Multi-Scale Sub-Image Search

Nicu Sebe Michael S. Lew D.P. Huijsmans
Leiden Institute of Advanced Computer Science,
Niels Bohrweg 1, 2333 CA, Leiden, The Netherlands
{nicu mlew huijsman}@wi.leidenuniv.nl

ABSTRACT

If an image should be retrieved by its subregions from a large image database, an immense number of possible queries will appear. Therefore, the index which encodes the spatial information of an image, should make only few assumptions about possible queries. In addition, this index has to consider different scales of objects in the image. In this paper, we propose a novel approach using a hierarchical index encoding image regions, gained by a fixed partition. The suggested index uses color features and is easy to implement. The index is tested on a database with more than 12,000 images.

1 INTRODUCTION

While most image retrieval systems retrieve images based on overall image comparison, users are typically interested in "object searching", where they can specify an "interesting" subregion (usually an interesting object) of an image, as a query. The system should then retrieve images containing this subregion (according to human perception) from a database. This task called *sub-image query*, is made challenging by a wide variety of effects (different viewing positions, camera noise, object occlusion, etc.) that cause the same object to have different appearances in different images. The system should also be able to solve the *localization problem*, i.e. it should find the location of the object in an image. The lack of a good image segmentation process suitable for large, heterogeneous image databases, implies that objects have to be located in unsegmented images, making the location problem difficult.

These simple considerations call for two important demands of an image index: (1) the index must have a hierarchical structure to cope with details of the image, and (2) the index should make nearly no assumptions about possible queries, since it is not possible to know in advance if the image should be retrieved by one subregion or by another. This is one of the drawbacks if the image is segmented explicitly in objects.

To process user queries, some systems use fixed grid segmentation methods [3, 1]. For further improvement on the efficiency and the effectiveness of content-based retrieval, a multiresolution matching approach [2] has been proposed. Here, the query image may be a hand-drawn sketch or (a potentially low-quality) scan of the image to be retrieved. However, in many situations, users are interested in, or can remember only local image contents, therefore, sub-image query processing is needed. Unfortunately, not many image database management systems can handle arbitrary-size sub-image queries based on color and spatial similarity. For the systems that can deal with sub-image queries of arbitrary size, multiresolution matching is not used.

This paper suggests a novel approach which leads to an index that (1) regards the hierarchical composition of an image, (2) makes nearly no assumption about possible queries and (3) is small. This is done by indexing only subregions and no specific objects. The image is partitioned in overlapping rectangles of different sizes. This allows to deal with the different scales of objects in the image. Each of these regions is represented by global information and, for efficiency reasons, instead of storing the whole global features, only the differences of the feature vectors are stored. The proposed index uses color features, but these ideas can be easily extended for other global features.

2 SUB-IMAGE QUERYING

The sub-image querying problem can be defined as follows: given as input query a sub-image $\mathcal{Q}$ of an image $\mathcal{I}$ and an image set $\mathcal{S}$, retrieve from $\mathcal{S}$ those images $\mathcal{Q}'$ in which query $\mathcal{Q}$ appears according to human perception (denoted $\mathcal{Q} \subseteq \mathcal{Q}'$). The problem is made more difficult than image retrieval by a wide variety of effects (such as changing viewpoint, camera noise, etc.) that cause the same object to appear different in different images.

2.1 Performance measures

Let $\mathcal{I}_i$ be the correct answer for $\mathcal{Q}_i$ and let $rank(\mathcal{I}_i)$ be the rank assigned by the retrieval method. We use the following measures: (1) *Average r-measure*, the averaged sum over all queries of the rank of the correct answer, $1/n \sum_{i=1}^{n} rank(\mathcal{I}_i)$, (2) *Average precision* of a method $1/n \sum_{i=1}^{n} 1/rank(\mathcal{I}_i)$, the sum over all queries of the precision at recall equal to 1. Note that a method is good if it has a low *r-measure* and a high *precision*.

2.2 Location problem

The location problem is the following: given a query image $\mathcal{Q}$ and an image $\mathcal{I}$ such that $\mathcal{Q} \subseteq \mathcal{I}$, find the "location" in $\mathcal{I}$ where $\mathcal{Q}$ is "present". Efficiency is required because huge amounts of data need to be processed.

The location problem can be solved by using a fixed rectangular partition. Using such partition, we obtain a precise localization of the sub-images that can be retrieved as matches for the user query. Problems appear when the desired sub-image covers parts of two or more neighboring possible locations drawn by the fixed grid. To overcome this problem we chose the rectangles to be overlapping. Moreover, we consider a partition formed of different sized rectangles in order to account for different scales of regions in the image.

3 IMAGE FEATURES

Color indexing is one of the dominant methods of the visual media retrieval methods. In color indexing, histogram methods [6] are often used but, even if they are relatively insensitive to position and orientation changes, they do not capture the spatial relationships of color regions and thus, have limited discriminating power. Some authors showed [5] that characterizing one dimensional color distributions with the first three moments is more robust and more efficient than working with color histograms. A further improvement can be achieved by taking the covariance and the mean of the color distribution in a multidimensional color space [4]. The color features representing the color distribution are: (1) the *average color* $\mu = (\mu_L, \mu_a, \mu_b)$ and, (2) the *covariance matrix* $[\sigma_{ij}]$ $(i, j \in \{L, a, b\})$ of the color channels. The La^*b^* color space is chosen because it is perceptually uniform. Since the covariance matrix is symmetric, only 6 entries have to be stored and considering the 3 mean entries, we obtain a nine dimensional global color feature $\nu_{color}(A)$.

To determine the similarity of two n-dimensional feature vector, a weighted L_1-norm, which is dependent of the database entries, is introduced:

$$|\nu(A)|_{db} = \sum_{i=1}^{n} \frac{|\nu_i(A)|}{\alpha(\nu_i)} \tag{1}$$

where $\alpha(\nu_i)$ are the standard deviations of the respective features over the entire database. Finally, the similarity of two color feature vectors $\nu_{color}(A)$ and $\nu_{color}(B)$ is given by $|\nu_{color}(A) - \nu_{color}(B)|_{db}$.

4 INTER HIERARCHICAL DISTANCE

Let A be a two-dimensional image pixel array partitioned into L subsets A_l. A convenient representation of the image information can be found by taking the distribution of the local features. This leads to a normalized probability density function represented by a histogram. Unfortunately, this approach suffers from complete lack of spatial information, which makes it difficult to index image regions properly. One solution is to extract the global features of image subregions $\nu(A_l)$, but this increases the index size dramatically. However, if only the differences of the global features of the image and its subregions are stored, then the spatial encoding is guaranteed without a major increase of the index size. We introduce a measure of the distance between the global features of the image and the features of its subregions: *inter hierarchical distance* (IHD) taken between feature vectors of different hierarchical levels of the image partition [1].

In the case of color features, a two dimensional IHD vector ν_{IHD} is used. The vector components are the L_1-norm of the differences of the mean and covariance elements, respectively:

$$\nu^l_{IHD,1}(A) = \sum_{i \in \{L,a,b\}} |\mu_i(A) - \mu_i(A_l)| \tag{2}$$

$$\nu^l_{IHD,2}(A) = \sum_{i,j \in \{L,a,b\}} |\sigma_{ij}(A) - \sigma_{ij}(A_l)| \tag{3}$$

where index l refers to the subregion A_l.

Let B be a region that serves as query, then all images A with regions A_l, similar to B, should be retrieved. For this, the IHD $\nu^B_{IHD}(A)$ for the image A and the region B is computed. This IHD is then compared with each entry of the set $\{\nu^l_{IHD}\}_{l=1\cdots L}$ of the image A. The minimum distance is used to rank the images A:

$$d_{reg}(A, B) = \min_{l=0\cdots L} |\nu^l_{IHD}(A) - \nu^B_{IHD}(A)|_{db} \tag{4}$$

Because the region B has also to be compared with the whole image A, the region $l = 0$ with $\nu^0_{IHD} = 0$, is included in the minimization.

5 MULTI-SCALE PARTITION

In Eq. (4), when two images are compared, the locations of the subregions A_l have to be known. To simplify

Method	Avg. r-measure	Avg. precision
Mean & Cov.	6.4	0.42
Mean	9.2	0.31
Cov.	8.7	0.33

Table 1: *Average r-measure* and *Average precision* for the first experiment.

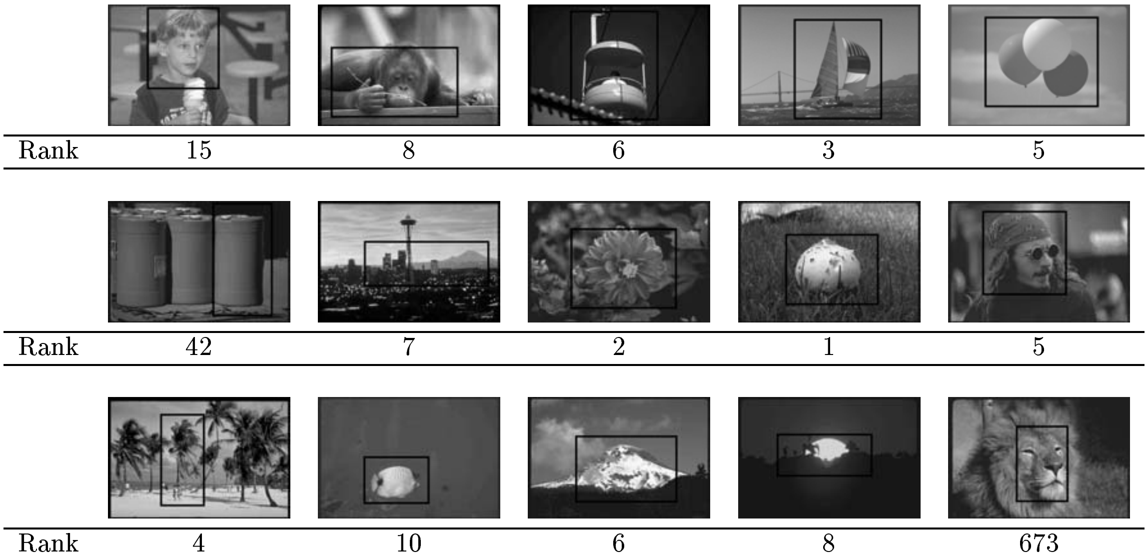

Figure 1: Variations: Rank of the original image using the combined index with two-dimensional IHD when the region indicated by the black frame serves as query

the comparison, a fixed partition for all images is chosen. The advantage is that the partition is robust and does not rely on any segmentation algorithm. In addition, the partition (1) has to cope with different scales of objects and regions in the image and (2) should not make any assumption about possible queries. To fulfill the former task, a partition with different sized rectangles is chosen. To accomplish the latter task, the regions are overlapping, ensuring that all objects of a certain scale are covered by a region of the same scale. We use a partition with three multi-scale levels. The highest level is the image itself. For the second level the image is sampled with 3x3 rectangles of half the side length of the image which yields to overlapping regions. The lowest level is composed by 5x5 rectangle regions which have a side length of one third of the image dimensions. Totally, the IHD's of 25+9=34 regions and the global color feature of the whole image are computed.

The overlapping leads to redundancy information in the IHD's, because some pixels are included in more than one region. This causes no problem since each image region is regarded as independent (Eq. (4)).

6 RESULTS

Due to space limitations we present only some of our experiments. All tests run on a database containing 12,012, 8 bit color JPEG images. They cover a wide range of nature scenes, animals, buildings, construction sites, texture and paintings.

We consider first a query set consisted of 50 queries, each with a unique correct answer. The queries represent various situations like different views of the same scene, large changes in appearance, small lighting changes, spatial translations. The images were retrieved and ranked by the distance d_{reg}. Table 1 contains the *average r-measure* and *average precision* obtained when on the same query the index was run with only one or the other vector component of the two-dimensional IHD. It shows that even one component can be used to achieve a reasonable performance, but the combination of both is more powerful.

In the second experiment, a variety of images is used (Figure 1). Each image is retrieved by the subregion indicated with the black frame. Although the regions does not correspond to the regions in the image partition, the images were retrieved. As long as the regions include several colors or saturated colors, the retrieval results are promising. The index is even able to deal with objects like fire and "sunsets", which would have been a real challenge for the methods which rely on an explicit segmentation of the image into objects. An example where the index fails completely, is the one with

Rank	2	6	50	15	156

Figure 2: Stability: Rank of the original image when the region indicated by the black frame serves as query

the face of the lion. The selected region includes colors that are very similar to the color of soil, therefore, the index retrieves many images which include regions of soil. Another problem is retrieving simply blue or green objects. In this case, all images with blue sky and green forest, respectively, are retrieved.

To illustrate the stability of the index we selected five different regions from an image with the head of an animal (Figure 2). We present the ranks of the original image when the regions indicated with the black frame serve as queries. Note that in the cases where the region is large enough, the index performs reasonable and seems to be stable. The index fails only in the cases where the details of the animal head are not characteristic enough.

In summary, we tried our method in different situations. First, we considered the situation where the index is used to retrieve similar images taken under different conditions, such as different views of the same scene, large changes in appearance, small lighting changes, spatial translations. It was proven that the best performances are obtained with the full index but, even the use of only one component of the index leads to promising results. Secondly, we considered various images that were retrieved using as queries, regions indicated by user. As long as the regions were characteristic enough, the index was able to retrieve, on average, the original images in top 10. Finally, in order to illustrate the stability of our index, we considered the situation where the user indicates different query regions from the same image. The index performed reasonable and seemed to be stable.

7 SUMMARY AND CONCLUSIONS

Not many image database management systems can handle arbitrary-size sub-image queries based on color and spatial similarity. For the systems that can deal with sub-image queries of arbitrary size, multiresolution matching is not used.

Our novel approach suggests an index which is based on a fixed multi-scale partition with overlapping rectangular regions. Each region is represented by global features and instead of storing them for all regions, only the global feature of the whole image and the IHD's of the subregions are stored. The index allows to retrieve images by relative arbitrary subregions. The proposed index uses color features. As tests have shown in a database containing more than 12,000 images, the index is suitable to retrieve images by multicolored subregions. The index is easy to be implemented, which makes it convenient to add it to existing retrieval systems which use color features.

The advantages of our approach come from the way the partition was chosen. Because different sized rectangles where chosen, the partition can cope with different scales of objects and regions in the image. On the other hand, the overlapping of the rectangles ensures that all objects of a certain scale are covered by a region of the same scale. Furthermore, out index does not rely on any segmentation algorithm.

8 ACKNOWLEDGMENTS

This research was supported with a grant from Philips in the Netherlands.

References

[1] A. Dimai. Spatial encoding using differences of global features. *Proc. SPIE, Storage and Retrieval for Image/Video Databases IV*, 1997.

[2] C. Jacobs, A. Finkelstein, and D. Salesin. Fast multiresolution image querying. *ACM SIGGRAPH*, pages 277–286, 1995.

[3] K-S. Leung and R. Ng. Multiresolution subimage similarity matching for large image databases. *Proc. SPIE, Storage and Retrieval for Image/Video Databases IV*, 1997.

[4] M. Stricker and A. Dimai. Spectral covariance and fuzzy regions for image indexing. *Machine Vision and Applications*, 10:66–73, 1997.

[5] M.A. Stricker and M. Orengo. Similarity of color images. *Proc. SPIE, Storage and Retrieval for Image/Video Databases III*, 2420:381–392, 1995.

[6] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.