Face Detection using Min-Max Features Enhanced with Locally Linear Embedding

Rahmat Hidayat [1], Fatin Nabila Jaafar [2], Ihsan Mohd Yassin [2,*], Azlee Zabidi [2], Fadhlan Hafizhelmi Kamaru Zaman [2], Zairi Ismael Rizman [3]

[1] *Department of Information Technology, Politeknik Negeri Padang, Indonesia*
[2] *Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[3] *Faculty of Electrical Engineering, Universiti Teknologi MARA, 23000 Dungun, Terengganu, Malaysia*

***Abstract* – Face detection is critical function in many embedded applications such as computer vision and security as it is widely used as preprocessor for face recognition systems. As a preprocessor, the face detection system needs to extract features from a region of interest and classify them quickly as either face or non-face. In our previous works, we have devised a feature representation method called Min-Max (MMX) feature that allows representation of a region of interest using a few data points based on the unique characteristics of vertical and horizontal summation of face regions. In this paper, we attempt to improve the classification accuracy of MMX by integrating a technique called Locally Linear Embedding (LLE), a powerful dimensionality and feature enhancement algorithm that has been used successfully in many pattern recognition tasks. To test the performance of the proposed enhancement, the LLE-treated features were compared with non-treated features using a Multi-Layer Perceptron (MLP) neural network classifier. The results indicate an increase (+1.2%) in classification accuracy of the MLPs, demonstrating the ability of LLE to enhance the representation of MMX features.**

***Keywords* – Face detection, Min-Max features (MMX), Locally Linear Embedding (LLE), Multi-Layer Perceptron (MLP), Artificial Neural Network**

DOI: 10.18421/TEM73-27
https://dx.doi.org/10.18421/TEM73-27

Corresponding author: Ihsan Mohd Yassin,
Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
Email: ihsan.yassin@gmail.com

Received: 08 July 2018.
Accepted: 15 August 2018.
Published: 27 August 2018.

1. Introduction

With the present and continuous threat of terrorism and criminal activities, surveillance and identity verification are receiving paramount attention from security practitioners. However, the need for them must be balanced with the perceived privacy concerns and obtrusiveness of the identification method [1].

Biometrics refers to identification of individuals based on his or her unique biological characteristics [2]. The area is expected to grow at an exponential rate from 2015 to 2024 according to research performed by [3] (Fig. 1.). Biometrics identification is divided into many methods, such as fingerprint [4], iris [5], gait [6], ear shape [7], facial recognition [8], and many others. Each method has their own advantages and weaknesses, but facial recognition is particularly useful in mass surveillance as it is non-obtrusive and can be integrated with current surveillance architectures with minimal effort [2].

Face recognition has received significant attention due to its functionality and purpose for a wide variety of applications. This is because human faces contain unique and naturally-identifiable characteristics for identifying a person's identity. Face recognition systems use computer software to identify a face by extracting important features and comparing them to a list of known individuals in a database. Recognition is a computationally-intensive task. Therefore, care must be given to select only the most prospective regions in an abstract image and focus the recognition only in those areas.

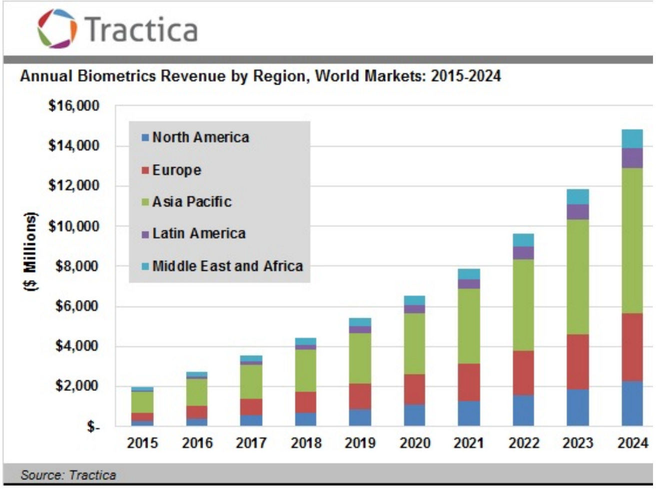

Fig. 1. Annual forecast of biometrics revenue by world regions [3].

This is achieved by integrating a face detection system, a lightweight system used specifically as a preprocessor that scans an image for prospective face regions and pass them on to the recognition system. Detection in unconstrained environments is a particularly difficult task, as it is affected by various ambiguities such as variations in pose, occlusion, expression, illumination, scale, color and texture [9]. Face detection systems, therefore must possess two important characteristics, namely speed (ability to scan images quickly for prospective regions) and accuracy (ability to accurately determine face regions to be passed on to the recognition system).

To achieve this, research by [2] presented a feature representation method called Min-Max features (MMX), that can easily find prospective face regions based on the summation of pixel intensities of vertical and horizontal regions in an area of interest. Comparisons with the benchmark face detection system by [10] showed that the MMX method was able to differentiate between face/non-face regions with good accuracy and much less data points to represent the data.

Locally Linear Embedding (LLE) is an unsupervised dimensionality reduction algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs [11]. It does this by taking advantage of the local geometry and then pieces it together to preserve the global geometry on a lower-dimensional space [11]. It has been proven to be very successful in supporting pattern classification tasks [12],[13],[14].

In this paper, we propose to apply Locally Linear Embedding (LLE) to the MMX features and compare the results with non-LLE implementation. We classify them using two separate MLP classifiers trained on LLE and non-LLE features. Then, some analysis was performed based on feature size and classification accuracy of the proposed approach.

The remainder of this paper is organized as follows: several relevant researches are presented in Section II, followed by the research methodology in Section III. Results and discussions are presented in Section IV. Finally, concluding remarks are presented in Section V.

2. Literature Review

Artificial Neural Network (ANN)

ANN is a computing paradigm in Artificial Intelligence (AI) which mimics the learning ability of humans. The structure of ANNs is like that of the biological brain, where interconnections between the neuron synapses are responsible for our ability to learn from experience and observations [15].

The two primary use cases for ANNs are pattern recognition (to classify data as belonging to a certain group or other(s), see [2], [16], [17] for examples), and function approximation (approximating the value of the output given certain input values, see [18], [19], [36],[37],[38]). For the discussion in this paper, we limit ANN to its first application. Additionally, ANN is one of the popular approaches for both face recognition and detection. In recent years, ANN is used largely in the fields of image processing (compression, recognition, encryption) and pattern recognition [20].

The ANN architecture consists of interconnected nodes called neurons [10]. The most common type of ANN is the Multi-Layer Perceptron (MLP). MLPs organize their nodes into layers (input, hidden and output) and training is performed with a gradient-based algorithm called backpropagation [21], [22]. The input layer receives values from the input data and relays it to the hidden layer, which performs some processing and activation of the hidden nodes guided by the activation function [10]. The output layer is responsible to receive the inputs from the hidden layer and approximate the output desired. The layers interconnections are weighted, and the weights are adjusted during the training process [22]. The advantages of ANN are adaptive learning, self-organization, and robustness [23].

Locally Linear Embedding (LLE)

Locally Linear Embedding (LLE) was proposed by [11] as a dimensionality reduction algorithm as an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs by taking advantage of the local geometry and then pieces it together to preserve the global geometry on a lower dimensional space. The algorithm has been extensively used in many pattern recognition tasks with significant success [24–27].

Relevant Research

This section describes several researches in the field of face detection.

In [28], a low-computation face detection method for the Raspberry Pi embedded system is presented. The proposed detection system relies on a fusion of methods (Viola-Jones, CamShift tracking, and Kalman filter) to perform detection and tracking. The Viola-Jones classifier is an ensemble classifier used for face detection. A combination of CamShift and Kalman filtering algorithm was used to track the faces once the Viola-Jones classifier found them. The result indicates that the frame processing and detection rate were sufficiently fast under limited computational resources in the embedded system.

Reference [29] compared the face detection performance between feature-based and image-based methods for real-time face detection on video. Two detection methods were studied, namely feature-based (consisting of low-level analysis, feature analysis and active shape models) and image-based. Classification was performed using the Viola-Jones classifier optimized using AdaBoost to discard non-important features. Various popular file formats were tested, including AVI, MOV and MP4 formats. The authors found that the 3GP file format achieved the best classification rate while MPG took the least amount of processing time.

In [9], a weighted decision tree-based method was used to upgrade a deformable part-based model to increase face detection robustness under occlusion. The deformable template consists of 11 subtrees representing face components. Each of the subtrees were weighted to emphasize certain features over others based on the results of a psychological experiment. The decision is based on the threshold of a scoring function. Experiments performed on the frontal faces in the XM2VTS database demonstrated that the method performs well under different poses and occlusions.

A Field Programmable Gate Array (FPGA)-based Viola-Jones face detection system was presented in [30] for fast face detection on hardware. The initial design was mapped to FPGA using the OpenCL platform. A nested parallel architecture was used to accelerate memory access and computing. The FPGA was further improved by optimizing the number and size of the cores under realistic hardware constraints to achieve real-time detection capabilities, while non-critical tasks were delegated to the Central Processing Unit (CPU). The authors discovered that the proposed implementation could accelerate the computation up to 3,000% without significant loss in recall or precision compared to a CPU-only based implementation.

Another FPGA-based face detection implementation was found in [31], which compared the Crude method and the Viola-Jones algorithm for face detection. In the Crude method, pixel-based detection was performed on three different facial regions (eyes, nose and mouth) and implemented using FPGA. Meanwhile, the Viola-Jones algorithm was used to convert the input image into an integral image, which was then classified using a committee machine-based classifier. The results indicate that the Crude method was faster than Viola-Jones method for feature extraction.

3. Methodology

Figure 2. shows the different stages of the project. Each stage is detailed in Section Hardware Description to Section Classification using MLP.

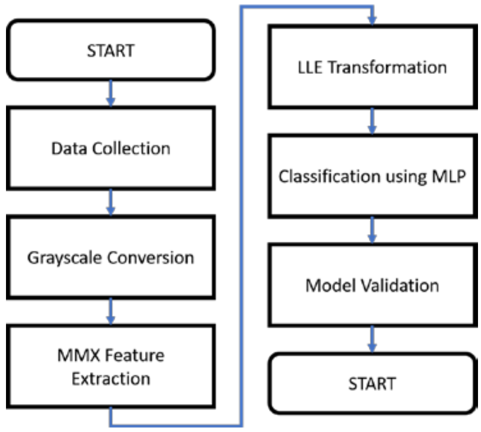

Fig. 2: Project Flowchart

Hardware Description

All experiments were implemented on an International Business Machines (IBM®)-compatible personal computer with an Intel Core i7 microprocessor running at 3.5 GHz and 12 GB of Random Access Memory (RAM). Testing was performed on MATLAB version R2016b with Microsoft Windows 7 Professional as the operating system.

Data Collection & Grayscale Conversion

Samples of face and non-face images were collected from the CBCL dataset and from various sources from the internet. All datasets were encoded using the 256-level grayscale color space. Each example was sized 20× 20 pixels, since investigation by [32] suggested that 20× 20 pixels windows were optimal for face detection. The Joint Picture Experts Group (JPEG) format was used to store the images. The number of faces and non-faces were equally distributed (250 faces and 250 non-faces) to ensure that MLP training results do not lean towards cases with the most examples. Several samples are shown in Figure 3. and Figure 4.

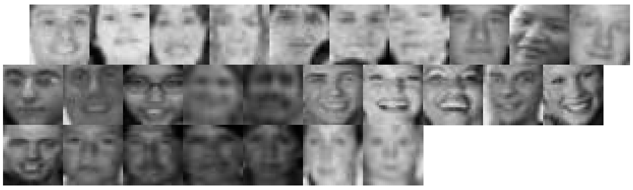

Fig. 3: Sample images used for classification (face class)

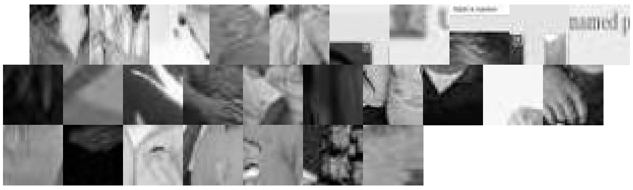

Fig. 4: Sample images used for classification (non-face class)

MMX Feature Extraction

The interested reader is referred to [2] for the MMX feature extraction process. Essentially, the MMX feature extraction method equally divides an image into three horizontal, and three vertical face regions. Then, the sum of horizontal and vertical face regions was calculated and combined to form the final MMX features.

Locally Linear Embedding (LLE) Transformation

After MMX feature extraction was performed, the features were transformed using LLE. The LLE method attempt to resolve a low-dimensional representation which approximates the high-dimensional features [33]. The method is exceedingly efficient as under certain circumstances, it is possible to derive an embedding solely from the geometric situation without recourse to scale, distance or connectivity between interval data [33].

The LLE algorithm relies on two parameters to perform dimensionality reduction. The first parameter is the nearest neighbors (k) to support the global geometry of the data. The second parameter, d_{max}, defines how much data is to be retained by LLE. Next, a set of weights is computed that can be used to reconstruct each data point. The embedding vector, Y, is computed next with the weights previously defined [33].

In the experiments performed, the LLE parameters were varied to determine the best configuration for the MMX features.

Classification using MLP

The next process is to perform classification using MLP. In this step, the number of hidden units was adjusted using a trial and error method to minimize the classification error. Both MLPs (for LLE and Non-LLE data) consisted of three layers (one input, one hidden, and one output layer). Since the problem posed here is a pattern classification, both the hidden and output layers were configured with the tangent-sigmoid activation function. The Early Stopping method was used to avoid overfitting. The Scaled Conjugate Gradient (SCG) algorithm was selected as the training algorithm as it demonstrates excellent performance in pattern classification tasks [34]. The performances of both MLPs were compared using the Confusion Matrix method, a standard visual method for examining the classification accuracy of binary classifiers.

4. Result and Discussion

MLP Results without using LLE (Benchmark)

The benchmark model uses only plain MMX features without subjecting them to LLE. Various hidden unit configurations were tested (five to 30, with increments of five). To avoid result differences due to initial weights settings, the initial random weights were locked using a predetermined seed in the Mersenne-Twister random number generator in MATLAB.

The results are shown in Figure 5. to Figure 10. All the hidden units tested produced good to excellent accuracy (above 80%), with some variation from 81.8% (minimum, at 15 hidden units) to 95.2% (maximum, at 20 hidden units). Based on the observations, it was found that MMX was already a powerful feature representation algorithm even before applying LLE features. This may be because it captures the relevant values and positions of dark and light areas in the face region (such as the eyes, nose, and mouth).

MLP Results using LLE

In this section, the MMX features were subjected to LLE prior to MLP training. Like Section MLP Results without using LLE (Benchmark), various hidden unit configurations were tested, with similar safeguards against initial weight differences. In addition to MLP hidden units, the LLE parameters d_{max} and k were also varied.

In the interest of space, the top five results with the best hidden unit, d_{max} and k are shown in Figure 11. to Figure 15. All the hidden units tested produced excellent accuracy (all classifiers registered above 94% accuracy), with smaller classification variations: 94.4% (minimum, at 30 hidden units, $d_{max} = 90$ and $k = 30$) to 96.4% (maximum, at 25 hidden units, $d_{max} = 80$ and $k = 20$). Based on the observations, it was found that LLE had helped improve the representation of the MMX features. This is because the LLE algorithm can represent the features in such a way that it maximizes the separability between face and non-face cases.

Furthermore, an additional advantage of using LLE is the number of features that can be explicitly controlled using the d_{max} parameter, which can specify the maximum number of dimensions for the LLE features. This flexibility could help adjust the features for classification in less powerful machines, such as Raspberry Pi and other types of microcontrollers.

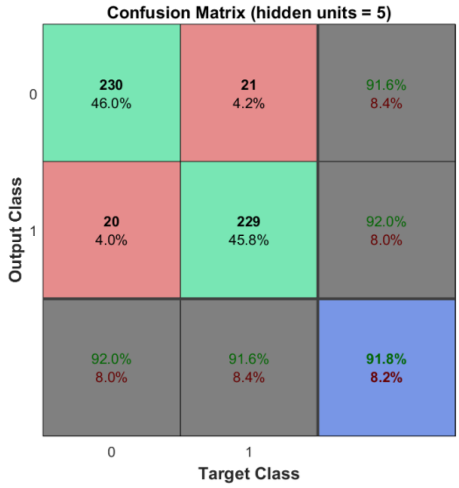

Fig. 5: MLP confusion matrix for hidden unit 5 (MMX only)

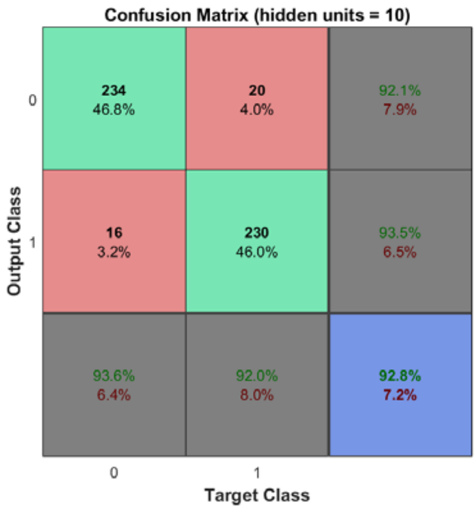

Fig. 6: MLP confusion matrix for hidden unit 10 (MMX only)

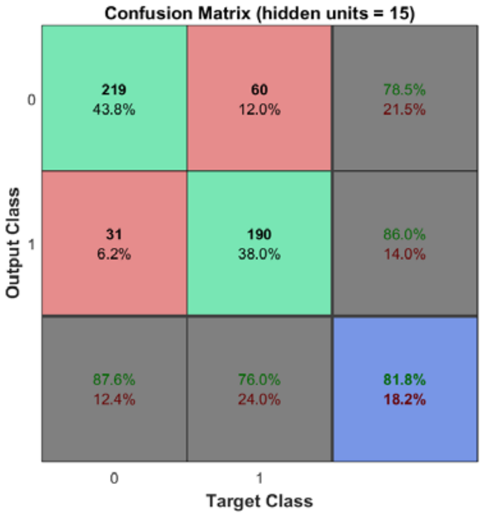

Fig. 7: MLP confusion matrix for hidden unit 15 (MMX only)

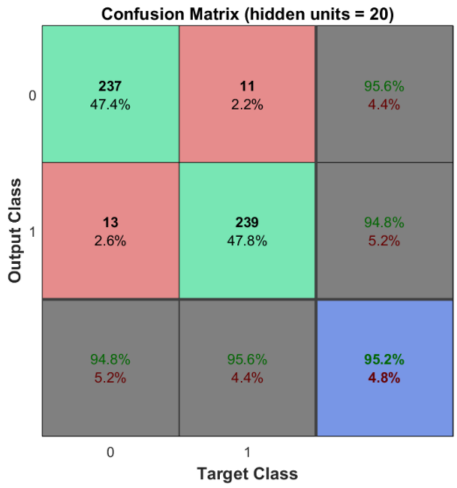

Fig. 8: MLP confusion matrix for hidden unit 20 (MMX only)

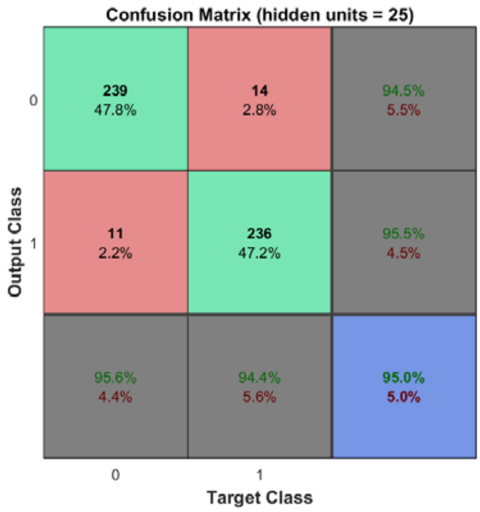

Fig. 9: MLP confusion matrix for hidden unit 25 (MMX only)

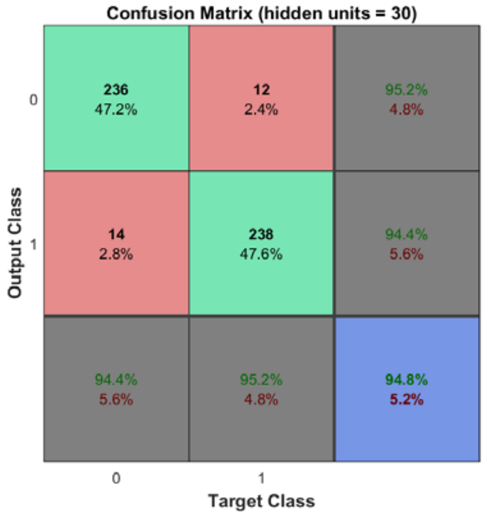

Fig. 10: MLP confusion matrix for hidden unit 30 (MMX only)

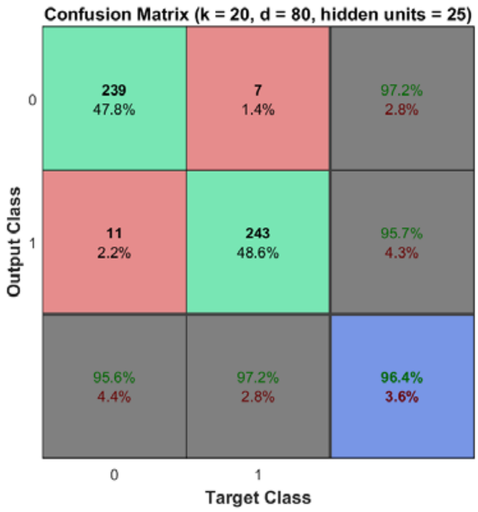

Fig. 11: MLP confusion matrix for LLE-treated features, hidden unit 25, $d_{max} = 80$ and $k = 20$

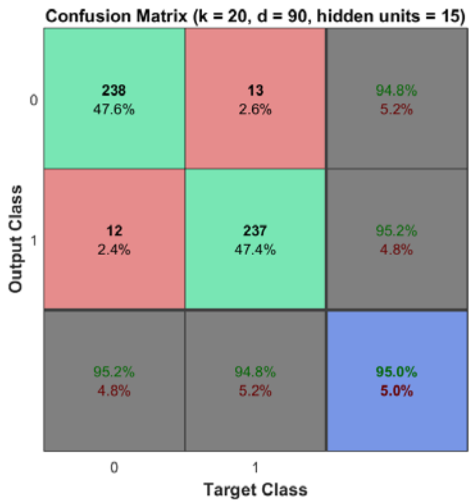

Fig. 12: MLP confusion matrix for LLE-treated features, hidden unit 15, $d_{max} = 90$ and $k = 20$

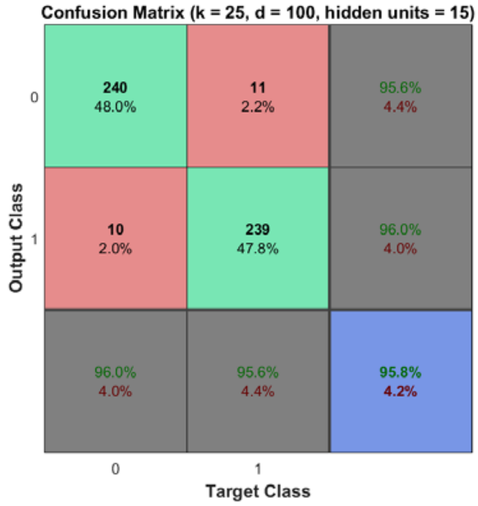

Fig. 13: MLP confusion matrix for LLE-treated features, hidden unit 15, $d_{max} = 100$ and $k = 25$

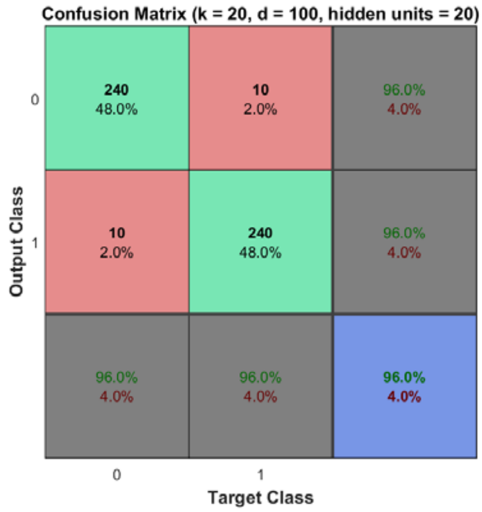

Fig. 14: MLP confusion matrix for LLE-treated features, hidden unit 20, $d_{max} = 100$ and $k = 20$

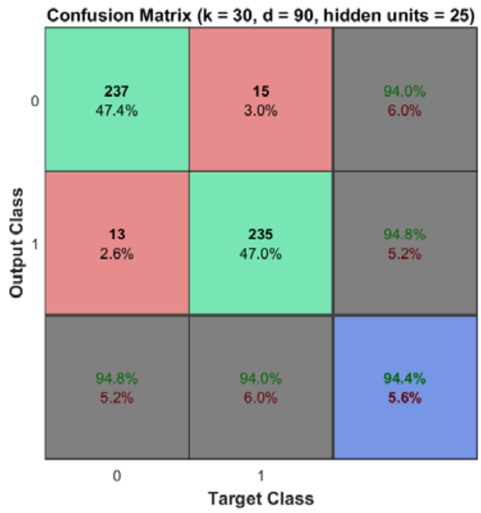

Fig. 15: MLP confusion matrix for LLE-treated features, hidden unit 25, $d_{max} = 90$ and $k = 30$

5. Conclusions

In this paper, LLE has been used to enhance MMX features for face detection. The results have demonstrated the ability of LLE to improve the MMX feature representation method and improve the classification accuracy of the face detection system [35]. Additionally, using the d_{max} parameter, the feature size can be explicitly controlled. This may prove useful in application areas where the feature size needs to be computationally inexpensive and fast.

References

[1] Chinomi, K., Nitta, N., Ito, Y., & Babaguchi, N. (2008, January). PriSurv: privacy protected video surveillance system using adaptive visual abstraction. In *International Conference on Multimedia Modeling* (pp. 144-154). Springer, Berlin, Heidelberg.

[2] Yassin, I. M. (2008). Face detection using artificial neural network trained on compact features and optimized using particle swarm optimization. *Faculty of Electrical Engineering, Universiti Teknologi MARA*.

[3] Tractica.(2017). Biometrics Market Revenue to Total $67 Billion Worldwide over the Next 10 Years.

[4] Chantal, M., Lee, S. W., & Kim, K. H. (2017, March). A Security Analysis and Reinforcement Design Adopting Fingerprints over Drawbacks of Passwords Based Authentication in Remote Home Automation Control System. In *Proceedings of the 6th International Conference on Informatics, Environment, Energy and Applications* (pp. 71-75). ACM.

[5] Nalla, P. R., & Kumar, A. (2017). Toward more accurate iris recognition using cross-spectral matching. *IEEE transactions on Image processing*, *26*(1), 208-221.

[6] Tang, J., Luo, J., Tjahjadi, T., & Guo, F. (2017). Robust arbitrary-view gait recognition based on 3d partial similarity matching. *IEEE Transactions on Image Processing*, *26*(1), 7-22.

[7] Abate, A. F., Nappi, M., & Ricciardi, S. (2017). I-Am: implicitly authenticate me person authentication on mobile devices through ear shape and arm gesture. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, (99), 1-13.

[8] Karczmarek, P., Kiersztyn, A., Pedrycz, W., & Dolecki, M. (2017). An application of chain code-based local descriptor and its extension to face recognition. *Pattern Recognition*, *65*, 26-34.

[9] Marčetić, D., & Ribarić, S. (2016, May). Deformable part-based robust face detection under occlusion by using face decomposition into face components. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016 39th International Convention on* (pp. 1365-1370). IEEE.

[10] Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(1), 23-38.

[11] Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, *290*(5500), 2323-2326.

[12] Dong, Y., Du, B., Zhang, L., & Zhang, L. (2017). Exploring locally adaptive dimensionality reduction for hyperspectral image classification: A maximum margin metric learning aspect. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(3), 1136-1150.

[13] Kanas, V. G., Zacharaki, E. I., Thomas, G. A., Zinn, P. O., Megalooikonomou, V., & Colen, R. R. (2017). Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma. *Computer methods and programs in biomedicine*, *140*, 249-257.

[14] Luo, F., Huang, H., Liu, J., & Ma, Z. (2017). Fusion of graph embedding and sparse representation for feature extraction and classification of hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, *83*(1), 37-46.

[15] Osuna, E., Freund, R., & Girosit, F. (1997, June). Training support vector machines: an application to face detection. In *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on* (pp. 130-136). IEEE.

[16] Zabidi, A., Yassin, I. M., Hassan, H. A., Ismail, N., Hamzah, M. M. A. M., Rizman, Z. I., & Abidin, H. Z. (2017). Detection of asphyxia in infants using deep learning convolutional neural network (CNN) trained on Mel frequency cepstrum coefficient (MFCC) features extracted from cry sounds. *Journal of Fundamental and Applied Sciences*, *9*(3S), 768-778.

[17] Hassan, H. A., Tahir, N. M., Zabidi, A., Yassin, I. M., & Karbasi, M. (2017). Classification of visualization exudates fundus images results using support vector machine. *Journal of Fundamental and Applied Sciences*, *9*(4S), 19-44.

[18] Indera, N. I., Yassin, I. M., Zabidi, A., & Rizman, Z. I. (2017). Non-linear autoregressive with exogeneous input (NARX) Bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators. *Journal of Fundamental and Applied Sciences*, *9*(3S), 791-808.

[19] Yassin, I. M., Abdul Khalid, M. F., Herman, S. H., Pasya, I., Wahab, N. A., & Awang, Z. (2017). Multi-Layer Perceptron (MLP)-Based Nonlinear Auto-Regressive with Exogenous Inputs (NARX) Stock Forecasting Model. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(3), 1098-1103.

[20] Moghaddam, B., & Pentland, A. (1997). Probabilistic visual learning for object representation. *IEEE Transactions on pattern analysis and machine intelligence*, *19*(7), 696-710.

[21] Yang, F., & Paindavoine, M. (2003). Prefiltering for pattern recognition using wavelet transform and neural networks. *Advances in imaging and Electron physics*, *127*, 126-208.

[22] Uwechue, O. A., & Pandya, A. S. (2012). *Human face recognition using third-order synthetic neural networks* (Vol. 410). Springer Science & Business Media.

[23] Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, *8*(1), 98-113.

[24] Mustapha, D., Zabidi, A., Sahak, R., Tahir, N. M., Yassin, I. M., Zaman, F. H. K., ... & Zan, M. M. M. (2017). Gait recognition using Kinect and locally linear embedding. *Journal of Fundamental and Applied Sciences*, *9*(3S), 755-767.

[25] Li, Z., Lai, Z., Xu, Y., Yang, J., & Zhang, D. (2017). A locality-constrained and label embedding dictionary learning algorithm for image classification. *IEEE transactions on neural networks and learning systems*, *28*(2), 278-293.

[26] Zhao, C., Wang, C., Hua, L., Liu, X., Zhang, Y., & Hu, H. (2017). Recognition of control chart pattern using improved supervised locally linear embedding and support vector machine. *Procedia engineering*, *174*, 281-288.

[27] Sun, W., Yang, G., Du, B., Zhang, L., & Zhang, L. (2017). A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 4032-4046.

[28] Soetedjo, A., & Somawirata, I. K. (2016, August). Implementation of face detection and tracking on a low cost embedded system using fusion technique. In *Computer Science & Education (ICCSE), 2016 11th International Conference on* (pp. 209-213). IEEE.

[29] Dutta, P., & Nachamai, M. (2016, January). Detection of faces from video files with different file formats. In *Microelectronics, Computing and Communications (MicroCom), 2016 International Conference on* (pp. 1-6). IEEE.

[30] Mohanty, A., Suda, N., Kim, M., Vrudhula, S., Seo, J. S., & Cao, Y. (2016, May). High-performance face detection with CPU-FPGA acceleration. In *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on* (pp. 117-120). IEEE.

[31] Abbas, S. A., & Vicithra, G. (2016, March). Actualization of face detection in FPGA using neural network. In *Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on* (pp. 634-638). IEEE.

[32] Lienhart, R., Kuranov, A., & Pisarevsky, V. (2003, September). Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *Joint Pattern Recognition Symposium* (pp. 297-304). Springer, Berlin, Heidelberg.

[33] Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, *290*(5500), 2323-2326.

[34] Mohamad, N., Zaini, F., Johari, A., Yassin, I., & Zabidi, A. (2010, May). Comparison between Levenberg-Marquardt and scaled conjugate gradient training algorithms for breast cancer diagnosis using MLP. In *Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium on*(pp. 1-7). IEEE.

[35] Zaman, F. H. K., Sulaiman, A. A., Yassin, I. M., Tahir, N. M., & Rizman, Z. I. (2017). Development of mobile face verification based on locally normalized Gabor wavelets. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(4), 1198-1205.

[36] Zabidi, A., Md Tahir, N., Mohd Yassin, I., & Rizman, Z. I. (2017). The performance of binary artificial bee colony (BABC) in structure selection of polynomial NARX and NARMAX models. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(2), 373-379.

[37] Mohd Yassin, I., Jailani, R., Ali, M., Amin, M. S., Baharom, R., Hassan, A., ... & Rizman, Z. I. (2017). Comparison between cascade forward and multi-layer perceptron neural networks for NARX functional electrical stimulation (FES)-based muscle model. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(1), 215-221.

[38] Yassin, I. M., Zabidi, A., Ali, M. S. A. M., Tahir, N. M., Abidin, H. Z., & Rizman, Z. I. (2016). Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm. *International Journal on Advanced Science, Engineering and Information Technology*, *6*(5), 630-637.