A Reliable Image Retrieval System Based on Spatial Disposition Graph Matching

E. Zagrouba¹, S. Ouni², W. Barhoumi³

Abstract – Content-based retrieval of multimedia information is one of the most difficult research topics in multimedia computing and information retrieval. However, the domain of image retrieval is still an open problem in comparison with the obtained results in the domain of text retrieval. In this paper, we present a reliable approach for image indexing and retrieval mainly based on the spatial positions of significant regions extracted from the studied images, in order to reduce the semantic gap between the query image and the retrieved ones. For modelling spatial configuration, we propose a novel graph-based data structure that we called Partial Graph of Relative Spatial Relations (PGRSR). The proposed graph can be considered as a better candidate for modelling semantically rich data. It permits to describe the degree of relative positioning between the significant regions not necessary adjacent. Thus, the image retrieval will consist principally to the resolution of a graph matching problem. For this, we present a novel algorithm of graph matching while proposing a measure of similarity between graphs allowing the ranking of the retrieved images. Besides, we combine both text and image features for image indexing and retrieval in order to more reduce the semantic gap. The efficiency of the proposed approach is illustrated while studying many scenarios on two different databases. The first database is homogenous and contains 120 images (composed of 560 regions) of countries flags. The second database contains 600 heterogeneous images (composed of 3220 regions) of diverse real-world indoor and outdoor scenes. The promising recorded results were evaluated while using the precision and recall assessment criteria. Copyright © 2007 Praise Worthy Prize S.r.l. - All rights reserved.

Keywords: Graph of Relative Spatial Relations, Image Indexing and Retrieval, Matching, Regions Segmentation, Textual Keywords

I. Introduction

In recent years, very large collections of images and videos have grown rapidly. The amount of image information is rapidly rising due to digital cameras and mobile telephones equipped with such devices. In parallel with this expansion, content-based retrieval and querying the indexed collections are required to access visual information. For designing an effective image retrieval system, we find it convenient to divide existing systems into two categories. The first category concerns specific image databases for which a ground truth is available. For example, most of the modern institutes (banks, supermarkets, universities, etc) have security cameras to supervise the area of the building. These systems produce a lot of image information which have to be stored in databases for future uses. Other application areas are, for example, medical imaging, remote sensing, cartographic systems and robotics systems. The second category includes databases with heterogeneous images where no ground truth is available or obvious. Examples include stock photography and the World Wide Web.

For this category, the user should be assumed to be an average user (not an expert). In this context, generic image signatures are computed in order to describe general visual appearance. In these systems, the traditional way of searching is based on keyword indexing, or simply by manual browsing. This approach has reached a certain maturity but present yet some limitations [1].

Recently, scientific researchers are more and more interested to the investigation of Content Based Image Retrieval (CBIR) techniques for indexing images using visual features such as colour, texture and shape. Most of the existing approaches, such as the system QBIC [2], are based on the query by example. This approach consists on sending the entire image to the search engine as a visual query. In this case the retrieval results express an overall global visual similarity. This approach provides approximate results when the focus of the research is a specific object or part of the image. Partial query formulation allows the user to specify which part of the image is the target of her interest and leads to a higher user satisfaction. In this context, partial query systems based on image regions have been
proposed to allow more specific queries what increases the user satisfaction. However, the number of systems [2] [3] [4] [5] that took in account the spatial disposition between regions is very limited, and the existing solutions suffer from the instability of images representation by regions adjacency relationships which also depend strongly on the quality of the used segmentation technique [6] [7] [8] [9] [10]. In addition, the combination of image and text information for the retrieval process is of great interest for semantic gap reduction. But, this hybrid text-image retrieval mode is not yet well considered [11] [12].

In this paper, we suggest to minimize the semantic gap by proposing a Partial Graph of Relative Spatial Relations (PGRSR). This graph is a better candidate for modelling semantically rich data. It permits to describe the degree of relative positioning between the different significant regions composing the image. Thus, we present a new algorithm for graph matching. Furthermore, we combine both text and image features for indexing and retrieval what allows us to record very promising results as well as for homogenous databases as for heterogeneous ones. The rest of the present paper is organized as follows. In the next section, we describe the architecture of the proposed system while detailing its components. The experimental results and performances study is reported in section 3. Finally, section 4 synthesizes some concluding remarks and some further works.

II. The Proposed System

The proposed system of content-based image retrieval ensures two main functionalities: image indexing and image retrieval (see Fig. 1). The indexing phase is responsible for extracting appropriate regions features, after images segmentation, and storing them into the image database. This phase is usually performed off-line. However, the retrieval phase is performed on-line and aims to extract the most similar images to a query one. In particular, it allows users to search images containing a set of objects respecting a particular spatial distribution. The final image retrieval phase supports one or more of the following options: textual query, visual query, visual query in defined spatial disposition and hybrid query (visual + textual). In what follows, we will detail the indexing phase and the retrieval phase in terms of two subsystems.

II.1. The Indexing Subsystem

Indexing is often used as identifying features within an image. Here, indexing data structures define structures permitting to accelerate the image retrieval. This phase is mainly composed of four stages:

- Region segmentation: The first step of the indexing subsystem permits to detect, in each image, the set of the regions which are visually distinguishable, while using a coarse segmentation. To realize it, we used our ameliorated version of the Fuzzy-C-Means (FCM) technique [13] which is an adopted segmentation approach performing an unsupervised and fast segmentation.

- Visual features extraction: After the segmentation in regions of each image, every detected region is characterised by a set of features reflecting its visual content in terms of colour (colour histograms [14]), texture (co-occurrence matrix [15]) and shape (invariant geometric moments [16]).

- Partial graph of relative spatial relations (PGRSR): In this stage each image is represented by a partial graph of relative spatial relations of the detected regions (PGRSR), which is a complete none-oriented, graph. In such PGRSR, image regions are represented by nodes and their relationships are represented by arcs (edges) between these nodes. Both of nodes and arcs are labelled by attributes corresponding respectively to the properties of the regions and to the relationships relating them. These relationships are the Relative Spatial Relation (RSR), which describes the degrees of inclusion and positioning of a region relatively to another one necessary neighbour region. These measures are defined by:
  - Degree of positioning on the right of a region \( R_i \) relatively to a region \( R_j \):
    \[
    P(R_i \rightarrow R_j) = \frac{1}{\text{Surf}(R_j)} \times \left| \text{card}\{ p \in R_i / \text{abs}(p) \in \left[ x\text{End}(R_j), x\text{End}(R_i) \right] \} \right|
    \]

  - Degree of positioning on the left of a region \( R_i \) relatively to a region \( R_j \):
    \[
    P(R_i \leftarrow R_j) = \frac{1}{\text{Surf}(R_i)} \times \left| \text{card}\{ p \in R_i / \text{abs}(p) \in \left[ x\text{Beg}(R_j), x\text{Beg}(R_i) \right] \} \right|
    \]

  - Degree of positioning in the top of a region \( R_i \) relatively to a region \( R_j \):
    \[
    P(R_i \uparrow R_j) = \frac{1}{\text{Surf}(R_i)} \times \left| \text{card}\{ p \in R_i / \text{ord}(p) \in \left[ y\text{Beg}(R_j), y\text{Beg}(R_i) \right] \} \right|
    \]

  - Degree of positioning in the bottom of a region \( R_i \) relatively to a region \( R_j \):
    \[
    P(R_i \downarrow R_j) = \frac{1}{\text{Surf}(R_i)} \times \left| \text{card}\{ p \in R_i / \text{ord}(p) \in \left[ y\text{End}(R_j), y\text{End}(R_i) \right] \} \right|
    \]
Fig. 1. Overview of the proposed system
Degree of inclusion of a region $R_i$ relatively to a region $R_j$ (5):

$$P(R_i \subset R_j) = \frac{1}{\text{Surf}(R_i)} \text{Surf}(R_i \cap \left[ [x_{\text{Beg}}(R_j), x_{\text{End}}(R_j)] \times [y_{\text{Beg}}(R_j), y_{\text{End}}(R_j)] \right])$$

where, $\text{Surf}(R)$ is the area of the region $R$, $\text{abs}(p)$ and $\text{ord}(p)$ are successively the abscissa and the ordinate of a pixel $p$ belonging to an image region.

The aforementioned measures $P(R_i \leftarrow R_j)$, $P(R_i \rightarrow R_j)$, $P(R_i \rightarrow R_j)$, and $P(R_i \subset R_j)$ represent the degree of positioning respectively on the left, in the top, on the right, in the bottom, and of the degree of inclusion. These relative space relations aim to calculate the position of each region $R_i$ relatively to another one $R_j$. Our spatial relations differ from the relations of Allen [3] which are deterministic: a region is either absolutely on the right, on the left, in top or in bottom of another region. Indeed, the proposed new way of positioning two regions, not obligatory adjacent, allows to minimize the effects of the meted problem during any process of regions matching being based on the graphs of adjacencies and on the relative positions of the regions (or Allen relations). The following figure (see Fig. 2) illustrates an example, which is very frequent, where the relations of Allen do not permit to properly define the space provision of two regions whereas our solution allows it robustly.

Fig. 2. Limits of spatial relations of Allen [3]

Fig. 3 illustrates an example of an image and its corresponding PGRSR and adjacent graph. We present the different degree of spatial disposition between two regions given by RSR. These degrees represent that the region $R_2$ is not completely in bottom of the region $R_8$ but have a part in the left of $R_8$ which is represented by the red area. However, adjacent relation shows that the region $R_2$ is in bottom of $R_8$ what is not completely true. These RSR are stable and permit to describe precisely the spatial disposition comparatively to the adjacent relation.

- **Classification**: The last step of the indexing subsystem is the classification process. This step allows the classification of the extracted regions, from the image database, considering their visual content defined by the two precedent steps (visual features extraction and PGRSR) and the set of textual keywords associated to each image. To do it, we used the classic method of Hierarchical Ascending Classification (CAH) [3]. In fact, we start by a visual classification of the regions, such that every gotten region category is represented by a prototype. These prototypes correspond to the nearest regions relatively to each class centroid. As a result, we construct with these prototypes the visual thesaurus. Besides, we realise semantic image categorisation, while describing each image by a set of key words allowing the textual thesaurus definition.

II.2. **The Retrieval Subsystem**

The retrieval processing is organized as follows. The interface allows the user to specify a query while organizing the desired regions in a well defined spatial partition respecting a specific semantic. The retrieval system support one or more of following options: textual query RT; visual query in a particular defined spatial disposition RVS; or hybrid query $R = RT \land RVS$. Indeed, the proposed system tries to be as supple as possible, while proposing to the user the formulation of its query from the visual and textual thesaurus. Given a query image, the adopted strategy for the retrieval of the most similar images in the database, consists mainly to spatial disposition graphs matching algorithm. In fact, every query is composed of a set $CR$ of selected regions categories. The result $R_f$ produced by the retrieval process is a set of images, $S_{rel}$, containing at least one region of every query categories. Then, the interest is focused on the visual retrieval in a defined spatial disposition, what corresponds to the refinement of the preliminary results of the visual retrieval. In fact, we integrate not only the composition of the query but also the spatial disposition respected by the concerned regions.

We have seen in the precedent section that the relative spatial relationship among the segmented regions in an image can be represented by PGRSR. Formally, the PGRSR is defined as follows:

- **Definition 1**: A PGRSR $G$ is a quadruplet $G = (V, E_{G}, \nu, \varnothing)$, where:
  - $V$ is a finite set of nodes
  - $E_{G} (\subseteq V \times V)$ is a finite set of edges between regions
  - $\nu : V \rightarrow A_{V}$ is the function generating node attributes
Fig. 3. Comparison between RSR and adjacent relations
\[ \xi : E_S \rightarrow A_{ES} \text{ is a the function generating spatial edge attributes} \]

Such that, a node \((v \in V)\) corresponds to a region and a spatial edge \((e \in E)\) represents a spatial relationship between two regions not necessary adjacent. The node attributes \(A_r\) define the corresponding region content in terms of colour, shape and texture. Finally, the spatial edge attributes \(A_{ES}\) indicate the relative spatial relationships between two nodes while specifying the degrees of positioning and inclusion between the two corresponding regions. Thus, for each image \(I\), we determine its \(PGRSR\) defined as a Matrix of Relative Spatial Relations (MRSR), such that:

\[ MRSR_I = \left( A_{ES}(i,j) \right), \quad 1 \leq i,j \leq |V| \]  

(6)

and:

\[ A_{ES}(i,j) = \begin{cases} P(R_i \leftarrow R_j), & P(R_i, R_j), \\ P(R_i \rightarrow R_j), & P(R_i \downarrow R_j), & P(R_i \subset R_j) \end{cases} \]

(7)

Moreover, the problem of retrieving images, which are similar to a given query, is transformed into a problem of computing the similarity degree between the correspondent graphs, which is also known as graph matching. In practice, there are several classes of graph matching problems [4] [17]. In the graph isomorphism problem, it is determined whether two given graphs are isomorphic to each other. An isomorphism is a bijective mapping between the nodes of the two studied graphs such that the structure of the edges is preserved. In the sub-graph isomorphism problem, given two graphs \(G_q\) (query) and \(G_I\) (stored PGRSR), it is found out whether \(G_I\) contains a sub-graph that is isomorphic to \(G_q\). There is always a sequence of “edit” operations that transform \(G_I\) to a sub-graph of \(G_q\). These edit operations take the form of node or edge insertions, deletions and substitutions (see Fig. 5). To describe our graph-based algorithm, we first define sub-graph isomorphism as follows (see Definitions 2, 3, 4).

- **Definition 2:** A function \(f: V \rightarrow V'\) is structure-preserving from a graph \(G=(V, E, v, \xi)\) to a graph \(\tilde{G}=(V', \tilde{E}, \tilde{v}, \tilde{\xi})\) if for any edge \(e=(v1, v2) \in E\) it exists an edge \(\tilde{e}=(\tilde{v}1, \tilde{v}2) \in \tilde{E}\) such that \(\tilde{\xi}(\tilde{e}) = \xi(e_{\tilde{e}})\).

- **Definition 3:** A bijective function \(f: V \rightarrow V'\) is a graph isomorphism from \(G\) to \(G'\) if it is structure-preserving from \(G\) to \(G'\) and \(f^{-1}\) is structure-preserving from \(G'\) to \(G\).

- **Definition 4:** An injective function \(f: V \rightarrow V'\) is a sub-graph isomorphism from \(G\) to \(G'\) if it exists a sub graph \(S \subseteq G'\) such that \(f\) is a graph isomorphism from \(G\) to \(S\).

Thus, the query is considered as a sub-graph isomorphism problem. Then, the problem of matching \(G_q\) to \(G_I\) returns to the problem of detecting an injection \(f\) from \(I_q\) to \(I\) which is structure-preserving. In order to overcome this problem, we propose a new graph-matching method (see Fig. 4). The first step of this method consists on resizing the graphs in order to have the same number of nodes (which corresponds to the minimum size of the two studied graphs). Afterwards, the number of edges will be also reduced what minimizes the number of comparisons too. Indeed, there are often some nodes in an image \(I\) that are not necessary representing the topological structure (they are not belonging to \(CR\)). These nodes must be removed. After, we calculate the difference between a considered image \(I\) and the query \(I_q\). The matrix of difference \(D_{I_q}\) is defined as follows (eq. (8)):

\[
D_{I_q}[i,j](k) = \left| MRSR_I \left[v_i, v_j\right](k) - MRSR_{I_q} \left[Crs(v_i), Crs(v_j)\right](k) \right|
\]

where, \(1 \leq i,j \leq |V|\) and \(Crs(v_i)\) is the correspondent region of the node \(v_i\) in the other graph.

This permits to estimate the distance between corresponding query and image nodes. However, if the PGRSR \(GI\) has nodes \(vn, ..., vm \in V\) which are belonging to the same category, we find \(G'_{ab}, ..., G'_ml\) sub-graphs isomorhisms. In this case, we choose the nearest sub-graph to the query graph \(G_q\). Besides, we eliminate the redundant category nodes which are not belonging to the selected sub-graph (the less value of \(SimGraph\) is the more similar subgraph). Finally, we compute the similarity degrees and we determine a threshold in order to keep only the most similar images ordered according to their degrees of similarity (see Fig. 4 and Fig. 5).

**Algorithm: Graph matching**

**Input:** a query example \(q\), composed by a set of categories \(CR\) and the set \(S_{vis}\) of visual retrieval result

**Output:** \(k\) most similar images

1: for each image \(i \in S_{vis}\) do
2: extract Matrix of Relative Spatial Relations \(MRSR_I\);
3: if \(G_q\) and \(G_i\) are \(\xi\) isomorphic then
4: if the set of categories \(I \in CR\) then
5: resize the matrix \(MRSR_i\);
6: end if
7: Compute the matrix \(D_{vis}\) of distances;
8: Eliminate the redundant regions of a category;
9: end if
10: find \(k\)-nearest neighbour \(G_i\);
11: done;

Fig. 4. The proposed algorithm for graph matching.
So therefore, the nearest graph to $G_q$ is the one corresponding to the image $I$ which minimizes the similarity between its graph and $G_q$ according to the following equation:

$$SimGrap(I_i, I_q) = \sum_{i,j=1}^{m} \sum_{k} \left| \text{Diff}[i,j](k) \right|$$

where, $i\neq j$ and $m=\text{card (CR)}$.

At the end, note that the result of the hybrid retrieval is the logical intersection of the visual retrieval result with the textual one. Therefore, the resulting images are semantically and visually similar to the input query image.

**III. Results and Evaluations**

Many scenarios of our application are investigated on two different databases. The first database is homogenous and contains 120 images (composed of 560 regions) of countries flags [18]. However, the second database contains 600 heterogeneous images (composed of 3220 regions) of varied real-world indoor
and outdoor scenes. Given the heterogeneous database, to search landscape, the user will look, for example, for the images with some sky, vegetation and sea. This can be translated into the following query composition: “images composed of a green region and a blue region”. Fig. 6 illustrates this query. The first and the second queries are composed by the same regions but in different spatial disposition. We can observe that the presence of green region and blue region is mostly matched by landscape. Each result lists the most four matching images, respectively from left to right. Besides, we can confirm that when the spatial disposition differs then the retrieval results change considerably. In the first list, the blue region present sky and in the second it presents sea. Thus, we can deduce that the spatial disposition performs strongly the introduction of semantic information in the retrieval result.

Moreover, the user can compose his query from visual and textual thesaurus. Fig. 7 shows results of textual retrieval. However, Fig. 8 illustrates the recorded results while using spatial disposition. In fact, the query image composed by the user indicates that he looks for images containing a dog in the left and a person (and/or an object) in the right with a specific characteristic (red color) such as red clothing.

In the next figure (see Fig. 9), hybrid text-image retrieval was done, while considering the intersection of visual and textual retrieval results. We can conclude easily that the final results are very promising after combining text and image information.

Moreover, to put in exergue the efficiency of the recorded results recorded with the proposed system, we evaluated the retrieval performance while using the recall and precision assessment criteria as it is the standard used measure in all existing CBIR systems [19] [20] [21]. On one hand, Recall measures the ability of retrieving all relevant or similar items in the database. It is defined as the ratio of the number of relevant or perceptually similar retrieved items (images) by the total number of relevant items in the used database. On the other hand, Precision measures the retrieval accuracy and is defined as the ratio of the number of relevant or perceptually similar items by the
total number of the retrieved items. The curves in Fig. 10 present a comparative study of the two different techniques of retrieval by composition without and with spatial disposition applied respectively on the flags database and on the heterogeneous database. It is clearly seen that introducing spatial disposition information provides better retrieval effectiveness.

IV. Conclusions and Further Works

In this paper, we presented the architecture of our system of content based image retrieval. It is based on an efficient method for indexing and retrieving images while comparing spatial disposition of a set of regions. Indeed, each image’s content is expressed by Partial Graph of Relative Spatial Relations (PGRSR). The proposed graph-based data structures provide a reliable solution for the description of relationships between regions. These relationships are used to describe the relative position of regions within an image. Moreover, we have proposed a new algorithm of graph matching. In addition, we have combined both text and image features for indexing and retrieval in order to reduce, as maximum as possible, the semantic gap between the query and the retrieved images. Experimental results and evaluations on two different real-world image databases show the effectiveness and the accuracy of the proposed approach and the importance of spatial disposition and textual refinements for the semantic retrieval.

As further works, we are now applying the proposed system on medical image indexing and retrieval in the final aim to propose an efficient tool for computer-aided diagnosis system for radiologists based on medical images treatments. Besides, the extension of the proposed system for video indexing and retrieval seems to be a very promising sees of research. Furthermore, considering its intrinsic structure, our system is naturally appropriate for research in great bases of images. The similarities for the regions are pre-calculated, and the search time is completely independent of the adopted descriptors. The major disadvantage is the manual annotation which remains painful. In order to facilitate this task, the last perspective aims at making the annotation of the regions semi-automatic by propagating the key words of the images annotated towards the images of similar photometric aspects. Finally, the improvement of the representative regions (prototypes) choice can be considered as an interesting prospect. Indeed, these representative regions, which are posted in the request user-interface, guide the user in the choice of the request categories. Their visual relevance takes part, therefore, with the effectiveness of the research. The categories being obtained by classification of the descriptors, we naturally defined the representative region of a category as being the nearest region to the prototype of the category (i.e. average of the category). This choice is simplest and most natural, but it would be
interesting to study other manners of representing each category, so that the user lay out the best possible outline. In fact, the best regions candidates should be visually significant to support those of sufficiently compact form to be visible. In addition, it would be interesting to study if the use of several regions allow the representation of a category in a more relevant way.

References


Authors’ information

1High Institute of Computer sciences
Unité de Recherche en Systèmes Intelligents en Imagerie et Vision Artificielle (URSIVA-ISI), 2 Rue Abou Rayhane Bayrouni, 2080 Ariana, Tunisia.

Faculty of Sciences of Tunis, Unité de Recherche en Systèmes Intelligents en Imagerie et Vision Artificielle (URSIVA-ISI), Campus Universitaire, 2002, Tunis, Tunisia.

1Higher School of Technology and Computer Sciences, Unité de Recherche en Systèmes Intelligents en Imagerie et Vision Artificielle (URSIVA-ISI), 45 Rue des Entrepreneurs, 2035, La Charguia, Tunisia.

Ezzedine Zagrouba has obtained his PhD degree from the Polytechnic National Institute of Toulouse (France) in 1994. Since 1995, he is a Computer Sciences full Professor in the Faculty of Sciences of Tunis (FST) and then in the Higher Institute of Computer Sciences (ISI) in Tunisia. His activities of research within the laboratory URSIVA-ISI are interesting in the theme of Image Analysis and Computer Vision.

Sonnia Ouni received her Master degree from the Faculty of Sciences of Tunis (FST), University of El Manar (Tunisia) in 2005. She is currently a PhD student in the FST and an Assistant in the same Faculty. Her research interests, within the laboratory URSIVA-ISI, focus mainly on the themes of Medical Image Analysis and Computer Vision.

Wald Barhoumi is born in June 1975 at Kairouan, Tunisia. He received his PhD degree from the National School of Computer Science (ENSII), University of Mannouba (Tunisia) in January 2006. He is currently an Assistant Professor at the Higher School of Technology and Computer Science (ESTI), Tunisia. His research interests, within the laboratory URSIVA-ISI, focus mainly on the themes of Medical Image Analysis and Computer Vision.

Copyright © 2007 Praise Worthy Prize S.r.l. - All rights reserved