Efficient Real-Time Face Detection For High Resolution Surveillance Applications

Xin Cheng, Ruan Lakemond, Clinton Fookes, Sridha Sridharan
Science and Engineering Faculty
Queensland University of Technology
Brisbane, Queensland 4113, Australia
Email: {x2.cheng,r.lakemond,c.fookes,s.sridharan}@qut.edu.au

***Abstract*—This paper presents an efficient face detection method suitable for real-time surveillance applications. Improved efficiency is achieved by constraining the search window of an AdaBoost face detector to pre-selected regions. Firstly, the proposed method takes a sparse grid of sample pixels from the image to reduce whole image scan time. A fusion of foreground segmentation and skin colour segmentation is then used to select candidate face regions. Finally, a classifier-based face detector is applied only to selected regions to verify the presence of a face (the Viola-Jones detector is used in this paper). The proposed system is evaluated using 640×480 pixels test images and compared with other relevant methods. Experimental results show that the proposed method reduces the detection time to 42 ms, where the Viola-Jones detector alone requires 565 ms (on a desktop processor). This improvement makes the face detector suitable for real-time applications. Furthermore, the proposed method requires 50% of the computation time of the best competing method, while reducing the false positive rate by 3.2% and maintaining the same hit rate.**

I. Introduction

As the demand for public security progressively increases, so does the demand for more robust and ubiquitous automated surveillance systems [1]. There is a growing interest in being able to detect and identify individuals in unconstrained environments without direct interaction with the subjects. The face is an important biometric for recognition in these environments [2]. Efforts to make face recognition more robust have lead to the recognition that 2D face recognition methods are not satisfactory [3] and have encouraged research in 3D face recognition. An example of a non-cooperative 3D face recognition system can be found in [4]. There the authors claim that in order to get acceptable 3D reconstruction accuracy, the face must be observed with a resolution high enough such that the distance between the outer eye corners is at least 100 pixels. The total resolution of the camera must be very high to deal with an environment where the position of the subject's head is not tightly constrained. In unconstrained, high resolution images, the visual front-end processes must be very time efficient to avoid excessive delays between a face appearing and the commencement of the recognition process. Face detection, in particular, must be very fast to provide good initialisation for the face tracking system.

A face detection method (referred to as the Viola-Jones detector in this paper) is proposed in [5] that is claimed to provide real-time performance when processing 384×288 images. This method cascades a large number of weak grey scale appearance-based classifiers into a strong classifier (an Adaptive Boosting or AdaBoost classifier [6]). Although it was claimed that this method could perform face detection 15 times faster than previous methods (while maintaining a reasonable error rate), the processing cost is still too large for real-time applications with high resolution video. When applied to a 640×480 image, the detection time of the Viola-Jones detector (using a typical configuration), run on an Intel Core 2 Duo 3.0GHz PC, is on average 565 ms per image.

Previous methods of increasing the computational efficiency of face detection have focussed on two areas. The first is to reduce the classifier cost by, for example, reducing the number of classifiers [7]. The second uses selective attention to reduce the image area processed [8]. This paper proposes a novel framework for face detection through selective attention. This method is differentiated from the existing methods in the selective attention family through a new framework for the fusion of skin and foreground segmentation techniques and includes an efficient sampling strategy. The method may be used in combination with any classifier.

II. Background

Face detection has attracted significant research interests over the past two decades. Hundreds of methods have been proposed [9]. Yang et al. [10] grouped the various methods into four categories: knowledge-based methods, feature invariant approaches, template matching methods, and appearance-based methods. Knowledge-based methods use pre-defined rules to detect a face based on human knowledge; feature invariant approaches aim to find face structure features that are robust to pose and lighting variations; template matching methods use face templates to judge if an image is a face; appearance-based methods learn face models from a set of representative training face images to perform detection. An appearance-based method proposed by Viola and Jones [11] is considered a robust and efficient baseline solution. Its performance is, however, still not satisfactory for real time applications. An extended work by Viola and Jones was proposed to achieve real-time performance [5], but only for detection in low resolution images (384×288 pixels). A

978-1-4673-2393-2/12/$31.00 ©2012 IEEE

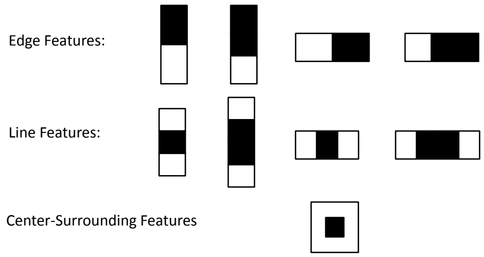

Fig. 1. Haar-like feature blocks: Edge features, line features and center-surrounding features

number of methods have been proposed for increasing the computational efficiency of [5]. The majority of these methods focus on more efficient classifiers or on selecting regions of attention before applying the classifiers.

A. Viola-Jones Face Detection

Viola and Jones proposed an object detection algorithm which uses Haar-like features to construct a set of weak classifiers [11]. Each weak classifier applies a simple threshold on one of the extracted features. The weak classifiers are then combined in a cascade structure to classify faces and non-faces.

1) Haar-Like Features: The Haar basis functions are a set of rectangular 2D features derived from the Haar wavelet. Some selected Haar-features are illustrated in Figure 1. The output of these features, when applied to grey scale images, is the difference between the sum of pixels in the white region and the black region.

The Haar-like features are applied to sub-windows of an image with varying scale and translations. There are a large number of Haar-like features extracted – for a sub-window of size 24×24 pixels, there will be approximately 10^5 features. A training process is used to selected a sub-set of extracted Haar-like features which can best separate the training data into positive and negative samples. An integral image is used to reduce the redundant computation involved in computing many Haar-like features across an image.

2) AdaBoost Algorithm: Adaptive Boosting is a method of constructing a strong classifier from a linear combination of weak classifiers. The classifier training process is as follows. The input is N example images and labels $(I_1, l_1), ..., (I_N, l_N)$. Each $I_i = I_i(x, y)$ is an example image and each l_i is a binary label indicating positive $(l = 1)$ or negative $(l = 0)$ examples. Define the number of positive labels to be M and the number of negative labels O. Let $g_j(I_i)$ be the j^{th} Haar-like feature extracted from image i. A set of weights, w, is initialised according to,

$$w_{0,i} \leftarrow \begin{cases} 0.5M^{-1} & \text{for} \quad l_i = 1 \\ 0.5O^{-1} & \text{for} \quad l_i = 0 \end{cases}, \tag{1}$$

For $t = \{1, ..., T\}$,

1) Train a classifier h_j on each feature g_j and evaluate the classifier error as $\varepsilon_j = \sum_{\forall i} w_{t-1,i} \left| h_j(I_i) - l_i \right|$.
2) Choose the classifier, h_t, with the lowest error, ε_t.
3) Update the weights according to,

$$w_{t,i} \leftarrow w_{t,i} \left(\frac{\varepsilon_t}{1 - \varepsilon_t} \right)^{1 - |h_t(I_i) - l_i|}, \tag{2}$$

4) Normalise the weights so that $\sum_{\forall i} w_{t,i} = 1$.

The final strong classifier is:

$$h(I) = \begin{cases} 1 & \text{for} \quad \sum_{t=1}^{T} \alpha_t h_t(I) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where $\alpha_t = -\log\left(\frac{\varepsilon_t}{1-\varepsilon_t}\right)$.

3) Cascaded Detector: The final detector is constructed by cascading a series of classifiers. Stages in the cascade are constructed by training classifiers to a target detection rate and false positive rate. Subsequent stages are trained using those samples which pass through all the previous stages.

With the initial classifier applied on all sub-windows in a test image, a large number of negative sub-windows are eliminated with very little processing. Subsequent classifiers eliminate additional negative sub-windows with additional computation. After a sufficient number of stages of processing, the number of sub-windows have been reduced radically and the remaining sub-windows are considered positive.

B. Improvements to Classifier Efficiency

A number of modifications and alternative methods have been proposed in an attempt to achieve a lower computation time compared to the original Viola-Jones detector. In [7] an evolutionary pruning method is proposed to form strong classifiers using fewer weak classifiers. The detection time is reduced to 57.7% of the original Viola-Jones detector. In [12] an efficient face detection algorithm is proposed for quasi-repetitive applications, such as video-conferencing. The algorithm caches a set of frame exemplars into a library. When a new frame is given, the method quickly search through the exemplar library. If a similar exemplar is found, the method skips the face detector, and reuses the previously detected object states. The authors claim that the method improved the frame processing time from 100 ms to 20 ms while processing 320×240 images. In [13] the detection speed is increased by using feature-centric evaluation on the first cascade stage. Feature-centric evaluation reuses feature evaluations across multiple candidate windows to avoid recomputation of the same features. In [14] the authors proposed a detection method using Locally Assembly Binary (LAB) features instead of Haar-like features. The authors claim that using LAB features in combination with Feature-centric evaluation is 20 times faster than the Viola-Jones method.

C. Selective Attention Methods

A number of selective attention methods have been proposed to accelerate face detection. In these methods, possible face regions are pre-detected. The Viola-Jones detector (or any other classifier) is applied to candidate regions only, in order to reduce the classifier scan time. In [8], [15], [16], [17], [18], [19], candidate regions are selected using colour segmentation methods. Regions with skin colour are determined as candidate regions. Similar methods using motion segmentation can be found in [20], [21], [22].

1) Skin Colour Segmentation: Colour is the most intuitive feature of human skin. Human skin colour has its own characteristic which can be used to distinguish skin regions and non-skin regions in an image. The colour distributions of faces in the YC_bC_r colour space are studied in [23]. It was found that the chrominance values of pixels in the facial region are narrowly distributed within certain ranges. The most representative range of C_b and C_r to define a skin colour is proposed to be,

$$\begin{array}{rcl} 77 & \leq C_b \leq & 127, \\ 133 & \leq C_r \leq & 173. \end{array} \tag{4}$$

In [8], an improved Viola-Jones detector combined with a skin colour model is proposed. The method determines possible face regions by a colour segmentation method, then applies Viola-Jones detector only on skin colour regions. The performance of this method is evaluated in Section IV-B. Similar methods using a variety of different colour threshold strategies can be found in [15], [16], [17], [18], [19].

More robust colour modelling methods such as Single Gaussian model [24] and Gaussian Mixture models [25] have also been proposed. In this paper, the simple threshold strategy in [23], which is a rectangular threshold in C_bC_r space, is preferred due to its computational simplicity and its good segmentation result on the data used.

2) Foreground Segmentation: A selective attention method based on foreground segmentation is proposed in [22], however, a performance comparison with the original Viola-Jones detector is not included. This method is implemented and evaluated in Section IV-B. Foreground segmentation is the process of classifying image regions into background regions and foreground regions. Robust methods for modelling the background, such as Single Gaussian Models [26], Gaussian Mixture Models [27] and the Codebook Method [28], have been proposed and are commonly used. In this paper, a Gaussian Mixture Model is used for background modelling.

The background image is modelled as a mixture of K Gaussian distributions at each pixel, each with mean μ_k, variance σ_k^2 and weight ω_k. When a new observation pixel x is given, the x is checked against the existing K Gaussian distributions, until a match is found. An observed pixel is considered matched to the background if its value falls within 2.5 standard deviations of the mean of one of the distributions of the background model. If the observation is matched with one existing model, the matched model is then updated using a learning rate factor α. That is, for mean μ_k,

$$\mu_{k,t+1} = \mu_{k,t} \cdot (1-\alpha) + x \cdot \alpha, \tag{5}$$

and for variance σ_k^2,

$$\begin{aligned} d &= x - \mu_{k,t+1}, && (6) \\ \sigma_{k,t+1}^2 &= \sigma_{k,t}^2 \cdot (1-\alpha) + (d \circ d) \cdot \alpha, && (7) \end{aligned}$$

where $\circ$ is an element-wise multiplication operator.

The weights of all models are then updated so that the weight of the matching model increases towards one and the weights of the other models decrease towards zero. After updating, the weights have to be normalised so that their sum remains one. The weight updating process can be expressed as,

$$\omega_{i,t}' = (1-\alpha)\omega_{i,t-1} + \alpha \cdot M_{k,t}, \tag{8}$$

$M_{k,t}$ is 1 for the model matched and 0 for remaining models. If there is no match between the observation and the existing models, the observation is injected into the background models by replacing model with the least weight.

The Gaussian distributions are classified into background and foreground by using the Stauffer and Grimson background test [27]. Firstly, the Gaussian distributions are sorted by the value of ω/σ in descending order. Then the first B distributions are chosen as the background model, where,

$$B = \arg\min_b \left(\sum_{k=1}^{b} \omega_k > T \right), \tag{9}$$

where T is a measure of the minimum portion of the data that should be accounted for by the background.

To date, no existing techniques have combined the strengths of both foreground segmentation and skin segmentation. This paper addresses this gap and the framework is outlined in the following section.

III. A New Selective Attention Method

In this section a novel selective attention face detection method is presented. The proposed method is outlined in Figure 2. Face candidate regions are detected by a four-step algorithm:

1) A sparse grid of pixels is sampled from the input image.
2) Foreground segmentation is applied.
3) A skin colour detector is applied to the foreground sample pixels
4) Connected foreground, skin-coloured pixels are grouped and used to generate candidate face regions.

The selected regions are then passed to the Viola-Jones detector.

A. Grid Sampling

In some applications, such as face recognition, there is a minimum face size requirement for face detection. This minimum size requirement makes it possible to reduce the detection resolution without missing faces that are of interest.

The minimum face size($F_{\min}$) is an application dependant variable. In this work, $F_{\min} = 40$ is chosen, as the minimum face size in the data set is 40×40 pixels. For an image I with resolution $w \times h$ pixels, the down-sampled image S can be expressed as,

$$S(x,y,z) = I\left(\frac{k}{2} + kx, \frac{k}{2} + ky, z\right), \tag{10}$$

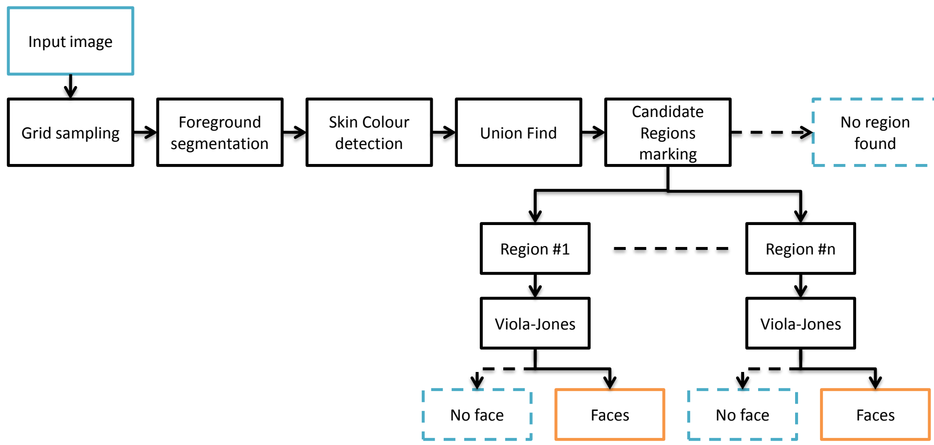

Fig. 2. Flow chart of the proposed selective attention face detection method.

where the domain of S is $x \in \left[0, \frac{w}{k}\right]$, $y \in \left[0, \frac{h}{k}\right]$, $z \in [0,2]$, z expresses channels in YC_bC_r colour space. The sampling period k is chosen based on the minimum face size $F_{\min}$ and $\frac{k}{2} \in \mathbb{N}$. To ensure the detection accuracy, k must be smaller than $F_{\min}$.

An experimental evaluation of the trade-off between detection performance and sampling scale factor was conducted and is reported in Section IV-A. The experimental result shows that the best performance is found at $k = 10$, which is a quarter of the minimum face size.

B. Foreground Detection

The foreground segmentation method described in Section II-C2 is applied to the down-sampled pixel grid in YC_bC_r colour space. Gaussian Mixture Models (GMM) are initialised at the sampled pixels of a background image.

The following constants are used in the GMM background model system:

- Each model consists of 5 Gaussian distributions $(K = 5)$.
- All channels in YC_bC_r colour space are modelled.
- The learning rate, α, is set to 0.005.
- The threshold, T, is set to 0.7.
- The initial weight, ω, of a new Gaussian distribution is set to 0.05 and the initial variance, σ^2, is set to 30.

Figure 3(a) shows an example of the foreground detection results. The foreground sample pixels are marked by dots, and background pixels are marked by crosses.

C. Skin Detection

The skin colour detection method described in Section II-C1 is applied only to the detected foreground pixels. The algorithm classifies each pixel into skin colour or non-skin colour using Equation 4. Figure 3(b) shows example results of the skin detection process applied to the output of the foreground segmentation step. The remaining pixels are are used to identify candidate face regions, since these pixels are in the foreground and are skin colour.

D. Face Candidate Region Selection

The Viola-Jones detector requires a rectangular image region as input. Candidate regions are produced from the foreground skin segmentation results by means of the following process. First, connected regions of pixels are found using the union-find algorithm [29]. A bounding box is then fit around each connected region. The vertices of each bounding box are computed as,

$$v = \left\{ \begin{array}{c} k^{-1}x_{\min} - (k+b)\,, k^{-1}y_{\min} - (k+b)\,, \\ k^{-1}x_{\max} + (k+b)\,, k^{-1}y_{\min} - (k+b)\,, \\ k^{-1}x_{\max} + (k+b)\,, k^{-1}y_{\max} + (k+b)\,, \\ k^{-1}x_{\min} - (k+b)\,, k^{-1}y_{\max} + (k+b) \end{array} \right\}. \tag{11}$$

Here $x_{\min}$ and $y_{\min}$ are the minimum coordinates and $x_{\max}$ and $y_{\max}$ are the maximum coordinates of the connected region. The scalar, $k+b$ is the width of a border placed around the region. It is composed of the sampling period, k, to cover the space between samples, and an additional border, b, that compensates for the border around the skin regions in the face images used to train the Viola-Jones detector. The implementation reported in this paper uses the values $k = 40$ (as discussed previously) and $b = 20$. Figure 3(c) illustrates the candidate region selected based on foreground and skin segmentation results.

E. Face detection

The Viola-Jones face detector is applied to the image regions selected using the proposed selective attention method. In the implementation reported in this paper, the Viola-Jones detector was configured using the following constants:

- The minimum search window size $F_{\min}$ is 40.
- The scale change step is 1.1.
- The search window shift step is 2 pixels

Figure 3(c) shows the output of the Viola-Jones detector as applied to a face candidate region marked by the outer rectangle. The detected face is marked by the inner rectangle.

(a) (b) (c)

Fig. 3. Output of the various stages of the proposed face candidate region extraction method. (a) Foreground segmentation results. Foreground pixels are marked by dots and background pixels are marked by crosses at the down-sampled locations. (b) Colour segmentation results. Skin colour pixels are marked by dots, non-skin pixels are marked by crosses. (c) Final face detection result. The outer rectangular area is the face candidate region. The inner rectangular area is the verified face region.

IV. Experimental Evaluation

Two experiments are reported in this section. The first experiment investigates the effects of the grid sampling period, k, on the time and detection performance of the proposed face detection method. The second experiment compares the proposed method to existing methods based on the Viola-Jones detector.

Unfortunately, the datasets most commonly used for face detection do not facilitate building a scene background model. A dataset was captured using a Sony Ipela SNC-RX550P camera placed in a fixed location viewing a doorway. Subjects were recorded as they passed through the doorway, without attracting their attention. The subjects' ethnicities include Western, South-East Asian, East Asian and Indian. A total of 490 frames were captured at 640×480 pixels resolution, including 100 frames with no face, 344 frames with one face and 46 frames with two faces. Unoccupied background frames were recorded to facilitate building a background model.

Each face detection method is applied to the captured images individually to detect the faces in the images. The detected regions are marked by rectangles. The detection accuracy is then verified manually. An Intel Core 2 Duo 3.0GHz desktop PC is used to execute the programs. The time cost of image file I/O is excluded from the process time monitoring. The hit rate, false positive rate and process time of each method are evaluated to compare the accuracies and efficiencies of each method.

A. Study on Sample Grid Density

The effect of sample grid period is investigated by evaluating the proposed method over a range of sampling periods. The time and detection performance results are plotted in Figure 4.

The detection rate reaches a maximum and the process time reaches a minimum at a sampling period of 10 pixels. This equates to 25% of the minimum face size. As the sampling period is increased beyond this point, the hit rate decreases rapidly as more faces are simply missed by the sparse sampling grid. At the same time, no further reduction in processing time is observed. This is due to the fact that as the sampling period increases, so too does the width of the minimum candidate region passed to the Viola-Jones detector (Equation 11). The grid sampling period does not appear to have a strong impact on the false positive rate. The sampling factor $k = 10$ is used in further experimentation.

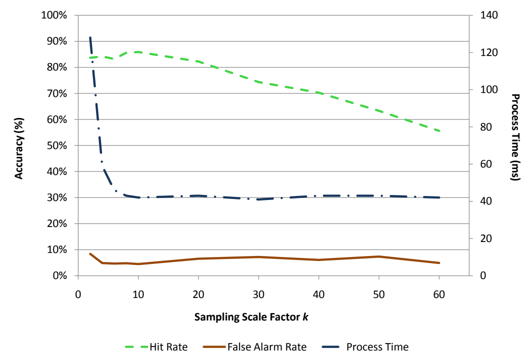

Fig. 4. Face detection performance and processing time performance of the proposed method evaluated over a range of grid sampling periods.

B. Comparison with Existing Methods

An experiment was conducted to evaluate the effect of various segmentation front-end systems on the performance of a Viola-Jones face detector. The following face detection systems are compared:

1) The Viola-Jones detector without any pre-processing.
2) A Viola-Jones detector with foreground segmentation.
3) A Viola-Jones detector with skin colour segmentation.
4) The proposed method – a Viola-Jones detector with a front-end using sparse sampling, foreground segmentation and skin colour segmentation.

The experimental results are listed in Table I.

The proposed method provides a 93% reduction in computation time compared to using only the Viola-Jones method and reduces the processing time by 51% compared to the best existing method. The average processing time is approaching real time requirements. Only a marginal reduction in hit rate is observed for all the segmentation front-end systems. The false positive rate is reduced significantly by colour segmentation and the proposed method provides an even greater

Method	Hit Rate	False Positive Rate	Time
Viola-Jones	**86.33**%	34.55%	565.2 ms
Foreground segmentation	84.65%	32.6%	144.3 ms
Colour segmentation	85.85%	7.73%	85.6 ms
Proposed method	85.61%	**4.52**%	**42.3 ms**

TABLE I
RESULTS OF EVALUATION COMPARING THE PROPOSED FACE DETECTION METHOD AND RELATED EXISTING METHODS.

reduction. Overall the proposed method achieves outstanding performance in terms of computational efficiency and false positive rate.

V. CONCLUSION AND FUTURE DIRECTIONS

An efficient face detection framework suitable for high resolution surveillance applications is proposed in this paper. A reduced sampling strategy, foreground segmentation and colour segmentation are employed to constrain the search space, before applying a classifier-based face detector. The new framework significantly reduces the computational time and false positive rate while maintaining the hit rate of a classifier-based face detector. It requires only 42ms (real-time), on average, while processing 640×480 pixels images, compared to 565ms for the Viola-Jones detector. The use of a sparse sampling method allows the computational cost to be controlled in accordance with the minimum face size, instead of the image resolution. This enables greater scalability for high-resolution applications.

Further opportunities exist to reduce face detection time. The Viola-Jones detector could be replaced with a more efficient classifier, such as the method in [14] (the proposed method is suitable for any classifier). Foreground segmentation can be performed in a coarse-to-fine method by detecting moving bodies before searching for a face.

REFERENCES

[1] C. Fookes, S. Denman, R. Lakemond, D. Ryan, S. Sridharan, and M. Piccardi, "Semi-supervised intelligent surveillance system for secure environments," in *IEEE Industrial Electronics Symposium*, Bari, 2010, special Session on Computational Intelligence for Safe and Secure Environments and Transport.

[2] R. Heitmeyer, "Biometric identification promises fast and secure processing of airline passengers," *The Int'l Civil Aviation Organization Journal*, vol. 55, no. 9, pp. 10–11, 2000.

[3] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002," in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, 2003, p. 44.

[4] G. Medioni, C. Jongmoo, K. Cheng-Hao, A. Choudhury, Z. Li, and D. Fidaleo, "Non-cooperative persons identification at a distance with 3d face modeling," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, 2007, pp. 1–6.

[5] P. Viola and M. Jones, "Robust real-time face detection," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, 2001, pp. 747–747.

[6] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," 1995.

[7] J. Jun-Su and K. Jong-Hwan, "Fast and robust face detection using evolutionary pruning," *Evolutionary Computation, IEEE Transactions on*, vol. 12, no. 5, pp. 562–571, 2008.

[8] G. Li and Y. Xu, "The study of face detection algorithm based on improved adaboost with skin color model," in *Photonics and Optoelectronic (SOPO), 2010 Symposium on*, 2010, pp. 1–3.

[9] C. Z. Zhang and Zhengyou, "A survey of recent advances in face detection," TechReport, 2010.

[10] Y. Ming-Hsuan, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 1, pp. 34–58, 2002.

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 1. Kauai, HI, USA: IEEE Comput. Soc, 2001, pp. I–511–I–518.

[12] T.-J. C. Minh-Tri Pham, "Detection caching for faster object detection," in *International Workshop on modeling People and Human Interaction (PHI'05)*. Beijing, China: Proc. IEEE, 2005.

[13] H. Schneiderman, "Feature-centric evaluation for efficient cascaded object detection," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, 2004, pp. II–29–II–36 Vol.2.

[14] Y. Shengye, S. Shiguang, C. Xilin, and G. Wen, "Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–7.

[15] L. Liying and G. Weiwei, "Study on face detection algorithm based on skin color segmentation and adaboost algorithm," in *Web Mining and Web-based Application, 2009. WMWA '09. Second Pacific-Asia Conference on*, 2009, pp. 70–73.

[16] M. Songyan and D. Tiancang, "Improved adaboost face detection," in *Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on*, vol. 2, 2010, pp. 434–437.

[17] W. Yan-Wen and A. Xue-Yi, "An improvement of face detection using adaboost with color information," in *Computing, Communication, Control, and Management, 2008. CCCM '08. ISECS International Colloquium on*, vol. 1, 2008, pp. 317–321.

[18] H. Yan and P. Qingsong, "Face detection based on adaboost and skin color," in *Information Science and Engineering (ISISE), 2009 Second International Symposium on*, 2009, pp. 407–410.

[19] L. Wen-Hsiang and L. Chang-Tsun, "Skin colour-based face detection in colour images," in *Video and Signal Based Surveillance, 2006. AVSS '06. IEEE International Conference on*, 2006, pp. 56–56.

[20] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 2003, pp. 734–741 vol.2.

[21] M. J. Jones and D. Snow, "Pedestrian detection using boosted features over many frames," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008, pp. 1–4.

[22] Z. Lijing and L. Yingli, "A fast method of face detection in video images," in *Advanced Computer Control (ICACC), 2010 2nd International Conference on*, vol. 4, 2010, pp. 490–494.

[23] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 4, pp. 551–564, 1999.

[24] M. Bernd and W. Mathias, "Segmentation and tracking of facial regions in color image sequences," N. N. King, S. Thomas, and S. Ming-Ting, Eds., vol. 4067. SPIE, 2000, pp. 731–740, visual Communications and Image Processing 2000 1.

[25] N. Oliver, A. P. Pentland, and F. Brard, "Lafter: Lips and face real time tracker," 1997, pp. 123–129.

[26] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: real-time tracking of the human body," in *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, 1996, pp. 51–56.

[27] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, 1999, p. 252 Vol. 2.

[28] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, 2005.

[29] R. Sedgewick, *Algorithms in C*. Reading, Mass: Addison-Wesley, 1998.