασμα), τέσσερις αντιστοιχίσεις γραμμών περιέχουν ικανή πληροφορία για τον υπολογισμό του πίνακα περιστροφής με χρήση ενός γραμμικού αλγορίθμου. Στην πραγματικότητα, ακριβής αντιστοίχιση γραμμών μία-προς-μία δεν είναι αναγκαία στην περίπτωση αυτή. Η πραγματοποίηση των αντιστοιχίσεων των 3Δ γραμμών αναφοράς με τις 2Δ γραμμές στο τρέχον καρέ, δίνει τη δυνατότητα υπολογισμού με ακρίβεια του διανύσματος μετατόπισης.

Ας εξετάσουμε τώρα την κατασκευή της σκηνής αναφοράς. Όπως αναφέρθηκε ήδη στην Ενότητα 8.2, η κατασκευή αυτή είναι στην πραγματικότητα εικονική, υπό την έννοια ότι δεν





**Σχήμα 8.2** Η γεωμετρία του προβλήματος. Το σύστημα συντεταγμένων κινείται με το επίπεδο του CCD (camera-based).

χρειάζεται καμία ρύθμιση της κάμερας στην σκηνή αυτή. Έστω ότι το επίπεδο προβολής (το παραλληλόγραμμο CCD) είναι παράλληλο στο  $XY$  επίπεδο του καρτεσιανού συστήματος συντεταγμένων σε βάθος  $Z = f$ , όπου το  $f$  υποδηλώνει το εστιακό μήκος. Έστω επίσης ότι το εστιακό κέντρο συμπίπτει με την αρχή των αξόνων  $(0, 0, 0)$  και ότι το κυανό πέτασμα είναι παράλληλο στο επίπεδο προβολής σε βάθος  $Z = z_0$ . Υποθέτοντας ότι το σύστημα συντεταγμένων κινείται μαζί με το παραλληλόγραμμο CCD, το παρακάτω απλοποιημένο μοντέλο προοπτικής προβολής ισχύει [155],

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (8-1)$$

όπου το  $(x, y)$  δίνει τις καρτεσιανές συντεταγμένες του σημείου  $(X, Y, Z)$  όταν προβληθεί στο επίπεδο προβολής. Η γεωμετρία του προβλήματος δίνεται στο Σχήμα 8.2. Το παγκόσμιο σύστημα συντεταγμένων είναι 'δεμένο' στο επίπεδο του CCD, ενώ το κυανό πέτασμα λαμβάνει τυχαίες θέσεις και προσανατολισμούς στον 3Δ χώρο. Στοχεύοντας στην απλοποίηση του μοντέλου προβολής και των σχετιζόμενων εξισώσεων [155], θεωρείται συχνά ότι το εστιακό μήκος είναι γνωστό και τίθεται ίσο με τη μονάδα ( $f = 1$ ). Αντιθέτως, όπως θα δειχθεί στη συνέχεια, στην περίπτωση μας το  $f$  μπορεί να υπολογιστεί ασφαλώς λύνοντας ένα σύστημα από γραμμικές εξισώσεις.

Η δομή του κυανού πετάσματος στη σκηνή αναφοράς είναι γνωστή και μπορεί να αναλυθεί στις εξής παραμέτρους,

- (1) το 3Δ βάθος του πετάσματος  $z_0$ ,
- (2) ένα σύνολο από ‘κατακόρυφες’ γραμμές  $X_v = \{X = x_i, i = 1 \cdots N\}$  και
- (3) ένα σύνολο από ‘οριζόντιες’ γραμμές  $Y_h = \{Y = y_j, j = 1 \cdots M\}$ .

Αυτά τα δύο σύνολα από 3Δ γραμμές στη σκηνή αναφοράς προβάλλονται σε δύο αντίστοιχα σύνολα 2Δ γραμμών στο τρέχον καρέ, τα  $E_v$  και  $E_h$  αντίστοιχα. Γενικά, τα σύνολα  $E_v$  και  $E_h$  περιέχουν 2Δ παραμέτρους γραμμών της μορφής  $(a, b)$ . Στη συνέχεια θα κάνουμε χρήση της καρτεσιανής αναπαράστασης των γραμμών στο τρέχον καρέ δηλ.  $y' = \alpha x' + \beta$ , όπου το  $(x', y')$  υποδηλώνει τις καρτεσιανές συντεταγμένες ενός σημείου στο τρέχον καρέ για ένα αντίστοιχο προβεβλημένο σημείο του χώρου  $(X', Y', Z')$  στην άγνωστη τρέχουσα 3Δ σκηνή. Όπως ορίστηκαν παραπάνω, τα  $(x, y)$  και  $(X, Y, Z)$  υποδηλώνουν τα αντίστοιχά τους στο καρέ αναφοράς και στην εικονική σκηνή αναφοράς.

Θα θεωρηθεί αρχικά ότι το εστιακό μήκος  $f$  είναι δεδομένο. Έτσι, ορίζουμε τις 2Δ παραμέτρους γραμμών  $(a, b)$  στα σύνολα  $E_v$  και  $E_h$  ως  $(a, b) = (\alpha, \frac{\beta}{f})$ . Συνδυάζοντας το τελευταίο με την εξίσωση (8-1), λαμβάνουμε

$$\begin{bmatrix} a & -1 & b \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = 0. \quad (8-2)$$

Το διάνυσμα  $\epsilon = [a \ -1 \ b]^T$  χρησιμοποιείται συχνά στα προβλήματα της Όρασης Υπολογιστών ως μία βολική διανυσματική αναπαράσταση γραμμής [149]. Όπως θα δειχθεί στη συνέχεια, η χρήση του μοναδιαίου διανύσματος  $\epsilon = \frac{\epsilon}{\|\epsilon\|}$  σ' αυτήν την κατεύθυνση επιτρέπει την εξαγωγή απλούστερων και περισσότερο συμπαγών εξισώσεων για τον υπολογισμό της 3Δ κίνησης της κάμερας. Με αυτό το σκεπτικό, ορίζουμε το σύνολο  $E_v = \{\epsilon_{vi}, i = 1 \cdots N\}$ , όπου  $\epsilon_{vi} = \frac{\epsilon_{vi}}{\|\epsilon_{vi}\|}$ ,  $\epsilon_{vi} = [a_{vi} \ -1 \ b_{vi}]^T$  και όπου η 2Δ γραμμή  $y' = a_{vi}x' + b_{vi}$  στο τρέχον καρέ αντιστοιχεί στην 3Δ γραμμή  $X = x_i$  στη σκηνή αναφοράς. Όμοια,  $E_h = \{\epsilon_{hj}, j = 1 \cdots M\}$ , με  $\epsilon_{hj} = \frac{\epsilon_{hj}}{\|\epsilon_{hj}\|}$  και  $\epsilon_{hj} = [a_{hj} \ -1 \ b_{hj}]^T$ , όταν η γραμμή  $y' = a_{hj}x' + b_{hj}$  αντιστοιχεί στην γραμμή  $Y = y_j$ .

Η κίνηση μίας γραμμής στον 3Δ χώρο δίνεται ως μία υπέρθεση μίας 3Δ περιστροφής και μίας 3Δ μετατόπισης. Έστω  $\mathbf{R}$  και  $\mathbf{T}$  ο  $3 \times 3$  πίνακας περιστροφής και το  $3 \times 1$  διάνυσμα μετατόπισης αντίστοιχα, τα οποία πρέπει να υπολογιστούν για να εκτιμήσουμε την 3Δ κίνηση. Τότε, η παρακάτω πρόταση υποδηλώνει ότι ο  $\mathbf{R}$  μπορεί να υπολογιστεί από όλα τα γνωστά στοιχεία των  $E_v, E_h$  χωρίς να είναι δεδομένη η ακριβής αντιστοίχισή τους στα στοιχεία των  $X_v, Y_h$ .

**Πρόταση 8.2:** Έστω  $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$  τα διανύσματα στήλες του πίνακα περιστροφής, δηλ.

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} \quad (8-3)$$

και ας ορίσουμε τον  $(n + m) \times 3$  πίνακα  $\mathbf{Q}$ , που κατασκευάζεται από  $n$  και  $m$  εξωτερικά γινόμενα προερχόμενα από τα σύνολα  $E_v$  και  $E_h$  αντίστοιχα, ως

$$\mathbf{Q} = \begin{bmatrix} \boldsymbol{\varepsilon}_{v1} \times \boldsymbol{\varepsilon}_{v2} & \boldsymbol{\varepsilon}_{v3} \times \boldsymbol{\varepsilon}_{v4} & \cdots & \boldsymbol{\varepsilon}_{h1} \times \boldsymbol{\varepsilon}_{h2} & \boldsymbol{\varepsilon}_{h3} \times \boldsymbol{\varepsilon}_{h4} & \cdots \end{bmatrix}^T. \quad (8-4)$$

Τότε,

(α) Το εξωτερικό γινόμενο οποιωνδήποτε δύο στοιχείων του  $E_v$  ( $E_h$ ) είναι σταθερό και ίσο με  $\pm \mathbf{r}_2$  ( $\pm \mathbf{r}_1$ ).

(β) Έστω  $\mathbf{Q} = \mathcal{U} \mathcal{S} \mathcal{V}^T$  η ανάλυση σε ειδικές τιμές του πίνακα  $\mathbf{Q}$ , με  $\mathcal{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$  και  $\text{diag}(\mathcal{S}) = [s_1 \ s_2 \ s_3]$ . Τότε  $s_1 = \max(\sqrt{n}, \sqrt{m})$ ,  $s_2 = \min(\sqrt{n}, \sqrt{m})$  και  $s_3 = 0$ . Επιπρόσθετα, εάν  $s_1 \equiv \sqrt{m}$ , τότε  $\mathbf{r}_1 = \mathbf{v}_1$ ,  $\mathbf{r}_2 = \mathbf{v}_2$  και  $\mathbf{r}_3 = \mathbf{v}_3$ , ενώ εάν  $s_1 \equiv \sqrt{n}$ , τότε  $\mathbf{r}_1 = \mathbf{v}_2$ ,  $\mathbf{r}_2 = \mathbf{v}_1$  και  $\mathbf{r}_3 = \mathbf{v}_3$ .

**Απόδειξη:**

(α) Η 3Δ μετακίνηση ενός σημείου  $(X, Y, Z)$  σε ένα σημείο  $(X', Y', Z')$  δίνεται από την εξίσωση

$$\begin{bmatrix} X' & Y' & Z' \end{bmatrix}^T = \mathbf{R} \begin{bmatrix} X & Y & Z \end{bmatrix}^T + \mathbf{T}. \quad (8-5)$$

Για μία κατακόρυφη γραμμή  $\boldsymbol{\varepsilon}_{vi}$  που ανήκει στο  $E_v$ , για σταθερά  $x_i, z_0$  και μεταβλητό  $Y$ ,  $\begin{bmatrix} X & Y & Z \end{bmatrix}^T \equiv \begin{bmatrix} x_i & Y & z_0 \end{bmatrix}^T$ . Χρησιμοποιώντας την εξίσωση (8-2), λαμβάνουμε

$$\boldsymbol{\varepsilon}_{vi}^T \begin{bmatrix} X' & Y' & Z' \end{bmatrix}^T = 0.$$

Από την εξίσωση (8-5), έχουμε

$$\boldsymbol{\varepsilon}_{vi}^T (\mathbf{R} \begin{bmatrix} x_i & Y & z_0 \end{bmatrix}^T + \mathbf{T}) \equiv \boldsymbol{\varepsilon}_{vi}^T (x_i \mathbf{r}_1 + Y \mathbf{r}_2 + z_0 \mathbf{r}_3 + \mathbf{T}) = 0.$$

Εφόσον η προηγούμενη εξίσωση ισχύει για κάθε  $Y$  στην  $\boldsymbol{\varepsilon}_{vi}$ ,

$$\begin{aligned} \boldsymbol{\varepsilon}_{vi}^T \mathbf{r}_2 &= 0 \quad \text{και} \quad \text{επίσης} \\ \boldsymbol{\varepsilon}_{vi}^T (x_i \mathbf{r}_1 + z_0 \mathbf{r}_3 + \mathbf{T}) &= 0. \end{aligned} \quad (8-6)$$

Έτσι, για κάθε ζεύγος γραμμών στο  $E_v$  (για παράδειγμα  $\boldsymbol{\varepsilon}_{v1}$  και  $\boldsymbol{\varepsilon}_{v2}$ ), εφόσον  $\boldsymbol{\varepsilon}_{v1} \perp \mathbf{r}_2$  και  $\boldsymbol{\varepsilon}_{v2} \perp \mathbf{r}_2$ , και το  $\boldsymbol{\varepsilon}_{v1}$  δεν είναι συνευθειακό με το  $\boldsymbol{\varepsilon}_{v2}$ ,

$$\boldsymbol{\varepsilon}_{v1} \times \boldsymbol{\varepsilon}_{v2} = \pm \mathbf{r}_2. \quad (8-7)$$

Η ισότητα ισχύει αφού από τον ορισμό τους  $\|\mathbf{e}_{v1} \times \mathbf{e}_{v2}\| = 1$ . Με όμοιο τρόπο, για κάθε γραμμή στο  $E_h$

$$\begin{aligned}\mathbf{e}_{hj}^T \mathbf{r}_1 &= 0 \quad \text{και} \quad \text{επίσης} \\ \mathbf{e}_{hj}^T (y_j \mathbf{r}_2 + z_0 \mathbf{r}_3 + \mathbf{T}) &= 0 ,\end{aligned}\tag{8-8}$$

και για κάθε ζεύγος γραμμών στο  $E_h$  (για παράδειγμα  $\mathbf{e}_{h1}$  και  $\mathbf{e}_{h2}$ ),

$$\mathbf{e}_{h1} \times \mathbf{e}_{h2} = \pm \mathbf{r}_1 .\tag{8-9}$$

(β) Από τον ορισμό του πίνακα  $\mathbf{Q}$  στην εξίσωση (8-4) και σύμφωνα με το μέρος (α),

$$\mathbf{Q}^T \mathbf{Q} = \sum_{k=1}^m \mathbf{r}_1 \mathbf{r}_1^T + \sum_{k=1}^n \mathbf{r}_2 \mathbf{r}_2^T = m \mathbf{r}_1 \mathbf{r}_1^T + n \mathbf{r}_2 \mathbf{r}_2^T + 0 \mathbf{r}_3 \mathbf{r}_3^T$$

και τα  $\mathbf{r}_1$ ,  $\mathbf{r}_2$ ,  $\mathbf{r}_3$  αποτελούν ένα σύνολο από ορθομοναδιαία δεξιά ειδικά διανύσματα (right singular vectors) του  $\mathbf{Q}$  που αντιστοιχούν στις ειδικές τιμές (singular values)  $\sqrt{m}$ ,  $\sqrt{n}$  και 0 αντίστοιχα. Τα δεξιά ειδικά διανύσματα βρίσκονται στον πίνακα  $\mathcal{V}$  με την ίδια σειρά, με την οποία βρίσκονται οι ειδικές τιμές στον  $\mathcal{S}$ .

Θα πρέπει να τονιστεί σε αυτό το σημείο ότι τα  $\mathbf{r}_1$ ,  $\mathbf{r}_2$ ,  $\mathbf{r}_3$  υπολογίζονται με μία αβεβαιότητα προσήμου. Από την άλλη μεριά, μπορεί να αποδειχθεί εύκολα ότι τα κατάλληλα πρόσημα υπολογίζονται από το γεγονός ότι αποτελούν στήλες πίνακα περιστροφής και παράλληλα από το ότι στην πράξη η αποδεκτή γωνία περιστροφής περιορίζεται στο διάστημα  $[-90^\circ, 90^\circ]$ . Ο.Ε.Δ.

Χάριν απλότητας, στην εξίσωση (8-4) υποθέτουμε ότι κάθε γραμμή υπεισέρχεται στον πίνακα  $\mathbf{Q}$  μόνο μία φορά. Λαμβάνοντας μία γραμμή περισσότερες από μία φορές, δεν δημιουργεί ωστόσο κανένα πρόβλημα στην παρακάτω ανάλυση.

Πρακτικά, η Πρόταση 8.2 υποδηλώνει ότι αφού του δύο σύνολα από γραμμές εξαχθούν στο τρέχον καρέ, τα οποία και αντιστοιχούν στις κατακόρυφες και οριζόντιες γραμμές της σκηνής αναφοράς, ο πίνακας περιστροφής υπολογίζεται βάσει του SVD ενός απλού πίνακα  $\mathbf{Q}$  (στην πραγματικότητα, λύνοντας ένα ομογενές γραμμικό σύστημα με τις άγνωστες παραμέτρους περιστροφής). Εφόσον η πληροφορία αντιστοίχισης δεν είναι αναγκαία, όλες οι διαθέσιμες γραμμές μπορούν να χρησιμοποιηθούν για τον υπολογισμό του  $\mathbf{R}$ . Αντιθέτως, από τις εξισώσεις (8-6) και (8-8), μπορεί κανείς να δει ότι ο υπολογισμός του διανύσματος μετατόπισης απαιτεί γνώση της αντιστοίχισης των γραμμών στην σκηνή αναφοράς και στο τρέχον καρέ, το οποίο αναμένεται και διαισθητικά. Στη συνέχεια, θα δειχθεί ότι η γνώση του  $\mathbf{R}$ , σε συνδυασμό με τη στρατηγική που ακολουθείται στην κατασκευή του κυανού πετάσματος, επιτρέπει και την εύρεση των αντιστοιχίσεων γραμμών και τον υπολογισμό του  $\mathbf{T}$ .

Όπως αναφέρθηκε στην Ενότητα 8.3, οι αντιστοιχίσεις γραμμών μεταξύ της σκηνής αναφοράς και του τρέχοντος καρέ μπορούν να πραγματοποιηθούν υπολογίζοντας τη θέση του ορατού τμήματος του πετάσματος πάνω στον αντίστοιχο δυαδικό χάρτη. Όπως όμως αναφέρθηκε και στην Ενότητα 8.2, έχοντας εξάγει όλες τις διαθέσιμες γραμμές του τρέχοντος καρέ και έχοντας εντοπίσει το μέγεθος του στοιχειώδους τετραπλεύρου (για να ξεπεραστεί η αβεβαιότητα μεγέθυνσης), παραμένει το πρόβλημα της αναγνώρισης του δυαδικού προτύπου όταν απουσιάζουν γραμμές στο περίγραμμα γειτονικών τετραπλεύρων του ίδιου χρωματικού τόνου. Έχοντας υπολογίσει τον πίνακα  $\mathbf{R}$ , πιθανή απουσία γραμμών στο τρέχον καρέ μπορεί να εντοπιστεί από την επόμενη πρόταση.

**Πρόταση 8.3:** Έστω  $\epsilon_{v1}, \epsilon_{v2}$  δύο τυχαία στοιχεία του  $E_v$  που να αντιστοιχούν στα στοιχεία  $x_1$  και  $x_2$  αντίστοιχα του  $X_v$ , για τα οποία μόνο το  $d_{21} = x_2 - x_1$  είναι γνωστό. Για ένα άλλο δεδομένο στοιχείο  $\epsilon_{v3}$  του  $E_v$ , η ποσότητα  $d_{31} = x_3 - x_1$  δίνεται από τη σχέση

$$d_{31} = d_{21} \frac{\left\| \frac{\epsilon_{v3}^T}{\epsilon_{v3}^T \mathbf{r}_1} - \frac{\epsilon_{v1}^T}{\epsilon_{v1}^T \mathbf{r}_1} \right\|}{\left\| \frac{\epsilon_{v2}^T}{\epsilon_{v2}^T \mathbf{r}_1} - \frac{\epsilon_{v1}^T}{\epsilon_{v1}^T \mathbf{r}_1} \right\|}. \quad (8-10)$$

Η ίδια εξίσωση ισχύει για οποιεσδήποτε τρεις γραμμές του  $E_h$  αντικαθιστώντας τους δείκτες  $v$  με  $h$  και το  $\mathbf{r}_1$  με  $\mathbf{r}_2$ .

**Απόδειξη:**

Για οποιεσδήποτε τρεις γραμμές του  $E_v$ , η δεύτερη εξίσωση της (8-6) δίνει

$$\begin{aligned} x_2 - x_1 &= \phi_2^T \mathbf{m} \\ x_3 - x_1 &= \phi_3^T \mathbf{m}, \end{aligned} \quad (8-11)$$

όπου  $\mathbf{m} \equiv -z_0 \mathbf{r}_3 - \mathbf{T}$ ,  $\phi_2 \equiv (\frac{\epsilon_{v2}^T}{\epsilon_{v2}^T \mathbf{r}_1} - \frac{\epsilon_{v1}^T}{\epsilon_{v1}^T \mathbf{r}_1})^T$  και  $\phi_3 \equiv (\frac{\epsilon_{v3}^T}{\epsilon_{v3}^T \mathbf{r}_1} - \frac{\epsilon_{v1}^T}{\epsilon_{v1}^T \mathbf{r}_1})^T$ , για ευκολία στο συμβολισμό.

Όμως, από την εξίσωση (8-7), έχουμε  $\epsilon_{v1} \times \epsilon_{v2} = \mathbf{r}_2$ , με  $\mathbf{r}_2 \perp \mathbf{r}_1$  και  $\mathbf{r}_2 \perp \mathbf{r}_3$ . Οπότε εξορισμού,  $\phi_2 \in \text{span}(\mathbf{r}_1, \mathbf{r}_3)$ . Ταυτόχρονα,

$$\phi_2^T \mathbf{r}_1 = (\frac{\epsilon_{v2}^T}{\epsilon_{v2}^T \mathbf{r}_1} - \frac{\epsilon_{v1}^T}{\epsilon_{v1}^T \mathbf{r}_1}) \mathbf{r}_1 = \frac{\epsilon_{v2}^T \mathbf{r}_1}{\epsilon_{v2}^T \mathbf{r}_1} \frac{\epsilon_{v1}^T \mathbf{r}_1}{\epsilon_{v1}^T \mathbf{r}_1} = 0.$$

Επομένως,  $\phi_2 \perp \mathbf{r}_1$  και τελικά  $\phi_2 \parallel \mathbf{r}_3$ . Οι ίδιες εξισώσεις ισχύουν για το  $\phi_3$ . Έτσι,  $\frac{\phi_2}{\|\phi_2\|} = \frac{\phi_3}{\|\phi_3\|} = \mathbf{r}_3$ . Επομένως,  $\phi_3^T \mathbf{m} = \frac{\|\phi_3\|}{\|\phi_2\|} \phi_2^T \mathbf{m}$ , ή,  $x_3 - x_1 = \frac{\|\phi_3\|}{\|\phi_2\|} (x_2 - x_1)$  το οποίο ολοκληρώνει την απόδειξη της εξίσωσης (8-10). Όμοια αποτελέσματα μπορούν να ληφθούν για γραμμές του  $E_h$  όπου τώρα χρησιμοποιείται η εξίσωση (8-8) και αντίστοιχα ισχύει  $\phi \in \text{span}(\mathbf{r}_2, \mathbf{r}_3)$ .

Ο.Ε.Δ.

Η Πρόταση 8.3 παρέχει έναν αποδοτικό τρόπο για το διαχωρισμό του ορατού μέρους του πετάσματος σε στοιχειώδη τετράπλευρα και την εξαγωγή του αντίστοιχου δυαδικού προτύπου. Αναγνωρίζοντας το τελευταίο επάνω στο δυαδικό χάρτη του πετάσματος, η αντιστοίχιση των γραμμών γίνεται δυνατή. Με άλλα λόγια και σύμφωνα με την ορολογία που έχουμε υιοθετήσει, για κάθε γνωστό στοιχείο του  $E_v$  (ή του  $E_h$ ) το αντίστοιχό του στοιχείο στο  $X_v$  (ή στο  $Y_h$ ) έχει πλέον βρεθεί.

Διαισθητικά μιλώντας, εάν ξεχάσουμε το δυαδικό πρότυπο του πετάσματος και υποθέσουμε ότι μόνο οι γραμμές αναφοράς ήταν διαθέσιμες στο καταγραφόμενο καρέ, η πληροφορία που θα χάναμε θα ήταν βασικά αυτή της 3Δ μετατόπισης. Με άλλα λόγια, θα ήταν τότε δυνατό να εξάγουμε την 3Δ περιστροφή αλλά δεν θα είχαμε καμία πληροφορία για την μετατόπιση, αφού οι γραμμές αναφοράς είναι πανομοιότυπες σε κάθε περιοχή του πετάσματος. Σύμφωνα με τα παραπάνω, η προτεινόμενη μεθοδολογία για την κατασκευή του κυανού πετάσματος όχι μόνο μας παρέχει εμμέσως ένα σύνολο από γραμμές αναφοράς για τον υπολογισμό της 3Δ περιστροφής αλλά επίσης επιτρέπει τον υπολογισμό και της 3Δ μετατόπισης. Θα πρέπει ακόμη να σημειωθεί εδώ, ότι για τον υπολογισμό της 3Δ μετατόπισης, η πληροφορία βάθους στη σκηνή αναφοράς είναι αναγκαία ( $z_0$ ). Γι' αυτό το λόγο, εάν είχαμε χρησιμοποιήσει 2Δ-σε-2Δ αντιστοιχίσεις γραμμών, μόνο ο υπολογισμός της κατεύθυνσης της μετατόπισης (ως ένα μοναδιαίο διάνυσμα) θα ήταν δυνατός (βλ. για παράδειγμα την [163]). Με αυτό το σκεπτικό, το διάνυσμα της 3Δ μετατόπισης  $\mathbf{T}$  υπολογίζεται από την Πρόταση 8.4.

**Πρόταση 8.4:** Έστω ότι τα  $\epsilon_{v1}, \dots, \epsilon_{vn}$  είναι  $n$  γνωστά στοιχεία του  $E_v$  που αντιστοιχούν στα  $x_1, \dots, x_n$  του  $X_v$ . Έστω επίσης ότι τα  $\epsilon_{h1}, \dots, \epsilon_{hm}$  είναι  $m$  γνωστά στοιχεία του  $E_h$  που αντιστοιχούν στα  $y_1, \dots, y_m$  του  $Y_h$ . Ορίζοντας τους  $3 \times 3$  πίνακες  $\mathbf{V} = \sum_{i=1}^n x_i \epsilon_{vi} \epsilon_{vi}^T$ ,  $\mathbf{H} = \sum_{j=1}^m y_j \epsilon_{hj} \epsilon_{hj}^T$ , και  $\mathbf{U} = \sum_{i=1}^n \epsilon_{vi} \epsilon_{vi}^T + \sum_{j=1}^m \epsilon_{hj} \epsilon_{hj}^T$ , μία εκτίμηση ελαχίστων τετραγώνων για το διάνυσμα μετατόπισης  $\mathbf{T}$  δίνεται από τη σχέση

$$\mathbf{T} = -\mathbf{U}^{-1} \mathbf{V} \mathbf{r}_1 - \mathbf{U}^{-1} \mathbf{H} \mathbf{r}_2 - z_0 \mathbf{r}_3 . \quad (8-12)$$

**Απόδειξη:**

Έστω  $\mathbf{m} = -z_0 \mathbf{r}_3 - \mathbf{T}$ . Από τις εξισώσεις (8-6) και (8-8) έχουμε

$$\begin{aligned} \epsilon_{vi}^T \mathbf{m} &= x_i \epsilon_{vi}^T \mathbf{r}_1 \\ \epsilon_{hj}^T \mathbf{m} &= y_j \epsilon_{hj}^T \mathbf{r}_2 . \end{aligned} \quad (8-13)$$

Λαμβάνοντας  $n$  και  $m$  τέτοιες εξισώσεις αντίστοιχα, καταλήγουμε σε ένα υπερορισμένο γραμμικό σύστημα της μορφής  $\mathcal{M}\mathbf{m} = \mathcal{N}$ , όπου  $\mathcal{M} = \begin{bmatrix} \epsilon_{v1} & \epsilon_{v2} & \cdots & \epsilon_{h1} & \epsilon_{h2} & \cdots \end{bmatrix}^T$ , ενώ  $\mathcal{N} = \begin{bmatrix} x_1\epsilon_{v1} & x_2\epsilon_{v2} & \cdots & y_1\epsilon_{h1} & y_2\epsilon_{h2} & \cdots \end{bmatrix}^T$ , που οδηγεί σε μία λύση ελαχίστων τετραγώνων της μορφής  $\mathbf{m} = (\mathcal{M}^T \mathcal{M})^{-1} (\mathcal{M}^T \mathcal{N})$ .

Μπορεί να επιβεβαιωθεί με πράξεις ότι  $\mathcal{M}^T \mathcal{M} \equiv \mathbf{U}$  και  $\mathcal{M}^T \mathcal{N} \equiv \mathbf{V}\mathbf{r}_1 + \mathbf{H}\mathbf{r}_2$ . Επομένως,  $\mathbf{T} = -z_0\mathbf{r}_3 - \mathbf{m} = -z_0\mathbf{r}_3 - \mathbf{U}^{-1}\mathbf{V}\mathbf{r}_1 - \mathbf{U}^{-1}\mathbf{H}\mathbf{r}_2$ , οπότε η Πρόταση 8.4 αποδεικνύεται. *Ο.Ε.Δ.*

Ο υπολογισμός του διανύσματος μετατόπισης ολοκληρώνει την εκτίμηση της 3Δ κίνησης της κάμερας ως προς τη σκηνή αναφοράς. Μπορεί τώρα να γίνει φανερό γιατί δεν υπάρχει απαίτηση για επίπονες ρυθμίσεις του συστήματος πριν από κάθε καταγραφή.

Ας θεωρήσουμε ξανά την περίπτωση του ‘camera ganging’ μεταξύ μίας πραγματικής κάμερας που καταγράφει μία σκηνή πραγματικού χρόνου και μίας εικονικής κάμερας που αποδίδει έναν εικονικό κόσμο. Αφότου οι παράμετροι 3Δ περιστροφής και μετατόπισης της πραγματικής κάμερας έχουν υπολογιστεί από το μόλις επεξεργασμένο καρέ και έχουν οδηγηθεί στην εικονική κάμερα, τότε η νέα θέση της εικονικής κάμερας υπολογίζεται ως η μετακίνησή της από την (γνωστή) σκηνή αναφοράς, βάσει μόνο των υπολογισμένων παραμέτρων περιστροφής και μετατόπισης. Έτσι η λειτουργία ‘camera ganging’ πραγματοποιείται χωρίς καμία αναγκαία ρύθμιση.

Σε θεωρητικές προσεγγίσεις, συχνά υποτίθεται ότι το εστιακό μήκος είναι γνωστό και τίθεται ίσο με τη μονάδα ( $f = 1$ ) για λόγους απλότητας. Όμως, αυτό δεν μπορεί να είναι γενικά αποδεκτό σε πρακτικές εφαρμογές όπου όχι μόνο το εστιακό μήκος δεν είναι μονάδα, αλλά επιπρόσθετα είναι άγνωστο και μεταβαλλόμενο με το χρόνο. Στη συνέχεια θα δειχθεί ότι στην περίπτωσή μας το  $f$  μπορεί να υπολογιστεί ακριβώς πριν από τον υπολογισμό των πινάκων  $\mathbf{R}$  και  $\mathbf{T}$ . Έχοντας επιλέξει τη χρήση 3Δ-σε-2Δ αντιστοιχίσεων χαρακτηριστικών, μεταξύ της σκηνής αναφοράς και του τρέχοντος καρέ, το άγνωστο εστιακό μήκος θα πρέπει να υπολογίζεται για κάθε καταγραφόμενο καρέ. Με άλλα λόγια, στην περίπτωσή μας μόνο ένας άγνωστος  $f$  εισέρχεται στις προκύπτουσες εξισώσεις. Η εξίσωση (8-2) μπορεί να ξαναγραφεί ώστε να εμπεριέχει το εστιακό μήκος ως

$$\begin{bmatrix} \alpha & -1 & \frac{\beta}{f} \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = 0. \quad (8-14)$$

Στην περίπτωση όπου το  $f$  δίνεται με το χρόνο, οι Προτάσεις 8.2, 8.3 και 8.4 δίνουν την ακριβή εκτίμηση για τα  $\mathbf{R}$  και  $\mathbf{T}$ , όπως αναφέρθηκε παραπάνω. Αντιθέτως, όταν το  $f$  είναι άγνωστο,

οι 2Δ παράμετροι των γραμμών  $(a, b)$  δεν μπορούν να υπολογιστούν από τα  $(\alpha, \beta)$ . Θαδειχθεί όμως ότι γενικά το  $f$  μπορεί να υπολογιστεί λύνοντας ένα απλό γραμμικό σύστημα.

Έστω, λοιπόν, για άγνωστο  $f$ , τα  $\hat{\mathbf{e}} = [\alpha - 1 \ \beta]^T$  και  $\hat{\mathbf{e}} = \frac{\hat{\mathbf{e}}}{\|\hat{\mathbf{e}}\|}$  για τη διανυσματική αναπαράσταση κάθε γραμμής. Έστω επίσης τα σύνολα  $\hat{E}_h, \hat{E}_v$ , τα αντίστοιχα των  $E_h, E_v$ , που κατασκευάζονται βάσει των  $\hat{\mathbf{e}}$  όταν το  $f$  είναι άγνωστο. Η Πρόταση 8.5 δίνει τότε μία γραμμική λύση για το  $f$ .

**Πρόταση 8.5:** Έστω  $\mathbf{q}_h$  ( $\mathbf{q}_v$ ) το εξωτερικό γινόμενο οποιωνδήποτε δύο στοιχείων του  $\hat{E}_h$  ( $\hat{E}_v$ ) όμοια με το μέρος (α) της Πρότασης 8.2. Έστω επίσης ότι το  $\mathbf{q}_{h12}$  ( $\mathbf{q}_{v12}$ ) περιέχει τα δύο πρώτα στοιχεία του  $\mathbf{q}_h$  ( $\mathbf{q}_v$ ) και το  $q_{h3}$  ( $q_{v3}$ ) το τρίτο. Τότε το  $f$  δίνεται, με μία αβεβαιότητα προσήμου, από την

$$f^2 = \frac{\mathbf{q}_{h12}^T \mathbf{q}_{v12}}{q_{h3} q_{v3}}. \quad (8-15)$$

**Απόδειξη:**

Έστω ότι το  $\mathbf{q}_h$  ( $\mathbf{q}_v$ ) δίνεται ως άνω, και ότι το αντίστοιχό του  $\mathbf{p}_h$  ( $\mathbf{p}_v$ ) λαμβάνεται από τα  $(a, b)$  όταν το  $f$  είναι κανονικοποιημένο στο 1. Σύμφωνα με την Πρόταση 8.2, ισχύει  $\mathbf{p}_v = \mathbf{r}_2$  και  $\mathbf{p}_h = \mathbf{r}_1$ , ή

$$\mathbf{p}_h^T \mathbf{p}_v = 0. \quad (8-16)$$

Διαισθητικά, η παραπάνω εξίσωση ισχύει για τα  $\mathbf{p}_h$  και  $\mathbf{p}_v$ , αφού, όταν το  $f$  είναι δεδομένο, μπορεί να ενσωματωθεί στη λύση μέσω των  $(a, b)$ . Από την άλλη, η εξίσωση (8-16) δεν ισχύει για τα αντίστοιχα  $\mathbf{q}_h$  και  $\mathbf{q}_v$ . Παρόλ' αυτά, μπορούμε να δούμε ότι τα  $\mathbf{p}_h$  και  $\mathbf{p}_v$  μπορούν να γραφούν μέσω των  $\mathbf{q}_h, \mathbf{q}_v$  και  $f$ , οπότε τότε το  $f$  μπορεί να ληφθεί μέσω της εξίσωσης (8-16). Πρακτικά, για  $\mathbf{q}_h = [q_{h1} \ q_{h2} \ q_{h3}]^T$  και  $\mathbf{p}_h = [p_{h1} \ p_{h2} \ p_{h3}]^T$ ,

$$\begin{bmatrix} q_{h1} \\ q_{h2} \\ q_{h3} \end{bmatrix} = \begin{bmatrix} \frac{p_{h1}}{f} \\ \frac{p_{h2}}{f} \\ p_{h3} \end{bmatrix} \frac{1}{(\frac{1}{f^2}(p_{h1}^2 + p_{h2}^2) + p_{h3}^2)^{1/2}}. \quad (8-17)$$

Οπότε συνδυάζοντας τις εξισώσεις (8-17) και (8-16), λαμβάνεται η εξίσωση (8-15), άρα και αποδεικνύεται η Πρόταση 8.5. *Ο.Ε.Δ.*

Πρακτικά, περισσότερα από δύο ζεύγη στοιχείων από τα  $\hat{E}_h$  και  $\hat{E}_v$  μπορούν να χρησιμοποιηθούν ώστε να λάβουμε μία καλύτερη εκτίμηση για το  $f^2$ . Σε αυτήν την περίπτωση,

$$f^2 = \frac{\sum_k (q_{hk3} q_{vk3} \mathbf{q}_{hk12}^T \mathbf{q}_{vk12})}{\sum_k (q_{hk3} q_{vk3})^2},$$



όπου το  $k$  αντιστοιχεί σε ένα ζεύγος στοιχείων των  $\hat{E}_h$  και  $\hat{E}_v$ . Όπως αναφέρθηκε παραπάνω, αφότου υπολογιστεί το  $f$ , λαμβάνονται τα σύνολα  $E_h$  και  $E_v$  (θέτοντας για κάθε γραμμή  $b = \frac{\beta}{f}$ ), και οι Προτάσεις 8.2, 8.3 και 8.4 δίνουν ακριβή λύση για τα  $\mathbf{R}$  και  $\mathbf{T}$ .

Η εκτίμηση του εστιακού μήκους με το χρόνο επιτρέπει να χρησιμοποιηθεί το προτεινόμενο σύστημα, χωρίς απαίτηση για οποιονδήποτε ηλεκτρομηχανικό εξοπλισμό (π.χ. αισθητήρες) ή για συστήματα καταγραφής ειδικών δυνατοτήτων.

## 8.5 Ασυμπτωτική ανάλυση σφάλματος

Όπως αναφέρθηκε στην Ενότητα 8.2.3, οι 2Δ παράμετροι των γραμμών  $(a, b)$  εξάγονται βάσει εκτίμησης ελαχίστων τετραγώνων (least squares line fitting). Υποθέτοντας ότι μία γραμμή εκτιμάται βάσει γενικά  $N$  σημείων  $(x_i, y_i), i = 1 \dots N$ , τότε λύνουμε το πρόβλημα

$$\begin{bmatrix} \mathbf{X} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{Y} + \mathbf{V}, \quad (8-18)$$

όπου το  $\mathbf{1}$  υποδηλώνει ένα διάνυσμα μεγέθους  $N \times 1$  που περιέχει μονάδες,  $\mathbf{X} = [x_i]$ ,  $\mathbf{Y} = [y_i]$  και το  $\mathbf{V} = [v_i]$  περιέχει τους παράγοντες θορύβου που εισάγονται στις μετρήσεις  $y_i$ . Στην παρακάτω ανάλυση, ο θόρυβος στις μετρήσεις  $y_i$  μοντελοποιείται ως μία ακολουθία λευκού θορύβου με μηδενικό μέσο. Τότε,

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} e_a \\ e_b \end{bmatrix}, \quad \text{με} \quad \begin{bmatrix} e_a \\ e_b \end{bmatrix} = \left( \begin{bmatrix} \mathbf{X}^T \\ \mathbf{1}^T \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{1} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{1}^T \end{bmatrix} \mathbf{V}, \quad (8-19)$$

όπου όλοι οι παράγοντες θορύβου υπεισέρχονται στο  $[e_a \ e_b]^T$ . Μετά από σχετικά πολύπλοκες αλλά ευθείς πράξεις, επιβεβαιώνεται ότι οι εκτιμήτριες των  $(a, b)$  είναι ισχυρά συνεπείς, δηλαδή ότι το σφάλμα είναι μηδενικού μέσου και η διασπορά του τείνει ασυμπτωτικά στο μηδέν όταν  $N \rightarrow \infty$ . Περισσότερο συγκεκριμένα,  $E\{e_a^2(N)\} \approx O(N^{-3})$ ,  $E\{e_a(N) e_b(N)\} \approx O(N^{-2})$  και  $E\{e_b^2(N)\} \approx O(N^{-1})$ .

Επανεξετάζοντας τον ορισμό του πίνακα  $\mathbf{Q}$  στην εξίσωση (8-4), έστω ότι το  $\mathbf{q}_j^T$  αντιπροσωπεύει την  $j$ -οστή  $3 \times 1$  γραμμή του. Όπως αναφέρθηκε και στην Πρόταση 8.2, ο υπολογισμός του πίνακα περιστροφής  $\mathbf{R}$  βασίζεται στον υπολογισμό των δεξιών ειδικών διανυσμάτων του  $\mathbf{Q}$ , ισοδύναμα των ιδιοδιανυσμάτων του  $\mathbf{Q}^T \mathbf{Q} = \sum_j \mathbf{q}_j \mathbf{q}_j^T$ . Θα δειχθεί ότι κάθε παράγοντας αυτού του αθροίσματος τείνει ασυμπτωτικά στην πραγματική του τιμή.

Υποθέτοντας ότι το  $\mathbf{q}_j$  είναι μία συνάρτηση των παραμέτρων των γραμμών  $(a_k, b_k)$ ,  $(a_l, b_l)$ , όπως στην εξίσωση (8-4), λαμβάνουμε

$$\mathbf{q}_j \mathbf{q}_j^T = \frac{1}{\sqrt{a_k^2 + b_k^2 + 1} \sqrt{a_l^2 + b_l^2 + 1}} \begin{bmatrix} (b_k - b_l)^2 & (a_l b_k - a_k b_l)(b_k - b_l) & (a_l - a_k)(b_k - b_l) \\ (a_l b_k - a_k b_l)(b_k - b_l) & (a_l b_k - a_k b_l)^2 & (a_l b_k - a_k b_l)(a_l - a_k) \\ (a_l - a_k)(b_k - b_l) & (a_l b_k - a_k b_l)(a_l - a_k) & (a_l - a_k)^2 \end{bmatrix} \quad (8-20)$$

Όλοι οι όροι στον παραπάνω πίνακα φαίνονται να είναι λόγοι ακολουθιών οι οποίες τείνουν ισχυρά στις πραγματικές τους τιμές όταν  $N \rightarrow \infty$ . Αυτό ισχύει, δεδομένων των τάξεων των προσδοκητών τιμών που αναφέρθηκαν παραπάνω, αλλά και του γεγονότος ότι οι όροι που συσχετίζονται με θόρυβο είναι έως και δεύτερης τάξεως σε όλα τα στοιχεία του πίνακα. Το τελευταίο μπορεί να επιβεβαιωθεί θεωρώντας για παράδειγμα το πάνω αριστερά στοιχείο του πίνακα στην περίπτωση όπου υπάρχει θόρυβος, όπου  $(\hat{b}_k - \hat{b}_l)^2 = \hat{b}_k^2 + \hat{b}_l^2 - 2\hat{b}_k \hat{b}_l$  με  $\hat{b}_k = b_k + e_{bk}$ ,  $\hat{b}_k^2 = b_k^2 + e_{bk}^2 + 2b_k e_{bk}$  και  $E\{e_{bk}(N)\} = 0$ ,  $E\{e_{bk}^2(N)\} \rightarrow 0$ , όπως ειπώθηκε ήδη παραπάνω.

Με στόχο να διερευνήσουμε εάν η διασπορά του θορύβου τείνει ασυμπτωτικά στο μηδέν και για τα στοιχεία του πίνακα  $\mathbf{Q}$ , αρκεί να αποδείξουμε ότι οι ποσότητες  $E\{e_a^4(N)\}$ ,  $E\{e_b^4(N)\}$ ,  $E\{e_a^2(N)e_b^2(N)\}$  τείνουν ασυμπτωτικά στο μηδέν όταν  $N \rightarrow \infty$  (με δεδομένη την έκφραση των στοιχείων του  $\mathbf{Q}$  στην εξίσωση (8-20)). Μετά από σχετικά επίπονες πράξεις και εκμεταλλευόμενοι τις ιδιότητες των cumulants και τις ιδιότητες του λευκού θορύβου (βλ. εξίσωση (8-23) παρακάτω), αποδεικνύεται ότι  $E\{e_a^4(N)\} \approx O(N^{-6})$ ,  $E\{e_a^2(N)e_b^2(N)\} \approx O(N^{-4})$  και  $E\{e_b^4(N)\} \approx O(N^{-2})$ . Για συντομία, μόνο μία ενδεικτική απόδειξη δίνεται εδώ και αφορά στην τελευταία ισότητα.

**Απόδειξη:** Θα δείξουμε ότι  $E\{e_b^4(N)\} \approx O(N^{-2})$ . Από την εξίσωση (8-19), μετά από μερικούς υπολογισμούς

$$e_b = \frac{\sum_i x_i^2 \sum_j v_j - \sum_i x_i \sum_j x_j v_j}{N \sum_i x_i^2 - (\sum_i x_i)^2}, \quad (8-21)$$

και

$$(N \sum_i x_i^2 - (\sum_i x_i)^2)^4 E\{e_b^4\} = E\{(\sum_i x_i^2 \sum_j v_j - \sum_i x_i \sum_j x_j v_j)^4\}. \quad (8-22)$$

Στην προσδοκητή τιμή του δεξιού μέλους της εξίσωσης (8-22) εμφανίζεται η ροπή τέταρτης τάξης  $m_4 = E\{v_j v_k v_l v_m\}$ . Μία περισσότερο συστηματική προσέγγιση σε αυτό το σημείο αποτελεί η χρήση του cumulant τέταρτης τάξης  $c_4(\tau_1, \tau_2, \tau_3)$  στην [94], ως:

$$c_4 \delta(\tau_1) \delta(\tau_2) \delta(\tau_3) = m_4(\tau_1, \tau_2, \tau_3) - \sigma^4 (\delta(\tau_1) \delta(\tau_2 - \tau_3) + \delta(\tau_2) \delta(\tau_3 - \tau_1) + \delta(\tau_3) \delta(\tau_1 - \tau_2)) \quad (8-23)$$

όπου  $\tau_1 = k - j$ ,  $\tau_2 = l - j$ ,  $\tau_3 = m - j$  και  $c_4 \triangleq c_4(0, 0, 0)$ . Για να εκμεταλλευτούμε την εξίσωση (8-23), το δεξί μέλος της (8-22) πρέπει να ξαναγραφεί κατάλληλα.

Για παράδειγμα, μετά την ανάπτυξη του, ο όρος  $(\sum_i x_i)^4 (\sum_j x_j v_j)^4$  δίνει:

$$\begin{aligned} E\{(\sum_i x_i)^4 (\sum_j x_j v_j)^4\} &= (\sum_i x_i)^4 \sum_j \sum_k \sum_l \sum_m x_j x_k x_l x_m E\{v_j v_k v_l v_m\} = \\ &= (\sum_i x_i)^4 \sum_j \sum_k \sum_l \sum_m x_j x_k x_l x_m (c_4 \delta(k-j) \delta(l-j) \delta(m-j) + \\ &+ \sigma^4 \delta(k-j) \delta(l-m) + \sigma^4 \delta(l-j) \delta(m-k) + \sigma^4 \delta(m-j) \delta(k-l)) . \end{aligned}$$

Μετά από ευθείς πράξεις καταλήγουμε στην

$$E\{(\sum_i x_i)^4 (\sum_j x_j v_j)^4\} = (\sum_i x_i)^4 \left( c_4 \sum_i x_i^4 + 3\sigma^4 (\sum_i x_i^2)^2 \right) , \quad (8-24)$$

που είναι  $O(N^{14})$ . Αυτό μπορεί να δειχθεί με απλό τρόπο παίρνοντας incrementals της μορφής  $x_i \equiv i \cdot x_0$ .

Με όμοιο τρόπο, έχει δειχθεί ότι όλοι οι υπόλοιποι όροι στο δεξί μέλος της εξίσωσης (8-22) είναι τάξης μικρότερης ή ίσης με  $O(N^{14})$ , ενώ  $(N \sum_i x_i^2 - (\sum_i x_i)^2)^4 \equiv O(N^{16})$ . Επομένως,  $E\{e_b^4(N)\} \equiv O(N^{-2})$  και  $E\{e_b^4(N)\} \rightarrow 0$  όταν  $N \rightarrow \infty$ . *O.E.Δ.*

## 8.6 Πειράματα και συζήτηση

### 8.6.1 Πειράματα με εικονική κάμερα

Στην ενότητα αυτή, ένας αριθμός πειραμάτων έχει πραγματοποιηθεί, ώστε να επιβεβαιωθεί η αποδοτικότητα του προτεινόμενου αλγορίθμου. Για τους σκοπούς των πειραμάτων, ένα επίπεδο κυανό πέτασμα κατασκευάστηκε σε ένα εικονικό περιβάλλον χρησιμοποιώντας ένα κατάλληλο εμπορικό πακέτο λογισμικού. Μία εικονική κάμερα χρησιμοποιήθηκε για να αποδώσει τμήματα του κυανού πετάσματος, για γνωστές παραμέτρους κίνησης της κάμερας, δηλ. άξονα και γωνία περιστροφής, 3Δ μετατόπιση και εστιακό μήκος. Βάσει της καταγραφόμενης ακολουθίας, η κίνηση της κάμερας υπολογίζεται ακολουθώντας τον προτεινόμενο αλγόριθμο και τα αποτελέσματα που λαμβάνονται συγκρίνονται με τα αντίστοιχα πραγματικά. Στα επόμενα, θα δειχθεί ότι οι παράμετροι κίνησης της κάμερας υπολογίζονται με τέτοια ακρίβεια, ώστε η ακολουθία που λαμβάνεται χρησιμοποιώντας τις υπολογιζόμενες παραμέτρους κίνησης (αντί των πραγματικών) είναι οπτικά πανομοιότυπη με την αρχική.

Θα πρέπει να σημειωθεί εδώ ότι δεν χρησιμοποιήθηκαν ηθοποιοί στις ακολουθίες των πειραμάτων με εικονική κάμερα, αφού ο διαχωρισμός χρώματος (chromakeying) είναι πέραν του

αντικειμένου αυτού του κεφαλαίου. Επίσης, όπως αναφέρθηκε και στην Ενότητα 8.1, ο διαχωρισμός χρώματος έχει μελετηθεί εις βάθος στην βιβλιογραφία (βλ. [53, 63]). Οι δύο τόνοι του μπλε που εμφανίζονται στα σχήματα είναι απλώς ενδεικτικοί, αφού επιλέγησαν ώστε να είναι οπτικώς διαχωρίσιμοι. Πρακτικά, οι τόνοι του μπλε καθορίζονται στα πλαίσια της τεχνικής διαχωρισμού χρώματος που υιοθετείται.

Για τα πειράματα που πραγματοποιήθηκαν, το κυανό πέτασμα κατασκευάστηκε σύμφωνα με τις πρακτικές οδηγίες της Ενότητας 8.3. Ας υποθέσουμε ότι ένας τοίχος φυσικών διαστάσεων  $340cm \times 528cm$  (που αντιστοιχεί σε κατακόρυφο  $\times$  οριζόντιο μήκος) είναι διαθέσιμος για το κυανό πέτασμα του συγκεκριμένου εικονικού στούντιο. Ας υποθέσουμε επίσης ότι τουλάχιστον το ένα έβδομο κατά την κατακόρυφη διάσταση του πετάσματος θα είναι μέσα στο οπτικό πεδίο της κάμερας, δηλ. περίπου  $50cm$ . Επιλέγοντας το μέγεθος του κάθε τετραπλεύρου να είναι  $10cm$ ,  $n = 5$  τετράπλευρα είναι ορατά κατά την κατακόρυφη διάσταση σε σύνολο  $2^n + n - 2 = 35$  τετραπλεύρων, τα οποία με τη σειρά τους μειώνονται σε  $N = 34$  για να χωρέσουν στη φυσική κατακόρυφη διάσταση του τοίχου. Υποθέτοντας ότι το ένα δέκατο-πέμπτο της οριζόντιας διάστασης είναι ορατό από την κάμερα, δηλαδή περίπου  $35cm$ , και θέτοντας το οριζόντιο μήκος κάθε τετραπλεύρου στα  $12cm$ ,  $m = 3$  τετράπλευρα είναι ορατά κατά την οριζόντια διάσταση.

Όπως αναλύσαμε στην Ενότητα 8.3, ένας μεγάλος αριθμός από τετράπλευρα  $M$  μπορεί να εισαχθεί στη διάσταση που επεξεργαζόμαστε δεύτερη (σε αυτήν την περίπτωση,  $(2^n - 1)^{(m-1)} + m - 1 = 963$  τετράπλευρα). Όμως, ο αριθμός των τετραπλεύρων περιορίζεται από το φυσικό οριζόντιο μήκος του τοίχου σε  $M = 528/12 = 44$  τετράπλευρα. Γενικά, το μικρότερο δυνατό τμήμα του τοίχου που υποτίθεται ορατό από την κάμερα θα πρέπει να επιλέγεται σε κάθε περίπτωση αρκετά μικρό, αφού το κυανό πέτασμα κατασκευάζεται μία φορά και πρέπει να είναι κατάλληλο για παραγωγές διαφόρων τύπων. Επιπρόσθετα, διαισθητικά, η οριζόντια διάσταση του ορατού μέρους θα πρέπει να επιλέγεται γενικά μικρότερη από την κατακόρυφη, αφού η επικάλυψη από το προσκήνιο αναμένεται γενικά να είναι μεγαλύτερη. Στο παράδειγμά μας, καταλήγουμε σε ένα ελάχιστο τμήμα διάστασης  $n \times m = 5 \times 3$  τετραπλεύρων, το οποίο αναγνωρίζεται μοναδικά σε ένα δυαδικό πίνακα  $\mathbf{B}$  με  $34 \times 44$  τετράπλευρα.

Εκ κατασκευής, οι γνωστές παράμετροι 3Δ δομής της σκηνής αναφοράς φυλάσσονται σύμφωνα με την Ενότητα 8.4, ως

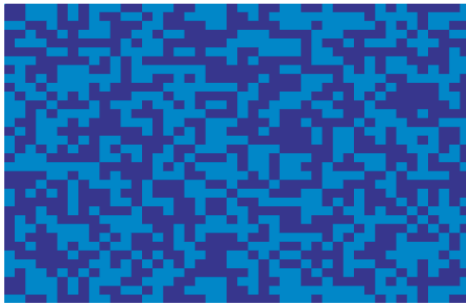
$$(1) \text{ το σύνολο των κατακόρυφων ευθειών } X_v = \left\{ -\frac{528}{2}, -\frac{528}{2} + 12, -\frac{528}{2} + 2 \cdot 12, \dots, \frac{528}{2} \right\},$$

$$(2) \text{ το σύνολο των οριζόντιων ευθειών } Y_h = \left\{ -\frac{340}{2}, -\frac{340}{2} + 10, -\frac{528}{2} + 2 \cdot 10, \dots, \frac{340}{2} \right\}$$

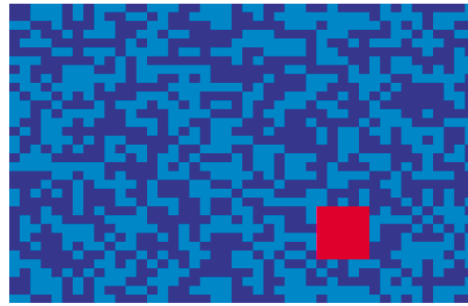
και

(3) το 3Δ βάθος του κυανού πετάσματος  $z \triangleq z_0$ .

Πρακτικά, οι κατακόρυφες και οριζόντιες γραμμές, όπως επίσης και το 3Δ βάθος ορίζονται βάσει της υποτιθέμενης σκηνής αναφοράς. Εφόσον αυτή η σκηνή αναφοράς δεν καταγράφεται ποτέ από την πραγματική κάμερα, οι παραπάνω παράμετροι δομής ορίζονται πρακτικά από την υποτιθέμενη αρχική θέση της εικονικής κάμερας σε σχέση με το κυανό πέτασμα. Με αυτό το σκεπτικό, θέτοντας την εικονική κάμερα στο γεωμετρικό κέντρο της εικόνας και με τις γραμμές σάρωσης παράλληλες στις οριζόντιες γραμμές αναφοράς, λαμβάνουμε τη συγκεκριμένη οριζόντια και κατακόρυφη κατάτμηση σε γραμμές. Ταυτόχρονα, θέτοντας την απόσταση μεταξύ του CCD και του κέντρου της οθόνης ίση με  $z_0$ , λαμβάνουμε το 3Δ βάθος. Για το παράδειγμα αυτό, υποθέτουμε ότι  $z_0 \equiv 2$ . Θα πρέπει να σημειωθεί ότι από εδώ και στο εξής όλα τα μήκη δίνονται σε τετραγωνικά εκατοστά (cm), οπότε αυτά θα παραλείπονται.



(α)



(β)

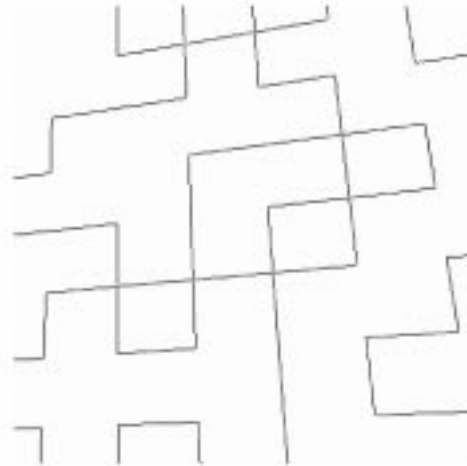
**Σχήμα 8.3** Παράδειγμα κυανού πετάσματος δύο χρωματικών τόνων (α) και καταγραφόμενο από την κάμερα τμήμα του πετάσματος (β).

Στο Σχήμα 8.3(α) παρουσιάζεται το κατασκευασμένο (εικονικό) κυανό πέτασμα. Για μία τυχαία κίνηση της κάμερας στον 3Δ χώρο, π.χ. για γωνία περιστροφής  $\alpha = 17^\circ$  και άξονα  $\mathbf{u} = [.760, .588, .277]^T$ , για μετατόπιση  $\mathbf{T} = [-120, 80, 110]^T$  και εστιακό μήκος  $f = 6$ , η κάμερα καταγράφει το τμήμα του κυανού πετάσματος που εμφανίζεται στο Σχήμα 8.4(α). Στο συγκεκριμένο παράδειγμα, η εικονική κάμερα καταγράφει  $576 \times 576$  non-interlaced έγχρωμες εικόνες με αναλογία όψεων (aspect ratio) ίση με 1.

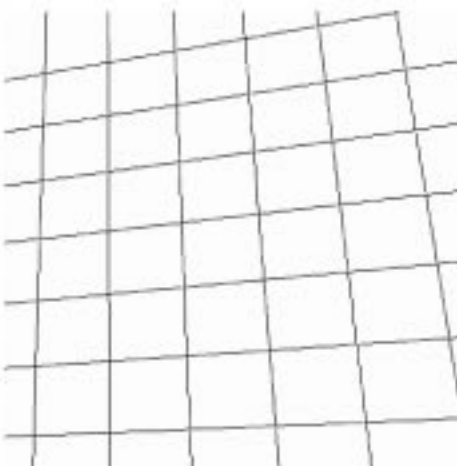
Μετά από εφαρμογή ενός βήματος εντοπισμού ακμών που χρησιμοποιεί τον τελεστή Sobel, λαμβάνεται ο δυαδικός πίνακας του Σχήματος 8.4(β). Οι ευθείες γραμμές που εντοπίζονται από τον μετασχηματισμό Hough, χρησιμοποιώντας πίνακα τιμών μεγέθους  $1024 \times 1024$ , εμφανίζονται στο Σχήμα 8.4(γ). Οι γραμμές που λαμβάνονται δίνονται ως είσοδος στον αλγόριθμο υπολογισμού της κίνησης της κάμερας (βλ. Ενότητα 8.4). Οι παράμετροι μεγέθυνσης και περιστροφής λαμβάνονται κατευθείαν από τις Προτάσεις 8.5 και 8.2 αντίστοιχα: για το εστιακό μήκος  $\hat{f} = 6.09$  και για την γωνία περιστροφής  $\hat{\alpha} = 17.2^\circ$  και τον άξονα  $\hat{\mathbf{u}} = [.762, .584, .280]^T$ .



(α)



(β)



(γ)



(δ)

**Σχήμα 8.4** Όλα τα βήματα της προτεινόμενης μεθοδολογίας για τον υπολογισμό της κίνησης της κάμερας στο Σχήμα 8.3(β), (α) το καταγραφόμενο καρέ, (β) οι ακμές που εντοπίστηκαν στο καρέ, (γ) οι εξαγόμενες γραμμές αναφοράς, (δ) το καρέ όπως αποδίδεται από τις υπολογιζόμενες παραμέτρους κίνησης.

Όπως μπορεί κανείς να διαπιστώσει, οι υπολογιζόμενες παράμετροι είναι πολύ κοντά στις πραγματικές.

Παρατηρώντας τα Σχήματα 8.4(α) και 8.4(γ), μπορεί κανείς να δει ότι η απουσία μεγάλων ομοιόμορφων περιοχών οδηγεί στον εντοπισμό όλων των γραμμών αναφοράς του καταγραφόμενου καρέ, η οποία είναι πρακτικά και η συνήθης περίπτωση (βλ. παρακάτω για την περίπτωση όπου απουσιάζουν γραμμές αναφοράς). Έτσι, ο υποπίνακας του  $\mathbf{B}$  που αντιστοιχεί στο ορατό

τιμήμα του κυανού πετάσματος μπορεί να εξαχθεί αμέσως:

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Πρακτικά, όπως εξηγήσαμε παραπάνω, ακόμη και ένας  $5 \times 3$  υποπίνακας του  $\mathbf{S}$  θα ήταν αρκετός για να αναγνωρίσουμε μοναδικά το τιμήμα επάνω στο δυαδικό χάρτη του πετάσματος. Επιπρόσθετα, όπως μπορεί κανείς να παρατηρήσει από το Σχήμα 8.3(β), όπου το ορατό τιμήμα έχει σημειωθεί με διαφορετικό χρώμα επάνω στο πέτασμα, μία πραγματική κάμερα θα κατέγραφε τουλάχιστον το διπλάσιο του μεγέθους αυτού σε μία ρεαλιστική καταγραφή. Έτσι, είναι μάλλον βέβαιο ότι ένας  $5 \times 3$  υποπίνακας του δυαδικού χάρτη  $\mathbf{B}$  θα είναι πάντα ορατός.

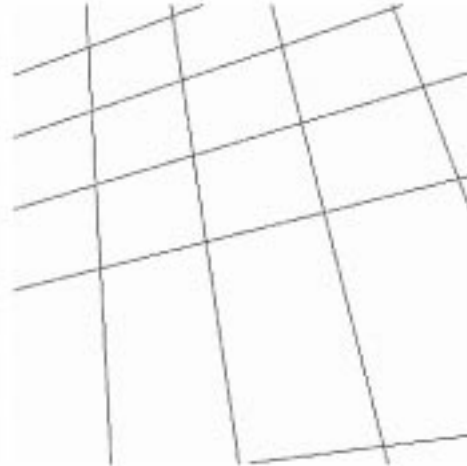
Η αναγνώριση του πίνακα  $\mathbf{S}$  πάνω στο δυαδικό χάρτη επιτρέπει την αντιστοίχιση μεταξύ των αρχικών 3Δ γραμμών και των εξαγόμενων 2Δ γραμμών του καρέ. Μετά την πραγματοποίηση της αντιστοίχισης, η Πρόταση 8.4 δίνει το 3Δ διάνυσμα μετατόπισης  $\hat{\mathbf{T}} = [-120.1, 79.6, 111.3]^T$ . Για να επιβεβαιώσουμε και οπτικά την απόδοση του προτεινόμενου αλγορίθμου, το καρέ του Σχήματος 8.4(α) αποδίδεται εκ νέου χρησιμοποιώντας τώρα τις υπολογιζόμενες παραμέτρους κίνησης για την εικονική κάμερα. Έτσι, το καρέ του Σχήματος 8.4(δ) θα πρέπει να συγκριθεί απευθείας με αυτό του 8.4(α). Συγκρίνοντας τα δύο αυτά σχήματα, μπορούμε να παρατηρήσουμε ότι η ίδια σκηνή αποδιδόμενη από μία δεύτερη εικονική κάμερα, βάσει των υπολογιζόμενων παραμέτρων κίνησης, θα ήταν συμβατή για σκοπούς σύνθεσης.

Στα επόμενα, θα επιδειχθεί η σημασία της Πρότασης 8.3. Σύμφωνα με την Ενότητα 8.3, αφότου η ακολουθία μέγιστου μήκους  $\mathbf{c}$  εξάγεται για τη μικρότερη πλευρά του πετάσματος, ο 2Δ δυαδικός χάρτης  $\mathbf{B}$  υπολογίζεται βάσει κατάλληλων συνδυασμών των ολισθήσεων του  $\mathbf{c}$ . Μπορεί κανείς να δει ότι κατάλληλοι συνδυασμοί από ολισθήσεις δεν επιτρέπουν την εμφάνιση στο πέτασμα μεγάλων ομοιόμορφων περιοχών από σκούρο ή ανοιχτό μπλε. Παρόλ' αυτά, το τελευταίο είναι δυνατόν στη γενική περίπτωση. Σε τέτοιες σπάνιες περιπτώσεις είναι η Πρόταση 8.3 που διασφαλίζει την επιτυχή εξαγωγή του δυαδικού προτύπου  $\mathbf{S}$ .

Το Σχήμα 8.5(α) απεικονίζει ένα τέτοιο καταγραφόμενο καρέ, ενώ οι γραμμές αναφοράς που εντοπίζονται δίνονται στο Σχήμα 8.5(β). Με απλή επισκόπηση των Σχημάτων 8.5(α,β), μπορούμε να δούμε ότι, για να εξάγουμε έναν ελάχιστο  $5 \times 3$  υποπίνακα, θα πρέπει να εντοπιστεί η γραμμή αναφοράς που έχει απορροφηθεί στη μεγάλη περιοχή μπλε ανοιχτού χρώματος. Σε αυτήν την περίπτωση, η εξίσωση (8-10) θα πρέπει να χρησιμοποιηθεί για όλες τις γραμμές



(α)



(β)

**Σχήμα 8.5** (α) Καρέ που καταγράφεται με απύσεις κάποιες από τις γραμμές αναφοράς και (β) οι γραμμές αναφοράς που εξάγονται.

που εξάγονται στο  $E_h$  - σημειώστε ότι τα  $f$  και  $\mathbf{R}$  έχουν ήδη ληφθεί σε αυτό το σημείο. Γενικά, οι γραμμές του  $E_h$  (ή  $E_v$ ) είναι διατεταγμένες χωρικά (το οποίο είναι απλό, αφού δεν συμβάλλουν μέσα σε ένα καρέ) και η απόστασή τους σε ζεύγη μπορεί να μετρηθεί από την εξίσωση (8-10). Όταν η απόσταση ξεπερνάει σημαντικά τη μονάδα, μία απύσα γραμμή έχει εντοπιστεί. Πρακτικά, η γραμμή η ίδια δεν είναι αναγκαίο να υπολογισθεί, φθάνει να εντοπιστεί το γεγονός της απουσίας της. Αντιθέτως, αφότου ο  $\mathbf{S}$  εξαχθεί, αρκεί η αντίστοιχη γραμμή του (για γραμμές του  $E_h$ ) ή η αντίστοιχη στήλη του (για γραμμές του  $E_v$ ) να ‘διπλασιαστεί’.

Ενδεικτικά αποτελέσματα για τον εντοπισμό γραμμών αναφοράς που απουσιάζουν δίνονται στον Πίνακα 8.2. Αρχίζοντας από την αριστερή πλευρά, οι στήλες 1-2 περιέχουν τις παραμέτρους των γραμμών που εξάγονται από τον μετασχηματισμό Hough, ενώ οι στήλες 3-4 περιέχουν τις αντίστοιχες παραμέτρους κανονικοποιημένες ως προς το υπολογιζόμενο εστιακό μήκος και τα μήκη των πλευρών του CCD. Η τελευταία στήλη απεικονίζει τις κανονικοποιημένες αποστάσεις των γραμμών στη σκηνή αναφοράς, όπως αυτές υπολογίζονται από την Πρόταση 8.3. Μπορεί να δει κανείς ότι, η κανονικοποιημένη απόσταση μεταξύ των γραμμών 4 και 5 (γραμμές 4 και 5 στον Πίνακα) είναι προσεγγιστικά ίση με δύο (γραμμή 5 στον Πίνακα), το οποίο υποδεικνύει ότι η γραμμή του προτύπου, που περιλαμβάνεται μεταξύ των γραμμών 4 και 5, θα πρέπει να διπλασιαστεί. Όμοια αποτελέσματα λαμβάνονται και για απύσες κατακόρυφες γραμμές.

Αν και η απουσία γραμμών είναι σπάνια (αν όχι αδύνατη) για την προτεινόμενη τακτική στην κατασκευή του κυανού πετάσματος, είναι μάλλον συνήθης περίπτωση για το κυανό πέτασμα που προτείνεται στις [182, 31]. Επιπρόσθετα, υποτίθεται ότι θα υπάρχει πάντα ένα ορατό



$A_h$	$B_h$	norm. $A_h$	norm. $B_h$	$D_h$
.372	306.65	.372	.626	
.341	219.27	.341	.448	1.00
.300	117.45	.300	.240	1.03
.251	1.74	.251	.004	1.02
.127	-288.59	.127	-.589	2.04

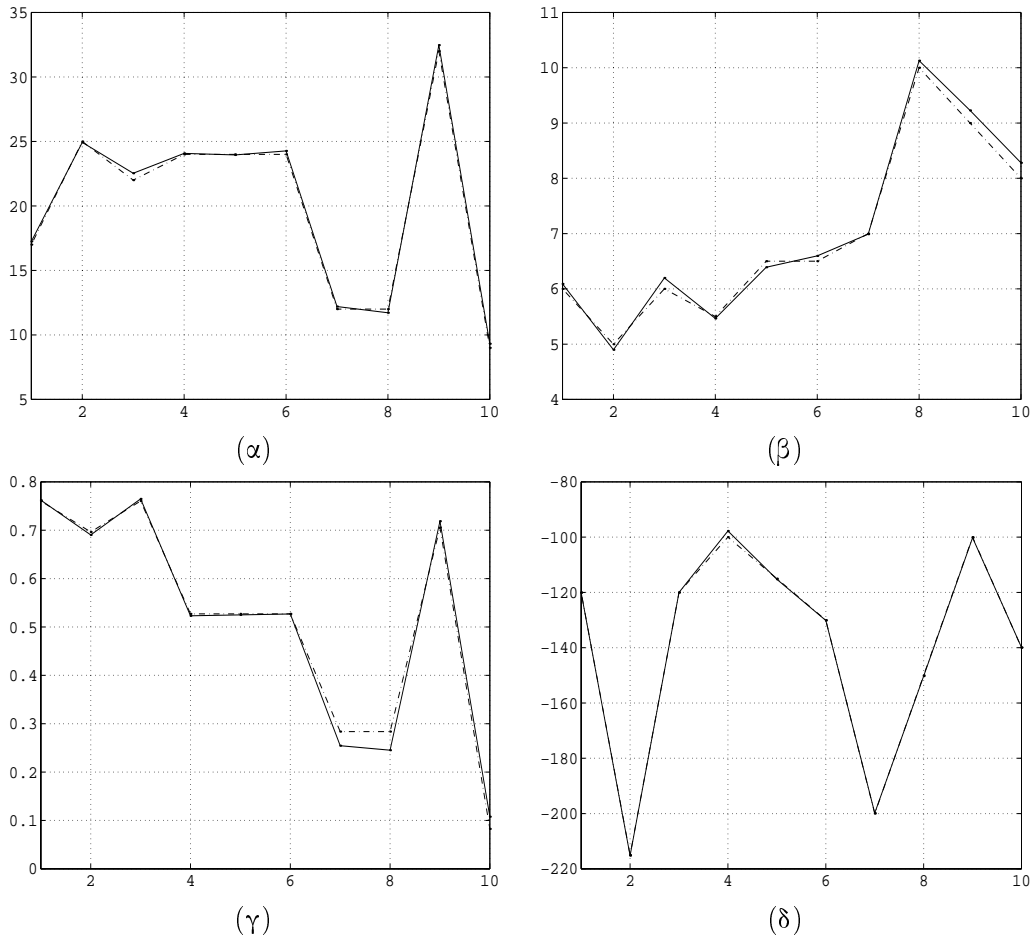
**Πίνακας 8.2** Εντοπισμός των γραμμών αναφοράς που απουσιάζουν. Παράμετροι γραμμών που υπολογίζονται (ξεκινώντας από αριστερά) από τον μετασχηματισμό Hough, μετά από κανονικοποίηση μεγέθυνσης και με τις πραγματικές διαστάσεις του CCD, κανονικοποιημένες αποστάσεις γραμμών σε ζεύγη όπως υπολογίζονται από την Πρόταση 8.3.

ζεύγος γραμμών στο καρέ, για το οποίο η απόσταση των αντίστοιχων γραμμών στη σκηνή αναφοράς θα είναι ίση με μονάδα (δηλ. ίση με την πλευρά ενός τετραπλεύρου). Αν και το τελευταίο είναι αλήθεια στην προσέγγιση που παρουσιάζεται στο παρόν κεφάλαιο, δεν ισχύει πάντα στις [182, 31]. Η απόσυρση της απαίτησης αυτής θα οδηγούσε κατευθείαν σε ανακριβείς μετρήσεις της μεγέθυνσης.

Η επίδοση του αλγορίθμου εξετάστηκε σε έναν μεγάλο αριθμό από όμοια πειράματα, χρησιμοποιώντας διαφορετικά κυανά πετάσματα και τυχαίες παραμέτρους κίνησης της κάμερας. Σε όλα τα πειράματα που εκτελέστηκαν, η υπολογιζόμενη γωνία περιστροφής και το εστιακό μήκος δεν διέφεραν περισσότερο από  $.3^\circ$  και  $.1$  αντίστοιχα από τα πραγματικά. Μερικά ενδεικτικά αποτελέσματα υπολογισμού κίνησης δίνονται στα Σχήματα 8.6(α-δ) για μία αποδιδόμενη ακολουθία από 100 καρέ, όπου οι παράμετροι κίνησης της κάμερας μεταβάλλονται τυχαία με το χρόνο. Στα Σχήματα 8.6(α-δ) μετρήσεις για τη γωνία περιστροφής, το εστιακό μήκος, την πρώτη συνιστώσα του άξονα περιστροφής και την πρώτη συνιστώσα του διανύσματος μετατόπισης αντίστοιχα (κανονικές γραμμές) δίνονται για κάθε 10 καρέ. Οι μετρήσεις αυτές θα πρέπει να συγκριθούν με τις πραγματικές τους τιμές (διακεκομμένες γραμμές).

### 8.6.2 Πειράματα με πραγματική κάμερα

Η επίδοση του αλγορίθμου εξετάστηκε και σε φυσικό περιβάλλον, χρησιμοποιώντας το εικονικό στούντιο - μινιατούρα του Σχήματος 8.7(α). Για το καρέ του παραδείγματος στο Σχήμα 8.7(β), το οποίο καταγράφηκε από επαγγελματική ψηφιακή κάμερα με CCD φυσικών διαστάσεων  $6.6mm \times 8.8mm$  και με διαστάσεις pixels  $752 \times 582$ , ο προτεινόμενος αλγόριθμος εφαρμόστηκε στο τμήμα που υποδεικνύεται από το παραλληλόγραμμο με το μαύρο περίγραμμα. Τα αντίστοιχα αποτελέσματα εντοπισμού ακμών απεικονίζονται στο Σχήμα 8.7(γ), ενώ το ανακατασκευασμένο καρέ, όπως το 'κατέγραψε' μία εικονική κάμερα από τις υπολογιζόμενες παραμέτρους κίνησης, δίνεται στο Σχήμα 8.7(δ). Συγκρίνοντας τα Σχήματα 8.7(β) και 8.7(δ), μπορεί κανείς να διαπιστώσει ότι οι υπολογιζόμενες παράμετροι είναι επίσης πολύ κοντά στις



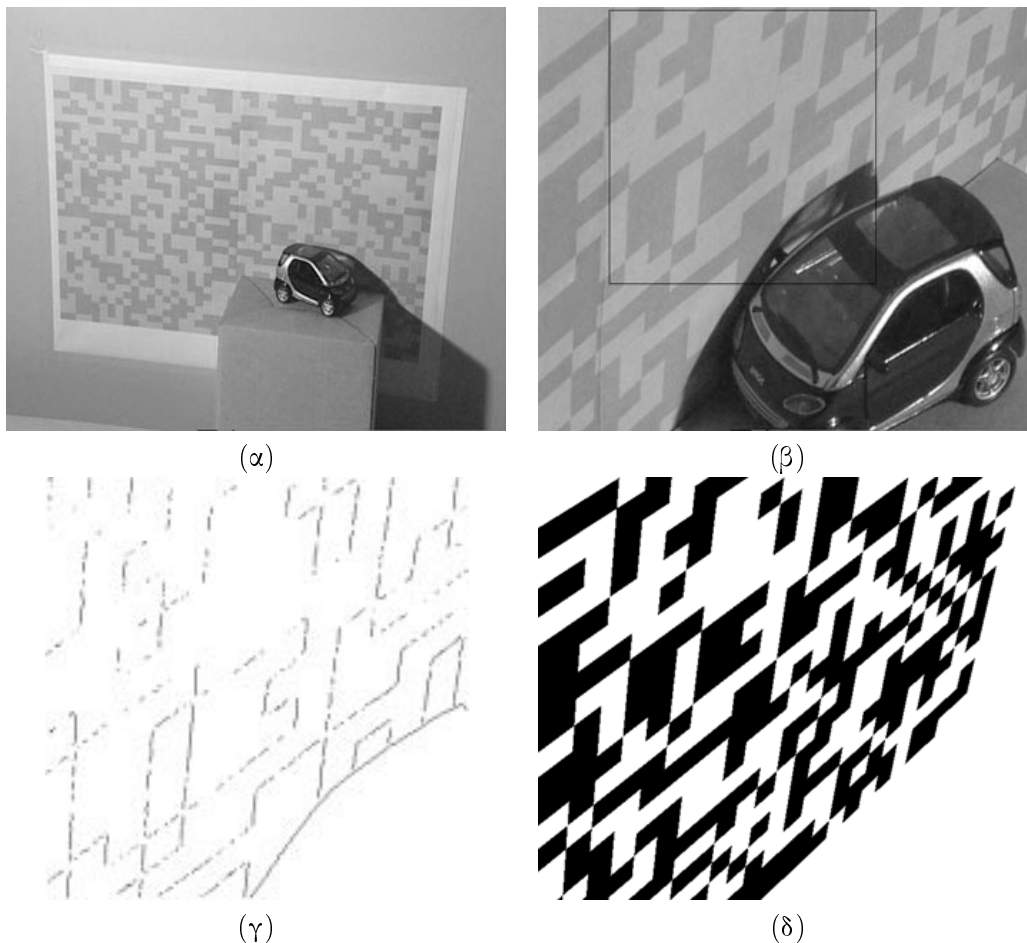
**Σχήμα 8.6** Εκτιμήσεις (α) της γωνίας περιστροφής, (β) του μήκους εστίασης, (γ) της πρώτης συνιστώσας του άξονα περιστροφής και (δ) της πρώτης συνιστώσας του διανύσματος μετατόπισης (απλές γραμμές) συναρτήσει του χρόνου, για μία τυχαία ακολουθία.

πραγματικές. Για λόγους πληρότητας δίνουμε εδώ τις υπολογιζόμενες παραμέτρους:  $\hat{\alpha} = 68.6204^\circ$ ,  $\hat{\mathbf{u}} = [-.563, -.805, .189]^T$ ,  $\hat{\mathbf{T}} = [-138, -28, 1400]^T$  και  $\hat{f} = 29.2414$  (μετρήσεις μήκους σε mm).

Τέλος, θα πρέπει να σημειωθεί ότι η ακρίβεια που επιτυγχάνεται από το προτεινόμενο σύστημα εξαρτάται μόνο από την επίδοση του μετασχηματισμού Hough. Επομένως, χρησιμοποιώντας ακόμη μεγαλύτερους πίνακες τιμών (look-up tables) το σφάλμα στις υπολογιζόμενες παραμέτρους κίνησης μπορεί να μειωθεί δραματικά. Το τελευταίο έχει να κάνει με τη διαθέσιμη υπολογιστική ισχύ αλλά και τις ενδεχόμενες απαιτήσεις επίδοσης σε πραγματικό χρόνο.

## 8.7 Συμπεράσματα και επεκτάσεις

Στο κεφάλαιο αυτό προτείναμε ένα πρότυπο σύστημα για τον υπολογισμό της κίνησης της κάμερας σε εικονικά στούντιο βάσει ενός κυανού πετάσματος δύο τόνων. Αναπτύχθηκε μία



**Σχήμα 8.7** Ένα φυσικό παράδειγμα: (α) το σύστημα-μινιατούρα ενός εικονικού στούντιο, (β) ένα καρέ που καταγράφεται από μία πραγματική κάμερα, (γ) οι ακμές που εξάγονται (δ) και το καρέ που παράγεται χρησιμοποιώντας τις παραμέτρους της κάμερας που υπολογίστηκαν.

νέα μεθοδολογία για την κατασκευή του κυανού πετάσματος και του δυαδικού του χάρτη, εκμεταλλευόμενη τα αρχέτυπα πολυώνυμα και τις ιδιότητές τους. Το προτεινόμενο κυανό πέτασμα μαζί με μία αποδοτική μέθοδο για τον υπολογισμό της κίνησης της κάμερας από 3Δ-σε-2Δ αντιστοιχίσεις γραμμών, που επίσης προτάθηκε εδώ, φαίνονται να εξάγουν ακριβείς εκτιμήσεις της κίνησης για επιτυχή σύνθεση εικόνων και βίντεο.

Στις κύριες συνεισφορές του κεφαλαίου συγκαταλέγονται τόσο η μεθοδολογία κατασκευής του κυανού πετάσματος όσο και η μέθοδος υπολογισμού της κίνησης της κάμερας. Όσον αφορά στην πρώτη, είναι μία ευέλικτη εναλλακτική λύση στα κυανά πετάσματα που κατασκευάζονται ευρεστικά, η οποία μπορεί να καλύψει τις ανάγκες κάθε εικονικού στούντιο. Όσον αφορά στη μέθοδο υπολογισμού της κίνησης της κάμερας, αυτή επιτυγχάνει να υπολογίσει όλες τις παραμέτρους κίνησης, συμπεριλαμβανομένου του εστιακού μήκους (που συναρτάται με την εστίαση της κάμερας). Πάνω από όλα, η προτεινόμενη μέθοδος καταργεί την ανάγκη επίπονων ρυθμίσεων πάνω στην κάμερα για κάθε καταγραφή, αφού όλες οι μετρήσεις ανάγο-

νται σε μία εικονική σκηνή αναφοράς και η σχετική κίνηση της κάμερας μπορεί να εξαχθεί επιτυχώς.

Τα αποτελέσματα που παρουσιάστηκαν στο κεφάλαιο αυτό προετοιμάζουν το έδαφος για περαιτέρω βελτιώσεις στα ακόλουθα:

(1) αποδοτική εξαγωγή της κίνησης της κάμερας σε οριακές περιπτώσεις, για παράδειγμα όταν υπεισέρχεται αλλοίωση στην καταγραφόμενη ακολουθία λόγω κίνησης ή ανεστίασης (motion or defocus blur) ή όταν το προοπτικό μοντέλο προβολής εκφυλίζεται σε παράλληλο (βλ. αναφορά [32] για μερικά πρώτα αποτελέσματα στο θέμα),

(2) κατάλληλη παρεμβολή μεταξύ διαδοχικών καρέ για μείωση του υπολογιστικού κόστους καθώς και για λόγους λείανσης (smoothing),

(3) χρήση των εξαγόμενων παραμέτρων καθώς και της μετρούμενης ανεστίασης για υπολογισμό της απόστασης των ηθοποιών από το κυανό πέτασμα (βλ. για παράδειγμα τις μεθόδους SFF - Shape From Focus [4]).

## ΚΕΦΑΛΑΙΟ 9

### Συμπεράσματα και Επεκτάσεις

Στα προηγούμενα κεφάλαια παρουσιάστηκαν ουσιαστικά πέντε νέες πρωτότυπες μέθοδοι εξαγωγής, μοντελοποίησης και χρήσης χαρακτηριστικών της δισδιάστατης και τρισδιάστατης δομής και κίνησης αντικειμένων σε ακολουθίες εικόνων. Ως προς την εξαγωγή χαρακτηριστικών δισδιάστατης δομής, προτάθηκε μια πρωτότυπη τεχνική εξαγωγής, μοντελοποίησης και σύγκρισης του χαρακτηριστικού κατανομής χρώματος από στατικές έγχρωμες εικόνες, καθώς και μια αναλλοίωτη αναπαράσταση περιγραμμάτων σημαντικών αντικειμένων για σκοπούς μοντελοποίησης, ταξινόμησης και ανάκτησής τους με βάση το σχήμα. Μια ταχεία και ακριβής μέθοδος για τον εντοπισμό και τη χωροθέτηση κύριων κινούμενων αντικειμένων προτάθηκε για τον προσδιορισμό της σύνθεσης και της δισδιάστατης κίνησης αντικειμένων σε ακολουθίες εικόνων. Όσον αφορά στην εξαγωγή της τρισδιάστατης δομής και κίνησης αντικειμένων από ακολουθίες τους, προτάθηκε μια πρωτότυπη μέθοδος εκτίμησης που αφορά σε στερεά αντικείμενα υπό παράλληλη προβολή. Μια ακριβής και εύρωστη τεχνική υπολογισμού της κίνησης της κάμερας από αντιστοιχίσεις γραμμών δόθηκε επίσης στο πλαίσιο καταγραφών σε εικονικά περιβάλλοντα.

Όπως αναδείχθηκε και στα επιμέρους κεφάλαια, οι προτεινόμενες τεχνικές παρουσιάζονται κατά κανόνα αποδοτικότερες από τις υπάρχουσες στη βιβλιογραφία μεθόδους, τουλάχιστον ως προς τις εφαρμογές για τις οποίες σχεδιάστηκαν και υλοποιήθηκαν. Σε κάθε περίπτωση, τα πειραματικά αποτελέσματα επαληθεύουν τις θεωρητικές προτάσεις και δείχνουν ότι οι τεχνικές αποτελούν τη βάση για περαιτέρω εξέλιξη των συστημάτων στην κατεύθυνση της κατανόησης των περιεχομένων των εικόνων και των ακολουθιών τους.

Αναλυτικότερα, το Κεφάλαιο 3 παρουσίασε ένα πρωτότυπο σύστημα για ανάκτηση εικόνων με βάση το περιεχόμενο, που βασίζεται στο χαρακτηριστικό της κατανομής χρώματος. Η προσέγγιση αυτή διαφέρει ουσιαστικά από τις παραδοσιακές τεχνικές ανάλυσης σε τετραδικά δένδρα αλλά και τις παραδοσιακές τεχνικές γράφων γειτνίασης, καταλήγοντας σε μία πιο γενική και αποδοτική αναπαράσταση του χρωματικού περιεχομένου. Η τεχνική που προτάθηκε μπορεί να αποτελέσει τη βάση για ένα σύνολο επεκτάσεων, όπως η χρήση κάποιας κατάλληλης πολυδιακριτικής υλοποίησης σε περισσότερα επίπεδα, η ενσωμάτωση στα κριτήρια συγκρίσεων χαρακτηριστικών επιπλέον του χρώματος προς την κατεύθυνση ορισμού ενός γενικότερου χαρακτηριστικού σύνθεσης και η χρήση κατάτμησης κίνησης στη λογική των κινούμενων αντικειμένων για πληροφορία βίντεο.

Στο Κεφάλαιο 4 προτάθηκε ένας ταχύς αλγόριθμος εντοπισμού και χωροθέτησης κινούμενων αντικειμένων. Ο αλγόριθμος αποτελεί μία ικανοποιητική εναλλακτική σε τεχνικές που χρησιμοποιούν κατάτμηση βασισμένη σε κίνηση, ιδιαίτερα σε εφαρμογές όπου το υπολογιστικό κόστος είναι κρίσιμο και δεν απαιτείται ακρίβεια σε επίπεδο pixel. Ταυτόχρονα, τα αποτελέσματα που λαμβάνονται μπορούν να χρησιμοποιηθούν ως άριστη αρχικοποίηση για περίπλοκες τεχνικές παρακολούθησης περιγράμματος, αντιμετωπίζοντας τα συνήθη προβλήματα των τελευταίων. Η συγκεκριμένη τεχνική ανοίγει ένα σύνολο από άμεσες επεκτάσεις όπως τη μελέτη της εισαγωγής της ως ένα υποσύστημα αρχικοποίησης σε μεθόδους παρακολούθησης περιγράμματος, τη μοντελοποίηση της πληροφορίας των εξαγόμενων πολυγώνων για σκοπούς σύγκρισης και ταξινόμησης στα πλαίσια του MPEG-7 και τη βελτίωση των λαμβανόμενων αποτελεσμάτων εισάγοντας περιορισμούς ‘διάδοσης’ με το χρόνο στα οριακά πολύγωνα.

Στο Κεφάλαιο 5 προτάθηκε μία διαδικασία κανονικοποίησης καμπυλών με τη χρήση της οποίας είναι δυνατόν να λάβει κανείς μία αναπαράσταση καμπύλης αναλλοίωτη σε μετασχηματισμούς affine, χωρίς πραγματική απώλεια πληροφορίας της αρχικής καμπύλης. Η διαδικασία αυτή μπορεί να χρησιμοποιηθεί ως ένα βήμα προεπεξεργασίας σε κάθε τεχνική αναπαράστασης σχήματος ή ταξινόμησης, αναγνώρισης και ανάκτησης αντικειμένων με βάση το περίγραμμο, αφού καταφέρνει επιτυχώς να αποσυνδέσει το πρόβλημα του ‘αναλλοίωτου σε μετασχηματισμούς affine’ από την εξαγωγή χαρακτηριστικών και το ταίριασμα προτύπων. Η προτεινόμενη κανονικοποίηση δείχνει πειραματική ευρωστέια στο θόρυβο και στις μικρές παραμορφώσεις του σχήματος. Επέκταση της διαδικασίας κανονικοποίησης είναι δυνατή προς αρκετές κατευθύνσεις, όπως ο χειρισμός συνόλων από συνδεδεμένες ή μη, ανοικτές ή κλειστές καμπύλες για σκοπούς οπτικής αναγνώρισης χαρακτήρων ή δεικτοδότησης βάσεων δεδομένων, η κανονικοποίηση δισδιάστατων δεδομένων εικόνας, τρισδιάστατων μοντέλων ή πολυδιάστατων δεδομένων και η κανονικοποίηση ως προς προοπτικούς μετασχηματισμούς.

Το Κεφάλαιο 7 παρουσίασε μία νέα μέθοδο για τον υπολογισμό των πινάκων περιστροφής που χαρακτηρίζουν την κίνηση ενός 3Δ στερεού αντικειμένου από τις παράλληλες 2Δ προβολές του. Πέραν της θεωρητικής λύσης στο πρόβλημα που δίνεται, παρέχονται εκφράσεις για την αλγοριθμική λύση του, οι οποίες επιτρέπουν να συμπεριληφθεί ένας τυχαία μεγάλος αριθμός από θορυβώδη διανύσματα κίνησης στους υπολογισμούς. Παράλληλα, τεκμηριώνεται η ισχυρή σύγκλιση των εκφράσεων στις πραγματικές παραμέτρους περιστροφής, ακόμη και παρουσία θορύβου στα πεδία κίνησης. Η ασυμπτωτική συμπεριφορά των εκτιμήσεων παρουσία θορύβου εξετάστηκε με θεωρητικό και πειραματικό τρόπο. Πιο σημαντικό ακόμη είναι το γεγονός ότι η θεωρητική ανάλυση και τα πειράματα υποδεικνύουν ότι η ποιότητα των μετρήσεων μπορεί να βελτιωθεί εάν μία συγκεκριμένη στρατηγική υιοθετηθεί στη επιλογή των διανυσμάτων κίνησης

που χρησιμοποιούνται στους εμπλεκόμενους πίνακες. Τα αποτελέσματα που παρουσιάστηκαν δίνουν τη βάση για σημαντικές επεκτάσεις όπως η βελτίωση των εκτιμήσεων των πινάκων περιστροφής χρησιμοποιώντας έναν κατάλληλο μηχανισμό βαρών στις εμπλεκόμενες διαδικασίες ελαχίστων τετραγώνων, η βελτίωση του ανακατασκευασμένου 3Δ σχήματος το οποίο τείνει να είναι αρκετά παραμορφωμένο, ο επαναυπολογισμός των διανυσμάτων κίνησης χρησιμοποιώντας τον προτεινόμενο αλγόριθμο ως περιορισμό στερεότητας και η βελτίωση των μετρήσεων σχήματος (βάθους) μέσω χρονικών επαναλήψεων.

Το Κεφάλαιο 8 προτείνει ένα πρότυπο σύστημα υπολογισμού της κίνησης της κάμερας σε εικονικά στούντιο βάσει ενός κυανού πετάσματος δύο τόνων. Αναπτύχθηκε μία νέα μεθοδολογία για την κατασκευή του κυανού πετάσματος και του δυαδικού του χάρτη, εκμεταλλευόμενη τα αρχέτυπα πολυώνυμα και τις ιδιότητές τους. Το προτεινόμενο κυανό πέτασμα μαζί με μία αποδοτική μέθοδο για τον υπολογισμό της κίνησης της κάμερας από 3Δ-σε-2Δ αντιστοιχίσεις γραμμών, που επίσης προτάθηκε, εξάγουν ακριβείς εκτιμήσεις της κίνησης για επιτυχή σύνθεση εικόνων και βίντεο. Τα αποτελέσματα που παρουσιάστηκαν προετοιμάζουν το έδαφος για περαιτέρω βελτιώσεις, όπως η αποδοτική εξαγωγή της κίνησης της κάμερας σε οριακές περιπτώσεις, για παράδειγμα όταν υπεισέρχεται αλλοίωση στην καταγραφόμενη ακολουθία λόγω κίνησης ή ανεστίασης ή όταν το προοπτικό μοντέλο προβολής εκφυλίζεται σε παράλληλο, η κατάλληλη παρεμβολή μεταξύ διαδοχικών καρέ για μείωση του υπολογιστικού κόστους καθώς και για λόγους λείανσης και η χρήση των εξαγόμενων παραμέτρων καθώς και της μετρούμενης ανεστίασης για υπολογισμό της απόστασης των ηθοποιών από το κυανό πέτασμα.

## Βιβλιογραφία



## Βιβλιογραφία

- [1] Aggarwal, J.K., and N. Nandhakumar. On the Computation of Motion from Image Sequences - A Review. *IEEE Proceedings*, 76:917–935, August 1988.
- [2] Aizawa, K., H. Harashima and T. Saito. Model-based analysis-synthesis image coding (MBASIC) system for a person's face. *Signal Processing: Image Communication*, 1:139–152, October 1989.
- [3] Arkin, E.M., L.P. Chew, D.P. Huttenlocher, K. Kedem and J.S.B. Mitchell. An Efficiently Computable Metric for Comparing Polygonal Shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(3):209–216, 1991.
- [4] Asada, N., H. Fujiwara and T. Matsuyama. Edge and Depth from Focus. *International Journal of Computer Vision*, 26(2):153–163, 1998.
- [5] Avrithis, Y., A. Doulamis, N. Doulamis and S. Kollias. An Adaptive Approach to Video Indexing and Retrieval Using Fuzzy Classification. In *International Conference on Very Low Bitrate Video Coding*, Urbana, IL, 1998.
- [6] Avrithis, Y., N. Doulamis, A. Doulamis and S. Kollias. Efficient Content Representation in MPEG Video Databases. In *IEEE Workshop of Content-Based Access of Image and Video Libraries*, Santa Barbara, CA, 1998.
- [7] Avrithis, Y., Y. Xirouhakis and S. Kollias. Affine Invariant Representation and Classification of Object Contours for Image and Video Retrieval. In *Third IMACS/IEEE Conference on Circuits, Systems, Communications and Computers (CSCC'99)*, Athens, Greece, July 1999.
- [8] Avrithis, Y., Y. Xirouhakis and S. Kollias. Affine-Invariant Curve Normalization for Shape-Based Retrieval. In *International Conference on Pattern Recognition (ICPR 2000)*, Barcelona, Spain, September 2000.
- [9] Avrithis, Y., Y. Xirouhakis and S. Kollias. Affine-Invariant Curve Normalization for Object Shape Representation, Classification and Retrieval. *To appear in Machine Vision and Applications*, 2001.
- [10] Avrithis, Y.S., A.D. Doulamis, N.D. Doulamis and S.D. Kollias. A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases. *Computer Vision and Image Understanding*, 75(1/2):3–24, 1999.

- [11] Ayer, S., H.S. Sawney and M. Gorkani. Model-based 2D and 3D Dominant Motion Estimation for Mosaicing and Video Representation. In *International Conference on Computer Vision*, pages 583–590, Boston, 1995.
- [12] Balslev, I. Noise Tolerance of Moment Invariants in Pattern Recognition. *Pattern Recognition Letters*, 19:1183–1189, 1998.
- [13] Bimbo, A.D., and P. Pala. Visual Image Retrieval by Elastic Matching of User Sketches. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(2):121–132, 1997.
- [14] Blake, A., R. Curwen and A. Zisserman. A Framework for Spatiotemporal Control in the Tracking of Visual Contours. *International Journal of Computer Vision*, 11(2):127–145, 1993.
- [15] Blonde, L., M. Buck, R. Galli, W. Niem, Y. Paker, W. Schmidt and G. Tomas. A Virtual Studio for Live Broadcasting: The Mona Lisa Project. *IEEE Multimedia*, 3(2):18–29, July 1996.
- [16] Blostein, S.D., and T.S. Huang. *Algorithms for Motion Estimation based on 3-D correspondences*. Motion Understanding. Martin and Aggarwal, Eds., Kluwer, Norewell, MA, 1988.
- [17] Blum, H. *A Transformation for Extracting New Descriptors of Shape*. Models for the Perception of Speech and Visual Form. W. Wathen-Dum, Ed., MIT Press, Cambridge, MA, 1967.
- [18] Bozdagi, G., A.M. Tekalp and L. Onural. An improvement to MBASIC algorithm for 3-D motion and depth estimation. *IEEE Trans. Image Processing*, 3:711–716, June 1994.
- [19] Briassouli, A., Y. Xirouhakis and A. Delopoulos. Recursive 3D Reconstruction under Orthography using Kalman Filtering. In *International Workshop on Synthetic-Natural Hybrid Coding and 3D Imaging (IWSNHC3DI'99)*, Santorini, Greece, September 1999.
- [20] Chang, S.-F., W. Chen, H.J. Meng, H. Sundaram and D. Zhong. A Fully Automated Content-Based Video Search Engine Supporting Spatiotemporal Queries. *IEEE Trans. on Circuits and Systems for Video Technology*, 8(5):602–615, 1998.
- [21] Cohen, F.S., Z. Huang and Z. Yang. Invariant Matching and Identification of Curves using B-Splines Curve Representation. *IEEE Trans. Image Processing*, 4(1):1–10, 1995.

- [22] Corridoni, J.M., A.D. Bimbo, M. Mugnaini, P. Pala and F. Turco. Pyramidal Retrieval by Color Perceptive Regions. In *IEEE Symposium on Visual Languages*, Isle of Capri, Italy, 1997.
- [23] Crowley, J.L., P. Stelmaszyk, T. Skodras and P. Puget. Measurement and Integration of 3D Structures by Tracking Edge Lines. *International Journal of Computer Vision*, 8(1):29–52, 1992.
- [24] Cutler, R., and L.S. Davis. Robust Real-Time Periodic Motion Detection, Analysis, and Applications. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):781–796, 2000.
- [25] Delopoulos, A., and Y. Xirouhakis. Robust Estimation of Motion and Shape based on Orthographic Projections of Rigid Objects. In *IEEE Image and Multidimensional Digital Signal Processing Workshop (IMDSP'98)*, pages 151–154, Alpbach, Austria, July 1998.
- [26] Delopoulos, A., M. Rangoussi, Y. Xirouhakis and A. Drosopoulos. Audio Pattern Recognition and Localization in Continuous Playout. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Baltimore, Maryland, USA, June 2001.
- [27] DeMenthon, D., and L.S. Davis. Exact and approximate solutions of the perspective-three-point problem. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:1100–1104, November 1992.
- [28] Deriche, R., and O. Faugeras. Tracking Line Segments. *Image and Vision Computing*, 8(4):261–270, 1990.
- [29] Doulamis, A.D., Y.S. Avrithis, N.D. Doulamis and S.D. Kollias. Interactive Content-Based Retrieval in Video Databases Using Fuzzy Classification and Relevance Feedback. In *IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, 1999.
- [30] Doulamis, N.D., A.D. Doulamis, Y.S. Avrithis, K.S. Ntalianis and S.D. Kollias. Efficient Summarization of Stereoscopic Video Sequences. *to appear in IEEE Trans. on Circuits and Systems for Video Technology*, 2001.
- [31] Drosopoulos, A., Y. Xirouhakis and A. Delopoulos. An Optical Camera Tracking System for Virtual Sets Applications. In *Visual, Modeling and Visualization Workshop (VMV'99)*, Erlangen, Germany, November 1999.

- [32] Drosopoulos, A., Y. Xirouhakis and A. Delopoulos. Optical Camera Tracking in Virtual Studios: Degenerate Cases. In *International Conference on Pattern Recognition (ICPR 2000)*, Barcelona, Spain, September 2000.
- [33] Egusa, Y., H. Akahori, A. Morimura and N. Wakami. An Application of Fuzzy Set Theory for an Electronic Video Camera Image Stabilizer. *IEEE Trans. Fuzzy Systems*, 3(3):351–356, 1995.
- [34] Ennesser, F., and G. Medioni. Finding Waldo, or focus of attention using local color information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):805–809, 1995.
- [35] Faugeras, O. *Three-Dimensional Computer Vision*. MIT Press, 1993.
- [36] Fieguth, P., and D. Terzopoulos. Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates. In *Computer Vision and Pattern Recognition*, pages 21–27, 1997.
- [37] Flickner, M., H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker. Query by Image and Video Content: The QBIC System. *IEEE Computer*, 28:23–32, 1995.
- [38] Frankel, C., M. Swain and W. Athitsos. Webseer: An image search engine for the world wide web. In *IEEE Conference on Computer Vision and Pattern Recognition, Demo Program*, San Juan, Puerto Rico, 1997.
- [39] Freeman, H. *Boundary Encoding and Processing*. Picture Processing and Psychopictorics. Lipkin and Rosenfeld, Eds., Academic Press, New York, 1970.
- [40] Funt, B., and G. Finlayson. Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.
- [41] Gallager, R.G. *Information Theory and Reliable Communication*. John Wiley & Sons Inc., 1968.
- [42] Gennery, D.B. Visual Tracking of Known Three-Dimensional Objects. *International Journal of Computer Vision*, 7(3):243–270, 1992.
- [43] Gevers, T., and A.W.M. Smeulders. Enigma: an image retrieval system. In *International Conference on Pattern Recognition*, Hague, Netherlands, 1992.

- [44] Gevers, T., and A.W.M. Smeulders. PicToSeek: a content-based image search engine for the World Wide Web. In *Visual Information Systems*, San Diego, CA, 1997.
- [45] Gevers, T., and A.W.M. Smeulders. Content-based image retrieval by viewpoint-invariant color indexing. *Image and Vision Computing*, 17(7):475–488, 1999.
- [46] Gibbs, S., C. Arapis, C. Breiteneder, V. Lalioti, S. Mostafawy and J. Speier. Virtual studios: an overview. *IEEE Multimedia*, 5(1):18–35, January 1998.
- [47] Golomb, S.W. *Shift register sequences*. Holden-Day, 1967.
- [48] Gong, Y., G. Proietti and C. Faloutsos. Image Indexing and Retrieval Based on Human Perceptual Color Clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, 1998.
- [49] Gu, C., and M. Kunt. Contour simplification and motion compensated coding. *Signal Processing*, 7(4-6):279–296, 1995.
- [50] Gupta, A., and R. Jain. Visual Information Retrieval. *Communications of the ACM*, 40(5):70–79, 1997.
- [51] Hampapur, A., A. Gupta, B. Horowitz, C.F. Shu, C. Fuller, J. Bach, M. Gorkani and R. Jain. Virage Video Engine. *SPIE Storage and Retrieval for Image and Video Databases V*, 3022:188–198, 1997. San Jose, CA.
- [52] Haritaoglu, I., D. Harwood and L.S. Davis.  $W^4$ : Real-Time Surveillance of People and Their Activities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):809–830, 2000.
- [53] Hayashi, M. Image Compositing Based on Virtual Cameras. *IEEE Multimedia*, 5(1):36–48, January 1998.
- [54] Hepper, D., and H. Li. Analysis of Uncovered Background Prediction for Image Sequence Coding. In *Picture Coding Symposium*, pages 192–193, 1987.
- [55] Hoetter, M. Differential Estimation of the Global Motion Parameters Zoom and Pan. *Signal Processing*, 16(3):249–265, March 1989.
- [56] Hsu, W., T.S. Chua and H.K. Pung. An Integrated Color-Spatial Approach to Content-based Image Retrieval. In *ACM Multimedia*, San Francisco, CA, 1995.

- [57] Hsu, Y.Z., H.H. Nagel and G. Rekers. New Likelihood Test Methods for Change Detection in Image Sequences. *Computer Vision, Graphics and Image Processing*, 26:73–106, 1984.
- [58] Hu, M.K. Visual Pattern Recognition by Moment Invariants. *IRE Trans. on Information Theory*, 8:179–187, 1962.
- [59] Huang, J., S.R. Kumar, M. Mitra, W.-J. Zhu and R. Zabith. Image Indexing Using Color Correlograms. In *IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997.
- [60] Huang, T.S., and A.N. Netravali. Motion and structure from feature correspondences: A review. *Proceedings of IEEE*, 82:252–269, February 1994.
- [61] Huang, T.S., and C.H. Lee. Motion and structure from orthographic projections. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:536–540, May 1989.
- [62] Huang, Z., and F.S. Cohen. Affine-invariant B-Spline Moments for Curve Matching. *IEEE Trans. Image Processing*, 5(10):1473–1480, 1996.
- [63] Hughes, D. Virtual-Studio-Ultimate 8. In *IRT Symp. on Virtual Studio Technique*, IRT, 1996, 1996.
- [64] Hung, Y.S., and H.T. Ho. A Kalman Filter Approach to Direct Depth Estimation Incorporating Surface Structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(6):570–575, June 1999.
- [65] Irani, M., and P. Anandan. A Unified Approach to Moving Object Detection in 2D and 3D Scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(6):577–589, 1998.
- [66] ISO/IEC JTC1/SC29/WG11. *MPEG-7: Context and Objectives (v.5). Doc. N1920*, 1997.
- [67] ISO/IEC JTC1/SC29/WG11. *MPEG-4 Video Verification Model Version 11.0. Doc. N2172*, 1998.
- [68] Jain, A.K. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.
- [69] Jain, A.K., Y. Zhong and S. Lakshmanan. Object Matching Using Deformable Templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(3):267–278, 1996.

- [70] Jiang, X., H. Bunke and D. Widmer-Kljajo. Skew correction of document images by focused nearest-neighbor clustering. In *International Conference on Document Analysis and Recognition*, pages 629–632, Bangalore, 1999.
- [71] Kankanhalli, M.S., B.M. Mehtre and H.Y. Huang. Color and Spatial Feature for Content-based Image Retrieval. *Pattern Recognition Letters*, 20:109–118, 1999.
- [72] Kankanhalli, M.S., B.M. Mehtre and J.K. Wu. Cluster-Based Color Matching for Image Retrieval. *Pattern Recognition*, 29(4):701–708, 1996.
- [73] Karpouzis, K., G. Votsis, Y. Xirouhakis, G. Stamou and S. Kollias. An Intelligent Multimedia System for Efficient Image Browsing and Retrieval. In *European Workshop on Content-Based Multimedia Indexing (CBMI'99)*, Toulouse, France, October 1999.
- [74] Kass, M., A. Witkin and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [75] Khotanzad, A., and Y.H. Hong. Invariant Image Recognition by Zernike Moments. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(5):489–497, 1990.
- [76] Kim, M., J.G. Jeon, J.S. Kwak, M.H. Lee and C. Ahn. Moving object segmentation in video sequences by user interaction and automatic object tracking. *Image and Vision Computing*, 19:245–260, 2001.
- [77] Lai, K.K., and R.T. Chin. Deformable Contours: Modeling and Extraction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(11):1084–1090, 1995.
- [78] Lai, S.-H., C.-W. Fu and S. Chang. A Generalized Depth Estimation Algorithm with a Single Image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(4):405–411, April 1992.
- [79] Lamport, L. *LATEX: A Document Preparation System*. Addison-Wesley Publishing Company, 1985.
- [80] Lei, Z., Y. Chan and D. Lopresti. Image Curvelet Feature Extraction and Matching. In *International Conference on Image Processing*, Santa Barbara, CA, 1997.
- [81] Li, Z.-N., O.R. Zaiane and B. Yan. C-BIRD: Content-Based Image Retrieval from Digital Libraries Using Illumination Invariance and Recognition Kernel. In *International Workshop on Database and Expert Systems Applications*, Vienna, Austria, 1998.

- [82] Lim, Y.W., and S.U. Lee. On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern Recognition*, 23(9):935–952, 1990.
- [83] Liu, Y., T.S. Huang and O.D. Faugeras. Determination of camera location from 2-D to 3-D line and point correspondences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(1):28–37, January 1990.
- [84] Longuet-Higgins, H.C. A Computer Algorithm for Reconstructing a Scene from two Projections. *Nature*, 293:133–135, September 1981.
- [85] Lowe, D.G. Robust Model-Based Motion-Tracking Through the Integration of Search and Estimation. *International Journal of Computer Vision*, 8(2):113–122, 1992.
- [86] Ma, W.Y., and B.S. Manjunath. Netra: A Toolbox for Navigating Large Image Databases. In *IEEE International Conference on Image Processing*, Santa Barbara, CA, 1997.
- [87] Makarov, A. Comparison of Background Extraction Based Intrusion Detection Algorithms. In *IEEE International Conference on Image Processing*, pages 521–524, Lausanne, Switzerland, 1996.
- [88] Makarov, A., J.M. Vesin and F. Reymond. Intrusion Detection Robust to Slow and Abrupt Lighting Changes. *Real-Time Imaging, SPIE*, 2661:44–54, 1996.
- [89] Mao, J., and A. Jain. Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models. *Pattern Recognition*, 25(2):173–188, 1992.
- [90] Marques, F., B. Llorens and A. Gasull. Prediction of Image Partitions Using Fourier Descriptors: Application to Segmentation-Based Coding Schemes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7(4):529–542, 1998.
- [91] Matas, J., R. Marik and J. Kittler. On representation and matching of multi-coloured objects. In *IEEE International Conference on Computer Vision*, Cambridge, MA, 1995.
- [92] McKenna, S.J., Y. Raja and S. Gong. Tracking colour objects using adaptive mixture objects. *Image and Vision Computing*, 17:225–231, 1999.



- [93] Mehtre, B.M., M.S. Kankanhalli and W.F. Lee. Content-based image retrieval using a composite color-shape approach. *Information Processing and Management*, 34(1):109–120, 1998.
- [94] Mendel, J.M. Tutorial in Higher Order Statistics (Spectra) in Signal Processing and System Theory. Theoretical Results and Some Applications. *Proceedings of the IEEE*, 79:278–305, 1991.
- [95] Messer, K., and J. Kittler. A region-based image database system using colour and texture. *Pattern Recognition Letters*, 20(11-13):1323–1330, 1999.
- [96] Metaxas, D., and D. Terzopoulos. Shape and Nonrigid Motion Estimation through Physics-Based Synthesis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):580–591, June 1993.
- [97] Meyer, F., and P. Bouthemy. Region-Based Tracking Using Affine Motion Models in Long Image Sequences. *CVGIP: Image Understanding*, 60(2):119–140, 1994.
- [98] Mokhtarian, F. Silhouette-Based Isolated Object Recognition through Curvature Scale Space. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(5):539–544, 1995.
- [99] Morita, T., and T. Kanade. A Sequential Factorization Method for Recovering Shape and Motion from Image Streams. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(8):858–867, August 1997.
- [100] Ngan, P.M. Motion Detection in Temporal Clutter. Technical report, Auckland, New Zealand, 1997. Technical Report 693, Industrial Research Limited.
- [101] Odobez, J.M., and P. Bouthemy. *Separation of Moving Regions from Background in an Image Sequence Acquired with a Mobile Camera*, pages 238–311. Video Data Compression for Multimedia Computing. Kluwer Academic Publisher, 1997.
- [102] Ogle, V.E., and M. Stonebraker. Chabot: Retrieval from a Relational Database of Images. *Computer*, 28(9):40–48, 1995.
- [103] ORAD. <http://www.orad.co.il>.
- [104] Ortega, J.M. *Matrix Theory: A second course*. Plenum Press, 1987.
- [105] Osher, S., and J. Sethian. Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on the Hamilton-Jacobi Formulation. *Journal of Computational Physics*, 79:12–49, 1988.

- [106] Ostuni, J., and S. Dunn. Motion from Three Weak Perspective Images Using Image Rotation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(1):64–69, January 1996.
- [107] Paragios, N., and R. Deriche. Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(3):266–280, 2000.
- [108] Paragios, N., P. Perez, G. Tziritas, C. Labit and P. Bouthemy. Adaptive Detection of Moving Objects Using Multiscale Techniques. In *IEEE International Conference on Image Processing*, pages 593–596, Lausanne, Switzerland, 1996.
- [109] Park, I.K., I.D. Yun and S.U. Lee. Color image retrieval using hybrid graph representation. *Image and Vision Computing*, 17(7):465–474, 1999.
- [110] Paschos, G., and I. Radev. Image Retrieval Based on Chromaticity Moments. In *International Conference on Image Analysis and Processing*, Venice, Italy, 1999.
- [111] Pass, G., and R. Zabih. Histogram Refinement for Content-based Image Retrieval. In *IEEE Workshop on Applications of Computer Vision*, pages 96–102, 1996.
- [112] Pavlidis, T., and F. Ali. Computer Recognition of Handwritten Numerals by Polygonal Approximation. *IEEE Trans. Systems, Man and Cybernetics*, 6:610–614, 1975.
- [113] Pentland, A., and B. Horowitz. Recovery of Nonrigid Motion and Structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):730–742, July 1991.
- [114] Pentland, A., B. Moghaddam and T. Starner. View-based and modular eigenspaces for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 1994.
- [115] Pentland, A., R.W. Picard and S. Sclaroff. Photobook: Tools for Content-Based Manipulation of Image Databases. *SPIE Storage and Retrieval for Image and Video Databases II*, 2185:34–37, 1994. San Jose, CA.
- [116] Pentland, A., R.W. Picard and S. Sclaroff. Photobook: Content-Based Manipulation of Image Databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.
- [117] Persoon, E., and K.-S. Fu. Shape Discrimination Using Fourier Descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(3):388–397, 1986.

- [118] Peterfreund, N. Robust Tracking of Position and Velocity with Kalman Filters. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(6):564–569, 1999.
- [119] Philip, J. Estimation of three-dimensional motion of rigid objects from noisy observations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(1):61–66, January 1991.
- [120] Picard, R.W., and F. Liu. A new Wold ordering for image similarity. In *IEEE International Conference on Acoustics, Sound and Speech Processing*, Adelaide, Australia, 1994.
- [121] Pless, R., T. Brodsky and Y. Aloimonos. Detecting Independent Motion: The Statistics of Temporal Continuity. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):768–773, 2000.
- [122] Poelman, C.J., and T. Kanade. A Paraperspective Factorization Method for Shape and Motion Recovery. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(3):206–218, March 1997.
- [123] Pratt, I. Shape Representation using Fourier Coefficients of the Sinusoidal Transform. Technical report, University of Manchester, 1996. Technical Report Series UMCS-96-7-1.
- [124] Puzicha, J., T. Hofmann and J.M. Buhmann. Histogram clustering for unsupervised segmentation and image retrieval. *Pattern Recognition Letters*, 20(9):899–909, 1999.
- [125] Quan, L., and T. Kanade. Affine Structure from Line Correspondences with Uncalibrated Affine Cameras. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(8):834–845, August 1997.
- [126] Ricquebourg, Y., and P. Bouthemy. Real-Time Tracking of Moving Persons by Exploiting Spatio-Temporal Image Slices. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2(8):797–808, 2000.
- [127] Rivlin, E., and I. Weiss. Local Invariants for Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(3):226–238, March 1991.
- [128] Roth, V. Content-based retrieval from digital video. *Image and Vision Computing*, 17(7):531–540, 1999.

- [129] Rothe, I., H. Susse and K. Voss. The Method of Normalization to Determine Invariants. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(4):366–375, 1996.
- [130] Rui, Y., A. She and T.S. Huang. *Image Databases and Multimedia Search. Series on Software Engineering and Knowledge Engineering 8*, pages 165–180. S.K. Chang Ed., World Scientific Publishing House in Singapore, 1998.
- [131] Rui, Y., T.S. Huang and S. Mehrotra. Content-Based Image Retrieval with Relevance Feedback in MARS. In *IEEE International Conference on Image Processing*, Santa Barbara, CA, 1997.
- [132] Salembier, P., and M. Pardas. Hierarchical Morphological Segmentation for Image Sequence Coding. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3:639–651, 1994.
- [133] Salembier, P., F. Marques, M. Pardas, R. Morros, I. Corset, S. Jeannin, L. Bouchard, F. Meyer and B. Marcotequi. Segmentation-based Video Coding System Allowing the Manipulation of Objects. *IEEE Trans. Circuits and Systems for Video Technology*, 7(1):60–73, 1997.
- [134] Samet, H., and A. Soffer. MARCO: MAP Retrieval by COntent. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):783–798, 1996.
- [135] Schmid, C., and R. Mohr. Local Grayvalue Invariants for Image Retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.
- [136] Sclaroff, S., L. Taycher and M. La Cascia. ImageRover: a content-based image browser for the World Wide Web. In *IEEE Workshop on Content-based Access and Video Libraries*, San Juan, Puerto Rico, 1997.
- [137] Seaborn, M., L. Hepplewhite and J. Stonham. PISARO: Perceptual Color and Texture Queries Using Stackable Mosaics. In *International Conference on Multimedia Computing and Systems*, Florence, Italy, 1999.
- [138] Sedgewick, R. *Algorithms*. Addison-Wesley Publishing Co., 1983.
- [139] Sethi, I.K., I. Coman, B. Day, F. Jiang, D. Li, J. Segovia-Jarez, G. Wei and B. You. Color-WISE: A system for Image Similarity Retrieval Using Color. *SPIE Storage and Retrieval for Image and Video Databases VI*, 3312:140–149, 1997.

- [140] Shapiro, L.S., A. Zisserman and M. Brady. 3D Motion Recovery via Affine Epipolar Geometry. *International Journal of Computer Vision*, 16:147–182, 1995.
- [141] Shen, D., and H. Ip. Generalized Affine Invariant Image Normalization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(5):431–440, 1997.
- [142] Sikora, T. The MPEG-4 Video Standard Verification Model. *IEEE Trans. Circuits and Systems for Video Technology*, 7(1):19–31, 1997.
- [143] Sistla, A., C. Yu and R. Venkatasubrahmanian. Similarity Based Retrieval of Videos. In *International Conference on Data Engineering*, Birmingham, UK, 1997.
- [144] Skifstad, K., and R.C. Jain. Illumination Independent Change Detection for Real World Image Sequences. *Computer Vision, Graphics and Image Processing*, 46:387–399, 1998.
- [145] Smith, J.R., and S.-F. Chang. VisualSEEk: a fully automated content-based image query system. In *ACM Multimedia*, November 1996.
- [146] Smith, J.R., and S.-F. Chang. *Querying by color regions using the VisualSEEk content-based visual query system*. Intelligent Multimedia Information Retrieval, IJCAI. M.T. Maybury, 1997.
- [147] Smith, J.R., and S.-F. Chang. SaFe: A General Framework for Integrated Spatial and Feature Image Search. In *IEEE Workshop on Multimedia Signal Processing*, 1997.
- [148] Soatto, S., and P. Perona. Reducing 'Structure From Motion': A General Framework for Dynamic Vision, Part 2: Implementation and Experimental Assessment. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(9):943–960, September 1998.
- [149] Spetsakis, M., and J. Aloimonos. Structure from Motion Using Line Correspondences. *International Journal of Computer Vision*, 4:171–183, 1990.
- [150] Steinbach, E., and B. Girod. Estimation of Rigid Body Motion and Scene Structure from Image Sequences Using a Novel Epipolar Transform. In *IEEE ICASSP*, pages 1911–1914, Atlanta, 1996.
- [151] Stricker, M., and A. Dimai. Color Indexing with Weak Spatial Constraints. *SPIE Storage and Retrieval for Still Image and Video Databases IV*, 2670:29–39, 1996.
- [152] Swain, M.J., and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

- [153] Swanson, M.D., and A.H. Tewfik. Affine-Invariant Multiresolution Image Retrieval using B-Splines. In *International Conference on Image Processing*, Santa Barbara, CA, 1997.
- [154] Tao, Y., and W. Grosky. Spatial Color Indexing: A Novel Approach for Content-Based Image Retrieval. In *International Conference on Multimedia Computing and Systems*, Florence, Italy, 1999.
- [155] Tekalp, M. *Digital Video Processing*. Prentice Hall, 1995.
- [156] Terzopoulos, D., and D. Metaxas. Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):703–714, July 1991.
- [157] Thompson, W.B., P. Lechleider and E.R. Stuck. Detecting Moving Objects Using the Rigidity Constraint. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2(15):162–166, 1993.
- [158] Tirakis, A., P. Katalagarianos, M. Papathomas and C. Hamilakis. Distributed audio-Visual Archives Network. In *IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, 1999.
- [159] Tomasi, C., and T. Kanade. Shape and Motion from Image Streams under Orthography: a Factorization Method. *International Journal of Computer Vision*, 9(2):137–154, 1992.
- [160] Torres, L., and M. Kunt. *Video Coding: The Second Generation Approach*. Kluwer, Boston, MA, 1996.
- [161] Tsai, R.Y., and T.S. Huang. Estimating three-dimensional motion parameters of a rigid planar patch. *IEEE Trans. ASSP*, 29(6):1147–1152, December 1981.
- [162] Tsai, R.Y., and T.S. Huang. Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6:13–27, January 1984.
- [163] Tsai, R.Y., T.S. Huang and W.L. Zhu. Estimating 3-D motion parameters of a rigid planar patch II: Singular value decomposition. *IEEE Trans. ASSP*, 30(4):525–534, 1982.

- [164] Tsang, P.W.M. A Genetic Algorithm for Affine Invariant Recognition of Object Shapes from Broken Boundaries. *Pattern Recognition Letters*, 18:631–639, 1997.
- [165] Tsechpenakis, G., Y. Xirouhakis and A. Delopoulos. 2D Motion Estimation in the Extraction of 3D Motion and Structure of Rigid Objects. In *International Workshop on Synthetic-Natural Hybrid Coding and 3D Imaging (IWSNHC3DI'99)*, Santorini, Greece, September 1999.
- [166] Tsechpenakis, G., Y. Xirouhakis and A. Delopoulos. Main Mobile Object Detection and Localization in Video Sequences. In *International Conference on Visual Information Systems (VISUAL2000)*, Lyon, France, November 2000.
- [167] Ullman, S. *The Interpretation of Visual Motion*. MIT Press, Cambridge, MA, 1979.
- [168] Votsis, G., A. Drosopoulos, G. Akrivas, V. Tzouvaras and Y. Xirouhakis. An MPEG-7 Compliant Integrated System for Video Archiving, Characterization and Retrieval. In *IASTED International Conference on Signal and Image Processing (SIP2000)*, Las Vegas, Nevada, November 2000.
- [169] Votsis, G., Y. Xirouhakis, K. Karpouzis and S. Kollias. Efficient Use of Content-Based Retrieval in Quick Browsing: A Realization. In *IEEE International Workshop on Multimedia Signal Processing (MMSP'99)*, Copenhagen, Denmark, September 1999.
- [170] Votsis, G., Y. Xirouhakis, S. Kollias and A. Stafylopatis. Feature Extraction and 3D Recovery of Human Faces in Video Sequences. In *The International Conference on Information Systems Analysis and Synthesis*, Orlando, Florida, July 1998.
- [171] Wang, J., W.-J. Yang and R. Acharya. Efficient Access to and Retrieval from a Shape Image Database. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, Santa Barbara, CA, 1998.
- [172] Wang, J., W. Yang and R. Acharya. Color Clustering Techniques for Color-Content-Based Image Retrieval from Image Databases. In *International Conference on Multimedia Computing and Systems*, Ottawa, Canada, 1997.
- [173] Wang, J.-Y., and F.S. Cohen. Part II: 3-D Object Recognition and Shape Estimation from Image Contours Using B-Splines, Shape Invariant Matching, and Neural Network. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(1):13–23, January 1994.
- [174] Watt, A., and M. Watt. *Advanced Animation and Rendering Techniques*. ACM Press, New York, 1992.

- [175] Weng, J., N. Ahuja and T.S. Huang. Optimal Motion and Structure Estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(9):864–884, September 1993.
- [176] Weng, J., T.S. Huang and N. Ahuja. Motion and structure from point correspondences with error estimation: Planar surfaces. *IEEE Trans. Signal Processing*, 39(12):2691–2717, December 1991.
- [177] Weng, J., T.S. Huang and N. Ahuja. Motion and structure from line correspondences. Closed-form solution, uniqueness and optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(3):318–336, March 1992.
- [178] Wildes, R. A Measure of Motion Saliency for Surveillance Applications. In *IEEE International Conference on Image Processing*, 1998.
- [179] Williams, P.S., and M.D. Alder. Segmentation of Natural Images for CBIR. In *International Conference on Pattern Recognition*, Brisbane, Australia, 1998.
- [180] Wixson, L. Detecting Salient Motion by Accumulating Directionally-Consistent Flow. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):774–780, 2000.
- [181] Xirouhakis, Y. *Estimating Structure and Motion of Three-Dimensional Objects from Video Sequences*. Diploma Thesis, National Technical University of Athens, Athens 1997.
- [182] Xirouhakis, Y., A. Drosopoulos and A. Delopoulos. A Novel Approach for the Estimation of Camera Motion in Virtual Studios Applications. In *International Workshop on Synthetic-Natural Hybrid Coding and 3D Imaging (IWSNHC3DI'99)*, Santorini, Greece, September 1999.
- [183] Xirouhakis, Y., A. Drosopoulos and A. Delopoulos. Camera Motion Estimation Using 3D-to-2D Line Correspondences. In *IEEE Nordic Signal Processing Symposium (NORSIG2000)*, Kolmarden, Sweden, June 2000.
- [184] Xirouhakis, Y., A. Drosopoulos and A. Delopoulos. Efficient Optical Camera Tracking in Virtual Sets. *IEEE Trans. Image Processing*, 10(4):609–622, April 2001.
- [185] Xirouhakis, Y., A. Tirakis and A. Delopoulos. An Efficient Graph Representation for Image Retrieval based on Color Composition. In *Third IMACS/IEEE Conference on Circuits, Systems, Communications and Computers (CSCC'99)*, Athens, Greece, July 1999.



- [186] Xirouhakis, Y., and A. Delopoulos. A Comparative Study on 3D Motion Estimation under Orthography. In *IEEE Nordic Signal Processing Symposium (NORSIG2000)*, Kolmarden, Sweden, June 2000.
- [187] Xirouhakis, Y., and A. Delopoulos. Image Retrieval on the basis of Color Composition Information. In *IASTED International Conference on Signal and Image Processing (SIP2000)*, Las Vegas, Nevada, November 2000.
- [188] Xirouhakis, Y., and A. Delopoulos. Least Squares Estimation of 3D Shape and Motion of Rigid Objects from their Orthographic Projections. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(4):393–399, April 2000.
- [189] Xirouhakis, Y., and A. Delopoulos. Using Graphs for the Representation of the Color Composition Attribute in Image Retrieval. *Submitted to Image and Vision Computing*, 2001.
- [190] Xirouhakis, Y., G. Tsechpenakis and A. Delopoulos. User Choices for Efficient 3D Motion and Shape Extraction from Orthographic Projections. In *6th IEEE International Conference on Electronics, Circuits and Systems*, Paphos, Cyprus, September 1999.
- [191] Xirouhakis, Y., G. Tsechpenakis and A. Delopoulos. Fast Mobile Object Detection and Localization in Video Sequences. *Submitted to Machine Vision and Applications*, 2001.
- [192] Xirouhakis, Y., G. Votsis and A. Delopoulos. Estimation of 3D Motion and Structure of Human Faces. In *Third European Robotics, Intelligent Systems, Control Conference*, Athens, Greece, June 1998.
- [193] Xirouhakis, Y., G. Votsis, K. Karpouzis and S. Kollias. Efficient Browsing in Multimedia Databases using Intelligent Agents and Content-Based Retrieval Schemes. In *IEEE International Workshop on Multimedia Signal Processing (MMSP'98)*, Los Angeles, CA, USA, December 1998.
- [194] Xirouhakis, Y., V. Mathioudakis and A. Delopoulos. An Efficient Algorithm for Mobile Object Localization in Video Sequences. In *Visual, Modeling and Visualization Workshop (VMV'99)*, Erlangen, Germany, November 1999.
- [195] Xirouhakis, Y., Y. Avrithis and S. Kollias. Image Retrieval and Classification using Affine Invariant B-spline Representation and Neural Networks. In *IEE Colloquium on Neural Networks in Multimedia Interactive Systems*, London, UK, October 1998.

- [196] Xu, G., and N. Sugimoto. A Linear Algorithm for Motion from Three Weak Perspective Images Using Euler Angles. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(1):54–57, January 1999.
- [197] Yalamanchili, S., W.N. Martin and J.K. Aggarwal. Extraction of Moving Object Descriptions Via Differencing. *Computer Graphics and Image Processing*, 18:188–201, 1982.
- [198] Yang, Z., and F.S. Cohen. Image Registration and Object Recognition Using Affine Invariants and Convex Hulls. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(7):934–946, 1999.
- [199] Zhang, R., P.-S. Tsai, J. E. Cryer and M. Shah. Shape From Shading: A Survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(8):690–706, August 1999.
- [200] Zhang, Z. Estimating Motion and Structure from Correspondences of Line Segments between Two Perspective Images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(12):1129–1139, December 1995.
- [201] Zhong, Y., A.K. Jain and M.-P. Dubuisson-Jolly. Object Tracking Using Deformable Templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(5):544–549, 2000.