2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

Simplified DFT for Hand Posture Recognition System

Hiroomi Hikawa† and Hirotada Fujimura‡

†Faculty of Engineering Science, Kansai University
3-3-35 Yamate-cho, Suita-shi, Osaka, 564-8680 Japan
Email: hikawa@ipcku.kansai-u.ac.jp
‡Genesis Technology Inc., Hyougo Japan

Abstract—This paper describes simple discrete Fourier transforms (DFT) to be used in the hand posture recognition system. Conventional DFT uses sine cosine, and square root functions, which requires high computational and hardware costs. The new DFT uses heavily quantized version of sine cosine functions that take only three values, resulting the generation of the function becomes easier and multiply add (MA) operation is replaced by simple add-sub-zero operation. For the calculation of the magnitude value, absolute and add operations are used instead of the square root function.

Due to the simple function, the recognition performance is degraded, but the simulation results shows that the deterioration of the recognition rate can be suppressed to about 1% by using training data set with the images in different positions.

1. Introduction

The use of hand gesture[1] provides an attractive alternative to cumbersome interface devices for human-computer interaction (HCI). Visual interpretation of hand gestures can make it possible to migrate the natural means that humans employ to communicate with each other into HCI. Generally hand gestures are either static hand postures[2][3] or dynamic hand gestures[4][5]. The hand sign recognition is a kind of the pattern recognition, which is a mapping process of the input vectors to a finite set of clusters, and each cluster is associated with a posture. Each input image is converted to a feature vector and its class is determined by searching the closest prototype of the cluster, which minimizes the distance to the input vector.

In [6] and [7], hand sign recognition systems based on the new classifier [8] has been proposed. In the system, input hand images are preprocessed through horizontal/vertical projection followed by discrete Fourier transforms (DFTs) that calculate the magnitude spectrum, which is used as the feature vector. Use of the magnitude spectrum makes the system very robust against the position changes of the hand image.

Our objective is the development of hardware-based posture classification system, but the use of DFT is not suitable for the hardware implementation. In this paper, simplified DFT is proposed and applied to the hand posture recognition system. The feasibility of the proposed method is verified by computer simulations.

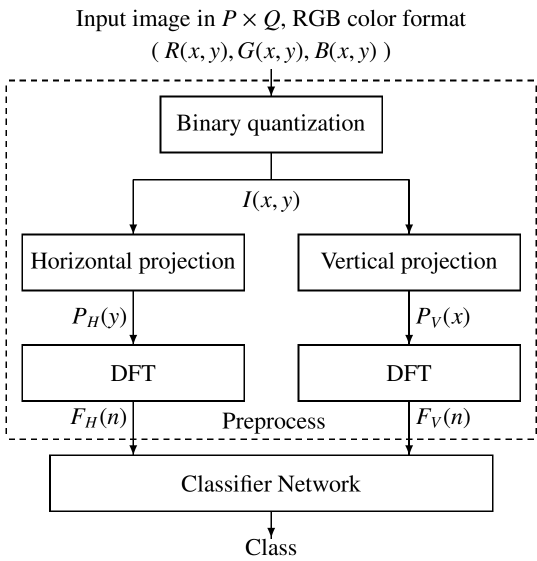


Figure 1: Hand sign recognition system.

2. Hand posture recognition system

The process flow of the hand sign recognition system [7] and signal examples are shown in Figures 1 and 2, respectively. Input image is $P \times Q$ pixels, RGB color format and the input image is preprocessed to obtain feature vectors. The preprocessing consists of a binary quantization, horizontal/vertical projections followed by two DFTs. The DFTs generate $F_H(n)$ and $F_V(n)$, which are the feature vector of the input image. The D dimensional feature vector is fed to the classifier network which finally identifies the hand sign. The network consists of the neurons with the vector distance evaluation method proposed in [8] and [6].

First, the hand portion is extracted from the input image, then it is converted to a binary image To make the hand shape extraction process easier, it is required for the users to wear a red glove, then the hand posture extraction and binarize process is realized by the following equation.

$$I(x,y) = g(\,Red(x,y),\ Green(x,y)+Blue(x,y)\,)\cdot g(\,Red(x,y),\ \rho\,) \tag{1}$$

where, $I(x,y)$ is the binary pixel value at (x,y) coordinate, $Red(x,y)$, $Green(x,y)$ and $Blue(x,y)$ are the intensity lev-

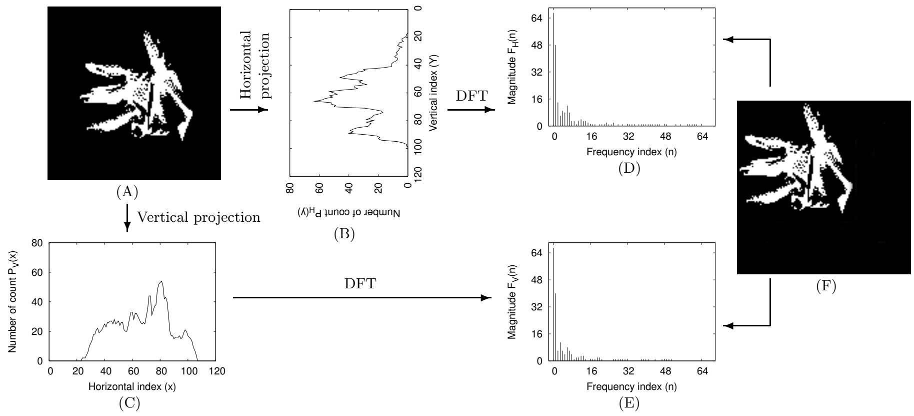


Figure 2: Example of the signals in the hand sign recognition system. (A) Binary image, (B) Horizontal projection $P_H(y)$, (C) Vertical projection $P_V(x)$, (D) magnitude spectrum $F_H(n)$, (E) magnitude spectrum $F_V(n)$, (F) The same hand posture in different positions yield identical $F_H(n)$, $F_V(n)$.

els of red, green and blue color at (x, y) of the input color image, respectively. ρ is a threshold parameter and $g(\cdot)$ is a threshold function.

$$g(x, \theta) = \begin{cases} 1 & \text{if } x \geq \theta \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

An example of the binary image is shown in Fig. 2(A).

$I(x, y)$ is then fed to horizontal and vertical projection subsystems that calculate two types of histograms, i.e., horizontal projection histogram $P_H(y)$ and vertical projection histogram $P_V(x)$.

$$P_H(y) = \sum_{x=0}^{P-1} I(x, y) \tag{3}$$

$$P_V(x) = \sum_{y=0}^{Q-1} I(x, y) \tag{4}$$

The horizontal and vertical projection histograms of the binary image of Fig. 2(A) are depicted in Fig. 2(B) and (C), respectively.

Following two DFTs calculate $F_H(n)$ and $F_V(n)$, the magnitude spectrums of $P_H(y)$ and $P_V(x)$, respectively. Two images with the same hand posture placed in different positions yields different projection histograms. However, the only difference between two histograms is their positions and their histogram shapes are identical. Thus the spectrum amplitude $F_H(n)$ and $F_V(n)$ calculated from images containing the same hand posture placed in different positions are all identical. As an example, two images (A) and (F) in Fig. 2 yield the same amplitude spectrums (D) and (E) as they are the same hand sign in different positions. In this way, the use of the amplitude spectrums $F_H(n)$, and $F_V(n)$ provides the position independent hand posture identification.

As can be seen in Fig. 2(D) and (E), the feature of the image concentrates in the lower frequency components. Thus the lower $D/2$ frequency components are taken from both of the spectrum data $F_H(n)$, $F_V(n)$ to form a D dimensional feature vector. The dimension of the vector is reduced from $P \times Q$ to D, and usually $D \ll P \times Q$.

The D-dimensional feature vector $\vec{x}$ is fed to the classifier network to determine the hand posture class to which the input image belongs. The classifier network consists of the neurons with range check functions. The detail of the classifier network is found in [6]-[8].

3. Simplified DFT

N-point DFT of the data sequence $x(n)$ is given by,

$$A(k) = \sum_{n=0}^{N-1} x(n) \cdot \cos(\frac{2\pi nk}{N}) \tag{5}$$

$$B(k) = \sum_{n=0}^{N-1} x(n) \cdot \sin(\frac{2\pi nk}{N}) \tag{6}$$

Then the spectrum amplitude is,

$$X(k) = \sqrt{A(k)^2 + B(k)^2} \tag{7}$$

Obviously equations (6) ~ (7) are not suitable for hardware implementation as they include complex functions and multiply and add (MA) operations. To simplify the equations, $\cos(\cdot)$ and $\sin(\cdot)$ functions are replaced by $\mathrm{tcos}(\cdot)$ and $\mathrm{tsin}(\cdot)$, respectively.

$$\hat{A}(k) = \sum_{n=0}^{N-1} x(n) \cdot \mathrm{tcos}(\frac{2\pi nk}{N}) \tag{8}$$

$$\hat{B}(k) = \sum_{n=0}^{N-1} x(n) \cdot \mathrm{tsin}(\frac{2\pi nk}{N}) \tag{9}$$

As shown in Fig. 3, $\mathrm{tcos}(\cdot)$ and $\mathrm{tsin}(\cdot)$ are the heavily quantized (tri-state) version of $\cos(\cdot)$ $\sin(\cdot)$ functions, and they take only three values, i.e., -1, 0 or $+1$. As the functions

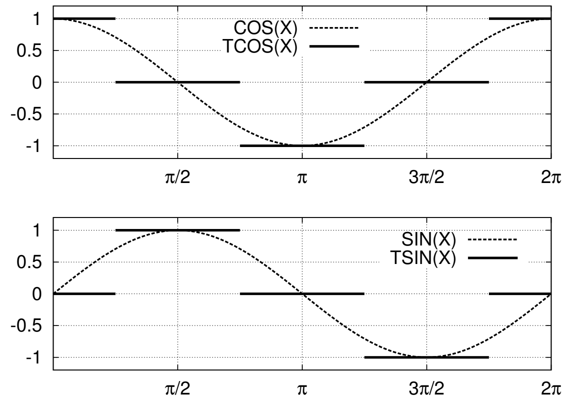


Figure 3: Tri-state cos, sin functions

(A) (B)

Figure 4: Posture images with different positions, (A) LT group, (B) RB group.

tcos(·) and tsin(·) take only three values, (8) and (9) can be realized with simple add or sub or zero (ASZ) operation.

(7) is also simplified as,

$$\hat{X}(k) \;=\; \mid \hat{A}(k) \mid \;+\; \mid \hat{B}(k) \mid \tag{10}$$

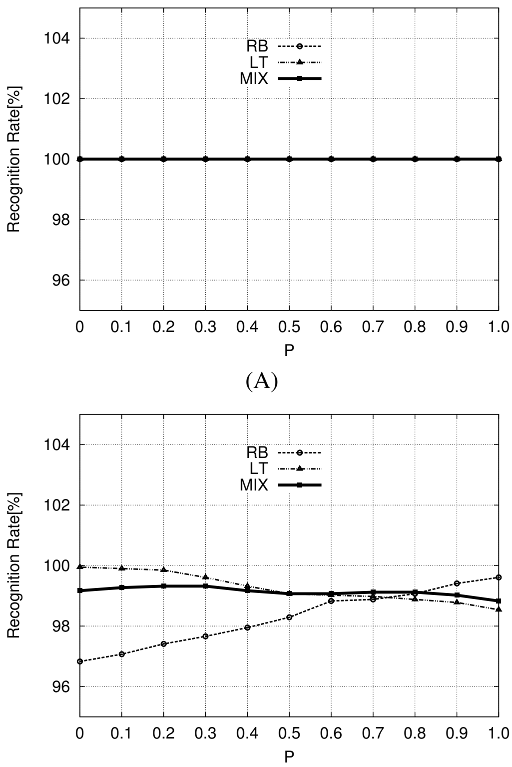
The computational cost of (10) is lower than that of (10).

4. Simulation

It is anticipated that the employment of the simplified DFT degrades the performance of the system, especially, the system's robustness against the position change of the hand image. The effect of the new DFT is investigated by computer simulations using hand posture image data set consisting of 41 classes, which are the static hand postures taken from Japanese hand signs. Each class data consists of 100 images, totaling 4,100 images. The data set can be also divided two groups. One is images placed in left-top corner of the frame (LT group), the other group is made of images in right-bottom corner (RB group). Fig. 4 shows some examples of the images. Note that the difference among images belonging to the same class is the position of the hand posture, and their hand shapes are all identical.

For the training of the classifier network, three types of the learning data, RB, LT, MIX are provided. The number of the training data for each class is 20.

RB : randomly selected 20 images from RB group,
LT : randomly selected 20 images from LT group,
MIX : 10 images are randomly selected from the both groups.



(A)

(B)

Figure 5: Position change vs. recognition rate, (A) with conventional DFT, (B) with new DFT.

Recognition rates are obtained by feeding test data to the system. The test data consists of images taken from the both groups. Here the mixture rate P is defined as follows.

$$P = \frac{N_{RB}}{N_{RB} + N_{LT}} \tag{11}$$

where, N_{RB} and N_{LT} are the numbers of RB and LT group images included in the training data, respectively. The number of the test data for each class is 50, i.e., $N_{RB}+N_{LT} = 50$.

The systems with conventional DFT and proposed DFT are trained three times with different training data sets, and their recognition rates are measured using test data sets with different mixture rate P.

The relations between the recognition rates of the two systems and the mixture rate P are shown in Fig. 5. Fig. 5(A) shows that the recognition rate of the system with conventional DFT is 100%. This is because the effect of the position difference is removed by conventional DFT, resulting very precise classification is performed by the classify network as the hand posture shapes of the same class are all identical.

Fig. 5(B) is the recognition rate of the system with the proposed DFT, which shows interesting nature of the system. As is expected, the recognition rates have correlation with the input data and the training data set. Trained with the RB data set, the recognition rate becomes better as P

Table 1: Average recognition rates

	system with conventional DFT	system with new DFT
Average recognition rate(%)	93.46	92.49

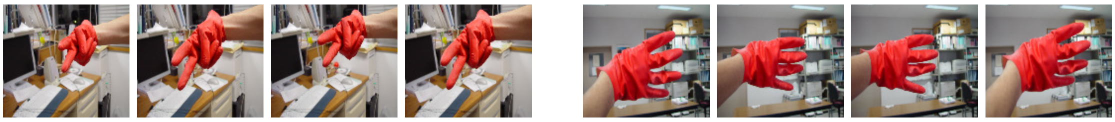


Figure 6: Input image examples.

increases (N_{RB} increases). On the other hand, recognition rate of the system trained with the LT data set increases as P approaches 0 (N_{LT} increases). In these two cases, the worst recognition rate is about 97%. However, the recognition rate with the MIX data set, deterioration is only 1%. Compared to the reduction of the hardware cost, 1% deterioration on the recognition performance is acceptable.

Next, the system with proposed DFT is applied to practical 41 static hand sign recognition. Each class has 100 images, and these images are slightly different both in position and shape even though they belong to the same class. Some example pictures used for the experiment are shown in Fig. 6. Randomly selected 50 images are used for the training and the remaining 50 images are used to obtain the recognition rate. This process is repeated five times and the average of the recognition rates are used for the evaluation. The results is summarized in Table 1. As the images used in the training are different in positions and shapes, the difference between the recognition performance between the systems with conventional and new DFTs are only about 1% as the previous simulation indicated. Also it should be noted that the recognition rate of 92% is considered to be very high.

5. Conclusions

This paper has described simple DFT to be used in the hand posture recognition system. Conventional DFT uses sine cosine, and square root functions. all of which require high computational and hardware costs. The new DFT uses heavily quantized version of sine cosine functions that take only three values, resulting the generation of the function becomes easier and MA operation is replaced by simple add-sub-zero operation. For the calculation of the magnitude value, absolute and add operations are used instead of the square root function.

Due to the simplified functions, the recognition performance is degraded, but the simulation results shows that the deterioration of the recognition rate can be suppressed about 1% by using training data set with the images in different positions. Compared to the reduction of the hardware cost, 1% deterioration of the recognition rate is acceptable.

The hardware implementation of the whole system is left for further research. The hardware solution of the proposed architecture will make it possible to recognize the hand posture in real-time. In addition, the hardware system can be extended to recognize the dynamic action, which is a sequence of hand postures.

References

[1] Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, pp.677-695, 1997.

[2] J. Triesch, C. von der Malsburg, "A System for Person-Independent Hand Posture Recognition against Complex Backgrounds," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.23, no.12, pp.1449-1453, 2001.

[3] K. Hoshino, T. Tanimoto "Realtime Hand Posture Estimation with Self-Organizing Map for Stable Robot Control," *IEICE Trans. on Information and Systems*, Vol.E89-D, no.6, pp.1813-1819, 2006.

[4] A. F. Bobick and A. D. Wilson, "A state-based approach to the representtion and recognition of gesture," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, no.12, pp.1325-1337, 1997.

[5] H-H. Yang, N. Ahuja, and M. Tabb, "Extraction of 2D motion trajectories and its application to hand gesture recogntion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, no.8, pp.1061-1074, 2002.

[6] H. Hikawa, S. Matsubara "Pseudo RBF Network for Position Independent Hand Posture Recognition System," *Proc. 2007 INNS-IEEE IJCNN 2007*, 2007.

[7] H. Fujimura, Y. Sakai, H. Hikawa, "Japanese Hand Sign Recognition System," *Proc. 2007 ICONIP 2007*, (in press).

[8] S. Matsubara, H. Hikawa, "Hardware Friendly Vector Quantization Algorithm," *Proc. IEEE ISCAS 2005*, pp.3623 - 3626, 2005.