# Sub-Image Search Using a Multi-Scale Approach

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

**Keywords:** sub-image search, fixed partition, global features, multi-scale

## Abstract

*Global feature approaches are commonly used methods to encode image information for retrieval purposes. However, they suffer from a complete lack of spatial information and therefore, are not suitable to be used in sub-image search. One the other hand, if an image should be retrieved by its subregions from a large image database then, 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 an 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

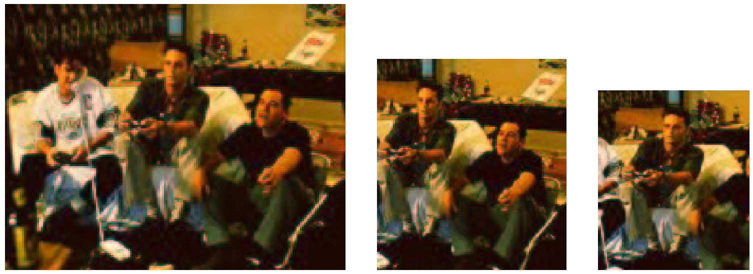
In recent years, more and more applications in digital libraries, geographical maps and medical image management, and many more, require effective and efficient means to access images based on their true content. Many of these systems uses query by example paradigm for retrieving images. This is, a sketch, an image montage or a subregion of an image serves as query which is matched against the images in the database. In this paradigm some difficulties arise due to the subjectivity of the user and the enormous number of possible queries. If for example, the image in Figure 1a should be retrieved by its subregions then, many different subregions could serve as query. One person wants to retrieve it using region 1b while another person uses region 1c, which is subregion of 1b, as query.

This simple example illustrates two important demands for an image index:

- the image is hierarchical organized: the image is a composition of objects, which are also a combination of even smaller objects. Therefore, the index must have a hierarchical structure to cope with details of the image and also, with the arrangement of the whole image.

- if the index is generated, nearly no assumptions about possible queries should be made. 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. Then, at retrieval time an immense number of possible object combinations have to be considered.

Minka and Picard [2] introduced a learning component in their system. They use positive and negative examples to learn which members of a set of image groupings (similarity measures) should be used. The image groupings occur



(a) (b) (c)

Figure 1: Different subregions (b) and (c) which could serve as queries for retrieving image (a)

within and across images and are based on color and texture cues. However, their system still requires the user to label various parts of the scene. This approach is a high-level approach and rely on user feed-back.

There are also low-level approaches. One is using the local grey-value invariants of characteristic image points, for indexing objects in images [3]. The problem is that the resulting index is relatively large and the method rely on characteristic image points extracting algorithm.

This paper suggest another low-level approach which leads to an index that regards (1) the hierarchical composition of an image, (2) needs 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 rectangles of different size. This allows to cope with the different scales of objects in an image. Each of these regions is represented by global information and instead of storing the whole global features, only the differences of the feature vectors are stored. The proposed index used color features but these ideas can be easily extended for other global features.

The remain of the paper is organized as follows. The global color feature that represent image regions is reviewed in section 2. The *inter hierarchical distance* (IHD) of global features that allows to reduce the data is introduced in section 3. The fixed hierarchical partition is given in section 4. The results and the performance of the combined index are presented in section 5. The paper conclusions are summarized in section 6.

## 2 Image features

Of the visual media retrieval methods, color indexing is one of the dominant methods because it has been shown to be effective in both academic and commercial arenas. In color indexing, histogram methods [6] are often used because they are feasible in terms of memory usage and provide sufficient accuracy. While histograms are useful because 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]. For the color features the $L*a*b*$ space is chosen because is perceptually uniform. The color features that represent the color distribution are the average color $\mu = (\mu_L, \mu_a, \mu_b)$ and the covariance matrix $[\sigma_{ij}]$ $(i, j \in \{L, a, b\})$ of the color channels. If the color components of a pixel $P$ are $P_L$, $P_a$ and $P_b$ respectively, then the index entries characterizing the color distribution of an image or an image region $A$ are:

$$\mu_i(A) = \frac{1}{N} \sum_{P \in A} P_i \qquad i, j \in \{L, a, b\} \quad (1)$$

$$\sigma_{ij}(A) = \frac{1}{N} \sum_{P \in A} (P_i - \mu_i(A))(P_j - \mu_j(A)) \quad (2)$$

where $N$ is the total number of pixels in the image region $A$. 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:

$$||\nu(A)||_{db} = \sum_{i=1}^{n} \frac{|\nu_i(A)|}{\alpha(\nu_i)} \qquad (3)$$

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

$$||\nu(A) \ - \ \nu(B)||_{db} \qquad (4)$$

# 3 Inter hierarchical distance for global features

Let $A = F(x,y)_{x,y=1\cdots N}$ be a two-dimensional image pixel array. The local feature $F$ can be the value of the n-color channels [6], the gradient, an edge map [1] or some invariant grey values [3].

The image is partitioned into $L$ subsets $A_l$, with

$$A = \bigcup_{l=1\cdots L} A_l \qquad (5)$$

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 kind of approach suffer from a complete lack of spatial information, which makes it difficult to index image regions properly. One way to overcome this is to extract the global features of image subregions $\nu(A_l)$. However, this increases the index size dramatically. 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 can now introduce a measure of the distance between the global features of the image and the features of its subregions. This distance is called *inter hierarchical distance* (IHD) because it is taken between feature vectors of different hierarchical levels of the image partition.

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

$$\nu^l_{IHD,1}(A) = \sum_{i=L,a,b} |\mu_i(A) - \mu_i(A_l)| \qquad (6)$$

$$\nu^l_{IHD,2}(A) = \sum_{i,j=L,a,b} |\sigma_{ij}(A) - \sigma_{ij}(A_l)| \qquad (7)$$

The index $l$ refers to the subregion $A_l$.

Consider now a region $B$ that serves as query, then all images $A$ with regions $A_l$, which are similar to $B$, should be retrieved. For this the IHD $\nu^B_{IHD}(A)$ for the image $A$ in the database 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 $A$. Finally, the minimum distance is used to rank the images $A$:

$$d_{reg}(A,B) = \min_{l=0\cdots L} ||\nu^l_{IHD}(A) - \nu^B_{IHD}(A)||_{db} \qquad (8)$$

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, too.

# 4 Hierarchical partition

If two images are compared (eq. (8)) then, the locations of the subregions $A_l$ have to be known. To simplify the comparison, a fixed partition for all images is chosen. An advantage of a fixed partition is that the partition is robust and does not rely on any segmentation algorithm. In addition the partition

- has to cope with different scales of objects and regions in the image
- should not make any assumption about possible queries

Query

| | | | | | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 4 | 7 | 15 |
| Mean | 2 | 3 | 5 | 10 | 24 |
| Cov. | 2 | 3 | 7 | 9 | 22 |

| | | | | | |
|---|---|---|---|---|---|
| Rank | 19 | 41 | 44 | 83 | 121 |
| Mean | 28 | 45 | 42 | 85 | 139 |
| Cov. | 24 | 50 | 49 | 82 | 145 |

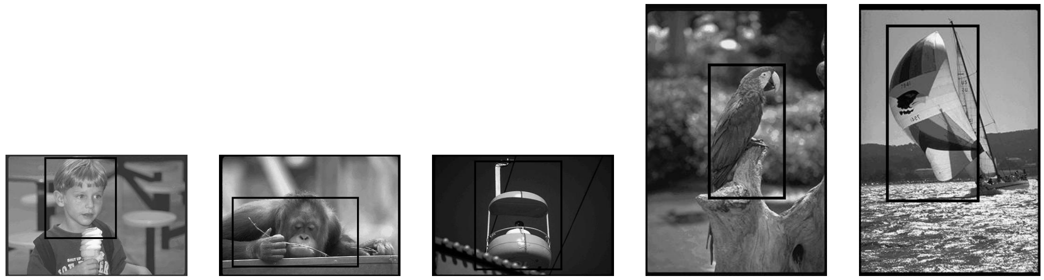
Figure 2: Query example.The selected images are the ones that are matching best the query "yellow tractor". The ranks of the images are given for the combined index using two-dimensional IHD (Rank), only the first component of the IHD (Mean) or only the second component of the IHD (Cov.)

| | | | | | |
|---|---|---|---|---|---|
| Rank | 2 | 6 | 50 | 15 | 156 |

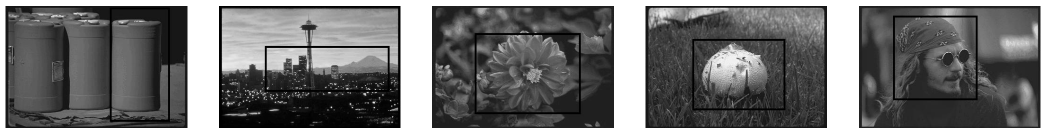
Figure 3: Stability: Rank of the original image when the region indicated by the black frame serves as query

To fulfill the former task a partition with different sized rectangles is chosen. To fulfill the latter task the regions are overlapping. The overlapping ensures that all objects of a certain scale are covered by a region of the same scale. For this work a partition with three scales, i.e. three hierarchical levels, is used. The highest level or the largest scale is the image itself. For the second level the image is sampled with rectangles of half the side length as the image. The image is sampled regularly with 3x3 rectangles 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.

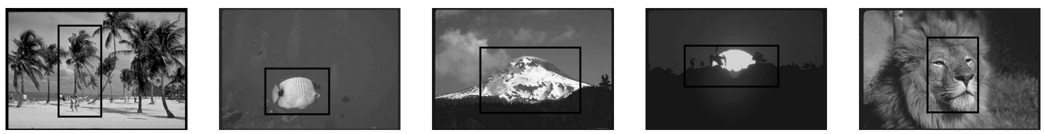
The overlapping of the rectangles leads to redundancy information in the IHD's, because some pixels are included in several regions. This redundancy is no problem since each image region is regarded as independent of the other regions in the same image (eq. (8)).



| | | | | | |
|---|---|---|---|---|---|
| Rank | 15 | 8 | 6 | 5 | 3 |



| | | | | | |
|---|---|---|---|---|---|
| Rank | 42 | 7 | 2 | 1 | 5 |



| | | | | | |
|---|---|---|---|---|---|
| Rank | 4 | 10 | 6 | 8 | 673 |

Figure 4: Variations: Rank of the original image when the region indicated by the black frame serves as query

# 5 Results

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.

The first example in figure 2 shows how the index can be used to retrieve a certain multicolored object. The images were retrieved and ranked by the distance $d_{reg}$. Only the rank and the images which shows an yellow tractor are depicted. Because the query image is not pure yellow, many yellow objects, as yellow flowers, etc., are higher ranked. 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.

To illustrate the stability of the index we selected five different regions from an image with the head of an animal (figure 3). 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 enough characteristic.

In figure 4 a variety of images is shown. Each image is retrieved by the subregion indicated by 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 given with 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.

The retrieved images are relative clear determined. In average the distance $d_{reg}$ between the first and the 50th rank increase about 100% and between the first and the 300th image about 260%.

# 6 Summary and Conclusions

If an image should be retrieved from a large database by its subregions, the index has to fulfill different criteria:

- it has to capture the different scales of the image
- it should not make many assumptions about possible queries

This paper suggest an index which is based on a fixed hierarchical partition with overlapping rectangular regions. Each region is represented by global features and instead of storing the global features of 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 proposed index is easy to be implemented, which makes it convenient to add it to existing retrieval systems which use color features.

The advantages of the fixed partition come from the way the partition it 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.

Unfortunately, since the partition do not make any assumption about the possible queries, some disadvantages can be drawn. The main one is related to the inter-scale problem. It may happened that the subimage that should be retrieved has the size that fit between two hierarchical scale levels considered. In this case the index cannot incorporate all the information from the subimage and therefore the matching is not accurate. Another problem is related to the fact that the partition is taken on a rectangular grid, so, if an user is interested in a particular object, the index will be able to retrieve only the bounding box around the object. A way to solve these problems will be to consider regions not only at a certain scale but also to provide the possibility for merging regions of different scales or splitting the larger regions into smaller regions in order to give a better answer for the user query.

# References


[1] A.K.. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, 1996.

[2] T.P. Minka and R.W. Picard. Interactive learning using a society of models. *Proc. of CVPR*, pages 447–452, 1996.

[3] C. Schmid and R. Mohr. Combining greyvalue invariants with local constraints for object recognition. *Proc. of CVPR*, pages 872–877, 1996.

[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.