Static Hand Gesture Recognition of Persian Sign Numbers using Thinning Method

Alaa Barkoky
Image processing Lab., Dept. of electrical engineering & computer science
Tarbiat Modares University
Tehran, Iran
barkoky@modares.ac.ir

Nasrollah M. Charkari
Image processing Lab., Dept. of electrical engineering & computer science
Tarbiat Modares University
Tehran, Iran
charkari@modares.ac.ir

Abstract—Sign language is the most natural way of communication for the people with hearing problems. One of its most appealing applications is developing interfacing of human-machine interaction with more effective. A hand gesture recognition system can provide an opportunity for deaf persons to communicate with normal people without the need of an interpreter or Intermediate. In this article, we propose a method to recognize the image-based numbers of Persian sign language (PSL) using thinning method on segmented image. In this approach, after cleaning thinned image, the real endpoints have been used for recognition. The method is qualified to provide real-time recognition and is not affected by hand rotation and scaling. Experimental results on 300 images show that our approach recognition rate is 96.6 % as average.

Keywords-Persian Sign Language; Gesture Recognition; Segmentation; Static Hand Gesture Recognition; Thinning.

I. INTRODUCTION

Human beings use different gestures and body movements to convey meaning and communicate with other people in their daily lives. Examples of meaningful gestures include greeting hello or goodbye using the hands, or indicating numbers using fingers.

Gesture is one of the most natural and meaningful forms in HCI1. In other words, a gesture is a meaningful concept of motions done by human and is of supreme importance in designing an intelligent and efficient human–computer interface [1].

HCI refers to dealing and interaction between human and computer. It is a set of interactions between human users on one hand and intelligent agent software on the other hand. This relation can be established by different members of the human body, including: hand with different state, emotional faces, movement of the human body or any combination of body’s organs, such as hand and arm [1]. Today, gesture recognition, and especially hand gesture recognition, has different and considerable applications, such as; vision and robotic [1-2], computer games, communication deaf people with computer [3-7], and emotional recognition.

Sign language recognition is a research area involving pattern recognition, computer vision, natural language processing and psychology [3]. Today, there exist a significant number of deaf people in different countries. They usually use different movements to establish communication between each other; and usually the movements depend on human cultural issues; therefore, it is not possible to define a common international language. For example, the representation of numbers 6 till 10 is different from Arabic and American sign languages than it is in PSL as shown as “Fig. 1”.

According to the literature, there is no certain work on the PSL numbers in HCI area. Our goal in this article is to present an approach to recognize PSL image-based numbers using thinning method and to be not affected by gesture rotation and scaling; and to make it a basis to recognize video-based ones later. Finally, the result will be a meaningful data for machine which can be treated in HCI.

Figure 1. Hand gestures from left to right correspond to numbers from 6 to 10, where (a) Arabic, (b) American and (c) Persian sign language.

The paper is organized as follows: section 2 briefly describes the related works; the proposed hand segmentation method is presented in section 3; section 4 explains our proposed hand gesture recognition algorithm and finally, experimental results and conclusions are presented in section 5 and 6, respectively.
II. RELATED WORKS

The Hand gesture recognition research has been grown rapidly since 2005. Hand gesture recognition has two branches: static gesture (image-based); and the other, on dynamic gesture (video-based). Dynamic hand gestures come from frames added with time, so that they can describe complex information; so, dynamic hand gestures are not easy to recognize.

Some interesting researches have been done in this area. Control robotic is one of the most important applications. Alijanpour et al. [2], proposed a method based on hand center tracking and used correlation coefficient for state matching; for device control using inner distance feature; six gestures including of turn right, turn left, etc. has been defined on this work. Fang et al. [7] have proposed a real-time hand gesture recognition method for robot control by finding the blob and ridge of hand gesture. There are some other approaches has proposed in this field, such as, Fuzzy Decision Tree [10] and Hidden Markov Model [13] to extract feature vectors and finally recognize the hand gestures.

On the other hand, some other important researchers have worked on different deaf people languages processing, Al-Jarrah and Halawani [4] presented a method for automatic translation of gestures of the manual alphabets in the Arabic sign language. They have designed a collection of ANFIS networks, each of which is trained to recognize one gesture. Incertis et al. [6] introduced an approach for deaf-people interfacing using computer vision. The recognition of static alphabetic characters of Spanish Sign Language was addressed. The proposed approach combines a number of norms to evaluate the distance of the current sign, to the sign models stored in a database. This solution leads to a largely selective criterion. Rokade et al [3] used thinned method on one to ten numbers of American Sign Language (ASL). The feature vector consists of angle and center point to the vertical extended line. As the feature vector contains some corners, the method required vertical hand images as input. Ren and Zhang [8] presented MEB-SVM to classify gestures and finally achieve the proper recognition. Karami et al [17] proposed a system for recognizing static gestures of alphabets in PSL using Wavelet transform and neural networks. The discrete wavelet transform is applied on the gray scale images. Finally, the extracted features are used to train a Multi-Layered Perceptron neural network. Other approaches such as; local linear embedding, Neural Network shape fitting, object based key frame selection, and Haar wavelet representations have been presented in [9], [11-12] and [5].

III. HAND SEGMENTATION METHOD

The original goal of this step is to extract the hand gesture region. Pixels corresponding to the gesture are set to white and the background to black color. The hand segmentation method in our work is plotted in “Fig. 2”.

A. Hand Detection

This step starts with image capturing by the camera under natural light conditions. Regardless of image size, is rescaled that considers an appropriate processing time and quality at the same time. The size which we have adopted is $150 \times 224$ pixels. We recall that input image maybe one or two hands. One to five numbers is presented by one hand and six to ten by two hands in PSL.

Several color spaces and color-based approaches have been proposing in the literature for skin detection applications. RGB is the most commonly used color space for storing and representing digital images. A pixel $(x, y)$ belongs to a skin if its $(R, G, B)$ component satisfies the following conditions [16]:

$$\begin{align*}
R > 95 & \quad \text{and} \quad G > 40 \quad \text{and} \quad B > 20 \\
\text{Max} \{R, G, B\} - \min \{R, G, B\} > 15 & \quad \text{and} \quad R - G > 15 \quad \text{and} \quad R > G \quad \text{and} \quad R > B
\end{align*}$$

(1)

On the other hand we know that the orthogonal color spaces like YCbCr, YIQ, YUV and YIS, reduce the redundancy present in RGB color channels and represent the color with statistically independent components [15]. As the luminance and chrominance components are explicitly separated, these spaces are favorable choice for skin detection [15]. The YCbCr space represents color as luminance $(Y)$ and chrominance $(Cb$ and $Cr)$ computed as a weighted sum of RGB values [14], where $Y$, $Cb$, and $Cr$ components and values to extract human skin [14] are found by the relations “2” and “3” respectively;

$$\begin{align*}
Y &= 0.299R + 0.587G + 0.114B \\
Cr &= 128 + 0.5R - 0.418G - 0.081B \\
Cb &= 128 - 0.168R - 0.331G + 0.5B
\end{align*}$$

(2)

$$85 < Cb < 135 \quad \text{and} \quad 135 < Cr < 180 \quad \text{and} \quad Y > 80$$

(3)

![Figure 2. Steps of hand segmentation method.](Image)

![Figure 3. (a) Original image, (b) YCbCr color space result, (c) RGB color space result and (d) result of combine both two color spaces.](Image)
Considering skin color differences between humans, we have suggested to merge "use OR operation" RGB color space result with YCbCr result to extract skin area and get better segmentation results. The previous figure "Fig.3" shows some results of both color spaces and combine them.

B. Reconstruction of hand area
In this step, filtering and morphology operations are performed to decrease noises and segmentation errors. A 5×5 median filter and morphology operation (close) that contains two steps, dilation, and erosion is applied to do that. "Fig.4" shows the filtering and morphology results.

C. Wrist-cropping
Our goal at this stage is separate the hand from arm, by a wrist cropping. Wrist cropping is essential, regardless of user shirt type; long or sleeves one and the segmented hand image includes the arm or not. The algorithm that we utilize is proposed in [18], and is based on sudden increase in the hand width from lower arm to the hand. "Fig. 5" shows the results.

IV. HAND GESTURE RECOGNITION ALGORITHM
In this stage and by getting segmented image as an input, we use the following algorithm that is shown in "Fig.6" to recognize the hand gesture.

A. Thinning and cleaning of segmented image
The first step is applying thinning method on input image. Thinning causes objects to change to lines. It removes pixels so that an object shrinks to a minimally connected stroke. Then, the end and joint points are found. According to 8-connectivity neighbors’ pixels, the endpoint is a point that contains only one connection and represents the terminal pixel of the thin segment. Joint point is a point on the thin segment that contains more than two 8-connectivity neighbors. Joint point is the meeting point of two or more thin segments. "Fig.7" shows the mentioned points.

It is important to note that, sometimes we observe additional thin segments (non-finger thin) resulting from topography of hand or sometimes from segmentation errors. To cope with this problem, we use the following steps:
- Step 1: The length of thin segments is calculated.
- Step 2: Thin segments are sorted by length. The longest one is presented with [A].
- Step 3: The thin segments lengths are compared with [A] and those that less than 40% of [A] are removed. 40% is defined as the ratio between the thumbnail and pinkie that it is normally more than 60%. We have used this ratio to include some special cases due to different hand sizes from one person to another. "Fig.8" shows the unclean and cleaned image.

B. Recognition of gesture
In the last step, the hand gesture will be recognized by the number of remaining endpoints in cleaned image properly. The following rule used to recognize and distinguish ten different classes:

\[
\text{If } \text{No. of endpoints in cleaned image is } \alpha \quad \text{Then } \text{the gesture number is } \alpha; \quad (4)
\]

For example if the number of remaining endpoints is five then the gesture will recognized as the number five. It is necessary to mention that, the lower section of the hand is common among almost all gestures and does not contain any useful information to distinguish one gesture from another. Therefore, we consider only the upper part of image to count endpoints, and the lowest joint point will be the separation point.

V. EXPERIMENTAL RESULTS
The proposed system was tested with 300 images, 30 samples for each number used in dataset. Considering that there is no certain dataset on PSL numbers, the dataset is
made by us and we’ve tried to make it contain different sample models such as rescaled and rotated samples between +45 and -45 degree. “Fig. 9” shows three different samples of number three that recognized correctly by our system. In general; the segmentation errors and recognition rate for different gestures are given in “Table 1”.

![Figure 9: Examples of (a) Rotated, (b) normal, and (c) rescaled gestures.](image)

Table 1. Experimental Results for Different Numbers.

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Segmentation Error (%)</th>
<th>Recognition Rate (%)</th>
<th>Gesture</th>
<th>Segmentation Error (%)</th>
<th>Recognition Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>4.9</td>
<td>96.6</td>
<td>Six</td>
<td>2.3</td>
<td>96.6</td>
</tr>
<tr>
<td>Two</td>
<td>4.1</td>
<td>96.6</td>
<td>Seven</td>
<td>2.2</td>
<td>96.6</td>
</tr>
<tr>
<td>Three</td>
<td>2.6</td>
<td>93.3</td>
<td>Eight</td>
<td>1.2</td>
<td>93.3</td>
</tr>
<tr>
<td>Four</td>
<td>1.3</td>
<td>96.6</td>
<td>Nine</td>
<td>0.9</td>
<td>96.6</td>
</tr>
<tr>
<td>Five</td>
<td>0.6</td>
<td>100</td>
<td>Ten</td>
<td>0.4</td>
<td>100</td>
</tr>
<tr>
<td><strong>Average Recognition Rate</strong></td>
<td><strong>96.62</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The average recognition rate is 96.62% which is a good result considering the diversity of data in dataset. Errors may result from the variety of factors, such as changing illumination, background clutter, and unclear signs that lead to error in segmentation and gesture recognition later.

According to what we noted in the introduction that there is no certain work on the PSL numbers in HCI area, and in addition; there is difference between sign gestures of PSL and other sign languages, it is difficult to compare the results of our work to other works results accurately. “Fig. 10” shows different classes of gestures numbers.

![Figure 10: Gestures of numbers 1 to 10 of Persian Sign Language.](image)

VI. CONCLUSION

In this paper, we have applied a simple and real-time method, which works suitably for recognize image-based numbers of PSL. In our algorithm, we utilized skin color information in segmentation phase, thinning method for feature extraction and areal endpoints in cleaned image for recognition. Dataset contained rescaled and rotated samples, which is considered a good advantage for our method.

The method is the base for our next work (video-based gestures). In addition; we can develop the algorithm required, to include some symbol gestures such as turn right, turn left, stop, that can be used for robotic control.

ACKNOWLEDGMENT

This research was supported by ITRC, Iran Telecommunication Research Center. We also thank Messrs. Chihada and Yousef for their assistance.

REFERENCES