Hand Gesture Recognition using Neural Networks and Moment Invariants

Nizar Saadi Dahir
Assistant Lecturer - Computing and IT Center - University of Kufa – IRAQ

ABSTRACT

Hand gesture recognition is a rich area of research and covers a wide scope of applications from human machine interaction (e.g in games) to Deaf people computer interface and from 3D animation to control of mechanical systems. In this work a new algorithm for hand gesture recognition is proposed and evaluated. This algorithm uses moment invariants for feature extraction and neural networks for classification. To evaluate the algorithm a subset from the ASL (American Sign Language) is used, this subset consists of the 26 American one-handed sign language alphabet as a training set for the system. The system then tested using images that are different from the training set in size, orientation and position and the results for these tests are presented and discussed.

1. Introduction

Several studies on gesture and sign language recognition have been published. These publications can be separated into three categories according to the signs they try to recognize.

1. In the first category, researchers propose methods to recognize static hand postures or the sign language alphabet [1-4]. They use images of the hands and extract feature vectors according to the static information of the hand shape.

2. The researchers in the second category [5, 6] collect sequential feature vectors of the gestures and, using the dynamic information, recognize letters with local movement, too. In these approaches, only movement due to changing hand postures is regarded, while path movement is ignored (movement made primarily with the shoulder or elbow).

3. The third category of researchers try to recognize sign language words and sentences [7-10]. In addition to local movement of the hands, signing includes also path movement of the hands. Therefore, most systems employ segmentation and tracking of the hands.

Also deaf people need to communicate with hearing people in everyday life. To facilitate this communication, systems that translate sign language into spoken language could be helpful. The recognition of the hand signs is the first step in these systems.

Other category of researches focused on the use of hand gesture recognition for human machine interaction to replace computer interfacing devices like mouse and pad in order to adapt computers to our body language [11] and such systems are proposed for computer games interface and for industrial control applications.

Practical hand gesture recognition systems in general must perform the following tasks

1. **Acquisition**: a frame is captured using digital camera.
2. **Segmentation**: in this stage each frame is processed separately before its analysis. The hand must be separated form the background. And this is usually done using the techniques used for skin recognition and segmentation through the modeling of the skin colour through the selection of an appropriate colour space and identifying the cluster associated with skin colour in this space for example using the YCbCr or the HSI colour spaces.
3. **Pattern Recognition**: once the user's hand has been segmented, features must be extracted from the segmented hand's gesture and these features are classified to the best match in a database of previously stored features for the different hand gestures that the system is required to recognize. These features must be transformation (translation, resize and orientation) invariant and they can be computed from the silhouette or the boundary of the object.
4. **Executing Action**: finally, the system carries out the corresponding action according to the recognized hand gesture. For example in sign alphabet recognition the system must type of speak or type the letter that corresponds to the detected gesture.

This work focuses on the third task, the feature extraction and classification. An algorithm that uses moment invariants for feature extraction and feed forward backpropagation neural network with adaptable learning rate for classification.

Moment invariants [12] used for many applications [13 ,14] and proven an acceptable invariance against translation, rotation and resizing. The purpose of this work is to evaluate their performance for recognizing hand gestures using Neural Networks as classifier.

2. The Proposed Hand Gesture Recognition Algorithm

The proposed hand gesture recognition algorithm consists of four main tasks as shown in figure 1:

(1) Creation of the library for gesture sign alphabet,

(2) Skin Segmentation

(2) Neural Network Training, and

(3) Testing and Evaluation.

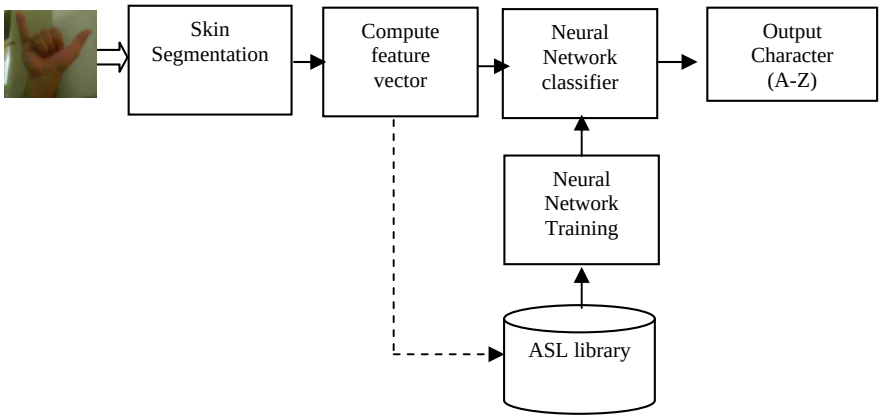


Figure 1: Hand gesture recognition algorithm block diagram.

2.1 The ASL sign alphabet library

American Sign Language is the language of choice for most deaf people. It is part of the "deaf culture" and includes its own system of puns, inside jokes, etc. ASL also has its own grammar that is different from English.

ASL consists of approximately 6000 gestures of common words with finger spelling used to communicate obscure words or proper nouns. Finger spelling uses one hand and 26 gestures to communicate the 26 letters of the alphabet. Some of the signs can be seen in Fig(2) below.

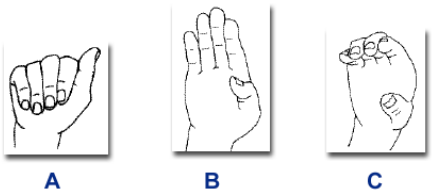
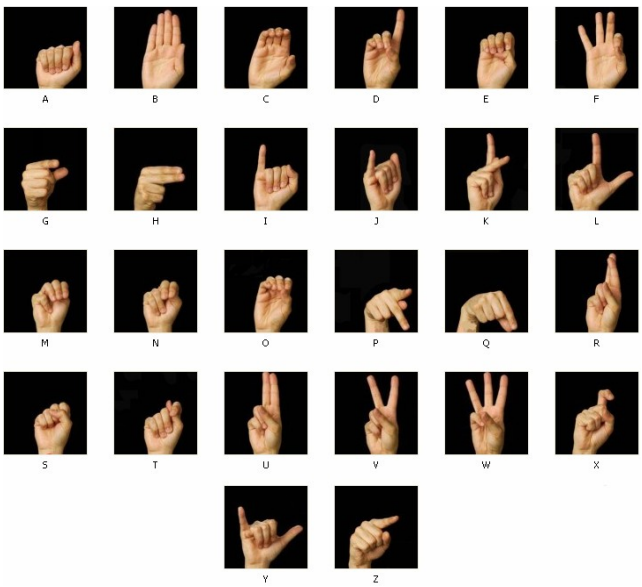


Figure 2: ASL examples

In this work I used the 26 finger spelling gestures to evaluated the proposed algorithm this set is shown in figure 3



2.2 Moment invariants

Let *F(x,* y) denote an image in the two-dimensional spatial domain. Geometric moment of order p+q is denoted as .

$$m_{p,q} = \sum_x \sum_y x^p y^q F(x,y) \qquad \ldots(1)$$

From eq (1) we can say that moment describe the shape of the object, take for example the first 5 moments (0 to 4) they can be described as follows:

- Moments of "00" describe summation of pixels in the whole shape
- Moments of 10, 01 describe the distribution of pixels along axis x, y
- The second moment is the variance,
- Skewness is the third moment that is a measure of the lopsidedness of the distribution; any symmetric distribution will have a third central moment, if defined, of zero.
- Kurtosis is The fourth moment which is a measure of whether the distribution is tall and skinny or short and squat, compared to the normal distribution of the same variance. Since it is the expectation of a fourth power,

For p,q=0,1,2 … The central moments are expressed as

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q F(x,y) \ldots(2)$$

Where $x_c = m_{1,0}/m_{0,0}$, $y_c = m_{0,1}/m_{0,0}$ and (x_c, y_c) is called the center of the region or object. The normalized central moments denoted $\eta_{p,q}$ are defined as.

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^{\lambda}} \qquad \ldots(3)$$

Where $\lambda = (p+q+2)/2$

For $p+q = 2,3,\ldots$ A set of seven transformation invariant moments can be derived from the second and third-order moments as follows [12]:

$\phi_1 = \eta_{2,0} + \eta_{0,2}$

$\phi_2 = (\eta_{2,0} - \eta_{0,2})^2 + 4\eta^2_{1,1}$

$\phi_3 = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta 03)^2;$

$\phi_4 = (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{2,1} + \eta_{0,3})^2;$

$\phi_5 = (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3})[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2];$

$\phi_6 = (\eta_{2,0} - \eta_{0,2})[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} + \eta_{0,3});$

$\phi_7 = (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{1,2} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3})[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2];$

These 7 feature are used to compute transformation invariant feature vector.

2.3 Skin Segmentation Algorithm

The feature vector is computed from a binary segmented image, to obtain this image we need a skin segmentation algorithm. Skin color pixel classification is used in this work. The aim of skin color pixel classification is to determine if a color pixel is a skin color or nonskin color. Good skin color pixel classification should provide coverage of all different skin types (blackish, yellowish, brownish, whitish, etc.) and cater for as many different lighting conditions as possible. In some cases, color classification is done using only pixel chrominance because it is expected that skin segmentation may become more robust to lighting variations if pixel luminance is discarded. Three representative color spaces which are commonly used in the image processing field are:

. **RGB:** Colors are specified in terms of the three primary colors: red (R), green (G), and blue (B).

. **HSV:** Colors are specified in terms of hue (H), saturation (S), and intensity value (V) which are the three attributes that are perceived about color. The transformation between HSV and RGB is nonlinear. Other similar color spaces are HIS,HLS, and HCI.

. **YCbCr:** Colors are specified in terms of luminance (the Y channel) and chrominance (Cb and Cr channels).
The transformation between YCbCr and RGB is linear. Other similar color spaces include YIQ and YUV.

Working with YCbCr color space in this work it's found that the ranges of Cb and Cr most representative for the skin color reference map are [14]:

$$77 \le Cb \le 127 \quad \text{and} \quad 133 \le Cr \le 173 \qquad \ldots(5)$$

The results of applying this techniques on some hand gestures are shown in Fig. 4.

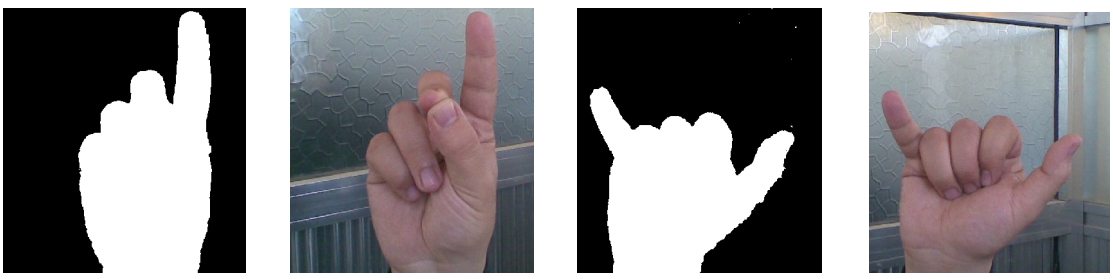


Fig 4 Skin segmentation results

2.4 Neural Network Structure

A feed-forward backpropagation neural network (NN)) with adaptable learning rate was used in the classification face. The NN have 3 layer; an input layer (7 neuron), a hidden layer (50 neuron) , and output layer (26 neuron).

The activation function used is the tan sigmoid function, for both the hidden and the output layer. The input to the neural network is the feature vector containing 7 component these are the 7 moment invariants, the NN has 26 output (A-Z) as shown in figure 5.

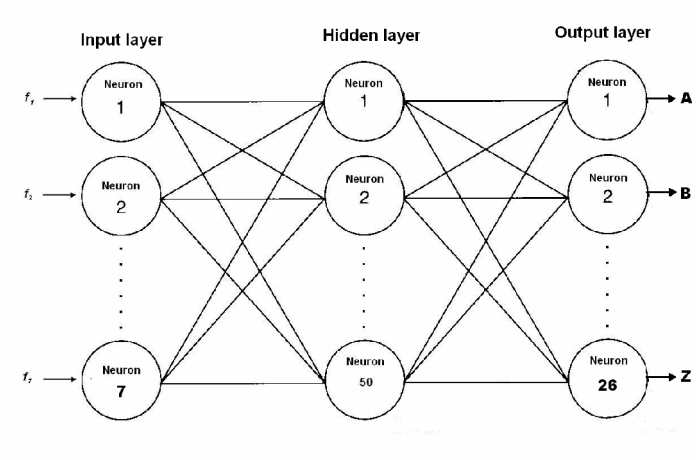


Figure 5 the Neural Network structure

3. Results and Discussion

To evaluate the hand gesture recognition algorithm the neural network was trained using the 26 gesture images shown in Fig. 3.

The training was done with using a SSE of 10e-5 as a goal the goal was reached in 51,328 training iteration. The NN performance is shown in figure 6.

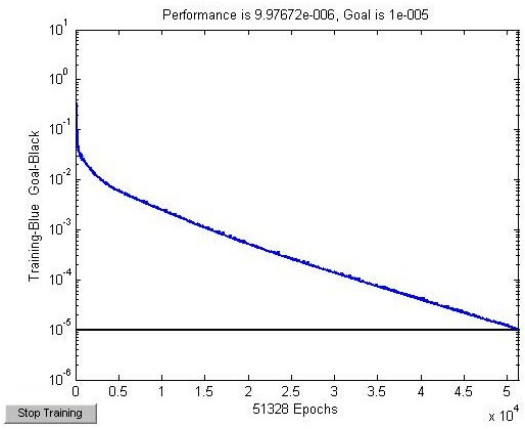


Figure 6 The Neural Networks Performance

To evaluate the system a testing set was generated by applying a spatial transformation on gesture images these transformations are translating, scaling, rotating, or applying the three transformations as shown in figure 7 bellow.

The criteria for evaluating the algorithm was the Classification accuracy (C.A.) which is given by:-

$$\text{C.A.} = \frac{\text{No. of correct classifications}}{\text{Total No. of classifications}} X 100\% \quad \ldots(5)$$

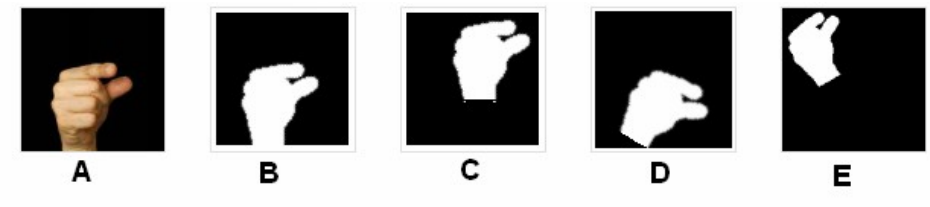


Figure 7 Generating the testing set a) Original
b) Binary image c) Translated
d) Rotated e) Translated, rotated, and scaled

The algorithm showed very high classification accuracy on images with translation and scaling, moderate with rotation and acceptable when all the transformations are applied to images. The results are shown in Table 1

Table 1 The system results

Image Transformation	C.A.
Translation	96.153%
Scaling	92.307%
Rotation (30°)	80..770%
Rotation, Scaling & translation	76.920%

From the results we can deduce that the use of moment invariants in general gave very encouraging results and that even applying all the transformations the system still giving acceptable results.

In this system the moment invariants was used. Moments invariants depend only on the object silhouette, and it ignores the internal details (e.g. the edges of the fingers in this case).

In the future work I will try to take the internal details into consideration by using the moments of the edges or by using different feature extraction strategy like fourier descriptors and a comparison with the current method will be made.

References

[1]J. Triesch and C. von der Malsburg. "A System for Person-Independent Hand Posture Recognition against Complex Backgrounds" IEEE Trans. Pattern Analysis and Machine Intelligence, 23(12):1449-1453, December 2001.

[2]H. Birk, T.B. Moeslund, and C.B. Madsen. "Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis." In 10th Scandinavian Conference on Image Analysis, Laeenranta, Finland, June 1997.

S. Malassiottis, N. Aifanti, and M.G. Strintzis. A Gesture Recognition System Using 3D Data. In Proceedings IEEE 1st International Symposium on

[3] Processing, Visualization, and Transmission, pp. 190{193, Padova, Italy, June 2002.

[4]S.A. Mehdi and Y.N. Khan. Sign Language Recognition Using Sensor Gloves. In Proceedings of the 9th International Conference on Neural Information Processing, volume 5, pp. 2204{2206, Singapore, November 2002.

[5]K. Abe, H. Saito, and S. Ozawa. Virtual 3-D Interface System via Hand Motion Recognition From Two Cameras. IEEE Trans. Systems, Man, and Cybernetics, 32(4):536{540, July 2002.

[6]J.L. Hernandez-Rebollar, R.W. Lindeman, and N. Kyriakopoulos. A Multi-Class Pattern Recognition System for Practical Finger Spelling Translation. In Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, pp. 185{190, Pittsburgh, PA, October 2002.

[7]Y. Nam and K. Wohn. Recognition of Space-Time Hand-Gestures Using Hidden Markov Model. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, pp. 51{58, Hong Kong, July 1996.

[8]B. Bauer, H. Hienz, and K.F. Kraiss. Video-Based Continuous Sign Language Recognition Using Statistical Methods. In Proceedings of the International Con- ference on Pattern Recognition, pp. 463{466, Barcelona, Spain, September 2000.

T. Starner, J. Weaver, and A. Pentland.

[9]3D Data Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video. IEEE Trans. Pat- tern Analysis and Machine Intelligence, 20(12):1371-1375, December 1998.

[10] C. Vogler and D. Metaxas. Adapting Hidden Markov Models for ASL Recognition by Using Three-dimensional Computer Vision Methods. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp. 156-161. Orlando, FL, October 1997.

[11] Elena S?nchez-Nielsen et.al "Hand Gesture Recognition for Human-Machine Interaction" Journal of WSCG, Vol.12, No.1-3, ISSN 1213-6972 WSCG'2004 Plzen, Czech Republic.

[12] M. K. Hu, "Visual pattern recognition by moment invariants," in *Computer Methods in Image Analysis*, IEEE Computer Society, Los Angeles, 1977.

Francesca Gasparini et. all "Skin Segmentation Using Multiple Thresholding" Universita Degli Studi Di Milano-Bicocca, Via Bicocca degli Rrcimboldi 8, 20216 Milano, Italy, 2004.

المستخلص

تعتبر عملية تمييز إشارات اليد مساحة غنية للبحث وتغطي مجالا واسعا من التطبيقات ذات العلاقة بالتفاعل بين الحاسبة والإنسان (الألعاب مثلا) للأشخاص الصم وتمثل التواصل باستخدام حركات ثلاثية الإبعاد للسيطرة على النظم الميكانيكية .

هذا البحث يعرض ويقيم خوارزمية جديدة لتمييز إشارات اليد هذه الخوارزمية تستخدم ثوابت لحظية لاستخراج الخواص باستخدام الشبكات العصبية لغرض التصنيف.

تم استخدام مجموعة جزئية أخرى من لغة ASL (لغة الإشارة الأمريكية)، وهذه المجموعة من الرموز والتي عددها 26 رمزا من الحروف الهجائية لليد الواحدة من لغة الإشارة الأمريكية كمجموعة تدريبية للنظام.

تم فحص النظام باستخدام مجموعة صور مختلفة عن مجموعة الإشارة التي استخدمت لتدريب المنظومة والتي تختلف في الحجم والاتجاه والموقع، كما تم أيضا عرض ومناقشة النتائج.