VECIMS 2004 - International Symposium on
Virtual Environments, Human-Computer Interfaces, and Measurement Systems
Boston, MA, USA, 12-14 July 2004

An Interface for Semantic Browsing of an Images Database

Alberto Amato[1], Vincenzo Di Lecce[2], Vincenzo Piuri[3]
[1] II Engineering Faculty of Politecnico di Bari - Viale del Turismo n.8 - 74100 Taranto – Italy
[2] DEE - Politecnico di Bari - Via Re David 200 - 70121 Bari – Italy – *corresponding author*
[3] DTI - University of Milan - Via Bramante 65, 26013 Crema (CR) - Italy

Abstract *– In this paper, the authors propose an interactive graphical interface suitable for semantic browsing of an image database. This interface allows user to explore the effects of features integration, multi-image queries and relevance feedback in enhancing the performance of an image retrieval system.*
Single image query is the retrieval model that is traditionally used, while the authors propose the implementation of a new methodology consisting of multiple query images. Features integration and feedback improve the performance of the proposed system significantly. These techniques require an interactive interface which is able to gather information from user and to show him the view of the image database closest to his requirements.

I. INTRODUCTION

The rapid and continuous development of the 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 favour 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 the modality of efficient image indexing and 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 gives to users an environment in which they, from different background, can perform queries with semantically relevant retrieved items and without much training.

In the latest few years the *content-based image retrieval* (CBIR) has appeared in the 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 image features, that are commonly used in CBIR to describe the image syntactic components and retrieve images from a database in automatic processing (no human assistance), are color, texture, shape and 2D spatial relationships [6, 7, 8, 9].

The performance of CBIR is considerably improving in the latest years, but they do not seem to be able to cover the semantic gap (the difference between the similarity idea of users and the similarity criterion used by image database system) yet. Indeed, CBIR differs from the classical data retrieval systems because image databases are essentially unstructured, indeed digitalized image consists purely of array of pixel intensities without any inherent meaning.

In order to overcome this problem, innovation in image retrieval systems are the introduction of relevance feedback concept and semantic clustering. Essentially, in the first approach user is seen as a part of query process and this becomes a cycling one that allows user to reach the desired results by means of successive approximations. With this approach, the process of querying the database should not be seen as an operation during which images are filtered and based on the illusory preexisting meaning, but, rather, as a process in which meaning is created through the interaction of user and image database [10].

The clustering technique is used to classify the images into different semantic clusters in order to narrow the queries to a specific category quickly.

The aim of this paper is to present a two layers interface that allows users to interact with an image retrieval system which is able to improve the performances of classical image databases using relevance feedback, local and global feature integration and multiple query model [11].

II. RELATED WORK

A large number of experimental systems have been developed to demonstrate the feasibility of the new CBIR techniques and in particular the effectiveness of the relevance feedback approach.

The noteworthy systems, from the point of view of the interface adopted are:

- **El Niño**: is the collective name of a group of search engines, interface tools, and communication and integration modules for managing image repositories developed by S. Santini and R. Jain. The answer of the database to a query is not a list of images, but a representation of the images distribution in the database. The way in which the images are arranged on the display and their mutual distance reflects their similarity as it is currently interpreted by the system. This feedback mechanism constitutes the fundamental interaction mode between user and database, but other collateral tools can be attached to the interface as visual concept and visual dictionaries [12].
- **CIRES:** is a robust image retrieval system that serves queries ranging from image containing conspicuous structure, such as buildings, bridges and towers, to image

containing purely natural objects, such as vegetation, water, sky and clouds. This system implements the multi-image queries, relevance feedback and features integration methods [11]. Using the multi-image queries the similarity assessment is done using the following formula:

$$D(X_j, S) = \min_k d(X_j, S_k)$$

Where X_j denotes the j^{th} image in an image database X, S represents the set of query images provided by user and D is the distance of the image X_j to the set S.

III. PROBLEM OVERVIEW

User sees an image retrieval system as an interface that gives him the possibility to explore image database. In the proposed evaluation, the exploration interface is composed by 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.
- Query space $\mathcal{Q}$: this space is the feature space endowed with a metric. This metric is derived by 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 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 features set associated to the images X and Y where $1 \leq j \leq n$, then the distance function can be written as:

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

  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 shows a typical graphic representation of a query space.

- Display space $\mathcal{D}$ is the space in which the images are presented to end-user. This space is often two or three dimensional and with it 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 $\mathcal{Q}$ 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 features space. In this work k is 20.

IV. SYSTEM OVERVIEW

In this work, the authors intend to investigate a two layers interface for a relevance feedback image retrieval system. The first layer allows user to select the query stimulus using the query by example method, while the second one is suitable to gather information from user in order to implement two kinds of relevance feedback.

The first kind of relevance feedback allows user to modify the weights (μ^i where $1 \leq i \leq \beta$ using the symbols of the previous section) of a similarity function that melts two different signatures, while the second one allows user to use the multiple query images system.

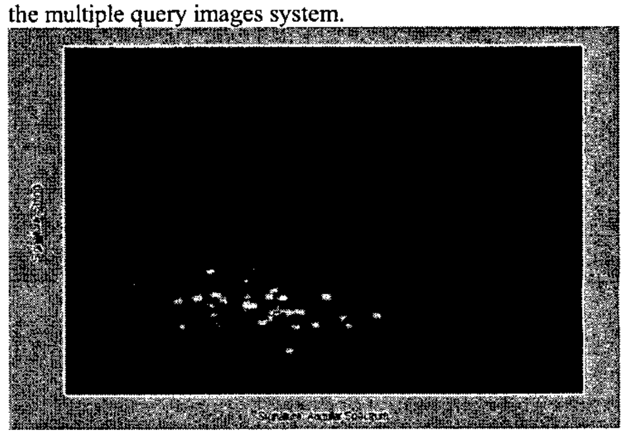

Figure 1 - A typical graphic representation of a query space. In this picture the coordinate X and Y of each point are respectively the distance of an image from the origin of the space computed using as signature angular spectrum and string distance. The luminosity of each pixel is directly proportional to the number of images that have the same distance from the origin of the space.

The system uses two signatures: angular spectrum [13] that is a global feature and a signature obtained by means of a new approach to local directionality feature that is a local one. Angular spectrum is related to the main orientation of the objects that are present in an image while local directionality is used to obtain a signature (that we call "string signature") that describes the shapes in an image by means of textual strings composed by the following characters: "-", "|", "\", "/", "o" [14]. The two features are tested and the results, in terms of effectiveness, are really good [13,14].

The authors have developed a retrieval environment in which there are:

- Three feature spaces (sets) $\mathcal{F}1 \subset \mathcal{R}^{70}$, $\mathcal{F}2 \subset \mathcal{S}^{11,11}$, where $\mathcal{S}$={"-","|","\","/","o"} and $\mathcal{F}3 \subset \mathcal{R}^{70} \mathrm{x} \mathcal{S}^{11,11}$. $\mathcal{F}1$ is defined by the signature angular spectrum, $\mathcal{F}2$ by the string signature obtained by the local directionality feature and $\mathcal{F}3$ is the Cartesian product of $\mathcal{F}1 \mathrm{x} \mathcal{F}2$.
- On these spaces three query space are defined by determining three different distance functions:
  - *Q1* that is obtained applying on $\mathcal{F}1$ a weighted Euclidean distance:

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

    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;
  - *Q2* that is obtained defining on $\mathcal{F}2$ 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 other with the minimal length [14].
  - *Q3* that is obtained defining on $\mathcal{F}3$ a weighted distance function f that, using the above symbols, is so defined:

$f: \mathcal{F}1 \mathrm{x} \mathcal{F}2 \mathrm{x} \mathcal{R}^{\beta} \rightarrow \mathcal{R}^{+} : \forall\ (x,y,\mu^{\beta}) \in \mathcal{F}1 \mathrm{x} \mathcal{F}2 \mathrm{x} \mathcal{R}^{\beta} \rightarrow f(x,y,\mu^{\beta}) \in \mathcal{R}^{+}$

- Two different display spaces $\mathcal{D}1$ and $\mathcal{D}2$. $\mathcal{D}1$ is used to display simultaneously *Q1* and *Q2*. For practical problems, $\mathcal{D}1$ displays only the results of a k-nearest neighboring operator applied both to *Q1* and *Q2*, where k=20. With $\mathcal{D}1$ user can interact to give to the system the relevant feedback in order to change the query point. $\mathcal{D}2$ is used to display the results of k-nearest neighboring operator applied to *Q3*. With $\mathcal{D}2$ user can interact to change the values of the parameters μ^{β} of the distance function f.

V. RETRIEVAL AND SEMANTIC CLUSTERING

In the traditional image retrieval systems, the similarity between two images is determined on the base of the distance between the feature vectors of the images in the feature space.

An interesting question is about the measurement method to utilize while computing the distance between two signatures. In this work a weighted Euclidean distance has been utilized for the angular spectrum signature while the String Distance method has been utilized for the string signature. Therefore, the content-based retrieval process is normally performed on individual features.

However, the low level features, utilized to indexing the image in the content-based retrieval systems, do not precisely represent the semantic of images. Thus, the feature vectors of some semantically irrelevant images may be located very close within the feature space. Let us consider two images representing sea and sky, using the feature vectors given by the color histogram, the distance between these images is very small, but they are not semantically similar. In this way, given a query whose feature vector is located in the neighborhood of the feature vectors of these two images, both images have a high probability of being retrieved togheter in response to the query. Thus, the indexing itself, based on the closeness of feature vectors in the feature space, may not sometimes provide satisfactory results.

To reduce this problem, researchers have been studying the clustering of database images. Clustering is an important 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 [15], [16]. In [17] 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 [10] that demonstrate that the goal of the interaction between the user and the image database is not so much to retrieve images based on a preexisting semantic but to create image semantic.

The system presented by this work is an interface that allows 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 [17].

VI. USER INTERFACE AND QUERY PROCESSING

The proposed system adopts a two layers interface: the first one is used to select the stimulus for the query, while the second one is employed to implement the mechanism of the relevance feedback. In the first layer (Figure 2) the system implements the query by example method. User can choose an image among those stored in the image database.

The so chosen image is used to endow the two feature spaces $\mathcal{F}1$ and $\mathcal{F}2$ with a metric as explained in section III.

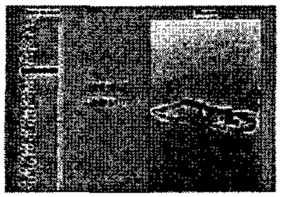

Figure 2 – The first layer of adopted system

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=20). In this way the two display spaces $\mathcal{D}1$ and $\mathcal{D}2$ are obtained. They are shown in figure 3.

User can act on $\mathcal{D}1$ and $\mathcal{D}2$ in order to realize two different relevance feedback methods. At the bottom of the windows there is a slider control on which user can act to choose the weight that the system will use in the similarity function to melt toghether the two signatures.

Various features may play different degrees of importance in making the decision about ordering database images.

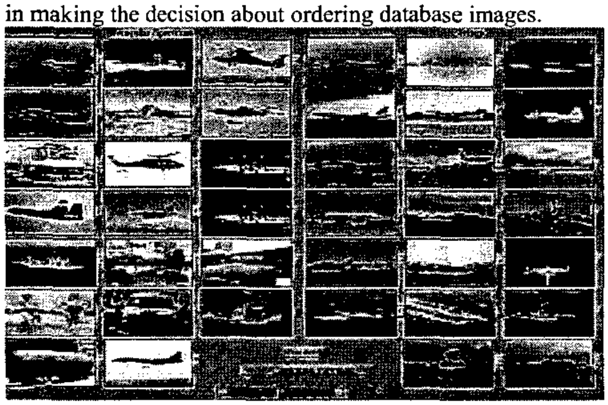

Figure 3 - The two display spaces after the first retrieval step

The ordering process must assign weights to the results obtained from the subqueries. In human perception, these feature classes do not have the same importance in distinguishing images. Thus, visual retrieval systems must consider the importance degree of each feature class to determine the overall similarity of database images to the query image.

The proposed system, like most of the current image content-based retrieval systems such as Photobook [18] and QBIC [3], uses a weighted linear method to combine the similarity measurements of different feature classes.

This first method of relevance feedback acts on $Q3$ and is used to deform this space (varying the scale of each axis) in order to change the similarity criterion exploited by the image database. This mechanism of relevance feedback is suitable to analyze the influence of the ratio between local and global visual features in the human similarity assessment criterion. Indeed, in view of the fact that the two used signatures are one global and one local, when user acts on the slider control in order to increase the weight of a feature in comparison with the other, he chooses that for his criterion of similarity, a kind of feature (local or global) is more useful than another one.

In figure 3 there are $\mathcal{D}1$ on the left side and $\mathcal{D}2$ on the right side. On right side of each image there is a check button on which user can act to indicate if that image is relevant or not relevant in relation to his similarity criterion. Using these elements user can place a query to the image database using the mechanism of the relevant feedback with multiple query images that is the second mechanism of relevant feedback which the authors are investigating. Indeed, once that user indicates all the images that he considers relevant to his similarity criterion, the system will compute the centroide for each set of relevant images.

The centroide 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 overcame this problem, a probabilistic approach has been utilized. Given a set of string signatures (those relative to the relevant images), each element of the centroide 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 Figure 1 a large number of clusters of images are clearly evident in certain zones of the query space. With the proposed method of relevance feedback it is possible to explore this space by means of small movements into the cluster containg the stimulus. The proposed interface allows user to fall into another neighborhood cluster by means of an appropriate selection of relevant images (tipically the farthest from the stimulus) in order to explore the entire image database.

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.

However, the use of more than one query images is useful for 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 user taste because it is explicitally chosen by him. So the proposed interface becomes a suitable tool for semantic browsing of the query space.

VII. RESULTS

The experiments have been performed on an image database of 7,474 images. The database consists of many classes, for example images of airplanes, ships, animals, landscapes, cars, pictures, and other.

Figure 2 and 3 present a query of airplanes. Figure 2 displays the query stimulus, while figure 3 presents the results obtained using just the query image. In *D1* 7 images of airplanes are visible (a precision of 7/20=35%), while in *D2* there are 8 images of airplanes (a precision of 8/20=40%).

Figure 4 shows the results of multi-image feedback when all the airplanes of figure 3 are used as feedback images. It is observable that the number of retrieved images containing airplanes in figure 4 increases to 13 in *D1* and to 16 in *D2*, a precisoin of 13/20=65% for *D1* and 16/20=80% for *D2*.

VIII. 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.

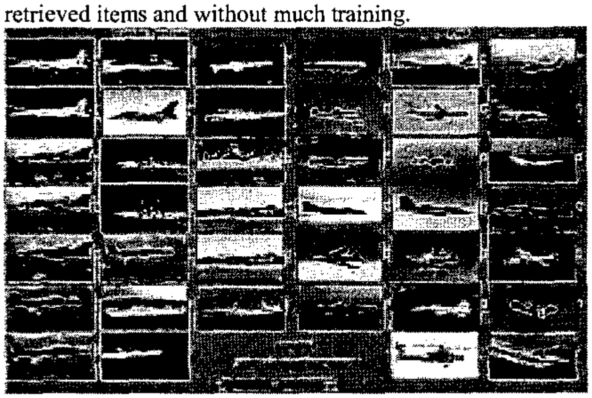

Figure 4 - *D1* and *D2* after a series of multi-image relevant feedback steps

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 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 user and to show him the view of the image database, which is semantically closest to his requirements. 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 two methods of relevance feedback, the multi-query model, and the integration of a local feature and a global one. The integration of these two types of features allows us to study the influence of the ratio between local and global visual features in the human similarity assessment criterion.

As the proposed results show, the introduction of the multi-query model boostes the performance of the retrieval system meaningfully.

Future research will focus on the evaluation of the relationship that exists between the clusters computed by using relevance feedback (human assistance) and those obtained by using automatic algorithms.

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] M. Swain and D. Ballard, Color Indexing, *International Journal of Computer Vision 7(11)* 1991.

[7] K. Hirata and T. Kato, Query by Visual Example – Content-Based Image Retrieval, *Proc. EDBT'92, Springer LNCS* 1992.

[8] A. Del Bimbo and P. Pala, Visual Image Retrieval by Elastic Matching of User Sketches, *IEEE Transactions on Pattern Analysis and Machine Intelligence 19(2)*, 1997.

[9] S.K. Chang and E. Jungert, Pictorial Data Management based upon the Theory of Symbolic Projectons, *Journal of Visual Language and Computing 2*, 1991.

[10] 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.

[11] 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.

[12] Simone Santini and Ramesh Jain, "Integrated Browsing and Querying for Image Databases", 1070-986X/00, IEEE, 2000.

[13] V. Di Lecce and A. Guerriero, "A Comparative Evaluation of retrieval methods for Duplicate search in Image database", Journal of Visual Languages and Computing, 2001, 12, 105-120.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] A. Pentland, R. W. Picard and S. Sclaroff, "Photobook: Content-based manipulation of image databases", Proc. SPIE Storage Retrieval Image Video Databases II, pp. 34-37, Feb. 1994.