Improving 2D Boosted Classifiers Using Depth LDA Classifier for Robust Face Detection

Mahmood Rahat[1], Masoom Nazari[2], Akram Bafandehkar[3] and Saeed Shiry Ghidary[4]

**[1] Department of computer engineering and Information Technology, Amirkabir University
Tehran, Iran**

**[2] Aazad Islamic University, Marand Branch
Marand, Iran**

**[3]Department of Computer Engineering, Dezful Branch, Islamic Azad University
Dezful, Iran**

**[4] Department of computer engineering and Information Technology, Amirkabir University
Tehran, Iran**

Abstract

Face detection plays an important role in Human Robot Interaction. Many of services provided by robots depend on face detection. This paper presents a novel face detection algorithm which uses depth data to improve the efficiency of a boosted classifier on 2D data for reduction of false positive alarms. The proposed method uses two levels of cascade classifiers. The classifiers of the first level deal with 2D data and classifiers of the second level use depth data captured by a stereo camera. The first level employs conventional cascade of boosted classifiers which eliminates many of nonface sub windows. The remaining sub windows are used as input to the second level. After calculating the corresponding depth model of the sub windows, a heuristic classifier along with a Linear Discriminant analysis (LDA) classifier is applied on the depth data to reject remaining non face sub windows. The experimental results of the proposed method using a Bumblebee-2 stereo vision system on a mobile platform for real time detection of human faces in natural cluttered environments reveal significantly reduction of false positive alarms of 2D face detector.

Keywords: *Face Detection, Human Machine Interaction, Stereo Vision, False Positive Error Reduction.*

1. Introduction

The task of face detection refers to finding location and boundary of all human faces in an image [1]. Many of machine vision applications use face detection as a primitive task such as Face Recognition, Face Expression Recognition, human detection and tracking. Two major applications of these methods are human machine interaction (HMI) and surveillance systems.

Face detection systems can be classified to three main categories based on the input data:

- 2D face detectors which use intensity images
- 3D face detectors which use 3D model of face
- Multimodal techniques which use a combination of 2D and 3D data.

Because of wide availability of 2D sensing devices, face detectors which use intensity images have acquired more attention. Early works on automatic face detection use different techniques such as correlation, template matching, subspace methods, deformable templates, etc [3-6]. To increase the detection accuracy other clues has been included to the algorithms. Face color is one promising feature that is used by many methods [7-10]. In these methods a color space is chosen (usually HSV) and the region of skin tone is used to detect faces. Variation of human face color, illumination changes, probability of detecting other part of human body as face and skin colored environment are some challenges in these methods. These challenges usually have been solved by applying secondary classifiers. Yang et al [8] used motion analysis, geometric features, and SVM method to verify detected faces. Sandeep et al [9] used edge information to detect the faces.

In [11] authors use examples to model the distribution of face and non face pattern. Six Gaussian clusters are used to compute density functions and a multilayer perceptrons is applied to detect the face. Ebr himpour et al [12] introduced a new implementation of mixture of expert systems called MMLPE which employs a multilayer perceptrons to divide the face detection problem to several subspaces for the experts. Each expert has the ability to detect face on a particular subspace. Finally a gating network selects proper expert(s) based on the problem subspace. Comprehensive survey on 2D face detection can be found in [1, 2].

Another approach for face detection is using 3D data and depth information as input. Colombo et al [13] use laser range data to create 3D model of face. A feature-based approach in combination with a holistic one is used for face detection. Using curvature analysis some face features like eyes and nose are detected. These features are processed by a PCA-based classifier to distinguish between faces and non faces. Malassiotis and Strintzis [14] project a color pattern on scene to compute range image. Their algorithm consists of two steps. The First step detects human face using global moment descriptors and prior geometric constraints. In the second step they estimate head pose using robust knowledge-based 3D feature detection and localization techniques. In contrary to the 2D methods, 3D face detectors have attracted less attention Because of high cost of 3D sensing devices.
Although it is known that 3D data is less sensitive to illumination changes [15, 16], still many issues exists which limit their use in face detection systems. Real time application of 3D sensors is challenging and some of them do not generate models good enough for detecting faces from distance.

There are algorithms which perform face detection using 2D images extremely rapidly with high accuracy. These systems usually suffer false positive in cluttered environment while have really low false negative rates. Meanwhile limiting the world to only 2D data is not acceptable. Many of hard problems on 2D world can be easily solved using 3D data and vice versa. For example a 3D face detector may easily distinguish between a splotchy curtain and a human face but 2D face detector may be fooled. In the other hand using only 3D data will increase computation time and many 2D promising features such as eye, eyebrow and nose would be lost. All of these will bring together the idea of using both 2D and 3D data to complement each other.
Boosting classifiers are one of the best choices for mixing two types of data for building multimodal face detector. Our work is based on the work of Viola and Jones [19]. The proposed system uses both 3D range data and 2D gray-level images captured by stereo camera. Two level cascades of classifiers are applied for face detection. Our major contribution is to reduce false detection of a boosted classifier on 2D images using depth data. First level of our system generates a salience map of face like regions. The detected face regions are verified using a heuristic classifier in conjunction with depth LDA.

The remainder of this paper is organized as follows. In section 2 our face detection system is described. Then experimental result is presented in section 3. Finally, section 4 draws conclusions and some motivations for future works.

2. System overview

The proposed method is motivated by the excellent work of Viola and Jones [19]. Their system has three key novelties. The first novelty is introducing a new image representation called integral image. Using this technique, they are able to calculate image features very quickly. The number of these features is far larger than the number of pixels and applying all of them on the classifier will increase computation time. The second novelty is using a variant of AdaBoost to select important features. Each selected feature form a weak classifier:

$$h_j(x) = \begin{cases} 1 & if\,(p_j f_j(x) < p_j \theta_j) \\ 0 & othervise \end{cases} \qquad (1)$$

$h_j(x)$ Is a weak classifier, $f_j(x)$ is a feature, θ_j is a threshold, p_j is a parity to indicate the direction of inequality sign and x is a sub window of an image.

They showed that a very small number of these features can form an efficient classifier. As a result, the third novelty they placed classifiers on a decision tree structure called cascade classifiers. This system can be used to detect faces extremely fast with high detection rate. This method is applied as the first level of our face detector as shown in Fig.1 In this level all of images in the input sub windows are intensity images. The system is improved by adding a secondary level which uses depth data to verify detected faces. This level contains two cascade classifiers which will be discussed in section 2.1 and 2.2. Both of them use depth data to reject any remaining non face sub windows. This technique can be used to improve efficiency of any other 2D face detector. The entire system is in the form of a decision tree. Rejecting of one sub window in each step will cause no further processing for that sub window. Only the output of the last 3D classifier will be considered as face.

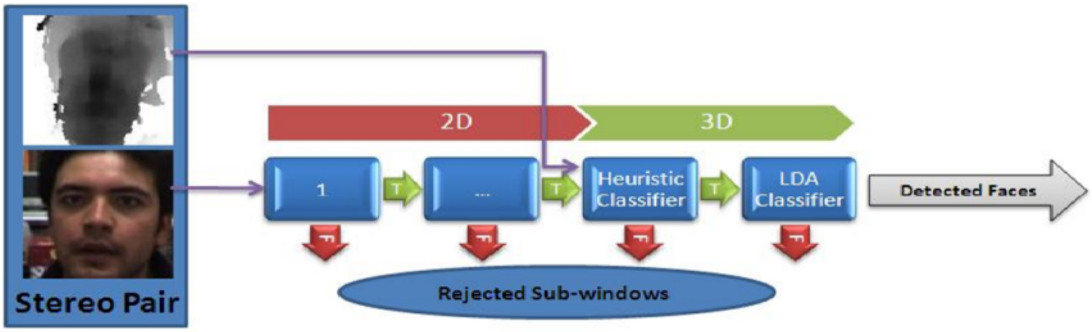

Fig. 1 Two levels of cascades classifiers used for face detection. 2D data is fed into a boosted classifier. After completion of the 2D layer, corresponding depth model of all remaining sub windows are computed. These depth models are used as the input to the 3D classifiers. Sub windows which verify through 3D classifiers would be determined as face.

2.1 Heuristic classifier

In the propose method there are two classifiers on 3D layer. The first one is a heuristic classifier which checks validity of a simple fact. The fact is that the near faces will have bigger size and the farther faces will have smaller sub window size. Assuming each sub window contains a face, the distance of detected face to camera using range images captured by stereo camera can be estimated.

The outputs of a stereo vision system can be two types of images. First one is an intensity image which can be called as 2D image and the second one is a 3D range image corresponding to the 2D image that known as depth image. In 2D images each pixel denotes intensity value sensed by camera and in depth images each pixel denotes distance of camera plane to the nearest object exposed to it in that point.

As the first step, corresponding depth models of all output sub windows from 2D layer are computed. Then the background of detected face image can be removed from depth pair of each 2D sub window using Peak analysis of the disparity histogram [15]. Detected face can be segmented out by applying thresholds on range images.

After removing background, a median filter with size 3 is applied to all pixels of range image to remove noise. Mean of all pixel value is used as face distance to camera.
Using an approximate function shown in Fig.2, a minimum and maximum acceptable face width for each distance to camera can be computed. This function is experimentally developed using a stereo camera. If width of the detected sub window does not have the value between minimum and maximum acceptable values, the heuristic classifier will reject that sub window; otherwise the sub window is passed to the next classifier. Fig.3 shows an example of camera frame processed by the proposed method. The upper red box on the right side of yellow box is rejected by the heuristic classifier. Distance of wall to camera is 1.3 meter. An acceptable face on this distance is about double of the size of upper red box.

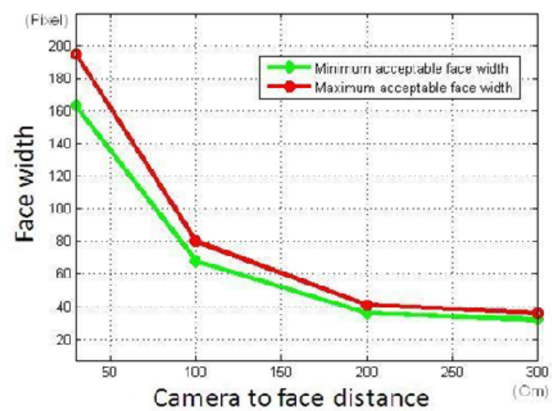

Fig. 2 A piecewise linear approximation of the face width using face distance to the camera. As face width may vary for different face pose a minimum and maximum acceptable face width are computed in four const distances to the camera. The intermediate values are estimated using a linear function.

2.2 The depth LDA classifier

LDA classifier (known as Fisher linear discriminant) [17, 18] applies a linear hyper plane defined by the equation (2) to distinguish between two classes of data.

$$w^t x + w_0 = 0 \qquad (2)$$

Where $w = s_w^{-1}(\mu_1 - \mu_2)$, $s_w = \frac{1}{2}(\Sigma_1 + \Sigma_2)$, μ_1 is first class mean, μ_2 is the second class mean, $\sum_1$ is the first class covariance, $\sum_2$ is the second class covariance matrix and $w_0 = -\frac{1}{2}(\mu_1 + \mu_2)s_w^{-1}(\mu_1 - \mu_2)$.

This classifier is named depth LDA because it addresses depth model of sub windows. After removing background from depth sub windows using technique described in section 2.1, a contrast stretching method is applied to each sub window to normalize them and make the range of all pixel data to the range of 0 to 255. Many depth models from face and non face sub windows are used to train the classifier. The trained classifier is capable of determining which sub window has the features of a depth model of a human face.

Each sub window which passes through this classifier will be considered as face. Detected faces have the features of a face on both 2D and depth data. In Fig.3 the yellow box has been passed through all the classifiers and will be determined as face. The lower red box is rejected by the depth LDA classifier for not having the shape of a human face depth model.

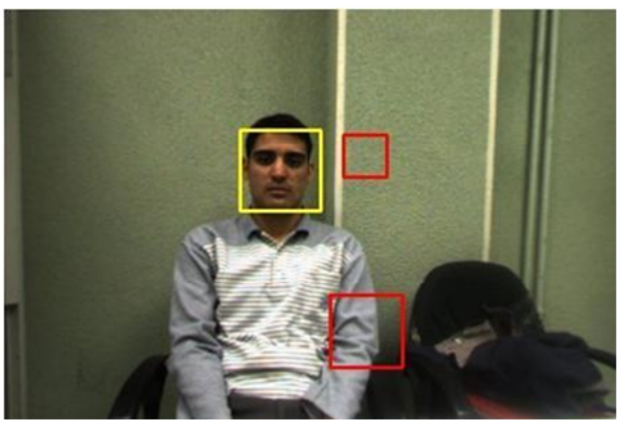

Fig. 3 Yellow box shows detected face. The corresponding Sub window is accepted by all classifiers on the cascade. Red boxes indicate that corresponding sub window is accepted by the 2D layer of proposed system but 3D layer has rejected them. Red boxes would not be considered as face.

3. Experimental results

The first layer of the proposed algorithm consists of a conventional boosted classifier. This algorithm has been already implemented in the Intel's OpenCV library. In our experiments, default Haar cascade face detector database of this library is used to implement first layer of the system. In the second layer a LDA classifier has been applied.

Fig.4 shows examples of face depth model and non face models captured by stereo camera. The first and second rows show depth models of faces with different pose and different distance to the camera. Depth models of faces which are near to the camera contain information on eyes holes and nose tip. In the other hand faces which are very far to the camera such as faces in the first two models of the second row contain less information about face features. These models are used to train a depth LDA classifier which can distinguish between face and non face in variety of pose and distances to camera. As there are only two fix classes; face and non face, the training process will be done off-line and computed projection vectors will be used to perform classification task very quickly.

Recently a work on stereo face detection has been done recently by Kosov et al [20]. They improved state of the art 2D face detectors by evaluating disparity map images captured by calibrated stereo camera. A principal component analysis approach on disparity images is used to improve efficiency of a 2D face detector. One important weakness of their work is detection of only frontal faces. Proposed method copes with this limitation by using LDA method to model the variations on two classes of data. Since PCA constructs the face space without using face class information, it does not perform well on the pose variation problem [21]. Also there are experimental evidences that assert, LDA outperforms PCA under varying illumination [22].

A Bumblebee-2 stereo vision camera was used to capture 2D and depth data required for face detection. This camera is installed on a mobile service robot as shown in Fig.5. The robot uses this system to detect human face and track it in natural cluttered environment.

Both 2D and depth data in the proposed method have to be applied. Thus there is no standard datasets for face detection Experiments. Thus two real-time experiments were conducted to verify the efficiency of the proposed algorithm. In these experiments, the proposed method is compared with Viola and Jones face detector [19] as a conventional face detector which is a benchmark algorithm for comparison task and also Kosov et al [20] face detector as a stereo method. The results reveal that the efficiency of our proposed method is higher than the mentioned algorithm when we consider all parameters of a good face detector (such as detection rate and false positive).

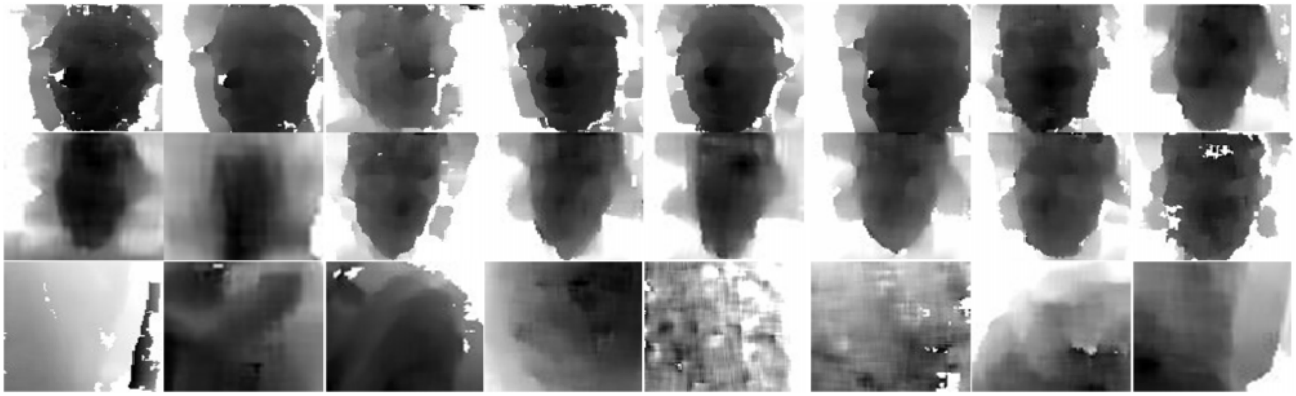

Fig. 4 Examples of face and non face samples used for training of the depth LDA classifier. The first two rows show face samples and the third row shows non face samples.

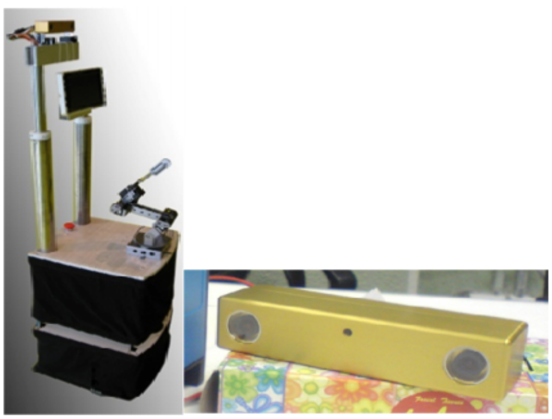

Fig.5 Bumblebee-2 stereo vision camera installed on Sourena service robot.

In the first experiment the robot has navigated in a cluttered environment looking for human faces. Three people are placed on the environment. Then 328 frames of the robot camera are processed by three face detectors. Table 1 shows the results. Detection rate of the proposed method is slightly less than the viola and Jones method because some actual face subs windows are incorrectly rejected in 3D layer. But in the proposed method the number of non face sub windows which are detected as face is significantly reduced which means that proposed method has very low false positive rate than the viola and Jones face detector. In the proposed method 96 percent of the detected faces were actually a face but only 53.6 percent of faces which were detected by Viola and Jones method were actually a face.

In the second experiment the robot has navigated in a cluttered environment with highly abnormal illumination. Three people are placed on –the mentioned environment. 435 frames of the robot camera are processed by three methods. Table 2 shows results of the second experiment. In this experiment 97.9 percent of faces which were detected by proposed method were actually a face while only 66.5 percent of the faces which were detected by Viola and Jones method were actually a face.

Table 1: First experiment is conducted in a cluttered environment in which 328 frame of the stereo camera is processed.

Method	Detection rate	Number of non face sub windows detected as face	Number of processed frames
Viola and Jones Ref 19	96.6	251	328
Ref 20	81	97	
Proposed method	92.3	12	

Therefore, by adding the proposed extra classifiers to the viola and Jones algorithm the number of non faces that selected as face significantly decreases.

Table 2: Second experiment is conducted in a cluttered environment with highly abnormal illumination in which 435 frame of the stereo camera is processed.

Method	Detection rate	Number of non face sub windows detected as face	Number of processed frames
Viola and Jones Ref 19	66.6	122	435
Ref 20	51.2	38	
Proposed method	64.7	5	

Fig.6 shows the output of proposed face detector on some test images captured by stereo camera. Note that the printed picture in fig.6.d is not considered as face because of not having 3D features of a human face however the statue in fig.6.e is considered as human face. This ability will enables service robots to distinguish between printed face images and real human faces.

a	b	C
d	e	F

Fig. 6 Test images processed by proposed face detection system.

4. Conclusions

In this paper, a new approach is presented for reducing false positive alarm of 2D face detector using depth data while keeping computation time low. In the first step, human faces are detected using a state of the art 2D face detector. Then the detected faces are verified using depth data captured by stereo camera. A recent psychophysical research [23] on human visual system states that Stereoscopic information about the three-dimensional structure of the face is one important feature to reduced viewpoint costs for face recognition tasks. Existence of Stereoscopic information will help humans to recognize faces across the pose. This idea can be extend for face detection. A human face can be detected more accurately using both 2D and 3D data. All the objects in the world have both 2D and 3D features which can be used for detection. Removing one of them, cause reduction of the accuracy of object detection systems. Two real-time experiments conducted to compare proposed face detection method with two prominent face detectors. To carry out these experiments a mobile robot equipped with stereo camera was used. The robot used this face detection method to detect human face and track it in natural cluttered environment. The experiments result showed that proposed method significantly reduces false positive alarm of 2D face detector while keeps detection rate high. Because of low computation time of the proposed method, our aim is working on an embedded stereo vision camera to apply this method on it which enable smaller robots such as low cost humanoid robots to detect human faces accurately in natural cluttered environments.

References

[1] Yang, M.-H., D. J. Kriegman, et al, "Detecting Faces in Images: A Survey." IEEE Transactions on pattern analysis and machine intelligence 24. 2002.

[2] Hjelmas, E, "Face Detection: A Survey." Computer Vision and Image Understanding 83, 2001, pp. 236–274.

[3] Pentland A,Moghaddam B, Starner T,. "View-based and modular eigenspaces for face recognition." In: Proceedings of the computer vision and pattern recognition, 1994, pp. 84–91.

[4] Yuille AL, "Deformable templates for face recognition." J Cognitive Neuroscience 3(1), 1991, pp. 59–70

[5] Chellappa R, Wilson CL, Sirohey S, "Human and machine recognition of faces: a survey." Proc IEEE 83(5), 1995, pp. 705–740.

[6] Samal A, Iyengar PA, "Automatic recognition and analysis of human faces and facial expression a survey." Pattern Recognition 25(1), 1992, pp. 65–77.

[7] D. Saxe and R. Foulds, "Toward Robust Skin Identification in Video Images." Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, 1996, pp. 379-384.

[8] Yang, J., X. Ling, et al, "A face detection and recognition system in color image series." Mathematics and Computers in Simulation 77, 2008, pp. 531–539.

[9] Sandeep, K. and A. N. Rajagopalan, "Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information." ICVGIP. 2002.

[10] J. Chun, K. Min, K. Park, and J. Son, "An Automatic Extraction of Face and Facial Feature from Face Images using Skin Color and Active Contour Model", Conf. on Imaging Science, Systems & Technology, 2003, pp. 428-433.

[11] Sung KK, Poggio T, "Example-based learning for view-based human face detection." IEEE Transactions on pattern analysis and machine intelligence 20(1), 1998, pp. 39–51.

[12] Ebrahimpour, R., E. Kabir, et al, "Face Detection Using Mixture of MLP Experts." Neural Processing Letters 26, 2007, pp. 69–82.

[13] Colombo, A., C. Cusano, et al, "3D face detection using curvature analysis." Pattern Recognition 39, 2006, pp. 444 – 455.

[14] Sotiris Malassiotis and M. G. Strintzis, "Robust real-time 3D head pose estimation from range data." Pattern Recognition 38, 2005, 1153 – 1165.

[15] WANG, J.-G., E. T. LIM, et al, "Real-time Stereo Face Recognition by Fusing Appearance and Depth Fisherfaces." VLSI Signal Processing 49, 2007, pp. 409–423.

[16] Tsalakanidou, F., D. Tzovaras, et al, "Use of Depth and Colour Eigenfaces for Face Recognition." Pattern Recogn. Lett 24, 2003, pp. 1427–1435.

[17] Webb, "A Statistical Pattern Recognition", 2nd edition. Wiley, New York, 2002

[18] Abdleazeem, S. and E. El-Sherif, "Arabic handwritten digit recognition." Springer, 2008

[19] Viola, P. and M. Jones, "Robust Real-time Object Detection." International Journal of Computer Vision, 2002.

[20] Kosov, S., K. Scherbaum, et al, "Rapid stereo-vision enhanced face detection." International Conference on Image Processing, 2009.

[21] Du, S. and R. Ward, "Face recognition under pose variations." Journal of the Franklin Institute 343, 2006, pp. 596–613.

[22] Belhumeur, P. N., J. P. Hespanha, et al, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection." IEEE Transactions on pattern analysis and machine intelligence 19, 1997.

[23] Burke, D., J. Taubert, et al, "Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces." Vision Research 47, 200 , pp. 2164–2169.

Mahmoud Rahat was born in Tehran, Iran, in July 1985. He received the BS degree in Computer science from Tehran University, Tehran, Iran and the MS degree in computer engineering from Amirkabir University, Tehran, Iran, in 2008 and 2010. His research interests include human and machine vision, neural networks, and pattern recognition.

Masoom Nazari was born in 1985. He received the B.Sc. degree in electronics engineering from Sabzevar University, sabzevar, Iran, in 2007, the M.Sc. degree in electronics engineering from the Electrical Engineering Department of Shahid Rajaee University, in 2010. His research interests include human and machine vision, neural networks, and pattern recognition.

Akram Bafandehkar was born in 1986. She received the B.Sc. degree in computer engineering from Azad University, Dezful, Iran, in 2008, the M.Sc. degree in Artificial Intelligence from the University of Shiraz, in 2011. She is a member of young research club since 2008 till now. Her research interests are machine vision, digital image processing, pattern recognition and machine learning.

Saeed Shiry Ghidary was born in 1967. He is Assist Prof. of Computer Engineering & Information Technology Department. He received the B.Sc. degree in electrical engineering from Electronic engineering, Amirkabir University, Tehran, Iran, in 1990, the M.Sc. degree in Computer architecture, Amirkabir University, Tehran, Iran, in 1994 and the Ph.D degree in Robotics, Kobe University, Kobe, Japan, in 2002. His research interests include fuzzy modeling, machine learning robotics and fuzzy modeling.