Fourier Descriptors and Neural Networks for Shape Classification

Terry McElroy
Elizabeth Wilson
Gretel Anspach

Raytheon Company
1001 Boston Post Road
Marlboro, MA 01752

Terry_L_McElroy@ccmail.ed.ray.com
bwilson@sud2.ed.ray.com
gretel@tlf360.ed.ray.com

ABSTRACT

There is a pressing need for sign language to English translation capability to supplement the shortage of sign language interpreters and to provide an aid for training. A modular hybrid design is underway to apply various techniques, including neural networks, in the development of a translation system that can facilitate communication between deaf and hearing people as part of an overall system to automatically translate American Sign Language to spoken English. The key features to be analyzed are hand motion, hand location with respect to the body, and handshape. In this paper, a neural network is used to recognize and classify alphanumeric handshapes using Fourier Descriptor coefficients as an input vector. The algorithm is described and results shown for applying this technique to experimental images.

1. INTRODUCTION

The focus of this work is on the automatic classification of the handshape in a video image of a person communicating in American Sign Language or Signed English. A modular hybrid design is underway to apply various techniques in the development of a translation system that will facilitate communication between deaf and hearing people.

Accurate handshape classification is the most difficult part of the feature extraction for this application because so many handshapes are similar and there is variation in the formation of the handshapes among different signers. The most promising results have been in the generation of Fourier Descriptor coefficients as inputs to a neural network classifier. The Fourier Descriptor coefficients are computed for handshape boundary curves resulting from preprocessing the image to generate a silhouette of the hand. The neural network classifier used is the Learning Vector Quantization (LVQ) architecture.

1.1 Related Work

A number of neural network architectures have been used for extracting similar features. Multilayer Perceptrons have been used to interpret handshape signals transmitted by a specially designed glove [3]. The goal of the work presented here is to classify the handshape in a passive manner without requiring the signer to wear any special equipment that would obstruct their natural signing style. The LVQ network has been used for face recognition [1] and an Adaptive Resonance Theory architecture has been used for silhouette classification [2].

 0-7803-2431-5/95 $4.00

Fourier Descriptors have been shown to be a useful feature in image processing. These descriptors have been used for shape discrimination noting that the coefficients are invariant to angular rotation [6]. Fourier processing has also been used to assist in handprinted character recognition [4] and three-dimensional object identification [7].

1.2 Complexity of This Application

The task of character recognition in ASL is vastly more difficult than recognizing print. While text location and shape is regular, people often use different styles (not unlike a spoken "accent"), varying signing speeds, and irregular sign shapes and locations. Children will form their handshapes with less clarity than adults. Fluent signers will form the handshapes with more ambiguity in casual conversation. In addition, while text is typically black characters on white background, the ASL translator must accommodate people of varying skin color, low and high contrast between hand location and clothing, and confusing or masking clothing.

This work embodies a unique approach of using Fourier Descriptors and Neural Networks to classify handshapes. Previous work by this group [8] focused on distinct handshapes and reported excellent results for different signers, cameras, and segmentation methods. The results, however, were limited to shapes that were very different from one another, such as "a", "b", "2", and "3" handshape groups. It was concluded that the combination of Fourier Descriptors and Neural Networks was a viable approach for handshape classification.

This paper describes an extension of this work to include all 36 alphanumeric handshapes (letters a-z and numbers 1-10). With these classification categories, many are similar ("a" and "e") and some are identical ("2" and "v"). In addition, some handshapes that look very different in three-dimensions are viewed similarly in their two-dimensional representation in video ("k" and "1" , "e" and "o"). The effects of preprocessing techniques and clothing variations on the classifier results were also investigated.

2. ALGORITHM DESCRIPTION

The basic algorithm is to preprocess the video image to yield Fourier Descriptor coefficients that encode the shape of the hand silhouette. These descriptors are then used as inputs to a neural network that classifies their shapes. The network is trained with various examples from different signers and is tested with new images from new signers.

For the experiments described here, the images were collected using a standard consumer video camera which was digitized into an 8-bit gray scale image using a Dipix P360F Frame Grabber board in an IBM 486 personal computer. To reduce noise and smooth the handshape boundary curve, the digitized images were processed with a 3 x 3 pixel median filter. The images were then subjected to a simple gray scale value threshold to eliminate the background and highlight the boundary of the hand.

The handshape boundary curves were considered as a counter-clockwise closed contour. The Fourier Descriptor coefficients were calculated using a 32-point FFT. The first and last points were not used because they did not contribute to discriminating the shapes. The remaining 30 coefficients became the input vector for the LVQ neural network.

The LVQ network [5] was trained using the LVQ2 and LVQ3 algorithms balancing the vectors two or three times to improve the classification accuracy. Recommended parameter values were accepted as part of the training. Test images different from the training images were applied to the resulting trained network for evaluation.

3. RESULTS

The images were generated with 4 subjects who used similar ASL handshapes. A dark background was used to provide high contrast between the hand and the background. In cases where the simple thresholding algorithm failed to produce a closed contour, the image required manual thresholding.

The handshapes that were captured and segmented automatically by the simple median filter and threshold application resulted in 49% classification accuracy. Of the 36 classes represented, only 24 are statistically unique with 12 of the classes being identical to other classes. Large misclassification for these cases was expected and observed. Clothing worn by the subjects were also viewed as a factor in affecting the wrist boundary condition. The handshapes that were captured and hand-segmented to produce consistent boundary curves resulted in 84% classification accuracy with the same statistical ambiguity for the 12 redundant classes.

An analysis is currently being performed to determine which subset of the handshape classes would be most appropriate for automatic classification. As reported previously [8], 100% classification accuracy has been achieved for both automatically segmented and hand-segmented handshapes for 4 distinct classes. There is also work under way to evaluate the effect of the preprocessing to produce the segmented contour and optimize the image presented for classification.

4. SUMMARY

Preliminary results have been encouraging for using Fourier Descriptor coefficients as the input to a neural network classifier to identify handshapes from video images of a person communicating in sign language. To evaluate the feasibility of such an algorithm, test images were preprocessed to isolate the area of the hand and produce a closed contour representing the hand. Fourier Descriptors were extracting using an FFT on the contour and used as input features to an LVQ neural network for subsequent classification.

REFERENCES

[1] H. Bouattour, F. Fogelman-Soulie, and E. Viennet, "Neural Nets for Human Face Recognition," *Proceedings of the IEEE International joint Conference on Neural Networks*, Vol. 3, 1992, pp. 700-704.

[2] G. A. Carpenter, S. Grossberg, and G.W. Lesher, "A What-and-Where Neural Network for Invariant Image Preprocessing," *Proceedings of the IEEE International Joint Conference on Neural Networks*, Vol. 3, 1992, pp. 303-308.

[3] S.S. Fels and G.E. Hinton, "Glove-Talk: A Neural Network Interface Between a Data-Glove and a Speech Synthesizer," *IEEE Transactions on Neural Networks*, Vol. 4, No. 1, 1993, pp. 2-8.

[4] G.H. Granlund, "Fourier Processing for Hand Print Character Recognition," *IEEE Transactions on Computers*, February 1972, pp. 195.

[5] T. Kohonen, "The Self-Organizing Map," *Proceedings of the IEEE*, Vol. 78, No. 9, September 1990, pp. 1464-1477.

[6] E. Persoon and K.S. Fu, "Shape Discrimination Using Fourier Descriptors," *IEEE Transactions of Systems, Man, and Cybernetics*, Vol. SMC-7, No. 3, March 1977, pp. 170.

[7] C.W. Richard Jr. and H. Hemami, "Identification of Three-Dimensional Objects Using Fourier Descriptors of the Boundary Curve," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-4, No. 4, July 1974, pp. 371.

[8] E. Wilson and G. Anspach, "Neural Networks for Sign Language Translation," *Applications of Neural Networks, SPIE Conference 1965*, April 1993.

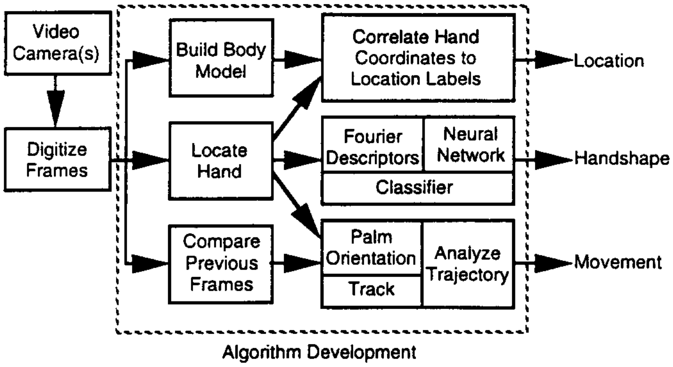

Figure 1. Block Diagram of Sign Language Parameter Extraction System

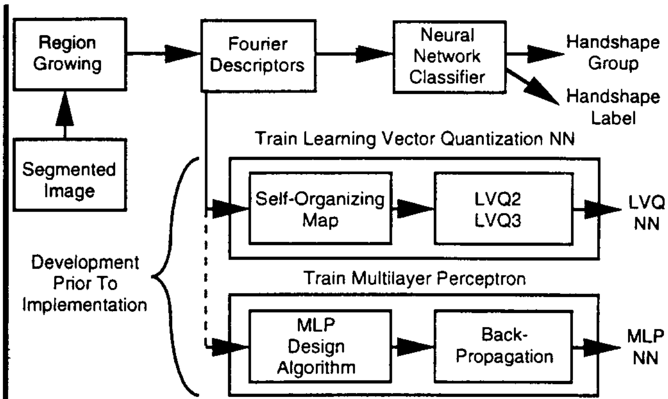

Figure 2. Handshape Algorithm

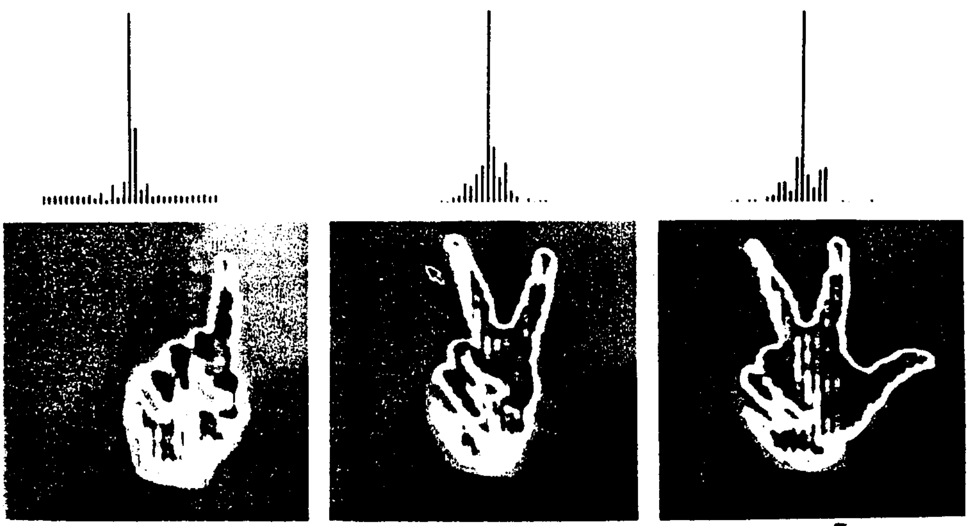

Figure 3. ASL Handshapes "1", "2" and "3" with Fourier Descriptors

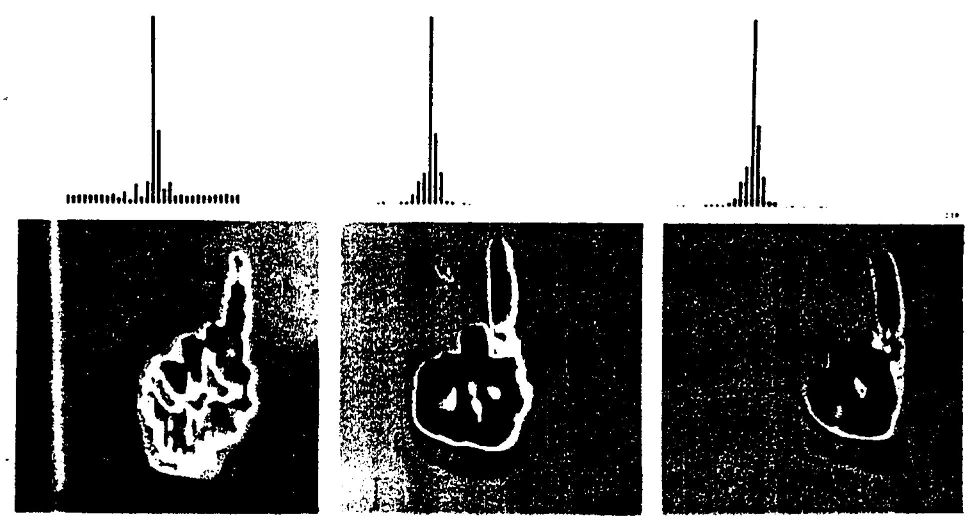

Figure 4. ASL Handshape "1" from Three Subjects with Fourier Descriptors