AN ONTOLOGY APPROACH TO OBJECT-BASED IMAGE RETRIEVAL

Vasileios Mezaris[1,2], *Ioannis Kompatsiaris*[2], *and Michael G. Strintzis*[1,2]

[1]Information Processing Laboratory
Electrical and Computer Engineering Department
Aristotle University of Thessaloniki
Thessaloniki 54006, Greece

[2]Informatics and Telematics Institute
1st Km Thermi-Panorama Rd,
Thessaloniki 57001, Greece
e-mail: strintzi@eng.auth.gr

ABSTRACT

In this paper, an image retrieval methodology suited for search in large collections of heterogeneous images is presented. The proposed approach employs a fully unsupervised segmentation algorithm to divide images into regions. Low-level features describing the color, position, size and shape of the resulting regions are extracted and are automatically mapped to appropriate intermediate-level descriptors forming a simple vocabulary termed *object ontology*. The object ontology is used to allow the qualitative definition of the high-level concepts the user queries for (*semantic objects*, each represented by a *keyword*) in a human-centered fashion. When querying, clearly irrelevant image regions are rejected using the intermediate-level descriptors; following that, a relevance feedback mechanism employing the low-level features is invoked to produce the final query results. The proposed approach bridges the gap between keyword-based approaches, which assume the existence of rich image captions or require manual evaluation and annotation of every image of the collection, 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 accelerated growth of digital media collections and in particular still image collections, both proprietary and on the Web, has established the need for the development of human-centered tools for the efficient access and retrieval of visual information. The very first attempts for image retrieval were based on exploiting existing image captions; although relatively simple and computationally efficient, this approach has several restrictions mainly deriving from the use of a restricted vocabulary that neither allows for unanticipated queries nor can be easily extended. Additionally, manual image annotation is necessary if no captions exist.

To overcome the limitations of the keyword-based approach, the use of the image visual contents has been proposed. This category of approaches utilizes the visual contents by extracting low-level indexing features for each image or image segment (region). Then, relevant images are retrieved by comparing the low-level features of each item in the database with those of a key-image that is either selected from a restricted image set or is supplied by the user (*query-by-example*) [1, 2, 3]. A major drawback of such approaches is that, in order to start a query, the availability of an appropriate key-image is assumed; occasionally, this is not feasible, particularly for image classes that are under-represented in the database.

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 section 2. Section 3 presents the low level indexing features. In section 4, the use of ontologies and their importance in associating low-level and high-level features in a flexible manner are discussed. The employed relevance feedback technique is presented in section 5. Section 6 contains an experimental evaluation of the developed methods, and finally, conclusions are drawn in section 7.

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 color image segmentation algorithm [4] to it, so as to break it down to a number of regions. Following that, a set of low-level indexing features is calculated for each formed region. These arithmetic features compactly describe the color, position, size and shape of the region.

The low-level indexing features are machine-centered rather than human-centered; for this reason, they are subsequently translated to intermediate-level descriptors qualitatively describing the region attributes, that humans are more familiar with. These intermediate level descriptors form a simple vocabulary termed *object ontology*. Since these descriptors only roughly describe the region, as opposed to the low-level features, they will be used only for ruling out regions that are irrelevant to the ones desired by the user in a given query, while accurate region ranking will still be based on the low-level features. Nevertheless, the whole system is designed so as to hide the existence of low-level features from the user; thus 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, such as object names or other keywords, to the images in the database, one has to additionally supply a definition 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 (Fig. 1), made of one or more keywords (objects) already described using the object ontology vocabulary or defined during the query procedure. The output of this query is a set of potentially relevant images, whose relevance can-

This work was supported by the EU IST project SCHEMA. The assistance of COST211 quat is also gratefully acknowledged.

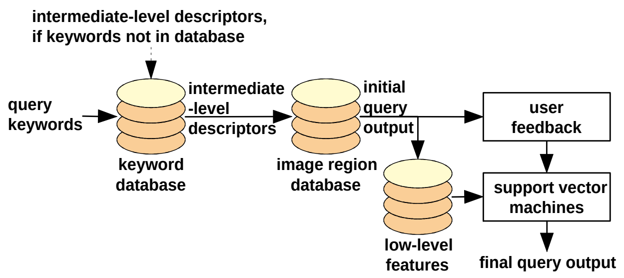

Fig. 1. Overview of the query process

not be quantitatively expressed at this point. Therefore, they are presented to the user at random order. The user then evaluates one or more pages of images, marking relevant image regions simply by checking the appropriate "relevant" box. By submitting this relevance feedback, one or two support vector machines are trained and subsequently rank according to relevance all regions returned by the initial query, using their low-level features. 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 [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 have been normalized so as to range from 0 to 1.

Let $\mathbf{p} = [p_x \; p_y]$ 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 $I_L(.)$, $I_a(.)$ and $I_b(.)$.

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, $I^N_{L,k}, I^N_{a,k}, I^N_{b,k}$, are defined as follows:

$$I^N_{L,k} = \frac{1}{100 \cdot A_k} \sum_{\mathbf{p} \in s_k} I_L(\mathbf{p}),$$

$$I^N_{q,k} = \frac{\frac{1}{A_k} \sum_{\mathbf{p} \in s_k} I_q(\mathbf{p}) + 80}{200}, \;\; q \in \{a, b\}$$

Similarly, the position descriptors $S^N_{x,k}, S^N_{y,k}$ are defined as:

$$S^N_{q,k} = \frac{1}{A_k \cdot p_{q,max}} \sum_{\mathbf{p} \in s_k} p_q, \;\; q \in \{x, y\},$$

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^N_k 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: $E^N_k = \frac{A_k}{p_{x,max} \cdot p_{y,max}}$. 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: $\varepsilon^N_k = e^{\varepsilon_k}$.

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

$$\mathbf{I}^{ID}_k = [I^N_{L,k}, I^N_{a,k}, I^N_{b,k}, S^N_{x,k}, S^N_{y,k}, E^N_k, \varepsilon^N_k], \tag{1}$$

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 qualitative descriptors to the region, to allow for association between low-level features and high-level concepts, 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.

4. ONTOLOGIES AND HIGH-LEVEL FEATURES

Ontologies are recently-introduced tools for structuring knowledge [5]. 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 objects [6]. Ontologies are primarily used in text retrieval. In this paper, an ontology termed *object ontology* is employed to allow the user to query a generic image collection, where no domain-specific knowledge can be employed, using high-level concepts (keywords representing semantic objects). High-level concepts, 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 high-level concepts and potentially relevant image regions. The simplicity of the employed object ontology serves the purpose of it being applicable to generic image collections without requiring the correspondence between image regions and relevant descriptors be defined manually.

The object ontology is presented in Fig. 2, where the possible descriptors for each of the employed object attributes (e.g. luminance) can be seen. Each one of these intermediate-level descriptors is mapped to an appropriate range of values of the corresponding low-level, arithmetic feature. The various value ranges for every low-level feature 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 fuzzyness to the descriptors; for example, both "low luminance" and "medium luminance" descriptors may be used to describe a single region.

Let $D_{q,z}$ be the q-th descriptor (e.g. "low luminance") of the z-th ontology attribute (e.g. "luminance") and $R_{q,z} = [L_{q,z}, H_{q,z}]$ be the range of values corresponding to that descriptor. Given the probability density function $pdf(x_z)$ of random variable x_z corresponding to the z-th element of region descriptor vector $\mathbf{I}$ (equation (1)), given the overlapping factor V expressing the degree of overlapping of adjacent value ranges, and given that value ranges should be equally populated, lower and upper bounds $L_{q,z}$, $H_{q,z}$ can be easily calculated according to equations (2) to (5),

$$L_{1,z} = L_z, \tag{2}$$

$$\int_{L_{q-1,z}}^{L_{q,z}} pdf(x_z)dx_z = \frac{1-V}{Q_z - V \cdot (Q_z - 1)}, \; q = 2, \ldots, Q_z, \tag{3}$$

$$\int_{L_{1,z}}^{H_{1,z}} pdf(x_z)dx_z = \frac{1}{Q_z - V \cdot (Q_z - 1)}, \tag{4}$$

$$\int_{H_{q-1,z}}^{H_{q,z}} pdf(x_z)dx_z = \frac{1-V}{Q_z - V \cdot (Q_z - 1)}, \; q = 2, \ldots, Q_z, \tag{5}$$

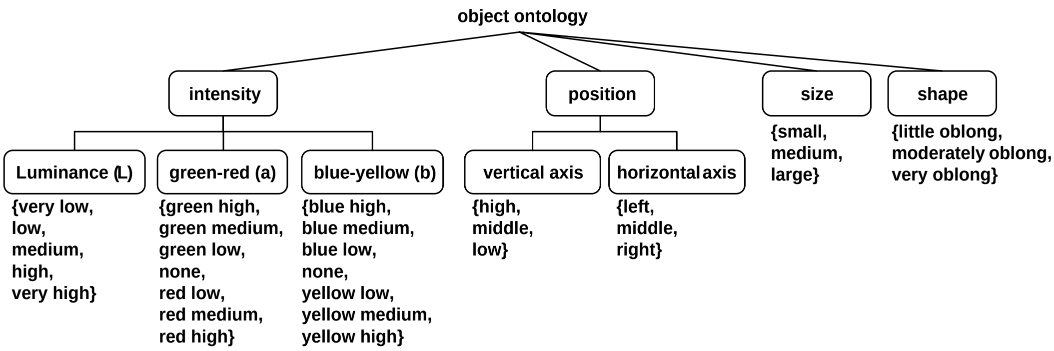

Fig. 2. Object ontology

where Q_z is the number of descriptors defined for the z-th ontology attribute (for example, for "luminance", $Q_z = 5$), and L_z is the lower bound of values of the random variable x_z. Note that for attributes "green-red" and "blue-yellow", the above process is performed twice: once for each of the two complementary colors described by each attribute, taking into account each time the appropriate range of values of the corresponding low-level feature. Lower and upper bounds for descriptor "none" of attribute "green-red" are chosen so as to associate with this descriptor a fraction V of the population of descriptor "green low" and a fraction V of the population of descriptor "red low"; bounds for descriptor "none" of attribute "blue-yellow" are defined accordingly. The overlapping factor V is defined as $V = 0.25$ in our experiments.

5. RELEVANCE FEEDBACK

Since the intermediate-level descriptors used for the formation of the initial query output roughly describe the desired region, they can only by used for excluding obviously undesirable regions, thus narrowing down the search to a set of potentially relevant image regions. Qualitative evaluation of the degree of relevance of each image region s_k has to be performed using the low-level descriptor vector $\mathbf{I}_k$ and additional user input in the form of a few manually evaluated image regions, a technique known as *relevance feedback*.

Under the proposed scheme, the first few image regions that are presented to the user as initial results of the query are evaluated by the user and are or are not marked as relevant; these are used to train a support vector machine (SVM); subsequently, a constrained similarity measure (CSM) [7] is utilized for ranking the images. The user intervention is restricted, under the proposed scheme, to marking appropriate image regions as relevant, as opposed to fine-tuning several unintuitive weights, which is common practice in query-by-example schemes.

Thus, all image regions contained in the initial query output are ranked, where the output of the constrained similarity measure serves as rank; they are subsequently presented to the user ordered by rank, in descending order.

In the case of two-keyword queries, two different SVMs are trained, each using the training set corresponding to one keyword, and the image rank is calculated as the sum of the two ranks calculated using the CSM for the two potentially relevant regions of each examined image.

6. EXPERIMENTAL RESULTS

The proposed algorithms were tested on a collection of 5000 images from the Corel library [8]. Application of the segmentation algorithm of [4] to these images resulted to the creation of a database containing 34433 regions, each represented by a low-level feature vector, as discussed in section 3. Following the creation of the region low-level-feature database, the mapping between these low-level features and the intermediate-level features defined by the object ontology was performed, as discussed in section 4.

The next step in experimenting with the proposed system was to use the object ontology to define keywords describing high-level concepts. Subsequently, these keywords were used to form and submit queries. Several experiments were conducted using single-keyword or dual-keyword queries, to retrieve images belonging to particular classes, e.g. images containing eagles, red cars, tigers, etc. In most experiments class population was 100 images; under-represented classes were also used. Performing ontology-based querying resulted in initial query results being produced by excluding the majority of regions in the database, that were found to be clearly irrelevant.

As a result, one or more pages of twenty randomly-selected, potentially relevant image regions were presented to the user to be manually evaluated; usually, evaluating two such pages was found to be sufficient. In all experiments, each query was submitted five times, to accommodate for varying performance due to different randomly chosen image sets being presented to the user for evaluation, for the purpose of relevance feedback. Precision-recall diagrams after the application of relevance feedback for two classes of queries are presented in Fig. 3, along with corresponding results using the query by example paradigm and global image histograms. On comparing the results of the two methodologies, it can be seen that the proposed scheme performs significantly better, in addition to being more flexible than conventional methodologies, as already discussed. Sample results after one round of relevance feedback are presented in Figs. 4 and 5; these can be further improved by repeating the relevance feedback procedure (Fig. 3).

7. CONCLUSIONS

An entirely novel methodology was presented for the flexible and user-friendly retrieval of color images, combining state of the art

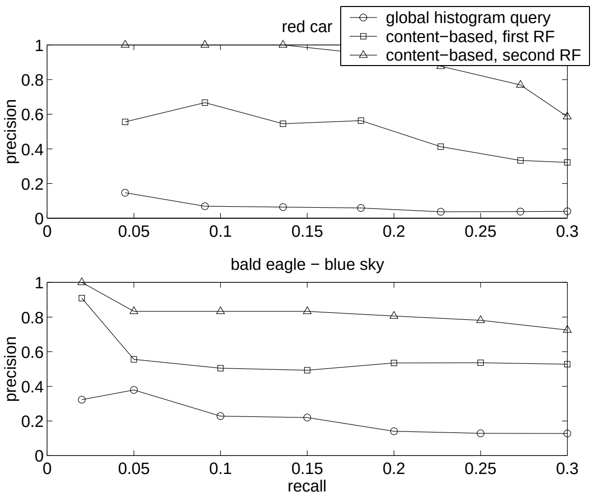

Fig. 3. Precision - Recall diagrams for one single- and one dual-keyword query, both after one and two rounds of relevance feedback, and comparison with global histogram method

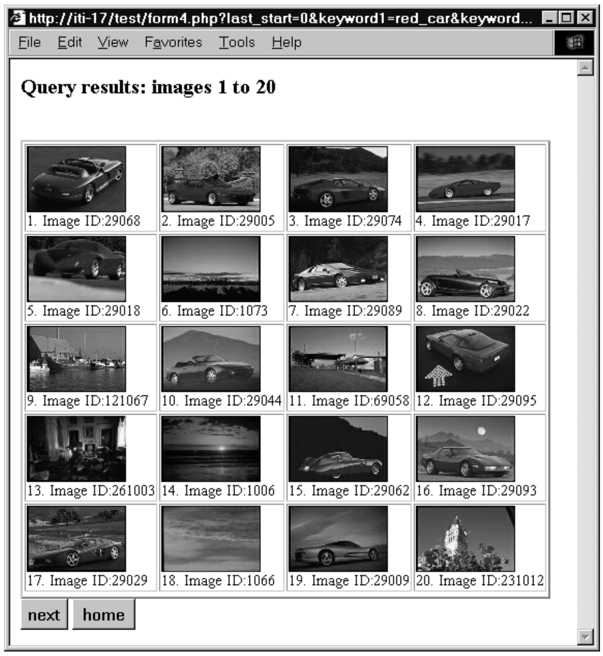

Fig. 4. Result images 1 to 20 for a "red car" query, after one round of relevance feedback

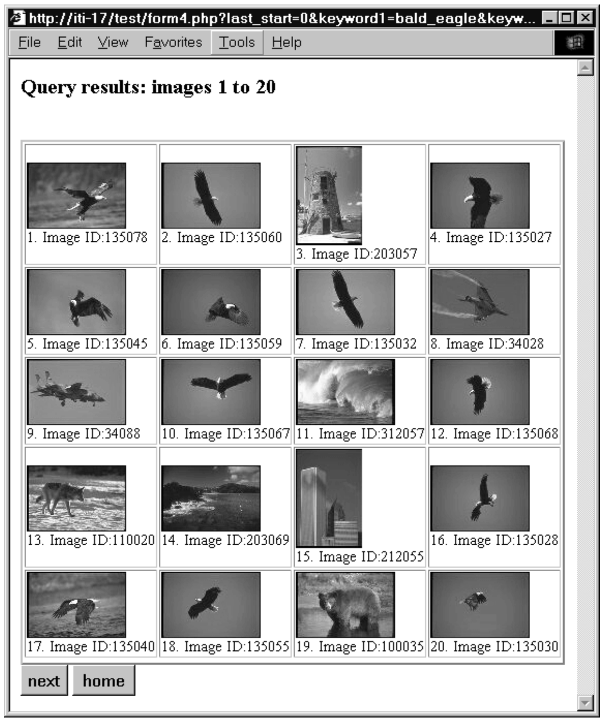

Fig. 5. Result images 1 to 20 for a "bald eagle - blue sky" query, after one round of relevance feedback

image analysis and knowledge organization tools. The resulting 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.

8. 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*, Thessaloniki, Greece, 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] 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.

[6] T. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, 1993.

[7] G.-D. Guo, A.K. Jain, W.-Y. Ma, and H.-J. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 811–820, July 2002.

[8] *Corel stock photo library*, Corel Corp., Ontario, Canada.