An interactive image retrieval system: from symbolic to semantic

Le Thi Lan* and Alain Boucher*,**

* International Research Center MICA
C10, Hanoi University of Technology, 1 Dai Co Viet, Hanoi, Vietnam
Tel: +84 4 8683087 Fax: +84 4 8683551 E-mail: Thi-Lan.Le@mica.edu.vn,
**Institut de la Francophonie pour l'Informatique
nha D, ngo 42, Ta Quang Buu, Hanoi, Vietnam
Tel: +84 4 869 6017 Fax +84 4 869 4509 E-mail: alain.boucher@auf.org

Abstract:
In this paper, we present a overview of content-based image retrieval (CBIR) systems: its results and its problems. We propose our CBIR system currently based on color and texture. From the CBIR systems, we discuss the way to add semantic values in image retrieval systems. There are 3 ways for adding them: concept definition, machine learning and man-machine interaction. Along with this we introduce our preliminary results and discuss them in the goal of reaching semantic retrieval. Different result representation schemes are presented. At last, we present our work to build a complete annotated image database and our image annotaion program.

Image Processing, Image Retrieval, Symbolic Values, Semantic Values, Image Annotation

1. INTRODUCTION

Image retrieval (IR) is one of the most exciting and fastest growing research area in the field of multimedia technology. There are 2 main approaches: the first one uses manual annotations and the second one uses automatic features extracted from the images. The first approach is based on manual textual image annotation. It works well, but it is a long and repetitive task for human as the image database becomes larger and larger. Furthermore, it is subjective to the culture, the knowledge and the feelings of each person. The second aprroach uses features extracted from the image such as color, texture, shape... It is independent of people. But it is difficult to design powerful features to represent images. We have constructed a CBIR system based on symbolic features (mostly color and texture) and we have realized that symbolic features alone cannot always satisfy the users. Therefore, we are now studying the problems of semantic image retrieval and adding semantic values in CBIR systems.

This paper is organized as follows: In the next section we review the CBIR systems and we present our system based mostly on color and texture. In section 3 we discuss the ways to head forward to semantic retrieval using the symbolic features and present our preliminary results. Section 4 discusses an inherent problem in images retrieval systems: evaluation. In section 5, we present our images database that we are building, and our image annotation program. In the last section, we give some conclusions.

2. CBIR: ONE DESCRIPTOR OR COMBINING DESCRIPTORS

A content-based image retrieval system is a system that uses features from the images such as color, texture and shape for computing the similarity between a pair of images. But an image can contain a lot of features. So how to choose the features and how to combine these features? These are always difficults but interesting problems in content-based image retrieval systems. They are getting a lot of attentions of the researchers. In this section, we review some methods that use these features.

2.1 Color

Color is one of the first features that is used for image retrieval. It has been proven as an efficient feature. Most of recently image retrieval systems analyze the images using color. The first method used color feature is histogram intersection that has been proposed by Swain and Ballard[23]. Color histograms are easy and fast to compute, robust to rotation and translation and have few constraints when applied on images. However color histograms have four inherent problems in indexing and retrieving images [3]. First, they are big in size, hence it is difficult to create an effective database indexing scheme. Second, they do not possess spatial relationships between color positions. Third, they are sensitive to small brightness changes. Finally, they are incompetent to support spatial matching of image contents. Therefore, recently reseach studies have been made to improve the problems associated with color historgrams. There are two approaches: the first one adds spatial information, the second one find other color spaces that is similar to human perception of color.

The first one adds some spatial information in color histograms. Stricker and Dimai[21] divided an image into five fixed overlapping blocks and extracted the first three moments of each block to create a feature vector. Pass and Zabih [17] added the spatial cohenrence into color histograms. A pixel is coherent if it belongs to sizable contiguous regions and incoherent otherwise. The pixels in a bin of color histogram are divided into two classes: coherent class and incoherent class. The comparison between two color histograms becomes the comparison between pixels in the corresponding classes. Huang et al. [4] proposed the correlogram. Since the correlogram requires a large storage and long time to compute, the authors used the auto-correlogram instead. Bin i-th gives the probability that two pixels with distance k have color value i.

The second approach tries to find other color spaces that are based on human perception of color. RGB is a color space used firstly in indexing and retrieving image systems. Smeuders et al. [20] presented some formulas for RGB color space. While Park et al. [15] proposed CIE LUV color space and weighted LUV quantization scheme. Their experimental results indicated that the weighted LUV quantization gives better performance than others. Gong et al [3] used HVC instead of RGB. According to the authors, this color space is more close to the human perception and gives results more accurate according to human perception.

2.2 Texture

Texture is a primitive visual cue that has been studied for over twenty years now. Various techniques have been developed for texture analysis.

In 1970, Haralick proposed a method based on gray level co-occurrence matrices. It is probably one of the most famous methods of texture analysis. The co-occurrence matrix $P^{Har}_{d,\alpha}(g,g')$ counts the numbers of pixel pairs (m,n) and (m',n') in an image that have intensity value of g and g' with the distance d in a relative direction α. In order to estimate the similarity between gray level co-occurrence matrices, Haralick proposed 14 statistical features extracted from them. In pratice, the four most relevant features that are widely used in literature are computed: energy, entropy, contrast and inverse difference moment. Partio et al. [16] adopted the co-occurrence matrices on a rock database. The authors compared performance of the co-occurrence matrices with Gabor and wavelet features and realized that the co-occurrence matrices performed better for the given rock image dataset.

The methods based on the Gabor filters are also developed. After applied the Gabor transform on an image, a texture region is characterized by a the mean and the standard deviation of the energy distribution of the transform coefficients. A feature vector is constructed using them as components. With the Gabor filters, we can choose the number of scales and the number of orientations. Ma and Manjunath [10][12] chose four scales and six orientations. The authors indicated that other representations such as using higher order or complex moments are also possible. But according to their experiments, the marginal improvement obtained by using higher order moments does not really justify the additional complexity.

2.3 Global features vs local features

A system based on global features does not always perform well. Suppose that in one image different parts have different color and texture characteristics, the feature vector extracted from the entire image would lost most of the local information from the original image. Therefore, there are some systems that compute the features globally, over the entire image. Besides, there are some other systems that compute their features locally. This choice is valid for a lot of different features being used, going from the histogram, the correlogram or the texture matrices. Two different approaches may be used to compute local features. The first approach divides the image using a grid and compute features separately for each square over the grid. In figure 1, the image is divided into 12 parts with the same size and the features are computed for each of parts, allowing to separate the sky from the beach in the comparison process.

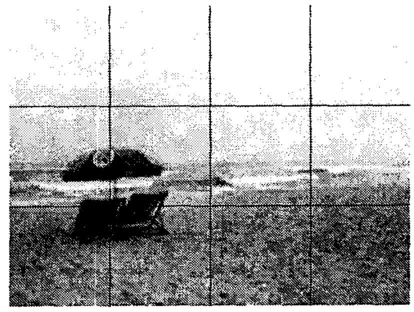


Figure 1. This image is divided into 12 parts with the same size and the features are computed locally for each part. This allows comparing separately the different parts, like the sky and the beach, for a more accurate result.

The second approach uses segmentation to divide the images into local zones and computes the features over all the extracted regions. Segmentation is a step that one may wish to avoid, because of all the related problems of choosing a "good" method of segmentation and having to change it or adapt it for each new image coming into the system. Nowadays, it is easily said that the general method of segmentation good for everything is a pitfall that we may want to avoid. However, dividing an image in regions of interest is almost necessary if we aim toward semantic information extraction from the image. This is why the concept of "weak segmentation" has been introduced and should be accepted. In this concept, one aims to segment the image in some general zones of interest, without looking for the exact computation of all exact regions (which is a pitfall in automatic segmentation because often you do not know if you are looking for the tree or for the individual leaves of the tree). In our work, we propose a general segmentation technique like the split-and-merge approach, and we constraint the technique to divide the image in 3 to maximum 10 regions. In figure 2, we segmented an image, the chose threshold is 10 and the chose minimum size of region is 9. While segmenting image, if the size of a region is less than the minimum size of region, this region is merged into another. The threshold and the minimum size of region is chose in order to get from 3 to 10 regions after segmentation. Doing this, we hope to capture the general meaning of the image, and we accept to loose the small details. Recalling that we already loose the important 3D information when processing 2D images, we can accept that those small details may be impossible to compare in the retrieving process.

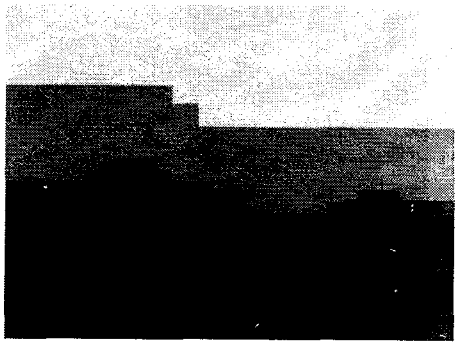


Figure 2. Example of segmented image using a split-and-merge algorithm, where the image is splitted into some general regions.

2.4 Other descriptors and combination of descriptors

The methods based on other features such as shape, salient points, invariants are also studied. The shape is an important visual feature and it is one of the primitive features for image content description. However, shape content description is a difficult task. Because it is difficult to define perceptual shape features and measure the similarity between shapes. Moreover, shape is often corrupted with noise, defection, arbitrary distortion and occlusion. So in recent CBIR systems, there are only few systems that use shape for image retrieval. The shape representation techniques are classified into contour-based methods and region-based methods based on whether shape features are extracted from contour only or are extracted form the whole shape region.

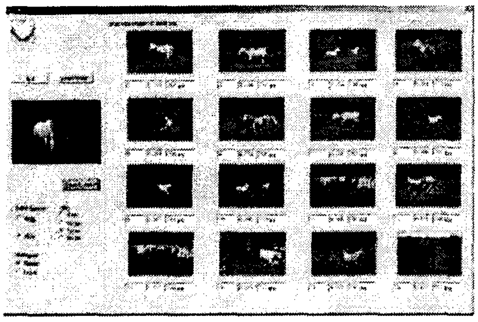
In practice, other features such as invariants, salient points are researched to improve the performance of CBIR systems [1][22][24]. But each method has both advantages and inconvenients. It can be developed well for certain images databases, but will not work properly with others.

Therefore, the combination of features is one of the most considerated approaches. Iqbal and Aggarwal [6][7] added image structure to color and texture in CIRES. Sciascio [19] introduced a system used color, object orientation and relative position as content features. This approach seems to be a good solution. But it is difficult to choose the adequate weight for each of features.

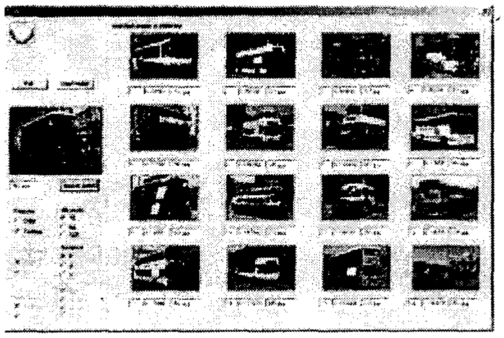
2.5 MICA CBIR system

At MICA, we are developing a content-based image retrieval system. The currently chosen features are the color and the texture. With the color, we used 2 color spaces: RGB and HSV and we have chosen the color histograms of Swain and Ballard[23]. This system supports both global histogram and local histogram. With the texture, we are computing co-occurrence matrices and created a features vector by extracting the four most relevant features from them. The user selects a request, chose a method with some parameters. The system computes features vectors and compares them to find results. The first sixteen results are presented with the names of results, the distances with the query and the ranks. The images used in this system come from Wang image database comprising 1000 images divided in 10 classes and available at http://wang.ist.su.edu/docs/related/. In section 5, we will present our work to build our own image database. In figure 3, we present some results. In figure 3a, we chose an image containing two horses and selected the method of color histogram with HSV color space, global features and 32 bins. The first fifteen images are relevant because they contain horses, the last result is not relevant. In figure 3b, we chose an image containing a bus and selected the method of texture. In order to compute the co-occurrence matrices, the chosen gray level is 16 and the chosen distance between two pixels is 1. The results are good, because the first sixteen results contain means of transport however the last result is not a bus.

The presentation of results is classical (list of images) and these results are similar to other works. We are developing this to demonstrate a CBIR system. In section 3, we will present a graphical representation that gives a more complete representation of the results for interaction and discuss semantic values that we hope to add in our system.



(3.a)



(3.b)

Figure 3. Examples of results from the MICA CBIR system. Images in this example are coming from Wang s image database.

3. USING SEMANTIC VALUES IN IR SYSTEMS

We state that semantic is not intrinsec to the images. Humans have semantic to describe objects, but it has been acquired through a long learning process during childhood, or for an unknown domain we can use references to understand it, or for an unknown object we actively explore it (by moving around it, touching it, testing it, ...). Moreover, we may use a lot of hypotheses to help the vision process (for example, general intuitions are that the sky is above and the ground is below and all the objects are subject to physic laws like gravity).

We need external sources to add semantic in the image retrieval process. In this section, we compare three techniques to add knowledge into a vision process: definition of a concept language, machine learning and interaction between the system and the user.

3.1 Concept definition language

The first technique to add high-level knowledge to a computer vision system is to add a knowledge base to it. The knowledge can take various forms and can be include in what we call here a concept definition language. Such definitions of languages have been used for image retrieval mostly to model textual queries as in [14].

Nowadays, an interesting advance to define knowledge in a computer vision is to use the domain of ontologies. As presented in [13]: "an ontology may be defined as the specification of a representational vocabulary for a shared domain of discourse which may incltde definitions of classes, relations, functions and other objects".

But still a problem often remains on how to link the high-level concepts defined in an ontology derived from humains vision and the low-level features that we can extract from images. Some work has been made in that direction. The work of Maillot at al. [11] in pattern recognition and image interpretation provides some tools and techniques to link general and domain-centered ontologies with images. This enables a path going from the low-level features through intermediate image concept and ending to high-level concepts understandable by a human expert.

In content-based image retrieval, some works exist using ontologies to retrieve image semantic. Several of them use the ontology to annotate with text the images, like in [5]. But other works use directly the image content, like in [13] where the authors present a system combining low-level features extraction and some high-level features defined using ontologies.

Specific knowledge declaration in a computer vision system can work mainly in specific or well-defined applications, due to the current difficulties to specify all knowledge for all possible domains of application. If the intermediate-level concepts may be defined general, the top-level concepts are often restricted to a specific application. But still, this technique can be used in image retrieval to identify some intermediate structures and concepts from the image. Declarative knowledge can be one key to obtain semantic, but it cannot be used alone.

3.2 Machine learning

Machine learning is a popular and efficient trend to identify concepts in images, as well in several domains of computer science in general. It is seen as one the best way to add high-level knowledge into a computarised system without imposing a human-centered vision of this knowledge, but using a human analogy of evolving through learning.

Machine learning has been used in many different ways to help the retrieval process. In [2] and [25], learning methods are used the concepts or the categories of objects contained in for content-based retrieval.

Machine learning is also a key to obtain semantic, but as all the possible queries from user cannot be learnt, machine learning cannot be used alone for image retrieval.

3.3 Man-machine interaction

A trend started several years ago to avoid the problem of defining all the universe knowledge into a computer vision system is to build semi-automatic system. This has already been the case for other fields of computer vision where high-level interpretation is needed and cannot be completely given by the computer system alone. But in content-based image retrieval, this aspect has become even more crucial. It is now admitted that a automatic system cannot process a single query and obtain the good answer. As Santini and Jain have shown in [18], the same query (here a painting of a portrait) can be used by the same user to retrieve different categories of images (here paintings or faces). To know the exact intention from the user, some interaction is necessary and the system can hope to understand the user's whish and give a correct answer [8]. In [18], the authors propose a model of interaction where the user can readjust its view of the displayed images to help the system for finding his exact query.

It is possible to combine interaction and learning to improve the retrieval process. The most common example of this is found in relevance feedback, a technique already used in text information retrieval. Some adaptations have been tried for content-based image retrieval [9].

In the case of semantic through interaction, it has to be noted that the semantic cannot be found in a knowledge base or in the system itself. One cannot extract from the system the semantic used in the previous queries. In a way, the system tests the user and tries to extract knowledge from his desires, and not necessarily from the images. But as this technique can produce, from a high-level point of view, acceptable results, we consider interaction as a semantic extraction technique, like knowledge definition or machine learning.

We have presented in this paper several ideas to deal with semantic within content-based retrieval. We have started to incorporate these ideas into our system of image retrieval, authough some functionalities are not yet implemented. In figure 4, we are using a graphical representation that display results of image retrieval. The query image is at the lower left of the view in red square, others are arranged in the distance (distance of color histograms and distance of co-occurrence matrices) order. The graphical representation gives a more complete representation of the results for interaction.

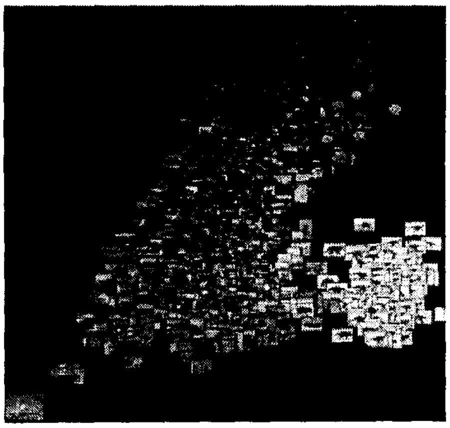


Figure 4. Graphical representation of the results that allow interaction and identify classes of images. Images in this example are coming from Wang s image database.

4. EVALUATION

The evaluation of image retrieval is an important but difficult topic. We want to have measures that satisfy two conditions. First, they are enough powerful for evaluating an image retrieval system and for comparing with other systems. Second, they are independent of human subjective evaluations. At this time, some measures are proposed but they do not satisfy both of conditions above.

In information retrieval, the main evaluation technique is based on two measures: recall and precision. Image retrieval is a subset of information retrieval. Therefore, recall and precision are also used for image retrieval evaluation. Suppose a data set D and a query q are given. Though the use of human subjectives, the data set can be divided into two sets: the set of images relevant for the query q , R(q) and the set of irrelevant images. Suppose that the query q is given to a data set and that it returns a set of images, A(q) as the answer. The precision and the recall are defined follows:

The **precision p** of the answer is the fraction of the returned images that is indeed relevant for the query.

The **recall r** is the fraction of relevant images that is returned by the query.

According to those difinitions, the precision and the recall are computed as the formulas below:

$$p = \frac{|A(q) \cap R(q)|}{|A(q)|}, \quad r = \frac{|A(q) \cap R(q)|}{|R(q)|}$$

Precision and recall are useful tools in information retrieval. But in image retrieval, they are not sufficient. According to Smeulders et al [20], there are two drawbacks when we choose the precision and the recall for evaluation in image retrieval. First, the selection of a relevant set in an image database is much more problematic than in a text database because of the more problematic definition of the meaning of an image. There is an idiom that says: **"an image worths about a thousand words"**. Second, image databases do not usually return an undiffrentiated set of "relevant" results, but a ranked list or results or some more complex configuration that shows the relation between the results of the query. In spite of these drawbacks, the precision and the recall are useful measurements in special circumstances. If the image database relies on the strong semantics provided by label or other textual description, precision and recall can be usually employed.

5. BUILDING AN ANNOTATED IMAGE DATABASE

Why do we need to build an image database? There are two main reasons. The first one is to have an image database obtained in real conditions, and not in controlled conditions for the database. We want to develop a system usable by the non-expert user taking the images he/she wants. The second one is that an image does not belong to a unique class. Usually, we take pictures contained more than one object, even if the focus is made for a specific object. The annotation and evaluation process should follow this idea. Also, we want to supply this image database for other research teams in order to compare the various existing image retrieval systems.

We are building a database containing 8800 images representing various subjects and scenes, such as flowers, animals, landscapes and people... These images are real images, taken in several different conditions, not special images taken in very strictly controlled conditions. Our goal is to provide the kind of images an average person can take in its daily life.. In the imagse, there are different objects, different scenes and different countries. It should allow to evaluate the system in real conditions, as opposed to other calibrated image database available. In figure 5, we present some images in our image database.

Figure 5. Examples of images from the MICA image database showing different type of scenes, in natural or urban conditions.

In order to use those images, they must be annotated and classified into somes classes. There is a problem while classifying them, because each image belongs to more than one classe. This is important and opposed this database to others. As it is often the case, an image contains more than a single object. Similarly, an image can belong to several classes, depending if we want the specific animal, or the type of scene or even the country where it has been taken. So we have written a program for annotating the images. This program can be used by non-expert people who are the final users aimed by the retrieval system. Several keywords have been chosen to represent the various objects and conditions in the images. The keywords must be specific enough to define the exact content of the images, but not too specific because we want to organize the images in classes (each image belongs to several classes) and one classe must have a minimum number of individuals to be validated. Figure 6 represents the interface of the image annotation system. The keywords are displayed as a tree of keywords so it is easy to select. Figure 7 represents some annotated images with the chosen keywords.

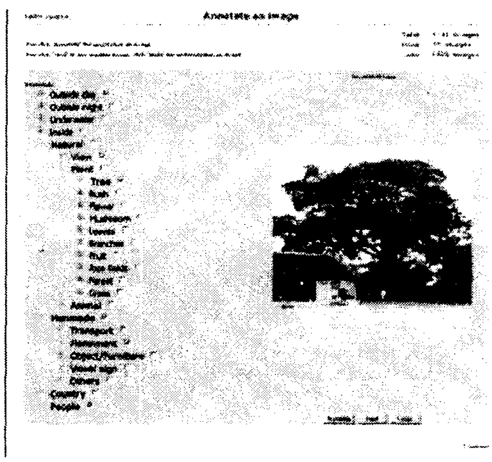
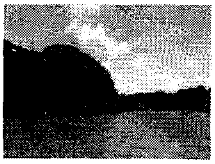


Figure 6. The image annotation system.

This program is aimed to be used not only for annotating images at MICA but also for other image databases. With different image database, we can construct different trees of keywords. It is even possible to annotate the same image database with different set of keywords that represents different applications aimed by the system.



Outside day, moutain, sky, clouds, water, Vietnam.

Outside day, animal, deer, grass

Outside day, bush, people, minorities, Vietnam.

Outside day, building, people, road, sky, Hong Kong (China).

Underwater, fish, water.

Outside day, tree, people, moto, car, building, Vietnam

Figure 7: Annotated images with chose keywords. These images are coming from the MICA annotated image database.

6. CONCLUSION

In this paper, we have review the Content-Based Image Retrieval (CBIR) systems and introduce our CBIR system based currently on color histogram and co-occurrence matrices of texture. Research on symbolic features for image retrieval is important and it has produced some results. But systems using symbolic features alone cannot work well for many image databases and real conditions of use by general users. Adding semantic to CBIR systems is an interesting approach. In section 3, we have reviewed the main techniques to add semantic in an image retrieval process: concept definion language, machine learning and man-machine interaction. We also emphasized the graphical representation that gives a more complete and intuitive representation of the results for interaction. At the moment, our system is based on some low-level features such as color, texture, but we need to do reseach on the general problem of semantic image retrieval to add semantic values into our system. Evaluation is important but it has an inherent problem, because it is lacking in objective measurements. Also, evaluation and system performance are closely related to the image database used. For this reason, we are developping our own image databases for research that used an extended annotation scheme in real image acquisition conditions.

REFERENCES

[1] R. Alferez and Y.F. Wang, "Image indexing and retrieval using image-derived, Geometrically and Illumination Invariant Features", in Proc. of the 2001 IEEE Multimedia Conference, Tokyo, Japan 2001.

[2] A. Dong and B. Bhanu. "Active Concept Learning for Image Retrieval in Dynamic Databases", *Proc. of International Conference on Computer Vision*, vol. 1, p. 90, 2003.

[3] Y. Gong, G.Proietti and C.Faloutsos, "Image Indexing and Retrieval Based on Human Perceptual Color Clustering", *International Conference on Computer Vision and Pattern Recognition*, June 1998.

[4] J. Huang, S.Kumar, M. Mitra, W.J. Zhu and R. Zabih, "Image Indexing Using Color Correlograms", *in Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR)*, San Juan, Puerto Rico, pp. 762-768, 1997.

[5] E. Hyvonen, A. Styrman and S. Saarela. "Ontology-Based Image Retrieval", *Proc. of International World Wide Web Conference*, May 2003.

[6] Q. Iqbal and J.K. Aggarwal, "CIRES: A System for content-based retrieval in digital image libaries"

[7] Q. Iqbal and J.K. Aggarwal, "Combining structure, color and texture for image retrieval: a performance evaluation", *Proc. Of International Conference on Pattern Recognition (ICPR)*, Quebec Canada, 2002.

[8] I. Keller, T. Meiers, T. Ellerbrock and T. Sikora. "Image Browsing with PCA-Assisted User-Interaction", *Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL)*, p. 102, December 2001.

[9] F. Long, H. Zhang and D.D. Feng. "Fundamentals of Content-Based Image Retrieval", in Multimedia Information Retrieval and Management – Technological Fundamentals and Applications, D. Feng, W.C. Siu and H.J. Zhang (ed.), Springer, 2002.

[10] W.Y. Ma and B.S. Manjunath, "Texture Features and Learning Similarity", *in Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 425-430, June 1996.

[11] N. Maillot, M. Thonnat and A. Boucher. "Towards Ontology Based Cognitive Vision", *Proc. Of International Conference on Vision Systems (ICVS)*, LNCS 2626, pp. 44-53, February 2003.

[12] B.S. Manjunath and W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", IEEE Transactions on Pattern Recognition and Machine Intelligence, vol. 18, pp. 837-842, 1996.

[13] V. Mezaris, I. Kompatsiaris and M. Strintzis. "An Ontology Approach to Object-Based Image Retrieval", *Proc. of International Conference on Image Processing (ICIP)*, vol. 2, pp. 511-514, September 2003.

[14] A. Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek and S. K. Ganapathy. "Matching and Retrieval Based on the

Vocabulary and Grammar of Color Patterns", Proc. of IEEE International Conference on Multimedia Computing and Systems (ICMCS), vol. 1, p. 9189-9195, June 1999.

[15] D. Park, J. Park, J.H.Han, "Image Indexing using Color Histogram in the CIELUV Color Space", *Proc. Of the 5th Japan-Korea Joint Workshop on Computer Vision*, Kyungbuk University, Korea, pp. 126-132, Jan 22-23, *1999.*

[16] M. Partio, B. Cramariuc, M. Gabbouj and A. Visa, "Rock texture retrieval using gray level co-occurrence matrix", *NORSIG-2002*, Norway October 2002.

[17] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval", *in Proceedings of the Third IEEE Workshop on Applications of Computer Vision*, Sarasota Florida, Dec 1996.

[18] S. Santini, A. Gupta and R. Jain. "Emergent Semantics though Interaction in Image Databases", IEEE Trans. on Knowledge and Data Engineering, vol. 13, no. 3, pp. 337-351, May/June 2001.

[19] E.D. Sciascio, "Similarity Evaluation In Image Retrieval Using Simple Features", *in Proc. of Storage and Retrieval for Image and Video Databases V*, February 8-14, San Jose, CA, USA, vol. 3022, pp. 467-477, 1997.

[20] A.W.M. Smeulders, M. Worring, S.Santini, A.Gupta and R. Jain, "Content-Based Image Retrieval at the End of the Early Years", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380, December 2000.

[21] M. Stricker and A. Dimai, "Color indexing with weak spatial constraints", *in SPIE Proceedings*, vol. 2670, pp. 29 - 40, 1996.

[22] Y. Sun and S. Ozawa, "A novel image retrieval algorithm by using salient points in wavelet domain", *in Proc. of ACCV*, 2004

[23] M.J. Swain and D.H. Ballard, "Color indexing", International Journal of Computer Vision, vol. 7, no. 1, pp. 11-22, 1991.

[24] J.Z. Wang, G. Wiederhold, O. Firschein, and S.X. Wei, "Content-based image indexing and searching using Daubechies wavelets", International Journal of Digital Libraries, pp. 311-328, Springer-Verlag,1998.

[25] M. Weber, M. Welling and P. Perona. "Towards Automatic Discovery of Object Categories", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, p. 2101, 2000.