Blocking Objectionable Images: Adult Images and Harmful Symbols*

Huicheng Zheng[1], Hongmei Liu[1,2], Mohamed Daoudi[1]

1. *MIIRE Group, LIFL/INT ENIC-Telecom Lille1, Rue G. Marconi, Cité Scientifique, 59655 Villeneuve d'Ascq, France. Email: (Daoudi, zheng)@enic.fr*
2. *Dept. of Electronics, Sun Yat-Sen Univ., Guangzhou 510275, P. R. China. Email: isslhm@zsu.edu.cn*

Abstract

This paper describes a practical objectionable image filtering system aimed at children's safer web access. It includes two image filters: adult image filter and harmful symbol filter. In adult image filter, we adopt statistical model for skin detection and neural network for adult image classification. The performance of the skin detection of our model outperforms that of the baseline model in [3]. Its elapsed time is about 0.18 second per. Compared with 6 mins in [1] and 10 seconds per image in [2], our system is more practical. In harmful symbol filter, we present an edge based Zernike moments method, which can capture the shape feature of symbol object effectively. Its elapsed time is about 0.13 second per image. Experimental results on large image database show that both of our filters can give promising performances.

1. Introduction

Protecting children from harmful content from Internet such as pornography and violence is increasingly a concerned research topic. Fleck et al. [1] detects naked people with an algorithm involving a skin filter and a human figure grouper. The WIPE system [2] uses Daubechies wavelets, moment analysis, and histogram indexing to provide semantically meaningful feature vector matching. Jones and Rehg [3] propose techniques for skin color detection and simple features for adult images detection. Bosson et al. [4] propose a pornographic image detection system that is also based on skin detection and the multi-layer perception (MLP) classifier.

In our adult image filter, the first step is skin detection. We build a model with Maximum Entropy Modeling (MaxEnt) [6] for the joint distribution of the input color images and the "skinness" images. This model imposes constraints on two-pixel marginal. We use Bethe tree approximation to eradicate the parameter estimation and the Belief Propagation (BP) algorithm to obtain exact and fast solution. Our model outperforms the baseline model in [3] in terms of pixel classification performance. Based on the output of skin detection, 9 features including global ones and local ones are extracted. We use a MLP to detect adult images based on these features. Our adult image filtering takes about 0.18 second per image in average. Compared with 6 mins in [1] and 10 seconds per image in [2], our system is more practical. Plenty of experimental results on 5,084 photographs show stimulating performance.

Symbols typically appear as mixed text and graphic icons which, when recognized, trigger an association of the object to which they are attached, with a given group or organization. The objectionable image filter systems in the literature seldom take symbols into consider. Among symbols, there are some violent and illegal ones, such as blood, drug, and nazi. If we can recognize these symbols, then we can combine with the text context to block access to such web sites. In this paper, we present a method for such system based on the image content using shape feature. Zernike moments are employed as a feature set. Taking account of the importance of the edge of an image to the human perception, we propose to compute Zernike moment on the edge of the symbol, without assumption that we know the grey scale of the background like in the logo or trademark applications. Experimental results on a database of 480 symbols demonstrate that the edge based Zernike moments method have promising performance.

* This work is partially supported by European Community IAP 2117/27572-POESIA http://www.poesia-filter.org, http://sourceforge.net/projects/poesia/

The paper is organized as follows. In section 2, we give a brief overview of the system. Adult image filter and symbol filter are presented in section 3 and 4 respectively. Section 5 summarizes the paper.

2. The structure of the system

Our object is blocking objectionable images, including adult images and harmful symbols. The system structure is shown in Fig.1. A captured image is first classified by a classifier discriminating natural images from artificial images. If an image is classified as natural image, it is passed to adult image filter, otherwise, passed to symbol filter. The natural image and artificial image classifier is discussed in [5]. In this paper, we will present our adult image filter and harmful symbol filter.

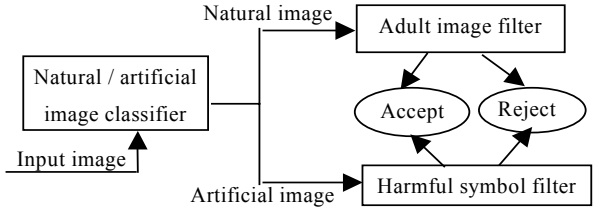

Fig 1. The structure of the system

3. The adult image filter

The structure of our adult image filter is shown in figure 2. The first step is skin detection, then feature vector is extracted from the output of skin detection. The MLP takes the feature vector as input and output a real number O_p, which can be compared with a threshold T to decide whether the input color image is an adult image.

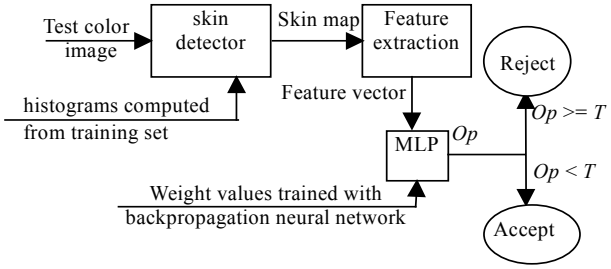

Fig.2 The structure of adult image filter

3.1. Skin detection

Based on the strong correlation between adult images and large skin patches, we have to design a skin detector. Jones and Rehg [3] implemented an independent statistical model to detect skin pixel individually. We call their model the baseline model. In this paper, we propose MaxEnt for skin detection.

Let us fix the notations. The set of pixels of an image is $\boldsymbol{S}$. We notate the RGB color space $\boldsymbol{C}=\{0,\ldots,255\}^3$. The color of a pixel $s\in\boldsymbol{S}$ is x_s, $x_s\in\boldsymbol{C}$. The "skinness" of a pixel s is y_s with $y_s=1$ if s is a skin pixel and $y_s=0$ if not. The set of neighbors of s is notated as $\cup(s)$. $<s, t>$ denotes a pair of neighboring pixels. The color image, which is the vector of color pixels, is notated x and the binary image made up of the y_s's is notated y.

The segmented Compaq Database [3] is a collection of samples $\{(x^{(1)},y^{(1)}),\ldots,(x^{(n)},y^{(n)})\}$ where for each $1\le i \le n$, $x^{(i)}$ is a color image and $y^{(i)}$ is the associated binary skinness image. We assume that the samples are independent of each other with distribution $p(x,y)$. The collection of samples is split into two equal parts, the training data and the test data.

We estimate $p(x,y)$ with MaxEnt. We define the following constraints $\boldsymbol{C}$: $\forall s \in\boldsymbol{S}$, $\forall t \in\cup(s)$, $\forall x_s \in C$, $\forall x_t \in C$, $\forall y_s \in\{0,1\}$, $\forall y_t \in\{0,1\}$, $p(x_s,x_t,y_s,y_t) = q(x_s,x_t,y_s,y_t)$. The quantity $q(x_s,x_t,y_s,y_t)$ is the empirical distribution observed from the training data. The MaxEnt solution $p(x,y)$ is the probability distribution with the maximum entropy (most uniform) under constraints $\boldsymbol{C}$. We approximate the pixel lattice with Bethe tree, then $p(x,y) \propto \prod_{<s,t>} \frac{q(x_s,x_t,y_s,y_t)}{q(x_s,y_s)q(x_t,y_t)} \prod_{s\in S} q(x_s,y_s)$, see [6]. We call it TFOM—tree approximation of first order model. We then use belief propagation algorithm to obtain exact and fast solution for the marginal $p(y_s|x)$ from $p(x,y)$ for $\forall s \in\boldsymbol{S}$, see [6] for a detailed account.

Our TFOM model outperforms the baseline model in [3] as shown in [6] in the context of skin pixel detection rate and false positive rate. At the same false positive rate 5%, the baseline model can detect 69% of skin pixels, while the TFOM model can detect 72% of skin pixels. The output of skin detection is a *skin map* indicating the probabilities of skin on pixels. In figure 3, skin maps of different kinds of people are shown.

Fig 3. Some images and their skin maps

3.2. Adult image detection

There are propositions for high-level features based on grouping of skin regions [1], but the actual

applications need high speed, so, along with [2][3], we are interested in simpler features. Skin regions of adult images have specific shapes and orientations. We use fit ellipses to roughly catch such information. Moreover, it is easy to calculate fit ellipses. We calculate two fit ellipses for each skin map---the Global Fit Ellipse (GFE) and the Local Fit Ellipse (LFE). GFE is for all detected skin and LFE is for the largest skin region.

The feature vector is composed of 9 features: 1) the average skin probability of the whole image, 2) the average skin probability inside the GFE, 3) number of skin regions in the image, 4) distance from the centroid of the largest skin region to the center of the image, 5) angle of the major axis of the LFE from the horizontal axis, 6) ratio of the minor axis to the major axis of the LFE, 7) ratio of the area of the LFE to that of the image, 8) average skin probability inside the LFE, 9) average skin probability outside the LFE. No effort was done to find the correlation among features.

Evidence from [4] shows that the MLP classifier offers a statistically significant performance over several other approaches such as the generalized linear model, the k-nearest neighbor classifier and the support vector machine. In this paper, we adopt the MLP classifier. The output of the MLP is a number $o_p \in [0,1]$, corresponding to the degree of adult. One can set a proper threshold to get the binary decision.

3.3. Experimental results

All experiments are made on the database containing




o_p=0.001452 o_p=0.000000 o_p=0.006828 o_p=0.899044

Fig.4 Images classified correctly **Fig.5** Misclassified images

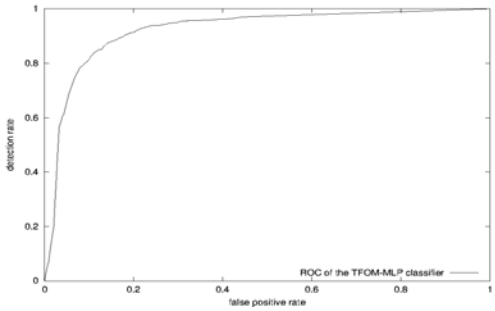

Fig. 6 ROC of classification

10,168 photographs, which are imported from the Compaq database and the Poesia database. It is split into two equal parts randomly, the training set and the test set, with 1,297 adult photographs and 3,787 benign ones in each set. Fig.4 shows some images from the test set with their MLP outputs, they are classified correctly. Fig.5 shows some misclassified images. The first adult image is not detected since the skin appears almost white due to over-exposure, so most of the skin is not detected. The second image is benign, but it is detected adult since the toy dog takes a skin-like color and the skin detector gives a very high false alarm. The ROC curve on the test set is shown in Fig.6. The average elapsed time is about 0.18 second per image.

4. The symbol filter

Our symbol filter is to block harmful symbols, such as blood, drug and symbols attached with a harmful group, for example, a nazi. In this paper, an edge based symbol filter using Zernike moment is presented. We first set up a training symbol and their features database, in which known harmful symbols, such as Nazi, KKK are contained. Then for an unknown symbol coming from web, our symbol filter decides whether this symbol is a harmful symbol by features matching algorithm. The problem is similar as retrieving a symbol from a database based on features of a symbol. Feature selection is one of the key problems in content-based image retrieval applications. The features used by various systems include: invariant moments and Fourier descriptors extracted from manually isolated distinct objects in [7]; histogram of edge directions and invariant moments in [8]; Zernike moment in [9]. The feature is required to be invariant to rotation, translation and scale. In this paper, Zenike moment magnitudes(ZMMs) are used as a feature set. ZMMs are robust to noise or small variance of a symbol, and have invariant characteristics. With a proper normalization method, scale invariance and translation invariance have also been achieved. Zernike moment has been used as shape descriptor in trademark and logo retrieval systems due to its many desirable properties. But in literature, the scale and translation invariance of Zernike moments are achieved under assumption that the object in the image is known. But this assumption is not valid for web application, so taking account of the importance of the edge of an image to the human perception, we propose to compute Zernike moment on the edge of the symbol, without assumption that we know the grey scale of the background like in logo or trademark applications.

The detailed description of Zernike moment can be found in [10]. Our symbol filter is composed of following parts:

1) Build training feature database offline. We extract ZMMs of each symbol in training database to build a

training feature database. This part can be implemented offline.

2) Extract ZMMs of unknown test symbol.

3) Classification. To decide whether a test symbol is in the training database. First, the nearest-neighbor classifier labels an unknown image represented by a feature vector with the label of the nearest neighbor among all the training samples. The distance between the test symbol and a training sample is measured using Euclidean distance. Compare the minimal distance with a predefined threshold, we can get a binary decision.

The feature extraction procedure is as follows:

1) transform input image to grey scale image.

2) get binarized edge image of the grey scale image, we consider that the object is composed of the edge pixels. The symbols and their edges are shown in figure 7.

3) normalize the binarized edge image to accomplish object scale invariance and move the origin of the image to the centroid of the object to obtain object translation invariance.

4) the scale and translation invariance stage does affect the first and second order Zernike features. The first order ZMM is going to be the same for all images and the second order ZMM is equal to zero. So these two ZMMs will not be included as one of the utilized feature. The extracted Zernike features start from the second order moments. We extract up to twelfth order Zernike moments corresponding to 47 features.

Fig.7 Some images and their edge images

The test set in our experiments include 375 harmful symbols, which are obtained by processing each symbol in training database, by rotations with different angles, scaling with different ratios, translations with different pixels and JPEG compression with different quality factors, and 105 benign symbols downloaded from web. The true negative rate for benign symbols is 0.89 and the true positive rate for harmful symbols is 0.85, respectively. The average elapsed time for each symbol is 0.13s.

5. Conclusion

This work aims at filtering objectionable images in Internet. Our main contributions are as follows:

1) We build a first order model for skin detection with MaxEnt that imposes constraints on two-site marginal. With Bethe tree approximation, parameter estimation of our MaxEnt is eradicated. We then use the BP algorithm to accelerate the calculation of one-site marginal of skinness. This model improves the performance of the baseline model in the previous work [3].

2) By using simple skin-based features and MLP classifier, our adult image filter gives promising performance.

3) Our adult image filter is more practical compared with those in [1] and [2] in terms of processing speed.

4) Pay attention to the harmful symbols in Internet, present an edge based Zernike moment method for blocking harmful symbols.

To improve the performance of our filters, we can use a face detector in adult image filter, and extract local features besides global features in symbol filter. In general, images tend to appear together and are surrounded by text in web pages, so combining with text analysis could improve the performance of our image filters. All of these are our future researches.

6. References

[1] M.M. Fleck, et al, " Finding naked people", in: Proc. European Conf. on Computer Vision, B. Buxton, R. Cipolla, Springer-Verlag, Berlin, Germany, 2:593-602, 1996.

[2] J. Z. Wang, et al " System for Screening Objectionable Images ", Computer Communications, (21)15:1355-1360, 1998.

[3] M.J. Jones, J.M. Rehg, " Statistical color models with application to skin detection ", Computer Vision and Pattern Recognition, 274-280, 1999.

[4] A. Bosson, et al, "Non-retrieval: blocking pornographic images", in: Proc. Intl. Conf. on the Challenge of Image and Video Retrieval, Lecture Notes in Computer Science, Springer-Verlag, London, 2383:60-70, 2002.

[5] H. Liu, M. Daoudi, B. Jedynak, "Classification of photographs and artificial images", submitted to Pattern Recognition.

[6] B. Jedynak, H. Zheng, M. Daoudi, "Statistical Models for Skin Detection", IEEE Workshop on Statistical Analysis in Computer Vision, in conjunction with CVPR 2003 Madison, Wisconsin, June 16--22, 2003.

[7] C.P.Lam, J.K.Wu, and B. Mehtre, "STAR - a system for trademark archival and retrieval", Proceedings 2nd Asian Conf. on Computer Vision, vol.3, pp214-217, 1995

[8] A.K.Jain and A.Vailaya, "Image retrieval using color and shape", Proceedings 2nd Asian Conf. on Computer Vision, Vol.2, pp529-533,1995

[9] W.Y.Kim, Y.S.Kim, "A region-based shape descriptor using Zernike moments", Signal Processing: Image Communication 16(2000) 95-100

[10] A.Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments", IEEE Trans. on PAMI, Vol.12, No.5, May 1990