
Shape Annotation for Intelligent Image Retrieval

Giovanna Castellano · Anna M. Fanelli ·
Gianluca Sforza · M. Alessandra Torsello

the date of receipt and acceptance should be inserted later

Abstract Annotation of shapes is an important process for semantic image retrieval. In this paper, we present a shape annotation framework that enables intelligent image retrieval by exploiting in a unified manner domain knowledge and perceptual description of shapes. A semi-supervised fuzzy clustering process is used to derive domain knowledge in terms of linguistic concepts referring to the semantic categories of shapes. For each category we derive a prototype that is a visual template for the category. A novel visual ontology is proposed to provide a description of prototypes and their salient parts. To describe parts of prototypes the visual ontology includes perceptual attributes that are defined by mimicking the analogy mechanism adopted by humans to describe the appearance of objects. The effectiveness of the developed framework as a facility for intelligent image retrieval is shown through results on a case study in the domain of fish shapes.

Keywords Shape annotation · Fuzzy shape clustering · Image annotation · Image retrieval · Semi-supervised clustering · Visual ontology.

1 Introduction

With the fast development of digital imagery ranging from real-world pictures to synthetic images, the description of the image content has become one of the biggest challenges in many application domains, ranging from image

G. Castellano · A.M. Fanelli · M.A. Torsello
Department of Informatics, University of Bari “A. Moro”
Via Orabona, 4 - 70126 Bari - Italy
E-mail: [giovanna.castellano,annamaria.fanelli,mariaalessandra.torsello]@uniba.it

G. Sforza
Department of Computer Science, University of Milano
Via Bramante,65 - 26013 Crema - Italy
E-mail: gianluca.sforza@unimi.it

indexing and retrieval [21], [58], [55] to biomedical image analysis [31]. While describing the content of a text is quite straightforward and can be done by computers according to the well known semantics of a language [60], [8], capturing and describing the content of an image is often a subjective task, due to the uncertain nature of visual content.

Much of the past research in image description was concentrated on the extraction of numerical features useful to represent *color, texture and shape* [51], [27], [64] and that can be easily processed by a computer. This kind of features provide the so called *low-level* description of the content of an image. More recently, many research efforts have been devoted towards finding *high-level* descriptions of visual content, giving a central role to the semantics expressed by images [59], [2].

Along with low-level features that represent the characteristics of images in a numeric form, high-level descriptions provide some form of knowledge which, once properly linked to numerical features, describes the visual content in a way that is compliant with the way humans adopt to describe what they observe.

In image retrieval systems the sole use of low-level visual features does not allow to express the actual semantic content embedded into images. Users may find difficult to formulate search queries by directly specifying low-level visual attributes. In effect, humans tend to recognize images and to express their content relying on high-level concepts, i.e. they usually formulate their queries in natural language by employing semantic concepts. In addition, the majority of users do not desire to retrieve images simply on the basis of similarity of appearance but they often need to search for images representing a particular type (or individual instance) of object, phenomenon, or event. This has led to draw a distinction between retrieval by primitive image features (such as colour, texture or shape) and semantic feature (such as the type of objects or events depicted by the image). According to [22] three distinct levels of retrieval can be distinguished:

- Level 1: retrieval by low-level features such as color, texture, shape or the spatial location of image elements; (e.g. “Find all images containing yellow or blue stars arranged in a ring”);
- Level 2: retrieval by derived attributes or logical features, involving some degree of inference about the identity of the objects depicted in the image; (e.g. “Find images of a passenger train crossing a bridge”);
- Level 3: retrieval by abstract attributes, involving complex reasoning about the significance of the objects or scenes depicted; (e.g. “Find images illustrating pageantry”).

One main problem of Level 1 retrieval, known as Content Based Image Retrieval, is the “semantic gap”, concerning “the lack of coincidence between the information automatically extracted from the visual data and the semantic meaning, i.e. the interpretation that the same visual data have for a user in a given situation” [58]. In an attempt to fill in the semantic gap, a variety of approaches have been proposed. One of the most successful approaches is

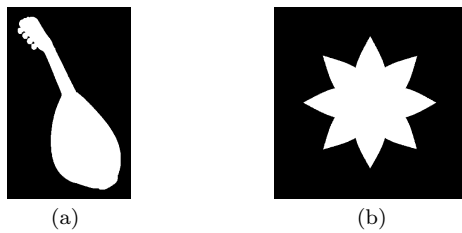


Fig. 1 Examples of ambiguous shapes: (a) may be annotated as a guitar or as a spoon (b) may be annotated as a flower or as a star

image annotation that consists in assigning labels (textual descriptions) to images or to their parts on the basis of visual features [28], [23]. Since manual annotation of images is a slow, error-prone and highly subjective process, increasing research efforts have been centered on the definition of automatic methods for image annotation based on low-level visual content [66], [24], [44].

To perform automatic image annotation, three main issues have to be faced: (i) the choice of low-level features to describe the visual content of the image, (ii) the definition of linguistic concepts characterizing the image domain and (iii) the annotation of the image content by means of linguistic concepts.

As concerns the choice of low-level features, several works have proved that visual features such as color, texture, and positioning, though important, are insufficient to convey the information that could be obtained through shape analysis of objects contained into images [7], [37], [55], [61]. Shape plays a critical role for the representation of objects contained into images becoming a key feature exploited in many applications of computer vision and image understanding for indexing and retrieval purposes. Indeed a considerable amount of information is contained in the boundaries of the objects, thus a definite opinion of the scientific community is that shapes should be considered as an essential mean to describe objects in an image. Moreover, it is recognized that shape is a feature strongly related to human perception since users usually perceive images as composed of individual objects identified by their shapes [46], [12]. For these reasons, in this work we consider the shape as primary feature to describe the objects contained in the images.

As concerns the definition of linguistic concepts describing the image domain, in many cases these are not completely known in advance. Hence, a mechanism to automatically identify linguistic concepts characterizing the image domain should be implemented. Linguistic concepts can be automatically derived by unsupervised learning applied to low-level features.

A commonly used approach is to apply clustering techniques to group visually similar shapes into clusters. Each cluster is then represented by a prototype or template and manually associated to linguistic concepts. Examples of this approach can be found in [38], [35], [15], [54], [33]. However, unsupervised clustering methods often generate inconsistent clusters including shapes that, although visually similar, actually represent different linguistic concepts (an example is given in fig. 1).

The presence of ambiguous shapes motivates the use of semi-supervised clustering that can identify linguistic concepts by learning from a combination of both labeled and unlabeled samples. Along with this idea, in [17] we proposed the use of a semi-supervised clustering algorithm, called SSFCM (Semi-Supervised FCM) to learn associations between shapes of objects and linguistic concepts by exploiting some domain knowledge expressed as a set of pre-labeled samples. Likewise, in the present work we adopt the SSFCM algorithm to automatically identify linguistic concepts describing the domain of shapes.

The third main issue in image annotation is the description of the visual content of an image. Recent progression in multimedia community has shown that ontologies are a powerful tool for describing visual content, especially in the domain of image retrieval [39], [42], [11]. In particular, the idea of visual ontology raised in literature as an effective tool to describe the image content from a semantic point of view [13], [41]. A visual ontology is an ontology whose structure is conceived to include several types of visual concepts which describe color, texture, shape or even the spatial visual content, as well as the relations between them. Several examples of visual ontologies have been proposed in literature so far. In [57] a visual ontology that combines low-level visual descriptors and domain knowledge is defined to describe multimedia content. In [34] a visual ontology is proposed that expresses knowledge by intermediate-level descriptors which can be more easily understood by humans. A visual ontology is proposed in [32] for describing digitized art images using type, style, concrete semantic (e.g., flower) and non objectionable semantics (e.g., warmth). In [40] the authors propose a shape ontology framework which integrates visual and domain information applied to bird classification.

Despite the increasing number of visual ontologies presented in literature, linking concepts to visual data by means of visual ontologies poses several problems that are still far from being solved. One key problem is how to obtain a deep and complete description of the semantic content conveyed by visual data. In many cases, using linguistic concepts alone is inadequate to completely express the semantics embedded in visual data [9]. Indeed, according to studies in cognitive psychology, the cognition process is based on different mental representations, such as symbols, images and schemata [25]. Therefore, the observer perception has to be taken into account for a complete description of visual data.

For this reason, in this work we propose a framework to describe shapes by means of both linguistic concepts and salient properties perceived by an observer. In particular, we propose a novel visual ontology for the description of shapes that is based on the idea of mimicking the analogy mechanism used by humans to describe parts of a shape. Specifically, our visual ontology provides a unified framework to gather several elements necessary for the description of shapes, namely: linguistic concepts expressing domain knowledge, prototype images that represent visual templates of shapes, and perceptual attributes describing parts of shapes by means of analogies, in a similar way that humans describe the appearance of objects.

The remainder of the paper is organized as follows. Section 2 presents the proposed framework and overviews its main phases. Section 3 reports results obtained by testing our framework as a facility for image retrieval. In section 4 the retrieval results of our system are compared with those obtained by Google Image Search. Conclusive remarks are drawn in section 5.

2 The proposed shape annotation framework

The idea underlying the proposed shape annotation framework for image retrieval is to exploit in a unified manner both the linguistic concepts characterizing the shape domain and the analogy mechanism adopted to describe shapes. In the proposed shape annotation framework, two main phases can be distinguished:

- Identification of linguistic concepts: a clustering process groups similar shapes into a number of semantic categories labeled by linguistic concepts characterizing the considered shape domain. Each category is represented by a shape prototype that is a visual template for that category to annotate shapes;
- Creation of a visual ontology: a novel ontology is built to describe the derived shape prototypes by means of perceptual attributes. Prototypes are divided into their salient parts and each part is described following the analogy mechanism usually adopted by humans to describe the appearance of objects.

Fig. 2 shows an overview of the proposed framework. It can be seen that the phase of identification of linguistic concepts provides as result a set of labels (linguistic concepts) related to the semantic categories underlying the considered shape domain and a set of shape prototypes (one for each category) that are manually annotated with the linguistic concepts. The second phase creates a particular visual ontology that provides a perceptual description of single parts of shape prototypes.

In the following the two phases involved in the shape annotation framework are described in more details.

2.1 Identification of linguistic concepts

To identify the linguistic concepts that characterize the considered shape domain, the salient objects contained into images have to be firstly detected. This can be done either manually or automatically. In the latter case, image segmentation has to be carried out. However, perfect identification of salient objects is not possible by unsupervised image segmentation and manual intervention is typically required to improve segmentation results. In this work, to avoid the complex phase of automatic image segmentation we assume that objects are manually detected from images or available in the form of shape contours.

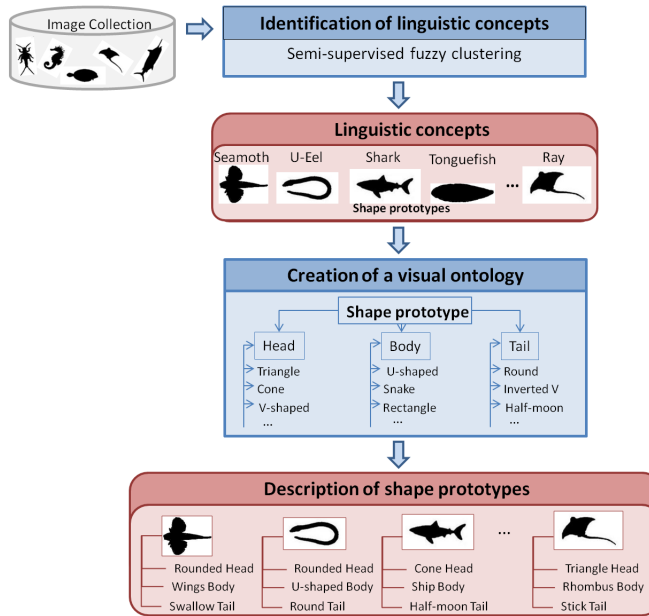


Fig. 2 Overview of our shape annotation framework

The contour of each shape is represented by means of Fourier descriptors that are well-recognized to provide robustness and invariance, obtaining good effectiveness in shape-based indexing and retrieval [6]. An object shape represented by means of its M Fourier descriptors is indicated by $\mathbf{x} = (x_1, x_2, \dots, x_M)$. In this work M is fixed to 32 by relying to our previous experimental experiences. Fourier descriptors of all the detected shapes are stored in a database.

Then a clustering process is applied to shape descriptors in order to group shapes deemed similar into a number of clusters. Each cluster represents a semantic category characterizing the shape domain. For each category, a shape prototype is derived that is considered as a visual template for that semantic category. The derived shape prototypes are manually annotated by domain experts by assigning them linguistic concepts representative of the domain categories.

To group similar shapes we employ SSFCM, a fuzzy clustering algorithm augmented by a semi-supervised mechanism firstly described in [49] that exploits knowledge about the semantic category of a limited number of shapes thus providing a useful guidance during the clustering process. SSFCM iteratively mines K clusters by minimizing the following objective function:

$$J = \sum_{k=1}^K \sum_{j=1}^N u_{jk}^m d_{jk}^2 + \alpha \sum_{k=1}^K \sum_{j=1}^{N_t} (u_{jk} - b_j f_{jk})^m d_{jk}^2 \quad (1)$$

where

$$b_j = \begin{cases} 1 & \text{if } \mathbf{x}_j \text{ is pre-labeled} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

f_{jk} denote the true membership values of the pre-labeled data to the categories, d_{jk} is the Euclidean distance between \mathbf{x}_j and the center \mathbf{c}_k of the k -th cluster, m is the fuzzification coefficient ($m \geq 2$) and α is a parameter that serves as a weight to balance the supervised and unsupervised components of the objective function. The higher the value of α , the higher the impact coming from the supervised component is. The second term of J captures the difference among the true memberships f_{jk} and the membership u_{jk} computed by the algorithm.

At the end of the clustering process, SSFCM provides a fuzzy partition matrix containing the membership degrees of each shape to each discovered cluster. For each cluster of shapes, a prototype is derived by selecting the shape with maximal membership degree to that cluster.

Finally, the derived prototypes are manually labeled by a domain expert who associates to each prototype a linguistic concept corresponding to a semantic category. The use of shape prototypes facilitates the annotation process, since only a reduced number of shapes (the prototypical ones) need to be manually annotated. Moreover, shape prototypes represent an intermediate indexing level that allows a faster retrieval process since a query is matched against prototypes, instead of the whole shape database, resulting in a speed up of the retrieval.

2.2 Creation of a novel visual ontology

Once the image domain is characterized by prototypical shapes and linguistic concepts, we design a visual ontology to add a mid-level description of shapes.

Specifically, we propose a visual ontology that employs perceptual concepts to express the appearance of a shape by considering the *analogy* as a mechanism for describing parts of shapes by resembling other shapes. The rationale behind our visual ontology is that the semantics of a shape is not only related to its physical aspect, but also to the visual concepts that the shape may suggest to the observer’s mind. For example, the shape of a red apple may be directly described as being spherical. But considering just the top of the apple in the proximity of its convexity, we may use analogy and associate this to something like the mouth of a volcano, with a column of smoke that raises up from the center, or even the shape of a 3D sin function as it is shown in fig. 3. This analogy-based mechanism can be applied to describe a shape as a whole, but also to describe its salient parts.

Hence, the proposed visual ontology includes two types of concepts to describe a shape:

1. Low-level visual concepts, that are used to describe geometric properties of shapes;



Fig. 3 The top of an apple may call forth other objects having analogous shape.

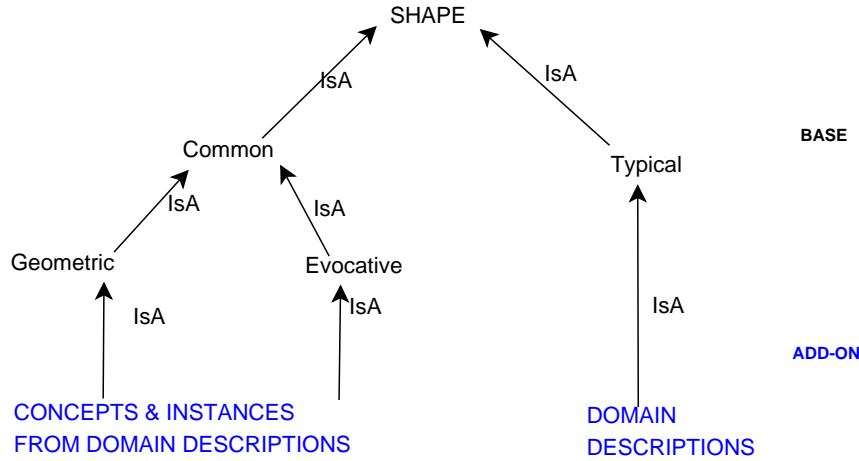


Fig. 4 The structure of our visual ontology

2. Medium-level visual concepts, that are objects whose shapes resemble the shape to be described.

These concepts correspond to two different types of visual attributes, that we call geometric and evocative attributes. *Geometric attributes* describe the geometry of a shape, while *evocative attributes* describe a shape in terms of objects having analogous shape. In fig. 4 we show the schema of our visual ontology. As it can be seen, shape knowledge is organized as a hierarchy of shape classes. The *Base* section of the ontology consists in the root concept *Shape* that is specialized into *Common* shapes and *Typical* shapes.

Common shapes collect all shapes that do not belong to a specific domain, while Typical shapes refer to objects belonging to a specific domain. Visual concepts associated to common shapes are distinguished in *Geometric* concepts (e.g. *rounded*, *pointed*) and *Evocative* concepts that describe a shape by means of analogy. Instances of common concepts (both geometric and evocative) are images that give a pictorial representation of the concepts by shapes.

Typical shapes refer to shapes of objects belonging to a specific domain, and they correspond to the prototypes derived by clustering, as described in the previous section.

The description of prototypical shapes is obtained by the following three-steps: (i) the shape is divided into meaningful parts; (ii) each part is described

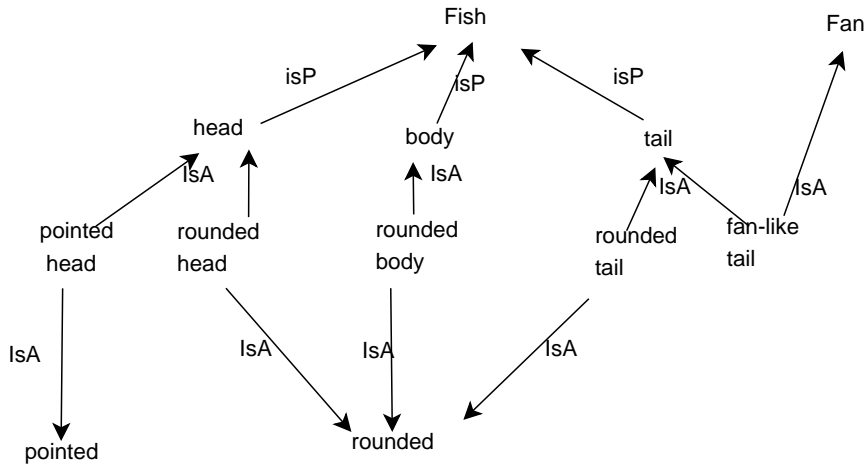


Fig. 5 The fish domain specification. Each term is intended suffixed by the word ‘shape’. Notations: IsP: sub-part, IsA: sub-class.

Table 1 High-level description of domain shape fish given by the visual ontology

<i>Shape</i>	fish
<i>Parts</i>	head, tail, body
<i>PartsOf</i> head	gills, eyes, mug
<i>ShapeAttributesOf</i> mug	
<i>Geometric</i>	pointed
<i>Evocative</i>	eagle beak

by visual concepts; (iii) each visual concept is associated with a visual instance (i.e. numerical descriptors used to represent the contour of the shape part). Hence our visual ontology contains knowledge to provide a perceptual description of single parts of each prototype. As an example, let us consider the domain of fish shapes. Each prototypical shape is divided into three parts, namely head, body and tail. A visual description is associated to each part and all the descriptions are merged to create the fish subtree in the visual ontology (fig. 5). This process can be iterated on each part of a fish shape, as shown in Table 1.

3 A case study

In order to verify the suitability of our shape annotation framework for image retrieval, a specific case study was considered. We developed a proof-of-concept that implements our visual ontology for the domain of fishes. Specifically, we considered a portion of the *Surrey* image set [6] that contains contours of 265 fish shapes belonging to 11 different semantic categories,

namely “Seamoth”, “Shark”, “Sole”, “Tonguefish”, “Crustacean”, “Eel”, “U-Eel”, “Pipefish”, “Swordfish” “Seahorse” and “Ray”.

3.1 Identification of linguistic concepts and prototypes

As a first phase, we derived a representation of the considered fish shapes by numerical descriptors. Starting from the coordinates of contour points of the shapes in the considered dataset, we computed Fourier descriptors and selected the first 32 coefficients for each shape. Such a number was empirically established during previous experiments on such dataset as a good trade-off between compactness and accuracy of shape representation.

Hence, we applied the SSFCM algorithm on shape descriptors to derive shape prototypes representative of a number of semantic categories characterizing the considered fish domain. To determine shape prototypes, 10 runs of the algorithm were performed starting from randomly initialized membership matrices. In all runs we fixed the fuzzification coefficient $m = 2$, the number of clusters $K = 11$ (equal to the number of semantic categories of the dataset), the parameter α to a value proportional to the percentage of labeled shapes. A set of 5% pre-labeled shapes was created by selecting randomly shapes among the shapes misclassified by running the unsupervised FCM. For each trial, we evaluated the Dominant Category Cardinality (DCC) for each derived cluster and we selected the trial with maximum average DCC value as the best run. Given clusters found by the best trial, for each cluster we selected the shape with maximum membership degree as a shape prototype representative of the corresponding semantic category. Finally, each shape prototype was manually annotated with a textual label expressing the linguistic concept related to the semantic category to which the prototype belongs. In fig. 6 we summarize the results obtained in the best trial. For each derived cluster, the figure reports the DCC value (expressed in terms of percentage), the respective shape prototype and the associated linguistic concept.

To assess the suitability of SSFCM for discovering shape prototypes, we performed a comparison with the *FCM* algorithm with no supervision. *FCM* was run by varying the number of clusters K from 9 to 15 (being 11 the number of categories) and we fixed the fuzzification coefficient $m = 2$. The experiments were repeated 10 times for each scenario. Table 2 summarizes the average DCC values obtained in the 10 performed trials, showing that SSFCM outperforms FCM in terms of the DCC values.

3.2 Creation of the visual ontology

Given the characterization of the fish domain in terms of linguistic concepts and prototypes, we created the visual ontology.

To construct the visual ontology, the shape of each prototype was segmented using the LabelMe tool [56]. Specifically, each prototype was divided












Cluster	DCC	Prototype	Linguistic concept
1	94.44%		Shark
2	100.00%		U-Eel
3	100.00%		Ray
4	84.62%		Eel
5	53.00%		Pipefish
6	55.00%		Swordfish
7	69.56%		Tonguefish
8	100.00%		Sole
9	91.00%		Seamoth
10	83.33%		Seahorse
11	55.55%		Crustacean

Fig. 6 The shape prototypes derived on the *Surrey* image set and the respective linguistic concepts

Table 2 Comparison of the obtained DCC values

	SSFCM	FCM
MEAN	75.64	73.74
ST. DEV.	2.56	2.95
MIN	72.70	69.30
MAX	80.60	77.60

into three parts, namely head, body and tail and each part was annotated by using geometric and evocative concepts. These annotations were provided by some volunteers who were required to give descriptions of shapes. All the geometric and evocative concepts provided by users were collected and used to construct the visual ontology.

Specifically, the visual ontology concepts were implemented using the *class* element of Protégé[47], a software developed in Java by Stanford University for editing ontology and acquiring knowledge. The root class of the ontology corresponds to the shape domain, that is Fishes in our example. Its sub-classes are associated to the parts of a fish shape: Head, Body and Tail. Each of them in turn branches into a set of classes, one per each kind of visual concept found for describing the shape of that part. These last classes implement both geometric and evocative concepts. Each part may have several kinds of shapes, depending on how its silhouette varies in the considered domain. For fishes we had 7 classes for Head and Tail, and 9 classes for Body. Fig. 7 depicts the schema of the resulting visual ontology. To realize connections between classes we used two types of *relations* in Protégé: *partOf* to link sub-classes directly with the root and *kindOf* to specialize them. For instance, each concept Head, Body and Tail is a *partOf* a Fish shape. Moreover, Triangle is a *kindOf* Head shape, or Ship is a *kindOf* Body shape and Fan is a *kindOf* Tail. All relations

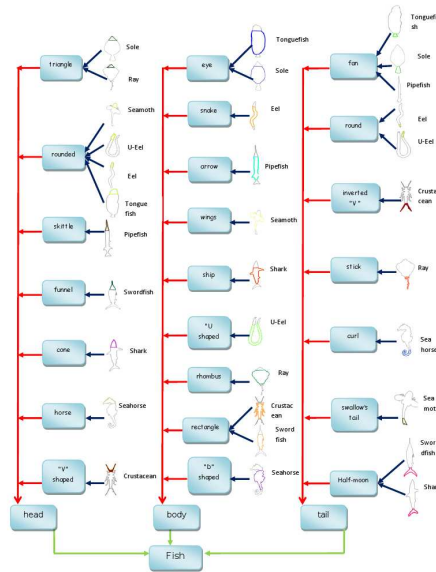


Fig. 7 The schema of the visual ontology for the fish domain

have their domain to the left side of the expression, and the co-domain to the right (e.g. Triangle is the domain and Head the co-domain of the relation previously described).

Successively, we included visual instances in the ontology. The visual instances are the shape prototypes derived for each semantic category, i.e. images used to visually identify the concepts of the ontology. For example, the prototype of a Ray with the head like a triangle is an instance of Triangle. Instances are implemented as individuals of the *Prototype* class of Protégé.

3.3 Image retrieval

One main benefit of describing an image is to create an intelligent and effective way to retrieve that image from a bunch of other images. Such a description will lead to more accurate results the more it is closed to the meaning associated by a human observer. The purpose of our ontology is twofold: collecting a set of terms commonly used to describe images, and linking the images having a similar look. Thus the visual ontology makes a bridge between images that are similar in the aspect and provides a way to retrieve them.

To test our framework as a facility for image retrieval, a search engine was developed to query the visual ontology. The form of a generic query is the following:

Get images from the *sample* image set where *properties* are valued

where *sample* represents a domain of shapes, and *properties* are the descriptions of the visual features given for each (even all) parts in which a shape is divided.

Queries over the considered image set can be resumed in the following ones:

1. Find images of fishes with a rounded tail;
2. Find images of fishes with rounded tail and the head like a triangle;
3. Find images of fishes with rounded parts.

All the queries work over the different parts of a fish shape: they check respectively one, two and all the three parts of the whole shape of a fish. The second query exploits the analogy mechanism to define the shape of a fish head to search for. These queries were translated in *SPARQL*, which is a W3C recommended language for making queries to a database of RDF meta data, which contains an archive of $\{Subject, Predicate, Object\}$ triples. Each element of the triple is represented by a URI (Uniform Resource Identifier) to be identified univocally inside the archive. By translating in *SPARQL* the three above queries, we obtained the following code:

```

1) SELECT ?FISHES {?FISHES <[URI]#Tail_shape> <[URI]#Rounded>}
2) SELECT ?FISHES {?FISHES <[URI]#Tail_shape> <[URI]#Rounded>.
      ?FISHES <[URI]#Head_shape> <[URI]#Triangle> }
3) SELECT ?FISHES {?FISHES <[URI]#Tail_shape> <[URI]#Rounded>.
      ?FISHES <[URI]#Head_shape> <[URI]#Rounded>.
      ?FISHES <[URI]#Body_shape> <[URI]#Rounded> }

```

where “.” is a concatenation operator which puts together two triples to be searched simultaneously into the ontology.

To develop the search engine we used *Apache Jena*, a Java framework for building Semantic Web applications [3]. In particular we used a query engine for Jena, namely ARQ, which exploits Jena libraries to work out *SPARQL* requests from the user. The developed search engine Java requires a textual input to start the search, and provides a set of images as a result. Valid inputs for the search engine are all the terms included in the ontology, i.e. domains, parts and their descriptions.

Summarizing, the developed search engine performs the online process of querying the ontology. As a result, the proposed shape annotation framework, intended as a facility for image retrieval, is composed of an online process and an offline process. A flow-chart describing the whole technique is depicted in fig. 8.

For simulation purposes, some images from the bird domain were also annotated and added to the visual ontology. While fish shapes were already available as a benchmark [1], bird shapes were derived after an edge detection process of the original color images¹. In particular, we considered about 100 bird images belonging to different bird shapes. Once the shape boundaries were extracted and represented by Fourier descriptors, we performed 10 runs

¹ <http://www.all-birds.com/Identify.htm>

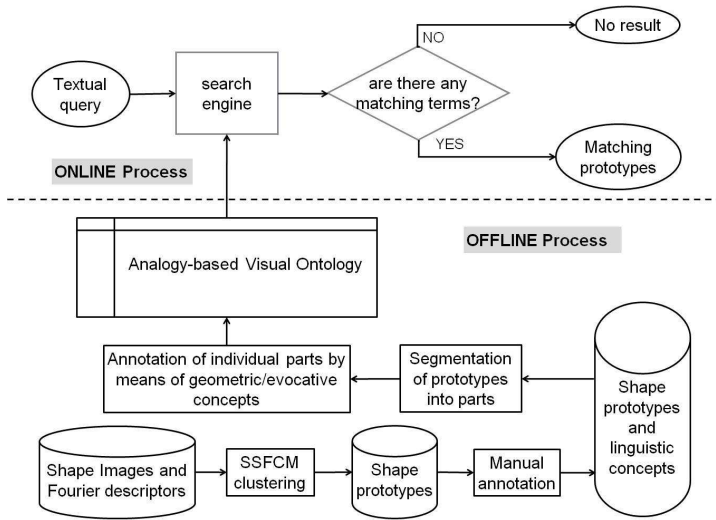


Fig. 8 Flowchart of the proposed framework intended as a facility for image retrieval

of SSFCM to identify prototypes and linguistic concepts of the bird domain. We selected the best trial and derived one prototype for each cluster. Prototypes of birds were then manually annotated by an expert using the following linguistic concepts: Gray Falcon, Blue Jay, Flycatcher, Harlequin Duck, Glossy Ibis and Goldfinch. Successively, each derived prototype was segmented into the salient parts (head, beak, wings and tail) with the use of LabelMe.

With the aim to show the suitability of the proposed framework, we provide a qualitative evaluation of the results by reporting the prototypes retrieved by our ontology in correspondence of some sample queries.

Firstly, a query using all the terms (domain, parts and descriptions) was considered:

Q1: Find fishes with a rounded head

Retrieval results obtained for query *Q1* in terms of prototypes satisfying the search terms are depicted in fig. 9.

As a second example, we considered the query:

Q2: Find shapes with a rounded head

This type of query is done when we do not know or we do not care about the shape domain we want to search into (fig. 10).

Doing the same query without specifying the part, we may look for shapes including at least one part that satisfies the search criteria such as:

Q3: Find shapes with rounded parts

Q4: Find shapes with parts like an half-moon

Q5: Find shapes with parts like a fan

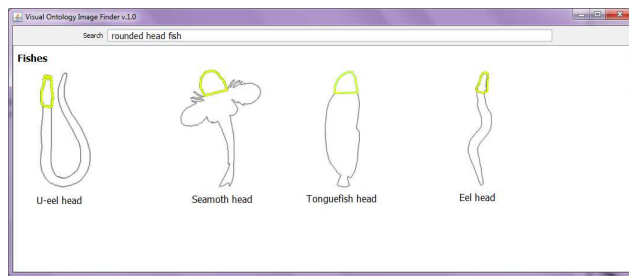


Fig. 9 Retrieval result for the query $Q1$

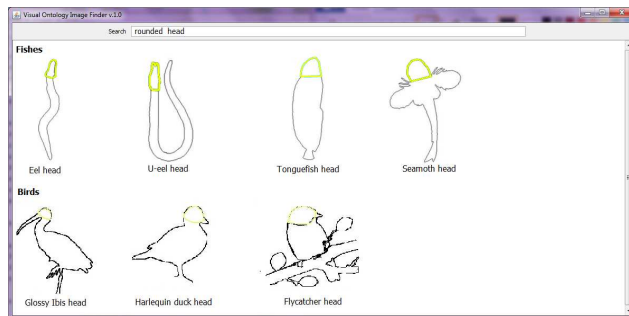


Fig. 10 Retrieval results for the query $Q2$

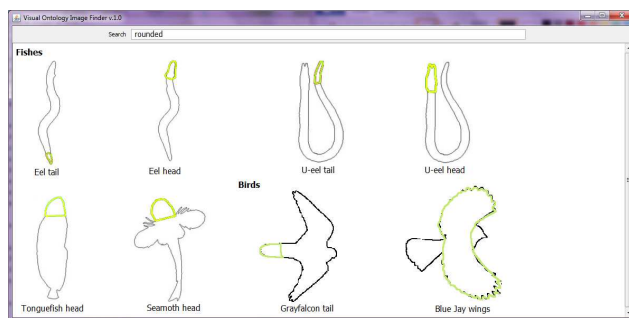


Fig. 11 Retrieval results for the query $Q3$

The retrieved prototypes obtained as result for these queries are shown in figures 11, 12, 13 respectively.

The retrieval results for queries $Q4$ and $Q5$ emphasize the main advantage of our ontology with respect to other visual ontologies, that is the analogy mechanism that allows to link parts of shapes to other shapes. In this way, shapes that belong to different domains (e.g. a bird and a fish) may result similar because they share the same evocative description for a part (e.g. half-moon, fan). This can be observed in fig. 12 and fig. 13. It should be noted that the same behavior can be observed for query $Q3$. The difference is that the property “rounded” is not an image but a descriptive concept. The retrieval

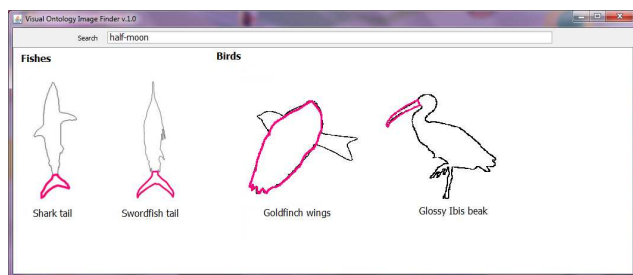


Fig. 12 Retrieval results for the query Q_4

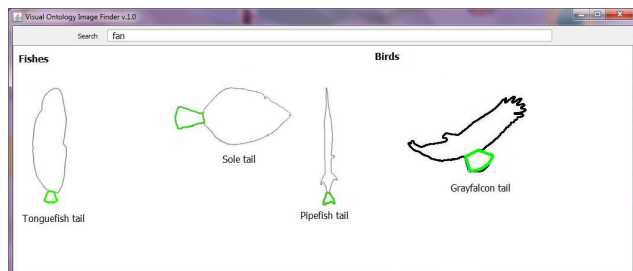


Fig. 13 Retrieval results for the query Q_5

result for such query is a set of shapes belonging to different domains without a considerable similarity as for queries Q_4 and Q_5 (see fig. 11).

Further, we performed a quantitative evaluation of the retrieval results of our framework by computing Precision and Recall measures for the queries previously examined. To obtain the retrieval results in correspondence of a query we exploited the composition of clusters derived in the phase of identification of linguistic concepts. Precisely, for each query, we firstly considered the prototypes retrieved by our framework. Hence, we considered as result for that query the set of all shapes that belong to clusters represented by the prototypes retrieved for that query. On this set of shapes we computed Precision and Recall values for that query. Hence, the results strongly depend on the quality of the clusters derived in the phase of identification of linguistic concepts. Table 3 reports Precision and Recall values obtained for each query. The values of precision and recall demonstrate that our strategy for the identification of linguistic concepts and related prototypes is effective and leads to good results of image retrieval.

4 Experimental comparison

To better assess the effectiveness of the proposed framework as a facility for image retrieval, in this section we present the results of a comparison with a well-known image retrieval system. It should be noted that, as stated in

Table 3 Precision and Recall values obtained for the sample queries

Query	Precision	Recall
<i>Q1</i>	0.87	0.93
<i>Q2</i>	0.89	0.94
<i>Q3</i>	0.86	0.90
<i>Q4</i>	0.87	0.92
<i>Q5</i>	0.86	0.92

[14], the issue of comparing different shape retrieval systems has been largely neglected in the research community due to the subjective character of such comparisons. The comparison among different retrieval systems is a difficult task and, often, not feasible since they work on various image domains and they adopt different search and annotation methods [45].

We point out that, to the best of our knowledge, no image retrieval system similar to ours, i.e. using annotations based on analogy, exists in the literature, hence a completely fair comparison was not possible. Being aware of this, we tried to accomplish an experimental comparison with the Google Image Search engine (<http://images.google.com/>), which is one of the most popular keyword based search engines. It should be noted that our system and Google Image Search use completely different methods for image annotation and retrieval. As well known, in Google the words attached to an image are from both surrounding text and top keywords in the query logs. Conversely, in our framework the words attached to an image are inherited from its prototype that is manually annotated according to human visual perception. Nevertheless, we chose Google Image Search for comparison since it is straight accessible and other researchers can easily compare our results with their experimental results.

For the comparison we considered two queries including geometric concepts, namely

Q_A : Fish round head

Q_B : Fish triangle head

and two queries including evocative concepts, namely:

Q_C : Fish fan tail

Q_D : Fish moon-like body

Each query was submitted to Google Image Search and data were collected by considering the set of images returned by Google for each query. To create a consistent dataset of fish shapes, a selection process was applied to the Google results in order to filter out spurious images (i.e. not depicting fishes) such as those depicted in fig.14. Also images including multiple fish shapes were removed. From the set of filtered Google images, we considered the top 50 images for each query. In this way, we collected a total of 200 images, each one containing a single shape of fish.

In order to create a reliable ground truth for this dataset, ten volunteers were asked to observe each fish image and classify it as relevant or not for each query, according to their perception. In this way, for each image, ten opinions were collected and the dominant opinion was finally considered as ground truth



Fig. 14 Spurious images removed from the results returned by Google Image Search (a) for query Q_A and (b) for query Q_B .

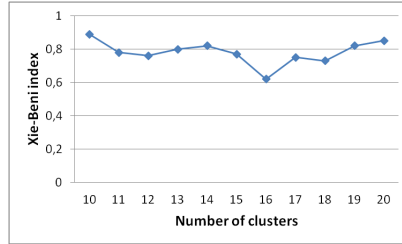


Fig. 15 The obtained values for the Xie-Beni index

for that image. Based on the ground truth, the number of relevant images for queries Q_A , Q_B , Q_C and Q_D is 101, 97, 36 and 93, respectively.

Successively, according to the scheme described in section 2, we created the knowledge base for our system, namely the prototypes and the visual ontology. All images were processed to extract shape contours and Fourier descriptors were computed to represent shape contours. Then the SSFCM algorithm was applied to derive a number of prototypes with related linguistic concepts. Different runs of SSFCM were performed by considering a different number of clusters in each run, ranging from 10 to 20. Results in terms of Xie-Beni validity index (fig. 15) show that the optimal number of clusters for this dataset is 16. For each cluster one prototype was derived and manually annotated with a linguistic concept. Fig. 16 shows the derived prototypes with their linguistic concepts. Next, the derived prototypes were divided into three parts (head, body and tail) and each part was annotated by using geometric and/or evocative concepts, according to the schema of our ontology.

Finally, the queries Q_A , Q_B , Q_C and Q_D were submitted to the search engine of our retrieval system and results were compared to the results of Google Image Search. Figures 17, 18, 19 and 20 depict results in terms of precision and recall computed according to the ground truth for each query. Plots of recall indicate also the maximum achievable value in correspondence of the Top 10, Top 20 and Top 30 retrieved images. It can be seen that our results are comparable or sometimes worse than Google results for queries using geometric concepts like *round* and *triangle*, while they are clearly better for queries using evocative concepts like *fan* and *moon-like*. To show this behavior, in fig. 21 and fig. 22 we summarize the top 10 results given by Google and by our system for query Q_B (using a geometric concept) and query Q_C (using an

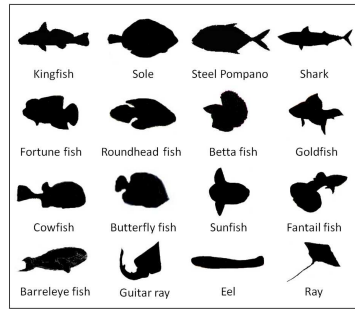


Fig. 16 The derived shape prototypes and their linguistic concepts

evocative concept), respectively. In these figures relevant images (according to the ground truth) are bordered in green, not relevant ones are bordered in red; for our system, the retrieved images include the prototypes (depicted as black silhouettes).

As shown in fig. 21 Google returns a high number of relevant images for query Q_B . This could be due to the common way adopted by the community to describe shapes, that is traditionally based on the use of geometric concepts. Moreover, for this query our system returns relevant prototypes (i.e. having a triangle shaped head) but 4 of the 10 returned images are not relevant. This is due to the automatic creation of clusters; indeed the clusters found by SSFCM are sometimes noisy and contain some shapes that differ from the prototypes in some parts. This may happen since during clustering the similarity among shapes is evaluated according to the whole contour rather than to the single parts.

A different behavior can be observed for query Q_C that contains the evocative concept *fan*. From fig. 22 it can be seen that our system succeeds in returning all relevant images, i.e. images containing fishes whose tail has a shape evocating a fan. Conversely, Google Image Search returns 5 images that are not relevant according to the ground truth. These images are retrieved by Google since their surrounding text contains the words “fan tail” (the name of that fish). However, according to the visual perception expressed by volunteers involved in the creation of the ground truth, such images are not relevant since they contain fishes whose tail does not evoke a fan.

On the overall, comparative results emphasize that for queries containing geometric concepts, our system and Google Image Search have a similar behavior, probably because geometric concepts are traditionally used to describe shapes. Conversely, when queries containing evocative concepts are considered, the results of our system seem to be better. This could be due to the fact that the mechanism of analogy, though more intuitive and immediate for an observer, is not very usual for describing digital images. Analogy-based annotations of images may improve the results of retrieval in the sense that they better reflect the human perception with respect to other types of descriptions used by state-of-art keyword-based image retrieval systems.

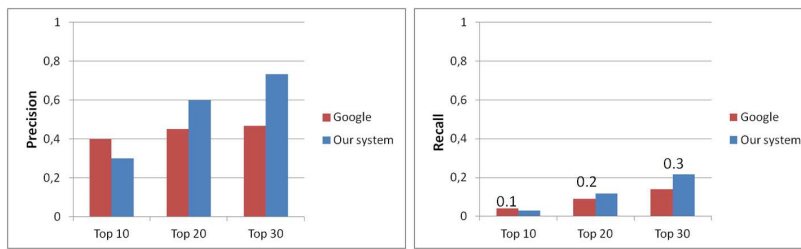


Fig. 17 Comparison of retrieval results for the query Q_A

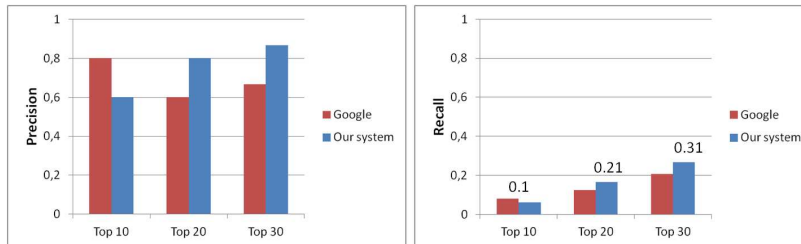


Fig. 18 Comparison of retrieval results for the query Q_B

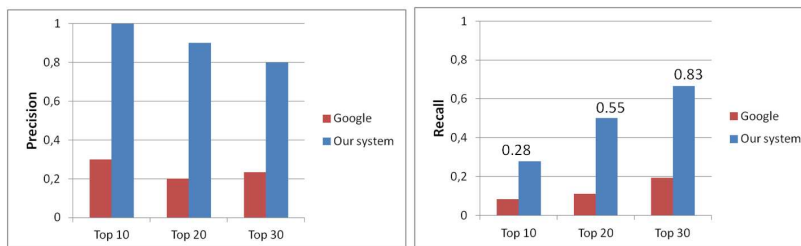


Fig. 19 Comparison of retrieval results for the query Q_C

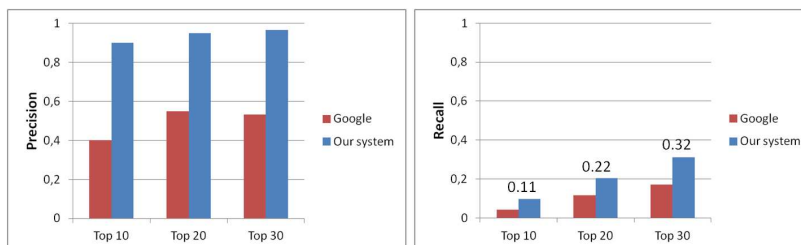


Fig. 20 Comparison of retrieval results for the query Q_D

5 Conclusions

In this paper, a shape annotation framework for intelligent image retrieval is proposed. Linguistic concepts and visual perceptual attributes are integrated together in a novel idea of visual ontology, that is properly designed to describe

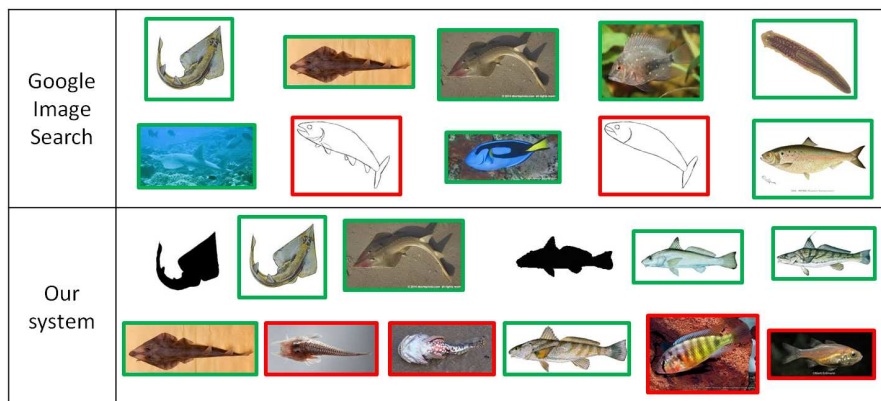


Fig. 21 Top 10 images returned by Google and by our system for query Q_B

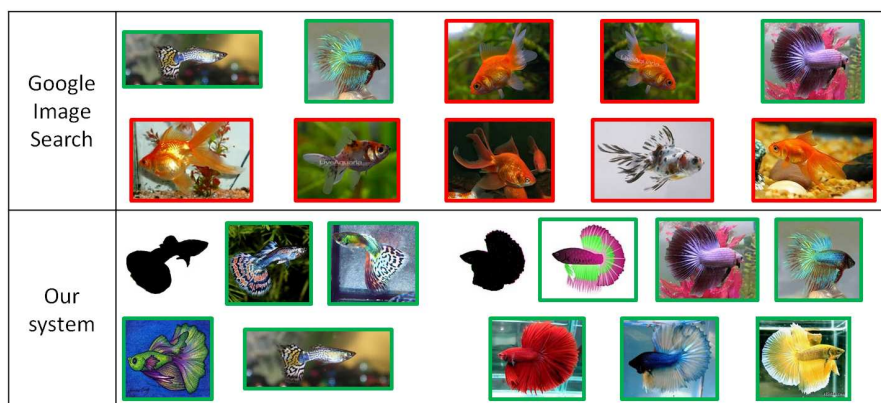


Fig. 22 Top 10 images returned by Google and by our system for query Q_C

shapes at different levels of representation. Sided to geometrical concepts, we consider evocative concepts that arise from recalling similar shapes using a mechanism of analogy to give a more immediate and natural description of the shape to be described.

Main features of the proposed framework are (i) the use of shape as the visual attribute for describing objects contained in images, (ii) the discovery of linguistic concepts and shape prototypes by an effective semi-supervised fuzzy clustering process, (iii) the design of a novel visual ontology to describe the derived prototypes by mimicking the analogy mechanism used by humans to describe parts of a shape.

As a proof-of-concept, a prototype of the visual ontology has been developed and tested on the fish shape domain to verify the effectiveness of the proposed framework for image retrieval. Preliminary results show that our framework is a good tool to retrieve relevant shapes starting from both linguistic and perceptual descriptions. These results foster the development of a

larger version of the ontology, including shapes from several different domains and enabling the annotation of large image collections.

Since the use of analogy-based perceptual descriptions to annotate images is fully innovative, assessing the validity of our framework through experimental comparisons with existing techniques is a hard task. Nevertheless we tried to compare our retrieval system with a popular image search engine, i.e. the Google Image Search system. Comparative results highlight that analogy-based annotations often provide results that are more adherent to the human perception of shapes. Hence the proposed framework represents a promising scheme for image annotation that is complementary to traditional annotation schemes. An integration of our framework with traditional annotation schemes could lead to the development of more powerful image retrieval systems.

Acknowledgements Funding for this work was provided by the Fondazione Cassa di Risparmio di Puglia (FCRP), that supported the Italian project “Annotazione di forme per la ricerca intelligente di immagini”.

References

1. Abbasi, S., Mokhtarian, F., Kittler, J.: Squid demo dataset: <http://www.ee.surrey.ac.uk/cvssp/demos/css/demo.html>
2. Al-Khatib, W., Day, Y.F., Ghafoor, A., Berra, P.B.: Semantic modeling and knowledge representation in multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):64–80, 1999.
3. Apache-Software-Foundation. <http://incubator.apache.org/jena/index.html>
4. Athanasiadis, T., Mylonas, P., Avrithis, Y., Kollias, S.: Semantic Image segmentation and object labeling. *IEEE Transaction on Circuits and Systems for Video Technology*. **17**(3) 298–312 (2007)
5. Ballan, L., Bertini, M., Del Bimbo, A., Serra, G.: Semantic annotation of soccer videos by visual instance clustering and spatial/temporal reasoning in ontologies. *Multimedia Tools and Applications*. **48**(2) 313–337 (2010)
6. Bartolini, I., Ciaccia, P., Patella, M.: WARP: Accurate retrieval of shapes using phase of Fourier descriptors and Time warping distance. *IEEE Trans. on Pattern Analysis and machine Intelligence*, **27**(1), 142–147 (2005)
7. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4), 509–522 (2002)
8. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic Web. *Scientific American*, 284(5), 28–37, (2001)
9. Bertini, M., Bimbo, A.D., Serra, G., Torniai, C.: Dynamic pictorially enriched ontologies for digital video libraries. *IEEE MultiMedia* 2009, 16, 2009.
10. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA (1981)
11. Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, Y., Staab, S., Strintzis, M.: Semantic annotation of images and videos for multimedia analysis. In A. Gmez-Prez and J. Euzenat, editors, *The Semantic Web: Research and Applications*, volume 3532 of *Lecture Notes in Computer Science*, pages 592–607, Springer Berlin / Heidelberg, 2005.
12. Borrás, A., Lladós, J.: Object Image Retrieval by Shape Content in Complex Scenes Using Geometric Constraints. *Pattern Recognition and Image Analysis*. Springer Berlin/Heidelberg. 325–332 (2005)
13. Bouet, M., Aufaure, M.-A.: New image retrieval principle: Image mining and visual ontology. In V. A. Petrushin and L. Khan, editors, *Multi-media Data Mining and Knowledge Discovery*, pages 168–184, Springer London, 2007.

14. Carlin, M.: Measuring the performance of shape similarity retrieval methods. *Computer Vision and Image Understanding*, **84**(1), 44–61 (2001)
15. Castellano, G., Fanelli, A. M., Torsello, M. A.: A fuzzy set approach for shape-based image annotation. In *Lecture Notes on Artificial Intelligence*, LNAI 6857. Springer-Verlag, 236–243 (2011)
16. Castellano, G., Fanelli, A. M., Torsello, M. A.: Fuzzy image labeling by partially supervised shape clustering. In *Lecture Notes on Artificial Intelligence*, LNAI 6882. Springer-Verlag, 84–93 (2011)
17. Castellano, G., Fanelli, A. M., Torsello, M. A.: Shape annotation by semi-supervised fuzzy clustering. *Information Sciences*, (289) 148–161, (2014).
18. Chander, V., Tapaswi, S.: Shape Based Automatic Annotation and Fuzzy Indexing of Video Sequences. In *Proc. of the 2010 IEEE/ACIS 9th International Conference on Computer and Information Science (ICIS '10)*, pp. 222–227 (2010)
19. Chen, Y., Wang, J.Z.: A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. **40**(9), 1252–1267 (2002)
20. Chuang, K., Tzenget, H.L., Chen, S., Wu, J., Chen, T.J.: Fuzzy c-means clustering with spatial information for image segmentation. *IEEE Trans. on Systems, Man and Cybernetics* 34(4), 1907–1916 (2004)
21. Datta, R., Dhiraj, J., Jia, L., Wang, J.Z.: *Image Retrieval: Ideas, Influences, and Trends of the New Age*. ACM Computing Surveys. 40, 1–60 (2008)
22. Eakins, J. P.: Towards intelligent image retrieval. *Pattern Recognition*. 35(1), 3–14 (2002)
23. Akbas, E., and Vural, F. Y.: Automatic Image Annotation by Ensemble of Visual Descriptors. In *Proc. of Conf. on Computer Vision (CVPR) 2007, Workshop on Semantic Learning Applications in Multimedia 1–8* (2007)
24. Fan, J., Gao, Y., Luo, H.: Hierarchical classification for automatic image annotation. In *Proc. of the 30th annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 111–118 (2007)
25. Gardner, H.: *The minds new science: A history of the cognitive revolution*. Basic Books, New York, NY, paperback edition, 1985.
26. Gonzalez, R.C., Woods, R.E. *Digital Image Processing*. Addison-Wesley, 1992.
27. Haralick, R.: Statistical and Structural Approaches to Texture, *Proceedings of the IEEE*,67(5), 786–804, 1979.
28. Inoue, M.: On the need for annotation-based image retrieval. In *Proc. of the Workshop on Information Retrieval in Context* pp. 44–46 (2004)
29. Yankov, D., Keogh, E.: Manifold Clustering of Shapes. In *Proc. of the 6th International Conference on Data Mining*, pp. 1167– 1171 (2006)
30. Hauptmann, A., Rong, Yan, Lin, W.-H., Christel, M., Wactlar, H.: Can High-Level Concepts Fill the Semantic Gap in Video Retrieval? A Case Study With Broadcast News. *IEEE Transactions on Multimedia* 9(5) 958–966 (2007)
31. Jan, J.: *Medical image processing, reconstruction, and restoration: concepts and methods*. Signal processing and communications. CRC Press, 2006.
32. Jiang, S.Q., Du, J., Huang, Q. M., Huang, T.J., Gao, W.: Visual ontology construction for digitized art image retrieval. *Journal of Computer Science and Technology*, 20(6) 855–860 (2005)
33. Klassen, E., Srivastava, A., Mio, W., Joshi, S. H.: Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3) 372–383 (2004)
34. Kompatsiaris, I., Mezaris, V., Strintzis, M.G.: Multimedia content indexing and retrieval using an object ontology. In *Multimedia Content and the Semantic Web* 339–371 (2005)
35. Kotschieder, P., Donoser, M., Bischof, H.: Beyond Pairwise Shape Similarity Analysis. In *Proc. of Asian Conference on Computer Vision (ACCV)*, pp.655–666 (2009)
36. Lew, M., Sebe, N., Djeraba, C., Lifl, F., Ramesh, J.: *Content-based Multimedia Information Retrieval: State of the Art and Challenges*. ACM Transactions on Multimedia Computing, Communications, and Applications. 1–19 (2006)
37. Li, D., Simske, S.: Shape retrieval based on distance ratio distribution. HP Tech Report. HPL-2002-251, 2002.

38. Liang, X., Zhuang, Q., Cao, N., Zhang, J.: Shape modeling and clustering of white matter fiber tracts using Fourier descriptors. In Proc. of the 6th Annual IEEE conference on Computational Intelligence in Bioinformatics and Computational Biology, pp. 292–297 (2009)
39. Liu, Y., Zhang, D., Lu, G., Ma, W.-Y.: A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282 (2007)
40. Liu, Y., Zhang, J., Tjondronegoro, D., Geve, S.: A shape ontology framework for bird classification. In Proc of 9th Biennial Conference of the IEEE Australian Pattern Recognition Society, pp. 478–484 (2007)
41. Liu, Y., Zhang, J., Tjondronegoro, D., Geve, S., Li, Z.: Mid-level concept learning with visual contextual ontologies and probabilistic inference for image annotation. In S. Boll, Q. Tian, L. Zhang, Z. Zhang, and Y.-P. Chen, editors, *Advances in Multimedia Modeling*, volume 5916 of *Lecture Notes in Computer Science*, pages 229–239, Springer Berlin/Heidelberg, 2010.
42. Maillot, N.E., Thonnat, M.: Ontology based complex object recognition. *Image and Vision Computing*, 26(1), 102–113 (2008)
43. Mylonas, P., Spyrou, E., Avrithis, Y., Kollias, S.: Using Visual Context and Region Semantics for High-Level Concept Detection. *IEEE Transactions on Multimedia* 11(2), 229–243 (2009)
44. Muda, Z.: Classification and Image Annotation for Bridging the Semantic Gap. In: *Summer School on Multimedia Semantics 2007*, University of Glasgow, Glasgow (2007)
45. Muller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications: clinical benefits and future directions. *International journal of medical informatics*, 73(1), 1–23 (2004)
46. Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., Yanker, P.: The Qbic project: Querying images by content using color, texture, and shape. In Proc. of SPIE Conference on Storage and Retrieval of Image and Video Databases, pp. 1–8 (1993)
47. Noy, N.F., Sintek, M., Decker, S., Crubezy, M., Ferguson, R.W., Musen, M.A.: Creating semantic web contents with protege. *IEEE Intelligent Systems*, (16), 60–71 (2001)
48. Pavlidis, T. *Algorithms for Graphics and Image Processing*. Computer Science Press, 1982.
49. Pedrycz, W., Waletzky, J.: Fuzzy clustering with partial supervision. *IEEE Transaction System Man Cybernetics*, 27(5), 787–795 (1997)
50. Pedrycz, W., Amato, A., Di Lecce, V., Piuri, V.: Fuzzy Clustering with partial supervision in organization and classification of digital images. *IEEE Transactions on Fuzzy Systems* 16(4) 1008–1025 (2008)
51. Plataniotis, K.N., Venetsanopoulos, A. N.: *Color Image Processing and Applications*. Springer, Berlin, 2000.
52. Platt, J., Cristianini, N., Shawe-Taylor, J.: Large margin DAGs for multiclass classification. *Advances in Neural Information Processing Systems*. 12 547–553 (2000)
53. Rafiei, D., Mendelzon, A.O.: Efficient Retrieval of Similar Shapes, *The Very Large Data Bases Journal*. 11(1) 17–27 (2002)
54. Rajpoot, N.M., Arif, M.: Unsupervised Shape Clustering using Diffusion Maps. *The Annals of the BMVA*. (5) 1–17 (2008)
55. Rui, Y., Huang, T., Chang, S.: Image retrieval: current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 10(4) 39–62 (1999)
56. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3) 157–173 (2008)
57. Simou, N., Tzouvaras, V., Avrithis, Y., Stamou, G., Kollias, S.: A visual descriptor ontology for multimedia reasoning. In Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS2005), Montreux, Switzerland, pp. 13–15 (2005)
58. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 22, 1349–1380 (2000)
59. Stamou, G., Kollias, S.: *Multimedia Content and the Semantic Web: Standards, Methods and Tools*. John Wiley and Sons, 2005.

60. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F.: Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1), 14–18 (2006)
61. Veltkamp, R., Tanase, M.: Content-based image retrieval systems: a survey. Technical Report, 2001.
62. Wang, H., Liu, S., Chia, L.T.: Does ontology help in image retrieval?: a comparison between keyword, text ontology and multi-modality ontology approaches. In *Proc. of the 14th annual ACM international conference on Multimedia*, pp. 109–112 (2006)
63. Wu, G., Chang, E., Li, C.: SVM binary classifier ensembles for image classification. In *Proc. of ACM Conf. on Information and knowledge Management*, pp. 395–402 (2001)
64. Yang, M., Kpalma, K., Ronsin, J.: A survey of shape feature extraction techniques. *Pattern recognition*, 43–90, (2008)
65. Zhang, D. and Lu, G.: A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval. In *Proc. Fifth Asian Conf. Computer Vision* 646–651 (2002)
66. Zhang, D., Islam, M. M., Lu, G.: A review on automatic image annotation techniques. *Pattern Recognition*. 45(1) 346–362 (2011)