A Model of Visual Attention for Natural Image Retrieval

Guanghai Liu*
College of Computer Science and Information Technology,
Guangxi Normal University
Guilin, 541004, China
e-mail: liuguanghai009@163.com

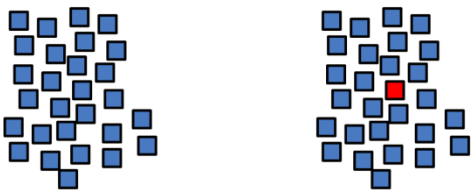
Dengping Fan
College of Computer Science and Information Technology,
Guangxi Normal University
Guilin, 541004, China
e-mail: fdp39@163.com

***Abstract*—in this paper, saliency textons model is proposed to encode color, orientation and saliency cue and spatial information as image features for CBIR, where the image representation is so called saliency textons histogram. Experimental results indicate that the performances of saliency textons histogram outperform Gabor filter and multi-texton histogram. The saliency textons histogram can combine color feature, edge feature and spatial layout together. Furthermore, saliency textons model can simulate visual attention mechanism.**

Keywords-Visual attention; Saliency model; Saliency textons histogram

I. INTRODUCTION

Visual attention mechanisms can help humans rapidly to select the key information from visual field. Recently, image retrieval has become a very popular topic in information science, where image retrieval can be classed into CBIR and objects retrieval, but how to extract features from the vast amount of image data is a challenging problem. Image retrieval can benefit from visual attention mechanisms by extraction the features of salient regions. For example shown in figure 1, humans can quickly recognize the differences between two scenes. In recent years, developing visual attention models to simulate visual attention mechanisms have been attracting more and more interest. Since human visual mechanisms remain elusive, it has becoming a hot topic how to build a visual attention model to well simulate human vision mechanism for natural image retrieval or content-based image retrieval.



(a) (b)

Fugure 1. An example of the significant difference between two scenes.

In order to build a visual attention model for CBIR, saliency textons model is proposed in this paper. In this model, saliency textons histogram is also proposed for image representation. It can outperform Gabor filter texture descriptor [1] and multi-texton histogram (MTH) [2], where both of them are originally developed for CBIR. Furthermore, saliency textons model can simulate visual attention mechanism to some extent.

In this paper, visual attention mechanisms are used for feature extraction and CBIR. This paper includes five sections. In section Ⅱ, there are some introductions of related works. The saliency textons model is presented in section Ⅲ. Performance comparisons among Gabor filter, multi-texton histogram and our algorithm are presented in section Ⅳ. In section V, we will conclude the paper.

II. RELATED WORKS

A. Visual Attention Models

Humans have a remarkable ability to implement various visual search tasks, but the mechanism of human visual system remains elusive. There are two main approaches to use the process of feature extraction derive from saliency models, it includes bottom-up model and top-down model. The characteristics of a visual scene are mainly used by bottom-up models, and it belongs to stimulus-driven model, whereas top-down models focus on cognition mechanisms [3].

Several visual attention models have been suggested over the past years. Majorities of those models derived from Treisman theory [4] and the guided search model [5]. Tsotsos et al have presented a visual attention model by using the concept of selective tuning, where a top-down hierarchy of WTA networks is used for tuning model neurons [6]. One of the most famous saliency models was proposed by Itti et al [7]. In this model, an color image is decomposed into three feature channels at multiple spatial scales by using linear filtering, and then the feature of color, intensity and orientation are extracted, where each feature is computed by center-surround operation which is to stimulate the mechanisms of visual receptive fields. In [8], a biologically plausible model has proposed by Walther and Koch. Meur et al have proposed a visual attention model by using a coherent computational approach [9]. Sun and Fisher have developed a visual attention model based on objects [10]. It has combined two new mechanisms about groupings and attention shifts [10]. Borji and Itti have introduced a saliency model by using the rarities property [11]. A review of saliency models can be found in [3] [12], and it will not be considered here.

In summary, those models have been studied in neuroscience and psychology, whereas they still have not been researched within the content-based image retrieval (CBIR) framework. Moreover, how to construct visual attention model

is still an open problem.

B. CBIR Techniques

The classic CBIR techniques are based on global features and local features. In global features-based algorithms, it pays more attention to the whole image, where color, texture and shape are used as visual content. Local features-based algorithms are mainly use keypoints features or salient patches features as image representation. Different kinds of algorithms have been proposed to extract features and used for CBIR.

There are dominant color descriptor, color layout descriptor, and scalable color descriptor in the MPEG-7 standard [1]. Many descriptors have been used for texture representation in current literatures [13-17], where include three texture descriptors in the MPEG-7 standard [16].

In order to improve the discrimination power, combing the texture features and color features can obtain better retrieval performance. There are some descriptors which can combine texture feature and color feature for CBIR or image classification [2] [18-21]. The classical shape features include moment–based descriptor [22] [23], frequency domain methods [24] [25], edge curvature area [40]. Besides, three shape descriptors are used in the MPEG-7 standard [16]. In many cases, it need image segmentation. Since it is very difficult to segment the precise shape, thus many researchers have adopted local features (e.g., key points, salient patches) instead of using the traditional shape features.

Various local features descriptors have been reported in the literature [26-30], where SIFT is the most famous local feature representation. In recent years, Bag-of-visual words have commonly used in objects recognition or scene categorization. In many cases, visual words can obtained by vector quantizing the SIFT descriptors or other local features descriptors. In [31], bag-of- visual words method is proposed to object-based image retrieval. The ideas of bag-of-words derive from text retrieval application, where the vocabulary is constructed by using the clustering algorithm. In [32], the hierarchical k-means algorithm is proposed to construct visual words, and it can result in better performance in image classification. The bag-of-visual words have two major limitations. One is the lack of any explicit semantic meanings, and other is visual words has the ambiguity. In order to address the above shortcomings, many researchers have embedded spatial information or semantic attributes into BOW histogram and therefore reduce those limitations [33-39]. In many cases, bag-of-words models are mainly focus on object-based image retrieval, object recognition and scene categorization, and not real content-based image retrieval.

There are extensive studies about developing visual attention models for object detection or scene categorization. However, developing visual attention models for CBIR need to be further studied, besides, CBIR can benefit from visual attention mechanisms. Thus, we have proposed the saliency textons model. This model can be considered as an improved version of our earlier work (multi-texton histogram) [2] by combining saliency information.

III. SALIENCY TEXTONS MODEL

A. Extraction of the Primary Visual Features

Generally speaking, the primary visual features are mainly included color, orientation and intensity information [3]. In many visual saliency models, "color double-opponent" system is commonly adopted. For example, it is used in Itti visual attention model [3]. Lab color space is colorimetric and device independent [40] [41].where L channel is measured the lightness information, and a and b channels are measured Chroma information. In the saliency textons model, Lab color space is adopted to extract the primary visual features.

In order to obtain color map, the L cannel is uniformly quantized into 10 bins, whereas 3 bins for both a and b channels, so that there are $10\times3\times3=90$ color combinations in color map. Let$\mathrm{M_C}(x,y)$ denotes the color combinations or color map, as $\mathrm{M_C}(x,y)=w, w\in\{0,1,..\mathrm{N_C}-1\}$, where $\mathrm{N_C}=90$.

In this paper, lightness information $\mathrm{L}(x,y)$is used to detect edge orientation $\mathrm{O}(x,y)$. Gradient image $\mathrm{g}(x,y)$ is obtained by using Sobel operation. Since the computational burden of Gabor filters is much high, we are not adopt Gabor filter for local orientation detection instead of using Sobel operation. After uniform quantization, we can obtain the edge orientation map $\mathrm{M_O}=v, v\in\{0,1,..,\mathrm{N_O}-1\}$, where $\mathrm{N_O}=18$.

B. The Saliecy Map

In order to detect the saliency map, color informationa(x,y), $\mathrm{b}(x,y)$ and gradient information $\mathrm{g}(x,y)$ are used to create Gaussian pyramid $\mathrm{a}(\sigma),\mathrm{b}(\sigma)$ and $\mathrm{g}(\sigma)$, respectively, where $\sigma\in[0\ldots4]$ is the scale. In Gaussian filter construction, the standard deviation of Gaussian kernel is 5.0. We use across scale subtraction "$\ominus$" to simulate the mechanisms of center-surround receptive fields, where two maps are used for this operation, one map at the center (c) scales , other map at surround (s) scales, and yields the so-called feature maps:

$$\mathrm{F}(c,s,a)=|\mathrm{a}(c)\ominus\mathrm{a}(s)| \tag{1}$$

$$\mathrm{F}(c,s,b)=|\mathrm{b}(c)\ominus\mathrm{b}(s)| \tag{2}$$

$$\mathrm{F}(c,s,g)=|\mathrm{g}(c)\ominus\mathrm{g}(s)| \tag{3}$$

After center-surround operation, we can obtain 12 feature maps.

In this paper, we denote $\overline{\mathrm{a}}$, $\overline{\mathrm{b}}$ and $\overline{\mathrm{g}}$ as the individual saliency maps at the scale $(\sigma=4)$ by using across-scale addition"$\oplus$", where the implementation of the across-scale addition consists of two steps, one is the reduction of each map, other is the addition in the manner of point-by-point. It is similar to the manner of Itti's saliency model [5].

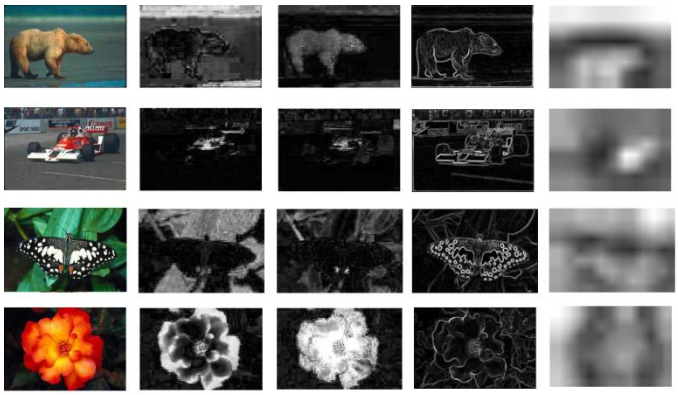
$$\overline{\mathrm{a}}=\bigoplus_{c=0}^{3}\bigoplus_{s=4}^{4}|w_d\times\mathrm{F}(c,s,a)| \tag{4}$$

$$\overline{b} = \bigoplus_{c=0}^{3} \bigoplus_{s=4}^{4} |w_d \times F(c,s,b)| \quad (5)$$

$$\overline{g} = \bigoplus_{c=0}^{3} \bigoplus_{s=4}^{4} |w_d \times F(c,s,g)| \quad (6)$$

Where w_d denotes the inhibition term, and $w_d = 0.01$. In order to obtain a single overall saliency map S, we need to combine $\overline{a}$, $\overline{b}$ and $\overline{g}$ together:

$$S = \frac{1}{3}(\overline{a} + \overline{b} + \overline{g}) \quad (7)$$



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

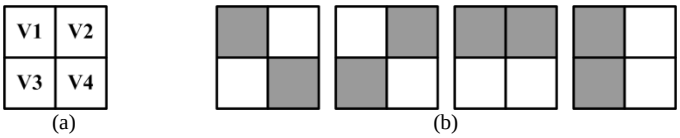
Fugure 2. The illustration examples of saliency textons detection and saliency texton map. (a)original image; (b) color information a(x, y); (c) color information b(x, y);(d) gradient image g(x, y)and (e) saliency map.

At last, the single overall saliency map S would be resized as the same as the original image. As can be seen from figure 2 (e), part of background of image is shielded and pop out the major objects by using the combination of color and gradient information.

In this paper, the single overall saliency map S is used for further processing and analysis.

C. Texton detection

In this model, we define saliency textons as the small saliency areas that have the similar local structures. In the prior works [2] [18] [20], several texton patterns have proposed for content-based image retrieval. In this model, four texton patterns are adopted to detect textons, and they are denoted as $T_1, T_2, T_3$and T_4, respectively. They are shown in figure.3 (b). Let there is a grid of size 2×2 in a saliency map S, it has four pixels and denoted as V_1, V_2, V_3 and V_4, respectively. If there are two pixels that have the same values, the grid can be considered as a texton. In order to denote the pixels which have the same values in 2×2 grid, gray color is highlighted in those pixels.

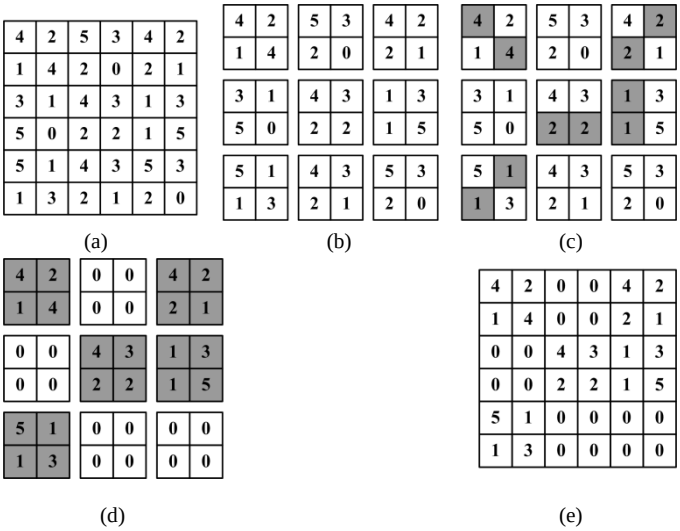


(a) (b)

Figure 3. Four texton types are adapted to detect texton: (a) grid and its pixels ;(b) four texton patterns derive from the grid.

The working mechanisms of texton detection include some steps:

Saliency map $S(x, y)$is divided into a set of 2×2 grid, then we begin to detect every grid by using four texton patterns which are shown in figure.3(b), and then it need to judged whether one of the four texton patterns occur in the grid. This grid can be considered as a saliency texton if that happens, and then the textons areas keep original pixels value. Otherwise all the pixels of this grid will be set to 0 values.



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

Figure 4. The working mechanism of the saliency textons detection: (a) original image, (b) grid divided, (c) textons detection, (d) saliency texton areas, and (e) saliency textons image.

The detection procedures are illustrated in figure 4. The areas highlighted in gray color in figure 4(d) are the textons areas, where the saliency textons image denotes as $T_S(x, y)$ of size $M \times N$ that is shown in figure 4(e).After saliency textons detection, image representation should be introduced. In stage of image representation, all the values of $T_S(x, y)$areprojected into 0 to 255.

D. Image Representation

In practice, how to integrate saliency information and spatial layout into image representation is a difficult problem. In order to address this problem, saliency texton histogram is propose for image representation. In saliency texton histogram, color, edge orientation, saliency information and spatial information are combined into one whole unit.

The saliency texton histograms of color image are defined as follows:

$$H_C(C(x,y)) = \begin{cases} \sum\sum T_S(x,y) \\ \text{where } C(x,y) = C(x+\Delta x, y+\Delta y) \end{cases} \quad (8)$$

$$H_\theta(\theta(x,y)) = \begin{cases} \sum\sum T_S(x,y) \\ \text{where } \theta(x,y) = \theta(x+\Delta x, y+\Delta y) \end{cases} \quad (9)$$

$$H = \text{conca}\{H_C, H_\theta\} \quad (10)$$

Where $\text{conca}\{.\}$ denotes the concatenation of H_C and H_θ, where H_C denotes the color co-occurring histogram within the texton image areas, leading to 90 dimension vector, and H_θ denotes the edge orientation co-occurring histogram within texton image areas, leading to 18 dimension vector. The total dimension of saliency texton histogram is 90+18=108.

IV. THE EXPERIMENTS OF CBIR

In order to demonstrate the performance of saliency texton histogram, Corel-5K dataset and Corel-10K dataset are used for comparisons, there are 50 categories and 100 categories on two datasets, respectively, where every category has 100 images, and the size of every image is 192×128 pixels.

In comparisons, 500 images and 1000 images have randomly selected from two dataset as queries, respectively. For fair comparisons, we have selected two CBIR methods, such as Gabor filter [1] and multi-texton histogram (MTH) [2], where MTH is our prior work. Gabor filter are implemented via three color channels in RGB color space, and result in 48×3=144 vector dimension. The vector dimension of our prior work MTH is 108. Indeed, all three algorithms have integrated color information into image representation.

A. Distance Metric

The commonly distance metrics or similarity metrics can be used to match images, such as Canberra distance [42] is one of the most popular distance metrics. It can be described as follow:

$$D(T,Q) = \sum_{i=1}^{108} \frac{|T_i - Q_i|}{|T_i| + |Q_i|} \quad (11)$$

If $D(T,Q)$ is small, the two images are similar in their image content.

B. Performance Measures

In CBIR experiments, the *Precision* and *Recall* can evaluate the performance of our algorithm [2] [18] [20] [21]. The definition of *Precision* and *Recall* can be described as follows:

$$P(N) = I_N / N \quad (12)$$

$$R(N) = I_N / M \quad (13)$$

In the calculation of precision and recall, the top positions is N=12, ever category has M=100 images. In the top N=12 positions, I_N is the number of similar images in those retrieved images. The average results of all queries are evaluated as the final performances.

C. Performance Comparisons

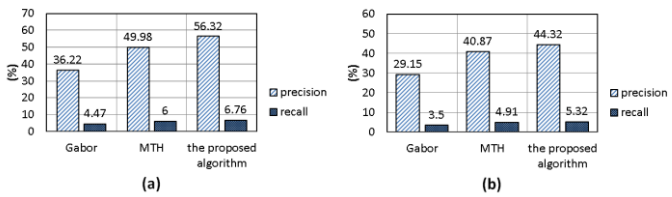
In table 1, there is the performance data on Corel-5k dataset, where the average precision of STH is from about54% to 57%. In order to obtain the best trade-off between vector dimension and the precision, 90 bins are selected for color quantization and 18 bins are selected for edge orientation quantization.

TABLE 1.THE PRECISION AND RECALL OF THE PROPOSED ALGORITHM ON COREL-5KDATASET.

Color bins	Edge orientation bins							
	Precision (%)				*Recall (%)*			
	6	12	18	36	6	12	18	36
180	53.50	54.78	55.65	57.22	6.42	6.57	6.68	6.87
160	45.08	47.48	47.48	48.15	5.41	5.70	5.70	5.78
90	53.08	55.05	56.32	56.93	6.37	6.61	6.76	6.83
45	51.77	53.88	54.62	54.72	6.21	6.47	6.55	6.57

The performance on the two dataset has showed in figure 5. According to the results of table 1 and figure 5, saliency texton histogram outperforms multi texton histogram and Gabor filter. If Corel-5K dataset is used for comparisons, the precision of saliency textons histogram is 20.1% and 6.34% higher than that of Gabor filer and multi-texton histogram, respectively. If Corel-10K dataset is used for comparisons, the precision of saliency textons histogram is 15.17% and 3.34% higher than that of Gabor filter and multi texton histogram, respectively. Gabor filter method achieves the worst performance.

As for the vector dimensionality comparisons, saliency textons histogram has 108 dimension vector, Gabor filter and multi texton has similar vector dimensionality to saliency textons histogram, where Gabor filter are too compute intensive [1][16].



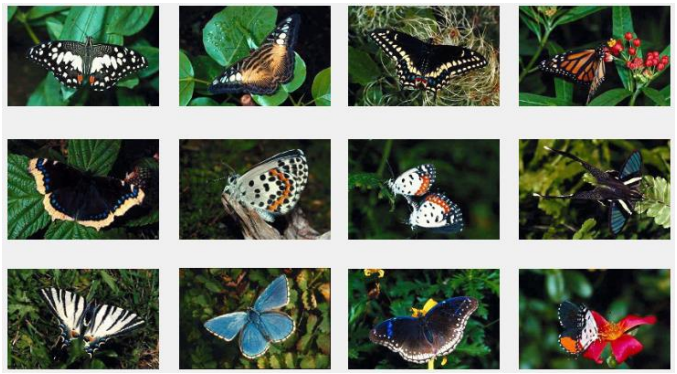
Fugure 5. The performance comparisons among Gabor, MTH and saliency textons histogram: (a) Corel-5K dataset, (b) Corel-10K dataset.

D. Performance Discussion

Figure 6 and 7 has shown two examples that they come from two Corel datasets. In Figure 6, a butterfly image is used as the query. As can be seen from the top 12 retrieved images, those butterfly images have similar texture and color features. In Figure 7, a racing car is used as the query. As can be seen from the top 12 retrieved images, all images are racing car images, where those racing car images have obvious shape features and significant direction-sense. Figure 6 and 7 are only used to indicate that saliency texton histogram can combine color, texture and shape features.

Gabor filter is one of the most popular methods in the extraction of texture features, where the fundamental characteristics of cortex cells can be captured by using Gabor function [43] [44] [45]. It is very suitable for texture representation and discrimination [1] [16]. But the texture feature is only one of the image attributes. Using texture

feature is not accurate enough, and it cannot improve the performance significantly.



Fugure 6. An examples of image retrieval by saliency texton histogram, where Corel-5K dataset is used for comparisons, a butterfly image is used as the query.

Fugure 7. An examples of image retrieval by saliency textons histogram, Corel-10K dataset is used for comparisons, a racing car image is used as the query.

The multi-texton histogram (MTH) is based on Julesz's texton conception [46]. It has encoded color information and edge orientation information as image representation, thus it can enhance the ability of image description. Nevertheless, the simulation of visual attention mechanism is ignored by multi-texton histogram. In practice, CBIR can benefit from visual attention by extracting the features of salient region.

The saliency textons histogram (STH) has adopted color feature and orientation feature in saliency areas by using two special histograms types. It has taken the advantage of MTH algorithm and has also overcome the shortcoming of MTH that it is not simulating visual attention mechanism. It has introduced saliency information into MTH algorithm and can simulate visual attention mechanism for CBIR. It is the reason why STH can outperform Gabor filter and MTH.

V. CONCLUSIONS

In this paper, the saliency textons model was proposed to encode color, orientation and saliency information and spatial layout as image representation for natural image retrieval or CBIR, where saliency textons histogram is used as image representation. In essence, saliency textons histogram has introduced saliency information into MTH algorithm and can simulate visual attention mechanism for CBIR. It outperforms multi-texton histogram and Gabor filter. It can combine color feature, texture feature, orientation cue and spatial information together.

ACKNOWLEDGMENT

This research was supported by the national natural science fund of china (no. 61202272). Besides, this research was supported by and Guangxi natural science foundation (no: 2011GXNSFB018070)

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] G-H Liu, L. Zhang, et al., "Image Retrieval Based on Multi-Texton Histogram," Pattern Recognition, vol.43, n.7, 2010, pp.2380-2389.

[3] A. Borji, L. Itti, "State-of-the-art in visual attention modeling," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.35, n.1, 2013, pp.185-207.

[4] A. Treisman, "A feature in integration theory of attention,"Cognitive Psychology, vol.12, n.1, 1980, pp.97-136.

[5] J.M. Wolfe, T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" Nature Reviews Neuroscience, vol.5, n.6, 2004, pp. 495-501.

[6] J.K. Tsotsos, S.M. Cuihane, etc, "Modeling visual attention via selective tuning," Artificial intelligence, vol.78, n.1, 1995, pp.507-545.

[7] 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, vol.20, n.11, 1998, pp.1254-1259.

[8] D. Walther, C. Koch, "Modeling attention to salient proto-objects," Neural Networks, vol.19, n.9, 2006, pp.1395-1407.

[9] 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, vol.28, n.5, 2006, pp. 802-817.

[10] Y. Sun, R. Fisher, "Object-based visual attention for computer vision,"Artificial Intelligence, vol.20, n.11, 2003, pp.77–123.

[11] A. Borji, L. Itti, "Exploiting local and global patch rarities for saliency detection," 2012 IEEE conference on computer vision and pattern recognition (CVPR), IEEE, vol.1, 2012, pp.478-485.

[12] A. Toet, "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.33, n.11, 2011, pp.2131-2146.

[13] R.M. Haralick, Shangmugam, Dinstein, "Textural feature for image classification,"IEEE Transactions on System, Man and Cybernetics, vol.3, n.6, 1973, pp.610-621.

[14] H. Tamura, S. Mori, T. Yamawaki, "Texture features corresponding to visual perception," IEEE Transaction on System, Man and Cybernetics, vol.8, n.6, 1978, pp.460–473.

[15] G. Cross, A. Jain, "Markov random field texture models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.5, n.1, 1983, pp.25–39.

[16] B. S. Manjunath, P. Salembier, and T. Sikora. Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley & Sons Ltd, New York, 2002, pp.187-260.

[17] T. Ojala, M. Pietikanen and T. Maenpaa, "Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, n.7, 2002, pp.971-987.

[18] G-H Liu, J-Y Yang, "Image retrieval based on the texton co-occurrence matrix,"Pattern Recognition, vol.41, n.12, 2008, pp.3521 – 3527.

[19] J. Luo and D. Crandall, "Color object detection using spatial–color joint probability functions," IEEE Transactions on Image Processing, vol.15, n.6, 2006, pp.1443–1453.

[20] G-H liu, Z-Y Li, L. zhang, Y. Xu, “Image retrieval based on micro-structure descriptor,” Pattern Recognition, vol.44, n.9, 2011, pp.2123-2133.

[21] G-H Liu, J-Y Yang, “Content-based image retrieval using color deference histogram,” Pattern recognition, vol.46, n.1, 2013, pp.188-198.

[22] M-K Hu, “Visual pattern recognition by moment invariants,” IRE Transactions on Information Theory,vol.8, n.2, 1962, PP.179-187.

[23] P-T Yap, R. Paramesran and S-H Ong, “Image Analysis using Hahn moments,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, n.11, 2007, pp.2057-2062.

[24] F.P. Kuhl, C.R. Giardina, “Elliptic Fourier features of a closed contour,” Computer Graphics Image Process, vol. 18, n.3, 1982, pp.236–258.

[25] C.T. Zahn, R.Z. Roskies, “Fourier descriptors for plane closed curves,” IEEE Transaction on. Computer, vol.21, n.3, 1972, pp.269–281.

[26] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol. 60, n.2, 2004, pp.91-110.

[27] Y. Ke, R. Sukthankar, “PCA-SIFT: a more distinctive representation for local image descriptors,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, vol.2, 2004, pp.506–513.

[28] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, n.10, 2005, pp.1615–1630.

[29] H. Bay, T. Tuytelaars, L.V. Gool, “SURF: speeded up robust features,” European Conference on Computer Vision (ECCV), Springer, vol.1, 2006, pp.404–417.

[30] K. Mikolajczyk, T. Tuytelaars, C. Schmid, et al., “A Comparison of Affine Region Detectors,” International Journal of Computer Vision. , vol.65, n1-n.2, 2005, pp.43-72.

[31] J.Sivic, A. Zisserman, “Video Google: A text retrieval approach to object matching in videos,” IEEE International Conference on In Computer Vision (ICCV), IEEE, vol.2, 2003, pp. 1470 – 1477.

[32] D. Nister, H. Stewenius, “Scalable Recognition with a Vocabulary Tree,” IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, vol.2, 2006, pp.2161 - 2168.

[33] S. Lazebnik, C. Schmid, and J. Ponce, “beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, vol.2, 2006, pp.2169 - 2178.

[34] E. Nowak, F. Jurie, and B. Triggs, “Sampling strategies for bag-of features image classification,” European Conference on Computer Vision (ECCV), Springer, vol.1, 2006, pp.490-503.

[35] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Lost in quantization: Improving particular object retrieval in large scale image databases,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, vol.1, 2008, pp.1-8.

[36] J. Sivic, A. Zisserman, “Efficient Visual Search of Videos Cast as Text Retrieval,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31, n.4, 2009, pp.591-606.

[37] Y. Su, F. Jure, “Improving image classification using semantic attributes,” International Journal of Computer Vision, vol.100, n.1, 2012, pp.59-77.

[38] J.C. van Gemert, C.J. Veenman, A.W.M Smeulders, J. M. Geusebroek, “Visual word ambiguity,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, n.7, 2010, pp.1271–1283.

[39] S. Lazebnik, C. Schmid, J. Ponce, “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, vol.2, 2006, pp.2169 – 2178.

[40] W. Burger, M.J. Burge, “Principles of Digital image processing: Core Algorithms,” Springer, London, 2009, pp. 32-124.

[41] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd edition. Prentice Hall, Upper saddle River, New Jersey, 2007, pp.282-344.

[42] G. N. Lance, W. T. Williams, “Mixed-Data Classificatory Programs I - Agglomerative Systems,” Australian Computer Journal, vol.1, n.1, 1967, pp.15-20.

[43] D. Hubel, T.N. Wiesel, “Receptive fields, Binocular interaction and functional architecture in the cat's visual cortex,’’ Journal of physiology, vol. 160, n.1, 1962, pp.106-154.

[44] J. G. Daugman, “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters,” Journal of the Optical Society of America A, vol. 2, n.7, 1985, pp.1160–1169.

[45] D.J. Field, “relations between the statistics of natural images and the response properties of cortical cells,” Optical society of America, vol. 4, n.12, 1987, pp.2379-2394.

[46] B. Julesz, “Textons, the elements of texture perception and their interactions,” Nature, vol.290, n.5802, 1981, pp.91–97.