

ONTOLOGIES FOR OBJECT-BASED IMAGE RETRIEVAL *

V. MEZARIS, I. KOMPATSIARIS AND M. G. STRINTZIS

*Information Processing Laboratory
Electrical and Computer Engineering Dept.
Aristotle University of Thessaloniki
Thessaloniki 54006, Greece
e-mail: strintzi@eng.auth.gr*

*Informatics and Telematics Institute
Centre for Research and Technology Hellas,
1st Km Thermi-Panorama Rd,
Thessaloniki 57001, Greece*

In this paper, a novel approach to image retrieval is presented. The proposed approach employs a time-efficient and fully unsupervised segmentation algorithm to divide images into regions. Low-level indexing features for each region are subsequently extracted. These arithmetic features are automatically mapped to appropriate qualitative descriptors of the *object ontology*. The object ontology is also used for describing semantic objects (keywords). By querying for images conforming to the qualitative description of the desired object, clearly irrelevant image regions are rejected; following that, a relevance feedback mechanism, based on neural networks, is invoked to rank the remaining, potentially relevant regions and produce the final query results. The proposed approach bridges the gap between keyword-based approaches, which assume the existence of rich image captions, and query-by-example approaches, which assume that the user queries for images similar to one that already is at his disposal.

1. Introduction

In recent years, the growth of still image collections has established the need for human-centered image retrieval tools. The first attempts for image retrieval were based on exploiting image captions for extracting keywords; newer approaches rely on utilizing the image visual contents to find images similar to a key-image that is either selected from a restricted image set or

*This work was supported by the EU IST project SCHEMA. The help of cost211quat is also gratefully acknowledged.

is supplied by the user (*query-by-example*)^{1,2,3}.

This paper addresses the problem of retrieval in generic image collections without imposing restrictions such as the availability of key-images or image captions. The general architecture of the proposed retrieval scheme is developed in Sec. 2. Section 3 presents the low level indexing features. In Sec. 4, the use of ontologies is discussed. Section 5 contains an experimental evaluation of the developed methods, and finally, conclusions are drawn in Sec. 6.

2. System Architecture

In this paper, an object-based approach to image retrieval has been adopted; thus, the process of inserting an image into the database starts by applying a segmentation algorithm⁴ to it, so as to break it down to a number of regions. Following that, low-level indexing features compactly describing the color, position and shape of each region are calculated. These are subsequently translated to intermediate-level descriptors qualitatively describing the region attributes, that humans are more familiar with. The intermediate-level descriptors that can be used for this qualitative description form a simple vocabulary termed *object ontology*. Since these intermediate-level descriptors roughly describe the region, they will be used only for ruling out regions that are irrelevant to the semantic objects desired by the user in a given query, while accurate region ranking will still be based on the low-level features. Nevertheless, the user has to manipulate only intermediate-level descriptors, in contrast to most other systems.

For the proposed system to be able to associate high-level concepts, i.e. semantic objects (each represented by a keyword, e.g. "tiger"), to the images in the database, one has to additionally supply a description of each keyword, formulated using the intermediate-level descriptors of the object ontology. In this way, keywords and image regions can be associated by comparing their intermediate-level descriptors; this is the first step in executing a query. The output of this step is a set of potentially relevant image regions, whose relevance cannot be quantitatively expressed at this point; these are presented to the user at random order. The user then evaluates one or more pages of image regions, checking the appropriate "relevant" box where appropriate. By submitting this relevance feedback, a back-propagation neural network⁵ is trained and subsequently ranks using the low-level features, all regions returned by the initial query. Corresponding

images are then presented to the user ordered by rank.

3. Low-level Indexing Features

As soon as the segmentation is performed, using the methodology developed in Ref. 4, a set of descriptors that will be useful in querying the database are calculated for each region. These region descriptors compactly characterize each region's color, position and shape. All descriptors are normalized so as to range from 0 to 1.

Let $\mathbf{p} = [p_x \ p_y]^T$ be a pixel and let s_k be an image region made of A_k pixels. The pixel intensity components in the CIE L*a*b* color space are denoted $I_L(\cdot)$, $I_a(\cdot)$ and $I_b(\cdot)$.

The **color and position descriptors** of a region are the normalized intensity and spatial centers of the region. In particular, the color descriptors of region s_k , $\bar{I}_{k,L}$, $\bar{I}_{k,a}$, $\bar{I}_{k,b}$, are defined as follows:

$$\bar{I}_{k,L} = \frac{1}{100 \cdot A_k} \sum_{\mathbf{p} \in s_k} I_L(\mathbf{p}), \quad \bar{I}_{k,q} = \frac{\frac{1}{A_k} \sum_{\mathbf{p} \in s_k} I_q(\mathbf{p}) + 80}{200}, \quad q \in \{a, b\}$$

Similarly, the position descriptors $\bar{S}_{k,x}$, $\bar{S}_{k,y}$ are defined as:

$$\bar{S}_{k,q} = \frac{1}{A_k \cdot p_{q,max}} \sum_{\mathbf{p} \in s_k} p_q, \quad q \in \{x, y\},$$

where $p_{x,max}$, $p_{y,max}$ are the image dimensions in pixels.

The **shape descriptors** of a region are its normalized area and eccentricity. The normalized area E_k^N is expressed by the number of pixels A_k that belong to region s_k , divided by the total number of pixels of the image. The eccentricity is calculated as $\varepsilon_k = 1 - \frac{\rho_1}{\rho_2}$, where ρ_1 , ρ_2 , $\rho_1 \geq \rho_2$ are the eigenvalues of the region covariance matrix. The normalized eccentricity is then defined as $\varepsilon_k^N = e^{\varepsilon_k}$.

The seven region descriptors defined above form a region descriptor vector \mathbf{F}_k^{ID} ,

$$\mathbf{F}_k^{ID} = [F_{k,1}^{ID} \ \dots \ F_{k,7}^{ID}] = [\bar{I}_{k,L} \ \bar{I}_{k,a} \ \bar{I}_{k,b} \ \bar{S}_{k,x} \ \bar{S}_{k,y} \ E_k^N \ \varepsilon_k^N], \quad (1)$$

where k is the region number and ID is a unique image identity. This region descriptor vector will be used in the sequel both for assigning intermediate-level descriptors to each region, to allow for association between keywords representing semantic objects and potentially relevant regions, and as input to the relevance feedback mechanism. In both cases the existence of these low-level features is not apparent to the end user.

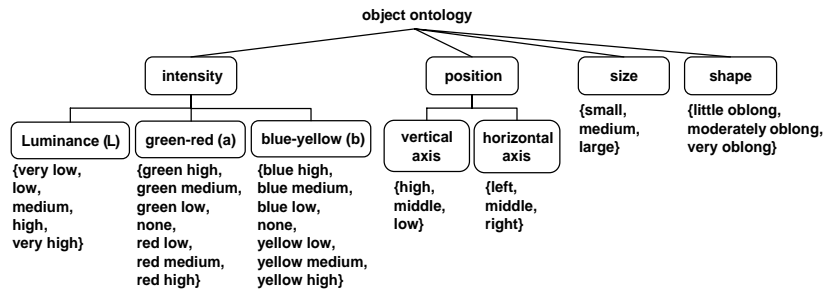


Figure 1. Object ontology

4. Object Ontology

Ontologies are tools for structuring knowledge⁶. An ontology may be defined as the specification of a representational vocabulary for a shared domain of discourse which may include definitions of classes, relations, functions and other objects⁷. Ontologies are primarily used in text retrieval. In this paper, an ontology termed *object ontology* is employed to allow the user to query an image collection using keywords representing semantic objects. Semantic objects, like “tiger”, are described using the intermediate-level descriptors of the object ontology. These descriptors are automatically mapped to the low-level features calculated for each region in the database, thus allowing the association of keywords and relevant image regions.

The object ontology is presented in Fig. 1, where the possible descriptors $D_{q,z}$, $q = 1, \dots, Q$ for each one of the object attributes A_z , $z = 1, \dots, 7$ (e.g. luminance) can be seen. Note that the number of descriptors Q is not equal for all attributes. Each one of these intermediate-level descriptors is statically mapped to an appropriate range of values of the corresponding low-level, arithmetic feature. The various value ranges $R_{q,z} = [L_{q,z}, H_{q,z}]$, $q = 1, \dots, Q$ for every low-level feature F_z , $z = 1, \dots, 7$ of the region descriptor vector \mathbf{F} (defined for a region in Eq. (1)), are chosen so that the resulting intervals are equally populated. This is pursued so as to prevent a descriptor from being associated with a majority of image regions in the database, because this would make it useless in restricting a query to the potentially most relevant images. Overlapping, up to a point, of adjacent value ranges, is used to introduce a degree of fuzziness to the descriptors; for example, both “low luminance” and “medium luminance” descriptors may be used to describe a single region.

5. Experimental Results

The proposed methodology was tested on a collection of 5000 images from the Corel library⁸. Application of the segmentation algorithm of Ref. 4 to these images resulted to the creation of a database containing 34433 regions. Using this database, several queries were submitted to retrieve images belonging to particular classes, e.g. images containing eagles, red cars, tigers, etc. Precision-recall diagrams after the application of relevance feedback for two classes of queries are presented in Fig. 2, along with corresponding results using the query by example paradigm and global image histograms. On comparing the precision-recall diagrams of the two methodologies, it can be seen that the proposed scheme performs significantly better, in addition to being more flexible than conventional methodologies. Additional results are presented in Fig. 3.

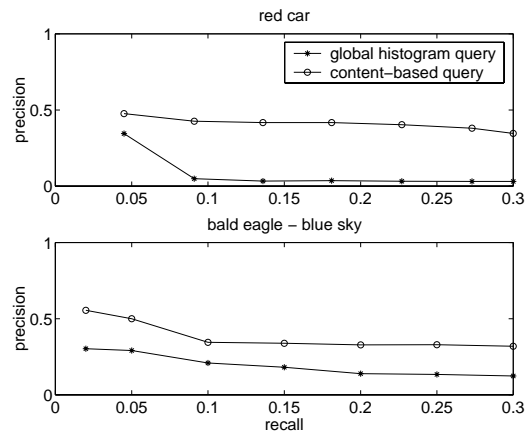


Figure 2. Precision - Recall diagrams for one single- and one dual-keyword query and comparison with global histogram method

6. Conclusions

A methodology was presented for the flexible and user-friendly retrieval of color images, combining image processing, knowledge structuring and machine learning tools. The proposed methodology overcomes the restrictions of conventional methods, such as the restricted vocabulary or the need for the availability of key-images, and requires no manual tuning of weights.

The resulting scheme is therefore appropriate for querying large collections of still images.

References

1. C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and effective querying by image content," *Journal of Intelligent Information Systems*, vol. 3, no. 3/4, pp. 231–262, 1994.
2. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, 2002.
3. I. Kompatsiaris, E. Triantafillou, and M. G. Strintzis, "Region-Based Color Image Indexing and Retrieval," in *Proc. IEEE ICIP*, October 2001.
4. V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "A Framework for the Efficient Segmentation of Large-format Color Images," in *Proc. IEEE ICIP*, Rochester, NY, USA, September 2002.
5. L.V. Fausett, *Fundamentals of neural networks*, Prentice Hall, 1994.
6. B. Chandrasekaran, J.R. Josephson, and V.R. Benjamins, "What are ontologies, and why do we need them?," *IEEE Intelligent Systems*, vol. 14, no. 1, pp. 20–26, Jan.-Feb. 1999.
7. T. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, 1993.
8. *Corel stock photo library*, Corel Corp., Ontario, Canada.

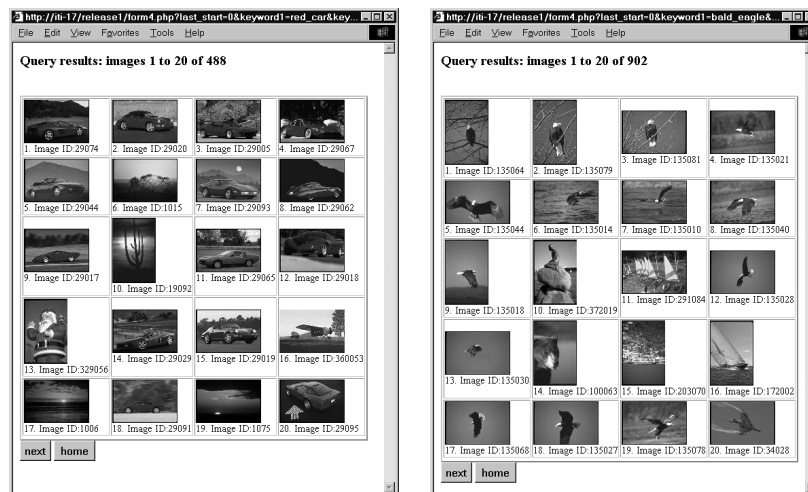


Figure 3. Results for "red car" and "bald eagle - blue sky" queries