HAND-WRITING RECOGNITION USING NEURAL MICRO-CLASSIFIERS NETWORK

Dr.Abdellatief Hussien Abou Ali
HICI, Computer Science Dept.,
El-Shorouk Academy, Cairo, Egypt,
Email: dr.abdullatif.hussein@sha.edu.eg; aabouali4@gmail.com;

ABSTRACT

In this study, a hand writing recognition methodology based on the neural binary micro-classifier network. The proposed methodology uses simple well known feature extraction methodology. The feature extraction used is the discrete cosine transformation low frequencies coefficients. The micro-classifier network is a deterministic four layers neural network, the four layers are: input, micro-classifier, counter, and output. The network provide confidence factor, and proper generalization is guaranteed. Also, the network allows incremental learning, and more natural than others. The recognition methodology was tested using the standard MNIST dataset. The experimental results of the methodology showed comparative performance taking in consideration the design advantages.

KEYWORDS: *Neural networks; Feature extraction; Image processing; DCT; MNIST; HAND WRITING;*

1. INTRODUCTION

Hand writing is the ability of a computer to properly acquire, interpret, and map to a symbol in a language set of symbols from sources such as paper documents, photographs, touch-screens and other input devices. The hand written text could be acquired from a piece of paper through scanners; digital cameras, electronic pen tipping, and touch screens. The acquisition could be online or offline. Hand writing applications normally uses Optical Character Recognition, OCR, as primary phase but it extends beyond to words and sentences analysis that could overcome the shortage of the first phase. That is, handwriting recognition principally entails optical character recognition. However, a complete handwriting recognition system also handles formatting, performs correct segmentation into characters and finds the most plausible words. Recognition is a collaboration effort in: neuroscience, computer vision, pattern recognition, digital image processing, linguistics, classification, and machine learning [1-5].

Optical character recognition requires preprocessing steps that may include: image enhancement, objects segmentation, detection of potential objects then character recognition. Images recognition is not a comparison process with recalled pre-stored known characters images. Character recognition is a lengthy search process within an object-image for a set of discriminating features, extracting and proper learning to these features form a training data for discriminations purposes[6]. The learning process induces a discriminating metrics and building boundaries between sets. After learning, the induced metrics or boundaries are used to assign the unknown image-character a character code from the language set of codes. Hand written characters recognition difficulties originates from the variance in the character-image, for same-object character in: scale, case, font, translation, illumination, clutters, orientation, and Imaging acquisition conditions. Hand-written characters recognition includes detection, feature extraction, as well as learning, and classification.

Hand written characters detection is a segmentation process for small objects, characters, from a word context that could exist in sentence context which in turn could be in a paragraph context that could be a part of larger document. The character detection based on Hypergraph model was proposed by Samrajya et al. in [7].

Hypergraph model treats an image as packets of pixels recombining these packets of different sizes a given word image. Which could be segmented into characters. Dawoud [4] introduce iterative cross section sequence graph (ICSSG) for the character segmentation. ICSSG tracks the characters growth at equally spaced thresholds. Lee and Verma [8] propose segmentation algorithm based on validity constrains for characters separators starting from words level. That is, word images heuristically separated based on pixel density considering upper and lower thresholds. Same concept is applied on word image to using expert based validation processes to extract character image. A pruning process is proposed with some knowledge about characters as well as size constrains proposed in [9]. Amit Choudharya et al. [10] proposed an approach contains several steps. These steps could be summarized as: noise removal (image enhancement), thresholding to turn to binary image in an inverted form, labeling process to segment character images, connected foreground object repeatedly extracted, insignificant and large objects dropped, and finally the remaining labels inverted and considered potential objects.

Feature selection and extraction goal is finding a minimal subset of features from a set of large possible features that if used properly will lead to minimal recognition errors. The factors affecting feature selection include: dimensionality, inclusion of in-class similarity, inclusion of dissimilarity of inter-classes, ease of finding and implementation, computation complexity, and robustness to former mentioned challenges. Feature selection is a focal point in the recognition process[]. Therefore, it gained the attention of the researchers' community. The features presented by the community include [6-16] include: moments, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Multi-dimensional Scaling (MDS), Self-organizing map (SOM), Active Shape Models (ASM), Gabor Wavelet Transforms (GWT), and Discrete Cosine Transform (DCT).

Classifiers, according to [17], are built based on similarity, probability or decision boundaries. Similarity classifiers use similarity metrics to measure the closeness to the class members or the class preset representative(s). The probability based classifiers uses in or out of class probability. Decision boundary based classifiers; basically find out the separating hyper-surfaces between the classes to find out classes polyhedrons that represent the class's containers. This process could be done through a training process from data sets through a training vehicle that evolves the surfaces and decision boundaries either implicit or explicit such as case of neural networks.

The classifying vehicles for image based objects are two main categories conventional and learning based. Convectional such as structural and statically searches for local features as well as the relationships and or uses statically patterns found from the training sets[17]. The second approach based on learning from known individuals such as: neural networks. Learning directly from raw images is complex and requires complex network structures, as well as learning algorithms of high time complexity [18]. So, either using image dimension reduction as well as filters highlights different features from raw images as in deep learning or applying features extraction ahead and learning classification in feature space. The later, classification, if properly implemented leads to less complex architecture and high generalization ability.

Artificial Neural networks, ANN, are massive parallel operated interconnected computing elements contains adapted parameters and has associated learning vehicle(s). The training process aims at setting the network parameters to hopefully the proper decision boundaries between the sets. The network basic computational element is the neuron. In the network, neurons organized in layers from input to output. The layers between input and output are called hidden layers. The neurons are interconnected, unidirectional, bidirectional, or both. Neurons interconnections could exist of the same layer, to forward layer, backward one, or combinations. The network topology, transfer function, and learning method are the focal points in network design. The networks topologies includes feed-forward and recurrent [19-23]. The transfer functions used in neural networks include: linear, sigmoidal, Gaussian, hyperbolic, and bi-radial. Networks learning could be supervise, unsupervised or reinforced. The learning could be, also, deep or shallow in structure. From the wildly used networks, multilayer perceptron, radial basis, Hopfield, deep networks, and self-organizing maps.

Neural networks application areas include: functions approximation, classification, forecasting, mapping, security alerts, marketing, classification, as well as recognition. There are many hardware realizations to networks [21]. Neural networks as a classifiers, in general, takes the classification burden to a learning process however there are still basic questions about the proper structure, evolution, and the correct generalization ability [22][23].

The efficiency of recognition depends on three basic parameters (i) an efficient invariant feature representation with respect to illumination, scaling, rotation, … etc. (ii) Classification technique that maps the feature vectors into their appropriate classes with minimal misclassification. (iii) Prober generalization abilities to unknown cases.

Human recognition development recognized when babies start classifying all males as father and all females as mother. That is, more or less building a separating hyper plane between the two sets from coarse features. This classification ability grows and develops by time. So, elder babies consider only males and females with common features with their father, and mother. A finer classification develops by time and association with names in a great complex unknown organization. However, in general, one can infer that adding a new person to some human life requires considering separating him from the previously known others and does not affect among the previous known. Also, one can easily infer that during human's recognition some sort of features recall happens. That means, there is some form of features memorization exist, not just the boundaries such as ANN, which is linked to that great recognition vehicle.

This study proposes hand written letters recognition methodology based on a binary classifier neural network. The binary classifier network built on learning separating hyper-planes between pair of classes proposed in [24-26]. The classifier doesn't require rebuild of knowledge when adding new classes to the system rather it integrate the knowledge of the new class to the network. The proposed network generalization ability guaranteed giving proper selection of the training set. The network structure is deterministic per problem. Moreover, adding members to a class requires rebuild of those class boundaries with others.

The remaining of this paper organized as following: Section two outlines the classifier neural network and the face recognition methodology under the study.

Section three contains tests and results. Section four is the study conclusion.

2. FACE RECOGNITION METHODOLGY AND CLASSIFER NETWORK

The proposed network and recognition methodology operates in three modes: initial learning, recognition, and incremental learning when needed. During the first mode, initial learning; the system learns the separation between the initial set of classes. In the second mode, operational or recognition, unknown face presented to the system to classify. In the third mode, incremental learning, a new classification abilities integrated to the current network or new subjects added to a class training set.

In the first mode, the sets of faces that are the subject of the initial recognition run through three stages: a preprocessing, feature extraction then set in classes- pairs to an elementary learning separation process. In this elementary learning, the learning algorithm simply finds the separating hyper-plane of its two classes if exist or a local-minima if not [24]. A successful first mode produces per class a set of hyper-planes separate a class from others. The elementary learning processes are parallel processes and don't include dependencies. The learning process sets its micro-classifier parameters/ weights. The micro-classifier output is similar to the flip-flops. That is, its output ($q, \bar{q}$). The micro-classifier, MC, assigned the classification of the two classes X, Y will have its output q will be 1, if and only if the presented pattern is seen to $\in X$ class side and vice versa. Figure (1) presents conceptual view of the micro classifier.

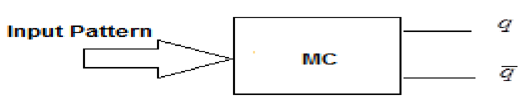

Figure (1) Micro-Classifier

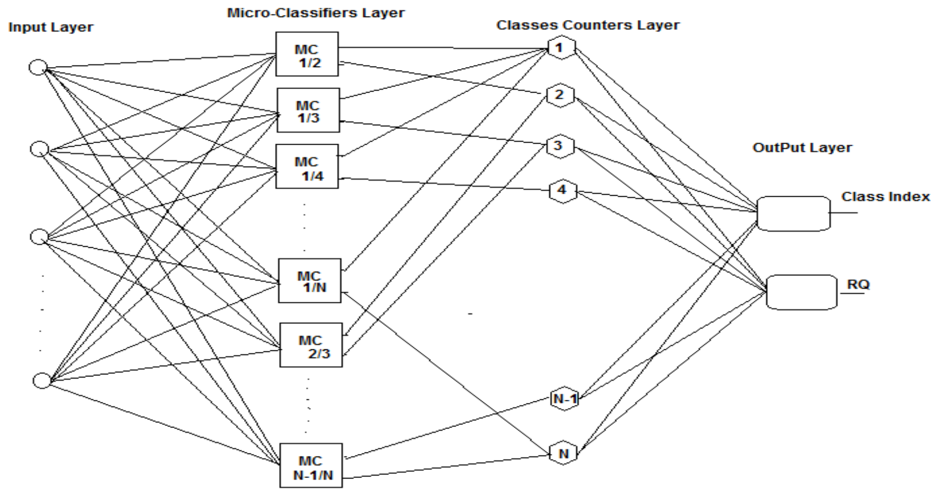

Figure (2) Classifier Network

In the second mode, an unknown face image runs through three stages: a preprocessing, feature extraction. The feature vector is applied, in parallel, to all MCs. The MCs are linked to classes counter array. The MC designated for classification between the two classes X, Y its outputs $q, \bar{q}$ connected X, Y counters respectively. The counter array designates the position of the feature vector with respect to class's polyhedrons. The classes counter array output is the input to the comparator. The comparator sets the class index, which is the class of the greatest vote and compute the recognition quality RQ .

In the third mode, adding more classification ability to the network, the training set of images for the classes run through the first two stages to get the new classes features vectors. Then, the system recalls the former classes- features. A learning process confined to the separating the new classes from the former classes is initiated. The outcome of the learning processes is used to add more configurations to the network without affecting the existing ones. The added configurations simply activate and set new MC's, and activate the new classes' counters. Also, as the system goes on a class with miss classification rate over the normal its polyhedron could be reset using more training elements using some agent in a similar process to the former one. Figure (3) shows the learning process of the proposed system.

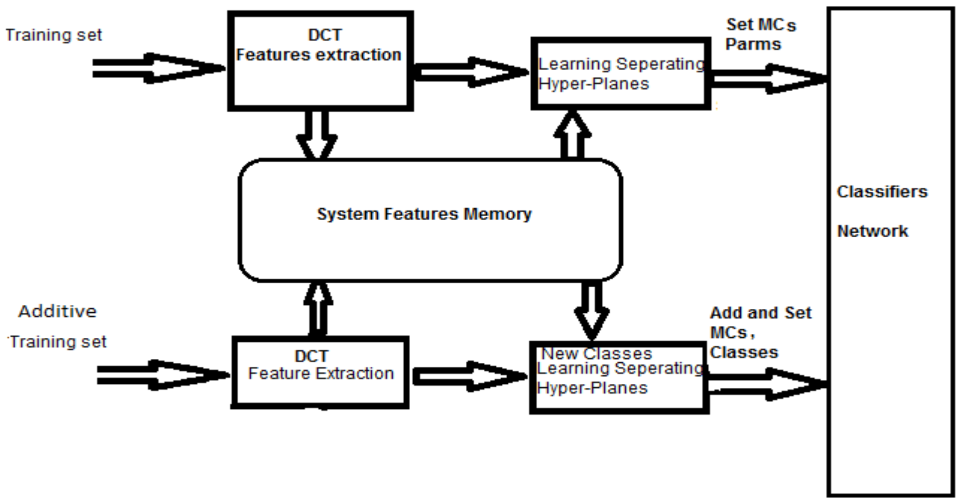

Figure (3) Classifier initial and incremental training

Micro-Classifiers

The proposed classifier network built based on 'micro-classifier' as basic building block. The micro-classifier concerns with classification of two classes of the set of classes. The micro-classifiers training process could be summarized as following:-

Assuming the two finite sets $X, Y \subset R^l \quad l,\ finite,$

$$X = \{x_1, x_2, x_3, \ldots\ldots, x_M\} \qquad and$$

$$Y = \{y_1, y_2, y_3, \ldots\ldots, y_N\} \qquad\qquad x_i, y_i \in R^l$$

The search process is for $\varphi_x^*, \varphi_y^* \in R^l$ such that

$$\|x_i - \varphi_x^*\| < \|x_i - \varphi_y^*\| \qquad and$$

$$\|y_i - \varphi_y^*\| < \|y_i - \varphi_x^*\| \qquad \forall x_i \in X, y_i \in Y,$$

The learning algorithm is based on the study in [24]. The algorithm evolves to representatives say φ_x^*, φ_y^* then: Having the φ_x^*, φ_y^* for the two sets the separating hyper plane equation is $(\varphi_x^* - \varphi_y^*) \bullet p = 0.5(\varphi_x^* - \varphi_y^*) \bullet (\varphi_x^* + \varphi_y^*)$

Where '$\bullet$' is the vector dot product operator, and p is plane point $\in plane$. The hyper plane divide the range of points either X or Y. The classifier for unknown vector Z considered X side if $(Z - 0.5(\varphi_x^* + \varphi_y^*)) \bullet (\varphi_x^* - 0.5(\varphi_x^* + \varphi_y^*)) >= 0$ otherwise considered Y side. The micro-classifiers increment the counters of the potential classes from their perspectives. The micro-classifiers act in parallel on its inputs. The second layer acts on the class's counters $clsc_i$ the vector assumed to be of class $n \quad iff\ clsc_n > clsc_i \quad \forall i \neq n \quad and \qquad RQ = ((clsc_n + 1)/m) > T$ where m is classes count, $T \in [0.5,1]$ is the recognition threshold, and RQ is the recognition quality.

Features Extractions

In general, the failure in feature extraction or the choice of wrong features collapses the recognition process. Features could be global, local, or both. The global feature operator applies to the entire image. However, Local uses chucks or window of the image. In local, window position is used together with operator outcome. Feature bases include: colors, edges, corners, textures, statistical, frequencies coefficients and combinations of them. The feature extraction could be done in spatial and frequency domains. DCT used effectively image and video encoding schemes such as JPEG, and MPEG. The basics on which such schemes count on DCT is the fact that lower frequencies contribute more significantly to images quality compared to higher ones. Figure(5) shows charater '6' from the MINTS standard dataset and Table (1) shows the 2D DCT for that character.

Discrete Cosine Transformation

The photoreceptors have over 1.5 hundred million signals while retinal level approximately receives 1 millions of these which are the subject of biological keeping and recognition [27]. Consequently, thinking of forms or transformations is natural. The transformations role concludes the high redundancies to more abstracted representations. Generally, human vision, and machine processing saturates at certain point of details. Therefore adding more details after saturation does not benefit, and could have negative effect. From the widely used transformations: Fourier, Discrete Cosine transformation (DCT), Karhunen-Loeve transform (KLT), Legendre moments, Hue moments, and others.

The discrete cosine transformation was widely used as a mean for image abstraction for the purpose of compression [28] as well as feature extraction [29-31]. The use of the DCT is used by holistic approaches in which transformation is done on the entire image. Also, it is used by the local based and block based approaches to overcome the computation complexity of the transform [28]. Some used combinations of both local and holistic for sake of more informative abstraction [32]. The use of the transform in image processing included both single (based on raw and column scans) and two dimensional transform. The two dimensional discrete cosine transform is a transformation from real to real domain. The outcome of the transformation is a real matrix of the same dimension of the original one, Table (1). The DCT has an inverse that could be used to retrieve the original image from the transformation frequency domain matrix.

The matrix elements of lower indices contain the low frequencies of the images. It has been reported that eliminating the highest frequencies is not significantly noted by the human vision and consequently does not affect computer vision for natural images of large sizes. That points to the fact that lowers frequencies more informative than other side of such cases.

The transformation for an image matrix $f(x, y)$ of dimensions N, M consequently is as following:-

$$C(u,v) = \alpha(u)\alpha(v)\sum_{y=0}^{M-1}\sum_{x=0}^{N-1} f(x,y)\cos(\frac{(2x+1)\pi u}{2N}) \quad \cos(\frac{(2y+1)\pi v}{2M})$$

$$where$$

$$\alpha(u) = \begin{cases} \sqrt{\dfrac{1}{N}} & u = 0 \\ \\ \sqrt{\dfrac{2}{N}} & u \neq 0 \end{cases}$$

The inverse transformation is:-

$$f(x,y) = \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} \alpha(u)\alpha(v)C(u,v)\cos(\frac{(2x+1)\pi u}{2N}) \quad \cos(\frac{(2y+1)\pi v}{2M})$$

3. TESTING AND RESULTS

The source images used in the testing process is the standard MNIST database []. The MNIST database contains Arabic Numerals hand written form write-diverse humans communities. The dataset contains 60000 handwritten image-numeral each is tagged with human recognition. Also, the dataset contains 10000 handwritten image-numeral for testing purposes. Many of the testing symbols are challenging[11]. These images are gray scaled 0 to 255. Samples from the dataset are in figure (4). Our testing grouped to four experiments. All experiments utilizes the full dataset both training and testing.

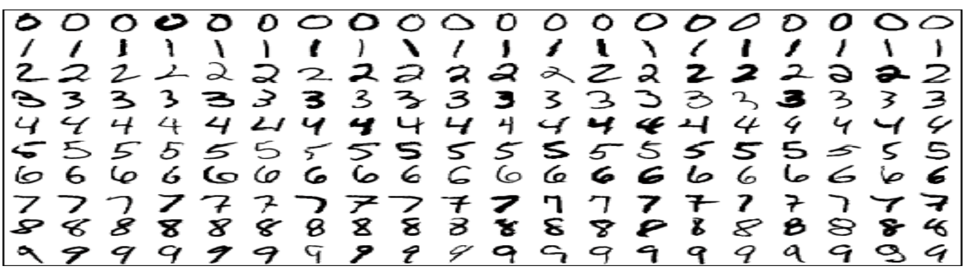

Figure (4) Samples of the standard testing dataset.

Table (1) contains the coefficients of the lowest 8 frequencies of the first two pictures of the first three subjects.

Table (1): 2D-DCT coefficients for figure (1) digit

1055.45	84.1287	-391.87	-53.77	-192.21	28.01	-77.34	-39.96	-33.4	-8.91
-411.38	-151.81	-70.37	82.87	157.98	28.37	94.40	15.58	57.81	-7.068
77.17	-77.97	-154.5	192.5	51.34	-22.65	35.73	-41.97	11.61	-8.31
150.68	-200.60	289.54	251.62	-216.62	-49.45	-97.02	-10.80	-17.94	-4.52
-203.85	30.73	77.67	-45.91	-196.13	-50.19	126.56	16.95	52.47	2.90
-32.29	60.41	75.33	-71.06	-16.42	150.08	8.444	-51.29	-13.14	-18.38
-20.0	-103.26	12.96	34.95	69.27	30.302	-38.48	-29.69	-0.90	14.58
0.539	24.700	59.99	-5.117	-97.0	17.51	-32.7	-8.02	34.83	-23.32

Figure (5) skinned out digit 6 from the dataset

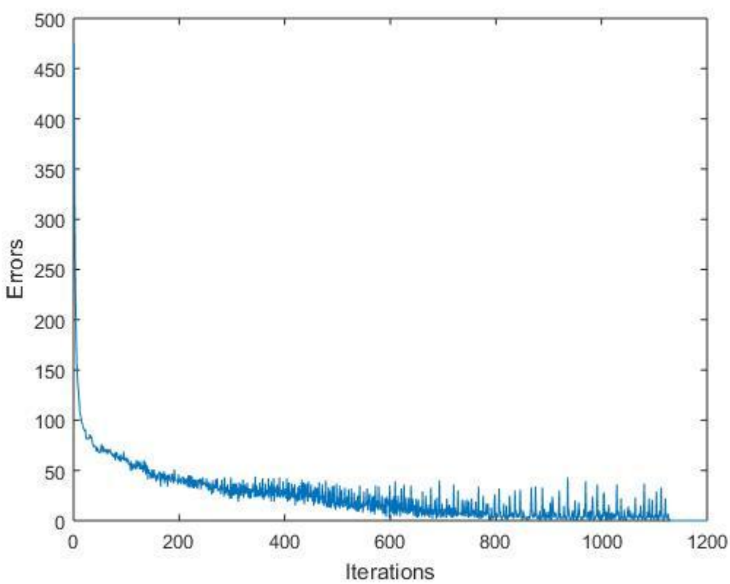

Figure (6) Two sets separation learning curve (classes 1,3)

Experiment 1:-

In this experiment we demonstrate the micro-classifiers learning performance as well as the effect of increasing the features set. The features vector formed in coincidence with zigzag sequence used in JPEG compression. Figure (6) show the learning curve of the first and third class. From the figure we can learn that the methodology reached full linear separation after a significant number of iterations. Taking in consideration that the subset used in training is over 12K elements and the feature vector dimension used is 60 therefore we consider that number of iteration normal. The learning curve below 50 error counts contain glitches that comes from the high learning rate used which is one. Figure (7) contains effect of increasing the size of the feature vector on the error percent. From the figure one can infer that low dimension vectors is not suitable and there is a point, in this cases 60, after which the rate of improvement is not significant i.e. some sort of saturation could be reached.

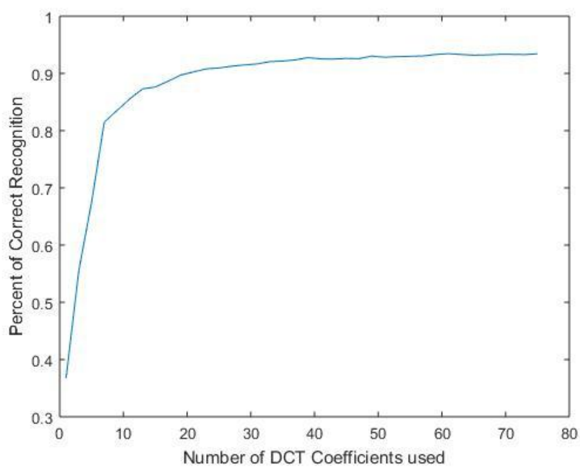

Figure (7) DCT coefficients performance

Experiment 2:-

In this experiment the characters the dataset used as without preprocessing and the learning process followed as per the design the trained network yielded recognition of 89% of the testing dataset. In the next run a preprocessing step for the character-image is added. The added step is skimmed to box only the characters and resizing to fixed size, 16x16 pixels. The result of the run came to recognition of 93%. Table (2) presents the inter-classes the training status after 2000 iteration or less.

Table (3) contains the first 20 elements neural counters reading while testing the 10000 elements of the dataset.

Table (2) Interclasses Learning Status

	1	2	3	4	5	6	7	8	9
0	100%	100%	100%	100%	99.8%	99.04%	100%	100%	100%
1		96%	97%	98%	95%	100%	97%	95%	100%
2			95%	100%	95%	97%	100%	96%	100%
3				100%	96%	100%	97%	94%	96%
4					100%	96%	95%	96%	93%
5						96%	100%	97%	99%
6							100%	100%	100%
7								99%	95%
8									94%

Table (3) first 20 Elements of testing 10000 character image voting counters values

0	1	2	3	4	5	6	7	8	9	Human Recognition
2	5	7	8	0	5	1	9	5	3	7
4	6	9	4	0	6	5	1	8	2	2
1	9	7	5	5	2	4	4	7	1	1
9	2	6	3	3	8	6	1	2	5	0
3	2	7	3	9	2	0	6	5	8	4
3	9	5	6	1	3	3	4	8	3	1
1	5	0	3	9	6	2	4	8	7	4
2	5	3	2	9	3	0	6	7	8	9
7	4	6	1	1	9	7	3	5	2	5
3	3	1	5	7	3	0	7	7	9	9
9	2	7	6	0	6	6	1	4	4	0
8	1	6	3	5	3	9	0	6	4	6
1	3	1	5	7	4	1	8	6	9	9

Experiment 3:-

In this experiment, the same preprocessing step was done and the dataset and testing data were rescaled to hypercube [-1, 1]. The recognition percent to testing dataset was 94%. There was insignificant increase in recognition percent approximately 1%.

There are many bench marks on MNIST dataset. The results of this study do not compete with the best. The results are between worst and the best. The bench marks includes: [47] of correct recognition reports percent's from 93% Gradient-Based Learning with 300 hidden neurons, and [49] which reports 99.5% percent using 1x29x29-20C4-MP2-40C5-MP3-150N-10N deep neural networks. The proposed network architecture is simple compared to that of deep networks. Also, the training and use more straight forward. Moreover, the architecture allows incremental learning and partial retraining when needed.

4.CONCLUSION

In this paper, hand-written recognition methodology was introduced. The methodology was based on micro classifiers network which contains four layers: input, micro-classifier, counter, and output. The micro-classifier concerns with classification of two classes of the problem. The micro-classifiers votes for the classes neurons of the next layer. The internal layer process of the classes-neurons evolves to a winner class and quality factor that set the output layer neurons.

The DCT transformation was used as feature extraction vehicle after median filter for noise removal. The proposed face recognition methodology was tested using the standard MNIST data set. The experimental results of the vehicle do not compete with best reported as per the percent of recognition however is somehow in the middle. Taking in consideration the feature extraction used , DCT, and the design preference then it could be considered comparable.

REFERENCES

[1]. R. Plamondon and S. N. Srihari, “On-line and off-line handwriting recognition: A comprehensive survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63–84, 2000. View at Publisher • View at Google Scholar • View at Scopus

[2]. S. N. Srihari, S.-H. Cha, H. Arora, and S. Lee, “Individuality of Handwriting,” Journal of Forensic Science, vol. 47, no. 4, pp. 856–872, 2002. View at Google Scholar

[3]. A. Graves and J. Schmidhuber, “Offline handwriting recognition with multidimensional recurrent neural networks,” in Proceedings of the 21st Intl. Conf. Neural Information Processing Systems (NIPS'08), pp. 545–552, 2008.

[4]. Y. LeCun et al., “Handwritten digit recognition with a back-propagation network in,” in Advances in neural information processing systems 2, pp. 396–404, Morgan Kaufmann Publishers, 1990. View at Google Scholar

[5]. P. Sermanet and Y. LeCun. Traffic sign recognition with multi-scale convolutional networks. In Proceedings of International Joint Conference on Neural Networks (IJCNN’11), 2011.

[6]. Bhumika B. Pate, Hinaxi M. Patel, ” Survey on Offline Character Recognition for Handwritten Gujarati Text”, International Journal of Computer Applications (0975 –8887), Volume 177 –No.6, November 2017.

[7]. P. Samrajya, M. Lakshmi, Hanmandlu, and A. Swaroop, "Segmentation of Cursive Handwritten Words using Hypergraph," TENCON 2006, IEEE Region 10 Conference, pp. 1-4, 2006.

[8]. Lee H, Verma B. A novel multiple experts and fusion based segmentation algorithm for cursive handwriting recognition. In: proceeding of joint international conference on neural networks, 2008, 2994 -2999.

[9]. Choudhary A, Rishi R. A Robust Character Extraction Technique for Machine Printed Cursive Words of Varying Length. In: proceeding of second international conference on signal,image processing , and pattern recognition, 2013.

[10]. Amit Choudharya, Rahul Rishib, Savita Ahlawatc, “A New Approach to Detect and Extract Characters from Off-Line Printed Images and Text”, Information Technology and Quantitative Management (ITQM2013), 2013.

[11]. Dan Cires¸an, Ueli Meier, J¨urgen Schmidhuber, ” Multi-column Deep Neural Networks for Image Classification”, Technical Report No. IDSIA-04-12, February 2012, IDSIA / USI-SUPSI,Dalle Molle Institute for Artificial Intelligence Galleria , Switzerland

[12]. Issam Dagher,”Incremental PCA-LDA algorithm”, International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (2), May 2010 .
[13]. J. Shermina,V. Vasudevan,”An Efficient Face recognition System Based on Fusion of MPCA and LPP”, American Journal of Scientific Research ISSN 1450-223X Issue 11(2010), pp.6-19.
[14]. Yun-Hee Han,Keun-Chang Kwak, ” Face Recognition and Representation by Tensor-based MPCA Approach”, 2010 The 3rd International Conference on Machine Vision (ICMV 2010).
[15]. Abdelatif Hussien A. Ali, “Face Recognition With Pre-moment processing”, Journal of the Advances in Computer Science, vol. 8, June 2015
[16]. Ajoy Kumar Dey, Susmita Saha, Avijit Saha, Shibani Ghosh, “A Method of Genetic Algorithm (GA) for FIR Filter Construction: Design and Development with Newer Approaches in Neural Network Platform”, International Journal of Advanced Computer Science and Applications, Vol. 1, No. 6, pp. 87-90, 2010.
[17]. A. Jain, R. Duin, and J. Mao. ,”Statistical pattern recognition: A review.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1):4–37, January 2000.
[18]. Steve Lawrence,C.Lee Giles,A.h Chung Tsoi,Andrew D. Back, ” Face Recognition: A Convolutional Neural Network Approach”, IEEE Transactions on Neural Networks, Vol. 8, Jan 1997.
[19]. Martin T Hagan , Howard B Demuth, Mark H Beale , Orlando De Jesús ,” Neural Network Design (2nd Edition) 2nd Edition”, Amazon, 2015
[20]. Abdellatief H. Abou ALI,” Non-Preemptive Multi-Constrain Scheduling for Multiprocessor with Hopfield Neural Network ”, International Journal of Computer Science and Information Security, Vol. 11 No. 3 (pp. 125-130), March 2013
[21]. Esraa Zeki Mohammed and Haitham Kareem Ali,” Hardware Implementation of Artificial Neural Network Using Field Programmable Gate Array”, International Journal of Computer Theory and Engineering, Vol. 5, No. 5, October 2013
[22]. L. Franco and S. A. Cannas, "Generalization and selection of examples in feedforward neural networks", Neural Comput., vol. 12, pp. 2405-2426, 2000.
[23]. Huan Xu and Shie Mannor. Robustness and generalization. Machine Learning, 86(3):391–423, 2012.

[24]. Abdellatief Hussien AbouAli “LEARNING LINEAR SETS SEPARATION”, International Journal of Engineering and Information Systems (IJEAIS), Vol. 1 Issue 9, November – 2017, Pages: 196-205

[25]. Abdellatief Hussien AbouAli, "Neural Micro-Based Classifier for Face Recognition",International Journal of Engineering and Information Systems (IJEAIS),Vol. 1 Issue 10, December – 2017, Pages: 32-44.

[26]. Abdellatief Hussien AbouAli, " Face Recognition Using Dct And Neural Micro-Classifier Network",International Journal of Engineering and Information Systems (IJEAIS), Vol. 2 Issue 3, March – 2018, Pages: 27-35.

[27]. Alice C. Parker; Adi N. Azar, A hierarchical artificial retina architecture , SPIE Europe Microtechnologies for the New Millennium, 2009, Dresden, Germany.

[28]. R.C.Gonzalez, R.E.Woods Gonzalez, Digital Image Processing, Addison- Wesley, 2012

[29]. SumanpreetKaur and Rajinder SinghVirk, DCT Based Fast Face Recognition Using PCA and ANN, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 5, May 2013.

[30]. Hazim Kemal Ekenel and Rainer Stiefelhagen ,’ Local appearance based face recognition using discrete cosine transform, the 13th Europe conference on signal processing 2005.

[31]. ZIAD M. HAFED AND MARTIN D. LEVINE, Face Recognition Using the Discrete Cosine Transform, International Journal of Computer Vision 43(3), 167–188, 2001.

[32]. Aman R. Chadha, Pallavi P. Vaidya, M. Mani Roja, Face Recognition Using Discrete Cosine Transform for Global and Local Features, Proceedings of the 2011 International Conference on Recent Advancements in Electrical, Electronics and Control Engineering (IConRAEeCE).