International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.2 (2014), pp.295-304
http://dx.doi.org/10.14257/ijsip.2014.7.2.27

Robotic Control using Hand Gesture Recognition baesd on FFT Analysis

Akshay Agarwal, Neha Singh and G. C. Nandi

Robotics & Artificial Intelligence Lab
Indian Institute of Information Technology
Allahabad, India
{akshay.agarwal854, nehasingh11091989, gcnandi}@gmail.com

Abstract

The study of gesture is still in its early childhood, gesture is present early in evolution, and is used to communicate before a child has the ability to enunciate. This paper describes the use of fast Fourier transforms both for preprocessing and feature extraction of hand gestures images. First the input gesture images are padded as the padding is necessary before applying Fourier transform to the image. Directional derivatives are then applied to the real part of the filtered images. Derivative angles are divided into the bins. These bins provide the final feature vectors. Distance based techniques were used for gesture classification. The real time simulation software WEBOTS is used for performing the classified gesture.

Keywords: *Gesture, Fast fourier transform; Distance metric; Webots robotic simulation software, HOAP-2 humanoid robot*

1. Introduction

Pattern recognition and gesture recognition plays a very vital role in today's research. Similarities of the hand shape in gesture; this paper aims to develop a real time hand gesture recognition system. Indian sign language and American Sign Language are the very famous examples for hand gesture recognition. Dynamic gesture has a potential application in human computer interaction. Dynamic gestures are a continuous motion of hand. Sensing of human emotion showing gesture is very important. In today's time more studies has done on gesture recognition. We aim to develop a gesture recognition system for the control of various household applications like TV, fridges, AC *etc*. American Sign Language is very old and famous form of gesture. It is very famous among English-speaking person like USA, Canada. American Sign Language has its own features and it was basically generated from French sign language. FSL is the very vast sign language committee. FSL have 50,000 to 100,000 signers approx. Indian Sign Language has its own grammatical and shape information. As ISL in concerned it consists of both static and dynamic hand gestures. Example, if a person wants to convey the message goes down. For this type message we want to perform the hand gestures using both the hand in a continuous manner. ISL is not the official sign language in India in spite of having million of deaf and dumb people. A lot of method has been proposed for recognizing the gesture. Previous methods the magnetic sensor, markers to separate the background from the hand gesture image. Extraction of visual information in the form of feature vector is very important in vision based recognition. In this paper we focus on the strategy to convert the image in Fourier domain then extract the feature vector from the hand gesture videos. In this paper we propose a method for gesture recognition using Fourier transform. For all REAL (as opposed to IMAGINARY or COMPLEX) images, the FT is

ISSN: 2005-4254 IJSIP

symmetrical about the origin so the 1st and 3rd quadrants are the same and the 2nd and 4th quadrants are the same. If the image is symmetrical about the x-axis (as the cosine images are) 4-fold symmetry results. Today's scenario demands a recognition system that translates the hand gesture for human computer interaction. Various similarity measurement techniques are then applied for gesture classification like Euclidean distance, Bhattacharya distance etc. Paper is divided into the following sections. Section II describes the related which is already done in this field. Section III describes the mechanism of ISL gesture acquisition. The Next section explores the gesture preprocessing technique with proper feature selection followed by evaluation technique. Section V expresses the recognition techniques with Bhattacharyya distance estimation among all the samples of ISL gestures. The Next section demonstrates the mimicry by the humanoid robot HOAP-2. Section VII presents the experimental results. The conclusion including future work has been attached in section VIII. References are included at the end of this paper.

2. Related Work

Tremendous amount of work has already been done for the recognition. The KHU-1 data glove based approach has been used for 3D hand motion and tracking and recognition of hand gestures. The recognition of Chinese Sign Language using a data glove based system has been done using only static symbols. Ishikawa M, Matsumura H, have proposed a recognition system based on a self organization method. The entire recognition process has been followed by acute measurement of finger joints using a data glove, extracting the gestures out from a sequence of hand shapes and proper aligning of data length. The most elegant process for recognition of Malay Sign Language using wireless data gloves is described. A data glove is used for recognition of continuous gesture considering huge number of vocabulary in Taiwanese Sign Language (TWL). The key glove concept presents a new input device for human hand motion recognition. A novel approach is being used for classification of Croatian sign language using K Nearest Neighbor method with Dynamic Time Wrapping (DTW) and Longest Common Subsequence (LCSS) for similarity measurement. In our approach dynamic hand gesture recognition has been done considering both hands simultaneously. It is the composition of spatio-temporal signal which could be analyzed in an elegant manner. The synthesis of motion data extracts salient features from hand motion signal for recognition purposes. A single video camera based recognition system deals with the recognition of German Sign Language with considering hand arm motion. Unlike the data glove based, vision based gesture recognition for human robot interaction has been entertained effectively. Another approach deals with the recognition of Chinese Sign Language towards implementation of human robot interaction. A probabilistic framework is applied for recognition and reconstruction of gesture for humanoid robot which is composed by PCA, ICA and HMM techniques. Kinesthetic demonstration by a human operator enables the humanoid robot to learn the gestures. Humanoid robot can be trained by applying imitation learning techniques. The portrait drawing is a challenging task performed by humanoid robot HOAP-2.

Table 1. Classification of Hand Gesture

Category	Type
Task	Pointing, Tracking, Gesture
Motion	Static, Dynamic
Input Information	Image by Camera
Image Feature	Natural image, Marked image
Dimension of data	2-d, 3-d
Used hand	One hand, Two hand
In view of psychology	Symbolic, Kinetographic, Iconic

3. ISL Dataset Construction

Background removal is very complex and computationally difficult work to perform. We used constant background while capturing the video and using different sizes at the rate of 30 frames per second. All the ISL gestures include various kinds of hand motions. A capturing device SONY handy cam with 2.5 mega pixel resolutions is used for capturing videos of several ISL gestures. In recognition purposes, limited specific dynamic ISL gestures have been considered as shown in Figure 1 under various light illumination conditions. One elementary approach for image processing tends to background uniformity where a dark background is chosen for dealing with gray scale images effectively. To have a controlled environment, the background uniformity has been kept while recording the videos in real time.

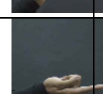
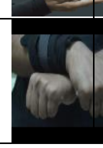
The definition of the Fourier transform is:-

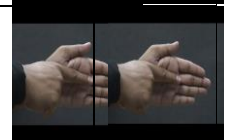
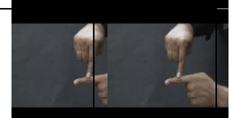
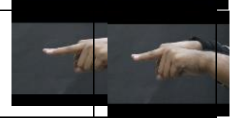
Fourier transform considers both the phase and amplitude in the image. Padding is necessary before applying the Fourier transform. We first pad the input image with zeros. Fourier transform of the image converts the input image into the complex image. The idea behind using the Fourier transform is that we can reconstruct the original waveform from the sum of the cosine and sine series.

$$Phase(F) = \text{atan}\,[Image(F)/Real(F)].$$

$$Magnitude(F) = \sqrt{Real(F)^2 + Imag(F)^2}.$$

Table 2 ISL dynamic and static gestures under different light illumination conditions

ISL	Training	Testing	start	end
Alone	25	10		
Arise	25	10		
Prisoner	25	10		

Prisoner	25	10	
Beside	25	10	
Flag	25	10	
Ascend	25	10	
Middle	25	10	
Marry	25	10	

4. Proposed Methodology

Steps of algorithm:-

Step 1:- Convert the RGB image frames into the grayscale frames.

Step 2:- Resize the image into the 150*140. Resizing helps in reducing the computation.

Step3:- Perform the padding operation on the input frames.

Step 4:- Apply the Fourier transform on the padded image frames.

Step 5:- Construct the filter function of the same size as input image.

Filter function may be low-pass, high-pass or the notch filter.

Step 6:- Apply the filter to the Fourier transformed image.

Step 7:- obtain the real part from this image frames. Finally crop the image to undo the padding function.

Now generate the angle histogram feature vector from the real image obtain after the Fourier domain filtering. Convolve the image with the spatial domain filters. Generate the angle histogram feature using the atan2(y,x) formulae. Where y is the convolution in the y direction and x is the convolution in x direction.

Step 8:- Create the feature vector F= (ahs1…….. ahs18) for all i, j, k. where i=number of classes, j= number of samples in the class and k= number of frames within a sample.

Step 9:- Find the mean of each feature vector.

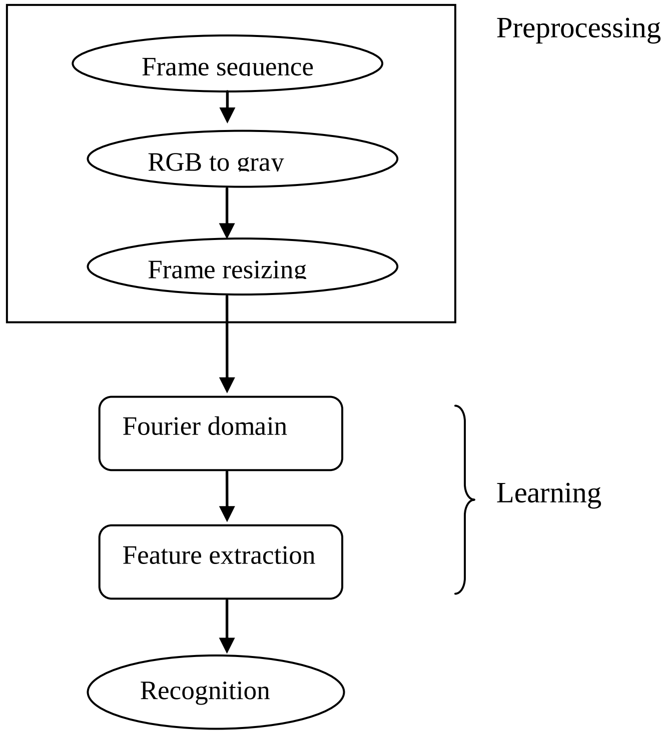


Figure 1. Methodology

Recognition algorithm steps:-

Step 1:- Apply all the steps explained above till the feature vector calculation. Let T= (aht1…… aht18) for all k.

Step 2:- Find the minimum distance using

$$D_{x,y} = \sqrt{\sum_{i=1}^{j} (d_{xi}^2 - d_{yi}^2)}$$

Between vector F and T.

Step 3:- Classify the gesture according to the minimum distance obtained from each training feature vector corresponding to a particular gesture.

3 basis kind of filters are famous in the Fourier domain I) low-pass filter, II) High-pass filter and III) notch filter. Low-pass filter sometime is also known as smoothing filter. It creates the blurred mage. The big advantage of using the Fourier is the fast computation of Fourier transform on the big images. High-pass filter is also called the sharpening filter. Notch filter is used to remove the repetitive spectral noise.

We have trained our gesture recognition system using twenty ISL hand gestures some of them are shown in the Figure 1. It is easy and very fast to apply the Fourier transform on the input gesture frames. Fourier calculation is computationally effective and fast on large scale images too. Advantage of using directional histogram as a feature is the efficiency in calculating the direction edge gradient derivative.

5. Webots and Hoap-2 Description

Webots is a programming and simulation software develops to program, simulate the mobile robot. Webots provides the environment to develop the complex robots and simulate the robot. Mobile robotics is a very demanding and complex research areas now a days. Webots have many facilities regarding the development the mobile robot like shape, sensor, texture *etc*. The robot controller can be tested on the given IDE. We require the real world to be compatible with the robotic environment. The joint angle value corresponding to different joints of a human are mapped to the robotic domain and write to the file. The file is called the CSV (comma separated values).The CSV file is the transferred to the robot. Webots provides the different programming language to simulate the humanoid robot to perform the desired function. HOAP is the abbreviation of the form Humanoid Open Architecture Platform. This is the very famous robot platform develop in Japan. The height and weight of the HOAP-2 robot is 48 cm and 6.8 kg respectively. HOAP-2 has a robotic body, PC and power supplies. The PC OS of the HOAP-2 work on RT-Linux. The controller program used for the simulation of the robot is used the CSV corresponding the different joint angle values for each joint. CSV file have 27 column belongs to the joints of the HOAP-2.

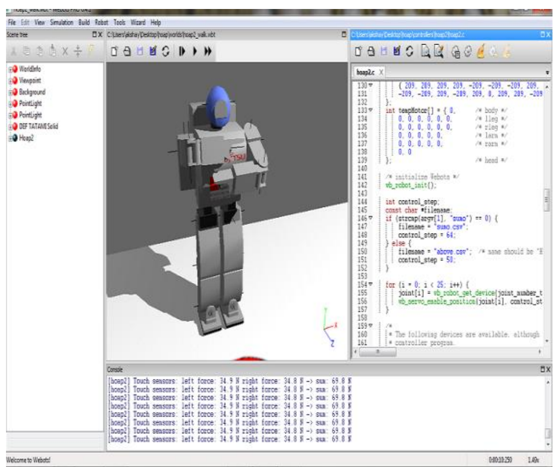


Figure 2. HOAP-2 Performing Gesture on Webots

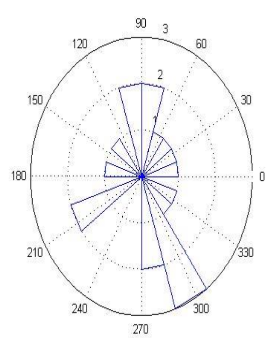


Figure 3. Polar PLot of BINs for the above GESTURE

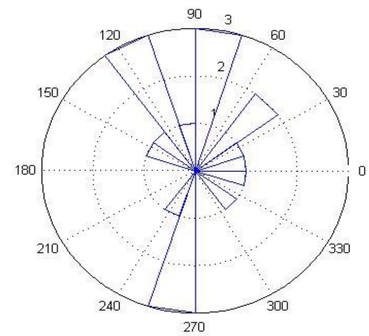


Figure 4. Polar Plot of Bins for the Advance Gesture

6. Classification and Result

The above defined steps in the methodology section are performed for the classification of the hand gesture dataset. Different distance measurement techniques are applied to calculate the distance between the trained gesture and test gesture. The techniques which we are adopted are Euclidean distance, City-block distance, Minkowski and Chebychev *etc*.

$$D_{st} = \sum_{j=1}^{n} |x_{sj} - y_{tj}|$$

City-block distance formulae

$$D_{st} = \sqrt[p]{\sum_{j=1}^{n} |x_{sj} - y_{tj}|}$$

Minkowski distance formulae

Table 3. Confusion Matrix of Result

Gesture	Ab ove	Acr oss	Adva nce	Afrai d	All_g one	All
Above	10	0	0	0	0	0
Across	0	9	0	0	1	0
Advance	0	0	10	0	0	0
Afraid	0	0	0	10	0	0
All gone	1	0	0	0	9	0
All	0	1	0	0	0	9

Where s, t are two points. Calculate distance among all the training classes and select the minimum as the matching class. Some gestures are correctly classified in the trained gesture. The proposed method gives the 99.24% accuracy over defined dataset of Indian sign language. For testing we applied the above defined distance measurement formulas.

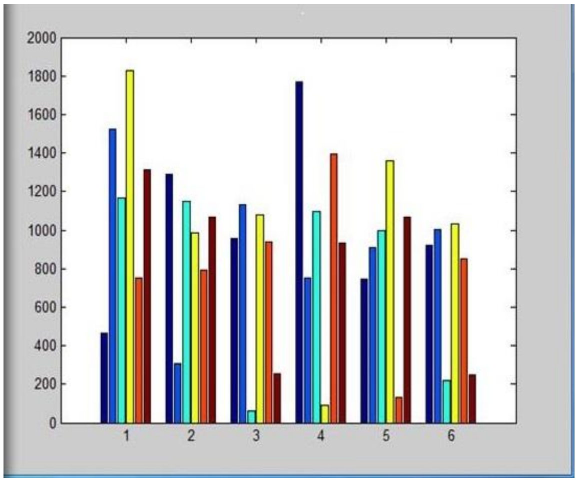


Figure 5.a.

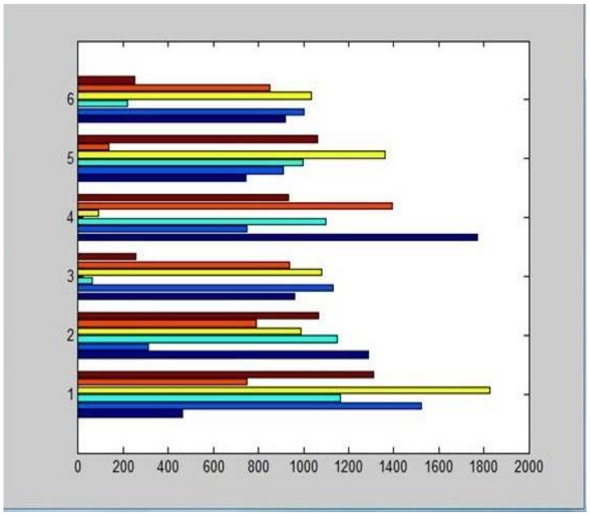


Figure 5.b.

Figure 5. Bar-charts Showing Distances

In the confusion matrix (table 2.) rows corresponds to the actual class and the column corresponds to the predicted class. In this paper we proposed the 18 bins division of the angle histogram of the gesture frames.

7. Conclusion and Future Work

In this paper we adopted the Fourier domain transformation technique for feature extraction. In the future work we will try to divide the angle data feature into 36 bins and test the result. Later we will work in discrete cosine transformation to extract the more relevant features from the gesture frames. The DCT is the acronym of discrete cosine Transformation. The result after applying the DCT to the image can be used as a feature vector. Discrete wavelet transformation is also the good techniques to find the relevant feature from the image. DWT considers both the phase and amplitude information of the image. Now a day's DWT transformation become a more advanced research areas for the image and video processing.

References

[1] Y. Huang, Z. Wu, L. Wang and T. Tan, "Feature Coding in Image Classification: A Comprehensive Study", Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 36, no. 3, **(2014)** March, pp. 493-506.

[2] S. Bhattacharya, B. Czejdo and N. Perez, "Gesture classification with machine learning using Kinect sensor data", Emerging Applications of Information Technology (EAIT), 2012 Third International Conference, **(2012)** November 30-December 1, pp. 348, 351.

[3] H. Venkateswara, V. N. Balasubramanian, P. Lade and S. Panchanathan, "Multiresolution Match Kernels for gesture video classification", Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference, pp. 1, 4, **(2013)**, July 15-19.

[4] K. Mey Chew, C. Yee Yong, R. Sudirman and N. H. Mahmood, "Left and right hand gesture classification using scatter diagram and Principle Component Analysis", Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference, **(2012)** December 17-19, pp. 795, 799.

[5] N. Amani, A. Shahbahrami and M. Nahvi, "A New Approach for Face Image Enhancement and Recognition", International Journal of Advanced Science & Technology, vol. 52, **(2013)**.

[6] K. Ruba Soundar and K. Murugesan, "Preserving global and local information-a combined approach for recognising face images", Computer Vision, IET, vol. 4, no. 3, **(2010)**, pp. 173-182.

[7] V. B. Semwal, S. A. Katiyar, P. Chakraborty and G. C. Nandi, "Biped model based on human Gait pattern parameters for sagittal plane movement", Control, Automation, Robotics and Embedded Systems (CARE), 2013 International Conference on, **(2013)** December 16-18, pp. 1, 5.

[8] V. B. Semwal, A. Bhushan and G. C. Nandi, "Study of humanoid Push recovery based on experiments", Control, Automation, Robotics and Embedded Systems (CARE), 2013 International Conference, **2013)** December 16-18, pp. 1, 6.

[9] M. Panwar, "Hand gesture recognition based on shape parameters", Computing, Communication and Applications (ICCCA), 2012 International Conference, **(2012)** February 22-24, pp. 1, 6.

[10] D. Mazumdar, A. K. Talukdar and K. K. Sarma, "Gloved and free hand tracking based hand gesture recognition", Emerging Trends and Applications in Computer Science (ICETACS), 2013 1st International Conference, **(2013)** September 13-14, pp. 197, 202.

[11] S. Shiravandi, M. Rahmati and F. Mahmoudi, "Hand gestures recognition using dynamic Bayesian networks", AI & Robotics and 5th RoboCup Iran Open International Symposium (RIOS), 2013 3rd Joint Conference, **(2013)** April 8, pp. 1, 6.