Kỷ yếu Hội nghị Khoa học Quốc gia lần thứ IX "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR'9)"; Cần Thơ, ngày 4-5/8/2016
DOI: 10.15625/vap.2016.0003

A NOVEL VISUAL CONTENT DESCRIPTOR USING COLOR AND SHAPE FEATURES

Thanh Le Hoang
Nha Trang University
thanhlh@ntu.edu.vn

ABSTRACT— *Images contain information in a very dense and complex form, which a human eye, after years of training, can extract and understand. The main goal is to extract form an image a set of composing objects or real life attributes. This paper presents a novel framework for combining color and shape information, to achieve higher retrieval efficiency. The combination of the color and shape features provide a robust feature set for image retrieval. Two proposed methods – PGLAC (Pyramidal gradient local auto-correlations) and GCDD (Global color distribution descriptor) will be used as visual descriptors extracting spatial information of shape in the image. The experimental results demonstrate the efficacy of the proposed methods.*

Keywords— *Content-based image retrieval, visual descriptors, Gradient Local Auto-Correlations.*

I. INTRODUCTION

A huge amount of image databases are added every minute and so is the need for effective and efficient image retrieval systems. Applications like medicine, entertainment, education, manufacturing, etc. make use of vast amount of visual data in the form of images. This envisages the need for fast and effective retrieval mechanisms in an efficient manner. A major approach directed towards achieving this goal is the use of low-level visual features of the image data to segment, index and retrieve relevant images from the image database. Recent content-based image retrieval (CBIR) systems based on features like color, shape, texture, spatial layout, object motion,…are cited in [1][2]. Of all the visual features, color and shape are the most dominant and distinguishing one in almost all applications. Though there are many techniques of search this paper will focus on color and shape features for CBIR.

CBIR system is a useful tool to resolve the above mentioned problem. A typical CBIR system performs two major tasks, the first one is the feature extraction, where a set of features is extracted to describe the content of each image in the database; and the second task is the similarity measurement between the query image and each image in the database, using the extracted features [3]. Generally the CBIR is performed using some low-level visual descriptors such as color-based, texture-based and shape-based descriptors, which extract feature vectors from the images.

There are many methods which combine more than one visual descriptor [, improving the image retrieval effectiveness. In [2], authors combine Linear Block Algorithm (LBA) for the global color feature extraction, Steerable Filter for the texture features extraction and Pseudo-Zernike Moments for extraction of the shape features which are rotation invariant. Authors of [3] combine color and texture features, in which the color features are extracted using the Color Layout Descriptor (CLD) and the texture features are obtained using the Gabor Filters. Another method that combines more than one visual descriptors is proposed in [6], in which the image is divided into six blocks then, the color space of each block is converted from RGB to HSV and a cumulative histogram is computed in order to obtain the color features, whereas to obtain the texture features of each block, four statistic features, such as energy, contrast, entropy and inverse difference from the Gray-Level Concurrence Matrix (GLCM), are computed.

II. PGLAC FEATURE EXTRACTION

A. Definition of GLAC

Gradient Local Auto-Correlations (GLAC) is proposed by [5]. Let I be an image region and $r = (x, y)^t$ be a position vector in I. The image gradient $(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})^t$ at each pixel can be rewritten in terms of the magnitude $n = \sqrt{\frac{\partial I^2}{\partial x} + \frac{\partial I^2}{\partial y}}$ and the orientation angle $\theta = arctan(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})$. As shown in the Figure 1(a), the orientation θ is coded into D orientation bins by voting weights to the nearest bins, and is described as a sparse vector $f \in R^D$, called the gradient orientation vector (in short, G-O vector).

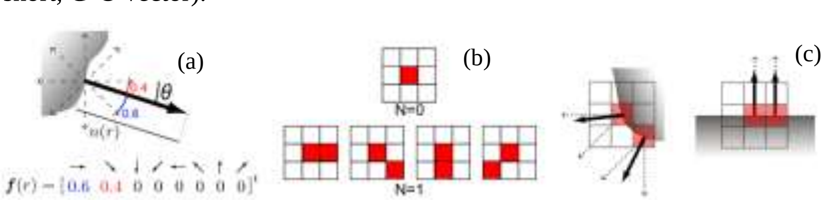

Figure 1. Image gradients are described by the G-O vectors, together with the gradient magnitudes (a). Then, by applying mask patterns (b), auto-correlations of G-O vectors are calculated, weighted by the gradient magnitudes (c) [5].

It is important that the image gradients are represented in terms of such quantized and sparse descriptors. By using the G-O vector f and the gradient magnitude *n*, the N^{th} order autocorrelation function of gradients in local neighbors is defined as follows:

$$R(d_0, \dots, d_N, a_1, \dots, a_n) = \int_I w[n(r), n(r + a_1), \dots, n(r + a_N)]\, f_{d_0}(r) f_{d_1}(r + a_1) \dots f_{d_N}(r + a_N) dr$$

Where a_i are displacement vectors from the reference point *r*, f_d is the *d-th* element of *f* and *w* is a (scalar) weighting function, e.g., min. Displacement vectors are limited to local neighbors because local gradients are supposed to be highly correlated [5].

The practical formulation of GLAC is given by:

$$0^{st} order: R_{N=0}(d_0) = \sum_{r \in I} n(r) f_{d_0}(r)$$

$$1^{th} order: \ R_{N=1}(d_0, d_1, a_1) = \sum_{r \in I} min[n(r), n(r + a_1)] f_{d_0}(r) f_{d_1}(r + a_1)$$

GLAC has high dimensionality ($D + 4D^2$), the computational cost is not large due to the sparseness of *f*. Moreover, the computational cost of GLAC is invariant with respect to the number of orientation bins, *D*, since the sparseness of *f* is invariant with respect to *D*.

B. PGLAC feature extraction

In this paper, the Pyramidal Gradient Local Auto-Correlations (PGLAC) is proposed as a spatial shape descriptor which represents the spatial distribution of edges and it is formulated as a vector representation. The operation of PGLAC consists of the following four steps:

1. Edge contour extraction: The contour of input image can be extracted using the Canny edge detector.
2. Cell division: The edge detected binary image is divided into cells at several pyramid levels. For example, in the first pyramid level, the edge image is divided into *2×2* cells and in the second pyramid level, each cell furthermore is divided into *2×2* sub-cells. The cell division is repeated until desirable resolution levels of pyramid.
3. GLAC calculation: The GLAC of each cell is calculated at each pyramid resolution level. The GLAC of each cell in the same pyramid level is concatenated to form a vector.
4. PGLAC extraction: The final PGLAC is a concatenation of all GLAC feature vectors generates in all pyramid levels.

III. GCDD FEATURE EXTRACTION

The Global Color Distribution Descriptor (GCDD) is designed to localize the visual contents based on color feature, in such a way that relative position as well as size of each object is considered as criteria for extraction.

It can be seen from the Figure 2, the pre-processing includes four different stages. Firstly, original images are partitioned into multiple non-overlapping regions by using image segmentation algorithms. Based on segmented images, the numbers of different color levels – which can be considered as components – as well as the average RGB value of each component in original images are returned as the output information of color extraction. Then, the regions with similar color feature of training images will be grouped into *k* different clusters by using *k*-means or random clustering. From the original idea of locating objects and distinguishing them based on size, images will be divided into four grid cells.

Now let us define what a GCDD feature vector is. According to our proposed model, the v_{ij} component of the feature vector denotes the size of *i-th* cluster occurring at the *j-th* grid cell. As a result, the total number of components in the vector is *4m*, where *m* is the total number of clusters obtained by *k*-mean clustering algorithm.

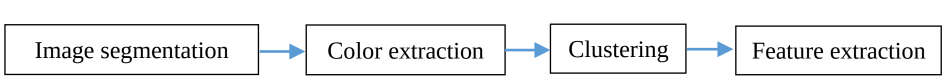

Figure 2. The pre-processing for GCDD feature extraction

IV. WEIGHTED COMBINATION OF FEATURES

Given a query image, a weighted combination of features is proposed to retrieve a set of relevant images *S* which satisfy the following formula:

$$sim(q, I_i) = \alpha.d_1 + (1-\alpha).d_2 \geq \lambda$$

Where d_1 and d_2 are the distance between the query q and image I_i in the dataset obtained by GCDD and PGLAC feature extraction respectively. α denotes the weight of a particular visual feature and λ is a threshold value predefined by user. In this paper, the optimal value of α is 0.3, which was obtained by running the experiments several times. Based on the set S including k images, we will determine the category of a query image as follows:

$$y(q) = argmax_k \sum_{I_j \in S} sim(q, I_j).y(I_j, c_k)$$

Where I_j is one of the neighbors of the query q in the training set, $y(I_j, c_k) \in \{0,1\}$ indicates whether I_j belongs to category c_k. This equation means the category with maximal sum of similarity will be the winner.

V. EXPERIMENTAL RESULTS

The training dataset used for our experiments consist of 2359 images which are chosen from the Corel6k benchmark database. The images can be classified into 30 different topics, thus each category includes about 80 images. To evaluate the retrieval performance of our models, a set of 548 images picked out randomly from the dataset will be considered as query images. The results obtained using shape and color-based for different categories is shown in the Table 1. Some example of retrieval result images with query image are shown in Figure 5a-b.

Figure 3. Sample image database: African antelope (a), Beautiful rose (b)

It can be observed from the Figure 4 that the performance obtained by using PGLAC feature extraction is much better than using GCDD descriptor. The figure also shows that the best results comes from the combination model, in which both shape and color features are considered by weighted linear combination. As mentioned in the Section 4, we set a greater shape-weight because the shape feature provides the most distinguishable information compared with the color distribution.

Table 1. The results of the combination model across different categories

Category	Accuracy	Category	Accuracy	Category	Accuracy
Antelope	0.925	Castles	0.900	Insects	0.975
Crystals	0.800	Caverns	0.725	Kitchens	0.933
Apes	0.975	Clouds	0.950	Lighthouses	0.847
Beads	0.965	Coastal	0.981	Lions	0.980
Bears	0.810	Contemporary	0.833	Marine life	0.910
Beverages	0.900	Crystallography	0.915	Models	0.825
Bonsai	0.933	Doors	0.900	Rose	1.000
Bridges	0.875	Fine dining	0.750	Rhinos	0.975
Canadian	0.900	Flora	0.975	Rome	0.810
Cards	0.836	Forests	0.950	Sunsets	0.900

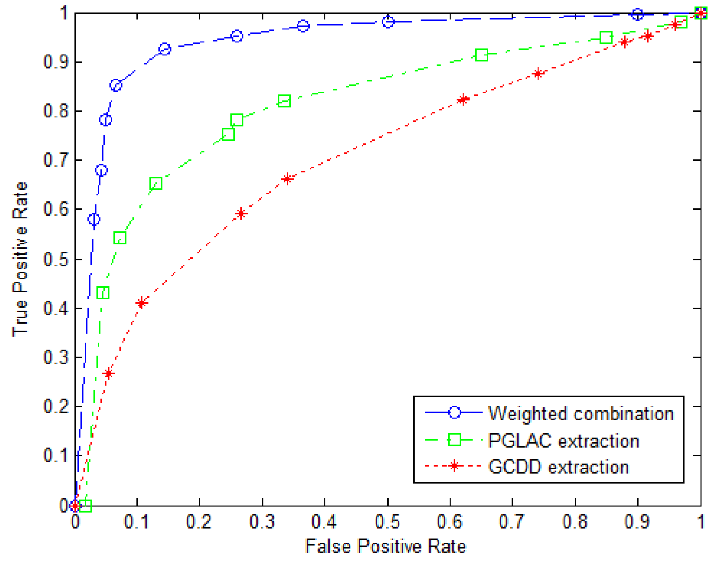

Figure 4. Comparing the performance of the proposed methods

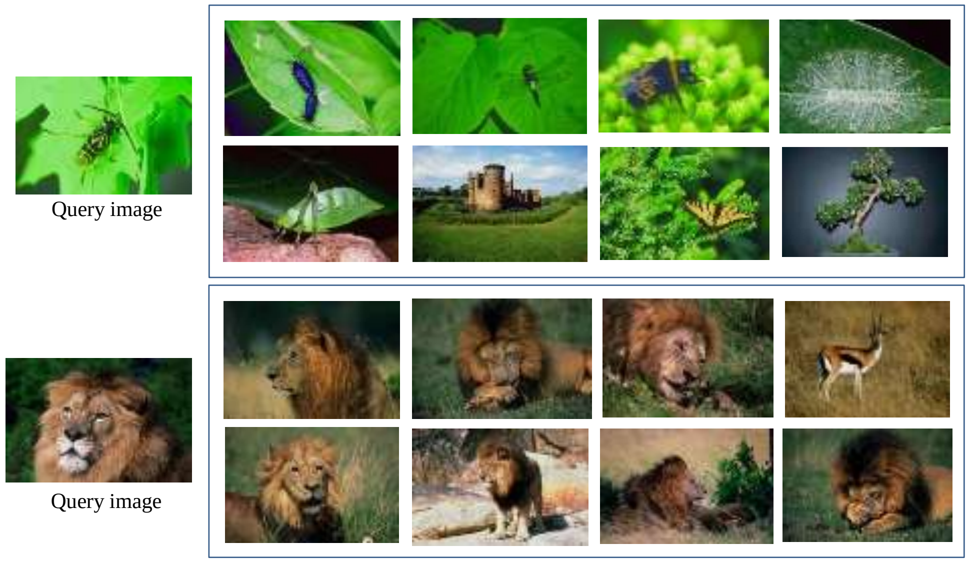

Figure 5. Results of retrieval for "insects" and "lions" image

VI. CONCLUSION

Context-based image retrieval with visual information is still a challenging problem which has not been solved satisfactorily by existing work. In this paper, we proposed a scheme which combines color and shape features for the content-based image retrieval task. The proposed scheme extracts both global and local color information using the GCDD and the shape feature based on boundary information of objects using the PGLAC. As can be seen from the experimental results, there are some gaps in our proposed models need to be improved for achieving a better model of visual information inference. Nevertheless, the experimental results also showed that our model is a very promising method for image retrieval. Especially, the experience gained from this study brings us to some valuable ideas for future work.

REFERENCES

[1] Penatti O., Valle E., Torres R., "Comparative study of global color and texture descriptors for web image retrieval", Journal of Visual Communication and Image Representation, vol. 23, issue 2, p.359-380, 2012

[2] Wang Y., Yu J., Yang Y., "An effective image retrieval scheme using color, texture and shape features", Computer Standards & Interfaces, vol. 33, issue 1, p.59-68, 2011

[3] Jalab H.A., "Image retrieval system based on color layout descriptor and global filters", in Proceeding of IEEE Conference on Open System, p. 32-36, 2011

[4] Pujari J., Hiremath P., "Content-based image retrieval based on color, texture and shape features", in Proceeding of International Conference on Advanced Computing and Communications, p.239-242, 2010

[5] T. Kobayasi, N. Otsu, "Image feature extraction Using Gradient Local Auto-correlations", in Proceeding of European Conference on Computer Vision 2008, p. 346-358, 2008

[6] Hafiane A., Zavidovique, "Local relational string and mutual matching for image retrieval", Information Processing & Manegement, vol. 44, issue 33, p.1201-1212, 2008

[7] Kavitha C., Prabhakara B., Govardhan, "Image retrieval based on color and texture features of the image subblocks", International Journal of Computer Applications, vol. 15, p.33-37, 2011

[8] Fierro A., Perez K., Nakano M., Benois, "Dominant color correlogram descriptor for content-based image retrieval", in Proceeding of the 6th International Conference on Graphic and Image Processing, 2014

[9] Yang B., Guo L., Jin L., Huang Q., "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition", in Proceeding of 16th IEEE International Conference on Image Processing, pp. 3305-3308, 2009

HƯỚNG TIẾP CẬN MỚI TRONG VIỆC MÔ TẢ NỘI DUNG HÌNH ẢNH BẰNG ĐẶC TRƯNG MÀU SẮC VÀ HÌNH DẠNG

Lê Hoàng Thanh

TÓM TẮT— *Truy vấn ảnh dựa trên nội dung là hệ thống truy vấn ảnh dựa trên việc tự động rút trích một số thông tin đặc trưng trong ảnh như: màu sắc, kết cấu, vị trí, hình dạng,... Vấn đề truy vấn ảnh trong cơ sở dữ liệu hiện đang rất được quan tâm và có nhiều phương pháp giải quyết khác nhau. Bài báo này giới thiệu một hướng tiếp cận mới trong việc kết hợp các thông tin màu sắc và hình dạng để nâng cao hiệu năng tìm kiếm. Hai phương pháp được đề xuất là PGLAC (Pyramidal gradient local auto-correlations) và GCDD (Global color distribution descriptor) được sử dụng để rút trích thông tin không gian của hình dáng trong ảnh. Kết quả thực nghiệm cho thấy tính hiệu quả của các phương pháp đề xuất.*