



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
Και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής &
Υπολογιστών

Εξαγωγή Χαρακτηριστικών και Κίνησης Αντικειμένων:
Εικονικά Στούντιο

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΑΘΑΝΑΣΙΟΥ Η. ΔΡΟΣΟΠΟΥΛΟΥ

Πτυχιούχου Πληροφορικής Πανεπιστημίου Ιωαννίνων (1998)

Αθήνα, Ιούλιος 2005



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής &
Υπολογιστών

Εξαγωγή Χαρακτηριστικών και Κίνησης Αντικειμένων: Εικονικά Στούντιο

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΑΘΑΝΑΣΙΟΥ Η. ΔΡΟΣΟΠΟΥΛΟΥ

Πτυχιούχου Πληροφορικής Πανεπιστημίου Ιωαννίνων (1998)

Συμβουλευτική Επιτροπή: Στέφανος Κόλλιας
 Γεώργιος Καραγιάννης
 Νικόλαος Μήτρου

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 1^η Ιουλίου 2005.

...
Σ. Κόλλιας
Καθηγητής Ε.Μ.Π.

...
Γ. Καραγιάννης
Καθηγητής Ε.Μ.Π.

...
Ν. Μήτρου
Καθηγητής Ε.Μ.Π.

...
Α. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

...
Π. Τσανάκας
Καθηγητής Ε.Μ.Π.

...
Π. Μαραγκός
Καθηγητής Ε.Μ.Π.

...
Αν. Ντελόπουλος
Επ. Καθηγητής
Αριστοτέλειου Πανεπιστημίου

Αθήνα, Ιούλιος 2005

...

ΑΘΑΝΑΣΙΟΣ Η. ΔΡΟΣΟΠΟΥΛΟΣ

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

© 2005 - All rights reserved

Περιεχόμενα

1	Εισαγωγή	1
1.1	Επαυξημένη Πραγματικότητα	1
1.1.1	Σύγκριση επαυξημένης και εικονικής πραγματικότητας	1
1.1.2	Πεδία εφαρμογής επαυξημένης πραγματικότητας	2
1.1.3	Προδιαγραφές συστημάτων επαυξημένης πραγματικότητας	3
1.2	Εικονικά Στούντιο	4
1.3	Ανάλυση της Διδακτορικής Διατριβής	6
2	Υπολογισμός της 3Δ κίνησης της Κάμερας σε Εικονικά Στούντιο	9
2.1	Εισαγωγή - Αναδρομή	9
2.2	Περιγραφή του συστήματος	12
2.2.1	Σχεδιασμός ενός συστήματος οπτικής παρακολούθησης για Εικονικά Στούντιο	12
2.2.2	Σχεδιασμός του κυανού πετάσματος	13
2.2.3	Εντοπισμός των δισδιάστατων χαρακτηριστικών	14
2.2.4	Υπολογισμός των παραμέτρων της τρισδιάστατης κίνησης της κάμερας	15
2.3	Κατασκευή του δυαδικού πίνακα του πετάσματος	16
2.3.1	Ένα ειδικό δισδιάστατο δυαδικό πεδίο	17
2.3.2	Ο αλγόριθμος κατασκευής	19
2.4	Υπολογισμός της θέσης και του προσανατολισμού της κάμερας	20
2.5	Ασυμπτωτική ανάλυση σφάλματος	28
2.6	Πειράματα και συζήτηση	30
2.6.1	Πειράματα σε περιβάλλον προσομοίωσης	31
2.6.2	Πειράματα σε πραγματικό περιβάλλον	37
2.7	Συμπεράσματα και επεκτάσεις	37
3	Εντοπισμός και Παρακολούθηση Αντικειμένων σε Εικονικά Στούντιο	41
3.1	Εισαγωγή - Αναδρομή	41
3.2	Περιγραφή του συστήματος	44
3.2.1	Το εικονικό στούντιο	44
3.2.2	Εντοπισμός ηθοποιού στο εικονικό στούντιο	45
3.3	Εντοπισμός και παρακολούθηση αντικειμένου	46
3.3.1	Εύρεση θέσης σημείου	46
3.3.2	Εκτίμηση του χώρου που καταλαμβάνει ένα αντικείμενο	48
3.3.3	Παρακολούθηση πολλαπλών αντικειμένων	49
3.3.3.1	Περίγραμμα αντικειμένου	49
3.3.3.2	Επικάλυψη	49

3.3.3.3	Προσέγγιση τομής με χρήση ενεργών περιγραμμάτων .	52
3.4	Συμπεράσματα και επεκτάσεις.....	52
4	Εξαγωγή χαρακτηριστικών προσώπου	55
4.1	Εισαγωγή - Αναδρομή.....	55
4.2	Μερικά από τα χρησιμοποιούμενα εργαλεία	57
4.2.1	Η κατάτμηση χαρακτηριστικών του προσώπου μέσω ανάλυσης ελάχιστου - μέγιστου	57
4.2.2	Μη επιβλεπόμενη κατάτμηση χρώματος - υφής	58
4.2.3	Κατάτμηση και ενεργά περιγράμματα	60
4.3	Η προτεινόμενη μέθοδος	61
4.3.1	Επισκόπηση του συστήματος	61
4.3.2	Κατάτμηση και καθορισμός των αρχικών εκτιμήσεων.....	63
4.3.3	Εμπλουτισμός και μεγέθυνση των αρχικών εκτιμήσεων	68
4.3.4	Καθορισμός αντικειμένων και υπολογισμός της επικρατούσας γωνίας	69
4.3.5	Επιλογή και ονοματισμός των βασικών χαρακτηριστικών	71
4.3.6	Υπολογισμός των παραμέτρων ορισμού και δυναμικής κίνησης του προσώπου	73
4.4	Αποτελέσματα προσομοιώσεων.....	73
4.5	Συμπεράσματα και επεκτάσεις.....	76
5	Αλληλεπίδραση Πραγματικού και Εικονικού Ηθοποιού σε Εικονικά Στούντιο	79
5.1	Εισαγωγή	79
5.2	Εξαγωγή χαρακτηριστικών για την εκτίμηση των εκφράσεων του προσώπου	80
5.3	Ολοκλήρωση του συστήματος - Ένα γενικό σενάριο	80
5.4	Συμπεράσματα και επεκτάσεις.....	85
6	Συμπεράσματα και Επεκτάσεις	87
	Βιβλιογραφία	89
	Κατάλογος Δημοσιεύσεων του συγγραφέα	97
	Βιογραφικό Σημείωμα	99

Κατάλογος Σχημάτων

1.1	Η ταξινόμηση της αναμειγμένης πραγματικότητας κατά τον Milgram ..	2
2.1	Το προτεινόμενο σύστημα υπολογισμού κίνησης σε μπλοκ-διάγραμμα. ..	16
2.2	Η γεωμετρία του προβλήματος. Το σύστημα συντεταγμένων κινείται με το επίπεδο του CCD (camera-based).	21
2.3	Το κυανό πέτασμα (α) Ένα κυανό πέτασμα κατασκευασμένο με δύο αποχρώσεις (β) Το τμήμα του κυανού πετάσματος που καταγράφεται από την κάμερα σημειωμένο με κόκκινο χρώμα.	32
2.4	Όλα τα βήματα της προτεινόμενης μεθοδολογίας για τον υπολογισμό της κίνησης της κάμερας στο παράδειγμα της ενότητας 2.6.1: (α) το καρέ που καταγράφηκε (βλ. Σχήμα 2.3(β)), το αποτέλεσμα της ανίχνευση ακμών, (γ) το πλέγμα γραμμών αναφοράς που εξάχθηκε (δ) το καρέ όπως ανακατασκευάστηκε χρησιμοποιώντας τις παραμέτρους κίνησης που υπολογίστηκαν.	33
2.5	Περίπτωση απουσίας γραμμών αναφοράς (α) Καταγεγραμμένο καρέ από το οποίο απουσιάζουν κάποιες από τις γραμμές αναφοράς του πλέγματος και (β) οι γραμμές αναφοράς που εντοπίζονται.	35
2.6	Εκτιμήσεις (απλή γραμμή) και πραγματικές τιμές (διακεκομμένη γραμμή) των παραμέτρων της 3Δ κίνησης της κάμερας συναρτήσει του χρόνου, για μία τυχαία ακολουθία (α) η γωνία περιστροφής, (β) η εστιακή απόσταση, (γ) η πρώτη συνιστώσα του άξονα περιστροφής και (δ) η πρώτη συνιστώσα του διανύσματος μετατόπισης.	36
2.7	Πείραμα σε φυσικό περιβάλλον: (α) το εικονικό στούντιο που κατασκευάστηκε υπό κλίμακα, (β) ένα καρέ που καταγράφεται από μία πραγματική κάμερα, (γ) οι ακμές που εξάγονται, και (δ) το καρέ που αποδίδεται από μία εικονική κάμερα χρησιμοποιώντας τις παραμέτρους της 3Δ κίνησης που υπολογίστηκαν.	38
3.1	Απεικόνιση του προβλήματος εντοπισμού σημείου (α) Οι κάμερες C_1 και C_2 βρίσκονται πάνω στην ευθεία $y = a$ (β) Η γενική περίπτωση κατά την οποία οι κάμερες C_1, C_2 βρίσκονται σε αυθαίρετες θέσεις	46
3.2	Υπολογισμός των άγνωστων γωνιών f_1 και f_2 βάσει των γωνιών λήψης των καμερών b_1 και b_2 και της γωνίας d που σχηματίζεται από τον άξονα x και την ευθεία C_1C_2	47
3.3	Εκτίμηση του χώρου που καταλαμβάνει ένα αντικείμενο υπολογίζοντας τη θέση των σημείων A_{1-6}	48
3.4	Ένα απλό σενάριο (α) Το καρέ της αριστερής κάμερας (β) Το καρέ της δεξιάς κάμερας (γ) Η πραγματική περιοχή που καλύπτει το αντικείμενο (ο) και η εκτίμηση της (x)	49

3.5	Περίπτωση επικαλυπτόμενων αντικειμένων (α) Το καρέ της κάμερας στα αριστερά (β) Το καρέ της κάμερας στα δεξιά (γ) Κάτοψη ($C1, C2$ οι δύο κάμερες)	50
3.6	Εντοπισμός των οριακών σημείων της τομής των αντικειμένων βάσει των αποστάσεων μεταξύ του τρέχοντος και προηγούμενων περιγραμμάτων (α) Απόσταση των περιγραμμάτων, αριστερόστροφα (β) Απόσταση των περιγραμμάτων, δεξιόστροφα (γ) Περίγραμμα των συνδεδεμένων αντικειμένων (με x σημειώνονται τα οριακά σημεία της τομής όπως υπολογίστηκαν)	51
3.7	Αρχικοποίηση και παραμόρφωση του snake	52
3.8	Η περίπτωση δύο επικαλυπτόμενων αντικειμένων	53
4.1	Η ροή της μεθόδου	62
4.2	Τρία αντιπροσωπευτικά καρέ εισόδου που απεικονίζουν πρόσωπα από μία φυσική ακολουθία βίντεο. Η μεθοδολογία που παρουσιάζεται αντιμετωπίζει εικόνες με ποικιλία στην τοποθέτηση του προσώπου (περιστροφή και προσανατολισμός), στις συνθήκες φωτισμού και στην κλίμακα απεικόνισης	63
4.3	Πάνω σειρά: Κάθετες και οριζόντιες προβολές ολοκληρώματος για τις εικόνες ακμών των τριών εικόνων εισόδου. Κάτω σειρά: Κάθετες και οριζόντιες προβολές ολοκληρώματος για τις <i>inverse grayscale</i> εικόνες των τριών παραδειγμάτων εισόδου	64
4.4	Συνδυασμός των προβολών ολοκληρώματος για τα τρία παραδείγματα ..	64
4.5	Το γινόμενο των οριζόντιων και κάθετων διανυσμάτων παράγει ένα χάρτη M , όπου εμφανίζονται κοιλάδες και λόφοι που σχηματίζουν τα διάφορα αντικείμενα	65
4.6	Οι αρχικές εκτιμήσεις M_0 των αντικειμένων για τα τρία αντιπροσωπευτικά παραδείγματα	65
4.7	Ο χρωματικός τόνος Hue των τριών παραδειγμάτων	66
4.8	Οι βελτιωμένες εκτιμήσεις για τα τρία παραδείγματα	68
4.9	Οι εκτιμήσεις μετά τον εμπλουτισμό και την εξάπλωση των αρχικών αντικειμένων	70
4.10	Αρχικοποίηση και τελική προσέγγιση μέσω των ενεργών περιγραμμάτων ..	70
4.11	Έξοδος του ασαφούς συστήματος που καθορίζει την επικρατούσα γωνία. Οι τριγωνικές συναρτήσεις συμμετοχής για κάθε αντικείμενο αθροίζονται, σχηματίζοντας συγκεντρώσεις γύρω από την επικρατούσα γωνία θ_{ad}	71
4.12	Οι εικόνες εισόδου περιστραμμένες στο επίπεδο, βάσει των αυτόματα υπολογισμένων γωνιών κλίσης	71
4.13	Το περιορισμένο, βάσει της επικρατούσας γωνίας, σύνολο των υποψήφιων αντικειμένων. Στο πάνω αριστερά μέρος κάθε εικόνας απεικονίζεται η επικρατούσα τριάδα αντικειμένων (όπου αυτή υπάρχει) περιστραμμένη κατά την επικρατούσα γωνία	72
4.14	Το κανονικοποιημένο γεωμετρικό πρότυπο για τα μάτια και το στόμα ..	72
4.15	Επιλογή και προσδιορισμός της ταυτότητας των χαρακτηριστικών του προσώπου που έχουν ανιχθευθεί	73
4.16	Τα σημεία που αντιπροσωπεύουν ένα υποσύνολο των παραμέτρων ορισμού του προσώπου FDP σύμφωνα με το πρότυπο MPEG-4	74

4.17	Δοκιμή με αρχική εικόνα (α) που περιέχει ανεπαρκώς καταταμημένο πρόσωπο, (β) τα αντικείμενα που ανιχνεύθηκαν και η μοντελοποίησή τους μέσω ενεργών περιγραμμάτων, (γ) τα χαρακτηριστικά του προσώπου που ανιχνεύθηκαν, (δ) βελτίωση της αρχικής κατάτμησης μέσω ενεργών περιγραμμάτων, (ε) αρχική εικόνα μετά την βελτιωμένη κατάτμηση (στ) τα χαρακτηριστικά του προσώπου όπως προσδιορίστηκαν τελικά.....	76
4.18	Δυναμική κίνηση ομοιώματος προσώπου βάσει των υπολογισμένων παραμέτρων FAP σύμφωνα με το πρότυπο MPEG-4: (α) ουδέτερη έκφραση, (β) έκφραση που αντιστοιχεί στο Σχήμα 4.16(γ)	77
5.1	Προσομοίωση εικονικού στούντιο: εκτίμηση θέσης στο επίπεδο του αντικειμένου και του ηθοποιού	81
5.2	Εκτίμηση θέσης στο επίπεδο του αντικειμένου και του ηθοποιού όταν το αντικείμενο δεν παρουσιάζεται ολόκληρο στα καρέ που καταγράφονται	82
5.3	Σκηνή στην οποία έχει εισαχθεί αυτόματα εικονικός ηθοποιός (νοσοκόμα). Στο (γ) φαίνεται η εκτίμηση για το χώρο που καταλαμβάνουν τα φυσικά αντικείμενα, καθώς και η θέση στην οποία τοποθετείται ο εικονικός ηθοποιός.....	83
5.4	Είσοδος για την αναγνώριση έκφρασης του προσώπου.	84
5.5	Ανάλυση εικόνας με στόχο την αναγνώριση της έκφρασης του προσώπου (α) Περιοχές ενδιαφέροντος για εξαγωγή χαρακτηριστικών του προσώπου (β) Ένα υποσύνολο των παραμέτρων προσδιορισμού των χαρακτηριστικών του προσώπου (FDP)	85

Κατάλογος Πινάκων

2.1	Κατασκευάζοντας το δυαδικό χάρτη του κυανού πετάσματος· άνω σειρά: η ακολουθία των διαφορών ολίσθησης, κάτω σειρά: η ακολουθία των ολισθήσεων.	19
2.2	Εντοπισμός γραμμών αναφοράς που απουσιάζουν από το πλέγμα. Οι στήλες από αριστερά προς τα δεξιά: παράμετροι γραμμών όπως προκύπτουν από τον μετασχηματισμό Hough, οι τελευταίες μετά από κανονικοποίηση (βάσει της μεγέθυνσης και των πραγματικών διαστάσεων του CCD), οι κανονικοποιημένες αποστάσεις μεταξύ των στοιχισμένων γραμμών σε ακολουθιακά ζεύγη όπως αυτές υπολογίζονται από την Πρόταση 2.3.	35

ΠΡΟΛΟΓΟΣ

Η παρούσα διδακτορική διατριβή είναι το αποτέλεσμα μίας εξαιρετικά ενδιαφέροντος και ανταποδοτικής πορείας στο εργαστήριο Ψηφιακής Εικόνας, Βίντεο και Πολυμέσων της Σχολής Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών, του ΕΜΠ. Κατά τη διάρκεια αυτής της πορείας, προσπάθησα να παρακολουθώ τις εξελίξεις, να κατανοήσω και να συμμετάσχω ενεργά στην ερευνητική κοινότητα που δραστηριοποιείται στα πεδία της Όρασης Υπολογιστών και της Κατανόησης του Περιεχομένου Ακολουθιών Εικόνων. Παράλληλα, επιχείρησα να ερευνήσω σε βάθος το πρόβλημα που μονοπώλησε το ενδιαφέρον μου και να βοηθήσω στην επίλυση του, με εργασίες όσο το δυνατόν πιο τεκμηριωμένες και με τη φιλοδοξία να αποτελέσουν μία βάση για τη μελλοντική εξέλιξη των τεχνικών αντιμετώπισής του.

Αφετηρία αλλά και καθοδηγητής σε όλη αυτή την πορεία υπήρξε ο κ. Στέφανος Κόλλιας, Καθηγητής ΕΜΠ, τον οποίο και ευχαριστώ για τη γνώση και τη συμπαράσταση που μου προσέφερε απλόχερα. Επίσης θα ήθελα να ευχαριστήσω θερμά τα μέλη της τριμελούς επιτροπής Γ. Καραγιάννη και Ν. Μήτρου, Καθηγητές ΕΜΠ. Ιδιαίτερη μνεία οφείλω στους κκ. Αναστάσιο Ντελόπουλο, Επίκουρο Καθηγητή ΑΠΘ, και Ιωάννη Ξηρουχάκη, Διδάκτωρ ΕΜΠ, γιατί ανέλαβαν πρώτοι να με ρίξουν στα βαθιά και να με ξεναγήσουν εκεί, ενώ έκτοτε με τιμούν και με τη φιλία τους. Η αξία και τα αποτελέσματα της συνύπαρξης και της συνεργασίας, μου αποκαλύφθηκαν με τη σημαντική συνεισφορά όλων των μελών του εργαστηρίου ΨΕΕΒΠ. Για τη συνεργασία στις επιστημονικές εργασίες, που δημοσιεύτηκαν και αποτελούν το περιεχόμενο της παρούσας διατριβής, θέλω να ευχαριστήσω τους Γεώργιο Βότση, Κωνσταντίνο Καρπούζη, Ιωάννη Αβρίθη και Γιώργο Ακρίβα, Διδάκτορες ΕΜΠ, Θέμη Μπαλωμένο, Αμαρυλλίς Ραουζαίου, Βασίλη Τζουβάρα και Σπύρο Ιωάννου, Υποψήφιους Διδάκτορες ΕΜΠ. Τέλος να ευχαριστήσω τους φίλους μου, που μου παρέχουν την προσωπική ηρεμία και τη συναισθηματική ισορροπία και να αφιερώσω την παρούσα διατριβή στην οικογένειά μου, καθοριστικό παράγοντα όπως και σε κάθε βήμα της ζωής μου.

Αθανάσιος Δροσόπουλος
Αθήνα, Ιούλιος 2005

ΠΕΡΙΛΗΨΗ

Στην παρούσα διδακτορική διατριβή παρουσιάζεται ένα πρωτότυπο οπτικό σύστημα εικονικού στούντιο. Συγκεκριμένα, προτείνεται μία μεθοδολογία κατασκευής ενός κυανού πετάσματος δύο αποχρώσεων, το οποίο περιλαμβάνει τα απαραίτητα χαρακτηριστικά για τον υπολογισμό της τρισδιάστατης κίνησης της κάμερας, χωρίς τη χρήση ηλεκτρομηχανικού εξοπλισμού. Παράλληλα, παρουσιάζεται μία αποτελεσματική λύση για τον υπολογισμό της τρισδιάστατης κίνησης της κάμερας βάσει αντιστοιχίσεων ευθειών. Τα αποτελέσματα προσομοιώσεων, πραγματικών πειραμάτων και της ασυμπτωτικής ανάλυσης σφάλματος επιβεβαιώνουν την επιτυχία του συστήματος. Με τον τρόπο αυτό παράγονται ακολουθίες μεικτής πραγματικότητας, χωρίς περιορισμούς στην κίνηση είτε των αντικειμένων και των ηθοποιών ή των καμερών. Στη συνέχεια, αντιμετωπίζεται το ζήτημα του εντοπισμού και της παρακολούθησης όλων των κινούμενων ή ακίνητων αντικειμένων και ηθοποιών. Μία απλή μέθοδος που εξασφαλίζει την παρακολούθηση σε πραγματικό χρόνο, επωφελούμενη και από τις ιδιότητες του πρωτότυπου εικονικού στούντιο, επιτρέπει την αυτόματη εισαγωγή ψηφιακών αντικειμένων και ηθοποιών. Τέλος, στην περαιτέρω προσπάθεια εμπλουτισμού της παραγόμενης ακολουθίας βίντεο εξετάζεται και παρουσιάζονται λύσεις σχετικά με τη δυνατότητα επικοινωνίας και αλληλεπίδρασης μεταξύ των πραγματικών και των εικονικών ηθοποιών. Παρουσιάζεται μία μεθοδολογία για την εξαγωγή των χαρακτηριστικών του ανθρώπινου προσώπου με στόχο την εκτίμηση της συμπεριφοράς του ατόμου και ειδικότερα της έκφρασης του προσώπου του, ώστε η τελευταία να χρησιμοποιηθεί για την ανάδραση του συστήματος προς τους ηθοποιούς. Η διατριβή καταλήγει με την παρουσίαση ενός γενικού σεναρίου επαυξημένου εικονικού στούντιο και πειραματικά αποτελέσματα σχετικά με τη συμπεριφορά των αναφερόμενων μεθόδων, υπό το σενάριο αυτό.

Λέξεις Κλειδιά

Εικονικό Στούντιο, Αρχέτυπα Πολυώνυμα, Υπολογισμός 3Δ Κίνησης Κάμερας, Εξαγωγή Χαρακτηριστικών Προσώπου, Εικονικός Ηθοποιός

ABSTRACT

Present thesis presents a prototype optical virtual studio system. In particular, a methodology for the construction of a two tone blue screen is presented, where the latter includes the appropriate features for the estimation of a camera's 3D motion parameters, without the use of any electromechanical equipment. In parallel, an efficient solution for the calculation of the camera's motion is introduced, based on line correspondences. The results of simulations, real experiments and error analysis prove the robustness of the system. In this way, mixed reality sequences are generated, without posing constraints on the object's, actor's, or camera's movement. The next step deals with the localization and tracking, over time, of moving or statical objects on the set. A simple method is presented that ensures tracking in real time, taking advantage of the properties of the prototype virtual studio, and allows for the automatic insertion of virtual objects and actors in the scene. Finally, in an attempt to further enhance the generated sequence, solutions concerning the communication and interaction between virtual and human actors are examined. A methodology is presented for the extraction of human facial features in order to extract indications concerning the actor's behavior, which in turn will provide the basis for the system's feedback. The dissertation concludes, presenting and discussing simulations performed under a general scenario of an augmented virtual studio.

Keywords

Virtual Studio, Primitive Polynomials, Camera Motion Estimation, Facial Features Extraction, Virtual Actor

Κεφάλαιο 1

Εισαγωγή

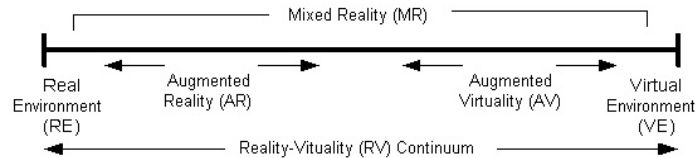
1.1 Επαυξημένη Πραγματικότητα

Η επαυξημένη πραγματικότητα (augmented reality) αποτελεί μία αναπτυσσόμενη περιοχή στην έρευνα για την εικονική πραγματικότητα. Ο κόσμος γύρω μας περιέχει ένα πλούτο πληροφοριών που είναι δύσκολο να αναπαρασταθεί σε έναν ηλεκτρονικό υπολογιστή. Αυτό αποδεικνύεται και από τους κόσμους που χρησιμοποιούνται στα εικονικά περιβάλλοντα. Αυτοί είναι είτε υπεραπλουστευμένοι, όπως τα περιβάλλοντα που χρησιμοποιούνται σε ψυχαγωγικά λογισμικά ή παιχνίδια, ή εναλλακτικά, το σύστημα που μπορεί να δημιουργήσει ένα περιβάλλον με μεγαλύτερο ρεαλισμό κοστίζει υπερβολικά, όπως ένας εξομοιωτής πτήσης. Ένα σύστημα επαυξημένης πραγματικότητας δημιουργεί μία σύνθετη όψη για το χρήστη. Είναι ένας συνδυασμός της πραγματικής σκηνής που βλέπει ο χρήστης και μίας εικονικής σκηνής που δημιουργείται από τον υπολογιστή και τελικά εμπλουτίζει το περιβάλλον με επιπρόσθετη πληροφορία. Τα πεδία εφαρμογής που περιγράφονται στην υποενοότητα 1.1.2 αποκαλύπτουν ότι ο εμπλουτισμός αυτός μπορεί να γίνει με διάφορους τρόπους. Σε όλες αυτές τις εφαρμογές, η επαυξημένη πραγματικότητα που παρουσιάζεται στο χρήστη εμπλουτίζει την αντίληψη και τις δυνατότητες του χρήστη μέσα στον κόσμο. Ο υπέρτατος στόχος είναι η δημιουργία ενός συστήματος που δε θα επιτρέπει στο χρήστη να ξεχωρίσει τον πραγματικό κόσμο από την εικονικά εμπλουτισμένη εικόνα του. Για το χρήστη αυτού του συστήματος κάθε σκηνή θα μοιάζει σαν μία πραγματική εικόνα.

1.1.1 Σύγκριση επαυξημένης και εικονικής πραγματικότητας

Η εικονική πραγματικότητα (virtual reality) είναι μία τεχνολογία που περιλαμβάνει ένα ευρύ φάσμα ιδεών. Ουσιαστικά ορίζει το πεδίο στο οποίο πολλοί ερευνητές και εμπορικοί φορείς εκφράζουν τη δουλειά τους. Ο όρος εικονική πραγματικότητα προέρχεται από τον Jaron Lanier, ιδρυτή μίας από τις πρώτες εταιρίες που εμπορευόταν συστήματα εικονικής πραγματικότητας. Ο όρος αυτό περιγράφει ένα τρισδιάστατο περιβάλλον που δημιουργείται από έναν υπολογιστή. Ο χρήστης συχνά μπορεί να αλληλεπιδράσει με αυτό το περιβάλλον μέσω απλών ή εξειδικευμένων συσκευών εισόδου-εξόδου. Στην πράξη, είναι πολύ δύσκολο να δημιουργηθεί ένα πειστικό περιβάλλον εικονικής πραγματικότητας, καθώς τα γραφικά και ο χειρισμός τους απαιτούν ισχυρή υπολογιστική δύναμη για να διατηρήσουν το ρεαλισμό στο απαιτούμενο υψηλό επίπεδο.

Η επαυξημένη πραγματικότητα ορίστηκε αρχικά ως το αντίθετο της εικονικής πραγματικότητας [91]. Αντί να εισάγεται ο χρήστης σε ένα συνθετικό περιβάλλον το οποίο



Σχήμα 1.1: Η ταξινόμηση της αναμειγμένης πραγματικότητας κατά τον *Milgram*

περιέχει μόνο πληροφορία, ο στόχος της επαυξημένης πραγματικότητας είναι ο εμπλουτισμός του πραγματικού κόσμου με δυνατότητες προβολής και διαχείρισης πληροφοριών. Για πολλούς, ένα από τα πιο σημαντικά χαρακτηριστικά είναι ο τρόπος με τον οποίο δίνεται η δυνατότητα να μετατοπιστεί το κέντρο της αλληλεπίδρασης. Το σύστημα αλληλεπίδρασης δεν είναι πια σε μία συγκεκριμένη τοποθεσία, αλλά σε όλοκληρο το περιβάλλον. Η αλληλεπίδραση δεν αποτελεί πλέον μία επικοινωνία που περιλαμβάνει ένα πρόσωπο και μία οθόνη υπολογιστή, αλλά διαχέεται σε όλο τον περιβάλλοντα χώρο και τα αντικείμενά του. Με τον τρόπο αυτό, η χρήση ενός συστήματος πληροφόρησης δεν είναι πλέον αποκλειστικά μία πράξη συνειδητή και σκόπιμη αλλά ενσωματώνεται στην γενικότερη σχέση του ατόμου με το περιβάλλον του.

Οι Milgram et al [57] όρισαν μία ταξινόμηση, προσδιορίζοντας τη σχέση μεταξύ εικονικής και επαυξημένης πραγματικότητας όπως φαίνεται στο Σχήμα 1.1. Ο πραγματικός κόσμος και ένα πλήρως εικονικό περιβάλλον βρίσκονται στα δύο άκρα, ενώ η μέση περιοχή ονομάζεται αναμειγμένη πραγματικότητα (*mixed reality*). Η συζήτηση σχετικά με τη σχέση αυτή επικεντρώνεται σε κάποιες βασικές διαφορές τους. Η πιο προφανής από αυτές είναι η διαδικασία με την οποία ο χρήστης εισέρχεται σε καθένα από τα περιβάλλοντα αυτά. Η εικονική πραγματικότητα επιδιώκει την είσοδο του χρήστη σε ένα αποκλειστικό περιβάλλον που θα ελέγχει τις αισθήσεις του. Αντίθετα, ένα σύστημα επαυξημένης πραγματικότητας εμπλουτίζει τον πραγματικό κόσμο και βασίζεται στην παρουσία του χρήστη στον κόσμο αυτό. Για να συμβεί αυτό απαιτείται η ύπαρξη ενός μηχανισμού που θα μπορεί να συνδυάζει τον πραγματικό με τον εικονικό κόσμο, κάτι που λείπει από τα συστήματα εικονικής πραγματικότητας. Ταυτόχρονα, τα ψηφιακά αντικείμενα θα πρέπει να είναι ευθυγραμμισμένα με ακρίβεια και σε όλες τις διαστάσεις με τον πραγματικό κόσμο. Η διατήρηση της ορθής σχέσης μεταξύ των ψηφιακών αντικειμένων και του πραγματικού περιβάλλοντος θα πρέπει να διατηρείται και κατά την κίνηση του χρήστη μέσα στο περιβάλλον. Η ανάπτυξη αυτών των τεχνολογιών αποτελεί ένα πολύ δραστήριο πεδίο έρευνας, όπως παρουσιάζεται εν συντομία στην υποενότητα 1.1.3.

1.1.2 Πεδία εφαρμογής επαυξημένης πραγματικότητας

Τα τελευταία χρόνια, συνδυάστηκαν οι δυνατότητες της επεξεργασίας ακολουθιών εικόνων σε πραγματικό χρόνο, τα συστήματα γραφικών για υπολογιστές και νέες τεχνολογίες απεικόνισης ώστε να γίνει δυνατή η προβολή ενός εικονικού αντικειμένου, σε σωστή ευθυγράμμιση με το τρισδιάστατο περιβάλλον στο οποίο κινείται ο χρήστης. Ερευνητές, που δουλεύουν με συστήματα επαυξημένης πραγματικότητας, τα προωθούν ως λύσεις σε ποικίλα πεδία εφαρμογών στα οποία, πολλές φορές, εφαρμόζονται και λύσεις εικονικής πραγματικότητας. Ένα τέτοιο παράδειγμα αποτελεί η ιατρική, πεδίο κατ'εξοχήν εφαρμογής τεχνολογιών απεικόνισης. Η δυνατότητα ψηφιακής ανακατασκευής και προβολής τμημάτων του ανθρώπινου σώματος για την ενίσχυση της διαδικασίας

σχεδιασμού μίας εγχείρησης, ή η δυνατότητα προβολής και ευθυγράμμισης των δεδομένων από μαγνητικές ή αξονικές τομογραφίες με τον ασθενή είναι από τις πρώτες εφαρμογές που εξετάζονται και σχεδιάζουν να βελτιώσουν την απόδοση των ιατρικών μηχανημάτων.

Είναι επίσης αναμενόμενο, το ότι η βιομηχανία της ψυχαγωγίας αποτελεί βασικό πεδίο έρευνας, εφαρμογής, αποτίμησης και βελτίωσης των συστημάτων επαυξημένης πραγματικότητας. Το δελτίο καιρού άλλωστε, όπως παρουσιάζεται εδώ και πολλά χρόνια, αποτελεί ένα απλοϊκό παράδειγμα επαυξημένης πραγματικότητας, καθώς παρακολουθούμε την παρουσίαση μπροστά από ψηφιακούς χάρτες ενώ, στο στούντιο, ο παρουσιαστής στέκεται μπροστά από ένα κυανό ή πράσινο πέτασμα (blue / green screen). Η εξέλιξη της τεχνολογίας αυτής έχει οδηγήσει στη δημιουργία των εικονικών στούντιο (virtual studios) τα οποία, με χρήση είτε ηλεκτρομηχανικών είτε οπτικών μεθόδων για την ευθυγράμμιση και απόδοση της τελικής εικόνας, αποτελούν αναπόσπαστο μέρος των μεγάλων κινηματογραφικών και τηλεοπτικών παραγωγών. Άλλα απλούστερα παραδείγματα αποτελούν οι ψηφιακές διαφημίσεις που ευθυγραμμίζονται και προβάλλονται πάνω σε ζωντανές αναμεταδόσεις, αθλητικών κυρίως, γεγονότων.

Τέλος, να αναφέρουμε πεδία εφαρμογής όπως η στρατιωτική εκπαίδευση και πρακτική, όπου έχουν κατασκευαστεί ειδικές συσκευές προβολής γραφικών με σημαντικές πληροφορίες για τους πιλότους ή χειριστές στρατιωτικού εξοπλισμού καθώς και για τη βελτίωση της εκπαίδευσης. Είναι, επίσης, ποικίλες οι εφαρμογές σε τομείς όπως το βιομηχανικό σχέδιο, η τηλερομποτική, η κατασκευή, συντήρηση και επιδιόρθωση εξοπλισμού, και πολλών άλλων που μπορούν να εμπλουτιστούν μέσω της προβολής ψηφιακών αντικειμένων στο πραγματικό περιβάλλον με σκοπό, μεταξύ άλλων, την προεπισκόπηση, την αποτίμηση, την πρόβλεψη και την εκπαίδευση.

1.1.3 Προδιαγραφές συστημάτων επαυξημένης πραγματικότητας

Στην παρούσα υποενότητα, θα σκιαγραφήσουμε τα τμήματα που συνθέτουν ένα σύστημα επαυξημένης πραγματικότητας. Παρά το εύρος των εφαρμογών για τις οποίες σχεδιάζονται και υλοποιούνται τέτοια συστήματα, αυτά διατηρούν κάποιες κοινές μονάδες. Τα πεδία της όρασης υπολογιστών, των γραφικών υπολογιστών και των διαπροσωπικών συμβάλλουν ενεργά στην πρόοδο των συστημάτων επαυξημένης πραγματικότητας. Ένα σύστημα εικονικής πραγματικότητας επιχειρεί να απορροφήσει πλήρως το χρήστη σε ένα περιβάλλον δημιουργημένο σε ηλεκτρονικό υπολογιστή. Για να γίνει εφικτή η εμβάθυνση του χρήστη θα πρέπει το σύστημα να διατηρεί τον ορθό συσχετισμό του χρήστη με το ψηφιακό περιβάλλον, αντιλαμβανόμενο και προσαρμοζόμενο στις κινήσεις και τις αλλαγές του. Ένα σύστημα επαυξημένης πραγματικότητας μπορεί να θεωρηθεί ως το απόλυτο σύστημα εμβάθυνσης του χρήστη. Ο στόχος είναι η συνεχής αντιστοίχιση των όποιων ψηφιακών προσθηκών στο πραγματικό περιβάλλον και στην οπτική γωνία του χρήστη.

Η σκηνή παρακολουθείται από μία συσκευή απεικόνισης, που στην προκειμένη περίπτωση είναι μία κάμερα. Η κάμερα προβάλλει προοπτικά τον 3Δ κόσμο σε ένα 2Δ επίπεδο. Οι εγγενείς και εξωτερικοί παράμετροι της συσκευής καθορίζουν επακριβώς το περιεχόμενο της προβολής στο επίπεδο της εικόνας. Τα ψηφιακά αντικείμενα μοντελοποιούνται σε ένα καρέ αναφοράς και το σύστημα γραφικών χρειάζεται να γνωρίζει τις τρέχουσες παραμέτρους της κάμερας, ώστε να αποδώσει σωστά αυτά τα αντικείμενα και να συγχωνεύσει τις δύο σκηνές. Οι ερευνητικές δραστηριότητες επικεντρώνονται στους τρόπους με τους οποίους επιτυγχάνεται και διατηρείται η ευθυγράμμιση των

δύο κόσμων, καθώς και ο τρόπος απόδοσης τους από τις συσκευές απεικόνισης. Τα κριτήρια απόδοσης των συστημάτων αυτών, που αναμένεται να λειτουργούν σε πραγματικό χρόνο, έχουν να κάνουν με το ρυθμό ανανέωσης της επαυξημένης εικόνας, καθώς και με την ακρίβεια και ρεαλιστικότητα της ευθυγράμμισης και συγχώνευσης της πραγματικής και της εικονικής ακολουθίας.

Σημαντικές για την απόδοση των συστημάτων αυτών είναι ο σχεδιασμός που γίνεται και οι τεχνολογίες που εφαρμόζονται σχετικά με την απεικόνιση της επαυξημένης ακολουθίας. Επιγραμματικά θα αναφερθούν οι τρεις επικρατέστερες μέθοδοι, με πρώτη την απεικόνιση της ακολουθίας σε μία οθόνη. Αυτή η προσέγγιση, που αναφέρεται και ως 'παράθυρο στον κόσμο', δε μπορεί, προφανώς, να απορροφήσει το χρήστη στο περιβάλλον, αλλά αποτελεί την πιο απλή και εύκολα υλοποιήσιμη, ειδικά κατά τη διάρκεια της ανάπτυξης και της δοκιμής του συστήματος. Για την ολοκλήρωση του συστήματος, απαιτείται η χρήση συσκευών που αναρτούνται στο κεφάλι του χρήστη, όπως αυτές που χρησιμοποιούνται στις εφαρμογές εικονικής πραγματικότητας. Υπάρχουν δύο κατηγορίες τέτοιων συσκευών, που διαχωρίζονται βάσει του τρόπου με τον οποίο αποδίδεται ο πραγματικός κόσμος. Στην πρώτη περίπτωση κάμερες στέλνουν την εικόνα στη συσκευή και αυτή προβάλλει τον πραγματικό κόσμο μέσω μίας ακολουθίας βίντεο, ενώ στη δεύτερη ο χρήστης βλέπει μέσα από τα γυαλιά τον πραγματικό κόσμο και η συσκευή αποδίδει την επιπλέον πληροφορία πάνω από αυτόν.

Όπως γίνεται εύκολα αντιληπτό, σημαντικό ρόλο στα συστήματα επαυξημένης πραγματικότητας παίζουν τα συστήματα παρακολούθησης ανθρώπων και αντικειμένων. Με αυτόν τον τρόπο γίνεται δυνατή η απόδοση του κόσμου από την οπτική γωνία του χρήστη και η ακρίβεια αυτών των συστημάτων καθορίζει και την τελική αποτίμηση. Η πιο καθιερωμένη μέθοδος για την παρακολούθηση της θέσης και της κίνησης, είτε των αντικειμένων είτε των καμερών που χρησιμοποιούνται, είναι η χρήση μαγνητικών αισθητήρων. Με τη χρήση ηλεκτρομηχανικού εξοπλισμού επιτυγχάνεται εύκολη παρακολούθηση των αντικειμένων μέσα από δοκιμασμένες μεθόδους, αλλά η ακρίβεια του τελικού αποτελέσματος δεν είναι η επιθυμητή, καθώς η ταχύτητα απόκρισής τους είναι περιορισμένη, ενώ υπεισέρχεται και ο αστάθμητος παράγοντας του θορύβου. Οι ερευνητές αναζητούν τα τελευταία χρόνια υβριδικές μεθόδους, ώστε να υπερβούν τα όρια που έχει θέσει η χρήση των αισθητήρων, θεωρώντας ότι η πιο πολλά υποσχόμενη τεχνική θα συνδυάζει τη χρήση αισθητήρων για μία χοντρική προσέγγιση και στη συνέχεια μεθόδους επεξεργασίας εικόνας για την τελική εκτίμηση.

1.2 Εικονικά Στούντιο

Τα εικονικά στούντιο (virtual studios) είναι μία επέκταση της παραδοσιακής τεχνολογίας του κυανού πετάσματος (blue screen). Αυτή η διαδικασία αφορά στο συνδυασμό δύο διαφορετικών ακολουθιών βίντεο μέσω της χρήσης συσκευών κατάτμησης χρώματος ελέγχου (chromakeyers). Η τεχνική αυτή χρησιμοποιήθηκε πολύ νωρίς στην τηλεόραση, συγκεκριμένα στα δελτία καιρού για την προβολή των μετεωρολογικών χαρτών ή στην εναλλαγή του φόντου σε διάφορες εκπομπές. Το πρόβλημα είναι ότι η κάμερα δε μπορεί να κινηθεί, καθώς το φόντο εισάγεται σε προκαθορισμένη θέση και κλίμακα σε σχέση με τον παρουσιαστή. Τα εικονικά στούντιο αποφεύγουν αυτό τον περιορισμό, καθορίζοντας ηλεκτρονικά τη θέση των αντικειμένων και παράγοντας μία κατάλληλα μετασχηματισμένη εικόνα, η οποία τελικά χρησιμοποιείται ως φόντο. Αυτό δημιουργεί την εντύπωση ότι κάποιος κινείται και μιλάει μέσα σε έναν εικονικό κόσμο.

Πολυάριθμοι τηλεοπτικοί σταθμοί έχουν αναγνωρίσει τα πλεονεκτήματα αυτής της

τεχνολογίας παραγωγής και τη χρησιμοποιούν σε ένα μεγάλο αριθμό προγραμμάτων. Παραδείγματα αποτελούν τα δελτία ειδήσεων ή τα δελτία καιρού όπως αναφέρθηκε και ανωτέρω. Τα εικονικά στούντιο χρησιμοποιούνται επίσης πολύ στην παραγωγή κινηματογραφικών ταινιών, για την όσο το δυνατόν ακριβέστερη μίξη στατικών ή κινούμενων γραφικών με την πραγματική εικόνα. Στην περίπτωση της τηλεόρασης πρέπει να ληφθεί υπόψη και ένα επιπλέον πρόβλημα, αυτό των δυνατοτήτων του λογισμικού και του υλικού σχετικά με την λειτουργία τους σε πραγματικό χρόνο. Στην περίπτωση του σήματος PAL αυτό σημαίνει ότι πρέπει να δημιουργούνται 50 εικόνες το δευτερόλεπτο ώστε να επιτυγχάνεται ρεαλιστική εντύπωση σε σκηνές όπου η κάμερα κινείται. Η τάση που επικρατεί είναι προς τη χρήση οικονομικότερου και χαμηλότερων δυνατοτήτων υλικού, το οποίο, στο προσεχές μέλλον, αναμένεται να έχει τις απαιτούμενες δυνατότητες στον τομέα των γραφικών.

Ανεξάρτητα από την προσέγγιση και τις επιδόσεις της εκάστοτε υλοποίησης του εικονικού στούντιο, το κύριο πλεονέκτημα παραμένει το ίδιο. Δεν απαιτείται αποθήκευση, συναρμολόγηση, ή αποσυναρμολόγηση των σκηνικών. Ο φωτισμός στο σκηνικό μπορεί να μειωθεί στο ελάχιστο, καθώς μόνο οι άνθρωποι και τα αντικείμενα στο προσκήνιο έχουν σημασία και όχι το φόντο. Επιπλέον, η χρήση των εικονικών στούντιο προσφέρει και έναν αριθμό νέων προοπτικών. Αντικείμενα μπορούν να κινούνται στον αέρα, εφαρμογές πολυμέσων μπορούν να ενσωματωθούν στο περιβάλλον και γενικά δημιουργούνται περισσότερο δυναμικές και αλληλεπιδραστικές καταστάσεις. Και σημαντικό πλεονέκτημά τους είναι ότι χρησιμοποιούνται για όλες τις παραγωγές, για διαφορετικά σκηνικά, χωρίς προβλήματα με τον φωτισμό, απαιτώντας μόνο ένα ψηφιακό αποθηκευτικό μέσο.

Ένα εικονικό στούντιο απαιτεί δύο επιπλέον μονάδες, επιπρόσθετα αυτής που χειρίζεται το chromakeying. Μία για τη μέτρηση της κίνησης και της κλίσης του φακού της κάμερας, και μία που θα φροντίζει για την παραγωγή της εικόνας του φόντου, προσαρμοσμένης στην οπτική γωνία της κάμερας. Χρειάζεται επίσης μία μονάδα καθυστέρησης του βίντεο για να αντισταθμίσει την καθυστέρηση που προέρχεται από την παραγωγή της εικόνας του φόντου. Τα περισσότερα συστήματα κατατάσσονται σε μία από δύο κατηγορίες, δισδιάστατα ή τρισδιάστατα. Σε ένα 2Δ σύστημα, η κίνηση της κάμερας περιορίζεται αναρτώντας την κάμερα σε ένα σταθερό σημείο και επιτρέποντας μόνο την κίνηση του φακού, με αντάλλαγμα σημαντική απλοποίηση των υπολογισμών που χρειάζονται για την παραγωγή και την ευθυγράμμιση του φόντου. Ο περιορισμός αυτός μειώνει την ευελιξία του συστήματος όμως, στην πράξη, πολλά προγράμματα δεν απαιτούν τη μετατόπιση της κάμερας κατά τη διάρκεια των γυρισμάτων. Ένα 3Δ σύστημα επιτρέπει την μετατόπιση της κάμερας και στις τρεις διαστάσεις, καθώς και όλες τις κινήσεις του φακού (pan, tilt, zoom). Αυτό απαιτεί ένα πιο περίπλοκο σύστημα παρακολούθησης της κάμερας καθώς και μία μονάδα απόδοσης των 3Δ γραφικών σε πραγματικό χρόνο για το φόντο. Υπάρχουν διάφορα συστήματα παρακολούθησης τα οποία είναι είτε μηχανικά (ειδικές βάσεις - τρίποδα, μαγνητικοί αισθητήρες κλπ.), ή βασίζονται σε τεχνικές ανάλυσης εικόνας (εντοπισμός σηματοδίων εντός ή εκτός του σκηνικού). Σε όλες τις περιπτώσεις χρειάζεται ένα σύστημα υψηλής ακρίβειας ώστε να εξασφαλίσει ότι οποιαδήποτε διαφορά στη θέση της εικονικής κάμερας και της πραγματικής να είναι αμελητέα στην τελική εικόνα (οι τυπικές απαιτήσεις αναφέρονται στον υπολογισμό της θέσης της κάμερας με ακρίβεια εκατοστού και της γωνίας λήψης με ακρίβεια εκατοστού της μοίρας).

Στην περίπτωση της απόδοσης της 3Δ σκηνής σε ποιότητα τηλεοπτικής μετάδοσης, σε συνεχή ρυθμό 50 ή 60 εικόνων το δευτερόλεπτο, απαιτείται η χρήση συστη-

μάτων γραφικών πολύ υψηλών απαιτήσεων, αν και οι τυπικοί προσωπικοί υπολογιστές αρχίζουν να γίνονται χρησιμοποιήσιμοι. Μία εναλλακτική στην ανάγκη για συνεχή απόδοση της εικόνας σε υψηλή ποιότητα και σε πραγματικό χρόνο είναι η απόδοση σε χαμηλότερη ποιότητα (ώστε να υπάρχει η δυνατότητα προεπισκόπησης κατά τη διάρκεια των γυρισμάτων) ενώ παράλληλα καταγράφονται οι παράμετροι κίνησης της κάμερας. Στη συνέχεια, μπορεί η απόδοση στην απαιτούμενη ποιότητα να γίνει μετά το γύρισμα, χρησιμοποιώντας εξοπλισμό χαμηλότερων επιδόσεων με αποτέλεσμα ίσης ή υψηλότερης ποιότητας εικόνα αλλά και σπατάλη σημαντικού χρόνου για την περάτωση της διαδικασίας.

1.3 Ανάλυση της Διδακτορικής Διατριβής

Στο Κεφάλαιο 2 παρουσιάζεται ένα πρωτότυπο οπτικό σύστημα εικονικού στούντιο. Συγκεκριμένα, προτείνεται μία μεθοδολογία κατασκευής του κυανού πετάσματος με τέτοιο τρόπο, ώστε να περιλαμβάνει τα απαραίτητα χαρακτηριστικά για τον υπολογισμό της τρισδιάστατης κίνησης της κάμερας, χωρίς τη χρήση ηλεκτρομηχανικού εξοπλισμού. Παράλληλα, παρουσιάζεται μία αποτελεσματική λύση για τον υπολογισμό της τρισδιάστατης κίνησης της κάμερας βάσει αντιστοιχίσεων ευθειών. Τα αποτελέσματα προσομοιώσεων, πραγματικών πειραμάτων και της ασυμπτωτικής ανάλυσης σφάλματος επιβεβαιώνουν την επιτυχία του συστήματος. Η πορεία της έρευνας έως την τελική ολοκλήρωση και τεκμηρίωση του συστήματος και των επιμέρους μονάδων του, παρουσιάστηκαν στις εργασίες [20, 92, 93, 21, 94].

Έχοντας επιλέξει την αντιμετώπιση στο βασικό πρόβλημα των εικονικών στούντιο, που έχει να κάνει με τον υπολογισμό των παραμέτρων των καμερών, μπορούν πλέον να παράγονται ακολουθίες μεικτής πραγματικότητας χωρίς περιορισμούς στην κίνηση είτε των αντικειμένων και των ηθοποιών ή των καμερών. Στο Κεφάλαιο 3 παρουσιάζεται το επόμενο βήμα που επελέγη στη θεώρηση του γενικότερου προβλήματος ενός εικονικού στούντιο, και που έχει να κάνει με τον εντοπισμό και την παρακολούθηση όλων των κινούμενων ή ακίνητων αντικειμένων και ηθοποιών. Μία απλή μέθοδος που επιτρέπει την παρακολούθηση σε πραγματικό χρόνο, επωφελούμενη και από τις ιδιότητες του πρωτότυπου εικονικού στούντιο, παρουσιάστηκε στην εργασία [23] και αποτελεί μέρος της [24].

Το τελευταίο βήμα στην προσπάθεια εμπλουτισμού του εικονικού στούντιο που παρουσιάσαμε, ήταν η δυνατότητα επικοινωνίας και αλληλεπίδρασης μεταξύ των πραγματικών ηθοποιών και των εικονικών αντικειμένων και ηθοποιών που μπορούν να εισαχθούν στην μεικτή ακολουθία, βάσει των αποτελεσμάτων του προηγούμενου Κεφαλαίου. Στο Κεφάλαιο 4 παρουσιάζεται μία μεθοδολογία για την εξαγωγή των χαρακτηριστικών του ανθρώπινου προσώπου, με στόχο την εκτίμηση της συμπεριφοράς του ατόμου και ειδικότερα της έκφρασης του προσώπου του. Το Κεφάλαιο βασίζεται στις εργασίες [88, 43] που έχουν προκύψει από συνεχή έρευνα και συνεργασία στο πεδίο αυτό, όπως φαίνεται και από τις εργασίες [22, 3, 44]. Τέλος, στο Κεφάλαιο 5 εφαρμόζονται οι τεχνικές εκτίμησης της έκφρασης του προσώπου του ηθοποιού και παρουσιάζονται πειραματικά αποτελέσματα για το επαυξημένο σύστημα εικονικού στούντιο υπό ένα γενικό σενάριο, όπως αυτό εμφανίζεται και στην εργασία [24].

Στα πλαίσια της έρευνας και της συνεργασίας με ερευνητικές ομάδες του εργαστηρίου, πραγματοποιήθηκαν και δημοσιεύτηκαν και άλλες εργασίες σε συναφή αντικείμενα, που όμως δε θεωρήθηκε σκόπιμο να συμπεριληφθούν στην τελική διατριβή ώστε να επιτευχθεί η απαιτούμενη συνοχή. Ενδεικτικά, οι εργασίες αυτές περιλαμβάνονται

στον Κατάλογο Δημοσιεύσεων του συγγραφέα.

Η συγγραφή του παρόντος πραγματοποιήθηκε με τη χρήση του L^AT_EX. Οι μεταφράσεις των όρων έχουν γίνει, κατά το δυνατόν, σύμφωνα με τις πρότυπες μεταφράσεις όρων πληροφορικής. Στις εισαγωγές των κεφαλαίων της διατριβής μπορεί να παρατηρηθούν μικρές επικαλύψεις στην προσπάθεια αυτά να αποτελούν ανεξάρτητες και κατανοητές ενότητες.

□

Κεφάλαιο 2

Υπολογισμός της 3Δ κίνησης της Κάμερας σε Εικονικά Στούντιο

2.1 Εισαγωγή - Αναδρομή

Η ραγδαία ανάπτυξη της επιστήμης των υπολογιστών έχει προσελκύσει τα τελευταία χρόνια ένα σημαντικό μέρος της βιομηχανίας της ψυχαγωγίας. Τα πρόσφατα επιτεύγματα στους τομείς της επεξεργασίας εικόνας, της τεχνολογίας βίντεο και των γραφικών με υπολογιστές, όπως επίσης και η αυξανόμενη υπολογιστική ισχύς που παρέχεται από τα σύγχρονα υπολογιστικά συστήματα, υιοθετούνται ευρέως για την εμπορική παραγωγή υλικού βίντεο. Τα εικονικά στούντιο (virtual studios) χρησιμοποιούνται σε ποικίλες παραγωγές της βιομηχανίας βίντεο, έχοντας ως αποτέλεσμα ιδιαίτερος ενδιαφέροντα οπτικά εφέ. Η πλέον συνήθης από αυτές τις παραγωγές είναι το δελτίο καιρού. Οι ακολουθίες βίντεο που παράγονται από τέτοια συστήματα προέρχονται βασικά από τη σύνθεση φυσικών και συνθετικών ακολουθιών, οι οποίες είτε υπάρχουν προ-αποθηκευμένες ή παράγονται σε πραγματικό χρόνο. Στην πλέον συνηθισμένη περίπτωση, η 'μικτή ακολουθία', η οποία αναφέρεται και ως εικονικό σύνολο (virtual set), αποτελείται από μία ακολουθία φυσικών εικόνων, που λαμβάνεται σε πραγματικό χρόνο, και μία αποθηκευμένη ακολουθία συνθετικής ή φυσικής οπτικής πληροφορίας [30]. Σε αυτό το πλαίσιο, τέτοια συστήματα προκαλούν αυξημένο ενδιαφέρον, ιδιαίτερα μετά τις οδηγίες του προτύπου MPEG-4 σχετικά με την αντικειμενοστραφή και συνθετική-φυσική-υβριδική κωδικοποίηση.

Ένα σύστημα εικονικού στούντιο αποτελείται από τρία βασικά υποσυστήματα, συγκεκριμένα τη μονάδα παρακολούθησης της κάμερας (camera tracking module), απόδοσης της εικόνας (rendering) και σύνθεσης των ακολουθιών (compositing). Το τελευταίο παραδοσιακά περιλαμβάνει ένα κυανό πέτασμα (blue screen) ως φόντο, μπροστά στο οποίο λαμβάνεται η φυσική ακολουθία, και μία τεχνική κατάτμησης χρώματος (chromakeying) για το διαχωρισμό του φόντου από το προσκήνιο. Η δράση που εκτυλίσσεται στο προσκήνιο συντίθεται τότε με την επιθυμητή ακολουθία φόντου. Με αυτόν τον τρόπο, ο μετεωρολόγος στην τηλεόραση εμφανίζεται να στέχεται μπροστά από έναν μετεωρολογικό / γεωγραφικό χάρτη, ενώ στην πραγματικότητα βρίσκεται μπροστά από ένα κυανό πέτασμα. Πρακτικά, η μονάδα σύνθεσης αντικαθιστά τις περιοχές του χρώματος κλειδιού (στην περίπτωση μας το κυανό) στο βίντεο που λαμβάνεται με το προ-αποθηκευμένο επιθυμητό φόντο. Οι παραδοσιακές τεχνικές chromakeying έχουν τροποποιηθεί και επεκταθεί, ώστε να μειωθούν οι ατέλειες και να βελτιωθούν τα αποτελέσματα της σύνθεσης (βλ. για παράδειγμα τις [33, 36]). Η μονάδα απόδοσης

είναι υπεύθυνη για την ευθυγράμμιση του λαμβανόμενου προσκηνίου με την επιθυμητή σκηνή φόντου. Το τελευταίο είναι κυρίως διαθέσιμο είτε ως προ-επεξεργασμένη ακολουθία είτε ως ένα μοντέλο εικονικού κόσμου. Στην πρώτη περίπτωση, η ακολουθία πρέπει να μετασχηματιστεί ώστε να είναι συμβατή με αυτή του προσκηνίου, ενώ στη δεύτερη περίπτωση, η εικονική σκηνή φόντου αποδίδεται από μία εικονική κάμερα η οποία κινείται αντίστοιχα με την πραγματική κάμερα που λαμβάνει την τρέχουσα σκηνή του προσκηνίου. Η μονάδα παρακολούθησης της κάμερας αποτελεί το πιο κρίσιμο μέρος ενός συστήματος εικονικού στούντιο καθώς καθορίζει την ευθυγράμμιση της ακολουθίας που λαμβάνεται με την ακολουθία του φόντου. Ο αποδοτικός υπολογισμός της κίνησης της κάμερας φροντίζει ώστε το βίντεο πραγματικού χρόνου και το διαθέσιμο φόντο να είναι οπτικά συμβατά. Μέχρι τις αρχές της δεκαετίας του '90, η μονάδα αυτή απουσίαζε από τα εικονικά στούντιο, περιορίζοντας τις δυνατότητες των συστημάτων σε εφαρμογές όπως τα τηλεοπτικά δελτία καιρού. Έκτοτε, διάφορες τεχνικές παρακολούθησης της κάμερας έχουν προταθεί ώστε να επεκτείνουν τις δυνατότητες των παραδοσιακών εικονικών στούντιο.

Τα συστήματα παρακολούθησης της κάμερας σε ένα εικονικό στούντιο εντάσσονται σε δύο ευρείες κατηγορίες: τα ηλεκτρομηχανικά και τα οπτικά, ενώ εμφανίζονται και συνδυασμοί των δύο. Διάφορα συστήματα εικονικών στούντιο έχουν αναπτυχθεί, ως πρότυπα ή ως εμπορικά προϊόντα, περιλαμβάνοντας τα Elset, 3DK και Synthevision στα ηλεκτρομηχανικά και τα Cyberset και Mindset στα οπτικά (βλ. την [30] για λεπτομέρειες). Η ηλεκτρομηχανική παρακολούθηση έχει υιοθετηθεί ευρέως, καθώς είναι ικανοποιητικής ακριβείας και εύκολα υλοποιήσιμη. Σε τέτοιες προσεγγίσεις, σερβομηχανισμοί ελέγχου χρησιμοποιούνται για τον έλεγχο της κάμερας όταν η τρισδιάστατη κίνησή της είναι προκαθορισμένη (ενεργά συστήματα / active systems). Όταν η τελευταία είναι άγνωστη, κατάλληλοι αισθητήρες (sensors) τοποθετούνται πάνω στην κάμερα, ώστε να ανιχνεύσουν την κίνησή της (παθητικά συστήματα / passive systems). Τα ηλεκτρομηχανικά συστήματα όμως, αναφέρεται ότι απαιτούν εκτενείς και χρονοβόρες διαδικασίες βαθμονόμησης (calibration), ενώ ταυτόχρονα η αποτελεσματικότητα των αισθητήρων επηρεάζεται σημαντικά από τις συνήθεις τυχαίες δονήσεις. Επιπρόσθετα, ο ειδικός αυτός εξοπλισμός μπορεί να είναι πολύ ακριβός, ιδιαίτερα όταν αυξάνονται οι επιθυμητοί βαθμοί ελευθερίας στην κίνηση της κάμερας.

Τα οπτικά συστήματα παρακολούθησης βασίζονται σε τεχνικές επεξεργασίας εικόνων για να εξάγουν την κίνηση της κάμερας από τα καρέ που λαμβάνονται από αυτή. Γι' αυτό το λόγο, το μονόχρωμο κυανό πέτασμα επεκτείνεται έτσι ώστε να περιλαμβάνει κατάλληλα χαρακτηριστικά αναφοράς για την παρακολούθηση στις δύο διαστάσεις, όπως σημεία ή ευθείες γραμμές. Παρόλο που, με αυτόν τον τρόπο, τα προβλήματα που προκαλούν η χρονοβόρα διαδικασία βαθμονόμησης και οι δονήσεις της κάμερας ξεπερνώνται, είναι εμφανές ότι τα οπτικά συστήματα αποτυγχάνουν όταν τα χαρακτηριστικά αναφοράς βρίσκονται εκτός εστίασης, καλύπτονται μερικώς από το προσκηνίο ή είναι ακόμη και εκτός οπτικού πεδίου της κάμερας. Αναφέρεται επιπλέον ότι τα σημάδια (markers) που τοποθετούνται στο πέτασμα προξενούν προβλήματα στη σύνθεση, αφού πρέπει να είναι δυνατό να διακριθούν από το κυανό φόντο [30].

Στο παρόν κεφάλαιο, εξετάζεται το πρόβλημα της κατασκευής ενός κατάλληλου κυανού πετάσματος ώστε να επιτρέπει τον άμεσο εντοπισμό του οπτικού πεδίου της κάμερας παράλληλα με τον υπολογισμό της 3Δ κίνησής της. Η συγκεκριμένη μέθοδος προτάθηκε στην εργασία [94], ενώ παλαιότερες προς την ίδια κατεύθυνση αλλά διαφορετικές προσεγγίσεις δόθηκαν στις εργασίες [20, 92].

Όπως θα φανεί στη συνέχεια του κεφαλαίου, το προτεινόμενο σύστημα οπτικής

παρακολούθησης εξαλείφει τα προαναφερθέντα συνήθη προβλήματα, οδηγώντας σε υψηλής ακρίβειας υπολογισμό της 3Δ κίνησης της κάμερας, χωρίς να επηρεάζεται από απότομες αλλαγές στο οπτικό πεδίο της κάμερας ή να δημιουργεί οποιεσδήποτε δυσκολίες στη σύνθεση. Η κύρια ιδέα για την κατασκευή του κυανού πετάσματος βασίζεται στη χρήση δύο κοντινών αποχρώσεων του μπλε χρώματος, στη θέση του παραδοσιακού μονόχρωμου πετάσματος. Τα δύο επίπεδα του μπλε κατανέμονται κατάλληλα στο πέτασμα του φόντου, ώστε να επιτρέπουν αναγνώριση του οπτικού πεδίου της κάμερας. Παρόμοιες προσεγγίσεις έχουν ακολουθηθεί στις [6] και [60], με την έννοια της επέκτασης του μονόχρωμου πετάσματος ώστε να περιέχει περισσότερα επίπεδα του μπλε. Στην [6], ένα μη ομοιόμορφο μπλε φόντο χρησιμοποιείται για να διευκολύνει τον 2Δ υπολογισμό κίνησης μέσω της μεθόδου οπτικής ροής (optical flow) που παρουσιάζεται στην [35], η οποία αναφέρεται ως αποτελεσματική ακόμη και για αντικείμενα εκτός εστίασης. Εν τούτοις, η χρήση μίας τέτοιας προσέγγισης υποφέρει εμφανώς από τις συνήθεις ανακρίβειες των μεθόδων 2Δ υπολογισμού κίνησης και περιπλέκει την επιλογή των μπλε σκιών (blue shadows) στην προσπάθεια εξαγωγής σκιών. Στο [60], 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 [60], το κυανό πέτασμα προκύπτει βάσει δοκιμής και σφάλματος ('trial and error') και η χρησιμοποιούμενη μέθοδος αναγνώρισης προτύπου, αν και δεν αναλύεται, φαίνεται να απαιτεί σημαντική υπολογιστική ισχύ.

Στο κεφάλαιο αυτό, το προτεινόμενο πέτασμα δύο τόνων του μπλε κατασκευάζεται βάσει της καλά θεμελιωμένης θεωρίας της αλγεβρικής κωδικοποίησης (algebraic coding), και βασίζεται σε αρχέτυπα πολυώνυμα (primitive polynomials) και δυαδικούς κώδικες μέγιστου μήκους (binary maximal length codes) [31] του μονοδιάστατου χώρου. Η 1Δ περίπτωση επεκτείνεται κατάλληλα στη 2Δ (ακριβέστερα, στην περίπτωση των επίπεδων επιφανειών) για την κατασκευή ενός κυανού πετάσματος επιθυμητού μεγέθους. Συγκεκριμένα, η προτεινόμενη μεθοδολογία περιλαμβάνει τον κατακερματισμό του φόντου σε παραλληλόγραμμα ίσου μεγέθους και τον χρωματισμό καθενός από αυτά με σκούρο ή ανοιχτό μπλε. Με αυτόν τον τρόπο, κάθε περιοχή του κυανού πετάσματος που υπερβαίνει ένα προκαθορισμένο μέγεθος (ανάλογο με το ελάχιστο οπτικό πεδίο της κάμερας) μπορεί να ταυτοποιηθεί μοναδικά πάνω σε ολόκληρο το πέτασμα. Ταυτόχρονα, τα σημεία και οι γραμμές αναφοράς μπορούν να εξαχθούν εμμέσως, ως οι τομές των παραλληλογράμμων. Οι πρώτες ιδέες για την κατασκευή του πετάσματος βάσει μέγιστων δυαδικών ακολουθιών παρουσιάστηκαν στις [20, 92].

Παράλληλα, δίνεται ιδιαίτερη προσοχή στην εξαγωγή των 2Δ χαρακτηριστικών του φόντου και στην μέθοδο εξαγωγής της 3Δ κίνησης της κάμερας. Τα χαρακτηριστικά αναφοράς (σημεία ή γραμμές) που χρησιμοποιούνται εξάγονται άμεσα από τις τομές των παραλληλογράμμων που σχηματίζουν το πέτασμα. Στην [20], η εκτίμηση της κίνησης της κάμερας πραγματοποιείται βάσει χαρακτηριστικών σημείων που εξάγονται από τα κοινά όρια των παραλληλογράμμων, τα οποία στη συνέχεια εισάγονται σε κατάλληλο αλγόριθμο υπολογισμού 3Δ κίνησης [87, 90, 74, 15] ανάλογο με το μοντέλο προβολής [78]. Στην [92], ο υπολογισμός της 3Δ κίνησης γίνεται βάσει 2Δ χαρακτηριστικών γραμμών λύνοντας ένα σύστημα γραμμικών ομογενών εξισώσεων ως προς τις άγνωστες παραμέτρους περιστροφής. Συγκρινόμενη με τις δύο προσεγγίσεις, η [94] παρουσιάζεται ανώτερη ως προς την ακρίβεια των αποτελεσμάτων. Στην προσέγγιση αυτή προτείνεται ένας νέος αλγόριθμος υπολογισμού 3Δ κίνησης βάσει 3Δ-σε-2Δ αντιστοιχίσεων γραμμών. Για παρόμοιες εργασίες, έχουν προταθεί στη βιβλιογραφία διάφοροι αλγόριθμοι, μεταξύ άλλων οι [65, 99, 76, 48], εστιάζοντας κυρίως στην

2Δ-σε-2Δ περίπτωση (για μία εκτενή ανάλυση στον υπολογισμό 3Δ κίνησης και δομής από αντιστοιχίσεις χαρακτηριστικών βλέπε [37, 27]). Ο προτεινόμενος αλγόριθμος αφορά στην περίπτωση των 3Δ-σε-2Δ αντιστοιχίσεων, για το πλέγμα γραμμών που δημιουργείται και εξάγεται από το προτεινόμενο κυανό πέτασμα. Μία εύρωστη και κομψή λύση παρουσιάζεται βάσει ανάλυσης ενός απλού πίνακα σε ειδικές τιμές (singular value decomposition).

Η προτεινόμενη μέθοδος κατασκευής του κυανού πετάσματος και ο αλγόριθμος υπολογισμού της 3Δ κίνησης συνδυάζονται αποδοτικά παρέχοντας βελτιωμένες εκτιμήσεις της 3Δ κίνησης της κάμερας. Στην Ενότητα 2.2, ο αλγόριθμος παρουσιάζεται με τη μορφή ακολουθιακών εργασιών και όλα τα ζητήματα που προκύπτουν υπογραμμίζονται και σχολιάζονται. Στην Ενότητα 2.3 περιγράφεται η προτεινόμενη μεθοδολογία για την κατασκευή του κυανού πετάσματος, όπως και οι απαιτούμενες επεκτάσεις της εφαρμοζόμενης θεωρίας των μονοδιάστατων μέγιστων δυαδικών ακολουθιών. Η Ενότητα 2.4 περιέχει τις θεωρητικές βάσεις για τον υπολογισμό των παραμέτρων της 3Δ περιστροφής και μετατόπισης της κάμερας βάσει των 3Δ-σε-2Δ αντιστοιχίσεων γραμμών και υποδεικνύεται πώς τα συνήθη προβλήματα απόκρυψης ή απουσίας χαρακτηριστικών ξεπερνώνται ενώ η Ενότητα 2.5 εισάγει την αντίστοιχη ασυμπτωτική ανάλυση σφάλματος. Η Ενότητα 2.6 δίνει στον ενδιαφερόμενο αναγνώστη πρακτικές αλγοριθμικές λεπτομέρειες και οδηγίες υλοποίησης, ενώ τέλος στην ενότητα 2.6 περιλαμβάνονται πειραματικά αποτελέσματα σε περιβάλλον προσομοίωσης αλλά και πραγματικό, τα οποία επιδεικνύουν την ευρωστεία της προτεινόμενης μεθόδου. Τέλος, η Ενότητα 2.7 συνοψίζει τις θεωρητικές και πρακτικές συνεισφορές της εργασίας αυτής. Σύντομες αποδείξεις όλων των θεωρητικών Προτάσεων δίνονται μέσα στο κείμενο.

2.2 Περιγραφή του συστήματος

Στο κεφάλαιο αυτό, ο προτεινόμενος αλγόριθμος παρουσιάζεται με τη μορφή ακολουθιακών εργασιών. Επιπρόσθετα, τα προβλήματα των συστημάτων οπτικής παρακολούθησης υπογραμμίζονται και σχολιάζονται σε σχέση με την προτεινόμενη προσέγγιση.

2.2.1 Σχεδιασμός ενός συστήματος οπτικής παρακολούθησης για Εικονικά Στούντιο

Τα συστήματα οπτικής παρακολούθησης στα εικονικά στούντιο 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 την παρακολούθηση της κίνησης, εκτός από το διαχωρισμό του φόντου. Υπό αυτή τη λογική, ενσωματώνονται στο κυανό πέτασμα κατάλληλα χαρακτηριστικά, τα οποία μπορούν να εντοπιστούν και να παρακολουθούνται κατά τη διάρκεια της λήψης.

Τα χαρακτηριστικά που θα υιοθετηθούν για την παρακολούθηση της κίνησης της κάμερας θα πρέπει να είναι αρκετά σε πλήθος, διαχωρίσιμα από το ομοιόμορφο κυανό φόντο και μεταξύ τους κατά κάποιο τρόπο, όπως για παράδειγμα να είναι διαφορετικής χρωματικής απόχρωσης ή διακριτής χωρικής διάταξης. Είναι εμφανές ότι η γνώση της πραγματικής ή έστω της σχετικής θέσης τους στο φόντο δεν επαρκεί, καθώς το οπτικό πεδίο της κάμερας είναι άγνωστο μεταβάλλεται στο χρόνο και η απόσταση μεταξύ τους καθορίζεται από το μοντέλο προβολής και την θέση της κάμερας στον 3Δ χώρο. Ταυτόχρονα, η ποσότητα της πληροφορίας που εισάγεται στο σύστημα για λό-

γους παρακολούθησης, δεν θα πρέπει να προξενεί δυσκολίες σύνθεσης (compositing). Συνολικά, τα παρακάτω ζητήματα θα πρέπει να ληφθούν υπόψη:

1. ο συνολικός αριθμός των αποχρώσεων του μπλε χρώματος θα πρέπει να διατηρηθεί σημαντικά χαμηλός ώστε να επιτρέπει αποδοτική κατάτμηση χρώματος (chromakeying) και ικανοποιητική εξαγωγή σκιών (αν απαιτείται),
2. τα χαρακτηριστικά που θα επιλεγούν θα πρέπει να είναι ευδιάκριτα ώστε να επιτρέπουν τον εντοπισμό (localization) του οπτικού πεδίου της κάμερας,
3. το σύνολο των χαρακτηριστικών θα πρέπει να περιέχει επαρκή πληροφορία τοπικά, ακόμη και όταν ένα σημαντικό μέρος του φόντου καλύπτεται από τη δράση στο προσκήνιο σε κάθε καρέ που λαμβάνεται,
4. η μέθοδος παρακολούθησης θα πρέπει να επιτρέπει απότομες κινήσεις της κάμερας προς όλες τις κατευθύνσεις (όλους τους βαθμούς ελευθερίας).

Μπορεί κανείς να δει ότι τα περισσότερα από τα υπάρχοντα συστήματα δεν ικανοποιούν το σύνολο των παραπάνω απαιτήσεων. Για παράδειγμα, τα συστήματα που βασίζονται στον υπολογισμό της οπτικής ροής εισάγουν δυσκολίες στην κατάτμηση χρώματος, λόγω του μεγάλου αριθμού αποχρώσεων του μπλε χρώματος που απαιτούν, ενώ τα συστήματα που βασίζονται σε σημάδια υποφέρουν από μερική απόκρυψη ή ακόμα και απουσία σημαδιών στο οπτικό πεδίο. Επιπρόσθετα, οι περισσότερες προσεγγίσεις υποθέτουν αργή και ομαλή κίνηση της κάμερας ή θέτουν περιορισμούς στους βαθμούς ελευθερίας που επιτρέπονται. Τέλος, μία ακόμη απαίτηση πρέπει να τεθεί, συγκεκριμένα

5. η μέθοδος παρακολούθησης θα πρέπει να επιτρέπει την πλήρη αλλαγή του οπτικού πεδίου της κάμερας.

Η τελευταία απαίτηση εμφανίζεται θεωρώντας το ακόλουθο σενάριο. Υποθέτουμε ότι μόνο μία κάμερα είναι διαθέσιμη για να καλύψει μία σκηνή όπου οι δύο ήρωες συνομιλούν, με αποτέλεσμα να τους μαγνητοσκοπεί εναλλάξ, επανειλημμένα. Στη γενική περίπτωση, θεωρούμε την ύπαρξη μικρότερου αριθμού μονάδων επεξεργασίας από τον αριθμό των καμερών. Σε αυτές τις περιπτώσεις το οπτικό πεδίο της κάμερας αλλάζει πλήρως μεταξύ δύο αλλαγών. Με άλλα λόγια δεν υπάρχουν χαρακτηριστικά που να μπορούν να παρακολουθηθούν μεταξύ των δύο σκηνών. Ίσως η καλύτερη λύση σε αυτές τις περιπτώσεις είναι το να έχει το σύστημα τη δυνατότητα να υπολογίζει την 3Δ θέση της κάμερας βάσει ενός καρέ (ως προς κάποια αναφορά), παρά την 3Δ κίνηση της κάμερας βάσει δύο ή περισσότερων καρέ.

Όπως θα φανεί στη συνέχεια, το προτεινόμενο σύστημα εκπληρώνει και τις πέντε παραπάνω απαιτήσεις.

2.2.2 Σχεδιασμός του κυανού πετάσματος

Στην παρούσα προσέγγιση, το κυανό πέτασμα κατασκευάζεται χρησιμοποιώντας δύο σχετικά κοντινές αποχρώσεις του μπλε. Με αυτόν τον τρόπο δεν εγείρονται προβλήματα στη μονάδα σύνθεσης, αφού τα επίπεδα του μπλε επιλέγονται επαρκώς κοντά μεταξύ τους στο χώρο των χρωμάτων (βλ. απαίτηση 1). Μετά τη διαίρεση του παραλληλόγραμμου τοίχου σε $N \times M$ παραλληλόγραμμα ίσου μεγέθους, το κάθε ένα από

αυτά βάφεται με τη σκούρα ή την ανοιχτή απόχρωση. Μ' αυτό τον τρόπο σχηματίζεται ένας $N \times M$ δυαδικός χάρτης, ο οποίος στη συνέχεια θα καλείται δυαδικός χάρτης κυανού πετάσματος. Σε κάθε καρέ, η κάμερα λαμβάνει ένα μικρό κομμάτι του κυανού πετάσματος που αντιστοιχεί σε ένα μέρος του δυαδικού χάρτη. Υποθέτοντας ότι ένας αυθαίρετος υποπίνακας μεγέθους τουλάχιστον $n \times m$ είναι πάντα ορατός από την κάμερα, αρκεί κάθε ένας από αυτούς να εμφανίζεται μόνο μία φορά σε ολόκληρο το κυανό πέτασμα (και αντίστοιχα σε ολόκληρο το δυαδικό χάρτη). Σε αυτό το πλαίσιο, όταν ένας συγκεκριμένος υποπίνακας εξάγεται από το ορατό μέρος του πετάσματος, εντοπίζεται η θέση του οπτικού πεδίου της κάμερας πάνω στο πέτασμα.

Σημειώνεται ότι με αυτή την αντιμετώπιση εκπληρώνονται άμεσα δύο ακόμα από τις απαιτήσεις. Δεδομένου ότι οποιοσδήποτε τυχαίος υποπίνακας του δυαδικού χάρτη μπορεί να εντοπιστεί και να αντιστοιχηθεί μοναδικά επάνω στο δυαδικό χάρτη, ο εντοπισμός του οπτικού πεδίου της κάμερας είναι άμεσος (βλ. απαίτηση 2). Επιπρόσθετα, με την κατάλληλη επιλογή του ελάχιστου υποπίνακα που είναι πάντοτε ορατός από την κάμερα, συμπεριλαμβάνοντας την περίπτωση της μέγιστης δυνατής μερικής επικάλυψης του φόντου, εξασφαλίζεται ότι το σύνολο των χαρακτηριστικών που εξάγονται περιέχει πάντα επαρκή πληροφορία για τον υπολογισμό της κίνησης (βλ. απαίτηση 3).

Ένα τέτοιο κυανό πέτασμα παρουσιάζεται στην εργασία [94], επεκτείνοντας στις δύο διαστάσεις τη θεωρία των ψευδοτυχαίων δυαδικών ακολουθιών μέγιστου μήκους (maximal length 1D pseudorandom binary sequences) που παράγονται από ένα κύκλωμα καταχωρητή ολίσθησης. Οι ακολουθίες αυτές παρουσιάζουν ένα σύνολο από χρήσιμες ιδιότητες όπως το ότι είναι κυκλικές και ότι σε μία τέτοια ακολουθία με συνολικό μήκος $2^n - 1$, κάθε υποακολουθία ή κατάσταση (state) μήκους n εμφανίζεται μόνο μία φορά. Αποδεικνύεται ότι, με τη χρήση των ακολουθιών μέγιστου μήκους και την επέκταση των ιδιοτήτων τους στο 2Δ χώρο, μπορεί να κατασκευαστεί κυανό πέτασμα δύο αποχρώσεων και επιθυμητού μεγέθους το οποίο να πληρεί τις προϋποθέσεις που θέσαμε παραπάνω.

2.2.3 Εντοπισμός των δισδιάστατων χαρακτηριστικών

Τα χαρακτηριστικά που χρησιμοποιούνται για τον υπολογισμό της κίνησης της κάμερας είναι οι γραμμές του πλέγματος που δημιουργείται στο κυανό πέτασμα (grid lines), οι οποίες λαμβάνονται από τον δυαδικό χάρτη του. Έτσι, ένα επιπλέον πλεονέκτημα της προτεινόμενης μεθόδου είναι ότι τα χαρακτηριστικά διακρίνονται χωρίς να μεγεθυνθεί η περιοχή του χρώματος κλειδιού και με τρόπο καλά ορισμένο και ακριβή, σε αντίθεση με τις προσεγγίσεις που χρησιμοποιούν σημάδια.

Για το τρέχον καρέ που καταγράφει η κάμερα, το ορατό τμήμα του πετάσματος παρέχεται από την κατάτμηση χρώματος (chromakeying). Τότε, τα περιγράμματα των τετραπλεύρων του πετάσματος εξάγονται χρησιμοποιώντας μία τεχνική εντοπισμού ακμών (edge detection) με ένα συνήθη τελεστή κλίσης, π.χ. Sobel. Στην πραγματικότητα, τα περιγράμματα των τετραπλεύρων εξάγονται μόνο για γειτονικά τετράπλευρα διαφορετικής απόχρωσης. Η ακόλουθη εφαρμογή του μετασχηματισμού Hough επιτρέπει τον εντοπισμό των γραμμών που αποτελούν το μετασχηματισμένο υποπλέγμα του κυανού πετάσματος. Μία απλή διαδικασία προσαρμογής των γραμμών (line fitting), με χρήση της μεθόδου ελαχίστων τετραγώνων, πραγματοποιείται τότε για την εξαγωγή των ακριβέστερων δυνατών παραμέτρων των γραμμών. Εφόσον στο ορατό μέρος του πετάσματος μπορεί να περιέχονται και ομάδες από γειτονικά τετράπλευρα ίδιας απόχρωσης, είναι πιθανόν κάποιες από τις γραμμές αναφοράς να μη δύναται να εξαχθούν

με την παραπάνω μέθοδο. Σε μία τέτοια περίπτωση, ο υποπίνακας που αντιστοιχεί στο ορατό μέρος του δυαδικού χάρτη δεν μπορεί ακόμη να εξαχθεί ολόκληρος.

2.2.4 Υπολογισμός των παραμέτρων της τρισδιάστατης κίνησης της κάμερας

Αν και, όπως αναφέρθηκε, υπάρχουν περιπτώσεις που το ορατό κομμάτι του δυαδικού χάρτη δεν μπορεί στη φάση αυτή της διαδικασίας να εξαχθεί, η περιστροφή στην οποία υπεβλήθη η κάμερα μπορεί να υπολογιστεί άμεσα. Αυτό συμβαίνει όταν οι επικρατέστερες 2Δ γραμμές επιλεγούν βάσει του δείκτη εμπιστοσύνης που υπολογίζεται από τον μετασχηματισμό Hough. Το σύνολο των γραμμών που έχουν εξαχθεί διαιρείται σε δύο υποσύνολα, βάσει των κλίσεών τους, ώστε αυτά να αντιστοιχούν στα υποσύνολα των οριζοντίων και κατακόρυφων γραμμών αναφοράς του πραγματικού κυανού πετάσματος. Για κάθε γραμμή, οι εξαγόμενες παράμετροι σε μονάδες pixel κανονικοποιούνται σε μονάδες πραγματικού χώρου χρησιμοποιώντας το λόγο των διαστάσεων του CCD (charge-coupled device) της κάμερας ως προς τις διαστάσεις του καρέ σε pixels.

Καθώς η πραγματική θέση των γραμμών αναφοράς στο κυανό πέτασμα είναι γνωστή, μπορούμε να λύσουμε ένα πρόβλημα υπολογισμού 3Δ κίνησης, βάσει των 3Δ-σε-2Δ αντιστοιχίσεων των γραμμών. Γί αυτό το σκοπό, μία εικονική 3Δ σκηνή αναφοράς κατασκευάζεται βάσει

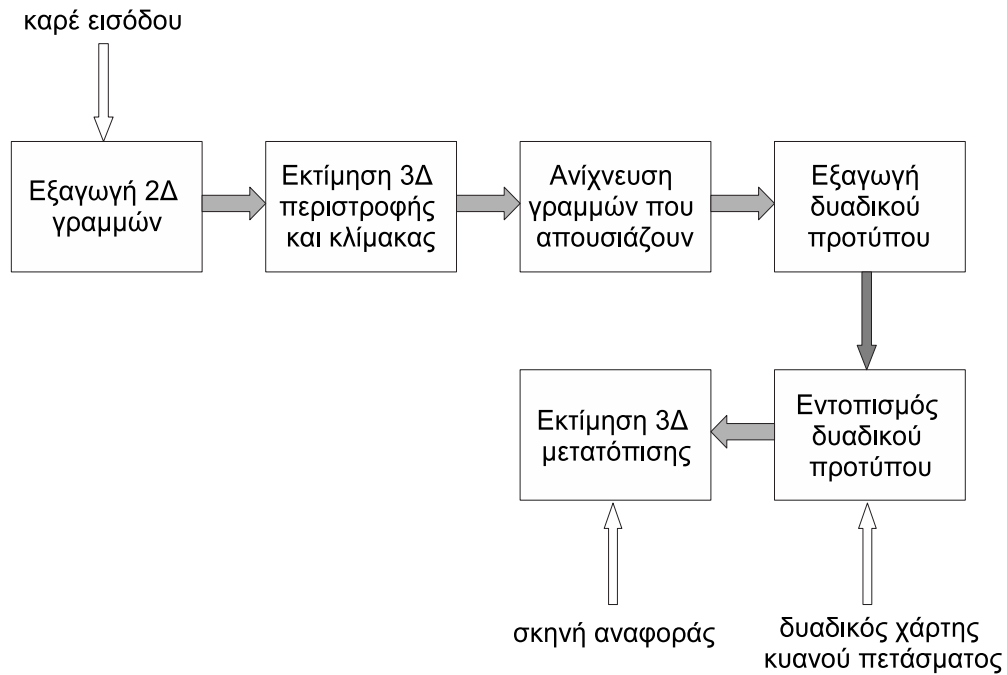
α) των οριζόντιων και κατακόρυφων μηκών των τετράπλευρων και

β) την απόσταση του κυανού πετάσματος από την κάμερα. Η τελευταία επιλέγεται τυχαία για την εικονική σκηνή αναφοράς.

Όπως μπορεί να δει κανείς στην Ενότητα 2.4, ο πίνακας περιστροφής υπολογίζεται τότε αποδοτικά από την ανάλυση σε ειδικές τιμές (singular value decomposition) ενός απλού πίνακα. Πρακτικά, η Πρόταση 2.2 της Ενότητας 2.4 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, σε αντίθεση με τον υπολογισμό της 3Δ περιστροφής, απαιτεί την ακριβή γνώση των 3Δ παραμέτρων των γραμμών στη σκηνή αναφοράς και την αντιστοίχιση των 3Δ γραμμών αναφοράς με τις 2Δ γραμμές που εντοπίζονται στο τρέχον καρέ. Σύμφωνα με τα παραπάνω, η αντιστοίχιση αυτή επιτυγχάνεται με την αναγνώριση του τμήματος του πραγματικού κυανού πετάσματος το οποίο λαμβάνουμε από την κάμερα στο τρέχον καρέ. Όμως, δεν είναι δυνατόν να καθορίσουμε το ακριβές δυαδικό πρότυπο, εφόσον κάποιες από τις μετασχηματισμένες γραμμές αναφοράς μπορεί, όπως αναφέρθηκε, να μη δύναται να εντοπιστούν από το τρέχον καρέ. Αυτές οι δύο δυσκολίες λύνονται παράλληλα καθώς, η Πρόταση 2.3 της Ενότητας 2.4 δείχνει πως για οποιεσδήποτε δύο γραμμές, η σχετική τους απόσταση στη σκηνή αναφοράς μπορεί να υπολογιστεί από τις αντίστοιχες τους 2Δ γραμμές. Με αυτόν τον τρόπο, οι απύσες γραμμές εντοπίζονται και το ορατό δυαδικό πρότυπο μπορεί να εξαχθεί. Η αντιστοίχιση μεταξύ των 2Δ και 3Δ γραμμών είναι τότε δυνατή μετά την ταυτοποίηση του προτύπου επάνω στο δυαδικό χάρτη. Το διάνυσμα της 3Δ μετατόπισης υπολογίζεται 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 2.2.4 στην Ενότητα 2.4).

Μία από τις κύριες συνεισφορές αυτού του κεφαλαίου, είναι ότι η προτεινόμενη αντιμετώπιση επιτρέπει και τον υπολογισμό του άγνωστου εστιακού μήκους (focal length)



Σχήμα 2.1: Το προτεινόμενο σύστημα υπολογισμού κίνησης σε μπλοκ-διάγραμμα.

f σε κάθε καρέ, βάσει των παραμέτρων των γραμμών που εξάγονται από αυτό. Πρακτικά, η Πρόταση 2.2.5 συμπληρώνει τις Προτάσεις 2.2.2, 2.2.3 και 2.2.4, οδηγώντας στον υπολογισμό όλων των αγνώστων παραμέτρων κίνησης, δηλ. 3Δ περιστροφής, μετατόπισης και μεγέθυνσης.

Σύμφωνα με τα παραπάνω, η προτεινόμενη προσέγγιση αντιμετωπίζει επιτυχώς και τις δύο τελευταίες απαιτήσεις, δηλαδή την ελεύθερη κίνηση της κάμερας (βλ. απαίτηση 4) και την αποδοτική εκτίμηση της κίνησής της ακόμη και για ολική αλλαγή του οπτικού πεδίου της (βλ. απαίτηση 5). Για την ακρίβεια, ο προτεινόμενος αλγόριθμος εκτιμά την 3Δ θέση και τον προσανατολισμό της κάμερας συναρτήσει του χρόνου, παρά την 3Δ κίνηση της. Το τελευταίο επιτυγχάνεται κύριως χάριν στο γεγονός ότι δε χρησιμοποιείται καμία τεχνική εκτίμησης κίνησης. Αντίθετα, τα χαρακτηριστικά που χρησιμοποιούνται εντοπίζονται από τη μοναδικότητα τους στο κυανό πέτασμα και, έτσι, δεν υπάρχει ανάγκη παρακολούθησής της κίνησής τους στο 2Δ χώρο. Ταυτόχρονα, φαίνεται ότι δεν απαιτούνται επίπονες διαδικασίες βαθμονόμησης (calibration) αφού η κατασκευή της σκηνής αναφοράς είναι εικονική. Η μοναδική τέτοια διαδικασία απαιτεί οι υποτιθέμενες αποστάσεις μεταξύ των γραμμών στη σκηνή αναφοράς να είναι ίσες με αυτές στο πραγματικό πέτασμα. Αυτό συνεπάγεται ότι το κυανό πέτασμα κατασκευάζεται με υψηλή ακρίβεια, αλλά έπειτα δεν απαιτούνται περαιτέρω ρυθμίσεις σε κάθε διαφορετικό γύρισμα. Τέλος, το προτεινόμενο σύστημα, ως ένα οπτικό σύστημα παρακολούθησης, χωρίς περιορισμούς στους βαθμούς ελευθερίας της κάμερας, επιτρέπει την καταγραφή βίντεο ακόμη και με κάμερα χειρός.

Στο Σχήμα 2.1 το προτεινόμενο σύστημα δίνεται σε μορφή ενός μπλοκ-διαγράμματος.

2.3 Κατασκευή του δυαδικού πίνακα του πετάσματος

Σε μία προσέγγιση οπτικού συστήματος υπολογισμού της κίνησης της κάμερας, η εξαγωγή όσο το δυνατόν περισσότερης ωφέλιμης πληροφορίας από την καταγραφόμενη

ακολουθία είναι ουσιώδης για το σύστημα παρακολούθησης. Στη γενική περίπτωση, η καταγραφόμενη πληροφορία αποτελείται από τα αντικείμενα του προσκηνίου και το κυανό πέτασμα που αποτελεί το φόντο. Εφόσον υπάρχει ήδη μία μεγάλη ποσότητα πληροφορίας στο φόντο, για να χρησιμοποιηθεί μόνο για το διαχωρισμό του από τα αντικείμενα του προσκηνίου βάσει χρώματος, ο αντικειμενικός σκοπός γίνεται η ενσωμάτωση στο πέτασμα πρόσθετης πληροφορίας, για χρήση στον υπολογισμό κίνησης της κάμερας. Δύο, κυρίως, τεχνικές εφαρμόζονται για αυτόν το σκοπό: η τοποθέτηση ορόσημων στο πέτασμα και η χρήση περισσότερων της μίας, κοντινών μεταξύ τους, αποχρώσεων του μπλε για την κατασκευή του.

Το παρόν κεφάλαιο παρουσιάζει την κατασκευή ενός κυανού πετάσματος αποτελούμενου από παραλληλόγραμμα, ίσου μεγέθους, χρωματισμένα σε μία από τις δύο χρησιμοποιούμενες αποχρώσεις του μπλε. Η τεχνική αυτή δεν εισάγει αξιοσημείωτες αλλαγές στην υλοποίηση ή την πολυπλοκότητα του αλγορίθμου διαχωρισμού του φόντου. Από την άλλη πλευρά, παρέχει έναν αποτελεσματικό τρόπο εισαγωγής δυαδικής πληροφορίας πάνω στο κυανό πέτασμα.

Ακολουθώντας, τα σύμβολα 1 και 0 θα δηλώνουν τον ανοιχτό και τον σκούρο τόνο του μπλε αντίστοιχα. Ένα κυανό πέτασμα, διαιρεμένο σε $N \times M$ παραλληλόγραμμα, μπορεί να οριστεί μαθηματικά από

α) τον αντίστοιχο $N \times M$ δυαδικό χάρτη **B** που περιέχει 1 και 0 στις κατάλληλες θέσεις, μαζί με

β) τις διαστάσεις του στον πραγματικό κόσμο.

Σε κάθε καρέ, η κάμερα καταγράφει ένα μικρό τμήμα του τοίχου, το οποίο αντιστοιχεί σε ένα τμήμα Π του πίνακα **B**. Το Π είναι μία συλλογή στοιχείων του **B**, που γενικά δεν αντιστοιχεί σε κάποιον παραλληλόγραμμο υποπίνακα του **B**. Ας θεωρήσουμε τον **S** ως τον μέγιστο υποπίνακα του **B**, του οποίου τα στοιχεία (παραλληλόγραμμα) ανήκουν όλα στο ορατό τμήμα Π του κυανού πετάσματος. Μπορεί να δει κανείς ότι, αν και μόνο αν ο **S**, ή οι υποπίνακές του που ξεπερνούν ένα προκαθορισμένο μέγεθος, εμφανίζεται μόνο μία φορά μέσα στον **B**, τότε το οπτικό πεδίο της κάμερας ως προς το πέτασμα μπορεί να καθοριστεί μοναδικά. Αυτό μπορεί να επιτευχθεί μόνο με την εισαγωγή και απόδειξη μίας κατάλληλης μεθοδολογίας για την ταξινόμηση των 1 και 0 στον πίνακα **B**. Θαδειχθεί ότι ένας τρόπος για να επιτευχθεί η μοναδικότητα σε όλους τους υποπίνακες του **B** που ισούνται ή ξεπερνούν ένα προδιαγεγραμμένο μέγεθος $n \times m$, δίνεται μέσα από τη χρήση των τεχνικών αλγεβρικής κωδικοποίησης (algebraic coding techniques).

Στην ακόλουθη υποενότητα δίνονται οι θεωρητικές κατευθυντήριες γραμμές για την κατασκευή ενός τέτοιου κυανού πετάσματος, ενώ στην Υποενότητα 2.3.2 σχηματίζεται ο αντίστοιχος αλγόριθμος.

2.3.1 Ένα ειδικό δισδιάστατο δυαδικό πεδίο

Γενικά, η δομή των πεδίων Galois (Galois fields) και οι ιδιότητες των αρχέτυπων πολυωνύμων (primitive polynomials) παρέχουν ένα αποδοτικό εργαλείο για την κατασκευή οποιουδήποτε κυκλικού κώδικα μέγιστου μήκους. Υποθέτουμε ότι το **h** είναι το ελάχιστο πολυώνυμο ενός αρχέτυπου στοιχείου στο $GF(p^n)$, όπου το p είναι ένας πρώτος αριθμός. Ένας κυκλικός κώδικας **c** με μήκος περιόδου $p^n - 1$, για τον οποίο το **h** είναι το αρχέτυπο πολυώνυμο ελέγχου βαθμού n (check polynomial), είναι γνωστός ως κώδικας/ακολουθία μέγιστου μήκους (maximal length code) [28]. Ακολουθώντας, όταν θα αναφέρεται η ακολουθία/κώδικας μέγιστου μήκους, θα εννοείται οποιαδήποτε

περίοδος του τελευταίου.

Οι κωδικές λέξεις (codewords) ενός τέτοιου κώδικα μπορούν να παραχθούν από ένα κύκλωμα καταχωρητή ολίσθησης (shift register). Εάν ο $R(\mathbf{h})$ είναι ο αντίστοιχος απλός καταχωρητής ολίσθησης με n θέσεις, τότε θα υπάρχει σύνδεση ανάδρασης στην θέση i , αν και μόνον αν $h_{i+1} = 1$. Ως αποτέλεσμα, ο καταχωρητής περνάει από όλες τις $p^n - 1$ καταστάσεις που παράγονται από τις n θέσεις ψηφίων σε σχέση με p διαχωρίσιμα ψηφία (έχοντας αποκλείσει τη μηδενική κατάσταση), και έτσι η ακολουθία εξόδου \mathbf{c} έχει περίοδο μήκους $p^n - 1$ [31].

Για $p \equiv 2$, η παραπάνω θεωρία παράγει 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 $(2^n - 1) \times 1$. Με αυτό τον τρόπο διασφαλίζεται ότι κάθε $n \times 1$ πρότυπο είναι μοναδικό σε μία τέτοια στήλη μεγέθους $(2^n - 1) \times 1$ του δυαδικού χάρτη του πετάσματος. Το επόμενο βήμα είναι η επέκταση αυτής της μονοδιάστατης 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 δημοσιεύσεις [20, 92], με τον πολλαπλασιασμό δύο κωδικών μέγιστου μήκους που λαμβάνονταν βάσει δύο αρχέτυπων πολυωνύμων τάξης n και m . Μία βελτιωμένη λύση εξασφαλίζεται στο κεφάλαιο αυτό, μέσω της εκμετάλλευσης της κυκλικής ιδιότητας των ακολουθιών μέγιστου μήκους· εννοώντας ότι όλες οι πιθανές ολισθήσεις ενός κώδικα \mathbf{c} είναι μέγιστου μήκους ακολουθίες που αποτελούνται από τις καταστάσεις του ίδιου καταχωρητή ολίσθησης.

Σ' αυτό το πνεύμα, η επόμενη πρόταση παρουσιάζει έναν κατάλληλο τρόπο για την κατασκευή ενός 2Δ δυαδικού χάρτη \mathbf{B} , όπου κάθε $n \times m$ (ή μεγαλύτερος) υποπίνακας \mathbf{S} είναι μοναδικός.

Πρόταση 2.1: Έστω \mathbf{c} ένας μέγιστου μήκους κώδικας 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 \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 των στηλών ($j, j + 1$).

Όμοια, έστω ότι το $\mathbf{d}_j \triangleq [d_j \ d_{j+1} \ \dots \ d_{j+m-2}]^T$ δηλώνει το διάνυσμα που περιέχει τις διαφορές ολίσθησης για m γειτονικές στήλες. 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) οι \mathbf{d}_j ($m - 1$)-άδες μπορούν να ταξινομηθούν βάσει μίας μέγιστης ακολουθίας που λαμβάνεται από οποιοδήποτε πολυώνυμο ελέγχου στο $\text{GF}((2^n - 1)^{m-1})$.

△

Απόδειξη:

α) Έστω ότι τα \mathbf{S}_1 και \mathbf{S}_2 είναι δύο $n \times m$ υποπίνακες του \mathbf{B} με τα πρώτα στοιχεία τους να συμπίπτουν με τα $b_{i_1 j_1}$ και $b_{i_2 j_2}$ αντίστοιχα. Μπορούμε να χειριστούμε χωριστά τις ακόλουθες περιπτώσεις:

1) 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, $\mathbf{S}_1 \neq \mathbf{S}_2$.

2) Όταν $j_1 \neq j_2$ και $\mathbf{d}_{j_1} \neq \mathbf{d}_{j_2}$, οι πρώτες στήλες των \mathbf{S}_1 και \mathbf{S}_2 μπορεί να είναι πανομοιότυπες. Όμως, αν για παράδειγμα τα \mathbf{d}_{j_1} και \mathbf{d}_{j_2} διαφέρουν στο k -οστό τους

d_j		0	1	1	2	3	1	...
s	1	1	2	3	5	1	2	...

Πίνακας 2.1: Κατασκευάζοντας το δυαδικό χάρτη του κυανού πετάσματος· άνω σειρά: η ακολουθία των διαφορών ολίσθησης, κάτω σειρά: η ακολουθία των ολισθήσεων.

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} , δηλαδή σε διαφορετικές καταστάσεις στο \mathbf{c} . Τότε, ισχύει πάλι ότι $\mathbf{S}_1 \neq \mathbf{S}_2$.

Είναι εμφανές ότι, αν η αποσυρθεί η απαίτηση $\mathbf{d}_{j_1} \neq \mathbf{d}_{j_2}$, 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), συνάγεται ότι ένας τρόπος να διασφαλιστεί 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} . Επιπρόσθετα, μπορεί να σημειωθεί ότι το \mathbf{d}_{j+1} είναι εκ κατασκευής μία ολισθημένη εκδοχή του \mathbf{d}_j . Επιπλέον, για να μεγιστοποιηθεί ο αριθμός των στηλών του \mathbf{B} , είναι απαραίτητες όλες οι πιθανές καταστάσεις του \mathbf{d}_j . Μία μέγιστη ακολουθία που περιέχει όλες τις πιθανές $(m - 1)$ -άδες των \mathbf{d}_j λαμβάνεται από οποιοδήποτε πολυώνυμο ελέγχου στο $\text{GF}((2^n - 1)^{m-1})$.

△

Η Πρόταση 2.1 πρακτικά δηλώνει ότι ο \mathbf{B} κατασκευάζεται βάσει ολισθημένων εκδοχών του \mathbf{c} που τοποθετούνται η μία δίπλα στην άλλη. Η κατάλληλη σειρά τοποθέτησης των τελευταίων καθορίζεται από τη μέγιστη ακολουθία όλων των δυνατών διαφορών ολίσθησης. Εφόσον η ακολουθία διαφορών ολίσθησης ληφθεί, η ακολουθία ολισθήσεων απορρέει θέτοντας την πρώτη ολίσθηση ίση για παράδειγμα με 0, δηλαδή $\mathbf{b}_1 \equiv \mathbf{c}_0$. Με αυτόν τον τρόπο, καθορίζονται όλες οι στήλες του και τελικά ο ίδιος ο \mathbf{B} .

2.3.2 Ο αλγόριθμος κατασκευής

Η κατασκευή του δυαδικού χάρτη του κυανού πετάσματος μπορεί να δοθεί περιληπτικά σε μερικά αλγοριθμικά βήματα. Αυτό το γεγονός δίνει σε κάποιον τη δυνατότητα να κατασκευάσει ένα κυανό πέτασμα με σχεδόν οποιαδήποτε επιθυμητή διάσταση είτε αναφερόμενοι σε φυσική διάσταση ή σε αριθμό παραλληλογράμμων. Τα βήματα του αλγορίθμου έχουν ως εξής:

1) Αποφάσισε το μέγεθος $n \times m$ του ελάχιστου υποπίνακα του \mathbf{B} που θα είναι ορατός από την κάμερα.

2) Βρες ένα αρχέτυπο πολυώνυμο βαθμού n στο $\text{GF}(2^n)$ και υπολόγισε τον αντίστοιχο κώδικα μέγιστου μήκους \mathbf{c} .

3) Θέσε μία αυθαίρετη περίοδο του \mathbf{c} (\mathbf{c}_0) ως την πρώτη στήλη του πίνακα \mathbf{B} .

4) Βρες ένα αρχέτυπο πολυώνυμο βαθμού $2^n - 1$ στο $\text{GF}((2^n - 1)^{m-1})$ και υπολόγισε τον αντίστοιχο κώδικα μέγιστου μήκους. Από αυτόν απορρέει η ακολουθία των διαφορών ολίσθησης.

5) Αρχίζοντας από την πρώτη στήλη του πίνακα \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 τώρα εξαχθεί.

Για να διευκρινίσουμε την προτεινόμενη μεθοδολογία, ο Πίνακας 2.1 περιέχει ένα τμήμα από την λαμβανόμενη ακολουθία με τις διαφορές ολίσθησης και το αντίστοιχο τμήμα των ολισθήσεων που απορρέουν από αυτή, για ελάχιστο ορατό τμήμα μεγέθους 3×3 . Πρέπει να ληφθεί υπ'όψιν ότι σε αυτό το παράδειγμα, ως εναρκτήριο ψηφίο στη

στήλη των ολισθήσεων χρησιμοποιήθηκε το '1' (\mathbf{c}_1) και ότι όλες οι πράξεις πραγματοποιήθηκαν με αριθμητική modulo-7 (γενικά modulo- $(2^n - 1)$).

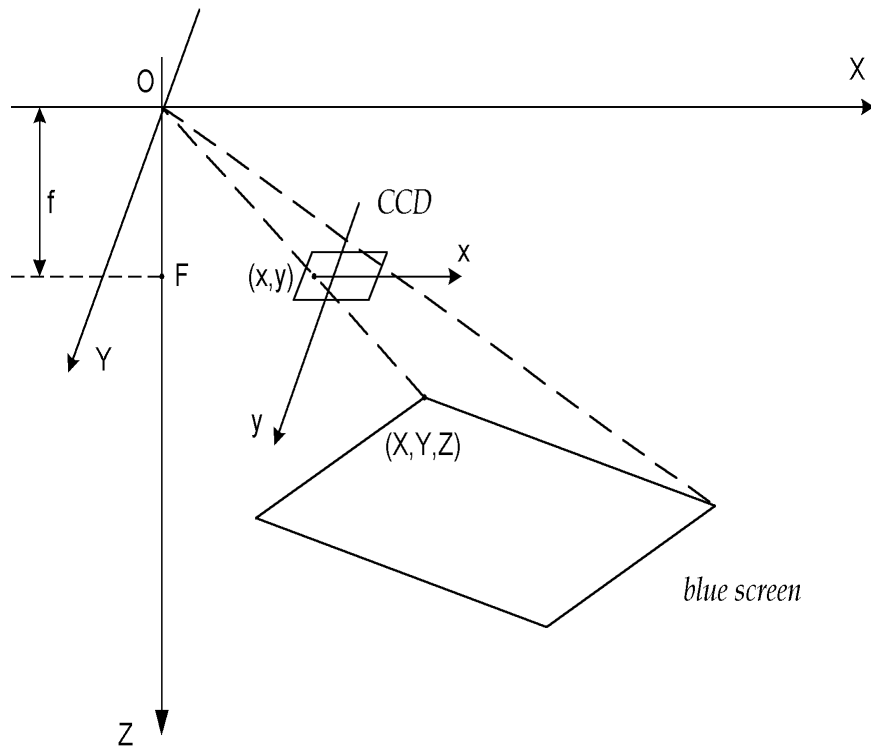
Είναι ενδιαφέρον να υπολογιστούν οι τελικές διαστάσεις του δυαδικού πίνακα, που παράγεται ακολουθώντας τα βήματα της προτεινόμενης μεθοδολογίας· δηλαδή δεδομένου ότι κάθε $n \times m$ πρότυπο πρέπει να είναι μοναδικό στον \mathbf{B} , να βρεθούν τα μέγιστα επιτρεπόμενα μήκη N και M . Μιας και κάθε στήλη αντιστοιχεί σε $2^n - 1$ καταστάσεις, περιέχει $2^n - 1 + (n - 1) = 2^n + n - 2$ παραλληλόγραμμα, αφού ακυρώσουμε την κυκλική ιδιότητα του \mathbf{c}_s σε κάθε στήλη. Με αυτόν τον τρόπο, ορίζουμε το N ως $N \triangleq 2^n + n - 2$. Παρόμοια, και σύμφωνα με τα παραπάνω, $(2^n - 1)^{m-1} - 1$ καταστάσεις απορρέουν από τις διαφορές ολίσθησης, ενώ καταργώντας την κυκλική ιδιότητα, οι διαφορές ολίσθησης αυξάνονται σε $(2^n - 1)^{m-1} + m - 2$. Το M δίνεται από τον ίδιο τον αριθμό των ολισθήσεων, οπότε ορίζεται ως $M \triangleq (2^n - 1)^{m-1} + m - 1$. Πρέπει εδώ να σημειωθεί ότι η παραπάνω μεθοδολογία θα μπορούσε, εναλλακτικά, να χρησιμοποιηθεί για να παράγει έναν πίνακα $M \times N$, με τον καθορισμό της πρώτης σειράς του \mathbf{B} και τον μετέπειτα υπολογισμό της ακολουθία των διαφορών ολίσθησης για τις σειρές του ίδιου πίνακα.

Η ανωτέρω θεωρία παρέχει μία καλώς ορισμένη και αποτελεσματική μέθοδο για την κατασκευή του δυαδικού χάρτη ενός κυανού πετάσματος, το οποίο 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 αναγνωριστεί λανθασμένα, λόγω πιθανών προβλημάτων που προκύπτουν σε συνθήκες πραγματικού χρόνου όπως ο φωτισμός, οι σκιές ή η μερική επικάλυψη. Είναι εμφανές ότι η προτεινόμενη μεθοδολογία παράγει δυαδικό χάρτη $M \times N$ με $M \gg N$. Σε αυτό το πλαίσιο, μόνο ένα υποσύνολο των διαθέσιμων, για παράδειγμα, στηλών είναι ικανό να χρησιμοποιηθεί. Η κατάλληλη επιλογή του τελευταίου μπορεί να επιτρέψει την ενσωμάτωση διαδικασιών εντοπισμού και διόρθωσης σφαλμάτων (error detection and correction).

2.4 Υπολογισμός της θέσης και του προσανατολισμού της κάμερας

Σύμφωνα με το σενάριο που περιγράφηκε στην Ενότητα 2.2, 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 σχηματίζει ένα πρόβλημα υπολογισμού της 3Δ κίνησης βάσει 3Δ-σε-2Δ αντιστοιχίσεις χαρακτηριστικών. Στον προτεινόμενο αλγόριθμο, τα χαρακτηριστικά επιλέγονται να είναι ευθείες γραμμές, συγκεκριμένα το πλέγμα από γραμμές που παρέχεται εμμέσως από την κατασκευή του δυαδικού χάρτη του κυανού πετάσματος.

Όπως εξηγείται στην Ενότητα 2.2, μερική επικάλυψη ή ολική αλλαγή των χαρακτηριστικών στο οπτικό πεδίο της κάμερας μεταξύ μεταβάσεων καθιστά οποιαδήποτε τεχνική υπολογισμού κίνησης, παρακολούθησης χαρακτηριστικών ή τοπικής συσχέτισης ανεφάρμοστη. Αντιθέτως, θεωρώντας την θέση της κάμερας στον 3Δ χώρο ως προς το επίπεδο του κυανού πετάσματος - αντί της 3Δ κίνησης της κάμερας - ο υπολογισμός των παραμέτρων της 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 [48], ενώ όταν



Σχήμα 2.2: Η γεωμετρία του προβλήματος. Το σύστημα συντεταγμένων κινείται με το επίπεδο του *CCD* (*camera-based*).

χρησιμοποιηθούν 2Δ-σε-2Δ αντιστοιχίσεις γραμμών τότε απαιτούνται τρεις όψεις [76]. Αυτό υποδηλώνει ότι επιλέγοντας την χρήση των γραμμών ως χαρακτηριστικών για την 'παρακολούθηση' μεταξύ δύο όψεων απαιτεί την εκ των προτέρων γνώση των 3Δ παραμέτρων των γραμμών τουλάχιστον σε μία από τις δύο διαθέσιμες όψεις. Υπό αυτή την έννοια, έχοντας εξάγει τις 2Δ παραμέτρους των γραμμών από το τρέχον καρέ, η 3Δ θέση της κάμερας μπορεί να υπολογιστεί θεωρώντας μία αυθαίρετη 3Δ σκηνή αναφοράς, όπου οι 3Δ παράμετροι των γραμμών, για τις ίδιες γραμμές, είναι γνωστές.

Ακόμη και σε αυτήν την περίπτωση, αναφέρεται στην [48] ότι γενικά οκτώ ή περισσότερες αντιστοιχίσεις γραμμών είναι απαραίτητες ώστε να υπολογιστεί ο πίνακας περιστροφής με χρήση ενός γραμμικού αλγορίθμου, ενώ απαιτούνται τρεις ή περισσότερες αντιστοιχίσεις γραμμών αν χρησιμοποιηθεί μία μη-γραμμική προσέγγιση. Ωστόσο, θαδειχθεί ότι με την κατάλληλη επιλογή του καρέ αναφοράς για γραμμές κάθετες μεταξύ τους που σχηματίζουν ένα πλέγμα (πάνω στο πέτασμα), τέσσερις αντιστοιχίσεις γραμμών περιέχουν ικανή πληροφορία για τον υπολογισμό του πίνακα περιστροφής χρησιμοποιώντας ένα γραμμικό αλγόριθμο. Στην πραγματικότητα, δεν απαιτείται ακριβής αντιστοίχιση των γραμμών μία προς μία σε αυτή την περίπτωση. Μετά την επίτευξη των αντιστοιχίσεων μεταξύ των 3Δ γραμμών στο καρέ αναφοράς και των 2Δ γραμμών στο τρέχον καρέ, ο υπολογισμός του πίνακα μετατόπισης μπορεί να πραγματοποιηθεί ευθέως.

Ας εξετάσουμε τώρα την κατασκευή της σκηνής αναφοράς. Όπως αναφέρθηκε ήδη στην Ενότητα 2.2, η κατασκευή αυτή είναι στην πραγματικότητα εικονική, υπό την έννοια ότι δεν χρειάζεται καμία βαθμονόμηση της κάμερας στην σκηνή αυτή. Έστω ότι το επίπεδο προβολής της εικόνας (το παραλληλόγραμμο του CCD) είναι παράλληλο στο επίπεδο XY του καρτεσιανού συστήματος συντεταγμένων σε βάθος $Z = f$, όπου το f

υποδηλώνει την εστιακή απόσταση. 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 συμφώνως προς το παραλληλόγραμμο CCD, το ακόλουθο απλό μοντέλο προοπτικής προβολής ισχύει [78],

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (2.1)$$

όπου το (x, y) δηλώνει τις καρτεσιανές συντεταγμένες του σημείου (X, Y, Z) όταν προβληθεί στο επίπεδο προβολής της εικόνας. Η γεωμετρία του προβλήματος δίνεται στο Σχήμα 2.2. Το παγκόσμιο σύστημα συντεταγμένων είναι πάντα ευθυγραμμισμένο με το επίπεδο του CCD, ενώ το κυανό πέτασμα λαμβάνει αυθαίρετες θέσεις και προσανατολισμούς στον 3Δ χώρο. Για να απλοποιηθούν το μοντέλο προβολής και οι σχετικές εξισώσεις [78], θεωρείται συχνά ότι η εστιακή απόσταση είναι γνωστή και τίθεται ίση με τη μονάδα ($f = 1$). Αντιθέτως, όπως θα δειχθεί στη συνέχεια, στην διατύπωσή μας το f μπορεί να υπολογιστεί αξιόπιστα λύνοντας ένα σύστημα γραμμικών εξισώσεων.

Η δομή του κυανού πετάσματος στη σκηνή αναφοράς είναι γνωστή και μπορεί να αναλυθεί στις ακόλουθες παραμέτρους:

- 1) το 3Δ βάθος του κυανού πετάσματος z_0 ,
- 2) το σύνολο των 'κατακόρυφων' γραμμών $X_v = \{X = x_i, i = 1 \cdots N\}$ και
- 3) το σύνολο των 'οριζόντιων' γραμμών $Y_h = \{Y = y_j, j = 1 \cdots M\}$.

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, καθένα από τα σύνολα E_v και E_h περιέχει 2Δ παραμέτρους γραμμών της μορφής (a, b) . Θα κάνουμε στη συνέχεια χρήση της καρτεσιανής αναπαράστασης της γραμμής για το τρέχον καρέ, $y' = ax' + b$, όπου το (x', y') υποδηλώνει τις καρτεσιανές συντεταγμένες ενός σημείου στο τρέχον καρέ για κάθε αντίστοιχο προβεβλημένο σημείο του 3Δ χώρου (X', Y', Z') στην τρέχουσα, άγνωστη 3Δ σκηνή. Όπως ορίστηκαν ανωτέρω, τα (x, y) και (X, Y, Z) υποδηλώνουν τα αντίστοιχά τους στο καρέ αναφοράς και στην εικονική σκηνή αναφοράς.

Θα θεωρηθεί αρχικά ότι η εστιακή απόσταση f είναι δεδομένη. Έπειτα, ορίζουμε τις 2Δ παραμέτρους (a, b) των γραμμών στα σύνολα E_v και E_h ως $(a, b) = (\alpha, \frac{\beta}{f})$. δηλαδή τα (a_{vi}, b_{vi}) και (a_{hj}, b_{hj}) για το i -οστό και j -οστό στοιχείο του E_v και E_h αντίστοιχα. Συνδυάζοντας το τελευταίο με την εξίσωση (2.1), λαμβάνουμε

$$\begin{bmatrix} a & -1 & b \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = 0. \quad (2.2)$$

Το διάνυσμα $\epsilon = [a \ -1 \ b]^T$ χρησιμοποιείται συχνά στα προβλήματα της όρασης υπολογιστών ως μία βολική διανυσματική αναπαράσταση γραμμής [76]. Όπως θα δειχθεί στη συνέχεια, η χρήση του μοναδιαίου διανύσματος $\epsilon = \frac{\epsilon}{\|\epsilon\|}$ σε 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. Ομοίως, ορίζουμε το σύνολο $E_h = \{\epsilon_{hj}, j = 1 \cdots M\}$, με $\epsilon_{hj} = \frac{\epsilon_{hj}}{\|\epsilon_{hj}\|}$ και $\epsilon_{hj} = [a_{hj} \ -1 \ b_{hj}]^T$, όπου 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 3×3 πίνακας περιστροφής και το 3×1 διάνυσμα μετατόπισης αντίστοιχα, τα οποία πρέπει να υπολογιστούν για να καθορίσουμε την 3Δ κίνηση. Έτσι, η ακόλουθη Πρόταση υποδηλώνει ότι ο \mathbf{R} μπορεί να υπολογιστεί βάσει όλων των γνωστών στοιχείων των E_v , E_h χωρίς να είναι δεδομένη η ακριβής αντιστοίχισή τους στα στοιχεία των X_v , Y_h .

Πρόταση 2.2: Έστω \mathbf{r}_1 , \mathbf{r}_2 , \mathbf{r}_3 τα διανύσματα στήλες του πίνακα περιστροφής, δηλαδή

$$\mathbf{R} = [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3] \quad (2.3)$$

και ας ορίσουμε τον $(n+m) \times 3$ πίνακα \mathbf{Q} , που σχηματίζεται βάσει n και m εξωτερικών γινομένων προερχόμενα από τα σύνολα E_v και E_h αντίστοιχα, ως

$$\mathbf{Q} = [\varepsilon_{v1} \times \varepsilon_{v2} \quad \varepsilon_{v3} \times \varepsilon_{v4} \quad \cdots \quad \varepsilon_{h1} \times \varepsilon_{h2} \quad \varepsilon_{h3} \times \varepsilon_{h4} \quad \cdots]^T. \quad (2.4)$$

Τότε,

1) το εξωτερικό γινόμενο οποιωνδήποτε δύο στοιχείων του E_v (E_h) είναι σταθερό και ίσο με $\pm \mathbf{r}_2$ ($\pm \mathbf{r}_1$).

2) έστω $\mathbf{Q} = \mathcal{USV}^T$ η ανάλυση σε ειδικές τιμές του πίνακα \mathbf{Q} , όπου $\mathcal{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]$ και $\text{diag}(\mathcal{S}) = [s_1 \quad s_2 \quad s_3]$. 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 αν $s_1 \equiv \sqrt{n}$, τότε $\mathbf{r}_1 = \mathbf{v}_2$, $\mathbf{r}_2 = \mathbf{v}_1$ και $\mathbf{r}_3 = \mathbf{v}_3$.

△

Απόδειξη:

1) Η 3Δ μετακίνηση ενός σημείου (X, Y, Z) σε ένα σημείο (X', Y', Z') δίνεται από την εξίσωση

$$[X' \quad Y' \quad Z']^T = \mathbf{R} [X \quad Y \quad Z]^T + \mathbf{T}. \quad (2.5)$$

Για μία κατακόρυφη γραμμή ε_{vi} που ανήκει στο E_v , για σταθερά x_i , z_0 και μεταβλητό Y , $[X \quad Y \quad Z]^T \equiv [x_i \quad Y \quad z_0]^T$. Χρησιμοποιώντας την εξίσωση (2.2), λαμβάνουμε

$$\varepsilon_{vi}^T [X' \quad Y' \quad Z']^T = 0.$$

Εφαρμόζοντας την εξίσωση (2.5), παίρνουμε

$$\varepsilon_{vi}^T (\mathbf{R} [x_i \quad Y \quad z_0]^T + \mathbf{T}) \equiv \varepsilon_{vi}^T (x_i \mathbf{r}_1 + Y \mathbf{r}_2 + z_0 \mathbf{r}_3 + \mathbf{T}) = 0.$$

Εφόσον η προηγούμενη εξίσωση ισχύει για κάθε Y στην ε_{vi} τότε,

$$\begin{aligned} \varepsilon_{vi}^T \mathbf{r}_2 &= 0 \quad \text{και} \quad \text{επισης} \\ \varepsilon_{vi}^T (x_i \mathbf{r}_1 + z_0 \mathbf{r}_3 + \mathbf{T}) &= 0. \end{aligned} \quad (2.6)$$

Υπό αυτή την έννοια, για κάθε ζεύγος γραμμών στο σύνολο E_v (για παράδειγμα ε_{v1} και ε_{v2}), εφόσον ισχύει ότι $\varepsilon_{v1} \perp \mathbf{r}_2$ και $\varepsilon_{v2} \perp \mathbf{r}_2$, και το ε_{v1} δεν είναι συνευθειακό με το ε_{v2} ,

$$\varepsilon_{v1} \times \varepsilon_{v2} = \pm \mathbf{r}_2. \quad (2.7)$$

Η ισότητα ισχύει εφόσον εξ' ορισμού έχουμε ότι $\|\varepsilon_{v1} \times \varepsilon_{v2}\| = 1$. Με παρόμοιο τρόπο, για κάθε γραμμή στο E_h

$$\begin{aligned} \varepsilon_{hj}^T \mathbf{r}_1 &= 0 \quad \text{και} \quad \text{επισης} \\ \varepsilon_{hj}^T (y_j \mathbf{r}_2 + z_0 \mathbf{r}_3 + \mathbf{T}) &= 0, \end{aligned} \quad (2.8)$$

και για κάθε ζεύγος γραμμών στο E_h (για παράδειγμα ε_{h1} και ε_{h2}),

$$\varepsilon_{h1} \times \varepsilon_{h2} = \pm \mathbf{r}_1. \quad (2.9)$$

2) Από τον ορισμό του πίνακα \mathbf{Q} στην εξίσωση (2.4) και σύμφωνα με το πρώτο μέρος της απόδειξης,

$$\mathbf{Q}^T \mathbf{Q} = \sum_{k=1}^m \mathbf{r}_1 \mathbf{r}_1^T + \sum_{k=1}^n \mathbf{r}_2 \mathbf{r}_2^T = m \mathbf{r}_1 \mathbf{r}_1^T + n \mathbf{r}_2 \mathbf{r}_2^T + 0 \mathbf{r}_3 \mathbf{r}_3^T$$

και τα $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ αποτελούν ένα σύνολο από ορθοκανονικά 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 ταξινομούνται στον πίνακα \mathcal{V} αντίστοιχα με την σειρά με την οποία βρίσκονται οι ειδικές τιμές στον \mathcal{S} .

Θα πρέπει να τονιστεί στο σημείο αυτό ότι τα $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ υπολογίζονται με μία αβεβαιότητα προσήμου. Εν τούτοις, μπορεί εύκολα να αποδειχτεί ότι τα σωστά πρόσημα καθορίζονται από το γεγονός ότι αποτελούν στήλες ενός πίνακα περιστροφής και παράλληλα από το ότι στην πράξη η αποδεκτή γωνία περιστροφής περιορίζεται στο διάστημα $[-90^\circ, 90^\circ]$.

△

Χάριν απλότητας, στην εξίσωση (2.4) υποθέτουμε ότι κάθε γραμμή περιέχεται στον πίνακα \mathbf{Q} μόνο μία φορά. Ωστόσο λαμβάνοντας μία γραμμή ακόμα και περισσότερες από μία φορές, δεν δημιουργεί κανένα πρόβλημα στην ακόλουθη ανάλυση.

Πρακτικά, η Πρόταση 2.2 δηλώνει ότι άπαξ και τα δύο σύνολα από γραμμές εξαχθούν με επιτυχία στο τρέχον καρέ, τα οποία και αντιστοιχούν στις κατακόρυφες και οριζόντιες γραμμές της σκηνής αναφοράς, ο πίνακας περιστροφής υπολογίζεται βάσει της ανάλυσης ιδιαιζουσών τιμών (SVD) ενός απλού πίνακα \mathbf{Q} (στην πραγματικότητα, λύνοντας ένα γραμμικό ομογενές σύστημα ως προς τις άγνωστες παραμέτρους περιστροφής). Εφόσον δεν είναι αναγκαία οποιαδήποτε πληροφορία αντιστοιχίσεων, όλες οι διαθέσιμες γραμμές μπορούν να χρησιμοποιηθούν για τον υπολογισμό του \mathbf{R} . Αντίθετα, από τις εξισώσεις (2.6) και (2.8), μπορεί κανείς να δει ότι ο υπολογισμός του πίνακα μετατόπισης απαιτεί γνώση της αντιστοίχισης των γραμμών στην σκηνή αναφοράς με αυτές στο τρέχον καρέ, το οποίο θα ανέμενε κανείς διαισθητικά. Ωστόσο, θα δειχθεί ότι η γνώση του \mathbf{R} , σε συνδυασμό με τη στρατηγική που παρουσιάστηκε για την κατασκευή του κυανού πετάσματος, επιτρέπουν και την εύρεση των αντιστοιχίσεων των γραμμών και τον υπολογισμό του \mathbf{T} .

Όπως αναφέρθηκε στην Ενότητα 2.3, οι αντιστοιχίσεις γραμμών μεταξύ της σκηνής αναφοράς και του τρέχοντος καρέ μπορούν να πραγματοποιηθούν με τον υπολογισμό της θέσης του ορατού τμήματος του κυανού πετάσματος πάνω στον δυαδικό του χάρτη. Όμως, όπως υποδείχτηκε και στην Ενότητα 2.2, έχοντας εξάγει όλες τις διαθέσιμες γραμμές από το τρέχον καρέ και έχοντας εντοπίσει τουλάχιστον το, προκαθορισμένα αποδεκτό, ελάχιστου μεγέθους κομμάτι του δυαδικού χάρτη (ώστε να υπερκεραστούν

οποιοσδήποτε αβεβαιότητες λόγω μεγέθυνσης), παραμένει μόνο το πρόβλημα της αναγνώρισης του δυαδικού προτύπου όταν απουσιάζουν γραμμές (ακμές) στο περίγραμμα γειτονικών τετραπλεύρων λόγω του ιδίου χρωματικού τόνου. Έχοντας υπολογίσει τον πίνακα \mathbf{R} , η πιθανή απουσία γραμμών στο τρέχον καρέ μπορεί να ανιχνευθεί χρησιμοποιώντας την ακόλουθη πρόταση.

Πρόταση 2.3: Έστω ε_{v1} , ε_{v2} δύο τυχαία στοιχεία του E_v που αντιστοιχούν στα 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 τυχαίο στοιχείο ε_{v3} του E_v , το $d_{31} = x_3 - x_1$ δίνεται από τη σχέση

$$d_{31} = d_{21} \frac{\left\| \frac{\varepsilon_{v3}^T}{\varepsilon_{v3}^T \mathbf{r}_1} - \frac{\varepsilon_{v1}^T}{\varepsilon_{v1}^T \mathbf{r}_1} \right\|}{\left\| \frac{\varepsilon_{v2}^T}{\varepsilon_{v2}^T \mathbf{r}_1} - \frac{\varepsilon_{v1}^T}{\varepsilon_{v1}^T \mathbf{r}_1} \right\|}. \quad (2.10)$$

Η ίδια εξίσωση ισχύει για οποιοσδήποτε τρεις γραμμές του E_h αντικαθιστώντας τους δείκτες v με τους h και το \mathbf{r}_1 με το \mathbf{r}_2 .

△

Απόδειξη:

Για οποιοσδήποτε τρεις γραμμές του E_v , η δεύτερη εξίσωση της (2.6) δίνει

$$\begin{aligned} x_2 - x_1 &= \phi_2^T \mathbf{m} \\ x_3 - x_1 &= \phi_3^T \mathbf{m}, \end{aligned} \quad (2.11)$$

όπου

$$\begin{aligned} \mathbf{m} &\equiv -z_0 \mathbf{r}_3 - \mathbf{T}, \\ \phi_2 &\equiv \left(\frac{\varepsilon_{v2}^T}{\varepsilon_{v2}^T \mathbf{r}_1} - \frac{\varepsilon_{v1}^T}{\varepsilon_{v1}^T \mathbf{r}_1} \right)^T \text{ και} \\ \phi_3 &\equiv \left(\frac{\varepsilon_{v3}^T}{\varepsilon_{v3}^T \mathbf{r}_1} - \frac{\varepsilon_{v1}^T}{\varepsilon_{v1}^T \mathbf{r}_1} \right)^T, \end{aligned}$$

για απλοποίηση του συμβολισμού.

Ωστόσο, από την εξίσωση (2.7), έχουμε ότι $\varepsilon_{v1} \times \varepsilon_{v2} = \mathbf{r}_2$, με $\mathbf{r}_2 \perp \mathbf{r}_1$ και $\mathbf{r}_2 \perp \mathbf{r}_3$. Συνεπώς, εξ ορισμού, $\phi_2 \in \text{span}(\mathbf{r}_1, \mathbf{r}_3)$. Ταυτόχρονα, ισχύει ότι

$$\phi_2^T \mathbf{r}_1 = \left(\frac{\varepsilon_{v2}^T}{\varepsilon_{v2}^T \mathbf{r}_1} - \frac{\varepsilon_{v1}^T}{\varepsilon_{v1}^T \mathbf{r}_1} \right) \mathbf{r}_1 = \frac{\varepsilon_{v2}^T \mathbf{r}_1}{\varepsilon_{v2}^T \mathbf{r}_1} \frac{\varepsilon_{v1}^T \mathbf{r}_1}{\varepsilon_{v1}^T \mathbf{r}_1} = 0.$$

Συνεπώς, $\phi_2 \perp \mathbf{r}_1$ και τελικά $\phi_2 \parallel \mathbf{r}_3$. Οι ίδιες εξισώσεις ισχύουν και για το ϕ_3 . Έτσι, $\frac{\phi_2}{\|\phi_2\|} = \frac{\phi_3}{\|\phi_3\|} = \mathbf{r}_3$. Συνεπώς, $\phi_3^T \mathbf{m} = \frac{\|\phi_3\|}{\|\phi_2\|} \phi_2^T \mathbf{m}$, ή, $x_3 - x_1 = \frac{\|\phi_3\|}{\|\phi_2\|} (x_2 - x_1)$ το οποίο ολοκληρώνει την απόδειξη της εξίσωσης (2.10). Όμοια αποτελέσματα μπορούν να αποκομισθούν για τις γραμμές του E_h όπου τώρα η εξίσωση (2.8) χρησιμοποιείται και προκύπτει αντίστοιχα ότι $\phi \in \text{span}(\mathbf{r}_2, \mathbf{r}_3)$.

△

Η Πρόταση 2.3 παρέχει έναν πλεονεκτικό τρόπο για την κατάτμηση του ορατού τμήματος του κυανού πετάσματος στα στοιχειώδη τετράπλευρα και την εξαγωγή του αντίστοιχου δυαδικού προτύπου. Με τον εντοπισμό του τελευταίου στο δυαδικό χάρτη του κυανού πετάσματος, επιτυγχάνεται η αντιστοίχιση των γραμμών. Δηλαδή, υπό την υιοθετημένη μαθηματική ορολογία, για κάθε γνωστό στοιχείο του E_v (ή του E_h) έχει βρεθεί το αντίστοιχό του στοιχείο στο X_v (ή στο Y_h).

Διαισθητικά, αν ξεχάσουμε το δυαδικό πρότυπο του κυανού πετάσματος και υποθέσουμε ότι μόνο οι γραμμές του πλέγματος που σχηματίζεται ήταν διαθέσιμες στο λαμβανόμενο καρέ, τότε η πληροφορία που κυρίως θα απουσίαζε θα ήταν αυτή της 3Δ μετατόπισης. Με άλλα λόγια, θα ήταν δυνατό να εξάγουμε την 3Δ περιστροφή αλλά δεν θα ήταν διαθέσιμη καμία επιπλέον πληροφορία για την μετατόπιση, καθώς το πλέγμα που σχηματίζουν οι γραμμές είναι παρόμοιο για κάθε λαμβανόμενο πρότυπο. Υπό αυτή την έννοια, η προτεινόμενη μεθοδολογία για την κατασκευή του κυανού πετάσματος, όχι μόνο μας παρέχει εμμέσως ένα πλεονεκτικό πλέγμα από γραμμές για τον υπολογισμό της 3Δ περιστροφής αλλά ταυτόχρονα επιτρέπει τον υπολογισμό της 3Δ μετατόπισης. Πρέπει επίσης να σημειωθεί σε αυτό το σημείο, ότι για τον υπολογισμό της 3Δ μετατόπισης είναι αναγκαία η πληροφορία για το βάθος στη σκηνή αναφοράς (z_0). Γί αυτό ακριβώς το λόγο, αν είχαμε θεωρήσει 2Δ-σε-2Δ αντιστοιχίσεις γραμμών, θα ήταν εφικτός μόνο ο υπολογισμός της κατεύθυνσης της μετατόπισης ως ένα μοναδιαίο διάνυσμα (βλέπε για παράδειγμα την [87]).

Με βάση τα παραπάνω, το διάνυσμα της 3Δ μετατόπισης \mathbf{T} υπολογίζεται από την Πρόταση 2.4.

Πρόταση 2.4: Έστω ότι τα $\varepsilon_{v1}, \dots, \varepsilon_{vn}$ είναι n γνωστά στοιχεία του E_v που αντιστοιχούν στα x_1, \dots, x_n του X_v . Έστω επίσης ότι τα $\varepsilon_{h1}, \dots, \varepsilon_{hm}$ είναι m γνωστά στοιχεία του E_h που αντιστοιχούν στα y_1, \dots, y_m του Y_h . Ορίζοντας τους 3×3 πίνακες $\mathbf{V} = \sum_{i=1}^n x_i \varepsilon_{vi} \varepsilon_{vi}^T$, $\mathbf{H} = \sum_{j=1}^m y_j \varepsilon_{hj} \varepsilon_{hj}^T$, και $\mathbf{U} = \sum_{i=1}^n \varepsilon_{vi} \varepsilon_{vi}^T + \sum_{j=1}^m \varepsilon_{hj} \varepsilon_{hj}^T$, μία εκτίμηση ελαχίστων τετραγώνων για το διάνυσμα μετατόπισης \mathbf{T} δίνεται από τη σχέση

$$\mathbf{T} = -\mathbf{U}^{-1}\mathbf{V} \mathbf{r}_1 - \mathbf{U}^{-1}\mathbf{H} \mathbf{r}_2 - z_0 \mathbf{r}_3. \quad (2.12)$$

△

Απόδειξη:

Έστω ξανά ότι $\mathbf{m} = -z_0 \mathbf{r}_3 - \mathbf{T}$. Από τις εξισώσεις (2.6) και (2.8) έχουμε ότι

$$\begin{aligned} \varepsilon_{vi}^T \mathbf{m} &= x_i \varepsilon_{vi}^T \mathbf{r}_1 \\ \varepsilon_{hj}^T \mathbf{m} &= y_j \varepsilon_{hj}^T \mathbf{r}_2. \end{aligned} \quad (2.13)$$

Λαμβάνοντας n και m αντίστοιχα τέτοιες εξισώσεις, καταλήγουμε σε ένα υπερ-ορισμένο γραμμικό σύστημα της μορφής $\mathcal{M}\mathbf{m} = \mathcal{N}$, όπου $\mathcal{M} = [\varepsilon_{v1} \ \varepsilon_{v2} \ \dots \ \varepsilon_{h1} \ \varepsilon_{h2} \ \dots]^T$, ενώ $\mathcal{N} = [x_1 \varepsilon_{v1} \ x_2 \varepsilon_{v2} \ \dots \ y_1 \varepsilon_{h1} \ y_2 \varepsilon_{h2} \ \dots]^T$, που οδηγεί σε μία λύση ελαχίστων τετραγώνων $\mathbf{m} = (\mathcal{M}^T \mathcal{M})^{-1} (\mathcal{M}^T \mathcal{N})$.

Μπορεί να επιβεβαιωθεί αριθμητικά ότι $\mathcal{M}^T \mathcal{M} \equiv \mathbf{U}$ και $\mathcal{M}^T \mathcal{N} \equiv \mathbf{V} \mathbf{r}_1 + \mathbf{H} \mathbf{r}_2$. Συνεπώς, $\mathbf{T} = -z_0 \mathbf{r}_3 - \mathbf{m} = -z_0 \mathbf{r}_3 - \mathbf{U}^{-1} \mathbf{V} \mathbf{r}_1 - \mathbf{U}^{-1} \mathbf{H} \mathbf{r}_2$, και έτσι η Πρόταση 2.4 έχει αποδειχθεί.

△

Ο υπολογισμός του διανύσματος μετατόπισης ολοκληρώνει τον υπολογισμό της 3Δ κίνησης της κάμερας ως προς τη σκηνή αναφοράς. Μπορεί τώρα να γίνει προφανές το γιατί δεν απαιτείται οποιαδήποτε διαδικασία βαθμονόμησης των καμερών πριν ξεκινήσει η κινηματογράφηση.

Ας μελετήσουμε εκ νέου την περίπτωση της σύνδεσης μεταξύ μίας πραγματικής κάμερας που καταγράφει μία σκηνή σε πραγματικό χρόνο και μίας εικονικής κάμερας που

οπτικοποιεί (render) έναν εικονικό κόσμο (camera ganging). Ουσιαστικά, η πραγματική κάμερα λειτουργεί επιτελικά, και η εικονική αντιγράφει την κίνησή της. Μόλις οι παράμετροι 3Δ περιστροφής και μετατόπισης της πραγματικής κάμερας καθοριστούν από το επεξεργασμένο καρέ και μεταβιβαστούν στην εικονική κάμερα, η νέα θέση της τελευταίας υπολογίζεται ως η μετακίνησή της από την (γνωστή) θέση που αντιστοιχεί στη σκηνή αναφοράς χρησιμοποιώντας μόνο τις τρέχουσες παραμέτρους περιστροφής και μετατόπισης. Με αυτό τον τρόπο η σύνδεση μεταξύ των καμερών πραγματοποιείται χωρίς καμία διαδικασία βαθμονόμησης ή επιπλέον ρύθμισης.

Σε θεωρητικές προσεγγίσεις, υποτίθεται συχνά ότι η εστιακή απόσταση είναι γνωστή και τίθεται ίση με τη μονάδα ($f = 1$) για λόγους απλότητας. Εν τούτοις, αυτό δεν είναι γενικά αποδεκτό σε εφαρμογές πραγματικού κόσμου, όπου όχι μόνο η εστιακή απόσταση διαφέρει της μονάδας, αλλά επιπρόσθετα είναι άγνωστη και μεταβάλλεται με το χρόνο. Θα δειχθεί ότι, στην περίπτωση μας, η εστιακή απόσταση f μπορεί να υπολογιστεί πριν από τον υπολογισμό των πινάκων \mathbf{R} και \mathbf{T} . Έχοντας επιλέξει τη χρήση 3Δ-σε-2Δ αντιστοιχίσεων των χαρακτηριστικών, μεταξύ μίας σκηνής αναφοράς και του τρέχοντος καρέ, η άγνωστη εστιακή απόσταση θα πρέπει να υπολογίζεται για κάθε καταγραφόμενο καρέ. Με άλλα λόγια, στη διατύπωσή μας μόνο ένας άγνωστος παράγοντας f υπεισέρχεται στις εξαχθείσες εξισώσεις. Η εξίσωση (2.2) μπορεί να ξαναγραφεί ώστε να ενσωματώσει την εστιακή απόσταση ως εξής

$$\begin{bmatrix} \alpha & -1 & \frac{\beta}{f} \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = 0. \quad (2.14)$$

Στην περίπτωση όπου το f δίνεται κατά τη διάρκεια του χρόνου, οι Προτάσεις 2.2, 2.3 και 2.4 παράγουν την ακριβή λύση για τα \mathbf{R} και \mathbf{T} , όπως επισημαίνεται ανωτέρω. Αντιθέτως, όταν το f είναι άγνωστο, οι 2Δ παράμετροι των γραμμών (a, b) δεν μπορούν να υπολογιστούν από τα (α, β) . Εν τούτοις, θα δειχθεί ότι, στη γενική περίπτωση, η απόσταση f μπορεί να υπολογιστεί λύνοντας ένα απλό γραμμικό σύστημα.

Ας θεωρήσουμε, υπό αυτή την έννοια, για άγνωστο f , τα $\hat{\epsilon} = [\alpha \ -1 \ \beta]^T$ και $\hat{\epsilon} = \frac{\hat{\epsilon}}{\|\hat{\epsilon}\|}$ για τη διανυσματική αναπαράσταση κάθε γραμμής. Έστω επίσης ότι τα \hat{E}_h , \hat{E}_v είναι τα αντίστοιχα σύνολα των E_h , E_v , που σχηματίζονται βάσει των $\hat{\epsilon}$ όταν το f είναι άγνωστο. Η Πρόταση 2.5 τότε παρέχει μία γραμμική λύση για το f .

Πρόταση 2.5: Έστω ότι το \mathbf{q}_h (\mathbf{q}_v) είναι το εξωτερικό γινόμενο οποιωνδήποτε δύο στοιχείων του \hat{E}_h (\hat{E}_v) όμοια με το πρώτο μέρος της Πρότασης 2.2. 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}) περιέχει το τρίτο. Τότε το f δίνεται, με αβεβαιότητα προσήμου, από την εξίσωση

$$f^2 = \frac{\mathbf{q}_{h12}^T \mathbf{q}_{v12}}{q_{h3} q_{v3}}. \quad (2.15)$$

△

Απόδειξη:

Έστω ότι το \mathbf{q}_h (\mathbf{q}_v) ορίζεται όπως στην Πρόταση 2.5, και ότι το αντίστοιχο του \mathbf{p}_h (\mathbf{p}_v) λαμβάνεται από τα (a, b) όταν το f είναι κανονικοποιημένο στο 1. Σύμφωνα με την Πρόταση 2.2, ισχύει ότι $\mathbf{p}_v = \mathbf{r}_2$ και $\mathbf{p}_h = \mathbf{r}_1$, ή αλλιώς

$$\mathbf{p}_h^T \mathbf{p}_v = 0. \quad (2.16)$$

Διαισθητικά, η παραπάνω εξίσωση ισχύει για τα \mathbf{p}_h και \mathbf{p}_v , εφόσον, όταν το f είναι δεδομένο, μπορεί να ενσωματωθεί στη λύση μέσω των (a, b) . Ωστόσο, η εξίσωση (2.16) δεν ισχύει για τα αντίστοιχα \mathbf{q}_h και \mathbf{q}_v . Παρ' όλα αυτά, είναι εμφανές ότι τα \mathbf{p}_h και \mathbf{p}_v μπορούν να γραφούν διαφορετικά μέσω των \mathbf{q}_h , \mathbf{q}_v και f , και τότε το f μπορεί να εξαχθεί με τη χρήση της εξίσωσης (2.16). Για την ακρίβεια, για $\mathbf{q}_h = [q_{h1} \ q_{h2} \ q_{h3}]^T$ και $\mathbf{p}_h = [p_{h1} \ p_{h2} \ p_{h3}]^T$

$$\begin{bmatrix} q_{h1} \\ q_{h2} \\ q_{h3} \end{bmatrix} = \begin{bmatrix} \frac{p_{h1}}{f} \\ \frac{p_{h2}}{f} \\ p_{h3} \end{bmatrix} \frac{1}{(\frac{1}{f^2}(p_{h1}^2 + p_{h2}^2) + p_{h3}^2)^{1/2}}. \quad (2.17)$$

Κατόπιν, συνδυάζοντας τις εξισώσεις (2.17) και (2.16), προκύπτει η εξίσωση (2.15), και αποδεικνύεται η Πρόταση 2.5.

△

Στην πράξη, περισσότερα από δύο ζευγάρια στοιχείων του \hat{E}_h (\hat{E}_v) μπορούν να χρησιμοποιηθούν ώστε να εξαχθεί μία πιο εύρωστη εκτίμηση του 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 ζευγάρι στοιχείων από το \hat{E}_h και το \hat{E}_v . Όπως αναφέρθηκε ανωτέρω, όταν καθοριστεί το f , προκύπτουν τα σύνολα E_h και E_v (θέτοντας, για κάθε γραμμή, $b = \frac{\beta}{f}$), και οι Προτάσεις 2.2, 2.3 και 2.4 δίνουν ακριβή λύση για τα \mathbf{R} και \mathbf{T} .

Η εκτίμηση της εστιακής απόστασης στο χρόνο επιτρέπει την υιοθέτηση του προτεινόμενου συστήματος, χωρίς να παρουσιαστεί ανάγκη για οποιονδήποτε ηλεκτρομηχανικό εξοπλισμό (π.χ. αισθητήρες) ή για εξεζητημένα συστήματα καμερών.

2.5 Ασυμπτωτική ανάλυση σφάλματος

Όπως αναφέρθηκε στην Ενότητα 2.2.3, οι 2Δ παράμετροι των γραμμών (a, b) εξάγονται προσεγγίζοντας τη γραμμή βάσει ελαχίστων τετραγώνων (least squares line fitting). Υποθέτοντας ότι μία γραμμή καθορίζεται γενικά βάσει N διακριτών σημείων (x_i, y_i) , $i = 1 \dots N$, τότε το πρόβλημα λύνεται ως εξής

$$\begin{bmatrix} \mathbf{X} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{Y} + \mathbf{V} \quad (2.18)$$

όπου

$\mathbf{1}$ υποδηλώνει ένα διάνυσμα μεγέθους $N \times 1$ που περιέχει άσους·

$\mathbf{X} = [x_i]$.

$\mathbf{Y} = [y_i]$.

$\mathbf{V} = [v_i]$ περιέχει τους παράγοντες θορύβου που υπεισέρχονται στις μετρήσεις y_i .

Στην ακόλουθη ανάλυση, ο θόρυβος που οφείλεται στις μετρήσεις των y_i μοντελοποιείται ως μία ακολουθία λευκού θορύβου με μηδενικό μέσο. Τότε,

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} e_a \\ e_b \end{bmatrix}, \quad \mu\epsilon$$

$$\begin{bmatrix} e_a \\ e_b \end{bmatrix} = \left(\begin{bmatrix} \mathbf{X}^T \\ \mathbf{1}^T \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{1} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{1}^T \end{bmatrix} \mathbf{V} \quad (2.19)$$

όπου όλοι οι παράγοντες λάθους περιλαμβάνονται στο $[e_a \ e_b]^T$. Μετά από πολύπλοκους αλλά ευθείς υπολογισμούς, επιβεβαιώνεται ότι οι εκτιμητές των (a, b) είναι ισχυρά συνεπείς, δηλαδή δεν προκύπτει σφάλμα πόλωσης (το σφάλμα είναι μηδενικού μέσου) και η διασπορά του σφάλματος τείνει ασυμπτωτικά στο μηδέν όταν $N \rightarrow \infty$. Ακριβέστερα, $E\{e_a^2(N)\} \approx \mathcal{O}(N^{-3})$, $E\{e_a(N) e_b(N)\} \approx \mathcal{O}(N^{-2})$ και $E\{e_b^2(N)\} \approx \mathcal{O}(N^{-1})$.

Επανεξετάζοντας τον ορισμό του πίνακα \mathbf{Q} στην εξίσωση (2.4), ας θεωρήσουμε ότι το \mathbf{q}_j^T είναι η j -οστή 3×1 του γραμμής. Όπως τέθηκε στην Πρόταση 2.2, ο υπολογισμός του πίνακα περιστροφής \mathbf{R} βασίζεται στον υπολογισμό των δεξιών ειδικών διανυσμάτων του \mathbf{Q} , ή, ισοδυνάμως, 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 (2.4), λαμβάνουμε την ακόλουθη εξίσωση

$$\mathbf{q}_j \mathbf{q}_j^T = \frac{1}{\sqrt{a_k^2 + b_k^2 + 1} \sqrt{a_l^2 + b_l^2 + 1}} \begin{bmatrix} (b_k - b_l)^2 & (a_l b_k - a_k b_l)(b_k - b_l) & (a_l - a_k)(b_k - b_l) \\ (a_l b_k - a_k b_l)(b_k - b_l) & (a_l b_k - a_k b_l)^2 & (a_l b_k - a_k b_l)(a_l - a_k) \\ (a_l - a_k)(b_k - b_l) & (a_l b_k - a_k b_l)(a_l - a_k) & (a_l - a_k)^2 \end{bmatrix} \quad (2.20)$$

Όλοι οι όροι στον παραπάνω πίνακα φαίνεται να είναι λόγοι ακολουθιών οι οποίες συγκλίνουν ισχυρά στις πραγματικές τους τιμές όταν $N \rightarrow \infty$. Αυτό ισχύει, δεδομένων των τάξεων των αναμενόμενων τιμών των προηγούμενων παραγράφων και του γεγονότος ότι οι όροι που συσχετίζονται με λάθη είναι έως και δεύτερης τάξης 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 αποδείχθηκε ήδη.

Με στόχο να εξετάσουμε εάν η διασπορά του θορύβου τείνει ασυμπτωτικά στο μηδέν και για τα στοιχεία του πίνακα \mathbf{Q} , αρκεί να αποδειχθεί ότι τα $E\{e_a^4(N)\}$, $E\{e_b^4(N)\}$, $E\{e_a^2(N) e_b^2(N)\}$ τείνουν ασυμπτωτικά στο μηδέν όταν $N \rightarrow \infty$ (δεδομένης της έκφρασης των στοιχείων του \mathbf{Q} στην εξίσωση 2.20). Μετά από σχετικά επίπονες πράξεις και εκμεταλλευόμενοι τις ιδιότητες των συσσωρεύσεων (cumulants) και τις ιδιότητες του υπεισερχόμενου λευκού θορύβου (βλέπε εξίσωση 2.23), αποδεικνύεται ότι $E\{e_a^4(N)\} \approx \mathcal{O}(N^{-6})$, $E\{e_a^2(N) e_b^2(N)\} \approx \mathcal{O}(N^{-4})$ και $E\{e_b^4(N)\} \approx \mathcal{O}(N^{-2})$. Ακολούθως, δίνεται μία ενδεικτική απόδειξη για την τελευταία εξίσωση.

Απόδειξη: Θα δείξουμε ότι $E\{e_b^4(N)\} \approx \mathcal{O}(N^{-2})$. Από την εξίσωση (2.19) και μετά από μερικούς υπολογισμούς προκύπτει ότι

$$e_b = \frac{\sum_i x_i^2 \sum_j v_j - \sum_i x_i \sum_j x_j v_j}{N \sum_i x_i^2 - (\sum_i x_i)^2}, \quad (2.21)$$

και

$$(N \sum_i x_i^2 - (\sum_i x_i)^2)^4 E\{e_b^4\} = E\{(\sum_i x_i^2 \sum_j v_j - \sum_i x_i \sum_j x_j v_j)^4\}. \quad (2.22)$$

Στην αναμενόμενη τιμή του δεξιού μέλους της εξίσωσης (2.22) εμφανίζεται η ροπή τέταρτης τάξης $m_4 = E\{v_j v_k v_l v_m\}$. Μία περισσότερο συστηματική προσέγγιση προέρχεται από τη χρήση του cumulant τέταρτης τάξης $c_4(\tau_1, \tau_2, \tau_3)$ στην [55], ως

εξής

$$c_4\delta(\tau_1)\delta(\tau_2)\delta(\tau_3) = m_4(\tau_1, \tau_2, \tau_3) - \sigma^4(\delta(\tau_1)\delta(\tau_2 - \tau_3) + \delta(\tau_2)\delta(\tau_3 - \tau_1) + \delta(\tau_3)\delta(\tau_1 - \tau_2)) \quad (2.23)$$

όπου

$$\tau_1 = k - j.$$

$$\tau_2 = l - j.$$

$$\tau_3 = m - j.$$

$$c_4 \triangleq c_4(0, 0, 0).$$

Για να εκμεταλλευτούμε την εξίσωση (2.23), το δεξί μέλος της εξίσωσης (2.22) πρέπει να ξαναγραφεί κατάλληλα. Για παράδειγμα, μετά την ανάπτυξή 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 υπολογισμούς καταλήγουμε στην

$$E\{(\sum_i x_i)^4 (\sum_j x_j v_j)^4\} = (\sum_i x_i)^4 \left(c_4 \sum_i x_i^4 + 3\sigma^4 (\sum_i x_i^2)^2 \right), \quad (2.24)$$

που είναι πολυπλοκότητας $\mathcal{O}(N^{14})$. αυτό μπορεί να προέρθει χρησιμοποιώντας, για παράδειγμα, incrementals της μορφής $x_i \equiv i \cdot x_0$.

Με παρόμοιο τρόπο, έχει δειχθεί ότι όλοι οι υπόλοιποι όροι στο δεξί μέλος της εξίσωσης (2.22) είναι μικρότερης ή ίσης τάξης με την $\mathcal{O}(N^{14})$, ενώ $(N \sum_i x_i^2 - (\sum_i x_i)^2)^4 \equiv \mathcal{O}(N^{16})$. Υπό αυτή την έννοια, $E\{e_b^4(N)\} \equiv \mathcal{O}(N^{-2})$ και $E\{e_b^4(N)\} \rightarrow 0$ όταν $N \rightarrow \infty$.

△

2.6 Πειράματα και συζήτηση

Στην ενότητα αυτή παρουσιάζεται μία σειρά πειραμάτων, που έχουν πραγματοποιηθεί σε περιβάλλον προσομοίωσης αλλά και σε πραγματικές συνθήκες, ώστε να επιβεβαιωθεί η αποδοτικότητα της προτεινόμενης προσέγγισης. Για τους σκοπούς των πειραμάτων, ένα 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 εστιακή απόσταση. Αντίστοιχα, στο φυσικό περιβάλλον, μία κάμερα κατέγραψε ένα εικονικό στούντιο που κατασκευάστηκε υπό κλίμακα. Βάσει της καταγραφόμενης ακολουθίας, η κίνηση της κάμερας υπολογίζεται στα πλαίσια του προτεινόμενου αλγόριθμου και τα αποτελέσματα που λαμβάνονται συγκρίνονται με τα αντίστοιχα δεδομένα των πειραμάτων. Θα δειχθεί ακολούθως ότι οι παράμετροι της κίνησης της κάμερας υπολογίζονται με ακρίβεια, και θα παρουσιαστεί η ακολουθία που λαμβάνεται όταν τροφοδοτήσουμε μία κάμερα με

τις παραμέτρους αυτές. Έτσι φαίνεται και οπτικά η επιτυχία της μεθόδου καθώς η ακολουθία που προκύπτει είναι πανομοιότυπη με την αρχική.

Θα πρέπει να σημειωθεί εδώ ότι οι δύο τόνοι του μπλε που εμφανίζονται στα σχήματα είναι ενδεικτικοί, υπό την έννοια ότι επελέγησαν ώστε να είναι οπτικά διαχωρίσιμοι. Πρακτικά, οι τόνοι του μπλε καθορίζονται στα πλαίσια της τεχνικής διαχωρισμού χρώματος που υιοθετείται. Όπως επισημάνθηκε και στην Ενότητα 2.1 ο διαχωρισμός χρώματος (chromakeying) είναι πέραν του αντικειμένου αυτού του κεφαλαίου. Ενδεικτικά, ο διαχωρισμός χρώματος έχει μελετηθεί εις βάθος στην βιβλιογραφία (βλ. [33, 36]).

2.6.1 Πειράματα σε περιβάλλον προσομοίωσης

Για τα πειράματα που πραγματοποιήθηκαν, κατασκευάστηκε ένα κυανό πέτασμα με τη μεθοδολογία που παρουσιάστηκε στην Ενότητα 2.3. Υποθέτουμε 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. Υποθέτουμε επίσης ότι τουλάχιστον το ένα έβδομο, κατά την κατακόρυφη διάσταση του κυανού πετάσματος, θα πρέπει να βρίσκεται μέσα στο οπτικό πεδίο της κάμερας, δηλαδή 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, διαδοχικά, μειώνονται σε $N = 34$ ώστε να χωρέσουν στην 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 του τετραπλεύρου στα $12cm$, καταλήγουμε σε $m = 3$ τετράπλευρα που πρέπει να είναι ορατά κατά την οριζόντια διάσταση.

Όπως εξηγήθηκε στην Ενότητα 2.3, ένας μεγάλος αριθμός τετραπλεύρων M μπορεί να συμπεριληφθεί στη διάσταση που επεξεργαζόμαστε δεύτερη (σε αυτήν την περίπτωση, $(2^n - 1)^{(m-1)} + m - 1 = 963$ τετράπλευρα). Ωστόσο, ο αριθμός των τετραπλεύρων περιορίζεται από την οριζόντια διάσταση του τοίχου, καταλήγοντας σε $M = 528/12 = 44$ τετράπλευρα. Γενικά, το μικρότερο δυνατό τμήμα του τοίχου που υποτίθεται ότι είναι ορατό από την κάμερα θα πρέπει να επιλέγεται επαρκώς μικρό, καθώς το κυανό πέτασμα μπορεί να κατασκευάζεται μία μόνο φορά, και θα πρέπει να μπορεί να εξυπηρετεί παραγωγές διαφόρων τύπων. Επιπρόσθετα, διαισθητικά, η οριζόντια διάσταση του ορατού τμήματος θα πρέπει να επιλέγεται γενικά μικρότερη από την κατακόρυφη, εφόσον η επικάλυψη από τα αντικείμενα που θα κινούνται στο προσκήνιο αναμένεται να είναι μεγαλύτερη. Στην περίπτωση μας, καταλήγουμε σε ένα ελάχιστο τμήμα αποτελούμενο από $n \times m = 5 \times 3$ τετράπλευρα, το οποίο αναγνωρίζεται μοναδικά σε ένα δυαδικό πίνακα \mathbf{B} μεγέθους 34×44 τετραπλεύρων.

Κατά τη διαδικασία της κατασκευής, οι γνωστές παράμετροι της 3Δ δομής στη σκηνής αναφοράς αποθηκεύονται, σύμφωνα με την Ενότητα 2.4, ως εξής

1) το σύνολο των κατακόρυφων ευθειών

$$X_v = \left\{ -\frac{528}{2}, -\frac{528}{2} + 12, -\frac{528}{2} + 2 \cdot 12, \dots, \frac{528}{2} \right\},$$

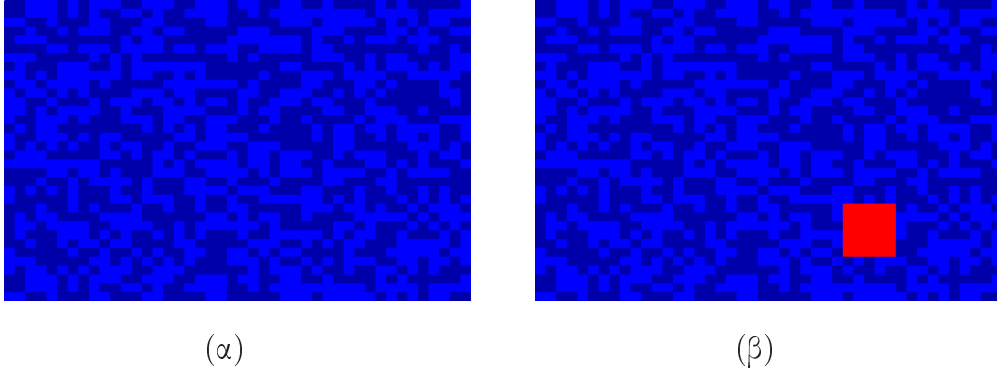
2) το σύνολο των οριζόντιων ευθειών

$$Y_h = \left\{ -\frac{340}{2}, -\frac{340}{2} + 10, -\frac{528}{2} + 2 \cdot 10, \dots, \frac{340}{2} \right\} \text{ και}$$

3) το 3Δ βάθος του κυανού πετάσματος

$$z \triangleq z_0 \equiv 2.$$

Πρακτικά, οι κατακόρυφες και οριζόντιες γραμμές, όπως επίσης και το 3Δ βάθος καθορίζονται βάσει της υποτιθέμενης σκηνής αναφοράς. Για την ακρίβεια, εφόσον η



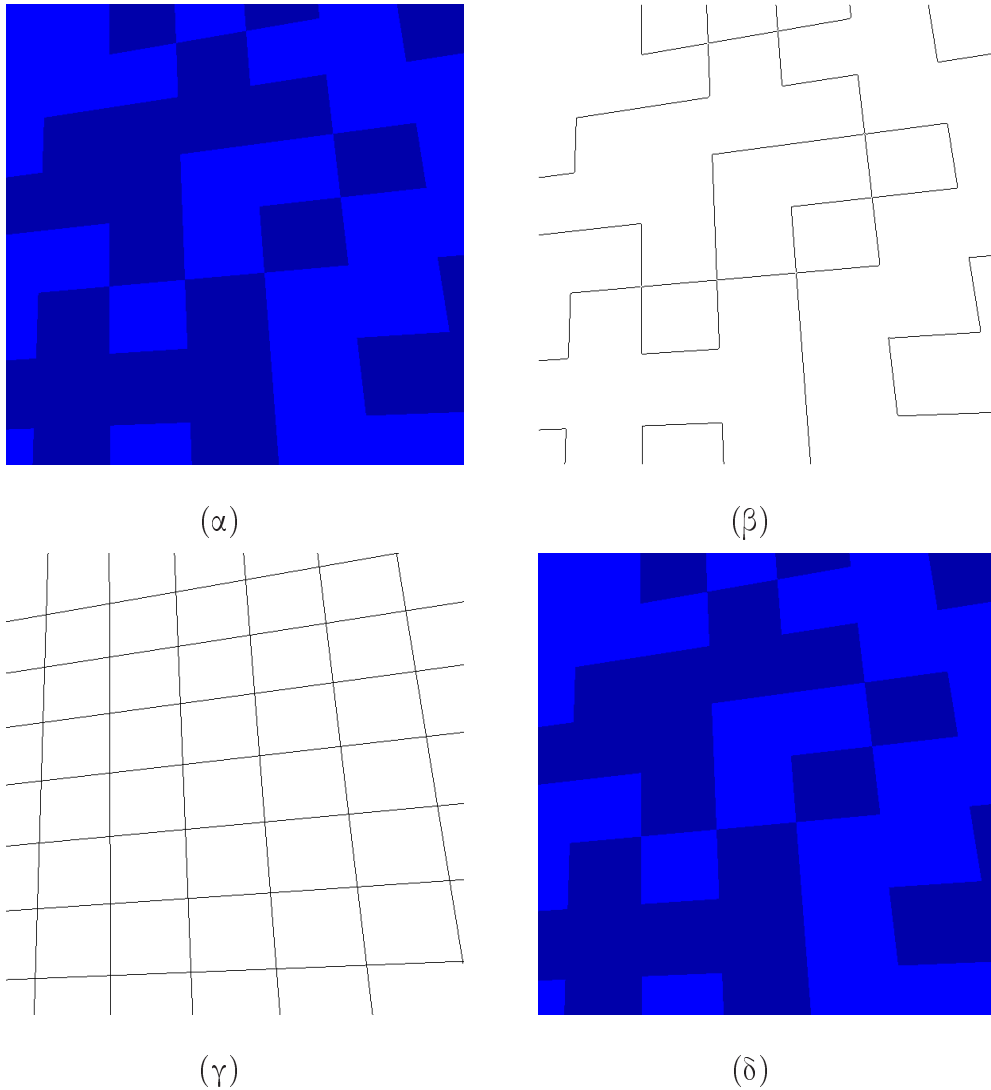
Σχήμα 2.3: Το κυανό πέτασμα (α) Ένα κυανό πέτασμα κατασκευασμένο με δύο αποχρώσεις (β) Το τμήμα του κυανού πετάσματος που καταγράφεται από την κάμερα σημειωμένο με κόκκινο χρώμα.

τελευταία δεν καταγράφεται ποτέ από την πραγματική κάμερα, οι ανωτέρω παράμετροι δομής ορίζονται πρακτικά από την υποτιθέμενη αρχική θέση της εικονικής κάμερας σε σχέση με τον τοίχο του κυανού πετάσματος. Υπό αυτή την έννοια, τοποθετώντας την εικονική κάμερα στο γεωμετρικό κέντρο της εικόνας και με τις γραμμές σάρωσης παράλληλες στις οριζόντιες γραμμές αναφοράς, λαμβάνουμε το συγκεκριμένο πλέγμα από οριζόντιες και κατακόρυφες γραμμές. Την ίδια στιγμή, ορίζοντας την απόσταση μεταξύ του CCD και του κέντρου της οθόνης ίση με z_0 , λαμβάνουμε το 3Δ βάθος.

Στο Σχήμα 2.3(α) απεικονίζεται το κατασκευασμένο (εικονικό) κυανό πέτασμα. Για μία αυθαίρετη 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 εστιακή απόσταση $f = 6$, η κάμερα καταγράφει το τμήμα του κυανού πετάσματος που σημειώνεται με κόκκινο χρώμα στο Σχήμα 2.3(β) και το συγκεκριμένο καρέ απεικονίζεται στο Σχήμα 2.4(α). Σε αυτό το συγκεκριμένο παράδειγμα, η κάμερα καταγράφει noninterlaced έγχρωμες εικόνες, μεγέθους 576×576 , με αναλογία όψεων (aspect ratio) ίση με 1.

Μετά το πέρας του βήματος που αφορά στην ανίχνευση ακμών, που χρησιμοποιεί τον τελεστή Sobel, λαμβάνεται ο δυαδικός πίνακας όπως αυτός φαίνεται στο Σχήμα 2.4(β). Οι ευθείες που εντοπίζονται από την μονάδα μετασχηματισμού Hough, χρησιμοποιώντας έναν πίνακα έρευνας (look-up table) μεγέθους 1024×1024 , εμφανίζονται στο Σχήμα 2.4(γ). Οι γραμμές που επικρατούν τροφοδοτούν τον αλγόριθμο υπολογισμού της κίνησης της κάμερας όπως περιγράφεται στην Ενότητα 2.4). Οι παράμετροι μεγέθυνσης και περιστροφής λαμβάνονται ευθέως με τη χρήση των Προτάσεων 2.5 και 2.2 αντίστοιχα· για την εστιακή απόσταση $\hat{f} = 6.09$ και για τη γωνία περιστροφής $\hat{\alpha} = 17.2^\circ$ και τον άξονα $\hat{\mathbf{u}} = [.762, .584, .280]^T$. Όπως μπορεί να διαπιστωθεί, οι υπολογιζόμενες παράμετροι είναι αξιοσημείωτα κοντά στις πραγματικές τιμές.

Παρατηρώντας τα Σχήματα 2.4(α) και 2.4(γ), γίνεται εμφανές ότι η απουσία μεγάλων ομοιόμορφων περιοχών οδηγεί στον εντοπισμό όλων των γραμμών αναφοράς του καταγραφόμενου καρέ, η οποία είναι πρακτικά και η συνήθης περίπτωση (βλέπε στη συνέχεια για την περίπτωση απουσίας γραμμών του πλέγματος). Έτσι, ο υποπίνακας του \mathbf{B} που αντιστοιχεί στο ορατό τμήμα του κυανού πετάσματος μπορεί να εξαχθεί



Σχήμα 2.4: Όλα τα βήματα της προτεινόμενης μεθοδολογίας για τον υπολογισμό της κίνησης της κάμερας στο παράδειγμα της ενότητας 2.6.1: (α) το καρέ που καταγράφηκε (βλ. Σχήμα 2.3(β)), το αποτέλεσμα της ανίχνευση ακμών, (γ) το πλέγμα γραμμών αναφοράς που εξάχθηκε (δ) το καρέ όπως ανακατασκευάστηκε χρησιμοποιώντας τις παραμέτρους κίνησης που υπολογίστηκαν.

ευθέως και είναι ο ακόλουθος:

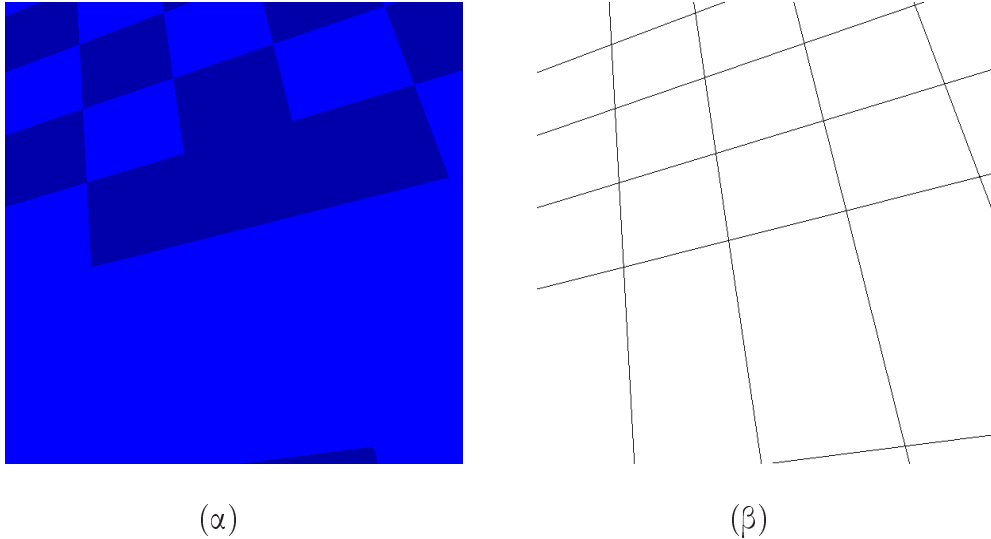
$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Για την ακρίβεια, όπως εξηγήθηκε ανωτέρω, ακόμη και ένας 5×3 υποπίνακας του \mathbf{S} θα επαρκούσε για να εντοπισθεί μοναδικά το τμήμα πάνω στο δυαδικό χάρτη του πετάσματος. Ο εντοπισμός του πίνακα \mathbf{S} πάνω στο δυαδικό χάρτη επιτρέπει την εύρεση των αντιστοιχίσεων μεταξύ των αρχικών 3Δ γραμμών και των 2Δ γραμμών που εξάγονται από το τρέχον καρέ. Μετά το πέρας της διαδικασίας αντιστοίχισης, η Πρόταση 2.4 αποφέρει το 3Δ διάνυσμα μετατόπισης $\hat{\mathbf{T}} = [-120.1, 79.6, 111.3]^T$. Για να επιβεβαιώσουμε οπτικά την αποδοτικότητα της προτεινόμενης μεθόδου, το καρέ του Σχήματος 2.4(α) αποδίδεται εκ νέου χρησιμοποιώντας τις εξαχθείσες παραμέτρους κίνησης και τροφοδοτώντας με αυτές την εικονική κάμερα. Υπό αυτή την έννοια, το καρέ του Σχήματος 2.4(δ) μπορεί να συγκριθεί απευθείας με αυτό του 2.4(α). Συγκρίνοντας τα δύο αυτά σχήματα, μπορεί να επισημανθεί ότι η ίδια σκηνή όπως αποδίδεται από μία διαφορετική εικονική κάμερα, χρησιμοποιώντας τις παραμέτρους κίνησης που υπολογίστηκαν, θα ήταν σωστά ευθυγραμμισμένη ώστε να χρησιμοποιηθεί κατά τη διαδικασία σύνθεσης.

Όπως αναλύθηκε στην Ενότητα 2.3, εφόσον η ακολουθία μέγιστου μήκους \mathbf{c} εξάγεται για τη μικρότερη πλευρά του κυανού πετάσματος, ο 2Δ δυαδικός χάρτης \mathbf{B} υπολογίζεται βάσει κατάλληλων συνδυασμών των ολισθήσεων του \mathbf{c} . Είναι γενικά εμφανές ότι, χρησιμοποιώντας κατάλληλους συνδυασμούς από ολισθημένες μέγιστες ακολουθίες αποτρέπεται η εμφάνιση μεγάλων ομοιόμορφων περιοχών από σκούρο ή ανοιχτό μπλε πάνω στο κυανό πέτασμα. Εν τούτοις, για τέτοιες σπάνιες περιπτώσεις, που σίγουρα θα παρουσιαστούν εκτός από τη γενική περίπτωση, η Πρόταση 2.3 εξασφαλίζει ότι το δυαδικό πρότυπο \mathbf{S} θα εξαχθεί επιτυχημένα.

Το Σχήμα 2.5(α) απεικονίζει ένα τέτοιο καταγεγραμμένο καρέ, ενώ οι γραμμές αναφοράς που εντοπίζονται φαίνονται στο Σχήμα 2.5(β). Παρατηρώντας απλά τα Σχήματα 2.5(α) και (β), είναι εμφανές ότι, για να εξαχθεί ένας ελάχιστος υποπίνακας μεγέθους 5×3 , θα πρέπει να εντοπιστεί η γραμμή του πλέγματος που έχει απορροφηθεί στη μεγάλη περιοχή που είναι χρωματισμένη με πιο ανοιχτό μπλε. Σε αυτήν την περίπτωση, η εξίσωση (2.10) πρέπει να χρησιμοποιηθεί για να ελέγξει όλες τις γραμμές που έχουν εξαχθεί στο E_h : σημειώστε ότι τα f και \mathbf{R} έχουν ήδη ληφθεί σε αυτό το σημείο. Γενικά, οι γραμμές του συνόλου E_h (ή του E_v) είναι διατεταγμένες χωρικά (διαδικασία τετριμμένη μιας και δεν συμβάλλουν μέσα σε ένα καρέ) και η μεταξύ τους απόστασή, ανά δύο, υπολογίζεται από την εξίσωση (2.10). Όταν μία εκ των αποστάσεων ξεπερνάει σημαντικά τη μονάδα τότε εντοπίζεται μία απύσχα γραμμή. Πρακτικά, η γραμμή αυτή δεν είναι αναγκαίο να εντοπισθεί, αφού φθάνει να εντοπιστεί μόνο η απουσία της. Ενώ τελικά, όταν εξαχθεί ο \mathbf{S} , αρκεί να διπλασιαστεί η αντίστοιχη γραμμή του (για γραμμές του E_h) ή η αντίστοιχη στήλη του (για γραμμές του E_v).

Στον Πίνακα 2.2 παρουσιάζονται ενδεικτικά αποτελέσματα για τον εντοπισμό γραμμών αναφοράς που απουσιάζουν από το πλέγμα που εξάχθηκε. Αρχίζοντας από αριστερά, οι δύο πρώτες στήλες περιέχουν τις παραμέτρους των γραμμών όπως αυτές εξάγονται από τον μετασχηματισμό Hough, ενώ οι επόμενες δύο στήλες περιέχουν



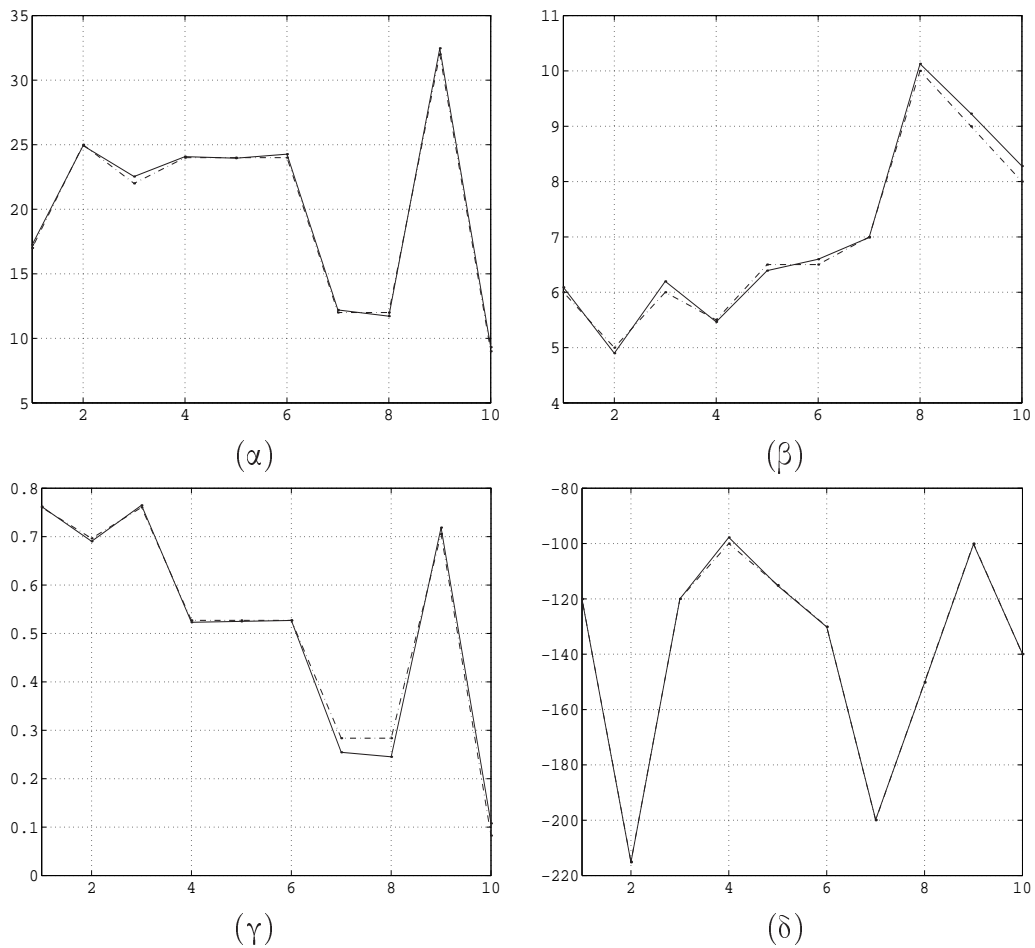
Σχήμα 2.5: Περίπτωση απουσίας γραμμών αναφοράς (α) Καταγεγραμμένο καρέ από το οποίο απουσιάζουν κάποιες από τις γραμμές αναφοράς του πλέγματος και (β) οι γραμμές αναφοράς που εντοπίζονται.

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 2.2: Εντοπισμός γραμμών αναφοράς που απουσιάζουν από το πλέγμα. Οι στήλες από αριστερά προς τα δεξιά: παράμετροι γραμμών όπως προκύπτουν από τον μετασχηματισμό *Hough*, οι τελευταίες μετά από κανονικοποίηση (βάσει της μεγέθυνσης και των πραγματικών διαστάσεων του *CCD*), οι κανονικοποιημένες αποστάσεις μεταξύ των στοιχισμένων γραμμών σε ακολουθιακά ζεύγη όπως αυτές υπολογίζονται από την Πρόταση 2.3.

τις αντίστοιχες παραμέτρους κανονικοποιημένες ως προς την υπολογισμένη εστιακή απόσταση και τα μήκη των πλευρών του *CCD*. Η τελευταία στήλη απεικονίζει τις κανονικοποιημένες αποστάσεις μεταξύ των γραμμών στη σκηνή αναφοράς, όπως αυτές έχουν εκτιμηθεί σε σχέση με την Πρόταση 2.3. Μπορεί να δει κανείς ότι, η κανονικοποιημένη απόσταση μεταξύ των γραμμών 4 και 5 (4η και 5η σειρά στον Πίνακα) είναι σχεδόν ίση με δύο (σειρά 5 στον Πίνακα), κάτι που υποδηλώνει ότι η γραμμή του προτύπου, που περιλαμβάνεται μεταξύ των γραμμών 4 και 5, θα πρέπει να αντιγραφεί μία φορά (δεδομένου ότι οι δύο γραμμές που θα προκύψουν πρέπει να περιέχουν την ίδια πληροφορία και ίση με αυτή που περιείχε η "διπλή γραμμή"). Παρόμοια αποτελέσματα λαμβάνονται και για απουσία κατακόρυφων γραμμών.

Αν και η απουσία γραμμών είναι σπάνια (αν όχι αδύνατη) για την προτεινόμενη μεθοδολογία κατασκευής του κυανού πετάσματος, είναι μάλλον συνηθής περίπτωση για το κυανό πέτασμα που προτείνεται στις [92, 20]. Επιπρόσθετα, υποτίθεται ότι θα υπάρχει πάντα ένα ορατό 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 [92, 20]. Η απόσυρση της απαίτησης αυτής θα οδηγούσε κατευθείαν σε ανακριβείς μετρήσεις της μεγέθυνσης.



Σχήμα 2.6: Εκτιμήσεις (απλή γραμμή) και πραγματικές τιμές (διακεκομμένη γραμμή) των παραμέτρων της 3Δ κίνησης της κάμερας συναρτήσει του χρόνου, για μία τυχαία ακολουθία (α) η γωνία περιστροφής, (β) η εστιακή απόσταση, (γ) η πρώτη συνιστώσα του άξονα περιστροφής και (δ) η πρώτη συνιστώσα του διανύσματος μετατόπισης.

Η επίδοση του αλγορίθμου εξετάστηκε σε έναν μεγάλο αριθμό από αντίστοιχα πειράματα σε περιβάλλον προσομοίωσης, χρησιμοποιώντας διαφορετικά κυανά πετάσματα και αυθαίρετες παραμέτρους κίνησης για την κάμερα. Σε όλα τα πειράματα που εκτελέστηκαν, η γωνία περιστροφής και η εστιακή απόσταση που υπολογίστηκαν δεν διέφεραν περισσότερο από $.3^\circ$ και $.1$ αντίστοιχα από τις πραγματικές τιμές. Ενδεικτικά αποτελέσματα του υπολογισμού των παραμέτρων της κίνησης παρουσιάζονται στο Σχήμα 2.6. Το πείραμα αφορά μία ακολουθία από 100 καρέ, στην οποία οι παράμετροι κίνησης της κάμερας μεταβάλλονται τυχαία με το χρόνο. Στα διαγράμματα 2.6(α-δ) απεικονίζονται οι τιμές που υπολογίστηκαν για τη γωνία περιστροφής, το εστιακό μήκος, την πρώτη συνιστώσα του άξονα περιστροφής και την πρώτη συνιστώσα του διανύσματος μετατόπισης αντίστοιχα, για κάθε 10 καρέ. Οι τιμές αυτές αντιστοιχούν στις κανονικές γραμμές ενώ απεικονίζονται με διακεκομμένες γραμμές και οι πραγματικές τιμές.

2.6.2 Πειράματα σε πραγματικό περιβάλλον

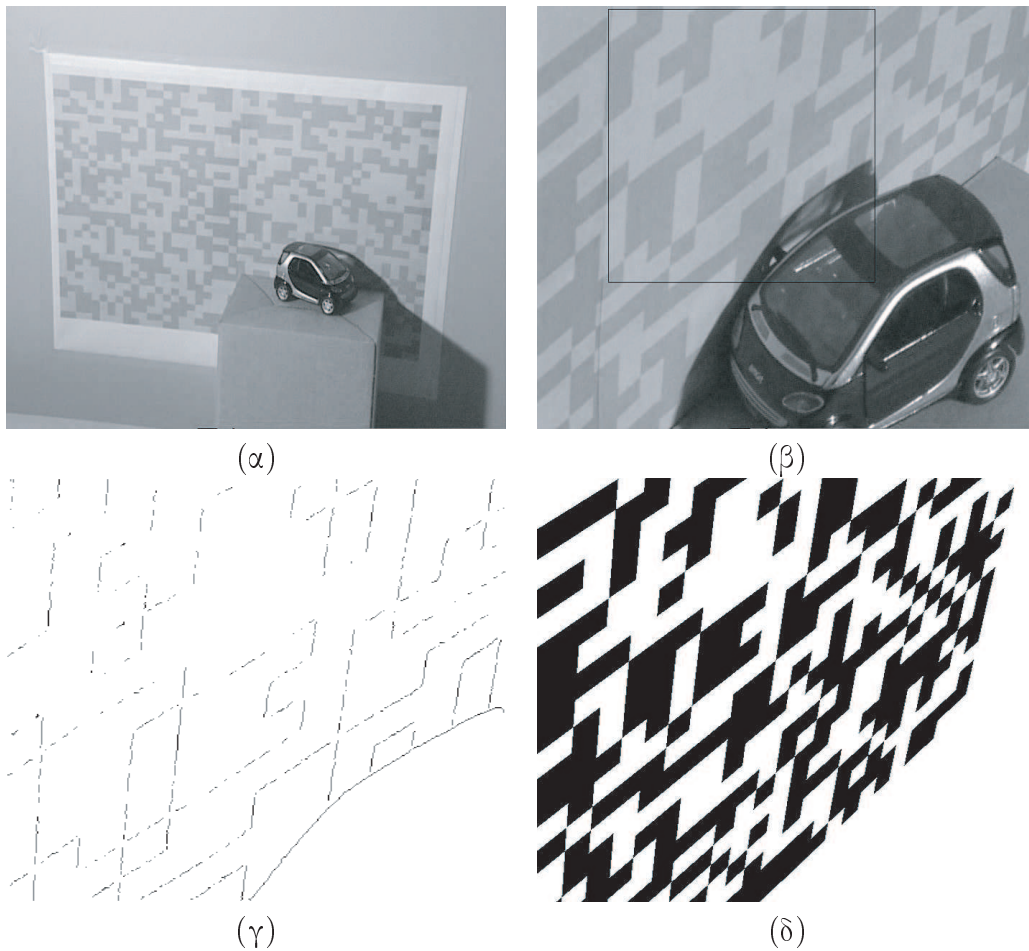
Η επίδοση του αλγορίθμου εξετάστηκε και μέσω πειραμάτων σε πραγματικό περιβάλλον, χρησιμοποιώντας ένα εικονικό στούντιο κατασκευάστηκε υπό κλίμακα και φαίνεται στο Σχήμα 2.7(α). Για το καρέ του παραδείγματος που απεικονίζεται στο Σχήμα 2.7(β), το οποίο καταγράφηκε από επαγγελματική ψηφιακή κάμερα με CCD φυσικών διαστάσεων $6.6mm \times 8.8mm$ και με διαστάσεις σε pixels 752×582 , ο προτεινόμενος αλγόριθμος εφαρμόστηκε στο τμήμα που υποδεικνύεται από το μαύρο παραλληλόγραμμο περίγραμμα. Τα αντίστοιχα αποτελέσματα της ανίχνευσης ακμών απεικονίζονται στο Σχήμα 2.7(γ), ενώ το ανακατασκευασμένο καρέ, όπως το απέδωσε μία εικονική κάμερα υπό τις εξαχθείσες παραμέτρους κίνησης, δίνεται στο Σχήμα 2.7(δ). Συγκρίνοντας τα Σχήματα 2.7(β) και 2.7(δ) μπορεί κανείς να συμπεράνει ότι οι παράμετροι που υπολογίστηκαν είναι επίσης αξιοσημείωτα κοντά στις πραγματικές τιμές. Για λόγους πληρότητας παραθέτουμε ακολούθως τις παραμέτρους που υπολογίστηκαν: $\hat{\alpha} = 68.6204^\circ$, $\hat{\mathbf{u}} = [-.563, -.805, .189]^T$, $\hat{\mathbf{T}} = [-138, -28, 1400]^T$ και $\hat{f} = 29.2414$ (οι μετρήσεις μήκους είναι σε χιλιοστά).

Τέλος, θα πρέπει να επισημανθεί ότι η ακρίβεια που επιτυγχάνεται από το προτεινόμενο σύστημα επηρεάζεται μόνο από την επίδοση του μετασχηματισμού Hough. Υπό αυτή την έννοια, χρησιμοποιώντας ακόμη μεγαλύτερους πίνακες εύρεσης (look-up tables), το σφάλμα στις παραμέτρους κίνησης που υπολογίζονται μπορεί να μειωθεί δραματικά. Το τελευταίο έχει να κάνει με την υπολογιστική ισχύ που είναι διαθέσιμη και είναι άμεσα εξαρτημένο από τις απαιτήσεις για επιδόσεις σε πραγματικό χρόνο.

2.7 Συμπεράσματα και επεκτάσεις

Στο παρόν κεφάλαιο προτείναμε ένα νέο σύστημα για τον υπολογισμό της κίνησης της κάμερας σε εικονικά στούντιο βάσει ενός κυανού πετάσματος δύο αποχρώσεων. Αναπτύχθηκε μία καινούρια μεθοδολογία για την κατασκευή του κυανού πετάσματος και του αντίστοιχου 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, αποδεικνύεται ότι εξάγουν ακριβείς εκτιμήσεις της κίνησης οδηγώντας σε επιτυχή σύνθεση εικόνων και βίντεο υπό διάφορες συνθήκες.

Οι κύριες συνεισφορές αυτού του κεφαλαίου είναι τόσο η προτεινόμενη μεθοδολογία κατασκευής του κυανού πετάσματος όσο και η μέθοδος υπολογισμού της κίνησης



Σχήμα 2.7: Πείραμα σε φυσικό περιβάλλον: (α) το εικονικό στούντιο που κατασκευάστηκε υπό κλίμακα, (β) ένα καρέ που καταγράφεται από μία πραγματική κάμερα, (γ) οι ακμές που εξάγονται, και (δ) το καρέ που αποδίδεται από μία εικονική κάμερα χρησιμοποιώντας τις παραμέτρους της 3Δ κίνησης που υπολογίστηκαν.

της κάμερας. Όσον αφορά στην πρώτη, αποτελεί μία ευέλικτη επιλογή έναντι των ευριστικών μεθοδολογιών κατασκευής κυανών πετάσματος, η οποία μπορεί να καλύψει τις ανάγκες κάθε εικονικού στούντιο. Σχετικά με τη μέθοδο υπολογισμού της κίνησης της κάμερας, αυτή επιτυγχάνει να υπολογίσει όλες τις παραμέτρους κίνησης, συμπεριλαμβανόμενης της εστιακής απόστασης (camera focus). Πάνω από όλα, η προτεινόμενη μέθοδος εξαλείφει την ανάγκη βαθμονόμησης της κάμερας, καθώς όλες οι μετρήσεις αναφέρονται σε μία εικονική σκηνή αναφοράς, και η σχετική κίνηση της κάμερας μπορεί να εξαχθεί αξιόπιστα.

Τα αποτελέσματα που στο κεφάλαιο αυτό παρέχουν τις βάσεις για περαιτέρω βελτιώσεις και εισαγωγή νέων μεθοδολογιών στα ακόλουθα ζητήματα:

1. αποδοτική εξαγωγή της κίνησης της κάμερας σε οριακές περιπτώσεις, για παράδειγμα όταν η εικόνα που λαμβάνεται εμφανίζεται θολή λόγω κίνησης ή ανεστίασης (motion or defocus blur) ή όταν το προοπτικό μοντέλο προβολής εκφυλίζεται σε ορθογραφικό (στην εργασία [21] παρουσιάστηκαν τα πρώτα αποτελέσματα για την αντιμετώπιση αυτών των περιπτώσεων,
2. αποτελεσματική παρεμβολή (interpolation) μεταξύ διαδοχικών καρέ με στόχο τη μείωση του υπολογιστικού κόστους και την εξομάλυνση (smoothing) της ακολουθίας,
3. εκμετάλλευση των εξαγόμενων παραμέτρων (πιθανόν και της μετρούμενης ανεστίασης) για τον υπολογισμό της θέσης των ηθοποιών μέσα στο στούντιο και την παρακολούθησή τους με στόχο την κίνησή τους παράλληλα με εικονικά αντικείμενα,
4. εξαγωγή χαρακτηριστικών της κίνησης των ηθοποιών (για παράδειγμα χειρονομίες) καθώς και των χαρακτηριστικών του προσώπου τους με στόχο την αλληλεπίδρασή τους με εικονικούς ηθοποιούς.

□

Κεφάλαιο 3

Εντοπισμός και Παρακολούθηση Αντικειμένων σε Εικονικά Στούντιο

3.1 Εισαγωγή - Αναδρομή

Πολυάριθμα ερευνητικά και τεχνολογικά επιτεύγματα στους τομείς της επεξεργασίας εικόνων, της τεχνολογίας βίντεο και των γραφικών για υπολογιστές έχουν γίνει ένθερμα αποδεκτά από τη βιομηχανία της ψυχαγωγίας τα τελευταία χρόνια. Οι παραγωγοί τηλεοπτικών και κινηματογραφικών προγραμμάτων πειραματίζονται διαρκώς χρησιμοποιώντας εικονικά στούντιο και τρισδιάστατο υλικό που παράγεται από υπολογιστές, παρουσιάζοντας αρκετά, πολύ ενδιαφέροντα οπτικά εφέ και μερικές καινοτόμες εφαρμογές (βλέπε, για παράδειγμα [4]). Εικονικά αντικείμενα μπορούν να αιωρούνται, κινούμενες εικόνες μπορούν να ενσωματώνονται σε πραγματικό χρόνο και συνδέσεις μπορούν να δημιουργούνται με περιβάλλοντα αλληλεπίδρασης, όπως το διαδίκτυο. Οι ακολουθίες βίντεο που παράγονται από τέτοια συστήματα είναι βασικά συνθέσεις φυσικών και συνθετικών ακολουθιών, οι οποίες είναι είτε αποθηκευμένες ή καταγράφονται σε πραγματικό χρόνο. Στην πιο συνηθισμένη περίπτωση, η συντεθειμένη ακολουθία, που συχνά αναφέρεται ως εικονικό σκηνικό (virtual set), αποτελείται από ζωντανές ακολουθίες βίντεο και ένα προ-φορτωμένο συνθετικό ή φυσικό περιβάλλον [30]. Υπό αυτό το σκεπτικό, παρουσιάζεται αυξημένη ερευνητική δραστηριότητα και ενδιαφέρον όσον αφορά σε τέτοιου είδους συστήματα, με τις κατευθυντήριες γραμμές του προτύπου MPEG-4, που αφορούν στην προσανατολισμένη σε αντικείμενα κωδικοποίηση και την δημιουργία υβριδικών ακολουθιών μέσω σύνθεσης φυσικών και συνθετικών, να είναι ένα αξιοσημείωτο παράδειγμα.

Ένα σύστημα εικονικού στούντιο αποτελείται από τρεις βασικές μονάδες, δηλαδή τη μονάδα εντοπισμού της κίνησης της κάμερας (camera tracking module), την μονάδα οπτικοποίησης (rendering) και την μονάδα σύνθεσης (compositing). Η τελευταία περιλαμβάνει ένα κυανό πέτασμα ως φόντο, μπροστά από το οποίο η κίνηση του προσκηνίου καταγράφεται από κάμερες, και μια τεχνική διαχωρισμού χρώματος (chromakeying) για την κατάτμηση του προσκηνίου από το παρασκήνιο. Εναλλακτικά ένα νέο σύστημα έχει παρουσιαστεί [68] που παρέχει ένα ομοιόμορφο χρώμα ελέγχου (key) χωρίς να απαιτείται χρήση συγκεκριμένου φωτισμού για το φόντο. Το σύστημα αποτελείται από ένα 'δαχτυλίδι' αποτελούμενο από φωτοδιόδους (LED), το οποίο τοποθετείται στο μπροστινό μέρος του φακού της κάμερας, και ένα ειδικό αντανακλαστικό (retro-reflective) ύφασμα που κρέμεται στο παρασκήνιο. Το πανί, που φαίνεται γκρι στο ανθρώπινο μάτι, περιλαμβάνει μικρά γυάλινα σωματίδια τα οποία, όταν το φωτεινό

δαχτυλίδι ανάβει, ανακλά το μπλε ή πράσινο χρώμα που εκπέμπουν οι φωτοδιόδοι.

Μετά την αφαίρεση του φόντου, η κίνηση στο προσκήνιο συνδυάζεται με το διαθέσιμο παρασκήνιο. Η μονάδα οπτικοποίησης είναι υπεύθυνη για την ευθυγράμμιση (alignment) του καταγραφόμενου προσκηνίου και της εικονικής σκηνής. Η τελευταία είναι είτε μία ακολουθία εικόνων που έχει ήδη οπτικοποιηθεί, ή ένα εικονικό μοντέλο/κόσμος. Στην πρώτη περίπτωση, η ακολουθία που θα χρησιμεύσει ως φόντο θα πρέπει να μετασχηματιστεί ώστε να είναι συναφής με την αντίστοιχη του προσκηνίου, ενώ στη δεύτερη περίπτωση, η εικονική σκηνή αποδίδεται οπτικά από μία εικονική κάμερα βάσει της κίνησης της πραγματικής κάμερας (βλέπε, για παράδειγμα [33]).

Η μονάδα παρακολούθησης της κίνησης της κάμερας είναι, όπως φαίνεται από τις παραπάνω παραγράφους, το πιο κρίσιμο τμήμα των σύγχρονων συστημάτων εικονικών στούντιο. Ο αποδοτικός υπολογισμός της κίνησης της κάμερας καθορίζει την ευθυγράμμιση της ζωντανής ακολουθίας με το διαθέσιμο παρασκήνιο. Αρχικά, η μονάδα αυτή απουσίαζε από τα εικονικά στούντιο, περιορίζοντας τις δυνατότητες των συστημάτων σε παραγωγές όπως τα τηλεοπτικά προγράμματα των δελτίων καιρού. Έκτοτε, διάφορες μέθοδοι παρακολούθησης έχουν προταθεί ώστε να επεκτείνουν τις δυνατότητες των παραδοσιακών εικονικών στούντιο.

Τα συστήματα παρακολούθησης της κίνησης της κάμερας που χρησιμοποιούνται στα εικονικά στούντιο ταξινομούνται σε δύο ευρείες κατηγορίες, τα ηλεκτρομηχανικά και τα οπτικά. Πολλές νέες λύσεις βασίζονται σε συνδυασμό συστημάτων των δύο προαναφερθέντων κατηγοριών δημιουργώντας τα λεγόμενα υβριδικά συστήματα. Αρχέτα συστήματα έχουν αναπτυχθεί, ως πρωτότυπα ή εμπορικά προϊόντα, υιοθετώντας κυρίως ηλεκτρομηχανικές προσεγγίσεις λόγω μεγαλύτερης ακρίβειας (βλέπε στην Ενότητα 2.2 για περισσότερα παραδείγματα). Σε γενικές γραμμές, σε αυτές τις περιπτώσεις χρησιμοποιούνται ειδικά ρυθμισμένοι μηχανισμοί που ελέγχουν την κίνηση και ακολουθούν προκαθορισμένες τροχιές (ενεργά συστήματα), είτε ενσωματώνονται ειδικοί αισθητήρες που επιστρέφουν μία εκτίμηση της κίνησης της κάμερας (παθητικά συστήματα). Η αδυναμία των ηλεκτρομηχανικών συστημάτων έγκειται στο ότι επιβαρύνουν την παραγωγή με αξιοσημείωτο οικονομικό κόστος καθώς και ότι παρουσιάζουν ευαισθησία σε τυχαίες δονήσεις ή απότομες κινήσεις.

Από την άλλη πλευρά, τα οπτικά συστήματα βασίζονται σε τεχνικές επεξεργασίας εικόνων για την εξαγωγή των παραμέτρων κίνησης βάσει του τρέχοντος καταγεγραμμένου καρέ. Για την αντιστοίχιση μεταξύ χαρακτηριστικών σε διαφορετικά καρέ της ακολουθίας χρησιμοποιούνται ειδικά σημάδια (markers), μία προσέγγιση η οποία πάσχει εμφανώς σε περιπτώσεις μερικής επικάλυψης ή και ολικής απουσίας των χαρακτηριστικών αυτών. Αναφέρεται επίσης ότι η εισαγωγή ευδιάκριτων χαρακτηριστικών στην κατασκευή του κυανού πετάσματος δημιουργεί και προβλήματα στην διαδικασία διαχωρισμού χρώματος και, κατά συνέπεια, στη διαδικασία της σύνθεσης των ακολουθιών [30]. Μία εναλλακτική προσέγγιση περιλαμβάνει μία μικρή, χαμηλού κόστους κάμερα που τοποθετείται πάνω στην κάμερα του στούντιο και καταγράφει ένα σύνολο ειδικών σημαδιών (markers) που έχουν τοποθετηθεί εκτός του σκηνικού. Η πιο επιτυχημένη υλοποίηση αυτής της τεχνικής είναι το σύστημα Free-d του BBC που χρησιμοποιεί μία δέσμη από φωτοδιόδους, τοποθετημένη πάνω στην κανονική κάμερα, και γραμμοκώδικες barcodes ως σημειωτές στην οροφή του στούντιο [79]. Τέλος, όπως παρουσιάστηκε στο Κεφάλαιο 2 το παραδοσιακό κυανό πέτασμα μπορεί να επεκταθεί σε ένα κυανό πέτασμα δύο αποχρώσεων. Με αυτό τον τρόπο, τα κατάλληλα χαρακτηριστικά αναφοράς για 2Δ παρακολούθηση, όπως σημεία ή ευθείες, ενσωματώνονται χωρίς να επηρεάζουν τα υπόλοιπα υποσυστήματα.

Τα τελευταία βήματα, που γίνονται όσον αφορά στις παραγωγές με χρήση εικονικών στούντιο, είναι προς την ενσωμάτωση εικονικών κινούμενων αντικειμένων στο εικονικό σκηνικό φτάνοντας μέχρι και την χρησιμοποίηση εικονικών ηθοποιών οι οποίοι είναι δυνατό να αλληλεπιδράσουν με τους πραγματικούς. Με τον τρόπο αυτό εμπλουτίζεται η εντύπωση ότι ο άνθρωπος κινείται και μιλάει σε έναν εικονικό κόσμο. Ένας από τους κύριους περιορισμούς στις σύγχρονες τεχνικές υψηλού επιπέδου για τα εικονικά στούντιο, σχετίζεται με την ανάγκη για αφεγάδιαστη ενοποίηση του συνθετικού και του φυσικού υλικού, που θα οδηγήσει σε μία ακολουθία που θα πληρεί τις ποιοτικές προϋποθέσεις των επαγγελματικών εκπομπών, ενώ η παραγωγή και η ανάμειξη (mixing) θα πρέπει να υλοποιούνται σε πραγματικό χρόνο. Η διαδικασία της σύνθεσης όμως, παραμένει βασισμένη κύριως σε μία απλή τεχνική ανάμειξης η οποία δεν περιλαμβάνει τις χωρικές και φυσικές σχέσεις μεταξύ των πραγματικών και των εικονικών αντικειμένων σε μία σκηνή.

Είναι εμφανές ότι, η παροχή μέσων μέτρησης της ακριβούς θέσης και του προσανατολισμού των φυσικών αντικειμένων και ηθοποιών πάνω στο σκηνικό θα προσφέρει μία ποικιλία νέων προοπτικών. Προς το παρόν, οι ηθοποιοί υποχρεώνονται να στέκονται σε συγκεκριμένες θέσεις στο χώρο ή να ακολουθούν προδιαγεγραμμένες πορείες, σε συγκεκριμένα μάλιστα χρονικά διαστήματα. Το ίδιο ισχύει και για τα εικονικά αντικείμενα που εμφανίζονται στη σκηνή και επιβαρύνουν σημαντικά τα καθήκοντα της παραγωγής πριν την κινηματογράφηση. Η παροχή της δυνατότητας αυτόματης εξαγωγής πληροφοριών για τη γεωμετρία μίας σκηνής κατά την κινηματογράφηση προϋποθέτει τον εντοπισμό και την παρακολούθηση της κίνησης των αντικειμένων και των ηθοποιών που στέκονται ή κινούνται σε αυτή. Υπάρχει μία ποικιλία προσεγγίσεων στην παρακολούθηση της κίνησης αντικειμένων, και συγκεκριμένα ανθρώπων, στη βιβλιογραφία με στόχο τη δημιουργία ακολουθιών εικόνων με πραγματικά και συνθετικά αντικείμενα (βλέπε για παράδειγμα [12][81]). Όμως, αφορούν κυρίως σε συνθήκες πραγματικού κόσμου και όχι σε εικονικά στούντιο, καθώς και σε διαδικασίες που δεν απαιτούν παραγωγές πραγματικού χρόνου. Στα εικονικά στούντιο, οι τεχνικές βασίζονται πάλι σε αισθητήρες, κυρίως, και λιγότερο σε ορατά ή αόρατα με γυμνό μάτι σημάδια. Επιπρόσθετα, οι κύριες συνεισφορές στο πεδίο αυτό αφορούν μηχανισμούς που ενεργοποιούνται μετά την παραγωγή για χρήση σε κινηματογραφικές ταινίες ή μαγνητοσκοπημένες τηλεοπτικές εκπομπές.

Τέλος, αυτές οι παραγωγές, που βασίζονται στη χρήση εικονικών στούντιο, πρέπει να υπομείνουν τις δυσκολίες που επιβάλλει η έλλειψη μίας ρεαλιστικής προεπισκόπησης της συνολικής σκηνής που θα παραχθεί. Καθώς κινηματογραφούν, οι σκηνοθέτες και οι ηθοποιοί πρέπει συχνά να φαντάζονται τη θέση και την εμφάνιση των εικονικών αντικειμένων που συμμετέχουν στη σκηνή. Αυτό δημιουργεί μία περίπλοκη σχέση μεταξύ της μείωσης του κόστους και των διευκολύνσεων και νέων προοπτικών που παρέχονται με τη χρήση των οπτικών εφέ, των επιβαρυνμένων καθηκόντων που αυτά δημιουργούν στην ομάδα παραγωγής, και της μερικής έλλειψης προεπισκόπησης της σκηνής την ώρα της κινηματογράφησης. Για τη διευκόλυνση της κατάστασης, θα πρέπει οι παραγωγοί να ξεπεράσουν αυτούς τους περιορισμούς, κάτι που υποδεικνύει την ανάγκη για τον εντοπισμό και την παρακολούθηση των αντικειμένων της σκηνής [32].

Στο κεφάλαιο αυτό, θα βασιστούμε στο σύστημα εικονικού στούντιο που προτάθηκε στο Κεφάλαιο 2 και θα παρουσιαστούν τρόποι για την ενσωμάτωση δυνατοτήτων παρακολούθησης αντικειμένων. Με αυτό τον τρόπο μπορούν να παραχθούν μέσα για την ενσωμάτωση εικονικών αντικειμένων και ηθοποιών σε μία σκηνή και να επιτραπεί

να αλληλεπιδράσουν με το φυσικό περιβάλλον της. Το υπόλοιπο του κεφαλαίου είναι οργανωμένο ως εξής. Στην Ενότητα 3.2 παρουσιάζονται οι βασικές μονάδες του συστήματος που εξετάζεται. Η Ενότητα 3.3 πραγματεύεται τον εντοπισμό ενός αντικειμένου βάσει των καρέ που καταγράφονται από δύο κάμερες και παρέχει το πλαίσιο εργασίας για την αντιμετώπιση της περίπτωσης πολλαπλών αντικειμένων στο σκηνικό. Τέλος, η Ενότητα 3.4 συνοψίζει τη συνεισφορά και τις επεκτάσεις της εργασίας αυτής.

3.2 Περιγραφή του συστήματος

3.2.1 Το εικονικό στούντιο

Στην προσέγγιση που παρουσιάστηκε στο Κεφάλαιο 2, το κυανό πέτασμα κατασκευάζεται χρησιμοποιώντας δύο σχετικά κοντινούς τόνους του μπλε. Με αυτόν τον τρόπο δεν δημιουργούνται προβλήματα στη σύνθεση, καθώς οι αποχρώσεις τους μπλε μπορούν να επιλεγούν επαρκώς κοντά στο χρωματικό μοντέλο (color space). Μετά τη διαίρεση του τοίχου, που αντιστοιχεί στο παρασκήνιο, σε $N \times M$ παραλληλόγραμμα, κάθε ένα από αυτά χρωματίζεται χρησιμοποιώντας πιο ανοιχτό ή πιο σκούρο μπλε. Υποθέτοντας ότι ένας τυχαίος υποπίνακας, μεγέθους τουλάχιστον $n \times m$, θα είναι πάντα ορατός από την κάμερα, αρκεί κάθε τέτοιος πίνακας - πρότυπο να εμφανίζεται μόνο μία φορά σε ολόκληρο το κυανό πέτασμα. Ένα τέτοιο πέτασμα παράγεται επεκτείνοντας τη θεμελιώδη θεωρία των ψευδοτυχαίων μονοδιάστατων δυαδικών ακολουθιών μέγιστου μήκους. Έτσι, εφόσον κάθε υποπίνακας του δυαδικού χάρτη του πετάσματος, που καταγράφεται από μία κάμερα, μπορεί να εξαχθεί, και να εντοπιστεί μοναδικά πάνω στο δυαδικό χάρτη, ο εντοπισμός του οπτικού πεδίου της κάμερας γίνεται με ευθύ τρόπο. Επιπλέον, με τον επιτυχή υπολογισμό του ελάχιστου υποπίνακα που είναι ορατός από την κάμερα, περιλαμβάνοντας την περίπτωση κατά την οποία παρουσιάζεται η μέγιστη δυνατή μερική συγκάλυψη του πετάσματος από τα αντικείμενα του προσκηνίου, εξασφαλίζεται ότι το σύνολο των χαρακτηριστικών που υιοθετείται θα περιέχει πάντα ικανή πληροφορία για την εκτίμηση της κίνησης της κάμερας.

Τα χαρακτηριστικά αναφοράς, που χρησιμοποιούνται με στόχο την παρακολούθηση της κίνησης της κάμερας, είναι πρακτικά οι το πλέγμα από ευθείες, που δίνεται έμμεσα από την εξαγωγή του δυαδικού χάρτη του πετάσματος. Ας υποθέσουμε, για το τρέχον καταγραφόμενο καρέ, ότι το ορατό τμήμα του φόντου που περιέχεται σε αυτό, είναι διαθέσιμο μέσω της τεχνικής διαχωρισμού χρώματος. Τότε, τα όρια μεταξύ των τετραπλεύρων εξάγονται μέσω μίας ανίχνευσης ακμών με ένα κοινό τελεστή. Η επακόλουθη εφαρμογή ενός μετασχηματισμού Hough επιτρέπει την ανίχνευση των γραμμών που αποτελούν το μετασχηματισμένο υπο-πλέγμα του πετάσματος, όπως αυτό καταγράφεται από την κάμερα. Εφόσον, στο ορατό τμήμα του φόντου, μπορεί να υπάρχουν συγκεντρώσεις τετραπλεύρων ίδιας απόχρωσης, μερικές ευθείες του πλέγματος μπορεί να μην ανιχνεύονται ακόμη. Σε αυτή την περίπτωση ο υποπίνακας που αντιστοιχεί στο ορατό τμήμα δεν μπορεί να εξαχθεί ακόμη.

Έστω και αν το ορατό τμήμα του δυαδικού χάρτη δεν έχει ακόμη εξαχθεί, η 3Δ περιστροφή που ακολούθησε η κάμερα μπορεί να υπολογιστεί με ευθύ τρόπο. Χρησιμοποιώντας το σύνολο που παρήγαγε ο μετασχηματισμός Hough επιλέγονται οι επικρατούσες 2Δ ευθείες, βάσει του βαθμού εμπιστοσύνης που τους ανατέθηκε κατά την ανίχνευση. Το παραγόμενο σύνολο ευθειών χωρίζεται σε δύο κατηγορίες, βάσει της κλίσης τους, που αντιστοιχούν στις οριζόντιες και κάθετες γραμμές του κυανού πετάσματος. Οι παράμετροι κάθε γραμμής, που υπολογίστηκαν σε μονάδες μέτρησης pixel,

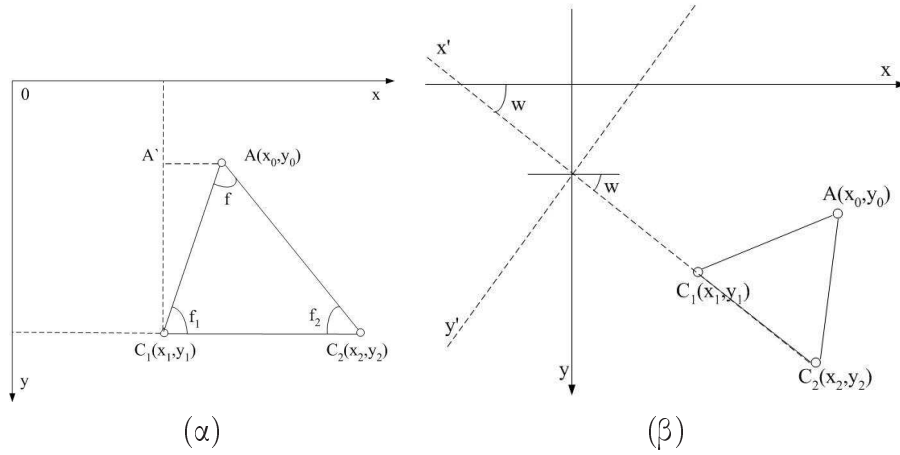
ανάγονται σε συντεταγμένες πραγματικού κόσμου χρησιμοποιώντας τα μήκη των πλευρών του CCD. Εφόσον η πραγματική θέση των γραμμών του πλέγματος στο πέτασμα είναι γνωστή, προκύπτει ένα πρόβλημα υπολογισμού 3Δ κίνησης βάσει 3Δ-σε-2Δ αντιστοιχίσεις χαρακτηριστικών γραμμών αναφοράς. Για αυτό το σκοπό, δημιουργείται μία εικονική σκηνή αναφοράς βάσει (α) του μήκους των πλευρών των τετραπλεύρων και (β) της απόστασης του κυανού πετάσματος από την κάμερα. Η τελευταία μπορεί να επιλεγεί αυθαίρετα για τη σκηνή αναφοράς. Ο πίνακας περιστροφής υπολογίζεται στη συνέχεια ικανοποιητικά, μέσω της ανάλυσης σε ιδιάζουσες τιμές ενός απλού πίνακα (SVD).

Ο υπολογισμός της 3Δ μετατόπισης, αντίθετα με τον υπολογισμό της 3Δ περιστροφής, απαιτεί την ακριβή γνώση των παραμέτρων των ευθειών στη σκηνή αναφοράς και επιπλέον την αντιστοίχιση μεταξύ των 3Δ γραμμών αναφοράς και των 2Δ γραμμών που ανιχνεύονται στο τρέχον καρέ. Η τελευταία διαδικασία επιτυγχάνεται με τον εντοπισμό του ορατού τμήματος του πετάσματος πάνω στο συνολικό. Το πρόβλημα που παραμένει έχει να κάνει με τις γραμμές που απουσιάζουν από το πλέγμα και εμποδίζουν την επιτυχή αναγνώριση του προτύπου. Τα δύο αυτά προβλήματα αντιμετωπίζονται παράλληλα εφόσον, για κάθε δύο ευθείες, η σχετική τους απόσταση στη σκηνή αναφοράς μπορεί να υπολογιστεί χρησιμοποιώντας τις αντίστοιχες τους 2Δ γραμμές. Με αυτό τον τρόπο, οι γραμμές που απουσιάζουν εντοπίζονται, και το ορατό δυαδικό πρότυπο εξάγεται πλήρως. Στη συνέχεια αντιστοιχίζονται οι 3Δ γραμμές αναφοράς με τις 2Δ του τρέχοντος καρέ και τελικά υπολογίζεται το διάνυσμα της 3Δ μετατόπισης βάσει του πίνακα περιστροφής, των αντιστοιχίσεων και του βάθους που έχει οριστεί για τη σκηνή αναφοράς. Μία από τις βασικές συνεισφορές της ανωτέρω διατύπωσης είναι ότι επιτρέπει και τον υπολογισμό της εστιακής απόστασης για κάθε καταγραφόμενο καρέ, καταλήγοντας στον υπολογισμό όλων των αγνώστων παραμέτρων κίνησης της κάμερας.

3.2.2 Εντοπισμός ηθοποιού στο εικονικό στούντιο

Όπως έχει αναφερθεί, τα οπτικά συστήματα παρακολούθησης στα εικονικά στούντιο εκμεταλλεύονται το γεγονός ότι, κάθε καταγραφόμενο καρέ περιέχει ήδη μία μεγάλη ποσότητα πληροφορίας σχετικά με το φόντο, η οποία μπορεί να χρησιμοποιηθεί και για την παρακολούθηση εκτός από την κατάτμηση του φόντου. Υπό αυτή την έννοια, κατάλληλα χαρακτηριστικά που ενσωματώνονται στο κυανό πέτασμα, εντοπίζονται και παρακολουθούνται στο χρόνο λόγω της διακριτότητας της φύσης του πετάσματος. Λόγω αυτού του γεγονότος, είναι εμφανές ότι η εμφάνιση του πετάσματος στο καταγραφόμενο καρέ μπορεί να βοηθήσει στον εντοπισμό των φυσικών αντικειμένων που βρίσκονται στη σκηνή χωρίς τη χρήση επιπλέον ειδικού εξοπλισμού. Συνολικά, τα ακόλουθα ζητήματα πρέπει να εξεταστούν:

- (i) δεν πρέπει να δημιουργηθεί ανάγκη παρακολούθησης επιπρόσθετων χαρακτηριστικών πάνω στα αντικείμενα που κινούνται στο σκηνικό,
- (ii) τα χαρακτηριστικά που θα χρησιμοποιηθούν θα πρέπει να είναι ευδιάκριτα σε κάθε καρέ ώστε να επιτρέπεται ο εντοπισμός των αντικειμένων,
- (iii) το σύνολο των χαρακτηριστικών θα πρέπει να περιέχει τοπικά αρκετή πληροφορία, ακόμα και όταν ένα σημαντικό τμήμα του αντικειμένου αποκρύπτεται από άλλο τμήμα του σκηνικού ή δεν βρίσκεται εντός του οπτικού πεδίου μίας ή περισσότερων καμερών, και



Σχήμα 3.1: Απεικόνιση του προβλήματος εντοπισμού σημείου (α) Οι κάμερες C_1 και C_2 βρίσκονται πάνω στην ευθεία $y = a$ (β) Η γενική περίπτωση κατά την οποία οι κάμερες C_1 , C_2 βρίσκονται σε αυθαίρετες θέσεις

- (iv) η τεχνική παρακολούθησης θα πρέπει να επιτρέπει απότομες αλλαγές στην κίνηση της κάμερας σε όλες τις κατευθύνσεις (όλους τους βαθμούς ελευθερίας).

Όπως θα δειχθεί ακολούθως, το προτεινόμενο σύστημα εκπληρώνει όλες τις ανωτέρω απαιτήσεις. Επιπλέον, θα γίνει εμφανές ότι με την καταγραφή και επεξεργασία του ιστορικού της παρακολούθησης κάθε αντικειμένου, μπορεί να επιτραπεί η πρόβλεψη και ο υπολογισμός εκ των προτέρων της κίνησης και της θέσης κάθε αντικειμένου στη σκηνή.

3.3 Εντοπισμός και παρακολούθηση αντικειμένου

3.3.1 Εύρεση θέσης σημείου

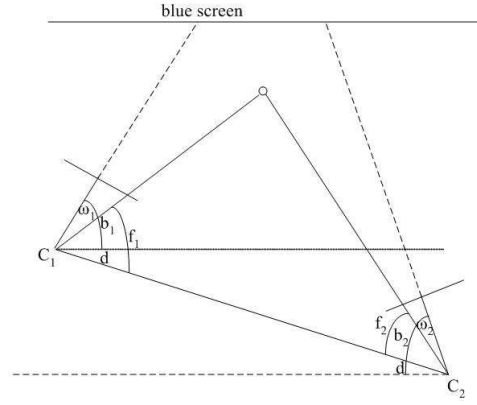
Σε αυτή την υποενότητα θα παρουσιαστεί μία μέθοδος για τον υπολογισμό της θέσης ενός σημείου στον 3Δ χώρο, χρησιμοποιώντας τις εικόνες που καταγράφονται από δύο διαφορετικές κάμερες. Η θέση των καμερών εξάγεται βάσει της αρχιτεκτονικής του εικονικού στούντιο όπως περιγράφηκε στην προηγούμενη ενότητα. Στο Σχήμα 3.1(α) απεικονίζεται η περίπτωση κατά την οποία οι κάμερες έχουν την ίδια συντεταγμένη στον άξονα y του καρτεσιανού συστήματος ενώ στο Σχήμα 3.1(β) φαίνεται η γενική περίπτωση. Το κυανό πέτασμα βρίσκεται τοποθετημένο πάνω στην ευθεία $y = 0$. Η γενική περίπτωση μπορεί να λυθεί εύκολα μετά την επίλυση της πρώτης περίπτωσης και μία απλή περιστροφή του συστήματος συντεταγμένων. Για να λύσουμε την απλή περίπτωση εφαρμόζουμε το νόμο των ημιτόνων για το τρίγωνο C_1C_2A :

$$\frac{C_1C_2}{\sin f} = \frac{AC_1}{\sin f_2} = \frac{AC_2}{\sin f_1} \Leftrightarrow AC_1 = C_1C_2 * \frac{\sin f_2}{\sin f} \Leftrightarrow AC_1 = |x_2 - x_1| * \frac{\sin f_2}{\sin f}$$

Επίσης, για το τρίγωνο $A'AC_1$:

$$\begin{aligned} \sin\left(\frac{\pi}{2} - f_1\right) &= \frac{A'A}{AC_1} \Leftrightarrow A'A = AC_1 * \sin\left(\frac{\pi}{2} - f_1\right) \\ \cos\left(\frac{\pi}{2} - f_1\right) &= \frac{A'C_1}{AC_1} \Leftrightarrow A'C_1 = AC_1 * \cos\left(\frac{\pi}{2} - f_1\right) \end{aligned}$$

Ως αποτέλεσμα, οι συντεταγμένες x_0 , y_0 του σημείου είναι:



Σχήμα 3.2: Υπολογισμός των άγνωστων γωνιών f_1 και f_2 βάσει των γωνιών λήψης των καμερών b_1 και b_2 και της γωνίας d που σχηματίζεται από τον άξονα x και την ευθεία C_1C_2 .

$$\begin{aligned} y_0 &= y_1 + A'C_1 = y_1 - \cos\left(\frac{\pi}{2} - f_1\right) * |x_2 - x_1| * \frac{\sin f_2}{\sin f} \\ x_0 &= x_1 + A'A = x_1 + \sin\left(\frac{\pi}{2} - f_1\right) * |x_2 - x_1| * \frac{\sin f_2}{\sin f} \end{aligned}$$

Περιστρέφοντας το σύστημα συντεταγμένων, στο επίπεδο xy , κατά w μοίρες, μπορούμε να εφαρμόσουμε τους προηγούμενους υπολογισμούς για τη γενική περίπτωση. Η γωνία w υπολογίζεται ως εξής:

$$w = \pi - \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$$

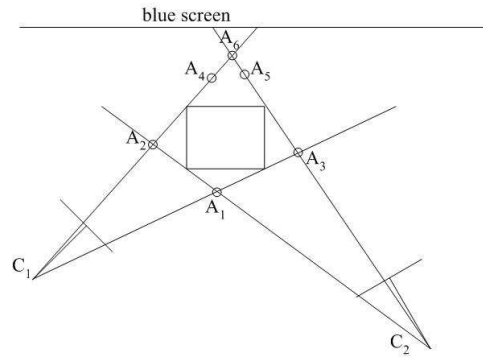
Οι συντεταγμένες του σημείου στο νέο σύστημα υπολογίζονται με τη χρήση των πινάκων μετατόπισης και περιστροφής T και R .

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \left(\begin{bmatrix} x \\ y \end{bmatrix} + T \right) * R \Leftrightarrow \begin{bmatrix} x' \\ y' \end{bmatrix} = \left(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -x_k \\ -y_k \end{bmatrix} \right) * \begin{bmatrix} \cos w & \sin w \\ -\sin w & \cos w \end{bmatrix}$$

Τέλος, οι γωνίες f , f_1 και f_2 υπολογίζονται βάσει της γεωμετρίας του Σχήματος 3.2. Έτσι προκύπτει ότι, $f_1 = b_1 - \omega_1 + d$ και $f_2 = b_2 - \omega_2 - d$, όπου οι b_1 και b_2 είναι η γωνίες λήψης των καμερών C_1 και C_2 αντίστοιχα, και η γωνία d σχηματίζεται από τις ευθείες $x'x$ και C_1C_2 . Οι γωνίες ω_1 και ω_2 εξάγονται ως εξής:

$$\omega_1 = \arctan\left(\frac{x_1 - x_0}{f.c.}\right), \omega_2 = \arctan\left(\frac{x_0 - x_2}{f.c.}\right),$$

όπου τα x_1 , x_2 , x_0 είναι οι κανονικοποιημένες ως προς τις πραγματικές διαστάσεις του CCD οριζόντιες αποστάσεις μεταξύ του pixel που αντιστοιχεί στο σημείο και των άκρων και του κέντρου της εικόνας, και το $f.c.$ αντιστοιχεί στην εστιακή απόσταση της κάμερας. Πρέπει να σημειωθεί στο σημείο αυτό ότι, ανάλογα με τις θέσεις του σημείου και των καμερών, ο συνδυασμός των γωνιών που χρησιμοποιούνται για τον υπολογισμό των f_1 και f_2 διαφέρει ως προς τα πρόσημα (συγκεκριμένα υπάρχουν τέσσερις περιπτώσεις) αλλά καθορίζεται εύκολα εκ των προτέρων. Τέλος η ίδια διαδικασία μπορεί να ακολουθηθεί για τον εντοπισμό του σημείου στο επίπεδο YZ ώστε να προκύψει η θέση του σημείου στον 3Δ χώρο.



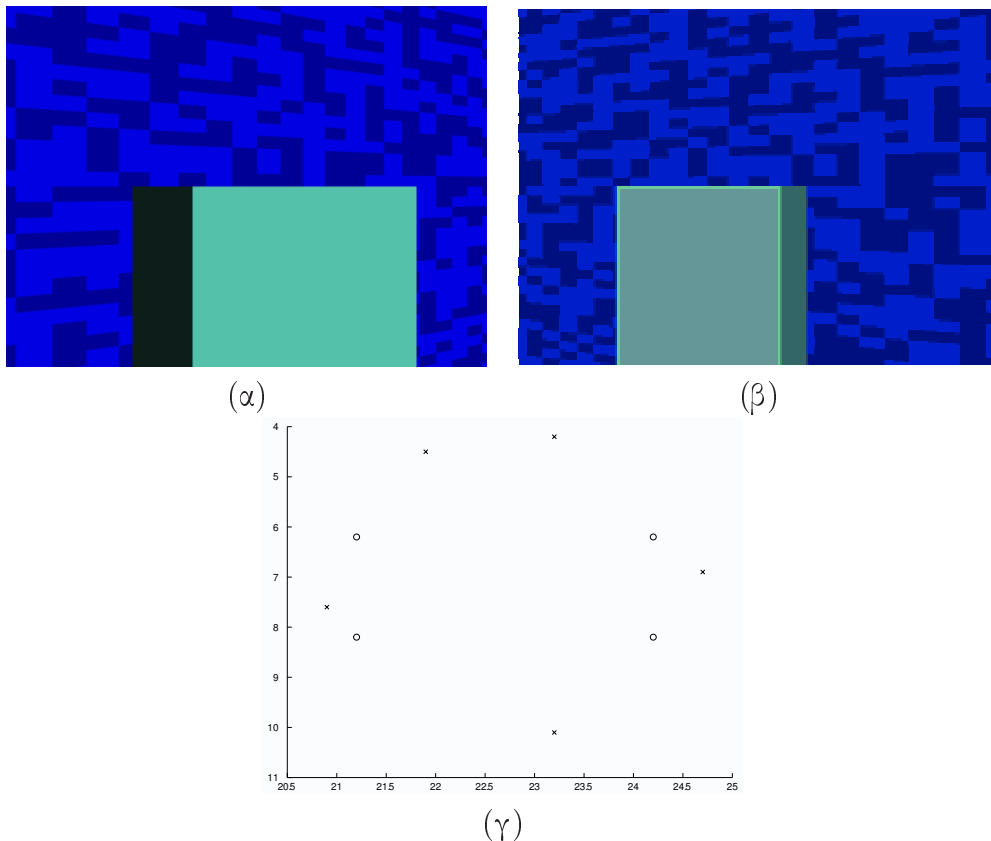
Σχήμα 3.3: Εκτίμηση του χώρου που καταλαμβάνει ένα αντικείμενο υπολογίζοντας τη θέση των σημείων A_1-6

3.3.2 Εκτίμηση του χώρου που καταλαμβάνει ένα αντικείμενο

Για την εκτίμηση της θέσης ενός σημείου στον 3D χώρο, με τη χρήση δύο καμερών, απαιτείται η αντιστοίχιση (correspondence) των προβολών του σημείου πάνω στα δύο καταγραφόμενα καρέ. Ανάλογα με την εκάστοτε εφαρμογή, διάφορες μεθοδολογίες έχουν προταθεί για να λύσουν το πρόβλημα, όπως η χρήση διακριτικών για τα ζητούμενα σημεία και η παρακολούθηση του σημείου στον χρόνο μέσω τεχνικών εκτίμησης κίνησης. Στο συγκεκριμένο πλαίσιο αυτής της εργασίας απαιτείται ένας συμψηφισμός μεταξύ της ανάγκης για ακριβή εκτίμηση της θέσης των αντικειμένων και της ανάγκης λειτουργίας του συστήματος σε πραγματικό χρόνο. Η μεθοδολογία που παρουσιάζεται παρέχει μία εύρωστη εκτίμηση σχετικά με την περιοχή που καταλαμβάνει ένα αντικείμενο πάνω στο σκηνικό, χωρίς να την ανάγκη επίπονων τεχνικών για την εύρεση αντιστοιχίσεων. Με αυτό τον τρόπο, η διαδικασία παραμένει απλή και γρήγορη, ενώ παρέχει μία ικανή εκτίμηση της περιοχής.

Στο Σχήμα 3.3 απεικονίζεται, προβεβλημένο στο επίπεδο xy , το σενάριο που περιγράφεται, με τις δύο κάμερες C_1 και C_2 , το αντικείμενο και τις ευθείες που σχηματίζονται συνδέοντας κάθε κάμερα με τα συνοριακά σημεία του αντικειμένου. Οι συντεταγμένες των σημείων A_1 , A_2 και A_3 υπολογίζονται όπως περιγράφηκε στην προηγούμενη υποενότητα, ενώ τα σημεία A_4 και A_5 επιλέγονται ώστε να ικανοποιούν τις εξισώσεις $A_2A_4 = A_1A_3$ και $A_3A_5 = A_1A_2$. Τα τελευταία δύο σημεία μπορούν να επιλεγούν υπό συνθήκες αντί του A_6 εφόσον, όπως δείχνουν και πειραματικά αποτελέσματα, βελτιώνουν την εκτίμηση της περιοχής, χωρίς να επηρεάζουν το αποτέλεσμα όταν αφορούν τα όρια της μη ορατής από τις κάμερες περιοχής που καλύπτει το αντικείμενο.

Στο Σχήμα 3.4 παρουσιάζονται κάποια πειραματικά αποτελέσματα για μία απλή περίπτωση· τα καρέ που καταγράφονται από τις δύο κάμερες απεικονίζονται στα 3.4(α) και 3.4(β) ενώ η πραγματική περιοχή που καταλαμβάνει το αντικείμενο στο επίπεδο XY (συμβολίζεται με o) και η εκτίμησή της (συμβολίζεται με x) φαίνονται στο 3.4(γ). Με τον τρόπο αυτό, εκπληρώνονται οι δύο πρώτες απαιτήσεις που ορίστηκαν στην υποενότητα 3.2.2, καθώς τα μόνα σημεία του αντικειμένου που ανιχνεύονται είναι τα συνοριακά του σημεία.



Σχήμα 3.4: Ένα απλό σενάριο (α) Το καρέ της αριστερής κάμερας (β) Το καρέ της δεξιάς κάμερας (γ) Η πραγματική περιοχή που καλύπτει το αντικείμενο (ο) και η εκτίμηση της (x)

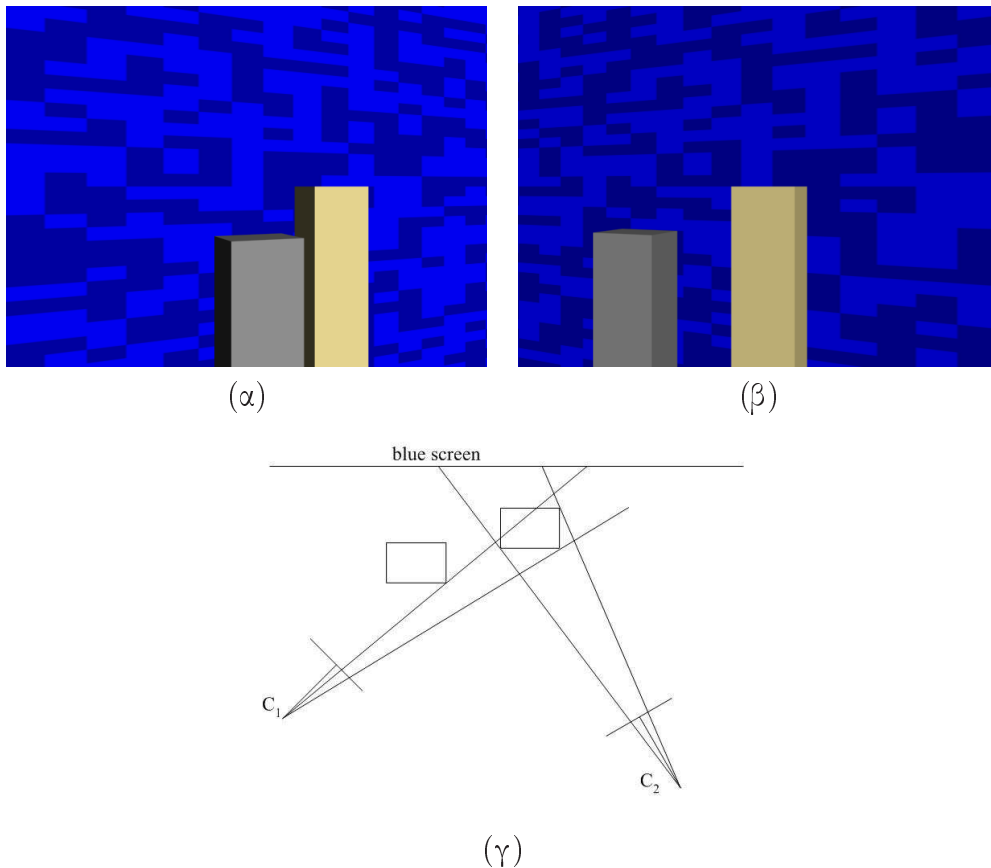
3.3.3 Παρακολούθηση πολλαπλών αντικειμένων

3.3.3.1 Περίγραμμα αντικειμένου

Η παραγωγή του περιγράμματος του αντικειμένου σε κάθε καρέ παρέχει ένα αποτελεσματικό εργαλείο για την παρακολούθηση του αντικειμένου καθώς και για την εκτίμηση της κίνησής του. Μία απλή μέθοδος υιοθετείται για την εξαγωγή του περιγράμματος ενός αντικειμένου από το καταγραφόμενο καρέ και για την διατεταγμένη αποθήκευση των συντεταγμένων των pixel που το αποτελούν. Η μάσκα του αντικειμένου, η οποία λαμβάνεται απλά μέσω της μονάδας διαχωρισμού χρώματος, χρησιμοποιείται μαζί με το αποτέλεσμα της ανίχνευσης ακμών για να καθορίσουν τα pixel που αποτελούν το περίγραμμα. Ξεκινώντας από το πιο αριστερά pixel και κινούμενοι πάντα με προτεραιότητα στα αριστερά, είναι δυνατόν να ακολουθηθεί η διαδρομή του περιγράμματος από pixel σε pixel και να αποθηκευτούν ακολουθιακά οι συντεταγμένες τους. Στην περίπτωση που το αντικείμενο ξεπερνάει τα όρια του καρέ δημιουργείται ένα νέο κλειστό αντικείμενο, με μία απλή προεργασία, χρησιμοποιώντας τα ορατά μέρος του αντικειμένου και τα όρια του καρέ. Ο σχεδιασμός και η υλοποίηση του αλγορίθμου είναι απλές διαδικασίες που και η παρουσίασή τους δεν κρίνεται σκόπιμη στα πλαίσια της παρούσας διατριβής.

3.3.3.2 Επι κάλυψη

Στη γενική περίπτωση, υπάρχουν στο σκηνικό περισσότερα του ενός αντικείμενα και κινούνται αυθαίρετα πάνω σε αυτό. Αναμένεται φυσικά ότι, σε διάφορες περιπτώσεις

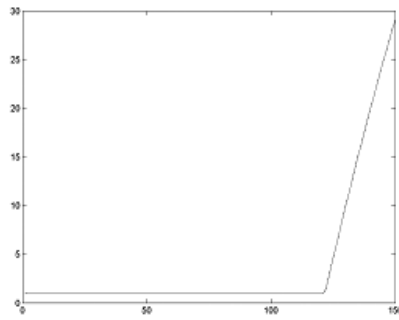


Σχήμα 3.5: Περίπτωση επικαλυπτόμενων αντικειμένων (α) Το καρέ της κάμερας στα αριστερά (β) Το καρέ της κάμερας στα δεξιά (γ) Κάτοψη (C_1, C_2 οι δύο κάμερες)

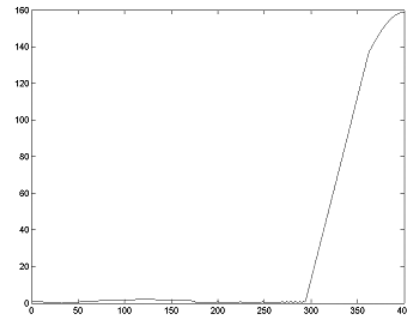
που ένα αντικείμενο κινείται, θα καλύπτει μερικά ή ολικά κάποιο άλλο ή αντίστροφα. Σε αυτή την περίπτωση πρέπει να αναπτυχθεί ένας μηχανισμός ώστε να συνεχιστεί να παρέχεται μία εκτίμηση για την περιοχή που καλύπτουν τα αντικείμενα. Πρέπει να σημειωθεί ότι, ενώ η μία κάμερα μπορεί να καταγράφει ένα αντικείμενο να επικαλύπτει ένα άλλο, είναι δυνατόν η άλλη κάμερα να καταγράφει ικανό χώρο μεταξύ των αντικειμένων ώστε αυτά να ξεχωρίζουν.

Αυτό μπορεί να παρατηρηθεί στο Σχήμα 3.5 όπου, το καρέ της κάμερας στα δεξιά μπορεί να επεξεργασθεί όπως περιγράψαμε στις προηγούμενες υποενότητες. Για το καρέ της πρώτης κάμερας, παρουσιάζεται μία τεχνική προεπεξεργασίας, που στοχεύει να διαχωρίσει τα δύο αντικείμενα ώστε να εξάγει τα περιγράμματά τους και να εκτιμήσει την περιοχή που καλύπτουν. Το πρώτο βήμα της μεθόδου είναι η αναγνώριση, κατά προσέγγιση, των ορίων της περιοχής της τομής των δύο αντικειμένων. Για το σκοπό αυτό χρησιμοποιούνται τα περιγράμματα των αντικειμένων που είχαν καταγραφεί σε προηγούμενα καρέ, όταν δεν υπήρχε επικάλυψη. Η σύγκριση μεταξύ αυτών των περιγραμμάτων και του περιγράμματος του ενιαίου αντικειμένου που έχει σχηματιστεί μπορεί να παρέχει ικανή εκτίμηση των ορίων της τομής.

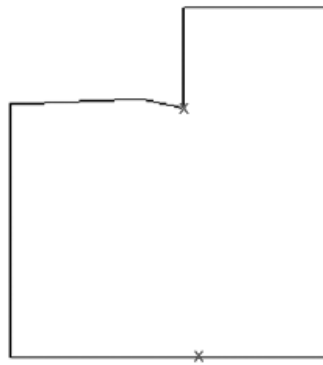
Το αρχικό σημείο για τη σύγκριση των περιγραμμάτων μπορεί να επιλεγεί εύκολα ώστε να ανήκει στο αντικείμενο που βρίσκεται στο προσκήνιο, καθώς αυτό είναι γνωστό από την κίνησή τους και τις εκτιμήσεις που έχουν πραγματοποιηθεί έως το σημείο αυτό. Το σημείο του περιγράμματος του ενιαίου αντικειμένου του τρέχοντος καρέ, που απέχει τη μικρότερη απόσταση από το επιλεγμένο σημείο, θα αποτελέσει το αρχικό σημείο για το τρέχον καρέ. Δημιουργείται, με τον τρόπο αυτό, μία προσεγ-



(α)



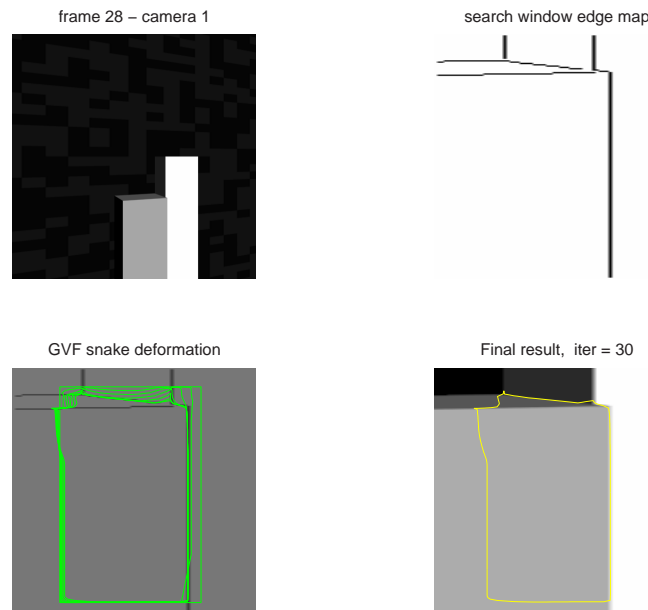
(β)



(γ)

Σχήμα 3.6: Εντοπισμός των οριακών σημείων της τομής των αντικειμένων βάσει των αποστάσεων μεταξύ του τρέχοντος και προηγούμενων περιγραμμάτων (α) Απόσταση των περιγραμμάτων, αριστερόστροφα (β) Απόσταση των περιγραμμάτων, δεξιόστροφα (γ) Περίγραμμα των συνδεδεμένων αντικειμένων (με x σημειώνονται τα οριακά σημεία της τομής όπως υπολογίστηκαν)

γιστική αντιστοίχιση μεταξύ των περιγραμμάτων. Παρακολουθώντας την απόσταση μεταξύ των αντίστοιχων σημείων των περιγραμμάτων του ενιαίου αντικειμένου και του αντικειμένου που βρισκόταν στο προσκήνιο σε προηγούμενο καρέ, μπορεί να εντοπιστούν οι περιοχές που περιλαμβάνουν τα όρια της τομής μεταξύ των αντικειμένων. Όπως φαίνεται και στο Σχήμα 3.6, η χρήση της απόστασης αυτής αποτελεί μία αποδοτική μέθοδο που παρέχει ακριβή εκτίμηση για την απλή περίπτωση δυό αντικειμένων σε σχήμα κιβωτίων. Σε αυτό το σημείο διάφορες μέθοδοι μπορούν να χρησιμοποιηθούν για να εκτιμήσουν την περιοχή που τα αντικείμενα καταλαμβάνουν ή ακόμα να ολοκληρώσουν τον διαχωρισμό και να υπολογίσουν τα νέα περιγράμματα. Αρχικά, χρησιμοποιήθηκαν πρόχειρες εκτιμήσεις των περιγραμμάτων, βάσει των ορίων της τομής, ώστε να εντοπιστούν τα ακραία σημεία των αντικειμένων και να ακολουθηθεί η προτεινόμενη μεθοδολογία για την περιοχή που καταλαμβάνουν. Με τον τρόπο αυτό, εκπληρώνονται και οι δύο τελευταίες απαιτήσεις της υποενότητας 3.2.2. Η εκτίμηση της θέσης του αντικειμένου περιέχει την απαραίτητη πληροφορία ακόμα και στην περίπτωση επικάλυψης ή απόκρυψης, ενώ η μεθοδολογία υποστηρίζει την επανέναρξη του υπολογισμού των θέσεων των αντικειμένων ακόμα και μετά από απότομες αλλαγές στο οπτικό πεδίο της κάμερας.



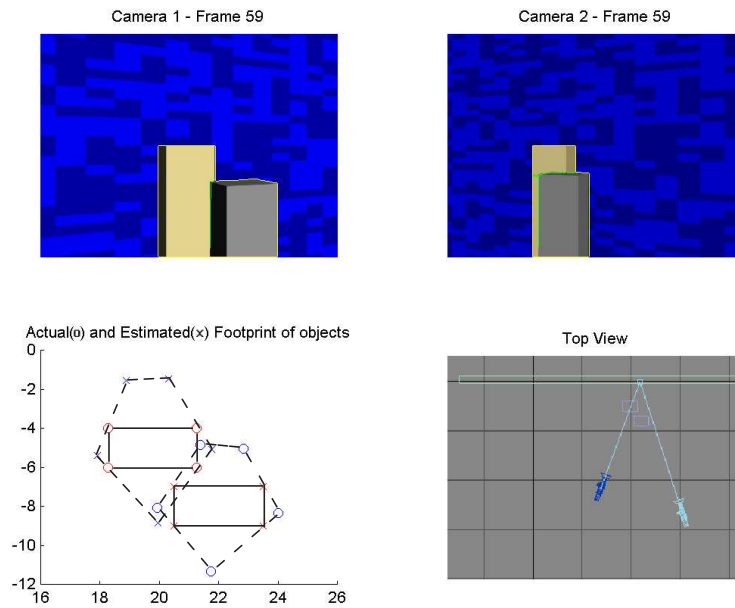
Σχήμα 3.7: Αρχικοποίηση και παραμόρφωση του *snake*

3.3.3.3 Προσέγγιση τομής με χρήση ενεργών περιγραμμάτων

Με στόχο να ολοκληρωθεί η προτεινόμενη μεθοδολογία, πρέπει να χρησιμοποιηθεί μία τεχνική για την ακριβέστερη δυνατή εξαγωγή των περιγραμμάτων των δύο αντικειμένων στην περίπτωση επικάλυψης που περιγράφηκε στην προηγούμενη υποενότητα. Η γνώση αυτή απαιτείται για το διαχωρισμό των αντικειμένων στα επόμενα καρέ. Τα ενεργά περιγράμματα επιλέχθηκαν βάσει του γεγονότος ότι έχουν ήδη εξαχθεί τα δύο συνοριακά σημεία της τομής, καθώς και για να αντιπαραβάλλουμε τα αποτελέσματα με αυτά μίας απλής ανίχνευσης ακμών. Τα προαναφερθέντα συνοριακά σημεία θα χρησιμοποιούν για την αρχικοποίηση της διαδικασίας, παρέχοντας το αρχικό παράθυρο για το *snake*. Το αποτέλεσμα της παραμόρφωσης του τελευταίου παρέχει το τμήμα που λείπει από το περίγραμμα των δύο αντικειμένων, όπως φαίνεται και στο σχήμα 3.7. Πρέπει να επισημανθεί ότι το τελικό περίγραμμά του αντικειμένου που καλύπτεται θα είναι μεν μερικό λόγω της επικάλυψης, αλλά η εκτίμηση της περιοχής που καλύπτει στο χώρο θα είναι επαρκής μιας και θα αφορά στο τμήμα του χώρου που είναι ορατό από τις κάμερες. Επιπρόσθετα, η εκτίμηση αυτή μπορεί εύκολα να βελτιωθεί μέσω της πρότερης γνώσης για το αντικείμενο και την περιοχή που καλύπτει. Στο Σχήμα 3.8 παρουσιάζονται αποτελέσματα για την ίδια ακολουθία, όπου και οι δύο κάμερες καταγράφουν τα αντικείμενα να επικαλύπτονται. Η εκτίμηση της τομής, που προσεγγίζεται με ενεργά περιγράμματα, αποτυπώνεται και στα δύο καρέ. Οι τελικά υπολογισμένες περιοχές που καλύπτουν τα αντικείμενα στο έδαφος παραμένουν αποδοτικές, παρά το γεγονός ότι και οι δύο κάμερες δεν έχουν πλήρη όψη των δύο αντικειμένων.

3.4 Συμπεράσματα και επεκτάσεις

Στο παρόν κεφάλαιο, παρουσιάστηκε ένα προηγμένο οπτικό εικονικό στούντιο που περιλαμβάνει ένα σύστημα εκτίμησης της κίνησης της κάμερας βάσει ενός κυανού πετάσματος δύο χρωματικών τόνων. Περιγράφηκε η μεθοδολογία για την παρακολούθηση των ηθοποιών και των αντικειμένων που βρίσκονται και κινούνται στο σκηνικό, επιτρέ-



Σχήμα 3.8: Η περίπτωση δύο επικαλυπτόμενων αντικειμένων

ποντας την αποτελεσματική προσθήκη επιπλέον, εικονικών, αντικειμένων στις σκηνές. Τα αποτελέσματα των πειραμάτων προσομοίωσης απεικονίζουν την αποδοτικότητα του ολοκληρωμένου συστήματος.

Στις μελλοντικές επιδιώξεις βρίσκεται η ολοκλήρωση ενός πραγματικού συστήματος που θα συνδέει τις υπάρχουσες μονάδες διαχωρισμού χρώματος και σύνθεσης με τις προτεινόμενες τεχνικές. Με αυτό τον τρόπο, η προτεινόμενη προσέγγιση θα δοκιμαστεί σε ένα πραγματικό εμπορικό στούντιο. Επιπρόσθετα η ολοκλήρωση θα επιτρέψει τη δημιουργία ενός πλήρως αυτοματοποιημένου συστήματος επιτρέποντας και το σχεδιασμό μίας μονάδας ελέγχου και αντιμετώπισης λαθών. Από θεωρητικής πλευράς, έχει ενδιαφέρον να εξεταστεί κατά πόσον η χρήση των εκτιμήσεων για τη θέση και την κίνηση των αντικειμένων μπορεί να επιτρέψει την επαλήθευση των εξαχθέντων παραμέτρων 3Δ κίνησης της κάμερας.

□

Κεφάλαιο 4

Εξαγωγή χαρακτηριστικών προσώπου

4.1 Εισαγωγή - Αναδρομή

Η κατανόηση του οπτικού περιεχομένου του ανθρώπινου προσώπου αποτελεί ιδιαίτερα ενεργό πεδίο έρευνας για πάνω από τρεις δεκαετίες. Οι ερευνητές έχουν προσπαθήσει πολύ, κατά τα χρόνια αυτά, να κωδικοποιήσουν το περιεχόμενο του προσώπου μέσω ποικίλων προσεγγίσεων (όπως το πρότυπο MPEG-4) και να χρησιμοποιήσουν την κωδικοποίηση αυτή για ταυτοποίηση ή αναγνώρισης συναισθημάτων [10, 13]. Οι εφαρμογές του ανήκουν σε μία ευρεία γκάμα, περιλαμβάνοντας τις τηλεπικοινωνίες αλλά και την επικοινωνία ανθρώπου-μηχανής. Τις περισσότερες φορές, το μείζον ζήτημα σε όλες τις προσεγγίσεις ήταν η κατάλληλη επιλογή των χαρακτηριστικών, έχοντας ως στόχο είτε την αναπαράσταση του προσώπου ή την αποτίμηση της συναισθηματικής του κατάστασης.

Το συγκεκριμένο πρόβλημα της αυτόματης επιλογής των χαρακτηριστικών του προσώπου αποτελεί μέρος του γενικότερου προβλήματος της αυτόματης κατάτμησης εικόνων. Η κατάτμηση μέσω αποκλειστικής χρήσης παραδοσιακών, χαμηλού επιπέδου (low-level), τεχνικών επεξεργασίας εικόνας συχνά απαιτεί αλληλεπίδραση σε μεγάλο βαθμό, ώστε να επιτευχθούν τα επιθυμητά αποτελέσματα. Αντίθετα, η αυτοματοποίηση συχνά εμπλέκει μοντελοποίηση ώστε να καθοδηγήσει την διαδικασία κατάτμησης. Η επιλογή του σωστού συνόλου μοντέλων για την όσο το δυνατόν πιο γενικευμένη αναπαράσταση των αντικειμένων είναι δύσκολη λόγω της σχηματικής πολυπλοκότητας και της ποικιλίας που παρατηρείται σε ένα σύνολο αντικειμένων που ανήκουν στην ίδια κατηγορία. Έχει αναφερθεί ότι, για αμετάβλητα, καλά καθορισμένα αντικείμενα (και τα χαρακτηριστικά του προσώπου μπορούν να θεωρηθούν ως τέτοια), είναι δυνατή η κατάτμηση της εικόνας σε μία καθολική, συνεπή ερμηνεία μέσω της χρήσης παραμορφώσιμων προτύπων (deformable templates), ενώ προηγουμένως χρησιμοποιούνται στατιστικά μοντέλα σχημάτων για την ενίσχυση των πιθανοτήτων των καθολικών παραμορφώσεων εντός της ίδιας κατηγορίας αντικειμένων [70]. Άλλες εργασίες στην κατάτμηση εικόνων περιλαμβάνουν στοχαστικές μοντελοπαγείς (model-based) προσεγγίσεις [18, 47, 61, 89, 98], μορφολογική ανάπτυξη περιοχής (region growing) [72], διάχυση ενέργειας (energy diffusion) [51], και διαχωρισμό γράφων (graph partitioning) [73]. Η κατάτμηση που δε βασίζεται στην μοντελοποίηση στοχεύει στο διαχωρισμό των περιοχών που παρουσιάζουν ομογένεια στο χρώμα ή/και στην υφή [19], αλλά γενικά δεν ικανοποιούν τις ανάγκες της σημασιολογικής κατάτμησης.

Η εύρωστη και ακριβής ανάλυση του προσώπου και η εξαγωγή των χαρακτηριστικών του παραμένει ένα σύνθετο πρόβλημα που έχει αντιμετωπιστεί με την εισαγωγή

υποθέσεων ή περιορισμών όσον αφορά την περιστροφή και τον προσανατολισμό του προσώπου, το ορατό του μέρος, τις συνθήκες φωτισμού και την κλίμακα της απεικόνισης. Αυτοί οι περιορισμοί αίρονται τελικά στη βιβλιογραφία [63, 5], καθώς οι ερευνητές αντιμετωπίζουν όλο και πιο ρεαλιστικά περιβάλλοντα, χωρίς να παραβλέπουν και τις πρωτοποριακές εργασίες της περιοχής [34, 69, 8, 40, 10].

Οι προσεγγίσεις που βασίζονται σε ακολουθίες εικόνων επιχειρούν να συνδυάσουν την χωρική με την χρονική πληροφορία ώστε να αποτιμήσουν και να βελτιώσουν την επίδοση της κατάτμησης [14, 26, 5]. Η χρήση στερεοσκοπικών μεθόδων και ειδικών επεξεργαστών [52] υπόσχεται γρήγορη παρακολούθηση των χαρακτηριστικών, δεδομένου ότι η αρχικές τους θέσεις είναι προκαθορισμένες. Η ιεραρχική εξαγωγή χαρακτηριστικών χρησιμοποιεί ένα συνδυασμό στατιστικών στοιχείων και συμπληρωματικά συγκεντρώνει αναπαραστάσεις πολλαπλών συνόλων Gabot για μία ολόκληρη ακολουθία [71, 54, 64]. Απλούστερες προσεγγίσεις στην περίπτωση ακολουθιών βίντεο χρησιμοποιούν ευριστικές μεθόδους με σκοπό την πρακτική υλοποίηση [49]. Η παράλληλη επεξεργασία χρησιμοποιώντας ενδείξεις για την κίνηση και το χώρο υλοποιεί μία συνεργασία για να αποδειχτεί ότι μπορεί η γρήγορη κατάτμηση είναι εφικτή με τη χρήση μόνο τυποποιημένων και συμπληρωματικών τεχνικών. Σε μία παλαιότερη δουλειά των Black et al [5], τοπικά παραμετρικά μοντέλα κίνησης εικόνων χρησιμοποιήθηκαν για την ανάκτηση και την αναγνώριση μη στερεής και αρθρωτής κίνησης ανθρώπινων προσώπων. Χρησιμοποιήθηκαν παραμετρικά μοντέλα ροής με σκοπό να παρέχουν μια περιεκτική περιγραφή της κίνησης του προσώπου (που, διαισθητικά, συνδέεται με την κίνηση των χαρακτηριστικών του προσώπου) χρησιμοποιώντας έναν μικρό αριθμό παραμέτρων. Η εργασία αυτή ασχολήθηκε με την κατανόηση των εκφράσεων του προσώπου, ωστόσο η εξαγωγή και η παρακολούθηση των χαρακτηριστικών ήταν η βάση για την επίτευξη οποιασδήποτε εκτίμησης για την έκφραση.

Η μοντελοπαγής εκτίμηση των χαρακτηριστικών του προσώπου χρησιμοποιεί ταίριασμα, γενικευμένων ή εξαρτώμενων από το υποκείμενο, μοντέλων υψής, αλλά απαιτεί υπολογιστικά δαπανηρές ενέργειες, όπως την οπτικοποίηση των μοντέλων (rendering) με τροποποιημένο σχήμα και θέση [25]. Έχει επιχειρηθεί επίσης και διερεύνηση για τη μείωση της υπολογιστικής πολυπλοκότητας, θεωρώντας την εξαγωγή χαρακτηριστικών προσώπου που βασίζεται σε μοντέλα ως ένα πρόβλημα βελτιστοποίησης [1].

Οι προσεγγίσεις υψηλότερου επιπέδου επιχειρούν να εκμεταλλευθούν ένα μεγάλο πλήθος στοιχείων, που σχετίζονται και ενδυναμώνονται από την γνώση των μοντέλων μέσω ενός πιθανοτικού πλαισίου [96]. Σε αυτό το πλαίσιο, οι ομαδοποιήσεις των χαρακτηριστικών που σχηματίζουν λογικές οντότητες που οργανώνονται με βάση την αντίληψη, εξετάζονται υπό την ανάθεση πιθανοτήτων για κάθε ομαδοποίηση και την ενίσχυσή τους μέσω τεχνικών λογικής Bayesian reasoning.

Η προσέγγιση που προτείνεται στο παρόν κεφάλαιο φιλοδοξεί να συμβάλει προς την κατεύθυνση της αντιμετώπισης περισσότερο ρεαλιστικών περιβαλλόντων με την θεώρηση γενικευμένων συνθηκών για την εικόνα, και στοχεύει στην ανάπτυξη περισσότερο δομημένων λύσεων. Παρουσιάζεται μια σταδιακή, εύρωστη ανάλυση των χαρακτηριστικών του προσώπου, που αντιμετωπίζει μεγάλες διαφοροποιήσεις στην εμφάνιση διαφόρων ατόμων, καθώς και στην περίπτωση του ίδιου προσώπου σε διαφορετικές στιγμές μέσα σε πραγματικές ακολουθίες βίντεο. Γίνονται χαλαρές υποθέσεις για την στάση του προσώπου ή τη θεωρητική θέση των χαρακτηριστικών πάνω σε αυτό. Η σταδιακή αποκάλυψη της πληροφορίας που σχετικά με το πρόσωπο υποστηρίζεται με διαδικασίες βελτιστοποίησης, ελαχιστοποίησης ενέργειας ή σφάλματος για κάθε στάδιο, παράγοντας εκ των υστέρων γνώση για το πρόσωπο και οδηγώντας σε μια

βήμα προς βήμα απεικόνιση των ζητούμενων χαρακτηριστικών. Κάτι τέτοιο έρχεται σε αντίθεση με το βασικό σκεπτικό άλλων λύσεων που προτείνονται στη βιβλιογραφία [25, 80, 66, 1, 56, 75], οι οποίες χρησιμοποιούν συγκεκριμένα μοντέλα αναπαράστασης χαρακτηριστικών, ή υποθέτουν ότι το πρόσωπο βρίσκεται σε κατακόρυφη θέση. Ωστόσο, συγχωνεύει νέα εργαλεία με κάποια που έχουν ήδη προταθεί στη βιβλιογραφία, κάτω από μια ενοποιημένη ιεραρχική προσέγγιση, κάθε στάδιο της οποίας αντιμετωπίζεται ως ένα ερώτημα βελτιστοποίησης.

Η προσέγγιση που παρουσιάζεται στο κεφάλαιο αυτό εστιάζει στη δομημένη εξαγωγή των χαρακτηριστικών του προσώπου σε ήδη καταταμημένες εικόνες προσώπου, ενώ η συμβολή της μπορεί να συνοψιστεί στα ακόλουθα:

- (i) Το πρόβλημα αντιμετωπίζεται κάτω από γενικευμένες συνθήκες απεικόνισης και υπό χαλαρές προϋποθέσεις. Αυτό το πρόβλημα έχει αντιμετωπιστεί στο παρελθόν για ακολουθίες βίντεο, για τις οποίες ωστόσο το πρώτο πλαίσιο (η αρχικοποίηση) θεωρήθηκε ως ένα λυμένο πρόβλημα.
- (ii) Η ανάγκη για εξαγωγή χαρακτηριστικών που δεν εξαρτάται από πιθανή περιστροφή του προσώπου έχει ως αποτέλεσμα τη δημιουργία ενός μηχανισμού ασαφούς λογικής για την αυτόματη ανίχνευση της επίπεδης γωνίας στρέψης του προσώπου.
- (iii) Ο στόχος της προτεινόμενης μεθοδολογίας είναι να αντιμετωπίζεται κάθε βήμα της διαδικασίας ως ένα ξεχωριστό πρόβλημα βελτιστοποίησης.
- (iv) Η ολοκλήρωση ενδιαφερουσών τεχνικών που προτείνονται στην βιβλιογραφία με νέες τεχνικές παρέχει περισσότερο ισχυρά εργαλεία για την επίτευξη του στόχου μας.

Στην Ενότητα 4.2 που ακολουθεί παρουσιάζονται και αναλύονται μερικές από τις μεθόδους που έχουν υιοθετηθεί. Συγκεκριμένα, παρουσιάζεται η αποτελεσματικότητά τους και συζητούνται τα μειονεκτήματά τους υπό προϋποθέσεις. Η προτεινόμενη μέθοδος κατάτμησης χαρακτηριστικών του προσώπου, που σχηματίζεται ως ένα τμηματικό πρόβλημα βελτιστοποίησης, εξετάζεται λεπτομερώς στην Ενότητα 4.3. Στη συνέχεια, στην Ενότητα 4.4 παρουσιάζεται μια σειρά από αποτελέσματα προσομοιώσεων πάνω σε πραγματικές ακολουθίες βίντεο που περιέχουν πρόσωπα. Τελικά, στην Ενότητα 4.5 γίνεται ανασκόπηση των συμπερασμάτων.

4.2 Μερικά από τα χρησιμοποιούμενα εργαλεία

4.2.1 Η κατάτμηση χαρακτηριστικών του προσώπου μέσω ανάλυσης ελάχιστου - μέγιστου

Οι προβολές ολοκληρώματος (integral projections) έχουν παίξει έναν σημαντικό ρόλο στην μακρά βιβλιογραφία της αναγνώρισης προσώπων και της εξαγωγής χαρακτηριστικών, ως εργαλείο εκτίμησης της θέσης των χαρακτηριστικών του προσώπου. Αυτή η τεχνική προτάθηκε αρχικά από τον Kanade στην πρωτοποριακή του εργασία [41] πάνω στην αναγνώριση ανθρώπινων προσώπων. Οι προβολές είναι εύκολα υλοποιήσιμες, ενώ ταυτόχρονα είναι ιδιαίτερα αποτελεσματικές στον καθορισμό της θέσης των χαρακτηριστικών, δεδομένου ότι το παράθυρο στο οποίο ενεργούν είναι κατάλληλα τοποθετημένο ώστε να αποφεύγονται παραπλανητικές παρεμβολές. Ο Kanade εφάρμοσε

την ανάλυση των προβολών σε δυαδικές εικόνες που προκύπτουν από την εφαρμογή ενός τελεστή Λαπλασε σε εικόνες υπό γκρι κλίμακα και με την υιοθέτηση στη συνέχεια ενός κατάλληλου κατωφλίου. Έκτοτε, πολλοί ερευνητές χρησιμοποίησαν το ίδιο εργαλείο κατά καιρούς. Οι Brunelli και Poggio [8] πραγματοποίησαν ανάλυση προβολής ακμών διαχωρίζοντας το χάρτη ακμών βάσει των κατευθύνσεων των ακμών, οριζοντίων και καθέτων. Οι περιοχές που υποδεικνύονταν από τα ολοκληρώματα συνιστούσαν μια καλή αρχική εκτίμηση για τη θέση στην οποία έπρεπε να εφαρμοσθούν τα πρότυπα (templates). Άλλοι ερευνητές ανακάλυψαν ότι μπορούν να αποκτηθούν ακριβέστερα αποτελέσματα με την εφαρμογή της ανάλυσης προβολής στην εικόνα εντάσεως (intensity image), λόγω της ομαλότητας των περιγραμμάτων αρκετών από τα χαρακτηριστικά των προσώπου [66, 42, 75, 86, 59]. Στην εργασία [66] ελαστικά μοντέλα φιδιών (rubber snake models) εφαρμόζονται στις περιοχές που υποδεικνύονται από τις προβολές ολοκληρωμάτων. Αυτό πραγματοποιείται με ακολουθιακό τρόπο ανάλογα με την παραμόρφωση του προτύπου σε διαφορετικές εποχές που χρησιμοποιούνται από τους Yuille et al [97]. Στην εργασία [75] οι συγγραφείς υπολογίζουν τις προβολές ως προς τον άξονα y του gray-level τοπογραφικού χάρτη. Ομαλοποιούν την προβολή αυτή με φιλτράρισμα μέσου (average filtering). Τα σημαντικά ελάχιστα καθορίζονται ελέγχοντας τις κλίσεις καθενός από τα ελάχιστα σε σχέση με τα γειτονικά τους μέγιστα. Κάθε σημαντικό ελάχιστο θεωρείται ως πιθανή κατακόρυφη θέση για κάποιο από τα χαρακτηριστικά του προσώπου. Για κάθε υποψήφιο κατά y, υπολογίζεται η ομαλοποιημένη τοπική προβολή κατά x, και τα ελάχιστα που προκύπτουν από αυτήν τη διαδικασία είναι υποψήφια χαρακτηριστικά σημεία.

Στις ανωτέρω αναφερόμενες εργασίες, οι ερευνητές εκμεταλλεύονται την εκ των προτέρων γνώση ότι τα πρόσωπα εντός των εικόνων βρίσκονται σε κατακόρυφη θέση, προσδιορίζοντας έτσι τα σημαντικά ελάχιστα των προβολών ως υποψήφιες εκτιμήσεις συγκεκριμένων χαρακτηριστικών ή ομάδας χαρακτηριστικών, συγκεκριμένα της ανώτερης ομάδας (φρύδια, μάτια), της μέσης (ρουθούνια) και της κατώτερης ομάδας (στόμα, πηγούνι). Σε περισσότερο ρεαλιστικά περιβάλλοντα, η υπόθεση για την θέση του προσώπου δεν ισχύει κατά κανόνα, γεγονός που αποκλείει και κάθε υπόθεση που γίνεται για το που βρίσκονται οι προαναφερθείσες ομάδες χαρακτηριστικών μέσα στην εικόνα. Έστω και έτσι, όλες οι σχετικές εργασίες δείχνουν πόσο ισχυρό είναι ένα απλό στη σύλληψή εργαλείο όπως αυτό.

4.2.2 Μη επιβλεπόμενη κατάτμηση χρώματος - υφής

Το πρόβλημα της μη επιβλεπόμενης κατάτμησης (unsupervised segmentation) είναι ασθενώς ορισμένο, καθώς τα σημασιολογικά αντικείμενα δεν αντιστοιχούν συνήθως σε ομοιογενείς σε χρώμα ή υφή περιοχές. Εργασίες στην κατάτμηση εικόνων περιλαμβάνουν στοχαστικές προσεγγίσεις βασισμένες σε μοντέλα [47, 89], μορφολογική ανάπτυξη περιοχών [72], μεθόδους εξάπλωσης ενέργειας [51], και διαχωρισμού γράφων [73].

Μια πρόσφατη εργασία που παρουσιάστηκε από τους Deng et al [19], στοχεύει στην κατάτμηση ομοιογενών περιοχών χρώματος - υφής. Σύμφωνα με αυτήν, τα χρώματα της εικόνας κβαντίζονται σε αντιπροσωπευτικές κλάσεις, παράγοντας έτσι ένα χάρτη κλάσεων, ο οποίος μπορεί να ειπωθεί ως ένα είδος σύνθεσης υφής. Η χωρική κατάτμηση πραγματοποιείται ευθέως σε αυτόν το χάρτη χωρίς να ληφθεί υπόψη όποια ομοιότητα μεταξύ των αντίστοιχων pixel.

Έστω Z το σύνολο όλων των N σημείων σε έναν χάρτη κλάσεων και m ο μέσος:

$$m = \frac{1}{N} \cdot \sum_{z \in Z} z \quad (4.1)$$

Ας υποθέσουμε ότι το Z κατατάσσεται σε C κλάσεις, $Z_i, i = 1, \dots, C$. Έστω m_i ο μέσος των N_i σημείων δεδομένων της κλάσης Z_i :

$$m_i = \frac{1}{N_i} \cdot \sum_{z \in Z_i} z \quad (4.2)$$

Έστω επίσης

$$S_T = \sum_{z \in Z} \|z - m\|^2 \quad (4.3)$$

και

$$S_W = \sum_{i=1}^C \sum_{z \in Z_i} \|z - m_i\|^2 \quad (4.4)$$

όπου S_T είναι η συνολική μεταβλητότητα και S_W είναι η μεταβλητότητα των σημείων που ανήκουν στην ίδια κλάση. Ας ορίσουμε επίσης:

$$J = (S_T - S_W)/S_W \quad (4.5)$$

Για μια εικόνα με κάποιες ομοιογενείς περιοχές, οι χρωματικές κλάσεις διαχωρίζονται περισσότερο η μία από την άλλη, και το J παίρνει μεγάλες τιμές. Εάν όλες οι κλάσεις διανέμονται ομοιόμορφα σε όλη την εικόνα, τότε η ποσότητα J παίρνει μικρές τιμές. Εάν υπολογίσουμε το J για κάθε κατατμημένη περιοχή, μπορούμε να ορίσουμε:

$$\bar{J} = \frac{1}{N} \cdot \sum_k M_k J_k \quad (4.6)$$

όπου το J_k υπολογίζεται για τη γειτονιά k , το M_k είναι ο αριθμός των σημείων στην γειτονιά k και N είναι ο συνολικός αριθμός των σημείων στο χάρτη κλάσεων. Σύμφωνα με τους Deng et al [19], το \bar{J} είναι μια έκφραση μίας συνάρτησης ενέργειας, η οποία έχει αποδειχθεί ότι είναι ένα καλό κριτήριο, το οποίο πρέπει να ελαχιστοποιηθεί με όλους τους πιθανούς τρόπους κατάτμησης της εικόνας, δεδομένου του αριθμού των περιοχών. Κατασκευάζοντας μια εικόνα της οποίας οι τιμές των pixel αντιστοιχούν σε τιμές J που υπολογίζονται σε μικρά παράθυρα με κέντρο τα pixel αυτά, οικοδομούνται οι λεγόμενες J-εικόνες. Όσο υψηλότερη είναι η τιμή του J , τόσο πιθανότερο είναι ότι το αντίστοιχο pixel βρίσκεται κοντά στο όριο μιας περιοχής.

Αυτό το νέο κριτήριο για κατάτμηση χρώματος - υφής είναι αρκετά πολύπλοκο ώστε να προϋποθέτει επεξεργασία σε μη πραγματικό χρόνο. Ωστόσο, εάν εφαρμοσθεί σε καλύτερα ορισμένα σενάρια κατάτμησης, οι υπολογιστικές απαιτήσεις της διαδικασίας αυτής μπορούν να μειωθούν σημαντικά.

4.2.3 Κατάτμηση και ενεργά περιγράμματα

Όπως και στο γενικό πεδίο της κατάτμησης έτσι και στην περίπτωση της εξαγωγής χαρακτηριστικών του προσώπου, έχουν χρησιμοποιηθεί συχνά παραμορφώσιμα μοντέλα (deformable models) και ενεργά περιγράμματα (active contours, snakes) [11, 17, 39, 45, 53, 77, 5, 80, 66]. Για παράδειγμα, οι Yuille et al [97] εφαρμόζουν παραμορφώσιμα πρότυπα για τη μοντελοποίηση χαρακτηριστικών του προσώπου. Αυτή η προσέγγιση επιτρέπει την ενσωμάτωση της γνώσης για συγκεκριμένα αντικείμενα στο μοντέλο. Ωστόσο, τέτοιες μέθοδοι απαιτούν την προσεκτική οικοδόμηση και παραμετροποίηση των προτύπων.

Τα φίδια (snakes), όπως αλλιώς ονομάζονται τα ενεργά περιγράμματα, ενσωματώνουν γνώση για την ομαλότητα ενός περιγράμματος καθώς και για την αντίσταση σε παραμορφώσεις. Μια συνήθης εκτίμηση ενός περιγράμματος επιτυγχάνεται με τον καθορισμό των δυνάμεων της εικόνας που έλκουν το μοντέλο φιδιού, ενώ την ίδια στιγμή, ενδογενείς δυνάμεις μπορούν να χρησιμοποιηθούν για τη συρρίκνωση ή τη διόγκωση του φιδιού προς κατευθύνσεις που δεν σχετίζονται με το περιεχόμενο της εικόνας. Ένα από τα κύρια πλεονεκτήματα των αναπαραστάσεων μέσω ενεργών περιγραμμάτων είναι η δυνατότητα ορισμού ποικίλων ιδιοτήτων τους μέσω της συνάρτησης ενέργειας τους, σε αναλογία με φυσικά συστήματα. Ο έλεγχος ενός φιδιού προκαλεί την ανάπτυξή του προς κατευθύνσεις όπου η ενέργειά του μειώνεται. Με τον καθορισμό μιας κατάλληλης συνάρτησης ενέργειας, μπορεί να δημιουργηθεί ένα ενεργό περίγραμμα που θα εξελιχθεί ώστε να συμπεριλαμβάνει συγκεκριμένες ιδιότητες.

Μία από τις πιο κοινές μορφές της συνάρτησης ενέργειας ενός ενεργού περιγράμματος είναι η ακόλουθη:

$$\begin{aligned}
 E_{snake} &= E_{int} + E_{ext} \\
 &= E_{elastic} + E_{image} \\
 &= K_1 \cdot \sum_{i=1}^N (d(i, i-1))^2 + K_2 \cdot \sum_{i=1}^N I(x_i, y_i)
 \end{aligned} \tag{4.7}$$

όπου με $d(i, i-1)$ συμβολίζεται η απόσταση μεταξύ δύο διαδοχικών σημείων ελέγχου (control points) και με $I(x_i, y_i)$ συμβολίζεται η ένταση στο pixel i . Οι σταθερές K_1 και K_2 επιλέγονται αυθαίρετα και ελέγχουν την επιρροή που έχει ο καθένας από τους δύο παράγοντες του αθροίσματος.

Η εσωτερική ή ελαστική ενέργεια είναι το κομμάτι που βασίζεται στις ενδογενείς ιδιότητες του φιδιού, όπως το μήκος του ή η καμπυλότητά του. Η εξωτερική ενέργεια ή ενέργεια της εικόνας εξαρτάται γενικά από τη δομή της εικόνας και από συγκεκριμένους περιορισμούς που επιθυμεί να θέσει ο χρήστης. Κάθε μία από τις ενέργειες αντιστοιχεί σε μια δύναμη, γεγονός που επιτρέπει την απλή υλοποίηση της δυναμικής του φιδιού. Σε κάθε βήμα της διαδικασίας, κάθε σημείο ελέγχου του φιδιού κινείται κατά ένα μέγεθος ανάλογο της δύναμης που ασκείται σε αυτό. Η εφαρμογή των δυνάμεων ώστε να κινηθούν τα σημεία ελέγχου είναι ένα είδος τεχνικής βελτιστοποίησης και υπάγεται στις μεθοδολογίες κατάβασης κλίσης (gradient descent).

Η επιτάχυνση της εξέλιξης των φιδιών μπορεί να επιτευχθεί μέσω τεχνικών δυναμικού προγραμματισμού, υπολογίζοντας άμεσα τη διάταξη του φιδιού που θα προκαλέσει ισορροπία μεταξύ των εσωτερικών δυνάμεων και ενός δεδομένου συνόλου εξωτερικών δυνάμεων. Κάτι τέτοιο επιτρέπει μεγαλύτερα βήματα στη διαδικασία, και είναι συνολικά πιο αποδοτικό, παρά το πρόσθετο υπολογιστικό κόστος που εισαγάγει.

Στην αρχική εργασία για τα ενεργά περιγράμματα [45], τα φίδια θεωρούνται ως βοήθημα για έναν ανθρώπινο χειριστή που επιχειρεί να μετρήσει τη δομή των εικόνων. Ο χειριστής θα κατηύθυνε το φίδι, για παράδειγμα, σε κάποια συγκεκριμένα κύτταρα σε μια ιστολογική εικόνα, ή σε έναν δρόμο μέσα σε μια εικόνα από δορυφόρο, και το φίδι θα προσαρμοζόταν στο αντικείμενο αυτό και θα παρείχε ένα ακριβές μέτρο του σχήματός του. Στην περίπτωση μας, όπως θα περιγραφεί ακολούθως, τα ενεργά περιγράμματα θα πρέπει να αρχικοποιούνται αυτόματα ώστε να προσαρμοστούν τελικά σε αντικείμενα που θα αποτελέσουν υποψήφια χαρακτηριστικά του προσώπου.

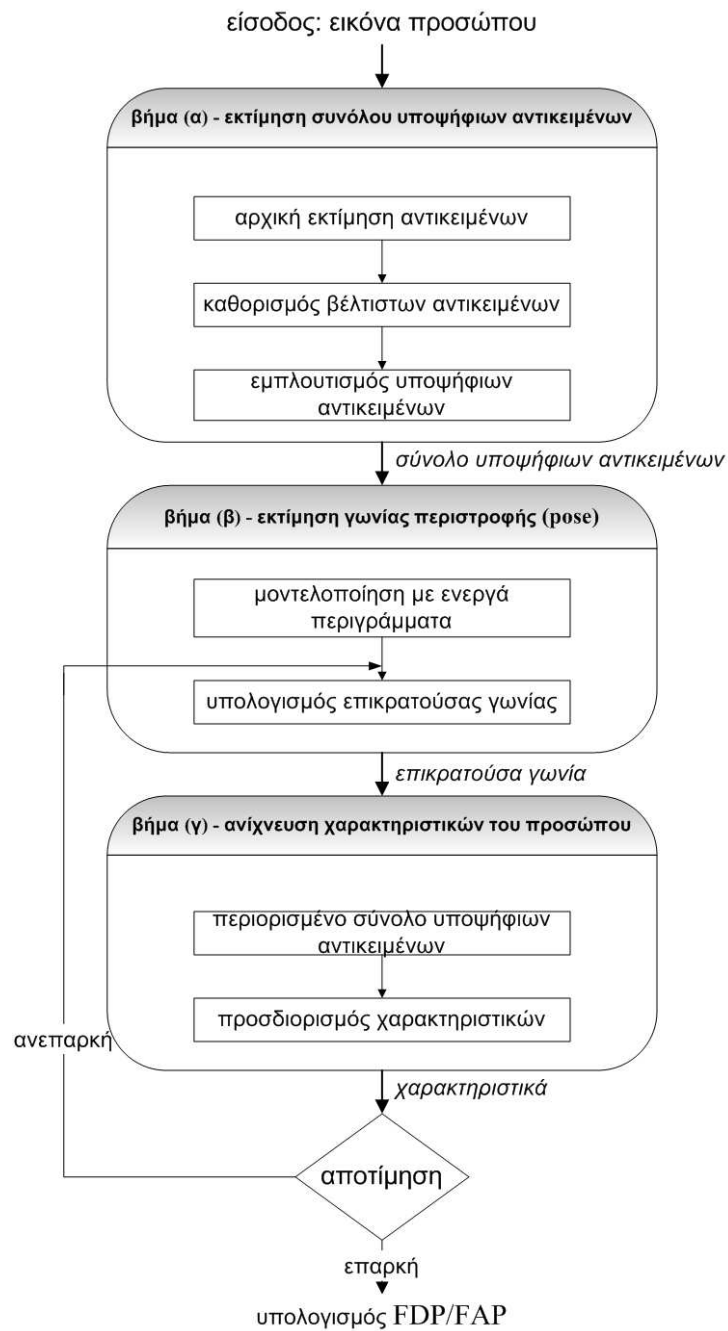
4.3 Η προτεινόμενη μέθοδος

4.3.1 Επισκόπηση του συστήματος

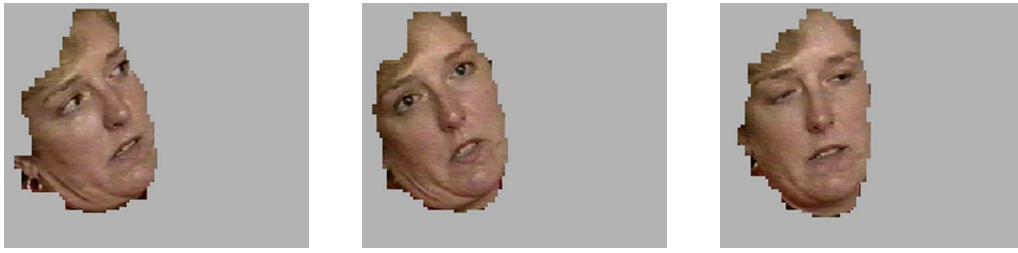
Σύμφωνα με την προτεινόμενη προσέγγιση, τα κυρίαρχα χαρακτηριστικά του προσώπου, όπως είναι τα μάτια και το στόμα, αντιμετωπίζονται ως μείζονες ασυνέχειες σε ένα κατατμημένο, αυθαίρετα περιστραμμένο πρόσωπο. Ο προσδιορισμός των χαρακτηριστικών αυτών απαιτεί μία προσεκτική επιλογή ανάμεσα σε όλες τις πιθανές ασυνέχεις, λόγω της αυθαίρετης γωνίας κλίσης του προσώπου, καθώς και πιθανής απόκρυψης τμημάτων του προσώπου. Η επιλογή βασίζεται σε κριτήρια που γενούν εκ των υστέρων γνώση για το πρόσωπο, όπως για παράδειγμα η επικρατούσα γωνία κλίσης των υποψήφιων χαρακτηριστικών και η θεώρηση των συμμετριών του.

Το Σχήμα 4.1 παρουσιάζει την προτεινόμενη προσέγγιση. Η ροή των βημάτων επεξεργασίας αποτελείται από τρία βασικά στάδια: (α) βελτιστοποιημένη κατάτμηση και εξαγωγή των κύριων χαρακτηριστικών του προσώπου, (β) εκτίμηση της επικρατούσας γωνίας κλίσης του προσώπου βάσει των εξαχθέντων χαρακτηριστικών, και (γ) προσδιορισμός των εξαχθέντων χαρακτηριστικών. Το πρώτο στάδιο της διαδικασίας παρέχει αρχικές εκτιμήσεις για τα χαρακτηριστικά του προσώπου, οι οποίες εισάγονται σε ένα επαναληπτικό σύστημα βελτιστοποίησης που εξάγει τις βέλτιστες εκτιμήσεις. Το στάδιο αυτό συμπεριλαμβάνει και μία διαδικασία περαιτέρω επεξεργασίας (που εκφράζεται επίσης ως ένα πρόβλημα ελαχιστοποίησης ενέργειας), η οποία βοηθάει στον σχηματισμό κλειστών αντικειμένων που αποτελούν τις εκτιμήσεις καθώς και στην απομάκρυνση μικρών τμημάτων θορύβου (artefacts). Το επόμενο στάδιο (β) αντιμετωπίζει ένα διαφορετικό ζήτημα, συγκεκριμένα τον καθορισμό της επικρατούσας γωνίας κλίσης του προσώπου, για περιστροφή στο επίπεδο της εικόνας (pose). Χρησιμοποιώντας ως είσοδο τις εκτιμήσεις του προηγούμενου βήματος (α), και μοντελοποιώντας τις μέσω ενεργών περιγραμμάτων υπολογίζει την επικρατούσα γωνία κλίσης με τη χρήση ενός συστήματος ασαφούς λογικής. Τέλος, το στάδιο (γ) που ακολουθεί εκμεταλλεύεται τα αποτελέσματα των προηγούμενων και, χρησιμοποιώντας τη γνώση για τη θέση των ματιών και του στόματος πάνω στο ανθρώπινο πρόσωπο, προσδιορίζει και επαληθεύει τα χαρακτηριστικά του προσώπου ανάμεσα στις εκτιμήσεις που έχουν εξαχθεί. Η επαλήθευση βασίζεται στη γνώση σχετικά με τις συμμετρίες του προσώπου και τη σχετική θέση μεταξύ των χαρακτηριστικών του, δίνοντας τη δυνατότητα να επαναληφθούν τα δύο προηγούμενα βήματα ώστε να η επιλογή της επικρατούσας γωνίας και ο ονοματισμός των χαρακτηριστικών να γίνουν με μεγαλύτερη ακρίβεια. Τελικά, μέσω της διαδικασίας που σκιαγραφήθηκε, μπορούν να εξαχθούν οι παράμετροι προσδιορισμού (FDP) και δυναμικής κίνησης (FAP) των χαρακτηριστικών σύμφωνα με το πρότυπο MPEG-4.

Ακολούθως περιγράφονται ένα προς ένα τα βήματα της προσέγγισής μας.



Σχήμα 4.1: Η ροή της μεθόδου



Σχήμα 4.2: Τρία αντιπροσωπευτικά καρέ εισόδου που απεικονίζουν πρόσωπα από μία φυσική ακολουθία βίντεο. Η μεθοδολογία που παρουσιάζεται αντιμετωπίζει εικόνες με ποικιλία στην τοποθέτηση του προσώπου (περιστροφή και προσανατολισμός), στις συνθήκες φωτισμού και στην κλίμακα απεικόνισης

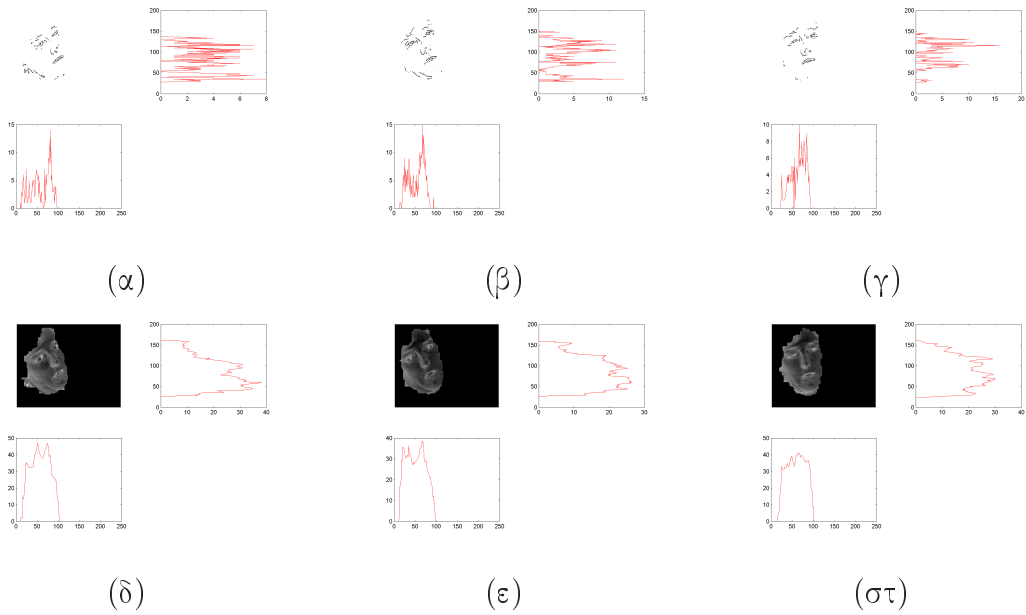
4.3.2 Κατάτμηση και καθορισμός των αρχικών εκτιμήσεων

Η είσοδος σε αυτή την υπομονάδα του συστήματος είναι μια εικόνα X μεγέθους $k \times l$ που περιέχει ένα πρόσωπο που έχει εξαχθεί από το φόντο, δηλαδή μια μάσκα εικόνας προσώπου. Αυτή η εικόνα θεωρείται ότι περιέχει μόνο πληροφορία που βρίσκεται εντός του κλειστού ορίου του προσώπου, το οποίο καθορίζεται από την περίμετρο του δέρματος του προσώπου που είναι ορατό. Στο Σχήμα 4.2 δίνονται τρία παραδείγματα εισόδου από μία φυσική ακολουθία βίντεο.

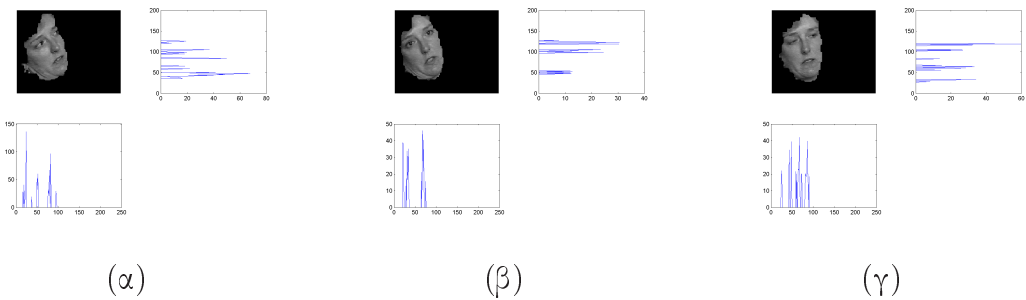
Η αρχική επεξεργασία, η οποία βασίζεται σε καθαρή στατιστική ανάλυση, λειτουργεί υπό την υπόθεση ότι όλα τα χαρακτηριστικά του προσώπου στην εικόνα εισόδου μπορούν να ειπωθούν ως ασυνέχειες πάνω σε μια, γενικά, ομαλή επιφάνεια. Φυσικά, θα πρέπει να αντιμετωπιστούν και διάφορα άλλα είδη ανωμαλιών στην επιφάνεια αυτή, όπως αυτές που δημιουργούνται από τις ανομοιότητες του φωτισμού που εξαρτώνται από την κατεύθυνση της πηγής φωτός και τη θέση της σε σχέση με το αντικείμενο, από περιπτώσεις μερικής απόκρυψης του προσώπου, όπου τμήματα άλλων αντικειμένων κρύβουν μέρος του αντικειμένου που εξετάζεται, καθώς και διαφοροποιήσεις ως προς τις γωνίες στρέψης του προσώπου, που γενικά μεταβάλλουν το σχήμα και την κατανομή των χαρακτηριστικών πάνω στο πρόσωπο. Η απόκρυψη αποτελεί σίγουρα ένα σύνθετο ζήτημα, για την αντιμετώπιση του οποίου απαιτείται γνώση του περιεχομένου. Σε τέτοιες περιπτώσεις χρησιμοποιείται ένας μηχανισμός ελέγχου συμμετρίας, ο οποίος παρουσιάζεται στη συνέχεια και χρησιμοποιεί τη γνώση που προκύπτει από τα ήδη εντοπισμένα χαρακτηριστικά.

Η στατιστική ανάλυση εικόνων του προσώπου, όπως περιγράφεται ακολούθως, παρουσιάζει ένα σημαντικό πλεονέκτημα συγκρινόμενη με άλλες προσεγγίσεις παρακολούθησης χαρακτηριστικών. Δεν απαιτεί τη γνώση της τοποθέτησης των χαρακτηριστικών του προσώπου εξαρχής, γεγονός που την αποσυνδέει από το πρόβλημα των διαφοροποιήσεων στις γωνίες στρέψης του προσώπου. Αυτό σημαίνει ότι πρώτα ανιχνεύονται τα σημαντικά χαρακτηριστικά του προσώπου, χωρίς το σύστημα να μπορεί να καθορίσει σε αυτό το στάδιο οτιδήποτε για την ταυτότητά τους.

Σε αυτό το πλαίσιο, χρησιμοποιήθηκε ο συνδυασμός θεμελιωδών εργαλείων επεξεργασίας εικόνας για τον καθορισμό των περιοχών του προσώπου που πιθανόν να περιέχουν τις περιοχές ενδιαφέροντος. Στην εικόνα της μάσκας του προσώπου πραγματοποιείται ανίχνευση ακμών και υπολογίζεται στη συνέχεια η κατακόρυφη και η οριζόντια προβολή ολοκληρώματος. Οι πρώτες τρεις εικόνες του Σχήματος 4.3 απεικονίζουν τα αποτελέσματα για το πρόσωπο της χρησιμοποιούμενης φυσικής ακολουθίας βίντεο.



Σχήμα 4.3: Πάνω σειρά: Κάθετες και οριζόντιες προβολές ολοκληρώματος για τις εικόνες ακμών των τριών εικόνων εισόδου. Κάτω σειρά: Κάθετες και οριζόντιες προβολές ολοκληρώματος για τις *inverse grayscale* εικόνες των τριών παραδειγμάτων εισόδου

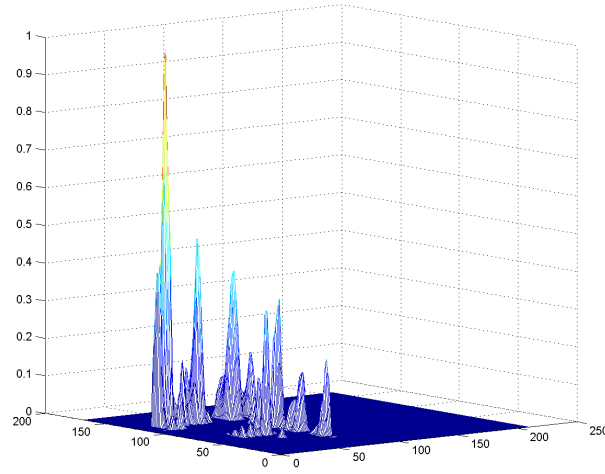


Σχήμα 4.4: Συνδυασμός των προβολών ολοκληρώματος για τα τρία παραδείγματα

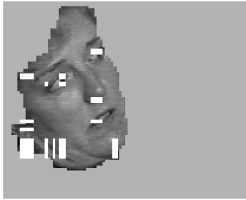
Υπό το ίδιο σκεπτικό, η πληροφορία στην ανάστροφης κλίμακα του γκρι (*inverse grayscale*) της εισόδου ολοκληρώνεται τόσο κατακόρυφα, όσο και οριζόντια, όπως φαίνεται στην κάτω σειρά του Σχήματος 4.3. Η αναστροφή θεωρήθηκε χρήσιμη τόσο για λόγους σύγκρισης, όσο και για λόγους συνδυασμού με τα αντίστοιχα γραφήματα της πάνω σειράς του Σχήματος 4.3. Οι προβολές και στις δύο περιπτώσεις υποδεικνύουν - με μια αυστηρή και με μία ευρύτερη έννοια, αντίστοιχα - τα μονοδιάστατα παράθυρα όπου μπορεί κανείς να εντοπίσει τις περιοχές ενδιαφέροντος.

Ένα τμήμα της πληροφορίας που περιέχει η εικόνα ακμών είναι άσχετο με τα χαρακτηριστικά του προσώπου και σχετίζεται με θόρυβο είτε με σκιάσεις. Έτσι, οι γραφικές παραστάσεις των εικόνων (α), (β) και (γ) του Σχήματος 4.3 φιλτράρονται, χρησιμοποιώντας ως περιέχουσες καμπύλες τις αντίστοιχες γραφικές παραστάσεις των (δ), (ε) και (στ) του Σχήματος 4.3. Με τον τρόπο αυτό τα παράθυρα μέσα στα οποία θα αναζητηθούν τα χαρακτηριστικά του προσώπου εκτιμούνται ακριβέστερα ενώ αποκλείεται μέρος του θορύβου που προκύπτει από την ανίχνευση ακμών. Τα αποτελέσματα παρουσιάζονται στο Σχήμα 4.4.

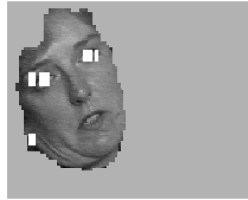
Τα παράθυρα \underline{P}_h και \underline{P}_v που προκύπτουν τελικά, για την οριζόντια και την κατακόρυφη προβολή αντίστοιχα, παρέχουν μονοδιάστατες ενδείξεις για τη θέση των



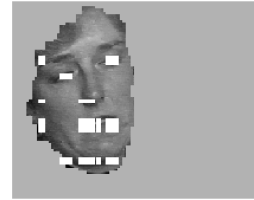
Σχήμα 4.5: Το γινόμενο των οριζόντιων και κάθετων διανυσμάτων παράγει ένα χάρτη M , όπου εμφανίζονται κοιλάδες και λόφοι που σχηματίζουν τα διάφορα αντικείμενα



(α)



(β)



(γ)

Σχήμα 4.6: Οι αρχικές εκτιμήσεις M_0 των αντικειμένων για τα τρία αντιπροσωπευτικά παραδείγματα

περιοχών ενδιαφέροντος μέσα στην εικόνα. Το επόμενο βήμα επιχειρεί την επέκταση των παραθύρων στις δύο διαστάσεις, καθώς και την κατωφλίωση (thresholding) και την δυαδικοποίηση (binarization) των τιμών του αποτελέσματος. Για να επιτευχθεί αυτή η επέκταση, υπολογίζεται το γινόμενο του οριζοντίου και του κατακόρυφου παραθύρου, παράγοντας ένα χάρτη M , όπως αυτός που φαίνεται στο Σχήμα 4.5.

$$M = \underline{P}_v^T \cdot \underline{P}_h \quad (4.8)$$

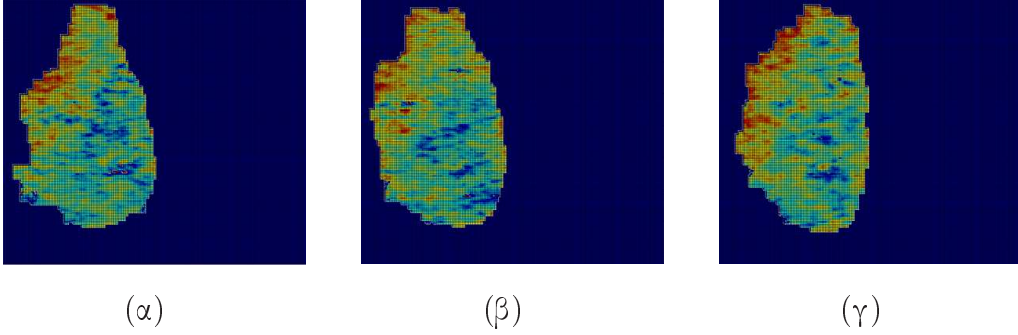
Πάνω σε αυτόν το χάρτη διακρίνονται συγκεντρώσεις pixel, ανάλογα με τις M -τιμές τους, σχηματίζοντας διαφορετικά αντικείμενα. Ακολουθώντας, για κάθε αντικείμενο ορίζουμε την αντίστοιχη αρχική εκτίμησή του ως το κλειστό υποσύνολο των στοιχείων του με μέγιστη περιοχή και ελάχιστη μεταβλητότητα χρωματικού τόνου (hue variance). Μία πρώτη εκτίμηση \underline{M}_0 , είναι ο δυαδικός χάρτης στο επίπεδο $z = 0$, όπως φαίνεται στο Σχήμα 4.6. Το παραπάνω εκφράζεται μαθηματικά ακολούθως. Ας θεωρήσουμε ότι, για λόγους απλοποίησης του συμβολισμού, η αρχική εκτίμηση περιλαμβάνει μόνο ένα αντικείμενο.

Έστω \underline{S} το διάνυσμα που περιέχει την πιθανότητα κάθε pixel:

$$x_i \in \underline{X}, i \in [1, \dots, k \times l] : m_0^i = 1 \quad (4.9)$$

να ανήκει σε μια αρχική εκτίμηση αντικειμένου:

$$\underline{S} = [s(x_1), s(x_2), \dots, s(x_N)] \quad (4.10)$$



Σχήμα 4.7: Ο χρωματικός τόνος *Hue* των τριών παραδειγμάτων

όπου m_0^i είναι η τιμή του i -οστού στοιχείου του \underline{M}_0 , N είναι ο αριθμός των pixel του \underline{M}_0 με τιμή τη μονάδα και $s(x_i) = p^i$ είναι η πιθανότητα το pixel x_i να ανήκει στην εκτίμηση του αντικειμένου.

Για μια δεδομένη εικόνα \underline{X} (με συγκεκριμένη χρωματική συνιστώσα, όπως στα παραδείγματα που δίνονται στο Σχήμα 4.7) και μια αρχική εκτίμηση \underline{M}_0 εντός αυτής, σκοπός μας είναι να βρεθεί το διάνυσμα πιθανότητας \underline{S} , καθώς και μία ακριβέστερη εκτίμηση \underline{M}_f . Αυτό μπορεί μαθηματικά να εκφραστεί ως ένα πρόβλημα μεγιστοποίησης μιας συνάρτησης πιθανότητας υπό όρους:

$$\{\hat{\underline{M}}_f, \hat{\underline{S}}\} = \arg_{\underline{S}} \max L(\underline{M}_f, \underline{S}/\underline{M}_0, \underline{X}) \quad (4.11)$$

όπου το L συμβολίζει τη λογαριθμική συνάρτηση πιθανότητας.

Η πυκνότητα πιθανότητας που περιέχεται στην πιο πάνω έκφραση μπορεί να γραφτεί, δεδομένης της εικόνας εισόδου, ως:

$$\begin{aligned} Pr(\underline{M}_f, \underline{S}/\underline{M}_0) &= \\ Pr(\underline{S}/\underline{M}_0) \cdot Pr(\underline{M}_f/\underline{M}_0, \underline{S}) &= \\ Pr(\underline{S}/\underline{M}_0) \cdot Pr(\underline{M}_f/\underline{S}) & \end{aligned} \quad (4.12)$$

όπου η ακριβέστερη εκτίμηση \underline{M}_f εξαρτάται ρητά από το διάνυσμα πιθανότητας \underline{S} . Ωστόσο, αφού ενδιαφερόμαστε για τη μεγιστοποίηση της συνάρτησης πιθανότητας, ο δεύτερος παράγοντας της παραπάνω έκφρασης παίρνει την τιμή της μονάδας μόνο όταν $\underline{S} \geq \underline{T}$, όπου \underline{T} είναι ένα κατάλληλο διάνυσμα κατωφλίου. Κάτι τέτοιο υποδηλώνει ότι αυτός ο παράγοντας μπορεί να εξαλειφθεί, με την ακόλουθη έννοια:

$$\begin{aligned} \hat{\underline{S}} &= \arg_{\underline{S}} \max \{\log Pr(\underline{S}/\underline{M}_0)\} \\ &= \arg_{\underline{S}} \max \{\log Pr(\underline{M}_0/\underline{S}) - \log Pr(\underline{M}_0) + \log Pr(\underline{S})\} \\ &= \arg_{\underline{S}} \min \{-\log Pr(\underline{M}_0/\underline{S}) - \log Pr(\underline{S})\} \end{aligned} \quad (4.13)$$

Υποθέτοντας ότι για ένα δεδομένο διάνυσμα πιθανότητας \underline{S} υπάρχει μόνο μια αρχική εκτίμηση \underline{M}_0 , υπάρχει ένας πίνακας R ο οποίος συνδέει γραμμικά τα δύο διανύσματα. Αυτό επιτρέπει την εξάλειψη του πρώτου από τους δύο όρους κάτω από την προϋπόθεση της ελαχιστοποίησης, λόγω του γεγονότος ότι η σχετική πιθανότητα είναι 1 εάν εκπληρώνεται η γραμμική σχέση και 0 αλλιώς. Μια τέτοια απλοποίηση σημαίνει ότι ο μοναδικός όρος, ο οποίος μπορεί να ελαχιστοποιηθεί για μια δεδομένη εικόνα εισόδου, είναι ο δεύτερος.

Από τη μία πλευρά, το \underline{S} είναι μια διανυσματική τυχαία μεταβλητή (RV), η οποία είναι ουσιαστικά μια συνάρτηση που εξαρτάται από τη χρωματική συνιστώσα μίας γειτονιά pixel της εικόνας εισόδου, καθώς και από την περιοχή που ορίζεται από αυτή τη

γειτονιά. Για την ακρίβεια, η συνάρτηση που πρέπει να ελαχιστοποιηθεί είναι ανάλογη της χρωματικής μεταβλητότητας (hue variance) και αντιστρόφως ανάλογη της περιοχής της γειτονιάς pixel. Αυτή η συνάρτηση είναι επίσης μια τυχαία μεταβλητή, εφόσον έχει τις ακόλουθες ιδιότητες [62]:

- Το πεδίο ορισμού της περιέχει το εύρος της μεταβλητής x_i , η οποία θεωρείται επίσης ως τυχαία μεταβλητή.
- Είναι μια συνάρτηση Baire, αφού για κάθε s , το σύνολο των x_i , R_s , τέτοιο ώστε $s(x_i) \leq s$ αποτελείται από την ένωση και την τομή ενός πεπερασμένου αριθμού διαστημάτων (στον άξονα x).
- Τα συμβάντα $s(x_i) = \pm\infty$ έχουν μηδενική πιθανότητα.

Από την άλλη πλευρά, τα στοιχεία του \underline{S} αντιπροσωπεύουν τις πιθανότητες ένα pixel x_i να καταταγεί σε μια από δύο κατηγορίες, δηλαδή να ανήκει ή όχι σε ένα αντικείμενο. Είναι γνωστό όμως, ότι τα pixel είναι έντονα συσχετισμένα (correlated) με τα γειτονικά τους. Αυτό το γεγονός έχει σαν συνέπεια την τοπική συσχέτιση ανάμεσα σε γειτονικά στοιχεία του \underline{S} . Αυτή η ιδιότητα χαρακτηρίζει τα Τυχαία Πεδία Markov (MRF) [7, 29].

Αυτές οι δύο παρατηρήσεις επιτρέπουν τη χρήση της κατανομή Gibbs με σκοπό τη μοντελοποίηση της $Pr(\underline{S})$, εφόσον αυτού του είδους η κατανομή είναι ικανή να εκφράσει ρητά ένα MRF σε πιθανοτικά πλαίσια:

$$Pr(\underline{S}) = F \cdot \exp\{-a \cdot \sum_{n \in N} G_n(x)\} \quad (4.14)$$

όπου F και a είναι παράγοντες κανονικοποίησης, n είναι οι γειτονιές, N είναι η ένωσή τους, και η $G_n(x)$ θα πρέπει να εκφραστεί με όρους ελαχιστοποίησης της μεταβλητότητας του χρωματικού τόνου και μεγιστοποίησης της περιοχής μίας γειτονιάς. Επιλέγεται:

$$G_n(x) = \sum_{i \in n} \{(H(x_i) - \overline{H}_n)^2 + (s(x_i) - m_0(x_i))^2\} \quad (4.15)$$

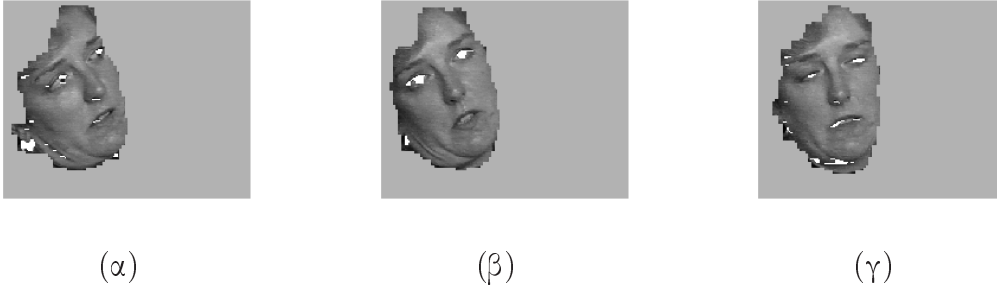
όπου i είναι ο δείκτης για κάθε pixel, H είναι η χρωματική συνιστώσα για την εικόνα X , \overline{H}_n είναι η μέση χρωματική συνιστώσα για την γειτονιά n , και $s(x_i)$ είναι η πιθανότητα στο pixel x_i . Ο πρώτος όρος του αθροίσματος μοντελοποιεί τη χρωματική μεταβλητότητα στη γειτονιά n . Ο δεύτερος όρος είναι ελάχιστος στην περίπτωση όπου οι πιθανότητες της γειτονιάς παραμένουν όσο το δυνατόν υψηλότερες, γεγονός που σημαίνει ότι η γειτονιά παραμένει όσο το δυνατόν μεγαλύτερη (μέγιστη περιοχή).

Από τις εξισώσεις 4.13, 4.14 και 4.15, προκύπτει:

$$\begin{aligned} \hat{\underline{S}} &= \arg_{\underline{S}} \max\{\log Pr(\underline{S})\} \\ &= \arg_{\underline{S}} \min\left\{\sum_{n \in N} \sum_{i \in n} \{(H(x_i) - \overline{H}_n)^2 + (s(x_i) - m_0(x_i))^2\}\right\} \end{aligned} \quad (4.16)$$

Αφού και οι δύο παράγοντες είναι τετραγωνικής μορφής, η εξίσωση 4.16 ορίζει μια κυρτή συνάρτηση (βλέπε [50]), γεγονός που υπονοεί ότι υπάρχει μια ολική λύση. Η ελαχιστοποίηση της 4.16 είναι άμεση, εάν υπολογιστεί η παράγωγος της έκφρασής της:

$$\begin{aligned} D &= \frac{\partial \{\sum_{n \in N} \sum_{i \in n} \{(H(x_i) - \overline{H}_n)^2 + (s(x_i) - m_0(x_i))^2\}\}}{\partial s} \\ &= \sum_{n \in N} \sum_{i \in n} \{2(H(x_i) - \overline{H}_n) \frac{\partial H(x_i)}{\partial s} + 2(s(x_i) - m_0(x_i))\} \end{aligned} \quad (4.17)$$



Σχήμα 4.8: Οι βελτιωμένες εκτιμήσεις για τα τρία παραδείγματα

όπου η παράγωγος $\partial H(x_i)/\partial s$ είναι είτε μηδέν στην περίπτωση που η τιμή του \underline{M}_f στο pixel x_i είναι ίδια από την κατάσταση j στην κατάσταση $j + 1$, ή $-H(x_i)$ στην περίπτωση όπου αυτή η τιμή αλλάζει. Μέσω της εξίσωσης 4.17 παρατηρούμε ότι υπάρχει όντως γραμμική σχέση μεταξύ των \underline{S} και \underline{M}_0 , όπως υποτέθηκε προηγουμένως. Λόγω του ότι το πρόβλημα βασίζεται στη δομή γειτονιάς, χρησιμοποιήθηκε η επαναληπτική τεχνική ICM (iterative conditional modes) για την εκτίμηση του βέλτιστου διανύσματος $\underline{\hat{S}}$. Το αντίστοιχο βέλτιστο διάνυσμα $\underline{\hat{M}}_f$ εκτιμάται μέσω του $\underline{\hat{S}}$ με την επιλογή ενός κατάλληλου διανύσματος κατωφλίου, όπως αναφέρθηκε. Παραδείγματα της εξόδου φαίνονται στο Σχήμα 4.8.

4.3.3 Εμπλουτισμός και μεγέθυνση των αρχικών εκτιμήσεων

Η έξοδος του προηγούμενου βήματος της μεθοδολογίας παρέχει μια ελάχιστη εκτίμηση για τη θέση των χαρακτηριστικών του προσώπου. Ο στόχος της τρέχουσας διαδικασίας είναι η βελτίωση του συνόλου των pixel που προέκυψε. Στον καθορισμό των αρχικών εκτιμήσεων των υποψήφιων χαρακτηριστικών, η χρωματική συνιστώσα της εικόνας εισόδου χρησιμοποιήθηκε απλά για να ορισθούν τοπικά οι μέγιστες περιοχές στις οποίες η χρωματική μεταβλητότητα ήταν ελάχιστη. Καθώς η χρωματική συνιστώσα ανταποκρίνεται στην καθαρά χρωματική πληροφορία (H) μιας εικόνας, θα εφαρμοστεί ένας εμπλουτισμός βασισμένος στην περιοχή καθενός από τα ελάχιστα αντικείμενα που καθορίστηκαν στο προηγούμενο στάδιο. Επίσης, η συνιστώσα V (value component) παρέχει σημαντική πληροφορία σε αυτό το στάδιο. Από όλη την πληροφορία σχετικά με τις ακμές που μπορεί να αντληθεί από τη V, θα κρατηθούν μόνο οι γραμμές που επικαλύπτονται μερικά με τις αρχικές εκτιμήσεις. Αυτού του είδους ο εμπλουτισμός θα παράγει αντικείμενα με όπες, τα οποία δεν περιγράφουν με ακρίβεια τα χαρακτηριστικά του προσώπου. Επιπρόσθετα, η ύπαρξη θορύβου παραμένει εμφανής. Για τους λόγους αυτούς, εφαρμόζεται μια σειρά μορφολογικών τελεστών που θα οδηγήσει στην περαιτέρω εξομάλυνση των περιοχών που έχουν εντοπισθεί. Οι μορφολογικές πράξεις περιλαμβάνουν την μορφολογία πλειοψηφίας (majority morphing) και το γέμισμα των κενών (filling of gaps) για την απομάκρυνση του θορύβου και τη δημιουργία κλειστών αντικειμένων, καθώς και τη γεφύρωση (bridging), για τη διατήρηση της συνεκτικότητας αντικειμένων που αποτελούνται από διαχωρισμένα αλλά πολύ κοντινά το ένα στο άλλο επιμέρους αντικείμενα. Τελικά, η διαδοχική εφαρμογή τέτοιων τελεστών έχει σαν αποτέλεσμα τον υπέρμετρο εμπλουτισμό της περιμέτρου των αντικειμένων, γεγονός που απαιτεί διάβρωση της περιμέτρου (perimeter erosion). Ο πυρήνας (kernel) για όλες τις μορφολογικές πράξεις προσαρμόζεται αυτόματα, ανάλογα με το μέγεθος και την πολυπλοκότητα των αντικειμένων, με έναν παράγοντα κλιμάκωσης που επιλέχθηκε εμπειρικά και επηρεάζεται από τον αριθμό των εφαρμοζόμενων μορφολογικών

τελεστών.

Η διαδικασία εμπλουτισμού μπορεί να παρουσιαστεί ως ένα πρόβλημα ελαχιστοποίησης ενέργειας, ως προς τους όρους E_H και E_{VM} , όπου ο πρώτος όρος αναφέρεται στον εμπλουτισμό κατά τη χρωματική συνιστώσα, και ο δεύτερος κατά τη συνιστώσα V και τη μορφολογική διαμόρφωση, όπως περιγράφηκε ανωτέρω.

Ο εμπλουτισμός βάσει της χρωματικής συνιστώσας βασίζεται στην ιδέα της ελαχιστοποίησης ενέργειας που προτάθηκε από τους Deng et al [19], μόνο που στην περίπτωση μας έχουμε μόνο δύο κλάσεις (δηλαδή μία κλάση ενδιαφέροντος) και η υπολογιστική πολυπλοκότητα μειώνεται σημαντικά λόγω του ότι

- οι υπολογισμοί περιορίζονται στις γειτονιές που ορίζονται από τις αρχικές εκτιμήσεις,
- η μείωση των χρωμάτων και ο χβαντισμός επιτυγχάνεται εύκολα, αφού για κάθε ένα από τα αντικείμενα της αρχικής εκτίμησης υπάρχει ένα κυρίαρχο χρώμα.

Συνεπώς, ο πρώτος όρος της ενέργειας εμπλουτισμού E_H παίρνει την ακόλουθη μορφή:

$$E_H = \frac{1}{N} \cdot \sum_{k \in W} M_k \cdot J_k \quad (4.18)$$

όπου η άθροιση πραγματοποιείται στις γειτονιές k , δηλαδή τα παράθυρα W που ορίζονται γύρω από τις αρχικές εκτιμήσεις που καθορίστηκαν στο προηγούμενο στάδιο. N είναι ο συνολικός αριθμός σημείων στον χάρτη των κλάσεων που έχει περιοριστεί βάσει του W . Μια άλλη ουσιαστική διαφοροποίηση από την προσέγγιση των Deng et al, είναι ότι για κάθε γειτονιά, η ποσότητα J_k υπολογίζεται για διαφορετικές κατατάξεις. Για την ακρίβεια, οι κλάσεις Z_i , $i = 1, 2$, όπως παρουσιάζονται στην εξίσωση 4.2, είναι διαφορετικές για κάθε k και εξαρτώνται από την παρουσία ή μη του κυρίαρχου χρώματος στο αντίστοιχο αντικείμενο της αρχικής εκτίμησης. Έτσι, η 4.18 ελαχιστοποιείται για όλα τα δυνατά μεγέθη παραθύρων, τα οποία καθορίζονται για κάθε αντικείμενο της αρχικής εκτίμησης από την περιοχή και τη δομή του.

Ο όρος E_{VM} ορίζεται ως ένα μοντέλο επικάλυψης μεταξύ της αρχικής εκτίμησης και της επέκτασής της μέσω της συνιστώσας V και των μορφολογικών φίλτρων:

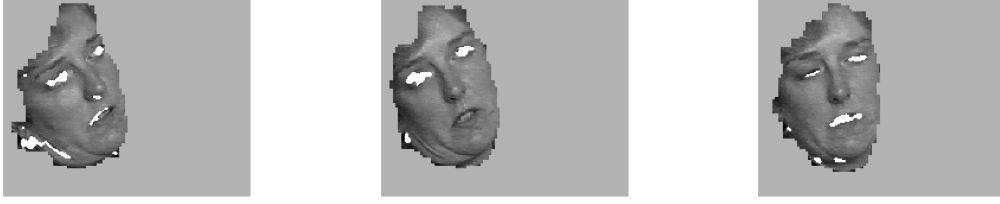
$$E_{VM} = \frac{A_S \cdot A_{VM}}{A_O^2} \quad (4.19)$$

όπου A_S είναι η περιοχή της αρχικής εκτίμησης, A_{VM} είναι η περιοχή της επέκτασης και A_O είναι η περιοχή της επικάλυψης. Όσο η τομή της περιοχής μεγαλώνει, τόσο μειώνεται η E_{VM} . Η ελαχιστοποίηση αυτού του λόγου μεγιστοποιεί την επικάλυψη μεταξύ της αρχικής εκτίμησης για το αντικείμενο και αυτής που προκύπτει μετά την επέκταση. Δεν εφαρμόζεται κανένα κατώφλι, λόγω του γεγονότος ότι τα αντικείμενα της αρχικής κατάτμησης μπορεί να είναι σημαντικά μικρότερα από το συνδυασμό της επέκτασης. Η σύγκλιση αυτού του όρου βασίζεται στην υπόθεση ότι υπάρχει συνεκτικότητα κατά τη συνιστώσα V για κάθε ζητούμενο χαρακτηριστικό.

Μερικά αποτελέσματα φαίνονται στο Σχήμα 4.9.

4.3.4 Καθορισμός αντικειμένων και υπολογισμός της επικρατούσας γωνίας

Τα βήματα που έχουν παρουσιαστεί μέχρι αυτό το στάδιο έχουν ως αποτέλεσμα ένα σύνολο αντικειμένων που είναι υποψήφια ως τα ζητούμενα χαρακτηριστικά εντός της

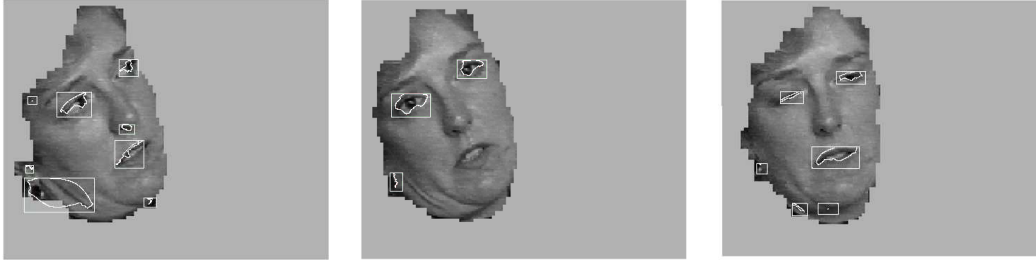


(α)

(β)

(γ)

Σχήμα 4.9: Οι εκτιμήσεις μετά τον εμπλουτισμό και την εξάπλωση των αρχικών αντικειμένων



(α)

(β)

(γ)

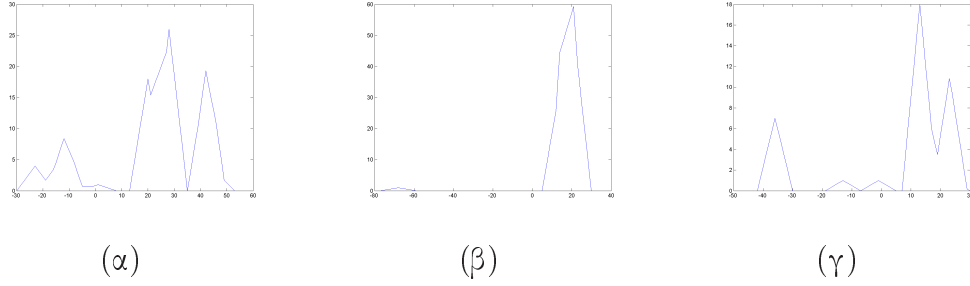
Σχήμα 4.10: Αρχικοποίηση και τελική προσέγγιση μέσω των ενεργών περιγραμμάτων

εικόνας του προσώπου. Ας συμβολίσουμε το σύνολο των υποψήφιων αντικειμένων:

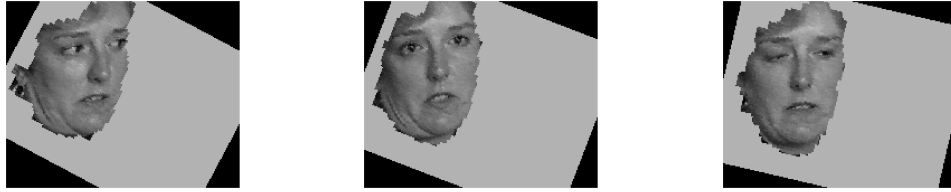
$$S_c = \{O_c : (O_c \in F) \vee (O_c \notin F)\} \quad (4.20)$$

όπου το υποψήφιο αντικείμενο O_c μπορεί να ανήκει ή όχι στο επιθυμητό σύνολο χαρακτηριστικών F . Κάθε αντικείμενο O_c μπορεί να μοντελοποιηθεί και να προσεγγιστεί με τη χρήση ενός ενεργού περιγράμματος, το οποίο είναι ικανό να περιγράψει το σχήμα του και την κατευθυντικότητα του στον 2Δ χώρο της εικόνας. Το ενεργό περίγραμμα απαρτίζεται από ένα σύνολο σημείων ελέγχου, τα οποία συνδέονται μεταξύ τους με ευθύγραμμα τμήματα σχηματίζοντας κλειστούς βρόχους. Για κάθε αντικείμενο O_c , δημιουργείται ένα αρχικό ενεργό περίγραμμα ως το ορθογώνιο που περιέχει το αντικείμενο στην εικόνα, όπως φαίνεται στο Σχήμα 4.10. Κατ' αυτόν τον τρόπο επιτυγχάνεται η αυτόματη αρχικοποίηση του ενεργού περιγράμματος. Η προσέγγιση του πραγματικού αντικειμένου επιτυγχάνεται από το συνδυασμό δύο ειδών δυνάμεων: τις εσωτερικές ή ελαστικές και τις εξωτερικές ή κλίσης [45], όπως περιγράφηκε στην ενότητα 4.2. Αυτές οι δυνάμεις εφαρμόζονται ακολουθιακά για έναν αριθμό επαναλήψεων που εξαρτάται από το μέγεθος και την πολυπλοκότητα του σχήματος του αντικειμένου. Για τα τρία παραδείγματά, τα τελικά μοντέλα ενεργού περιγράμματος για όλα τα αντικείμενα O_c φαίνονται στο Σχήμα 4.10.

Όπως αναφέρθηκε ανωτέρω, τα μοντέλα ενεργού περιγράμματος είναι κατευθυντικά από τη φύση τους. Κάτι τέτοιο σημαίνει ότι για κάθε O_c μπορούμε να ορίσουμε μία γωνία $theta_i$ που υποδηλώνει την κατεύθυνση του αντικειμένου στον 2Δ χώρο. Η επικρατούσα γωνία $theta_d$ για όλα τα αντικείμενα υπολογίζεται μέσω ενός απλού ασαφούς υποσυστήματος. Για κάθε αντικείμενο O_c , ανατίθεται στη γωνία του $theta_i$ μια τριγωνική συνάρτηση συμμετοχής (membership function), της οποίας το πλάτος είναι ανάλογο με το μέγεθος του αντικειμένου. Αν προστεθούν όλες οι συναρτήσεις συμμετοχής στον άξονα των $theta$, εμφανίζεται συνήθως μια συγκέντρωση γύρω από μία συγκεκριμένη γωνία. Έτσι, η γωνία που συγκεντρώνει το μέγιστο πλάτος καλείται επικρατούσα γωνία $theta_d$. Το ίδιο πρόβλημα μπορεί να ειπωθεί ως ένα ζήτημα



Σχήμα 4.11: Έξοδος του ασαφούς συστήματος που καθορίζει την επικρατούσα γωνία. Οι τριγωνικές συναρτήσεις συμμετοχής για κάθε αντικείμενο αθροίζονται, σχηματίζοντας συγκεντρώσεις γύρω από την επικρατούσα γωνία θ_{eta_d}



Σχήμα 4.12: Οι εικόνες εισόδου περιστραμμένες στο επίπεδο, βάσει των αυτόματα υπολογισμένων γωνιών κλίσης

ελαχιστοποίησης σφάλματος ως εξής:

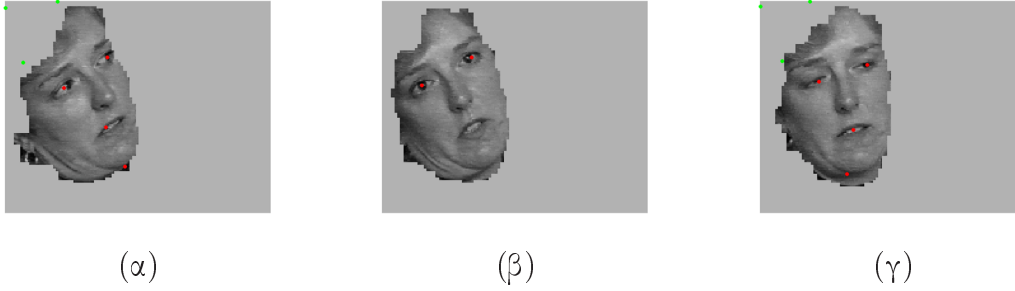
$$E_{ang} = \sum_{i=1}^{N_O} [C - f_m(O_i)] \cdot F_O \quad (4.21)$$

όπου N_O είναι ο αριθμός των αντικειμένων στο σύνολο S_c , f_m είναι η συνάρτηση συμμετοχής για τα αντικείμενα O_i και τις αντίστοιχες γωνίες, ενώ F_O είναι ένας παράγοντας που εντοπίζει αντικείμενα μη συγκρινόμενα με την πλειοψηφία (outlier factor). Αυτό το σφάλμα αντικατοπτρίζει την επιλογή του ασαφούς συστήματος (αθροιστική μεγιστοποίηση των συναρτήσεων συμμετοχής), συμπεριλαμβανομένης της ανίχνευσης στοιχείων εκτός συνόλου. Για μια δεδομένη συνάρτηση συμμετοχής f_m , εξαρτάται από τη γωνία του ενεργού περιγράμματος και το πλάτος της f_m για το αντικείμενο O_i . Επιλέχθηκε η τριγωνική συνάρτηση συμμετοχής με πλάτος που είναι γραμμικά ανάλογο της περιοχής του αντικειμένου, έτσι ώστε να εκφράζει το βαθμό αβεβαιότητας ως προς την εκτίμηση της γωνίας, η οποία υπολογίζεται αυτόματα για κάθε υποψήφιο αντικείμενο. Ο βαθμωτός παράγοντας F_O που περιλαμβάνεται, εξαρτάται από το σχετικό μέγεθος του αντικειμένου.

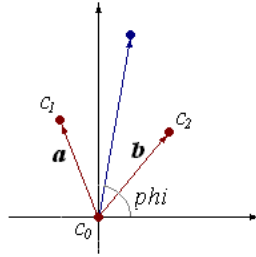
Η γωνία που ελαχιστοποιεί την 4.21 χρησιμοποιείται για την περιστροφή των καταμημένων προσώπων των εικόνων εισόδου. Τα αποτελέσματα για την εκτίμηση της επικρατούσας γωνίας και την περιστροφή των προσώπων απεικονίζονται στα Σχήματα 4.11 και 4.12 αντίστοιχα.

4.3.5 Επιλογή και ονοματισμός των βασικών χαρακτηριστικών

Η επικρατούσα γωνία χρησιμεύει ως ένα κριτήριο για περαιτέρω περιορισμό του συνόλου S_c . Χρησιμοποιώντας την τυπική απόκλιση της κατανομής των γωνιών στο ασαφές



Σχήμα 4.13: Το περιορισμένο, βάσει της επικρατούσας γωνίας, σύνολο των υποψήφιων αντικειμένων. Στο πάνω αριστερά μέρος κάθε εικόνας απεικονίζεται η επικρατούσα τριάδα αντικειμένων (όπου αυτή υπάρχει) περιστραμμένη κατά την επικρατούσα γωνία



Σχήμα 4.14: Το κανονικοποιημένο γεωμετρικό πρότυπο για τα μάτια και το στόμα

σύστημα, ορίζεται μια περιοχή γύρω από την επικρατούσα γωνία, η οποία υλοποιεί αυτόν τον περιορισμό. Τα αντικείμενα που βρίσκονται μέσα σε αυτό το παράθυρο γωνιών σχηματίζουν το νέο σύνολο. Ας συμβολίσουμε με

$$S'_c = \{O'_c : (O'_c \in F) \vee (O'_c \notin F)\} \quad (4.22)$$

το νέο περιορισμένο σύνολο. Τα αποτελέσματα για τις εικόνες των παραδειγμάτων φαίνονται στο Σχήμα 4.13.

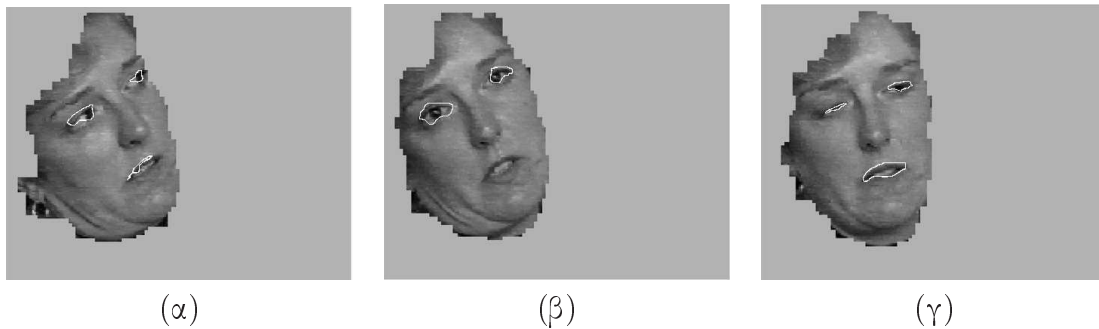
Μέχρι αυτό το στάδιο του αλγορίθμου, έχουν γίνει χαλαρές υποθέσεις για τη φύση των χαρακτηριστικών χωρίς να χρησιμοποιηθεί εκ των προτέρων γνώση για αυτά. Το τρέχον στάδιο θα επιχειρήσει τον προσδιορισμό των ζητούμενων χαρακτηριστικών, δηλαδή των ματιών και του στόματος. Αυτά θεωρούνται αρκετά για το σχηματισμό της βάσης για την περαιτέρω εξαγωγή των δευτερευόντων χαρακτηριστικών, εφόσον μπορούν να παρέχουν εκ των υστέρων γνώση για την πραγματική θέση άλλων χαρακτηριστικών του προσώπου, όπως είναι τα φρύδια, η μύτη, καθώς και μεταβατικά χαρακτηριστικά (ρυτίδες γύρω από τη μύτη, τα μάτια κλπ), εάν αυτά είναι ορατά ;;. Σε αυτήν την προσπάθεια, χρησιμοποιείται γενική γνώση για την γεωμετρική εξάρτηση μεταξύ των βασικών χαρακτηριστικών σε ένα κανονικοποιημένο (μετά από την περιστροφή στο επίπεδο) πρόσωπο, όπως φαίνεται στο Σχήμα 4.14.

Η κανονικοποιημένη τριάδα των υποψήφιων αντικειμένων που είναι πλησιέστερη προς το γεωμετρικό πρότυπο που εκφράζεται στην εξίσωση 4.23 και απεικονίζεται στο Σχήμα 4.14 θεωρείται ως η νικήτρια τριάδα.

$$E_{sym} = (|a| - |b|) \cdot |90^\circ - \phi| / \max(|a|, |b|) \quad (4.23)$$

Η τριάδα που ελαχιστοποιεί την έκφραση αυτή εκτιμάται εξετάζοντας τον αριθμό των πιθανών τριάδων, ο οποίος είναι:

$$\binom{n}{3} = \frac{n!}{3!(n-3)!} \quad (4.24)$$



Σχήμα 4.15: Επιλογή και προσδιορισμός της ταυτότητας των χαρακτηριστικών του προσώπου που έχουν ανιχνευθεί

όπου n είναι η διάσταση του S'_c .

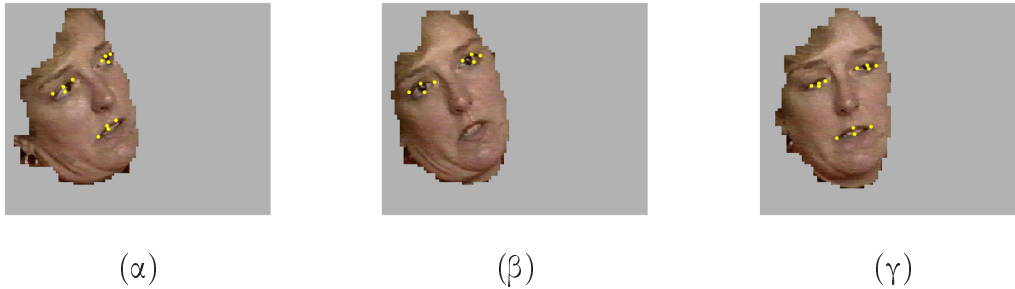
Είναι ωστόσο πιθανή και η περίπτωση κατά την οποία δεν είναι παρόντα και τα τρία χαρακτηριστικά. Κάτι τέτοιο θα μπορούσε να οφείλεται σε μερική απόκρυψη, ή στην αποτυχία του αλγορίθμου να τα περιλάβει όλα στο σύνολο S'_c , ή ακόμα και στην ανεπάρκεια του συνόλου ($n < 3$). Στην περίπτωση όπου $n = 2$, ή γενικά στην περίπτωση ζεύγους υποψηφίων, όταν αυτά βρίσκονται στην ίδια παράλληλη με το οριζόντιο άξονα (μετά την περιστροφή στο επίπεδο της εικόνας και την κανονικοποίηση), ακόμα και στην περίπτωση μικρών αποκλίσεων, θεωρείται ότι αντιπροσωπεύουν τα μάτια. Για $n = 1$ και για μοναδικούς υποψήφιους, δεν μπορεί να υπάρχει ασφαλής πρόβλεψη για τον τύπο τους χαρακτηριστικού. Επιπρόσθετα, στην περίπτωση όπου $n = 1$ δημιουργείται και το ερώτημα του πώς ένα χαρακτηριστικό επιβάλλει μια επικρατούσα γωνία. Στις περιπτώσεις αυτές μπορεί να επαναληφθεί η διαδικασία με μεγαλύτερη ανοχή σε πιθανά σφάλματα και δίνοντας βαρύτητα σε όποια χαρακτηριστικά έχουν ήδη ανιχνευθεί. Τα αποτελέσματα του προσδιορισμού των χαρακτηριστικών φαίνονται στο Σχήμα 4.15.

4.3.6 Υπολογισμός των παραμέτρων ορισμού και δυναμικής κίνησης του προσώπου

Όταν επιλεγούν και ονοματιστούν τα κατάλληλα αντικείμενα, μπορεί να υπολογιστεί ένα υποσύνολο των παραμέτρων δυναμικής κίνησης των ομοιωμάτων των προσώπων FAPs (Facial Animation Parameters) και των παραμέτρων ορισμού του προσώπου FDP (Facial Definition Parameter) (όπως αυτές ορίζονται από το πρότυπο MPEG-4 [58]). Η εξαγωγή των FDP βασίζεται στον υπολογισμό των οριακών σημείων των αντικειμένων ανάλογα με τον προσανατολισμό που προκύπτει από την επικρατούσα γωνία. Στη συνέχεια, έχοντας τις συντεταγμένες των σημείων, μετρούνται οι μετατοπίσεις τους μεταξύ διαδοχικών καρέ. Τέλος, τα αποτελέσματα αυτών των μετρήσεων μοντελοποιούνται μέσω των FAP χρησιμοποιώντας έτσι στην αναπαράσταση των εκφράσεων του προσώπου. Τα σημεία FDP για τις εικόνες των παραδειγμάτων παρουσιάζονται στο Σχήμα 4.16.

4.4 Αποτελέσματα προσομοιώσεων

Κάθε μία από τις προαναφερθείσες διαδικασίες βελτιστοποίησης εφαρμόστηκε σε καρέ πραγματικών ακολουθιών βίντεο, όπου η κατάτμηση προσώπων έχει ήδη πραγματοποιηθεί [2, 84].



Σχήμα 4.16: Τα σημεία που αντιπροσωπεύουν ένα υποσύνολο των παραμέτρων ορισμού του προσώπου *FDP* σύμφωνα με το πρότυπο *MPEG-4*

Το σύστημα υλοποιήθηκε σε υπολογιστή με επεξεργαστή Pentium III και για τις προσομοιώσεις χρησιμοποιήθηκε το περιβάλλον της MATLAB. Όλες οι στατιστικές επίδοσης αναφέρονται για μη βελτιστοποιημένο (optimized) κώδικα. Για να υποστηριχθεί η αποδοτικότητα του συστήματος, ο κώδικας έτρεξε για πάνω από 100 περιπτώσεις που προέρχονται από διάφορες φυσικές ακολουθίες βίντεο που περιέχουν πρόσωπα σε ποικίλες στάσεις και θέσεις και με διάφορες εκφράσεις προσώπου: εικόνες, εργαστηριακές εγγραφές, ακόμα και τηλεοπτικές εκπομπές [13].

Τα Σχήματα 4.3 έως 4.8 παρουσιάζουν όλα τα στάδια της φάσης του καθορισμού των αρχικών εκτιμήσεων για τα αντικείμενα για έναν αριθμό χαρακτηριστικών καρέ της χρησιμοποιούμενης βάσης. Η βελτιστοποίηση των αρχικών εκτιμήσεων έχει ως αποτέλεσμα τους ακριβέστερους χάρτες M_f , οι οποίοι φαίνονται στο Σχήμα 4.8 για τις εικόνες του παραδείγματός μας. Το Σχήμα 4.9 απεικονίζει την επέκταση των αρχικών αντικειμένων που προκύπτει από την ελαχιστοποίηση της ενέργειας διαστολής, όπως περιγράφεται μαθηματικά από τις εξισώσεις 4.18 και 4.19. Στο Σχήμα 4.10 φαίνονται οι αρχικοποιήσεις και οι τελικές εκτιμήσεις των ενεργών περιγραμμάτων. Τα Σχήματα 4.11 και 4.12 αναφέρουν την εκτίμηση της επικρατούσας γωνίας, δηλαδή τις γωνίες περιστροφής για κάθε ένα από τα παραδείγματα. Το περιορισμένο (εκλεπτυσμένο) σύνολο υποψήφιων αντικειμένων S'_c και οι κανονικοποιημένες τριάδες που επικράτησαν παρουσιάζονται στο Σχήμα 4.13. Τα πρωτογενή χαρακτηριστικά επιλέγονται και προσδιορίζονται όπως φαίνεται στο Σχήμα 4.15. Τελικά, τα χαρακτηριστικά σημεία που προκύπτουν υπολογίζονται και απεικονίζονται στο Σχήμα 4.16, και σχηματίζουν ένα υποσύνολο του συνόλου FDP κατά το MPEG-4.

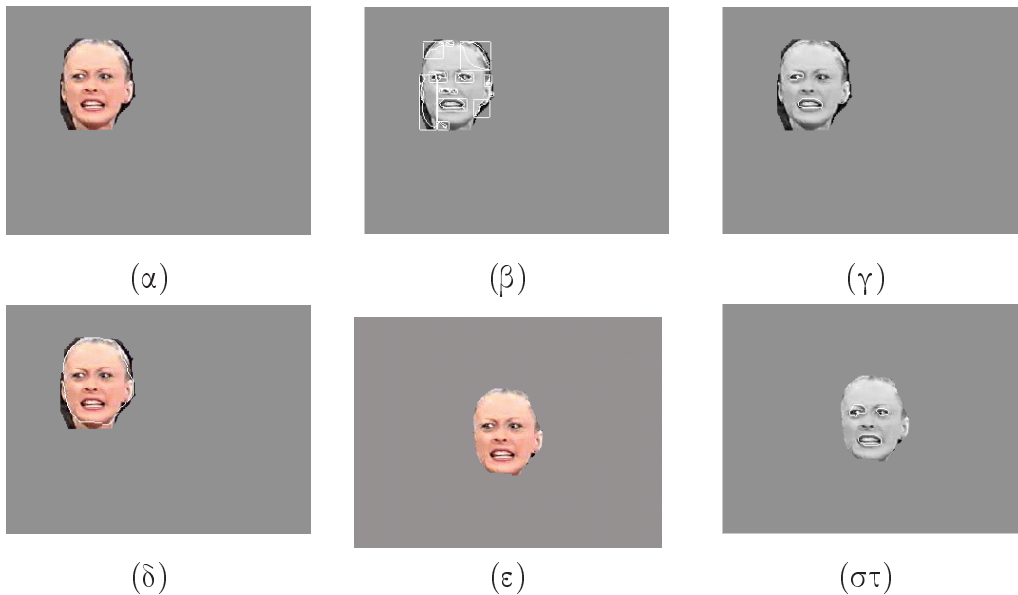
Τα αποτελέσματα που εξήχθησαν με την αυτοματοποιημένη αυτή διαδικασία συγκρίθηκαν με τα αντίστοιχα που εξάγονται και ονοματίζονται χειρονακτικά από το υλικό εισόδου του συστήματος. Για να μετρηθεί η συνέπεια των αποτελεσμάτων, εφαρμόστηκε μια σειρά από μετρικές για κάθε χαρακτηριστικό, καθώς και για το σύνολο και τη διάταξη των χαρακτηριστικών, και στις δύο περιπτώσεις (αυτόματη και χειρονακτική). Συγκεκριμένα, υπολογίστηκε η Ευκλείδεια απόσταση μεταξύ των δύο περιπτώσεων για κάθε χαρακτηριστικό σημείο, για το κέντρο βάρους κάθε χαρακτηριστικού αντικειμένου, καθώς και για τη σχετική θέση όλων των κέντρων βάρους. Τέλος λαμβάνεται υπόψη η απουσία λόγω αδυναμίας εξαγωγής κάποιων χαρακτηριστικών καθώς και η παρουσία μικρών αντικειμένων λόγω θορύβου. Αυτά τα τρία επίπεδα των μέτρων και ο παράγοντας εξωπραγματικού αποτελέσματος εφαρμόστηκαν από κοινού για την αξιολόγηση του βαθμού επιτυχίας της διαδικασίας στο σύνολο των εξεταζόμενων καρέ. Τα αποτελέσματα δείχνουν ότι μόνο το 4% των συνόλων χαρακτηριστικών χαρακτηρίστηκε λανθασμένα (mislabelled). Για τις εικόνες αυτές, εκτιμάται ότι η αποτυχία οφείλεται κυρίως στη μη ακριβή κατάτμηση του προσώπου, πράγμα που σημαίνει ότι συμπεριλαμ-

βάνονταν αντικείμενα που δεν αναμένονταν να υπάρχουν στη μάσκα του προσώπου. Για το 7% των εικόνων εισόδου τα σύνολα χαρακτηριστικών δεν ανιχνεύθηκαν καθόλου. Σε αυτές τις περιπτώσεις, το σύστημα παρουσίασε την ικανότητά αντίληψης της αδυναμίας του να παρέχει μια αξιόπιστη λύση, μην παρουσιάζοντας καμία λύση. Αυτό είναι αποτέλεσμα του μηχανισμού συμμετρίας όπως περιγράφηκε ανωτέρω. Για το υπόλοιπο 89% των εικόνων εισόδου, σχεδόν το 9% των χαρακτηριστικών (στόμα) δεν εντοπίστηκαν ενώ ήταν παρόντα, γεγονός που οφείλεται ξανά κυρίως στην αβεβαιότητα του συστήματος για το αν το χαρακτηριστικό που λείπει ανήκει πράγματι στο σύνολο που αναζητείται. Μόνο ένα 1% των χαρακτηριστικών τοποθετήθηκε λανθασμένα, ενώ τα υπόλοιπα χαρακτηριστικά εντοπίστηκαν ικανοποιητικά, με μια λογική ανοχή απόστασης. Η αποτελεσματικότητα του αυτόματου μηχανισμού στροφής στο επίπεδο εκτιμήθηκε ξεχωριστά. Τα ποσοστά αποτυχίας για την εκτίμηση της γωνίας περιστροφής, με μια ανοχή, συμβαδίζουν με αυτά του λάθος χαρακτηρισμού, δηλαδή 4%, και οφείλονται στους ίδιους λόγους.

Η μεθοδολογία που παρουσιάστηκε στην ενότητα αυτή, εστιάζεται στην εξαγωγή χαρακτηριστικών του προσώπου από εικόνες με ήδη εντοπισμένα και καταταμημένα πρόσωπα. Εν τούτοις, προβλήματα μπορούν να παρουσιαστούν, ειδικά όταν χρησιμοποιούνται απλές μέθοδοι εντοπισμού του προσώπου, για παράδειγμα βασισμένες στην ανίχνευση του χρώματος του δέρματος. Αυτό μπορεί να οδηγήσει στην λανθασμένη εξαγωγή των χαρακτηριστικών λόγω απουσίας περιοχών και χαρακτηριστικών του προσώπου, ανακριβών περιγραμμάτων, ασύνδετων περιοχών, ή λόγω του ότι μέρος των μαλλιών ή του φόντου έχουν συμπεριληφθεί στην περιοχή του προσώπου. Η ευρωστέια του συστήματος δοκιμάστηκε και υπό τις προαναφερθείσες συνθήκες. Το Σχήμα 4.17(α) δείχνει μία εικόνα προσώπου που έχει καταταμηθεί χωρίς ακρίβεια και η οποία χρησιμοποιήθηκε ως είσοδος στο σύστημα. Στο Σχήμα 4.17(β) απεικονίζονται οι αρχικοποιήσεις και οι τελικές εκτιμήσεις μετά την εφαρμογή των ενεργών περιγραμμάτων, ενώ είναι εμφανής η αύξηση του αριθμού των υποψήφιων αντικειμένων. Σχεδόν κάθε τμήμα του φόντου ή των μαλλιών που έχει συμπεριληφθεί θεωρείται υποψήφιο χαρακτηριστικό του προσώπου. Τα πρωτεύοντα χαρακτηριστικά που επιλέχθηκαν και προσδιορίστηκαν από το σύστημα φαίνονται στο Σχήμα 4.17(γ): μόνο το δεξί μάτι και το στόμα ανιχνεύθηκαν σωστά.

Εξετάστηκαν δύο μέθοδοι για την αντιμετώπιση περιπτώσεων ανακριβούς αρχικής κατάτμησης του προσώπου. Η πρώτη αφορούσε την επανάληψη των βημάτων (β) και (γ) της μεθόδου όπως αυτά φαίνονται και στο Σχήμα 4.1, αυξάνοντας σταδιακά την συμβολή στους υπολογισμούς του δεύτερου βήματος των δύο χαρακτηριστικών που ανιχνεύθηκαν εξ' αρχής. Η δεύτερη προσέγγιση χρησιμοποιεί αρχικά ένα ενεργό περίγραμμα [85] για τη βελτίωση της αρχικής κατάτμησης του προσώπου, και στη συνέχεια προχωράει στην εφαρμογή της μεθοδολογίας που παρουσιάστηκε. Το Σχήμα 4.17(δ) δείχνει την προσέγγιση του προσώπου μέσου του ενεργού περιγράμματος, ενώ το Σχήμα 4.17(ε) δείχνει το πρόσωπο που χρησιμοποιείται ως είσοδο του συστήματος μετά τη βελτιωμένη κατάτμηση. Τέλος, το Σχήμα 4.17(στ) παρουσιάζει την σωστή εξαγωγή των χαρακτηριστικών όπως αυτή επιτυγχάνεται και με τις δύο ανωτέρω αντιμετώπισεις.

Τέλος, παρατίθενται κάποια πειραματικά αποτελέσματα στην κατεύθυνση της δυναμικής κίνησης ομοιωμάτων του προσώπου (animation), στα πλαίσια του προτύπου MPEG-4, μέσω της χρήσης των παραμέτρων FDP και FAP. Βάσει των [83, 67], χρησιμοποιήθηκε η εκτίμηση των FDP μεταξύ της ουδέτερης έκφρασης και αργά αλλά συνεχώς εξελισσόμενων ποικίλων εκφράσεων, ώστε να υπολογιστούν οι παράμετροι FAP. Για τα πειράματα αυτά δημιουργήθηκε ένα ομοίωμα του προσώπου μέσω ενός

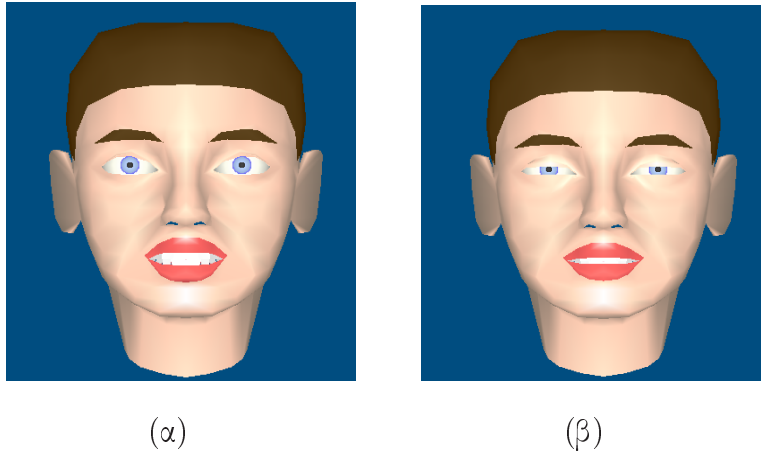


Σχήμα 4.17: Δοκιμή με αρχική εικόνα (α) που περιέχει ανεπαρκώς κατατεμημένο πρόσωπο, (β) τα αντικείμενα που ανιχνεύθηκαν και η μοντελοποίησή τους μέσω ενεργών περιγραμμάτων, (γ) τα χαρακτηριστικά του προσώπου που ανιχνεύθηκαν, (δ) βελτίωση της αρχικής κατάτμησης μέσω ενεργών περιγραμμάτων, (ε) αρχική εικόνα μετά την βελτιωμένη κατάτμηση (στ) τα χαρακτηριστικά του προσώπου όπως προσδιορίστηκαν τελικά

μοντέλου που αναπτύχθηκε στα πλαίσια ερευνητικού προγράμματος (Project ACTS MoMuSys, <http://www.iso.ch/ittf>). Χρησιμοποιώντας τις παραμέτρους FDP που εξάγονται σύμφωνα με τη μέθοδο, όταν αυτή εφαρμόζεται σε καρέ μίας ακολουθίας (όπως αυτά στα Σχήματα 4.2 και 4.15, (α) - (γ)), και θεωρώντας ένα από τα καρέ ως αντιπροσωπευτικό της ουδέτερης έκφρασης, δημιουργήθηκαν τα αντίστοιχα animation για το πρόσωπο όπως αυτά φαίνονται στο Σχήμα 4.18.

4.5 Συμπεράσματα και επεκτάσεις

Στην παρούσα ενότητα παρουσιάστηκε μια προσέγγιση σταδιακής εμπιστοσύνης που αφορά στην εξαγωγή χαρακτηριστικών του προσώπου από καρέ πραγματικών ακολουθιών βίντεο. Η προτεινόμενη μεθοδολογία αντιμετωπίζει μεγάλες παραλλαγές στην απεικόνιση διαφορετικών ατόμων, καθώς και του ίδιου ατόμου σε διαφορετικές στιγμές μέσα σε πραγματικές ακολουθίες βίντεο. Υπό αυτήν την έννοια, το πρόβλημα της εξαγωγής χαρακτηριστικών αντιμετωπίζεται υπό γενικευμένες συνθήκες απεικόνισης. Το σύστημα εξάγει περιοχές του προσώπου οι οποίες ξεχωρίζουν από στατιστικής άποψης και σχηματίζει ένα αρχικό σύνολο περιοχών που είναι πιθανό να περιέχουν πληροφορία σχετικά με τα χαρακτηριστικά που ενδιαφέρουν. Το σύστημα καθορίζει την επικρατούσα γωνία ανάμεσα σε όλα τα αντικείμενα του συνόλου (η οποία θεωρείται ως η γωνία περιστροφής του προσώπου στο επίπεδο της εικόνας) με τη χρήση ενός απλού ασαφούς συστήματος. Το σύνολο αντικειμένων περιορίζεται χρησιμοποιώντας την επικρατούσα γωνία. Μια εξαντλητική αναζήτηση ανάμεσα στα υποψήφια αντικείμενα στοχεύει να τα ταιριάζει στο πρότυπο που αντιπροσωπεύει την γεωμετρία του ανθρώπινου προσώπου. Τα χαρακτηριστικά που επιλέγονται και προσδιορίζονται θα χρησιμοποιηθούν για τον υπολογισμό τελικά των παραμέτρων ορισμού του προσώπου σύμφωνα με το πρότυπο MPEG-4. Η σταδιακή αποκάλυψη της πληροφορίας που αφορά



Σχήμα 4.18: Δυναμική κίνηση ομοιώματος προσώπου βάσει των υπολογισμένων παραμέτρων *FAP* σύμφωνα με το πρότυπο *MPEG-4*: (α) ουδέτερη έκφραση, (β) έκφραση που αντιστοιχεί στο Σχήμα 4.16(γ)

στα χαρακτηριστικά του προσώπου υποστηρίζεται από μία διαδικασία βελτιστοποίησης για κάθε βήμα, που παράγει εκ των υστέρων γνώση για το πρόσωπο και οδηγεί σε μια βήμα προς βήμα οπτικοποίηση των χαρακτηριστικών που αναζητούνται.

Κατά την προσέγγισή αυτή, εντοπίζονται βασικά χαρακτηριστικά του προσώπου, όπως τα μάτια και το στόμα. Μελλοντικές επεκτάσεις συμπεριλαμβάνουν την διερεύνηση μίας αξιόπιστης, ιεραρχικής εξαγωγής δευτερευόντων και μεταβατικών χαρακτηριστικών του προσώπου, καθώς και την εύρωστη παρακολούθηση, σε πραγματικό χρόνο, για ακολουθίες βίντεο. Επίσης, εξετάζεται και υλοποιείται η χρήση της προσέγγισης αυτής για τη δημιουργία συναισθηματικά πλούσιων διεπαφών μεταξύ ανθρώπου και υπολογιστή, όπου ο τελευταίος μπορεί να αναλύει το λόγο, την έκφραση και τις κινήσεις του ανθρώπου για να εξάγει συμπεράσματα για την συναισθηματική του κατάσταση και να αντιδράσει αναλόγως.

□

Κεφάλαιο 5

Αλληλεπίδραση Πραγματικού και Εικονικού Ηθοποιού σε Εικονικά Στούντιο

5.1 Εισαγωγή

Εφόσον αντιμετωπιστεί επιτυχώς το κύριο θέμα ενός προηγμένου συστήματος εικονικού στούντιο, όπως αυτό περιγράφηκε στις προηγούμενες παραγράφους, μία πληθώρα βελτιώσεων γίνεται διαθέσιμη. Αρχικά, για παράδειγμα, παθητικοί εικονικοί ηθοποιοί, με περιορισμένη αυτονομία που λειτουργούν βάσει θεμελιώδους τεχνητής νοημοσύνης, μπορούν να τοποθετηθούν στο σκηνικό, κινούμενοι αποφεύγοντας άλλα αντικείμενα ή ακολουθώντας τους πραγματικούς ηθοποιούς. Κάνοντας ακόμα ένα βήμα, οι εικονικοί ηθοποιοί θα πρέπει να έχουν τη δυνατότητα να αλληλεπιδράσουν αυτόνομα με το περιβάλλον και ειδικά με τους ανθρώπους. Αυτό μπορεί να βασιστεί σε ανάδραση οπτικής πληροφορίας από τους ανθρώπους, όπως η στάση του σώματος, η κατεύθυνση του βλέμματος, οι κινήσεις των χεριών και οι εκφράσεις του προσώπου. Με τη σειρά τους, οι εικονικοί ηθοποιοί θα πρέπει να έχουν τη δυνατότητα να μεταδώσουν το γεγονός της παρουσίας τους, να αντιληφθούν τον πιθανό ρόλο του ανθρώπου στη σκηνή, τις κινήσεις του απέναντί τους και, τελικά, να επικοινωνήσουν και να αλληλεπιδράσουν. Πολλοί ερευνητές δραστηριοποιούνται στον τομέα των εικονικών ηθοποιών από διάφορες προσεγγίσεις και απόψεις και στοχεύουν σε μία ποικιλία εφαρμογών. Ενδεικτικά, αναφέρουμε την [82] που στοχεύει στην δημιουργία μίας εφαρμογής στην οποία ένας άνθρωπος θα μπορεί να παίζει ντάμα με ένα εικονικό αντίπαλο, ή την [9] που στοχεύει στη δημιουργία ενός αλληλεπιδραστικού περιβάλλοντος αφήγησης ιστοριών. Τέλος σε εργασίες όπως η [38] γίνονται βήματα για την εγκαθίδρυση των βάσεων και ενός πλαισίου για τους εικονικούς ηθοποιούς χρησιμοποιώντας τις επιστήμες της συμπεριφοράς για να καταλήξουν σε ρεαλιστικότερα αποτελέσματα.

Ανάμεσα σε όλες αυτές τις προοπτικές, στο παρόν κεφάλαιο θα εξεταστούν και θα παρουσιαστούν βασικά εργαλεία για την αλληλεπίδραση μεταξύ των πραγματικών και των εικονικών ηθοποιών, βάσει της εκτίμησης της συμπεριφοράς τους και ειδικότερα της έκφρασης του προσώπου τους. Υπάρχει μία μακρά ιστορία ενδιαφέροντος για το πρόβλημα της αναγνώρισης των εκφράσεων του προσώπου ενός ανθρώπου, και έχουν παραχθεί εκτεταμένες μελέτες στην αντίληψη του προσώπου κατά τη διάρκεια των τελευταίων είκοσι χρόνων. Η προσέγγιση στην αναπαράσταση των συναισθημάτων που υιοθετείται στο παρόν κεφάλαιο βασίζεται στο χώρο ενεργοποίησης - συναισθήματος

[13], που αποτελεί ένα αποτελεσματικό εργαλείο, ταυτόχρονα απλό και ικανό να συλλάβει μία ευρεία γκάμα σημαντικών θεμάτων σχετικά με τη συμπεριφορά. Τα εργαλεία που δημιουργούν την αντιστοίχιση μεταξύ των χαρακτηριστικών του προσώπου, τις εκφράσεις του προσώπου και τελικά των συναισθημάτων παρέχονται από του πλαίσιο που έχει θέσει το πρότυπο MPEG-4. Το τελευταίο ορίζει μία αντιστοίχιση μεταξύ των παραμέτρων δυναμικής κίνησης των ομοιωμάτων των προσώπων FAPs (Facial Animation Parameters) και των κινήσεων συγκεκριμένων χαρακτηριστικών σημείων (FPs) των παραμέτρων ορισμού του προσώπου FDP (Facial Definition Parameter) [58].

Η Ενότητα 5.3 περιέχει μεθόδους για την αλληλεπίδραση μεταξύ των πραγματικών και εικονικών στοιχείων μίας σκηνής, μέσω της γνώσης που αποκτάται σχετικά με τη θέση τους, μαζί με τα δεδομένα που εξάγονται από τις εκφράσεις ενός ανθρώπου. Παράλληλα, παρουσιάζονται πειραματικά αποτελέσματα για τις μονάδες του συστήματος, ενώ η ολοκλήρωση του τελευταίου εξετάζεται υπό ένα βασικό, γενικό σενάριο.

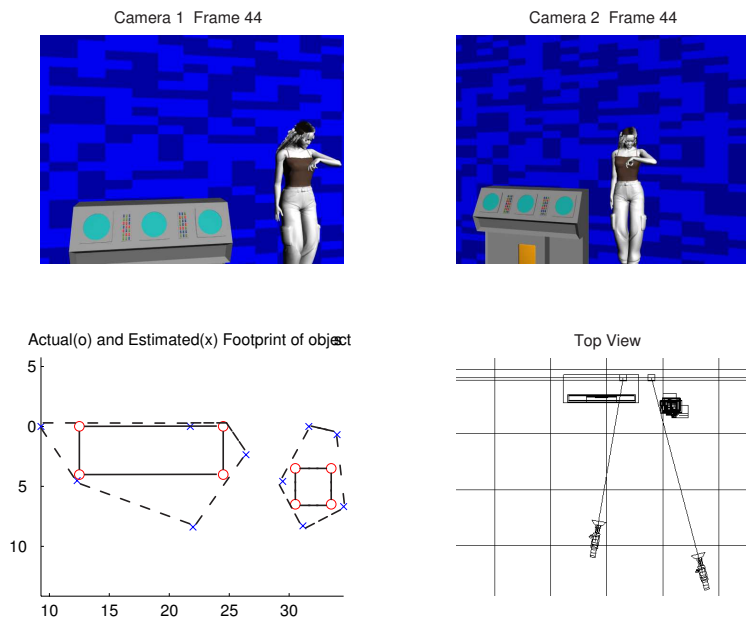
5.2 Εξαγωγή χαρακτηριστικών για την εκτίμηση των εκφράσεων του προσώπου

Για τη διευκόλυνση στην ενσωμάτωση των εικονικών ηθοποιών σε μία σκηνή ανάμεικτης πραγματικότητας (mixed reality), πρέπει να σχεδιαστεί ένα σύστημα αποφάσεων. Βάσει της διαδικασίας που περιγράφηκε στην προηγούμενη υποενότητα, το σύστημα αυτό μπορεί να παρέχει πληροφορία και να λαμβάνει αποφάσεις, χωρίς ανθρώπινη παρέμβαση, ώστε να καθορίζει την κίνηση ενός εικονικού ανθρώπου. Ο στόχος αυτής της μονάδας όμως είναι ο εμπλουτισμός του εικονικού ηθοποιού, ώστε αυτός να έχει τη δυνατότητα να αντιληφθεί και να επεξεργαστεί πληροφορίες που προέρχονται από το πρόσωπο του πραγματικού ηθοποιού και, ακολούθως, να επικοινωνήσει την αντίδρασή του. Αυτό απαιτεί μία ικανοποιητική καταγραφή του προσώπου του ανθρώπου ώστε να αποτιμηθεί η έκφραση του προσώπου του.

Η τεχνική εξαγωγής των χαρακτηριστικών του προσώπου, σε αυτό το σύστημα, βασίζεται, όπως αναλύθηκε στο Κεφάλαιο 4, σε μία ιεραρχική, εύρωστη μεθοδολογία, που αντιμετωπίζει επιτυχώς μεγάλες αποκλίσεις στην εμφάνιση διάφορων προσώπων, καθώς και στο ίδιο πρόσωπο σε διαφορετικές του εκφάνσεις, εντός πραγματικών ακολουθιών εικόνων. Ελαστικές υποθέσεις γίνονται όσον αφορά στον προσανατολισμό (pose) του προσώπου καθώς και στη γενική θέση των χαρακτηριστικών πάνω σε αυτό. Υποστηρίζεται σταδιακή αποκάλυψη της πληροφορίας σχετικά με το πρόσωπο με στόχο την βελτιστοποίηση σε κάθε βήμα της ιεραρχικής μεθοδολογίας, παράγοντας εκ των υστέρων γνώση και οδηγώντας σε μία σταδιακή οπτικοποίηση των χαρακτηριστικών που αναζητούνται.

5.3 Ολοκλήρωση του συστήματος - Ένα γενικό σενάριο

Ο εντοπισμός των ηθοποιών στο εικονικό σκηνικό παρέχει τη δυνατότητα ανάπτυξης ποικίλων εφαρμογών, που αφορούν στην αλληλεπίδραση τους με τον εικονικό κόσμο. Εκτός των προφανών, που περιλαμβάνουν την ελεύθερη κίνηση των ηθοποιών γύρω από ψηφιακά αντικείμενα ή την τοποθέτηση τέτοιων αντικειμένων γύρω τους, μπορούμε να κάνουμε ένα επιπλέον βήμα και να εξεταστούν οι πιθανότητες δημιουργίας ενός εικονικού ηθοποιού που θα έχει την ικανότητα να κινείται σωστά στο σκηνικό. Επιπρόσθετα, θα πρέπει να έχει την ικανότητα να αλληλεπιδρά με έναν πραγματικό



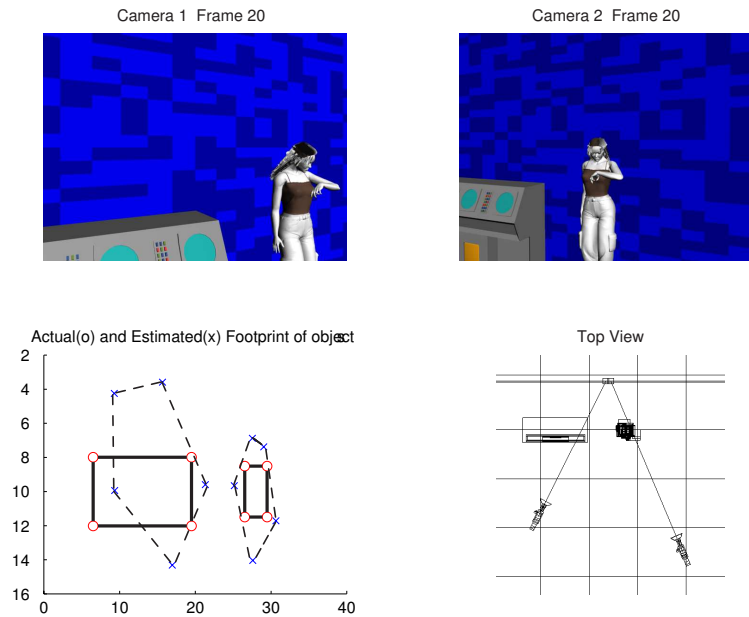
Σχήμα 5.1: Προσομοίωση εικονικού στούντιο: εκτίμηση θέσης στο επίπεδο του αντικειμένου και του ηθοποιού

ηθοποιό, ελαχιστοποιώντας την ανάγκη για εξωτερική ανθρώπινη παρέμβαση. Για να σχηματιστεί πιο σαφής εικόνα του προβλήματος, θα σχεδιαστεί και θα υλοποιηθεί στην παρούσα ενότητα ένα σχετικά απλό σενάριο, ώστε να εξεταστούν οι εκάστοτε δυσκολίες και οι τεχνικές αντιμετώπισής τους. Για το σκοπό αυτό θα στηθεί μία προσομοίωση ενός εικονικού στούντιο και θα ακολουθηθεί ένα σενάριο που θα λάμβανε χώρα στον πραγματικό κόσμο.

Όπως και στην Ενότητα 2.6.1 του Κεφαλαίου 2, κατασκευάζεται ένα κυανό πέτασμα δύο τόνων σε ένα εικονικό περιβάλλον χρησιμοποιώντας ένα κατάλληλο εμπορικό λογισμικό. Δύο εικονικές κάμερες χρησιμοποιούνται για να αποδώσουν τα αντίστοιχα τμήματα του πετάσματος, για γνωστές παραμέτρους κίνησης, δηλαδή άξονα και γωνία περιστροφής, 3Δ μετατόπιση και εστιακή απόσταση. Βάσει των καταγραφόμενων ακολουθιών η κίνηση των καμερών υπολογίζεται μέσω του αλγορίθμου που παρουσιάστηκε και επαληθεύεται ότι τα αποτελέσματα είναι σύμφωνα με τη θεωρία του Κεφαλαίου 2. Στη συνέχεια, τοποθετούνται αντικείμενα σε προκαθορισμένες θέσεις μπροστά από το πέτασμα, τα οποία θα χρησιμεύσουν ως υποκατάστατα για τα φυσικά αντικείμενα και τους πραγματικούς ηθοποιούς.

Το Σχήμα 5.1 απεικονίζει την προσομοίωση του εικονικού στούντιο, όπου ένας ηθοποιός (εικονικός άνθρωπος που συμβολίζει τον πραγματικό ηθοποιό) και ένα μηχάνημα κινηματογραφούνται μπροστά από ένα κυανό πέτασμα. Τα καρέ που καταγράφονται από δύο από τις κάμερες του στούντιο φαίνονται μαζί με μία κάτοψη του σκηνοθέτη. Τέλος, παρουσιάζεται η έξοδος του συστήματος που παρέχει τις εκτιμήσεις για την περιοχή που καλύπτουν ο ηθοποιός και το μηχάνημα. Τα αποτελέσματα που λαμβάνονται είναι αποτελεσματικά, μη απαιτώντας περαιτέρω επεξεργασία, καθώς ο σκηνοθέτης μπορεί να τοποθετήσει με ασφάλεια ένα εικονικό αντικείμενο μπροστά, πίσω ή ενδιάμεσα των δύο πραγματικών αντικειμένων.

Στο Σχήμα 5.2 παρουσιάζεται η περίπτωση κατά την οποία ένα από τα φυσικά αντικείμενα επικαλύπτεται μερικά ή δεν καταγράφεται εξ' ολοκλήρου στα καρέ και των δύο καμερών. Είναι εμφανές ότι, παρά το ότι η εκτίμηση για την περιοχή που καλύ-

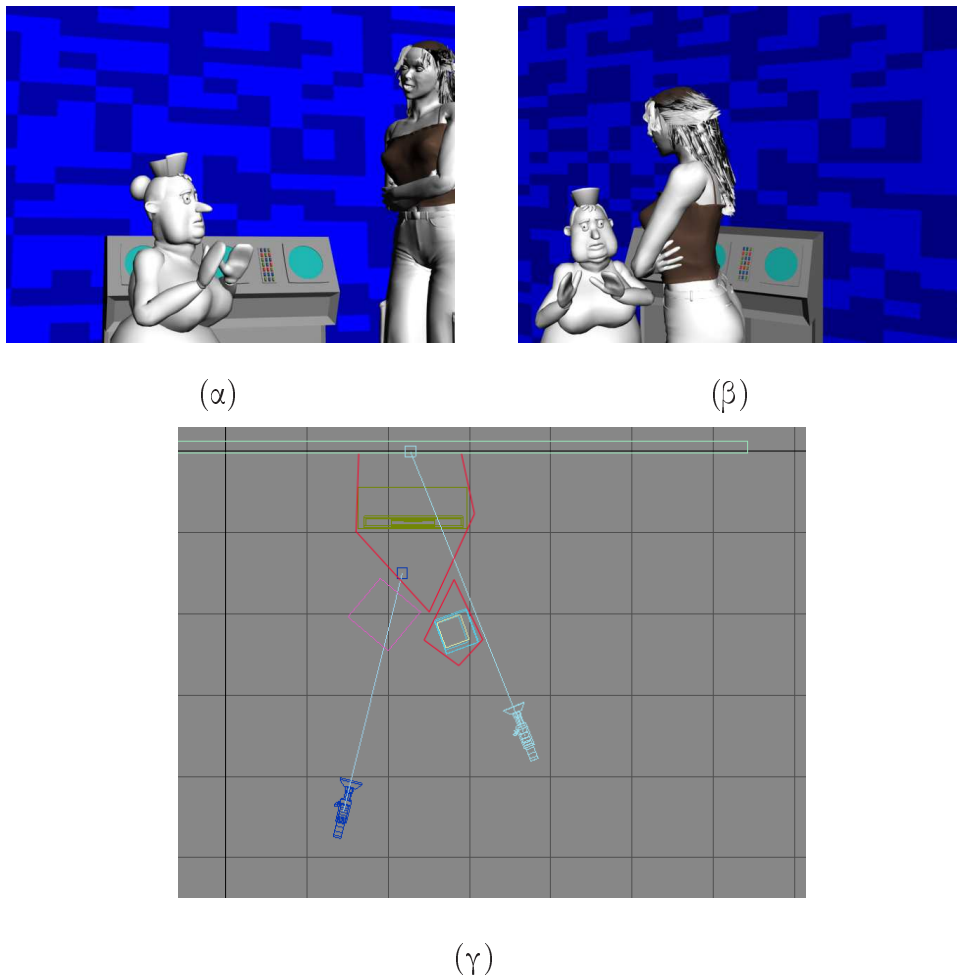


Σχήμα 5.2: Εκτίμηση θέσης στο επίπεδο του αντικειμένου και του ηθοποιού όταν το αντικείμενο δεν παρουσιάζεται ολόκληρο στα καρέ που καταγράφονται

πει το αντικείμενο δεν περιλαμβάνει όλο το αντικείμενο, το αποτέλεσμα επιτρέπει την ορθή προσθήκη και τον χειρισμό όποιων εικονικών αντικειμένων χρειαστούν. Αυτό συμβαίνει καθώς μας ενδιαφέρει το ορατό μέρος των αντικειμένων που ενσωματώνονται στη σκηνή και το οποίο, λόγω της εκτίμησης, δε είναι δυνατό να τοποθετηθεί σε αναντιστοιχία με τα φυσικά αντικείμενα. Είναι επίσης αρκετά λογικό να υποθεθεί ότι το επόμενο βήμα προς την βελτίωση του συστήματος θα πρέπει να περιλαμβάνει μία μονάδα που θα παρακολουθεί στο χρόνο τα αντικείμενα που κινούνται στο σκηνικό, ενισχύοντας τις εκτιμήσεις που γίνονται για αυτά κατά την κινηματογράφηση. Αυτές θα πρέπει να περιλαμβάνουν την περιοχή που καταλαμβάνει κάθε αντικείμενο, την κατεύθυνσή του, την ταχύτητα και το εμβαδό της ορατής του επιφάνειας, δημιουργώντας έτσι ένα σύνολο εκτιμήσεων που θα εξάγονται πριν την επεξεργασία του τρέχοντος καρέ.

Με τα ληφθέντα δεδομένα, ένας σκηνοθέτης μπορεί να τοποθετήσει ένα οποιοδήποτε εικονικό αντικείμενο στον διαθέσιμο χώρο και να ελέγξει την θέση του σε σχέση με τον πραγματικό κόσμο. Είναι βέβαια ενδιαφέρον να εξεταστεί η περίπτωση κατά την οποία αυτό το αντικείμενο είναι ένας εικονικός άνθρωπος που θα πρέπει να κινείται και να αλληλεπιδρά. Όπως φαίνεται στο Σχήμα 5.3, αυτό είναι εφικτό σε μία αντιμετώπιση καρέ-καρέ, όπου ο εικονικός ηθοποιός (συγκεκριμένα η νοσοκόμα) εισάγεται στην σκηνή. Όμως, για να δημιουργηθεί ένα βίντεο ή μία ζωντανή εκπομπή, αυτός ο ηθοποιός θα πρέπει να αντιδρά σε ερεθίσματα του πραγματικού κόσμου και να κινείται αντίστοιχα στον ελεύθερο χώρο αντί να επανατοποθετείται διαρκώς. Για το σκοπό αυτό έχει υλοποιηθεί μία πρότυπη διαδικασία ελέγχου που, δεδομένων των εκτιμήσεων για τις περιοχές που καλύπτουν τα πραγματικά αντικείμενα, αποφασίζει κατάλληλα για την ύπαρξη ανάγκης να κινηθεί ο εικονικός ηθοποιός ενώ, επίσης αποφασίζει και για την κατάλληλη κατεύθυνση.

Τέλος, η παρούσα εργασία εστιάζει στην ανάπτυξη ενός ευφυούς συστήματος που θα μπορεί να ελέγχει τις αντιδράσεις του εικονικού ηθοποιού, παρέχοντας ταυτόχρονα τα μέσα για να αλληλεπιδράσει αυτός με τους αντίστοιχους πραγματικούς. Αυτό



Σχήμα 5.3: Σκηνή στην οποία έχει εισαχθεί αυτόματα εικονικός ηθοποιός (νοσοκόμα). Στο (γ) φαίνεται η εκτίμηση για το χώρο που καταλαμβάνουν τα φυσικά αντικείμενα, καθώς και η θέση στην οποία τοποθετείται ο εικονικός ηθοποιός.

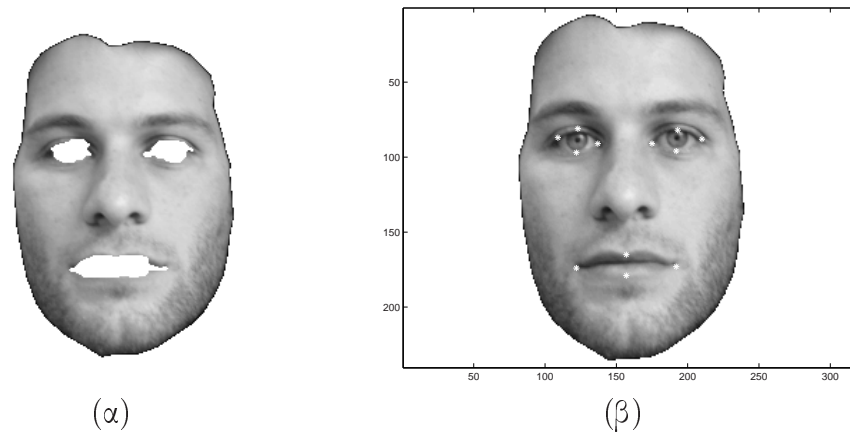


Σχήμα 5.4: Είσοδος για την αναγνώριση έκφρασης του προσώπου.

επιτυγχάνεται, σε πρώτο στάδιο, με την αναγνώριση εκφράσεων του προσώπου από το καταγραφόμενο καρέ μίας κάμερας που περιέχει μία όψη του προσώπου (για παράδειγμα βλέπε Σχήμα 5.4), βάσει ανάλυσης της εικόνας. Στο σενάριο που εξετάζεται στην παρούσα ενότητα, το καρέ που χρησιμοποιείται ως είσοδος για την εκτίμηση της έκφρασης του προσώπου, καταγράφεται από μία τρίτη κάμερα που παρακολουθεί αποκλειστικά το πρόσωπο του ηθοποιού. Ανάλογα με το στήσιμο και τον εξοπλισμό του εκάστοτε στούντιο, καθώς και της ανάγκης για επεξεργασία σε πραγματικό χρόνο, μία τέτοια εικόνα μπορεί να προκύψει μέσω κατάκτησης των καρέ που καταγράφονται από τις ήδη υπάρχουσες κάμερες. Όπως έχει προταθεί και στο Κεφάλαιο 4, τα χαρακτηριστικά του προσώπου του ηθοποιού αντιστοιχούν σε συγκεκριμένες εκφράσεις οι οποίες είναι δυνατόν να εξαχθούν. Πρέπει να σημειωθεί στο σημείο αυτό ότι, το πρόσωπο του εικονικού ανθρώπου που αναπαριστά τον πραγματικό ηθοποιό στο συγκεκριμένο πείραμα (Σχήμα 5.4) αντικαθίσταται στη συνέχεια από ένα πραγματικό πρόσωπο (Σχήμα 5.5) ώστε να παρουσιαστεί αποτελεσματικά η προτεινόμενη μεθοδολογία.

Στο Κεφάλαιο 4 παρουσιάζεται μία προσέγγιση σταδιακής εμπιστοσύνης για την εξαγωγή των χαρακτηριστικών του προσώπου σε πραγματικά καρέ βίντεο. Το πρόβλημα αντιμετωπίζεται υπό γενικευμένες συνθήκες απεικόνισης και με την υιοθέτηση ελαφρών υποθέσεων. Τα βασικά χαρακτηριστικά του προσώπου, όπως τα μάτια και το στόμα, θεωρούνται ως σημαντικές ασυνέχειες σε ένα πρόσωπο που έχει περιστραφεί αυθαίρετα. Οι ευδιάκριτες περιοχές του προσώπου αυτού σχηματίζουν ένα αρχικό σύνολο περιοχών που πιθανόν περιέχουν πληροφορία για τα χαρακτηριστικά που μας ενδιαφέρουν. Η επεξεργασία των περιοχών αυτών παράγουν κλειστά αντικείμενα (βλέπε Σχήμα 5.5(α)) και η επιλογή γίνεται βάσει κριτηρίων που γενούν εκ των υστέρων γνώση για το πρόσωπο, όπως οι επικρατούσες γωνίες των υποψήφιων χαρακτηριστικών καθώς και θεωρήσεις σχετικά με τη συμμετρία.

Η προτεινόμενη διαδικασία παράγει, σύμφωνα με το πρότυπο MPEG-4, παραμέτρους feature definition(FDP) και animation (FAP), οι οποίες χρησιμοποιούνται για την ανάλυση της έκφρασης και την αναγνώριση των συναισθημάτων (βλέπε Σχήμα 5.5(β)). Παρακολουθώντας τις κινήσεις του προσώπου που αντιστοιχούν σε κινήσεις των FDP και / ή των FAP στο χρόνο, είναι δυνατόν να εξαχθούν στοιχεία για τις εκφράσεις του ανθρώπου καθώς και για τα συναισθήματά του. Όπως παρουσιάζεται και στην εργασία [43], έχει αναπτυχθεί μία μέθοδος μετατροπής της ποσοτικής πληροφορίας των χαρακτηριστικών σε μία εκτίμηση της έκφρασης του προσώπου. Αυτός ο μετασχηματισμός βασίζεται σε μία αρχιτεκτονική ασαφούς λογικής που λαμβάνει υπόψη τη γνώση για την αναπαράσταση και τα φυσικά χαρακτηριστικά της ανθρώπινης



Σχήμα 5.5: Ανάλυση εικόνας με στόχο την αναγνώριση της έκφρασης του προσώπου (α) Περιοχές ενδιαφέροντος για εξαγωγή χαρακτηριστικών του προσώπου (β) Ένα υποσύνολο των παραμέτρων προσδιορισμού των χαρακτηριστικών του προσώπου (*FDP*)

έκφρασης. Ακολουθώντας αυτή την προσέγγιση και ενσωματώνοντας τη στο προτεινόμενο εικονικό στούντιο, καθιστά δυνατό τον εμπλουτισμό τέτοιων παραγωγών, διευκολύνοντας τη χρήση των εικονικών ηθοποιών και δημιουργώντας ενδιαφέρουσες αλληλεπιδράσεις με τους πραγματικούς ηθοποιούς. Συγκεκριμένα στο παράδειγμα μας, θα μπορούσε να χρησιμοποιηθούν οι παράμετροι προσδιορισμού των χαρακτηριστικών του προσώπου (*FDP*) που εξάχθηκαν για τη δυναμική κίνηση των χαρακτηριστικών του προσώπου του εικονικού ηθοποιού. Εναλλακτικά, θα μπορούσαμε να προχωρήσουμε περαιτέρω στην ανάλυση της έκφρασης του προσώπου βάσει των *FDP*, εξάγοντας συμπεράσματα σχετικά με τη συμπεριφορά του ανθρώπου, και έτσι να οδηγήσουμε τον εικονικό ηθοποιό να επιλέξει μία από τις προκαθορισμένες του αντιδράσεις.

5.4 Συμπεράσματα και επεκτάσεις

Στο παρόν κεφάλαιο, παρουσιάστηκε ένα προηγμένο οπτικό εικονικό στούντιο που περιλαμβάνει ένα σύστημα εκτίμησης της κίνησης της κάμερας βάσει ενός κυανού πετάσματος δύο χρωματικών τόνων. Περιγράφηκε η μεθοδολογία για την παρακολούθηση των ηθοποιών και των αντικειμένων που βρίσκονται και κινούνται στο σκηνικό, επιτρέποντας την αποτελεσματική προσθήκη επιπλέον, εικονικών, αντικειμένων στις σκηνές. Τέλος προτάθηκε μία προσέγγιση για τον εμπλουτισμό της αλληλεπίδρασης μεταξύ πραγματικών και εικονικών ηθοποιών, συμπεριλαμβανόμενου ενός πλαισίου εργασίας για την αναγνώριση των εκφράσεων του προσώπου βάσει της εξαγωγής των χαρακτηριστικών του. Τα αποτελέσματα των πειραμάτων προσομοίωσης απεικονίζουν την αποδοτικότητα του ολοκληρωμένου συστήματος.

Στις μελλοντικές επιδιώξεις βρίσκεται η ολοκλήρωση ενός πραγματικού συστήματος που θα συνδέει τις υπάρχουσες μονάδες διαχωρισμού χρώματος και σύνθεσης με τις προτεινόμενες τεχνικές. Με αυτό τον τρόπο, η προτεινόμενη προσέγγιση θα δοκιμαστεί σε ένα πραγματικό εμπορικό στούντιο. Επιπρόσθετα η ολοκλήρωση θα επιτρέψει τη δημιουργία ενός πλήρως αυτοματοποιημένου συστήματος επιτρέποντας και το σχεδιασμό μίας μονάδας ελέγχου και αντιμετώπισης λαθών. Από θεωρητικής πλευράς, έχει ενδιαφέρον να εξεταστεί κατά πόσον η χρήση των εκτιμήσεων για τη θέση και την κίνηση των αντικειμένων μπορεί να επιτρέψει την επαλήθευση των εξαχθέντων παραμέτρων 3Δ κίνησης της κάμερας. Επίσης θα πρέπει να προστεθούν επιπλέον δε-

δομένα στην μεθοδολογία εκτίμησης της συμπεριφοράς του ατόμου όπως ανάλυση του λόγου και των χειρονομιών των ηθοποιών.

□

Κεφάλαιο 6

Συμπεράσματα και Επεκτάσεις

Στο κυρίως τμήμα της παρούσας διατριβής, παρουσιάστηκε η διαδικασία σχεδιασμού και υλοποίησης ενός εμπλουτισμένου εικονικού στούντιο, για την παραγωγή ακολουθιών εικόνων επαυξημένης πραγματικότητας. Αρχικά, πραγματοποιήθηκε η επέκταση του παραδοσιακού κυανού πετάσματος, ώστε να περιέχει δύο χρωματικές αποχρώσεις, επιτρέποντας την αποτελεσματική εξαγωγή των παραμέτρων 3Δ κίνησης των καμερών, χωρίς τη χρήση επιπλέον ηλεκτρομηχανικού εξοπλισμού. Η μεθοδολογία που προτάθηκε και τεκμηριώθηκε επιτρέπει την κατασκευή ενός κυανού πετάσματος, αυθαιρέτου μεγέθους, και την εξαγωγή των παραμέτρων μίας κάμερας που το καταγράφει. Στη συνέχεια, εκμεταλλευόμενοι τις ιδιότητες του συστήματος αυτού, επιτυγχάνεται ο εντοπισμός και η παρακολούθηση των αντικειμένων και των ηθοποιών που κινούνται στο σκηνικό, με στόχο την αυτοματοποίηση και τον έλεγχο της εισαγωγής επιπλέον εικονικών στοιχείων, εκτός του εικονικού μοντέλου του κόσμου που περιβάλλει το σκηνικό. Η παραγωγή τέτοιων σύνθετων ακολουθιών μεικτής πραγματικότητας αποτελεί την αιχμή της τεχνολογίας στις παραγωγές βίντεο και αποσκοπεί στην όσο το δυνατόν πιο αυθόρμητη ενσωμάτωση του χρήστη στο περιβάλλον επαυξημένης πραγματικότητας. Η αναζήτηση για την καλύτερη αλληλεπίδραση μεταξύ ηθοποιών και εικονικών ηθοποιών οδήγησε στην διερεύνηση του προβλήματος εξαγωγής των χαρακτηριστικών του ανθρώπινου προσώπου και της περαιτέρω επεξεργασίας των αποτελεσμάτων ώστε να παραχθούν εκτιμήσεις για τη συμπεριφορά του και ειδικότερα για τις εκφράσεις του προσώπου του. Με τον τρόπο αυτό, δίνεται η δυνατότητα για αλληλεπίδραση στο εικονικό στούντιο περιορίζοντας τη χρήση προκαθορισμένων κινήσεων, χειρονομιών ή διαλόγων.

Το αντικείμενο της εργασίας εξελίσσεται σχεδόν ταυτόχρονα με την έρευνα στους αντίστοιχους τομείς λόγω της επιτυχίας των αποτελεσμάτων του. Στο σύνολο της, επιχειρήσαμε να παρουσιάσουμε μεθόδους πλήρως βασισμένες στην πληροφορία που λαμβάνεται από την εικόνα που καταγράφεται, απεγκλωβίζοντας έτσι την παραγωγή από τη χρήση επιπλέον ηλεκτρομηχανικού εξοπλισμού. Παράλληλα, κύριο ενδιαφέρον αποτελεί η τεκμηρίωση και μαθηματική στήριξη των μεθόδων που παρουσιάζονται ώστε να προσφέρουν τη βάση για την ανάπτυξη τέτοιων συστημάτων. Όπως ήταν αναμενόμενο, αντιμετωπίσαμε πληθώρα προβλημάτων λόγω της αδυναμίας υλοποίησης ενός τέτοιου συστήματος σε πραγματική κλίμακα και με το σύνολο των δυνατοτήτων που αυτό παρέχει. Όμως, τα αποτελέσματα των προσομοιώσεων, των πειραμάτων σε μικρότερης κλίμακας υλοποίηση αλλά και της ανάλυσης σφάλματος όταν αυτή χρειάστηκε, δίνουν τις ενδείξεις για τις δυνατότητες των τεχνικών. Άλλωστε, τμήματα ή και το σύνολο των προτεινόμενων μεθοδολογιών επιχειρείται να ενσωματωθούν στα

πραγματικά συστήματα παραγωγής που κατασκευάζονται από εταιρίες που δραστηριοποιούνται στο χώρο της ψυχαγωγίας.

Στόχος για τις μελλοντικές επεκτάσεις του συστήματος αποτελεί η υλοποίηση ενός πλήρως αυτοματοποιημένου συστήματος αλληλεπίδρασης μεταξύ των ανθρώπων και των εικονικών ηθοποιών, ώστε να παρέχει και μία πλατφόρμα δοκιμών και εξέλιξης των τεχνικών αυτών. Το ενδιαφέρον όμως παραμένει στις ερευνητικές επεκτάσεις που παρουσιάζονται και έχουν να κάνουν με την αντιμετώπιση συνθηκών πραγματικού χρόνου και οριακών συνθηκών καταγραφής. Συγκεκριμένα, εξετάζεται η αποδοτική εξαγωγή των παραμέτρων κίνησης της κάμερας υπό οριακές συνθήκες, όταν για παράδειγμα παρουσιάζεται πτώση της ποιότητας της καταγραφόμενης εικόνας λόγω ανεστίασης ή πολύ γρήγορης κίνησης ή όταν το προοπτικό μοντέλο προβολής εκφυλίζεται σε παράλληλο. Επίσης εξετάζονται προβλήματα που έχουν να κάνουν με την ταχύτητα απόκρισης του συστήματος, και τεχνικές όπως παρεμβολή μεταξύ διαδοχικών καρτέ και εκτιμήσεις για την κίνηση και το χώρο που θα καταλάβει ένα αντικείμενο σε μελλοντικά καρτέ. Η αξιοποίηση της περιοχής της όρασης υπολογιστών και της κατανόησης του περιεχομένου ακολουθιών εικόνων σε συγκεκριμένες συνθήκες παράγει αξιοσημείωτα ερευνητικά αποτελέσματα και, κυρίως, εμφανίζει τις δυσκολίες τελειοποίησης τέτοιων υλοποιήσεων, οδηγώντας στην προσπάθεια συνολικής τεκμηρίωσης των εφαρμοζόμενων τεχνικών καθώς και των προτεινόμενων βελτιώσεων.

□

Βιβλιογραφία

- [1] AHLBERG, J., AND LI, H. Representing and Compressing Facial Animation Parameters Using Facial Action Basis Functions. *IEEE Trans. Circuits and Systems for Video Technology* 93, 3 (Apr. 1999), 405–411.
- [2] AVRITHIS, U., N. T., AND KOLLIAS, S. Broadcast News Parsing Using Visual Cues: A Robust Face Detection Approach. In *IEEE International Conference on Multimedia and Expo* (New York City, USA, July 2000).
- [3] BALOMENOS, T., ET AL. Emotion Analysis In Man-Machine Interaction Systems. In *Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI04)* (Geneva, Switzerland, June 2004).
- [4] BAZZONI, G., ET AL. The ORIGAMI project: advanced tools and techniques for high-end mixing and interaction between real and virtual content. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on* (June 2002), pp. 306–312.
- [5] BLACK, M., Y. Y. A. J., AND FLEET, D. Learning Parameterized Models of Image Motion. In *Proc. of IEEE CVPR* (1997), pp. 561–567.
- [6] BLONDE, L., ET AL. A Virtual Studio for Live Broadcasting: The Mona Lisa Project. *IEEE Multimedia* 3, 2 (1996), 18–29.
- [7] BOUMAN, C., AND SHAPIRO, M. A Multiscale Random Field Model for Bayesian Image Segmentation. *IEEE Trans. Image Processing* 3, 2 (Mar. 1994), 162–177.
- [8] BRUNELLI, R., AND POGGIO, T. Face Recognition: Features versus Templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 15, 10 (Oct. 1993), 1042–1052.
- [9] CAVAZZA, M., ET AL. Multi-modal Acting in Mixed Reality Interactive Storytelling. *IEEE Multimedia* 11, 3 (2004).
- [10] CHELLAPPA, R., C. L. W., AND SIROHEY, S. Human and Machine Recognition of Faces: A Survey. *Proceedings of the IEEE* 83, 5 (May 1995), 705–740.
- [11] COHEN, L. On Active Contour Models and Balloons. In *Proc, CVGIP: Image Understanding* (Mar. 1991), pp. 211–218.
- [12] COMPORT A., E. M., AND CHAUMETTE, F. A Real-time Tracker for Markerless Augmented Reality. In *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'03)* (Tokyo, Japan, Oct. 2003), pp. 36–45.

- [13] COWIE, R., ET AL. Emotion Recognition in Human Computer Interaction. *IEEE Signal Processing Magazine* 18, 1 (Jan. 2001).
- [14] DARRELL, T., AND PENTLAND, A. Recognition of Space-Time Gestures using Distributed Representation. Τεξη. Ρεπ. 197, M.I.T. Media Laboratory, 1993.
- [15] DELOPOULOS, A., AND XIROUHAKIS, Y. Robust Estimation of Motion and Shape based on Orthographic Projections of Rigid Objects. In *IEEE Image and Multidimensional Digital Signal Processing Workshop (IMDSP'98)* (Alpbach, Austria, July 1998), pp. 151–154.
- [16] DELOPOULOS, A., M. R. Y. X., AND DROSOPOULOS, A. Audio Pattern Recognition and Localization in Continuous Playout. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing* (Baltimore, Maryland, USA, June 2001).
- [17] DELBIMBO, A., AND PALA, P. Visual Image Retrieval by Elastic Matching of User Scetches. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19, 2 (Feb. 1997), 121–132.
- [18] DELIGNON, Y., A. M., AND PIECZYNSKI, W. Estimation of Generalized Mixtures and Its Application in Image Segmentation. *IEEE Trans. Image Processing* 6, 10 (1997), 1364–1376.
- [19] DENG, Y., AND MANJUNATH, B. Unsupervised Segmentation of Color-Texture Regions in Images and Video. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23, 8 (Aug. 2001).
- [20] DROSOPOULOS, A., Y. X., AND DELOPOULOS, A. An Optical Camera Tracking System for Virtual Sets Applications. In *Visual, Modeling and Visualization Workshop (VMV'99)* (Erlangen, Germany, Nov. 1999).
- [21] DROSOPOULOS, A., Y. X., AND DELOPOULOS, A. Optical Camera Tracking in Virtual Studios: Degenerate Cases. In *International Conference on Pattern Recognition (ICPR 2000)* (Barcelona, Spain, Sept. 2000).
- [22] DROSOPOULOS, ., ET AL. Emotionally-rich Man-machine Interaction Based on Gesture Analysis. In *Human-Computer Interaction International 2003* (Crete, Greece, June 2003).
- [23] DROSOPOULOS, A., AND KOLLIAS, S. Object Localization in a Virtual Studio Environment. In *International Conference From Scientific Computing to Computational Engineering, IC-SCCE* (Athens, Greece, Sept. 2004).
- [24] DROSOPOULOS, A., AND KOLLIAS, S. Enhanced Interaction in an Optical Virtual Studio Using Object Tracking and Facial Analysis. *submitted for publication* (2005).
- [25] EISERT, P., AND GIROD, B. Model-Based Estimation of Facial Expression Parameters from Image Sequences. In *Proc. of IEEE International Conference on Image Processing* (1997).

- [26] ESSA, I., AND PENTLAND, A. Coding, Analysis, Interpretation and Recognition of Facial Expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19, 7 (July 1997), 757–763.
- [27] FAUGERAS, O. *Three-Dimensional Computer Vision*. MIT Press, New York, 1993.
- [28] GALLAGER, R. G. *Information Theory and Reliable Communication*. John Wiley & Sons Inc., 1968.
- [29] GEMAN, S., AND GEMAN, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 6 (1984), 721–741.
- [30] GIBBS, S., ET AL. Virtual Studios: An overview. *IEEE Multimedia* 5, 1 (1998), 18–35.
- [31] GOLOMB, S. W. *Shift Register Sequences*. Holden-Day, New York, 1967.
- [32] GRAU, O., T. P., AND THOMAS, G. A. A Combined Studio Production System for 3-D Capturing of Live Action and Immersive Actor Feedback. *IEEE Trans. Circuits Syst. Video Technol.* 14, 3 (2004), 370–380.
- [33] HAYASHI, M. Image Compositing Based on Virtual Cameras. *IEEE Multimedia* 5, 1 (1998), 36–48.
- [34] HARMON, L., ET AL. Machine Identification of Human Faces. *Pattern Recognition* 13, 2 (1981), 97–110.
- [35] HOETTER, M. Differential Estimation of the Global Motion Parameters Zoom and Pan. *Signal Processing* 16, 3 (1989), 249–265.
- [36] HUGHES, D. Virtual-studio-ultimate 8. In *IRT Symp. Virtual Studio Technique* (Munich, Germany, Jan. 1996).
- [37] HUANG, T. S., AND NETRAVALI, A. N. Motion and Structure from feature correspondences: A review. *Proceedings of IEEE* 82 (Feb. 1994), 252–269.
- [38] IGLESIAS, A., AND LUENGO, F. A new based-on-artificial-intelligence framework for behavioral animation of virtual actors. In *Computer Graphics, Imaging and Visualization, 2004. CGIV 2004. Proceedings. International Conference on* (July 2004), pp. 245–250.
- [39] JAIN, A.K., Y. Z., AND LAKSHMANAN, S. Object Matching Using Deformable Templates. *IEEE Trans. Pattern Analysis and Machine Intelligence* 18, 3 (Mar. 1996), 267–278.
- [40] JIA, X., AND NIXON, M. Extending the Feature Vector for Automatic Face Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 17, 12 (Dec. 1995), 1167–1176.
- [41] KANADE, T. Picture Processing by Computer Complex and Recognition of Human Faces. *Tech. Rep., Dept. of Information Science, Kyoto University* (1973).

- [42] KARPOUZIS, K., G. V. N. T., AND KOLLIAS, S. Compact 3D Model Generation based on 2D Views of Human Faces: Application to Face Recognition. *Machine Graphics and Vision* 7 (1998), 75–85.
- [43] KARPOUZIS, K., ET AL. Facial and Body Feature Extraction for Emotionally-Rich HCI. In *3D Modeling and Animation*, N. Sarris and M. G. Strintzis, Ed. Idea Group, 2004.
- [44] KARPOUZIS, K., ET AL. Facial and Body Feature Extraction for Emotionally-Rich HCI. In *Encyclopedia of Information Science and Technology*, Mehdi Khosrow-Pour, Ed. Idea Group Publishers, 2004.
- [45] KASS, M., A. W., AND TERZOPOULOS, D. Snakes: Active Contour Models. *International Journal of Computer Vision* 1, 4 (1988), 321–331.
- [46] KOLLIAS, S., AND PIAT, F. The PHYSTA project. In *International Conference on Multimedia Computing and Systems (ICMCS'99)* (Florence Italy, June 1999).
- [47] LANGAN, D.A., J. M., AND ZHANG, J. Cluster Validation for Unsupervised Stochastic Model-Based Image Segmentation. *IEEE Trans. Image Processing* 7, 2 (1997), 180–244.
- [48] LIU, Y., T. S. H., AND FAUGERAS, O. D. Determination of camera location from 2d to 3d line correspondences. *IEEE Trans. Pattern Analysis and Machine Intelligence* 12, 1 (Jan. 1990), 28–37.
- [49] LOW, B., AND IBRAHIM, M. A Fast and Accurate Algorithm for Facial Feature Segmentation. In *Proc. of IEEE International Conference on Image Processing* (1997).
- [50] LUENBERGER, D. *Liner and Nonlinear Programming*. Addison-Welsey, Reading, 1984.
- [51] MA, W., AND MANJUNATH, B. Edge Flow: a Framework of Boundary Detection and Image Segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (1997), pp. 744–749.
- [52] MATSUMOTO, Y., AND ZELINSKY, A. Real-Time Stereo Face Tracking System for Visual Human Interfaces. In *Proc. of IEEE Workshop on Real-Time Analysis and Tracking of Face and Gesture in Real-Time Systems* (Kerkyra, Greece, Sept. 1999).
- [53] MCINERNEY, T., AND TERZOPOULOS, D. Topologically Adaptable Snakes. In *Proc. International Conf. Computer Vision* (1995), pp. 840–845.
- [54] MCKENNA, S., ET AL. Tracking Facial Feature Points with Gabor Wavelets and Shape Models. In *Int. Conf on Audio- and Video-based Biometric Person Authentication* (1997).
- [55] 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 (1991), 278–305.

- [56] METAXAS, D. Deformable Model and HMM-Based Tracking, Analysis and Recognition of Gestures and Faces. In *Proc. of IEEE Workshop on Real-Time Analysis and Tracking of Face and Gesture in Real-Time Systems* (Kerkyra, Greece, Sept. 1999).
- [57] MILGRAM, P., AND KISHINO, F. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems E77-D*, 12 (1994), 1321–1329.
- [58] Overview of the MPEG-4 standard, 2002.
- [59] NIKOLAIDIS, A., AND PITAS, I. Facial Feature Extracion and Pose Determination. In *Proc. of the NOBLESSE Workshop on Nonlinear Model Based Image Analysis* (1998).
- [60] ORAD. <http://www.orad.co.il/>.
- [61] PANJWANI, D., AND HEALEY, G. Markov Random Field Models for Unsupervised Segmentation of Textured Color Images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 17, 10 (Oct. 1995), 939–954.
- [62] PAPOULIS, A. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Singapore, 1991.
- [63] PENTLAND, A., B. M., AND STARNER, T. View-Based and Modular Eigenspaces for Face Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition* (1994).
- [64] POTZSCH, M., N. K., AND VON DER MALSBURG, C. Improving Object Recognition by Transforming Gabor Filter Responses. *Network: Computation in Neural Systems* 7, 2 (May 1996), 341–347.
- [65] QUAN, L., AND KANADE, T. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19, 8 (Aug. 1997), 834–845.
- [66] RADEVA, P., AND MARTIA, E. Facial Features Segmentation by Model-Based Snakes. In *Int. Conference on Comp. Analysis and Image Processing* (Prague, 1995).
- [67] RAOUZAIIOU, A., N. T. K. K., AND KOLLIAS, S. Parameterized facial expression synthesis based on MPEG-4. *Eurasip Journal on Applied Signal Processing* 2002, 10 (2002), 1021–1038.
- [68] THE REFLECMEDIA WEBSITE. <http://www.reflecmedia.com/>, 2002.
- [69] SAMAL, A., AND IYENGAR, P. Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. *Pattern Recognition* 25, 1 (1992), 65–76.
- [70] SCLAROFF, S., AND LIU, L. Deformable Shape Detection and Description via Model-Based Region Grouping. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23, 5 (May 2001).

- [71] SENIOR, A. Recognizing Faces in Broadcast Video. In *Proc. of IEEE Workshop on Real-Time Analysis and Tracking of Face and Gesture in Real-Time Systems* (Kerkyra, Greece, Sept. 1999).
- [72] SHAFARENKO, L., M. P., AND KITTLER, J. Automatic Watershed-Segmentation of Randomly Textured Color Images. *IEEE Trans. Image Processing* 6, 11 (1997), 1530–1544.
- [73] SHI, J., AND MALIK, J. Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22, 8 (2000), 888–905.
- [74] SHAPIRO, L. S., A. Z., AND BRADY, M. 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision* 16 (1995), 147–182.
- [75] SOBOTTKA, K., AND PITAS, I. A Novel Method for Automatic Face Segmentation, Facial Feature Extraction and Tracking. *Signal Processing: Image Communication* 12 (1998), 263–281.
- [76] SPETSAKIS, M., AND ALIMONOS, J. Structure from motion using line correspondences. *International Journal of Computer Vision* 4 (1990), 171–183.
- [77] STAIB, L., AND DUNCAN, J. Boundary Finding with Parametrically Deformable Models. *IEEE Trans. Pattern Analysis and Machine Intelligence* 14, 11 (Nov. 1992), 1061–1075.
- [78] TEKALP, M. *Digital Video Processing*. Prentice Hall, Englewood Cliffs, NJ, 1995.
- [79] THOMAS, G.A., J. JIN, T. NIBLETT, AND C. URQUART. A versatile camera position measurement system for virtual reality TV production. In *International Broadcasting Convention (IBC 97)* (Amsterdam, Netherlands, Sept. 1997), pp. 284–289.
- [80] TIAN, Y., T. K., AND COHN, J. Multi-State Based Facial Feature Tracking and Detection, Tech. Rep. CMU-RI-TR-99-18. *Robotics Institute, Carnegie Mellon University* (Aug. 1999).
- [81] TOKLU, C., A. E., AND TEKALP, A. Two-Dimensional Mesh-Based Mosaic Representation for Manipulation of Video Objects with Occlusion. *IEEE Trans. Image Processing* 9, 9 (Sept. 2000), 1617–1630.
- [82] TORRE, R., S. B. P. F., AND THALMANN, D. Interaction between real and virtual humans: playing checkers. In *Proc. Eurographics Workshop On Virtual Environments 2000* (Amsterdam, Netherlands, June 2000).
- [83] TSAPATSOULIS, N., ET AL. Emotion Recognition and Synthesis based on MPEG-4 FAPs. In *MPEG-4 Facial Animation*, Igor Pandzic, R. Forchheimer, Ed. John Wiley and Sons, UK, 2002.
- [84] TSAPATSOULIS, N., Y. A., AND KOLLIAS, S. Facial Image Indexing in Multimedia Databases. *Pattern Analysis and Applications* 4, 2/3 (2001), 93–107.

- [85] G. TSECHPENAKIS, Y. X., AND DELOPOULOS, A. A Multiresolution Approach for Main Mobile Object Localization in Video Sequences. In *International Workshop on Very Low Bitrate Video Coding (VLBV01)* (Athens, Greece, Oct. 2001).
- [86] TSEKERIDOU, S., AND PITAS, I. Facial Feature Extraction in Frontal Views using Biometric Analogies. In *Proc. of the IX European Signal Processing Conference* (1998), pp. 315–318.
- [87] TSAI, R. Y., T. S. H., AND ZHU, W. L. Estimating 3-D motion parameters of a rigid planar patch II: singular value decomposition. *IEEE Trans. Acoustic, Speech, Signal Processing* 30, 4 (Apr. 1982), 525–534.
- [88] VOTSIS, G., ET AL. 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, Nov. 2000).
- [89] WANG, J. Stochastic Relaxation on Partitions with Connected Components and Its Application to Image Segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 20, 6 (1998), 619–636.
- [90] WENG, J., T. S. H., AND AHUJA, N. Motion and structure from point correspondences with error estimation: Planar surfaces. *IEEE Trans. Signal Processing* 39, 12 (Dec. 1991), 2691–2717.
- [91] WELLNER, P., W. M., AND R.GOLD. Computer-augmented environments: Back to the real world. *Communications of the ACM* 36, 7 (1993), 24–27.
- [92] XIROUHAKIS, Y., A. D., AND DELOPOULOS, A. A Novel Approach for the Estimation of Camera Motion in Virtual Studios Applications. In *International Workshop on Synthetic-Natural Hybrid Coding and 3D Imaging (IW-SNHC3DI'99)* (Santorini, Greece, Sept. 1999).
- [93] XIROUHAKIS, Y., A. D., AND DELOPOULOS, A. Camera Motion Estimation Using 3D-to-2D Line Correspondences. In *IEEE Nordic Signal Processing Symposium (NORSIG2000)* (Kolmarden, Sweden, June 2000).
- [94] XIROUHAKIS, Y., A. D., AND DELOPOULOS, A. Efficient Optical Camera Tracking in Virtual Sets. *IEEE Trans. Image Processing* 10, 4 (Apr. 2001), 609–622.
- [95] XIROUHAKIS, Y., AND DELOPOULOS, A. Least Squares Estimation of 3D Shape and Motion of Rigid Objects from their Orthographic Projections. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22, 4 (Apr. 2000), 393–399.
- [96] YOW, K., AND CIPOLLA, R. A Probabilistic Framework for Perceptual Grouping of Features for Human Face Detection. In *Proc. 2nd Int. Conference on Automatic Face and Gesture Recognition* (Vermont, USA, 1996).
- [97] YUILLE, A.L., D. C., AND HALLINAN, P. Feature Extraction from Faces Using Deformable Templates. *International Journal Computer Vision* 8, 2 (1992), 99–111.

- [98] ZHU, S., AND YUILLE, A. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 18, 9 (Sept. 1996), 884–900.
- [99] ZHANG, Z. Estimating motion and structure from correspondences of line segments between two perspective images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 17, 12 (Dec. 1995), 1129–1139.

Κατάλογος δημοσιεύσεων του συγγραφέα

ΔΙΕΘΝΗ ΠΕΡΙΟΔΙΚΑ ΚΑΙ ΒΙΒΛΙΑ

XIROUHAKIS, Y., DROSOPOULOS, A., AND DELOPOULOS, A. Efficient Optical Camera Tracking in Virtual Sets. *IEEE Trans. Image Processing* 10, 4 (Apr. 2001), 609-622.

VOTSIS, G., DROSOPOULOS, A., AND KOLLIAS, S. A Modular Approach to Facial Feature Segmentation on Real Sequences. *Signal Processing: Image Communication* 18, 1 (Jan. 2003), 67-89

KARPOUZIS, K., RAOUZAIΟΥ, A., DROSOPOULOS, A., IOANNOU, S., BALOMENOS, T., TSAPATSOULIS, N., AND KOLLIAS, S. Facial Expression and Gesture Analysis for Emotionally-rich Man-machine Interaction. In *3D Modeling and Animation: Synthesis and Analysis Techniques*, N. Sarris, and M. Strintzis, Eds., Idea Group, (2004)

KARPOUZIS, K., DROSOPOULOS, A., IOANNOU, S., RAOUZAIΟΥ, A., TSAPATSOULIS, N., AND KOLLIAS, S. Facial and Body Feature Extraction for Emotionally-Rich HCI. In *Encyclopedia of Information Science and Technology*, Mehdi Khosrow-Pour, Ed., Idea Group Publishers, 1180-1185, (2004)

DROSOPOULOS, A., AND KOLLIAS, S. Enhanced Interaction in an Optical Virtual Studio Using Object Tracking and Facial Analysis. submitted for publication, (2005)

ΔΙΕΘΝΗ ΣΥΝΕΔΡΙΑ

XIROUHAKIS, Y., DROSOPOULOS, A., AND DELOPOULOS, A. 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, Sept. 1999)

DROSOPOULOS, A., XIROUHAKIS, Y., AND DELOPOULOS, A. An Optical Camera Tracking System for Virtual Sets Applications. In *Visual, Modeling and Visualization Workshop (VMV'99)* (Erlangen, Germany, Nov. 1999)

XIROUHAKIS, Y., DROSOPOULOS, A., AND DELOPOULOS, A. Camera Motion Estimation Using 3D-to-2D Line Correspondences. In *IEEE Nordic Signal Processing Symposium (NORSIG2000)* (Kolmarden, Sweden, June 2000)

DROSOPOULOS, A., XIROUHAKIS, Y., AND DELOPOULOS, A. Optical Camera Tracking in Virtual Studios: Degenerate Cases. In *International Conference on Pattern Recognition (ICPR 2000)* (Barcelona, Spain, Sept. 2000)

VOTSIS, G., DROSOPOULOS, A., AKRIVAS, G., TZOUVARAS, V., AND XIROUHAKIS, Y. 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, Nov. 2000)

DELOPOULOS, A., RAGGOUSI, M., XIROUHAKIS, Y., AND DROSOPOULOS, A. Audio Pattern Recognition and Localization in Continuous Payout. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing* (Baltimore, Maryland, USA, June 2001)

BALOMENOS, T., DROSOPOULOS, A., RAOUZAIIOU, A., KARPOUZIS, K., AND KOLLIAS, S. Use of the Finite Element Method in 3D Structure and Motion Estimation of Non-rigid Objects. In *SPIE Electronic Imaging* (Santa Clara, California, USA, Jan. 2003)

DROSOPOULOS, A., MPALOMENOS, T., IOANNOU, S., KARPOUZIS, K., AND KOLLIAS, S. Emotionally-rich Man-machine Interaction Based on Gesture Analysis In *Universal Access in HCI: Inclusive Design in the Information Society*, C. Stephanidis, Ed., Lawrence Erlbaum Associates, vol. 4, 1372-1376, (June 2003)

BALOMENOS, T., RAOUZAIIOU, A., IOANNOU, S., DROSOPOULOS, A., KARPOUZIS, K., AND KOLLIAS, S. Emotion Analysis In Man-Machine Interaction Systems In *Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI04)* (Geneva, Switzerland, June 2004)

DROSOPOULOS, A., AND KOLLIAS, S., Object Localization in a Virtual Studio Environment In *International Conference From Scientific Computing to Computational Engineering, IC-SCCE* (Athens, Greece, Sep. 2004)

BALOMENOS, T., RAOUZAIIOU, A., IOANNOU, S., DROSOPOULOS, A., KARPOUZIS, K., AND KOLLIAS, S. Emotion Analysis in Man-Machine Interaction Systems In *Machine Learning for Multimodal Interaction, Lecture Notes in Computer Science*, S. Bengio, and H. Bourlard, Eds., Springer-Verlag, vol. 3361, 318-328, (2005)

□

Βιογραφικό Σημείωμα

Ο Αθανάσιος Ηλία Δροσόπουλος γεννήθηκε στη Λαμία το 1976. Το 1998 του απενεμήθη το πτυχίο από το τμήμα Πληροφορικής του Πανεπιστημίου Ιωαννίνων. Το 2005 περάτωσε τη διδακτορική του διατριβή στη σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβειου Πολυτεχνείου, υπό την επίβλεψη του καθηγητή Στέφανου Κόλλια. Τα κύρια ερευνητικά του ενδιαφέροντα περιλαμβάνουν την εκτίμηση 3Δ δομής και κίνησης, την εξαγωγή χαρακτηριστικών του προσώπου και τα συστήματα εικονικών στούντιο.

