Pornography Detection Using Support Vector Machine

Yu-Chun Lin (林語君) Hung-Wei Tseng (曾宏偉) Chiou-Shann Fuh (傅楸善)
Department of Computer Science and Information Engineering
National Taiwan University
Taipei 106, Taiwan
sbb@bunny.idv.tw bunny@bunny.idv.tw fuh@csie.ntu.edu.tw

Abstract

As Internet grows quickly, pornography, which is often printed into a small quantity of publication in the past, becomes one of the highly distributed information over Internet. However, pornography may be harmful to children, and may affect the efficiency of workers.

In this paper, we design an easy scheme for detecting pornography. We exploit primitive information from pornography and use this knowledge for determining whether a given photo belongs to pornography or not. In the beginning, we extract skin region from photos, and find out the correlation in skin region and non-skin region. Then, we use these correlations as the input of support vector machine (SVM), an excellent tool for classification with learning abilities. After a period of training SVM model, we achieved about 75% of accuracy, 35% of false alarm rate, and only 14% of mis-detection rate. Moreover, we also provide a simple tool based on our scheme.

1 Introduction

Pornography is ubiquitous all over the Internet. However, the universality of pornography becomes a serious problem, because Internet is an easy-access medium, that is, anyone who knows how to open Web browsers, use search engine and go to portal, even a 10-year-old child, can view pornography by keying in several key words without any forbiddance. Otherwise, some companies forbid their employees to go on Internet to view pornography during office hour, because these companies think it will influence the efficiency of working. Thus, applications for preventing persons, especially minors, from accessing pornography over Internet become more and more important.

By the common components in pornography such as naked people, sexual organs, and sexual intercourse, we can determine if the image is porn graph or not by detecting such contents. However, traditional image matching techniques have their disadvantages. The checking of keywords or large ban database is currently used in Internet firewall to prevent the access of pornography, but these simple schemes are useless for smart pornography provider.

We design a series of processes for detecting pornography without the traditional disadvantages. We extract the skin region from photos first, and then find out the attributes which can be vital evidences for pornography, and finally, we use SVM, an excellent classification tool based on Statistic Learning Theory with reasonable training time, to predict the input images to be pornography or not.

In this paper, we focus on designing a simple, time-efficient scheme for detecting pornography. We also implement a simple application based on our scheme. Our experimental result shows 75% of accuracy, 35% of false-alarm rate, and only 14% of mis-detection rate.

The rest of this paper is organized as follows. Section 2 provides basic concepts of traditional techniques for preventing pornography web site browsing, support vector machine, and color spaces. Section 3 describes our scheme for detecting pornography in detail. Section 4 states our experimental methodology. Section 5 shows our application and experimental result. Section 6 concludes this paper.

2 Backgrounds

2.1 Detecting Pornography

Pornography is often composed of naked persons, special shots of sexual organs, or a picture of sexual intercourse. Furthermore, web sites contain pornography often with some keywords as their names. Therefore, some techniques are often used in preventing the access of pornography over Internet based on such characteristics [1].

1. Examine words in URL (Uniform Resource Locator).
2. Build a huge database to reject sources.
3. Check the content of web pages (by database or psychological analysis).

These techniques would be inefficient and useless under some scenarios. For the first scheme, if we use non-obscene words as site URL, the technique becomes ineffective. The second technique, building reject database, the obscene web sites change their URL addresses quite often, and this technique can not detect newly opened obscene web sites.

The third technique is the most accurate technique for detecting sites containing pornography, since the technique indeed checks the contents of web sites. But classical methods by this concept have some disadvantages. Techniques using pattern recognition might be slow in the matching procedure, and techniques using only skin filter and the percentage of skin in the whole photo may be fast, but it may not be accurate, no one will consider bikini girls as pornography. Both schemes for detecting pornography lack learning abilities, that is, they can not update their knowledge to detect pornography, and decrease the false-alarm and mis-detection rates by themselves. Some papers suggest using neural networks, but the training process for neural networks is quite slow.

2.2 Support Vector Machine

Support Vector Machine (SVM) [10] has similar origin as neural network. It is widely used in classification. Before using SVM in classification, we have to give SVM sufficient information for training a model that can be used in the future. SVM builds up this model based on statistics learning theory, and the process of building up a model and tuning parameters can be finished in reasonable time.

Each record of information is a vector of attributes that should be as representative as possible for that record of data. Since each record is formed as a vector, it is easy to understand why SVM is called SVM.

2.3 Color Spaces

Skin is the most widely used primitive in human image processing research, with applications ranging from face detection and person tracking to pornography filtering. Skin detection techniques can be both simple and accurate, and so can be found in many commercial applications.

The first issue is to choose a color space which is proper for easy recognition of skin pixels. The wanted color space obviously should be the one in which skin pixel cluster is as compact as possible. Besides, it is generally agreed that intensity has a large effect on (red, green, blue) triples and that it is desirable to remove an intensity component before subsequent analysis, but there is much less agreement on how the remaining two color parameters should be defined. Here are some of possible color spaces for skin filtering. Most of them are described conceptually; refer to other resources for further research.

1. Red, Green, Blue (RGB) [1]:

The original color space from input photos. Some easy human perspective on porn could be included to support main skin filter.

2. Hue, Saturation, and Value (HSV) [1][2][3]:

Essentially a description of color space in cylindrical polars, the HSV color space forms a hexacone, with black at the main vertex and white at the center of the base. The central axis of the hexacone gives the V coordinate, while (H, S) is a polar description of a point on the "color wheel".

HSV is the major color model for advanced analysis in our system (the original RGB color space is preserved)

$$\begin{cases} v = \max(r, g, b) \\ s = \dfrac{d}{v} \\ h = \begin{cases} \dfrac{g-b}{6d}, & \text{if } r = v \\ \dfrac{2-r+b}{6d}, & \text{if } g = v \\ \dfrac{4-g+r}{6d}, & \text{if } b = v \end{cases} \\ d = \max(r, g, b) - \min(r, g, b) \end{cases} \quad (2.3.1)$$

3. Log opponent (I, R_g, B_y) [4]

$$\begin{cases} I = \log g \\ R_g = \log r - \log g \\ B_y = \log b - \dfrac{\log g + \log r}{2} \end{cases} \quad (2.3.2)$$

This is an attempt to model human vision system opponent color representation. The contention is that at least one of the log-opponent channels is insensitive to melanin content.

4. Normalized RGB [2] (Two of $\bar{r}, \bar{g}, \bar{b}$)

The aim of this color space is to remove variations in color due to either illumination angle of color.

$$\begin{cases} \bar{r} = \dfrac{r}{r+g+b} \\ \bar{g} = \dfrac{g}{r+g+b} \\ \bar{b} = \dfrac{b}{r+g+b} \end{cases} \quad (2.3.3)$$

5. YCbCr [9]:

YCbCr is a color space that separates red, blue, and green into luminance and chrominance information. Y is in relation to luminance, and Cb, and Cr are in relation to chrominance. The transform between RGB and YCbCr is

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.29900 & 0.58700 & 0.11400 \\ -0.16874 & -0.33126 & 0.50000 \\ 0.50000 & -0.41869 & -0.08131 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 0x80 \\ 0x80 \end{bmatrix} \quad (2.3.4)$$

3 Pornography Detecting Scheme

In this section, we will introduce the scheme we use to detect pornography.

We detect pornography by the following process:

1. Skin filtering
2. Attributes generating
3. SVM-prediction

Of course, before we apply such schemes to real world application, a process of training SVM model is necessary.

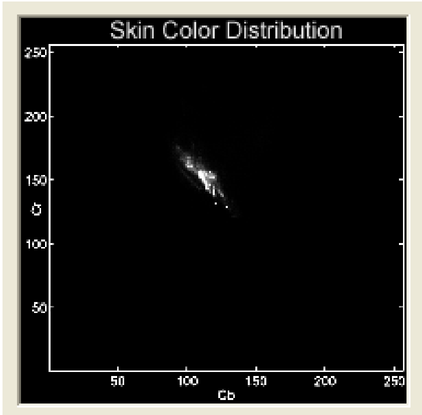

Figure 3.1 The skin distribution in Cb-Cr plane.

We will describe each process of our scheme in the following paragraphs in detail.

3.1 Skin Filtering

For detecting human image, skin is one of the most important features we have to extract. For pornography, this feature is more important since pornography contains significant portion of skin region. Therefore, a simple and precise scheme to extract skin region from whole image is important in designing a pornography detection system.

Skin pixel detection can be simply based on the claim that a pixel of RGB (Red, Green, Blue) color is skin when R > G or R > B, or perhaps both [1][2]. Such a criterion shows some independence of the regional illumination level, and can be applied very efficiently. However, its limitations are wide generalities, where the color red can appear on non-skin objects.

RGB and Hue-Saturation-Value (HSV) are two most common color spaces discussed for detecting skin region.

However, an issue arises, that is, how to extract skin region from every kind of skin colors of different races. Chai and Bouzerdoum [9] claim that skin model based on Cb and Cr values can provide good coverage of all human races. Therefore, we choose YCbCr as our color space for extracting skin region. YCbCr consists of three values, *Y*, *Cb*, and *Cr*. Intensity *Y* is the value in relation to the brightness of the color pixel, *Cb* and

Cr are the chrominance information. Chai and Bouzerdoum [9] describe that the apparent difference among skin colors is due to the brightness, that is, if we want to distinguish skin for all races, the *Y* value is ignorable. Therefore, we just define a skin color distribution in *Cb-Cr* plane as shown in Figure 3.1.

The most common formats of pictures such as GIF (Graphic Interchange Format) and JPEG (Joint Photographic Experts Group) use RGB as their default color spaces. To detect pornography on Internet, we have to translate these images based on RGB into YCbCr color space before skin filtering. Equation 2.3.4 has shown the relationship between these two color spaces, and we use such equation in our application.

3.2 Attributes Generation

In Section 3.1, we described how we use the color distribution in color space for skin region detection. But the skin region only is not enough for determining pornography. The selection of meaningful feature vectors is very important [8]. We describe all the metrics we use to evaluate pictures.

After skin filtering [5][7], we can get so-called skin region, and we use the percentage of skin region in the whole picture as the first metric of attribute. Though bikini girls may be misunderstood in the metric, we believe most pornography contains a large area of skin.

After determining skin pixels in skin filtering, we can exploit more color features from such pixels. By the skin pixels' color distribution in color space [1][3] (Figure 3.2), we can apply pattern-matching technique over a smaller area with small patterns from an obscene database. The obscene database consists of some common but unique pornography skin pixel distribution. Furthermore, we can also add some "back-areas" those often appear in pornography to database. If there is a matching, that means the photo is more likely to be a porno graph, and we can add its weight for determining pornography. This weight becomes a new attribute of a picture, called pornography-weight. Since we choose Cb-Cr color space for detecting picture's color distribution, we collect various kinds of skins pattern first. Each skin pixel distributes a piece of weight to the skin-weight-map. After a mass of samples are processed, a final skin-weight-map could be regarded as a bitmap which contains different gray intensity on its each CbCr pair, this map is used to calculate pornography-weight. Thus, more suspected pixels would prove the detected picture as a porn graph.

Skin pixel geometric distribution [4] could be regarded the skin region's position and size relationship. First, in whole picture we can calculate the percentage of skin pixels in skin region (and its neighboring non-skin region).

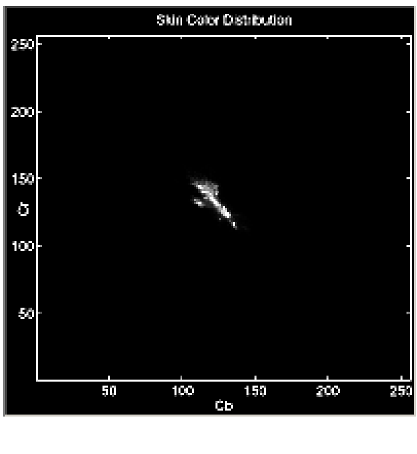

Figure 3.2 A color distribution for its skin region.

Second, the largest region of skin regions, and finally the distances between these regions, since the regions regarded as skin should not part from one another too far. Here traditional K-mean algorithm [5] could be used for grouping regions. These aspects of geometry information could be a good component for special obscene contents in photos. The attribute is called skin-area-correlation.

Skin pixel correlation [5] is the set of relations between skin regions to its nearby non-skin regions. These neighboring regions are sometimes useful information for determining obscenity. Hair is one of the useful information that can be used in detecting pornography. Moreover, this contributes another attribute called hair-inside-body.

The skin filtering process can divide photo into background and skin region. By finding the percentage of "skin pixels" in such skin regions and the smoothing and color correlation of skin region, we can do more accurate SVM-prediction. These two attributes are very important for separating bikini girls from the pornography.

To sum up, skin percentage, pornography-weight, skin area geometric distribution, skin pixels in skin correlation, hair-inside-body, skin-region-smoothness are the six attributes for us to determine whether a photo is porno graph or not.

3.3 SVM-training and SVM-prediction

After we get the six attributes of pictures, we start the prediction process. But before prediction, we must train an SVM model first [10].

To train a SVM model, we have to obtain sufficient classified data as the input of SVM-train [10]. We manually select hundreds of pornography and non-porno graphs from several web sites, and extracted the six attributes. The data we input to SVM-train contain the six attributes and a label that stands for porn or non-porn images. Finally, we input these data again for cross-validation, the process that helps us tune the accuracy of prediction.

After the process described above, we can obtain an SVM model for our system to use.

4 Experimental Methodology

We test the system by using 520 photos. We divide the test set by two main categories: naked people and other obscene contents (such as genitalia). Photos with naked people contain mostly skin area, while the latter contains just some. Implicit pornography usually more complicated and varying color feature distributions. These two categories differ from each other on photos color features.

We train our database for "typical and meaningful" photos first, in order to shorten searching complexity and match explicitly.

After training we have two sets of testing model, one of them consists of elaborately selected photos, the amount of confusing and noisy challenging photos is designed properly for all kinds of situation. Another set of photo sets are collected from the Internet and by scanning or re-photographing images from books or magazines. They show a very wide range of contents.

Photos are in the format of BMP (*.bmp), JPEG (*.jpg/*.jpeg), or GIF (*.gif), which the only three kinds of image exchange format for our developed system.

Images of various complexions (generally White people, Asians, or Blacks) are included into the set. For testing false-alarm rate, we have photos with body with clothes, or other covering objects, or with unexpected background. Even more, overlapped people or obscenity contents are tested. For testing misdetection rate, we try to provide different and implicit obscene photos.

In the aspect of random collecting photos sets, the only appropriate way is to use large numbers of images, drawn from wide variety of sources. Our collections are as follows:

1. Natural photos: animals, landscapes, natural scenes, etc.
2. Magazine photos: peoples, vogue, etc.
3. A set of obscene photos grabbed from few single web sites (no external link traced)

5 Results

5.1 Porn Detector

As shown in Figure 5.1, this is the application

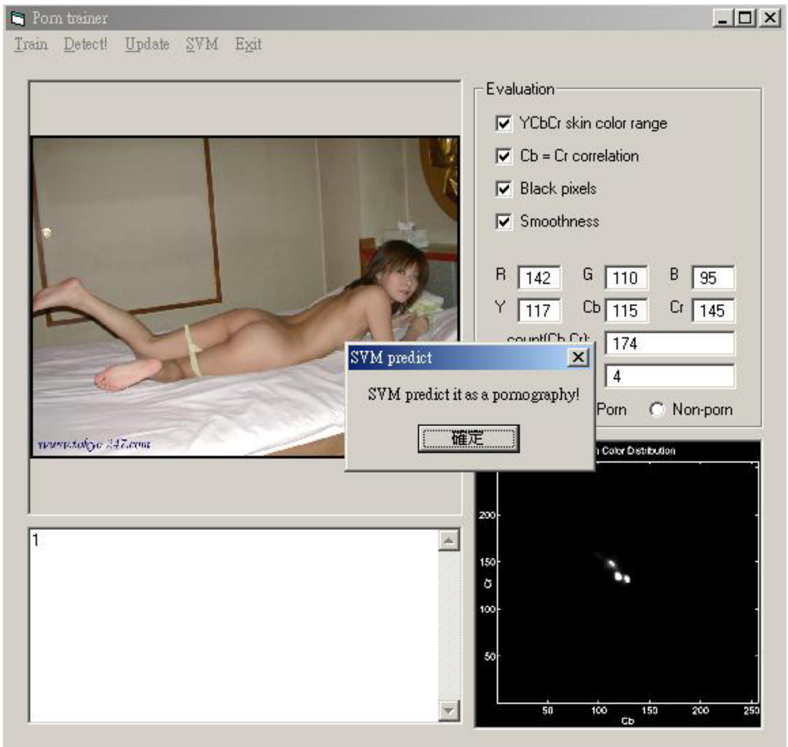

Figure 5.1 The interface of Porn-trainer and the porn graph is detected correctly.

we implemented based on the pornography detecting scheme. Figure 5.1 also shows that the input porn graph is correctly detected.

The basic application has the basic function of detecting pornography based on our initial SVM predict model. The application also has the abilities of training and update by the integration with SVM training libraries. That is, users can enhance the precision of prediction by training the origin SVM model with more new data.

For the convenience of research, the window also shows the values of several attributes we use for evaluating pornography. This simple application also provides the function of batch mode, which is quite convenient for training or detecting large number of files.

Integrating the application into some other applications like web browsers can be used in binding improper accesses of pornography.

5.2 Results

After testing 520 photos (with 236 pornography), we have the total accuracy of 74.6154% (388/520), false-alarm rate of 34.9823%(99/284), and miss-predict rate of 13.9831%(33/236). The accuracy and miss-predict rate is perfect when comparing with previous works, and the low miss-predict rate represents that this scheme is useful for prevent improper access of pornography. However, the false-alarm rate is still too high, we might consider some normal images as pornography quite often.

Figure 5.2 shows a mis-detection example. The input picture is considered as a porn graph, but our program still considers it as non-pornograpy. It is because that the girl in the picture has not take off her brassiere and underpants yet. In the point of view from our program, the girl is similar to bikini girls. This kind of mis-detection is hard to prevent unless we can extract the differences between this kind of picture and bikini girls.

Figure 5.3 shows a false-alarm example. The girl dressed in uniform-skin-color clothes. Since the girl is dressed in skin-color-clothes, our program will consider the skin-color-clothes as a large area of skin, and because of the clothes is uniform colored, it will increase some weights in our attributes. To detect the kind graph correctly, we have to find out a color space that can clearly find the difference between skin and skin-color-clothes.

We think that if we have more domain knowledge of pornography that can help us to extract more useful information in detecting pornography. The results can be better.

6 Conclusions

The pornography detection is important for Internet, but there is still no effective method for preventing improper access of such information.

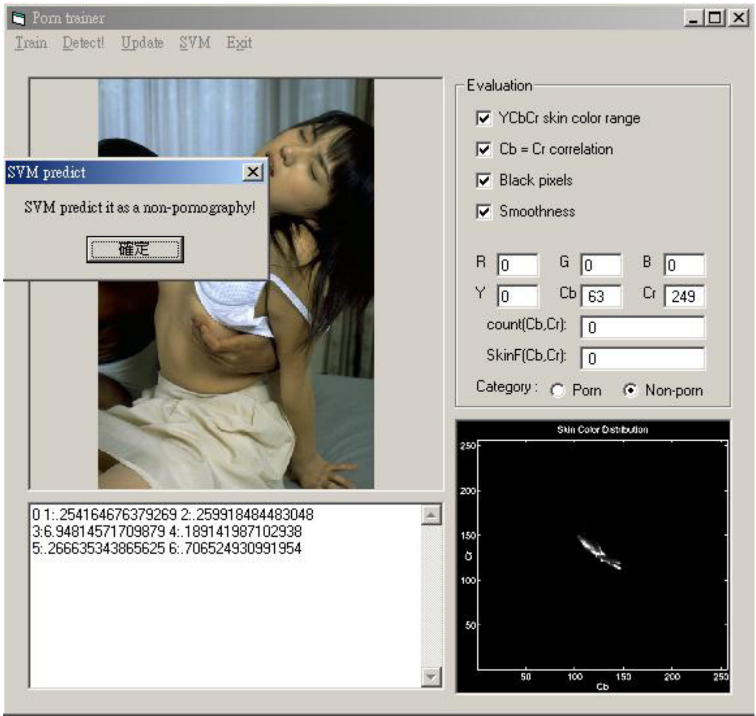

Figure 5.2 Porn graph, but detected as non-porn.

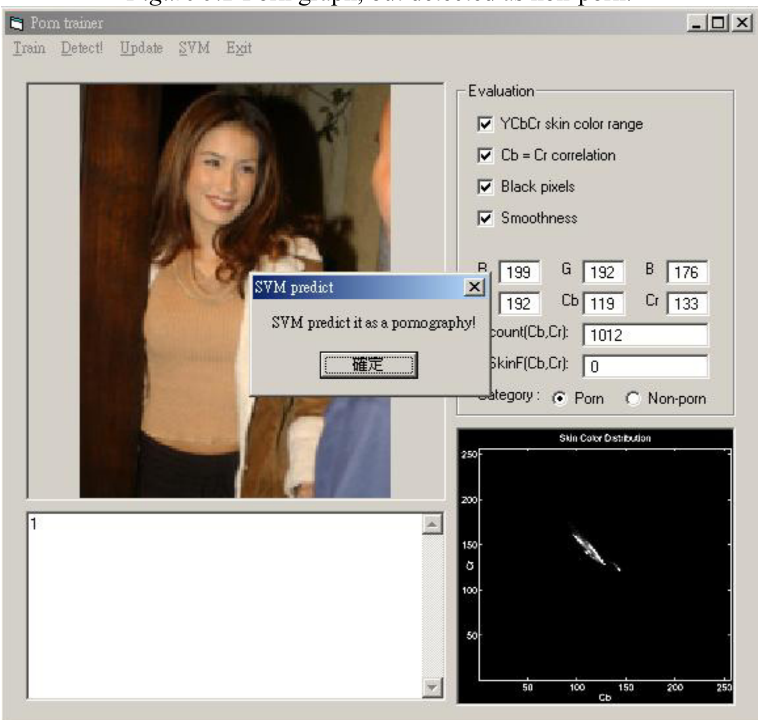

Figure 5.3 Non-porn graph, but detected as porn.

In this paper, we designed a simple and effective scheme on such work. Our scheme shows high total accuracy of 74.62% (388/520), false-alarm rate of 34.86% (99/284), and low mis-detection rate of 13.98% (33/236). The scheme also shows its infinite ability of advance since it is integrated with machine learning functions. By the using of more accurate feature selecting, we believe that the results can be better in the future.

7 Future Works

Pornography detection highly depends on the domain knowledge in this kind graph. With the help of domain knowledge, we can generate more accurate attributes for detecting pornography.

In the future, we will improve the scheme in the following aspects:

1. Generate more respective attributes from pornography.
2. The tuning of prediction model.
3. The more proper classification tools than SVM.

Reference

[1] Y. Chan, R. Harvey, and D. Smith, "Building Systems to Block Pornography," In Eakins, J., Harper, D., Eds.: *Challenge of Image Retrieval*, BCS Electronic Workshops in Computing Series, pp. 34-40, 1999.

[2] D. A. Brown, I. Craw, and J. Lewthwaite, "A SOM Based Approach to Skin Detection with Application in Real Time Systems," *Proceedings of the British Machine Vision Conference*, Manchester, UK, pp. 491-500, 2001.

[3] A. Bosson, G. C. Cawley, Y. Chan, and R. Harvey, "Non-Retrieval: Blocking Pornographic Images," *Proceedings of International Conference on the Challenge of Image and Video Retrieval, Lecture Notes in Computer Science*, Springer-Verlag, London, Vol. 2383, pp. 60-70, 2002.

[4] D. A. Forsyth and M. M. Fleck, "Automatic Detection of Human Nudes," *International Journal of Computer Vision*, Vol. 32 , No. 1, pp. 63-77, 1999.

[5] M. M. Fleck, D. A. Forsyth, and C. Bregler, "Finding Naked People," *Proceedings of European Conference on Computer Vision*, Cambridge, UK, Vol. 2, pp. 592-602, 1996.

[6] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Vols. I and II, Addison Wesley, Reading, MA, 1993.

[7] D. A. Forsyth and M. M. Fleck. "Identifying nude pictures". In *IEEE Workshop on the Applications of Computer Vision*, Sarasoto, FL , pp. 103-108, 1996.

[8] J. Z. Wang, J. Li, G. Wiederhold, and O. Firschein, "System for Screening Objectionable Images," *Computer Communications*, Vol. 21, No. 15, pp. 1355-1360, 1998.

[9] D. Chai and A. Bouzerdoum, "A Bayesian Approach to Skin Color Classification in YCbCr Color Space," *Proceedings of IEEE Region Ten Conference*, Kualu Lumpur, Malaysia, Vol. 2, pp. 421-424, 2000.

[10] C. C. Chang and C. J. Lin, "Training Nu-support Vector Regression: Theory and Algorithms," *Neural Computation*, Vol.14, pp. 1959-1977, 2002.