MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Headdress Detection Based on Saliency Map for Thangka Portrait Image

Lu Yin
School of Mathematics and Computer Science
Northwest University for Nationalities
Lanzhou, China.
E-mail: superthumb@163.com

Weilan Wang*
School of Mathematics and Computer Science
Northwest University for Nationalities
Lanzhou, China.
E-mail:weilanchina@gmail.com

Abstract

This paper applies the saliency map of attention model on Thangka image detection for the first time. Making contrast with some detection methods, an automatic detection algorithm is proposed to segment the headdress area of buddha in Thangka portrait image. Firstly, attention model is applied to calculate salient value of each pixel. Then, through using the detection method for back-light area of buddha, a new algorithm which fits for Thangka headdress is proposed. Meanwhile, this research also gives new thought to automatic image segmentation, not only for Thangkaka image, but also for other kinds of research objects. And, experiments show that the detection result is effectively approached to our purpose.

1. Introduction

Image detection and segmentation are important research direction in image processing field. As the pretreatment processing of high level image applying, such as image recognition and semantic image retrieval, it also captured lots of attention during the past decades. Meanwhile, many mature models have been proposed to solve this task, including graph cut[1], neural net, clustering algorithm and so on. The trend of image segmentation is developing from low level to high level, tough to accuracy, manual to automation. However, there is not any method or algorithm can achieves perfect segmentation result and suits for all kinds of images yet.

A model of attention mechanism based on human vision was proposed by Lauren Itti [2] in 1998. This model used control strategy based on bottom-to-up method and applied "feature integration theory"[3] to explain and analyze human visual research strategies. Recently, the theory of attention model is effectively used for natural images and is applied to detect and recognize natural objects. However, as we know, Thangka image has plentiful features itself, such as colorful visual effect, complicated background, irregular shape, and subtle painting[4]. So, how to use classical model to detect the salient area in Thangka image is the key point of the paper.

2. Related Work

Thangka image is a valuable artistic painting of Tibetan history and have plenty of religion information which are very useful for people. For example, headdress, gesture, religious tools, religious seats are major research objects and also defined as interested area and key point for image recognition and retrieval. During recent years, researchers used Thangka as the studying object and made some research outcomes in the area of image processing and pattern recognition. For example, the creation of Thangka image database, the digital image inpainting of Thangka [5] and Thangka image semantic retrieval[6,7,8].

Contrasting with other features of Thangka image, headdress is more conspicuous and important to research. In some earlier research, headdress of Thangka image has already been chosen for the interested area as showed in[7]. Firstly, it delimitated headdress in a small area by manual way, and then extracted features and realized image recognition by using matching algorithm of contour feature.

However, as the development of digital Thangka image research, especially in high level semantic retrieval field, manual segmentation is not good enough for the image preprocessing. Therefore, this paper proposed an automatic image segmentation and detection algorithm which preferably satisfied the need.

3. Saliency Map Based on Attention Model

Lauren Itti proposed a visual attention system, inspired by the behavior and neural architecture of the early primate visual system, and called widely attention. This system built several sets composed by pixels which have conspicuous feature in the area where the eyes of human beings interested in. These sets of pixels called saliency map. It could detected and extracted the interested area by searching salient pixels in saliency map. Contrasting with some traditional algorithms which need people to define interested area by manual way, this model has great ad-vantages in the field of image processing and pattern recognition.

The Itti saliency map model applied a bottom-to-up attention mechanism[9]. Firstly, it constructed a Gaussian pyramid of nine spatial scales by using low-pass filter and subsampling the input image. And then, the model proceeded nonuniform sampling on each scale of the pyramid. By extracting three visual features including intensity, color and orientation, this model applied "center-surround" operations to obtain corresponding feature of saliency map under different scales, and combined multiple feature maps into a whole one. At last, Itti model used a mechanism called "winner take all" to detect the conspicuous area.

Although Itti model is close to the visual mechanism of human, but the calculation and time complexity is heavy for computer system and difficult to realize. Hence, there is another saliency map algorithm based on contrasting theory[10] was proposed. It simply extracted color feature in LUV space, and calculated saliency map in only one scale. This model has low complexity but lacks of accurate salient value of pixels. Thereby, this paper used an algorithm combined the advantages of the two models to detect our object. Firstly, we constructed a three scales Gaussian pyramid. And then, we used contrasting method and “center-surround” theory inside each scale and made integration operations between three scales. At last, every pixel value in the final saliency map represented the conspicuous feature that akin to the visual mechanism of human beings.

3.1. Feature Extraction of Intensity and Color

First of all, making preprocessing of input images and down sampling them into 1/4 and 1/16 of the original size. Then, building the Gaussian pyramid by combining the three images of different scales.

I(x) is defined as the intensity value of each pixel. So, formula (1) is used to finish gray process and compute the intensity of pixels.

$$I(x) = 0.299r(x) + 0.587g(x) + 0.114b(x) \quad (1)$$

Where, x is current pixel and r, g, b indicate the three color channels of image respectively.

After that, in order to imitate the color visual mechanism, we used the r, g, b value of a pixel to creat four color channels, which include R(x), G(x), B(x) and Y(x) indicate red, green, blue and yellow respectively.

$$\begin{aligned} R(x) &= r(x) - [g(x) + b(x)]/2 \\ G(x) &= g(x) - [r(x) + b(x)]/2 \\ B(x) &= b(x) - [g(x) + r(x)]/2 \\ Y(x) &= [r(x) + g(x)]/2 - |r(x) - g(x)|/2 - b(x) \end{aligned} \quad (2)$$

Through realizing the contrasting method, a neighborhood window was built for each pixel to complete feature contrast inside every scale of Gaussian pyramid. The size of neighborhood window is 3×3.

As shown in formula (3), the paper applied “center-surround” theory to imitate the “double-opponent channels” system in human visual pallium.

$$\begin{aligned} RG(x,y) &= |(R(x) - G(x)) - (R(y) - G(y))|/2 \\ BY(x,y) &= |(B(x) - Y(x)) - (B(y) - Y(y))|/2 \end{aligned} \quad (3)$$

Where, x is current pixel, and y is one of the pixels in neighborhood window. RG is the double-opponent feature between red and green, while BY is the double-opponent feature between blue and yellow.

Consulting by researcher[11], the contrasting value of color is defined as following formula (4):

$$C(x,y) = \sqrt{\eta_{RG}^2 RG^2(x,y) + \eta_{BY}^2 BY^2(x,y)} \quad (4)$$

Where, ŋRG and ŋBY represent the weight of two double-opponent features. And the specific value of ŋRG and ŋBY are calculated by formula (5).

$$\eta_{RG} = \frac{R(x)+R(y)+G(x)+G(y)}{R(x)+R(y)+G(x)+G(y)+B(x)+B(y)+Y(x)+Y(y)} \quad (5)$$

$$\eta_{BY} = \frac{2\sqrt{B(x)^2+B(y)^2+Y(x)^2+Y(y)^2}}{3\times 255}$$

Meanwhile, the contrasting value of intensity is defined as following formula (6).

$$I(x,y) = |I(x) - I(y)| \quad (6)$$

From above, combining the features obtained from formula (4) and formula (6), the integration contract value of color and intensity can be calculated by formula (7):

$$S_{CI}(x,y) = \sqrt{C(x,y)^2 + I(x,y)^2} \quad (7)$$

3.2. Feature extraction of orientation

Orientation feature is important information for the shape of image. This paper consulted Itti model and chose Gabor filter to calculate orientation feature from 4 directions, include $\{0, \pi/4, \pi/2, 3\pi/4\}$. Then, we obtained 4 orientation feature images. For each pixel, the salient value can be calculated by formula (8):

$$S_O(x,y) = \sum_{\theta \in \{0,\pi/4,\pi/2,3\pi/4\}} O(x,y,\theta) \quad (8)$$

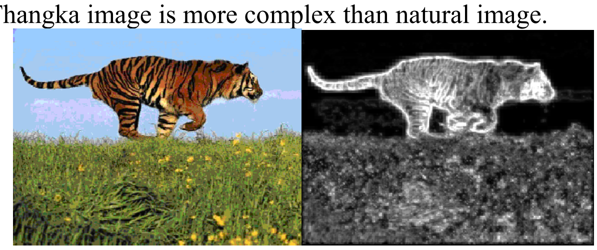
3.3 Integration of multiple features

After defining the three features contain intensity, color and orientation, this model used formula (9) to calculate the final salient value for every pixel.

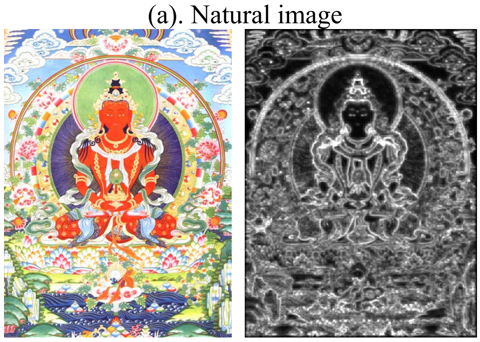
$$SP(x) = \sum_{L=1}^{3} \sum_{y \in \theta} (\gamma_{CI} S_{CI}^L(x,y) + \gamma_O S_O^L(x,y)) \quad (9)$$

Where, $L \in \{1,2,3\}$ indicates there are three scales in Gaussian pyramid. We computed sub-saliency-map on each scale, and integrated them together to get the final saliency map. Here, θ is neighborhood window of x, while γ_{CI} and γ_O are weight coefficients for intensity color feature and orientation feature respectively. And, they are set to 1, $\gamma_{CI} = \gamma_O = 1$.

In summary, the easier that one area can excite people’s attention, the bigger salient value it will obtain, so this area should be brighter than others. In our experiment, two images are chosen to display the effect of saliency map as shown in Figure 1. It is easy to find that the saliency map of tiger in (a)’s right is much brighter than sky, and headdress part in the saliency map of Thangka image in (b)’s right is brighter, but of course the saliency map of Thangka image is more complex than natural image.



(a). Natural image



(b). Thangka image

Figure 1. Effect figure of Saliency map

4. Detection of Salient Area

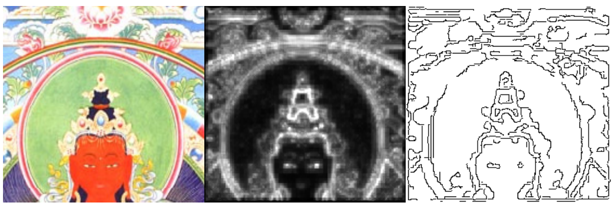
The main purpose of the paper is to extract headdress area from Thangka image. Headdress of Thangka has plentiful religion information and important research value.

Recently, there are two main categories of methods for the detection of salient area. First one is the famous "Winner take all" theory in Itti model. It combined neural net system and defined the pixel with maximum salient value as the FOA(focus of attention). Then, Itti model considered the FOA pixel and its corresponding conspicuous area is the winner, and should be chose as the salient area. But, as shown in Figure1 (b), Thangka salient map is so conspicuous and bright, so that there are many pixels achieved the maximum salient value—255 at the same time. Thereby, it is difficult to make sure which area is the winner.

The second category for detection is proposed by using algorithm[12]. This model used canny operator to detect all edges of saliency map at first, then applied a weight formula to compute the prime edge in the image.

$$SE(e_i) = \lambda_L L(e_i) + \lambda_S SP(e_i), i = 1 \ldots N \quad (10)$$

In this detection formula, N is the number of edges, e_i represents each edge created by canny operator, i=1…N. $L(e_i)$ and $SP(e_i)$ are the length and salient value of every e_i . So, this formula chose the length and salient value as features to detect prime edges. Coefficient λ_L and λ_S are 0.3 and 0.7 respectively. Obviously, the bigger the value of $SE(e_i)$ is, the higher the probability which ensure e_i become prime edge will be. However, as we mentioned before, the image content of Thangka is very rich and vivid, so that the edges will be too disordered and complex to confirm that which canny edge is the prime one. Contrast figure is shown in Figure 2.



(a)Original headdres (b)Saliency map (c)Canny edges

Figure 2. Canny edges of Thangka saliency map

Therefore, this paper proposed a new algorithm to extract headdress area that fit for Thangka image itself. The idea of this algorithm is inspired by the back-light area of Thangka headdress. Back-light of headdress in Thangka image contains abundant religion information. It indicates the sanctity and status of buddha. Generally, it locates in the upper-middle field of Thangka image and its salient value is lower than the surrounding areas. So, we decide to detect headdress through finding the area of Thangka back-light.

Firstly, the algorithm automatically delimitate a general area which can contain headdress area inside. Then, applying salient map method to detect back-light area and eliminating redundant image information. The specific process is described in the following algorithm.

Algorithm description:

Step 1. Automatically cutting a rough frame from image and ensure this area can cover headdress.

Step 2. Normalizing headdress image into regular size and making partition processing, each size of subblock is defined as 3×3.

Step 3. Calculating T1 (average salient value of all pixels) inside every subblock, and painting the subblock whose T1 is lower than 63 into green.

Step4. Detecting the back-light area of headdress:

Here, this step defined 4 rules for back-light detection. By matching the four rules one bye one, we chose the final area as our back-light if it satisfied all rules.

Rule 1. The area should contains 140 subblocks at least.

Rule 2. The area satisfied rule1 can not contain any subblocks which are on the outside layer edge of saliency map. If all candidate areas are failed to satisfy rule2, we delete the outside layer edge and turn back to rule1 to rejudge all areas again.

Rule 3. The area should have symmetry feature. (The subblocks of one area should locate both on the two sides of middle axis uniformly).

Rule 4. If the number of candidate areas are still more than 1, then, the area with most subblocks win and will be chose as the back-light area of headdress.

Step 5. Scanning subblocks of saliency map from bottom to up row by row and marking the pixels which contained by the back-light area. Displaying all these marked pixels in original Thangka image correspondingly and painting other pixels into white.

5. Experiment Result and Analysis

Through experiments and testing, finally, all thresholds involved in the algorithm that described above is obtained. Besides some thresholds which have already given, we define HI and WI is the height and width of image respectively. Then, the boundaries of headdress frame in Step1 are shown in table1. Meanwhile, normalization size of headdress frame and the number of subblocks are also mentioned.

Table 1. Threshold table

Title	Threshold
Left boundary	4/15HI
Right boundary	11/15HI
Up boundary	29/30WI
Bottom boundary	2/3WI
Normalization size	150×120
Subblock size	3×3
Number of subblocks	2000

Finally, three Thangka images are picked to show the whole process of headdress detection. There are five images in every process. As shown in Figure 3, including original image, headdress area, headdress saliency map, back-light detection and final detection result from the top down respectively.

6. Conclusion

This paper offered a new thought for image segmentation and interested area detection. The new detection algorithm based on gray image which calculated by saliency map method. It is good at detecting the key object in a relatively simple area which around by complex background. And, it also can be applied into other research field, such as the recognition of traffic sign, license plate of car, the detection of human face and eyes.

This is the first time we adopt attention model to detect interested area of Thangka image. It is pleasant to see that the experiments could achieve automatic process by using detection algorithm of headdress. How to find the back-light of buddha is very important step. Although we make partition process which works this problem out by setting several rules, it also brings some affiliated results that do not satisfy us. For example, the final edge of headdress area seems not smooth because of partition processing. So, how to improve the segmentation effect is the new focus for our further research in the future. Besides, how to extract features of headdress to realize automatic recognition based on the detection method proposed by this paper is also a focus of our attention.

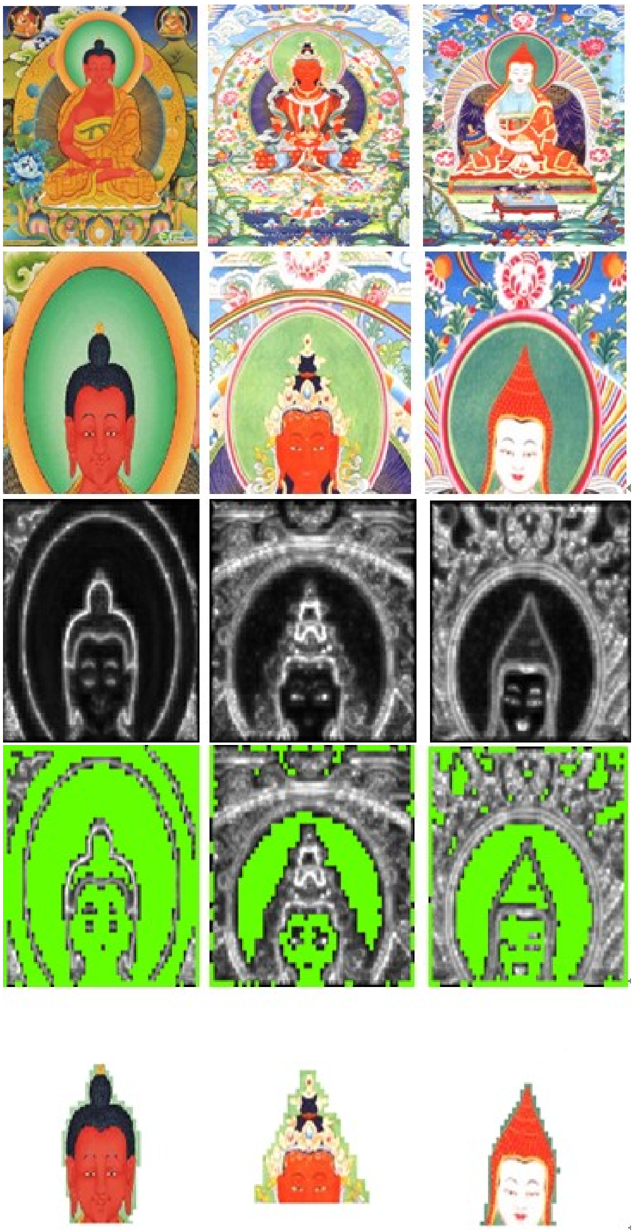


Figure 3. Process of headdress detection

* Weilan Wang, Corresponding author. Email:weilanchina@gmail.com

ACKNOWLEDGEMENTS

This work was supported in part by the Nature Science Foundation of China under Grant No.60875006, and Science and Research Foundation of Gansu province under Grant No. 096RJZA112.

References

[1] Yuri Y. Boykov, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images", Proceedings of "Internation Conference on Computer Vision", Vancouver, Canada, July 2001.

[2] Itti L, Koch C, Niebur E. "A model of saliency based visual attention for rapid scene analysis". IEEE Trans. On Pattern Analysis and Machine Intelligence, 1998,20(11): 1254-1259.

[3] Itti L, Koch C. "Computational modelling of visual attention". Nature Reviews Neuroscience, 2001,2:pp.194-202.

[4] Lu Yin, Weilan Wang, Danchun Yang. "Study on how to distinguish Thangka and non-Thangka image". International MultiConference of Engineers and Computer Scientists 2010 Vol II. Hong Kong. March, 2010. pp 1476-1480.

[5] Huaming Liu, Weilan Wang and Hui Xie. "Thangka Image Inpainting Using Adjacent Information of Broken Area". International Multi Conference of Engineers and Computer Scientists 2008. Hong Kong, March, 2008. pp.646-649.

[6] Weilan Wang. "Study of Thangka Image Retrieval and Multimedia Presentation Management System". Fifth In-ternational Conference on Intelligent Information Hiding and Multimedia Signal Processing, Japan 12-14 September 2009, pp.981-984.

[7] Jianjun Qian, Weilan Wang. "Main Feature Extraction and Expression for Religious Portraint Thangka Image". The 9th International Conference for Young Computer Scientists, 18-21 Nov,2008, pp.803-807.

[8] JianjunQian, Weilan Wang. "Religious portrait Thangka image retrieval based on gesture feature". Chinese Conference on Pattern Recognition, 4-6 Nov,2009, pp.568-572.

[9] Itti L, Koch C. "Feature combination strategies for saliency-based visual attention systems". Journal of Electronic Imaging, 2001,10(1):pp.161-169.

[10] Ma Y F, Zhang H J. Contrast-based image attention analysis by using fuzzy growing. In: proc. of Int. Conf. on ACM Multimedia (ACM Multimedia 03), Berkeley, CA, USA, Nov,2003: pp.374-381.

[11] Yaoru Sun, Fisher R. Object-based visual attention for computer vision. Artificial Intelligence, 2003,146: pp.77-123

[12] Songhe Feng. "Research on Perception Oriented Image Retrieval and Automatic Image Annotation". Dissertation. Beijing Jiaotong university.2009. (In Chinese)