INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

ROBUST AND REAL TIME FACE TRACKING USING PARTICLE FILTER BASED ON PROBABLISTIC FACE MODEL

F.Hourali[1], Dr. M. Sedaaghi[2]

[1] *Department of Electrical and Computer Engineering, Esfarayen University of Technology, Esfarayen, Iran.*
Email: *Hourali.f@gmail.com*
[2] *Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran.*
Email: *Sedaaghi@sut.ac.ir*

Abstract: - This paper presents an algorithm for real time and robust human face tracking against pictures and other objects. It is based on Haar-like features, skin segmentation and motion information for face detection. Face tracking is performed using particle filter which depends on skin color and probabilistic face model. Basically, the employed features for face detection are Haar-like. We employ PCA to extract the most significant features. The extracted features are then used as learning vectors of neural networks. Neural networks classify the objects to face or non-face. Face detector locates faces from the face candidates determined in first frame by using skin color information. Then the detected face is tracked using particle filter and based on the probabilistic face model which is updated using the information related to variability with respect to head rotation, illumination, facial expression and occlusion. The proposed method is done in real time and achieves high performance. Also it is robust against illumination variations and geometric changes. Experiment results show that this algorithm is the best in comparison with other existing methods specially when there is occlusion. Also the robustness of the proposed method, when the tracked face is close to unicolor objects, is proved.

Keywords*:* Face detection, skin segmentation, Haar-like features, face tracking, probabilistic model.

1. Introduction

Face detection and tracking in video sequences have many applications in face identification, human computer interaction and facial expression. The first step in these applications is to detect face. The next step is tracing the detected face using a face position predictor and analysing the track, for instance, to understand its behaviour in the consecutive frames. The Purpose of the simplest form of face tracking is to locate the face position in every frame of the sequence. Additional data, e.g., orientation, extension or the precise contour of the faces at each frame are required in other applications like face recognition or facial expression identification. The information about the face position in previous frame(s) are used to predict the location of the detected face in next frames. Suandi et al. have employed dynamic Bayesian network (DBN) to detect the human face in color video sequence. Pupils, mouth center and skin region are the features utilized by them to compute the evidence for DBN inference in (Suandi et al, 2008). Kabakli et al. have proposed a face and facial feature tracking algorithm using skin color segmentation, connected component labeling and morphology to extract the head tilt angle (Kabakli et al, 2004). Zheng and Bhandarkar have developed a method for face detection and face tracking using a boosted adaptive particle filter. Their proposed face detection algorithm is based on adaboost classification. Also their face tracking method is performed using an adaptive particle filter (Zheng & Bhandarkar, 2009). Face detection and tracking presented by Verma et al. is based on a probabilistic method. A combination of face probability provided by the detector and the temporal information resulted from the tracker is used to produce an almost robust method (Verma et al, 2003).

The face detection algorithm proposed by Vasant et al. in (Vasant et al, 2007) uses a feature set that is Haar-like. It was originally developed by Viola and Jones (Viola &Jones, 2001). Also they have used a cascade of boosted decision tree classifiers to classify faces against non-face objects. In this paper, face tracking is performed employing continuous detection with the detected objects in two frames and using a greedy algorithm for mapping.

A hierarchical face detector is proposed in (Yang et al, 2007). It is a combination of a template matching algorithm and 2D PCA (principal component analysis). They use two different classifiers for classification. The rough classifier filters the most of the non-face and the second one uses 2D PCA algorithm to detect the face based on the result of the first classifier.

All of the aforementioned methods have good performances in the controlled conditions. However, they are not robust in other circumstances and especially when there is a quick movement in video sequences. In this paper, an algorithm based on Haar-like features (proposed by (Viola &Jones, 2001)), motion information and skin color segmentation is introduced for human face detection. Also human face tracking against the still picture (of human) and other objects is performed using particle filter and based on a probabilistic face model and skin color. The proposed algorithm has consistently provided a high performance and satisfies the following requirements: (1) it has the ability to automatically determine the initial position and size of the human face against picture and other objects; (2) it is insensitive to face orientation and scale changes; and (3) it is also insensitive to lighting condition modifications and locally occlusion in face region. In addition, the algorithm is computationally simple so that it can be executed in real-time. The remainder of the paper is organized as follows: Face detection which is used for locating the human face against picture and other objects, is described in Section 2. In this section, face detection is explained with details including skin color modelling and segmentation, extracting Haar-like features and employing them as inputs to neural networks for classification.

The proposed face tracking method is described in Section 3. In Section 4, implementation and experimental results are demonstrated using the proposed algorithm while the results are compared with other existing methods. Finally, conclusions will be given in Section 5.

2. Face detection

The proposed method for face detection is based on skin color segmentation (Section 2.1) and Haar-like features (Fig. 1, adapted from Viola &Jones, 2001)). Efficient features are then selected using PCA (Section 2.2). These features are used for training the neural networks (Section 2.2).

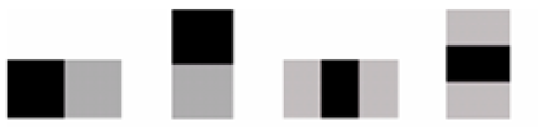


Figure 1: Haar-like features (adapted from (Viola &Jones, 2001)).

2.1 Skin segmentation method

The first step of skin segmentation is to make a skin model. The input image has RGB format and is sensitive to lighting conditions because the brightness and color information are coupled together but it is not suitable for color segmentation under unknown lighting conditions. Therefore, color system transformation is required for skin's color segmentation. We employ YCbCr color system (Y represents the luminance component while Cb and Cr represent the chrominance components of a color image).

The color distribution of skin colors of different people was found to be clustered in a small area of the chromatic color space. The Y component is discarded as it is related to brightness, but Cb and Cr components are used because they contain the color information. A manually selected skin samples from color images were used to determine the color distribution of human skin in chromatic color space. The 2D Joint Gaussian model and the 2D histogram of Cr Cb values of sample skin image are shown in Fig.2. After skin model is produced, the foreground segmented regions will resemble the skin (binary image). Before image can be skin segmented, skin likelihood value is computed for each pixel by computing Mahalanobis distance from mean. Then, skin likelihood values are normalized and shown as grayscale image where whiter areas have higher probability of being skin areas than the darker parts. After getting a grayscale image of skin likelihood, it is then necessary to threshold the image into a binary image. Since skin color varies from person to person, an adaptive threshold process is necessary. If the threshold value is too low, the amount of the segmented skin regions will increase.

Based on the fact that when a threshold value decreases and is too low, the increase in the number of skin regions sharply shows a jump, an adaptive threshold model needs to be created. The optimal threshold value is determined by finding the point when the change in the number of segmented images is a minimum. It is desirable to keep as many pixels as possible since this stage is at the beginning of the whole system. After an optimal threshold is set, all pixel values having likelihood values higher than the threshold are set to 1 and the rest of the pixels are set to 0 (Sayed & Saad, 2006). Fig. 3 shows the result of applying this algorithm on a sample image.

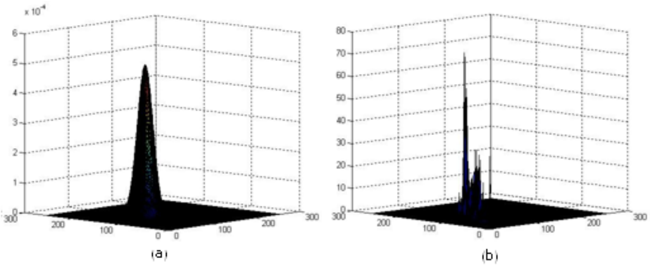


Figure2: (a) the 2D Joint Gaussian Model. (b) the 2D Histogram of Cr Cb values of sample skin images.

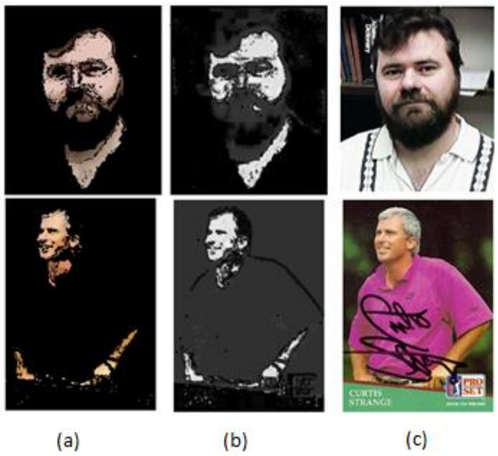


Figure 3: (a): Original color image.(b):probabilistic regions of skin.(c): Skin segmented regions.

2.2 Feature extraction

In this step, human face is detected against pictures and other objects using the face candidates. First, the detector discerns Haar-like features. Then, it selects the features that can be used to judge whether a detected feature is a face region or a non-face one using PCA. Face detection is based on the simple rectangular features that were presented by Viola and Jones (Viola &Jones, 2001). It measures the differences between the regional averages at various scales, orientations, and aspect ratios. The rectangular features can be rapidly evaluated at any scale. The selected features are applied as input feature vectors to neural networks. The training data are face and non-face images: each image is normalized to 24 × 24 pixels. The experiments demonstrate that they provide useful information and improve the performance of accurate classification. After detecting all faces in the first frame the human face is recognized among pictures, using the motion information.

3. Proposed face tracking method

We present the details of the proposed face tracking algorithm in this section. It is based on particle filtering using the face model and skin color. The model can dynamically be updated in size and content to adapt to temporal changes of the face's scale and orientation. It is performed using the information related to variability with respect to head rotation, facial expression, illumination and occlusion. The method for updating the face model is firstly described (see Section 3.1), next we describe our approach for drawing particles in motion parameter space and predicting the most likely object location with the help of the trained model and skin color information (see Sections 3.2 and 3.3).

3.1 Updating the face model using R-SVD

The batch method for updating the face model is computationally inefficient and it might not be possible to execute it at each frame. Therefore, we consider an incremental subspace method based on the described method in (Skocaj & Leonardis, 2002) and the sequential Karhunen-Loeve algorithm (Lindenbaum, 2000) to update the subspace. Given a d × n matrix $Z_a = [Z_1, Z_2, ..., Z_n]$ where each column Z_i is an observation (a d dimensional image vector), the singular value decomposition (SVD) of $Za = U_a \Sigma_a V_a^T$ is computed. When a d × m matrix of new observations $Z_b = [Z_{n+1}, Z_{n+2}, ..., Z_{n+m}]$ is available, the R-SVD algorithm efficiently computes the SVD of the larger matrix Zr = [Za | Zb] = $U_r \Sigma_r V_r^T$. The mean value of Zr can then be computed as follows:

$$\bar{Z}_r = \frac{n}{m+n}\bar{Z}_a + \frac{m}{m+n}\bar{Z}_b \tag{1}$$

Where $\bar{Z}_r$, $\bar{Z}_a$ and $\bar{Z}_b$ denote the means of Za, Zb and Zr, respectively. Using $U_a \Sigma_a V_a^T$ and E (defined in Eq. 2), $U_r \Sigma_r V_r^T$ can be obtained.

$$E = \left[Z_b - \bar{Z}_r 1_{1\times m} \middle| \sqrt{\frac{n}{n+m}} (\bar{Z}_a - \bar{Z}_b) \right] \tag{2}$$

Where 11×m is an m dimensional unit vector. Using Gram-Schmidt algorithm on $[U_a \mid E]$, the orthonormal matrix $\acute{U} = [\, U_a | \tilde{E} \,]$ is obtained. Considering the matrix $V' = \begin{bmatrix} V_a & 0 \\ 0 & I_{(m+1)} \end{bmatrix}$, where I(m+1) is an identity matrix of size (m + 1) × (m + 1), $\acute{\Sigma}$ can be obtained as follows(Levy & Lindenbaum,2000).

$$\acute{\Sigma} = \acute{U}^T [Z_a | E] V' = \begin{bmatrix} U_a^T \\ \tilde{E}^T \end{bmatrix} [Z_a | E] \begin{bmatrix} V_a & 0 \\ 0 & I_{(m+1)} \end{bmatrix} = \begin{bmatrix} U_a^T Z_a V_a & U_a^T E \\ \tilde{E}^T Z_a V_a & \tilde{E}^T E \end{bmatrix} \tag{3}$$

Computing SVD of $\acute{\Sigma} = \tilde{U}\tilde{S}\tilde{V}^T$, the SVD of Zr will be:

$$U_r \Sigma_r V_r = \acute{U}(\tilde{U}\tilde{S}\tilde{V}^T)\acute{V}^T = (\acute{U}\,\tilde{U})\tilde{S}(\tilde{V}^T \acute{V}^T) \tag{4}$$

Based on the R-SVD method and the sequential Karhunen-Loeve algorithm, the SVD computation of larger matrix Zr = [Za | Zb] can be performed efficiently.

3.2 Particle filtering for object tracking

Bayesian filtering is used to formulate the problem of tracking as follows.

$$p\,(x_t\,|\,z_{1:t}) \propto p\,(z_t\,|\,x_t) \int p\,(x_t\,|\,x_{t-1})\,p\,(x_{t-1}\,|\,z_{1:t})\,dx_{t-1} \tag{5}$$

Where $x_t$and $z_t$denote the hidden state of the object of interest and observation vector at discrete time t, respectively. All the observations up to current time step can be denoted as $z_{1:t}$ = $z_1 \ldots z_t$. Using Eq. 5, the posterior distributions (p (x_t | $z_{1:t}$)) is computed given a dynamic model p (x_t | x_{t-1}) (describing the state propagation) and an observation model p (z_t | x_t) (describing the likelihood that a state xt causes the measurement z_t). Particle filter operates by approximating the posterior distribution using a collection of weighted samples St = {($x_t^{(n)}, \pi_t^{(n)}$) | n = 1...N}, where each sample$x_t^{(n)}$ represents hypothesized state of the target and the weights are normalized such that $\sum_n \pi_t^{(n)} = 1$.

. Re-sampling the particles should be according to their weights. The performance of particle filters rely on importance sampling and the nature of the proposal distribution. More information about particle filters can be found in (Hongwei et al, 2008).

3.2.1 Dynamic model

The location of face in an image frame can be represented by a canonical box such as a square with variable scales and angles. Dynamic model describes the state propagation. In this work the state at time t consists of six parameters Xt = (x_t, y_t, θ_t, s_t, α_t) where x_t, y_t, θ_t, s_t, α_t denote, x & y translations, rotation angle, scale, aspect ratio of the each sample at frame t. Each parameter in X_t is modeled independently by a Gaussian distribution around its counterparts in X_{t-1}. Specifically, the dynamic model, utilized for tracing the motion, can be represented as follows.

$$P(X_t|X_{t-1}) = N(x_1; x_0; \sigma_x^2)N(y_1; y_0; \sigma_y^2)N(\theta_1; \theta_0; \sigma_\theta^2)N(s_1; s_0; \sigma_s^2)N(\alpha_1; \alpha_0; \sigma_\alpha^2) \tag{6}$$

Where N(x; μ,σ^2) denotes the normal distribution.

3.2.2 Observation model

In this section, our goal is to use a representation to describe the human face being traced. Image observations are achieved using a probabilistic interpretation of face's space and skin's color model. X_t predicates the image patch Z_t if it was generated from a subspace of the face model spanned by $\phi$and centered at μ and when this patch contains pixels with skin color. We use two different measurements proposed by Sung and Poggio (Sung & Poggio, 1998) to determine which sample is generated from the face model's subspace. The first value is Mahalanobis distance between the sample and the model's centroid in a lower-dimensional subspace and spanned by the clusters largest few eigenvectors. The second value is a normalized Euclidean distance between the sample and its projection in the lower-dimensional subspace. Therefore the probability of a sample being generated from such subspace is inversely proportional to these distances. Let x (image observation be a high dimensional data sample) and y as the corresponding low dimensional variable which is projected on the face's space. The relationship between them is as follows.

$$y = \phi_M^T \tilde{x} \tag{7}$$

Where $\tilde{x}$= x − μ is the mean-normalized image vector and $\phi_M$is the sub-matrix of ϕ (the eigenvector of Σ) containing the principal eigenvectors. Considering the observation noise, v, as Gaussian with covariance σ^2 we will have:

$$x = \phi_y + \mu + v \tag{8}$$

We assume that the mean μ and covariance matrix Σ of the face space have been estimated using the approach described in previous section. In general, the likelihood of an input pattern x in normal distribution is defined as:

$$P(x|\Omega) = \frac{\exp[\frac{-1}{2}(x-\bar{x})^T\Sigma^{-1}(x-\bar{x})]}{2\pi^{N/2}|\Sigma|^{1/2}} \tag{9}$$

Therefore the likelihood of the generation of the observed sample x from the face model (p(x)) can be computed as follows.

$$P(x) = N(\mu, \phi M \phi^T + \sigma^2 I) \tag{10}$$

Where M is the covariance matrix of y and I is an identity matrix. For estimating the Gaussian distributions, we use the approaches proposed in (Moghaddam & Pentland, 1995) and (Skocaj and Leonardis, 2002). The distances information is obtained as the likelihood of state vector x_t:

$$\pi(x_t^f) = N(x_t; \mu, \phi\phi^T + \sigma^{2} I) N(x_t; \mu, \phi\Sigma^{-1}\phi^T) \tag{11}$$

Where $\sigma^{2} I$ corresponds to the additive Gaussian noise in the observation process and ϕ is the matrix of singular values corresponding to the columns of ϕ. After the image windows corresponding to the most likely particle are accumulated, the Eigen basis and mean of the face model are update using the algorithm in Section 3.1. With considering equations (11), we can update the samples' weights as follows.

$$\pi(x_t^P) = N(x_t; \mu, \phi\phi^T + \sigma^{2} I) N(x_t; \mu, \phi\Sigma^{-1}\phi^T) \tag{12}$$

4. Experimental results

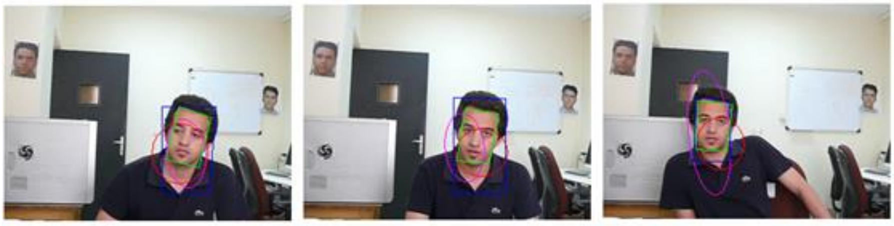
To evaluate the performance of the proposed tracking algorithm, we have provided lots of videos. Each of them, consists of 320 × 240 frames recorded at 25 frames per second. The first step to detect the human face against pictures and other objects is skin segmentation. After determining face candidates, we use Haar-like features, applied as inputs to neural networks and motion information to perform the final face detection. In the next step, face tracking is performed using particle filter based on the face model and skin color information. The face detection is done every other frame. During tracking, the face model is updated with respect to changes in head angle, illumination, facial expression and occlusion.

The proposed method can process 5.5 frames per second on a standard 2 GHz computer. Experimental results are provided to demonstrate the efficiency of the proposed method. In fact, the face detection and tracking method show robustness under varying illumination conditions, occlusion and geometric changes (such as viewpoint and scale changes) and at the same time entails a significantly reduced computational complexity. As a qualitative benchmark, evaluation results of the proposed algorithm (indicated with a green box) comparing with color histogram-based mean shift algorithm (Zivkovic & Krose, 2008) (pink ellipse), Kalman-filter-based procedure (Ho Kim & Doo Kang, 2007) (blue box) and improved Camshift algorithm (Bo Li et al,2011)(red ellipse) are depicted in Figs. 4-6. Results of implementing the proposed method on the first and second image sequences, shown in Figs. 4 and 5, indicate that it provides a comparable performance to Kalman filter and color histogram-based mean shift and cam shift tracker under changes in pose and expressions and scale of the tracked face. Especially when there are fast motions and partially occlusion occurs in the face region under varying illumination condition (Fig. 6), the proposed method achieves the goal despite using a tighter target window around the tracked face. This is while the color histogram-based mean-shift, cam shift and Kalman filter-based trackers perform poorly, experiencing significant drift off the traced face. Mean shift and cam shift based trackers do not have a good performance specially when the tracked face has fast motions and it is close to equicolor objects. It is because of the appearance model of the mean shift and cam shift tracker which are based on histograms of pixel intensities and are not adapted over time. Therefore the proposed algorithm is a robust method for tracking the human face in various conditions by determining the precise location and rotation angle of face.

(a) (b) (c)

Figure 4: The results of implementing the proposed face tracking algorithm (indicated with a green box) comparing with other methods (extended mean shift algorithm (Zivkovic & Krose, 2008) (pink ellipse),

improved Camshift algorithm (Bo Li et al,2011)(red ellipse) and Kalman-filter-based algorithm (Ho Kim & Doo Kang, 2007) (blue box) under changes of pose and expressions and scale by implementing the explained method on the input frames 28, 250, 465 for different face distances and orientation).



(a) (b) (c)

Figure 5: The same results in a different illumination condition from the previous video sequence for input frames 5, 259, 435.

(a) (b) (c)

Figure 6: Results of applying the proposed algorithm on a video sequence in which occlusion occurs locally in the face region and the traced face is close to an equicolor object for input frames 27, 253, 453.

5. Conclusion

We have presented a method for robust and real time face detection and tracking against human pictures and other objects. In face detection approach, a combinational model of skin's color information from three color plates is used to segment the skin regions. Then Haar-like features are fed into the neural networks and the final detection is performed. The face model is updated with respect to changes in head angle, illumination, facial expression and occlusion. Then the detected face is tracked using the particle filter based on the face model, color information and alternative face detection. The proposed algorithm is proved to be efficient compared with other methods specially when occlusion occurs locally in the face region or tracked face is close to equicolor objects.

REFERENCES

[1] Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp, 2001, A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, pp.174–188.
[2] Chiang, Tai, Yang, Huang, and Huang, 2003, A novel method for detecting lips, eyes and faces in real time. *Elsevier. Trans.Real-Time Imaging*, pp.277–287.
[3] C.O. Conaire, N.E. O'Connor, and A.F. Smeaton, 2007, Detector adaptation by maximising agreement between independent data sources. In *CVPR*, pp. 1–6.
[4] T. Kabakli, T. Celik, H. Ozkaramanli, M. Uyguroglu, and A.M. Kondoz, 2004, Robust face tracking using color segmentation aided with connected component labeling and morphology. In *Signal and Image Processing*, pp. 442–445.

[5] Jong-Ho Kim and Byoung-Doo Kang, 2007, Real-time face tracking system using adaptive face detector and kalman filter. In *Human-Computer Interaction. HCI IntelligentMultimodal Interaction Environments*, Springer Berlin / Heidelberg, pp. 669–678.

[6] Levy and Lindenbaum, 2000, Face detection and tracking in a video by propagating detection probabilities. *IEEE. Trans. on Image Processing*, pp.1371–1374,

[7] Hongwei Li, Yi Wu, and Hanqing Lu, 2008, Visual tracking using particle filters with gaussian process regression. In *Advances in Image and Video Technology*, Springer Netherlands, pp. 261–270.

[8] B.Moghaddam and A.P. Pentland, 1995, Probabilistic visual learning for object detection. In *ICCV95*, pp.786–793.

[9] E. Sayed and M. Saad, 2006, Frontal-view face detection in the presence of skin-tone regions using a new symmetry approach. *Computer Science and technology*, pp.127–135.

*[10]*D. Skocaj and A. Leonardis, 2002, weighted incremental subspace learning. In *Workshop on Cognitive Vision*.

[11] S. A. Suandi, S. Enokida, and T. Ejima, 2008, Face pose estimation from video sequence using dynamic bayesian network. In*Motion08*, J.Dittmann,J.Fridrich and P.Wohlmacher, Authentication,Secrecy and Steganalysis. pp. 1–8.

[12] Kah-Kay Sung and Tomaso Poggio, 1998, Example-based learning for view-based human face detection. *IEEE Transactions on pattern analysis and machine intelligence*, pp.88–101.

[13] M. Vasant, S. Padmanabhan, K. Valentina, B. Matthew, G. Dmitry, and K. Rangachar, 2007, A baseline algorithm for face detection and tracking in video. *Optics and Photonics for Counterterrorism and Crime Fighting III, Proc. SPIE*, pp.78–85.

[14] R. C. Verma, C. Schmid, and K. Mikolajczyk, 2003, Face detection and tracking in a video by propagating detection probabilities. *IEEE. Trans. on Pattern Analysis and Machine Intelligence*, pp.1215– 1228.

[15] P. Viola and M. Jones, 2001, Robust real-time face detection. *International Journal of Computer Vision*, pp.747–753.

[16] Jian Yang and Chengjun Liu, 2007, Horizontal and vertical 2dpca-based discriminant analysis for face verification on a large-scale database. *IEEE Transactions on Information Forensics and Security*, pp.781–792

[17] W. Zheng and S. M. Bhandarkar, 2009, Face detection and tracking using a boosted adaptive particle filter. *Visual Communication and Image Representation*, pp.9–27.

[18] Z. Zivkovic and B. Krose, 2008, An em-like algorithm for color-histogram-based object tracking. In *CVPR04*, pp. 798–803.

[19] Yi-Bo Li , Xiao-ling Shen ,Shan-shan Bei ,2011, Real-time tracking method for moving target based on an improved Camshift algorithm, Mechatronic Science, Electric Engineering and Computer (MEC), 2011 International Conference, pp.978 – 981.

A Brief Author Biography

Fatemeh Hourali- was born in Shahrood, Iran. She received his M.Sc. degree from the Dept. of Electrical Engineering, Sahand University of Technology in 2009. Since 2009, Fatemeh Hourali is Head of Electrical Engineering Department for 2011-2015. She has published more than 13 conference and journal papers. She is now an Instructor leading Electrical Engineering Group at the Department of Engineering, Esfarayen University of technology. Her research interests include Image and Video processing, computer Vision, Pattern Recognition and Neural Networks.

Mohammad Hossein Sedaaghi- was born in Tehran, Iran. He received the B.Sc. and M.Sc. degrees from the Sharif University of Technology, Tehran, IRAN, in 1986 and 1987, respectively. In 1998, he received the Ph.D. degree from Liverpool University. He is now a professor at Sahand University of Technology, Tabriz. His research interests include signal/image processing, pattern recognition, machine learning and biometrics.