International Journal of Engineering Research in Electronic and Communication Engineering (IJERECE) Vol 3, Issue 4, April 2016

FPGA Design For Location Independent Hand Posture Recognition System

[1] Aleena Lal, [2] Anas A S
[1] PG Student [VLSI & ES] [2] Assistant professor,
Department of ECE, TKM Institute of Technology, Kollam
[1] aleenalal99@gmail.com [2] 7anas5@gmail.com

***Abstract*:-- In the present era of smart electronics, hand sign recognition has a significant role. A hand sign recognition algorithm based system employing a hybrid SOM-Hebb network is designed. The input posture image is pre-processed and feature vectors are extracted from them. These are then mapped on to lower dimensional neurons on the self-organizing map (SOM). The input image is compared with the hand posture images in the database. The algorithm offers a stagnant response irrespective of the change in location of the input image. The work is to develop a design that supports real- time FPGA implementation using onboard camera. The set of default images in database used as the current input. The input image is pre-processed and converted to a binary image. The feature vectors are extracted using Discrete Fourier Transform (DFT) and fed to the hybrid classifier network for recognition. FPGA implementation enhances its use in portable embedded applications. The system offers better recognition accuracy and faster response. The system is designed for robustness against change in location of input image. The algorithm is coded using VHDL and simulated in Xilinx ISE 13.2 and MATLAB 2013.**

***Index Terms*— Discrete Fourier Transform (DFT), Field Programmable Gate Array (FPGA), Hebbian learning, Self-Organizing Map (SOM).**

I. INTRODUCTION

The influx of smart devices in the society has enhanced the significance of Human Computer Interaction (HCI) in the routine of mankind. With the massive development of ubiquitous computing, current user interaction approaches with keyboard, mouse and pen are not sufficient. The applications on PC platform like interactive entertainments and augmented reality requires more natural and intuitive interface. The relatively small size of mobile and hand held devices leads to limited input space as well as encumbered experience with tiny keyboard or touch screen. Hand gestures are frequently used in people's daily life. It provides a natural and intuitive communication modality for human-computer interaction. Hence the use of hand gesture is an attractive alternative to cumbersome interface devices for HCI.

Gestures recognition is a powerful and simple means of communication among humans. Gesture recognition involves interpretation of human gestures using mathematical algorithms. The real-time processing of hand sign for HCI is an essential feature. It requires a powerful full specification PC which is not feasible in real-time embedded applications. Field-programmable gate arrays (FPGAs) are key components in implementing high performance digital signal processing (DSP) systems particularly in the areas of digital communication, networking, video and imaging. The focus is on a FPGA based approach to develop a posture recognition algorithm suitable for hardware architecture. The gesture recognition can be either dynamic or static posture recognition. The focus here is on a static hand gesture / hand posture recognition algorithm. The key issues in posture recognition are in terms of real-time performance, recognition accuracy, transformations, complex background as well as varying locations. The approaches can be either data-glove based or vision based hand sign recognition. The data-glove based methods use sensor devices for digitizing hand and finger motions into multi-parametric data. These devices are quite expensive and bring much tedious experience to the users. On the other hand, the vision based methods require only a camera, thus realizing a natural human-computer interaction without using any extra devices. The artificial vision systems that are implemented in software and/or hardware complement the biological vision. The vision based recognition can be appearance-based or 3D hand model based. The current recognition approach is appearance based hand tracking which also uses colored gloves in-order to make image processing task easier.

The input hand images are pre-processed through horizontal/vertical projections that obtain horizontal/vertical histograms followed by Discrete Fourier Transforms (DFTs). DFTs calculate the magnitude spectrum of the histograms

and the magnitude spectrum is used as the feature vector. The use of magnitude spectrum makes the system robust to change in position of the hand posture image. A hybrid network based classifier is applied to the hand sign recognition system. The hybrid network is made of SOM (Self-Organizing Map) and Hebbian learning network.

The system is based on a hybrid network classifier to identify the different posture class. The SOM-Hebb classifier yields the finest performance depending on the number of classes. SOM uses an unsupervised learning algorithm. It performs a non-linear mapping from a given higher-dimensional input space to a lower-dimensional map of neurons. It is necessary to interpret the obtained mapping properly in order to make the final classification accurate. The hybrid network employs a single-layer feed-forward neural network to interpret the mapping done by SOM. The network is trained with Hebbian learning algorithm so as to perform category acquisition. The classifier is trained with 24 American Sign Language hand signs used as the database. The posture recognition system is designed for FPGA based implementation

.

II. LITERATURE REVIEW

Gesture recognition is a promising sub-field in the discipline of computer vision. There has been extensive research done on the topic of posture recognition and several techniques were developed for effective HCI. Extensive research has been done on different posture recognition algorithm that uses an artificial neural network technique to identify the class of an image.

The hand gesture recognition approaches for HCI can be mainly classified as data-glove based and vision based techniques. The data-glove-based methods use sensor devices for digitizing hand and finger motions into multi-parametric data [11]. They are accurate and fast but very expensive. The vision-based methods only require a camera [10], thus achieving natural interaction between humans and computers. A number of approaches to the video based recognition of hand gestures has been introduced in recent years [14]. The vision-based recognition of hand gestures fall into two categories as 3D-model-based and appearance-model-based. 3D kinematic hand models calculate hand parameters, and hand postures are determined by comparing input frames and the 2D appearance projected by the 3D-hand models. The 3D models exactly describe hand movements and shapes, but are computationally expensive to use. Appearance-based models, on the other hand, directly infer gestures from visual images, extract image features to model the visual appearance of hand images, and compare these features with features extracted from video frames. Thus, appearance based approaches have advantages in the real-time processing because it is easier to extract the 2D image features. The visual appearance of hands is detected by segmenting out the hand regions from the image background .Skin colour based recognition is a popular method but it is sensitive to lighting conditions and no other skin-like objects should be in the images. Inexpensive colour-coded gloves used for hand segmentation in several systems which simplifies the filtering out of other objects and backgrounds in which unique colour coding assigned to fingers and palm.

Pre-processing and feature extraction are the two most important tasks to be done on the input images. Image binarization is a common task in most of the image processing tasks. In several applications as posture recognition, document analysis, etc., binary images can be used as input to algorithms performing useful tasks. Projection histograms [2], [5], [6] are suitable means of extracting unique and powerful feature vectors in the pre-processing. A plot on the pixel distribution in horizontal and vertical directions of an image gives the projection histogram. It converts the original 2-D image signals to one-dimensional signals for further processing.

Principal component analysis (PCA) is a classical form of statistical feature extraction and dimension reduction technique. It is very easy to apply and the features can be extracted efficiently using linear algorithms. Locality preserving projections (LPPs) enhance the discriminative power of PCA by using known or estimated similarity measures as used in Sign Language Translator [9]. Several feature extraction techniques as Scale Invariant Feature Transforms (SIFTs) accompanied by Adaboost Learning algorithm for adaptively selecting best features, Haar Transforms, etc implemented in various posture recognition systems. Discrete Fourier Transform (DFT)[12] is one of the efficient feature extraction methodology suitable for FPGA based posture recognition system. The digital signal processing tools like Discrete Fourier Transform can be used for processing the one dimensional signal.

The next area of concern is to use these extracted feature vectors for posture recognition. The extracted features are compared with the pre-trained values stored in the database. Artificial Neural Networks (ANN) plays a promising role in developing the matching algorithm in this scenario. Artificial neural networks [11] are among the most powerful learning models. They have the versatility to approximate an extended range of complex functions that represent multidimensional input-output maps. Neural networks also have an inherent adaptability, and can perform robustly in noisy environments. Self-Organizing Map(SOM)[13] signifies a class of neural network algorithms in the unsupervised-learning category. The central property of the SOM is that it forms a non-linear projection of a higher

dimensional data manifold onto a regular, low-dimensional (usually 2D) grid.

Majority of the work on hand posture recognition were done on PC based platform. The prior hand posture recognition techniques were particularly PC based and they do not support real-time embedded applications. There were several limitations in the optical hand posture recognition as in image processing, feature extraction, background variation, etc. To switch to embedded application, FPGA is a better platform. The algorithm implemented on FPGA [3], [8] functions as the hardware circuit. Also the independent modules formed by logic gates are used to accomplish the complex computational tasks. The parallelism architecture in FPGA enables real time processing [4]. Another advantage is that the use of FPGA is close to Application-Specific Integration.

III. PROPOSED SYSTEM

Hand gesture recognition system uses hand gesture for interfacing between computer and human. This is a technique for a human computer interface through hand posture recognition that recognizes 24 static gestures from the American Sign Language hand alphabet. The system compares any of the features of the hand image and finds a best match for recognition. The input hand image in the RGB

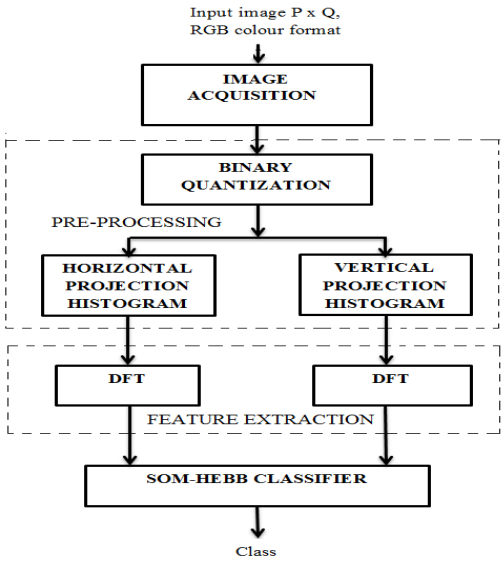

Fig. 1. Block Diagram

colour format consists of P x Q pixels. The image is pre-processed to extract the feature vectors. The pre-processing consists of binary quantization, horizontal/vertical projections that calculate histograms, followed by two discrete fourier transforms (DFTs). The preprocessing generates $F_H(n)$ and $F_V(n)$, which are the feature vectors of the input image. The D dimensional feature vectors are fed to the SOM classifier network trained by Hebbian learning algorithm that finally identifies the hand sign. The block diagram for the hand posture recognition system is shown in Fig 1.

A. Database Description

All the operations are performed on colour image. The database [1] uses hand gesture of American Sign Language (ASL) as shown in Fig 2. The set of the ASL hand signs consists of 24 static postures and two dynamic hand gestures. The letters J and Z have been discarded for their dynamic content. Since the system can only handle static images, the system is trained to recognize the 24 static hand signs, thus H=24. The size of the images taken as 256 x 256 (P = 256), and the threshold value for the binary quantization as zero (ρ=0). The dimension of the feature vector taken is 32 (D=32). The number of training iterations for the SOM includes 1024 epochs and 512 epochs for the Hebb network.

The RGB colour image of some of the hand posture images are taken and used as default input images in the database. The input image consists of P x Q pixels. The system designed demands the user to wear a glove with specified colour coding. It simplifies the filtering out of other objects and background from the input image. The system demands a glove which has the finger portions covered red and palm white. The hand region is segmented by detecting the red region. In the image acquisition part, the input hand image of P x Q pixel size in RGB format is read and resized to 256 x 256 pixel orientation. It is segmented separating the red

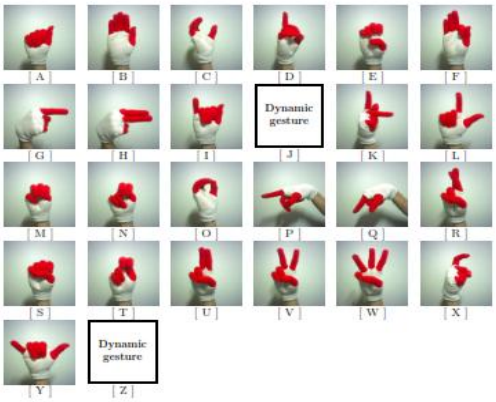

Fig. 2. American Sign Language

portion indicating the fingers using Fuzzy tool box. The R, G and B values are separated from it and converted to ASCII format so that they compatible in different platforms and stored as text file.

B. Pre-Processing

Binary Quantization: The input colour image of particular hand posture is converted to binary image before further processing. The hand region is segmented by detecting the red region as mentioned and each image pixel is quantized to a binary value. The quantization of the image is done using the following equation

$$I(x,y) = g\big(R(x,y), G(x,y) + B(x,y)\big).g\big(R(x,y),\rho\big) \quad (1)$$

where I(x, y) is the binary pixel value, and R(x, y), G(x,y) and B(x,y) are the colour component values of a pixel at the (x,y) co-ordinates. The ρ is a threshold parameter and g(x, ρ) is a threshold function. The binary quantization is done such that g(x, ρ) is 1 if x ≥ ρ and 0 otherwise.

Projection Histogram: The next step in pre-processing is to obtain the histogram values of the binarized image information. Here, the horizontal and vertical projection histograms of I(x, y) are calculated. The projection maps the binary image into 1-D array called histogram. The histogram value is obtained by summing the pixel values along a particular direction. Horizontal histogram generated by summing the pixel values in horizontal direction and vertical histogram by summing along the vertical direction. The respective horizontal projection histogram $P_H(y)$ and vertical projection histogram $P_V(x)$ are defined by

$$P_H(y) = \sum_{x=0}^{P-1} I(x,y) \quad (2)$$

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

C. Feature Extraction

After pre-processing, the next step is feature extraction which is the most vital part in any image processing applications. The feature vectors are the magnitude spectra of the horizontal as well as the vertical projection histograms of the hand image. The use of magnitude spectrum as the feature vector enables the posture recognition system to be location independent. The feature extraction is done using the Discrete Fourier Transform technique. The magnitude spectra $F_H(n)$ and $F_V(n)$ of $P_H(n)$ and $P_V(n)$ computed by activating two DFTs. Normal DFT calculation has issue of high computational cost since nonlinear functions and multipliers are used. To overcome this limitation, the $F_H(n)$ and $F_V(n)$ are calculated sequentially using two DFT circuits as shown

$$A(k) = \sum_{n=0}^{P-1} x(n)\cos\left(\frac{2\pi nk}{P}\right) \quad (4)$$

$$B(k) = \sum_{n=0}^{P-1} x(n)\sin\left(\frac{2\pi nk}{P}\right) \quad (5)$$

Each of the 256 pixel values is modelled using 16 point FFT algorithm and the filter coefficients for FFT calculation obtained using FDA Tool Box in MATLAB. The magnitude spectrum is given by

$$X(k) = \sqrt{A(k)^2 + B(k)^2} \quad (6)$$

The DFT result is used as the feature vector. The two DFT circuits calculate two frequency components which serve as input to the classifier. The magnitude spectra $F_H(n)$ and $F_V(n)$ of the same hand posture images placed in different locations are identical. They lack phase information for hand posture0 s position. Hence the objective behind the work to make the system robust to change in position of input hand postures is satisfied. Each feature vector element to be fed to the classifier is defined as

$$\xi_i = \begin{cases} F_H(i), 0 \le 1 \le (D/2) \\ F_V\big(i-(D/2)\big), (D/2) \le 1 \le (D) \end{cases} \quad (7)$$

D. SOM - Hebb Classifier

A hybrid network classifier is used in the system for posture identification. The figure outlines the SOM-Hebb vector classifier used in the system. The hybrid network consists of SOM, and a single layer feed-forward neural network trained with the Hebbian learning algorithm. The classier reads the D-dimensional vectors from pre-processing, and classifies them into H classes.

SOM: The operation of the SOM is divided into two phases. The weight map is trained with a set of input vectors in the learning phase. After that, the map is used in the recall phase. In the learning phase, the input vectors are given to SOM in multiple iterations. The distances to all weight vectors are calculated for each input vector and winner neuron is determined. The weight vector of the winner neuron has the shortest distance to the input vector.

Hebbian learning network: The Hebbian network identifies categories from the winner neuron. One winner neuron is determined from a winner search for each input vector, and the winning neuron is assigned to the input vector. From the winning neuron, the class to which the input vector belongs can be recognized. Each neuron represents a single cluster in the input vector space, and a single posture class may consist of multiple clusters. Therefore, neurons belonging to the same class must be selected and they are associated with that class. Training vectors and teaching data indicate the class of the given vectors which are sequentially fed to the network

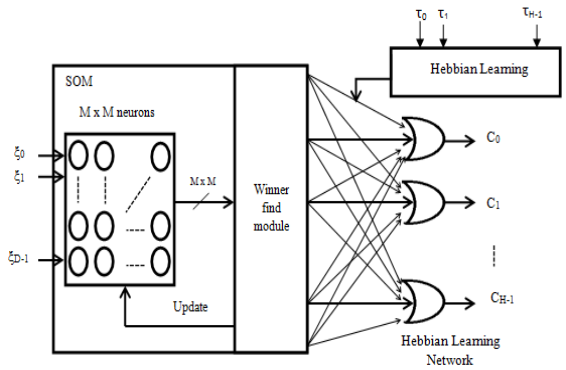

Fig. 3. SOM-Hebb Classifier

during the learning phase. A training vector makes one of the neurons a winner. Then, the winner neuron is associated with the corresponding class that is indicated by the teaching data if a strong correlation is found between them. The correlation is detected by the simultaneous activation of the neurons i and t which is counted during the training phase. If the count exceeds a threshold, then the neuron i is connected to a particular class j. A single hand posture class can be connected to multiple neurons because the class may consist of more than one cluster in the input vector space.

IV. EXPERIMENTAL RESULTS

The modules are designed using VHDL language in Xilinx ISE Design Suite 13.2 and MATLAB 2013. The simulation of the design is performed using ModelSim SE 6.2c to verify the functionality of the design.

The simulation result for a particular hand gesture is shown in Fig. 4 and the output is then exported to Matlab arena is shown in Fig. 5.

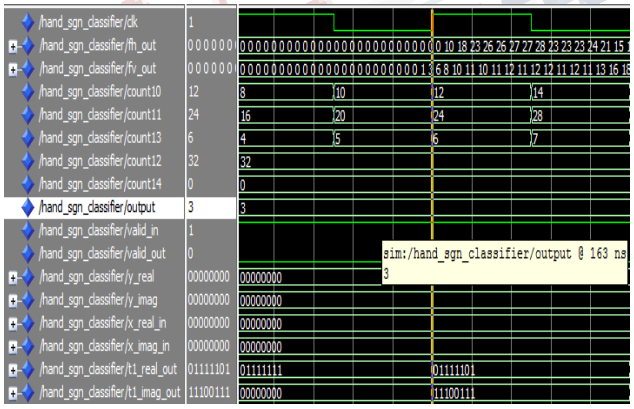

Fig. 1. Simulation result for a particular hand gesture

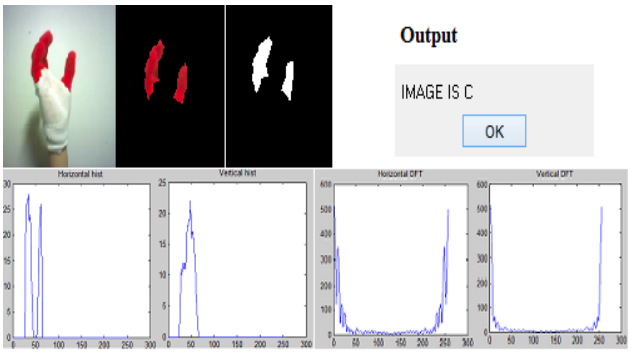

Fig. 5. MATLAB output for a particular hand gesture

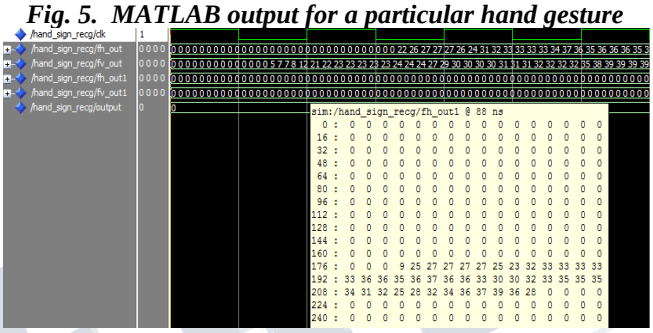

Fig. 2. Simulation result for same hand posture at two different locations

The simulation result for the same hand gesture in two different quadrants giving the same feature values irrespective of the change in location of input image is shown in Fig. 6.

V. CONCLUSION

The gesture recognition is a domain of great interest today due to its multiple application possibilities. A hand posture recognition algorithm that supports FPGA implementation for embedded real-time application is designed. The posture recognition system uses any hand sign as the input and is recognized using a hybrid network classifier. It is a vision based recognition method that uses a colour-coded glove. The system employs DFT technique for feature extraction of the input image. The feature vector is extracted from the magnitude spectrum which is invariant to the change in location of input images. This enables the recognition robust to change in location of hand signs. The hybrid classifier uses SOM-Hebb learning for class identification. The system deals with only the static hand posture recognition. The simulation results prove the accuracy of performance of the design. The algorithm offers better recognition accuracy and gives the gesture recognition irrespective of the change in location of the input image.

REFERENCES

[1] Hiroomi Hikawa, Member, IEEE, and Keishi Kaida, " Novel FPGA implementation of Hand sign recognition system with SOM-Hebb classifier", IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 1, Jan 2015.

[2] Xiaofei Zheng, Zhe Gao and Wei Luo, "Feature of statistical projection algorithm based image retrieval", AMIS Technology, vol. 7, no. 9, pp. 221229, Mar 2014.

[3] C. Li and W. Chen, "A novel FPGA-based hand gesture recognition system,Journal of Advanced Computer Science and Technology", vol. 7, no. 9, pp. 221229,May 2012.

[4] K Manikantan, S Ramachandran, Science Publishing Corporation DFT domain Feature Extraction using Edge-based Scale Normalization for Enhanced Face Recognition.

[5] Sajilal Divakaran, "Spectral Analysis of Projection Histogram for Enhancing Close matching character Recognition", in Malayalam International Journal of Computer Science Information Technology (IJCSIT) Vol 4, No 2, April 2012

[6] Mohammed Javed, P. Nagabhushan, B.B. Chaudhuri Extraction of Projection Profile, "Run-Histogram and Entropy Features Straight from RunLength" Compressed Text-Documents

[7] Y. Taki, H. Hikawa, S. Miyoshi, and Y. Maeda, "Hand sign recognition system based on hybrid network classifier", in Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN), Jun. 2009, pp. 30743081.

[8] H. Hikawa, H. Fujimura, and D. Sato, " Hand sign recognition algorithm for hardware implementation", IEICE Trans. Inform. System, vol. J92-D, no. 3, pp. 405416, 2009.

[9] N. Gamage, K. Y. Chow, and R. Akmeliawati, "Static hand sign recognition using linear projection method", in Proc. 4th Int. Conf. Auto. Robots Agents (ICARA), Feb. 2009, pp. 403407.

[10] Y. C. Ham and Y. Shi, "Developing a smart camera for gesture recognition in HCI applications", in Proc. 13th IEEE Int. Symp. Consumer Electron. (ISCE), May 2009, pp. 994998.

[11] S. Oniga, A. Tisan, D. Mic, A. Buchman, and A. Vida-Ratiu, "Hand posture recognition system using artificial neural networks implemented in FPGA", in Proc. 30th Int. Spring Seminar Electron. Technology, 2007, pp. 507512.

[12] Mark S. Nixon and Alberto S. Aguado, "Feature Extraction and Image Processing", vol. 30, 3rd New York, NY, USA: Springer-Verlag, 2002.

[13] T. Kohonen, Self-Organizing Maps (Information Sciences), vol. 30,3rd ed. New York, NY, USA: Springer-Verlag, 2001.

[14] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction": A review,,IEEE Trans.Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 677695, Jul. 1997.