An Image Retrieval Interface Based On Dynamic Knowledge

Alberto Amato[1], Vincenzo Di Lecce[1]*, Vincenzo Piuri[2]
[1] DEE – Politecnico di Bari – Via Re David 200 – 70121 Bari – Italy
E-mail: {a.amato, dilecce}@poliba.it
[2] DTI – University of Milan – Via Bramante 65, 26013 Crema (CR) – Italy
E-mail: piuri@dti.unimi.it

Abstract

Focus of this work is the presentation of an interface for an image retrieval system that makes users aware of the overall organization of the image database. The main features of this system are the multi-images query approach and the use of a dynamic knowledge base in order to improve the system performances.

The multi-images query is utilized because often a single image does not fully cover the semantic query idea of the user, so, this work presents a system that allows the user to submit multi query images.

Another relevant feature of the proposed system is the use of a dynamic knowledge based approach. This knowledge base is composed of two sets of clusters: the first one is obtained with an automatic clustering algorithm and the second one is built dynamically by the user when using the system.

The results presented in this paper demonstrate that this approach provides accurate retrieval results showing acceptable interaction speed that can be compared with existing methods..

1. Introduction

The rapid and continuous development of Information Technology has enormously increased the number of computer applications. Indeed, the computer is becoming more and more present in every day life for many people. In order to favor this computer intrusion in every day life, it is mandatory to study and realize a considerable number of human machine interfaces suitable to provide the access based on human spontaneous communication abilities [1, 2].

Image databases are an interesting framework for the study of user interfaces. The main problems in the creation and management of large image databases regard:

1. **the modality of efficient image indexing**: In the latest few years the *content-based image retrieval* (CBIR) has appeared in literature [3, 4]. This approach is based on the process of retrieving the desired image from a large collection on the basis of a set of visual features that can be automatically extracted from the images themselves (syntactic data representation) [5]. The main problem of this approach is the semantic gap. It is the difference between the similarity idea of user and the similarity criterion used by the image database system. Indeed, user's query are typically semantic (e.g. "show me an image of a red car") and they are not on the low level (e.g. "show me a predominantly red image"). As a result, most of these image retrieval systems have poor performances for (semantically) specific queries [6].
2. **the creation of an interface that allows user to explore the content of the database**: the point of question is how to develop a system that provides an environment where the users, from different background, can perform queries with semantically relevant retrieved items and without much training.

In this work, the authors propose a system that is able to tackle both the problems.

As regards the modality of efficient image indexing, the natural evolution of the CBIR systems is the relevance feedback approach. In this approach the user is seen as a part of query process and this becomes a cycling one that allows the user to reach the desired results by means of successive approximations. From this point of view, the authors propose two improvements to the classical approach:

1. The multi-images query system: often, a single image is not able to cover all the aspects of the human semantic query. As a result of this fact, the high-level concepts will have to be represented using a combination of various images. For this

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)

reason, the authors propose to use the multi-query system. This system utilizes more than one image to formulate a query. This kind of method is able to specify the image content much better than the single-query model [7].

2. The integration of a dynamic knowledge base in order to improve the effectiveness of the retrieval system.

As regards the creation of an interface that allows user to explore the content of the database the authors propose an interface that make users aware of the overall organization of the database. This is a three layer interface that allows users to interact with the image retrieval system, which is able to improve the performances of classical image databases using relevance feedback, multi-images queries and a dynamic knowledge base.

2. Problem Overview

The user sees an image retrieval system as an interface that gives him the possibility of exploring image database. In the proposed evaluation, the system is composed of a dynamic knowledge base and an exploration interface.

The dynamic knowledge is composed of two sets of clusters: the first one is obtained with an automatic clustering algorithm and the second one is built by the user dynamically when using the system.

The exploration interface is composed of three spaces and a set of operators applied to them. The spaces are:

- Features space $\mathcal{F}$: the traditional retrieval methods are based on the low-level features extraction from image (color, shape, texture, position, etc.), in order to generate a data vector (often called signature). This data vector contains data, which are easier to handle than those of the original image. However it is able to represent the image in all the retrieval processes. $\mathcal{F}$ is an *n*-dimensional space, which is defined by the coefficients used to represent the images in all the retrieval processes. This space is topological, but a metric is not defined on it. In this work the authors use two feature spaces (sets): the first one ($\mathcal{F}1$) is defined by the signature angular spectrum while the second one ($\mathcal{F}2$) is defined by the symbolic string signature. The angular spectrum is a global signature (namely referred to the entire image) that localizes the orientation of the main objects present in the image. The symbolic string is a local signature (namely referred to parts of the image) that conveys information relative to the edge of the objects present in the image.
- Query space Q : this space is the feature space endowed with a metric. This metric is derived by the user's query so that the distance between an image and the origin of the space defines the degree of "dissimilarity" of that image from the current query. The distance function used in this space depends on a number β of parameters $\in \mathcal{R}$. These parameters can be used to represent the current degree of database approximation to the user's similarity idea (semantic space). Let μ_i be the parameters that define the distance function where $1 \leq i \leq \beta$, and let x^j, y^j be the feature set associated to the images X and Y where $1 \leq j \leq n$, then the distance function can be written as:

$$f : \mathcal{F} x \mathcal{F} x \mathcal{R}^{\beta} \rightarrow \mathcal{R}^{+} : \forall\ (x^j, y^j, \mu^{\beta}) \in \mathcal{F} x \mathcal{F} x \mathcal{R}^{\beta} \rightarrow f(x^i, y^i, \mu^{\beta}) \in \mathcal{R}^{+}$$

At this time, the query space is a metrical one, and so it is possible to define other operations in it. An example of these operations is the distance from origin that, given a feature set x^j, returns its distance from the origin of the query space (the query itself):

$$D_0(x^j)=f(0,x^j,\mu^{\beta})$$

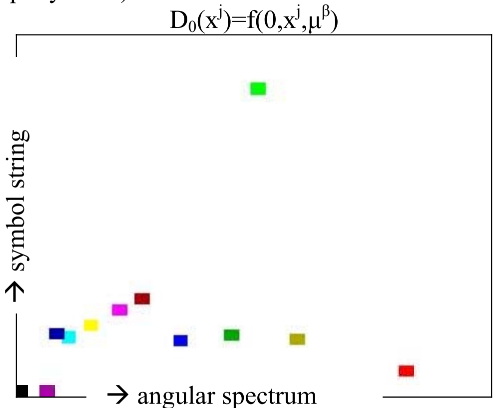

Figure 1.Cartesian representation of the automatic knowledge base and his position in compareson with the user query. In this representation, the user's query is the origin of the space, the other visible points are the centroids computed by the automatic clustering algorithm. For each point, the x coordinate is its distance from the origin obtained using the angular spectrum feature and the y coordinate is its distance from the origin obtained using the symbol string.

In this work the query space is defined by determining two different distance functions:

1. $Q1$ that is obtained applying on $\mathcal{F}1$ a weighted Euclidean distance:

$$dw(n,m) = w_{n,m} * d(n,m)$$

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

where the weight $w_{n,m}$ depends on the distance between the largest term of the reference image signature and the largest term in the first band (terms 1:10) of the examined signature;

2. *Q2* that is obtained defining on *F2* the string dist function. This distance is achieved by using the transform list concept. A transform list is a list of strings, where each string, except for the last one, can be changed to the following string by adding, deleting or replacing a character. The length of a transform list is the count of strings minus 1 (that is the count of operations to transform these two strings). The distance between two strings is the length of a transform list from one string to the others with the minimal length [8].

- Display space *D* is the space in which the images are presented to end-user. This space is often two or three-dimensional and with it the user can interact to modify some system parameters (in this work query and distance function). One of the most important problems of this space is the way in which displayed images are arranged. Displaying the entire database is impracticable, so the retrieval system applies to the query space *Q* a k-nearest neighboring operator. This operator returns the k images that are closest to the query:

$$N(k)=\{x^j:|\{y^j:D_0(y^j)<D_0(x^j)\}|<k\}$$

  It is noteworthy that this operator is applied to the query space because a metric is not defined on the feature space.

  In this work the authors use a display space composed of a three-layer structure. The first layer shows the results of a k-nearest neighboring operator applied both to *Q1* and *Q2* (where k=15) to the user. The second layer shows the user a Cartesian representation of the automatic knowledge base and his position in comparison with the user query. The third layer shows to the user the images present in the clusters (both in the clusters obtained automatically and in the clusters created by the user). In each layer there is a system that allows the user to select some images to enrich his query.

With the display space the user can interact to give the relevant feedback to the system in order to change the query point.

3. Clustering

Clustering is a well-known technique used to discover inherent structure inside a set of objects. Clustering algorithms attempt to organize unlabeled pattern vectors into clusters or "natural groups" so that points within a cluster are more similar to each other than to points belonging to different clusters.

If the database images are successfully classified into different semantic clusters, a visual query can be quickly narrowed to a specific category. Therefore, image retrieval can proceed effectively and efficiently. However, different semantic clusters can have arbitrary shapes and overlaps, so, it may be impossible to detect and specify them accurately.

In literature, many clustering methods of database images have been proposed [9, 10]. In [11] the authors propose the use of a "human domain expert" to categorize the semantic of the image database assuming that the semantic of the image database is well defined. For each semantic cluster so obtained, the centroid is computed and then it is utilized as "templates" to represent the entire cluster.

Often the semantic of the image database is not well defined because the meaning of an image is a subjective concept. In literature there are many works [12] that demonstrate that the goal of the interaction between the user and the image database is not so much retrieving images based on a preexisting semantic as creating image semantic.

The system presented in this work has an interface that allows the user to discover semantic cluster using the relevant feedback approach. Indeed, the centroids, computed during the relevance feedback phase, are surely correlated with the user's semantic concept and so they can be utilized as "templates" to represent the entire semantic cluster as shown in [11].
The user is aided in this process by an a-priori knowledge base of the image database obtained using an automatic clustering algorithm.

This algorithm is based on the k-means for the angular spectrum signature and on the k-medoid for the string signature.

4. User Interface And Query Processing

In order to select the initial query point the system implements the query by example method using the interface shown in figure 2. The user can choose an image among those stored in the image database or browse the system in order to find another image.

The so chosen image is used to endow the two feature spaces *F1* and *F2* with a metric as explained in section II.

In this way two query spaces *Q1* and *Q2* are obtained. Now the retrieval system applies to each query space *Q1* and *Q2* a k-nearest neighboring operator (with k=15).

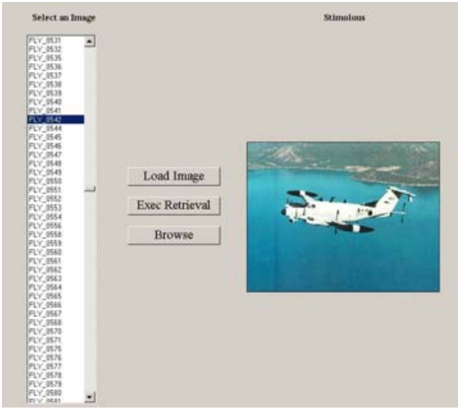

Figure 2 - The first layer of adopted system

In this way the second layer of the two display spaces $\mathcal{D}1$ and $\mathcal{D}2$ is obtained. It is shown in figure 3.

In this figure, there are $\mathcal{D}1$ on the left side and $\mathcal{D}2$ on the right side.

Around each image there are the following elements:

- **On right side** of each image there is a check button on which the user can act to indicate if that image is relevant or not relevant in relation to his similarity criterion.
- **On bottom side** of each image there are two optional buttons: a green one and a yellow one. The yellow button is present if the image is element of an automatically computed cluster. The green button is present if the image is element of a cluster built by the user in a previous retrieval session. If the user pushes one of these buttons, the third layer of the interface is shown (figure 4). If the user pushes a yellow button, this layer shows the first 21 images that constitute the selected automatically computed cluster. If the user pushes a green button, this layer shows the first 21 images that constitute the selected cluster manually built. In figure 4 it is visible a check button that has the same function of that one present in the first layer of the interface.

Using these elements the user can place a query to the image database using the mechanism of the relevant feedback with multiple query images. These images can be selected both from those shown in the first layer and from those shown in the third layer. Once that the user indicates all the images that he considers to be relevant to his similarity criterion, he can push the button “FeedBack” that is present in the bottom part of the first layer. At this time, the system will compute the centroid for each set of relevant images (namely a centroid for the angular spectrum signature and one for the string signature). The so computed centroids are used to generate the second layer (figure 1) in which a Cartesian representation of the automatic knowledge base and his position in comparison with the user query is shown. In this representation, the user’s query is the origin of the space, the other visible points are the centroids computed by the automatic clustering algorithms. For each point, the x coordinate is its distance from the origin obtained using the angular spectrum feature and the y coordinate is its distance from the origin obtained using the symbol string.

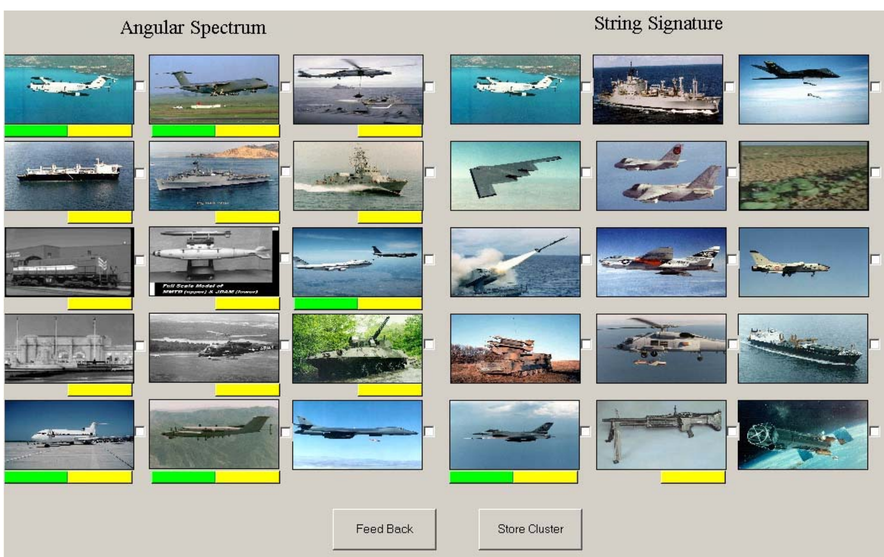

Figure 3 - the second layer of the two display spaces $\mathcal{D}1$ and $\mathcal{D}2$

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC’05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

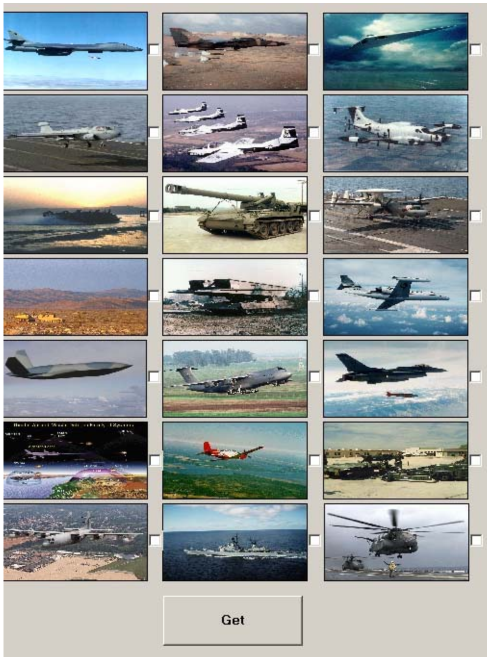

Figure 4 - The third layer of the display interface.

If the user clicks on a centroid in this layer of the interface, the third layer shows the nearest 21 images, which are elements of that cluster. At this time the user can act in the layer three to enrich his query with other relevant images.
At this time the user has composed a set of images that he consider to be relevat to his semantic query idea. Now the system computes the centroid of this semantic cluster.

The centroid relative to the set given by the angular spectrum method is computed in a natural way. It consists in computing the mean vector of the vectors (signatures) associated to each relevant image. It is clear that the same method is not applicable to the string signature, indeed, each element of this signature represents the prevailing direction of the edges drawn in a given zone of the original image. In order to overcome this problem, a probabilistic approach has been utilized. Given a set of string signatures (those relative to the relevant images), each element of the centroid signature is computed analyzing the corresponding element of every relevant signature and keeping the most frequent one.

The so obtained centroids are used as a new stimulus (one for the angular spectrum and one for the string signature) for the query in an iterative process.

When the user is satisfied of the results that he has achieved, he can push the button "Store Cluster" (in the bottom part of the first layer) in order to save the semantic cluster, that he has built with his query, in the dynamic knowledge of the system.

5. Results

The experiments have been performed on an image database of 6,240 images. The database consists of many classes, for example images of airplanes, ships, animals, landscapes, cars, pictures, and others.

Various experiments were conducted to evaluate the retrieval performance of the proposed relevance-feedback algorithm measuring precision and convergence speed.

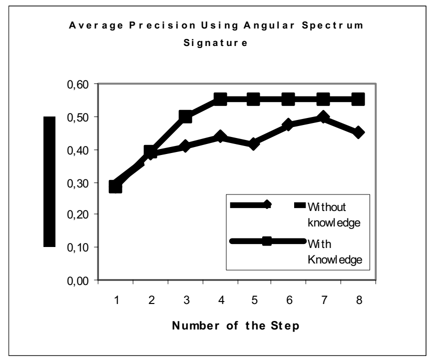

Figure 5 - Average Precision versus iteration of relevance feedback obtained using the angular spectrum signature

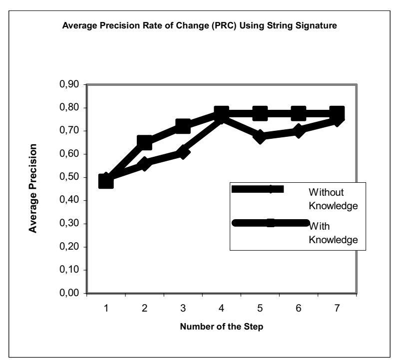

Figure 6 - Average Precision versus iteration of relevance feedback obtained using the symbol string signature

Precision is defined as the ratio between the number of relevant images and the total number of images returned by the system.

Convergence speed is measured in terms of Precision Rate of Change (PRC) using the following formula:

$$\frac{\partial P}{\partial s}$$

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

where P is the precision and s is the number of the feedback step.

In order to evaluate the effects of the proposed knowledge based retrieval system, a comparison performance evaluation between it and another system is provided in the figures 5 and 6. The other system has all the features (multi-query relevance feedback, angular spectrum and symbol string signature) of the proposed knowledge based retrieval system except for the use of the knowledge base.

Figure 5 and 6 show the results of a preliminary set of tests in terms of average precision trend.

6. Conclusions

With the rapid development of information technologies, the amount of multimedia information increases explosively. This involves a growing need for systems that allow users, from different background, to perform queries with relevant retrieved items and without much training.

In the latest years various techniques have been proposed for retrieval in image databases. CBIR systems are obtaining good results but they seem to be unable to cover the semantic gap.

Relevance feedback seems to be a promising technique because the user interaction may provide some information, to an image retrieval system, that is difficult to obtain automatically.

In this work the authors have implemented an interface suitable for gathering information from the user and to make him aware of the overall organization of the image database. In this way, the proposed interface can be seen as a tool for semantic browsing of the query space.

The search engine of the proposed system uses a dynamic knowledge base to improve the effectiveness of the retrieval. Indeed, the proposed system has a visual interface that allows the user to build semantic clusters and to store them in the knowledge base of this system. The interface has a direct method to retrieve them quickly.

Most of the image retrieval systems support only the single query model. In this query model, the image database has to find the images which are similar to a given stimulus. But, often a direct mapping between the low-level features of a single image and the high-level user concepts may not exist. For this reason the system uses the multi query images model that is useful for a detailed knowledge representation.

An advantage of the proposed multi-query model is in using the cluster centroid of a set of images as query stimulus. Of course, this set of images is semantically correlated to the user taste because it is explicitly chosen by him. So the proposed interface becomes a suitable tool for semantic browsing of the query space.

As the preliminary results shown, the introduction of the dynamic knowledge base improves both the average precision and the convergence speed of the system.

7. References

[1] L.Romary and N.Bellalem "Multimodal Interfaces for Multimedia Applications: Consequences on the Interface Design" *ERCIM News* No.46, July 2001.

[2] E.M.Petriu, G.Patry, T.E.Whalen,A.Al-Dhaher and V.Z.Groza. "Intelligent Robotic Sensor Agents for Environment Monitoring", *Visms2002 International Symposium on Virtual and Intelligent Measurement Systems*, Mt.Alyeska Resort, AK,USA,18-20 May 2002

[3] M. Flickner et al., Query by Image and Video Content: The QBIC System, *IEEE Computer*,Vol. 28,No. 9, September 1995.

[4] V.E. Ogle and M. Stonebraker, Chabot: Retrieval from a Relational Database of images, *IEEE Computer* September 1995.

[5] John P Eakins and Margaret E Graham, Content-based Image Retrieval A report to the JISC Technology Applications Programme, Institute for Image Data Research, University of Northumbria at Newcastle, January 1999.

[6] Vailaya, M. A. T. Figueiredo, A.K. Jain, H. J. Zhang, Image Classification for Content-Based Indexing, IEEE Transaction on Image Processing, Vol. 10, No. 1, January 2001.

[7] Qasim Iqbal and K. Aggarwal, "*Feature Integration, Multi-image Queries and Relevance Feedback in Image Retrieval"*, Invited Paper, To Appear, 6th International Conference on Visual Information Systems (VISUAL), Miami, Florida, Sep. 24-26, 2003.

[8] Alberto Amato, Vincenzo Di Lecce, "*Edge Detection Techniques In Image Retrieval: The Semantic Meaning Of Edge*", EC-VIP-MC 2003 4th EURASIP Conference, 2-4 July 2003, Zagreb, Croatia.

[9] M. Ester, H. Kriegel et Al., A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Proc. Second Int'l Conf. KDD, pp. 226-231, 1996.

[10] E.J. Pauwels, P. Fiddelaers and L. Van Gool, DOG-Based Unsupervized Clustering for CBIR, Proc. Second Int'l Conf. Visual Information Systems, pp. 13-20, Dec. 1997.

[12] G. Sheilholeslami, W. Chang and A. Zhang, SemQuery: Semantic clustering and Querying on Heterogeneous Features for Visual Data, IEEE Transaction on Knowledge and Data Engineering, Vol. 14, No. 5, September/October 2002.

[13] Simone Santini, Amarnath Gupta and Ramesh Jain, Emergent Semantics through Interaction in Image Databases, IEEE Transactions On Knowledge and Data Engineering, Vol. 13, No. 3, May/June 2001.

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 **IEEE**