# New Methods for Image Retrieval

Zoran Pečenović, Minh Do, Serge Ayer, Martin Vetterli
Laboratory for Audio-Visual Communications,
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne, Switzerland
{Zoran.Pecenovic, Minh.Do, Serge.Ayer, Martin.Vetterli}@epfl.ch

## Abstract

Image Retrieval (IR) is one of the most exciting and fastest-growing research areas in the field of multimedia technology. We present here a highlight of recent research for IR. Some trends and probable future research directions are presented. We expose the major problems that we have recognized: the lack of a good measurement of visual similarity, the little importance accorded to user interaction and feedback, and the neglect of spatial information. Answering these concerns, we describe the solutions implemented by recent IR systems. We also present the current image retrieval projects in our laboratory, which are motivated to a large extent by these same considerations.

## 1 Introduction

Large and distributed collections of scientific, artistic, and commercial data comprising images, text, audio and video abound in our information-based society. To increase human productivity, however, there must be an effective and precise method for users to search, browse, and interact with these collections and do so in a timely manner.

The fundamental operation of yesterday's databases was *matching*: determining whether a data element is the same, in some predefined sense, as the query. Today, with complex multimedia data, matching is not expressive enough, and database systems will move to systems in which the fundamental operation is *similarity assessment*. This reflects the preference in image retrieval of general users, who want to retrieve a number of similar images and then use them to iteratively refine their queries. Therefore IR systems should be designed to be an effective and efficient tool for browsing and navigating in image databases.

We first present a brief overview of existing systems and of research work in the field. The presented systems are those which seem to promote the most relevant issues. Then we develop the general motivations and directions of research. In section 4 we briefly expose the work underway in our laboratory.

## 2 State-of-the-art IR systems

Image retrieval is a very fast growing research area in the last few years. Famous early examples include the QBIC system from IBM [1] which allows users to retrieve images based on color, texture, layout and by a sketch; the Photobook system by MIT Media Lab [2] which is very powerful for retrieving images from homogeneous collections; the Virage system by Virage company [3] which can be tailored to many applications; the Chabot system from UC Berkeley [4]. These systems provide interactive human-machine interfaces for image searching and browsing,

The most recent version of Photobook includes FourEyes [5]. This system has a distinguishing feature of benefiting from user interaction to help segmentation, retrieval and annotation of an image database. Data is dynamically organized into groups according to relevance feedback from users. In order to classify images, instead of using just one model, FourEyes employs a "society of models".

Other IR systems incorporate automatic image segmentation to allow more accurate retrieval. VisualSEEk [6] proposed a feature back-projection scheme to extract salient image regions and therefore the system is able to provide joint content-based and spatial search capacity. Carson *et al.* [7] employed a so called "blobworld" representation which is based on segmentation using EM algorithms on combined color and texture features. In another system, NETRA [8], images are segmented into homogeneous regions using a technique called "edge flow" at the time of ingest into the database. Image features that represent each of these regions are computed for indexing and searching.

Some recent IR systems exploited wavelet inspired approaches. Jacobs *et al.* [9] proposed a fast image querying system which uses spatial information and visual features represented by dominant wavelet coefficients. Another system, WaveGuide [10], uses a joint feature set of texture, color and shape which are all based on significant wavelet coefficients. Content descriptors are extracted from a wavelet coding scheme through the successive approximation quantization (SAQ) stage.

The above are only a few of the best known approaches, much work is being carried out on specific areas used by these systems, in particular by the computer-vision and pattern recognition specialists, for developing better segmentation, classification and interpretation algorithms of the image content. An example of a more complete bibliography on the state-of-the-art, can be found in [11], as well as on various dedicated sites on the World Wide Web (WWW).

# 3 Image retrieval: directions & trends

In this section we try to subjectively identify some of the current trends in the research for image retrieval systems.
A common ground in most of current IR systems is to exploit low-level features such as color, texture and shape, which can be extracted by a machine automatically. While semantic-level retrieval would be more desirable for users [12], given the current state of technology in image understanding, this is still very difficult to achieve. This is especially true when one has to deal with a heterogeneous and unpredictable image collection such as from the WWW.
As mentioned before, current research fights to bridge the gap between low-level, statistical, descriptions and high-level semantic content. Thus methods inspired by artificial intelligence [13], textual retrieval [14, 15], and psychology & human-computer interaction [16, 17], are starting to influence the research. Synthetically, image retrieval starts off by the design of a robust, meaningful and flexible feature set to characterize all plausible images in the collection. Then clever manipulation of the features tries to uncover some higher-level similarity between the query and the database candidates. An interactive, iterative, and user-oriented query process then improves on the raw results. This is schematically shown on Figure 1. Each of the elements presented is studied by groups of specialists, but rarely, the whole system is examined.

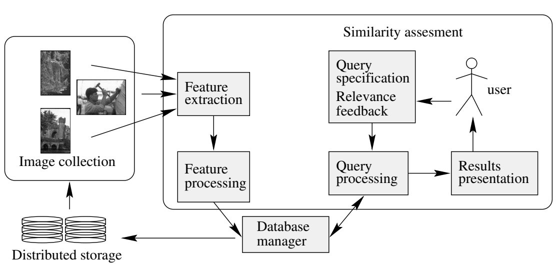


Figure 1: Typical image retrieval process.

Early IR systems [1, 2, 3, 4] mainly relied on a *global* feature set extracted from images (see Table 1 for typical features). For instance, color features are commonly represented by a global histogram. This provides a very simple and efficient representation of images for the retrieval purpose. However, the main drawback with this type of systems is that they have neglected spatial information. More recent systems have addressed this problem. Spatial information is either expressed explicitly by the segmented image regions [6, 7, 8] or implicitly via dominant wavelet coefficients [9, 10].

Most systems use the *query by example* approach, where the user selects one or several images, and the system returns the ones judged similar. An alternative way of querying the image database based on content, is by allowing the user to *sketch* the desired image's color/texture layout, thus abstracting himself, the objects searched for [18, 19, 1]. Other more targeted systems allow the user to specify spatial constraints on the dominant objects. All of these methods suffer somewhat from the drawback that the system relies on the users abilities and does not adapt to his/her needs.

Table 1: Overview of commonly used features in IR.

| | |
|---|---|
| Color | histograms, color co-occurrence histograms |
| Shape | segmentation & contour extraction followed by : contour matching, moments, template matching |
| Texture | directionality, periodicity, randomness, Fourier-domain characteristics, random fields |
| Others | wavelet coefficients, eigenimages, edge-maps of user made sketch, image context vectors |

Another active research direction is to speed-up the retrieval process. As discussed above, since image searching is only based on primitive-features, the results might not meet the user's expectation at the first result. Therefore IR systems must support *interactive querying*, i.e. letting users view the results quickly, refine their queries and try again. This requires the retrieval process to be fast, even with a large database (typically over 10000 images). In this approach the user must be able to specify in a natural way what he/she wants, and do this using the entire range of available features, either in conjunction or separately. Other issues become of capital importance, such as transmission times (even more critical in video-applications), security and the distributed nature of today's databases.
First attempts in speeding-up the retrieval process were based on pre-filtering, using high dimensional database structures such as R-trees, or on the Principal Component Analysis ([20, 21]). However, those techniques only alleviate the high dimensionality of the feature data set but neglect one important fact: the semantic structure of natural images. More recent approaches gear toward this direction. Among those are tree-structure vector quantization [22] and multi-scale search [23, 10]. These techniques organize images or the search process into a hierarchical structure. Images are considered in multiple resolutions –or semantic levels– where the search can be applied in the coarse resolution first (see Figure 2). High resolution search is selectively guided by results of the low level search. Note that this is exactly what happens in human eyes when searching for information [24]. Wavelets provide a powerful and efficient mathematical tool to process images at multiple scales.
There are still several aspects, in our opinion, which have received little attention from the image retrieval research community. First of all, is the consistence of similarity measurement with human perception. The recent work done by Santini and Jain [25] establishes some connection with psycho-

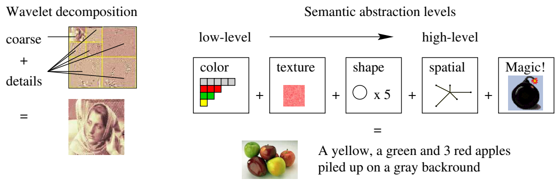


Figure 2: Multi-level representations

logical studies. More empirical evidences, and a method of assessing the agreement between humans and machines, can be found in [16]. However, in most of current IR systems, the similarity measurements are still either *arbitrarily defined* or obtained from a training procedure using a *pre-selected* query set. The similarity measurement is even more complex if one wants to combine several models of similarity operating at the same time to obtain a global measurement. The weight given to different models depends on the particular task at hand [25]. This is also applied when the system is used by different users. Therefore the kind of systems which can interactively learn from relevance feedback like FourEyes [5] should be emphasized. This might require the extracted features to be flexible enough so that the similarity metric could be *dynamically* modified during the query time.

The lack of a coherent and complete way of expressing the user's information needs and query process, still has to be realized by the IR research community. Using feedback from image producers (museums, archives, photographers, and scientific imagery) and image consumers (private users, intranets, news agencies, scientists and scholars), would give the community the grasp on what their true goals are.

Another aspect, which has also been raised elsewhere [9], is the lack of a common database or benchmark for testing and assessing the performance of IR systems. Apart from the relevance of retrieved images, other factors should be considered as well, such as retrieval time, storage overheads, flexibility, etc. Results reported from recent work are various and somewhat subjective.

# 4 Research at LCAV

The current research at the Laboratory for Audio-Visual Communications (LCAV) is mainly targeted at a) the development of novel methods for similarity metrics and for integration of incommensurate descriptors, and b) at the design of robust, efficient and effective feature sets.

## 4.1 IR using Latent Semantic Indexing

The initial work at the LCAV for IR [15] was based on a method adapted from the text retrieval literature. This method, called Latent Semantic Indexing (LSI), uses a term by document occurrence matrix as a representation of the information content of the collection. This matrix is then approximated by a lower rank truncated Singular Value Decomposition (SVD). This approach alleviates noise in term usage and implicitly solves the problems of polysemy and synonymy. This method has given very promising results in text retrieval, information filtering, thesaurus construction and other term-document oriented information processing tasks.

In our case, "terms" or "words" are content descriptors (features) which are extracted from images. Currently these features are based on color, texture, layout, and meta-data descriptors, later on wavelet-based features will be added (see below). One of the interesting capabilities of LSI is that the nature of the features used is irrelevant, thus allowing the integration of visual, and non-visual descriptors of the image into a unique and user-invisible index. Therefore if textual description of the images is available (typical case in existing databases, CD-collections or WWW), it can be easily integrated with other content-based and visual features.

The first version of the system has been implemented. Its user interface is written in Java so that image queries can be made from anywhere via WWW. Query construction allows the use of relevance feedback. Rather than a single example image, users can specify a set of relevant images and non-relevant images (e.g. among the result images from the last query). For more details and results of the system, see [15].

The work in progress is concentrated on the study of faster approximation techniques, like the wavelet-packet optimal approximation. This method chooses from a vast collection of basis the one that best approximates the input matrix. The criterion of *optimality* can be any additive function on the basis set. Another direction of research is targeted at the study of the interaction between features. Unlike the more traditional approaches, we do not want to have a restricted and specific set of descriptors, rather we prefer to consider the broadest possible "library" of features, from which to choose the *optimal* set. With this goal in mind we are establishing an information-theory based method for detecting redundancy and importance in the feature sets. In conjunction with the above techniques, we are studying a metric for the distinguishing power of a given set of features. This arsenal of novel techniques allows us to decide – perhaps even at query time – which descriptors the system should consider and in parallel to exploit the user's relevance judgment to guide the convergence of the similarity metric.

## 4.2 Invariant feature extraction using wavelet maxima

Automatic feature extraction is an important part of an IR system. As mentioned in section 3, most current feature extraction techniques suffer from the problem that they do not retain any spatial information. Some more recent systems exploited wavelet basis coefficient to cope with this problem [9, 10]. In addition, wavelet decomposition provides a very good approximation of images and the underlying multi-resolution ability allows the retrieval process to be done progressively.

The main drawback with wavelet bases is their lack of translation invariance. This is because the wavelet coefficients are obtained by sampling uniformly the continuous wavelet transform via a dyadic scheme [24]. An obvious cue to this problem is to apply a non-subsampled wavelet transform, ie. skip the down-sampling step. However this creates a highly redundant representation and we have to deal with a large amount of feature data.

To reduce the representation size, to facilitate the retrieval process while maintaining translation invariance, an alternative approach is to use an adaptive sampling scheme. This can be achieved via the wavelet maxima transformation [26, 27], where the sampling grid is automatically translated when the signal is translated. Wavelet maxima have been shown to work well in detecting edges which are likely key features in a query. Moreover this method provides flexibility in choosing filters and the size of extracted data. By varying the applied filers, one could control the amount of data to be recorded.

We are currently experimenting with this approach and results will be reported soon.

# 5 Conclusion

This paper presented a brief review of recent methods for image retrieval. The mentioned systems were categorized and we highlighted their ability to express and exploit spatial information either via automatic image segmentation or wavelet decomposition. Further emphasis was made upon the novel techniques of speeding-up retrieval processes using hierarchical searches, and wavelet approaches. We also tried to stress the major advantages and shortcomings of the existing research, both in particular cases and globally. We expressed the concern for tighter collaboration between the three parties involved in image retrieval applications: image producers, image consumers and system designers. We have insisted on the open questions in the domain, like good measurements of visual similarity, robust features, the importance of the user in the query process, and the gap between image understanding and image retrieval.

We have also briefly presented our current research, under the aspects which seemed to us as the most important in the problematic context exposed above.

# References


[1] M. Flickner et al. Query by image and video content: The QBIC system. *Computer*, pages 23–32, September 1995.

[2] R.W. Piccard A. Pentland and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.

[3] J.R. Bach et al. The Virage image search engine: An open framework for image management. In *Storage and Retrieval for Image and Video Databases III*, volume 2420 of *SPIE*, pages 76–87, 1995.

[4] V. E. Ogle and M. Stonebraker. Chabot: Retrieval from a relational database of images. *Computer*, pages 40–48, September 1995.

[5] T.P. Minka and R.W. Piccard. A society of models for video and image libraries. Technical Report 349, M.I.T. Media Laboratory Perceptual Computing Section, 1996.

[6] J.R. Smith and S.-F. Chang. VisualSEEk: a fully automated content-based image query system. In *Proc. The Fourth ACM International Multimedia Conference*, pages 87–98, November 1996.

[7] H. Greenspan C. Carson, S. Belongie and J. Malik. Region-based image querying. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, Puerto Rico, June 1997.

[8] W. Y. Ma and B. S. Manjunath. NETRA: A toolbox for navigating large image databases. In *IEEE International Conference on Image Processing*, 1997.

[9] A. Finkelstein C.E. Jacobs and D.H. Salesin. Fast multiresolution image querying. In *Computer graphics proceeding of SIGGRAPH*, pages 278–280, Los Angeles, 1995.

[10] K.-C. Liang and C.-C. Jay Kuo. WaveGuide: A joint wavelet image description and representation system. 1998. to appear.

[11] A. Rosenfeld. Image analysis and computer vision. *Computer Vision and Image Understanding*, 66:33–93, April 1997.

[12] J.P. Eakins. Automatic image content retrieval - are we getting anywhere? In *Proc. of Third International Conference on Electronic Library and Visual Information Research*, pages 123–135, May 1996.

[13] T. Minka. An image database browser that learns from user interaction. Master's thesis, MIT Media Laboratory, 1995.

[14] S. I. Gallant and M. F. Johnston. Image retrieval using image context vectors: first results. In *Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 82–94, 1995.

[15] Z. Pecenovic. Intelligent image retrieval using Latent Semantic Indexing. Master's thesis, Swiss Federal Institute of Technology, Lausanne, Vaud, April 1997.

[16] D. McG. Squire and T. Pun. Assessing agreement between human and machine clusterings of image databases. *Pattern Recognition*, accepted, to be published 1998.

[17] M. Richeldi and P. L. Lanzi. ADHOC: A tool for performing effective feature selection. In *Proceedings of the International Conference on Tools with Artificial Intelligence*, pages 102–105, 1996.

[18] K. Hirata and T. Kato. Query by visual example. In *EDBT'92*, pages 56–71, 1992.

[19] M. Egenhofer. Spatial-query-by-sketch. In *IEEE Symposium on Visual Languages*, pages 60–67, 1996.

[20] C. Faloutsos et al. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3:231–262, 1994.

[21] R. Ng and A. Sedighian. Evaluating multi-dimentional indexing structures for images transformed by Princpal Component Analysis. In *Storage and Retrieval for Image and Video Databases IV*, volume 2670 of *SPIE*, pages 50–61, 1996.

[22] C.A. Bouman J.-Y. Chen and J.P. Allebach. Fast image database search using tree-structure VQ. In *Proc. of ICIP97*, pages 827–830, Santa Barbara, October 1997.

[23] C.A. Bouman J.-Y. Chen and J.P. Allebach. Multiscale branch and bound image database search. In *Proc. SPIE/IS&T Conf. on Storage and Retrieval for Image and Video Databases*, San Jose, February 1997.

[24] S. Mallat. Wavelets for a vision. *Proceeding of the IEEE*, 33:604–614, 1996.

[25] S. Santini and R. Jain. Similarity matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1996. submitted.

[26] S. Mallat and W.L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. Info. Theory*, 38:617–643, March 1992.

[27] Z. Cvetkovic and M. Vetterli. Discrete-time wavelet extrema representation: design and consistent reconstruction. *IEEE Trans.Signal Processing*, 43:681–693, March 1995.