A Fuzzy Object Retrieval System for Image Understanding

A. Del Bimbo
Dip. di Sistemi e Informatica
Università di Firenze
Via S. Marta 3, I-50139 Firenze, Italy

D. Luca ReLLA
Dip. di Scienze dell'Informazione
Università degli Studi di Milano
Via Moretto da Brescia 9, I-20133 Milano, Italy

Abstract
This work addresses the problem of object recognition in the field of image understanding systems. In this context, the purpose of the paper is twofold:

a) Firstly, we show how the problem can be approached exploiting information retrieval techniques. That is, given a collection of object models, the target object is regarded as a query and object recognition is regarded as a query processing task aimed at locating the models which best fulfill the given one. So the proposed system manages an input image and returns a ranked set of relevant objects computing for each of them a degree of support depending on multiple sources of evidence. Since the problem is fuzzy in nature, the inference mechanism is based on a fuzzy set theoretical framework.

b) Secondly, we show how the object-oriented paradigm provides the right framework to represent effectively and efficiently the knowledge about the objects. Each real world entity in the domain is mapped on a collection of composite objects, hierarchically organized, each level representing a specific abstraction level at which the entity is regarded.

Finally, we discuss implementation issues with reference to the retrieval component and related techniques that have been devised to improve the computational efficiency of the system.

1 Introduction
Object recognition is a fundamental research area in image understanding. Generally speaking, it requires to match the features of a given object against those of some predefined object models. However, in several application fields, as aerial recognition, environment monitoring, mobile robots etc., the need for establishing a general purpose image understanding system requires to manage large data bases with different kinds of information. In all these cases, the retrieval from a large set of complex objects is involved. Therefore modeling in a flexible and efficient way both the domain knowledge and the retrieval process becomes a central issue to the problem [Mo88].

Modeling information for visual recognition is a hard task since the generic model takes the form of a complex composite object. In fact:

- Spatial data can be represented in different ways, including skeleton, shape, texture, surface and volume-based representations [Mo88]. The choice is directly related to the specific processing environment and the objectives of the recognition task. Therefore a generic visual model of each object has to take into account all of them.
- Different resolutions and points of view of the same object give raise to possible different descriptions. This fact should be considered in the model, explicitly or implicitly.
- In order to perform recognition and understanding, knowledge from other sources than the visual sensor is generally needed. This includes the observer's general knowledge which refers to items not related to vision, and is valid for all the scenes or for the specific application domain [Mu88].

In this work, an object-oriented framework is adopted to represent effectively visual and non-visual knowledge about the objects in the domain. On the other hand, the recognition task can be cumbersome since a lot of models could require to be inspected before reaching a solution. This is mainly due to the uncertainty that is present at several levels in the process:

- During the image creation process. Here information is merged with spurious data, (like those arising from surface reflectance), and modified by external factors, such as the observer's point of view and the sensor capacity.
- After edge detection and segmentation. Some of the resulting spatial tokens do not correspond to physical objects, and have to be considered as "segmentation noise". Moreover viewing information at different res-
olution levels, results in a different capability to identify part-subpart relationships among objects [Bur88].

- During classification. Spatial tokens can be classified at different abstraction levels depending on the amount of information available from the steps above. In general, several abstraction levels of the same object model have to be activated in order to perform satisfactory classification.

- During retrieval. A crucial point is to establish a relationship between the object to be recognized and the models in the collection to determine whether a model is pertinent or not. In terms of decision theory, this selection process is essentially a decision under uncertainty.

In the area of information retrieval, van Rijsbergen [Rij86] and Croft [Cro88] have proposed a model of retrieval based on plausible inference. The retrieval process is regarded as the problem of determining an implication relationship between a document and a query and assessing the plausibility of that implication. The approach investigated in our prototype system is based on a fuzzy theoretical framework [Luc90]. Moreover, in order to improve the response time of the system, a bounded strategy is followed to reduce the set of models to be evaluated and the search procedure is stopped as soon as a minimum number of object models has been examined which is enough to guarantee the retrieval (recognition) of the target object. In the following, we denote with model the stored reference object and with object the elements to be recognized in the image.

Essentially, we propose an approach in which object-oriented modeling of image data with fuzzy information retrieval are integrated.

The paper has the following structure: Section 2 gives an overview of system architecture and main functionalities. In Section 3 the Object-Oriented paradigm is presented and used for modeling domain knowledge. Section 4 presents the fuzzy retrieval model and discusses in detail the search strategy to locate and rank the relevant models. In Section 5 results are summarized and further research perspectives are outlined.

2 System Architecture

Design and implementation issues are discussed with reference to a prototype system under development at University of Florence which combines both the features of two previous systems, respectively in the field of fuzzy information retrieval [Luc91], and in the field of object-oriented image understanding [Cap90].

Fig. 1 reports the system organization and the basic building modules. When images are loaded, the content is processed by the Image Analysis component with the support of the knowledge base. Every spatial token resulting from the segmentation can be regarded as a specific sub-image of the original grabbed image. It is mapped on a collection of objects, hierarchically organized, stored in the database by comparing its relevant features. Object models run according to a task dependent object list and evaluate their confidence to represent the spatial token.
cess terminates when some hypotheses with the required confidence level are reached [Fox87].

From the user's point of view, the object to be recognized is regarded as a query and the retrieval component acts as a filter which returns all of the models which best match the descriptors of the target object [Bos88].

The order in which models fulfilling the query are returned to the user reflects an assessment made automatically by the system about the degree of relevance to the query of the retrieved models. So the effect of this filtering capability is to restrict the attention to a manageable subset of relevant models.

For shortness, we have packaged into the module Knowledge & Object Management many common facilities including initialization, loading, deleting, update, save, restore and so forth [Pur87]. All these functions are not described here since they are of less interest in this context.

3 Object-Oriented Knowledge

The object-oriented paradigm is gaining large acceptance in recent years in programming languages and knowledge bases [Ste86, And87]. The basic concept is the class which represents the definition of an object type. It defines a set of variables which encode object attributes and a set of functions usually called methods that represent object behavior. Each variable and method has an associated type that is used as a constraint. Classes act as templates according to which objects are stamped out. Objects are therefore instances of classes. Values of attributes are references to objects and can be computed by activating the methods encapsulated in the object or stated explicitly. Central to the paradigm is the concept of class-inheritance, namely, a more specific class which inherits some properties (attributes and methods) from a more general class.

In the system the object-oriented model is used for knowledge structuring and organization as well as for arranging system operations.

Knowledge is broken down into classes representing data types. Abstraction levels are modeled through class inheritance. Every object is assigned to a class, when created. Therefore each object belonging to the same class has exactly the same set of variables and methods, i.e. uniform structure. This results in a modular and flexible system architecture which is easily adaptable to multiple environments.

3.1 Multiple type hierarchies

Recognition has to combine two kinds of information related to the entity and the application context respectively. With this goal, two type-hierarchies have been defined.

The former, entity type-hierarchy concerns the definition of abstraction levels at which knowledge of each model can be regarded. This is shown in fig. 2. Here information of the entity visual appearance as well as information stemming out from other sources than the sensor, like entity behavior and possible relationships with other entities, are collected.

![Figure 2: The entity type-hierarchy]

To support that, the entity-model type is specialized into visual model and non-visual model types. Further specializations of these knowledge types have been hypothesized. Regarding visual model types, different representations of the visual appearance of the entity are known and can be used for recognition. Examples are skeleton, shape, volume... representations. Moreover, each visual representation can be regarded at several abstraction levels depending on the approach to perception used in the recognition. Each level corresponds to a partial approximated view of the entity. Each abstraction level should take into account that different resolutions
and point of views are usually needed. On the other hand, we support the ability to define several non-visual model types. This complies with the need for modeling the different roles that each entity can play, depending on the context. Each role, roughly speaking, corresponds to a different possible behavior and to specific relationships with other entities.

The second type-hierarchy, application type hierarchy, takes into account the abstraction levels at which the application scenario can be viewed. Each class in this hierarchy represents some entity category, limited to a certain level of generality (see fig.3). Leaf classes are descriptions of entities that have a one-to-one correspondence with real world entities. Structures of these classes comply with the general structure of the above discussed entity type hierarchy.

3.2 Object hierarchies

Only one object exists per abstraction level of the application type-hierarchy. This is the model of the entity according to the view at that level of abstraction. Each entity model is made up by a collection of several objects in the system and modeled as a complex composite object. These objects are grouped into an object hierarchy that therefore corresponds to a real world entity. Each leaf of this object hierarchy can be regarded as a perspective played by the entity. In this way we partition information of the entity into parts of independent interest. This conforms to the concept of object specialization and complies with the fact that real world entities can play different roles in different environments thus changing relationships with other entities.

We should notice that for each entity model there are several visual model objects (one for each kind of representation at a specific resolution and orientation), and several non-visual model objects (one for each role played by the entity in the environment). Each of them can be used as a separate conceptual focus in the recognition process.

Fuzzy set techniques managed by the system allow us to model the inherent uncertainty. Each attribute is accessed through a specific function that embeds the uncertainty associated with the attribute value.

In the system, object hierarchies have been linked to entity class hierarchies. This gives the chance of maintaining links between objects implementing representations of the same entity at different abstraction levels.

4 Fuzzy Object Retrieval

4.1 Retrieval Model

In the following the basic notions on fuzzy sets are assumed. For a detailed discussion on this topic see [Kan86, Zad88]. Let \( O \) be a finite set of object models of cardinality \( |O| = n \), \( Q \) be a finite set of objects of cardinality \( |Q| = m \) as resulting from the Image Analysis component, and \( D \) be a finite set of descriptors of cardinality \( |D| = k \). We assume a binary fuzzy indexing relation \( I \) of the form:

\[
I = \{\mu_I(o, d) \mid (o, d) \in O \times D\}
\]

with a membership function \( \mu_I : O \times D \rightarrow [0, 1] \) indicating for each pair \((o, d)\) to what degree the descriptor \( d \) represents the object \( o \).

For each object \( o \in O \), on the basis of the indexing relation \( I \), it is possible to define the object description \( I_o \) as a fuzzy subset in \( D \):

\[
I_o = \{\mu_{I_o}(d) \mid d \in D; \mu_{I_o}(d) = \mu_I(o, d)\}
\]

Applying the same indexing relation \( I \), for each object to be recognized \( q \in Q \), it is possible to define the query
description $I_q$ as a fuzzy subset in $D$, in the form:

$$I_q = \{\mu_{I_q}(d)/d \in D; \mu_{I_q}(d) = \mu_I(q, d)\}$$

It is possible to express now the retrieval rule in the form of a binary fuzzy relation as well:

$$R = \{\mu_R(q,o)/\{q,o\} | q \in Q; o \in O\}$$

with a membership function from an aggregation scheme of confidences computed at different abstraction levels as described in the following subsection.

Given an object $q \in Q$, on the basis of the retrieval relation $R$, we can define the retrieved set $R_q$ as a fuzzy subset of the model set $O$:

$$R_q = \{\mu_{R_q}(o)/o \in O; \mu_{R_q}(o) = \mu_R(q,o)\}$$

where $\mu_{R_q}(o)$ represents the strength (certainty) of the relationship between the model $o$ and the target object $q$. As a result, the recognition task can be regarded as a re-ordering of the collection with respect to the values of $\mu_{R_q}(o)$.

In order to limit the response to only those models which are characterized by the highest scores, we can get an $\alpha$-level of the set $R_q$ by extracting elements with a membership value greater than a fixed threshold $\alpha \in [0, 1]$:

$$R_q(\alpha) = \{\mu_{R_q}(o)/o \in O; \mu_{R_q}(o) \geq \alpha; o \in O; \alpha \in [0, 1]\}$$

This is the set of object models which best fulfill the target object $q$. A ranked output can be returned by arranging the retrieved models in decreasing order according to the degree of their membership in $R_q(\alpha)$:

$$\text{Ord}(R_q(\alpha)) = \mu_{R_q}(o) / o_1, \ldots, \mu_{R_q}(o_{\alpha-1}) / o_{\alpha-1},$$

where: $\mu_{R_q}(o_{\alpha-1}) \geq \mu_{R_q}(o_i); i = 2 \ldots \alpha$

### 4.2 Confidence function

Given the object $q$ to be recognized, during the evaluation process, we have to compute for each model $o$ in the collection a value $\mu_{R_q}(o) \in [0, 1]$ reflecting the grade of relevance of the model $o$ with respect to the query $q$. For composite objects, such a membership function results from the aggregation scheme presented in the following.

Inspecting a visual model for its structure forces activation of visual models corresponding to each of its components. Reasoning is based upon assigning a certain weight to each of the geometrical, positional and relational properties of the object subparts [Lee87]. Weights $w_i$ are strongly dependent on the task and can be defined at initialization time. The retrieval function $\mu_{R_q}(o)$ for the composite object $o$ with $n$ global attributes and $p$ subparts $S_j$, (in total $n + p$ attributes), has been defined according to an averaging aggregation function:

$$\mu_{R_q}(o) = \frac{1}{p} \left( \frac{\sum_{i=1}^{n} w_i T(f_r_i, f_g_i) + \sum_{j=1}^{p} w_j \mu_{R_q}(S_j)}{N} \right)$$

with $T$ a function matching the $n$ reference feature attributes $(f_r_i)$ with the corresponding $n$ attribute values from the grabbed image $(f_g_i)$. $N$ and $P$ are normalization factors, with $N = \sum_{i=1}^{n} w_i$ and $P = 1 + \sum_{j=1}^{p} w_j$ respectively.

The overall confidence value is compared with a task-dependent threshold and when exceeded the object is retrieved (recognized).

### 4.3 Object Selection

Now we discuss how to evaluate each model with respect to the query. Since the objective is to locate only the set of top ranking models, we can try to minimize the number of models that are to be evaluated in order to improve the efficiency of the search procedure. Informally this is described in the following.

a) A side effect of the adopted object-oriented model is the grouping of individual objects in application type-hierarchies that we shall refer to hereafter as clusters. If we assume that each cluster contains on average $B$ objects, it follows that a collection of $|O|$ objects will give rise to a set of clusters $N = |O|/B$. The query is matched first against the root objects of the clusters. Individual objects in the clusters will be matched against the query if and only if the degree of relevance of the root is sufficiently large (it is not under a given threshold). Accordingly, a saving in processing has been achieved since a single query-cluster matching computation has replaced $B$ query-objects computations.

b) Non-visual knowledge can be exploited in order to prune in advance the number of clusters involved in the previous step. Non-visual knowledge is related to different roles that the entity can play in the context. This kind of information is modeled through classes of the entity type hierarchy and represented by role-objects in the entity object hierarchy. An index over the role-objects is maintained in order to improve the selection process.
Each role-object implements a separate perspective on the entity. Essentially, it maintains information concerning behavior and relationships with other entities. Role-objects accounting for contextual information act as a conceptual focus in the retrieval process so providing discrimination on stored models. In fact each entity can play a different role depending on the specific environment. As an example, looking for a car entity, different motion laws and proximity relationships with other entities can be supposed in the following roles: at-toll-station, on-highway, on-road, ... From the previous example, if we have a picture from an highway context, bicycles are strongly improbable. The effect being that the number of clusters to be accessed is reduced. Essentially, non visual knowledge acts as an index to visual knowledge and defines the starting subset of clusters to be analyzed.

c) Once a cluster has been selected and accessed, an expansion activity takes place. The system works by spreading activation from the high level nodes down to the lower levels in the application type-hierarchy [Coh87]. Each node is assigned a special activation weight which depends on the starting activation weight and the weight associated to the nodes traversed in the activation process. The set \( P_i \) of models appearing in the hierarchy descending from the root of the selected cluster represents the set of models which possibly satisfy the query. It is the multiset of objects which share at least one descriptor in common with the query and, hence, have a non-zero value for the confidence function. The cardinality of the set \( P_i \) may be large and this requires the useless calculation of many low confidence values, with the result that a large fraction of the model collection must be inspected. To improve the efficiency of such a procedure, heuristics can be introduced to constrain the search algorithm in order to favor particular pathways through the hierarchy and to terminate search along other ones. The procedures that have been developed here derive from previous work that was described by Lucarella [Luc88, Luc89].

Assume for the moment we are interested in locating the best model, that is the model with the greatest relevance value. The search begins by descending through the tree toward the leafs. For each arc not taken an upperbound is calculated and this value is stored along with the pointer to the arc. Such a value gives an upperbound on the relevance to the request of any possible model descending from that arc. When the search advances through a path, models in the nodes encountered along the way are examined, the confidence for each of them is computed and an overall assessment is reached aggregating different values. The resulting value along with the referenced model is maintained as the current best matching model, that is the most promising candidate for recognition. At any point in the search such a variable stores the reference to the best model seen so far along with the associated estimated confidence. Next we compare the confidence value for the best matching model with the maximum upperbound. If the upper-bound is greater, then the associated arc is worth while being explored since it might lead to a better model. Conversely, if the confidence value of the current best match is greater then the maximum upperbound, than the search can stop since there is no use to proceed. Such a procedure can be extended in a straightforward way if it is not required the best matching model but the set of best matching models as it is often the case. Further technical details can be found in [Luc89].

5 Conclusion

In this paper we have proposed a model in which object recognition for image understanding is regarded as a query processing task in presence of uncertainty. Under this assumption, an object-oriented model for knowledge representation and a fuzzy retrieval model for object recognition have been proposed in a unified framework. The object oriented paradigm provide facilities to represent complex composite objects and different levels of abstraction as needed for image understanding. Fuzzy set theory makes available a sound mathematical framework to take into account the different contributions of features in object representation and query characterization, to combine weighted confidence values, and to assess different grades of relevance of models to a given object.

System performance has been identified as a critical point particularly with reference to general systems in which object recognition from large model collections is required and heuristics have been suggested to reduce as much as possible the set of models to be evaluated.

Further research will be aimed at carrying out experiments on a larger scale in order to better evaluate the proposed approach and the associated retrieval strategy.

References


