DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232

Visual Attention Model and Relevant Feedback based Image Retrieval

Zhijiang Li[a,b], Jiaxian Long[a], Chuan Dong[c]
[a]School of Printing and Packaging, Wuhan University, Wuhan, 430079, P.R.China
[b]School of Design, University of Leeds, Leeds, LS2 9JT, UK
[c]Internet Center, Hubei Mobile Communication Company Limited, Wuhan, 430040, P.R.China Mobile

Abstract

To improve the efficiency and accuracy of Content Based Image Retrieval (CBIR) for specific images, a new method is presented in the paper. The method focuses on 3 key problems. Firstly, considering the impaction of saliency point near the attention focus, an improved saliency region extraction algorithm is proposed to locate object of interest more accurately. Then, the construction of Bag-of-Features (BoF) feature vector is improved by our visual attention model to extract features more effectively. Finally, Particle Swarm Optimization (PSO) is introduced to optimize the learning process of the feedback model based on Support Vector Machine (SVM) to boost the accuracy and efficiency of the image retrieval. Experiments and comparison between typical algorithms based on Caltech 101 dataset and self-collection dataset demonstrate that the method proposed in this paper can improve the accuracy and efficiency of content based image retrieval.

Keywords： *Content Based Image Retrieval (CBIR); Visual Attention; Saliency Region; Relevance Feedback; Particle Swarm Optimization (PSO); Support Vector Machine (SVM); Bag-of-Features (BoF)*

1 Introduction

The explosive growth of image database and the research on machine learning has made CBIR increasingly significant. The application of CBIR range from personal collections to medical and scientific images classification and searching. Since 2009, image retrieval has been considered as the most perspective innovative application [1].

In general, CBIR systems use low level features like color, texture, shape and spatial location of objects to retrieve images from databases. However, different retrieval requests consider different features or objects in the image. The extraction of the Region of Interest (ROI) becomes a basis of relative methods.

In addition, because of the sematic gap between low level feature and sematic representation, traditional CBIR methods can't obtain satisfied retrieval result. With the development of statistical learning methods and increasing volume of kinds of image databases, CBIR based on statistical learning methods become more and more important.

Literatures indicate that few method or feature is suitable for most images in CBIR [2]. The problems are: (1) Precision. Users usually have to spend a lot of time to make selection in the retrieval results. (2) Interactive mode. Different application request need different feedback, even based on the same dataset. So, how to get a useful feedback and how to select features based on different applications are problems should be faced. (3) Relevance between low level features and high level sematic. The research in low level features extraction and application is almost full-fledged now, but how to use the features efficiently to form an optimized model still deserves further study.

Therefore, it's very important to locate ROI accurately to obtain valid features for retrieval, provide friendly and efficient interaction model for users, and search images reflect user's attention effectively.

This paper presents a method aimed at the problems above to improve the CBIR accuracy and efficiency. Major contributions of the paper are first, we improve the saliency region extraction algorithm considered the impaction of saliency point near the attention focus and relevance feedback; second, optimize the POS-SVM-RF algorithm in feature selection, training parameter optimization and fitness evaluation; third, a global framework of CBIR is proposed based on Visual Attention Model and Relevant Feedback Mechanism. The experimental results show that the proposed method can meet requirements of a user more accurately with limited friendly interactions.

The remainder of this paper is organized as follows. Firstly, overview of related work is presented in Section 2. The proposed work is described in Section 3. Results are given in Section 4. Conclusions are drawn in Section 5.

2 Related Works

Researches has indicated that global feature based CBIR methods [3, 4, 5] are difficult to obtain interested result accurately. It's necessary to find a local method to describe and match features in ROI. Visual attention mechanisms by using the saliency information will be useful in the process [6].

Except the local features extraction, the methods about how to use the features to query images are researched more and more extensive. The machine learning algorithms were introduced to consider the query process as a supervised learning problem or a classification problem. Thus, image retrieval can be looked as a problem about dichotomy classification between relevant images and irrelevant images through a classification model based on the dataset has been labeled by users.

In the following subsections, a comprehensive review of visual attention models and machine learning algorithms based image retrieval methods are given.

2.1 Visual Attention Models

Visual Attention Model was first proposed by Itti and Koch[7,8]. Thereafter, lots of researches were presented. Walther [9] improved the attention focus region detection based on saliency region. Wu et.al. [10] proposed an interest points selection method by topology coherence for multi-query image retrieval. Zheng et.al. [11] explored a shape prior method based on interest points for medical images. Hiremath and Pujari [12] presented a method for salient points determination based on color saliency. Pedrosa et.al.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232

[13] proposed a method based on the shape description of salience points. Yang et.al. [14] proposed a novel method using local visual attention feature. Most methods focused on extracting the most interested region in specific images via point set. While, in the methods, the impaction of saliency point near the attention focus were usually ignored, which would form the shift of saliency region. Besides, as a subjective evaluation simulation, it's a complex process to simulate visual physiology and psychology. If the subjective reflection could be considered through the process of relevance feedback, it will improve the visual attention model greatly.

2.2 Relevance Feedback (RF)

To understand the subjective meaning of a visual query, Relevance Feedback (RF) is introduced as an effective method for CBIR [15]. RF items from text retrieval area, which changes the retrieval mode from one-shot search to multi-times interaction through the introduction of human interaction. According to the retrieval model adopted in the algorithms, RF algorithms are categorized into three classes: distance-based approach, probabilistic approach, and machine learning based approach [16]. In the kinds of RF methods, relevance are usually evaluated by direct or indirect methods, and in the direct methods, binary relevance scale were used more popular than multiple relevance scale.

2.3 Machine Learning based CBIR

Recent years, increasing number of machine learning algorithms are introduce in CBIR, such as Bayes network [17], SVM [18], neural network [19]. The methods are useful in the improvement of the CBIR algorithms. But most of them are suitable for big sample set which is relative difficult to obtain in reality. Then, SVM as a popular method suitable for smaller samples set is usually introduced to realize the CBIR. But SVM is sensitive to the amount of number of positive training examples and the average accuracy is not the best [20]. To boost the deficiency of SVM, PSO algorithm was introduced [21,22] to minimize the user interaction by minimizing the RF number. However, the main drawback of PSO is that the swarm may prematurely converge, which traps the particles in local optima rather than global optima. Therefore, it's deserved to optimize the POS-SVM-RF methods further in feature selection, training parameter optimization and fitness evaluation.

3 Methodology of the Proposed Work

An image usually contains many redundant information or multiple objects. User always want to search the image including his most interested objects [23], which makes the ROI extraction important in CBIR. Then, the paper considered object region extraction for image database and query images and RF is introduced as an interaction mechanism in the iterations to reflect the subjective intention in image query to improve the performance of retrieval system.

Therefore, this study has focused on developing a new approach for enhancing the performance of CBIR using Visual Attention Models and RF by integrating user interaction in saliency region selection and retrieval. It consists of three processes: (1) Improved Visual Attention Model provide relative accurate initial display result and weight adjust method based on RF; (2) Local features descriptor based on Bag-of-Features (BoF) for the extraction of feature vector; (3) Optimization in feature selection and fitness evaluation for PSO-SVM-RF to speed up the convergence of iteration. Architecture of the overall method is described in Figure 1. Detailed mechanism of each part is presented in the following sub-sections.

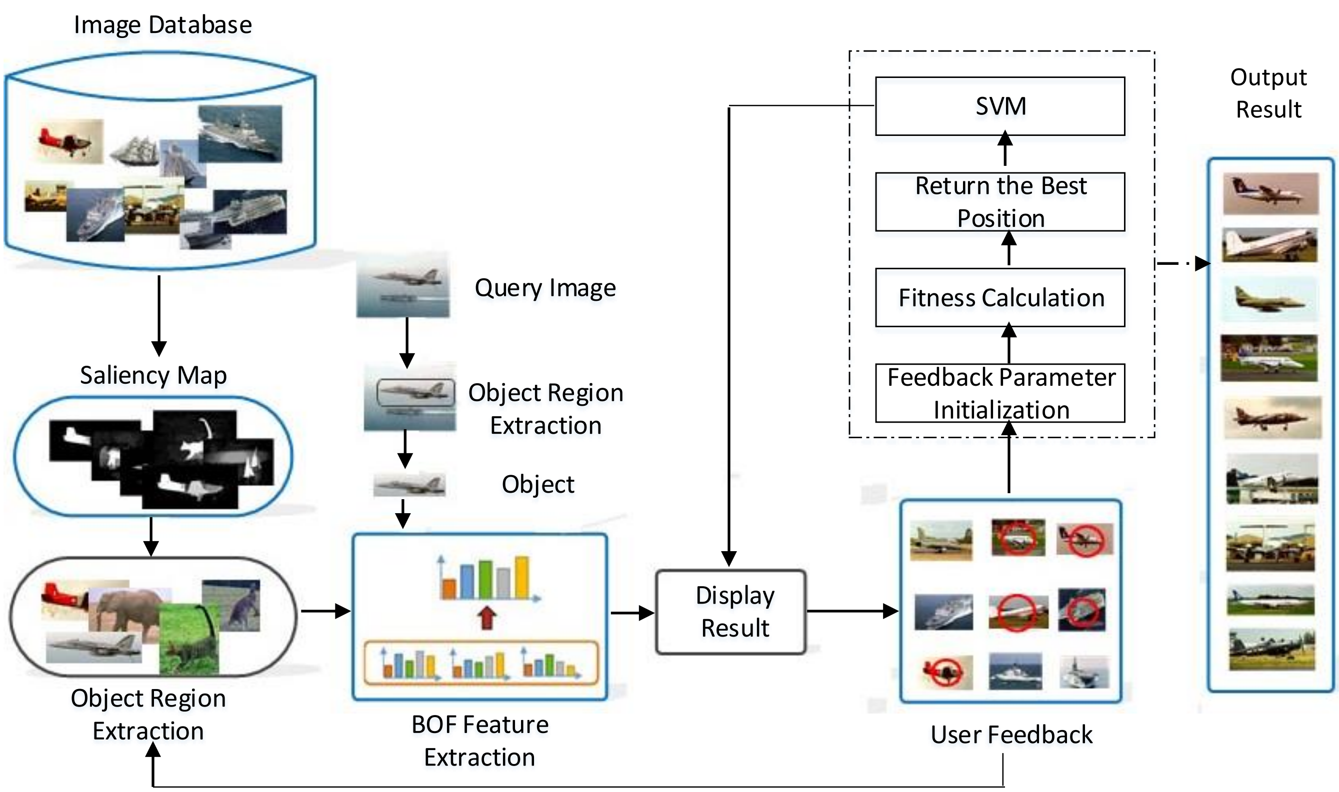


Fig.1 Architecture of the proposed method

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232

3.1 Object Region Extraction

Selective visual attention provides an effective mechanism to serialize perception of complex scenes in both biological and machine vision systems. The method can be formed in two direction: bottom-up salient region selection and top-down salient region selection and the former is better in expansibility.

Several computational models of visual attention have been suggested. An important model was presented by Itti et al. [7,8] based on feature-integration theory of attention [24] and saliency map [25], which verified that 95% maximal saliency points was in the object region labeled by users. The model was referenced extensively by other researches. The method proposed in the paper is also based on the algorithms proposed by Itti and Walther [9].

3.1.1 Early Visual Features Extraction

Early visual features of a color images can be extracted in intensity, orientation and color by linear filtering. Intensity channel can be calculated as the average of R, G and B. Local orientation information is obtained from Intensity channel using oriented Gabor pyramids in 4 directions {0°, 45°, 90°, 135°} [26]. Color channel is suggested to be obtained by the difference map between R and G, Y and B according to the equation below in Itti model to present different contrast.

$$M_{RG} = \frac{r - g}{\max(r, g, b)} \tag{1}$$

$$M_{BY} = \frac{b - \min(r, g)}{\max(r, g, b)} \tag{2}$$

Literature [27] introduced direction and contrast in color feature channel based on the color perception mechanism of living being. In the view of biology, human vision system can perceive vision information hierarchically alone the ventral pathway of Retina-Lateral Geniculate-V1-V2-PIT-IT. And in area V1, RG, BY and intensity channels are formed by double-opponent neurons, where the RG and BY are antagonism colors with contrast sensitivity. To enrich the color feature, a receptive field model suggested by Shapley et.al. [27] has been introduced in the paper as the following equation.

$$\begin{aligned} R(x, y, \lambda) = {} & a_L L(\lambda)\, r_L(x, y) + \\ & a_M M(\lambda)\, r_M(x, y) + \\ & a_S S(\lambda)\, r_s(x, y) \end{aligned} \tag{3}$$

In the equation, (x, y) is the coordination of the image. L(λ), M(λ), S(λ) are R, G, B channels according to spectral response function. r_L, r_M, r_S are spatial sensitivity distribution of each input. They reflects the shape of receptive field and could be simulated by Difference of Gaussians (DOG). a_L, a_S, a_M are coefficient to adjust the result. Fig. 2 shows the experiment result of extraction of early visual features based on the method above.

Then, nine spatial scales are created using dyadic Gaussian pyramids [26], which progressively low-pass filter and subsample the input image, yielding horizontal and vertical image-reduction factors ranging from 1:1 (scale zero) to 1:256 (scale eight) in eight octaves [7]. In order to highlight the object region better based on the feature images, each feature is computed by a set of linear "center-surround" operations to obtain the contrast information between local center and background [7].

3.1.2 Saliency Map Generation

The saliency map can be obtained by the combination of all feature through a combination strategy to provide evidence for attention selection and transfer. The paper uses the local iteration method [7], which includes normalization of all feature maps, iterative convolution with Gaussian Difference Function as the formula below, and weighted overlay of color, intensity, orientations map in each scale, to obtain a saliency map. Fig. 3 shows the component and combination of a saliency map generation.

$$\begin{cases} DOG(x, y) = \dfrac{c_{ex}^2}{2\pi\sigma_{ex}^2} \exp\left(-\dfrac{x^2 + y^2}{2\sigma_{ex}^2}\right) - \dfrac{c_{inh}^2}{2\pi\sigma_{inh}^2} exp\left(-\dfrac{x^2 + y^2}{2\sigma_{inh}^2}\right) \\ F = \left[E + F * DOG - C\right] \end{cases} \tag{4}$$

Where, DOG(x,y) is 2d Gaussian Difference Function, σ_{ex} and σ_{inh} are exciting bandwidth and inhibition bandwidth respectively, the default value are 2% and 25% of the image width. c_{ex} and c_{inh} are exciting constant and inhibition constant respectively, the default value are 0.5 and 1.5.

The experiment indicated that Itti model can express the local region with larger saliency value such as the outline of the planes in Fig. 3. However, because the saliency map was combined by low level features, some saliency points with relative lower saliency value would be ignored, which would make the saliency region incomplete. The paper proposed a compensation method based on a global saliency information as following.

(1) Color space conversion from RGB to Lab.

(2) Calculate the average value of each channel in Lab color space, labeled as avgl, avga, avgb respectively.

(3) Noise reduction for each channel.

(4) Calculate the global saliency value for each pixel as following formula.

$$\mathrm{salmap}_i = \left(L_i - \mathrm{avgl}\right)^2 + \left(a_i - avga\right)^2 + \left(b_i - avgb\right)^2 \tag{5}$$

(5) Combine with the saliency obtained by Itti model.

Fig. 4 shows the result of our experiment. The completeness of the saliency region is improved.

3.1.3 Object Region Extraction based on RF

In the process of most saliency region extraction, considered the impaction of the saliency of neighbor points near the attention point, a method based on saliency region weight calculation is proposed. In the method, the weight of each saliency region is calculated as following formula and the weight should be sorted in descending order. The region with biggest weight is the most saliency region. Every time the user feedback is completed, the weight sequence would be changed according to the user interaction situation.

$$\mathrm{weight}_i = \frac{\mathrm{area}_i}{\sum_{i=1}^{n} \mathrm{area}_i} \tag{6}$$

The algorithm can be described as following.

(1) For each saliency region which the focus is confirmed, sums the pixel number where the saliency value beyond that pre-set threshold. Take the value as the area of the saliency region.

(2) If the area is too small, the region should be removed.

(3) Calculate the ratio of each saliency region area and the total area of all saliency regions to reset the weight of each saliency region.

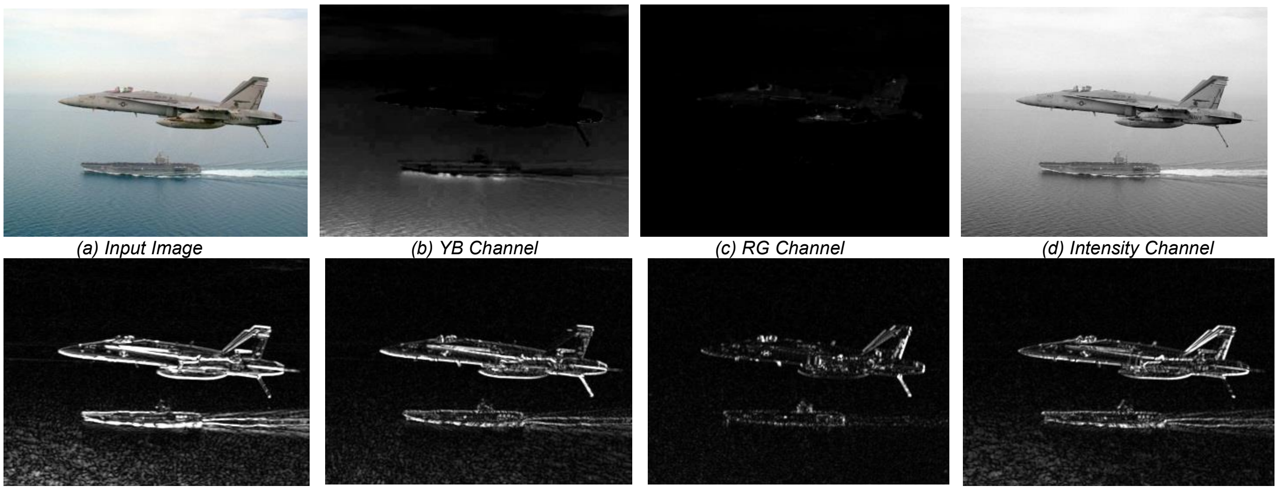
©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232

(4) When the user feedback information has been submit, the irrelevant region should be ignored in the weight sequence and the weight sequence should be reset.

(5) The saliency region with the biggest weight would be recommended as the object to be matched in the retrieval.

According to the framework in Fig.1, the weight - sorted saliency regions is the initial saliency measure for following processes in the framework. Once the user feedback is finished, the weight – sorted saliency regions will be changed to be provided as the new feature regions for the following SVM-PSO process.

Fig.5 shows the initial visual attention region extraction result based on our algorithm. The result indicates that the saliency region selection method proposed in the paper can provide a relative accurate initial value for feature matching in the retrieval.



(a) Input Image (b) YB Channel (c) RG Channel (d) Intensity Channel

(e) 0° Gabor Filtering Result (f) 45° Gabor Filtering Result (g) 90° Gabor Filtering Result (h) 135° Gabor Filtering Result

Fig.2 Feature Channels

(a) Color CM (b) Intensity CM (c) Orientations CM (d) Saliency Map

Fig.3 Component and Combination of a Saliency Map

(a) Input Images (b) Method in the paper (c) Method by Itti

Fig.4 Saliency Map Comparison

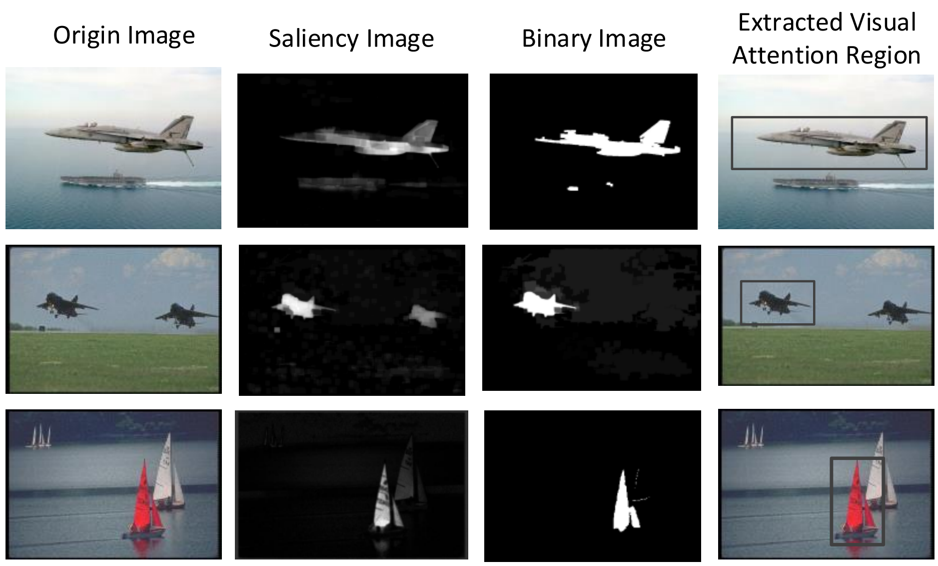


Fig.5 Experiments of Visual Attention Region Extraction

3.2 Feature Vector Extraction

Traditional low level feature (color, shape, texture) can describe global image content, but it is relative difficult to keep invariance when the illumination, size, angle of the image varies and to describe and distinguish local content of the image accurately. Hence, the paper select a local feature descriptor based on BoF to construct feature vector of the extracted object as following steps.

(1) Take SURF (Speeded Up Robust Features) as the local feature descriptor.

(2) Use clustering algorithm for each image feature descriptor and set the centroid of each cluster as visual vocabulary. K-means clustering is a usual method in BoF algorithm, which takes extreme value of target function as the criterion in iteration adjustment, uses Euclidean distance as the similarity measurement, set square error and criterion function as cluster target function, to find the optimal solution of the target function.

(3) Feature coding. That's a process of assigning a most intuitionistic visual word or multiple adjacent visual words to describe the matching degree between visual word and local feature.

According to the framework showed in Fig.1, when the object region has been changed in each iteration based on the user feedback, the feature vector of new object regions should be re-calculated as the dataset of the improved SVM-PSO-RF method in next sub-section.

3.3 Improved SVM-PSO-RF based Image Retrieval

Since an image can be described as a feature vector, all of the images in the database can be described as a vector space. Each feature vector can be looked as a particle. Then, the process of finding optimal solution via PSO algorithm is equivalent to a process of finding a positive image corresponding to optimal feature vector in feature vector space.

Because of the quick convergence and applicability for small sample set, the paper combined the SVM algorithm and PSO algorithm. After user feedback, feedback parameters would be initialized, particle fitness would be calculated, and the best position for each particle would be refreshed. The new positions can be used in SVM classifier to calculate the distance between image and classification plane, and output ranked retrieval result. The detail is presented in the following sub-sections.

3.3.1 Optimal Feature Selection in PSO Algorithm

In the feature selection for images in the database, the paper used particle evolution direction to guide the feature optimization. Firstly, supposed that user had labeled N positive images in the feedback. Then, local features of the N images would be extracted. Thirdly, average of each dimension would be calculated as initial individual optimal feature vector according to following equation.

$$\text{pBest} = \frac{1}{N_{positive}}\left(\sum_{i=1}^{N} X_i^1 + \sum_{i=1}^{N} X_i^2 + \cdots + \sum_{i=1}^{N} X_i^f\right) \quad (7)$$

Where, f is the dimension of the features, X is the particle.

Then, the fitness of particle evolution can be defined as:

$$\text{fitness} = \text{Distance}(X_i, pBest_i) \quad (8)$$

When the particle evolution matched the iteration terminal, global optimal position pBest is the optimal feature vector. Then, the relevancies between all of the positive images and corresponding feature vector in all of the dimensions would be calculated. If all of the vectors are relevant in a certain dimension, a bigger weight would be given in this dimension and vice versa. Then, the weight of feature dimension for query image and images in database can be adjusted to find optimal feature dataset for SVM classification.

3.3.2 SVM Parameters Optimization based on PSO

Although SVM algorithm is effective in the classification with high dimension and small sample set, the parameter selection is still deserved to be researched further. In the paper, an optimization for fitness function based on PSO is suggested to find optimal feature sub-set.

Fitness function is the key factor for PSO. In PSO algorithm, the space position of particle can be represented as a group of parameters in SVM, which includes kernel function parameter, error control coefficient ε, and penalty factor c. The particle fitness means the quality of training result in this group of parameters. The paper select average absolute error as fitness function as following formula.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232

$$f_{MAPE} = 1/n\sum_{i=1}^{N}\left|\left(y_i - \overline{y_i}\right)/y_i\right| \quad (9)$$

Where, N is the training sample number, y_i is the output, and $\overline{y_i}$ is the prediction result. The iteration will be terminated when the prediction error match a given value or the iteration times meet the upper limitation.

3.3.3 The workflow of PSO based SVM

Literature has indicated that when the number of positive samples after feedback isn't big enough, the SVM classifier will be not stable [28], especially when the number of positive samples and negative samples are unbalanced or positive samples are labeled by wrong. To overcome the problem, the paper take PSO to optimize the positive/negative sample number selection in the training like the method in literature [22].

Then, a workflow of PSO based SVM is presented in the paper as following steps.

(1) Extract the features in training set and test set, and calculate the number and ratio of positive sample in the test set as the prediction result.

(2) Initialize the parameters (e.g particle initiate position and velocity) of PSO based on the object extraction result and feature vectors.

(3) Set the pBest as current position.

(4) Set global optimal position gBest as current position of the particle with lowest fitness.

(5) SVM training. Calculate the fitness for every particle, refresh the pBest and gBest for particles. If current fitness is superior to pBest, then refresh pBest. If current fitness is superior to gBest, then refresh gBest.

(6) Keep refreshing particle position and velocity according to formula following until iteration time match the upper limitation or the evolution matched the iteration terminal.

Logically, the PSO based SVM method in the paper can overcome the blindness in SVM parameter selection. It's more instructive than common cut-and-try method.

4 Experiments

Experiment system is based on two datasets, which includes Caltech 101 dataset with 9146 pictures in 101 categories and self-collection dataset with 674 pictures (each picture has multiple objects) in 4 categories.

The paper select images in 10 categories (each category has 100 images) in Caltech 101 dataset to construct image feature database and select images in 10 categories (each category has 20 images) in self-collection dataset as test set.

Four experiments are designed to verify the validity of the method the paper proposed.

The first experiment is a qualitative experiment to verify that the feedback of visual attention region proposed in the paper is effective. Fig.6 shows a query image with 2 typical objects – human and traffic sign. After a feedback, the traffic sign was focused and the second retrieval result is more precise.

The second experiment is to demonstrate the initial saliency region extraction as Fig.7. Based on the feature dataset which is constructed from Caltech 101 dataset and feature vectors dataset for query images from self-collection dataset, retrieval experiments have implemented for every query image and the average precision and recall ratio are calculated as a basis for comparison in following experiments.

The third experiment is to demonstrate the effect of RF as Fig.8. Only one feedback is requested in the experiment. The experiment result is compared with the first experiment and basic SVM + Feedback. The experiment can indicate that RF is an effective mechanism for CBIR and the method proposed in the paper can achieve higher average precision than basic SVM + Feedback.

To analysis the impact of different feedback times on retrieval performance, the fourth experiment is designed. The precision of different types of image retrieving after a certain number of times feedback are collected. Fig.9 shows current experiment result based on proposed CBIR framework. The experiment can indicate that the method proposed in the paper can improve the retrieval average precision obviously (the average precision improved obviously at the first two times of feedback), and can converge quickly (the average precision change little after the second feedback).

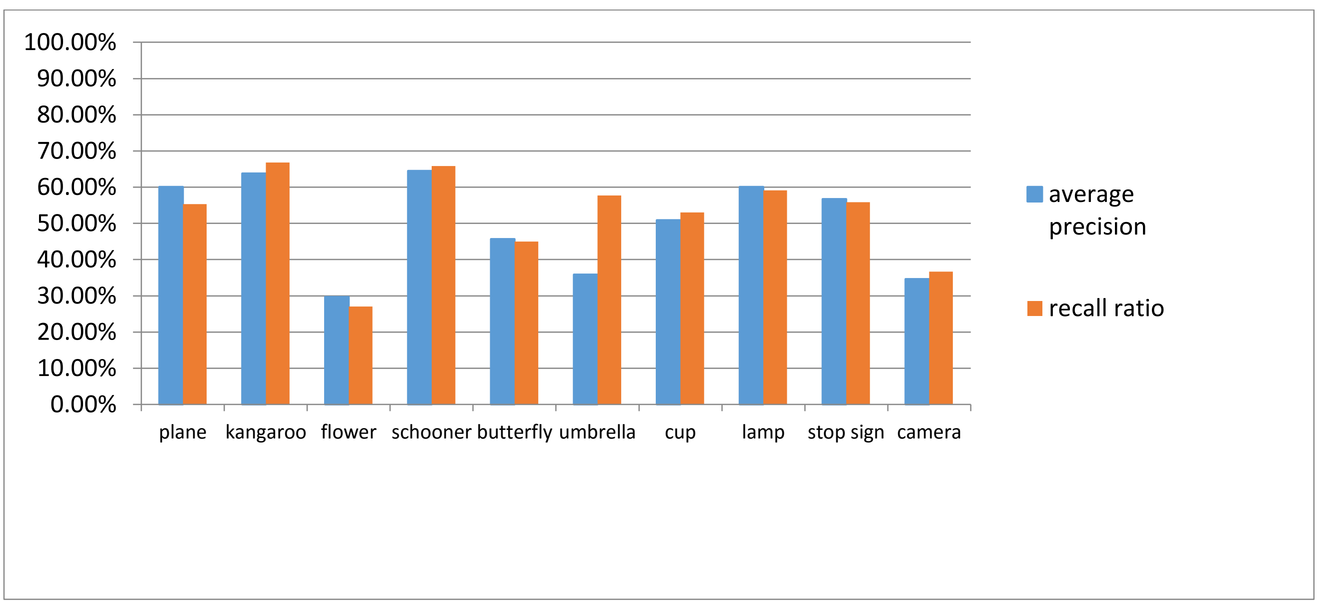
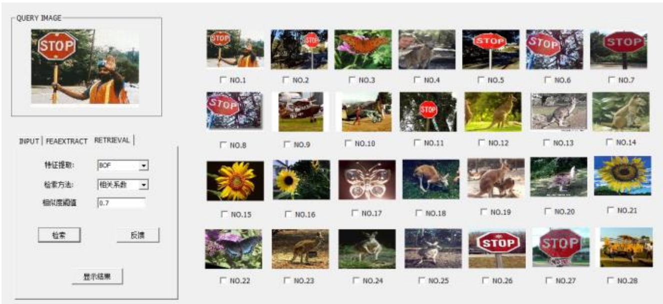
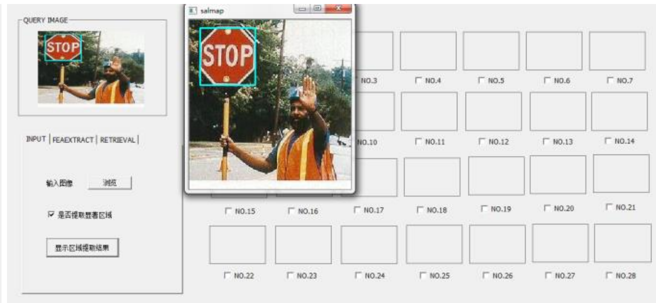


Fig. 6 Initial experiment based on object region extraction and BoF algorithm

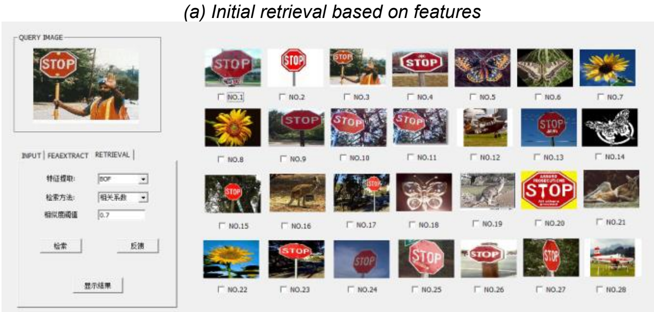
DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232



(a) Initial retrieval based on features



(b) Visual attention region feedback



(c) Retrieval result after feedback

Fig. 7 Visual attention region feedback experiment

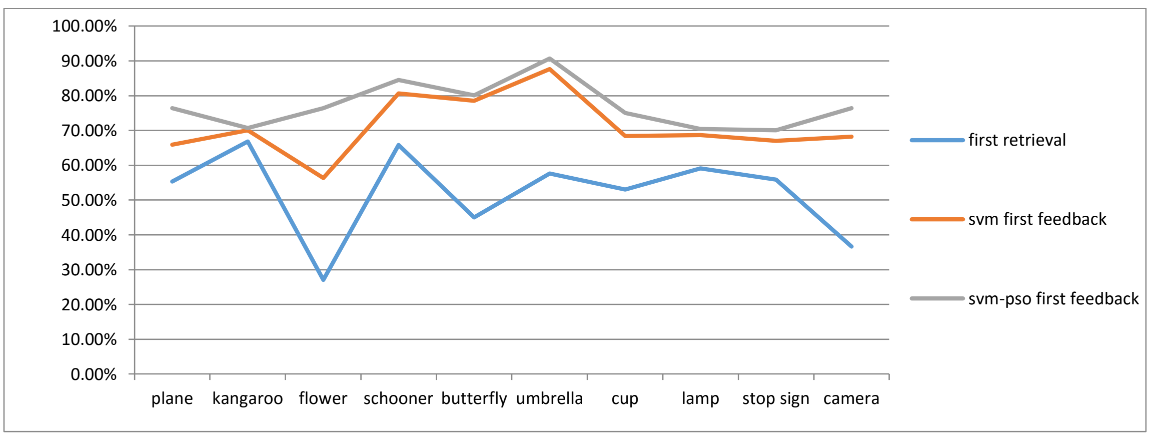


Fig. 8 Relevance feedback experiment

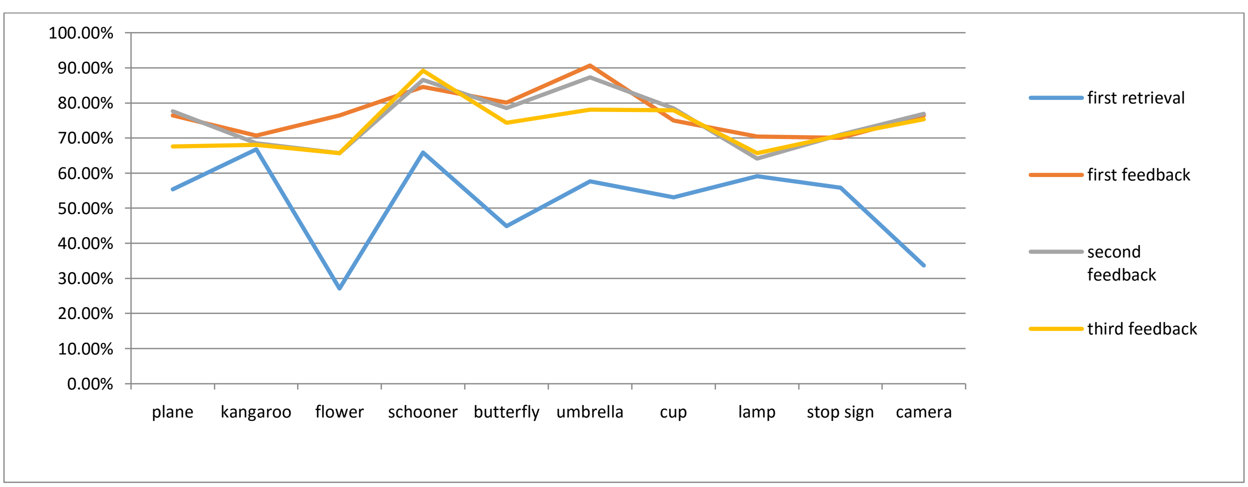


Fig.9 Experiment result of current algorithm

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232

5 Conclusion

In this paper, a novel framework for CBIR based on visual attention model and relevance feedback is proposed. The framework extends the relevance feedback to the object extraction for query image and images in database and to the retrieval algorithm based on SVM. The extension is effective in the retrieval of multiple objects in one image.

In the framework, the object extraction method based on visual attention model is improved on early visual features extraction (color feature channel), the saliency map generation through the compensation of global saliency information, and a weighted object region extraction method combined with RF is proposed. Additional, improved SVM-PSO-RF based image retrieval with the optimal feature selection in PSO algorithm, and SVM parameters optimization based on PSO is proposed and a workflow is presented in the paper.

Three experiments can verify that the method proposed in the paper is an effective, fast convergent method in CBIR, it can logically work in the image retrieval for images including multi-objects.

However, the proposed method is still deserved to be improved for multi-objects image or other complex images. More experiments are also requested to investigate and help to boost the algorithm.

Acknowledgement

This work was supported by National Natural Science Foundation of China under Grant No. 41322010.

References

[1] W. Hu, N. Xie, L. Li, X. Zeng, 2011. A survey on visual content-based video indexing and retrieval [J], IEEE Trans. Syst. Man Cybernet. Part C: Appl. Rev. 41 (6):797–819.

[2] B.A. Ceyhun, L.R. Daniel, N. Sandy, F.B. Christopher, G. Hayit, A. Burak, 2011. Content based image retrieval in radiology: current status and future directions [J], J. Digit.Imaging, 24 (2): 208–222.

[3] Gianluigi Ciocca, Claudio Cusano, Simone Santini, Raimondo Schettini, 2011. Halfway through the semantic gap: prosemantic features for image retrieval [J], Information Sciences, 181: 4943–4958.

[4] Xiang-Yang Wang, Yong-Jian Yu, Hong-Ying Yang, 2011. An effective image retrieval scheme using color, texture & shape features [J], Computer Standards and Interfaces, 33 (1): 59–68.

[5] Nishant Shrivastava, Vipin Tyagi, 2014. Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching [J]. Information Sciences, 259: 212－224.

[6] Guang-Hai Liu, Jing-YuYang, ZuoYong Li, 2015. Content-based image retrieval using computational visual attention model [J]. Pattern Recognition, 48: 2554–2566.

[7] Itti,Koch C, Niebur E, 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence [J], 20(11):1254-1259.

[8] Itti L, 2000. Models of bottom-up and top-down visual attention [D]. California Institute of Technology.

[9] Dirk Walther, 2006. Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics [D]. California Institute of Technology.

[10] Wu X, Kashino K, 2014. Interest point selection by topology coherence for multi-query image retrieval [J]. Multimedia Tools & Applications, 74(17): 7147-7180.

[11] X. Zheng, M. Zhou, X. Wang, 2007. Interest point based medical image retrieval, in: Medical Imaging and Informatics (MIMI 2007), pp. 118–124.

[12] P.S. Hiremath, J. Pujari, Content based image retrieval using color boosted salient points and shape features of an image [J], Int. J. Image Process. 2 (1) (2008)10–17.

[13] G.V. Pedrosa, C.A.Z. Barcelos, M.A. Batista, 2011. An image retrieval system using shape salience points [C], in: 2011 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2797–2800.

[14] Hong-Ying Yang, Yong-Wei Li, Wei-Yi Li, Xiang-Yang Wang, Fang-Yu Yang, 2014. Content-based image retrieval using local visual attention feature [J], J. Vis. Commun. Image R. 25: 1308–1323.

[15] Meihac C, 1999. Relevance feedback and category search in image database[C]. IEEE International Conference on Multimedia Computing and System, Florence, June 7-11, 512-517.

[16] Wu Hong, Lu Hanqing, Ma Songde, 2005. A Survey of Relevance Feedback Techniques in Content-Based Image Retrieval [J]. Chinese Journal of Computers, 28 (12): 1969 – 1979.

[17] Cox I, Miler M I, Minka T P, Yianilos P N, 1998. An optimized interactions strategy for bayesian relevance feedback[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Santa Barmara, 553-558.

[18] Wang JZ, Li J, 2002. Learning-based Linguistic Indexing of Pictures with 2-D MHMMS [J]. ACM conference on Multimedia, 45-67.

[19] Zhang Lei, Lin Fuzong, Zhang Bo, 2002. A Forward Neural Based Relevance Feedback Algorithm Design in Image Retrieval [J]. Chinese Journal of Computers, 25(7): 673-680.

[20] E. Yildizer, A. M. Balci, M. Hassan, and R. Alhajj, 2012. Efficient content based image retrieval using multiple support vector machines ensemble [J], Expert Systems with Applications, 39(3): 2385–2396.

[21] M. Broilo and F. G. B. De Natale, 2010. A stochastic approach to image retrieval using relevance feedback and particle swarm optimization [J], IEEE Transactions on Multimedia, 12(4): 267–277.

[22] Muhammad Imran, Rathiah Hashim, Abd Khalid Noor Elaiza, and Aun Irtaza, 2014. Stochastic Optimized Relevance Feedback Particle Swarm Optimization for Content Based Image Retrieval [J], the Scientific World Journal, Volume 2014, Article ID 752090, 12 pages.

[23] Xiaoxiao Wang, 2012. Research on Image Retrieval Technology based on Relevance Feedback [D]. Xiamen: Xiamen University, China.

[24] Treisman AM, Gelade G., 1980. A feature-integration theory of attention [J]. Cognitive Psychology, 12(1): 97-136.

[25] C.Koch, S.Ullman, 1985. Shifts in selective visual attention: towards the underlying neural circuitry [J]. Human Neurobiology, 4(4):219-227.

[26] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C.H. Anderson, 1994. Over-complete Steerable Pyramid Filters and Rotation Invariance [C]. Proc. IEEE Computer Vision and Pattern Recognition, pp. 222-228.

[27] Shapley R, Hawken M S, 2006. Spatial and temporal properties of cone signals in alert macaque primary visual cortex [J]. The Journal of neuroscience, 26(42):10826-10846.

DOI: 10.2352/ISSN.2470-1173.2016.2.VIPC-232

[28] Li J, Wang JZ, 2003. Automatic linguistic indexing of pictures by a statistical modeling approach [J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(9): 1075-1088.

Author Biography

Zhijiang Li received his BE in printing engineering from Wuhan University of China (1998) and his PhD in photogrammetry and remote sensing from Wuhan University (2005). Since then he has worked in the School of printing and packaging, Wuhan University. Currently, he is a Visiting Research Fellow in School of Design, University of Leeds. His work has focused on the capture, retrieval, reproduction and publishing of color images and graphs.

Chuan Dong received her BS in graphic communications（2012）and her MS in computer science and technology （2015）from Wuhan University. Since then she has worked in Internet Center at Hubei Mobile Communication Company Limited in China. Her work has focused on the application maintenance and management.

Jiaxian Long received his B.E. in printing engineering from the University of Wuhan (2015). He is now a Master of the Light Industry Technology and Engineering in Wuhan University, and a member in the lab of Visual cognition and information communication. His work has focused on the visual saliency detection and image retrieval.