A ROTATION INVARIANT APPROACH ON STATIC-GESTURE RECOGNITION USING BOUNDARY HISTOGRAMS AND NEURAL NETWORKS

Simei G. Wysoski, Marcus V. Lamar†, Susumu Kuroyanagi , Akira Iwata

Dept. of Electrical and Computer Eng., Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya-shi, 466-8555 Japan
E-mail: simei@mars.elcom.nitech.ac.jp
†Dept. of Electrical Eng., Federal University of Parana
CP19011 CEP 81531-970 – Curitiba Brazil

ABSTRACT

The appropriate selection of feature extraction method plays an important role in designing a pattern recognition system. The proper representation of features contributes to a significant improvement of the classifier performance and also to the reduction of time processing. The purpose of this study is to present a description of hand's posture features based on boundary histograms. The use of histograms aims to deal with two problems: the chain small magnitude circular-shift problem caused by posture rotation and, to attenuate the non-linearity caused by shape differences when performing gesture postures. We also present a fast search start point algorithm for the boundary chain that gives a rotation invariance property to the system. The performance was evaluated using 26 postures of American Sign Language, and a comparison with other algorithms is presented. As result we obtained a robust method to be used in large-scale applications using neural networks.

1. INTRODUCTION

Many researchers have extensively studied gesture recognition, but it is still a challenging and interesting problem in the computer science field due to the complexity and the vagueness of the patterns. Not only due technical reasons but it also presents a human appeal that suggests the use of gesture is one of the most primitive and efficient ways of communication. Even without any systematic and structured language system, it is possible for two people to communicate each other by performing only some basic gestures.

Regarding the process of capturing gestures' information let us to consider the human eyes and the human natural vision system. People can easily using eyes, identify objects, have a sense of the object's shape, detect the object's trajectory and movements. In order to try to simulate these human natural abilities, an automatic view-based hand gesture recognition system has, generically, its tasks divided in three distinct stages: Image Capture/Pre-processing, Feature Extraction and Classification.

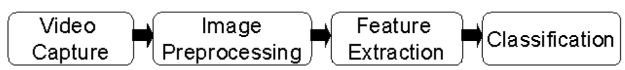

Fig. 1 Hand Gesture Recognition System

Various methods of representing hand's posture information have been proposed. Some of them make use of auxiliary devices such as color gloves or finger marks to simplify the problem [1][2][3][4]. These approaches allow the system to perform a detailed posture description, capturing inclusive information of finger and palm separately. Alternatively, some methods propose application of skin color filter for hand's segmentation [5][6], thus providing a more natural and practical way rather the ones mentioned earlier.

In this study, we focus the discussion on the feature extraction phase that follows the object's segmentation, thus the hand region is assumed have been extracted correctly. Various methods have been proposed to represent the hand's shape. Birk [5] described a method that submits pixel's intensity information to a PCA transformation. Followed by PCA transformation, a method to reduce the dimensionality of the features was implemented. In the same work a rotation algorithm based in main axes was performed giving to the system rotation invariance to the image plane.

Other approaches have been proposed by the use of boundary shape information. Boundary chain can be represented in many ways, such as Chain Codes, Signature, Chord's size and others [7][8]. In these methods, only the hand's external shape information is captured. Gupta [8] described a robust system that recognizes 10 postures using chord's size. After alignment using circular shifting (become rotation invariant), DP matching algorithm is used as classifier by measuring the discrepancy between the chord's size chain templates and the test samples.

In this study, chord's size measures are represented as histogram to give a robust set of invariant-to-shift rotation features to be presented to a neural network classifier.

The paper is organized as follows. Section 2 provides a detailed description of the system. The database set and its properties are mentioned in Section 3 followed by experiments results. The conclusion of this study is presented in Section 4.

2. SYSTEM DESCRIPTION

2.1. Video Capture and Pre-processing

The first step consists of a camera to capture image containing hand-posture information. The image is presented to a skin color detection filter [9], and is followed by performing erosion, dilation and union of the pixels as preprocessing to clean up the spurious pixels [10]. The result is presented to a clustering process to find the remaining groups in the image. The boundary of each group is then extracted using an ordinary contour-tracking algorithm. Dividing iteratively the image in grids [7], the boundary is normalized in size as shown in Fig. 2. Boundary normalization in size gives to the system, the distance invariance property between camera and hand.

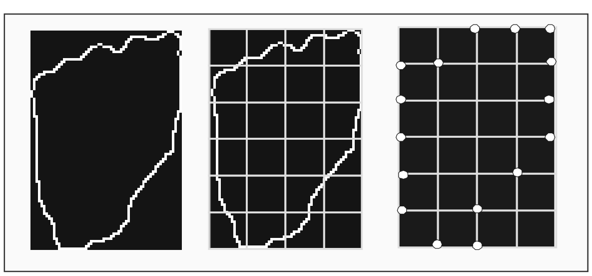

Fig. 2 Grid line division and boundary representation

2.2. Boundary description

After obtaining a boundary with a specific number of pixels, the boundary is represented as chord's size chain as described in details in [8] and shown in Fig.3. In the chord's size calculation a window size *w* need to be set. The advantage of using chord's size resides on the independency of the axis *x*-*y* coordinate system and, the dimensional reduction to the half of the number of features. However, the use of chain representation should be followed by a careful consideration on determining the chain start point (Chain Alignment Problem)

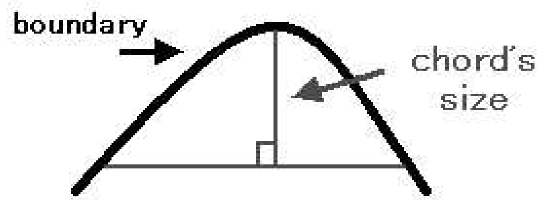

Fig.3 Chord's size representation

2.3. Classification

Gupta [8] used circular shifting for chain's alignment and to improve the generalization capacity. In his study, a non-linear alignment method (DP Matching algorithm) is employed for classification. Non-linear alignment aims to deal with posture differences among different users and also among gestures performed by a same person with different shapes.

DP Matching algorithm presented to be robust and even they need to perform a comparison with all the classes' templates for every vector to be classified, they are not time consuming when working with modest number of classes. When the number of class is relatively small, the method will give sufficient performance in reasonable processing time. However, in a large-scale problem with many classes, the necessity of creating the DP Matching matrix for each class' templates results in a more complex problem, and the system many times, become non-feasible computationally or its generality decreases. When a real gesture set of postures is taken into consideration, the size of the database and the complexity of postures lead to the necessity of a robust classifier with reasonable time processing and good generalization power. Thus, in this study we evaluate the performance of neural network in solving the classification problem, which is known to have promising properties [11].

The non-linear properties inherent in the hand-posture dataset appear in boundary chord's size representation as non-linear expansion or compression in the chain segments. It generates small localized shifting over all of the chain representation. This property makes the order of features in the input vector be shifted, which is difficult to be handled by neural network. To overdue this problem, we propose a modification to the method by the inclusion of a histogram-based description. Describing the chord's size chain as histograms, non-linear characteristics that produced a shifting in the chain, in the new representation, cause an amplitude change in the histogram elements.

2.4. Feature Extraction

The use of histograms is proposed based on the Peripheral Direction Contributivity (PDC) [12], a feature extraction method used to acquire the complex features in Japanese Kanji character recognition. Different from PDC, in this study, the image is divided in a specific number of regions *N*. The regions are divided in a radial form, according to a specific angle as shown in Fig. 4. As a result, the line that divides two regions starts in the center and finishes in the image's boundary.

For the pixels located in each region, a histogram of boundary chord's size is calculated and the whole feature vector consists of a sequential chain of histograms.

The use of histogram allows attenuating the influence of small variations in the start point detection, to be explained in the next section. However, histogram does not exclude its utilization since large magnitude shift variations can overcome completely the histogram range.

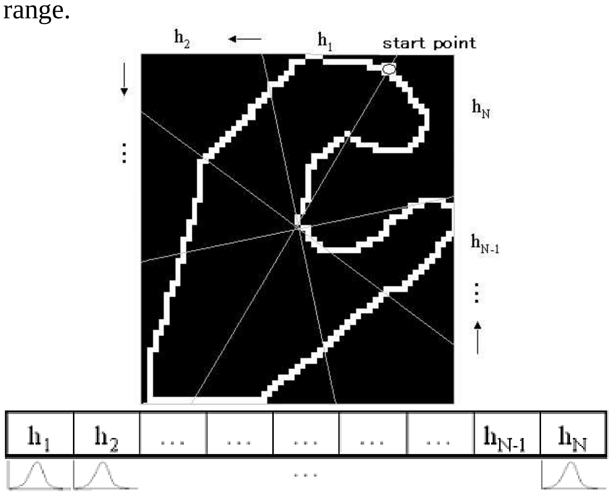

Fig. 4 Image divided in regions and feature vector model based on histograms

2.5. Start-point problem and Rotation-Invariance

In this section we intend to discuss the hand's posture rotation in the image plane when using the boundary chord's size histogram. A pure planar rotation in the image does not change the posture contour. Consequently, the boundary description by chord's size remains unchanged. It is only necessary to appropriately, set an invariant beginning for the chord's size vector or in other words, define the chain start point.

One method to define the start point is to find the principal axes of the region using eigenvectors. One of the boundary points that cross the main axes is a good cue to be used to start the chain. However, eigenvector calculation, similarly with a recursive method of circular shifting alignment, has a high computational cost.

We propose to use also the chord's size calculation on defining the start point. The approach is to use as start point the maximum chord's size value in the chain. To this method be effective, the value of window size *w* used to calculate chord's size must be considered. Increasing *w*, the chord's size amplitude is also increased, increasing consequently the signal-to-noise ratio [8]. Using a small window, the chord's size retrieves a local representation of a segment and the probability of a chord's size having the same values in different parts in the chain increases. Increasing the window value, chord's size chain gives a more global meaning of the entire object. It's important to mention that this method is susceptible to errors. In a square shape object, for example, there is 4 maximum points. Similarly, in a circular shape object, the maximum value is undetermined. However, these risks can be ignored, due the shape characteristics of sign gesture postures as it is showed in the following experiments results.

Also, different classes that have same shapes, differing each other by rotation information cannot be discriminated using rotation invariant methods. In ASL it occurs with postures ***H*** and ***U*** (see Fig 7). Postures ***H*** and ***U*** have same boundary shapes with different posture direction. ***H*** has vertical principal axes while ***U*** has horizontal ones. To deal with this problem we constraint the start-point region of search in an angle of 90°, as shown in Fig. 5. In this case, we will have a partially rotation invariant system.

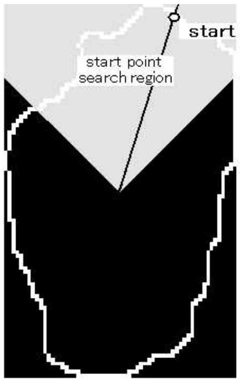

Fig. 5 Start point search region. The white region denotes the region of start point searching

3. DATABASE AND EXPERIMENTS

To test the system performance was created a database using 26 static postures of the American Sign Language as shown in Fig. 7.

From every posture, 40 pictures (RGB with 320x240 pixels) were taken and divided in 2 parts: 20 of them used for train and 20 for test. In order to facilitate hand's segmentation a homogeneous background was applied. The gestures were taken in a frontal view from a distance camera–hand of approximately 50 cm.

The boundary representation of the postures was normalized in size of 256 pixels and the window size *w* was set as 31 for features calculation. Window size equal to 255 was used when searching the start-point.

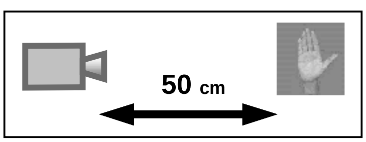

Fig. 6 Distance camera-hand

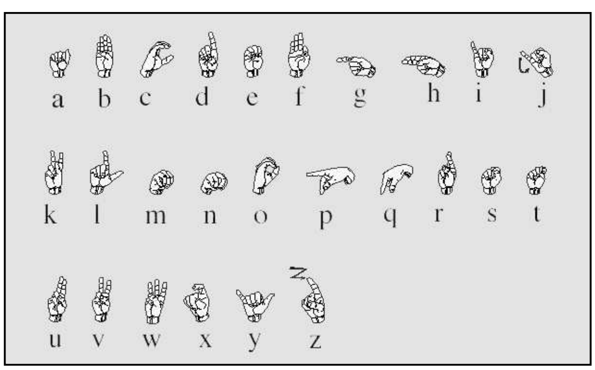

Fig. 7 26 static postures from ASL used in the experiments

Table 1 presents five experiments conducted using different features format and different classifiers. All these methods present rotation invariance characteristic.

Table 1 Experiments description

Method Number	Features format	Features Number	Classifier	Start Point Detection
1	Chord's size FFT	80	MLP	–
2	Chord's size	256	DP Matching	Circular Shift
3	Chord's size	256	MLP	MAX(Chord's size)
4	Chord's size Histogram	256	DP Matching	MAX(Chord's size)
5	Chord's size Histogram	20(histograms)x 7(resolution)	MLP	MAX(Chord's size)

The first experiment used FFT frequencies as features. Fourier transform was used due its time shifting property that can be stated as: "*A shift in time does not alter the magnitude of the Fourier coefficients*"[13]. Hence, any alignment algorithm needs to be used. Despite Fourier transform aligned the vectors, the classifiers showed be submitted to a more difficult task, mainly because most of the signal energy is presented in low frequencies in the Fourier spectrum. Using multilayer perceptron (MLP) neural network for classification, the best performance reached was 96.7%, as presented in Fig. 8.

The second and third experiments used chord's size chain as feature information. In the second experiment, every vector with 256 elements was one-by-one linearly aligned using circular shifting. Using MLP neural network, as expected, the results showed that, the localized shifts in the input neurons were not well assimilated. After performing many experiments with different number of neurons on the hidden layer, MLP did not correctly classify more than 94.0%. On the other hand, with the same features representation (chord's size chain) and using DP matching algorithm the recognition rate was 98.8%. These results demonstrate that DP matching can better handle the localized shifts resulted of differences in the same class gesture postures.

At last, we evaluated the use of chord's size represented in histograms as proposed in section 2.4 as feature extraction algorithm, and a MLP neural network for classification. In addition, the alignment, instead of using Circular Shifting, was held using the faster algorithm, based on finding the biggest chord position, as described in the section 2.4.1. The performance is shown to be similar to that of the anterior method, and correctly classified 98,7%, of the test dataset. However, as positive point, the computational time was reduced and it required less memory. Table 2 shows the performance for different number of histograms (image divisions) and also for different number of histogram resolution.

Table 2 Recognition rate for different number of histograms and different histogram resolution

Features: Chord's size Histogram			
Classifier: Multi Layer Perceptron			
Number of Histograms	Histogram Resolution		
	4	5	7
8	97.5%	97.3%	96.5%
16	95.6%	96.9%	97.9%
20	98.1%	97.9%	98.7%
24	97.9%	97.1%	98.3%
28	96.7%	97.7%	97.1%
32	97.9%	97.1%	96.9%
36	97.9%	98.1%	98.1%

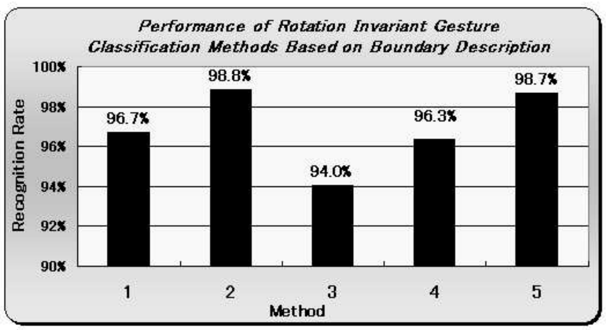

Fig. 8 Comparison between 5 different rotation invariant experiments described in Table 1

4. CONCLUSION

We presented a rotation invariant approach gesture recognition system using boundary chord's size histograms for features description and multilayer perceptron (MLP) neural network as classifier. The system has fulfilled three significant requirements as a recognition system: computational time, memory requirement and classification accuracy.

A fast search method for the boundary chain start-point definition was proposed. Results showed it to be robust in the gesture posture classification problem.

The use of chord's size histograms features followed by the MLP classifier achieved the similar performance to that was obtained by chord size feature and DP matching. However, the use of neural network has advantages in the required memory and the substantial decrement of computational time. These are important advantages to consider in large-scale problems with several numbers of classes.

As the future work, we will evaluate the performance of the histogram-based features when the hand posture performs a non-planar rotation. Additionally, the histogram representation will be improved by including internal features of the objects, aiming to give more significant information for gesture discrimination.

5. ACKNOWLEDGEMENT

The first author is supported by a scholarship from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

6. REFERENCES

[1] K-H. Jo, Y. Kuno, and Y. Shirai, "Manipulative Hand Gesture Recognition Using Task Knowledge for Human Computer Interaction", *Proceedings of the 3th International Conference on Automatic Face and Gesture Recognition*, pp. 468-473, Japan, 1998.

[2] Y. Iwai, T. Hara, N. Yachida, "Gesture Recognition based on Subspace Method and Hidden Markov Model", *Proceedings of the IROS 97*, Grenoble, France, 1997.

[3] M.V.Lamar, Md.S.Bhuiyan, A.Iwata, "Hand Alphabet Recognition using Principal Component Analysis and Neural Networks", *Proc. of International Joint Conference on Neural Networks*, Vol.4, pp.2839-2844, Washington, USA, July 1999.

[4] James Davis, Mubarak Shah, "Visual Gesture Recognition", *Vision, Image and Signal Processing*, April 1994.

[5] Hendrik Birk, Thomas B. Moeslund, Claus B. Madsen, "Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis", *10th Scandinavian Conference on Image Analysis*, Lappeenranta, Finland, 1997

[6] A. Lanitis, C. J. Taylor, T.F. Cootes and T. Ahmed, "Automatic Interpretation of Human Faces and Hand Gestures Using Flexible Models", *Proc. Intl. Workshop on Automatic Face and Gesture-Recognition*, pp.98-103, 1995.

[7] Gonzales, Rafael C. *Digital Image Processing*, Addison-Wesley Publishing Company, New York, 1994.

[8] Gupta, Lalit. Suwei Ma, "Gesture-Based Interaction and Communication: Automated Classification of Hand Gesture Contours", *IEEE Transactions on Systems, man and Cybernetics*, PartC: Applications and Reviews, Vol.31, No. 1 February 2001.

[9] Jie Yang, Alex Waibel, "A real-time face tracker", Proceedings of WACV, Sarasota, Florida, 1996.

[10] William K. Pratt, *Digital Image Processing*, Wiley-Interscience, New York, 1991.

[11] Christopher M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, United Kingdom, 1996

[12] N.Hagita, S.Naito, and I.Masuda, "Chinese Character Recognition by Peripheral Direction Contributivity Feature", *IEICE Trans. on Information and Systems*, Vol.J66-D, No.10, pp.1185-1192, Oct.1983

[13] Brigham, E. Oran. *The Fast Fourier Transform*, Prentice-Hall, Inc.,Englewood Cliffs, USA, 1974.