

A NOVEL OS-ELM CLASSIFIER BASED ON HOG FEATURE EXTRACTION AND VIOLA JONES ALGORITHM FOR FACIAL RECOGNITION

^{1†}Ankit Rajpal, Department of Computer Science, DeenDayalUpadhyaya College, University of Delhi, India, ankit.rajpal@jeee.org

²Anurag Mishra, Department of Electronics Science, DeenDayalUpadhyaya College, University of Delhi, India, anurag_cse2003@yahoo.com

³Khushwant Sehra, University School of Information Communication & Technology, Guru Gobind Singh Indraprastha University, India, sehrakhushwant@gmail.com

ABSTRACT

The automatic human-machine interaction is an important area on which the world research community is presently focusing. This has taken a significant space in developing newer schemes for efficient cyber physical systems. Face recognition on real time scale is an important component of this domain. In the present paper, a fast single layer feed-forward neural network known as Online Sequential Extreme Learning Machine (OS-ELM) is exploited as a classifier for face recognition problem. In the beginning, the Viola Jones algorithm is employed to detect and extract the faces from the datasets. These faces are subjected to Histogram of Oriented Gradients (HOG) feature detection to extract relevant image features which will form the classifier dataset. Two standard datasets – AT&T and YALE are used in this work. The OS-ELM based classifier is novel and successful to carry out a very high degree of classification after training and testing in the millisecond time span. Thus, it is found to be compatible with real time events such as face recognition.

Keywords: Online Sequential – ELM, Viola Jones Algorithm, Histogram of Oriented Gradients (HOG), Multi – Class Classification, One Against One (OAO), One Against All (OAA).

INTRODUCTION

The newly emerging area of cyber physical systems has many potential applications. Pattern recognition in general and real time facial recognition in particular as a cyber investigation tool is very promising and cutting edge technological application in this domain. So much so that a very large scale research activity is going on worldwide with several research groups doing fantastic research work in this domain. Increased demand for secured systems has led the researchers to find solutions in terms of accessing control, verifying identities, securing cyber physical systems, copyright protection and content authentication (Mishra, Rajpal & Bala, 2018) and developing new surveillance systems that are robust and impenetrable (Zhao, Chellappa, Phillips & Rosenfeld, 2003), (Turk & Pentland, 1991), and (Gumus, Kilic, Sertbas & Ucan, 2010).

There is a strong requirement of the reliable systems that are not only robust / secure but are also capable of recognizing individuals accurately. To this end, a number of technological advancements have been carried out. Finger and iris based biometric authentication and its convergence to machine learning techniques has paved the way to developing automatic face recognition (Jain & Li, 2011). There is, however, one difference between the two. While the biometric authentication such as iris recognition is very advanced and individualistic in nature along with its being intrusive, the face recognition is equally if not more advanced but is essentially convenient, natural and also non-intrusive in its implementation (Chude-Olisah, Sulong, Chude-Okonkwo & Hashim, 2013). There are several facial recognition schemes available at present, yet they do not provide the precision and efficiency needed for a reliable recognition. Thus, there is a clear paucity of such schemes as a pre-requisite to developing precise and time efficient cyber physical systems and for computer vision applications.

This problem can be classified into two basic categories – first is related to facial recognition on a global scale limited to the face of the subject and the second is related to detecting and classifying facial expressions. There are, however, several common aspects between these two categories. These include – acquisition of properly

recognizable images and constructing their datasets, extraction and pre-processing of the relevant features followed by its classification thus leading either to recognition of the face itself or the facial expression in particular. For this purpose, several classifiers have been employed with varying degree of training and testing accuracy and time consumption in this process. This also depends upon the size of the image datasets, the number and type of images included in the datasets along with their orientation and light illumination levels. As far as academic research is concerned, several standard image datasets are used in practice. These include AT&T and YALE datasets which are used in this experimental work. This is similar to other research groups who have worked on the same datasets. There are multiple subjects considered in these datasets along with their images with varying orientations. However, one aspect is very important. This is related to the variations in the illumination levels and its possible effect on the performance of the classifier employed for the purpose. This issue is found to greatly reduce the performance of the face detection schemes in general and is therefore extremely crucial for the research community. The existing systems are highly sensitive to light variations (Du & Ward, 2005). Facial images taken under a condition where illumination was un-controlled suffered a non-uniform illumination. To cope up with this, certain adaptive techniques are being used. These adaptive techniques use normalization of the illumination by amending it to restore the image features to its original form. These normalization techniques include – Logarithmic Transform (LT), Gamma Correction (GC) and Histogram Equalization (HE) etc.(Chude-Olisah, Sulong, Chude-Okonkwo & Hashim, 2013), (Rujirakul & So-In, 2018).

As far as classifiers are concerned, there are facial recognition schemes reported widely in the literature which employ Linear Discriminant Analysis (LDA) and multi-class support vector machines (Gan & He, 2009), (Lihong, Ying, Yushi, Cheng & Yi, 2009). Besides, these schemes, Extreme Learning Machine (ELM) and its kernelized variant K - ELM are used for facial recognition schemes by Zong et al. (Zong & Huang, 2011), (Zong, Zhou, Huang & Lin, 2011). Improvement in performance of the facial classification using ELM based on facial views was reported by Iosifidis et al. (Iosifidis, Tefas & Pitas, 2015). Rujirakul et al. (Rujirakul & So-In, 2018) used Histogram Equalization (HE) along with Principal Component Analysis (PCA) in conjunction with ELM for facial recognition. On similar lines, Independent Component Analysis (ICA) method is used along with the hybrid architecture of Standard Particle Swarm Optimization (PSO) and ELM to obtain the facial recognition rates upto 93% (Wang, Zhang & Miao, 2018). Lately, the local difference binary (LDB) descriptors and fuzzy logic with Histogram of Oriented Gradients (HOG) were fused for developing efficient facial recognition systems (Salhi, Kardouchi & Belacel, 2013), (Wang, Zhang & Miao, 2018).

In this paper, the Histogram of Oriented Gradients (HOG) features are extracted from the face of the given subject by applying Viola – Jones algorithm to identify the regions corresponding to the face. This feature selection method is used as a pre-processing technique to modulate the overall brightness of a face image to a pre-defined canonical form which essentially discards the consequence of varying illumination. This is a very important step and for each image in the dataset, the HOG extracts crucial features which form the basis for training of the neural machine. The construction of the datasets out of these HOG based features is described in detail in Section 4. These datasets are supplied to the OS-ELM in the multi-class classifier mode by selecting either 80:20 or 70:30 split ratios of the data points for training and testing respectively.

The paper is structured as follows. Section 2 describes the basics of ELM. Section 3 gives an outline of Viola – Jones Algorithm. Section 4 gives a brief introduction of HOG feature selection and extraction method. Section 5 gives the proposed methodology and detailed scheme for facial recognition. The results are compiled and analyzed in Section 6 and finally, the paper is concluded in Section 7.

ONLINE SEQUENTIAL EXTREME LEARNING MACHINE (OS – ELM)

The ELM is a batch learning algorithm which assumes that the training examples are fed for training in one go. However, the training examples may be available one-by one or chunk by-chunk. To address such needs, Liang et al.(Liang, Huang, Saratchandran & Sundararajan, 2006) proposed Online Sequential ELM (OS-ELM), based on recursive least-squares (RLS) algorithm, as a modification of the primitive ELM (Huang, 2004), (Huang, Zhu & Siew, 2006). The OS – ELM works in two phases, namely, the boosting phase and the sequential learning phase.

In the boosting phase, we initiate with a mini – training set T_0 comprising of L training instances and initializes the hidden nodes parameters randomly. We then compute the initial hidden layer output matrix G_0 followed by computation of initial output weight $\beta^{(0)}$ using the equations (1) and (2) after which k is set to 0.

$$G_0 = [g(a_1 \cdot x_i + b_1), \dots, g(a_L \cdot x_i + b_L)]^T, i = 1, \dots, L \quad (1)$$

The boosting phase data will be discarded once the boosting phase is completed.

$$\beta^{(0)} = H_0 G_0^T T_0, \text{ where } H_0 = (G_0^T G_0)^{-1} \text{ and } T_0 = [t_1 \dots, t_L]^T \quad (2)$$

Second phase of OS – ELM i.e. Sequential Learning Phase commences either on a one – by – one or chunk – by – chunk (with fixed or varying size) basis as desired. This phase computes the hidden layer output matrix G_{k+1} and a new output weight $\beta^{(k+1)}$ for each subsequent chunk $k+1$ using the equations (3) – (5).

$$G_{k+1} = [g(a_1 \cdot x_i + b_1), \dots, g(a_L \cdot x_i + b_L)]^T, i = L + 1, L + 2, L + 3, \dots \quad (3)$$

Once a piece of data is used, it is discarded and not used any more.

$$H_{k+1} = H_k - \frac{H_k G_{k+1} G_{k+1}^T H_k}{1 + G_{k+1}^T H_k G_{k+1}} \quad (4)$$

$$\beta^{(k+1)} = \beta^{(k)} + H_{k+1} G_{k+1} (t_i^T - G_{k+1}^T \beta^{(k)}) \quad (5)$$

The time complexity of the OSELM must be calculated for its two phases, namely, the initial boosting phase and the sequential learning phase.

Boosting Phase:

1. The first step in the boosting phase involves the computation of initial hiddenlayer output matrix with the time complexity of $O(LN_L m) \approx O(L^2 m)$, where L is the number of hidden layer neurons, N_L is the number of initial training instances and m is size of feature set. Note that, N_L is taken to be L in this phase.
2. The computation of output weight matrix of the initial network has the time complexity equal to $O(L^3 + L^2 N_L + L N_L) \approx O(L^3)$.

The two steps of the boosting phase has the time complexity equal to $O(L^2 m + L^3) \approx O(L^2 m)$.

Sequential Learning Phase:

For each newly arrived training instance,

1. The time taken in the computation of hidden layer output matrix is $O(Lm)$, where L is the number of hidden layer neurons, m is the number of features of the current training instance.
2. The time complexity of the hidden layer output matrix is $O(L^3)$. Note that, the time complexity of the sequential learning phase for single training instance is equal to $O(Lm + L^3)$. Hence, the time taken considering all the newly arrived training instances must be equal to $O((N - N_L)Lm + L^3) \approx O(N(Lm + L^3)) \approx O(NLm)$

Therefore, the total time complexity of OS-ELM comprising of two above phases is equal to $O(NLm + L^2 m) \approx O(Nm)$.

VIOLA-JONES ALGORITHM

Viola-Jones algorithm is an object detection framework put forward by Paul Viola and Michael Jones in 2001 (Viola & Jones, 2004). Viola Jones object detector is based on a binary classifier that produces a positive output when the search window consists of the desired object otherwise it returns a negative output. The classifier may be used a number of times as the window slides over the image under test.

The binary classifier used in the algorithm is realized using several layers of hierarchy forming an ensemble classifier (Mathworks, 2012). The said classifier operates by classifying images based on value of simple features. This is observed to operate much faster than a system which basis classification on a pixel – based system (Viola & Jones, 2004). The Viola Jones algorithm exercises control over three features as dictated by Viola et al. in (Viola & Jones, 2004) viz., Two – Rectangle feature, Three – Rectangle feature, and Four – Rectangle feature. The framework put forward by the group is noted to have following stages: (i) Haar Feature Selection, (ii) Integral Image Generation, (iii) Adaboost Training, and (iv) Cascading Classifiers. This is represented as a flowchart in Fig. 1.

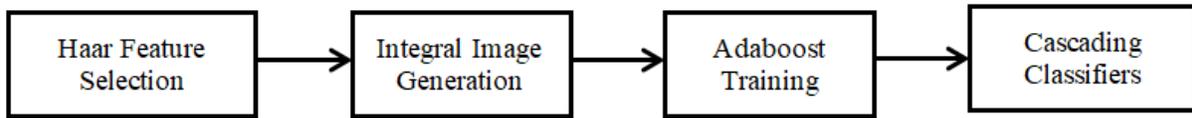


Figure 1. A depiction of the Viola – Jones Algorithm involving four important stages

The Haar feature selection is performed through Haar basis function that is based on the three features as listed above and generally includes pixel summation of involved adjacent rectangular areas and then calculates the difference between these sums. A depiction of Haar features relative to the corresponding detection window is shown in Fig. 2.

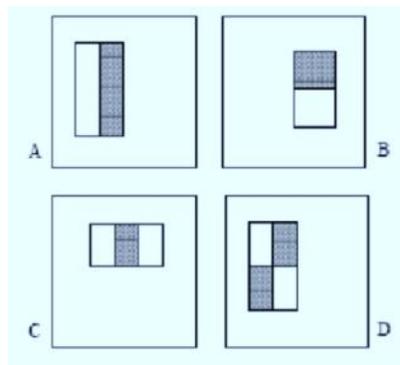


Figure 2. Depiction of rectangle features shown relative to the detection window

The integral image is then created which is used to evaluate the rectangular features in a constant time. Since the number of features can vary greatly, Adaboost or Adaptive Boosting algorithm is used to select best features and to train the classifiers that use them. This is responsible for creation of a “strong” classifier which is viewed as a linear weighted combination of simple “weak” classifiers. Finally, in cascading, each stage consisting of “strong” classifiers are grouped into several stages. Each stage is responsible for determining whether a sub – window consists of a face or not as depicted in Fig. 3. The algorithm described here is implemented using a MATLAB inbuilt routine as described in (Mathworks, 2013).

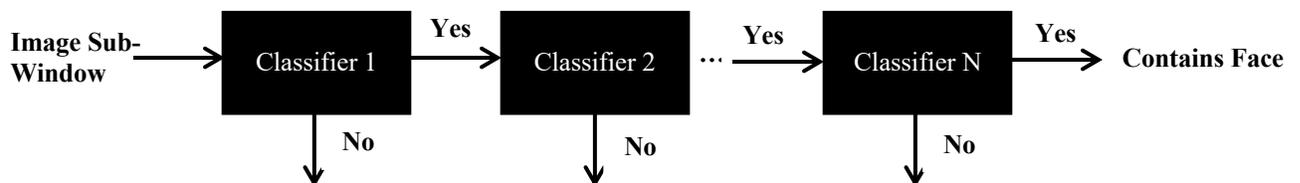


Figure 3. Depiction of the working flow of classifiers in Viola – Jones algorithm

HISTOGRAMS OF ORIENTED GRADIENTS

Histograms of oriented gradients (HOG) finds applications in object and pattern recognition domain as it is capable of extracting crucial information even from the images that are obtained under garbled environments (Korkmaz, Akçiçek, Bınoł & Korkmaz, 2017). It is therefore well suited for tackling the facial recognition problem. The feature extraction process of HOG is based on extracting information about the edges in local regions of a target image (Korkmaz, Akçiçek, Bınoł & Korkmaz, 2017). In other words, the HOG feature extraction is primarily the characterization of the orientation and magnitude values of the pixels in an image (Dalal, & Triggs, 2005). That is, it defines an image in terms of groups of local histograms that point to local regions of an image.

The features of HOG can be seen on a grid of rose plots spaced uniformly. The grid dimensions depend upon the size of the cell and image. Thus, every rose plot depicts the gradient orientations distributed in a HOG cell. In a cell histogram, the length of the petals in a rose plot refers to the contribution of every orientation. For the gradient directions, the plot indicates the directions of the edges that are normal. MATLAB inbuilt routine are applied using HOG feature extraction (Mathworks, 2013).

Thus, in a portion of image with 9 cells (Fig. 4), the HOG feature extraction routine takes input a block of $(m \times n)$ cells and arranges them in a vector as depicted in Fig. 5.

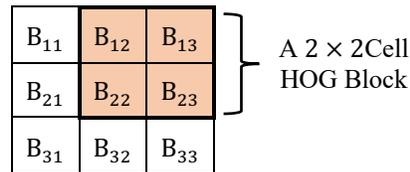


Figure 4. A portion of image realized using equal sized cells. Each cell consists of pixel values of that image portion

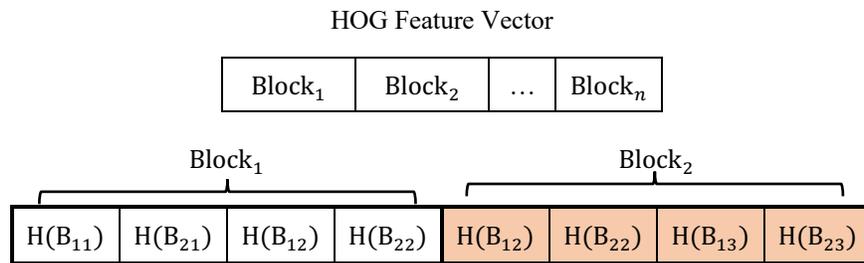


Figure 5. Depiction of composition of HOG feature vector. $H(B_{ij})$ represents the cell histogram at (i, j) position.

ALGORITHMIC DESCRIPTION OF THE PROPOSED SCHEME

The facial recognition methodology adopted in this work is depicted in Fig. 6.

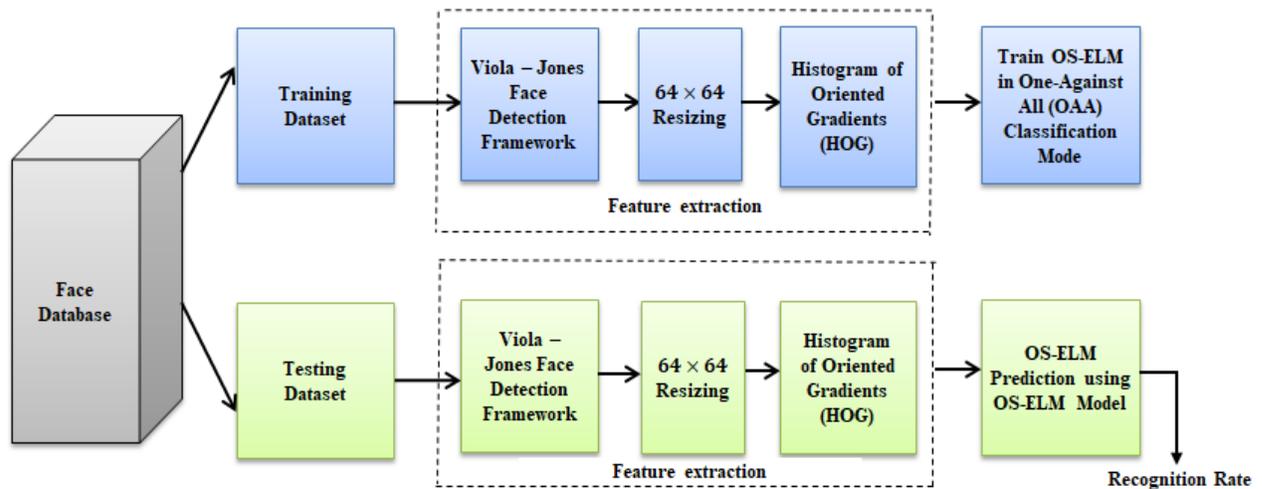


Figure 6. Block diagram depicting the face recognition approach employing OS-ELM in OAA mode along with Viola Jones object detection framework and HOG feature selection

Zong et al. in (Zong & Huang, 2011) have used two different approaches for their work. These are: (1) One – Against – All (OAA) and (2) One – Against – One (OAO) multi – class classification approaches using the primitive ELM. Given the multi – label dataset consisting of α different classes, OAA methodology takes into consideration α binary classifiers trained in such a way to distinguish each class and remaining classes. On the contrary, the OAO approach deals with one binary classifier to distinguish one pair of classes. This results in a total of $(\alpha - 1) * \alpha/2$ binary classifiers. As per the results compiled by Zong et al. (Zong & Huang, 2011), (Zong, Zhou, Huang & Lin, 2011) and also the related work using Linear Discriminant Analysis (LDA) and multiclass Support Vector Machine (SVM) (Gan & He, 2009), (Lihong, Ying, Yushi, Cheng & Yi, 2009), the OAA approach is found to give better performance as compared to the OAO approach. We, therefore, follow the former OAA based approach to deal with the classification of HOG feature based datasets for facial recognition.

Note that the face datasets are first split randomly into training and testing components as depicted in Fig. 6. To each image in dataset, the Viola – Jones algorithm is applied initially to detect the region which contains useful information pertaining to a subject’s facial features. The Viola – Jones algorithm is implemented by using the MATLAB routine given in (Mathworks, 2012). The extracted regions are resized to a uniform size of 64 by 64. This is done to ensure that the number of HOG features extracted in the subsequent stage is similar. This is important to ensure the uniformity of the processing for all the subjects of the datasets under consideration. The extraction of HOG features is carried out by another Matlab routine given at (Mathworks, 2013). These features are returned in the form of a row vector and consist of 1764 features for the proposed recognition scheme. All such row vectors corresponding to each image in training dataset is stacked one over the other to form the final dataset. This final dataset is supplied to the OS-ELM for multi – class classification as depicted in Fig. 7. In this dataset, a total of 1764 extracted features are considered along with the label fixed in column 1, thus making it a total of 1765 column dataset. There are 160 such rows pertaining to different facial samples of different subjects encapsulated within the dataset for training. Thus, overall the dataset has a size of 160×1765 . The dataset depicted in Figure 7 is considered as per the 80:20 split ratio into training and testing samples. A similar 70:30 split among training and testing samples is also considered in this experiment. Once the OS-ELM gets trained, the images in testing dataset are subjected to same processing and are supplied to the trained OS-ELM model to carry out classification. The recognition rate is computed by obtaining the total number of correct hits to the total images under the testing dataset.

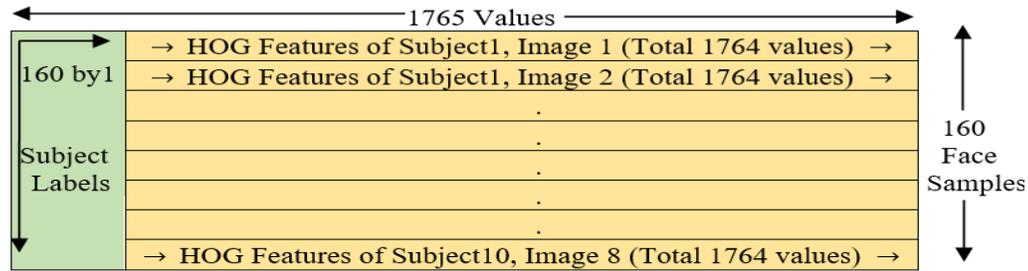


Figure 7. Depiction of the final dataset that is to be supplied to the OS-ELM considering an AT&T database with 80:20 split, giving 160 face images for training dataset [Note: The images are partitioned randomly into testing and training datasets. The depiction is for understanding purposes.]

EXPERIMENTAL RESULTS AND COMPARISONS

The proposed facial recognition scheme is tested on standard face recognition datasets viz. AT&T (ORL) and YALE (YALE). The AT&T consists of ten different grayscale images of each 40 distinct subjects with varying illumination levels, facial expressions and the images taken at different time instants. The size of each image is 92×112 with 8-bit per pixel representation. These images are available in portable gray map (PGM) format. The second dataset is YALE which consists of 11 grayscale images for each 15 distinct individuals. Again these images are available with different facial expressions, varying illumination levels and with miscellaneous eye wear. The size of each image in the YALE dataset is 243×320 with 8-bit per pixel representation. However, all these images are available in graphics interchange format (GIF). Some sample face images from both the datasets are shown in Fig. 8.

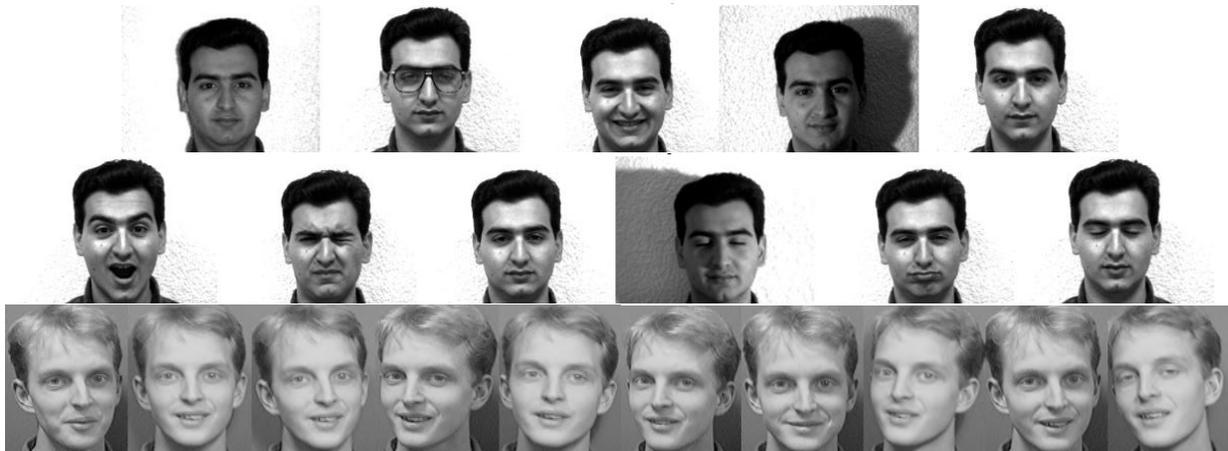


Figure 8. The first and second rows depict the face samples of a particular subject from YALE dataset, while the third row depicts face samples of another subject from AT&T Dataset

For each dataset, the images are split into training and testing samples as per 80:20 and 70:30 splitting ratios. The simulations are carried out using MATLAB 9.4 running on the Windows 10 Home Edition with an 8GB of memory and an i5 7300 HQ (2.50 GHz) processor. The results are tabulated in Table 1 and Table 2 and are depicted in the form of curve plots in Fig. 9. The recognition rate mentioned in Table 1 and Table 2 and the depiction in Fig. 9 correspond to the average recognition rate so obtained after 20 iterations. This is done to average out the error that emanates due to not so good generalization capabilities of OS-ELM. The OS-ELM is a very fast neural network which operates in milliseconds regime as shown by the data compiled in Table 1 and Table 2. The results presented in Table 1 and Table 2 also show a pattern that as the number of hidden neurons is increased, the recognition rates also increase to cross the 90% marker, but this happens only at the cost of the training time span. The training time span does increase but it still remains within the scale of few milliseconds.

Table 1. Effect of Variation of number of Hidden Neurons on the Recognition Rate and Testing Time Span for AT&T Dataset for 80:20 and 70:20 splitting ratios

No. of Hidden Neurons (L)	80:20 Split			70:30 Split		
	Recognition Rate (%)	Training Time (s)	Testing Time (s)	Recognition Rate (%)	Training Time (s)	Testing Time (s)
100	81.81	0.0625	0.2500	78.41	0.0156	0.1094
150	85.25	0.0156	0.0469	80.54	0.0625	0.1563
200	84.06	0.0156	0.0625	76.70	0.0625	0.0469
250	79.94	0.0938	0.1563	67.46	0.0781	0.2188
1000	94.06	1.3594	0.5625	93.00	1.3125	0.4063

Table 2. Effect of Variation of number of Hidden Neurons on the Recognition Rate and Testing Time Span for YALE Dataset for 80:20 and 70:20 splitting ratios

No. of Hidden Neurons (L)	80:20 Split			70:30 Split		
	Recognition Rate (%)	Training Time (s)	Testing Time (s)	Recognition Rate (%)	Training Time (s)	Testing Time (s)
100	98.29	0.0156	0.0938	98.41	0.0313	0.2188
150	99.41	0.0625	0.1875	99.52	0.0625	0.2656
200	99.45	0.0781	0.1875	99.49	0.0625	0.3906
250	99.58	0.0781	0.1406	99.52	0.125	0.4219
1000	100	1.5156	0.625	100	1.5938	0.8438

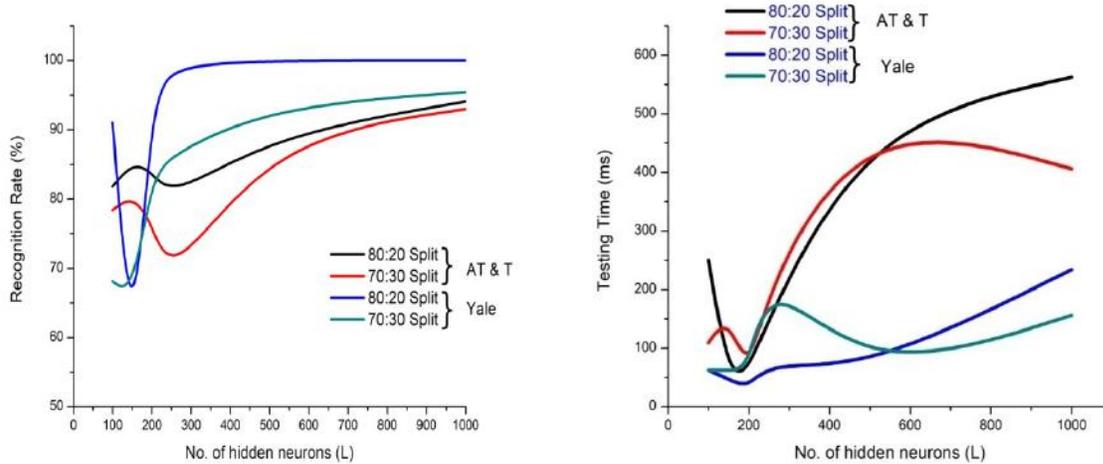


Figure 9. Curves depicting the effect of hidden neurons on (a) Recognition Rate, and (b) Testing Time for the two datasets under 80:20 and 70:30 splitting ratios

It is observed from the results summarized in Table 1 and Table 2 that the recognition rate outcomes are better in case of 80:20 split in comparison to 70:30 splitting ratio. This too even better in terms of recognition rate of more than 90% and the training and testing time spans computed thereof in case of YALE dataset. As the grayscale images saved in YALE dataset are available in GIF file format, it is suggested that the images which are captured at different orientations with varying illumination need to be stored in this format. There is another similarity between the results pattern obtained for AT&T and YALE datasets. In both the cases, the recognition rate dips at some threshold value of the number of hidden neurons, and then it is found to be monotonically increasing. The threshold in case of AY&T and YALE are found to be different. In case of former it is 250 while for latter, it is 150 (Fig. 9). Comparatively, overall results obtained for YALE are far superior to those of AT&T. Further, among YALE, the 80:20 splitting is better placed than 70:30 split ratio. This is true both for the recognition rate and the training and testing time spans. Therefore, according to us, the optimized value of $L = 150$ for which both recognition rate (%)

and testing time (ms) are better placed in YALE dataset in comparison to AT&T dataset. The maximum value of the recognition rate and the testing time span is obtained for $L = 1000$.

In second phase of this experimental work, to evaluate the performance of the proposed facial recognition scheme, we compare our results with some state – of – the – art methods for different datasets under Table 3 and Table 4.

Table 3. Comparison of recognition rate and running time spans for some state – of – the art facial recognition techniques for AT&T dataset taking 1000 Hidden Neurons

AT&T Dataset (70:30 Split)	Recognition Rate (%)	Training Time(ms)
PCA – NN (Zong, 2011)	91.67	-
LDA – NN (Zong, 2011)	95.50	-
DLA – NN (Zong, 2011) 32×32	97.62	-
PCA – ELM (Zong, 2011) Resizing	97.11	6380
LDA – ELM (Zong, 2011)	96.46	7680
DLA – ELM (Zong, 2011)	98.44	4540
SPSO – ELM (Wang &Guo, 2017)	93.57	102×10^3
Fuzzy – HOG (Salhi, 2013)	92.60	-
LDB – HOG (Wang, 2018)	92.00	-
Proposed Work 32×32 Resizing	95.00	1218
Proposed Work 64×64 Resizing	93.00	1325

Table 4. Comparison of recognition rate and running time spans for some state – of – the art facial recognition techniques for YALE dataset taking 1000 Hidden Neurons

YALE Dataset (70:30 Split)	Recognition Rate (%)	Training Time (ms)
PCA – NN [(Zong, 2011)]	60.50	-
LDA – NN (Zong, 2011)	83.83	-
DLA – NN (Zong, 2011)	87.25	-
PCA – KELM (Zong, 2011)	85.58	39
LDA – KELM (Zong, 2011) 40×40	85.92	23
PCA – ELM (Zong, 2011) Resizing	82.08	2210
LDA – ELM (Zong, 2011)	83.75	1240
DLA – ELM (Zong, 2011)	89.04	1680
LDB – HOG (Wang, 2018)	92.00	-
Proposed Work 40×40 Resizing	93.33	1203
Proposed Work 64×64 Resizing	100	1593

A close observation of the results summarized in Table 3 and Table 4 confirms that the proposed facial recognition scheme outperforms several other existing schemes. This is more particularly observed in case of Table 4 which compiles the data for the YALE dataset. The results in this case are clearly better both in terms of recognition rate and the testing time spans wherever available. This is primarily due to the use of OS-ELM based classification which is not only faster but also yields better results in terms of recognition percentage. These outcomes are supplemented due to the use of GIF format of the images saved in the YALE dataset if they are compared with those obtained for AT&T. We, therefore, conclude that our facial recognition technique not only gives better results in terms of the recognition rate (%), but our testing time span is also measured in ms domain thereby suggesting that all necessary procedures – pre-processing, feature extraction and classification of images are carried out in real time. This comparison is also done with other existing schemes involving the architectures of ELM obtained by amalgamating PCA, LDA and LDB etc. The OS-ELM classifier produces better outcomes as compared to all these schemes. This is possible only due to the use of a combination of several efficient existing methods in this work. The combination is that of the Viola-Jones Algorithm for object identification, HOG based feature selection and the use of Online Sequential Extreme Learning Machine (OS-ELM) for multi-class classification. The proposed combination of these procedures along with their packaged real time execution brings in the desirable novelty of the proposed facial recognition scheme.

CONCLUSIONS

A novel facial recognition scheme as part of a global cyber physical system development is proposed in this paper. This scheme makes use of OS-ELM based classifier which classifies the HOG feature vectors obtained by using Viola-Jones Algorithm. The novelty of this scheme is in its real time implementation for facial recognition without compromising on the recognition rate in comparison to several other existing schemes. Two standard facial image datasets – AT&T and YALE are considered in the work. These datasets are profoundly used by the research community all over the world. We conclude that the OS-ELM classifier is an optimized one for this purpose which gives better results in comparison to existing schemes. Our results are additionally supplemented by the use of GIF images of the YALE dataset for which results are even better in comparison to those for AT&T. Overall, we find that a very high recognition rate (%) is achieved in the millisecond time scale more particularly for the YALE dataset. Thus, the proposed scheme is proved to be practically implementable and effective.

REFERENCES

- Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4), 399-458.
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 71-86.
- Gumus, E., Kilic, N., Sertbas, A., & Ucan, O. N. (2010). Evaluation of face recognition techniques using PCA, wavelets and SVM. *Expert Systems with Applications*, 37(9), 6404-6408.
- Jain, A. K., & Li, S. Z. (2011). *Handbook of face recognition*. New York: springer.
- Chude-Olisah, C. C., Sulong, G., Chude-Onkonkwo, U. A., & Hashim, S. Z. (2013, October). Illumination normalization for edge-based face recognition using the fusion of RGB normalization and gamma correction. In *2013 IEEE International Conference on Signal and Image Processing Applications* (pp. 412-416). IEEE.
- Du, S., & Ward, R. (2005, September). Wavelet-based illumination normalization for face recognition. In *IEEE International Conference on Image Processing 2005* (Vol. 2, pp. II-954). IEEE.
- Zong, W., & Huang, G. B. (2011). Face recognition based on extreme learning machine. *Neurocomputing*, 74(16), 2541-2551.
- Zong, W., Zhou, H., Huang, G. B., & Lin, Z. (2011, June). Face recognition based on kernelized extreme learning machine. In *International Conference on Autonomous and Intelligent Systems* (pp. 263-272). Springer, Berlin, Heidelberg.
- Iosifidis, A., Tefas, A., & Pitas, I. (2015). Enhancing ELM-based facial image classification by exploiting multiple facial views. *Procedia Computer Science*, 51, 2814-2821.
- Rujirakul, K., & So-In, C. (2018, January). Histogram equalized deep PCA with ELM classification for expressive face recognition. In *2018 International Workshop on Advanced Image Technology (IWAIT)* (pp. 1-4). IEEE.
- Wang, Y., Li, H., & Guo, Y. (2017, December). Face recognition based on ICA and SPSO-ELM. In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (pp. 602-606). IEEE.
- Zhang, G. Y., Peng, S. Y., & Li, H. M. (2008, July). Combination of dual-tree complex wavelet and SVM for face recognition. In *2008 International Conference on Machine Learning and Cybernetics* (Vol. 5, pp. 2815-2819). IEEE.
- Gan, J. Y., & He, S. B. (2009, July). Face recognition based on 2DLDA and support vector machine. In *2009 International Conference on Wavelet Analysis and Pattern Recognition* (pp. 211-214). IEEE.
- Lihong, Z., Ying, S., Yushi, Z., Cheng, Z., & Yi, Z. (2009, June). Face recognition based on multi-class SVM. In *2009 Chinese Control and Decision Conference* (pp. 5871-5873). IEEE.

- Salhi, A. I., Kardouchi, M., & Belacel, N. (2013, January). Histograms of fuzzy oriented gradients for face recognition. In *2013 International Conference on Computer Applications Technology (ICCAT)* (pp. 1-5). IEEE.
- Wang, H., Zhang, D., & Miao, Z. (2018, July). Fusion of LDB and HOG for Face Recognition. In *2018 37th Chinese Control Conference (CCC)* (pp. 9192-9196). IEEE.
- Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on neural networks*, *17*(6), 1411-1423.
- Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, *70*(1-3), 489-501.
- Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Real-time learning capability of neural networks. *IEEE Trans. Neural Networks*, *17*(4), 863-878.
- Huang, G. B. (2004). The Matlab code for ELM. Available at: <http://www.ntu.edu.sg/home/egbhuang>.
- Mishra, A., Rajpal, A., & Bala, R. (2018). Bi-directional extreme learning machine for semi-blind watermarking of compressed images. *Journal of information security and applications*, *38*, 71-84.
- Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.
- Lo, C., & Chow, P. (2012, December). A high-performance architecture for training viola-jones object detectors. In *2012 International Conference on Field-Programmable Technology* (pp. 174-181). IEEE.
- MathworksMatlab.: Detection objects using the Viola – Jones algorithm. *MathworksMatlab Documentation R2018b*, (2012) Available at: <https://in.mathworks.com/help/vision/ref/vision.cascadeobjectdetector-system-object.html>
- Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *international Conference on computer vision & Pattern Recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE Computer Society.
- Korkmaz, S. A., Akçiçek, A., Bınoğlu, H., & Korkmaz, M. F. (2017, September). Recognition of the stomach cancer images with probabilistic HOG feature vector histograms by using HOG features. In *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)* (pp. 000339-000342). IEEE.
- MathworksMatlab.: Extract histogram of oriented gradients (HOG) features. *MathworksMatlab Documentation R2018b*, (2013) Available at: https://in.mathworks.com/help/vision/ref/extracthogfeatures.html?s_tid=doc_ta
- AT&T Laboratories Cambridge.: The AT&T Dataset (formerly 'The ORL Dataset of Faces'). Available at: http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.zip
- YALE Face Dataset. Available at: <http://cvc.cs.YALE.edu/cvc/projects/YALEfaces/YALEfaces.html>