On "Shapes" of Colors for Content-Based Image Retrieval

Renato O. Stehling*
Institute of Computing
State Univ. of Campinas
Brazil
renato.stehling@dcc.unicamp.br

Mario A. Nascimento†
Dept. of Computing Science
Univ. of Alberta
Canada
mn@cs.ualberta.ca

Alexandre X. Falcão‡
Institute of Computing
State Univ. of Campinas
Brazil
afalcao@dcc.unicamp.br

ABSTRACT

Color is a commonly used feature for realizing content-based image retrieval (CBIR). Towards this goal, this paper presents a new approach for CBIR which is based on well known and widely used color histograms. Contrasting to previous approaches, such as using a single color histogram for the whole image, or local color histograms for a fixed number of image cells, the one we propose (named Color Shape) uses a variable number of histograms, depending only on the actual number of colors present in the image. Our experiments using a large set of heterogeneous images and pre-defined query/answer sets show that the Color Shape approach offers good retrieval quality with relatively low space overhead, outperforming previous approaches.

Keywords

Image similarity retrieval, image databases, image metadata, histograms

1. INTRODUCTION

In general, the goal of content-based image retrieval (CBIR) is to retrieve images similar to an image/sketch provided by the user. In order to achieve this, we believe that one needs: (1) a domain-independent characterization of the visual content of the images; (2) image processing techniques to automatically extract such visual characteristics; (3) a compact yet representative abstraction for these characteristics; (4) a similarity function to effectively compare images and; (5) indexing techniques to efficiently access relevant images in the database. The first two requirements are related to image metadata extraction. The last three are related to CBIR efficiency and effectiveness. A great challenge in this area is to find the best compromise between these conflicting requirements.

*Supported by a Scholarship from FAPESP.

†Partially supported by a Research Grant from NSERC.

‡Partially supported by a Research Grant from CNPq.

ACM Multimedia Workshop Marina Del Rey CA USA

This paper deals with the first four requirements. Our goal is to propose and evaluate a new technique for image abstraction and retrieval using color distribution. Our main motivation is based on the fact that global histograms are not able to capture adequately image content and, for this reason, are not adequate for the large majority of practical applications. In addition, alternative approaches pose a large overhead in terms of metadata representation/storage. Our main contribution is a histogram-based representation which yields a lower overhead while being able to capture more information about an image's color distribution. The paper is organized as follows: Section 2 describes and compares three approaches for color-based image abstraction; Section 3 presents a similarity metric, which is common to all compared approaches, and which will be used in the experiments described in Section 4. Finally, Section 5 presents our conclusions and directions to future work.

2. COLOR-BASED APPROACHES FOR CONTENT-BASED IMAGE RETRIEVAL

This section describes three approaches for color-based image abstraction for CBIR. We focus on the first three issues highlighted in previous section.

Global Color Histogram (GCH) – The global color histogram is a simple and well known approach to encode the color information of an image for CBIR [12], [7]. A GCH is described by a set of bins (one for each color), each one with a height given by the function $p(c_k) = n_k/n$, where c_k is the k^{th} color in quantized color space, n_k is the number of pixels whose color is c_k and n is the total number of pixels in the image. The most simple quantization technique is to decompose the color space in equal size regions. For instance, the RGB (cubic) color space may be decomposed in several smaller cubes. The image color distribution is represented without any additional information about spatial location or shape of homogeneous colored regions. Images completely different from the perspective of users may have the same GCH. Hence, retrieval of images based on GCH are prone to yield a large number of false hits, specially when used with a large database of heterogeneous images.

Grid – The goal of the Grid approach is to encode more information about images when compared to GCHs. In addition to the color distribution, there is the notion of spatial location of the colors. This is achieved by decomposing an image into fixed size cells (which usually do not overlap) and, for each cell, extracting a local color histogram. Each local histogram is obtained as in GCH, but considering only each cell instead of the whole image. Thus, an image divided into 64 cells is represented by 64 local color histograms, one for each cell. This type of image decomposition may be considered a primitive type of image segmentation. The most relevant problem with the Grid approach is the space overhead related to the potentially large number of color histograms used to abstract an image. The Grid approach has been used with some variations in several CBIR approaches. The research presented in [5], [10] and [2] use the Grid approach as we describe above, by changing only the number of cells. The work in [11] decomposes images in three levels. The first one is the whole image itself. The second level is a 3x3 grid and the third level is a 5x5 grid. This decomposition results in 34 regions plus the image (level 1). The regions in this approach are of different sizes (according to their level) and overlap in different levels. The approach in [9] is similar to the previous approach. They used a quadtree of three levels to decompose the image. Another approach based in quadtrees was proposed in [8]. Finally, the approach presented in [1] decompose an image in five regions: the center and the four corners.

Color-Shape Histograms – The main contribution of this paper is an approach we call Color-Shape (CS). This approach may be considered a complement to the Grid approach, and to the best of our knowledge, it is original in the way that an image's colors are encoded. Our idea was to reduce the Grid space overhead taking advantage of the fact that only a relatively low number of colors are present in most images. Using the CS idea, colors not present in an image need not to be represented, therefore making the image abstraction more compact, yet representative. In the Grid approach, absent colors are represented by empty bins within the fixed number of local color histograms. To alleviate this overhead, we use what we call *color-shape histograms* (CSHs). Consider an image partitioned into non-overlapping cells as in the Grid approach. A CSH for a given color C is a set of bins (one for each image cell), where the bins' values are described by the function $p(cell_k) = n_k/n$. In this function, $cell_k$ is the k^{th} cell of the image, n_k is the number of pixels in the $cell_k$ with color C and n is the total number of pixels in the image. An image composed by N colors is thus described by N CSHs, each one describing the spatial distribution of one color. In this type of decomposition, if a color is not present in an image, an entire histogram needs not to be represented, nor stored. Our experiments have showed that typically the number of colors present in an image is much less than the total number of colors in the used color space. Moreover, CS approach is able to represent directly the area occupied by each color in the image, whereas the Grid approach needs to be further processed to obtain such information. Thus, CS combines the information described by Grid and GCH in an elegant way, while likely reducing the space overhead.

3. SIMILARITY METRIC

As mentioned before, the three compared approaches are based on histograms. It is then possible to use a single similarity metric for the three approaches, therefore addressing the fourth issue presented in Section 1. The chosen metric is based in L1 (Equation 1), where $h_q[i][j]$ and $h_d[i][j]$ represent the j^{th} bin of the i^{th} histogram used to represent the query image (h_q) and the database image (h_d), respectively. We assume that the histograms bins are normalized with respect to the image size, i.e., number of pixels.

$$D(h_q[i], h_d[i]) = \sum_{j=1}^{m} |h_q[i][j] - h_d[i][j]| \quad (1)$$

Recall that in the case of GCH, there is only one histogram to be considered, while for the Grid and CS approaches there are a fixed and a variable number of histograms, respectively. Furthermore, recall that a color not present in an image does not yield a CSH. Even though it does not need to be stored for the purpose of the metric computation, all of its (virtual) bins are assumed to be of zero height.

$$D_n(h_q[i], h_d[i]) = \frac{D(h_q[i], h_d[i])}{a_q[i] + a_d[i]} \quad (2)$$

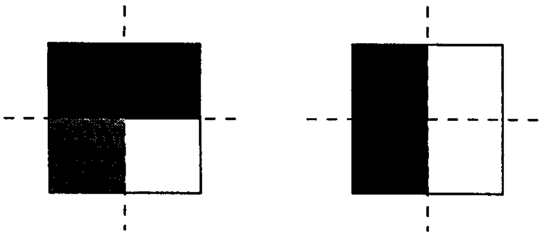
To normalize the result obtained with the L1 metric, we divide it by the sum of the areas (number of pixels) of the regions described by each histogram ($a_q[i] + a_d[i]$). These areas are also normalized with respect to the image sizes. The normalized metric is shown in Equation 2. Note that if we use GCHs, the denominator in Equation 2 is always 2, as usual and expected. However, in the CS approach, the areas being compared are, in general, smaller than the whole image, in fact, they are equal to the percentage of the color being compared in each image. This case requires one to use the sum of the areas explicitly to normalize the L1 result (notice that the sum is not known *a priori*). So far, D_n measures the distance between two histograms. The similarity between two histograms is the complement of the distance D_n. The similarity S between two images (Equation 3) is the sum of the similarity of the histograms used to the describe the images weighted by $w[i]$:

$$S(h_q, h_d) = \sum_{i=1}^{n} w[i] \times (1 - D_n(h_q[i], h_d[i]) \quad (3)$$

The goal of the weight values is to normalize the similarity between two images and describe the importance of each compared histogram. We chose $w[i] = min(a_q[i], a_d[i])$. Clearly, the sum of the $w[i]$s is at most 1.

The two images in Figure 1 will be used to exemplify the application of the similarity metric in the aforementioned approaches: q is the query image, and d is the database image to be compared against q. For simplicity, we have divided the images into 4 cells (2x2 grid) to spatially locate colors. The cells are compared from top to bottom, left to right, and the color space has only three colors: black, white and gray.

The GCH for q could be represented as $h_q = [0.5, 0.25, 0.25]$ meaning that there are 50% of black, 25% of white and 25% of gray pixels, respectively. Similarly, $h_d = [0.5, 0.5, 0.0]$.



Query Image (q) Database Image (d)

Figure 1: Images used to exemplify the application of the similarity metrics

Using Equation 3 we compute: $S_{GCH}(q,d) = 1 \times (1 - \frac{0+0.25+0.25}{2}) = 0.75$

As for the Grid approach, the normalized distance for the upper left cell is $D_n^{ul} = 0$ – both cells have only black pixels – while for the upper right, lower left and lower right corner those would be $D_n^{ur} = 1, D_n^{ll} = 1$ and $D_n^{lr} = 0$, respectively. In addition, the weights $w[i]$ are all the same, as all the cells have the same relative size. Thus, we have: $S_{Grid}(q,d) = 0.25 \times (4 - (D_n^{ul} + D_n^{ur} + D_n^{ll} + D_n^{lr})) = 0.5$

Finally, using the CS approach, we compare the three CSHs (for colors black, white and gray) using Equation 2. We have 25% of black pixels in upper left and right cell of q, and in the upper and lower left cells of d (recall that the quantity of pixels in a cell is normalized with respect to the image size), hence: $D_n^{black} = \frac{|0.25-0.25|+|0.25-0|+|0-0.25|+|0-0|}{0.5+0.5} = 0.5$. Likewise we obtain $D_n^{white} = 0.33$, and $D^{gray} = 1$.

Lastly, the normalized CSHs distances are complemented and weighted according to the Equation 3, notice that unlike in the GCH and Grid approaches, the weights $w[i]$ are now variable, depending on the areas occupied by each color. Therefore, the similarity between the images according to their CSHs is: $S_{CSH}(q,d) = 0.5 \times 0.5 + 0.25 \times 0.66 + 0 \times 0 = 0.42$

It is interesting to note that $S_{GCH} > S_{Grid} > S_{CSH}$, what is expected, given that the CSH uses more information (with not much more overhead) than the other approaches. As we will see in results obtained, this more accurate similarity measure will yields a higher retrieval quality.

4. EVALUATION OF RETRIEVAL

The dataset used was a set of 20,000 JPEG images from a stock CD by Corel Corp. The images content are heterogeneous but, at the same time, there are small subsets where the images are very similar, allowing one to distinguish them from the others without ambiguity. Some of these subsets was used to evaluate query results. The RGB color space, quantized in 64 colors, was used. For the Grid and CS approaches, each image was decomposed in 64 cells (an 8x8 grid). In the GCH approach, each image was represented by one color histogram with 64 bins, one for each of the quantized colors. In the Grid approach, each image was described with 64 local color histograms, one for each cell. In the CS approach, each image was described by a variable number of CSHs, depending on the number of colors present in the images. Each CSH histogram had 64 bins, one for each image cell.

The database creation and query times were not measured. At this point, our focus is only the quality evaluation of the image retrieval. The efficiency of the retrieval is an aspect which is tightly related to indexing structures/techniques, which is subject of further research. Out of the 20,000 images in the data set we chose 15 images to be used as query images. The answer sets for these query images were also built *a priori* in order to facilitate our evaluation. We call these sets the "required result" (RRset) for each query image. The average number of images in such RRsets was 11.27. The RRset allows to comparatively evaluate the three approaches, however, it does not allow an individual and qualitative analysis of each approach.

As better the retrieval, as higher the images in the RRset will rank in the returned answer set. Therefore, for each query image we measured the average rank (AVGR) of its RRset in the returned answer set. We also computed the similarity between the query image and the images in the RRset—note that each approach may yield a different similarity measure, and calculated the ratio of the average rank of the RRset relative to the ideal rank, $(|RRset| - 1)/2$, where $|RRset|$ denotes the cardinality of the RRset). We call this ratio the normalized AVGR (NAVGR). This evaluation technique was based on the one used for QBIC [6].

Table 1 summarizes the experiment results. The GCH and Grid approaches use a constant number of histograms (1 and 64, respectively) per database image. GCH is the most compact image abstraction. In the CS approach, the number of CSH is variable, according to the number of colors in an image. The maximum number of histograms is 64. In average, only 28.71 histograms were needed per database image (even though). Therefore, the CSHs required 55% less space than the Grid histograms. Note that this means potential savings of over 50% in space when the image abstraction is stored for further processing, e.g., indexing, which is a desirable feature of the proposed approach.

For each RRset, we determined its average similarity with the query image in each approach. The GCH results the higher average similarity. The Grids similarity is 30% smaller than GCH similarity, and the CSs similarity is 15% smaller than the Grids similarity. The high similarity values from GCH are due to a gross representation, which ultimately yields a large number of false-hits. As the database size increases and becomes more heterogeneous, the probability of completely different images (from the viewpoint of the user) have a higher similarity increases even further. Although the Grid and CS approaches may be affected in the same way, within these approaches this effect is minimized simply because there is more information to distinguish images, e.g., some spatial information. The main motivation to use GCH was to use it as a yardstick to compare the Grid and CS approaches.

The Grid approach has the higher representation overhead, measured in terms of the number of histograms. It uses more information than GCH to represent an image thus its NAVGR is considerably better than the GCH's NAVGR.

Table 1: Summary of the experimental results

	GCH	Grid	Color-Shape
Average number of histograms per image	1	64	28,71
Average RRset Similarity to Query Image	76%	53%	45%
Average Normalized Average Rank (NAVGR)	22,58	7,34	4,59

Although the CS approach uses less histograms to represent an image than the Grid, it combines the global information of GCH with the local information of Grid in an elegant way. These two types of information yield in a smaller (therefore better) NAVGR for the CS approach. Because of the amount of information used to compare two images, the RRset similarity for the CS approach is in the average, 41% smaller than GCH similarity and 15% smaller than the similarity obtained using the Grid Approach. In other words, a more strict similarity yields in less false-hits.

The NAVGR reflects the number of times a given approach is less efficient than the ideal approach (i.e., one that ranks the answer set perfectly). For example, if a user queries an image and expect to retrieve 10 similar images, an approach with $NAVGR = 5$ should return, in average, 50 images in order to present the 10 images expected by the user. As it can be seen in Table 1, the Grid approach yields a NAVGR 68% smaller than in GCH approach. More importantly, however, the NAVGR obtained by using CSHs is 38% smaller than if we were using the Grid Approach.

5. CONCLUSIONS AND FUTURE WORK

We have compared three approaches to content based image retrieval: global color histogram (GCH), Grid and Color-Shape. The Grid and Color-Shape approaches decompose an image in predefined cells to extract spatial information from the images. The comparison used only image-based queries. The GCH is the more used approach in practice. There are also, several variation of the Grid. The Color-Shape is a new interesting approach because it combines the information of Grid and GCH in an elegant way that reduces space overhead in 55% (relative to Grid approach). It is also the best approach in terms of average rank, 38% smaller than Grid and 80% smaller than GCH. Our contributions with this work were the Color-Shape approach, and the similarity metric that may be used in any histogram-based CBIR to compare two images.

The next experiments will concentrate in using Color-Shape histograms, which seem to be quite promising. We will investigate the effect of the numbers of cells and colors, the effect of different color spaces, for example, HSV and L*u*v* [3], alternative similarity metrics and alternative quantization schemes. Experiments using object-based queries and the study of possibly using automatic image segmentation techniques to decompose images in a more consistent way will also be pursued. To improve query speed, we will investigate indexing structures, e.g., the M-tree [4]. The M-tree is a promising approach because it indexes distances between images and thus does not suffer from the problems related to indexing points in high dimensional spaces, to which histograms are usually mapped to.

6. REFERENCES

[1] A.R. Appas, A.M. Darwish, A.I. El-Desouki, and S.I. Shaheen. Image indexing using composite regional color channels features. In *Proc. of SPIE - Storage and Retrievalfor Image and Video Databases VII*, volume 3656, pages 492–500, 1999.

[2] J. Ashley, R. barber, M. Flickner, J. Hafner, D. Lee, W. Niblack, and D. Petkovic. Automatic and semi-automatic methods for image annotation and retrieval in qbic. In *Proc. of SPIE - Storage and Retrievalfor Image and Video Databases III*, volume 2420, pages 24–35, 1995.

[3] A. Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann, 1999.

[4] P. Ciaccia, M. Partella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proc. of the* 23^{th} *VLDB*, pages 426–435, 1997.

[5] A. Dimai. Spatial encoding using differences of global features. In *Proc. of SPIE - Storage and Retrievalfor Image and Video Databases IV*, volume 3022, pages 352–360, 1997.

[6] C. Faloutsos, W. Equitz, M. Flickner, W. Niblack, D. Petkovic, and R. Barber. Efficient and effective querying by image content. *J. of Intelligent Information Systems*, 3(3/4):231–262, 1994.

[7] B. Funt and G. Finlayson. Color constant color indexing. *IEEE TPAMI*, 17(5):522–529, 1995.

[8] L.J. Guibas, B. Rogoff, and C. Tomasi. Fixed-window image descriptors for image retrieval. In *Proc. of SPIE - Storage and Retrievalfor Image and Video Databases III*, volume 2420, pages 352–362, 1995.

[9] J. Malki, N. Boujemaa, C. Nastar, and A. Winter. Region queries without segmentation for image retrieval by content. In *Proc. of Proc. of* 4^{th} *Intl. Conf. on Visual Information Systems*, pages 115–122, 1999.

[10] E. Di Sciascio, G. Mingolla, and M. Mongiello. Content-based image retrieval over the web using query by sketch and relevance feedback. In *Proc. of VISUAL'99*, pages 123–130, 1999.

[11] N. Sebe, M.S. Lew, and D.P Huijsmans. Multi-scale sub-image search. In *Proc. of* 7^{th} *ACM Int. Conf. on Multimedia (Part 2)*, pages 79–82, 1999.

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