VOL. 3, NO. 8 August, 2012 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

Development of a New Arabic Sign Language Recognition Using K-Nearest Neighbor Algorithm

[1] Prof.reyadh Naoum, [2] Dr.Hussein H. Owaied, [3] Shaimaa Joudeh
[1] Dean of Faculty of Information Technologies - Middle East University
[2] Associated prof - Faculty of Information Technologies - Middle East University
[1] reyadh_naoum@yahoo.com, [2] owaied@yahoo.com, [3] shm-joudeh@hotmail.com

ABSTRACT

This paper presents a new Arabic sign language recognition using K-nearest Neighbor algorithm. The algorithm is designed to work as a first level detection upon a series of steps to bring the captured character images into actual spelling. The algorithm acts in a high performance execution which is exactly needed for such type of systems. K-Nearest Neighbor Algorithm and feature extraction are the guidelines of the recognition system, because hand gestures is treated as a block of curves needed to be extracted in the best fit with a predefined character set in the knowledge base. The specific image preprocessing to form a new idea of histogram and a histogram transition table is formed as a hashed string of transformation of block histogram sequence using K-Nearest Neighbor Algorithm. Preparing the knowledge base as a sequence of characters for one time and will and fast easily compared to detecting the character input.

Keywords: *Arabic Sign Language, Artificial Neural Network, Image Processing, OCR, Sign Language Processing, Text-based Image Processing.*

1. INTRODUCTION

Sign language is a highly visual-special linguistically complete natural language. It is typically the first language of communication for deaf people. However, the signers still have serious problems of communicating with speaking persons, who are not sign users [1]. The communication difficulty adversely affects the life and interpersonal relationships in the deaf community. Deaf individual communicate with speaking people usually via interpreters or text writing [2]. Although interpreters can help the communication between deaf and hearing persons, they are often expensive and have negative effect on independency and privacy. Note writing is used by many deaf people to communicate with someone who is seated nearby, but it is awkward while walking, standing at a distance, and when more than two persons are in a conversation [3].

Artificial neural networks have been widely used in sign language recognition research. Murakami and Taguchi investigated the use of recurrent neural nets for Japanese Sign Language recognition. Although it achieved a high accuracy of 96%, their system was limited only to 10 distinct signs. Kramer and Leifer developed an ASL finger spelling system using a Cyber glove, with the use of neural networks for data segmentation, feature classifier, and sign recognition. Using a tree-structured neural classifying vector quantizes, a large neural network with 51 nodes was developed for the recognition of ASL alphabets. They claim a recognition accuracy of 98.9% for the system [4].

In some interactive computer system oriented applications, it is required to track the position or orientation of a hand that is prominent in the image. Relevant applications might be computer games, or interactive machine control. In such cases, a description of the overall properties of the image may be adequate. Image moments, which are fast to compute, provide a very coarse summary of global averages of orientation and position. If the hand is on a uniform background, this method can distinguish hand positions and simple pointing gestures [5].

2. METHODS FOR ARABIC SIGN LANGUAGE RECOGNITION

There are many methods have been used for sign language processing in the following subsection some of those methods are used in sign language processing and specially in the phase of recognition [6].

2.1 Large Object Tracking

The large-object-tracking method makes use of a low-cost detector/processor to quickly calculate moments. This is called the artificial retina chip. This chip combines image detection with some low-level image processing (named artificial retina by analogy with those combined abilities of the human retina). The chip can compute various functions useful in the fast algorithms for interactive graphics applications

2.2 Shape Recognition

Most applications, such as recognizing particular static hand signal, require a richer description of the shape of the input object than image moments provide. If the hand signals fell in a predetermined set, and the camera views a close-up of the hand, we may use an example-based approach, combined with a simple method top analyze hand signals called orientation histograms.

These example-based applications involve two phases; training and running. In the training phase, the user shows the system one or more examples of a specific hand shape. The computer forms and stores the corresponding orientation histograms. In the run phase, the computer compares the orientation histogram of the current image with each of the stored templates and selects the category of the closest match, or interpolates between templates, as appropriate. This method should be robust against small differences in the size of the hand but probably would be sensitive to changes in hand orientation.

2.3 Glove based Analysis

Glove based analysis, employs sensors (mechanical or optical) attached to a glove that transducers finger flexions into electrical signals for determining the hand posture. Glove based analysis uses a hardware equipment which uses the electricity or electromagnetic interference to gain information about the hand state which will be enough to provide a vision of a hand gesture. The relative position of the hand is determined by an additional sensor. This sensor is normally a magnetic or an acoustic sensor attached to the glove. For some data glove applications, look-up table software toolkits are provided with the glove to be used for hand posture recognition.

In the project Glove-Talk II, Fells and Hinton used three neural networks and several input devices to translate hand gestures to speech. One neural network was used for the vowel/consonant decider, and two others were used for the individual vowel selector and the consonant selector. Their system is very extensive. However, the training time is as long as over 100 hours before the system is able to perform intelligibly. Waldron and Kim used neural networks to recognize 14 ASL signs using different networks for hand shapes and for hand orientation and position. The limited sign vocabulary was divided into a standard set of motions, which was recognized by another network. The overall accuracy of this system was 86%. Although the use of neural networks can provide reliable recognition of hand shapes and a limited sign vocabulary, it is not a feasible method in the cases of a large sign vocabulary and recognition at the sentence level.

Full ASL recognition systems (words, phrases) incorporate data gloves. Takashi and Kushiro discuss a Data glove-based system that could recognize 34 of the 46 Japanese gestures (user dependent) using a joint angle and hand orientation coding technique. From their paper, it seems the test user made each of the 46 gestures 10 times to provide data for principle component and cluster analysis. A separate test was created from five iterations of the alphabet by the user, with each gesture well separated in time [7]. While these systems are technically interesting, they suffer from a lack of training. Excellent work has been done in support of machine sign language recognition by Spelling and Parish, who have done careful studies on the bandwidth necessary for a sign conversation using spatially and temporally sub-sampled images. Point light experiments (where “lights” are attached to significant locations on the body and just these points are used for recognition), have been carried out by Poizner. Most systems to date study isolate/static gestures. In most of the cases those are finger spelling signs [8].

2.4 Vision Based Analysis

Vision based analysis, is based on the way human beings perceive information about their surroundings, yet it is probably the most difficult to implement in a satisfactory way. Several different approaches have been tested so far. One is to build a three-dimensional model of the human hand. The model is matched to images of the hand by one or more cameras, and parameters corresponding to palm orientation and joint angles are estimated. These parameters are then used to perform gesture classification. A hand gesture analysis system based on a three-dimensional hand skeleton model with 27 degrees of freedom was developed by Lee and Kunii. They incorporated five major constraints based on the human hand kinematics to reduce the model parameter space search. To simplify the model matching, specially marked gloves were used. [9]

Analysis of drawing gestures, usually involves the use of a stylus as an input device. Analysis of drawing gestures can also lead to recognition of written text. The vast majority of hand gesture recognition work has used mechanical sensing, most often for direct manipulation of a virtual environment and occasionally for symbolic communication. Sensing the hand posture mechanically has a range of problems, however, including reliability, accuracy and electromagnetic noise. Visual sensing has the potential to make gestural interaction more practical, but potentially embodies some of the most difficult problems in machine vision. The hand is a non-rigid object and even worse self-occlusion is very usual.

Automatic capture and analysis of human motion is a highly active research area due both to the number of potential applications and its inherent complexity. The research area contains a number of hard and often ill-posed problems such as inferring the pose and motion of a highly articulated and self-occluding non-rigid 3D object from images. This complexity makes the research area challenging from a purely academic point of view. From an application perspective computer vision-based methods often provide the only non-invasive solution making it very attractive [10].

3. THE PROPOSED METHOD

The proposed method is a hybrid method of using K-Nearest Neighbors (K-NN) and Artificial Neural Network. The K-Nearest Neighbors (K-NN) algorithm is a nonparametric method in that no parameters are estimated as, for example, in the multiple linear regression models. Instead, the proximity of neighboring input (x) observations in the training data set and their corresponding output values (y) are used to predict (score) the output values of cases in the validation data set. These output variables can either be interval variables in which case the K-NN algorithm is used for prediction while if the output variables are categorical, either nominal or ordinal, the K-NN algorithm is used for classification purposes.

In pattern recognition, the k-nearest neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. K-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification.

The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor.

The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its k nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. (A common weighting scheme is to give each neighbor a weight of 1/d, where d is the distance to the neighbor. This scheme is a generalization of linear interpolation.)

The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The k-nearest neighbor algorithm is sensitive to the local structure of the data.

Nearest neighbor rules in effect compute the decision boundary in an implicit manner. It is also possible to compute the decision boundary itself explicitly, and to do so in an efficient manner so that the computational complexity is a function of the boundary complexity.

As a pattern recognition technique, sign language must be implemented against one or more of programming patterns and techniques, in such problem a sequence of operations is needed to transfer the gained images from the raw image format to form the character gained from the questioned image, those operations are listed below.

a. Image Questioning and capturing.
b. Image Clipping and narrowing.
c. Image Masking.
d. Image Histogram Generation.
e. Image Histogram Comparison.

These representations will be suitable for this kind of projects; as a result the system will implement each character as a sequence of characters called the Fix Increase Decrease (FID) Rule produced by implementing character histogram based on the black color distributions over a set of discrete values.

After implementing this histogram function this histogram will be compared with each character gained from webcam a questioning engine, after that the character will be retrieved to be appended to form the whole word the deaf person need to express. As a result this system should enhance the ability for deaf people to be able to act as a normal person has the ability to interact with other people. In the following subsections are the details descriptions.

3.1 Image Questioning and Capturing

This I the process for claiming the image as a ping frame from a questioning device, for example here the webcam is the questioning device for getting the image frame, by getting the frame this frame will be in a raw format which is not in the correct format and size to be processed by the next processing steps, so it is converted by a specific API wrapper that has the ability to interact with windows APIs to get the image in the specific suitable format.

3.2 Image Clipping and Narrowing

Image clipping is very important step in the preprocessing phase, because there is the need to compare the histogram in a proportional way, this is done by comparing a histogram express the block to the a questioned image, to perform this goal there is a need to clip the image, by doing that the image is converted to black/white representation.

After doing so; the image should be narrowed by using four lines approaching the black pixels, by doing this the image will express the clipping area just, by doing this we are able to compare the image whatever the scale of the image.

3.3 Image Masking

Image masking is the process for converting the image into black histogram, this process is necessary for generating the histogram by taking a phase of 5 pixels width. This process is done by counting the pixels numbers in each phase generating the next idea which is the next step for generating the histograms.

3.4 Image Histogram Generation

This step is responsible for generating the histogram to be compared later by the next process to generate a character; each character in any language is expressed using a knowledge base histogram, by using the dot counting technique the phase length will be generated and after that there will be a suitable way to draw a character histogram. Two histograms is needed for each character to express the horizontal and vertical behaviors, this is the way to detect the character surface behavior using the K-Nearest Neighbor algorithm.

3.5 Image Histogram Comparison

This is the way for detecting the character by using class classification and detection according to the stored histogram character representation among several representations by comparing a set of classification to get the vest intersecting class to get the best fit character. Using this

VOL. 3, NO. 8 August, 2012 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

algorithm the character will be claimed in character format and then be detected in the interface which is the main aim for this paper.

4. EVALUATION

Two cases will be considered for testing with this system, which are Hits and colored Glove.

1. **Hit Evaluation**: these criteria to be evaluated all the processes will be how many hits the system performs and how many the system failure, is called misses, as seen in table 1 and figure 1.

Table 1: Representing Character hit/miss percentage according to colored glove

Colored Glove	Hit Rate
Nacked Hand	50%
Red Hand	75%
Black Hand	65%
White Hand	80%

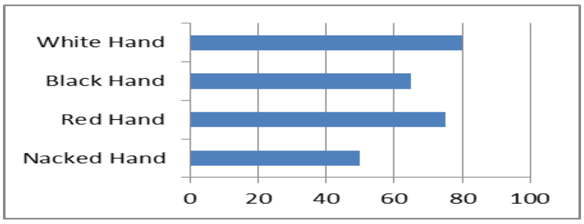
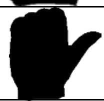
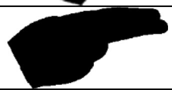


Fig 1: Graph Representing Character hit/miss percentage according to colored glove

2. **Glove Color:** because the user of this system should wear a glove to be extracted by the camera, the testing and the application for the Arabic characters set shown in table 2 and figure 2.

Table 2: Containing Arabic Characters Evaluation

Character	Image	Hit Rate %
ا		70
ب		65
ت		65
ث		70
ج		50
ح		50
خ		50
د		90
ذ		90
ر		65
ز		65
س		50
ش		95
ص		50
ض		80
ط		90
ظ		90
ع		85
غ		85
ف		50
ق		50

VOL. 3, NO. 8 August, 2012 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

ك		50
ل		90
م		90
ن		90
ه		50
و		80
ي		95

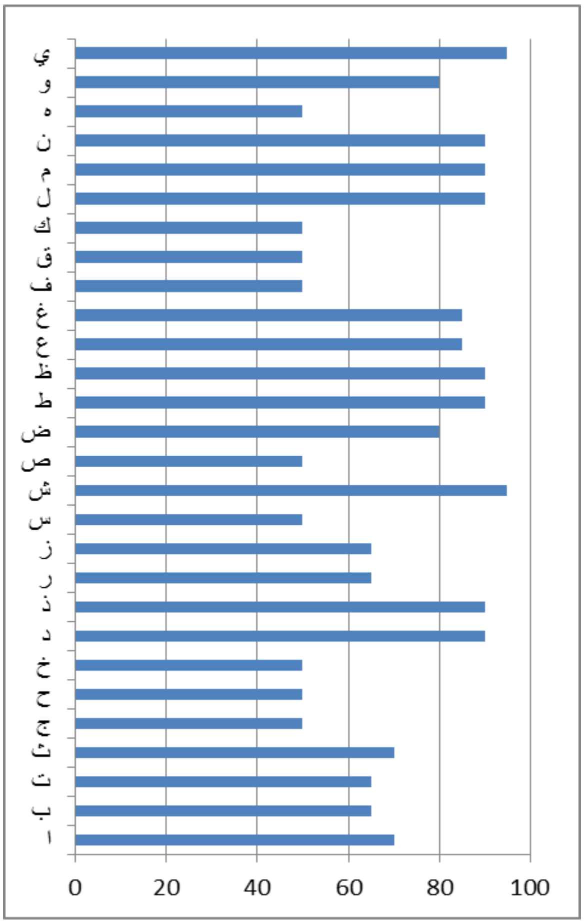


Fig 2: Graph representing the hit/miss rate for each character in Arabic language

5. CONCLUSION

The proposed algorithm can be combined with a glove based analysis technique to form a very high performance sign language recognition technique. Therefore may be used as an online sign language recognition tool for Sign Language Recognition in general. This project can be enhanced by adding a self-learning engine to the engine by combining many images (Histograms) and apply selection criteria for discarding less rate configuration for each character. Also will enhance the algorithms needed for sign language recognitions such as neural network and fuzzy logic and other techniques. This project will enhance the ability for deaf people to act as normal person, it will be better to invent a new technique that helps those people to understand the English words as a sign language which is called “ASL Synthesis”.

REFERENCES

[1] Vaishali S. Kulkarni and S.D.Lokhande, 2010, Appearance Based Recognition of American Sign Language Using Gesture Segmentation, International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010, 560-565.

[2] HONGGANG WANG, MING C. LEU AND CEMIL OZ, 2006, American Sign Language Recognition Using Multi-dimensional Hidden Markov Models, JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 22, 1109-1123 (2006).

[3] M. AL-Rousan, K. Assaleh, and A. Tala’a, "Video-based Signer-independent Arabic Sign Language Recognition Using Hidden Markov Models", Applied Soft Computing, 9(3), pp. 990-999, 2009.

[4] O. Al-Jarrah, and A. Halawani, "Recognition of Gestures in Arabic Sign Language Using Neuro-Fuzzy Systems", 2006, Past and Current Trends in Sign Language Research”, Language & Communication, 26(2), pp. 168-192, 2006.

[5] E. Keogh, 2002, “Exact indexing of dynamic time warping,” in International Conference on Very Large Data Bases, 2002, pp. 406–417.

[6] Xiang Ma, Dan Schonfeld and Ashfaq Khokhar, 2008, Distributed Multi-Dimensional Hidden Markov Model: Theory and Application in Multiple-Object Trajectory Classification and Recognition, SPIE-IS&T Vol. 6820 68200O-1.

[7] K. Assaleh, and M. Al-Rousan, 2005, "Recognition of Arabic Sign Language Alphabet Using Polynomial Classifiers", EURASIP Journal on Applied Signal Processing (JASP), 2005(13), pp 2136-2145, 2005.

[8] E. Keogh, 2002, “Exact indexing of dynamic time warping,” in International Conference on Very Large Data Bases, 2002, pp. 406–417.

[9] Ho-Sub Yoon, Jung Soh, Younglae J. Bae, and Hyun Seung Yang, 2001, Hand gesture recognition using combined features of location, angle and velocity, Pattern Recognition Volume 34, Issue 7, 2001, Pages 1491–1501.

[10] Chang-Yi Kao, Chin-Shyurng Fahn, 2011, A Human-Machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion, Procedia Engineering Volume 15, 2011, Pages 3739–3743.