Fast Image Retrieval Based on Attributes of the Human Visual System

Georgios Panitsidis[1], *Konstantinos Konstantinidis*[1], *Vasileios Vonikakis*[1], *Ioannis Andreadis*[1†] *and Antonios Gasteratos*[2]

[1]Democritus University of Thrace
Laboratory of Electronics
Department of Electrical & Computer Engineering
GR-67100, 12 Vas. Sofias
GREECE
[†]Tel. +302541079566, Fax: +302541079564
†E-mail: iandread@ee.duth.gr

[2] Democritus University of Thrace
Laboratory of Robotics and Automation
Department of Production & Management Engineering
GR-67100, 12 Vas. Sofias
GREECE
Tel. +302541079359, Fax: +302541079343
E-mail: agaster@pme.duth.gr

ABSTRACT

In this paper we present a new method for content-based image retrieval (CBIR), based on the retinal signal processing of the Human Visual System (HVS). A center-surround operator similar to the receptive fields of the ganglion cells of the retina is employed to create a new form of color histogram, the Center-Surround Histogram (CSH). Unlike other proposed color histograms, the CSH takes into consideration only the visual signal surrounding the zero-crossings of an image. This reduces the processed amount of visual information and minimizes the computational burden. Furthermore, a combination of spatial and chromatic information of the image is also achieved. The method is compared to other contemporary methods for image retrieval, exhibiting better results in shorter computational times.

1. INTRODUCTION

The blistering pace of advancement in multimedia technology, has led to vast image, video and audio databases, which are rapidly growing. These vast digital libraries produced by various applications such as medicine, military, entertainment and education belong to either private or public (World Wide Web) databases and are in large need to be indexed. This has led to an increasing interest on the research and development of automatic content-based image retrieval systems. Effective retrieval of image data is an important building block for general multimedia information management. For an image to be searchable, it has to be indexed by its content. Color can provide powerful information about the content of an image. Among the methods that use color as a retrieval feature, the most popular one is probably that of color histograms [1, 2, 3]. The histogram is a global statistical feature which describes the intensity distribution for a given image [4]. Its main advantage is that low computational cost is required for its manipulation, storage and comparison. Moreover, it is insensitive to rotation and scale of the image scene and to any displacement of objects in the image. On the other hand it is also somewhat unreliable as it is sensitive even to small changes in the context of the image. Swain and Ballard [3] proposed a simple, but at the same time cunning, method of matching images through the intersection of their color histograms. Despite its simplicity, the results and performance of this method are fairly good. Other low-level features widely used by researchers for indexing and retrieval of images, except color [1, 2, 3], are texture [5, 2, 6] and shape [5, 2, 7]. In order to exploit the strong aspects of each of these features while constructing an optimum and robust CBIR system, a plethora of methods, introduced over time, have been based on combinations of these features [7, 8, 9].

Ultimately in CBIR, what all researchers are trying to achieve is to approximate the way humans distinguish the similarity between two images, and effectively apply it on large databases. Thus, in the quest for better CBIR many human vision attributes have been repeatedly used over time, such as region based [8] and dominant color systems [10].

In this paper, we incorporate the way that the HVS manipulates visual information at its first levels, in a new 'smart' histogram. The novelty of the proposed histogram is that it reduces the processed visual information by using only the colored area surrounding the zero-crossings of an image. These areas are defined by a center-surround operator, analogous to the ganglion cells of the retina. The proposed histogram contains only the chromatic information of these areas. Hence it is defined as Center Surround Histogram (CSH). This approach is an approximation of the way that the HVS processes color areas. Although the proposed CSH includes information from the whole image, it is not global in a sense that only a selected subset of the pixels in the image is taken into account. As a result, the contribution of the proposed method is that it significantly improves execution time in comparison to other contemporary color histogram methods and reduces storage demands.

Additionally, it exhibits much more effective results in retrieval. Following the histogram creation procedure which is proposed in the next section, the Matusita distance metric is used to measure the similarity between the query image and the images stored in the database, therefore presenting the most similar images as requested by the user.

The rest of the paper is organized as follows: Section 2 briefly describes the way that the HVS processes visual information at its basic levels and perceives surfaces. This is the biological background upon which, the proposed method is based. Section 3 describes the construction of the proposed 'smart' histogram. The performance and execution time of the proposed method along with comparative results is presented in section 4 and the conclusive remarks are made in the final part of this paper.

2. VISUAL SIGNAL PROCESSING IN THE HVS

In contradiction to artificial visual systems, the retina does not act as a conventional camera; it processes the visual signal before transmitting it to the primary visual cortex through the optic nerve. The optic nerve is constructed by the axons of the ganglion cells which are the only outputs of the retina. The transmission from the retina to the primary visual cortex imposes a considerable bottle-neck problem; a huge amount of visual information must be transmitted through a channel of limited bandwidth. For this reason, the processing of visual signal in the retina aims, among others, at minimizing any redundancies. Consequently, only the changes in color or brightness (namely edges) are transmitted to the brain. In other words, the ganglion cells send only an image of edges to the brain along with the color signal that is detected around them [11]. The HVS compensates for this by using a 'filling-in' mechanism. This mechanism diffuses the visual signal, extracted by ganglion cells across the edges, in order to fill the missing information. By this approach the HVS perceives surfaces and at the same time significantly reduces the required visual information that is processed [12, 13].

The receptive fields of mammalian ganglion cells are reported to have a center-surround organization [11, 14]. This means that their receptive fields are concentrically organized into antagonistic zones. Illumination of the surround decreases the response to illumination at the center, or vice versa. This is often referred to as 'spatial antagonism' or 'lateral inhibition'.

Center-surround operators have been modeled as a Difference of Gaussians (DoG) [15]. Convolution with a center-surround operator results to the extraction of the second derivative of the image intensity. Consequently the center-surround operator highlights two distinct regions, for every intensity transition: a negative and a positive one. The zero-crossing which is located

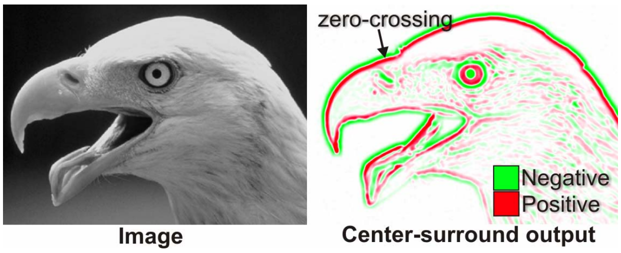

Fig. 1. Output of a center-surround operator for a real image.

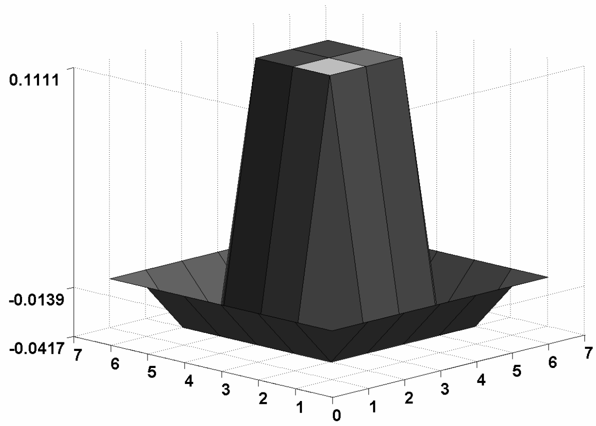

Fig. 2. The 3-dimensional representation of the mask.

between the two regions marks the true edge of the transition, c.f. Fig. 1. The visual signal located in the positive and negative regions is the one that the HVS uses in the filling-in procedure and also used in the proposed histogram construction technique.

3. HISTOGRAM CONSTRUCTION TECHNIQUE

In this method the HSV (Hue-Saturation-Value) color space is used. This color space has been defined to provide a more "intuitive" way of selecting colors, and it forms the basis for most color selection tools. In HSV, H stands for hue and contains all the chromatic information, S represents the contribution of the color white to the image pixel (Saturation), and V represents the value of the illumination of the certain pixel (Value). All shades of the various colors, even the grey levels can be expressed through the combination of these three components. The particular color space was selected for a number of reasons, but mainly due to the fact that nearly all the chromatic information is gathered in one component (Hue) providing greater efficiency. Furthermore, the Value component strictly represents the brightness of an image, which is essential for the proposed method as it renders the application of an edge detection method on the Value component more effective.

The first stage of the proposed technique is to apply a center-surround operator on the Value component of the given image. This procedure is accomplished using the center-surround mask, c.f. Fig. 2. This mask is an approximation of the center-surround operator extensively described in Section 2. As with all the operators of this kind, the sum of the center and its surround is zero, making it sensitive only to intensity transitions and not to uniform areas. Additionally, the symmetry of the operator allows for a non-directional extraction of the zero-crossings and the regions around them.

The dimensions of the particular center-surround operator were selected after extensive experimentation. The aim of this test was to define the spatial scale which is more common in every-day images and detects the significant edges in an image, while minimizing the computational burden. Center-surround operators of 3×3 size are sensitive to high spatial scales. As a result, they tend to extract many zero-crossings that are not part of significant edges e.g. textures. Similarly, sizes exceeding 15×15 tend to extract only zero-crossings of lower spatial scales, leaving many edges undetected. Additionally, when convolved with an image, they increase the computational burden to undesirable levels. The dimensions that achieve a gainful trade-off between edge extraction and computational burden were found to be the 7×7 for the surround and the 3×3 for the center. This scale is usually sensitive to all the significant edges of an image, not sensitive to high scales, such as textures and at the same time small enough to contain the computational burden of the convolution to desirable levels.

The center-surround operator is applied to the Value component of the image (ranging between 0-255). The result after the convolution is an image with values between –255 and 255. These values define the positive and negative regions around a zero-crossing (edge). The novelty of the proposed method is that these values are used as weights in the CSH generation procedure. The CSH is a histogram of 256 bins, containing visual information only from the Hue component H (ranging between 0-255) of the HSV color space. For every pixel (i,j) of an image of size $m \times n$, the output of the center-surround operator $CS(i,j)$ is used as a function to define the degree of membership of the Hue component $h(i,j)$ to the CSH. This is described by equation 1, where $\delta(\cdot)$ is the unitary impulse response.

$$hist(H) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left|\frac{CS(i,j)}{255}\right| \cdot \delta\big(h(i,j) - H\big), \qquad (1)$$

Pixels with zero values belong to a uniform area, and thus, are not taken into consideration. Pixels with absolute values near zero are located near a weak intensity transition. Consequently, they have a small participation degree in the CSH. On the other hand, pixels with absolute values near 255 are located near a strong intensity transition and for that reason they have a high participation degree in the CSH.

4. EXPERIMENTAL RESULTS

The database used for the development of our method, consists of 1040 pictures collected from the internet, digital cameras and scanners. A variety of human faces, landscapes, animals, sport events, science fiction graphics and other categories is included in the database which may easily confuse an image retrieval system. The experiments were performed using Mathworks' Matlab software on an AMD 3200+ 64-bit processor with 1 GB of RAM. All the images were scaled to 200x200-pixel size using the nearest-neighbor method in order to make the algorithms faster and to avoid later normalization of the histograms resulting in loss of color quantity information.

The last component of the image retrieval system which performs the comparison of the previously presented CSHs between the query image and the images in the database is the metric. Following many tests and simulations using the same database, it was concluded that the best similarity metric for the proposed method was the Matusita distance [16]. The Matusita Distance is a separability measure which provides a reliable criterion presumably because as a function of class separability it behaves much more like probability of correct classification. It is expressed by the equation shown below:

$$M(H_Q, H_C) = \sqrt{\sum_i \left(\sqrt{H_Q(i)} - \sqrt{H_C(i)}\right)^2}, \quad (2)$$

where H_Q is the query histogram, H_C is the histogram to be compared and *(i)* is the number of bins.

The comparison is executed bin by bin. A great distance between two bins signifies a great difference between the images for that specific color.

In order to assess the proposed method's effectiveness and efficiency, its precision and time cost performances were compared to two contemporary methods, i.e.: Swain's and Ballard's [3] and the global histogram creation method in the L*a*b* color space. Using Swain and Ballard's technique a histogram was created from the "opponent" color space (derived from RGB via linear combination of R, G and B) by subdividing the three derived components rg, by and wb into 8, 8 and 4 sections respectively, resulting in a 256 bin histogram.

In order to attain the assessment of the proposed system's precision and recall performance, a significantly larger image database was used which contained a diverse collection of about 78000 images [17]. Following numerous tests using a wide range of image sets, it is shown that the proposed method proves very efficient despite the diversity of images in such a large database.

The measure utilized to compare the proposed method against the ones mentioned above, is the retrieval performance (precision) percentage, which in this case is the proportion of relevant images retrieved (similar to the query image) in respect to the 20 first retrieved images [18].

The numerical comparison of the three methods and the significant advantage which the proposed method holds above the rest can be seen through the retrieval precision in Table 1. Swain and Ballard's method ranges from 40% to 95%, the straightforward L*a*b* histogram from 10% to 95%, and finally the proposed method spans from a low precision percentage of 55% to the highest 100% stating the clear advantage in precision of the proposed method over the rest assessed.

Table 1. Retrieval Performance (precision) Percentage for five sample image sets

	Methods		
Image Set	Swain & Ballard	L*a*b* Histogram	CSH
1	80%	90%	90%
2	65%	65%	100%
3	70%	30%	75%
4	95%	95%	90%
5	40%	10%	55%
Time (sec)	1850	32264	6286

Moreover, a noteworthy advantage of the proposed method is its execution time (6286 seconds) as exhibited in Table 1. It is only 3.4 times slower than Swain and Ballard's which requires only a simple transformation of the images to the opponent color space, but 5.1 times faster than the conventional L*a*b* histogram method. This is due to the fact that it does not examine the entire image but only part of it.

In order to further appraise the robustness of the compared systems, salt & pepper noise of density 0.15 was added to the query images. Moreover, several filters were applied to the images of the data sets, such as a blurring, rendering, sharpening, sketching and even a cropping filter. Last, the query images were rotated by 90^{o} counterclockwise. The outcome of these tests decreased the precision percentages of the proposed method insignificantly, proving that although the images were severely altered, the results were very satisfying and the images were successfully retrieved.

5. CONCLUSIONS

This paper proposes a solution to the problem of image retrieval in large databases where the need of fast and accurate indexing has become a necessity. A new center-surround histogram construction method has been proposed, motivated by the way that the HVS manipulates visual information in its first stages. The CSH is created in the HSV color space and only the utmost important color information which is located in the neighborhood of the edges is considered. This reduces the total execution time of the retrieval method since only a portion of the full amount of the pixels in the image is regarded. The results of the proposed method were more accurate and the execution time was shorter than those of other popular and contemporary methods, proving that the proposed system poses an advancement in the area of content based image retrieval.

REFERENCES

[1] K. Konstantinidis, I. Andreadis, "Performance and Computational Burden of Histogram based Color Image Retrieval Techniques", *Journal of Computational Methods in Sciences and Engineering*, vol. 5, pp. 1-7, 2005.

[2] A. Del Bimbo, Visual Information Retrieval. San Francisco, California: Morgan Kaufman Publishers Inc., 1999.

[3] M. J. Swain and D. H. Ballard, "Color Indexing", *Int. Journal of Computer Vision*, vol.7, pp. 11-32, 1991.

[4] R. C. Gonzalez and R. E. Woods, Digital Image Processing. Reading, Massachusetts: Addison-Wesley, 1992.

[5] V. Castelli and L. D. Bergman, Image Databases: Search and Retrieval of Digital Imagery. John Wiley & Sons, Inc., 2001.

[6] A. Gasteratos, P. Zafeiridis and I. Andreadis, "An Intelligent System for Aerial Image Retrieval and Classification", In: *Lecture Notes in Artificial Intelligence*. G. A. Vouros and T. Panayiotopoulos, Eds. Berlin Heidelberg: Springer-Verlag, 2004. pp. 63-71.

[7] A. K. Jain and A. Vailaya, "Image Retrieval Using Color and Shape", *Pattern Recognition*, vol. 29, pp. 1233-1244, 1996.

[8] Y. C. Cheng and S. Y. Chen "Image classification using color, texture and regions", *Image and Vision Computing*, vol. 21, pp. 759-776, 2003.

[9] G. Pass and R. Zabih, "Comparing Images Using Joint Histograms", *Multimedia Systems*, vol. 7(3), pp. 234-240, 1999.

[10] Y. Liang, H. Zhai and P. Chavel, "Fuzzy Color Image Retrieval", *Optics Communications*, vol. 212, pp. 247-250, 2002.

[11] R. W. Rodieck, The first steps in seeing. Sinauer Associates, 1998.

[12] S. Grossberg, "Visual Boundaries and Surfaces", *The Visual Neurosciences*, vol. 2, pp. 1624-1639, MIT Press, 2004.

[13] S. Grossberg, and S. Hong, "A neural model of surface perception: Lightness, anchoring, and filling-in", *Spatial Vision*, 2005, in press.

[14] R. W. Rodieck and J. Stone, "Analysis of receptive fields of cat retinal ganglion cells", *Journal of Neurophysiology*, vol. 28, pp. 833-849, 1965.

[15] R. W. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli", *Vision Research*, vol. 5, pp. 583-601, 1965.

[16] K. Fukunaga, Introduction to Statistical Pattern Recognition. Academic Press Inc., 1990.

[17] B. C. Russell, A .Torralba, K. P. Murphy, W. T. Freeman, "LabelMe: a database and web-based tool for image annotation", MIT AI Lab Memo AIM-2005-025, September 2005.

[18] H. Muller, W. Muller, D. McG. Squire, S. Marchand-Maillet, "Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals", *Pattern Recognition Letters*, vol. 22(5), pp. 593-601, 2001.