Color Model Based Real-Time Face Detection with AdaBoost in Color Image

Yuxin Peng*, Yuxin Jin* ,Kezhong He*,Fuchun Sun*, Huaping Liu*, Linmi Tao*
*Department of Computer Science and Technology, Tsinghua University, BeiJing, China
Email: $\{pyx05; jyx05\}$@mails.tsinghua.edu.cn
$\{hkz\text{-}dcs; fcsun; hpliu; linmi\}$@mail.tsinghua.edu.cn

***Abstract*— Face detection plays an important role in Computer Vision. In the past few years, much effort has been put into this field. Some approaches are based on Color Model which is highly expensive. Viola and Jones proposed a real-time approach based on AdaBoost, but it suffered from the risk of overfitting. In this paper, we propose a method to take advantage of Color Model and AdaBoost. A novel Face Region Model will be proposed so as to eliminate a large quantity of non-face regions in only several milliseconds. Skin Strong Classifiers will be trained based on color. Thereafter, the remained regions will be presented into AdaBoost System for refining. Experiment shows that our approach is even faster. Its error rate is lower than or at least equal to Viola and Jones'.**

***Index Terms*— Color Model, Integral Skin Map, Face Region Model, Face Detection, AdaBoost.**

I. Introduction

Face Detection is a basic and important problem. It can be applied into face recognition, User Interface (UI), security and surveillance etc.. In these applications, robust and real-time face detector is required.

Various approaches focused on the skin color due to its special attributions. By clustering a number of skin pixels, the candidate localizations of face can be identified. [3] proposed a color model based on skin and non-skin pixels' histogram and concludes that the model based on histogram is better than on Gaussian Model. However, its result was not satisfactory. [1] introduced light compensation and extraction of eyes, mouths features. But it was highly expensive and not suitable for detecting profile view. [2] proposed a method for skin detection based on Markov Random Field. But the false alarm rate was high.

With the development of pattern recognition, AdaBoost was introduced to face detection. In 2001, Viola and Jones [5] proposed a robust and real-time scheme for face detection. Based on their contribution, [6] extended the scheme into multi-view face detection. Unlike the fact that color model is easily influenced by illumination, current AdaBoost is robust to the change of illumination because it is based on grayscale texture ignorant of the color information. However, due to the loss of information in color, more weak classifiers should be trained to classify the faces from non-faces. In result, such method encounters serious overfitting problem according to [4].

Our approach takes the advantage of both color information and AdaBoost.

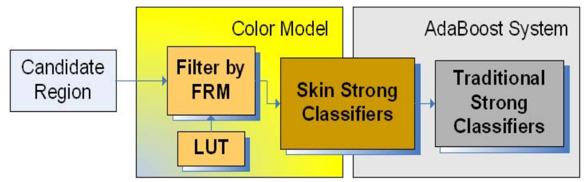


Fig. 1. Frame Work of our approach. The Candidate Region will be firstly filtered by FRM, and then passed to AdaBoost System consisting.

Before AdaBoost System, all sub-windows will be filtered by a novel Face Region Model (FRM) based on a Skin Pixel Model (SPM). A novel Skin Strong Classifier will be trained under AdaBoost System. These methods are used to make a fast elimination of a large quantity of sub-windows with the help of Integral Skin Map (ISM). Experiment shows promising result in high detection rate, low false positive rate and great efficiency.

In Section 2, Color Model including SPM, FRM, and ISM will be introduced. Section 3 will give a brief introduction of our AdaBoost training strategy and structure. Experiment results are given in Section 4. Conclusion is drawn in Section 5. Fig. 1 shows the frame work of our approach.

II. Color Model

Unlike many previous works [1], our skin model does not aim to act as a classifier distinguishing faces from non-faces. We just use the skin model to eliminate as many apparent non-face regions as possible and run in a sufficiently short time. Thus, our approach does not include the step to compensate the illumination's effect and the illuminated pixels will be chosen as ordinary training examples. For real-time application, based on a large quantity of skin and non-skin pixels, we construct a fast Look-Up-Table (LUT) of discriminant function value of single pixel. As to face detection, it is required to identify which region can be viewed as face candidate.

Our Color Model includes two sub-models: Skin Pixel Model (SPM) and Face Region Model (FRM). To further reduce the computation load, Integral Skin Map (ISM) similar to Integral Image [5] will be proposed at the end of this section. The survived regions after elimination by FRM will be feed forwarded to AdaBoost Face Detection.

978-1-4244-1625-7/07/$25.00

A. Skin Pixel Model

Our Skin Pixel Model (SPM) is built based on the histogram of training skin pixels' values in different channels in RGB space.

We divide the range of R,G,B into n_r,n_g,n_b bins respectively for different channels. Then, our histogram over $h[b_r, b_g, b_b]$ pixel's value is a 3-dimensional table. Generally, the skin pixel value can be denoted as $\mathbf{x} = \{x_r, x_g, x_b\}$. We count the value of histogram as:

$$\begin{aligned} h[b_r, b_g, b_b] = \sum_{\mathbf{x}} I(&\frac{b_r}{n_r} \le \frac{x_r}{256} < \frac{b_r+1}{n_r}, \\ &\frac{b_g}{n_g} \le \frac{x_g}{256} < \frac{b_g+1}{n_g}, \\ &\frac{b_b}{n_b} \le \frac{x_b}{256} < \frac{b_b+1}{n_b}) \\ 0 \le x_r, x_g, x_b \le 256& \\ I(\cdot) = \begin{cases} 1 & \text{if condiction } \cdot \text{ is true} \\ 0 & \text{otherwise} \end{cases}& \end{aligned} \tag{1}$$

Intuitively, a simple SPM can be directly converted from the histogram by:

$$P_h(\mathbf{x} \in [b_r, b_g, b_b]) = \frac{h[b_r, b_g, b_b]}{N_{\mathbf{x}}} \tag{2}$$

where $\mathbf{x} \in [b_r, b_g, b_b]$, $N_{\mathbf{x}}$ is total number of training pixels.

We collect more than 1.5 billion skin pixels to calculate the histogram. The large quantity of training pixels can make the result statistically valid. No Gaussian Model has been trained based on the training pixels, according to [3] Gaussian Model performs worse than histogram when the pixel examples' number is sufficiently large.

Equation (2) is the conditional distribution probability function of pixel value when the pixel belongs to skin. In Bayesian, it is the likelihood for pixel to be skin.

$$p(\mathbf{x}|skin) = P_h(\mathbf{x} \in [b_r, b_g, b_b]) \tag{3}$$

In fact, we want to decide whether the given pixel is skin or non-skin: given $\mathbf{x}$, the probability for $\mathbf{x}$ to be skin is $P(skin|\mathbf{x})$. From Bayesian Theory, we know:

$$\begin{aligned} P(skin|\mathbf{x}) &= p(\mathbf{x}|skin)P(skin)/Z \\ P(\neg skin|\mathbf{x}) &= p(\mathbf{x}|\neg skin)P(\neg skin)/Z \end{aligned} \tag{4}$$

where $\neg skin$ means non-skin, Z is a normalize factor.

Only if $P(skin|\mathbf{x}) > P(\neg skin|\mathbf{x})$, $\mathbf{x}$ will be decided as skin pixel. To make such decision, another histogram based on non-skin pixels will be calculated (2 billion pixels) and artificially set a priori probability $P(skin) = P(\neg skin) = 0.5$. Actually, the *a priori* probability $P(skin)$ must be greatly less than $P(\neg skin)$, because skin color is only a minority in the whole color space. If we use the *true* probability, the decision will always be non-skin. We just need a fast elimination of apparent non-skin pixel instead of a precise decision, it is reasonable to set $P(skin) = 0.5$.

This stage not only output the decision whether $\mathbf{x}$ is skin, but also gives its discriminant function value as skin pixel against non-skin pixel.

$$g(\mathbf{x}) = \mathbf{log}\frac{p(skin|\mathbf{x})}{p(\neg skin|\mathbf{x})} \tag{5}$$

Based on these, a fast LUT is constructed whose entry's value is a 2-tuple.

$$\begin{aligned} \mathbf{f}_{LUT}(\mathbf{x}) &= \{I(\mathbf{x} \in skin), g(\mathbf{x})\} \\ I(\mathbf{x} \in skin) &= I(P(skin|\mathbf{x}) > P(\neg skin|\mathbf{x})) \end{aligned} \tag{6}$$

where $\mathbf{x} \in [b_r, b_g, b_b]$.

LUT is previously computed and saved. During the process of detection, it will not be calculated again. With the LUT, we can establish Face Region Model.

B. Face Region Model

For face detection, we should first localize the face. [1] proposed method to group the skin pixels into skin region which is time-consuming (44-116 milliseconds for 150×220 image). We propose a novel Face Region Model (FRM) to decide whether a rectangle region is face region or not. With the help of Integral Skin Map (ISM), this approach will be very fast.

One region is composed of multiple pixels. Denote region as $\mathbf{r} = (\mathbf{x}_0, \mathbf{x}_1, \dots \mathbf{x}_n)$.

Assuming each pixel is independent, the discriminant function value of the region to be skin against non-skin region should be(Log discriminant function is easy for computing):

$$\begin{aligned} g(\mathbf{r}) &= \mathbf{log}\frac{\prod_{i=0}^{n} p(skin|\mathbf{x}_i)}{\prod_{i=0}^{n} p(\neg skin|\mathbf{x}_i)} \\ &= \sum_{i=0}^{n} \mathbf{log}\frac{p(skin|\mathbf{x}_i)}{p(\neg skin|\mathbf{x}_i)} \\ &= \sum_{i=0}^{n} g(\mathbf{x}_i) \end{aligned} \tag{7}$$

We collect more than 10000 frontal face regions and 600000 non-face regions (size 24×24) shown in Fig. 2, and calculate $g(\mathbf{r})$ for each face region to obtain the distribution of $g(\mathbf{r})$ over faces and non-faces samples. Based on the distribution we can decide the lower- and upper-threshold. The value of $g(\mathbf{r})$ of most non-faces is very small. Regions with too large $g(\mathbf{r})$ should also be rejected. Typically, human face region includes many non-skin pixels such as eyes, hair or lip. Thus, a large $g(\mathbf{r})$ value may come from the arm, leg or even yellow mud rather than face so that to set an upper-threshold is meaningful.

Merely dependent upon region's discriminative function value is not enough to eliminate a large quantity of non-face regions. It is likely that there are only several pixels which contain less than 30% pixels in one region but their discriminative function value is too large so that they will be accepted as face candidate. In order to avoid such condition, we give thresholds on the following value:

$$pn_{skin}(\mathbf{r}) = \sum_{i=0}^{n} I(\mathbf{x}_i \in skin)/n \tag{8}$$

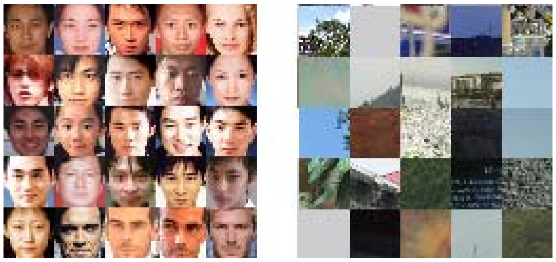


Fig. 2. Face and Non-face regions for training FRM.

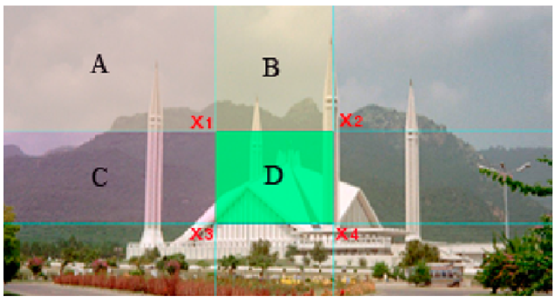


Fig. 3. Integral Skin Map.

The region which satisfies the constraints on both (7) and (8) will be further delivered into AdaBoost Detector. (7) and (8) only involve the addition of the value previously saved in LUT, thus the computation load is small.

C. Integral Skin Map

Integral Image was firstly introduced in [5] to accelerate the computation of the sum of pixels' grayscale value in rectangle region for exhaustive search in an image.

Inspired by this idea, we extend its usage to establish Integral Skin Map (ISM).

Each pixel in an image has a 2-tuple vector according to SPM equation (6). We compute two Integral Images based on $I(\mathbf{x}_i \in skin)$ and $g(\mathbf{x})$ which constitute ISM. With the help of ISM, the computation of (7) and (8) only involves 8 additions. As shown in Fig. 3, calculating $pn_{skin}(\mathbf{r})$ and $g(\mathbf{r})$ for region D is just to calculate $ISM(\mathbf{x}_4) - ISM(\mathbf{x}_2) - ISM(\mathbf{x}_3) + ISM(\mathbf{x}_1)$.

It will greatly increase the efficiency and gain much in time. In addition, we can boost the features extracted based on ISM which will be introduced in next Section.

III. ADABOOST FACE DETECTION

Viola and Jones [5] proposed a rapid AdaBoost detection method using Haar-like features. However it cannot overcome the problem of overfitting. According to theory [4], overfitting may come from the large number of weak classifiers. The difficult negative examples will not be rejected until the number of stages is sufficiently high which results in serious overfitting problem. Our color model can eliminate many negative examples which are difficult for AdaBoost.

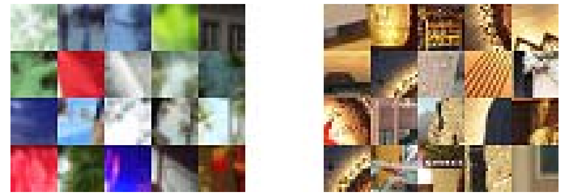


Fig. 4. Left: Difficult Negative Examples for AdaBoost but easy for Color Model. Right: Difficult Negative Examples for Color Model.

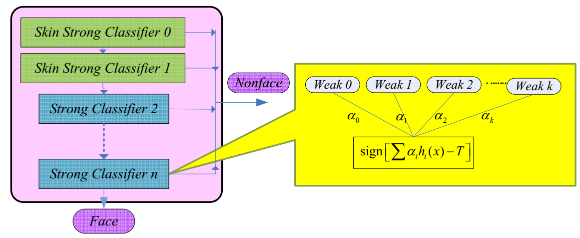


Fig. 5. Detector Structure.

A. Training Negative Examples after Filtering by Color Model

Traditionally, the Haar-like features are extracted based on grayscale image. Experiment shows that the difficult negative examples for AdaBoost are those with the shape and grayscale texture similar to faces. But, they differ greatly from positive examples in color space. Left column of Fig. 4 shows the difficult negative examples for AdaBoost but easy for Color Model.

As mentioned in Section 2, the regions will be filtered by Color Model before they can be sent to AdaBoost Detector. Thus, during training phase, picked from a large quantity of images absent of face regions, only those regions surviving elimination by Color Model are used as negative examples for training a Face Detector, shown in Fig. 5. Our experiment shows that by focusing on these regions, fewer stages and weak classifiers are trained to insure 99.5% detection rate and 50% false positive rate for each stage and the training will not stop until all negative examples are rejected.

Worth to mention, in Viola and Jones' approach [5], first layer can also fast reject 50% non-faces sub-windows. But, the remained sub-windows will be more difficult for the later layers to reject. Our approach can eliminate many difficult sub-windows for the approach in [5] with high efficiency and accuracy, then the remained may be also easy for AdaBoost System.

B. Boosting Weak Classifiers with Haar-like Features on ISM

ISM is introduced in Section 2-C used as an accelerator for color filter. Moreover, features can be extracted from it and thereafter be boosted under AdaBoost scheme.

The output of FRM is only a 2-tuple $[pn_{skin}(\mathbf{r}), g(\mathbf{r})]$ which loses much information about the texture of the region. Skin pixels in a human face are not randomly located in the region, but are highly regularized. Pixels of eyes or hair are non-skin. The eyes do not locate at the bottom of face. Thus, such

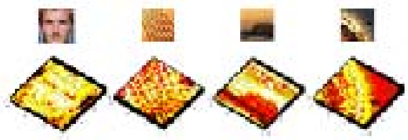


Fig. 6. Skin image for face and non-face regions based on $g(\mathbf{r})$.

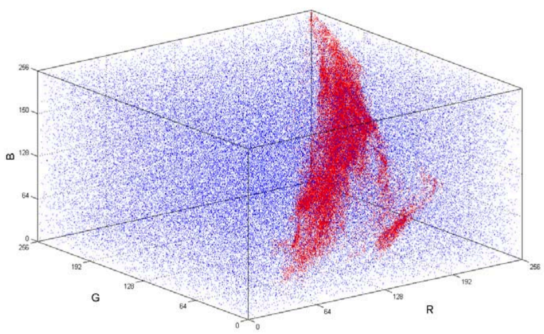


Fig. 7. Distributions of the skin and non-skin pixels in RGB color space.

information should be preserved and useful in training. With $g(\mathbf{x})$, we construct Skin Images shown in Fig. 6.

Fig. 6 shows the skin image of faces and non-faces surviving the elimination by Color Model. The pixels with higher $g(\mathbf{x})$ are represented brighter in the figure. Although the color of pixels in these non-faces is similar to skin, the regions cannot be viewed as faces because of the different Skin Image.

Haar-like features are extracted on the skin image and then strong classifier will be trained based on these features' value and reject a lot of negative examples before further operation. Only the regions accepted by the skin strong classifier will be passed to the traditional strong classifiers [5] based on the grayscale image.

IV. Experiment

A. Color Model

SPM is built by evaluating equation (6) on 1.5 billion pixels from skin texture and 2.0 billion pixels from non-skin texture. Fig. 7 shows the distribution of the two kinds of pixels in RGB space. Red points stand for skin pixels, while blue points for non-skin.

In practice, we divide the range of each channel into 16 bins.

Based on SPM, 2-tuple $[pn_{skin}(\mathbf{x}), g(\mathbf{x})]$ of 10000 face and 600000 non-face regions are calculated. Fig. 8 shows the distribution of $-g(\mathbf{r})$ and ROC curve for FRM Filter. By setting different lower- and up-threshold on FRM, detection rate and false positive rate are changing. In our test, $pn_{skin}(\mathbf{r})$ is set to range from $40\% \sim 95\%$ and threshold on $-g(\mathbf{r})$ is set to insure 95% detection rate with 22.7% false positive rate.

B. Adaboost System for Face Detection

After filtering by FRM, an AdaBoost System is trained according to the structure shown in Fig. 5. But distinct from [5], our system will train additional strong classifiers based on $g(\mathbf{x})$ accelerated by ISM.

We trained three skin strong classifiers with 307 features and fifteen traditional strong classifiers with 4071 features.

As to detection phase, the images with about 380×260 size are exhaustively searched for the square sub-window whose size ranges from 24×24 to 256×256 with increasing scalar factor 1.2. Filtered by Color Model, the remaining sub-windows will be passed to AdaBoost System.

C. Result

Our experiment is conducted on a P4 2.0GHZ, 512 RAM computer. The examples for faces or non-faces are all collected from Internet. Some of them are shown in Fig. 2. We tested our approach on the photos downloaded from [9] which includes 328 frontal faces in 168 images. We can process about 22 frames with size about 380×260 per second. The consumed time is dramatically less than that proposed in [1] (about 1∼10 seconds for 150×220 images) . In order to show our approach's superiority, we also do experiments with the approach proposed in [5] and apply FRM filter before AdaBoost System. The average time, detection rate and false positive rate/number of the comparison is listed in Table. I. Fig. 9 shows some results. The comparison of ROC curves is shown in Fig. 10.

TABLE I

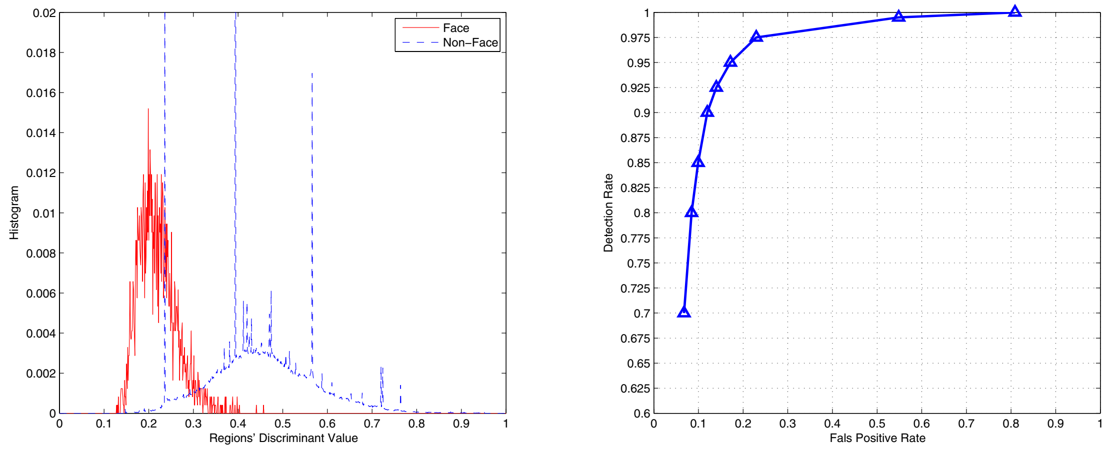
List of consumed time, detection rate and false positive rate/number in average

Approach	Stage	Time	DR	FR/FN
Our Approach	Computing ISM	3.5ms	–	–
	FRM Filter	1.5ms	97.46%	24.32%
	AdaBoost System	39.2ms	91.46%	14
FRM+ Viola and Jones	Computing ISM	3.5ms	–	–
	FRM Filter	1.5ms	97.46%	24.32%
	AdaBoost System	44.1ms	90.24%	18
Viola &Jones	AdaBoost System	67.28ms	90.85%	51

The less computational time attributes to the elimination of large quantity of sub-windows including difficult samples for AdaBoost by FRM Filter. Thus, only a small fraction of sub-windows will be passed to AdaBoost System. Our approach does not only succeed in the less computing time but also in the better generalization effect, because less weak classifiers are trained.

V. Conclusion

This paper proposes a method for face detection in color images which is based on Color Model and refined by AdaBoost System subsequently. But it is not a simple combination of two existing approaches. With the Skin Pixel Model (SPM), novel Face Region Model (FRM) is proposed to act as a filter discriminating faces from non-faces coarsely. This step can be efficient with novel Integral Skin Map (ISM). Furthermore, Haar-like features can be extracted from the ISM



(a) Distribution of $-g(\mathbf{r})$ for face and non-face region. $-g(\mathbf{r})$ has been normalized so that it ranges from 0.0∼1.0.

(b) ROC Curve of FRM Filter.

Fig. 8. FRM Model.

Fig. 9. Some results. Top: Results of our approach. Bottom: Results without filtering by Color Model and Skin Strong Classifiers.

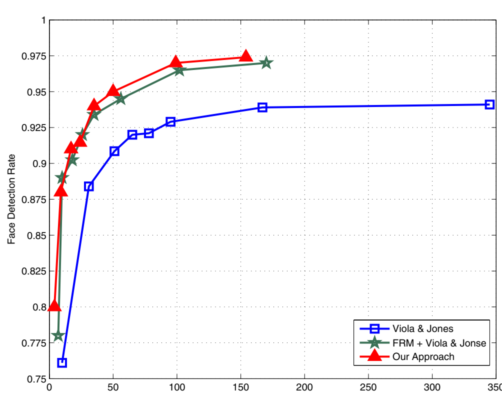


Fig. 10. ROC Curve.

which records the information of pixels' discriminant value. Thereafter, a novel Skin Strong Classifiers are trained based on the discriminant value and integrated into the whole scheme of AdaBoost System.

Primary experiment shows our approach's superiority in efficiency and accuracy. Although most previous works on Color Model are deficient to handle the illumination changing and different racial faces, ours is not sensitive to such influence because our SPM is established based on variant skin and non-skin pixels and FRM is just used as a coarse filter. Risk of overfitting is reduced because many difficult samples for traditional AdaBoost can be eliminated by FRM and Skin Strong Classifiers.

Acknowledgement

This work is supported by the National High-tech Research and Development Plan (No.2007AA04Z232).

References

[1] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K.Jain, "Face detection in color images," *IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI 2002)*, vol.24, no.5, pp.696-706, May 2002.

[2] K. Chenaoua, A. Bouridane, "Skin detection using a Markov Random Field and a new color cpace," *IEEE International Conference on Image Processing(ICIP 2006)*, pp. 2673-2676, October 2006.

[3] Michael J. Jones, James M. Rehg, "Statistical color models with application to skin detection," *IEEE Conf. Computer Vision and Pattern Recognition(CVPR)*, vol.1, pp.274-280, June 1999.

[4] Yoav Freund and Robert E. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Science*, vol. 55, no.1, pp. 119-139, August 1997.

[5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Conference Computer Vision and Pattern Recognition (CVPR 2001)*, vol. 1, pp. 511-518, 2001.

[6] M. Jones and P. Viola, "Fast multi-view face detection," *TR2003*.

[7] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," *IEEE International Conference on Computer Vision and Pattern Recognition(CVPR 2000)*,vol. 1, pp. 746-751, June 2000.

[8] S. Z. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, H. Shum, "Statistical learning of multi-view face detection," *7th European Conference on Computer Vision (ECCV 2002)*, Copenhagen, Denmark, May 2002.

[9] Http://www.fotop.net/micoele.