Content-Based Image Retrieval Based On Visual Attention And The Conditional Probability

Guang-Hai Liu
College of Computer Science and Information Technology, Guangxi Normal University, China
Corresponding author: liuguanghai009@163.com

Abstract— a novel content-based image retrieval framework was presented in this paper. This framework is used to encode primary visual feature and saliency information as natural image features by simulating visual attention mechanism and using the conditional probability. In this framework, the color volume is used as a novel feature to detect saliency areas. Besides, a novel generalized visual feature representation method, namely the conditional probability histogram, is proposed to describe natural image features. It can integrate primary visual features and saliency information into one whole unit. Experimental results indicate that the proposed algorithm outperform our prior works, namely multi-texton histogram and color difference histogram.

Keywords-image retrieval; HSV color space; visual attention; the conditional probability

I. INTRODUCTION

Image retrieval has become a very extensively investigated topic in the field of information retrieval, and it can be classed into two groups: content-based image retrieval (CBIR) and objects-based image retrieval according to the concept of similar images. Generally speaking, CBIR often adopt global feature to describe image content, whereas local features are usually used to describe image contents in objects-based image retrieval. In order to extract global features, several color, texture and shape features are recommended in MPEG-7 standard in many years ago [1] [2]. With development of multimedia technology, how to extract features from the vast amount of image data is a challenging problem. Fortunately, the visual attention mechanisms can help Humans to quickly recognize the conspicuity object in an image or a scene, and this mechanism can made human pay more attention to the conspicuity areas and to discard the unimportant areas.

Since Treisman proposed the feature integration theory of attention in 1980 [3], several computational models of visual attention have been suggested over the past years. Indeed, the feature integration theory of attention is the base of majorities of computational visual attention models [3-13], whereas they have not been investigated within content-based image retrieval (CBIR) framework. Moreover, how to construct visual attention model is still an open problem in the field of compute vision.

To address this problem, we propose a novel content-based image retrieval framework which using color volume for saliency areas detection and then the conditional probability is adopted to distinguish the grade of saliency areas, finally, histogram-based method is used as image representation for CBIR.

II. THE PROPOSED CBIR FRAMEWORK

In image content analysis and understanding, color, intensity and orientation play an important role in describe image content, which are commonly used in many computational models of visual attention [3-13],. It is well known that visual perceptual difference between two colors in uniform color space will be related to a measure of Euclidean distance [14]. Furthermore, saliency information plays an important role in pre-attention of visual perception and can provide important information for further processing.

A. Extraction of the primary visual features

Most of the existing saliency model frameworks often use a combination of the primary visual features to generate saliency maps and describe image content [4]. In many color spaces, HSV color space could mimic human color perception well, thus it is adopted to extract the primary visual features. In order to construct histogram representation, such as color-based , orientation-based and intensity–based histogram, H, S and V color channels are uniformly quantized into 6, 3 and 3 bins, respectively, so that it results in total 6×3×3=54 color combinations. Let $M_C(x,y)$ be the color combination or color map, as $M_C(x,y) = w, w \in \{0,1,..N_C - 1\}$, where $N_C = 54$.

Intensity information comes from the Value component $V(x,y)$. After uniform quantization, we can obtain the intensity map $M_I(x,y)$, as $M_I(x,y) = s,\ s \in \{0,1,\dots,N_I - 1\}$, where $N_I = 16$.

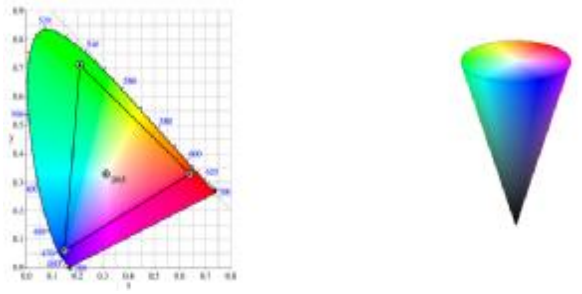
$V(x,y)$ is used to detect edge orientation $O(x,y)$ by using Sobel operation. After uniform quantization, we can obtain the edge orientation map $M_O = v, v \in \{0,1,..,N_O - 1\}$, where $N_O = 36$.

$M_C(x,y)$, $M_I(x,y)$ and $O(x,y)$ are used to construct histogram representation in the stage of describing image contents.

B. The saliency model

According to Treisman's feature integration theory, visual attention can be categorized as pre-attentive stage and focused attention stage [3]. The features are used before integration stages can be considered as the primary visual features. Thus, we need to define the visual features to detect saliency areas. The shape of CIE chromaticity diagram looks like a horseshoe [14], which is shown in fig 1

(a), so it is very difficult to calculate the color volume of Lab color space. It is one of the reasons why HSV color space is used to compute color volume. The shape of HSV color space can be interpreted as cylinder coordinate system which is shown in fig 1 (b).



(a) (b)

Figure.1. the shape of CIE chromaticity diagram (a) and the shape of HSV color space (b).

It is well known that the cylinder volume cv can be defined as $cv = \pi \cdot r^2 \cdot h'$, where r denotes the radius of cylinder, and h' denotes the height of cylinder. Let there is a random dot (h, s, v) in cylinder coordinate system, the cylinder volume derived from this dot can be defined as:

$$cv_1(x,y) = \pi \times s(x,y)^2 \times v(x,y) \times \frac{h(x,y)}{360}$$

Where $s(x,y) \in [0,1]$, $v(x,y) \in [0,1]$ and $h(x,y) \in [0,360]$. When HSV color space to be transformed into Cartesian coordinate system, the color volume derived from this dot (h, s, v) can be defined as:

$$cv_2(x,y) = s(x,y) \times \cos\big(h(x,y)\big) \times s(x,y) \times \sin\big(h(x,y)\big) \times v(x,y)$$

cv_1 and cv_2 are used to create Gaussian mid $cv_1(\sigma)$ and $cv_2(\sigma)$, where $\sigma \in [0 \ldots 4]$ is the scale. The standard deviation of Gaussian kernel is $\delta = 5.0$ in Gaussian filters construction. Deriving from Itti saliency model [5], center-surround receptive fields are simulated by across scale subtraction " $\ominus$ " between two maps at the center (c) scale and surround (s) scale, and yield the so-called feature maps:

$$F(c, s, cv_1) = |cv_1(c) \ominus cv_2(s)|$$

$$F(c, s, cv_2) = |cv_2(c) \ominus cv_2(s)| \quad (4)$$

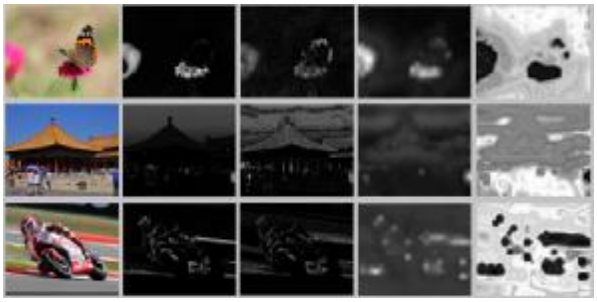
After center-surround operation, we can obtain 6 feature maps, and then they are combined into a conspicuity maps $\overline{cv_1}$ and $\overline{cv_2}$ at the scale $(\sigma = 4)$. They are obtained through across-scale addition" $\oplus$ ", which consists of reduction of each map to scale $(\sigma = 4)$ and point-by-point addition. It is similar to the manner of Itti's saliency model [5].

$$\overline{cv_1} = \sum \begin{Bmatrix} 2 & 4 \\ \oplus & \oplus \\ c = 0 & s = 3 \end{Bmatrix} \mathcal{N}(F(c,s,cv_1))$$

$$\overline{cv_2} = \sum \begin{Bmatrix} 2 & 4 \\ \oplus & \oplus \\ c = 0 & s = 3 \end{Bmatrix} \mathcal{N}(F(c,s,cv_2))$$

$$\overline{cv} = \frac{1}{2}(\overline{cv_1} + \overline{cv_2})$$

As with Itti's model, $\mathcal{N}(.)$ denotes the normalization operator [5] and projects all values into the range [0, 1]. At last, $\overline{cv}$ would be resized until it has the same size as the original image by using the bilinear interpolation manner.



(a) (b) (c) (d) (e)

Figure.2. the illustration examples of saliency areas detection. (a) Original image, (b) color volume cv_1, (c) color volume cv_2 (d) saliency map and (e) the conditional probability map.

At last, the overall saliency map $\overline{cv}$ would be resized until it has the same size as the original image by using the bilinear interpolation manner, then the finally saliency map denoted as $f(x,y)$. In order to construct histogram representation, the values of $\overline{cv}$ are projected into the range of $[0,255]$.

C. *The conditional probability*

In probability theory, a conditional probability measures the probability of an event given that another event has occurred [15]. If the event of interest is A and the event B is known or assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as $P(A|B)$.

We denote the values of saliency map $f(x,y)$ as $f(x,y) = w, w \in \{0,1,\ldots,L-1\}$. Let (x, y) and $(x + \Delta x, y + \Delta y)$ be two neighboring pixel locations in image, where Δx and Δy are the offsets of x coordinate axis and y coordinate axis, respectively. Denote by $f(x,y) = A$ and $f(x + \Delta x, y + \Delta y) = B$, then the conditional probability of A given B can be defined as follow:

$$P(A|B) = \frac{P(AB)}{P(B)}$$

In order to easily describe the image content, the case of the conditional probability $P(A|B)$ where $A = B$, is adopted in our algorithm. Indeed, our algorithm focus on "the conditional probability of A given B, where $A = B$", or "the probability of A under the condition B, where $A = B$". As can be seen from figure 2(e), the maps of the conditional probability $P(A|B)$ can distinguish the grade of saliency areas.

D. Image representation

Inspired by the prior work, namely micro-structure descriptor [21], the prior probability is adapted to image representation. The basic idea of our image representation is to generate three histograms considering intensity map, orientation map and color map of the original image via a very special type by using the conditional probability of saliency areas information. The conditional probability histogram of a full color image is defined as follows:

$$H = conca\{H_C, H_\theta, H_I\} \tag{9}$$

Where

$$H_C\big(C(x,y)\big) = \begin{cases} \sum\sum |P(A|B)| \\ where\ \ C(x,y) = C(x+\Delta x, y+\Delta y) \end{cases} \tag{10}$$

$$H_\theta\big(\theta(x,y)\big) = \begin{cases} \sum\sum |P(A|B)| \\ where\ \theta(x,y) = \theta(x+\Delta x, y+\Delta y) \end{cases} \tag{11}$$

$$H_I\big(I(x,y)\big) = \begin{cases} \sum\sum |P(A|B))| \\ where\ \ I(x,y) = I(x+\Delta x, y+\Delta y) \end{cases} \tag{12}$$

Where conca{.} denotes the concatenation of H_C , H_θ and H_I, where H_C, H_θ and H_I denotes the conditional probability histogram within the saliency areas using color, edge orientation and intensity as constraints respectively. The total dimension of conditional probability histogram is 54+36+16=106.

Using the above method to processing the intensity map, orientation map and color map, respectively, we can obtain the conditional probability histogram of them and denote as H_C, H_θ and H_I. In this manner, the primary visual features (e.g. intensity, color and orientation) and saliency information are integrated into one whole unit.

III. The Experiments and Results of CBIR

In order to demonstrate the performance of the proposed algorithm, Corel-10K dataset and GHIM-10K dataset are used in our experiments. On Corel-10K dataset, every category contains 100 images of size 192×128 or 128×192 in JPEG format. All Corel images come from Corel Gallery Magic 20, 0000. On GHIM-10K dataset, every category contains 500 images of size 400×300 or 300×400 in JPEG format, and all images are collected by the author (Guang-Hai Liu) from web.

Besides, two other existing CBIR algorithms including multi-texton histogram (Liu et al. 2010) and color difference histogram (Liu et al. 2013) are used for comparisons. Both of them are originally developed for CBIR by the author.

In the experiments, we have randomly selected 1000 images and 1000 images from Corel-10K and GHIM-10K dataset as query images, respectively. Canberra distance metric (Lance et al. 1967) is used to measure image similarity.

A. Performance metrics

In the field of image retrieval, two terms are the most common measurements used for evaluating the retrieval performance, namely $Precision$ and $Recall$ [18]. In our prior works, they are also used for evaluating the performance of CBIR [16-17] [20-22]. They are defined as follows:

$$P(N) = {}^{I_N}/_{N} \tag{13}$$

$$R(N) = {}^{I_N}/_{M} \tag{14}$$

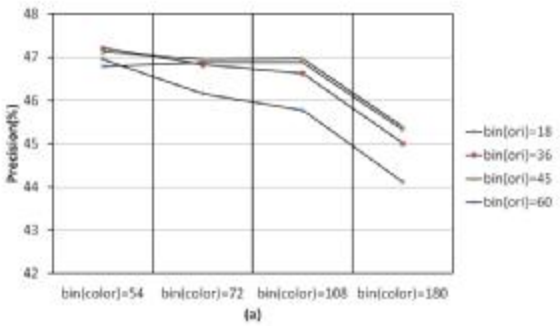
Where I_N ,N and M are the three very important parameters. I_N is the number of images retrieved in the top N positions that are similar to the query image, M is the total number of images in the database similar to the query, and N is the total number of images retrieved.

In our image retrieval system, $N = 12$ and $M = 100$ on Corel-10K dataset, whereas $N = 12$ and $M = 500$ on GHIM-10K dataset. The final performances are evaluated by the average results of all queries.

B. Determining the quantization levels of visual features

The proposed algorithm is histogram-based method, how to determine the quantization levels of visual features is an important problem. Our algorithm is adopted HSV color space for color quantization because it could mimic human color perception well. Different uniform quantization levels of color, intensity and orientation are used to test the performance of the proposed algorithm.

There may be too many combinations about the quantization levels of color, intensity and orientation. In order to confirm the best quantization levels, the quantization levels of color from 54 to 180 bins, the quantization levels of intensity is from 16 to 64 bins, the quantization levels of orientation is from 18 to 60 bins.



(a)

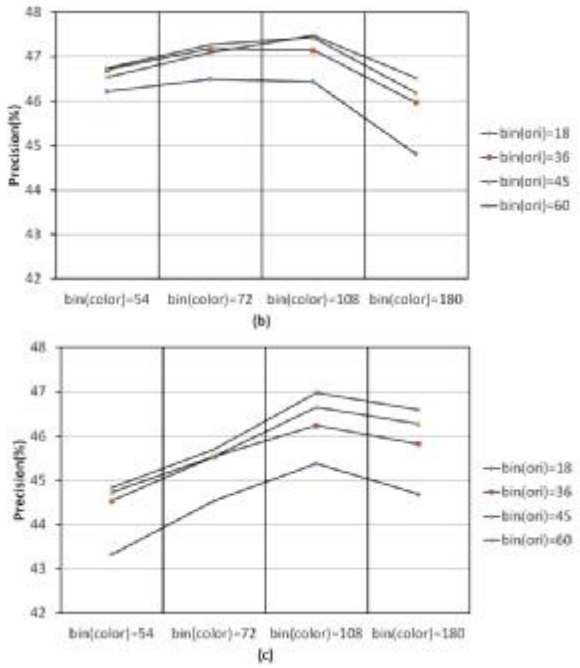


Figure.3. the accuracy of the proposed algorithm with different gray quantization levels on Corel-10K dataset in HSV color space, where bin () denotes the quantization levels of a component. (a) The quantization levels of intensity is 16 bins ($bin\ (inten) = 16$), (b) the quantization levels of intensity is 32 bins ($bin\ (inten) = 32$) and (c) the quantization levels of intensity is 64 bins ($bin\ (inten) = 64$).

As can be seen from figure 3, when $bin(inten) = 16$, $bin(color) = 54$ and $\mathrm{bin(ori)} = 36$, the accuracy of the proposed algorithm is more than 47%. If $bin(inten) = 32, bin(color) = 108$, when the quantization levels of orientation $bin(ori) \in \{36,45,60\}$, the accuracy of the proposed algorithm is more than 47%, whereas in the case of $bin(inten) = 64$, the accuracy of the proposed algorithm is below 47%. It is very important that selecting proper vector dimensions which obtain good retrieval performance while not requiring a great amount of storage space and computation burdens.

In order to obtain the best trade-off between the performance and vector dimension, we think that $bin(inten) = 16$, $bin(color) = 54$ and $bin(ori) = 36$ are more suitable for the proposed algorithm in HSV color space。

C. *Performance comparisons*

In order to implement performance comparisons, our prior works are adopted. They are multi-texton histogram (MTH) [16] and color difference histogram (CDH) [17]. Both of them are developed for content-based image retrieval. As can be seen from figure 4, the performance of the proposed algorithm significantly outperforms our prior works according to the results of precision and recall.

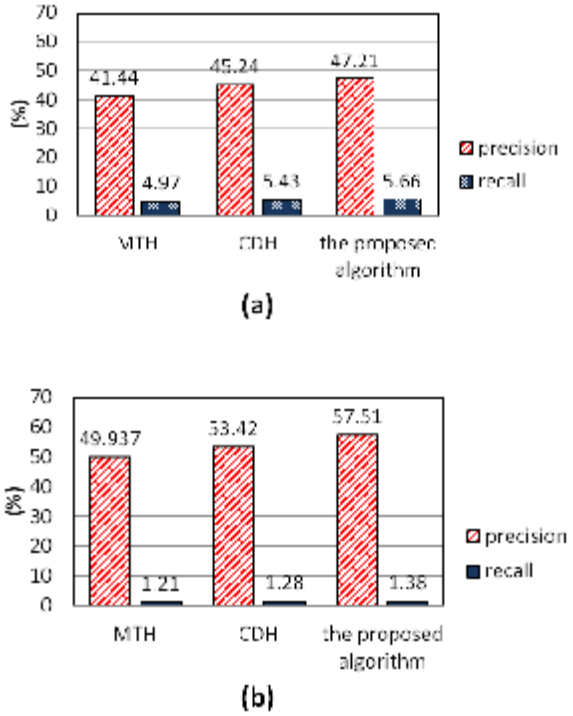


Figure.4. the performance comparisons with the proposed algorithm and our prior works (color difference histogram (CDH) and multi-texton histogram (MTH)), (a) Corel-10K dataset and (b) GHIM-10K dataset.

Color, intensity and orientation are important features in the image representation. Multi-texton histogram encodes color and edge orientation as feature representation [16]. Color difference histogram is the concatenation of three special histograms types involving color, intensity and orientation information [17]. Intuitively, both Multi-texton histogram and color difference histogram have the discrimination power of color, texture and shape features. Figure 4 have shown two retrieval examples on GHIM-10K dataset.

Nevertheless, the most important factors ignored by multi-texton histogram and color difference histogram are the simulation of visual attention mechanism and embedding saliency information into features representation. This ignoring will inevitably weaken the discrimination power [22].

(a)

(b)

Figure.5. two retrieval examples on the GHIM-10K dataset, where the top-left image is the query image, and the similar images include the query image itself, (a) Butterfly and (b) Imperial Palace of Beijing.

As mentioned before, the visual attention mechanisms can help Humans to quickly recognize the conspicuity object in an image or a scene, and this mechanism can made human pay more attention to the conspicuity areas and to discard the unimportant areas. The proposed algorithm has contained this function to some extent. Besides, this framework can encode color, intensity, orientation and saliency information as natural image feature through a series of processes, thus the discrimination power of our algorithm has greatly improved. For these reasons, the proposed algorithm can obtain better performances than color difference histogram and multi-texton histogram.

IV. CONCLUSIONS

In this paper, we present a novel content-based image retrieval framework to encode the primary visual features and saliency information as natural image features by simulating visual attention mechanism and using the conditional probability.

In this framework, color volume is used as a novel visual feature to detect saliency areas and then the prior probability is used to distinguish the grade of saliency areas. Besides, a novel generalized saliency feature representation method, namely the conditional probability histogram, is proposed to describe natural image features. Experimental results indicate that the performances of the prior probability histogram outperform multi-texton histogram and color difference histogram on both Corel-10k dataset and GHIM-10k dataset.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Fund of China (No. 61463008, No. 61202272).

REFERENCES

[1] B.S. Manjunathi and W.Y. Ma,"Texture Features for Browsing and Retrieval of Image Data," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.18, n.8, 1996, pp.837-842.

[2] T. Skora. The MPEG-7 visual standard for content description-an overview, IEEE Transactions on circuits and systems for video technology, 11(6) (2001)696-702.

[3] A. Treisman. A feature in integration theory of attention. Cognitive Psychology, 12(1) (1980) 97-136.

[4] A. Toet. Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(11) (2011)2131-2146.

[5] L.Itti, C.Koch, E. Niebur, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11) (1998)1254-1259.

[6] A. Borji, L. Itti. State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1) (2013)185-207.

[7] X. Hou and L.Zhang. Saliency detection: a spectral residual approach. IEEE conference computer vision and pattern recognition,(2007) 1-8.

[8] D. Walther, C. Koch. Modeling attention to salient proto-objects. Neural Networks, 19(9) (2006) 1395-1407.

[9] A. Borji, L. Itti. Exploiting local and global patch rarities for saliency detection. 2012 IEEE conference on computer vision and pattern recognition. (2012) 478-485.

[10] J.Harel,C.Koch, and P.perona. Graph-based visual saliency, Proceedings of Neural Information Processing Systems (NIPS), 2006.

[11] O.L. Meur, P.L. Callet, etc. A coherent computational approach to model bottom-up visual attention. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(5) (2006) 802-817.

[12] L. Zhang, M.H. Tong, T.K. Marks, etc,. SUN: A Bayesian framework for saliency using natural statistics, Journal of vision, 8(7) (2008):32, 1-20.

[13] Y. Sun, R. Fisher. Object-based visual attention for computer vision. Artificial Intelligence, 20 (11) (2003) 77–123.

[14] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd edition. Prentice Hall, 2007. .

[15] https://en.wikipedia.org/wiki/Conditional_probability.

[16] G-H Liu, L. Zhang, et al., Image Retrieval Based on Multi-Texton Histogram, Pattern Recognition 43(7) (2010)2380-2389.

[17] G-H Liu, J-Y Yang. Content-based image retrieval using color deference histogram, Pattern recognition, 46(1) (2013)188-198.

[18] G. N. Lance, W. T. Williams. Mixed-Data Classificatory Programs I - Agglomerative Systems. Australian Computer Journal 1(1) (1967)15-20.

[19] C.J. van Rijsbergen. Information retrieval (2nd ed). London: Butterworths, 1979.

[20] G-H Liu, J-Y Yang. Image retrieval based on the texton co-occurrence matrix. Pattern Recognition 41(12) (2008) 3521 – 3527.

[21] G-H liu, Z-Y Li, L. zhang, Y. Xu, Image retrieval based on micro-structure descriptor, Pattern Recognition 44(9) (2011)2123-2133.

[22] G-H Liu, J-Y Yang, et al., Content-based image retrieval using computational visual attention model, Pattern Recognition, 48(8) (2015) 2554-2566.