IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 9, Issue 3, Ver. II (May - Jun. 2014), PP 24-28
www.iosrjournals.org

A Systolic Architecture for Object Recognition Based On Hmax Model

T. Daslin Pon Santhosh[1], P. Kannan[2]
[1]*PG Scholar, Department of ECE, PET Engineering College, India*
[2]*Professor, Department of ECE, PET Engineering College, India*

Abstract: *The Recognition of objects is considered as difficult one in Image Processing. It includes recognition of face, iris, cars, airplanes, etc. For object recognition, many neural networks have been proposed. But these neural network methods have high computational complexity which require more computation time. Hierarchical Model and X model (HMAX model) is a biologically based model which is used to improve all complexities in the object recognition. This HMAX model need less number of training examples and it is the state-of-the-art system in object recognition task. The basic HMAX model involves the operation of Gabor filter for object recognition. The recognition based on Gabor filter is done in MATLAB. The Gabor filter is used to provide the features of an image of different orientations. It provides independent variables. In the proposed system, the object recognition is done by means of Linear Discriminant Analysis (LDA). It is then implemented in Field Programmable Gate Array (FPGA) which has the advantage of less area, less required power and also provide less computational time.*
Index Terms: *Object recognition, Neural network, HMAX model, Gabor filter, LDA, FPGA.*

I. Introduction

In Image Processing, Object Recognition plays a major role. This object recognition involves the recognition of face, leaves, fingerprint, vehicle, etc. and also it involves all kind of image and video tracking. This systolic architecture could be configured to partition the image and distribute the resulting sections to multiple sections of processing and the processed data provide high computation.

D. Le Ly and P. Chow [5] introduce the Restricted Boltzmann Machine in the year 2010.A RBM is a generative, stochastic neural network architecture consisting of two layers of nodes representing visible and hidden variables. This work focuses on the family of RBMs where both the visible and hidden variables have binary states. There are weighted connections between every node in opposite layers, and no connections between any nodes in the same layer. Biases are represented by setting the first node.

S. Himavathi, D. Anitha, and A. Muthuramalingam [7], introduce the concept of multiplexing the layers of Neural Networks, in order to implement a given NN of any size with minimum hardware. Layer multiplexing is the implementation of the single largest layer (i.e., layer with maximum neurons) with each neuron having the maximum number of inputs and a control block to coordinate them. Here the largest layer has eight neurons and the largest number of inputs to a neuron in the network is eight. The largest layer with each neuron having the maximum number of inputs is implemented in this method. A start signal is given to the control block to initiate the operation of the network. The results of the neurons are computed parallel and sent to the layer control block to be provided as inputs to the next layer. Once the complete network is computed, the end of computation (EOC) signal is issued to latch the output of the NN.

A.Savich, M. Moussa, and S. Areibi, introduce a MLP network [9] contains neurons structured in parallel layers. The layers are numbered from 0 to M and the neurons are numbered from 1 to N. When it used with the error back propagation (BP) algorithm, each neuron contains two key arithmetic functions which perform forward and backward computations. In the forward computation, it uses internal weights associated with each neuron for calculating the neuron's output. The backward computation step compares the network's overall output to a target, computes an error gradient, and propagates the error through layers by adjusting each neuron's weights to correct for it. However, when reviewing the mathematical properties of the relationship of outputs from inputs, a network with number of layers can be fit into a network with two layers, namely one hidden and one output layer, without loss of these properties. MLP-BP networks should have a high precision. Range can be limited as long as inputs and outputs are normalized since weights should be limited for learning progress.

In the paper, Fast Neuromimetic Object Recognition Using FPGA Outperforms GPU Implementations by Garrick Orchard, Jacob G. Martin, R. Jacob Vogelstein, and Ralph Etienne-Cummings, the performance of Gabor filter is obtained. But in that it need a very highly modified FPGA hardware for recognition. Else it provide more complexity and also take more time to recognize.

To overcome these, here the LDA method is implemented in FPGA instead of Gabor filter. Linear Discriminantanalysisexplicitlyattemptsto model thedifference between theclasses of data. LDAisapowerfulface recognitiontechniquethat overcomesthelimitationofPrinciple component analysistechnique by applyingthelineardiscriminant criterion.This criteriontriestomaximizetheratio of thedeterminant ofthe between-classscatter matrixof the projectedsamples tothedeterminant of the with-in classscattermatrixof the projectedsamples. Lineardiscriminantgroupimagesofthe sameclass and separates images of different classesofthe images. It provide less computational complexity than Gabor when it is implemented in FPGA. The implementation of FPGA provide less consumption of power also need less area for recognition.

II. System Model

2.1 Recognition based on Gabor filter:

2.1.1. Gabor filter

A Gabor filter is a linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. A Gabor filter is a linear filter used for edge detection in image processing which is named after Dennis Gabor. Gabor filter frequency and orientation representations are similar to those of human visual system, for texture representation and discrimination it has been found to be remarkably appropriate.

A sinusoidal plane wave has been modulating a 2D Gabor filter which is a Gaussian kernel function in the spatial domain. From one parent wavelet all filters can be generated by dilation and rotation, thus the Gabor filters are self-similar. The overall proposed methodology [1] is given in figure 2. From that figure, Gabor filter is of second block. It has four orientations. With four different orientations of Gabor filter, features of the image are extracting and are combined. In the HMAX model of object recognition, input images (128 x 128 or 160 x 160 grey scale pixel images) are used and the features are created based on orientations 0°, 45°, 90°, and 135° and sizes from 7 x 7 to 29 x 29 pixels. At each pixel of the input image, filters of each size and orientation are centered. By applying the four orientations and 16 different scales, the various features of the input image is created.

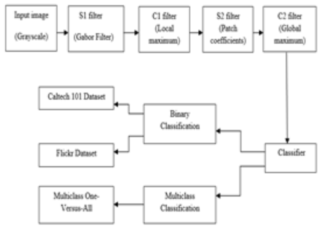


Figure: 1 Block diagram for object recognition based on Gabor filter

2.1.2. Maximum Response

In the next step, filter bands are defined, i.e., groups of Gabor filters of a certain size range (7 x 7 to 9 x 9 pixels; 11 x 11 to 15 x 15 pixels; 17 x 17 to 21 x 21 pixels; and 23 x 23 to 29 x 29 pixels). Within each filter band, a pooling range is defined (i.e.) the features having maximum match within it is taken. Else are rejected which provide simplicity and reduction in the space requirement. The maximum response of the features created is obtained and it is then compared with the dataset which is already created. The maximum response is obtained by setting the threshold value.

2.1.3. Recognition

The training dataset is already trained and stored. The test set image is given and training of that image produce some output arrays. The maximum of these output arrays are taken and then this maximum array of test image is compare with the threshold of the stored dataset. If this equals then it recognize the corresponding image. Else if the array value is not exactly equal to the threshold but somewhat related to the threshold means then also the corresponding image is recognized from the dataset.

2.1.4. Classification

Boosting classifier is better than SVM for the binary classification problem. We used the gentle boosting algorithm with weak learners consisting of tree classifiers, each with a maximum of three decision branches. We used 1280 weak learners in the classifier, each computed in series. For multiclass classification, a linear one-versus-all SVM classifier was chosen. This is a simple linear classifier, but is memory-intensive in its requirement for storing coefficients. In order to not restrict the FPGA implementation to only binary problems or only multiclass problems, the classifier was implemented separately on a host PC.

a. Caltech 101 Dataset

The binary task constituted discriminating the class in question (airplanes, cars, faces, leaves, or motorbikes) from the background class. In each case, half the images from the class in question and half images from the background class were used for training. The remaining images from both the class in question and the background class were used for testing. In each case, 10 trials were run.

b. Flickr Dataset:

The binary Minaret classification task was performed on a dataset containing 662 images of Minarets and 1332 background images. The Minaret (positive) images were obtained from Flicker by searching for “Minaret”, while the negative images were obtained by periodically downloading the most recently uploaded Flicker image.

c. Multiclass One-Versus-All

It uses 15 training examples per category. Testing was performed using 50 examples per category or as many images as remained if fewer than 50 were available. Each of the categories was weighted such that it contributed equally to the result. This is a 102-category problem including the background category. In this, it build N different binary classifiers (i.e.) it has two or more number of multipliers which are connected together to form a multiclass classification. The classification is done by means of taking the argument of maximum value of the created features if the input image. For the ith classifier, let the positive examples be all the points in class i, and let the negative examples be all the points not in class i. Let f_i be the ith classifier. It is then classify with the following equation,

$$f(x) = \arg \max_i f_i(x)$$

where, f(x) is classified output of the image.

2.2 Recognition based on Linear Discriminant Analysis:

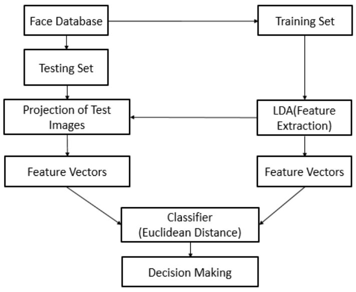


Figure: 2 Block diagram for face recognition based on LDA

2.2.1 Algorithmusedin LDA:
Step-1:

The LDA method needatrainingset composed of arelatively large groupof subjects with diverse facial characteristics. Theappropriateselectionof thetrainingsetdirectly determinesthe validity ofthefinal results.The database shouldcontainseveralexamplesof face images for each subject in the training set and at least one exampleinthetestset. Theseexamplesshould represent different frontal views of subjects with minorvariationsin viewangle.They shouldalso include different facial expressions, different lighting and backgroundconditions, andexampleswithand without glasses.Itisassumedthatallimagesare already normalized to m ×n arrays and that they contain onlythefaceregionsand notmuchofthe subjects' bodies.

Step-2:

For each image and sub image, starting withthe two dimensionalm × narrayofintensityvaluesI(x,y), weconstructthevector expansion$\Phi \in R_{m\times n}$.This vectorcorresponds totheinitialrepresentationofthe face.Thusthe setofallfacesinthefeaturespaceis treated asahigh-dimensionalvector space.

Step-3:

By defining all instances of the same person's face as beinginoneclassandthefacesofdifferentsubjects as beingin different classes for all subjects inthe trainingset,weestablishaframeworkforperforming a clusterseparationanalysisin the feature space.

Also, havinglabeledallinstancesinthetrainingset and having definedalltheclasses,wecomputethe within-class and between-class scatter matrices.

Now with-in class scatter matrix 'S_W' and the betweenclassscattermatrix'S_b' aredefinedas follows:

$$s_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (\Gamma_i^j - \mu_j)(\Gamma_i^j - \mu_j)^T \qquad (1)$$

where , Γ_i^j is the ith samples of class j, μ^j is the mean of class j, c is the number of classes, N_j is the number of samples in class j.

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T \qquad (2)$$

where, μ represents the mean of all classes.

Then the subspace for LDA is spanned by a set of vectors W= [W1, W2… Wd], satisfying

$$W = \arg\max = \bmod \left[\frac{W^T S_b W}{W^T S_w W}\right] \qquad (3)$$

The withclassscattermatrixrepresents how face images are distributed closely with-in classes and betweenclassscatter matrixdescribeshowclasses areseparatedfrom eachother.Whenfaceimagesare projectedinto thediscriminant vector W.

Face images should be distributed closely with-in classes and should be separated between classes, as much as possible. In other words, these discriminant vectors minimize the denominator and maximize the numerator in equation (3). W can therefore be constructed by the Eigen vectors of $S_w^{-1} S_b$. PCA tries to generalize the input data to extract the features and LDA tries to discriminant the input data by dimension reduction.

III. Experimental Results

Gabor filter:

In Gabor filter, the face is recognized by means of first the Gabor filter is created and Fast Fourier Transform (FFT) is taken. Then the input image is taken and the FFT is applied. After that convolve both the transformed input image and the transformed filter which is created already. Now the more orientations and scales of the input image is created. It is then compared with the trained database and recognition is performed. But it is more complicated to implement in FPGA because of more scaling and orientations.

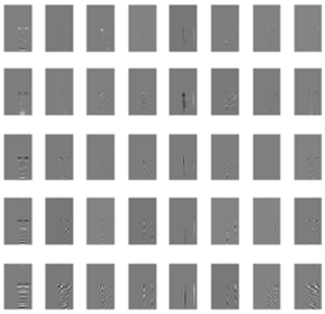
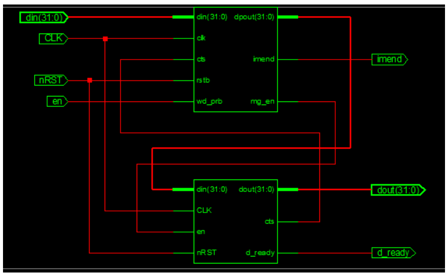


Figure: 3Gabor filtered image

The above figure shows the Gabor filtered image and then the maximum of these features are taken up and from that the recognition is take place (figure 4) by means of the Caltech 101 classifier, minaret classifier and Multiclass One-Versus-All.

Figure: 4Recognized image using Gabor filter

LDA:



Device Utilization Summary (estimated values)			
Logic Utilization	Used	Available	Utilization
Number of Slice Registers	2485	19200	12%
Number of Slice LUTs	9867	19200	51%
Number of fully used Bit Slices	2058	10294	19%
Number of bonded IOBs	69	220	31%
Number of BUFG/BUFGCTRLs	16	32	50%
Number of DSP48Es	4	32	12%

Device utilization summary:
Selected Device: 5vlx30ff324-3

Slice Logic Utilization		
Number of Slice Register	2485 out of 19200	12%
Number of Slice LUT	9867 out of 19200	51%
Number used as Logic	5451 out of 19200	28%
Number used as Memory	4416 out of 5120	86%
Number used as RAM	4416	
Slice Logic Distribution		
Number of Bit Slices used	10294	
Number with an unused Flip Flop	7809 out of 10294	75%
Number with an unused LUT	427 out of 10294	4%
Number of fully used Bit Slices	2058 out of 10294	19%
IO Utilization		
Number of IOs	69	
Number of bonded IOBs	69 out of 220	31%

IOB Flip Flops/Latches	32	
Specific Feature Utilization		
Number of BUFG/BUFGCTRLs	16 out of 32	50%
Number of DSP48Es	4 out of 32	12%

Timing Summary:

Speed Grade	3
Minimum period	6.498ns (Maximum Frequency: 153.900MHz)
Minimum input arrival time before clock	2.158ns
Maximum output required time after clock	3.135ns
Maximum combinational path delay	No path found

IV. Conclusion

By Gabor filter, the image is recognized properly, but the drawback is it is more complicated to implement in FPGA. Because it has large number of orientations and for that it need FPGA kit of latest version which is more expensive. To overcome this, Linear Discriminant Analysis method has been successfully applied to face recognition which is based on a linear projection from the image space to a low dimensional space. But the major drawback of applying LDA is that it may encounter the small sample size problem. When the small sample size problem occurs, the within-class scatter matrix becomes singular. Since the within-class scatter of all the samples is zero in the null space of S_w, the projection vector that can satisfy the objective of an LDA process is the one that can maximize the between-class scatter.

But face image data distribution in practice is highly complex because of illumination, facial expression and pose variation. The kernel technique is used to project the input data into an implicit space called feature space by nonlinear kernel mapping. Therefore kernel trick is used taking input space and after that LDA performed in this feature space, thus a non-linear discriminant can be yielded in the input data.

Acknowledgement

Apart from our efforts, the success of any work depends on the support and guidelines of others. We take this opportunity to express my gratitude to the people who have been supported us in the successful completion of this work. We owe a sincere prayer to the LORD ALMIGHTY for his kind blessings without which this would not have been possible. We wish to take this opportunity to express our gratitude to all who have helped us directly or indirectly to complete this paper.

References

[1] Garrick Orchard, Jacob G. Martin, R. Jacob Vogelstein, and Ralph Etienne-Cummings, Fellow, IEEE “Fast Neuromimetic Object Recognition Using FPGA Outperforms GPU Implementations,” IEEE Trans. Neural Netw. And Learning Syst, VOL. 24, NO. 8, AUGUST 2013.

[2] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, “Robust object recognition with cortex-like mechanisms,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 3, pp. 411–426, Mar. 2007.

[3] J. Kim, M. Kim, S. Lee, J. Oh, K. Kim, and H. Yoo, “A 201.4 GOPS 496 mW real-time multi-object recognition processor with bio-inspired neural perception engine,” IEEE J. Solid-State Circuits, vol. 45, no. 1, pp. 32–45, Jan. 2010.

[4] V. Bonato, E. Marques, and G. Constantinides, “A parallel hardware architecture for scale and rotation invariant feature detection,” IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 12, pp. 1703–1712, Dec. 2008.

[5] D. Le Ly and P. Chow, “High-performance reconfigurable hardware architecture for restricted boltzmann machines,” IEEE Trans. Neural Netw., vol. 21, no. 11, pp. 1780–1792, Nov. 2010.

[6] M. Pearson, A. Pipe, B. Mitchinson, K. Gurney, C. Melhuish, I. Gilhespy, and M. Nibouche, “Implementing spiking neural networks for real-time signal-processing and control applications: A modelvalidated FPGA approach,” IEEE Trans. Neural Netw., vol. 18, no. 5,pp. 1472–1487, Sep. 2007.

[7] S. Himavathi, D. Anitha, and A. Muthuramalingam, “Feedforward neural network implementation in FPGA using layer multiplexing for effective resource utilization,” IEEE Trans. Neural Netw., vol. 18, no. 3, pp. 880–888, May 2007.

[8] D. Geman and S. Geman, “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,” IEEETransactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, pp. 721–741, 1984.

[9] R. Chellappa, C. Wilson, and S. Sirohey, Human and Machine Recognition of Faces: A Survey, Proc. IEEE, vol. 83, no. 5, pp. 705- 740, 1995.

[10] P.N.Belhumeur and D.J. Kriegman, “What is the Set of Images of an Object under all Possible Lighting Conditions”, IEEE Proc. Conf. Computer Vision and Pattern Recognition, 1996.

[11] K.Etemad and R. Chellappa,"Discriminant analysis for face recognition of human face images" Journal of Optical society of America A/Volume 14, pp 1724- 1733, Aug1997.

[12] P. N. Belhumeur, J.P.Hespanha and D.J. Kriegman, "Eigen faces Vs Fisher faces: Recognition using Class specic linear projection” IEEE Trans. Pattern Anal. Machine Intell, Vol 19, pp 711-720, July1997

[13] Juwei Lu,Kostantinos N. Plataniotis and Anastasios N. Venet sanopoulos, "Face Recognition Using LDA- Based Algorithms" IEEE Transactions in Neural Network,Vol.14 No.1,January 2003.

[14] ORL face database: AT &T Laboratories, Cambridge,U.K..[Online]. Available: http://www. Cam-orl. co. uk / facedatabse.html.

[15] Aleix M.Matinez and Avinash C. Kak," PCA versus LDA"IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No. 2, February 2001.

[16] M. Turk and A. Pentland, “Face Recognition Using Eigenfaces,” Proc. IEEE Conf on Computer Vision and Pattern Recognition, 1991, pp. 586 – 591.

[17] Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen,Jian-Cheng Huang, A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA ,Proc., International Journal of Information Technology Vol. 11 No. 11 2005 .

[18] Xue-wen Chen, Thomas Huang, Facial expression recognition: A clustering-based approach, Proc., ELSEVIER, Pattern recognition letters, Volume, June 2003, Pages 1295–1302.

[19] Jianke Li, Baojun Zhao, Hui Zhang, Face Recognition Based on PCA and LDA Combination Feature Extraction, Proc.IEEE,Information Science and Engineering (ICISE), 2009 1st International Conference on Dec. 2009,pp. 1240 – 1243.

[20] H.Y. Mark Liao, C.C. Han, G.J. Yu, H.R. Tyan, M.C. Chen, L.H. Chen, Face recognition using a face-only database: a new approach, Proceedings of the third Asian Conference on Computer Vision, Hong Kong, Lecture Notes in Computer Science, Vol. 1352, 1998, pp. 742-749.

[21] B. Moghaddam, A. Pentland, Probabilistic visual learning for object representation, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 696-710.

[22] M.Turk,A.Pentland,Eigenfaces for recognition, J.Cognitive Neurosci. 3 (1) (1991) 71-86.