

Adult Image Detection Using Statistical Model and Neural Network*

Huicheng Zheng¹

Mohamed Daoudi¹

Bruno Jedynak^{2,3}

¹ MIIRE LIFL-INT / ENIC-Telecom Lille1,

Cité Scientifique, rue G. Marconi, 59658 Villeneuve d'Ascq cedex - France

{zheng, daoudi}@enic.fr

² Center for Imaging Science, The Johns Hopkins University, U.S.A.

³ Laboratoire de Mathématiques Appliquées, USTL, Bât M2, Cité Scientifique,
59655 Villeneuve d'Ascq, France

bruno.jedynak@jhu.edu

*This work was partially supported by European project Internet Action Plan Contract Number POESIA-2117/27572 www.poesia-filter.org

Abstract

This work is aimed at the detection of adult images appear in Internet. Skin detection is of the paramount importance in the detection of adult images. In our previous work [1], we built a maximum entropy model for skin detection. The output of skin detection is a grayscale skin map with the gray level indicating the belief of skin. Two fit ellipses are then calculated from the skin map—the Global Fit Ellipse and the Local Fit Ellipse. We then calculate several simple features from the skin map and fit ellipses. A multi-layer perceptron classifier with backpropagation is trained for these features. We have done plenty of experiments. The Receiver Operating Characteristics (ROC) curve calculated from a large test set of images shows stimulating performance for such simple features. The elapsed time is about 1 second for a 256×256 image. Compared with [2], which takes about 6 minutes on a workstation for the figure grouper in their algorithm to process a suspect image passed by the skin filter, our algorithm is more practical.

Key words

Skin detection, Maximum entropy modeling, Adult image detection, Fit ellipse

1 Introduction

Images are an essential part of today's World Wide Web. The statistics of more than 4 million HTML webpages reveal that 70.1% of webpages contain images and that on average there are about 18.8 images per HTML webpage[3]. On the other hand, images are also contributing to harmful (e.g. pornographic) or even illegal (e.g. paedophilic) Internet content. So effective filtering of images is important in an Internet filtering solution.

To block adult content some representative companies as

Net Nanny and SurfWatch operate by maintaining lists of URL's and newsgroups and require constant manual updating. Abundant literature is available, but the Internet is very rapidly evolving, not only quantitatively. Each day, 3 million pages are appearing on the Web. Detection based on image content analysis has the advantage to process equally all the images without the need for frequent updating, so will produce more effective filtering.

By taking advantage of the fact that there is a strong correlation between images with large patches of skin and adult images we have to develop a skin detector. Skin color offers an effective and efficient way to detect the adult image content. There is already some work on this track.

The WIPE [4] system developed by Wang, Li, Wiederhold and Firschein uses a manually-specified color histogram model as a prefilter in an analysis pipeline. Input images whose average probability of skin is low are accepted as non-offensive. Images that contain considerable skin pass on to a final stage of analysis where they are classified using wavelet features. The algorithm uses a combination of Daubechies wavelets, normalized central moments, and color histograms to provide semantically-meaningful feature vector matching.

Forysth's [2] research group has designed and implemented an algorithm to screen images of naked people. Their algorithms involve a skin filter and human figure grouper. The skin color model used by Fleck, Forsyth and Bregler consists of a manually specified region in a log-opponent color space. Detected regions of skin pixels form the input to a geometric filter based on skeletal structure. As indicated in their paper, 52.2% sensitivity and 96.6% specificity have been obtained for a test set of 138 images with naked people and 1401 assorted benign images. However, it takes about 6 minutes on a workstation for the fig-

ure grouper in their algorithm to process a suspect image passed by the skin filter. Most of the people in the images used in the experimental protocol are Caucasians and a small number of images are Blacks or Asians.

Jones and Rehg [5] propose techniques for skin color detection by estimating the distribution of skin and non-skin color in the color space using labeled training data. To detect adult images, some simple features are extracted. The discrimination performance based solely on skin is rather good for such simple features.

Bosson et al. [6] propose a pornography detection system which is integrated in a commercial system. This system is also based on skin detection. They compared the generalised linear model, the k -nearest neighbor classifier, the multi-layer perception (MLP) classifier and the support vector machine and found that the MLP gives the best classification performance.

Our approach is as follows. The first main step is skin detection. We build a model with Maximum Entropy Modeling (MaxEnt) for the skin distribution. This model imposes constraints on color gradients of neighboring pixels. Parameter estimation as well as optimization cannot be tackled without approximations. With Bethe tree approximation parameter estimation is eradicated and the Belief Propagation (BP) algorithm permits to obtain exact and fast solution for skin probabilities at pixel locations. This model is referred to as TFOM for Tree First Order Model. The output of skin detection is a grayscale *skin map* with the gray levels being proportional to the skin probabilities. The second main step is pattern recognition. Two fit ellipses are calculated from the skin map—the Global Fit Ellipse and the Local Fit Ellipse. We calculate several simple features from the skin map and fit ellipses which form a pattern. A MLP classifier is trained on 5,084 patterns from the training set. In the test phase, the MLP classifier takes a quick decision on the input pattern in one pass.

The rest of this paper is organized as follows: in section 2 we present the skin detection module. Section 3 is devoted to feature extraction and pattern recognition. In section 4, some experimental results are presented. Section 5 concludes this paper.

2 Skin detection

MaxEnt is a method for inferring models from a data set. See [7] for the underlying philosophy. It works as follows: (1) choose relevant features (2) compute their histograms on the training set (3) write down the maximum entropy model within the ones that have the feature histograms as observed on the training set (4) estimate the parameters of the model (5) use the model for classification. This plan has been successfully completed for several tasks related to speech recognition and language processing. See for example [8] and the references therein. In these applications the underlying graph of the model is a line graph or even a tree but in all cases it has no loops. When working with images, the graph is the pixel lattice. It has indeed many

loops. A breakthrough appeared with the work in [9] on texture simulation where (1)–(4) were performed for images and (5) replaced by simulation.

In the paper [1] we adapt this methodology to skin detection as follows: in (1) we specialize in colors for two adjacent pixels given “skinness”. We choose RGB color space in our approach. In practice we know from [5][6] that the choice of color space is not critical given a histogram-based representation of the color distribution and enough training data. In (2) we compute the histogram of these features in the Compaq manually segmented database. Models for (3) are then easily obtained. In (4) we use the Bethe tree approximation, see [10]. It consists in approximating locally the pixel lattice by a tree. The parameters of the MaxEnt models are then expressed analytically as functions of the histograms of the features. This is a particularity of our features. In (5) we pursue the approximation in (4): we use the BP algorithm, see [11], which is exact in tree graph but only approximative in loopy graphs. Vezhnevets et al. [12] recently compared some most widely used skin detection techniques and conclude that the proposed MaxEnt model gives the best performance in terms of pixel classification rates. The output of skin detection is a grayscale skin map with the gray levels being proportional to the skin probabilities. We show the output of the skin detection in Fig. 1, where on the left is the original color image, on the right the corresponding skin map.



Figure 1 – Left: original color image. Right: the corresponding skin map.

3 Adult image detection

3.1 Feature extraction

There are propositions for high-level features based on grouping of skin regions[2] that might distinguish adult images from those not, but here we have a requirement to process the images speedily so, along with [5][4], we are interested to try simpler features.

We first binarize the skin map by simple thresholding. We then implement morphological open/close operations to remove noise and connect broken regions. Small skin regions are considered insignificant and discarded. Many of our features are based on the fit ellipses[13] calculated on the skin map, since they could meet our requirement for sim-

plicity and capture some important shape information. We observed from experiments that for approaches based on skin detection, portraits have a tendency to be detected as adult images since generally portraits expose plenty of skin as adult ones. The fit ellipses will hopefully at least help discriminate portraits from adult images. We will calculate two fit ellipses for each skin map—the Global Fit Ellipse (GFE) and the Local Fit Ellipse (LFE). The GFE is computed on the whole skin map, while the LFE only on the largest skin region in the skin map.

We extract 9 features from the skin map and fit ellipses. The first 3 are global: (1) average skin probability of the whole image, (2) average skin probability inside the GFE and (3) number of skin regions in the image. The other 6 features computed on the largest skin region of the input image are (1) distance from the centroid of the largest skin region to the center of the image, (2) angle of the major axis of the LFE from the horizontal axis, (3) ratio of the minor axis to the major axis of the LFE, (4) ratio of the area of the LFE to that of the image, (5) average skin probability inside the LFE, (6) average skin probability outside the LFE. No effort was done to find the correlation between features.

3.2 Pattern recognition

The feature extraction steps described in the previous subsection produce a feature vector for each image. The task is then to find the decision rule on this feature vector that optimally separates adult images from those not. Evidence from [6] 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.

Our MLP classifier is a semilinear feedforward net with one hidden layer as in [14]. This net outputs a number between 0 and 1, with 1 for adult image and 0 not. The learning procedure starts off with a random set of weight values. For each training pattern p , the net evaluates the output o_p in a feedforward manner. To decrease the error between the output o_p and the true target t_p , the net calculates the corrections of the weight values using the backpropagation procedure. This procedure is repeated for all the patterns in the training set to yield the resulting corrections for all the weights for that one iteration. In a successful learning exercise, the system error will decrease with the number of iterations, and the procedure will converge to a stable set of weights, which will exhibit only small fluctuations in value as further learning is attempted. In the test phase, for each test pattern, the net calculates the output in one pass. We then set a threshold T , $0 < T < 1$, to get the binary decision.

4 Experiments

All experiments are made using the following protocol. The database contains 10,168 photographs, which are imported from the Compaq database and the Poesia database. It is split into two equal parts randomly, with 1,297 adult

photographs and 3,787 other photographs in each part. One part is used as the training set while the other one, the test set is left aside for the ROC curve computation. This is calculated by varying the threshold T . Figure 2 shows the resulting ROC curve. The elapsed time is about 1.51×10^{-5} second/pixel, i.e., about 1 second for a 256×256 image. Compared with [2], which takes about 6 minutes on a workstation for the figure grouper in their algorithm to process a suspect image passed by the skin filter, our algorithm is more practical.

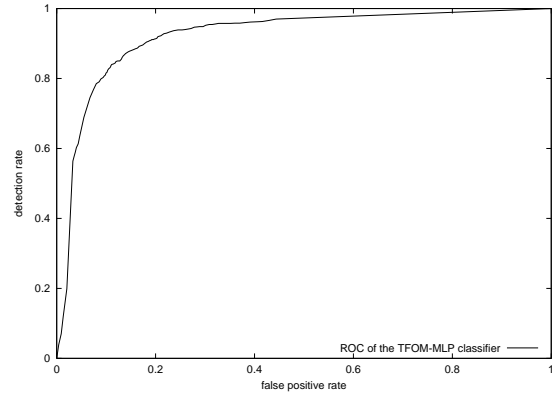


Figure 2 – ROC curve of the TFOM-MLP adult image detection.

There are some false alarms worth a look as shown in Fig. 3. The toy dog is detected adult since it takes a skin-like color and the average skin probabilities inside the GFE and the LFE are very high. The portrait is declared adult since it exposes a lot of skin and even the hair and the clothes take skin-like colors. We believe skin detection based solely on color information of one image cannot do much more, so maybe some other sorts of information is needed to improve the adult image detection performance. For example, some kind of face detector could be implemented to improve the results. Moreover, adult images in webpages tend to appear together, and are surrounded by text, which could be an important clue for the adult content detector.

5 Summary and conclusions

This work is aimed at filtering adult images appear in Internet. The first step of our approach is skin detection. Maximum entropy modeling is used to model the distribution of skinness from the input image. We build a First Order Model that introduces constraints on color gradients of neighboring pixels. We then use Bethe tree approximation to eradicate parameter estimation. It is then called TFOM for Tree First Order Model in this paper. The Belief Propagation algorithm could be further implemented to accelerate the processing.

The output of skin detection is a grayscale skin map with the gray levels being proportional to the skin probabilities. We use the fit ellipses to catch the characteristics of

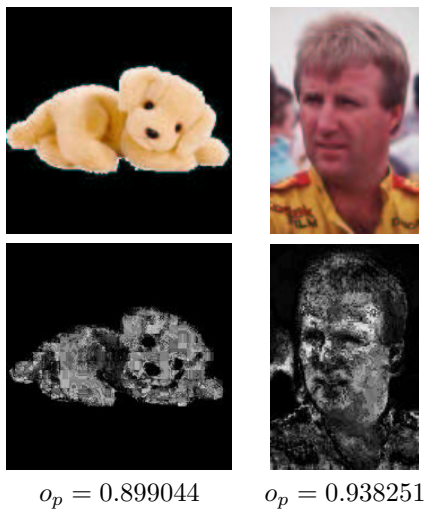


Figure 3 – First row: original images. Second row: the corresponding skin maps. Third row: the corresponding outputs of the MLP. These results show a toy dog and a portrait detected as adult, which is false.

skin distribution. Two ellipses are calculated for each skin map—the Global Fit Ellipse (GFE) and the Local Fit Ellipse (LFE). A set of 9 simple features are then computed from the skin map and fit ellipses. A multi-layer perceptron classifier is trained for these features. It is a semilinear feedforward net with backpropagation.

We have done plenty of experiments. A ROC curve computed from 5,084 test images shows stimulating performance for such simple features. To improve the results one can use a face detector. Moreover, adult images tend to appear together and are surrounded by text in webpages, which could improve the performance of adult image detection.

References

- [1] B. Jedynek, H. Zheng, and M. Daoudi. Statistical models for skin detection. In *IEEE Workshop on Statistical Analysis in Computer Vision, in conjunction with CVPR 2003*, Madison, Wisconsin, June 2003.
- [2] M.M. Fleck, D.A. Forsyth, and C. Bregler. Finding naked people. In *Proc. European Conf. on Computer Vision*, pages 593–602. B. Buxton, R. Cipolla, Springer-Verlag, Berlin, Germany, 1996.
- [3] B. Starynkevitch, M. Daoudi, and et al.. Poesia software architecture definition document. Technical report Deliverable 3.1, POESIA consortium, December 2002. http://www.poesia-filter.org/pdf/Deliverable_3_1.pdf.
- [4] James Ze Wang, Jia Li, Gio Wiederhold, and Oscar Firschein. System for screening objectionable images. *Images, Computer Communications Journal*, 1998.

- [5] M.J. Jones and J. M. Rehg. Statistical color models with application to skin detection. In *Computer Vision and Pattern Recognition*, pages 274–280, 1999.
- [6] A. Bosson, G.C. Cawley, Y. Chian, and R. Harvey. Non-retrieval: blocking pornographic images. In *Intl. Conf. on the Challenge of Image and Video Retrieval*, volume 2383 of *Lecture Notes in Computer Science*, pages 50–60, London, 2002. Springer-Verlag.
- [7] E. Jaynes. Probablity theory: The logic of science. <http://omega.albany.edu:8008/JaynesBook>.
- [8] A. Berger, S. Della Pietra, and V. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71, 1996.
- [9] S.C. Zhu, Yingnian Wu, and David Mumford. Filters, random fields and maximum entropy (frame): towards a unified theory for texture modeling. *International Journal of Computer Vision*, 27(2):107–126, 1998.
- [10] Chihsin Wu and Peter C. Doerschuk. Tree approximations to markov random fields. *IEEE Transactions on PAMI*, 17(4):391–402, April 1995.
- [11] J. S. Yedida, W. T. Freeman, and Y. Weiss. Understanding belief propagation and it’s generalisations. Technical report TR-2001-22, Mitsubishi Research Laboratories, January 2002.
- [12] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Graphicon-2003, 13th International Conference on the Computer Graphics and Vision*, Moscow, Russia, September 2003.
- [13] R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley, 1992.
- [14] Y.-H. Pao. *Adaptive Pattern Recognition and Neural Networks*, pages 121–129. Reading. Addison-Wesley, Massachusetts, 1989.