

Persian Sign Language Recognition Using Radial Distance and Fourier Transform

Bahare Jalilian

Department of Computer Engineering, Kermanshah Science and Research Branch, Islamic Azad University
Kermanshah, Iran
baharejalilian@gmail.com

Abdolah Chalechale

Department of Computer Engineering, Razi University
Kermanshah, Iran
chalechale@razi.ac.ir

Abstract — This paper provides a novel hand gesture recognition method to recognize 32 static signs of the Persian Sign Language (PSL) alphabets. Accurate hand segmentation is the first and important step in sign language recognition systems. Here, we propose a method for hand segmentation that helps to build a better vision based sign language recognition system. The proposed method is based on YCbCr color space, single Gaussian model and Bayes rule. It detects region of hand in complex background and non-uniform illumination. Hand gesture features are extracted by radial distance and Fourier transform. Finally, the Euclidean distance is used to compute the similarity between the input signs and all training feature vectors in the database. The system is tested on 480 posture images of the PSL, 15 images for each 32 signs. Experimental results show that our approach is capable to recognize all 32 PSL alphabets with 95.62% recognition rate.

Index Terms — Persian sign language, hand gesture recognition, Gaussian model, centroid distance, Fourier transform, Euclidean distance

I. INTRODUCTION

Sign language is a non-verbal visual language that used by deaf. In order to improve communication between deaf and hearing people, research in automatic sign language recognition is needed. One application of sign language recognition systems are the use on Internet because many deaf people are unable to use the World Wide Web and communicate by e-mail in the way normal people do, since they commonly have great difficulties in reading and writing. The reason for this is that normal people learn and perceive written language as a visual representation of spoken language. For deaf people, however, this correspondence does not exist, and letters—which encode phonemes—are just symbols without any meaning.

The major part of sign language is comprised of gestures and postures hand that imply meaningful

information. In gesture recognition, sign Language is translated for dynamic hand motions. In posture recognition, sign Language is recognized for static alphabets and numbers. Sign language recognition system is useful for the hearing impaired to communicate with the normal people.

Research on automatic sign recognition has been started since around 1995 [1]. Many different techniques for hand gestures recognition have been analyzed, such as fuzzy logic [2], Hidden Markov Models (HMMs) [3], neural networks [4, 5] and support vector machine [6]. Tsai and Huang used Support Vector Machine (SVM) to recognize the static sign and apply HMM model to identify the dynamic signs in the Taiwanese Sign Language (TSL) [6]. Mehdi and Niaz khan [7] have proposed an American Sign Language Recognition based on sensor glove to capture the signs.

Artificial neural networks are used to recognize the sensor values coming from the sensor glove. Lee and Tasi have used 3D data and neural network to interpret Taiwan sign language [8]. Assaleh et al [9] have proposed the first continuous (Arabic Sign Language) ArSL system that was able to recognize ArSL using HMM. Hough transform and neural network have been used for recognition American Sign Language [10]. In [11] a system for Arabic Sign Language Recognition is designed that based on a Gaussian skin color model to detect the signer's face and a Hidden Markov Model to recognition signs. Wavelet transform and neural network have been used for recognition dynamic signs of American Sign Language [12].

A system for Arabic Sign Language Recognition is provided which the input image is converted into YCbCr color space and a skin profile is used to detect the skin color from the YCbCr image. Principal Component Analysis algorithm is used to compose the feature vectors for signs and gestures library [13]. Kiani Sarkaleh et al [14] proposed a Persian sign language recognition system which is capable of recognizing 8 Persian signs by discrete wavelet transform and a multi layered Perceptron (MLP) Neural Network (NN) to classify the selected images. In this system background of all images was black. Karami et al [15] designed a

system to recognize static signs of the Persian Sign Language (PSL) alphabets, background of all input images were uniform and black. Hand images are converted to grayscale images. Their system is based on the wavelet transform and neural networks.

Paulraj et al [16] introduced a very simple Malaysian Sign Language recognition system based on the area of the objects in a binary image and Discrete Cosine Transform (DCT) for extracting the features from the video sign language. In preprocessing stage the movie frames are converted into indexed image format and Average filter is applied on these images and the unwanted noises are removed. A simple sign language recognition system was developed using the skin color segmentation, moment invariants for features extracting and neural network model [17]. In most of approaches, input hand images have been assumed that have uniform and plain background.

The signs can be either static or dynamic. A static sign is a particular hand shape and pose which represented by a single image. A dynamic sign is a moving gesture that represented by a sequence of hand images. The Persian sign language consists of approximately 1075 gestures of the common alphabets and words that the basic Persian sign alphabet is composed of 37 static and dynamic gestures [19]. Our approach focuses on the recognition of 32 static hand images. In this paper, we propose a system to interpret static gestures of alphabets in Persian sign language (PSL). First of all, an effective hand segmentation method is presented to detecting hand region in a

complex background with changing illumination condition. The hand segmentation method begins by modeling human skin color in YCbCr color space using a database of skin pixels. Skin color distribution is modeled as a single Gaussian model. Similarly a non-skin or background model is built using a database of non-skin pixels. Thus a Skin Probability Image is obtained in which the gray level of each pixel represents the probability of the corresponding pixel in the input image to represent skin. Next, hand gesture features are extracted by radial distance and Fourier transform.

The rest of the paper is organized as follows. A detail description of the proposed PSL recognition system is presented in Section II. In Section III, we exhibit our experimental results with discussion. Finally, the conclusions and further work are presented in Section IV.

II. THE SUGGESTED SYSTEM

The suggested system is provided to recognize 32 static sign of the Persian sign language (PSL) alphabets. These signs are shown in Figure 1. We assumed that the input images include exactly one hand. Our system has two main phases: hand segmentation phase (hand region detection) and the feature extraction phase. The block diagram of the proposed recognition system is shown in Figure 2 and its main steps are discussed in the following.

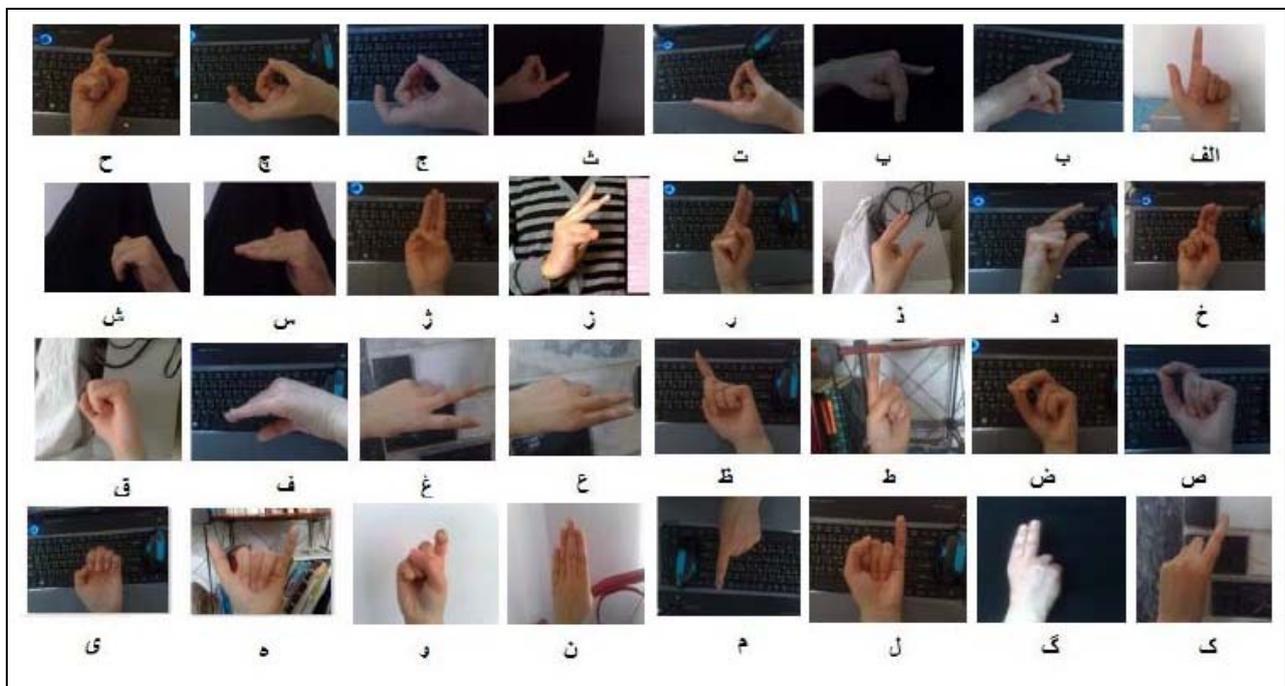


Figure 1. The Persian Sign Language Alphabets.

A. Hand Segmentation

Image Pre-processing and detection of the hand

region is necessary for image enhancement and for getting good results in sign language recognition systems. The proposed system introduces a method for

segmenting hands using YCbCr color space, single Gaussian model, Bayes rule and morphology operators. The block diagram of this phase is shown in Figure 3.

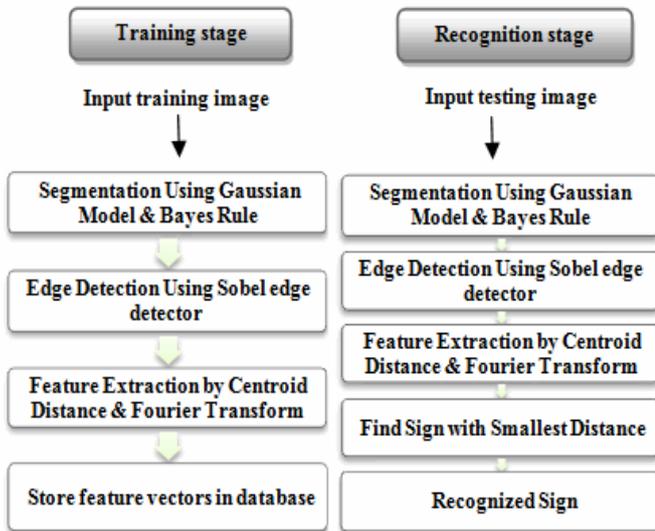


Figure 2. The Block Diagram of the Proposed Recognition System

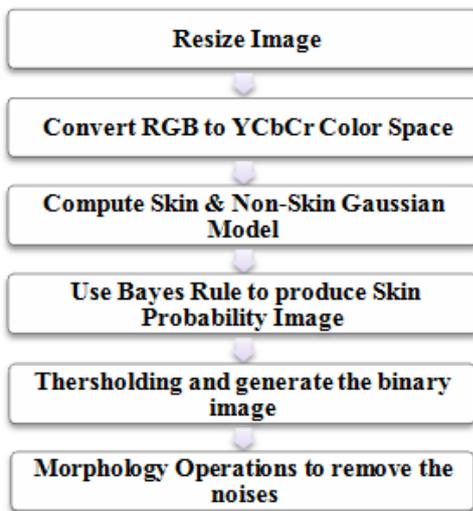


Figure 3. The Block Diagram of the Segmentation Phase in the Proposed Recognition System

1) *YCbCr Color Space Conversion*

One important factor that should be considered while building a statistical model for color is the choice of a right color space. For most images, the RGB color space is considered as the default color space. In order to convert into other color spaces, we can apply linear or non-linear transformation on the RGB components. In this algorithm, after that color images are resized to 128 by 128 pixels, the input RGB image is converted in to YCbCr images due to the fact that RGB color space is more sensitive to different light conditions so we need to transform the RGB values in to YCbCr. The color space transformation is assumed to decrease the overlap

between skin and non-skin pixels to classify skin-pixel and to provide robust parameter against varying illumination conditions. RGB values can be transformed to YCbCr color space using (1):

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

If only the chrominance component is used, segmentation of skin colored regions becomes powerful in this process. Therefore, the variations of luminance component are eliminated as much as possible by choosing the CbCr plane (chrominance components) of the YCbCr color space to build the model. Research has shown that skin color is clustered in a small region of the chrominance space [4], as shown in Figure 4.

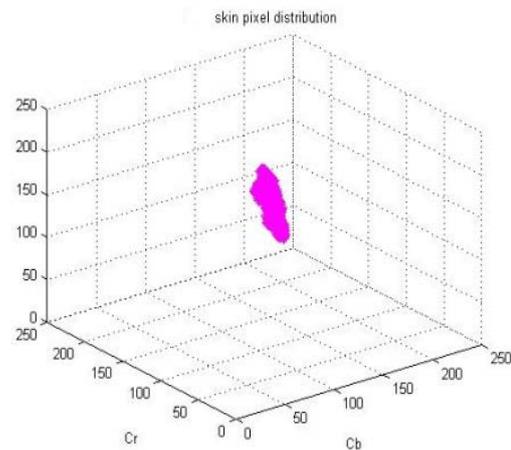


Figure 4. Skin color Distribution in YCbCr Color Space

2) *Single Gaussian Model*

The skin color distribution in CbCr plane is modeled as a single Gaussian model. According to section 2.1 the reason for using a single Gaussian model is the localization of skin color to a small area in the CbCr chrominance space. This step begins with the modeling of skin and non-skin color using a database of skin and non-skin pixels, respectively. A database of labeled skin pixels is used to train the Gaussian model. Some of skin images from the database are shown in Figure 5. The mean and the covariance of the database characterize the model. Images containing human skin pixels as well as non-skin pixels are collected. The skin pixels from these images are carefully cropped out to form a set of training images.

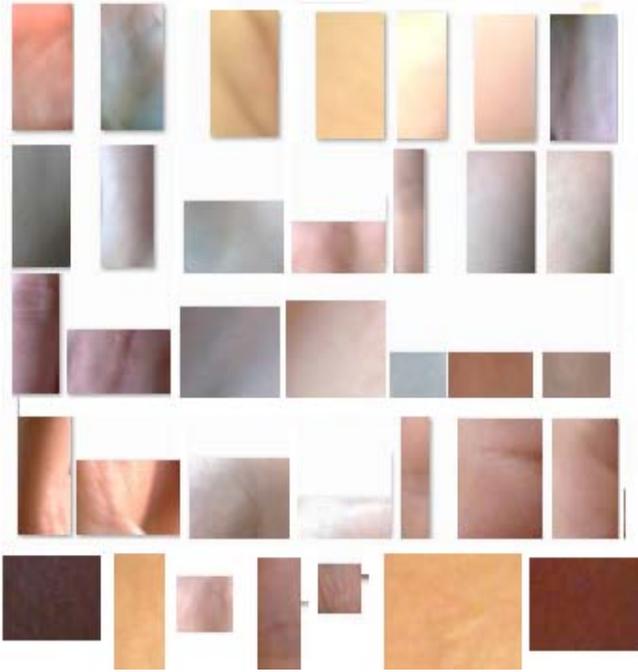


Figure 5. Some of Skin Images from the Skin Database

Let $c = [Cb Cr]^T$ denote the chrominance vector of an input pixel. Then the probability that the given pixel lies in the skin distribution is given by (2):

$$p(c / skin) = \frac{\exp\left[-\frac{1}{2}(c - \mu_s)^T \Sigma_s^{-1}(c - \mu_s)\right]}{2\pi\sqrt{|\Sigma_s|}} \quad (2)$$

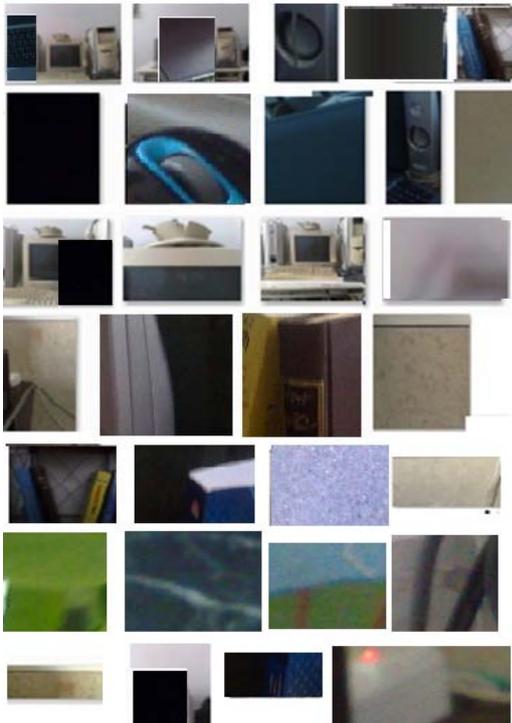


Figure 6. Some of Non-Skin Images from the Non-Skin Database

where μ_s and Σ_s represent the mean vector and the covariance matrix of the training pixels, respectively. Thus the mean and the covariance have to be estimated from the training data to characterize the skin color distribution as illustrated by (3) and (4). In these equations, n is number of samples in training set.

$$\mu_s = \frac{1}{n} \sum_{j=1}^n c_j \quad (3)$$

$$\Sigma_s = \frac{1}{n-1} \sum_{j=1}^n (c_j - \mu_s)(c_j - \mu_s)^T \quad (4)$$

Then, a Gaussian model similar to skin model is built for non-skin pixels also which is called the non-skin or the background model. A database of background region is used to train the non-skin model. Some of non-skin images from the database are shown in Figure 6. Then the probability that the given pixel lies in the non-skin distribution is $p(c/non-skin)$. The skin model and non-skin model are used to obtain the Skin Probability Image of an input color image. Once the skin color and background are modeled using the single Gaussian, these can be used to calculate the probability of an input pixel representing skin, i.e. $p(skin/c)$, where c is the input color value. $p(c/skin)$ is again used to compute the required probability $p(skin/c)$. To compute this probability, the Bayes rule is used [19]:

$$p(skin / c) = \frac{p(c / skin)P(skin)}{p(c / skin).P(skin) + p(c / non - skin).P(non - skin)} \quad (5)$$

To calculate the probability, $p(skin/c)$, for each input pixel, The probabilities $p(skin)$ and $p(non-skin)$ can be estimated from skin and non-skin image in the training database [19]. In this study, for training set we assumed that all pixels are belong to the skin or non-skin clusters. Hence, we used:

$$p(skin) = p(non - skin) = 0.5 \quad (6)$$

$$p(skin / c) = \frac{p(c / skin)}{p(c / skin) + p(c / non - skin)} \quad (7)$$

Thus, the two conditional probabilities and the above ratio are computed pixel-by-pixel to give the probability of each pixel representing skin given its chrominance vector c . This results in a gray level image where the gray value at a pixel indicates the probability of that pixel representing skin. This is called the Skin Probability Image (SPI) given by (8):

$$SPI(i, j) = p(skin / c_{ij}) \quad (8)$$

Where a is a proper scaling factor and c_{ij} is the chrominance value of *pixel* (i, j). Here a is chosen to be 255 so that the highest probability value results in a gray level of 255 in the Skin Probability Image. Then gray

image obtained is converted into a binary image by Otsu thresholding.

3) Morphology Operations

The binary image obtained in the previous section may contain white pixels at non-skin regions (background) where the background color resembles the color of skin or black pixels at hand region. These noises may be caused due to bad lighting conditions or existing pixels similar to skin pixels in those regions. In order to detect the hand clearly, it has been further implemented morphology operations, to fill up the black pixels on the segmented hand and white pixels on background. There are two operations involve namely dilation and erosion. Firstly, dilation operation is performed. Dilation adds pixels to fill up any missing pixels in hand region. Secondly, erosion operation is performed. Erosion removes any white pixels which do not belong to the hand region. This stage is performed to improve the result of hand segmentation.

Figure 7 shows the images obtained after applying each of steps in the proposed segmentation algorithm. Figure 7(a) is original image; Figure 7(b) is Skin Probability Image. Figure 7(c) is the binary image after Otsu thresholding. Figure 7(d) is the hand detected after morphology operation.

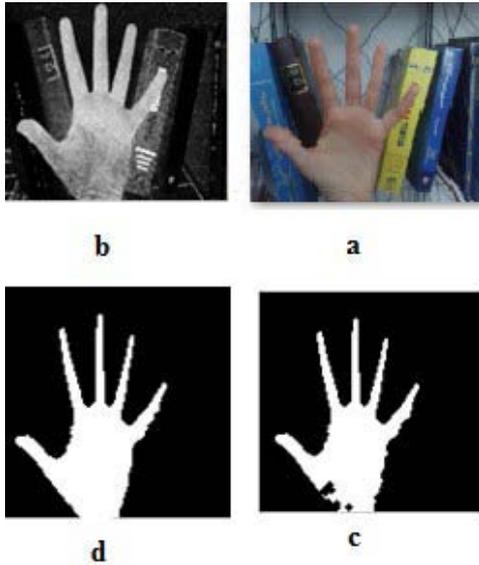


Figure 7. (a): Original Image, (b): Skin Probability Image, (c): Segmented Hand After Otsu thresholding (d): The Hand Detected After Morphology Operations.

B. Feature Extraction

In this phase, first, we apply Sobel edge detector to the hand segmented image, then features are extracted from edges of the hand region. Next, we use radial distance model to obtain a 1-D functional representation of a boundary shapes (signatures) and to build feature vectors [20]. Radial distance technique is based on the distance from the centroid of the shape to the boundary edge pixels as a function of angle, as shown in Figure 8.

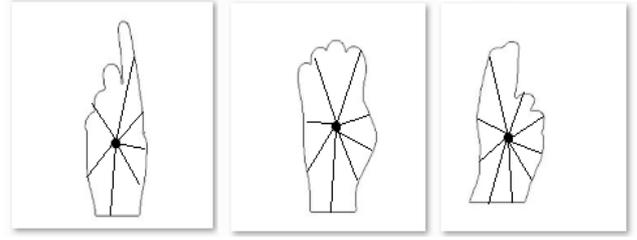


Figure 8. Implemented Radial Distance Technique on Hand Region.

The boundary of a shape consists of a series of edge points. A "radial" is a straight line joining the centroid to an edge point. The "centroid" is located at the position (x_c, y_c) such that x_c and y_c are calculated, respectively, using (9) and (10).

$$x_c = \frac{\sum_{u=0}^{N-1} x(u)}{N} \quad (9)$$

$$y_c = \frac{\sum_{u=0}^{N-1} y(u)}{N} \quad (10)$$

Lengths of a shape's radial from its centroid, $r(u)$, are computed by (11). $(x(u), y(u))$ are coordinates of edge points and (x_c, y_c) is coordinate of centroid.

$$r(u) = \sqrt{(x(u) - x_c)^2 + (y(u) - y_c)^2} \quad (11)$$

The signatures generated by this technique are invariant to translation, but they do depend on slope and scaling (hand' size). Normalization with respect to slope and scaling can be achieved by Fourier transform using (12).

$$a_n = \frac{1}{n} \sum_{t=1}^n r(u) e^{\frac{-j2\pi n u}{N}} \quad n = 1, \dots, N \quad (12)$$

Where, N is the number of selected points on the boundary of the hand shape. The computed coefficients, a_n , are divided by the maximum coefficient. We considered only first $N/2$ coefficients of Fourier for the final feature vector, that is:

$$fv = \left[\frac{|a_1|}{|a_0|}, \frac{|a_2|}{|a_0|}, \dots, \frac{|a_{n/2}|}{|a_0|} \right] \quad (13)$$

Therefore, each sign is expressed by these computed coefficients as its feature vector.

C. Classification and Recognition Stage

A feature vector is computed for each image sign or gesture in the training set and stored in the training data set. When the system receives a new sign, it segments

proposed approach needs no constraint such as gloves, sensors and illuminations. In the future, we intend to extend the proposed method to construct a complete Persian sign language recognition system including both static and dynamic signs in order to help deaf and hearing impaired people to communicate with others.

REFERENCES

- [1] M.B. Waldron and S. Kim, "Isolated ASL sign recognition system for deaf persons," *IEEE Transactions on Rehabilitation Engineering*, 3, 261–271, 1995.
- [2] C.S. Lee, G.T. Park and J.S. Kim, "Real-time Recognition System of Korean Sign Language based on Elementary Components," *1997 IEEE Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*. Vol. 3, pp. 1463 – 1468.
- [3] M. AL-Rousan and K. Assaleh, "Video-based signer-independent Arabic sign language recognition, using hidden Markov models," *Appl Soft Comput* 9(3), 2009, pp.990–999.
- [4] E. Stergiopoulou and N.Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," *Engineering Applications of Artificial Intelligence* 22, 2009, pp.1141–1158.
- [5] P. Vamplew and A. Adams, "Recognition of sign language gestures using neural networks," *Australian Journal of Intelligent Information Processing Systems*, 1998, vol. 5, 94–102.
- [6] B.L. Tsai and C.L. Huang, "A Vision-Based Taiwanese Sign Language Recognition System," *IEEE International Conference on Pattern Recognition*, 2010, Vol. 5, pp.3683-3686.
- [7] S.A. Mehdi and Y. Niaz Khan, "Sign Language Recognition Using Sensor Gloves," *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02)*, 2002, Vol. 5, pp.2204-2206.
- [8] Z. H. Lee and C. Y. Tsai, "Taiwan sign language (TSL) recognition based on 3D data and neural networks", *Expert Systems with Applications*, 2009, vol.36, 1123-1128.
- [9] K. Assaleh, T. Shanableh, M. Fanaswala and F. Amin, "Continuous Arabic sign language recognition in user dependent mode", *J Intell Learn Syst Appl* 2, 2010, pp. 19–27.
- [10] Q. Munib, M. Habeeb, B. Tahruri and A.I-Malik, "American sign language (ASL) recognition based on Hough transform and neural networks", *Expert Systems with Applications*, 2007, vol.32, pp. 24-37.
- [11] M. Mohandes and M. Deriche, "Image Based Arabic Sign Language Recognition", *the 8TH International Symposium on Signal Processing and ITS Applications*, 2005, pp. 86-89.
- [12] S. Kumar, D. K. Kumar and A. Sharma, "Visual hand gesture classification using wavelet transforms, *International Journal of Wavelets, Multiresolution and Information Processing*, 2003, vol. 1, pp. 373-392.
- [13] E.E. Hemayed and A. S. Hassanien, "Edge-based recognizer for Arabic sign language alphabet (ArS2V -Arabic sign to voice)," *IEEE International Computer Engineering Conference*, 2011, pp.121-127.
- [14] A. K. Sarkaleh, F. Poorahangaryan, B. Zanj and A. Karami, "A Neural Network Based System for Persian Sign Language Recognition", *IEEE International Conference on Signal and Image Processing Applications*, 2009, pp.145-149.
- [15] A. Karami, B. Zanj, A. K. Sarkaleh, "Persian sign language (PSL) recognition using wavelet transform and neural networks." *Expert Systems with Applications* 38, 2011, pp. 2661–2667.
- [16] M.P Paulraj, S. Yaacob H. Desa and C.R. Hema, "Extraction of Head and Hand Gesture Features for Recognition of Sign Language", *International Conference on Electronic Design*, 2008, pp.1-6.
- [17] M.P Paulraj, S. Yaacob and M. Azalan, R. Palaniappan, "A Phoneme Based Sign Language Recognition System Using Skin Color Segmentation", *6th International Colloquium on Signal Processing & Its Applications*, 2010, No. 4, pp. 86-90.
- [18] I. Bahadori, Persian sign language collection for the deaf. Education and research office rehabilitation research group. Iran Welfare Organization, 1992.
- [19] V. Vezhnevets, V. Sazonov and A. Andreeva, "A survey on pixel-based skin color detection techniques, In *Proceedings of the GraphiCon 03*, 2003, pp. 85–92.
- [20] K.L. Tan, and L. F. Thiang, "Retrieving similar shapes effectively and efficiently," *Multimedia Tools and Applications*, Kluwer Academic Publishers, Netherlands. 2003, pp.111–134.

Bahare Jalilian, She is currently working toward the M.Sc. degree in the Islamic Azad University, Kermanshah Science and Research Branch. She obtained her B.S. in Computer Engineering from Razi University of Kermanshah, Iran in 2011. Her research interests include computer vision, digital image processing and pattern recognition.

Abdolah Chalecale, Born in Kermanshah, Iran, received his B.S. and M.Sc. degrees in Electrical Engineering (Hardware) and Computer Engineering (Software) from Sharif University of Technology, Tehran, Iran. He received his Ph.D. degree from Wollongong University, NSW, Australia in 2005 and currently is with Razi University, Kermanshah, Iran. His research interests include image processing, machine vision and human-machine interactions.