IOP Conf. Series: Materials Science and Engineering **341** (2018) 012026 doi:10.1088/1757-899X/341/1/012026

Locally Linear Embedding of Local Orthogonal Least Squares Images for Face Recognition

Fadhlan Hafizhelmi Kamaru Zaman

Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

fadhlan@salam.uitm.edu.my

Abstract. Dimensionality reduction is very important in face recognition since it ensures that high-dimensionality data can be mapped to lower dimensional space without losing salient and integral facial information. Locally Linear Embedding (LLE) has been previously used to serve this purpose, however, the process of acquiring LLE features requires high computation and resources. To overcome this limitation, we propose a locally-applied Local Orthogonal Least Squares (LOLS) model can be used as initial feature extraction before the application of LLE. By construction of least squares regression under orthogonal constraints we can preserve more discriminant information in the local subspace of facial features while reducing the overall features into a more compact form that we called LOLS images. LLE can then be applied on the LOLS images to maps its representation into a global coordinate system of much lower dimensionality. Several experiments carried out using publicly available face datasets such as AR, ORL, YaleB, and FERET under Single Sample Per Person (SSPP) constraint demonstrates that our proposed method can reduce the time required to compute LLE features while delivering better accuracy when compared to when either LLE or OLS alone is used. Comparison against several other feature extraction methods and more recent feature-learning method such as state-of-the-art Convolutional Neural Networks (CNN) also reveal the superiority of the proposed method under SSPP constraint.

1. Introduction

Since over two decades ago, face recognition has received unprecedented attentions from across numerous research fields such as pattern recognition, computer science, neuroscience, artificial intelligence and many more. The reasons for its popularity is two-fold – largely because of (1) the intricate, unsolved mystery behind how human recognize faces as well as (2) the wide range of practical applications that can typically benefit from face recognition. Previous researchers reported that in face recognition there are numbers of facial features that represent the spatial and configural information of the face which help human brain recognize faces [1-5]. Applications such as access control, security, image database investigations, and surveillance [6] can highly benefit from face recognition due to its non-intrusiveness and high accuracy under constrained condition [7]. However, reliability of face recognition in unconstrained environment is somehow inferior to other intrusive biometric features such as fingerprints and iris. This is because face recognition always suffers under unconstrained variations in face image, for instance change in lighting, pose, facial expressions and age. These are non-trivial problems and images collected from unconstrained environments such as images from CCTV and Webs have a very profound effect of these variations. Over the years, many previous works have focused on how to deal with such variations to increase the reliability of face recognition in unconstrained environments.

Facial appearances can commonly be represented in two ways, either by component-based (local) or holistic (global) representations which have their own advantages and weaknesses. Component-based method is more suitable to be used when the face variations are local, such as variations in expressions, while holistic method is more suited to when the variations are global, such as variations in pose. By holistic representation, a vector of pixel intensities is used to represent the whole image while component-based representation uses multiple group of vectors of pixel intensities to represent various components or regions on face. These components can be evenly divided, or segregated per several common pre-defined regions, and it is either unweighted or linearly/non-linearly weighted at classifier level, for better classification results especially when used with ensembles of classifiers. In both types of representations, dimensionality reduction, or sometimes referred to subspace projection, are usually applied to extract the facial feature from the pixel intensities and to reduce the curse of dimensionality effect. The dimensionality reduction can be linear, non-linear or manifold-learning methods, and can be applied holistically for simpler implementation but with higher processing cost, or locally for better computation performance but slightly complex implementation [8].

Some of popular linear dimensionality reduction methods that has been used in extracting facial features are Eigenfaces [9], Fisherfaces [10], and Local Binary Patterns (LBP) [11]. Some current variants of these methods are discussed in [12, 13]. Non-linear methods such as Kernel Principal Component Analysis (KPCA) [14] which extends PCA to nonlinearity by projecting the data into a higher-dimensional feature space via the kernel trick and Independent Component Analysis (ICA) have also been used in face recognition. On the other hand, manifold-based techniques, such as local linear embedding (LLE) [15], ISOMAP [16] and curvilinear component analysis (CCA) [17], and their linear variants (e.g., locality preserving projection (LPP) [18] and preserving projection (OLPP) [19]) has been used extensively. Investigation on the performance of several linear, non-linear and manifold methods in face recognition have been reported extensively in [20].

Based on findings reported in [20] the applications of these dimensionality projection methods particularly LLE has improved recognition performance, and LLE has been successfully used previously in face recognition [20-22]. However as pointed in [8], the computation cost has increased with increasing complexity of each projection method. We propose in this paper to use Orthogonal Least Square (OLS) [23] to overcome this problem. OLS is known as orthogonal constraint which can preserve more discriminant information in the subspace and it is very popular in other fields and applications especially in system identification [24]. Recently, a unified two-stage orthogonal least squares methods instead of the fast recursive-based methods has been proposed for neural network construction [25]. OLS also has been shown to be able to produce considerable improvement in neural networks training compared to the randomly initialized networks [26]. Similarly, decomposition of the recursive OLS procedure into Sequential Partial Orthogonalization (SPO) has been used for extreme machine learning in [27]. Other recent use of OLS includes to approximate sparse solution [28] and recovery of sparse signals [29]. The amount of work focusing on the use of OLS in feature extraction specifically used for face recognition is limited. Only just recently, Zhao et al. proposed an orthogonal least squares regression model which is used for the extraction of facial features in face recognition [30]. However, in this paper only OLS is used for feature extraction, without fusion with any other subspace method.

In this paper, we specifically focus on increasing computation performance of LLE while maintaining its face classification accuracy. We use locally-applied OLS as pre-processing step prior the application of LLE to reduce the number of features and in doing so we propose a new compact projected face images called Local Orthogonal Least Square (LOLS) images. The feature extraction approach used in this paper is component-based approach since most of the face datasets used for experiments encompasses local variations and component-based method delivers better performance under this constraint. Another constraint implemented in this paper is Single Sample Per Person

(SSPP) [31, 32] constraint – this is where only single image is used for training (gallery), imitating real-world condition where in most applications only single image is available as reference (e.g. identification card, passport, and license). This paper is arranged as follows: Section 2 will describe our proposed method in detail. Section 3 presents experimental results which is followed by discussions in Section 4, and we conclude the paper in Section 5.

2. Methodology

In this section, we discuss the process of computing LOLS images by locally applying OLS on face images. Then we use component-based application of LLE on LOLS images to acquire LLE-LOLS features for face classification. Finally, we classify the face using ensembles of classifiers based on ensembles of Cosine Similarity Metric Learning [33].

2.1. Local Orthogonal Least Squares (LOLS) Images

Orthogonal least square algorithm (OLS) is an algorithm which implements the forward selection method for subset model selection and capable of estimating the parameter estimators. The OLS algorithm transforms the set of regressors P_i into orthogonal basis vectors [23]. Here the OLS is used as discriminant analysis method, such that it improves the model sparsity and control the model complexity [30]. As described in [23], the orthogonal polynomials can be calculated by applying Classical Gram-Schmidt (CGS) (this method is susceptible to round-off error) and Modified Gram-Schmidt (MGS) procedures. The MGS procedure calculates A one row at a time and orthogonalizes P as follows: at the kth stage make the columns subscripted $k+1,\dots,M$ orthogonal to the kth column and repeat the operations for $k=1,\dots,M-1$. Specifically, denoting $p_i^{(0)}=p_i, i=1,\dots,M$, then

$$\left.\begin{array}{c}\left.\begin{array}{c} w_k = p_k^{(k-1)} \\ \alpha_{ki} = \dfrac{\langle w_k, p_i^{(k-1)}\rangle}{w_k, w_k}, i = k+1,\dots,M \\ p_i^{(k)} = p_i^{(k-1)} - \alpha_{ki} w_k, i = k+1,\dots,M \end{array}\right\} k = 1,2 \dots, M-1 \\ w_M = p_M^{(M-1)} \end{array}\right\} \tag{1}$$

The elements of g are computing by transforming $z^{(0)}=z$ in a similar way

$$\left.\begin{array}{c} g_k = \dfrac{\langle w_k, z^{(k-1)}\rangle}{w_k, w_k} \\ z^{(k)} = z^{(k-1)} - g_k w_k \end{array}\right\} k = 1,2,\dots,M \tag{2}$$

We locally apply OLS on each of facial component that is extracted using a sliding window of size *W* that traverse across the surface of the facial image by certain amount of stride *S*. These facial components are projected into smaller component called LOLS component. LOLS component has smaller dimension than the original facial component which would result into much smaller final LOLS images – a combination of all LOLS component that would yield a reduced representation of original image. This process is illustrated in Figure 1.

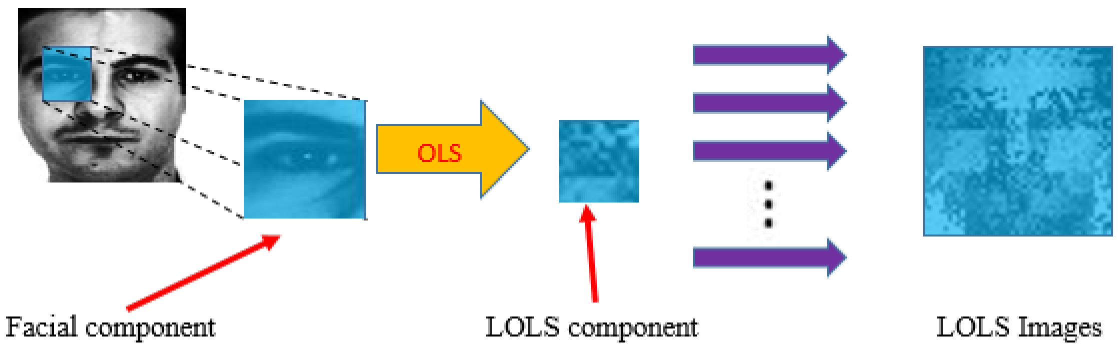


Figure 1. The process of extraction of OLS features from each facial component. The combination of all LOLS components will result into a complete facial image with reduced dimension.

Based on Figure 1, the resulting LOLS image still resembles the original face, however with some details missing. The missing details are those feature that has been eliminated by the OLS, which may not contribute much to the discrimination of the facial identity. This is actually similar to the implementation of Linear Discrimination Analysis (LDA) [34] on face images. The LOLS images acquired in this step is further reduced using LLE, as described in the following section.

2.2. Locally Linear Embedding of LOLS Images(LLE-LOLS)

LLE [15] is a nonlinear manifold method which maps nonlinear high dimensional multivariate data onto lower dimensional subspace having single global coordinate scheme without local minima optimization. The dimension of LOLS facial components acquired from previous steps are further reduced by projection into LLE subspace as illustrated in Figure 2. First we define the number of LLE neighborhood K_{LLE} for each data point x_i , and then by minimizing the cost function $\phi(w)$ as constraint least-square problem, the weight of w_{ij} of x_i from its neighbor x_j is computed by (3). After that, by minimizing the cost function $\phi(y)$, we calculate the projected vector y_j, which is solved as eigenvalue problem in (4).

$$\phi(w) = \min_w \sum_i \left\| x_i - \sum_j w_{ij} x_j \right\|^2 \tag{3}$$

$$\phi(y) = \min_y \sum_i \left\| y_i - \sum_j w_{ij} y_j \right\|^2 \tag{4}$$

where in the case of data x_j is not the neighborhood of x_i, $\sum_j w_{ij} = 1$ and $w_{ij} = 0$. The reduction of LOLS components into LLE vectors take into consideration the lateral LOLS components that belong in the training data. This is similar to the Locally Lateral Subspace (LLS) strategy adopted in [21] which delivers better performance with considerably lower memory requirements. The final features acquired from this step is LLE-LOLS vectors, which is practically has significantly lower dimension than the original image (around more than 80%-dimensional reduction in feature, as pointed out in later sections) while maintaining the salient facial features of the original image.

2.3. Face Classification using Ensembles of Cosine Similarity Metric Learning

We use an ensemble of classifier strategy similar to the one adopted in [21] to classify the acquired LLE-LOLS features. This includes finding the contribution of each LLE-LOLS components on final classification termed as component's confidence, *c*. This confidence is calculated based on the distance vector d_m between component *m* and all lateral components from *P* gallery (training) images. Given that h is the minimum values in d_m, such that $h = \arg\min \{d_m\}$, vector of component's confidence c_m can be computed from modification of softmax normalization in (5).

$$c_m = \log\ (h+2)\,/\log\ (d_m+2) \qquad (5)$$

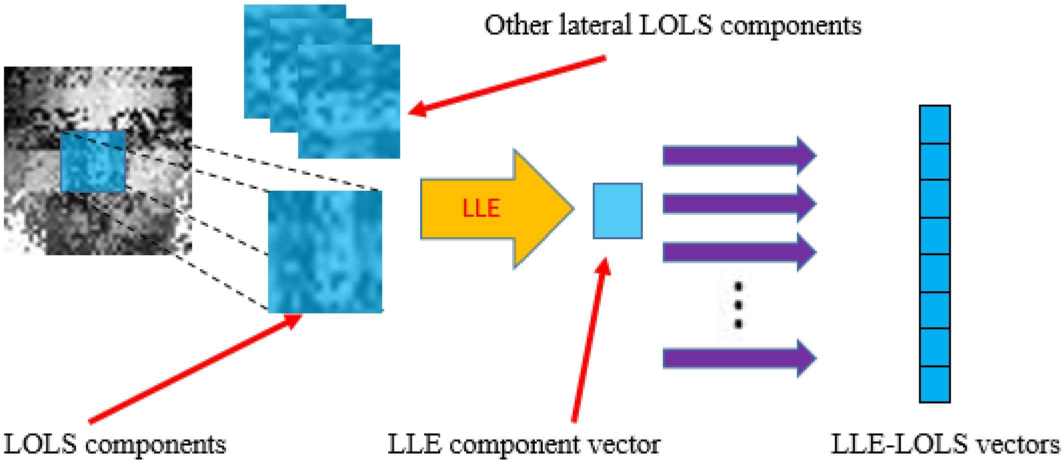


Figure 2. The process of extraction of LLE features from each LOLS component. The combination of all LLE components will result into LLE-LOLS features for face classification. This strategy used in this paper follows LLS strategy proposed in [21].

As shown in (5), maximum confidence is assigned to the nearest feature and others are assigned exponentially lesser confidence per their distance. Then, the component confidence vector, c is calculated using (5). The process is recurring for all available component features and the class label is assigned based on the summation of component confidence $C = \sum_{m=1}^{M} c_m$. For example, if linearly weighted summation of confidence is used, the final classification label of image i would be given to class with $\arg max\{C\}$. Given two vectors of attributes A and B where A_i and B_i are the components of vector A and B respectively, the distance metrics d used in this paper can be calculated using equations (6) which is based on the Cosine Similarity Metric Learning [33].

$$d\ = 1 - \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\ \sqrt{\sum_{i=1}^{j} B_i^2}} \qquad (6)$$

3. Results and Discussions

In this section, we present the results and discussions on our experiments on the performance of LLE-LOLS in face classification. For benchmarking purpose, we use publicly available datasets such as ORL, AR, YALE B, and FERET face datasets. Details on datasets used in this paper is summarized in Table 1.

Table 1. Datasets distribution of gallery and probe images for experiments conducted in this paper

Dataset	SSPP Gallery	Probes
AR	100 images of 100 persons	25 images per person
YALE B	38 images with $0^{\circ} \leq (\varphi, \theta) < 12°$	2376 images with $(\varphi, \theta) \geq 12°$
ORL	40 images of 40 persons	5 to 6 images per person
FERET	1196 images from *fa*	2111 images from *fb, fc, dup I* and *dup II*

The AR dataset [20] contains images taken in two sessions with 13 images per session where 14 images per subject depict expressions and illumination variation, 12 images per subject depict the subject as wearing sunglasses and scarf. We use 100 subjects of 50 males and 50 females for the experiment following procedures adopted in [21]. YALE B dataset contains images having different illumination conditions with light direction with respect to camera axis defined by azimuth angle φ and elevation θ. The cropped version is used which contains 2414 full-frontal images from 38 subjects [18]. ORL dataset [17] consists of 40 different subjects with 10 different images each. The images were taken at different times, varying the lighting, facial expressions and facial details in an upright, frontal position with some side movement. For the experiment on FERET dataset, five subsets used namely *fa* (gallery), *fb* (1,195 expression-variant images), *fc* (194 illumination-variant images), *dup I* (722 images taken later in time), and *dup II* (234 images, which is a subset of *dup I*), following the standard FERET evaluation protocol [19].

All images used in this paper are normalized to 63x63 pixels and they are histogram-equalized before LOLS feature extraction is applied on the images. The size of sliding window is set constant for all images that is W = 7x7 with stride size *S*=7. Samples of LOLS images obtained from our proposed method is illustrated in Figure 3. In this illustration, these images are reduced to 45x45 LOLS images after we apply the proposed LOLS algorithm. Other parameters used in experiments are as follows: Retained features for each facial component and neighbourhood for LLE are f_{LLE} = 8 and K_{LLE} = 100 respectively, while retained features for PCA and ICA are also 8. Parameters for Convolutional Neural Network (CNN) are *number of filter* = 1000, *filter size* = 7, *max pooling* = 2, *stride* = 2, *initial learning rate* = 0.01. The learning method used is stochastic gradient descent while the iterations are set to be maximum of 1000 iterations. All experiments are in accordance to SSPP constraint where only 1 image per subject is used for training. We analyse the results of experiments from two different aspects which is classification accuracy and computation time. These are discussed in detail below.

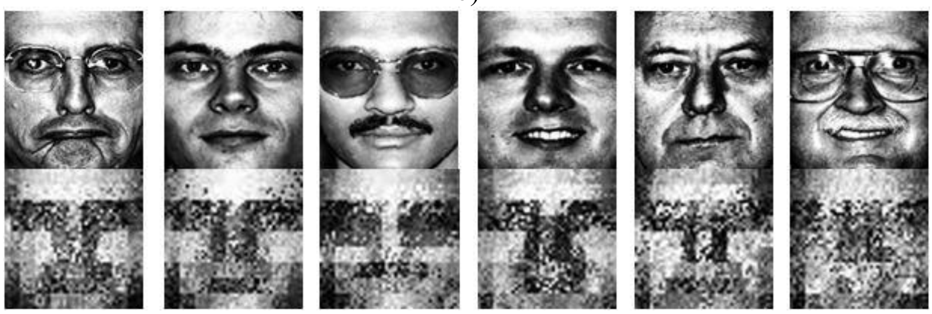
3.1. Face classification performance (classification accuracy).

Face classification accuracy results from the experiment are tabulated in Table 2, where the best classification accuracy of our proposed method is compared against several other popular face extraction methods. Baseline method denotes classification using only pixel intensities as features. Based on results shown, for AR dataset, LLE-LOLS produces the best result of 83.08% classification accuracy which is slightly better than LLE method, which is at 82.23%. This is also significantly higher than Baseline accuracy, which is at a mere 65.08% accuracy. The number of dimensions reduced are also significant when compared to Baseline – the reduction is as much as 83.67%, which means that 83.67% of the original features are 'removed' from the final LLE-LOLS features. Based on the result in Table 2, it is also shown that OLS managed to reduce 38.77% of dimension from original features before further processing by LLE. Based on the result, the best number of OLS retained feature for AR dataset is 26 for each facial component.

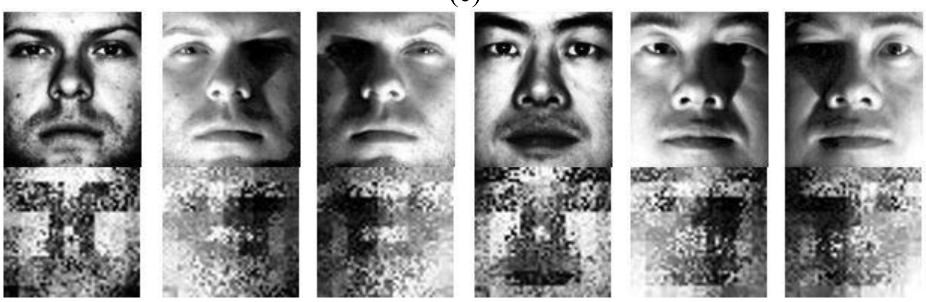
Furthermore, results on YaleB face classification demonstrates the superiority of LLE-LOLS method compared to other methods. It delivers 94.75% classification accuracy which the highest out of all tested methods. It produces even better results than the recent Orthogonal Least Squares Regression (OLSR) [30] method which delivers only 89.14% accuracy. Besides, the CNN method fails to deliver good result since the SSPP constraint affect CNN's ability to generalize from the limited sample. For both AR and YaleB datasets, CNN delivers 66.39% and 54.42% classification accuracy respectively. On two other datasets, the ORL and FERET datasets, the proposed LLE-LOLS method also delivers the best result – 94.03% and 56.47% classification accuracy respectively. LLE-OLS also produces better results when compared against OLSR method, which shows that LLE-OLS is more than 3% better than OLSR in terms of accuracy for ORL dataset. The result for ORL and FERET also demonstrates that CNN once again fall short of expectation where the results were severely affected by limitation of samples. In all tested datasets, we observed that OLS alone cannot produce good results on face recognition, while LLE alone can be good but LLE-OLS manages to improve LLE further. LLE-LOLS produced best results and shows improvement in classification accuracy when compared to results obtained from just using either OLS or LLE in feature extraction stage. The difference is as much as 3% (LLE-LOLS compared against LLE in ORL dataset) and 19% (LLE-LOLS compared against OLS in AR dataset) respectively. For YaleB, ORL and FERET datasets, the best number of OLS features used in LLE-LOLS method is 34, 26 and 20 features per facial component respectively.

(a)

b)



(c)



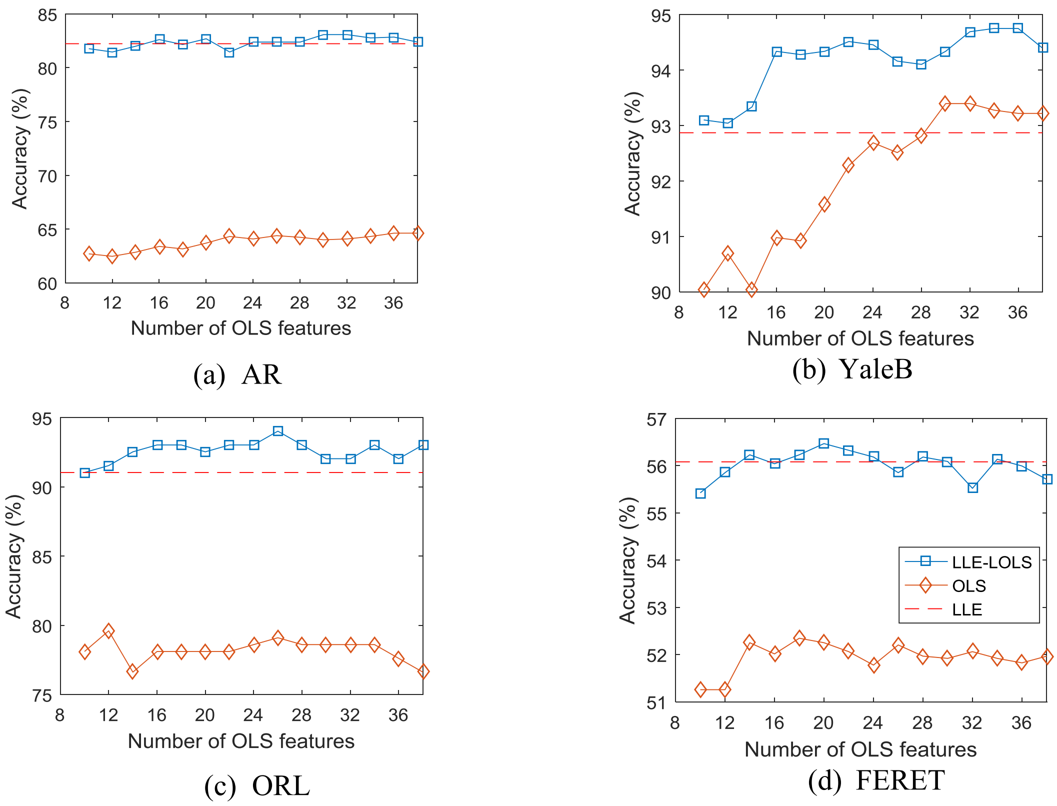
(d)

Figure 3: Sample of images from datasets used in this paper, namely (a) AR, (b) ORL, (c) FERET and (d) YaleB. These images are normalized to 63x63 and the subsequent images shown after each row are the equivalent 45x45 LOLS images.

Table 2. Face recognition results for several methods as compared to our proposed method. All experiments are conducted in adherence to SSPP constraint

Dataset	Method	Number of Retained Local Features	Reduced Dimension (%)	Classification Accuracy (%)
AR	Baseline	49	0	65.08
	PCA	8	83.67	76.31
	ICA	12	75.51	76.77
	CNN [35]	49	0	66.39
	OLS	36	26.53	64.62
	LLE	8	83.67	82.23
	Weighted LLE [21]	8	83.67	80.27
	LLE-LOLS	26 8	38.77 (OLS) 83.67 (LLE)	**83.08**
YaleB	Baseline	49	0	93.81
	PCA	8	83.67	79.30
	ICA	12	75.51	92.69
	CNN [35]	49	0	54.42
	OLS	30	38.77	93.39
	OLSR [36]	NA	NA	89.14
	LLE	8	83.67	92.87
	LLE-LOLS	34 8	30.61 (OLS) 83.67 (LLE)	**94.75**
ORL	Baseline	49	0	78.11
	PCA	8	83.67	91.54
	ICA	12	75.51	86.07
	CNN [35]	49	0	86.07
	OLS	12	75.51	79.61
	OLSR [36]	NA	NA	90.43
	LLE	8	83.67	91.04
	LLE-LOLS	26 8	46.94 (OLS) 83.67 (LLE)	**94.03**
FERET	Baseline	49	0	51.87
	PCA	8	83.67	51.25
	ICA	12	75.51	52.44
	CNN [35]	49	0	43.15
	OLS	18	63.26	52.34
	LLE	8	83.67	56.08
	LLE-LOLS	20 8	59.18 (OLS) 83.67 (LLE)	**56.47**

Figure 4 further illustrates the comparison between the classification accuracy of LLE-LOLS, LLE and LOLS. It shows the classification accuracy when different number OLS features are used during feature extraction. For AR dataset, the classification accuracy of LLE-LOLS is not too sensitive towards different number of OLS features used. The LLE-LOLS accuracy is significantly higher than OLS, but just slightly better than LLE. For YaleB dataset, LLE-LOLS is sensitive towards different number of OLS features used where large fluctuations can be observed when different number of features are used. Its classification accuracy is also much better than both LLE and OLS method. In both ORL and FERET datasets, the LLE-LOLS method is just slightly affected by number of OLS features, while the best accuracy for LLE-OLS is better than LLE and OLS accuracy.



(a) AR (b) YaleB (c) ORL (d) FERET

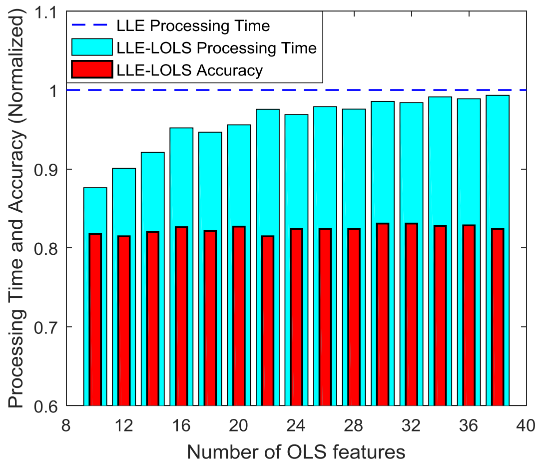
Figure 4. The performance of face classification accuracy (in percentage) for (a) AR, (b) YaleB, (c) ORL and (d) FERET datasets when different number of OLS feature are used in feature extraction stage. Shown together in this figure is the accuracy of LLE face classification.

3.2 Face classification performance (processing time).

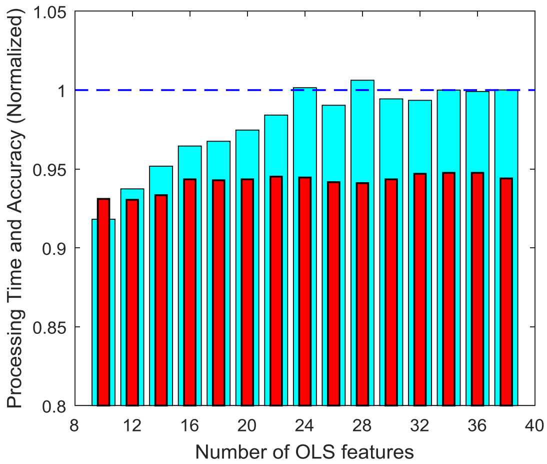
In terms of processing time, we measure the reduction in processing time achieved by LLE-LOLS method when compared against when only LLE is used for feature extraction. Figure 4 shows the results in detail. In this figure, we plot the processing time of LLE-LOLS which is normalized against the processing time taken by LLE. In the same figure, we plot the classification accuracy (normalized to 1.0 – where 1.0 is the highest accuracy) to further emphasize the relationship between the processing time and face classification accuracy. According to Figure 5, for AR and YaleB and FERET datasets, the processing time of LLE-LOLS are significantly less than the processing time of LLE by considerable amount. In fact, the reduction of processing time is as much as 10% when compared against LLE processing time. Even though smaller number of OLS features consumes lower processing time, the classification accuracy is not severely affected by the small number of features. This is highly desirable since if the classification accuracy is not sensitive to number of features used, we could opt for better processing time should such requirement arise. However, for ORL dataset, OLS features higher than 18 would increase the processing time of LLE-LOLS, making it to be higher than LLE processing time. In fact, higher number of features does not guarantee better performance, however it does increase the processing time required. It is thus observed that the proposed method

IOP Conf. Series: Materials Science and Engineering **341** (2018) 012026 doi:10.1088/1757-899X/341/1/012026

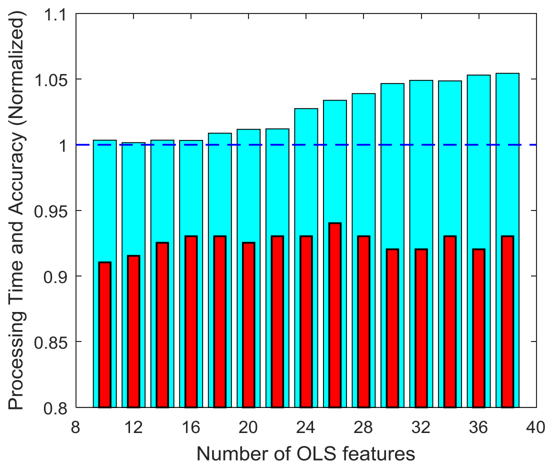
works best when large datasets are used (i.e. AR, YaleB and FERET) while smaller datasets such as ORL does not benefit much in terms of processing time.



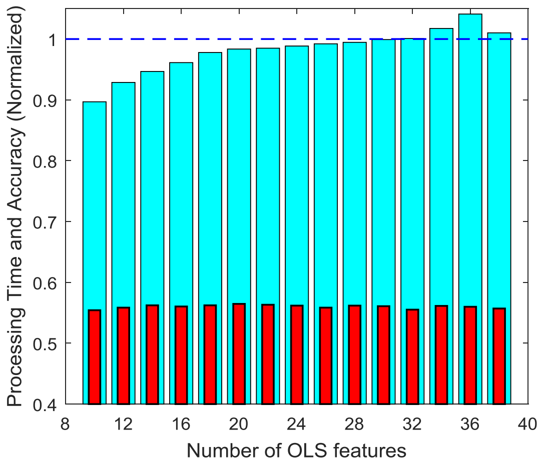
(a) AR



(b) YaleB



(c) ORL



(d) FERET

Figure 5: The normalized processing time of LLE-LOLS (normalized against LLE processing time) for (a) AR, (b) YaleB, (c) ORL and (d) FERET datasets when different number of OLS feature are used in feature extraction stage. Shown together in this figure is the LLE-LOLS face classification accuracy.

4. Conclusions

In this paper, we present a method to improve LLE processing time in face extraction stage while improving or at least maintaining the face classification accuracy. We propose a method called LLE-LOLS method where components of face images are dimensionally reduced using OLS which results into LOLS images. These LOLS images are further reduced in dimension using conventional LLE method. The LLE-LOLS features practically has significantly lower dimension than the original image (around more than 80%-dimensional reduction in feature) while maintaining the salient facial features of the original image, as proven by the classification accuracy. In all tested datasets, LLE-LOLS produced best results when tested against several popular face recognition methods, and shows improvement in classification accuracy when compared to results obtained from just using either OLS or LLE in feature extraction stage. The difference is as much as 3% (LLE-LOLS compared against

LLE in ORL dataset) and 19% (LLE-LOLS compared against OLS in AR dataset) respectively. Comparisons against state-of-the-art CNN also reveals the robustness of our proposed method against limitation in number of available samples for training. In terms of processing time, LLE-LOLS manages to improve the processing time in almost all tested datasets, where improvement of as much as 10% in processing time is achieved. To further improve this method, it is recommended to investigate the aptness of this method to be used with other popular feature extraction methods such as Gabor Wavelets or Scale Invariant Feature Transform (SIFT) features.

Acknowledgements
The authors would like to thank Dr. Aleix M. Martinez for providing the AR Face Dataset. Some portions of the research in this paper also use Extended Yale Face Database B. The authors would like to thank the FERET Technical Agent, the US National Institute of Standards and Technology (NIST) for providing the FERET dataset. The work presented here is sponsored by Universiti Teknologi MARA, Malaysia under grant 600-RMI/DANA 5/3/LESTARI (20/2015).

References
[1] A. W. Young, D. Hellawell, and D. C. Hay, "Configural information in face perception," *Perception,* vol. 16, pp. 747-759, 1987.
[2] A. W. Young, F. Newcombe, E. H. F. de Haan, M. Small, and D. C. Hay, "Face perception after brain injury: Selective impairments affecting identity and expression," *Brain,* vol. 116, pp. 941-959, 1993.
[3] J. W. Tanaka and J. A. Sengco, "Features and their configuration in face recognition," *Memory and Cognition,* vol. 25, no. 5, pp. 583-592, 1997.
[4] A. Schwaninger, C. Wallraven, D. W. Cunningham, and S. D. Chiller-Glaus, "Processing of facial identity and expression: a psychophysical, physiological, and computational perspective.," *Progress in Brain Research,* vol. 156, pp. 321-343, 2006.
[5] J. Sadr, I. Jarudi, and P. Sinha, "The role of eyebrows in face recognition,," *Perception,* vol. 32, pp. 285-293, 2003.
[6] D. N. Parmar and B. B. Mehta, "Face Recognition Methods & Applications," *Computer Technology & Applications,* vol. 4, no. 1, pp. 84-86, 2013.
[7] Y.-Z. Lu, "A Novel Face Recognition Algorithm for Distinguishing Faces with Various Angles," *International Journal of Automation and Computing,* vol. 5, no. 2, pp. 193-197, 2008.
[8] F. Kamaruzaman, A. A. Shafie, and Y. M. Mustafah, "Robust Face Recognition Against Expressions and Partial Occlusions," *International Journal of Automation and Computing,* 2015.
[9] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.
[10] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, 1997.
[11] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 28, no. 12, pp. 2037-2041, 2006.
[12] H. V. Nguyen, L. Bai, and L. Shen, "Local Gabor binary pattern whitened pca: a novel approach for face recognition from single image per person," in *Lecture Notes in Computer Science, Proceedings of the Third International Conference on Advances in Biometrics*, Alghero, Italy, 2009, vol. 5558, pp. 269-278.
[13] F. Hafiz, A. A. Shafie, and Y. M. Mustafah, "Face Recognition From Single Sample Per Person by Learning of Generic Discriminant Vectors," *Procedia Engineering,* vol. 41, pp. 465-472, 2012.

[14] H. Hoffmann, "Kernel PCA for novelty detection," *Pattern Recognition,* vol. 40, no. 3, pp. 863-874, 2007.

[15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science,* vol. 290, no. 5500, pp. 2323-2326, 2000.

[16] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science,* vol. 290, no. 5500, pp. 2319-2323, 2000.

[17] P. Demartines and J. Hérault, "Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets," *IEEE Trans. Neural Networks,* vol. 8, pp. 148-154, 1997.

[18] X. Niyogi, "Locality preserving projections," *Neural information processing systems,* vol. 16, p. 153, 2004.

[19] D. Cai, X. He, J. Han, and H. J. Zhang, "Orthogonal Laplacianfaces for Face Recognition," *IEEE Transactions on Image Processing,* vol. 15, no. 11, pp. 3608-3614, 2006.

[20] W. Huang and H. Yin, "On nonlinear dimensionality reduction for face recognition," *Image and Vision Computing,* vol. 30, pp. 355-366, 2012.

[21] F. K. Zaman, A. A. Shafie, and Y. M. Mustafah, "Robust face recognition against expressions and partial occlusions," *International Journal of Automation and Computing,* journal article vol. 13, no. 4, pp. 319-337, 2016.

[22] N. Mekuz, C. Bauckhage, and J. K. Tsotsos, "Face recognition with weighted locally linear embedding," in *Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*, 2005, pp. 290-296.

[23] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control,* vol. 50, pp. 1873-1896, 1989.

[24] L. Mirea, "Dynamic multivariate B-spline neural network design using orthogonal least squares algorithm for non-linear system identification," in *2014 18th International Conference on System Theory, Control and Computing (ICSTCC)*, 2014, pp. 720-725.

[25] L. Zhang, K. Li, E. W. Bai, and G. W. Irwin, "Two-Stage Orthogonal Least Squares Methods for Neural Network Construction," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 26, no. 8, pp. 1608-1621, 2015.

[26] M. Lehtokangas, J. Saarinen, K. Kaski, and P. Huuhtanen, "Initializing Weights of a Multilayer Perceptron Network by Using the Orthogonal Least Squares Algorithm," *Neural Computation,* vol. 7, no. 5, pp. 982-999, 1995.

[27] N. Wang, M. J. Er, and M. Han, "Parsimonious Extreme Learning Machine Using Recursive Orthogonal Least Squares," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 25, no. 10, pp. 1828-1841, 2014.

[28] A. Hashemi and H. Vikalo, "Sparse linear regression via generalized orthogonal least-squares," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016, pp. 1305-1309.

[29] J. Wang and P. Li, "Recovery of Sparse Signals Using Multiple Orthogonal Least Squares," *IEEE Transactions on Signal Processing,* vol. 65, no. 8, pp. 2049-2062, 2017.

[30] H. Zhao, Z. Wang, and F. Nie, "Orthogonal least squares regression for feature extraction," *Neurocomputing,* vol. 216, pp. 200-207, 12/5/ 2016.

[31] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognition,* vol. 39, no. 9, pp. 1725-1745, 2006.

[32] F. H. K. Zaman, "Single Sample Face Recognition Using a Network of Spiking Neurons," PhD Thesis, Kulliyyah of Engineering, International islamic University Malaysia, Kuala Lumpur, 2015.

[33] H. V. Nguyen and L. Bai, "Cosine Similarity Metric Learning for Face Verification," in *Computer Vision – ACCV 2010: 10th Asian Conference on Computer Vision, Queenstown,*

New Zealand, November 8-12, 2010, Revised Selected Papers, Part II, R. Kimmel, R. Klette, and A. Sugimoto, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 709-720.

[34] W. Zuo, K. Wang, D. Zhang, and H. Zhang, "Combination of two novel LDA-based methods for face recognition," *Neurocomputing,* vol. 70, no. 4–6, pp. 735-742, 2007.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems,* 2012.

[36] H. Zhao, Z. Wang, and F. Nie, "Orthogonal least squares regression for feature extraction," *Neurocomput.,* vol. 216, no. C, pp. 200-207, 2016.