Combining multiple segmentation algorithms and the MPEG-7 eXperimentation Model in the Schema Reference System *

Vasileios Mezaris[1,2], Haralambos Doulaverakis[2], Raul Medina Beltran de Otalora[3], Stephan Herrmann[3], Ioannis Kompatsiaris[2], and Michael G. Strintzis[1,2]

[1]Information Processing Laboratory, Electrical and Computer Engineering Department
Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

[2]Informatics and Telematics Institute (ITI)
Centre for Research and Technology Hellas (CERTH), Thessaloniki 57001, Greece

[3]Institute for Integrated Systems
Munich University of Technology, Munich D-80290, Germany

Abstract

The SCHEMA Network of Excellence aims to bring together a critical mass of universities, research centers, industrial partners and end users, in order to design a reference system for content-based semantic scene analysis, interpretation and understanding. In this paper, recent advances in the development of the SCHEMA reference system are reported, focusing on the application of region-based image retrieval using automatic segmentation. More specifically, the integration of four segmentation algorithms and the MPEG-7 eXperimentation Model with the reference system are discussed, along with the motivation behind these and various other choices that were made during the development of the reference system. Experimental results for this system, as well as results for an earlier version of it employing proprietary descriptors, are shown using a common collection of images. Comparative evaluation of these versions, both in terms of retrieval accuracy and in terms of time-efficiency, allows the evaluation of the reference system as a whole as well as the evaluation of the usability of different components integrated with it, such as the MPEG-7 eXperimentation Model. These results illustrate the efficiency of the proposed system, as well as its suitability in serving as a test-bed for evaluating and comparing different algorithms and approaches pertaining to the content-based and semantic manipulation of visual information.

* This work was supported by the EU project SCHEMA "Network of Excellence in Content-Based Semantic Scene Analysis and Information Retrieval", www.schema-ist.org (IST-2001-32795).

1. Introduction

The efficient access and retrieval of visual information has emerged in recent years as an important research direction. This is due to the continuously accelerated generation and distribution of digital media and in particular still images. Many approaches to image retrieval have appeared in the literature, ranging from content-based ones to approaches exploiting other modalities such as text, while at the same time the importance of the media retrieval task has also motivated the introduction of the relevant MPEG-7 International Standard [4]. The latter is formally named "Multimedia Content Description Interface".

Despite initial attempts for image retrieval, based on exploiting text (e.g. image captions), the Query-by-Example paradigm was soon established as the prevalent methodology for addressing the problem of image retrieval from generic collections. The Query-by-Example paradigm has been explored in conjunction with both *coarse granularity* and *fine granularity* preprocessing of the image data, where the latter signifies the analysis of the image to meaningful regions while the former involves the processing of images as a whole [1]. With recent works documenting the superiority of fine granularity approaches over coarse granularity ones [3], the remaining two most important issues when building a content-based image retrieval system are the method to be used for segmenting the images to meaningful regions and the features that should be extracted for these regions and employed for matching. These issues are highly dependent upon the image data to be used, thus no universal solution exists.

This work presents a more modular and generic fine granularity approach to addressing the problem of image retrieval: the Schema Reference System (SchemaXM) [11]. This is a content-based image retrieval system that employs multiple segmentation algorithms and a plethora of standardized MPEG-7 indexing descriptors [6], thus being able to counter the imperfection of segmentation methods and the variability of the employed image data. Additionally, this architecture can facilitate the comparative evaluation of different integrated modules, particularly since the Schema reference system allows the easy integration of additional algorithms and tools with it.

The paper is organized as follows: in section 2 the SCHEMA reference system is presented, introducing the use of four segmentation algorithms and the MPEG-7 eXperimentation Model. Section 3 contains experimental evaluation of the reference system and comparison with an earlier version of it. Finally, conclusions are drawn in section 4.

2. Reference System

2.1. Visual medium analysis

As already mentioned, the SCHEMA Reference System has adopted a fine granularity approach to image indexing and retrieval, thus requiring the use of a segmentation algorithm for decomposing the images to meaningful regions. The use of a segmentation algorithm for region-based image retrieval has several advantages, mainly deriving from the fact that the user of an image retrieval system typically queries for objects similar to one such depicted in a key-image, rather than simply for similar images. Thus, using image segments (regions) that are closer to the notion of objects than the entire images themselves can significantly improve the accuracy of the retrieval process. The imperfection of any segmentation algorithm is, however, a limiting factor in such schemes. To counter this drawback and to provide at the same time a test-bed for the comparison of different segmentation algorithms in terms of their suitability for the application of content-based image retrieval, a number of different segmentation modules have been integrated with the SCHEMA reference system. The different segmentation masks produced by these for a given image are simultaneously presented to the user, to allow for the one most suited to the user needs at the given time to be employed for initiating the query.

There have been four segmentation algorithms integrated so far with the reference system. All were previously integrated in the Qimera framework [7, 8], which provides common input/output formats, thus facilitating their rapid subsequent integration with the reference system. These algorithms are the following:

- Pseudo Flat Zone Loop algorithm (PFZL), contributed by Munich University of Technology - Institute for Integrated Systems. The PFZL algorithm uses as homogeneity criterion the difference between two neighboring pixels; it proceeds by gradually labelling homogeneous, connected areas, and checking if these, at some iteration, become unconnected.

- Modified Recursive Shortest Spanning Tree algorithm (MRSST), contributed by Dublin City University. This is based on the original RSST algorithm, modified so as to initially avoid merging regions with very different colors and at a second stage to encourage the creation of large connected regions.

- K-Means-with-Connectivity-Constraint algorithm (KMCC), contributed by the Informatics and Telematics Institute / Centre for Research and Technology - Hellas. In the employed variant, color, position and texture features are used. The algorithm is combined with a procedure for accelerating its execution, based on partial pixel reclassification using a Bayes classifier.

- Expectation Maximization algorithm (EM) in a 6D colour/texture space, contributed by Queen Mary University of London. Under the EM technique, the distribution of color and texture features in each region is modelled as a mixture of Gaussians, and an iterative procedure is employed for estimating the parameters of these models.

A more detailed description and references to these algorithms can be found in [9]. Results of applying the aforementioned algorithms, as they are employed in the reference system, to a color image are shown in Fig. 1

2.2. Graphical User Interface functionality

The Graphical User Interface of the Reference System is built around the Query-by-Example paradigm. Thus, the user is first presented with several image categories (Fig. 2(a)); after selecting a category, a few images belonging to it are presented to the user for the purpose of initiating the query (Fig. 2(b)). The user selects an image and is subsequently presented with the segmentation masks generated for this image by the integrated segmentation algorithms (Fig. 2(c)). After selecting a region of the image, using any of the available segmentations, similarity search is performed in the image collection and its results are presented (Fig. 2(d)). Any of the result images (Fig. 2(d)) can then be employed for initiating a new query using the corresponding segmentations, as in (Fig. 2(c)).

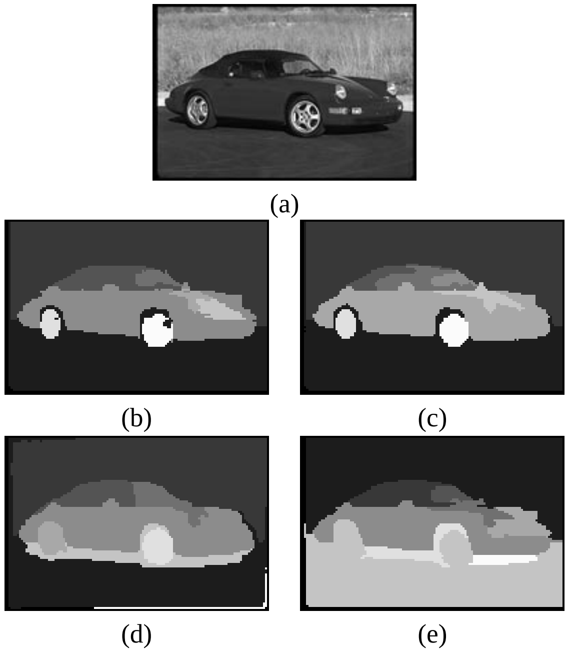

Figure 1. Segmentation example: (a) original image, (b)-(e) segmentation results by (b) MRSST, (c) KMCC, (d) EM, (e) PFZL algorithms.

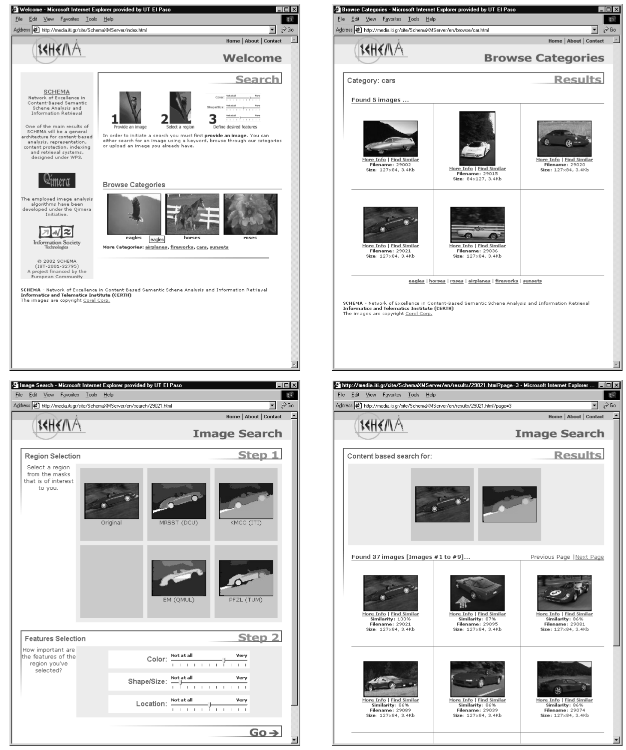

Figure 2. Graphical User Interface of the Schema Reference System: (a) category selection, (b) query image selection, (c) query region selection using any segmentation algorithm, (d) presentation of query results, which can be used for initiating a new query.

2.3. Indexing and retrieval using MPEG-7 XM

The set of MPEG-7 descriptors that where selected for describing regions in the Schema reference system are summarized in table 1. These descriptors are instantiated using an abstract descriptor, the so called MultiDescriptor for images. This abstract descriptor module simply encapsulates the memory management for the selected descriptors and allows calling the extraction and the matching of all the descriptors as if they where parts of a single descriptor. Figure 3 shows the architecture of the matching process using the abstract descriptor (marked in blue). When instantiating the MultiDescriptor objects, the corresponding objects of the selected descriptors are also instantiated. This behavior is indicated by the grey (curved) arrows. When creating the processing chain, the descriptor modules are also connected to their processing chains (marked with black arrows).

An essential problem when combining matching values of different visual descriptors is the fact that the distance functions are not normalized in any way. To alleviate this problem, the MultiDescriptor module performs a simple normalization of the working point of the matching functions. The working point is the value of the distance between two descriptions that, if not exceeded, signifies that the two descriptions should be treated as similar. If, on the opposite, the distance value exceeds this threshold, the two descriptions are assumed to be different. The working points for the individual descriptors were determined experimentally. Subsequent to normalization using the working points, i.e making the latter equal to 1.0 for all descriptors, the different distance functions can be scaled in a linear way. Thus, in order to generate a single overall matching value for each region, a weighted linear combination of the individual distance values is simply calculated.

2.4. Implementation issues

The original MPEG-7 eXperimentation Model is a simple command line program; for doing a similarity search using the selected visual descriptors, the program reads the descriptions for the entire image collection from the .mp7 files containing the descriptor bit stream. Additionally, it extracts the description of the image used for submitting the query

Color descriptors	Color Layout
	Color Structure
	Dominant Color
	Scalable Color
Texture descriptors	Edge Histogram
	Homogeneous Texture
Shape descriptors	Contour Shape
	Region Shape

Table 1. MPEG-7 Descriptors used

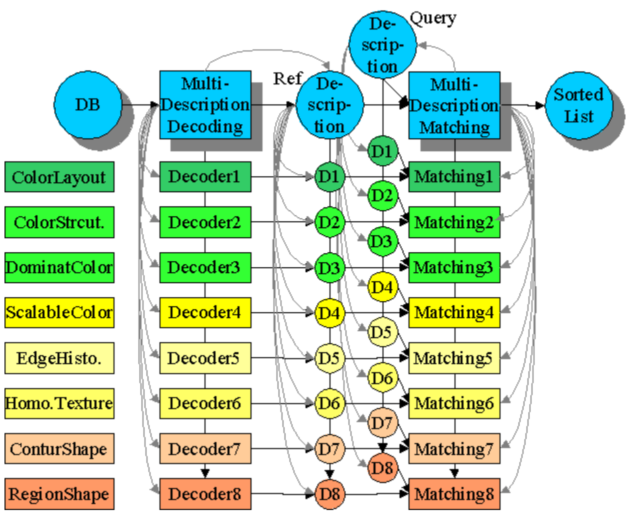

Figure 3. Architecture of the MultiDescriptor module, i.e. of the matching chain using multiple descriptors.

(key-image) during query execution. As a result, for every search process the descriptions database is accessed and decoded and the key-image is processed, leading to unnecessary overheads in the query execution process. To accelerate the search procedure, a different approach is adopted in SchemaXM.

First, if the query image is already part of the database, the query description is not extracted again from the image data. This is achieved by restructuring the MPEG-7 XM search and retrieval application using the Visual XM Library and taking advantage of the modularity of the XM software.

Secondly, the MPEG-7 XM is used as a server that keeps running in the background, accepting new queries and delivering the corresponding retrieval results. This removes the need for continuously reading and decoding the MPEG-7 bitstreams, since these tasks can then be performed only once during server initiation and be omitted at query time. The use of the MPEG-7 XM as a server required the introduction of extensions to the XM software, an approach also investigated in [2].

Using the two aforementioned methods for integration with the MPEG-7 XM resulted in significant improvement of the time-efficiency of the SchemaXM system as compared to using the original MPEG-7 XM software for the purpose of integration, as discussed in the following section.

3. Experimental Results

The Schema reference system was tested, for the purpose of performance evaluation, on a collection of 2000 still images of the Corel gallery [5]. These images were pre-assigned to 20 categories (e.g. eagles, flowers, cars, etc.), each containing 100 images, while subcategories (e.g. red cars) were manually defined whenever necessary. This category / subcategory membership information was employed only for evaluating the performance of the reference system by calculating precision-recall diagrams for specific query categories. Note that the term *precision* is defined as the fraction of retrieved images which are relevant, and the term *recall* as the fraction of relevant images which are retrieved [3]. The aforementioned precision-recall curves were produced by averaging the results of five different queries for objects of the same category, to allow for objective results to be extracted.

In order to comparatively evaluate the retrieval efficiency of the Reference System, precision-recall diagrams are also presented for an earlier version of the system employing proprietary descriptors and the Euclidean distance for indexing and retrieval [10]. These descriptors include:

- Linearized color histograms in the RGB color space, quantized to 8 bins per color component and normalized so that the sum of the bin values for each color component equals to 1, thus resulting in 24 color features.
- Coordinates of the center of gravity of the region in the image grid, normalized by division with the image dimensions in pixels (2 position features).
- Size of the region, expressed as the ratio of the number of pixels assigned to it over the total number of pixels of the image.
- Normalized eccentricity, calculated using the covariance matrix of the region and Principal Component Analysis.

The corresponding results for both versions are presented in Fig. 4, where it can be seen that the reference system employing the MPEG-7 XM for indexing feature extraction and for matching, performs consistently better than

System	Time(sec)
Schema reference system v.1.0	< 1
MPEG-7 eXperimentation Model (original version)	15
SchemaXM (with MPEG-7 XM running as a server)	6

Table 2. Average query execution times

the earlier version of it employing proprietary descriptors. However, this improvement is achieved, to some extent, at the expense of the time required for query execution. As can be seen in table 2, the SchemaXM requires on the average 6 seconds to process a query. This is a significant improvement as compared to using the original version of MPEG-7 XM, achieved due to the use of the underlying MPEG-7 XM retrieval module as a server; however, SchemaXM still requires considerably more query execution time than the earlier version of the reference system. It should be noted that such time-efficiency results are indeed expected, since the MPEG-7 XM MultiDescriptor search application employs a plethora of complex descriptors and correspondingly complex matching functions, as opposed to using simple proprietary descriptors and the Euclidean distance.

4. Conclusions

Recent advances in the development of the SCHEMA reference system were reported in this paper. In the presented system, the availability of a variety of analysis tools and the combination of various indexing descriptors are employed for alleviating the imperfection of any segmentation algorithm and for countering the variability of image data, respectively, thus greatly improving the usability of the system in real-world applications. Furthermore, the use of a variety of analysis tools enables their comparative evaluation in terms of their suitability for use in a content-based image retrieval system. This, along with the possibility of integrating additional such tools with the SCHEMA reference system, illustrates an additional potential use of it: as a testbed for evaluating and comparing different algorithms and approaches.

Future research will concentrate on further improving the time-efficiency of SchemaXM by implementing a sophisticated indexing mechanism for the descriptor database, on enabling more extensive interactivity by means of relevance feedback, and on introducing the use of other modalities in combination with content-based features, in order to improve the retrieval accuracy.

References

[1] W. Al-Khatib, Y. Day, A. Ghafoor, and P. Berra. Semantic modeling and knowledge representation in multimedia databases. *IEEE Trans. on Knowledge and Data Engineering*, 11(1):64–80, Jan.-Feb. 1999.

[2] BUSMAN IST Project. http://busman.elec.qmul.ac.uk/.

[3] 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*, 24(8), 2002.

[4] S.-F. Chang, T. Sikora, and A. Puri. Overview of the MPEG-7 standard. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6):688–695, June 2001.

[5] Corel stock photo library. Corel Corp., Ontario, Canada.

[6] MPEG-7 XM software. http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html.

[7] N. O'Connor, T. Adamek, S. Sav, N. Murphy, and S. Marlow. QIMERA: A Software Platform for Video Object Segmentation and Tracking. In *Proc. Workshop on Image Analysis For Multimedia Interactive Services*, London, UK, Apr. 2003.

[8] N. O'Connor, S. Sav, T. Adamek, V. Mezaris, I. Kompatsiaris, T. Lui, E. Izquierdo, C. Bennstrom, and J. Casas. Region and Object Segmentation Algorithms in the Qimera Segmentation Platform. In *Proc. Third Int. Workshop on Content-Based Multimedia Indexing (CBMI03)*, 2003.

[9] Schema Deliverable D3.1. http://www.schema-ist.org/SCHEMA/files/document/30-03-2004/D3.1.pdf.

[10] Schema reference system v.1.0. http://media.iti.gr/site/Schema.

[11] Schema reference system v.2.0. http://media.iti.gr/site/SchemaXM.

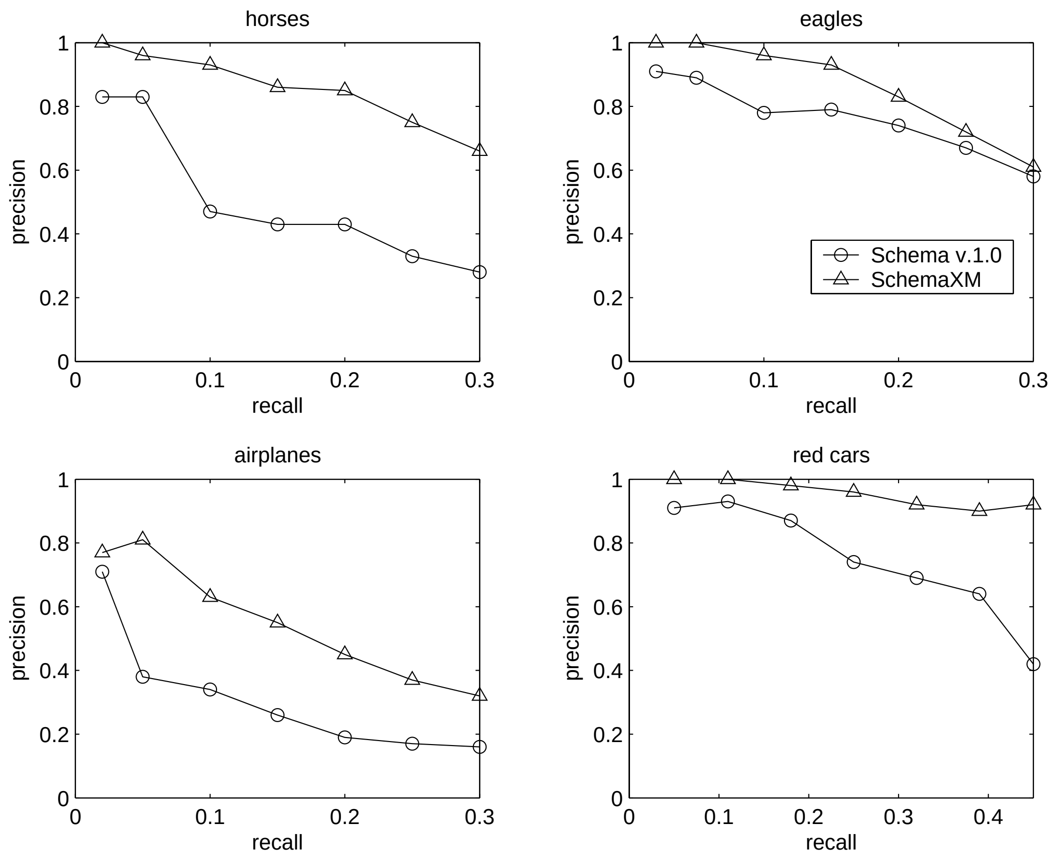

Figure 4. Precision-Recall diagrams for comparing between the two versions of the Schema reference system.