Journal of ELECTRICAL ENGINEERING, VOL. 60, NO. 4, 2009, 204–208

LVQ FOR HAND GESTURE RECOGNITION BASED ON DCT AND PROJECTION FEATURES

Ahmed Said Tolba — Mohamed Abu Elsoud — Osama Abu Elnaser *

The purpose of this paper is to set out an investigation into the use of handshape for the development of a gesture recognition algorithm. We present a system that performs automatic gesture recognition without using data gloves or colored gloves. Both the Discrete Cosine Transform and Projection features are used to match the input pattern against a standard gesture Database using a LVQ, algorithm. Experiments were based on 24 hand-shapes, produced by Thomas Moeslund [1]under special conditions and possible consideration of scaling, Translation, Rotation, Color and Illumination variance. A training set of 480 imaged (20 occurrence of each of the 24 hand-shape) and testing set of 480 images (20 occurrence of each of the 24 hand-shape) has been used. The correct recognition rate is approximately 83.13 % to 84.17 % for diagonal projection features and approximately 85.36 % to 87.50 % for Cols & Rows projection features. It reaches 91.04 % for DCT features.

K e y w o r d s: gesture recognition, discrete cosine transform

1 INTRODUCTION

A primary goal of gesture recognition research is to create a system which can identify specific human gestures and use them to convey information for device control. Applications vary from remote control mechanisms to interactive computer games.

In recent years various approaches to gesture recognition have been proposed. Gupta *et al* [2] presented a method of performing gesture recognition by tracking the sequence of contours of the hand using localized contour sequences. Chen *et al* [3] developed a dynamic gesture recognition system using Hidden Markov Models (HMMs). Patwardhan *et al* [4] recently introduced a system based on a predictive eigentracker to track the changing appearance of a moving hand. Kadir *et al* [5] describe a technique to recognize sign language gestures using a set of discrete features to describe position of the hands relative to each other, position of the hands relative to other body locations, movement of the hand, shape of the hand. While some of these approaches display impressive results, many exploit controlled environments and a compromise between vocabulary size and recognition rate.

To achieve accurate gesture recognition over a large vocabulary we need to extract information about the hand shape. We propose a system which is based mainly on some preprocessing steps that extract and transform the bitmap image from its original form to a new form will be suitable as an input to the recognition process. This is achieved by using two different feature extraction techniques; Image Projection and Discrete Cosine Transform. Once the suitable features have been extracted, classification subspaces are created by performing Learning Vector Quantization (LVQ).

The remainder of this paper is organized as follows: The architecture of the system is introduced in Section 2. The algorithm used to pre-process the input gesture and prepare it for further processing is reported in Section 3. Feature extraction methods are discussed in Section 4. The gesture recognition technique including hand shape recognition and neural network classifier are described in Section 5. In section 6, some experiments will be discussed and finally we provided our conclusions.

2 SYSTEM ARCHITECTURE

We propose a recognition system architecture that takes into account the functioning of sign languages and measurement limitations due to capture device used and variability state of the surrounding environment. The general form of vision-based gesture recognition system architecture is illustrated in Figure 1 and composed several basic stages

- Pre processing Stage
  This includes a series of steps that Scaling, Filtering and Converting the hand gesture to Monochrome format.
- Feature Extraction Stage
  This includes methods for transforming the Monochrome image into set significant features.
- The recognition of target object
  By matching the input pattern against the standard gesture Database using a LVQ algorithm.

* Faculty of Computer Science and Information Systems, Mansoura University, Egypt; Moh_soud@mans.edu.eg

ISSN 1335-3632

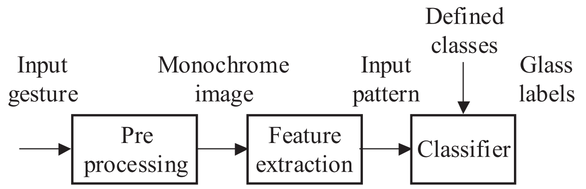

Fig. 1. Vision-based gesture recognition system architecture

Fig. 2. DCT Stages

3 HAND IMAGE PRE PROCESSING

The input hand gesture is preprocessed and converting it into a suitable format for easily further processing stages. This process is described in detail in [6] and illustrated in algorithm.1 that includes a series of steps like Scaling, Filtering and Converting the hand gesture to Monochrome format.

Algorithm 1 Basic Pre-Processing Steps

Step 1: Acquire the Hand Gesture Image.
Step 2: Convert Hand image to a Bitmap Gray scale.
Step 3: Hand is colored normalized using histogram equalization technique.
Step 4: Convert the gray bitmap image to Monochrome bitmap image.
Step 5: Hand objects are centered using the center of the bounding box.
Step 6: Scaling the image to 25×25 Pixels.
Step 7: Convolve Image with a 3×3 Gaussian Kernel to reduce noise.

4 FEATURE EXTRACTION

This stage is concerned with transforming the bitmap of the Monochrome image into a new form which is suitable as an input to the neural network to be trained without loss of the information represented on the original image. Two different feature extraction techniques Discrete Cosine Transform and image Projection are presented.

4.1 Discrete Cosine Transform

Many researches have been proposed for using DCT as in [7],[8]and [9]. This algorithm operates on three stages as shown below:

4.1.1 DCT Transformation

The DCT is closely related to the Fourier Transform. It takes a set of points from the spatial domain and transforms them into an identical representation in the frequency domain. The discrete cosine transform of a two-dimensional signal is calculated according to the following Equation 1:

$$\mathrm{DCT}(i,j) = \frac{1}{\sqrt{2N}} C(i)C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \mathrm{Pixel}(x,y) \times \cos\frac{(2x+1)i\pi}{2N} \cos\frac{(2y+1)i\pi}{2N}\,,$$

$$C(x) = \begin{cases} \frac{1}{\sqrt{2N}} & \text{if } x = 0 \\ 1 & \text{if } x > 0\,. \end{cases} \tag{1}$$

A more efficient form of the DCT can be calculated using matrix operations. To perform this operation, we first create an N-by-N matrix known as the Cosine Transform matrix, C according to Equation 2.

$$C_{i,j} = \begin{cases} \frac{1}{\sqrt{N}} & \text{if } i = 0\,, \\ \frac{2}{\sqrt{N}} \cos\frac{(2y+1)i\pi}{2N} & \text{if } i > 0\,. \end{cases} \tag{2}$$

Once the Cosine Transform matrix has been built, we transpose it by rotating it around the main diagonal (CT). Once these two matrices have been built, we can take advantage of the alternative definition of the DCT function according to Equation 3:

$$\mathrm{DCT} = \mathrm{C} * \mathrm{Pixels} * \mathrm{CT} \tag{3}$$

4.1.2 Quantization

DCT doesn't actually perform compression. It prepares for the quantization stage of the process. Quantization is the process of reducing the number of bits needed to store an integer value.

$$\text{Quantized value}(i,j) = \mathrm{DCT}(i,j)/\mathrm{quantum}(i,j)\,. \tag{4}$$

$$\mathrm{Quantum}(i,j) = 1 + \big((1+i+j) * \text{quality}. \tag{5}$$

4.1.3 Zigzag Sequence

It is the conversion of the two dimension matrix into a vector. The result vector is treated as a set of extracted features; by obtaining the previous vector the process of feature extraction has been completed.

4.2 Image Projections

In this technique we present two methods, one move along the rows and the columns, the second moves along the diagonals of the image.

$$\begin{array}{ccccc|c} 3 & 2 & 2 & 2 & 4 & \\ \hline 1 & 0 & 1 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 5 \\ 0 & 1 & 0 & 1 & 1 & 3 \end{array}$$

Fig. 3. Rows and Cols projection features extraction

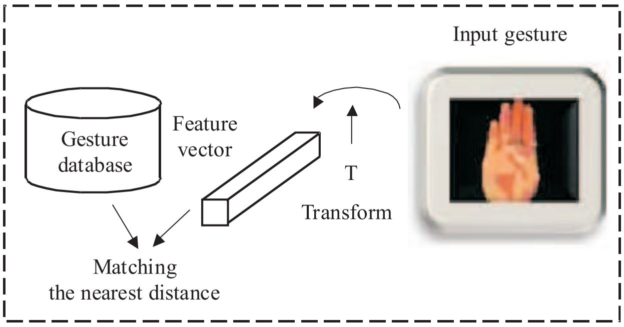

Fig. 4. Hand-Shape recognition process

4.2.1 Rows and Cols Projection

After converting the image to monochrome one, rows and column projections are calculated. This step work as follow

- Count ones (1's) *ie* (Black pixels) in the rows of the image and, also Count ones (1's) *ie* (Black pixels) in the columns of the ima ge so this step will return two vectors:
  - Row vector where each element in this vector is the count of black pixel of the corresponding row of the monochrome image, and
  - Columns vector where each element in this vector is the count of black pixel of the corresponding columns of the monochrome image. These steps are illustrated as given in Fig. 3:

$$\text{Rows projections vector} = 3,\ 1,\ 1,\ 5,\ 3 \text{ and}$$
$$\text{columns projections vector} = 3,\ 2,\ 2,\ 2,\ 4\,.$$

- Append the rows projections and the columns projections: 3,2,2,2,4), (3,1,1,5,3.

4.2.2 Diagonals Projection

After converting the image to monochrome one, data along diagonals are conceder. This step work as follow:

- Scan the image and conceder the diagonals, this step will return two vectors
  - The first represent data that have the indices $(0,0),(1,1),(2,2),\ldots,(24,24)$.
  - The second represent data that have the indices $(0,24),(1,23),(2,22),\ldots,(24,1)$.
- Append the rows projections and the columns projections

5 HAND SHAPE RECOGNITION

5.1 Related Work

To date many approaches have been proposed for hand shape recognition. These include

(i) Shamaie [10] introduced a PCA based approach.

(ii) Just *et al* [11] introduced a hand shape system based on the Modified Census Transform MCT.

(iii) A method to classify hand postures against complex cluttered background was proposed by Triesch and von der Malsburg [12] using elastic graph matching.

(iv) Yuan *et al* [11] developed their system by determining a new Active Shape Model (ASM) kernel based on the shape contours.

(v) Chen *et al* [3] present a method of classifying the static hand poses by using the Fourier Descriptor to characterize the spatial features of the hands boundary.

Many of these approaches for hand shape recognition display sub-optimal results due to the highly deformable nature of the hand. Once again a compromise is determined between small vocabulary and accurate recognition. Due to the complex temperament of the hand, any hand shape classifier needs to be able to cope with small rotations and translation transformations.

We now have the backbone for hand shape recognition system which use supervised training algorithm LVQ The system is constructed as follows:

Training – Generating a transformation subspace for each hand shape.

Testing – Project the test image into each of the subspaces to find the subspace with the nearest perpendicular distance. This subspace will be representative of one particular hand shape. The full description of Hand-shape Gesture recognition process can be summarized in the Fig. 4.

5.2 Learning vector quantization

Learning vector quantization (LVQ) is a method for training competitive layers in a supervised manner. A competitive layer automatically learns to classify input vectors. However, the classes that the competitive layer finds are dependent only on the distance between input vectors. If two input vectors are very similar, the competitive layer probably will put them in the same class. There is no mechanism in a strictly competitive layer design to say whether or not any two input vectors are in the same class or different classes.LVQ networks, on the other hand, learn to classify input vectors into target classes chosen by the user.

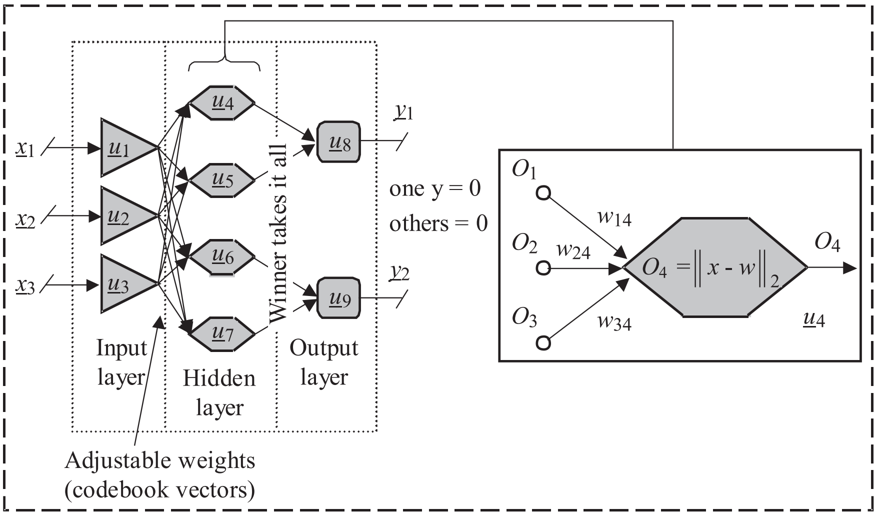

Fig. 5. LVQ architecture

Table 1. Network Structure

Technique	Input Size	Hiddens Neurons	Epochs	Alpha
DCT	200	50, 100	1000, 1500, 2000	0.05, 0.1
Projection	50	50 , 100	1000, 1500, 2000	0.05, 0.1

5.2.1 Architecture of the LVQ Classifier

The LVQ is essentially a Kohonen network without the topological structure.

In order to classify real hand images we create a train set for each hand-shape as described earlier. However, this time all the training images will be preprocessed using the techniques described above. Similarly all test images will traverse through the same pre-processing steps.

The primary objectives of our experiments were to determine the structure of the network being trained, so we must identify a set of parameters as shown in Tab. 1

5.2.2 Basic LVQ Learning Algorithm

```
public void Train( )
{
    do
      {
      for(int classIndx =0; classIndx <
            classesNames.Length;classInex++)
         {
         for(int ptrnIndx =0; ptrnIndx <
               in_Ptrns[classIndx].Length; ptrnIndx++)
            {
         pattern = PrepareInput();
         winner = WinnerNeuron(pattern);
```

Table 2. The system recognition accuracy using individual LVQ based on: Diagonals Projection; Rows and Cols features; DCT features

ALPHA	Hiden Neurons	Epochs	Diagonals projection Recognition-rate Training %	Diagonals projection Recognition-rate Testing %	Row-Col features Recognition-rate Training %	Row-Col features Recognition-rate Testing %	DCT features Recognition-rate Training %	DCT features Recognition-rate Testing %
0.05	50	1000	90.42	83.13	94.79	86.04	97.29	90.36
		1500	90.83	83.54	95.00	86.04	97.50	90.83
		2000	90.83	83.54	95.00	85.83	97.29	90.83
	100	1000	90.00	83.33	95.43	85.36	97.08	89.38
		1500	90.00	83.33	95.43	86.25	97.29	90.42
		2000	90.00	83.33	95.65	86.76	97.29	90.63
0.10	50	1000	90.83	83.96	95.63	85.83	96.67	90.36
		1500	91.04	84.17	96.04	86.04	97.08	91.04
		2000	90.83	84.17	96.04	85.42	97.29	91.04
	100	1000	90.63	83.13	95.21	86.46	96.25	90.00
		1500	91.04	84.17	95.42	87.50	96.88	90.36
		2000	90.83	84.17	96.04	86.88	97.08	89.78

```
        winner.Update((name==winner.name),pattern);
          }
        }
      alpha = 0.5 * alpha ;
      }while(iteration < epochFactor);
}
```

6 RECOGNITION EXPERIMENTS

Table 2 illustrates the variation of the system recognition accuracy based on diagonals projection, Rows and Columns Projection and DCT features respectively at feature level and individual LVQ classifier at match level.

7 CONCLUSIONS

In this paper we have presented the results of developing a real hand-shape classification system. A supervised training algorithm; Learning Vector Quantization (LVQ) technique has been used for classification of hand gesture images. We introduced a chain of relatively complex preprocessing steps to remove user dependant features such as color and scale. A series of tests were then performed to experimentally evaluate our technique. During the course of our experiments we endeavored to identify particular techniques and parameter values that improved the accuracy of our system. Two Techniques has been proposed Discrete Cosine Transform (DCT) and Image Projection as a feature extraction Techniques to reformulate input data to suitable format accepted by the Neural Network. The results showed that the using of DCT features is better than using the image projection features.

Acknowledgements

The authors would like to thank prof. Thomas Moeslund, Aalborg University, Denmark, for making his gesture database available to us.

References

[1] THOMAS: Moeslund's Gesture Recognition Database, [Online], http://www.prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html (Last Accessed 2006).

[2] GUPTA, L.—MA, S.: Gesture-Based Interaction and Communication: Automated Classification of Gesture Contours, IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and reviews, **31** No. 1 (2001).

[3] CHEN, F.-S.—FU, C.-M.—HUANG, C.-L.: Hand Gesture Recognition Using a Real-Time Tracking Method and Hidden Markov Models, Image and Vision Computing **21** (2003), 745–758.

[4] PATWARDHAN, K. S.—DUTTA, ROY, S.: Hand Gesture Modeling and Recognition Involving Changing Shapes and Trajectories, Using a Predictive EigenTracker, Pattern Recognition, (Article in Press), 2006.

[5] KADIR—BOWDEN—ONG, E.—ZISSERMAN: Minimal Training, Large Lexicon, Unconstrained Sign Language Recognition, In Proc. BMVC04, Vol. 2, pp. 849–858, 2004.

[6] THOMAS, C.—GEORGE, A.—JUNWEI, H.—ALISTAIR, S.: Real Time Hand Gesture Recognition Including Hand Segmentation and Tracking, School of Computer Applications, Dublin City University, Ireland, 2007.

[7] RAKSHIT, S.—MONRO, D.: Robust Iris Feature Extraction And, Department of Electronic and Electrical Engineering, University of Bath, BA2 7AY, UK.

[8] ZHENGJUN, P.—HAMID, B.: High Speed Face Recognition Based On Discrete Cosine Transform And Neural Networks, Science & Technology Research center, and Departement of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire, 1999.

[9] ZHENGJUN, P.—ROD, A.—HAMID, B.: Dimensionality Reduction of Face Image Using Discrete Cosine Transforms for Recognition, Science & Technology Research center, and Department of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire, 1999.

[10] SHAMAIE, A.: Hand Tracking and Bimanual Movement Understanding, PHD Thesis, Dublin City University, 2003.

[11] JUST, A.—RODRIGUEZ, Y.—MARCEL, S.: Hand Posture Classification and Recognition using the Modified Census Transform, in Proc. IEEE International Conference on Face and Gesture, 2006, pp. 351–356.

[12] TRIESCH, J.—Von der MALSBURG, C.: Classification of Hand Postures Against Complex Backgrounds Using Elastic Graph Matching, Image and Vision Computing **20** (2002), 937–943.

[13] YUAN, Y.—BARNER, K.: An Active Shape Model Based Tactile Hand Shape Recognition with Support Vector Machin, in Proc 40th Annual Conf Information Sciences and Systems, 2006.

Received 15 November 2008

Ahmed Said Tolba, (Prof) is Dean of Arab Open University, Kuwait. He was the dean of faculty of computer science and information system, Mansoura University, Egypt until September 2007. BSc of electrical engineering, Mansoura University, Egypt at 1978. Obtained master degree in 1981 Mansoura university and PhD in Artificial Intelligance from Wuppertal university, Germany in 1988.

Mohamed Abu El Soud (Dr) Assistant Professor at Computer science Department, Faculty of computer science and information System, Mansoura University, Egypt, since Jan. 2006. BSc of electronic engineering, from Mansoura University, Faculty of Engineering, at 1991. Master degree of communication engineering, Mansoura University, 1996. PhD degree of communication engineering, Mansoura University, 2005.

Osama Abu El Naser born in Kafr El sheikh, 1984, graduated from Faculty of Computer science and Information, Computer Science Department, Mansoura University, Egypt, with grade "Excellent with Honor", 2005. Master of computer Science, in 2009. Now he is working as a demonstrator, Faculty of Computers and information, Computer Science Department, Mansoura University