Emotion Recognition from Scrambled Facial Images via Many Graph Embedding

Richard Jiang, Tony Ho, Ismahane Cheheb, Somaya Al-Maadeed, Ahmed Bouridane

***Abstract* — Facial expression verification has been extensively exploited due to its wide application in affective computing, robotic vision, man-machine interaction and medical diagnosis. With the recent development of Internet-of- Things (IoT), there arouses the needs of mobile-targeted facial expression verification, where face scrambling has been proposed for privacy protection during image/video distribution over public network. Consequently, facial expression verification needs to be carried out in a scrambled domain, bringing out new challenges in facial expression recognition. An immediate impact from face scrambling is that conventional semantic facial components become not identifiable, and 3D face models cannot be clearly fitted to a scrambled image. Hence, the classical facial action coding system cannot be applied to facial expression recognition in scrambled domain. To handle with chaotic signals from face scrambling, in this paper, we propose an new approach – Many Graph Embedding (MGE) to discover discriminative patterns from the subspaces of chaotic patterns, where the facial expression recognition is carried out as a fuzzy combination from many graph embedding. In our experiments, the proposed MGE was tested on three scrambled facial expression datasets: JAFFE, MUG and CK++. In our experiment, we evaluated our algorithm via 2-fold, 4-fold and 10-fold cross validation. The benchmark results demonstrated that the proposed method can apparently improve the recognition accuracy, making our method a promising candidate for the scrambled facial expression recognition in the emerging privacy-protected IoT applications.**

***Index Terms* — Facial expression, emotion recognition, user privacy, many graph embedding.**

1. INTRODUCTION

THE past decade has witnessed many new developments in facial expression analysis due to its wide application in robotic vision [1, 2], forensics[3], affective computing[4], man-machine interaction[4, 5], and even medical diagnosis[2]. Especially, recent advances in the paradigm of Internet-of-Things (IoT) has made it available to take face photos anywhere from mobile devices, bringing out a wide needs of effective facial expression analysis over IoT devices. For example, a patient at home can send their facial images to their GP's mobile device for an automated diagnosis of their facial muscles [2] from their facial expression analysis; a passenger back to home can send a photo of their facial expression images from their mobile to their intelligent home to interact with a smart home-care system [5]. With these potential applications, facial analysis over mobile devices has become an emerging need of this function.

In the context of mass internet technology, privacy [6-12] has become a widely concerned issue in web-based video streaming. As one practical solution to this privacy issue, face scrambling [6-12] is emerging as a practical technique to protect the privacy

The manuscript was received _______, revised _______.

Richard Jiang and Ahmed Bouridane are with Computer and Information Science, Northumbria University, Newcastle upon Tyne, United Kingdom.

S. Al-Maadeed is with the Department of Computer Science and Engineering, Qatar University, Doha 110353, Qatar.

M. Emre Celebi is with Department of Computer Science, Louisiana State University in Shreveport, LA, USA.

Correspondence e-mail: richard.jiang@unn.ac.uk.

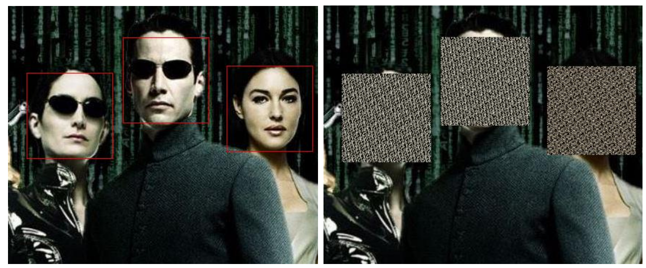


Fig.1. Facial expression becomes inscrutable after scrambled. While it does protect the privacy of the subjects/users, it brings out new challenges in facial expression recognition.

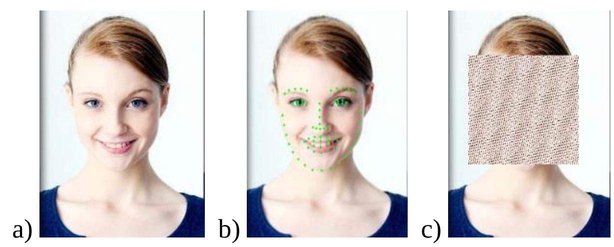


Fig.2. Semantic facial components, as shown in b), can be easily detected and utilized from original facial images. But these approaches may not be feasible in the scrambled domain.

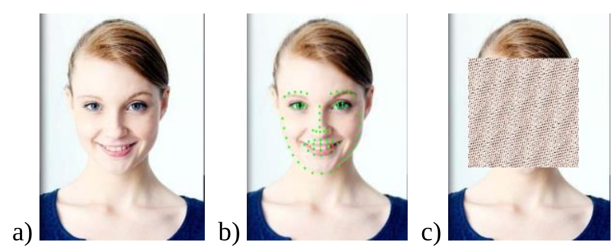


Fig.3. Who is smiling? (Answers at the end of the paper).

legally during video distribution over public internet. As shown in Fig.1, by scrambling faces detected in videos, the privacy of subjects can be respected in modern security technology.

Facial expression recognition has been researched extensively in the last decade. It can be carried out over multimodal features [15], multi-view images [16], or multiple frames [17, 18]. Usually, facial expression recognition is associated with facial action coding system (FACS) [18, 19], and landmarks tracking [20] can be used to assist the face expression recognition. In these approaches, sematic features [21-23] were exploited as effective cues for facial action recognition, while facial image is considered as being represented by semantic components such as eyes and mouth [21, 22] that can form the basic FACS units.

As shown in Fig.2-b), it is not difficult to detect these semantic components automatically in a face. However, after scrambled, as shown in Fig.2-c), a scrambled face has a very different appearance from its original facial image and it becomes extremely hard to match it with a 3D or semantic facial model. In the scrambled domain, semantic facial components simply become chaotic patterns. In this context, it becomes unavailable to exploit landmarks or 3D models for a better accuracy. In this case, as it has been discussed in [24, 25], an easy and straightforward way is to use the traditional data-driven approaches, where chaotic signals are treated simply as a set of data points spreading over manifolds.

Various data-driven face recognition algorithms have been well developed in the past several decades for image-based face expression recognition. In early days, linear dimensionality reduction [29-30] was introduced to this challenge, such as principal component analysis (PCA) [29], independent component analysis (ICA) [29], and Fisher's linear discriminant analysis (FLD) [30]. With kernel methods (KM) [30], these methods can be extended to a reproducing kernel Hilbert space with a non-linear mapping, and extended as k-PCA and k-FLD. Later, nonlinear manifold learning [31~34] have brought out a number of new methods for face recognition, such as Laplacianface [33] and Tensor subspace [34]. These approaches have been successfully applied to data-driven face recognition. However, for face recognition in scrambled domain, we need a robust approach to handle with chaotic signals in the scrambled domain, which are random and beyond the human perception.

Facial expression recognition has mostly been relevant to the challenge of dimensionality reduction [26-27]. Recently, multi-view based manifold learning [35-39] has been proposed to handle with the complexity of data structure, where it is believed multiple-view discriminative structures need to be discovered while a manifold may have different geometric shapes in different subspaces. With the hope to utilize this approach for chaotic signals, in this paper, we propose a new approach called *Many Graph Embedding* (MGE) to handle with this new challenge of chaotic signal recognition in scrambled domain.

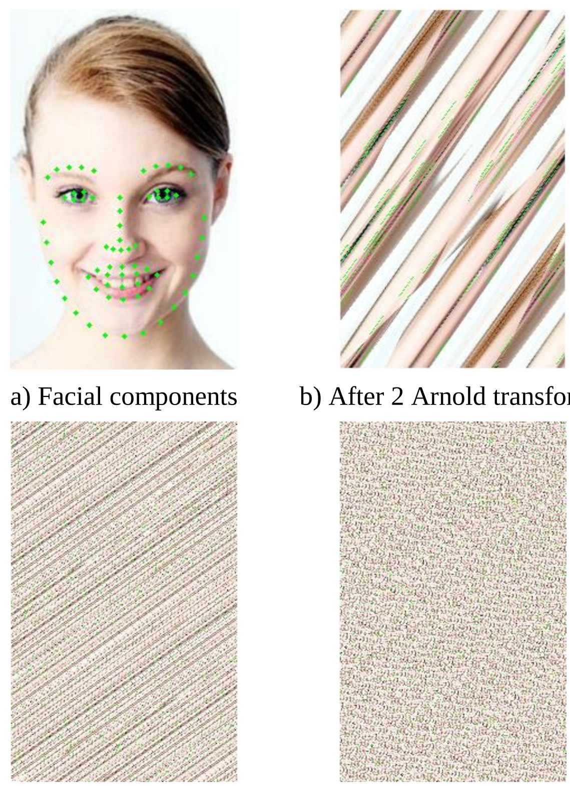
In the following sections, a preliminary about facial image scrambling method is introduced in section 2, and challenges from chaotic pattern analysis are discussed in section 3. In section 4, we present the proposed *Many Graph Embedding* (MGE) method. In section 5, we exploited MGE for scrambled facial expression classification. In section 6 gives the experimental results on three facial expression datasets, and conclusions are drawn in section 7.

2. CHAOTIC PATTERNS FROM FACE SCRAMBLING

There are several ways to perform facial image scrambling, such as simple masking or cartooning [8]. However, this kind of scrambling will simply lose the facial details, and hence facial expression recognition becomes unachievable in this case. This is not welcome by many applications that require recognizing a facial action in the scrambled domain.

Arnold transform [13, 14] is a kind of recoverable scrambling method. Scrambled faces can be unscrambled by several manual tries. Arnold scrambling algorithm has the feature of both simplicity and periodicity. In this work, we have chosen Arnold transform based scrambling as our specific test platform.

Arnold transform is proposed by V. I. Arnold in the research of ergodic theory. It is also called cat-mapping before it is applied to digital images. It has been widely used in visual systems where it is favored as a simple and efficient scrambling method. In the Arnold transform, a pixel at the point (x, y) is shifted to another point (x', y′) as:



a) Facial components   b) After 2 Arnold transforms

c) After 5 Arnold transforms   d) After 7 Arnold transforms

Fig.3. Face scrambling by Arnold transform. While we can easily identify semantic components and group pixels together to represent these semantic regions, it becomes extremely hard to find which dimensions (pixels) can be grouped together to form semantic subspaces (regions) after scrambled.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \bmod N, \qquad (1)$$

which is called 2D Arnold scrambling, which is called two-dimensional Arnold scrambling. Here, x and y are the coordinates of the pixel; N is the height or width of the square image processed; x' and y' are the coordinate of the scrambled image. Considering the feedback, iterative process of Arnold transform can be applied as the following:

$$P_{xy}^{k+1} = AP_{xy}^{k}, \quad P_{xy}^{k} = (x, y)^{T} \qquad (2)$$

Here, a pixel $(x, y)^{\mathrm{T}}$ after the *k-th* Arnold transform is the input, $P_{xy}{}^{k+1}$ in the left is the output for the k+1th Arnold transform. k represents the time of iterations, where k = 0, 1, 2 and so on.

By the replacement of the discrete lattice for transplantation, Arnold transform produces a new image after all pixels of the original image have been traversed. In addition to simple, easy to come true, Arnold scrambling also has the character of being cyclic and reversible.

Fig.3-a) shows a face with its facial components (i.e., eyes, nose and mouth) detected automatically by 3D model fitting. Fig.3-b) shows the scrambled face after two operations of Arnold transform, where it can be seen that facial components have drastic displacements. Fig.3-c) and d) shows the scrambled faces after five and seven operations of Arnold transform. In comparison with Fig.3-b), the scrambled faces in Fig.3-c) and d) are more difficult to identify by human eyes. In this work, we use seven operations of Arnold transform to scramble all faces.

In many IoT based applications, it may not be allowed to unscramble detected faces due to privacy-protection policy. Moreover, unscrambling may involve parameters (such as the initial shift coordinates) that are usually unknown by the online software. Facial expression recognition in the scrambled domain then becomes a necessity in these IoT applications.

As we can see from Fig.3, before scrambling, facial components can be easily identified by human eyes. After scrambling, the images become chaotic signals, and it is hard to figure out eyes and noses. In the concept of manifold learning, each pixel is a feature dimension, and pixels around eyes could form a subspace that denotes a semantic component for discriminative purpose. In the scrambled domain, we do not know which pixels or dimensions can be grouped together to form a semantic-meaning subspace for facial expression recognition. Hence, we need a more clever way in dimensionality analysis for facial expression recognition in the scrambled domain.

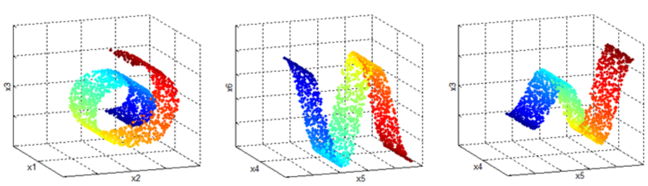
3. Many Manifold Problem in Scrambled Faces

In many real world applications such as face recognition and image classification, the data is often of very high dimensionality. Procedures that are computationally or analytically manageable in low-dimensional spaces can become completely impractical in a space of several thousands dimensions. This has been well known in machine learning as a notorious issue --- "*Curse of Dimensionality*" [28]. To tackle with this challenge, various techniques [29-39] have been developed for reducing the dimensionality of the feature space, in the hope of obtaining a manageable problem. Especially for face classification, dimensionality reduction (DR) has become an important step. Among various DR algorithms, we have linear subspace methods such as PCA, ICA and FLD, and nonlinear manifold learning methods such as LLE, Laplacian Eigenmap , non-negative matrix, tensor subspace analysis [34], and local Fisher discriminant analysis (LFDA) [24]. These approaches usually assume there is an underlying discriminative structure to discover, which leads to the paradigm of manifold learning.

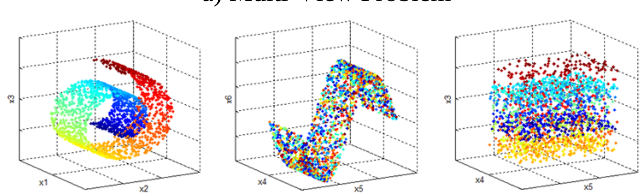
Recently, multi-view problem has been investigated by the research community, where it is advocated that the same manifold can have different shapes in different subspaces, as shown in Fig.4-a). Foster *et al* have employed canonical correlation analysis (CCA) approach [35] to derive the low dimensional embedding of two-view data and compute the regression function based on the embedding. Hedge *et al* [36] propose a multiple projection approach from the same manifold. Hou *et al* [37] used the pairwise constraints to derive embedding in mulitple views with linear transformation. Xia *et al* [38] combined spectral embedding with multi-view issue. Han *et al* [39] proposed a sparse unsupervised dimensionality reduction to obtain a sparse representation for multi-view data.

In the multi-view problem, as shown in Fig.4-a), though a manifold has different shapes in different subspaces, these shapes can always be unified as the same manifold in a higher-dimensional subspace. However, this could not be always true. As shown in Fig.4-b), while the sequence of data points in the second subspace is shuffled, the combination of two submanifolds simply creates a noisy-like distribution, as shown in the right side figure of Fig.4-b). This means two submanifolds cannot be merged. In this case we have to treat it as a multiple or even many manifold problem, where more than one sub-manifold structures need to be discovered from numerous possible subspaces.

In our facial recognition in scrambled domain, facial images



a) Multi-View Problem



b) Multiple Subspace Problem

Fig.4. Multi-view dataset and multi-manifold dataset. When the sequence of data points in the second subspace is shuffled, two sub-manifolds become independent to each other, and cannot be unified in a higher dimensional subspace, which is not fully addressed in conventional multi-view methods.

become chaotic signals, as shown in Fig.3. In this real-world issue, its underlying discriminative structures could be more like the case in Fig.4-b), where multiple or even many manifold structures need to be discovered. When this "many manifold" issue is associated with scrambled faces, an immediate challenge is how to find out which pixels/dimensions can be grouped together to form a meaningful discriminative subspaces. In original faces, we can easily do so by identifying semantic regions and take advantage of those semantic approaches. In chaotic scrambled images, there is no way to do so. This makes chaotic pattern classification a new challenge that different from conventional visual pattern analysis.

4. MANY GRAPH EMBEDDING FOR CHAOTIC PATTERNS

4.1 Problems to address

As shown in Fig.4-b), we have a new assumption that there are multiple graph (manifold) structures hiding in a given dataset, especially for a dataset of chaotic patterns. These graphs have not only different geometric shapes but also different edge-node connections. As a result, they cannot be simply combined in a higher dimensional space and treated as a multi-view issue in lower dimensional subspaces.

For example, in scrambled face analysis, each pixel is considered an independent dimension and we can have $W \times H$ dimensions for each scrambled image. We will then have many possible combinations of these dimensions to form a subspace, which can be approximated as C_N^K. Here, N is the total number of pixels, and K is the dimensionality of the chosen subspaces. It then becomes a typical *NP*-hard problem to find out all discriminative subspaces.

Given a dataset $X=\{x_1, x_2, \dots x_i \dots x_N\} \in R^D$, we may assume there are K discriminative graph structures $G^k \sim \{G^1, G^2, \dots, G^K\}$ underlying the given dataset X. While each graph corresponds to a subspace $\in R^l$ constructed from a subset of dimensions, the total possible k could include all numerated combinations $\sum_l C_D^l$, which is an extremly large number. To learn an embedding from each subspace, we will have,

$$\{Y^k\} \sim \arg\min_{\phi^k} \sum_k b_k \left(\phi^{k^T} X \left(D^k - Q^k\right) X^T \phi^k\right) \qquad (3)$$

Here, b_k can be a binary value to denote if there is a clear embedded structure in a given subspace.

As illustrated in Fig.4-b), an embedding in a subspace may be fully independent to another embedding in a different subspace, especially when their dimensional sets are not overlapped with each other. Based on this approximation, we can apply a "divide and conquer" strategy and estimate an embedding from each subspace independently, and Eq.(3) can then be reformulated as,

$$y^k \sim \arg\min_{\phi^k} \phi^{k^T} x^k \left(D^k - Q^k\right) x^{k^T} \phi^k \qquad (4)$$

Here, x^k has a decimated dimensionality of a subspace R^l. As a result, the projections $\tilde{Y} = \{y^k, b_k\}$ will be a matrix of $K \times d$, while $d<l$. The above formulation can be seen as a simple extension or repetitive application of the single graph embedding method.

After such simplification, though it seems we can repetitively carry out a process of graph embedding in each subspace simply one subspace by one subspace, it is however not practical to search through all possible subspaces. The number of such subspaces could be very large, making it computationally unavailable to carry out such an exhaustive search.

Besides, it is also difficult to clearly assert whether or not a subspace contains an obvious embedding for discriminative purpose. As a consequence, it is an invincible challenge that is hard to be handled directly for the "many manifold" challenge. In this work, we proposed a random approach to generate a random selection of subspaces and provide a probability solution to the search of subspaces.

4.2 Random Graph Generation

When a search problem in a large search space is concerned, in computer science, a widely applied heuristic search strategy is called *Monte Carlo*, where random threads are generated to search in a vast space with the hope to converge on the best or second best decisions. This strategy has been well exploited in many areas such as particle filter for object tracking. Similarly, when many graphs are concerned in our "*Many Manifold*" challenge, we randomly select a number of subspaces to testify whether or not a discriminative graph embedding can be found in a specific subspace.

Typically the random selection of a subspace can be described by the following procedure:

$$x \to z^n(x) : z^n(x) = w(n)^T x \qquad (5)$$

Here, $w^{\{n\}}$ is a randomized diagonal binary matrix with constraints,

$$\sum_k w_{kk}^n = r_0 \text{, with } w_{pq}^n = 0 \text{ when } p \neq q \qquad (6)$$

In practical, dimensions with $w_{kk}^n=0$ can be simply removed from further procedure, and hence $z \in R^m$ and $m<d$. With this constraint of randomized dimension selection, we have the selection ratio as (r_0/d), and the random selection is repeated $N_r \sim (d/r_0)$ times, where N_r stands for the plurality of the graph structures.

In this paper, we aim to use this randomness procedure to produce a set of random graphs to find the embedding. For a randomly selected feature z_i, we simply apply *k-NN* algorithm and obtain local graph G^n from a random selection w,

$$w^n \to G^n : \{z_i^n\}, \{s_{ij}^n\} \qquad (7)$$

Where $z^n \in R^m$, and $\{s_{ij}\} \in B^{N \times N}$ forms a $N \times N$ adjacent matrix. As a consequence, the above procedure can be repeated N_r times and obtain N_r sets of local structure based graphs.

4.3 Embedding of Many Graphs

Given we have obtained multiple random graphs $G^n \sim \{z^n, S^n\}$, the embeddings for each graph can be estimated in the same way as described in section 3,

$$y^n = \phi^{n^T} z^n \qquad (11)$$

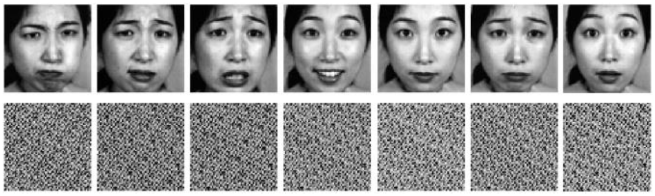
Similar to Eq.(2), we have,

$$y^k \sim \arg\min_{\phi^k} \phi^{k^T} z^k \left(D^k - Q^k\right) z^{k^T} \phi^k \qquad (12)$$

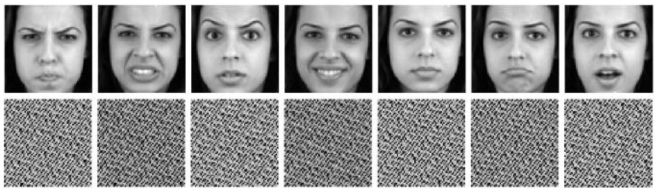
For all random graphs, we will find their embedding,

$$Y = \Phi^T Z \qquad (13)$$

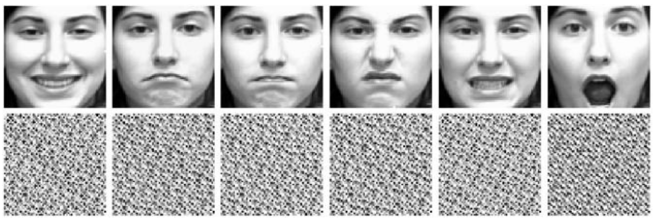
Where $Z \sim \{z^n\} \in B^{l \times N}$, $\Phi \sim \{\phi^n\} \in B^{m \times l \times Nr}$, $Y \sim \{y^n\} \in B^{m \times N \times Nr}$. Here, Y is the final projection result of the input Z from the



a) Samples in JAFFE database and their scrambled images



b) Samples in MUG database and their scrambled images



c) CK+ dataset and their scrambled images

Fig.5. Facial images in JAFFE, MUG and CK++ datasets.

proposed *Many Graph Embedding*. It is worth to mention that the final output Y is actually a matrix of $l \times N$, where N stands for the number of graphs.

4.4 Facial Expression Classification

It is noted that in Eq.(13), each data point will have a matrix of Y_{ij} calculated from MGE projection. From probabilistic view, on each manifold (or subspace) ϕ^n, given a query x_q, we can find its likelihood (namely distance) to a known data point or a class landmark x_c as $P(x_c|x_q,\phi^n)$ via the embedding ϕ^n by measuring the distance,

$$P(x_c \mid x_q, \phi^n) \propto \left\| Y_{i,n}^{q} - Y_{i,n}^{c} \right\|, \quad (14)$$

Then the final likelihood over all embeddings can be estimated as,

$$\widetilde{P}(x_c \mid x_q) = \frac{\sum_n P(x_c \mid x_q, \phi^n)}{\sum_c \sum P(x_c \mid x_q, \phi^n)}, \quad (15)$$

Then the decision rule is to assign x_q to class c for which the likelihood is the maximum:

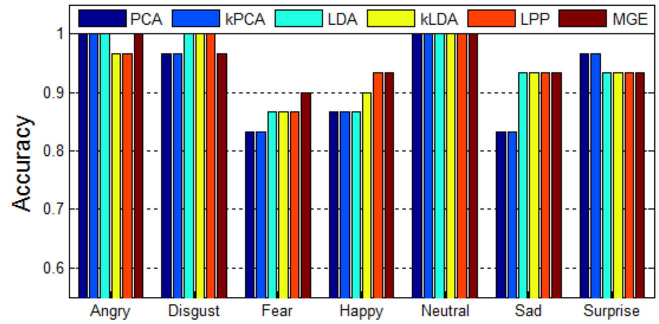
$$c \sim \arg\max_{x_c} \frac{\sum_n P(x_c \mid x_q, \phi^n)}{\sum_c \sum_n P(x_c \mid x_q, \phi^n)} \quad (16)$$

Here, c stands for the class label assigned to the query x_q.

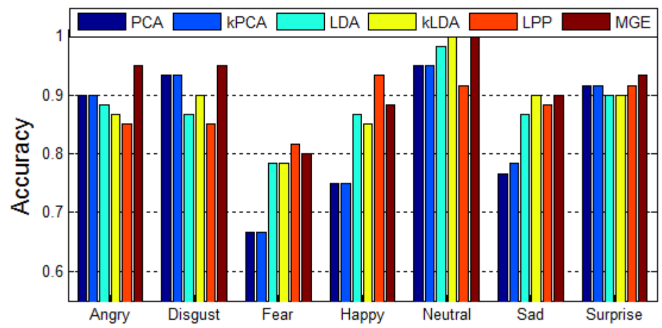
5. Experiments

5.1 Experimental Conditions

To investigate the performance of the proposed scheme, we carried out systematic experiments on three databases: JAFFE database [40], MUG expression database [41] and Cohn-Kanada database [42]. Each subject in these databases has seven facial expressions, i.e., angry (AN), disgust (DI), fear (FE), happy (HA), neutral (NE), sad (SA), and surprise (SU). The JAFFE database contains facial images of 10 Japanese females, where each has 3 sample images per each expression. In total, there are 210 grayscale facial expression images in this database, each of pixel resolution 256×256. The MUG database contains image sequences of facial expressions belonging to 86 subjects comprising 35 women and 51 men. Each image is of resolution 896×896. In our experiments, we use images of 45 subjects, where 4 typical images were selected from image sequences per subject per expression, totaling 1260 images. The CK+ includes both posed and non-posed (spontaneous) expressions and additional types of metadata. In our experiments, we use images of 55 subjects, where each subject has 4 images per expression and in total 1320 images are included for our tests. Fig.5 shows the selected representative facial expression images in these three databases, respectively.



k=1



k=2

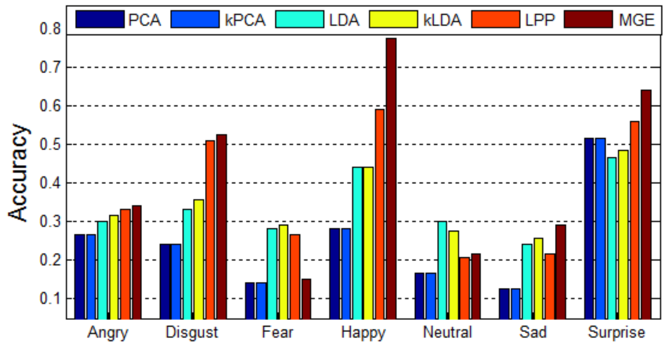
a) *Leave-k-sample-out* test results per expression

k	PCA	kPCA	LDA	kLDA	LPP	MGE
1	92.38	92.38	94.29	94.29	94.76	**95.24**
2	84.05	84.30	87.86	88.57	88.10	**91.67**

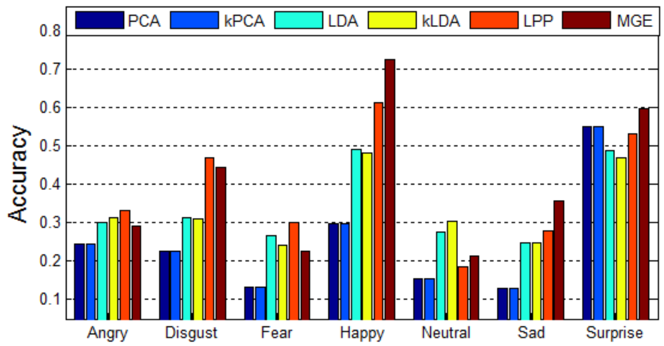
b) Overall accuracy on all expressions per *k* test

Fig.6. Overall *leave-one-sample-out* results on JAFFE.

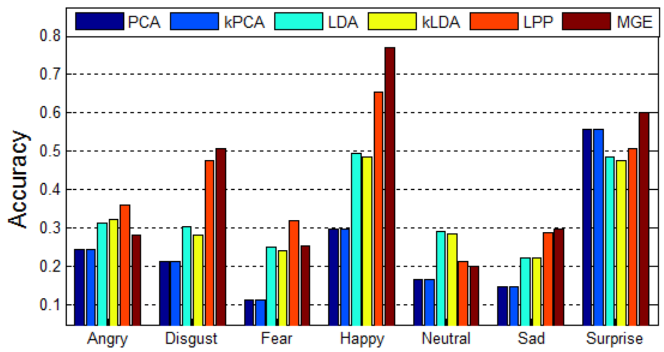
In our benchmark test, we implemented our codes on Matlab, and ran on a PC with 2.7GHz dual-core Intel CPU. Our benchmark tests aim to verify whether or not the proposed MGE can improve the accuracy on scrambled face recognition. Our approach is a pure data-driven face classification method. Hence, similar to Ref.[24], we compared our approach with a number of typical data-driven methods, including Eigenface [29], Fisherface [30], kPCA[30], kLDA[30], and Laplacianface (LPP) [33]. In the evaluation of the proposed scheme, we simply use the nearest neighbor classifier because any involvement of



k=2



k=5



k=7

a) *Leave-k-sample-out* test results per expression

k	PCA	kPCA	LDA	kLDA	LPP	MGE
2	24.88	24.88	33.81	34.64	38.33	**42.02**
5	24.76	24.76	34.00	33.76	38.76	**40.71**
7	24.90	24.90	33.74	33.03	40.20	**41.53**

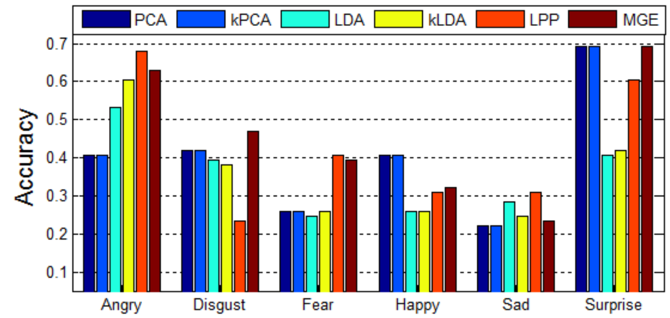
b) Overall accuracy on all expressions per *k* test Fig.7. Overall *leave-k-subject-out* results on MUG.

any other methods may blur the comparison and we then cannot easily assert if the enhancement comes from our MGE method or any other underlying more complicated classifiers.

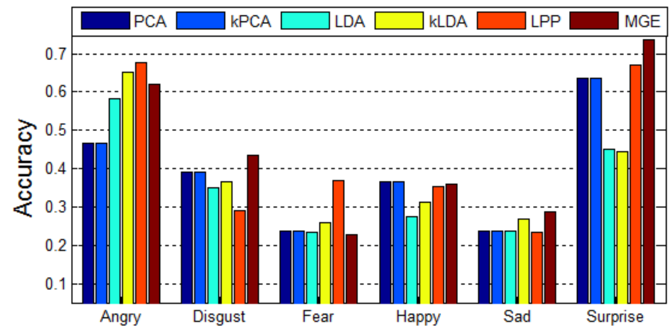
5.2 Validation on JAFFE dataset

The validation on JAFFE dataset is based on a typical test scheme called *leaving-k-sample-out*, where *k* samples per class/subject are left out as test samples and the rest are kept as train samples.

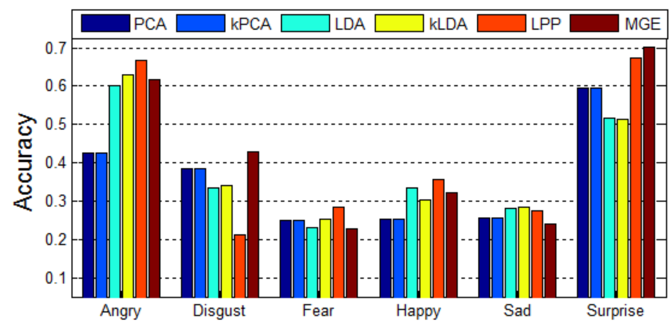
For a leaving *k* out scheme, there usually are C_N^k choices. In our experiment, we just chose *k* consecutive faces from *N* samples/subjects. As a result, we have *N* sets of tests in turn for



k=3



k=7



k=12

a) *Leave-k-sample-out* test results per expression

k	PCA	kPCA	LDA	kLDA	LPP	MGE
3	40.12	40.12	35.39	36.21	42.40	**45.68**
7	38.90	38.90	35.45	38.36	43.30	**44.36**
12	36.11	36.11	38.27	38.68	41.10	**42.28**

b) Overall accuracy on all expressions per *k* test Fig.8. Overall *leave-k-subject-out* results on CK+.

a *leave-k-out* experiment. The final accuracy is given by the average of all *N* tests.

Fig.6-a) gives the results for *leave-k-sample-out* tests with *k* equal to 1 and 2, respectively. As it is shown, there are obvious degradations in accuracy for all methods when *k* is increased from 1 to 2. This is because fewer samples were used to train the classifier. However, the proposed MGE attained the best average accuracy (95.24% and 91.67%) in both tests for all facial expressions, as shown in the Table in Fig.6-b).

5.3 Validation on MUG dataset

MUG dataset has more faces to use for tests. Our validation on MUG dataset is based on a similar test scheme called *leaving-k-subject-out*, where samples from *k* subjects are left out as test samples and the rest are kept as train samples. Usually this could be a bit more challenging than *leave-k-sample-out* while different subjects may have different ways for the same facial expression.

Fig.7-a) gives the results for *leave-k-subject-out* tests with various k values. As it is shown, different from the test on JAFFE, the degradations over k is small when k is varied from 2 to 7. This is largely because there are sufficient faces in the dataset for training. It is also seen that the proposed MGE method consistently attained the best average accuracy over all facial expressions, as shown in the table in Fig.7-b).

5.4 Validation on CK+ dataset

CK+ dataset has posed facial expression images, making it more challenging than previous two datasets. We still chose the *leave-k-subject-out* test scheme for our experiment on CK+ dataset. Fig.8-a) gives the results for leave-k-subject-out tests with various k values. Fig.8-b) shows the overall results per expression and it can be found that MGE attains the best accuracy in all k tests as well.

6. CONCLUSION

In conclusion, we identified a new challenge in scrambled facial expression recognition originated from the emerging IoT-orientated biometric security paradigm, and proposed a novel method which considers as many graphs as possible to find the discriminative subspaces for scrambled signals. In our experiments, the validation shows that the proposed MGE method can work well on the chaotic patterns of scrambled faces, and consistently attained higher accuracy in all our tests on facial expression datasets, making our method a promising candidate for emerging IoT applications.

REFERENCES

[1] L. D. Riek, R. N. M. Watson, "The Age of Avatar Realism", IEEE Robotics & Automation Magazine, Vol.17, Issue 4, 2010, 37 - 42.

[2] D. Jayatilake, T. Isezaki, Y. Teramoto, K. Eguchi, K. Suzuki, "Robot Assisted Physiotherapy to Support Rehabilitation of Facial Paralysis", IEEE Trans Neural Systems and Rehabilitation Engineering, Vol. 22 , Issue 3, 2014, Page(s): 644 - 653.

[3] Y. Balasubramanian, K. Sivasankaran; S. P. Krishraj, "Forensic video solution using facial feature-based synoptic Video Footage Record", IET Computer Vision, Vol.10, Issue 4, 2016, 315 - 320.

[4] D. McDuff, R. E. Kaliouby, R. W. Picard, "Crowdsourcing Facial Responses to Online Videos", IEEE Trans Affective Computing, Vol 3, Issue 4, 2012 , Page(s): 456 – 468.

[5] S. Fleck, W. Strasser, "Smart Camera Based Monitoring System and Its Application to Assisted Living", Proceedings of the IEEE, On page(s): 1698 - 1714 Volume: 96, Issue: 10, Oct. 2008

[6] A. Melle, J.-L. Dugelay, "Scrambling faces for privacy protection using background self-similarities," Proc. 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp.6046-6050.

[7] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, T. Toft, "Privacy-Preserving Face Recognition," Proc. Ninth Int'l Symp. Privacy Enhancing Technologies (PETS '09), 2009, pp.235-253.

[8] T. Honda, Y. Murakami, Y. Yanagihara, T. Kumaki, T. Fujino, "Hierarchical image-scrambling method with scramble-level controllability for privacy protection," Proc. IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS), 2013, pp.1371-1374.

[9] A. Erdlyi, T. Bart, P. Valet, T. Winkler, B. Rinner, "Adaptive Cartooning for Privacy Protection in Camera Networks". Proc. International Conference on Advanced Video and Signal Based Surveillance, 2014, pp.6.

[10] F. Dufaux, "Video scrambling for privacy protection in video surveillance: recent results and validation framework," Proceedings of SPIE, Vol. 8063, 2011, pp.14.

[11] T. Winkler, B. Rinner, "Security and Privacy Protection in Visual Sensor Networks: A Survey," ACM Computing Surveys, Vol.47, Issue 42, 2014, pp.1.

[12] A. Erdlyi, T. Bart, P. Valet, T. Winkler, B. Rinner, "Adaptive Cartooning for Privacy Protection in Camera Networks". Proc. International Conference on Advanced Video and Signal Based Surveillance, 2014, pp.6.

[13] Y. Wang, T. Li, "Study on Image Encryption Algorithm Based on Arnold Transformation and Chaotic System," Proc. 2010 International Conference on Intelligent System Design & Engineering Application, 2010, pp.449-451.

[14] Z. Tang, X. Zhang, "Secure Image Encryption without Size Limitation Using Arnold Transform and Random Strategies," Journal of Multimedia, Vol. 6, No. 2, April 2011, pp.202-206.

[15] Mariooryad, S. ; Busso, C., "Exploring Cross-Modality Affective Reactions for Audiovisual Emotion Recognition", IEEE Trans Affective Computing, Vol 4, Issue 2, 2013, Page(s): 183 - 196

[16] Wenming Zheng, "Multi-View Facial Expression Recognition Based on Group Sparse Reduced-Rank Regression", IEEE Trans Affective Computing, Vol 5 , Issue 1, 2014 , Page(s): 71 - 85

[17] Yongqiang Li ; Shangfei Wang ; Yongping Zhao ; Qiang Ji, "Simultaneous Facial Feature Tracking and Facial Expression Recognition", IEEE Trans. Image Processing, Vol 22, Issue 7, 2013, Page(s): 2559 - 2573.

[18] Bihan Jiang ; Valstar, M. ; Martinez, B. ; Pantic, M., "A Dynamic Appearance Descriptor Approach to Facial Actions Temporal Modeling", IEEE Trans Cybernetics, Vol 44, Issue 2, 2014, Page(s): 161 - 174

[19] P. Ekman and W. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.

[20] Perakis, P. ; Passalis, G. ; Theoharis, T. ; Kakadiaris, I.A. "3D Facial Landmark Detection under Large Yaw and Expression Variations", IEEE Trans Pattern Analysis and Machine Intelligence, Vol 35, Issue 7, 2013, pp.1552 – 1564.

[21] R. Hsu, A. Jain, "Semantic face matching", IEEE International Conference on Multimedia and Expo, 2002, pp.145.

[22] T. F. Cootes, G. J. Edwards, C. J. Taylor, "Active appearance models", IEEE Trans. Pattern Analysis & Machine Intelligence, Vol. 23, No. 6, 2001, pp.681.

[23] Taheri, S.; Patel, V.M.; Chellappa, R., "Component-Based Recognition of Faces and Facial Expressions", IEEE Trans Affective Computing, Vol 4, Issue 4, 2013, Page(s): 360 - 371.

[24] Y. Rahulamathavan, R. C.-W. Phan, J. A. Chambers, D. J. Parish, "Facial Expression Recognition in the Encrypted Domain Based on Local Fisher Discriminant Analysis", IEEE Tran. Affective Computing, Vols. 4, Issue 1, 2013, 83-92.

[25] R. Jiang, S. Al-Maadeed, A. Bouridane, D. Crookes, M. E. Celebi, "Face Recognition in the Scrambled Domain via Salience-Aware Ensembles of Many Kernels", IEEE Trans. Information Forensics & Security, Vol.11, No.8, Aug 2016, pp.1807-1817.

[26] Ping Liu, Shizhong Han, Zibo Meng, Yan Tong, "Facial Expression Recognition via a Boosted Deep Belief Network", CVPR 2014.
[27] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust Face Recognition via Sparse Representation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, Feb.2009. Also see CVPR 2014.
[28] Dong Chen, Xudong Cao, Fang Wen, Jian Sun. Blessing of Dimensionality: High-dimensional Feature and Its Efficient Compression for Face Verification. Computer Vision and Pattern Recognition (CVPR), 2013.
[29] B. Draper, K. Baek, M. Bartlett, J. Beveridge, “Recognizing faces with PCA and ICA”, Comput. Vision Image Understanding, Vol.91,Issue 1-2, 2003, pp.115.
[30] M. H. Yang, “Kernel Eigenfaces vs. kernel Fisherface: face recognition using kernel methods”, International Conference on Automatic Face and Gesture Recognition, 2002, pp.215.
[31] B. Tenenbaum, V. Silva, J. Langford, “A global geometric framework for nonlinear dimensionality”, Science, Vol.290, No.5500, 2000, pp.2319.
[32] M. Belkin, P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering", Advances in Neural Information Processing Systems, 2001, pp.14.
[33] X. He, S. Yan, Y. Hu, P. Niyogi, H. J. Zhang, “Face Recognition Using Laplacianfaces”, IEEE Trans. Pattern Analysis & Machine Intelligence, Vol. 27, No. 3, Mar. 2005, pp.1.
[34] Xiaofei He, Deng Cai and Partha Niyogi, "Tensor Subspace Analysis", Advances in Neural Information Processing Systems 18 (NIPS), Vancouver, Canada, Dec. 2005.
[35] D. Foster, S. Kakade, T. Zhang, Multi-view dimensionality reduction via canonical correlation analysis, TTI-C Technical Report, TTI-TR-2008-4, 2008.
[36] Hedge et al., Random Projections for Manifold Learning, NIPS 2008
[37] Hou C, Zhang C, Wu Y, et al. Multiple view semi-supervised dimensionality reduction. Pattern Recognition, 2010, 43(3): 720-730.
[38] T. Xia, D. Tao, T. Mei, and Y. Zhang. Multiview spectral embedding. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 40(6):1438–1446, 2010.
[39] Y. Han, F. Wu, D. Tao, J. Shao, Y. Zhuang, and J. Jiang. Sparse unsupervised dimensionality reduction for multiple view data. IEEE Transactions on Circuits and Systems for Video Technology, 22(10):1485–1496, 2012.
[40] M. Lyons, J. Budynek, and S. Akamatsu, “Automatic Classification of Single Facial Images,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 12, pp. 1357-1362, Dec. 1999.
[41] Niki Aifanti and Anastasios Delopoulos, "Linear subspaces for facial expression recognition", Signal Processing: Image Communication, 29, (1), pp. 177-188, 2014 Jan.
[42] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I., “The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression”, Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis, San Francisco, USA, 94-101.