Detection of Composite Geometric Structures through Possibilistic Reasoning

Giorgos Akrivas and Stefanos Kollias
National Technical University of Athens
Department of Electrical and Computer Engineering
Image Video and Multimedia Laboratory
Iroon Politechniou 9 Athens Greece
Phone: +30-10-772-4352, Fax: +30-10-772-2492
email: gakrivas@image.ntua.gr, stefanos@softlab.ntua.gr

ABSTRACT: In this paper, we deal with the problem of modelling and recognizing composite objects. Recognition involves two stages: first, recognition of the components via their visual features and second recognition of correct positions of the components with respect to the composite object geometric model. The geometric model we use allows for deformable composite objects, as well as indefinite number of components. Furthermore, we use a description logic in order to model the conceptual aspects of the part - whole relationship. The uncertainty involved in recognizing the various components, as well as the uncertainty involved in recognizing the correct positions is modelled through possibility theory. A maximal specificity criterion is introduced for merging components into composite objects.

KEYWORDS: Composite objects, Description logics, Possibilistic Reasoning, Triangulation, Partonomy

INTRODUCTION

Lately, the problem of multimedia understanding has received attention, because of its great importance in the fields of intelligent retrieval, human - computer interaction and computer vision. By *multimedia understanding*, we mean bridging the *semantic gap* between the signal segments, extracted by standard signal processing and the entities that a human recognizes in a multimedia document, i.e. objects and events.

The MPEG-7 standard for description of multimedia content [1] provides two frameworks, which rely on different sides of the semantic gap: the *structural* description schemes, which provide for a description of the features of the signal segments, and the *conceptual* description schemes, which provide for a description of the semantic entities encountered in a multimedia document. What is missing is a framework, providing signal - level models for *classes* of semantic entities (e.g. the class of all humans, as opposed to a specific person encountered in a document), which can be used to map signal segments to semantic entities, through their features.

This paper deals with the problem of modelling *composite* objects, i.e. objects which consist of other objects (*components*), which must obey to certain classification (classes they must belong to) and geometric constraints.

Classes and knowledge about them are encountered in the more general framework of *ontologies*. An ontology usually introduces *concepts*, *properties* of concepts, *relations* between concepts and *rules* that concern concepts and relations. A legal world description is a collection of individuals, which belong to the concepts and the relations and obey the rules of the ontology. Reasoning (e.g. validation, subsumption and recognition) is performed by mapping an ontology to a first order logic.

Description logics [2][3] are structured fragments of first order logic, with desirable properties such as decidability. Thus, they are considered important candidates for reasoning within ontologies, and ontological languages such as OIL and OWL-DL can be mapped to a description logic. The basic notions of a description logic are *concepts*, *roles* and *individuals*. Knowledge in a system consists of a set of terminological axioms (TBox), i.e. definitions of concepts with respect to other concepts and roles and assertions (ABox), i.e. individuals and the concepts and roles they belong to.

Although the part-whole relation between a composite object and its components can be thought of as a specific kind of role, lately there has been a movement towards promoting it with to “first – class citizen status” in description logics. A review of various approaches in this subject can be found in [4].

In this work, we use a particular description logic which provides for part - whole roles, proposed by Lambrix et al. in [5]. Important features of this language is typing of parts (parts belonging to specific concepts, or being specific individuals) and ordering and quantity constraints between parts. Moreover, the authors propose algorithms for checking whether a set of individuals *composes* a certain composite object, and for finding, out of a set of individuals, which individuals are needed in order to compose a composite object (*completion*). The process of creating a new individual based on evidence about composition is called *assembly*. According to this framework, a set of assertions about individuals can compose a composite object when they contain the proper individuals, the proper classes of parts, the individuals satisfy the quantity and ordering constraints, no other individuals are contained, and the number of individuals is at least two. A by-product of the process of completion is the construction of a part-of hierarchy for individuals.

We apply the completion - assembly framework in the following application: a given multimedia document has had some of its segments extracted and classified, based on feature and/or component evidence. Moreover, the position of the semantic entities is known. Based on this information, composite objects are assembled if they contain the correct components in the correct positions.

Having the components in the correct geometric position requires a generalization of the notion of ordering, used in [5]. In section 2, a geometric model for components is introduced. Moreover, recognition in multimedia is a process which involves uncertainty. This issue is covered in section 3. Some integrity constraints, which composite objects must obey to are discussed in section 4. Finally, our concluding remarks are in section 5.

GEOMETRIC MODELLING OF COMPOSITE OBJECTS

As mentioned in the introduction, the process of completion includes incrementing the set of components until they compose the concept. We will refer to each step of this process as *merging* a component with a composite object. Each individual which is merged in the set of components must, apart from belonging to the correct concept, be in the correct position. This section defines how this can be inferred. We constrain ourselves to still images (therefore, we model objects only, not temporal activities). The geometric model we propose is sufficient for 2D modelling and is invariant to Euclidean transformations (translation, rotation and scaling).

The geometric model is based on the notion of *triangulation*. A triangulation of a set of points is a tessellation of the interior of their convex hull into triangles, whose vertices coincide with the points. The relative positions of components can thus be given by the geometric characteristics (edges and angles) of the respective triangulation. Moreover, the angles (and edges, with a scale parameter) of a triangle are invariant to Euclidean transformations.

For *rigid* 2D objects, two points are necessary for defining their size and orientation. Thus, it is desirable that more than one feature points have been recognized in a visual object. Moreover, after the merge with the composite object, feature points that remain in the boundary of the triangulation become feature points of the composite object.

The advantages of referring to triangles instead of an arbitrary coordinate system is that it allows for local, instead of global comparisons, and that it allows for certain kinds of deformations, by omitting certain sides or angles of the triangles. We note that allowing for the points of a coordinate system to become fuzzy it too fluid, and does not permit us to define connection relations among components.

A triangle can be defined by two edges and one angle. If the angle is left undefined, we obtain rotation. If a side is left undefined, we obtain translation. We will refer to components with ill-defined triangles as *connected* components.

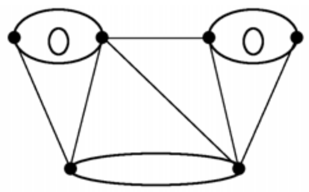

Figure 1: Triangulation over the feature points of the components of a face

Finally, when the number of components varies, as in the case of the books in a book-shelve, we will call them *free* components. We constrain free components to two special cases: those of being in a straight line and those of having equal positions to each other. These two cases correspond to the above two kinds of deformation mentioned above.

UNCERTAINTY ISSUES

In this section, we handle the degree of validity of merging a component into a composite object.

Mapping of visual objects to semantic entities through their features clearly involves uncertainty, because an object can match with many visual models. Merging of a visual objects with a composite semantic entity additionally involves a

different type of uncertainty, that of having the components in the correct position, according to the geometric model of the composite object.
The above two uncertainties correspond to the degree of compatibility of a model to a given region. An additional kind of uncertainty is whether a small degree of certainty must lead to the decision that the region actually corresponds to no visual model stored in the knowledge base. Again, we distinguish between two kinds of uncertainty, namely the uncertainty in visual recognition, and the uncertainty of composite object recognition. The latter corresponds to the degree of whether the degree of detecting a correct component of a composite object is sufficient evidence to merge it with it, rather than allowing it to be separate than the composite object.
We consider the visual matching value as a possibility distribution and the geometric matching value as a conditional possibility distribution.
Let a visual object o be described in terms of its visual features $f(o)$ and its position $\mathbf{x}(o)$. Moreover, let S denote the set of concepts. Visual appearance (which can be visual features and/or visual components), in combination with the semantic entities' models yields the *possibility distribution function* $\pi_{f(o)}: S \rightarrow [0,1]$, or, more simply, π_o. For $s \in S$, $\pi_o(s)$ corresponds to the degree, to which $f(o)$ is compatible with the visual description of semantic entity s [6]. Moreover, since the number of semantic entities, for which we have descriptions is limited, there is always the possibility that the object does not correspond to any of them. We can formulate this fact by $\pi_o(\text{"unknown"}) = 1, \forall o$. Therefore, the possibility distribution function is described via a normal fuzzy set.
The possibility of o being in the correct position for merge can be formulated as a conditional possibility distribution $\pi(m \mid s)$, where $m \in \{0,1\}$ denotes the two hypotheses, that of merge (1) and that of not merge (0).
No matter how perfectly the position matches with the geometric model, there is always the possibility that the object is unrelated to the composite object. Actually, since no geometry is involved, the possibility value of it being in the "correct" position in the 0 event is always 1. More formally, $\pi(0 \mid s) = 1, \forall s$. Moreover, an unrecognized object cannot be considered a component. Therefore, $\pi(1 \mid \text{"}unknown\text{"}) = 1, \forall s$.
Since the visual features of the object does not interact with its position, the possibility distributions are independent. Therefore, the joint possibility distribution is simply the minimum [7]: $\pi(m,s) = \min\{\pi(s), \pi(m \mid s)\}$
Thus, if $\pi(1 \mid s) \geq \pi(s)$, then $\pi(0,s) = \pi(1 \mid s)$, i.e. the possibility of merge is equal to the possibility of not merge. Thus, possibility alone cannot help a decision on merging. Let us examine how this problem is solved by human vision. When we see two eyes in the correct position, we are strongly inclined to believe that they belong to a human head. Let us now consider the problem where o looks like either an eye or an egg. Moreover, let us suppose that the possibility of merge equals the possibility of not merge. We notice that the decision to merge the two objects into a head reduces uncertainty about the semantic identity of o. Thus, we can argue that reduction of uncertainty is an important factor in deciding whether to merge or not. Consequently, we propose a decision on the possibility of merge if this reduces the uncertainty of o.
Under uncertainty theory [7], the uncertainty (non-specificity) of a possibility distribution is:

$$U(\pi) = \sum_{i=2}^{n} \pi_i \log_2 \frac{i}{i-2} \tag{1}$$

where the possibilities π_i are sorted, such that $\pi_i \geq \pi_{i-1}$.
The posterior possibility distribution, after the hypothesis of merge can be computed as a bayesian revision [8] by:

$$\pi(s \mid m) = \frac{\pi(s)\pi(m \mid s)}{\pi(s) \circ \pi(m \mid s)} \tag{2}$$

where $\circ$ is a $\max$ - product fuzzy composition.
The possibility distribution of s thus becomes narrower, and hence more specific, because many values of s are eliminated. However, a bayesian revision does not take into account the possibility value of $\pi(1 \mid s)$. Since this is opposite to the desired effect, we modify the revised distribution by:

$$\pi'(s \mid m) = \pi(s \mid m) \vee (\pi(s) \wedge \neg \pi(m \mid s)) \tag{3}$$

where $\wedge$ and $\vee$ are a fuzzy t-norm and a fuzzy t-conorm, respectively and $\neg$ is a fuzzy complement.
Thus, a zero value for $\pi(1 \mid s)$ will result into $\pi'(s \mid 1) = \pi(s)$.
Using uncertainty theory, we can measure the uncertainty of the possibility distributions π' (merge) and π (not merge), and decide to merge or not on grounds of minimizing uncertainty. Moreover, this holds as long as uncertainty is

low (and thus can be increased). When uncertainty is already high, then component evidence cannot increase it any further, but naturally this is not a reason to decide not to merge.

ONTOLOGICAL CONSTRAINTS

The geometric relation of section 2 can be considered, according to [4] as a *horizontal* relationship, i.e. a relation between the components. *Vertical* relationships, i.e. relationships between the components and the composite object which refer to the existence of the whole and the parts can also be defined: *essential* parts (the whole cannot exist without the part), *dependant* parts (the part cannot exist without the whole) and *exclusive* parts (the part belongs at most to one whole).

A part being essential or dependant influences the confidence of recognition: the confidence of recognizing the whole (e.g. a human face) is influenced by the confidence of recognizing an essential part (e.g. the eye), and influences the confidence of recognizing a dependant part (e.g. the hair).

Considering objects in the plane, we have the constraint that two objects cannot occupy the same space. This has the following implications:

- visual parts are always exclusive
- only feature points belonging to the boundary of the composite object are considered, and only objects adjacent to them are considered for merge. Therefore, one begins by triangulating the whole image, and clustering adjacent objects into composite objects

DISCUSSION

In this paper, a geometric model was introduced in the part-whole-enabled description model of Lambrix et al., in order to extend the completion - assembly framework toward visual parts. Moreover, possibility theory was used to answer the following question: if the position of an object is ``rather wrong'' is there sufficient evidence to merge the object into a composite object? This being a work in an early stage, numerous issues remain open.

One group of issues is the geometric model. Supporting 2D shapes only enables us to make a domain closure assumption (the only objects that exist are the ones that can be seen). In a 3D world, an object can occlude a part of itself, as well as parts of other objects. Therefore, one must infer the existence of a composite object by detecting just part of its components. Moreover, the spatial relations between the components are no longer invariant.

Another issue is the relationship between the visual features (such as color) of the components and the ones of the whole. For example, a very small face may have eyes that are too small to be recognized alone, but they may influence the visual features of the face, in a way that not only allows it to be detected, but the eyes to be located. Moreover, the geometric model can estimate the possible location of the undetected components, thus guiding the process that detects components out of visual features.

Finally, we think that there is a class of composite objects where the integrity constrain proposed in the previous section must be dropped. For example, let us consider the concept ``person holding hands with another person''. Obviously, a person can hold hands with two persons at once, and therefore be part of two concepts.

A proposal to construct description logics for image recognition appears in [9]. There, composite shapes are constructed out of simpler ones by using linear transformations on them. A complex shape is satisfiable if no overlap occurs between the components. While the authors avoid the definition of a new kind of role, as [5] does, we believe that using global, instead of local constraints for the geometric relationship between parts is a limitation, since it does not provide for notions such as deformation and the introduction of unlimited number of parts. Ontological problems occur too, for example, subsumption is defined in a way that would infer that a part is subsumed by (is a kind of) its composite object.

REFERENCES

[1] ISO/IEC JTC 1/SC 29 M4242, Text of 15938-5 FDIS Information Technology -- Multimedia Content Description Interface -- Part 5 Multimedia Description Schemes , October 2001.

[2] The official web site of description logics, http://dl.kr.org

[3] Franz Baader, Diego Calvanese, Deborah McGuinnes, Danielle Nardi and Peter Patel Schneider, 2003, “Description Logics Handbook”, CUP

[4] Alessandre Artale, Enrico Franconi, Nicola Guarino and Luca Pazzi, “Part – whole relations in object – centered systems: An overview”, Data & Knowledge Engineering 20 (1996) 347-383, Elsevier Science

[5] Patric Lambrix, Lin Padgham, “Conceptual modelling in a document management environment using part – of reasoning in description logics”, Data & Knowledge Engineering 32 (2000) 51-86, Elsevier science
[6] Didier Dubois, Henri Prade, “Bayesian conditioning in possibility theory”, Fuzzy Sets and Systems 92 (1997) 223-240, Elsevier Science
[7] Klir G. and Bo Yuan, 1995, “Fuzzy Sets and Fuzzy Logic, Theory and Applications”, New Jersey, Prentice Hall
[8] Stephane Lapointe, Bernard Bobe, “Revision of possibility distributions: A Bayesian inference pattern”, Fuzzy Sets and Systems 116 (2000) 119-140, Elsevier Science
[9] Eugenio Di Sciascio, Francesco M. Donini, “Description Logics for Image Recognition: a preliminary proposal”, in Patrick Lambrix, Alexander Borgida, Maurizio Lenzerini, Ralf Moller, Peter F. Patel-Schneider (Eds.): Preceedings of the 1999 International Workshop on Description Logics (DL’99), Linkoping, Sweden, July 30 - August 1, 1999