Journal of Information Hiding and Multimedia Signal Processing ©2011 ISSN 2073-4212
Ubiquitous International Volume **2**, Number **2**, April **2011**

Feature-based Face Detection Against Skin-color Like Backgrounds with Varying Illumination

Wu-Chih Hu[1], Ching-Yu Yang[1], Deng-Yuan Huang[2], and Chun-Hsiang Huang[1]

[1]Department of Computer Science and Information Engineering,
National Penghu University of Science and Technology, Taiwan
300, Liu-Ho Rd., Makung, Penghu 880, Taiwan
wchu@npu.edu.tw; chingyu@npu.edu.tw; d94405051@npu.edu.tw

[2]Department of Electrical Engineering, Dayeh University, Taiwan
168, University Rd., Dacun, Changhua 515, Taiwan
kevin@mail.dyu.edu.tw

Received January 2010; revised August 2010

Abstract. *A three-stage scheme for real-time reliable face detection is presented. The proposed three-stage scheme is a feature-based method that is mainly based on skin color and facial features. Skin regions are obtained using a YCbCr skin-color model in the first stage. In the second stage, a face template measure is used to obtain face candidates and then a suitable face box is used to effectively remove non-face regions from the face candidates. Finally, facial features are measured to detect faces from face candidates in the third stage. Experimental results show that the proposed method has good performance in the face detection of faces in various poses, faces in skin color-like backgrounds, faces under varying illumination, and faces of various races.*
Keywords: Face Detection, Varying Illumination, Skin Region, Skin-color like Background

1. **Introduction.** Visual surveillance has received increasing attention. Visual surveillance has a wide range of applications, such as event detection and description [1, 2], object tracking [3, 4], person or vehicle counting [5-7], person or face tracking [8, 9], and security [10]. Recently, the demands of surveillance system in crossroads have greatly increased with an increase in pedestrian safety. In the past years, the applications of surveillance system in crossroads were primarily used in traffic events or theft cases. Besides these cases, the applications of surveillance system in crossroads need to be further extended. Therefore, intelligent surveillance system in crossroads has significantly attracted much attention of researchers in recent years. Face detection plays an important role in the intelligent surveillance system in crossroads. Face detection has attracted a lot of attention due to its wide variety of applications, such as face recognition [11, 12], facial expression recognition [13], facial expression analysis [14], speaker recognition [15], VoIP (voice over IP) security [16], and video surveillance.

Face detection techniques can be generally classified into four categories [17]: template matching-based methods, knowledge-based methods, machine learning-based methods, and feature-based methods. Template matching-based methods use the similarity between an input image and the template. These methods are easy to implement, but they are scale-dependent, rotation-dependent, and computationally expensive.

Knowledge-based methods use the knowledge about a basic face, such as the ellipse shape and the triangle feature, to obtain the final region of a face. Knowledge-based methods can greatly reduce computational cost, but they are rotation-dependent.

Machine learning-based methods use a lot of training samples of faces and non-faces to learn to identity faces. Machine learning-based methods have high detection rates, but their accuracy depends on the training samples. Most of these methods focus on frontal faces or faces with fixed orientations.

Feature-based methods use low-level features (such as color, edge, shape, and texture) to detect faces. Feature-based methods are scale-independent, rotation-independent, and fast. Yang et al. [18] found that skin color, which is scale-independent and rotation-independent, can greatly reduce computational cost of face detection. Kakumanu et al. [19] presented a survey of skin-color modeling and detection methods. However, skin-color methods do not robustly detect skin regions in the presence of skin-color like backgrounds and varying illumination.

Some researchers have integrated the machine learning-based method and the feature-based method [17, 20]. Skin regions are obtained using skin-color detection, and then classifiers (AdaBoost algorithm) based on Haar-like features (local features) and/or Gabor features (global features) are used to detect faces from skin regions. This scheme can greatly reduce computational complexity and robustly detect faces in skin-color like backgrounds. However, the accuracy depends on the training samples.

In the present study, three-stage scheme is proposed for the face detection of pedestrians in intersection monitoring. The proposed method is a feature-based method that is mainly based on skin color and facial features. Lighting compensation, image resizing, the YCbCr skin-color model, morphological processing, and fast 4-connected component labeling are used to obtain skin regions in the first stage. In the second stage, a face template is used to detect face candidates from the skin regions. Since non-face regions may be included in the obtained face candidates, such as necks and arms, a suitable face box is used to remove the non-face regions from the face candidates. In the third stage, a facial feature measure is used to detect faces from face candidates. In the facial feature measure, the luminance variation and the histogram bin-based skin distribution are used to identify face candidates as either faces or non-faces.

Experimental results show that the proposed method has good performance in face detection. Using the proposed method, the poor face detection in the presence of skin-color like backgrounds and varying illumination can be greatly improved.

The rest of this paper is organized as follows. The proposed three-stage scheme of face detection is described in Section 2. Section 3 presents experimental results and their evaluations. Finally, the conclusion is given in Section 4.

2. Three-stage Scheme of Face Detection.

A feature-based three-stage scheme is proposed for face detection. The first stage is used to obtain skin regions. Face candidates are detected in the second stage. Finally, faces are detected from the face candidates in the third stage.

2.1. Skin-region detection.

In order to detect faces under varying illumination, reference white [21] is used for lighting compensation. In this method, the top 5% of luminance values in the image is regarded as the reference white if the number of these pixels is sufficiently large (> 100). The RGB components of the original image are then adjusted so that the average gray value of the reference-white pixels is linearly scaled to 255. Let $i \in [l_u,\ 255]$ be the top 5% gray levels and f_i be the number of pixels of gray level i in the image. Therefore, the modified RGB components can be estimated using Eq. (1) and Eq. (2).

$$Y_{new} = (Y_{old}/M_{top}) \times 255 \ , \ Y \in \{R, G, B\} \tag{1}$$

$$M_{top} = \sum_{i=l_u}^{255} i \cdot f_i \bigg/ \sum_{i=l_u}^{255} f_i \tag{2}$$

Furthermore, the low-low (LL) subband of the discrete wavelet transform (DWT) is used to resize the original image to 1/16 of its original size to reduce the computational cost. In the resized image, the YCbCr skin-color model is used to detect skin regions. Garcia and Tziritas [22] found that the YCbCr model has a better cluster of skin color than do other skin-color models. They used two sets of eight conditions depending on two areas of the color space in order to approximate the distribution in the light and dark extreme cases, as defined in Eq. (3) and Eq. (4).

$$\begin{array}{l} if \ (Y > 128) \\ \theta_1 = -2 + \frac{256 - Y}{16}; \ \theta_2 = 20 - \frac{256-Y}{16}; \ \theta_3 = 6 \ ; \ \theta_4 = -8 \\ if (Y \leq 128) \\ \theta_1 = 6 \ ; \ \theta_2 = 12 \ ; \ \theta_3 = 2 + \frac{Y}{32}; \ \theta_4 = -16 + \frac{Y}{16} \end{array} \tag{3}$$

and

$$\begin{array}{l} Cr \geq -2\,(Cb + 24) \ ; Cr \geq -\,(Cb + 17) \ ; Cr \geq -4\,(Cb + 32)\,; \\ Cr \geq 2.5\,(Cb + \theta_1) \ ; Cr \geq \theta_3; Cr \geq 0.5\,(\theta_4 - Cb) \ ; \\ Cr \leq \frac{220-Cb}{6}; Cr \leq \frac{4}{3}\,(\theta_2 - Cb) \end{array} \tag{4}$$

The opening operation of morphological processing is used to obtain more accurate contours of these skin regions. The dilation operation with a 3×3 structuring element is applied to skin regions followed by the erosion operation with the same structuring element.

Finally, fast 4-connected component labeling is used to obtain complete skin regions. The fast 4-connected component labeling can quickly label each isolated region. The size of each isolated region is calculated, and then an isolated region is removed if its area is smaller than the given threshold Th_c.

The fast 4-connected component labeling works as follows. Global-block and row-block are used to record attributes of 4-connected components. First, we horizontally scan the whole image from left to right and from top to bottom to label horizontally connected components in the global-block. Next, we construct the index of obtained labels of horizontally connected components; they are initially set to zero in the row-block. Then, we scan the labels of horizontally connected components in the whole image from left to right and from top to bottom. We update the index of labels of horizontally connected components by increasing the index of isolated labels by one. Finally, we check the pixel under the scanned label. If its index is zero, we rewrite the index to be the same index of the scanned label; otherwise, the index is preserved. Figure 1 shows a 320×240-pixel test image. The execution times using the fast 4-connected component labeling and the traditional method are 4 ms and 280 ms, respectively. This demonstrates that fast 4-connected component labeling greatly outperforms the traditional method in terms of execution time.

Using image resizing and opening operation of morphological processing, face regions and regions of complex backgrounds whose color is similar to that of skin can be effectively separated to achieve the robust face detection.

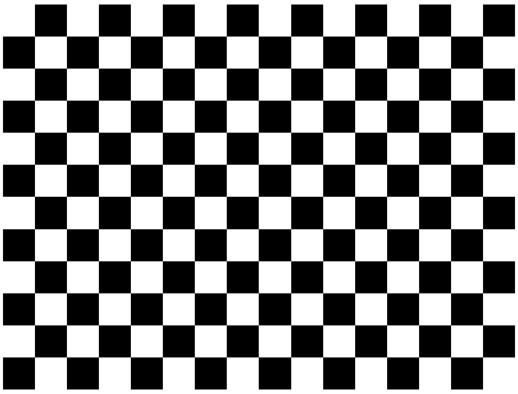

Figure 1. Test image for evaluating the performance of 4-connected component labeling

2.2. **Face candidate detection.** The face template measure is used to detect face candidates from skin regions. First, a bounding box of the skin region in the LL subband image is found, and then the coordinates of the bounding box are mapped into ones in the original image.

We use an ellipse region, W_e, as the face template, as shown in Figure 2, where W_r is the bounding box of the skin region. A skin region is considered as a face candidate if it meets the following two conditions: (i) the major axis of ellipse region W_e must be vertical, and (ii) sensitivity S_e (true positive rate) [23] must be larger than the threshold Th_s. Sensitivity S_e is defined in Eq. (5), where TP and FN are the total number of true positive pixels and false negative pixels, respectively.

$$S_e = TP/(TP + FN) \tag{5}$$

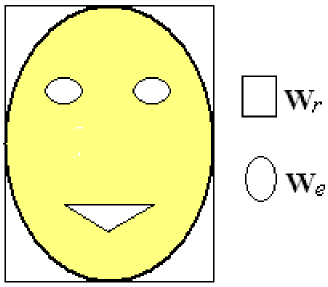

Figure 2. Face template

However, some non-face regions may exist in the obtained face candidates, such as necks and arms. The non-face regions must be removed from face candidates before face detection. A suitable face box is used to effectively remove non-face regions from face candidates. First, we scan the Y-axis in the face candidate to calculate the histogram. Next, 75% of the median of the histogram is set as the threshold Th_m. Finally, the skin region is removed if its histogram is smaller than Th_m. The size of a suitable face box (square box) can be obtained by scanning the X-axis in the preserved skin region. Figure 3(a) shows the bounding box of a face candidate, Figure 3(b) shows the obtained histogram, and Figure 3(c) shows the suitable face box.

2.3. **Face detection.** The facial feature measure is used to detect faces from face candidates obtained in the second stage. In the facial feature measure, the luminance variation

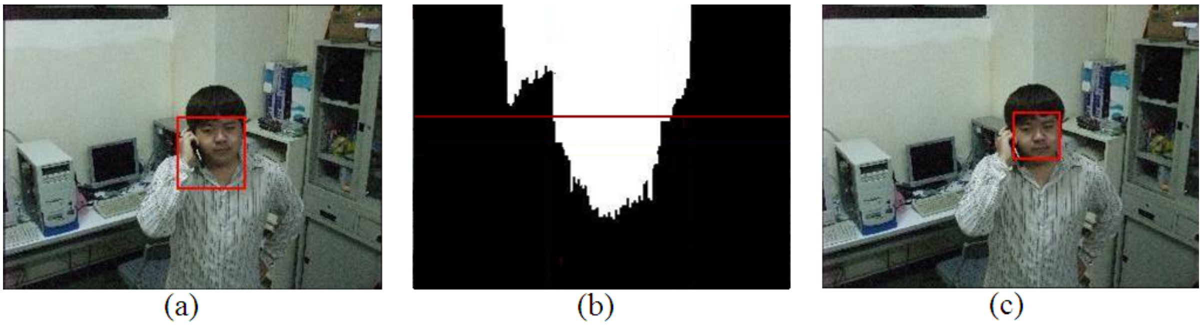

Figure 3. Suitable face box of the face candidate. (a) Bounding box of a face candidate; (b) histogram; (c) suitable face box

of the face [24] and the histogram bin-based skin distribution of the face are used to detect faces from face candidates. The face region usually has higher luminance variation than those of other skin regions, such as arms and legs. In the suitable face box of a face candidate, luminance variation is defined in Eq. (6), where σ and μ are the standard deviation and the mean of luminance, respectively.

$$\lambda = \sigma/\mu \tag{6}$$

The histogram bin-based skin distribution of the face divides 256 gray levels into 16 histogram bins, with each histogram bin containing 16 gray levels. From observation and experience, the skin distribution of a non-face region is concentrated in some histogram bins, as shown in Figure 4, where the first and second columns are the cases of a non-face region, and the third column is the case of a face region.

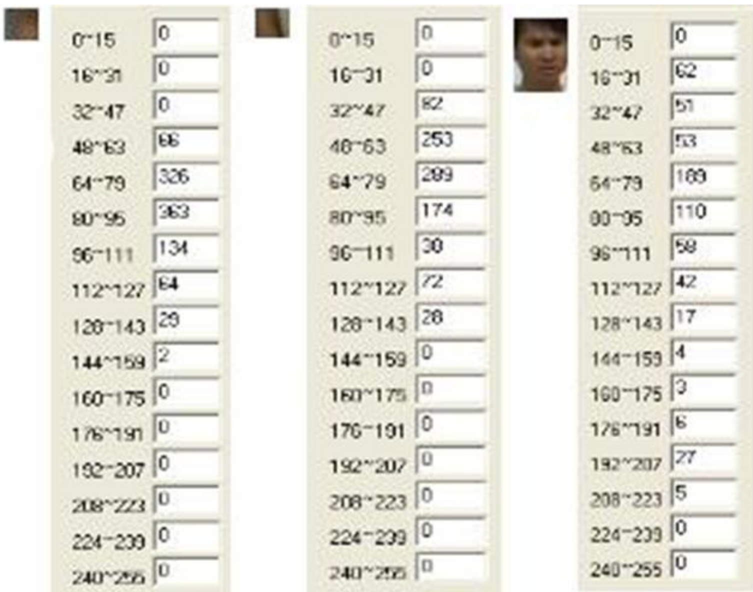

Figure 4. Histogram bin-based skin distribution

Therefore, face candidates are considered as faces if they meet the following two conditions: (i) luminance variation $\lambda \geq Th_\lambda$, and (ii) the number of histogram bin-based skin distribution is larger than Th_n, where Th_λ and Th_n are the given thresholds.

3. **Experimental Results.** The experimental results show that the proposed method performs well. Experiments were conducted on a computer with an Intel(R) Core(TM)2 Duo E4300 CPU and 2GB of RAM. The algorithms were implemented in BCB (Borland C++ Builder) 6.0. In the experiments, the thresholds $Th_c = 25$, $Th_s = 0.7$, $Th_\lambda = 1.008$, and $Th_n = 8$ were set from experience.

Faces in various poses and scales selected from the FERET image database [25] were used to evaluate the performance of the proposed method. In the experiment, 1609 face images were selected from the directories data\montages\misc\no_mustache_M_Asian and data\montages\misc\no_mustache_F_Asian, with 1038 and 571 faces of Asian males and females, respectively. Figure 5(a) and Figure 5(b) show the samples of Asian male and Asian female faces, respectively. The detection rate, missing rate, and false alarm rate are $P = 96.4\%$, $M = 3.6\%$, and $FA = 0.7\%$, respectively. The proposed method also can obtain good face detection of non-Asian faces, as shown in Figure 6.

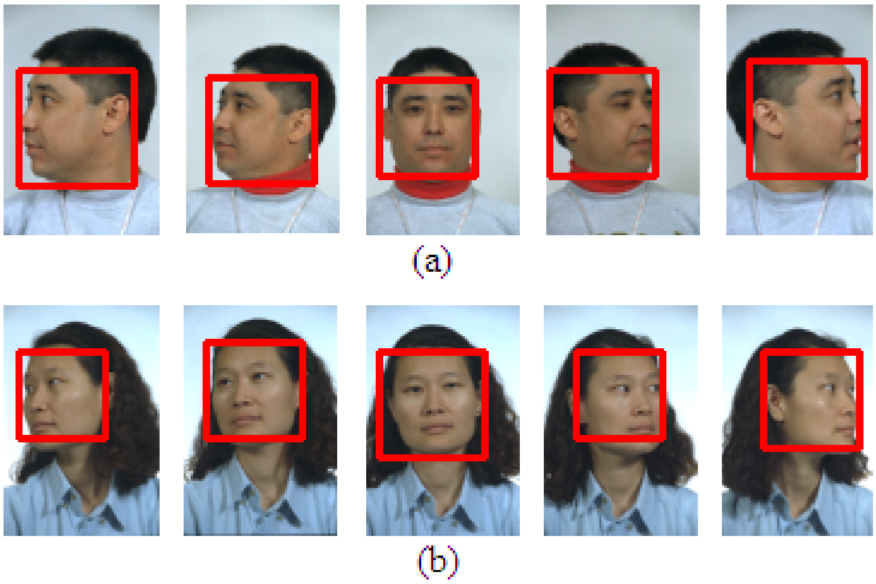

FIGURE 5. Samples of Asian male and female face with various poses and scales from the FERET image database. (a) Asian male; (b) Asian female

Moreover, the proposed method can obtain good face detection of faces in skin-color like backgrounds, as shown in Figure 7, where images on the left are the skin-region detection and images on the right are the face detection. All faces are detected in Figure 7(a) and Figure 7(b), and only one non-face is misjudged in Figure 7(b).

Finally, face detection was simulated using a test sequence consisting of 500 frames with a size of 320×240 pixels. There were 945 faces in the test sequence. Figure 8 shows the face detection results under varying illumination, where Figure 8(a)$\sim$Figure 8(c) are the cases of normal illumination, dark illumination, and light illumination, respectively. The detection rate, missing rate, and false alarm rate are $P = 98.4\%$, $M = 1.6\%$, and $FA = 0.1\%$, respectively. Furthermore, the average execution time using the proposed method is only 15.9 ms for each frame in the test sequence.

For face detection under extreme illumination, such as back lighting, adaptive skin-color model switching [26] can be used to replace the YCbCr skin-color model in the first stage of the proposed method to obtain skin-color regions.

The proposed method has the advantages of a feature-based method while overcoming the disadvantages. The proposed method achieves fast detection of faces in various poses,

FIGURE 6. Face detection of non-Asian faces from the FERET image database

(a)

(b)

FIGURE 7. Face detection in skin-color like backgrounds

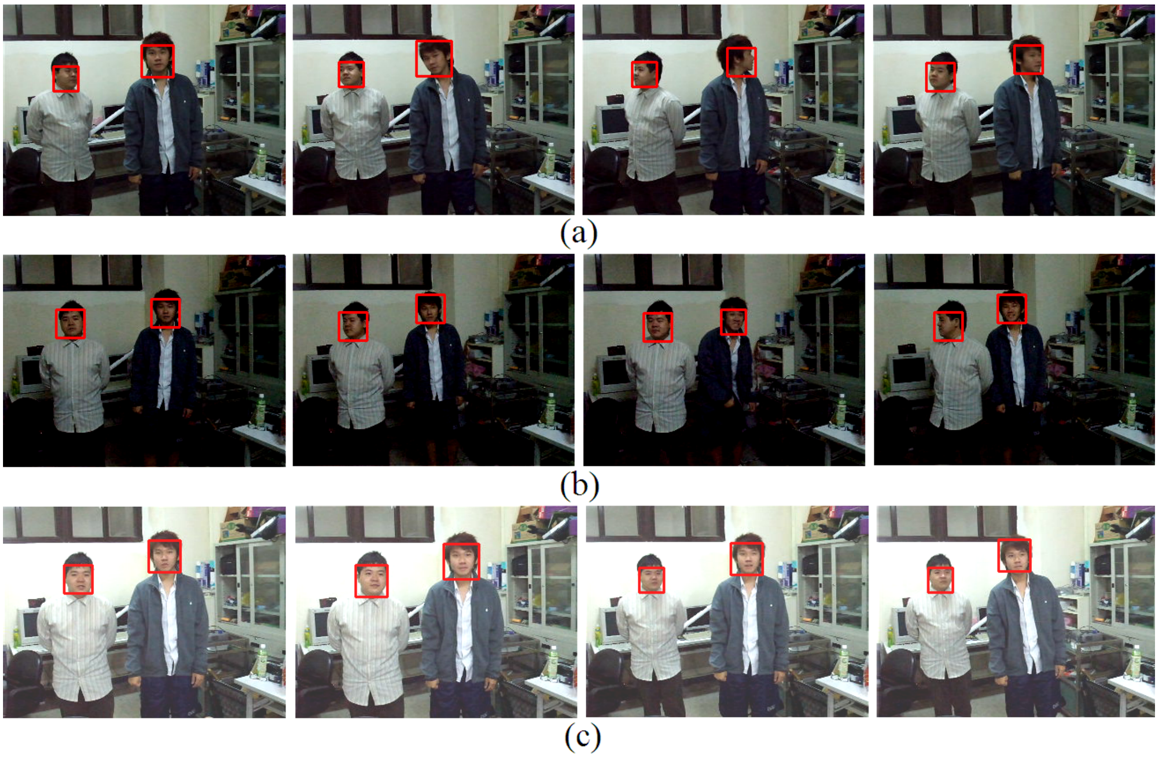

FIGURE 8. Face detection under varying illumination. (a) Normal illumination; (b) dark illumination; (c) light illumination

faces in skin-color like backgrounds, faces under varying illumination, and faces of various races.

4. **Conclusion.** A feature-based three-stage scheme was proposed to achieve real-time and reliable face detection. In the first stage, lighting compensation using the reference white, image resizing using the discrete wavelet transform, the YCbCr skin-color model, morphological processing, and fast 4-connected component labeling are used to detect skin regions. Next, a face template measure and suitable face box are used to obtain face candidates in the second stage. Finally, facial features (luminance variation and histogram bin-based skin distribution) are measured to detect faces from face candidates.

Experimental results show that the proposed face detection has good performance in the detection rate, missing rate, and false alarm rate. Furthermore, the average execution time was only 15.9 ms for each frame in the test sequence. The proposed face detection method can achieve good results of faces in various poses, faces in skin-color like backgrounds, faces under varying illumination, and faces of various races.

Acknowledgment. This work is partially supported by National Science Council, Taiwan under grant NSC96-2221-E-346-005. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] H. Veeraraghavan, N. Papanikolopoulos, and P. Schrater, Learning dynamic event descriptions in image sequences, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-6, 2007.

[2] M. H. Y. Liao, D. Y. Chen, C. W. Sua, and H. R. Tyan, Real-time event detection and its application to surveillance systems, *Proc. of IEEE International Symposium on Circuits and Systems*, pp. 509-512, 2006.

[3] C. Poppe, S. de Bruyne, T. Paridaens, P. Lambert, and R. van de Walle, Moving object detection in the H.264/AVC compressed domain for video surveillance applications, *Journal of Visual Communication and Image Representation*, vol. 20, pp. 428-437, 2009.

[4] W. C. Hu and C. Y. Yang, An improved full search algorithm with adaptive template block for fast and accurate object tracking, *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 11, pp. 5115-5130, 2010.

[5] T. Y. Chen, T. H. Chen, and D. J. Wang, A cost-effective people-counter for passing through a gate based on image processing, *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 3, pp. 785-800, 2009.

[6] C. C. Lien, Y. L. Huang, and C. C. Han, People counting using multi-mode multi-target tracking scheme, *Proc. of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kyoto, Japan, 2009.

[7] T. H. Chen, Y. F. Lin, and T. Y. Chen, Intelligent vehicle counting method based on blob analysis in traffic surveillance. *Proc. of the 2nd International Conference on Innovative Computing, Information and Control*, Kumamoto, Japan, 2007.

[8] W. L. Zheng, and S. M. Bhandarkar, Face detection and tracking using a boosted adaptive particle filter, *Journal of Visual Communication and Image Representation*, vol. 20, pp. 9-27, 2009.

[9] B. Sugandi, H. Kim, J. K. Tan, and S. Ishikawa, Real time tracking and identification of moving persons by using a camera in outdoor environment, *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 5, pp. 1179-1188, 2009.

[10] S. Ferrando, G. Gera, and C. Regazzoni, Classification of unattended and stolen objects in video-surveillance System, *Proc. of IEEE International Conference on Video and Signal Based Surveillance*, pp. 21-26, 2006.

[11] J. B. Li, J. S. Pan, and Z. M. Lu, Kernel optimization-based discriminant analysis for face recognition, *Neural Computing and Applications*, vol. 18, no. 6, pp. 603-612, 2009.

[12] J. B. Li, J. S. Pan, and Z. M. Lu, Face recognition using Gabor-based complete kernel Fisher discriminant analysis with fractional power polynomial models, *Neural Computing and Applications*, vol. 18, no. 6, pp. 613-621, 2009.

[13] L. Ma and K. Khorasani, Facial expression recognition using constructive feedforward neural networks, IEEE *Trans. Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 34, no. 3, pp. 1588-1595, 2004.

[14] S. Krinidis, and I. Pitas, Statistical analysis of human facial expressions, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 3, pp. 241-260, 2010.

[15] H. Sayoud, and S. Ouamour, Proposal of a new confidence parameter estimating the number of speakers-an experimental investigation, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 2, pp. 101-109, 2010.

[16] R. Nishimura, S. Abe, N. Fujita, and Y. Suzuki, Reinforcement of VoIP security with multipath routing and secret sharing scheme, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 3, pp. 204-219, 2010.

[17] Y. W. Wu and X. Y. Ai, Face detection in color image using AdaBoost algorithm based on skin color information, *Proc. of 2008 International Workshop on Knowledge Discovery and Data Mining*, pp. 339-342, 2008.

[18] M. H. Yang, D. J. Kriegman, and N. Ahuja, Detecting faces in images: a survey, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58, 2002.

[19] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, A survey of skin-color modeling and detection methods, *Pattern Recognition*, vol. 40, pp. 1106-1122, 2007.

[20] H. Y. Chen, C. L. Huang, and C. M. Fu, Hybrid-boost learning for multi-pose face detection and facial expression recognition, *Pattern Recognition*, vol. 41, pp. 1173-1185, 2008.

[21] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, Face detection in color images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-706, 2002.

[22] C. Gracia and G. Tziritas, Face detection using quantized skin color regions merging and wavelet packet analysis, *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 264-277, 1999.

[23] E. J. Carmona, J. Martínez-Cantos, and J. Mira, A new video segmentation method of moving objects based on blob-level knowledge, *Pattern Recognition Letters*, vol. 29, pp. 272-285, 2008.

[24] T. W. Chen, S. C. Hsu, and S. Y. Chien, Automatic feature-based face scoring in surveillance systems, *Proc. of the 9th IEEE International Symposium on Multimedia*, pp. 139-146, 2007.

[25] P. J. Phillips, H. J. Moon, S. A. Rizvi, and P. J. Rauss, The FERET evaluation methodology for face recognition algorithms, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp. 1090-1104, 2000.

[26] D. Y. Huang, W. C. Hu, and M. H. Hsu, Adaptive skin color model switching for face tracking under varying illumination, *Proc. of the 4th International Conference on Innovative Computing, Information and Control*, Kaohsiung, Taiwan, pp. 326-329, 2009.