On Conversion from Color to Gray-scale Images for Face Detection

Juwei Lu
Vidient Systems
4000 Burton Drive, Santa Clara,
CA 95054, USA

K.N. Plataniotis
The Edward S. Rogers Sr. Department of
Electrical and Computer Engineering,
University of Toronto, Canada

Abstract

The paper presents a study on color-to-gray image conversion from a novel point of view: face detection. To the best knowledge of the authors, research in such a specific topic has not been conducted before. Our work reveals that the standard NTSC conversion is not optimal for face detection tasks, although it may be the best for use to display pictures on monochrome televisions. It is further found experimentally with two AdaBoost-based face detection systems that the detect rates may vary up to 10% *by simply changing the parameters of the RGB-to-Gray conversion. On the other hand, the change has little influence on the false positive rates. Compared to the standard NTSC conversion, the detect rate with the best found parameter setting is* 2.85% *and* 3.58% *higher for the two evaluated face detection systems. Promisingly, the work suggests a new solution to the color-to-gray conversion. It could be extremely easy to be incorporated into most existing face detection systems for accuracy improvement without introduction of any extra cost in computational complexity.*

1. Introduction

Many learning-based face detection methods such as those using neural networks and AdaBoost techniques [14, 15, 6] usually work on gray-scale images. Thus given a color input image, $\mathcal{I}$, it has to be first converted to a gray-scale image, $\mathbf{J}$, which is then able to be fed to and processed by these face detection methods. Let $\mathbf{R}$, $\mathbf{G}$, $\mathbf{B}$ be the three color channels of the image $\mathcal{I}$. Classically, the gray-scale image $\mathbf{J}$ is obtained by a linearly weighted transformation:

$$\mathbf{J}(x,y) = \alpha \cdot \mathbf{R}(x,y) + \beta \cdot \mathbf{G}(x,y) + \gamma \cdot \mathbf{B}(x,y) \quad (1)$$

where α, β and γ are the weights corresponding to the three color channels, $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{B}$, respectively, and (x,y) are the pixel location in the input image. The most popular method selects the values of α, β and γ by eliminating the hue and saturation information while retaining the luminance. To this end, a color pixel is first transformed to the so-called NTSC color space from the RGB space by the standard NTSC conversion formula:

$$\begin{bmatrix} \mathbf{Y}(x,y) \\ \mathbf{I}(x,y) \\ \mathbf{Q}(x,y) \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} \mathbf{R}(x,y) \\ \mathbf{G}(x,y) \\ \mathbf{B}(x,y) \end{bmatrix} \quad (2)$$

where $\mathbf{Y}$, $\mathbf{I}$ and $\mathbf{Q}$ represent the NTSC luminance, hue, and saturation components, respectively. Then the luminance is used as the gray-scale signal: $\mathbf{J}(x,y) = \mathbf{Y}(x,y)$. Thus we have

$$\alpha = 0.299,\ \beta = 0.587,\ \gamma = 0.114. \quad (3)$$

The luminance in the NTSC color space is the best for use to display pictures on monochrome (black and white) televisions. However, we found that it may not be optimal to be utilized in face detection. The reason behind is not difficult to see. By analyzing the RGB histogram of a huge number of skin-color pixels, it has been observed that the distribution of the skin-color pixels in the RGB color space is strongly biased toward to the red [3]. The observation indicates that the red takes the major proportion in the skin-color signal. Thus, it is apparently inappropriate for face detection to use the parameter setting of Eq.3, where the red signal is significantly suppressed, while the green one is enhanced too much. This essentially results in a weakened gray-scale face signal.

In this work, we research the RGB-to-gray conversion of Eq.1, which to the best knowledge of the authors, has not been studied from the perspective of face detection, and attempt to find new values of (α, β, γ), which are optimal for face detection. Our research results reveal that 1) to get a strong gray-scale face signal, one should select the values of (α, β, γ) that satisfy $\alpha > \beta > \gamma$ so as to enhance face signals while suppress other signals, which could be considered to be noise in face detection; 2) A face detection system with the best found parameter setting $\alpha^* = 0.85$, $\beta^* = 0.10$ and $\gamma^* = 0.05$ outperforms the same system with the traditional setting (Eq.3) up to at least 2% in detection rate with little influence on false positive rate. The

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

results are obtained with two AdaBoost-based face detection systems. One is from the well-known Intel OpenCV library [7, 12]. Another is developed by the authors [8].

2. Statistical Study on Skin-Color Pixels

In the last two decades, the skin color of human face has been widely studied in literature [16, 3, 4, 11, 13, 5]. A great deal of experiments reveal that the human skin-color distribution tends to cluster in a small region in various color spaces, although in reality skin-colors of different people appear to vary over a wide range. Also, one further found that the skin-color distribution is strongly biased toward the red direction [3]. Motivated by the founding, we revisit the RGB-to-gray conversion of Eq.1, which to best knowledge of the authors, has not been studied from the perspective of face detection.

2.1. Histogram Analysis

In this work, we built a color image database, which contains a set of 834 color images with 555 obtained from the Learning-Based Multimedia Group, University of California, Santa Barbara [17] and the remaining collected by the authors. These images cover a wide range of skin pixel variations in lighting conditions, ethnic groups, and acquisition tools *etc.* . The skin-color regions in each image are manually labeled. This constitutes a skin-color database of 50 million pixels and a nonskin-color database of 400 million pixels. Using the two databases, we generate two 3D RGB histograms, the skin-color histogram $\mathbf{H}_s(R,G,B)$ and the nonskin-color histogram $\mathbf{H}_n(R,G,B)$. For the purpose of visualization, each 3D histogram is decomposed into three 1D marginal histograms corresponding to the three color channels, R, G, B, respectively by the following transformations:

$$\mathbf{H}_{s,X} = \sum_{Y=0}^{255}\sum_{Z=0}^{255}\mathbf{H}_s(X,Y,Z), \tag{4}$$

$$\mathbf{H}_{n,X} = \sum_{Y=0}^{255}\sum_{Z=0}^{255}\mathbf{H}_n(X,Y,Z), \tag{5}$$

where X denotes any one of the three color channels, and Y, Z denote the other two.

The 1D histograms of skin-color and nonskin-color pixels are shown in Fig.1:Left and Right, respectively. Apparently, it can be seen from the histograms that among the skin-color signal, the strongest part is the red component, next is green, and the weakest is blue. In comparison with this, the nonskin-color signal is almost distributed with equal strength in the R, G, B channels. Thus, we have reason to believe that for face detection, it is inappropriate to use the weighting parameters of Eq.3, where the signal in the most important channel (red) is significantly suppressed, while those in the other two channels of minor importance are enhanced too much. To address the problem, we propose here a new parameter selection criterion, which suggests

$$\alpha > \beta > \gamma > 0, \text{ subject to } \alpha+\beta+\gamma = 1. \tag{6}$$

If all the nonskin-color signals are considered to be noise, the motivation behind Eq.6 is rather straight forward, that is, enhancing the signal-noise ratio.

2.2. Discriminant Analysis

Rather than traditional signal processing, the problem of converting color to gray-scale could be considered as dimension reduction from a discriminative learning point of view. To this end, the linear discriminant analysis method (LDA) is applied to find the most discriminant mapping from the 3-D color space to the 1-D gray-scale subspace:

$$\phi_{opt} = \arg\max_{\phi}\left\{\phi\mathbf{S}_b\phi^T\right\}, \quad \phi = [\alpha,\beta,\gamma] \tag{7}$$

where $\mathbf{S}_b$ is the between-class scatter matrix of skin- and nonskin-color pixel samples. It is not difficult to see that the optimal mapping ϕ_{opt} is the eigenvector of $\mathbf{S}_b$, which corresponds to its maximal eigenvalue. For the data sets described in Section 2.1, it is found to have

$$\phi_{opt} = [0.4331, 0.3147, 0.2523] \tag{8}$$

The result is consistent with the conclusion (Eq.6)from the histogram analysis.

3. Experimental Results

3.1. Description of Face Detection Systems

In this section, we introduce two AdaBoost-based face detection systems to be used to perform an empirical analysis to demonstrate the above conclusion. One system is developed by the authors [8]. It has exhibited excellent performance in practical applications. The system consists of four key components: a skin-color based classifier, an edge based classifier, an AdaBoost based classifier, and a post processor. By simply removing the skin-color based classifier, we can obtain a gray-scale version of the efficient face detection system (hereafter EFD-Gray), which is used in this work. Compared to the Viola-Jones method [15], the major improvement of the proposed EFD-Gray method is that we extend the haar-like features to 3×3 rectangular features, the mother template of which is shown in Fig.2. The scalar feature corresponding to the template is computed by a linearly weighted combination of the pixel sums of the nine gray rectangles,

$$f_n = \sum_{i=1}^{3}\sum_{j=1}^{3}\beta_{ij}\cdot sum(\mathbf{W}_{ij}); \tag{9}$$

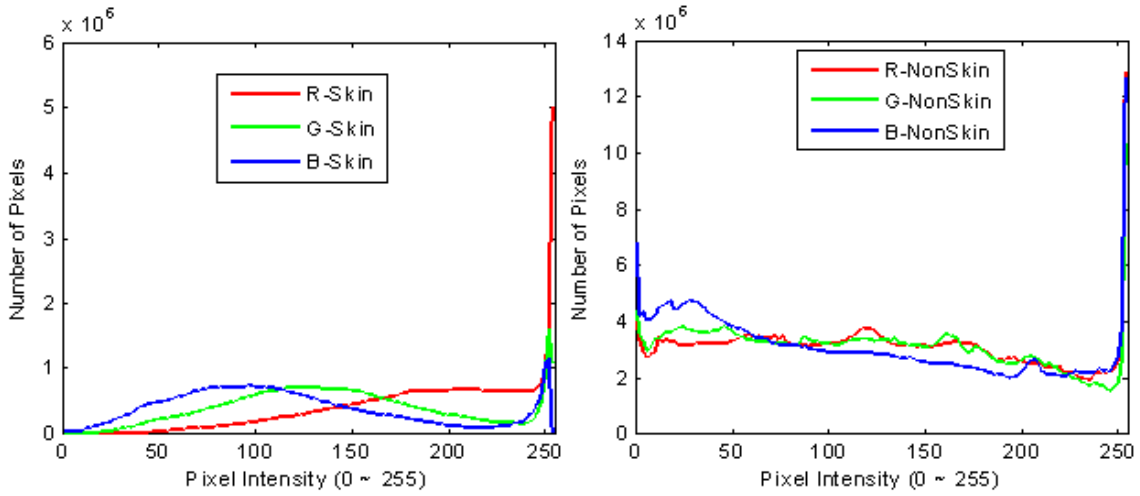

Figure 1. Left: histogram of skin-color pixels; Right: histogram of nonskin-color pixels.

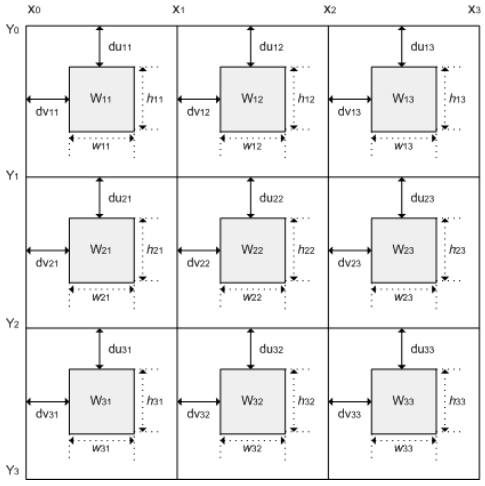

Figure 2. The mother template of the 3×3 rectangular features used to replace the haar-like features in EFD-Gray [8]. New features can be generated by adjusting the template parameters: $\{X_k\}_{k=0}^{3}$, $\{Y_k\}_{k=0}^{3}$, $\{du_{ij}\}_{i,j=1}^{3}$, $\{dv_{ij}\}_{i,j=1}^{3}$, $\{w_{ij}\}_{i,j=1}^{3}$, $\{h_{ij}\}_{i,j=1}^{3}$, and $\{\beta_{ij}\}_{i,j=1}^{3}$(rectangle combination weights).

where $sum(\mathbf{W}_{ij})$ denotes the pixel sum of the gray rectangle $\mathbf{W}_{ij}$. By adjusting the template parameters, we can create thousands of novel features. Some have been shown to be particularly efficient to learn rotated frontal and profile faces [8]. In addition, the feature learning is enhanced by introducing a cross-validation mechanism in the boosting process [10].

The other face detection system is from the well-known Intel OpenCV library (hereafter OpenCVFD) [7, 12]. It is actually a slightly revised variant of the Viola-Jones method [15]. There are several trained boosting cascades available in the OpenCV libary. In this work, we adopt two cascades built with stump-type weak classifiers for frontal face detection. One has a file name of *haarcascade_frontalface_default.xml* (hereafter OpenCVFD-DAB), trained with discrete AdaBoost [1], and the other has a file name of *haarcascade_frontalface_alt.xml* (hereafter OpenCVFD-GAB), trained with the so-called "gentle AdaBoost" [2].

3.2. Optimal Parameter of RGB-to-Gray Conversion for Face Detection

Using the two face detection systems, EFD-Gray and OpenCVFD, we attempt to experimentally verify the conclusion of Section 2, and find the best values of the weighting parameters of RGB-to-Gray conversion, denoted as $(\alpha^*, \beta^*, \gamma^*)$, for face detection. To this end, we generated a database of 1901 color images containing 3249 human faces. These images, most collected from internet, cover a wide range of facial variations in illuminations, expressions, races, ages, resolutions, color casting, occlusions, rotations in plane (roughly falling in $[-45°, +45°]$), and rotations out of plane (roughly falling in $[-60°, +60°]$).

Let M_t and N_t be the total number of the test images and the total number of the faces in the images, respectively. The detect rate ϑ and false positive rate κ are measured by

$$\vartheta = \frac{\text{number of faces detected}}{N_t}, \tag{10}$$

$$\kappa = \frac{\text{number of false positives}}{M_t}. \tag{11}$$

To reduce the variance of performance evaluation, we created five EFD-Gray systems, each one being trained with a different set of training samples. All the EFD-Gray results reported below have been averaged over the five systems. Similarly, the OpenCVFD results have been averaged over

the two systems: OpenCVFD-DAB and OpenCVFD-GAB. It should be noted at this point that the training samples are gray-scale images obtained by using the standard NTSC conversion, Eq.3.

Both the detect rate and false positive rate are functions of the weighting parameters, (α, β, γ). To find the best parameter values for face detection, we conduct an exhaustive search on the domain of (α, β, γ), which are subject to $\alpha + \beta + \gamma = 1$ with $\alpha \in [0, 1]$, $\beta \in [0, 1]$, and $\gamma \in [0, 1]$. The search step for each parameter is set as 0.05, and the EFD-Gray and OpenCVFD systems are applied to each valid combination of (α, β, γ) to calculate its corresponding $\vartheta(\alpha, \beta, \gamma)$ and $\kappa(\alpha, \beta, \gamma)$. For the sake of visualization and analysis, we use marginal detect rates and false positive rates obtained by

$$\vartheta(\alpha) = \frac{1}{O} \sum_{\beta} \sum_{\gamma} \vartheta(\alpha, \beta, \gamma), \tag{12}$$

$$\vartheta(\beta) = \frac{1}{P} \sum_{\alpha} \sum_{\gamma} \vartheta(\alpha, \beta, \gamma), \tag{13}$$

$$\vartheta(\gamma) = \frac{1}{Q} \sum_{\alpha} \sum_{\beta} \vartheta(\alpha, \beta, \gamma), \tag{14}$$

$$\kappa(\alpha) = \frac{1}{O} \sum_{\beta} \sum_{\gamma} \kappa(\alpha, \beta, \gamma), \tag{15}$$

$$\kappa(\beta) = \frac{1}{P} \sum_{\alpha} \sum_{\gamma} \kappa(\alpha, \beta, \gamma), \tag{16}$$

$$\kappa(\gamma) = \frac{1}{Q} \sum_{\alpha} \sum_{\beta} \kappa(\alpha, \beta, \gamma), \tag{17}$$

where O, P and Q are normalization constants. The obtained marginal detect rates and false positive rates are shown in Fig.3 for EFD-Gray and Fig.4 for OpenCVFD, respectively. It is interesting to observe that the two figures appear very similar.

It can be seen from Fig.3:Left that the detect rate of EFD-Gray increases from 68.3% to 78.3% as α increases. However, it drops dramatically from 76.4% to 63.8% and slightly from 73.9% to 72.2% as γ and β increases, respectively. Such a phenomena is consistent with our analysis on skin-color pixels. On the other hand, it can be observed from Fig.3:Right that the changes of the parameters have little influence on the false positive rate. The maximal difference between the false positive rate is around 0.3%, 0.4%, and 0.5% as α, β, and γ change, respectively. Similar phenomena can be also observed from the results of OpenCVFD shown in Fig.4. Such a consistency allows us to have some optimistic reasons to believe that our finding is somewhat independent of the face detection methods, although it needs to be demonstrated with more methods.

Considering the tradeoff between the detect rate and false positive rate, it is found that the optimal parameter configuration is around $\alpha^* = 0.85$, $\beta^* = 0.10$ and $\gamma^* = 0.05$. It corresponds to a detect rate of 78.25% with a false positive rate of 4.05% for EFD-Gray, and a detect rate of 68.17% with a false positive rate of 5.26% for OpenCVFD, respectively. In comparison with this, the traditional configuration of Eq.3 produces a detect rate of 75.4% and a false positive rate of 4.42% for EFD-Gray, and a detect rate of 64.59% and a false positive rate of 5.13% for OpenCVFD, respectively. An increase of 2.85% or 3.58% in detect rate seems not to be very significant. However, it should be noted that it is achieved by simply adjusting the weighting parameters in the RGB-to-gray conversion without making any change to the face detection methods and introducing any extra computational costs. Also, we conducted similar experiments on other test image sets, and high consistency was observed from the obtained results.

3.3. Discussions and Future Works

On the other hand, it can be seen that the optimal mapping $\phi_{opt} = [0.4331, 0.3147, 0.2523]$ found by LDA is far away from the one obtained by the exhaustive search. The former corresponds to a detect rate of 75.96% with a false positive rate of 4.23% for EFD-Gray, and a detect rate of 65.35% with a false positive rate of 4.60% for OpenCVFD, respectively. They are only slightly better than the results of the standard NTSC conversion. The limited success may be attributed to the linear nature of LDA, which requires each class of samples are subject to a Gaussian distribution. It is a rather strong condition, which is hard to be met in real-world face detection tasks. Thus nonlinear discriminant analysis such as kernel LDA [9] could be a more suitable solution.

Also, it should be noted at this point that both the EFD-Gray and OpenCVFD detectors are trained with the gray-scale images obtained by using the standard NTSC conversion, Eq.3. We have the reason to believe that an even bigger improvement could be expected if both training and test images are converted with the optimal parameter, $(\alpha^*, \beta^*, \gamma^*)$.

In addition, it is interesting for future works to extend the proposed idea to other color spaces such as YCbCr space and HSV. Many researchers have shown that human faces lie on a much more compact manifold in these color spaces, compared to other distracting objects. Thus there is a real hope for further performance improvements.

4. Conclusion

In the work, we revisit the conversion of RGB to gray-scale images from the perspective of face detection, which to the best knowledge of the authors has not been studied before. Our work reveals that most energy of a color facial

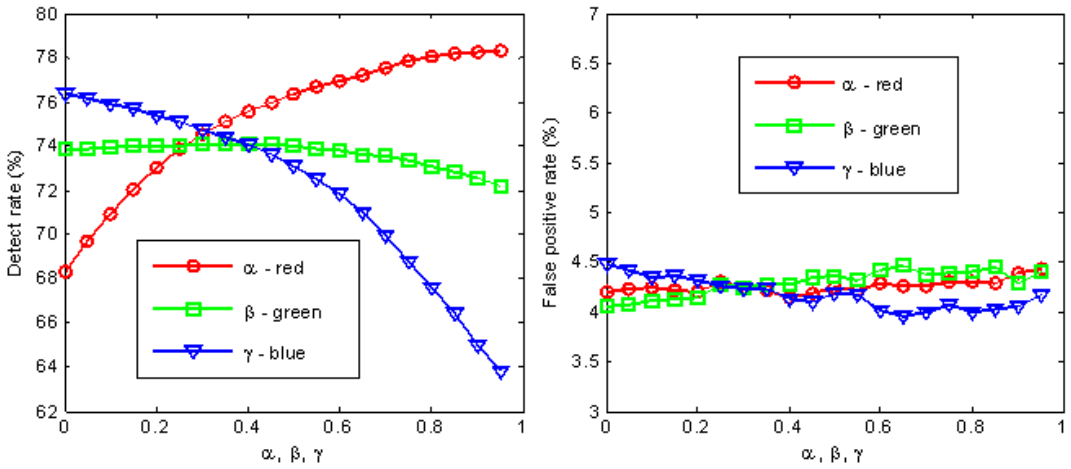

Figure 3. Test results obtained with the proposed face detection system (EFD-Gray). Left: Detect rate as functions of α, β, or γ; Right: False positive rate as functions of α, β, or γ.

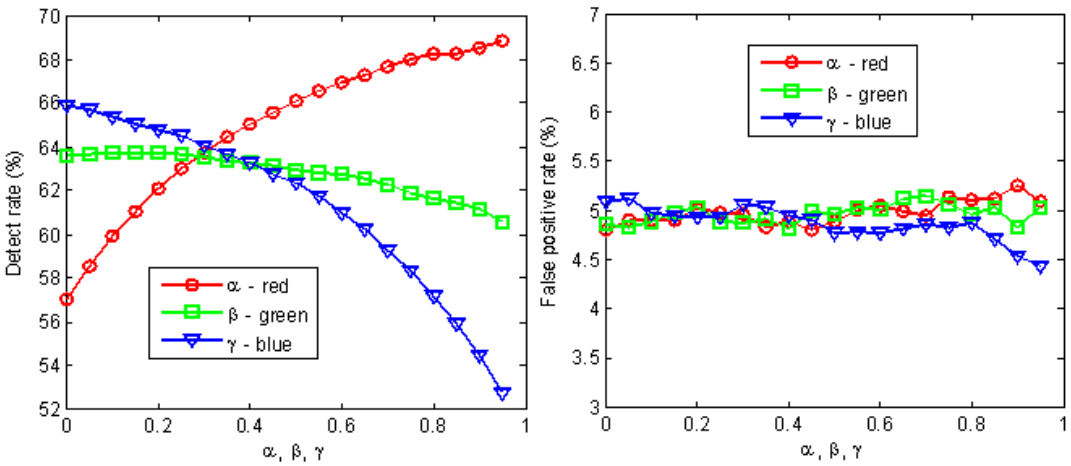

Figure 4. Test results obtained with OpenCV face detection system. Left: Detect rate as functions of α, β, or γ; Right: False positive rate as functions of α, β, or γ.

signal is distributed in the red channel, and the selection for the weighting parameters in the RGB-to-Gray conversion should meet the criterion of $\alpha > \beta > \gamma$ subject to $\alpha + \beta + \gamma = 1$ so as to maximize the signal-noise ratio in face detection. It is further found in the experiment that the detect rate is able to increase approximately up to 10% as α increases, while the change of false positive rate is less than 0.5%. Compared to the standard NTSC conversion, the detect rate with the best found configuration of (α, β, γ) is 2.85% and 3.58% higher for the two evaluated here systems, EFD-Gray and OpenCVFD, respectively.

The RGB-to-gray solution proposed here is extremely simple and efficient. Moreover, it is not difficult to see that the solution is independent of the face detection methods, although it was only evaluated with the EFD-Gray and OpenCVFD systems in the work. We expect that by incorporating such an easy solution, the face detect accuracy of most existing face detection systems could get improved without introduction of any extra computational cost.

References

[1] Y. Freund and R. E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting". *Journal of Computer and System Sciences*, 55(1):119–139, 1997. 3

[2] J. Friedman, T. Hastie, and R. Tibshirani. "Additive logistic regression: a statistical view of boosting". *Annals of Statistics*, 2000. 3

[3] M. J. Jones and J. M. Rehg. "Statistical color models with application to skin detection". *International Journal of Computer Vision*, 46(1):81 – 96, January 2002. 1, 2

[4] C. Jones-III and A. Abbott. "Optimization of color conversion for face recognition". *EURASIP Journal on Applied Signal Processing*, pages 522–529, 2004. 2

[5] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. "A survey of skin-color modeling and detection methods". *Pattern Recognition*, 40(3), 2007. 2

[6] S. Z. Li and Z. Zhang. "FloatBoost learning and statistical face detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1112–1123, September 2004. 1

[7] R. Lienhart, A. Kuranov, and V. Pisarevsky. "Empirical analysis of detection cascades of boosted classifiers for rapid object detection". In *Proceedings of the 25th Pattern Recognition Symposium, DAGM'03*, pages 297–304, Madgeburg, Germany, 2003. 2, 3

[8] J. Lu. "Method and apparatus for detecting faces in digital images". *United States Patent Application No. 20080107341*, Published in 8 May 2008. 2, 3

[9] J. Lu, K. Plataniotis, and A. Venetsanopoulos. "Face recognition using kernel direct discriminant analysis algorithms". *IEEE Transactions on Neural Networks*, 14(1):117–126, January 2003. 4

[10] J. Lu, K. Plataniotis, A. Venetsanopoulos, and S. Z. Li. "Ensemble-based discriminant learning with boosting for face recognition". *IEEE Transactions on Neural Networks*, 17(1):166–178, January 2006. 3

[11] B. Martinkauppi and M. Pietikinen. Facial skin color modeling. In S. Li and A. Jain, editors, *Handbook of Face Recognition*, pages 109–131. Springer, 2005. 2

[12] OpenCV-MainPage. Open source computer vision library (opencv) main page. *Available at: http://www.intel.com/technology/computing/opencv*, last viewed in December 2006. 2, 3

[13] S. L. Phung, A. Bouzerdoum, and D. Chai. "Skin segmentation using color pixel classification: Analysis and comparison". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):148–154, January 2005. 2

[14] H. A. Rowley, S. Baluja, and T. Kanade. "Neural network-based face detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:23–28, 1998. 1

[15] P. Viola and M. J. Jones. "Robust real-time face detection". *International Journal of Computer Vision*, 57:137–154, May 2004. 1, 2, 3

[16] J. Yang and A. Waibel. "Tracking human faces in real-time". *Technical report (CMU-CS-95-210), Carnegie Mellon University*, 1995. 2

[17] Q. Zhu, K.-T. Cheng, C.-T. Wu, and Y.-L. Wu. "Adaptive learning of an accurate skin-color model". In *Proceedings of the IEEE 6th International Conference on Automatic Face and Gesture Recognition*, Seoul, Korea, May 2004. 2