HETEROGENEOUS IMAGE RETRIEVAL SYSTEM BASED ON FEATURES EXTRACTION AND SVM CLASSIFIER

Rostom Kachouri
Research unit on Computers, Imaging, Electronics and Systems, ENIS, BP W, 3038 Sfax, Tunisia
Informatics, Integrative Biology and Complex Systems, 40 Rue de Pelvoux, 91020 Evry Cedex, France

Khalifa Djemal, Hichem Maaref
Informatics, Integrative Biology and Complex Systems, 40 Rue de Pelvoux, 91020 Evry Cedex, France

Dorra Sellami Masmoudi, Nabil Derbel
Research unit on Computers, Imaging, Electronics and Systems, ENIS, BP W, 3038 Sfax, Tunisia

Keywords: CBIR, SVM, QUIP-tree, feature extraction, heterogeneous image database.

Abstract: Image databases represent increasingly important volume of information, so it is judicious to develop powerful systems to handle the images, index them, classify them to reach them quickly in these large image databases. In this paper, we propose an heterogeneous image retrieval system based on feature extraction and Support vector machines (SVM) classifier.

For an heterogeneous image database, first of all we extract several feature kinds such as color descriptor, shape descriptor, and texture descriptor. Afterwards we improve the description of these features, by some original methods. Finally we apply an SVM classifier to classify the consequent index database.

For evaluation purposes, using precision/recall curves on an heterogeneous image database, we looked for a comparison of the proposed image retrieval system with an other Content-based image retrieval (CBIR) which is QUadtree-based Index for image retrieval and Pattern search (QUIP-tree). The obtained results show that the proposed system provides good accuracy recognition, and it prove more better than QUIP-tree method.

1 INTRODUCTION

Several methods ensuring image recognition were developed. But these techniques are often developed for one kind of image and present difficulties for recognition in an heterogeneous image database.

Different applications domains like medical domain, industrial domain etc, demonstrate a real need for image recognition in large databases. To this end we can distinguish two main types of image databases: the specific database where the images show a natural similarity (the same type of images, the same content presented in a different situation, etc), and heterogeneous databases, which can contain different types and image content. One of the important steps in a recognition system is the image description. Indeed, this step is based on a priori knowledge of the image content on the one hand and on the modeled descriptors for a specific type of image. Methods based on this concept gave satisfaction for specific databases. The relevance of this description strategy becomes almost ineffective when image databases are heterogeneous. It is within this framework, that the system we present is registered. A content image recognition system is typically composed of two main phases, images description and extracted features classification allowing effective recognition.

In fact, in an heterogeneous image database, images are various categories, and we can find a big difference between them. So a unique feature or a unique feature kind, can not be relevant to describe the whole image database. In this paper, we present an heterogeneous image recognition system, to this aim, several kinds of features was used and improved for this purpose, such as color descriptor, shape descriptors and texture descriptors. The used and improved features should be efficient and relevant to describe heterogeneous images. A better images description allows to obtain a satisfactory images classification.

Since the Nineties, Support vector machines (SVMs) did not cease arousing the interest of several researcher communities of various expertise fields.

Kachouri R., Djemal K., Maaref H., Sellami Masmoudi D. and Derbel N. (2008).
HETEROGENEOUS IMAGE RETRIEVAL SYSTEM BASED ON FEATURES EXTRACTION AND SVM CLASSIFIER.
In *Proceedings of the Fifth International Conference on Informatics in Control, Automation and Robotics - SPSMC*, pages 137-142
DOI: 10.5220/0001490601370142

Such as (Schokopf et al., 1999) which was applied SVMs to insulated handwritten figures recognition, and (Osuna et al., 1997) which was applied SVMs to face recognition. In the majority of cases, SVM performance exceeds those of already established traditional models.

So, for classification, SVMs is used in our retrieval system. SVMs originally formulated for two-class classification problems, have been successfully applied to diverse pattern recognition problems and have become in a very short period of time the standard state of-the-art tool. The SVMs, based on the Structural Risk Minimization (SRM), are primarily devised in order to minimize the upper bound of the expected error by optimizing the trade-off between the empirical risk and the model complexity (Burges, 1998). To achieve this, they construct an optimal hyperplane to separate binary class data so that the margin is maximal.

To evaluate this image retrieval system, we compare it with an other Content-based image retrieval (CBIR) system: the QUadtree-based Index for image retrieval and Pattern search (QUIP-tree).

QUIP-tree indexing structure permits to store the visual characteristics of the various areas in the image. Database images are first of all compared globally with the query image. Then, if its global similarity with the query image is lower than a certain similarity threshold, the under-areas of homologous images are compared, so on until reaching the bottom level (Genevire et al., 2004) (Kachouri et al., 2007).

The paper is organized as follows: Section II describes the CBIR system Structure, and the SVM approach. Section III deals with the different features used in our system, and details the basic improvements done. Experimental results, with a brief description of the QUIP-tree technique are presented in section IV. Finally we conclude in section V.

2 CBIR SYSTEM

In this section, we first review the CBIR theory and describe its system Structure. Then we briefly outline the SVM classifier, and QUIP-tree technique.

2.1 Content-based Image Retrieval

CBIR is today ubiquitous in computer vision. Similarity queries on feature vectors have been widely used to perform content-based retrieval of images. In fact nowadays, CBIR systems allow image access according to their visual characteristics such as color, texture, shape, etc,..., by means of similarity measures. The smaller the similarity distance is, the closer the two images are.

The typical CBIR system architecture, is composed essentially of two stages. The first one is Off Line, where is carried out the feature extraction of each database image, and the storage of each feature in an index database. The second one is On Line, where is carried out the recognition (classification) by computing similarity measures between the query image signature and the index in the corresponding image database.

There are several popular CBIR systems such as: IBMs QUERY-BY-IMAGE-CONTENT (QBIC) which allows to index images using divers features. Visual SEEK (Smith and Chang, 1996) developed by Smith and Chang in the university of COLUMBIA. Surfimage developed in 1995 by INRIA, which is more sophisticated than the other commercial systems. In this paper, we propose a new CBIR system destined for heterogenous image database.

2.2 Support Vector Machines

SVM is a supervised classification method. The supervised classification, supposes that there is already an image classification. So it uses necessarily training methods which from images already classified, allow classifying new images. For image indexing systems, supervised classification allows to build a model which will classify as well as possible new images, from a classified image database.

First, in the Off Line stage: we use a training image database, which is represented by visual descriptors. With the labeled training database images, SVM learns a boundary (i.e., hyper plane) separating the relevant images from the irrelevant images with maximum margin. The images on a side of boundary are considered as relevance, and on the other side are looked as irrelevance.

Second, in the On Line stage: using the built model (boundary computed in the first stage), SVM allows to classify an evaluation image database, which must be also represented by visual descriptors.

SVM have recently attracted a lot of researchers from the machine learning and pattern classification community for its fascinating properties such as high generalization performance and globally optimal solution (Burges, 1998). In SVM, original input space is mapped into a higher dimensional feature space in which an optimal separating hyper-plane is constructed on the basis of SRM to maximize the margin between two classes, i.e., generalization ability.

2.2.1 The Separable Case

Given a set of labeled images (x_1,y_1), ..., (x_n,y_n), x_i is the feature representation of one image, $y_i \in \{-1,+1\}$ is the class label (-1 denotes negative and +1 denotes positive).
The goal is to find a boundary such as all the elements, with the same annotation, are on the same side. So we must find a vector w and a real b such as:

$$y_i(w.x_i + b) > 0, \forall i \in [1,n] \quad (1)$$

we can take so, a decision function:

$$f(x) = sign(w.x + b) \quad (2)$$

This decision function is invariant by scale change, so we choose to find the boundary which verify $wx + b = \pm 1$ for nearest elements to margin, what amounts minimizing $\|w\|^2$ such as:

$$y_i(w.x_i + b) \geq 1, \forall i \in [1,n] \quad (3)$$

Using the Lagrangian, the problem amounts maximizing W on α, and the decision function is written as follows:

$$f(x) = sign(\sum_{i=1}^{n} y_i \alpha_i x.x_i + b) \quad (4)$$

We note that if we omit the sign operator in the decision function, we obtain a belonging measurement to the required category.

2.2.2 The Non Separable Case

The above algorithm for separable data, when applied to non-separable data, will find no feasible solution. So a flexible margin may be introduced, by accepting bad classification for certain elements. This amounts to raising each α_i by a constant C.
Moreover, linear separation is not adapted to all problems, and it is often preferable to introduce a kernel $k(x,x')$ which replaces the scalar product $x.x'$.
The classification function can be written as:

$$f(x) = sign(\sum_i \alpha_i y_i.k(x_i,x) + b) \quad (5)$$

2.2.3 Choice of Kernel

The first kernel investigated for the pattern recognition problem were the following:

$$k(x,y) = (x.y + c)^d \quad \text{Polynomial} \quad (6)$$

$$k(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \quad \text{Gaussian} \quad (7)$$

$$k(x,y) = tanh(x.y + \theta) \quad \text{Sigmoidal} \quad (8)$$

The most commonly used kernel is the gaussian one. Since it allows to exploit the distance d placed into exponential:

$$k(x,y) = e^{-\frac{d(x-y)^2}{2\sigma^2}}$$

3 USED AND IMPROVED FEATURES

Feature (content) extraction is the basis of CBIR. Recent CBIR systems retrieve images based on visual properties.

As we use an heterogeneous image database, images are various categories, and we can find a big difference between their visual properties. So a unique feature or a unique feature kind, cannot be relevant to describe the whole image database. Then in this paper we are interested by divers visual feature extraction such as color, shape, texture.

3.1 Color Features

Color is one of the most important image indexing features employed in CBIR because it has been shown to be effective in both the academic and commercial arenas. Some of the popular methods to characterize color information in images are Color average and color histograms.

3.1.1 Color Average

The color average of an image is defined by $\bar{x}$, as follows:

$$\bar{x} = (\bar{R}_{(avg)}, \bar{G}_{(avg)}, \bar{B}_{(avg)})^t \quad (9)$$

where: $\bar{Color}_{(avg)} = \frac{1}{N}\sum_{p=1}^{N} Color(p)$. N is the total number of pixels in the image.

3.1.2 Color Histograms

Color Histograms are useful because they are relatively insensitive to position and orientation changes. So, despite they are so simple, they are the most commonly used color feature representation. We extract this feature just by computing the occurrence of each gray levels for R, G, and B color planes of the image.

3.2 Shape Features

Shape is a very important descriptor in image database. Generally, shape descriptor indicate the general aspect of an object, which is its contour.

3.2.1 Invariant Moments

Invariant moments are important shape descriptors in computer vision. They are obtained from quotients and powers of moments. One moment is a sum on all image pixels weighted by polynomials related to the pixel positions.

In 1962, HU derived seven bi-dimensional invariant moments (Hu, 1962).

This moments are invariant to scale, rotation and translation.

3.2.2 Sobel Filter

Sobel filter is used for contour detection. So, it is supposed that the image areas are homogeneous and that the contour can be detected on the basis of gray levels discontinuity.
First, we apply Sobel masks to obtain the directional gradients according to x and y:

$$G_x(i,j) = h_x(i,j) \otimes I(i,j), G_y(i,j) = h_y(i,j) \otimes I(i,j) \tag{10}$$

Where $I(i,j)$ is the image gray level information and $h_x(i,j)$, $h_y(i,j)$ are Sobel masks:

$$h_x(i,j) = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, h_y(i,j) = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

then, gradient norm is computed as follow:

$$G(i,j) = \sqrt{G_x(i,j)^2 + G_y(i,j)^2} \tag{11}$$

3.3 Texture Features

Multiresolution approaches to texture analysis have gained wide acceptance over the years as they effectively describe both local and global information (Julesz et al., 1978). For this we use in this paper the $Wavelet\ texture\ features$.

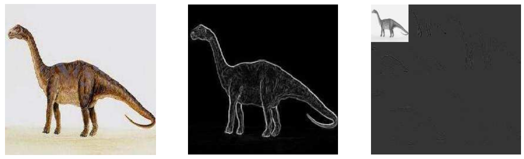
3.3.1 Daubechies Wavelet

Texture features are extracted from Daubechies wavelet coefficients of a two-level decomposition. Daubechies proposed an orthogonal wavelet construction with compact support. Daubechies wavelet has different lengthes called wavelet orders. Daubechies wavelet order, which is always even, is the number of null moments, it is related to the number of oscillations, more there is null moments, more Daubechies wavelet oscillates and so there are more regularities. Indeed, Daubechies wavelet, having M null moments, verify :

$$\Phi(x) = \sqrt{2} \sum_{k=0}^{2M-1} h_{k+1} \Phi(2x-k) \tag{12}$$

$$\Psi(x) = \sqrt{2} \sum_{k=0}^{2M-1} g_{k+1} \Psi(2x-k) \tag{13}$$

with $g_k = (-1)^k . h_{k-1}$, $k = 1, 2, ..., 2M$



a) b) c)

Figure 1: a) Dinosaur, b) Dinosaur gradient norm and c) Dinosaur Daubechies wavelet coefficients of a two-level decomposition.

Wavelet coefficients are $c^l_{ij}(x,y)$, where l is the decomposition level.

Fig. 1 shows Dinosaur image, its gradient norm, and its Daubechies wavelet transformation of a two-level decomposition.

3.4 Feature Improvement

To improve the feature size and description, we applied original modifications to some obtained feature coefficients:

3.4.1 Sobel Coefficients

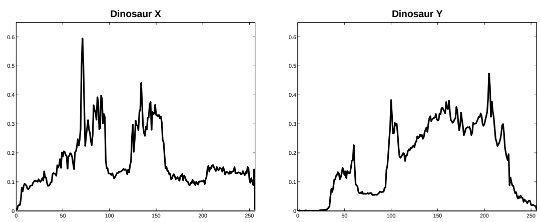
As the coefficient number in the gradient norm is the same as the pixel number in the image, we compute the gradient norm projection according to x and y, in order to reduce this feature size:

$$P_{Xi} = \frac{1}{maxG_{i,j}} \sum_j G(i,j), \text{ and } P_{Yj} = \frac{1}{maxG_{i,j}} \sum_i G(i,j) \tag{14}$$

Despite, this new form is a reduced form of the Sobel feature, it preserves the same properties of the old one.

3.4.2 Moment Coefficients

To obtain more efficient shape description by this feature, we do not use simple moments, which is computed on image pixels, but we compute moments from the gradient norm matrix obtained on sobel feature.



a) b)

Figure 2: New form of Dinosaur Sobel feature: a) The gradient norm projection according to X and b) The gradient norm projection according to Y.

So the particularity of our method, is that it combines Sobel with moments, in a new shape feature description.

3.4.3 Wavelet Coefficients

The lowest frequency coefficients $c^2_{00}(x,y)$ are not inherently useful for texture analysis. Therefore, a direction-independent measure of the high-frequency signal information is obtained by filtering the raw coefficients $c^2_{00}(x,y)$ with the Laplacian.

The texture features are obtained by computing the subband energy of all wavelet coefficients (including the Laplacian filtered $c^2_{00}(x,y)$ coefficients):

$$e^l_{ij} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}|c^l_{ij}(m,n)|^2, \tag{15}$$

where M and N are the dimensions of coefficients $c^l_{ij}(x,y)$. (see Ref. (Serrano et al., 2004) for details).

Table 1: Dinosaur and Rose texture features: subband energy of all Daubechies wavelet coefficients of a two level decomposition.

Second level decomposition				
Images	e^2_{00}	e^2_{01}	e^2_{10}	e^2_{11}
Dinosaur	226.584	11.699	8.868	6.025
Rose	252.829	12.941	7.914	4.965

First level decomposition			
Images	e^1_{01}	e^1_{10}	e^1_{11}
Dinosaur	5.184	3.755	2.494
Rose	4.141	2.458	1.294

4 EXPERIMENTS

In this section we present, first, a brief description of the QUIP-tree technique, used for comparison purpose. Then we evaluate our proposed system.

4.1 Quadtree-based Index for Image Retrieval and Pattern Search

QUIP-tree is an unsupervised classification method. The unsupervised classification, is used when images are not classified. So it is a process by which images are divided into different clusters such as images of the same cluster are as similar as possible and images of different clusters are as dissimilar as possible.

First, in the Off Line stage: we decompose database images into n quadrants, (where n is multiple of four), and we represent them by a visual descriptor by means of quadtree. Then a similarity measure is applied to calculate distance between images. Finally, a clustering of image database is applied.

Second, in the On Line stage: Image query is also decomposed into quadtree structure, after that we compare this query image with image database cluster centers to identify candidate clusters. So query image will be compared, at the end, with only images which belong to candidate clusters to finally find out similar images.

For more details see Ref. (Genevire et al., 2004), (Manouvrier et al., 2005), and (Kachouri et al., 2007).

4.2 System Evaluation

For evaluation, we tested our proposed image retrieval system, on an heterogenous image database composed of eight clusters: a collection of 400 images (50 images by cluster). The used heterogeneous database contains images having large difference in colors, shapes, and textures. Some samples are shown in Fig. 3.

To quantitatively evaluate the performances of this system, we have carried out the following tests. Queries representing different clusters are picked from the image database. Then, for each query image, a list of similar images is found in the image database, using SVM classifier.

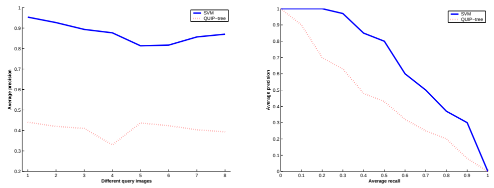
For evaluation purposes, we compare the results of our image retrieval system with other well known classification techniques QUIP-tree (see Fig. 4. (a)).

We subsequently computed the retrieval efficiency using the standard retrieval benchmarks: precision and recall (Bimbo, 2001). Let the total number of images retrieved for a query be 50, and let $x1$ be the number of images retrieved that are similar to the query. Let $x2$ be the actual number of images similar to the query in the image database. Evaluation standards recall and precision are defined as follows:

$$precision = \frac{x1}{50} \times 100\%, \text{ and } recall = \frac{x1}{x2} \times 100\% \tag{16}$$

The criteria of precision and recall are often represented like graphs called precision/recall curves. In these decreasing curves, the precision is represented in terms of recall values. Ideally precision is equal to 1 for all recall values (see Fig. 4. (b)).

Figure 3: Samples of the used heterogeneous image database.



a) b)

Figure 4: a) Average precision graph for SVM and QUIP-tree using a combination of color, shape, and texture descriptor and b) Precision/recall curves.

Since QUIP-tree is based on a computation of similarity / dissimilarity, it is efficient, only for small dimensions (only one or two same kind features). So, in (Kachouri et al., 2007), QUIP-tree proved more better than SVMs method in term of recognition rate results according to different image request, because the descriptors used for comparison are simple features (color histogram and color average), which do not permit to build a reliable model of SVMs, and image database used for tests contains synthetic images, where there are only a color variation between the different database images.

But, as soon as dimension is increased, by using more features (in order to improve description), the QUIP-tree retrieval accuracy decreases significantly, from where the favor of SVMs which in such case pass to a larger dimension, using a kernel.

Indeed, by comparing the results of our retrieval system based on SVM classifier with those of QUIP-tree, we find that in all experimental results the SVM retrieval accuracy is better than the QUIP-tree one (as shown in Fig. 4).

Fig. 5 shows the first twelve retrieval results for an example of two query image, using our proposed image retrieval system. The image displayed first is the query and ranking goes from left to right and top to bottom.

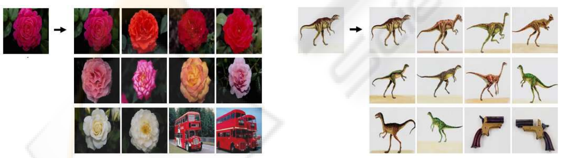


Figure 5: Retrieval results for two query image using our proposed image retrieval system.

5 CONCLUSIONS

In this paper, we have presented an heterogeneous image retrieval system based on feature extraction and SVM classifier. To evaluate this system, several kinds of features are used and improved, such as color, shape, and texture features.

The improved features have allowed obtaining a satisfactory image description. The relevance of this description is tested through an SVM classifier. A comparison with QUIP-tree technique is carried out.

As we use a real heterogenous image database, and several kinds of features to indexing images, SVMs prove more better than QUIP-tree method in term of retrieval accuracy and precision/recall curves.

Moreover, in QUIP-tree method, we calculate all distances between each image request and the other database images; whereas, with SVMs, once the model is built, each image request will be just evaluated. So, for consequent database images the SVMs answer is faster than the QUIP-tree one.

The obtained results show that the proposed system provides good accuracy recognition.

REFERENCES

Bimbo, A. (2001). Visual information retrieval. *Morgan Kaufmann Publishers*.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discovery*, 2(2):121–167.

Genevire, J., Maude, M., Vincent, O., and Marta, R. (2004). Indexation multi-niveau pour la recherche globale et partielle d'images par le contenu. In *BDA*.

Hu, M. (1962). Visual pattern recognition by moment invariants. *IEEE Transactions information Theory*, 8:179–187.

Julesz, B., Gilbert, E., and Victor, J. (1978). Visual discrimination of textures with identical third-order statistics. *Biol. Cybern.*, 31:137–140.

Kachouri, R., Djemal, K., Sellami-Masmoudi, D., Maaref, H., and Derbel, N. (2007). On the heterogeneous image retrieval with quip-tree. In *SSD*.

Manouvrier, M., Rukoz, M., and Jomier, G. (2005). *Spatial Databases : Technologies, Techniques and Trend, Quadtree-Based Image Representation and Retrieval*, chapter 4, pages 81–106. IDEA Group Publishing, Information Science Publishing and IRM Press.

Osuna, E., Freund, R., and Girosi, F. (1997). Training support vector machines: an application to face detection.

Schokopf, B., Burges, C., and Smola, A. (1999). *Introduction to support vector learning*, chapter 1. Advances in Kernel Methods - Support Vector Learning.

Serrano, N., Savakisb, A., and Luoc, J. (2004). Improved scene classification using fficient low-level features and semantic cues. *Pattern Recognition*, 37:1773–1784.

Smith, J. and Chang, S. (1996). Tools and techniques for colour image retrieval. In *IS T/SPIE Proceedings*, volume 2670, pages 426–437, San Jose, CA, USA.