# A Hybrid Approach to Detect Adult Web Images

Qing-Fang Zheng, Ming-Ji Zhang, and Wei-Qiang Wang

Institute of Computing Technology, Chinese Academy of Science, Beijing, China
{qfzheng, mjzhang, wqwang}@jdl.ac.cn

**Abstract.** This paper presents a hybrid approach to discriminate benign images from adult images. Different from previously published works, our approach combines face detection and adaptive skin detection. First, face detection using haar-like features is performed, then skin color model is learned on-line by incorporating the information of color distribution in the face regions. Based on the result of face detection and adaptive skin detection, a set of semantically high-level features are extracted. These features have human-oriented meanings which can effectively discriminate adult images from benign images. Our two main contributions are i) a set of high-level features for adult or benign image classification and ii) a novel adaptive skin detection algorithm in still images. Experimental results are reported to demonstrate the strength of our approach.

## 1 Introduction

The Internet is one of the greatest inventions of all times, but it has also become a playground for pornographers. According to a commercial report [1], the number of pornographic web pages on the Internet in year 2003 has increased nearly 1800 percent compared with 14 million pages five years ago. Exposure to the sea of pornography can lead to many social problems including cyber-sex addiction. It is now an urgently necessary task to prevent people, especially children, from accessing this type of harmful material. A direct solution to Internet porn images filtering is to evaluate an image's content before displaying it. This paper presents a new approach to separate benign images from adult images by analyzing a set of semantically high-level features obtained from face detection and adaptive skin detection.

To our knowledge, there are no previously published works that used high-level features such as face to perform benign-or-pornographic images classification. Related works concerning this field mostly used low-level visual features. Wang et al. employed a combination of Daubechies' wavelets, normalized central moments and color histograms as a feature vector and matched it against a small number of features obtained from a training database [2]. Chan et al. used three simple features: the ratio of skin area to image area, the ratio of the largest skin segment to the image area and the number of segments in the image [3]. The features used by Jones and Rehg are [4]: percentage of pixels detected as skin;

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3332, pp. 609–616, 2004.

average probability of the skin pixels size of the largest connected pixels; size of the largest connected component of skin; number of connected components of skin; percent of novel pixels, height and width of the image. In Image Guarder system develop by Zeng, a combination of skin feature, texture feature and shape feature are used [5]. Forsyth developed an algorithm that involved a skin filter and a human figure grouper to find naked people in the image [6].

Although low-level features are easy to compute, they are insufficiently accurate because of semantic gap: Human's interpretation of image's content as pornographic or not is so abstract that there is no simple computational transformation that will map low-level image features to human perception. We propose to use high-level features to achieve pornographic or non-pornographic image classification. The features we used are all related to human face. Our approach takes advantage of research achievements of face detection technique because face detection is now being extensively studied and some effective and efficient face detectors have been introduced in the literature. There are two main contributions of our work. The first is a set of semantically high-level features for image classification and the second is a novel adaptive skin detection algorithm in still images which is robust to illumination conditions and not biased by human races.

The rest of the paper is organized as follows. In section 2, we will detail our hybrid approach including face detection, on-line skin color modeling, adaptive skin detection, feature extraction and image classification. In section 3, we will provide experimental result. Conclusion and future directions will be discussed in section 4.

## 2 Our Hybrid Approach

In this section, we describe our approach to separate benign images from pornographic images by analyzing image content. We will start with the overall architecture, followed by a detailed discussion of the main components.

### 2.1 General Overview

The basic process flow of the proposed approach is as follows. First, face detection is performed on the input image using a set of over-complete haar-like features. Once face is detected, color distribution of face region is computed and used for on-line human skin color modeling. And then adaptive skin detection is performed on the whole image. Based on the result of face detection and skin detection, a set of semantically high-level features are extracted and fed into a decision tree classifier. The whole procedure can be visualized in Fig. 1.

### 2.2 Semantically High-Level Features

High-level features gain advantages over low-level features in that they narrow down the semantic gap between human perception and raw image data. To interpret an image's content, we must identify some important features or objects

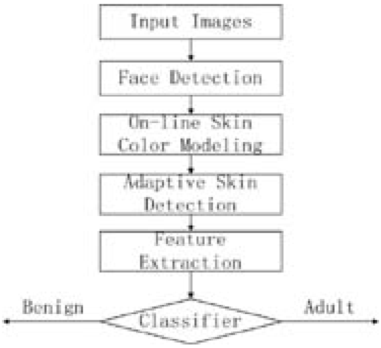


**Fig. 1.** Overall process of our approach

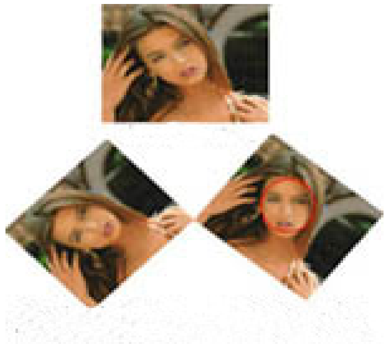

**Fig. 2.** Face detection, the red circle marks the detected face region.

in the image. Face is the most distinctive part of human body. Detection of faces can allow the observer to form a hypothesis about the presence of humans in the scene and other measures can be taken to verify whether or not they are nudes. We empirically choose five features for benign-or-pornographic image classification. These features are all related to face:

(1) Face Number: the number of faces in the image. Pornographic images usually do not contain too many faces. We define $F1$ as:

$$F1 = number\ of\ face\ in\ the\ image \tag{1}$$

(2) Face Area: the area of face regions. Images in which face regions cover too much area may be full-face portraits. We define $F2$ as:

$$F2 = \frac{number\ of\ pixels\ in\ face\ region}{number\ of\ pixels\ in\ whole\ image} \tag{2}$$

(3) Face Position: Images which present face in the center are usually benign. We define $F3$ as distance between the center of face region and the center of image normalized by the minimum length of image axis:

$$F3 = \frac{distance(center_{face}, center_{image})}{min(image.width, image.height)} \tag{3}$$

(4) Skin Ratio I: the ratio of whole image skin area to face area;

(5) Skin Ratio II: the ratio of the largest skin segment to the face area. The higher the Skin Ratio I and Skin Ratio II are, the more likely the images are pornographic. We define $F4$ and $F5$ as follows:

$$F4 = \frac{number\ of\ skin\ pixels\ in\ whole\ image}{number\ of\ pixels\ in\ face\ regions} \tag{4}$$

$$F5 = \frac{number\ of\ skin\ pixels\ in\ largest\ skin\ segement}{number\ of\ pixels\ in\ face\ regions} \tag{5}$$

### 2.3 Face Detection

Face detection plays an important role in the our image classification approach. Our face detector is based largely on the work of Paul Viola [7]. A cascade of boosting classifiers is built on an over-complete set of haar-like features. In each stage of the cascade, a variant of AdaBoost is used to integrate the feature selection and classifier design in one boosting procedure. Only positive result from the previous classifier needs further more complex evaluation. By adopting this simple-to-complex strategy, most non-face candidates are rejected in earlier stage of cascade with little computation costs. Details of the face detection algorithm can be found in [7]. In order to successfully detect rotated faces, each input image is passed to face detector three times. The first time is the original image, the next two times are its rotated variants with rotation angle of 45 degrees anticlockwise and clockwise respectively. We do this because we find that faces in adult images are usually within a rotation angle of no more than 90 degrees from vertical direction. Fig. 2 illustrates our face detection method.

### 2.4 Online Skin Color Modeling

Skin detection is an important technique for identifying adult images because of the fact that there is a strong correlation between image with large patches of skin and adult images. However, accurate skin detection is a non-trivial task. In traditional skin detection schemes, very often a static skin color model is learned off-line and each image pixel is checked whether or not its color value satisfies the learned model. Skin color varies greatly between different human races. To make things worse, skin color, as measured by camera, can change when illumination condition changes. Therefore, skin detection that uses a static skin color model is certain to fail in unconstrained imaging conditions. Here we propose a novel adaptive skin detection method based on the result of face detection. Once face is detected, color distribution in face region is used as useful context information for on-line skin color model building. The proposed method is robust to imaging conditions and not biased by human ethnicity.

Our skin detection method takes advantage of the fact that the face and body of a person always share same colors. Color distribution of face regions can provide useful cues to detect skin regions of other body parts. We choose to build skin color model on YCbCr color space and use a normal distribution $N(\mu, \sigma)$ to represent the distribution of each skin color component (Y,Cb,Cr). Color values of image pixels in face region are viewed as an ensemble of skin color samples:

$$\Omega = \{\{y_1, cb_1, cr_1\}, \{y_2, cb_2, cr_2\}, \cdots, \{y_K, cb_K, cr_K\}\} \tag{6}$$

Then the mean and variance of each normal distribution can be computed. For distribution of Y component:

$$\mu_y = \frac{1}{K}\sum y_j \qquad \sigma_y^2 = \frac{1}{K-1}\sum (y_j - \mu_y)^2 \tag{7}$$

K is the size of $\Omega$, i.e., number of pixels in face regions.

The distribution parameters of Cb and Cr components can be computed similarly. Pixel outside of face region is classified as skin pixel if it satisfies the following requirements:

$$\|y-\mu_y\| \leq a_y\sigma_y \;\; and \;\; \|cb-\mu_{cb}\| \leq a_{cb}\sigma_{cb} \;\; and \;\; \|cr-\mu_{cr}\| \leq a_{cr}\sigma_{cr} \tag{8}$$

$y$,$cb$,$cr$ are the color values of the pixel to be classified and $a_y$,$a_{cb}$,$a_{cr}$ are threshold values to be adaptively determined which will be described in section 2.5.

### 2.5 Adaptive Skin Detection

We do not use fixed threshold values because through experiment we find fixed thresholds can not separate skin region from non-skin region which have colors similar to skin. Instead, we select threshold values adaptively by taking the texture property of human skin region into consideration. Skin region is usually homogeneous and has smooth texture. Using texture characteristics of skin region to find an optimal threshold for skin segmentation was proposed by Phung [8]. Two main aspects can differentiate Phung's method from ours. One is the fact that Phung's method performs on skin score map computed from Bayesian decision theory while our method performs on original images, the other is the region homogeneity measures. The homogeneity measures we used are:

$$\sigma^y_{region} < 0.5\mu^y_{region} \;\; and \;\; \sigma^{cr}_{region} < 0.4\mu^{cr}_{region} \;\; and \;\; \sigma^{cb}_{region} < 0.4\mu^{cb}_{region} \tag{9}$$

$\sigma^y_{region}$ is the standard deviation of Y color component in the region, $\mu^y_{region}$ is the mean of Y color component in the region. $\sigma^{cr}_{region}$, $\mu^{cr}_{region}$,$\sigma^{cb}_{region}$, $\mu^{cb}_{region}$ have similar meanings.

Firstly, each image pixel is classified using equation (8) with initial threshold values $a_y = 2.5$, $a_{cb} = a_{cr} = 2.0$. After this coarse skin detection, we can get some skin regions. For each large enough region, we check the homogeneity property. If it is homogeneous, it is considered as true skin region. If it's not homogeneous, the threshold values are decreased by a factor of 0.9 respectively, and skin detection using the new threshold values is performed on the region. This process continues until the skin regions become homogeneous. Fig. 3 gives some experimental result. The top row is coarse skin detection result, the bottom row is the final result. The original images are not presented for offensive reasons.

When more than one faces are detected in the image, each face region is used to construct a skin color model, and each model is used to perform skin detection on the whole image. The final result is a logical OR combination of each of the detected regions obtained respectively with each skin color model.

### 2.6 Image Classification

After face detection and skin detection, high-level features described in section 2.2 can be easily extracted. We used these features to do pornographic-or-not images classification. Although there are many sophisticated classification methods

**Fig. 3.** Skin detection result. The top row are coarse skin detection results, the bottom row are the final refined results. For offensive reasons, original images are not presented here.

such as Support Vector Machine (SVM) and Artificial Neural Network (ANN), we opt for using decision tree[9] for its simplicity and its particular efficiency when our features have human-oriented meanings. Fig. 4 shows the tree we used. Nodes in the tree involve testing a particular feature by comparing its value with a constant threshold. The tree calls firstly for a test on $F1$, and the first two branches correspond to the two possible outcomes. If $F1$ is bigger than a threshold, the outcome is benign. If outcome is the other branch, a second test is made, this time on $F2$. Eventually, whatever the outcome of the tests, a leaf of the tree is reach that dictates the classification result.

## 3 Experimental Results

This section reports testing result of our approach. The performance of our approach is tested on a database of 2196 images, among which 451 images are manually labeled as adult images. Benign images are 1119 human images and other 626 images including animal, plant, landscape images. Adult images and human images are downloaded from the Internet. Other images are from Corel image library. It should be pointed out that adult images and human images we used all contain frontal or near-frontal view faces because our face detector currently can not detect faces with too much out-of-image-plane rotation. The test result is shown in Table 1. For offensive reasons, we do not list the adult images here. Fig. 5 gives some images that successfully detected as benign images

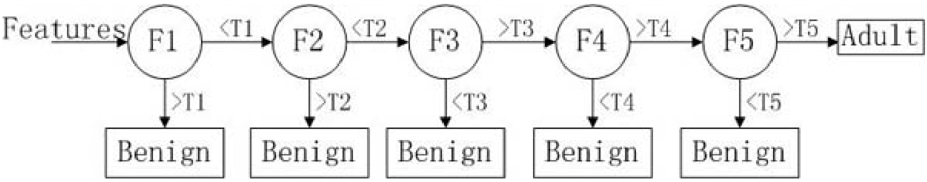


**Fig. 4.** Schematic depiction of image classification process. $T1$, $T2$, $T3$, $T4$ and $T5$ are predefined thresholds.

**Table 1.** Testing Result

| Images | Detected as Adult Images | Detected as Benign Images |
|---|---|---|
| Adult Images (451) | 412 (91.35%) | 39 (8.65%) |
| Benign Human Images (1119) | 99 (8.47%) | 1020 (91.53%) |
| Other Benign Images (626) | 36 (5.75%) | 590 (94.25%) |

a) b) c) d)

**Fig. 5.** Some example images successfully detected as benign images because many faces are detected in a) and b), c) contains large area of face region and d) present face in the center. These images can be wrongly classified as adult images only using low-level visual features.

by our method. These images can be easily classified as pornographic images using low-level features, because these images present large area skin patches.

## 4 Conclusion and Future Directions

In this paper, we propose an approach to automatically discriminate benign images from adult images. Our approach takes advantage of achievements of face detection research. Based on the result of face detection and adaptive skin detection, our approach achieves image classification by analyzing a set of semantically high-level features. Face detection plays a dominant role in our approach. In future work, we will focus on more robust face detector. And we will pay attention to the combination of low-level and high-level visual feature to more effectively detect adult images.

**Acknowledgement.** This work has been financed by the National Hi-Tech R&D Program (the 863 Pro-gram) under contract No.2003AA142140. The authors want to thank Wei Zeng for his constructive suggestions.

## References

1. N2H2 Inc. N2H2 Reports Number of Pornographic Web Pages Now Tops 260 Million and Growing at an Unprecedented Rate. Http://www.n2h2.com.
2. Wang, J., Wiederhold, G., Firshein, O.: System for screening objectionable images using Daubechies' wavelets and color histograms. IDMS'97. Volume 1309, Spring-Verlag LNCS (1997), pp. 20-30
3. Chan, Y., Harvey, R., Smith, D.: Building systems to block pornography. In Eakins, J., Harper, D., eds.: Challenge of Image Retrieval, BCS Electronic Workshops in Computing series (1999), pp.34-40
4. Jones, M.J., Rehg, J.M.: Statistical Color Models with Application to Skin Detection, Technical Report, Cambridge Res. Lab., Compaq Comput. Corp., 1998
5. Zeng, W., Gao, W., Zhang, T., Liu, Y.: Image Guarder: An Intelligent Detector for Adult Images. Asian Conference on Computer Vision. ACCV2004, Jeju Island, Korea, Jan.27-30,2004, pp198-203
6. Fleck, M.M., Forsyth, D.A., Bregler, C., Finding naked people. ECCV. Volume II Spring-Verlag, (1996), pp.593-602
7. Viola, P., Jones, M. J.: Rapid Object Detection Using A Boosted Cascade of Simple Features. IEEE Conference on Computer Vision and Pattern Recognition, Jauai, Hawaii, 2001
8. Phung, S.L., Chai, D., Bouzerdoum, A.: Adaptive Skin Segmentation in Color Images. ICASSP 2003
9. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools with Java Implementations, Morgan Kaufmann, San Francisco, 2000