DYNAMIC HAND GESTURE RECOGNITION SYSTEM USING NEURAL NETWORK

Chitralekha Mahanta, T. Srinivas Yadav and Hemanta Medhi
Department of Electronics and Communication Engineering, Indian Institute of Technology Guwahati
781039 Guwahati, India

Keywords: MPEG-7 art shape descriptor, Radial basis function, Particle filter.

Abstract: Vision-based hand gesture recognition enabling computers to understand hand gestures as humans do is an important technology for intelligent human computer interaction. In this paper, a recognition system for dynamic hand gestures is proposed. In dynamic hand gesture recognition, hand is segmented by using background subtraction method. MPEG-7 ART based shape descriptors are used to extract spatial information. Our approach is based on particle filter to extract trajectory features. After collecting suitable features, Radial Basis Function neural network is used for classification. Gesture recognition rate is in the range of 80% to 98%.

1 INTRODUCTION

Gesture recognition is an important area of research in the field of computer vision. Gesture recognition pertains to recognizing meaningful expressions of motion by a human, involving the hands, arms, face, head or body with the intention of conveying meaningful information or interacting with the environment. In addition to the theoretical aspects, any practical implementation of gesture recognition typically requires the use of different imaging and tracking devices or gadgets.

Gestures can be static (the user assumes a certain pose) or dynamic (with prestroke, stroke, and poststroke phases). Direct use of the hand as an input device is an attractive method for providing natural human computer interaction (HCI)(Pavlovic et al., 1997). Vision-based techniques, while overcoming this constraint, need to contend with other problems related to occlusion of parts of the user's body. Vision-based devices can handle properties such as texture and color for analyzing a gesture, while tracking devices cannot. Vision-based techniques can also vary among themselves in: the number of cameras used, their speed and latency, the structure of environment (restrictions such as lighting or speed of movement), any user requirements (whether user must wear anything special), the low-level features used (edges, regions, silhouettes, moments, Histograms), whether 2-D or 3-D representation is used. Gesture recognition has wide-ranging applications in human-computer interaction, sign-language communication, video surveillance, dance/video annotations, forensic identification and the likes.

The paper is organized as follows: In Section 2, dynamic hand gesture recognition system is introduced. Simulation results are shown in Section 3. Finally our work is concluded in Section 4.

2 DYNAMIC HAND GESTURE RECOGNITION SYSTEM

This section is concerned with the recognition of dynamic hand gestures. Since the movement of the hand conveys important information about the message users try to communicate, in addition to the spatial information, temporal features which represent the motion of the hand are necessary for recognizing dynamic hand gestures. The recognition strategy as shown in Figure 1 uses a combination of static shape recognition (performed using MPEG-7 ART shape descriptors), particle filter based hand tracking and a neural network based classifier. For temporal information, the hand is tracked by using the particle filter(Ionescu et al., 2005). After obtaining temporal information, integration of tracking algorithm and neural networks is done for recognizing dynamic hand gestures, which are varying in global motions.

Mahanta C., Srinivas Yadav T. and Medhi H. (2011).
DYNAMIC HAND GESTURE RECOGNITION SYSTEM USING NEURAL NETWORK.
In *Proceedings of the 1st International Conference on Pervasive and Embedded Computing and Communication Systems*, pages 253-256
DOI: 10.5220/0003299102530256

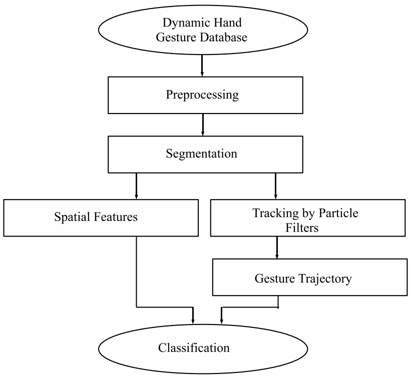


Figure 1: Dynamic Hand Gesture Recognition System.

2.1 Preprocessing

The main aim of preprocessing is to eliminate noise present in the captured images. Noise creates problems in hand segmentation and tracking phase, and hence reduces the performance of the overall system. Mostly salt and pepper noise and additive Gaussian noise are present (Gonzalez and Woods, 2002) in the images. These types of noises can be removed by using median filter and mean filter respectively.

We have collected 20 dynamic gestures which are captured by using Panasonic handycam having 3.1 mega pixel picture quality and have been chosen for our experiments. Among them, 8 are ASL gestures and 12 are control commands. Here the preprocessing is done frame by frame. Some of the dynamic gestures which are collected from (Lee and Kim, 1999) are shown in Figure 2.

2.2 Segmentation

In order to realize high-precision hand gesture recognition, it is necessary that the hand region is extracted correctly from the background. The main goal of this step is to segment the moving hand from the background. Hand region extraction method is implemented by using background subtraction algorithm. One major advantage of this method is that it extracts the gesture region without future frame images.

- **Initialization of "Target Image" and "Background Image".** Here "Target image" is defined as an image from which the gesture region is extracted at the present moment, "Background image" is estimated by eliminating the moving object. First, 1st frame image is used as background image. We are going to extract the gesture region from the images on and after 2nd frame, and thus 2nd frame image is used as the target image.

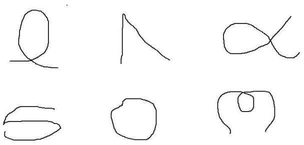


Figure 2: Gesture Moving Patterns.

- **Frame Difference.** Image of a candidate of the gesture region is calculated by the following equation. The difference between the "Target image" and "Background image" is calculated and segmented, and this can be written as

$$F(x,y) = |T(x,y) - B(x,y)| \tag{1}$$

$$FD(x,y) = \begin{cases} 255 \quad if \quad F(x,y) \geq Th \\ 0 \qquad\qquad otherwise \end{cases} \tag{2}$$

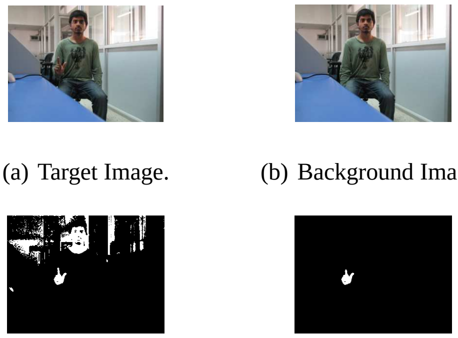
where $T(x,y)$ represents the Y value of the pixel with position (x,y) in target image, $B(x,y)$ represents the Y value of the pixel with position (x,y) in the background image, and Th is the threshold value to adjust the sensitivity of detecting the candidate of the gesture region. In our case, the threshold value equals to the maximum intensity value divided by 2.

- **Division of Gesture Region.** The gesture image obtained may contain other objects. But we need only the hand part. For this, at first edge image is calculated from the target image through the process of edge detection. After obtaining the boundary line, the image of the candidate of the gesture region is divided into two or more regions.

- **Gesture Region.** Now multiply the original image with the image obtained in the above step, then the gesture region is extracted by considering the regions in which 2/3 or more is flesh-colored pixel. Later we have removed small regions which are caused by the minute movement which is not regarded as gesture methods which were used for the static case. Some of the results are shown in Figure 3

2.3 Feature Extraction

Feature extraction is a crucial module in any Computer vision (CV) system. The implementation of this module has a considerable effect on the robustness of the system. For recognizing global hand motions, in addition to the spatial information of the hand, temporal features are also necessary. Tracking is used to



(a) Target Image.

(b) Background Image.

(c) Division of Gesture region.

(d) Gesture region.

Figure 3: Different Stages of Segmentation Algorithm.

obtain these motion features of a dynamic hand gesture video. It is the process of estimating the trajectory of hand, as the hand moves in the area of interest. The tracking algorithm has to estimate the state of the system at any time instant, given a set of observations.

2.3.1 Spatial Features using MPEG-7 Shape Descriptors (Azhar and Amer, 2008)

For spatial information, MPEG-7 ART shape descriptors are used. The visual part of MPEG-7 standard defines three descriptors with different properties: region based, contour based and 3D spectrum shape descriptors.

Region based shape descriptors are used here. These descriptors take into account all pixels constituting the shape, boundary and inner pixels. Conceptually, a descriptor works by decomposing the shape into a number of orthogonal 2-D basis functions, defined by Angular Radial Transform (ART). The MPEG-7 ART descriptor employs a complex Angular Radial Transformation defined on a unit disk in polar coordinates.

2.3.2 Tracking of Hand by Particle Filters (Arulampalam et al., 2002)

Tracking objects efficiently and robustly in complex environment is a challenging issue in computer vision. Often dynamic image frames of these hand regions are tracked to generate suitable features. Particle-filtering-based tracking and its applications in gesture recognition systems became popular very recently. The key idea is to represent probability densities by a set of samples. As a result, it has the ability to represent a wide range of probability densities, allowing real-time estimation of nonlinear,non-Gaussian dynamic systems.

2.4 Classification and Recognition

Spatial feature vectors and features obtained from gesture trajectory are given as input to the neural network classifier to classify different dynamic gestures. Radial basis function (RBFs) neural network(Michie et al., 1974) is used for classification. A total of 20 dynamic hand gestures have been chosen for our experiments. Among them, 8 belong to ASL gestures and 12 belong to control commands.

A radial basis function network is an artificial neural network that uses radial basis function as the activation function. Radial basis function (RBF) networks typically have three layers: an input layer, a hidden layer with RBF activation function and an output layer. The hidden units provide a set of functions that constitute an arbitrary basis for the input patterns (vectors) when they are expanded into the hidden-unit space; these functions are called radial-basis functions. Different types of radial basis functions could be used, but the most common is the Gaussian function.

3 SIMULATION RESULTS

In this paper we discussed the dynamic hand gesture recognition system using radial basis function neural network. A total of twenty dynamic hand gestures have been chosen for our experiments. The complete system works at a frame rate of about 25 frames/s. 20 dynamic gestures are captured by using Panasonic handycam having 3.1 mega pixel picture quality. Among them, 8 are ASL Gestures and 12 are control commands as shown in Figure 2. The hand region is extracted by using background subtraction algorithm. One major advantage of this method is that it extracts the gesture region without future frame images. Spatial features are extracted by using MPEG-7 shape descriptors. For temporal information, the hand is tracked by using particle filter. Trajectory features are extracted only for key points. Spatial and trajectory features are combined and given as inputs to the classifier. Radial basis function neural network is used as classifier. 8 ASL gestures are used for testing. Table 1 presents the results of different dynamic gestures using RBFs. Table 2 presents the results of 8 ASL gesture recognition using RBFs. Recognition rates ranged from 80% to 98%.

Table 1: Recognition rates(%) of dynamic gestures using RBFs.

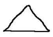
Moving Pattern	No. of attempts	Misclassified	Recognition rate(%)
	20	1	95
	20	2	90
	20	2	90
	20	3	85
	20	4	80
	20	1	95
	20	1	95
	20	4	80
	20	2	90
	20	3	85
	20	4	80
	20	1	95

Table 2: Recognition rates(%) of ASL gestures using RBFs.

S. No.	Gestures	Classified	Misclassified	Recognition rate(%)
1	East	55	1	98.18
2	Up	52	1	98.08
3	Away	48	2	95.29
4	Circle	54	1	97.14
5	Down	48	2	95.3
6	Admonish	55	1	98.18
7	Adrift	53	3	85.4
8	West	40	8	80

4 CONCLUSIONS

Dynamic hand gesture recognition using neural network is presented in this paper. The hand region is extracted by using background subtraction algorithm. Video sequences are collected from a stationary camera with complex backgrounds. Spatial features are extracted by using MPEG-7 shape descriptor. Particle filter is used to track a hand under complex background. After obtaining the temporal information, the tracking algorithm is integrated with the neural network for recognizing the dynamic hand gestures, which are varying in global motions. RBFs are used for classification of different dynamic hand gestures. Recognition rates range from 80% to 98%.

REFERENCES

Arulampalam, M. S., Maskell, S. S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear non gaussian bayesian tracking. In *IEEE Trans. Signal Processing*.

Azhar, H. and Amer, A. (2008). Chaos and mpeg-7 based feature vector for video object classification. In *IEEE Trans. Pattern Classification*.

Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Prentice Hall of India, India, 2nd edition.

Ionescu, B., Coquin, D., and Lamber, P. (2005). Dynamic hand gesture recognition using the skeleton of the hand. In *EURASIP Journal on Applied Signal Processing*.

Lee, H.-K. and Kim, J. H. (1999). An hmm-based threshold model approach for gesture recognition. In *IEEE Trans. Pattern Analysis and Machine Intelligence*.

Michie, D., Spiegelhalter, D. J., and Taylor, C. C. (1974). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood.

Pavlovic, V. I., Sharma, R., and Huang, T. S. (1997). Visual interpretation of hand gestures for human-computer interaction: A review. In *IEEE Trans. Pattern Analysis and Machine Intellegence*.