Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007

Pre-classification Module for an All-Season Image Retrieval System

Hong Fu, Zheru Chi, Dagan Feng, Weibao Zou, King Chuen Lo and Xiaoyu Zhao

***Abstract* — From the study of attention-driven image interpretation and retrieval, we have found that an attention-driven strategy is able to extract important objects from an image and then focus the attentive objects while retrieving images. However, besides the images with distinct objects, there are images which do not show distinct objects. In this paper, the classification of "attentive" and "non-attentive" image is proposed to be a pre-process module in an all-season image retrieval system which can tackle both kinds of images. In this pre-classification module, an image is represented by an adaptive tree structure with each node carrying normalized features that characterize the object/region with visual contrasts and spatial information. Then a neural network is trained to classify an image as an "attentive" or "non-attentive" category by using the Back Propagation Through Structure (BPTS) algorithm. Experimental results indicate the reliability and feasibility of the pre-classification module, which encourages us to conduct further investigations on the all-season image retrieval system.**

I. INTRODUCTION

In our previous study [1], we proposed an attention-driven image interpretation method to pop out visually attentive objects from an image iteratively by maximizing a global attention function. In the method, an image is interpreted as containing several perceptually attended objects as well as the background, where each object is measured by an attention value. The attention values of attentive objects are then mapped to importance measures so as to facilitate the subsequent image retrieval. An attention-driven matching algorithm is proposed based on a retrieval strategy emphasizing attended objects. Experiments show that the retrieval results from our attention-driven approach compare favorably with conventional methods, especially when important objects are seriously concealed by the irrelevant background.

However, besides the images with distinct objects, there are images which do not show distinct objects. Examples of these two classes of images are shown in Fig. 1. The first class is the so-called "attentive image", as shown in Fig. 1(a). These images contain distinct objects, such as "flower", "human face", "statuary", etc. If one submits such an image, he/she usually wants to retrieve images with the similar objects, not caring about the background. Obviously, an attention strategy is suitable for handling these attentive images. The second class is the so-called "non-attentive image", as shown in Fig. 1(b). Different from the first category, there is no "major character" in non-attentive images. For these non-attentive images, although a set of objects/regions and the background can be obtained using the attention-driven image processing, it is difficult to determine important objects. In other words, laying emphasis on any object may lead to an undesirable retrieval result. Therefore, a retrieval strategy which fuses all the factors in the query is more suitable for non-attentive images.

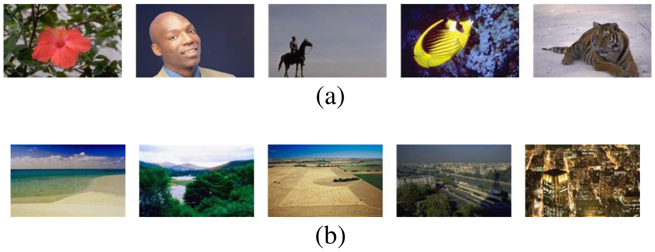


Figure 1: Two classes of images. (a) Examples of attentive images; (b) Examples of non-attentive images.

In order to tackle both types of images, an all-season system was designed as illustrated in Fig. 2. First, a pre-classification step is carried out to classify an image into attentive or non-attentive category. Then the desirable retrieval strategy is employed to perform the retrieval task. For attentive images, an algorithm emphasizing attentive objects is adopted. For non-attentive images, an algorithm that fuses all objects in the query image is used. More favorable retrieval results are expected by using this combined system.

Manuscript received Feb 12, 2007.

Hong Fu, Zheru Chi, Dagan Feng, Weibao Zou, King Chuen Lo and Xiaoyu Zhao are with Center for Multimedia Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
(e-mail: enhongfu@eie.polyu.edu.hk, enzheru@inet.polyu.edu.hk, enfeng@polyu.edu.hk, hkzouwb@yahoo.com, enkclo@polyu.edu.hk, enxyzhao@eieserver.eie.polyu.edu.hk).

Dagan Feng is also with School of Information Technologies, The University of Sydney, NSW 2006, Australia.

1-4244-1380-X/07/$25.00 ©2007 IEEE

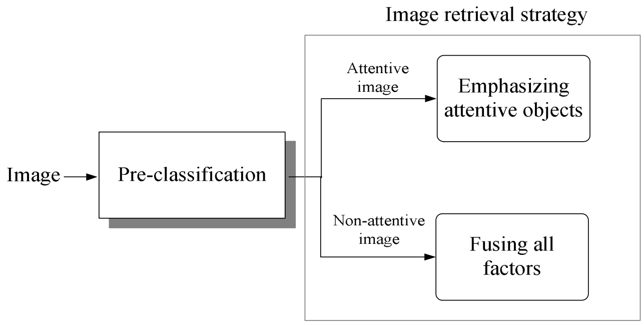


Figure 2: Block diagram of an all-season image retrieval system.

The rest of this paper is organized as follows. Section II discusses the nature of the pre-classification technique to classify attentive and non-attentive images. Then the tree structure and node features are introduced in Section III and IV, respectively. Experimental results are presented in Section V. Finally, concluding remarks are given in Section VI.

II. Nature of the Problem

Although there have been many investigations on image classification in the past decades [2-5], separating attentive images from non-attentive images is not a trivial problem. A typical image classification system is normally used to classify images with different concepts or different themes, such as "mountain", "flower", "human", "trees", etc. Visual features such as color and texture are rather effective to characterize these classes. However, our classification problem is significantly different from the others. The difficulties in our classification problem are due to: (1) there is no deterministic criterion to define two classes and it is a tough task even for human beings and (2) direct color and texture features are of little use for such a classification task. After a careful study, we found that attentive images are distinguished from non-attentive images in terms of two factors:

- Overall region/object arrangement of an image, and
- The difference between an item (region or object) and its surroundings.

Based on these two observations, we adopt a tree structure scheme to represent the overall arrangement of an image and use difference-based measures to characterize tree nodes, which will be discussed in the next section.

III. Layout of the Objects: Tree Structure Image Representation

Attentive and non-attentive images have different layouts of objects. In an attentive image, some objects often lie at dominant locations while the others are located at less important locations. For example, for the first image shown in Fig. 1(a), the flower as the main object is in the center of the image and the leaves, being the background, surrounds the flower. On the contrary, there are no dominant objects in the first image shown in Fig. 1(b). All the objects, including the sky, the sea and the sand, have a similar priority. In summary, the overall arrangement of the objects, or the structure of the image is helpful to classify attentive and non-attentive images. Therefore, we use a tree structure scheme which organizes all the objects to represent an image.

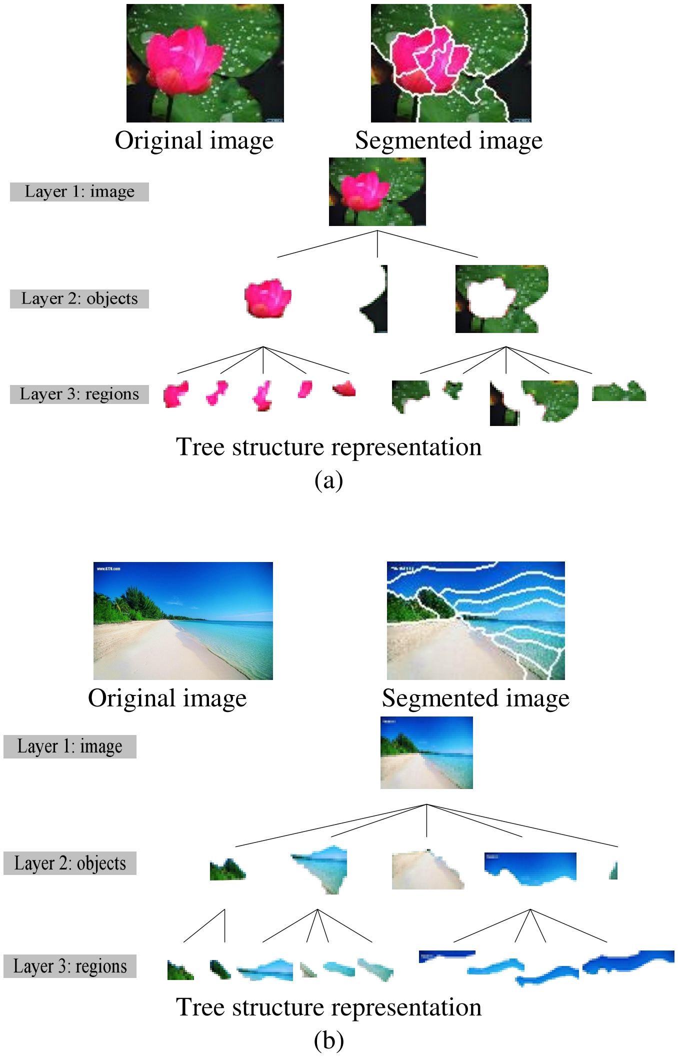


Figure 3: Examples of tree representation of (a) an attentive images and (b) a non-attentive image. (The objects in layer 2 are extracted by our attention-driven image interpretation algorithm[1]. The regions in layer 3 are obtained by JSEG [13], an image segmentation method.)

Attention-driven image interpretation is applicable to both attentive and non-attentive images. The difference is that the objects extracted from the non-attentive images may not be the important ones. Based on the result of the attention-driven image interpretation, we can construct a tree structure to represent the layout of an image. The tree structure has three layers: image layer, object layer and region layer. The bottom layer is the region layer, in which each node corresponds to one segmented region. The merged regions constitute the objects in the middle layer, in which each node represents one object, such as lotus, leaves, sand, sky, etc. Finally, the top layer is formed by combining all the objects into one whole image. This tree structure representation characterizes the overall arrangement of the objects and their regions. Two examples, including one attentive image and one non-attentive image, are shown in Fig. 3. We can see that the number of nodes in each layer is normally different for different images.

IV. Features of nodes

As we discussed before, the "direct" visual features such as color or texture cannot characterize the attentiveness of an image. Therefore, we use seven difference-based features $f_1, f_2, ..., f_7$ to characterize each node in the tree structure.

1) f_1, f_2, f_3 : relative attention values in terms of the boundary color matrix, region color matrix and texture matrix. It is similarly defined as in the relative attention value discussed in [1]. Here we separate the attentive value into three feature components, including two color components and one texture component. The relative attention value represents the saliency between an item (object or region) and its surroundings, which are helpful for detecting attentive patterns.

2) f_4, f_5 : normalized location of an item. f_4 and f_5 are the relative center coordination of an item. We define the bottom-left corner of an image as (0, 0) and the top-right corner as (1, 1). The items near the center of an image are usually related to attentive objects while the items near the boundary of an image might be less attentive.

3) f_6 : normalized area of an item. The area of the whole image is defined as 1. f_6 is the area of an region as a fraction of the whole image. Very small or very large items might not be an attentive object. The items of a reasonable size are more likely to be an important object.

4) f_7 : normalized length of outer-image boundary of an item.

$$f_7 = \frac{\text{length of outer - image boundary of the item}}{\text{length of outer boundary of the item}}$$

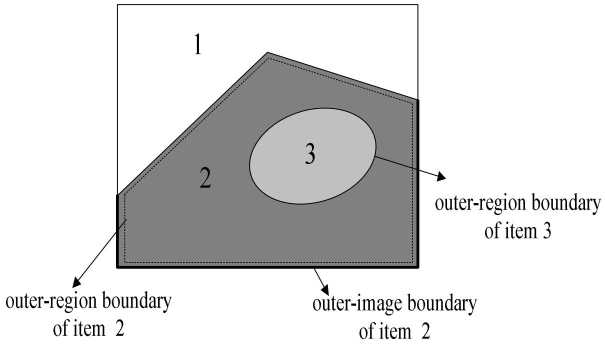

1
2
3
outer-region boundary of item 3
outer-region boundary of item 2
outer-image boundary of item 2

Figure 4: Illustration of feature f_7 : relative length of the outer-image boundary of items.

As illustrated in Fig. 4, f_7 of item 3 is zero and f_7 of item 2 is a fraction between 0 and 1. Based on our observations, when one takes photos, one would like to include the whole contour of an important object in the picture as much as possible. So the items with smaller f_7 's are more likely to be important objects.

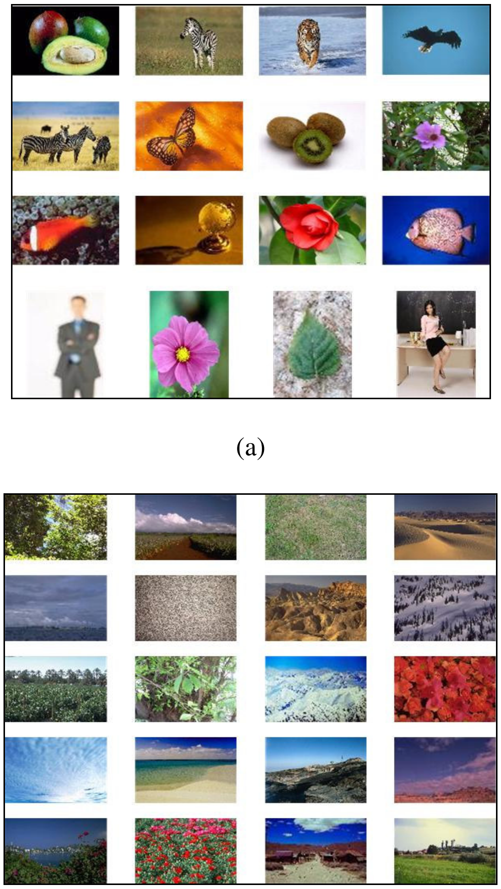
V. Experiments and discussion

The tree representation of attentive and non-attentive images is of an adaptive nature, meaning that the tree structure varies in different images. In order to classify attentive and non-attentive images, a special neural network and its training algorithm, called "Back Propagation Through Structure (BPTS)", that can handle adaptive structural patterns is employed. After it was first proposed by Goller and Kuchler in 1996 [6], several researchers have contributed great efforts to further enhance the training algorithms [7,8] and apply it to solve various classification problems [9-12]. It is recognized that this neural network is able to generalize both the node features and the structural information encoded in the tree representation.

Table 1: Training and test results with different numbers of hidden nodes

Number of hidden nodes	Classification rate (%)	
	On training set	On test set
5	84.6	84.4
10	85.9	82.4
15	85.6	84.9
20	86.5	84.7
25	86.5	84.4
30	86.0	83.9
35	86.8	84.8

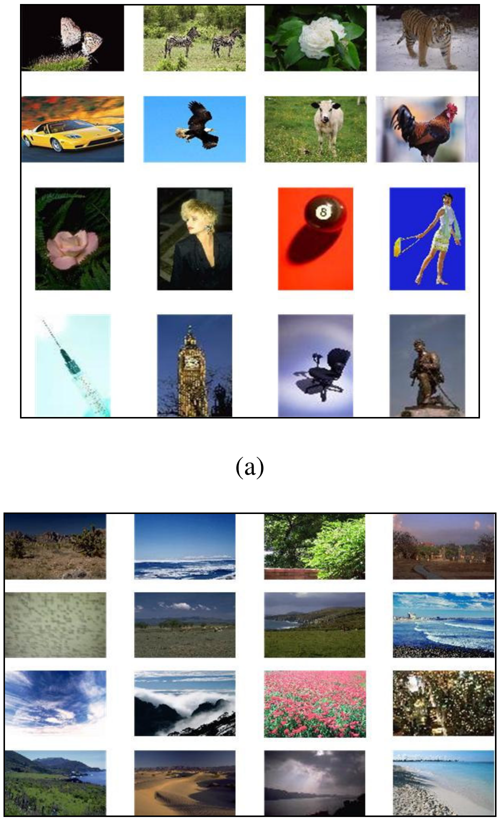
We prepared 500 attentive images and 500 non-attentive images for training the neural network. Other 756 images including 378 attentive images and 378 non-attentive images are used as test samples. Some examples of the training and test images are shown in Fig. 5 and Fig. 6, respectively. During the training process, the tree structure is traversed by a three-layered perceptron, whose inputs contain both the features of the current node and the outputs from the child nodes if any, as shown in Fig. 7. The target output of the perceptron is a two-dimensional vector $Y = [y_1 \quad y_2]^T$ with $Y = [1 \quad 0]^T$ representing the attentive class and $Y = [0 \quad 1]^T$ representing the non-attentive class. The learning process is done with the BPTS algorithm. In this investigation, different numbers of hidden nodes were tested in order to determine the size of network. Experimental results are given in Table 1, which shows that the performances are not sensitive to the number of hidden nodes. We chose a network with 15 hidden nodes in the experiment.



(a)

(b)

Figure 5: Some training samples of (a) attentive and (b) non-attentive images.



(a)

(b)

Figure 6: Some test samples of (a) attentive and (b) non-attentive images.

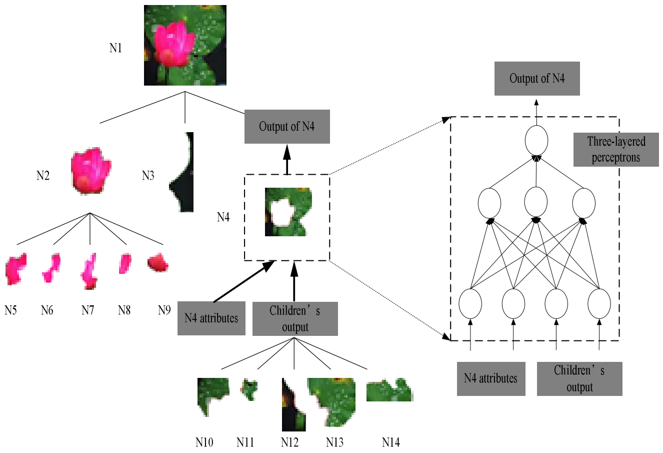


Figure 7: Illustration of the tree structure encoding with a three-layered perceptron.

VI. CONCLUSION AND FUTURE WORK

Based on the study of attention-driven image interpretation and retrieval, we have shown that an attention-driven strategy is able to extract important objects from an image and then focus on the attentive objects while retrieving images. In this paper, the classification of “attentive” and “non-attentive” image is proposed to be a preprocessing module in an all-season image retrieval system which can tackle both attentive and non-attentive images. In this pre-classification module, an image is represented by an adaptive tree structure with each node carrying normalized features that characterize the object/region with visual contrasts and spatial information. Then a neural network is trained to classify an image as an “attentive” or “non-attentive” category by using the Back Propagation Through Structure (BPTS) algorithm. Experimental results indicate the reliability and feasibility of the pre-classification module. Future work will focus on the implementation of the all-season image retrieval system.

ACKNOWLEDGMENTS

We are grateful for the support from a Polyu grant (Project No.: 1-BBZ9) and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China. (Project No. PolyU 5229/03E).

REFERENCES

[1] H. Fu, Z. Chi, and D. Feng, “Attention-Driven Image Interpretation with Application to Image Retrieval,” Pattern Recognition, Vol. 39, No. 9, pp. 1604-1621, 2006.

[2] Chapelle, O., Haffner, P., Vapnik, V.N. (1999). “Support Vector Machines for Histogram-Based Image Classification,” IEEE Transactions on Neural Networks, Vol. 10, Issue 5, Sept, pp. 1055-1064.

[3] Delopoulos, A., Tirakis, A., Kollias, S. (1994). “Invariant Image Classification Using Triple-Correlation-Based Neural Networks,” IEEE Transactions on Neural Networks, Vol. 5, Issue 3, May, pp. 392-408.

[4] Etemad, K. and Chellappa, R. (1998). “Separability-based Multiscale Basis Selection and Feature Extraction for Signal and Image Classification,” IEEE Transactions on Image Processing, Vol. 7, Issue 10, Oct., pp. 1453-1465.

[5] Li, J. and Wang, J. (2003). “Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 10, October, pp. 1-14.

[6] Golloer, G. and Kuchler, A. (1996). “Learning Task-Dependent Distributed Representations by Backpropagation through Structure,” IEEE International Conferences on Neural Networks, pp. 347-352.

[7] Cho, S.Y. and Chi, Z. (2005). “Genetic Evolution Processing of Data Structures for Image Classification,” IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 2, February, pp. 216-231.

[8] Cho, S.Y., Chi, Z., Siu, W.C., and Tsoi, A.C. (2003a). “An Improved Algorithm for Learning Long-Term Dependency Problem in Adaptive Processing of Data Structures,” IEEE Transactions on Neural Networks, Vol. 14, No. 4, July, pp. 781-793.

[9] Cho, S.Y., Chi, Z., Wang, Z., and Siu, W.C. (2003b). “An Efficient Learning Algorithm for Adaptive Processing of Data Structure,” Neural Processing Letters, Vol. 17, No. 2, April pp.175-190.

[10] Frasconi, P., Gori, M. Sperduti, A. (1998). “A General Framework for Adaptive Processing of Data Structures,” IEEE Transactions on Neural Networks, Vol. 9, pp. 768-785.

[11] Wang, Z., Chi, Z., Feng, D., and Tsoi, A.C. (2003). "Content-Based Image Retrieval with Relevance Feedback Using Adaptive Processing of Tree-Structure Image Representation,” International Journal of Images and Graphics, Vol. 3, No. 1, pp. 119-143.

[12] Zou, W., Lo, K.C. and Chi, Z. (2006). “Structured-Based Neural Network Classification of Images Using Wavelet Coefficients,” Proceedings of Third International Symposium on Neural Networks (ISNN), Lecture Notes in Computer Science 3972: Advances in Neural Networks-ISNN2006, Springer-Verlag, Chengdu, China, May 28-June 2, part II: pp. 331-336.

[13] Y. Deng, and B. Manjunath. (2001). “Unsupervised Segmentation of Color-Texture Regions in Images and Video”, IEEE Transactions on Patten Analysis and Machine Intelligence, Vol. 23, No. 8, pp. 800-810.