International Journal of
Soft Computing And Software Engineering (JSCSE)
ISSN: 2251-7545
Vol.1,No.1, 2011
DOI: 10.7321/jscse.v1.n1.7

A Novel Classificatory Filter for Multiple Types of Erotic Images Based on Internet

* Xu Zhang[1], Baolong Guo[2], Yunyi Yan[3], Chunman Yan[4]

School of Electro-mechanical Engineering, Xidian University, Xi'an, China
[1]*xuzhang@mail.xidian.edu.cn*
[2]*blguo1199@163.com*
[3] *yyyan@xidian.edu.cn*
[4]*yancha02@163.com*

Abstract. **In the condition of which various types of erotic images are available on internet, so big an image database make the retrieval precision and recall rate stay low. Considering the fact that the present erotic image filtering algorithms has a higher false-positive rate when the image is of dark brightness and color distortion, a kind of classificatory filter for multiple types of erotic images on Internet was proposed. The filter consists of three parts: skins-color model matching layer, dark erotic image detecting layer and a decision-making layer based on SVM. The experiment results show that the model can simply realization, overcome the shortness of the traditional methods, effectively improve the detection correction, and fit the requirements of application on Internet.**

Keywords: Image processing; Image retrieval; Erotic Image filtering;

1. Introduction

Nowadays, with the rapid progress of the network, more and more people acquire the information from the internet. This has been an indispensable part of daily life. But at the same time, some information such as erotic images in the internet has more harm to the children. So we must research and develop efficient tools to detect and filter out erotic images.

The image whose content include human erotogenic zone is defined to be erotic. Currently, by technology of computer vision and image understanding, such as the skin color models, the naked skin color regions' segmentation, the geometric feature detection and texture content [1-4], we can detect and filter some erotic images.

In 1966, Fleck has researched on erotic images' identification using the technology of computer vision and image understanding. The naked skin color area's segmentation was used to determine whether an image contained any erotic information. At present, the methods such as skin color model, texture content, posture judgment have been widely used in the identification of erotic images and obtained a higher recognition rate. Yiding Wang et al using the method of constructing the mixture skin color model in different color space to detect naked skin color regions [5], M. Felek et al combine color and texture properties to obtain an effective mask for skin regions [6]. Zhouyao Chen et al proposed a fusion algorithm to combine the results obtained from text as well as image classification [7]. Xuanjing Shen et al presents an algorithm based on the face and trunk detection, and through the recognizing the erotogenic-part in the image, implemented the algorithm and construct a simple and fast pornographic images classifier system based on a decision tree[8].

Using the color features, shape features, and threshold analysis algorithms, these methods change the traditional train of thought of the filtering method. But on real internet, because of that the types of

International Journal of
Soft Computing And Software Engineering (JSCSE)
ISSN: 2251-7545
Vol.1,No.1, 2011
DOI: 10.7321/jscse.v1.n1.7

normal images are very complex, they still stay a low correction when the image is of a dark brightness or color distortion in some complex illumination conditions. In this paper, we introduce a novel classificatory filter for multiple types of erotic images based on Internet, which is consist of the skin-color model matching layer, the dark erotic image detecting layer and the decision-maker layer. In Fig.1, the structure has shown that our filter consists of three parts: skin-color model matching layer, dark erotic image detecting layer and weighted decision mechanism. Detailed description will be introduced in the following sections.

Figure 1. Filter structure

2. Skins-Color Model Matching Layer

This layer is to find naked skin color regions. Generally, there are two kinds of methods: pixel-based method and region-based method. In order to obtain a high calculate speed, we take the pixel-based method in this layer. The procedures are as follows:

1) Following the equation (1), image should be normalized firstly [5, 15]. Normalized image is the basis for establishment of skin-color model.

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B} \tag{1}$$

2) Transform the normalized RGB color space to YIQ color space if the pixels meet the demand of equation (2) and we can get vector I with equation (3).

$$\begin{cases} 0.333 < r < 0.664 \\ 0.246 < g < 0.398 \\ r > g, g \geq 0.5 - 0.5r \end{cases} \tag{2}$$

$$\mathrm{I} = 0.596\mathrm{R} - 0.275\mathrm{G} - 0.321\mathrm{B} \tag{3}$$

3) Transform the normalized RGB color space to YUV color space [16] with equation (5) if the pixels meet the equation (4).Then, judge whether θ meets the demand of equation (6).

$$20 < I < 70 \tag{4}$$

$$\theta = |V/U| * 180/3.14 \tag{5}$$

International Journal of
Soft Computing And Software Engineering (JSCSE)
ISSN: 2251-7545
Vol.1,No.1, 2011
DOI: 10.7321/jscse.v1.n1.7

$$100 < \theta < 150 \tag{6}$$

If pixels satisfy the conditions (2), (4) and (6), we can define them as skin-color pixels and get a skin-color array.

4) Pick out the skin-color area and set other pixels to 0. Then we denote the sum of original pixels by on and skin-color area pixels by sn and calculate the value of ar which represent the proportion of skin-color area in original image by the equation (7).

$$ar = \text{int}[(sn / on) \times 100 + 0.5] \times 100\% \tag{7}$$

5) If $ar > 15\%$, based on the shape characteristic of human body, we will select the minimum enclosing rectangle of the skin-color area by maximum diameter, which is the max distance among the pixels in skin-color area.

$$\begin{cases} d_{\max} = \max\{d \mid d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\} \\ (x_i, y_i), (x_j, y_j) \in f(x, y) \end{cases} \tag{8}$$

We set image as $f(x, y)$. Then, (x_i, y_i) and (x_j, y_j) represent for coordinates of pixels i and j respectively. Maximum diameter is keep to the equation (8).

6)Calculate the ar again to get a new ar value. If $ar > 20\%$, we will do face detection. The specific procedure is introduced in reference [10].

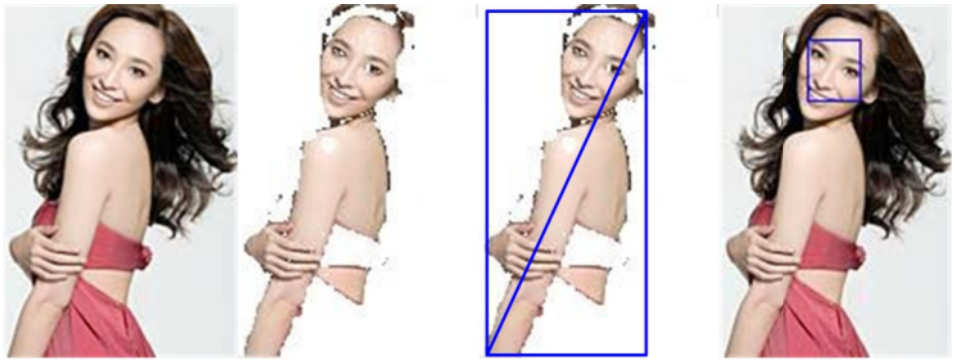

Figure 2. Skin-color and face detection

After all of the procedures given above, we define the image as normal or suspect- erotic by the judgment whether it includes the human face. Fig.2 shows an example, in which, (a) is the Original image, (b) shows a detecting result of the Skin-color area, (c) represents an enclosing rectangle of maximum diameter which is expressed by blue rectangle and diameter, and (d) is a Face detection that blue rectangle represents the face region.

International Journal of
Soft Computing And Software Engineering (JSCSE)
ISSN: 2251-7545
Vol.1,No.1, 2011
DOI: 10.7321/jscse.v1.n1.7

3. Dark erotic image detecting layer

Some compressed images on internet are lack of brightness and color distortion due to artificial or mechanical reason. The most typical one is dark brightness image. Many algorithms using skin-color model will be invalid and fail to detect. The reason is that the brightness of image is too dark to create a skin-color array. By integrated application of clipping, median filter, edge detection and threshold analysis, dark erotic image detecting layer can effectively solve this problem.

3.1 Image pre-processing

The main contents of the image almost focus on the internal area, so clipped images can decrease the computation when the amount of samples is too large. Besides, the image which contrast is great can mislead detection, such as the image that the internal area is black and the around is white. Furthermore, noise can also influence the precision of detection. So we need to image pre-processing. The steps are as follows:

1) Assume that the images' size is $M \times N$, we clip the image from starting position is $(row, column)$, termination point is $((5M+3N)/15, 0.85M)$.

2) Use a method based on adaptive wavelet to de-nosing image. The specific procedure is introduced in reference [18].

3.2 Feature detection of erotic image

According the characters that the human skin's texture Features are relatively smooth and only a few of connected objects are found in the erotic image, we will detect the erotic images by the analysis of threshold in the next procedures:

1) Transform the clipped images' RGB color space to HSI color space. We can get the vector I by equation (9).

$$I = \frac{1}{3}(R+G+B) \tag{9}$$

2) Calculate the pixels' mean M and standard deviation V of vector I. If the values are out of the range 35.5<M<71&9<V<12.6 , we will define the image as a non-erotic.

3) Use an improved sobel operator to make edge detection of images. The detailed method is introduced in reference [18]. Fig.3 shows an experiment example.

4) Label connected components and record the number of connected objects that found in binary image. Then calculate the proportion of the edge Pixels to the total pixels in the further process. Through the experiment, we found that the image is erotic in the condition as follows

$$\begin{cases} p<7.7\%, 20<S<41 \\ 8.4\% \le p<13.3\%, 31<S<91 \end{cases} \tag{10}$$

In equation (10), we assume the proportion of the edge pixels to the total pixels is p and the number of connected objects is S.

International Journal of
Soft Computing And Software Engineering (JSCSE)
ISSN: 2251-7545
Vol.1,No.1, 2011
DOI: 10.7321/jscse.v1.n1.7

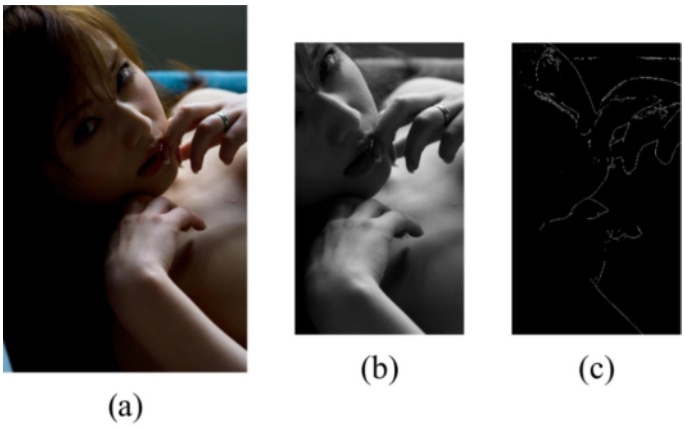

Figure 3. Dark images filtering process

Fig.3 shows an experiment result of dark images filtering process. (a), (b), (c) represent the original image, vector I of clipped image, and the image after edge detection respectively. If images' features satisfy these conditions in this layer, we can make a decision that the image is suspect- erotic.

4. Decision-Maker Layer Based on SVM

The support vector machine (SVM) is a widely used tool in classification problems. The SVM trains a classifier by solving an optimization problem to decide which instances of the training data set are support vectors, which are the necessarily informative instances to form the SVM classifier. Suppose that there are m instances of training data. Each instance consists of a (x_i, y_i) pair where $x_i \in IR^{IN}$ is a vector containing attributes of the i th instance, and $y_i \in \{+1, -1\}$ is the class label for the instance. To classify two classes of data, we need solve the quadratic programming optimization problem of equation (11), which is subject to equation (12).

$$\arg\min_{w,b,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{m} \xi_i \tag{11}$$

$$yi(f(x_i)) \geq 1 - \xi_i, \xi_i \geq 0, for\ i = 1, ..., m$$
$$f(x_i) = w \cdot x_i + b \tag{12}$$

In the functions above, minimizing $\| w \|^2 / 2$ corresponds to maximizing the margin between $w \cdot x + b = 1$ and $w \cdot x + b = -1$. We suppose the suspect-erotic image to 1, and non-erotic image to -1. Using the optimal hyperplane of equation (13), we can get optimization results.

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} \alpha_i^* y_i K(x, x_i) + b^*) \tag{13}$$

5. Experimental Results

In order to evaluate the performance of the proposed method, some intensive experiments are designed. Considering that there are no public datasets, 2138 erotic images and 8549 normal images have been collected manually, in which 1000 normal images are from Corel image database. The types of normal images include landscape picture, people photos， animal and other kinds of picture, such as remote sensing images, aerial images and so on. The ratio of normal images to erotic images is 4:1.

International Journal of
Soft Computing And Software Engineering (JSCSE)
ISSN: 2251-7545
Vol.1,No.1, 2011
DOI: 10.7321/jscse.v1.n1.7

Because the traditional erotic image filters detect images in once time, some images which features are among the classificatory interval will be misjudgment easily. Using the classificatory filter for multiple types of erotic images based on Internet this paper introduced, we can effectively filter the erotic images according to each type of images' feature. By detecting 10687 images of set, we found that our method is better than the others. The results have shown in table 1.

Table 1. Results of image detection

Methods	Filtering Results		
	Correct rate (%)	Recall rate (%)	Filtering time(seconds)
Our	93.2741	82.4524	30.5435
Ref.5	78.5824	75.1753	25.2629
Ref.7	69.8433	64.3067	33.6647
Ref.9	87.4826	72.2915	28.6437
computer : Core 2 Q8200 CPU, 2.33GHz, 4.0GB memory			

6. Conclusions

In summary， Skins-Color Model Matching layer, Dark erotic image detecting layer and Decision-making layer based on SVM are effectively combined in this paper. The proposed algorithm has the learning ability to improve the system's filtering capabilities. Compared with other methods, the proposed classificatory filter for multiple types of erotic images based on Internet displays the advantage that can effectively detect the erotic images when the image is of a dark brightness or color distortion in some complex illumination conditions. Further research will focus on: (1) improving the recognition accuracy of the erotic images by applying evolutionary computing, (2) enlarging the model to handle more complex data flows, (3) improve the detection speed while guaranteeing accuracy.

7. Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grants Grant No. 61003196 & No. 60802077 and the Fundamental Research Funds for the Central Universities under Grants K50510040007 & K50510040004.

References

[1] Yu-Ting Pai, Lite Lee, Shanqjang Ruan, Yen-Hsiang Chen, Mohanty, S.P. and Kougianos, E, "Honeycomb model based skin color detector for face detection," 15th International Conference on mechatronics and machine Vision in Practice, pp.11-16, 2008.
[2] Jun Tang, "A color image segmentation algorithm based on region growing," 2nd International Conference on Computer Engineering and Technology, pp.634-637, 2010.
[3] Zhengya Xu, Hong Ren Wu, Bin Qiu and Xinghuo Yu, "Mechatronics and machine vision in practice," IEEE Transactions on Image Processing," pp.1742-1759, 2009.
[4] Zhizhong Wang and Junhai Yong, "Texture analysis and classification with linear regression model based on wavelet transform," IEEE Transactions on Image Processing, pp.1421-1430, 2008.
[5] Yiding Wang and Jingnan Gu, "A method of erotic images filtering in real internet. Proceedings of the eighth international conference on machine learning and cybernetics," pp.1477-1481, 2009.
[6] M Felek, Forsyth D and A Bregler C, "Finding naked people," Proceedings of the 4th European Conference on Computer Vision and Pattern Recognition, pp. 593-602, 1996.

DOI: 10.7321/jscse.v1.n1.7

[7] Zhouyao Chen, Ou Wu, Mingliang Zhu and Weiming Hu, "A novel web page filtering system by combining texts and images," Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pp.732-735,2006.
[8] Xuanjing Shen, Wei Wei and Qingji Qian, "The filtering of internet images based on detecting erotogenic-part," Third International Conference on Natural Computation, pp.732 -736, 2007.
[9] Xiaohui Zhao, "Research of self-adaptive chroma space model skin-color algorithm based on illumination and its application to erotic webpage filtering," 2010 International Conference on Networking and Digital Society, pp.13-17.
[10] Shuicheng Yan, Huan Wang, Jianzhuang Liu, Xiaoou Tang and Huang T.S, "Misalignment-robust face recognition," IEEE Transactions on Image Processing, pp.1087-1096, 2010.
[11] Huang Min, Shu Huazhong and Jiang Jing, An algorithm of key-frame extraction based on adaptive threshold detection of multi-features. , International Conference on Test and Measurement, pp.149-152.2009
[12] Chunxi Ma, Lei Yang, Wenshuo Gao and Zhonghui Liu, "An improved sobel algorithm based on median filter," 2nd International Conference on Mechanical and Electronics Engineering, pp.88-92, 2010.
[13] Y Wan and H Tong, "URL assignment algorithm of crawler in distributed system based on hash," IEEE International Conference on Networking," Sensing and Control, pp. 1632-1635, 2008.
[14] Kengpei Lin and Ming-Syan Chen, "On the Design and Analysis of the Privacy-Preserving SVM Classifier," IEEE Transactions on Knowledge and Data Engineering, pp.1704-1717, 2010.
[15] Vladimir V, Vassili S and Alla A, "A survey on pixel based skin color detection techniques," International Conference on Information Science and Engineering, pp. 85-92, 2003.
[16] HSU Rein-Lien and M Abdel-Mottaleb, "Face detection in color image," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 696-706, 2002.
[17] Jinfeng Yang and Xuanjing Shen, "Research on technologies of content-based erotic image filtering and its application," Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, pp. 13-16, 2006.
[18] Yang Liu, Guo Baolong and Ni Wei, Algorithm of Contourlet Multi-threshold Image Denosing Based on Layer Structure. Computer Engineering, 32(20), pp.180-182, 2006.