Face Recognition in the Scrambled Domain via Salience-Aware Ensembles of Many Kernels

Face Recognition in the Scrambled Domain via Salience-Aware Ensembles of Many Kernels

Richard Jiang, Somaya Al-Maadeed, Ahmed Bouridane, Danny Crookes *IEEE, Senior Member* and M. Emre Celebi *IEEE, Senior Member*

***Abstract* — With the rapid development of internet-of-things (IoT), face scrambling has been proposed for privacy protection during IoT-targeted image/video distribution. Consequently in these IoT applications, biometric verification needs to be carried out in the scrambled domain, presenting significant challenges in face recognition. Since face models become chaotic signals after scrambling/encryption, a typical solution is to utilize traditional data-driven face recognition algorithms. While chaotic pattern recognition is still a challenging task, in this paper we propose a new ensemble approach – Many-Kernel Random Discriminant Analysis (MK-RDA) to discover discriminative patterns from chaotic signals. We also incorporate a salience-aware strategy into the proposed ensemble method to handle chaotic facial patterns in the scrambled domain, where random selections of features are made on semantic components via salience modelling. In our experiments, the proposed MK-RDA was tested rigorously on three human face datasets: the ORL face dataset, the PIE face dataset and the PUBFIG wild face dataset. The experimental results successfully demonstrate that the proposed scheme can effectively handle chaotic signals and significantly improve the recognition accuracy, making our method a promising candidate for secure biometric verification in emerging IoT applications.**

***Index Terms* — Facial biometrics, face scrambling, many manifolds, many kernels, random discriminant analysis, mobile biometrics, Internet-of-Things, user privacy.**

I. INTRODUCTION

WITH rapid developments in Internet-of-Things (IoT) technology, face recognition [1~4] has recently found a new use in web-based biometric verification, man-machine interaction, internet medical diagnosis, video conferencing, distance learning, visual surveillance, and psychological evaluation. In the context of mass internet technology, privacy [5~15] has become an issue of wide concern in web-based video streaming. As a result, face scrambling [5] is emerging as a practical technique to protect privacy legally during video distribution over the public internet. By scrambling faces detected in private videos, the privacy of subjects can be respected, as shown in Fig.1.

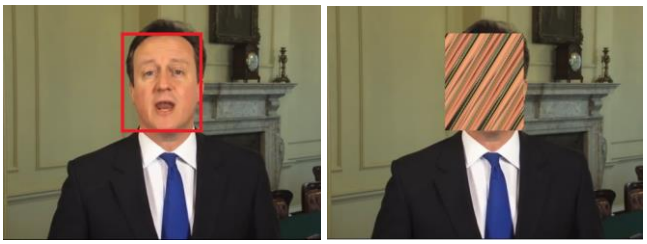


Fig.1. A detected face in video scrambled by using the Arnold transform.

Fig.2. Semantic approaches such as using AAM [18]~[25] for facial emotion estimation cannot be applied in the scrambled domain.

Compared with full encryption methods, face scrambling is a compromise choice because it does not really hide information, since unscrambling is usually achievable by simple manual tries even though we do not know all the parameters. It avoids exposing individual biometric faces without really hiding anything from surveillance video. As shown in Refs.[5~14], scrambling has recently become popular in the research field of visual surveillance, where privacy protection is needed as well as public security. Another advantage of face scrambling over encryption is its computing efficiency, and usually it is far simpler than complicated encryption algorithms. In many business cases such as public surveillance, the purpose is limited to only privacy protection from unintentional browsing of user data. Hence, full encryption becomes unnecessary in this context.

There are many ways to perform face scrambling. For example, scrambling can be done simply by masking or cartooning [8]. However, this kind of scrambling will simply lose the facial information, and hence subsequent face recognition or verification becomes unsuccessful in this case. Especially for security reasons, it is obviously not a good choice to really erase human faces from surveillance videos. In

The manuscript was received _____, revised _____.

Richard Jiang and Ahmed Bouridane are with Computer Science & Digital Technologies, Northumbria University, Newcastle upon Tyne, UK.

Somaya Al-Maadeed is with Department of Computer Science & Engineering, Qatar University, Doha, Qatar.

Danny Crookes is with ECIT Institute, School of Electronics, Electrical Engineering & Computer Science, Queen's University Belfast, Belfast, UK.

M. Emre Celebi is with Department of Computer Science, University of Central Arkansas, Conway, AR, USA.

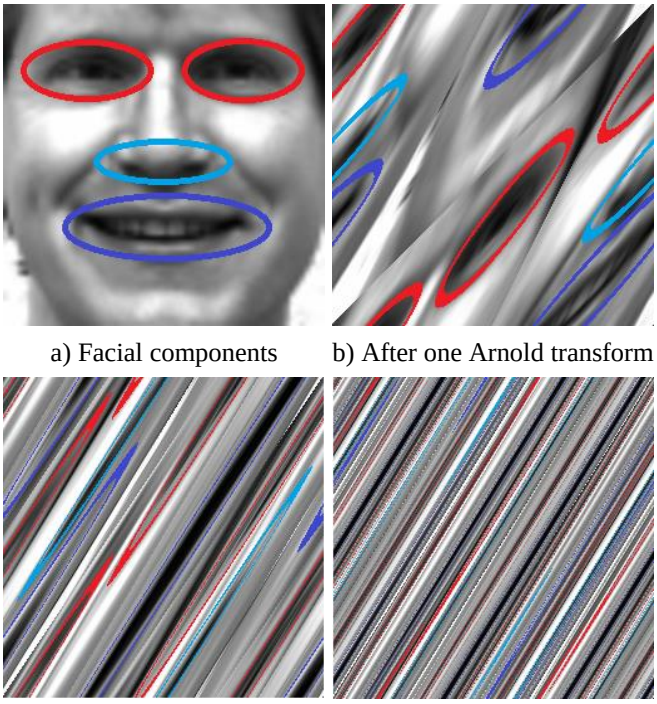
Correspondence e-mail: richard.jiang@acm.org.

comparison, the Arnold transform [13, 14], as a basic step in many encryption algorithms, is a kind of recoverable scrambling method. Scrambled faces can be unscrambled by several manual tries. Hence, in this work, we have chosen Arnold transform based scrambling as our specific test platform.

Face recognition has been extensively researched in the past decade and significant progress has been seen towards better recognition accuracy in recent reports [15~21]. These approaches usually exploit semantic face models [22~23] where a face is considered as an integration of semantic components (such as eyes, nose and mouth), and hence semantic related sparse features or local binary patterns (LBP) can be effectively used to improve the recognition accuracy. Beyond 2D facial modelling, 3D models [23] can also be exploited for better accuracy by taking advantage of 3D face alignment.

However, as shown in Fig.2, a scrambled face has a very different appearance from its original facial image. While we can easily match a 3D model to a normal facial image, it becomes extremely hard to do so after the face has been scrambled. In the scrambled domain, semantic facial components simply become chaotic patterns. In this context, it becomes difficult to exploit landmarks or 3D models for better accuracy. As shown in Fig.2, while face models can be easily fitted with a facial image, it becomes impossible after a face is scrambled into chaotic patterns. As has been discussed in [15], one straightforward way is to use traditional data-driven approaches, where chaotic signals are treated simply as a set of data points spread over manifolds.

Various data-driven face recognition algorithms have been developed over several decades. In the early days, linear dimensionality reduction [24~27] was used for this challenge, such as principal component analysis (PCA) [24], independent component analysis (ICA) [24], and Fisher's linear discriminant analysis (FLD) [25]. With kernel methods (KM) [26], these methods can be extended to a reproducing kernel Hilbert space with a non-linear mapping, and extended as k-PCA and k-FLD. Recent progress on nonlinear manifold learning [27~32] has produced a number of new methods for face recognition, such as Laplacianface [30] and Tensor subspace [31]. These approaches have been successfully used for data-driven face recognition. However, for face recognition in the scrambled domain, we need a robust approach to handle chaotic signals in the scrambled domain, which appear random and beyond human perception.



a) Facial components b) After one Arnold transform

c) After 2 Arnold transforms b) After 3 Arnold transforms

Fig.3. Face scrambling by the Arnold transform.

In recent research, multi-kernelization [32, 33] has been proposed to handle the complexity of data structure, where it is believed multiple-view discriminative structures [34, 35] need to be discovered where a manifold may have different geometric shapes in different views. With the hope of utilizing this approach for chaotic signals, in this paper we propose a new approach called Many Kernel Random Discriminant Analysis (MK-RDA) to handle this new challenge of chaotic signal recognition in the scrambled domain. We also propose a mechanism to incorporate a salience model [36] into MK-RDA for pattern discovery from chaotic facial signals, since it is believed that semantic features are usually salient and useful for facial pattern classification.

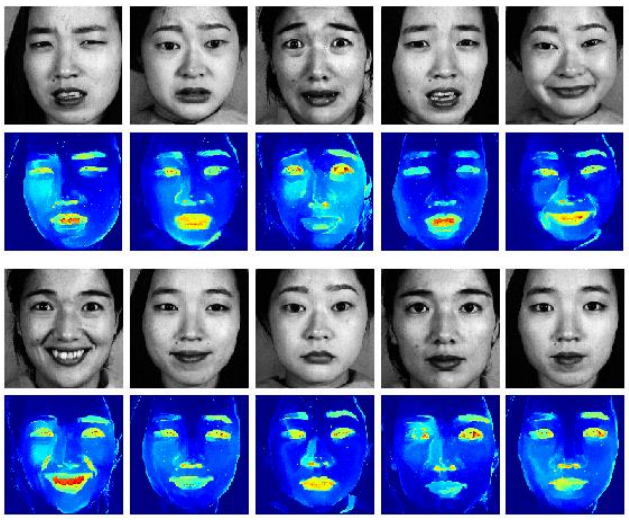
In the following sections, facial image scrambling using the Arnold transform is introduced in section II, and the semantic mapping of facial components for robust feature extraction in the scrambled domain is described. In section III, we introduce the background and motivation of our "many kernel" ensemble method, and present our many-kernel random discriminant analysis. In Section IV, we present the framework using MK-RDA with the salience model for chaotic facial pattern verification. Section V gives the experimental results on three face datasets, and conclusions are drawn in Section VI.

II. Facial Components in the Scrambled Domain

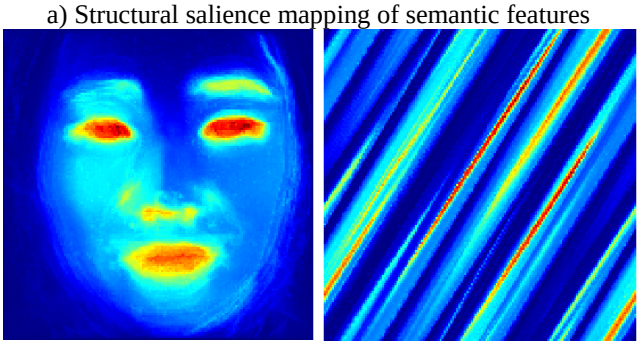
A. *Face Scrambling*

In many IoT applications, it is not encouraged to hide any information by encryption; on the other hand, it is legally required to protect privacy during distribution and browsing. As a result, scrambling becomes a compromise choice because it doesn't really hide information (unscrambling is usually achievable by simple manual attempts), but it does avoid exposing individual faces during transmission over the internet. Additionally, scrambling usually has much lower computation cost than encryption, making it suitable for simple network-targeted applications using low power sensors.

Among various image scrambling methods, the Arnold scrambling algorithm has the feature of simplicity and periodicity. The Arnold transform [11, 12] was 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 surveillance systems where it is favored as a simple and efficient scrambling method which nevertheless retains some spatial coherence. In this paper, we use this scrambling method to set up the test environment of our algorithm in the scrambled face domain.



a) Structural salience mapping of semantic features



b) Summarized semantic map c) Scrambled semantic map

Fig.4. Semantic salience of facial images

In the Arnold transform, a pixel at point (x, y) is shifted to another point (x', y') by:

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

which is called two-dimensional Arnold scrambling. Here, x and y are the coordinates of the original pixel; N is the height or width of the square image processed; x' and y' are the coordinates of the scrambled pixel. The Arnold transform can be applied iteratively as follows:

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

Here, the input is the original image after the *k-th* Arnold transform, and P_{xy}^{k+1} on the left is the output of the k+1th Arnold transform. k represents the number of iterations, where k = 0, 1, 2 and so on.

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

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

As we can see from Fig.3, before scrambling, facial components can easily be identified by the human eye. After scrambling, the images become chaotic signals, and it is hard to figure out eyes and noses. Since semantic facial components are considered important cues for face recognition, we need to find a way to incorporate semantic approaches into the scrambled domain to attain higher matching accuracy.

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

B. Semantic Facial Components

Fundamentally a 2-D face image is the projection of a real 3-D face manifold. This viewpoint leads to model-based face recognition, where semantic facial components (such as eyes, nose, and lips) are modeled by their parameters. A very frequently applied face model is the active appearance model (AAM) [20]~[23]. 3D facial information is better for describing the semantic facial components in the presence of illumination and pose changes, where 2-D descriptors sometimes turn out to be less effective. Hsu and Jain [23] have advocated that such semantic facial components constitute the meaning of a face and decisively form the basis of face recognition.

Along this roadmap, template-based face description [21] has been considered to emphasize the importance of semantic facial components. In our human perception system, concept-level semantic features are more meaningful than pixel-level details. A good emotion estimation model usually relies on the importance of semantic features. Changes in a single pixel or sparse set of pixels should not distort the final decision.

Though semantic approaches have attained great success in facial analysis, they need a robust scheme to map a 2D image into its semantic feature space or 3D deformable model. This computation is not trivial and usually cannot be afforded by many real-world applications such as mobile computing platforms. Besides, the detection of semantic features can be sensitive to different conditions, and hence produces extra errors in face classification. To take advantage of semantic features without worrying about its computing complexity, in this paper we introduce a salience-aware method into our facial analysis.

C. Semantic Salience Mapping of Facial Images

Since semantic components are important cues to identify a specific face, we need to find a way to introduce these factors in statistic face modelling. In this paper, we propose to use salience learning for semantic facial mapping, and incorporate the learned semantic map into a random forest method for face recognition.

As shown in Fig.4-a), facial components are usually salient features in a facial image. In this paper, we employ the *Deep Salience* model [39] for sematic feature mapping. Unlike other models based on color salience using pixel contrast, this deep salience model bases its algorithm on structural salience, and

hence can easily find the semantic components as its salient features, as shown in Fig.4-a). This fits well with our purpose to exploit semantic components in a facial image.

We then apply a Gaussian mixture model to summarize the learned salience maps of the training dataset, where the salience distribution is considered as a mixture of Gaussian functions,

$$p(x \mid \lambda) = \sum_i w_i g(x \mid \mu_i, \sigma_i) \qquad (3)$$

where $g(x \mid \mu_i, \sigma_i)$ is the normalized Gaussian distribution with mean μ_i and variance σ_i. In our work, we use a two-class GMM model and estimate the probability of a pixel being salient or non-salient. Learning with GMM mixtures can find optimized Gaussian distribution parameters in the GMM model, and consequently produce a distribution map $S=p(x|\lambda)$ from Eq.(2), which is referred to as the semantic importance map in this paper.

Fig.4-b) shows the estimated semantic importance map learned from Fig.4-a), which highlights semantic features such as eyes, nose and mouth. This importance map represents the importance of each feature subspace in terms of its relation to semantic features. Fig.4-c) shows the scrambled semantic map. Once we have the semantic salience map of the training dataset, we can then use it to guide the feature sampling to favor semantic features.

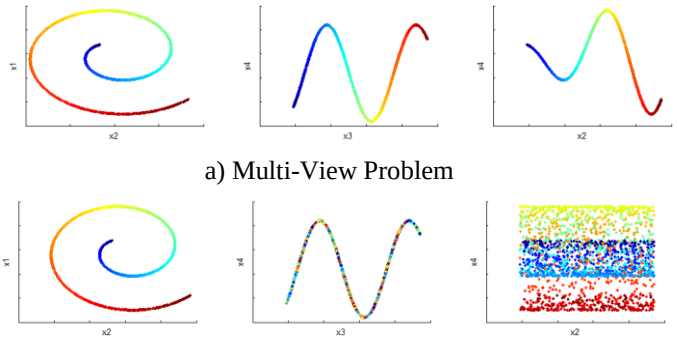
III. Ensembles of Many-Kernel Discriminant Analysis

A. Background on Multi-Kernel Approaches

In many real world applications such as face recognition and image classification, the data often has very high dimensionality. Procedures that are computationally or analytically manageable in low-dimensional spaces can become completely impractical in a space having several thousand dimensions. This has been well known in machine learning as a notorious issue --- the "*Curse of Dimensionality*" [1~3]. To tackle this challenge, various techniques [1~12] have been developed for reducing the dimensionality of the feature space, in the hope of obtaining a more manageable problem. Dimensionality reduction has become an especially important step for face classification.

Various algorithms have been developed for image-based face recognition. In this paradigm, dimensionality reduction [19] has always been a primary concern. As mentioned previously, methods developed for this challenge include principal component analysis (PCA) [24], independent component analysis (ICA) [24], and Fisher's linear discriminant analysis (FLD) [25]. With kernel methods (KM) [26], these methods can be extended to a reproducing kernel Hilbert space with a non-linear mapping, and extended as k-PCA, k-ICA and k-FLD. Recent progress on nonlinear manifold learning [27]~[31] has led to a number of new methods for face recognition, such as Laplacianface [35], Tensor subspace [36], non-negative matrix [37], and local Fisher discriminant analysis (LFDA) [38,22]. These approaches usually assume there is an underlying discriminative structure to discover, which leads to the paradigm of manifold learning.

Recently, the multi-view problem has been noticed by the research community, where the same manifold can have different shapes in different subspaces, as shown in Fig.5-a).



a) Multi-View Problem

b) Multiple Manifold Problem

Fig.5. Multi-view dataset and multi-manifold dataset. When the sequence of data points in the second subspace is shuffled, the two sub-manifolds become independent of each other, and cannot be unified in a higher dimensional subspace.

Foster et al. have employed canonical correlation analysis (CCA) [32] to derive the low dimensional embedding of two-view data and to compute the regression function based on the embedding. Hedge et al [33] propose a multiple projection approach from the same manifold. Hou et al [34] used the pairwise constraints to derive embedding in multiple views with linear transformation. Xia et al [35] combined spectral embedding with the multi-view issue. Han et al. [36] proposed a sparse unsupervised dimensionality reduction to obtain a sparse representation for multi-view data. Lin et al [37] proposed multiple kernel learning of a manifold, where various kernel spaces are constructed with different sets of parameters. Zien et al [38] considered multiple kernels with regards to multi-class cases.

In the multi-view problem, as shown in Fig.5-a), although a manifold has different forms in different subspaces, these forms can always be unified as the same manifold in a higher-dimensional subspace. However, this may not always be true. As shown in Fig.5-b), when the sequence of data points in the second subspace is shuffled, the combination of two submanifolds simply creates a noisy-like distribution. This means two submanifolds cannot be merged at all. In this case we have to treat it as a multiple or even "many manifold" problem, where multiple manifold structures need to be discovered.

In our facial recognition in the scrambled domain, facial images become chaotic signals, as shown in Fig.1 and Fig.2. In this real-world case, its underlying discriminative structures could be more like the case in Fig.5-b), where multiple manifold structures need to be discovered. In this paper, we include this case in our consideration and propose a new many-kernel approach to handle its complexity. Before we go further, we give an introduction to kernel based analysis.

B. Preliminary on Kernel based Discriminant Analysis (KDA)

For a set of data points $\{x_i\} \in R^N$, we may select a set of data points as the landmarks $\{L_j\}$ that can characterize this dataset. A data point on the manifold then can be located by its kernel distance to the landmarks:

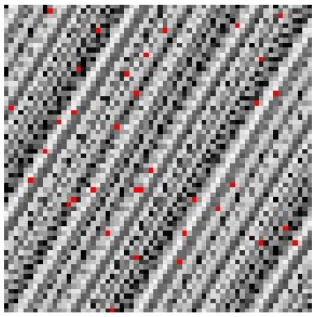
$$x_i \to \kappa_i = \{K(x_i, L_j)\} \qquad (4)$$

Hence, each data point is represented in the constructed kernel

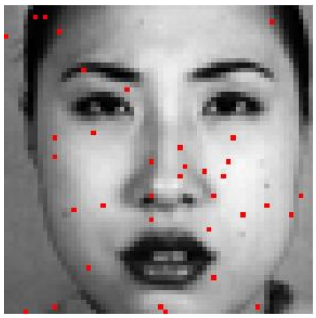
List I. Random Generation of Many Kernels

Input:
$\{x_i\}$ – Dataset;
L_K – Number of kernels;
Output:
$\{\kappa_i^k\}$ – Constructed "many kernel" representations;

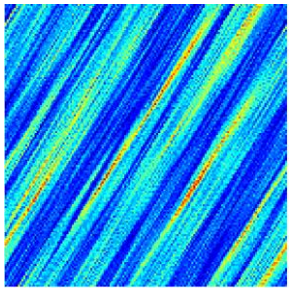
Process:
Loop for L_K times
  Generate random selection $\{l_j^m\}$
  Select K_L landmarks from $\{x_i\}$
  Loop for each data point x_i
    Compute its kernel representation κ_i^k based on $\{l_j^k\}$.
  End Loop
End Loop
Return $\{\kappa_i^k\}$.



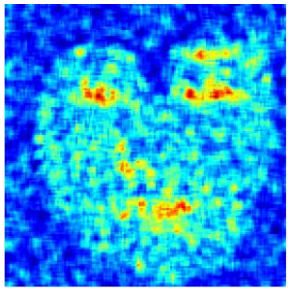
a) Random feature selection in scrambled domain guided by the salience map in Fig.4-c).



b) The corresponded pixels on the original facial image.



c) Actual hit rates in scrambled domain.



d) Unscramble the hit map back to facial domain.

Fig.6. Selecting kernel subspaces toward semantic features.

space as $\kappa_i \in R^M$, where M is the number of selected landmarks. Following this, we then simply apply Fisher's linear discriminant analysis in the kernel space:

$$\arg\max_{\Phi} \frac{\left\|\Phi^T S_B \Phi\right\|}{\left\|\Phi^T S_W \Phi\right\|} \qquad (5)$$

where Φ is the projection matrix, and S_B is the between-class covariance matrix:

$$S_B = \sum_{c=1}^{Kc} n_c \left\|\mu_c - \bar{\mu}\right\| \qquad (6)$$

and S_w is the within-class covariance matrix:

$$S_W = \sum_{c=1}^{Kc} \sum_{i=1}^{n_c} \left\|\kappa_i - \mu_c\right\| \qquad (7)$$

By optimizing over Eq.(7), we then have the Eigen projection matrix W, and each data point is then represented by its new coordinates in the KDA space:

$$y_i = \Phi \kappa_i \qquad (8)$$

Here, Φ is an Eigen matrix $\in R^{D \times M}$, $y_i \in R^D$, and D is usually a number smaller than M as well as smaller than the number of classes in the training dataset $\{x_i\}$.

C. *Many Kernels for the Many Manifold Problem*

Though it has been assumed in many methods that there is only one underlying manifold structure, it is obvious that there can often be multiple manifolds underlying many real-world datasets, as shown in Fig.5-b). However, the discovery of the underlying manifold structures is an inverse engineering problem that could be very complex, and often intractable.

For example, consider selecting M dimensions out of the feature space $\in R^N$: there are $K=N!/\{M!(N-M)!\}$ such choices that can be made, and within each selection an independent sub manifold may be discovered. For example, when N=10 and M=5, K will be 252. For a facial image, there could be 64×64=4096 dimensions, and M could be any number. Hence, the estimation of possible subspaces becomes an NP-hard problem that cannot be handled exhaustively in realistic computing time. Hence, the discovery of "many manifolds" becomes a major challenge that has not yet been fully appreciated.

In this work, to address the challenge shown in Fig.5-b), we propose a randomization strategy to generate "many kernels" and try to cover as many manifolds as possible in a given dataset by chance, which reduces the complexity of the "many manifolds" problem from its exponential computing time to something manageable.

D. *Many Kernels from Random Feature Selection*

If we have K data points $\{x_i\}$, then typically the random selection of subspaces can be easily attained by generating a list of random numbers l_k, and selecting K_L features to construct the new datasets:

$$z_j^k \sim x_j(l_k) \qquad (9)$$

Here, $\{z_j\} \in R^{K_L}$. Then we can construct a kernel space based on this randomly selected subspace:

$$x_i \to \kappa_i^k = \left\{K\left(z_i^k, z_j^k\right)\right\} \qquad (10)$$

We can repetitively redo the above randomization process, and as a result, we can easily construct as many kernels as we want.

If we have L_K kernels and each kernel has K_L dimensions, then for each data point x_i, we will have the kernel representation $\{\kappa_i^k\}$ actually as an $L_K \times K_L$ matrix. To guarantee the kernelized dimensions are not too much more than the original data dimensions, we add a constraint:

$$L_K \times K_L \sim N, \qquad (11)$$

which means the "many kernel" process will not increase or decrease the dimensions. This process is outlined in List I.

E. *Many-Kernel Random Discriminant Analysis*

The purpose of this many-kernel strategy is to find the underlying discriminative structures in each subspace. After we obtain the many kernel based representation κ_i^k, we can then apply discriminant analysis over each kernel subspace and find

its discriminative projection.

For a set of training data and its kernel representation $\{\kappa_i^k\}$, we can calculate its within-class covariance at its *k-th* kernel subspace as:

$$S_W^k = \sum_{c=1}^{Kc}\sum_{j=1}^{n_c}\left\|\kappa_j^k - \mu_c^k\right\| \qquad (12)$$

and its between-class covariance matrix:

$$S_B^k = \sum_{c=1}^{Kc} n_c\left\|\mu_c^k - \bar{\mu}^k\right\| \qquad (13)$$

To find the most discriminative features, we can maximize its between-class covariance over its within-class one by finding a projection matrix Φ^k:

$$\Phi^k \sim \arg\max_{\Phi^k}\frac{\left\|\Phi^{k^T} S_B^k \Phi^k\right\|}{\left\|\Phi^{k^T} S_W^k \Phi^k\right\|} \qquad (14)$$

By optimizing over Eq.(10), we then have the Eigen projection matrix $\Phi^k \in B^{D\times K_L}$. For each data point κ_i^k, we can then have its discriminant projection in its *k*-th subspace:

$$y_i^k = \Phi^k \kappa_i^k \qquad (15)$$

For each kernel subspace, we can obtain the kernel discriminant projection for each data point. As a result, we will have the L_K projection:

$$Y \sim \left\{y_i^k\right\} \qquad (16)$$

where Y will be a matrix $\in B^{D\times L_K}$.

IV. Facial Semantic Aware Ensembles of Many Kernels

A. Salience-Biased Feature Space Reconstruction

Unsurprisingly, salient features usually play an important role in face classification. Therefore, rationally we can expect a mechanism to give salient features more weight than others. In this work, we consider a biased strategy to reconstruct the feature space to favor semantic salient features.

Considering a scrambled facial image x as a vector of facial features/signals $\{f_1, f_2..., f_k, ...\}$, and a semantic salience map $S\sim\{s_1, s_2..., s_k, ...\}$ learned from training (as shown in Fig.4-c), we can then construct a new feature space by replicating each feature according to its semantic importance. Assuming the maximum multiplicative factor as K_s, the repetition of each feature is then defined as:

$$k_i = \text{int}\left\{1 + K_s\frac{s_i}{\max_i\{s_i\}}\right\} \qquad (17)$$

Here, k_i means how many times the *i*-th feature/signal will be repeated, and s_i is the salience value of the *i*-th signal shown in Fig.4-c). Consequently, we have a new set of features:

$$\chi_{new} = \{\underbrace{f_1,..., f_1}_{k_1},..., \underbrace{f_k,..., f_k}_{k_k},.....\} \qquad (18)$$

With the above multiplicative process, salient features will have a higher likelihood to be chosen in the randomized selection process in Eq.(9).

We then can apply the random selection to select subspaces from the reconstructed feature space χ_{new} to form the "many kernels" for MK-RDA. Fig.6 shows the results of such a salience-guided selection using the scrambled salience map in Fig.4-c). We can see that with the salience guiding, semantic facial features will be more likely to be used to form our kernels subspaces.

B. Salience-Aware MK-RDA

After the feature space is reconstructed, we can apply *MK-RDA* on the reconstructed datasets $\{\chi_i\}$ instead of $\{x_i\}$, and we have:

$$Y \to \Psi\times K\left(z(\chi_i), z(\chi_j)\right), \Psi \sim \left\{\Phi^k\right\} \qquad (19)$$

At the end, we will have Y as a matrix $\in B^{D\times L_K}$.

For any two data points x_1 and x_2, their distance in the projected subspaces can be calculated as:

$$d_{ij} = \left\|y_i^k - y_j^k\right\|^2, \qquad (20)$$

Here $\|\cdot\|$ denotes the Euclidean norm.

For data classification, the likelihood of a data point belonging to a class c can be estimated from its distances to all training data points in the *k-th* learned kernel subspace:

$$P\left(c\,|\,\Phi^k\right) = \frac{\hat{P}\left(c\,|\,\Phi^k\right)}{\sum\hat{P}\left(c_k, \Phi^k\right)} \qquad (21)$$

Here, $P(c|\Phi^k)$ denotes the estimated probability in the *k-th* kernel projection Φ^k that an input data point x belongs to a class c (c = 1, 2,..., n_c). For all kernels, the discriminant function is defined as:

$$\varphi_c(x) = \frac{1}{K}\sum_{k_t}\tilde{P}\left(c\,|\,\Phi^m\right) \qquad (22)$$

and the decision rule is to assign x to class c for which $\varphi_c(x)$ is the maximum.

C. Overview of the Salience-Aware Scheme

Fig.7 gives an overview of the proposed salience-aware scheme for scrambled face verification. Given a training dataset, faces are forwarded to the training procedure. The offline procedure then learns its semantic salience map. Following this, the database is scrambled and the feature space is reconstructed by multiplying salient features according to their semantic salience weights. Random sampling is then applied to select features sparsely to construct as many kernels as is allowed, and discriminant analysis is used to learn a kernel subspace for each kernel.

After a scrambled facial image is input as a test, the input is

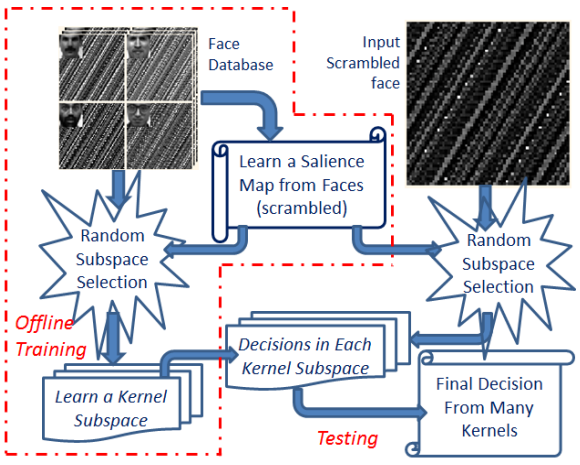
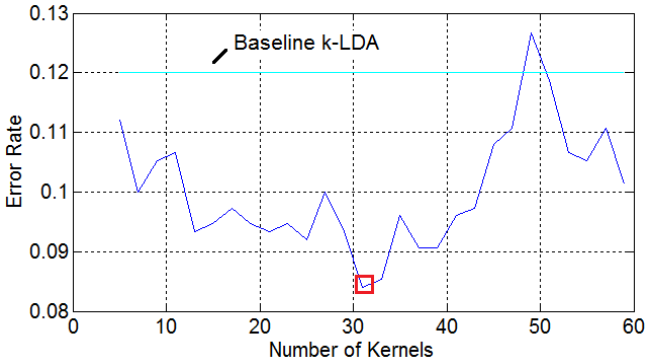
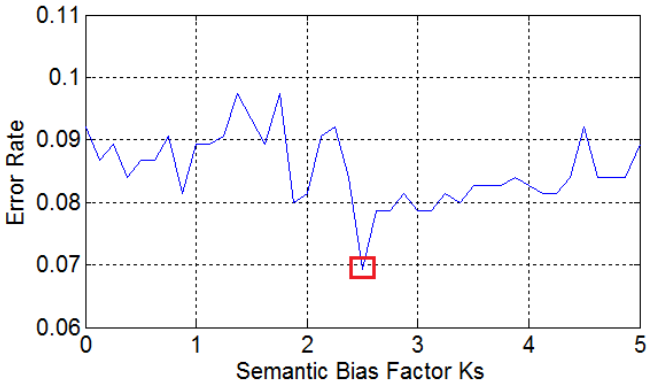


Fig.7 Overview of the proposed salience-aware scheme

a) A small face dataset --- Yale dataset.



b) Number of kernels in MK-RDA



c) Semantic weight factor *Ks*

Fig.8 Parameters in Salience-Aware MK-RDA

projected into each kernel subspace, and the distance to each training sample is computed. The decision procedure is based on the combination of all kernel subspaces via Eq.(22).

It is noted that we can have unscrambled images (mainly for statistic salience learning) in the offline training because offline training is carried out centrally with authorities'/business supervisors' permission and will not undermine users' privacy. Privacy protection is mainly an issue with distribution over the internet.

In this scheme, the training procedure can be carried out offline. The online verification then becomes purely a data-driven process. In the test procedure, all test images and semantic maps are scrambled for privacy protection, and no original face will be utilized for recognition purposes. Hence, it is similar to other data-driven approaches, and is simple and straightforward.

D. Discussion of Salience-Aware MK-RDA

Before we proceed to our benchmark experiments, there are two questions that need to be answered. First, in the MK-RDA mechanism, what is the best L_K to choose? Namely, how many kernels are enough? Second, in the above salience-aware mechanism, can such a salience biased mechanism really help attain better accuracy in face recognition? Here, we design an experiment to find out the answers to these two questions.

For this experiment, we chose the Yale face dataset [40] for our tests. In the Yale dataset, each of the 15 subjects has 11 sample faces with different expression, illumination and glasses configuration. We only choose 6 faces with different expressions for our test, as shown in Fig.8-a). With this small dataset, we carried out the face recognition tests by splitting the small dataset into training and test datasets, where the training dataset has five subjects and test dataset has the rest. We then varied L_K, the number of kernels, and K_s, the max weight of salience map, in our experiments. We then examined which set of parameters gives the best error rates. Fig.8 shows the results of our experiment.

Fig.8-b) gives the experiment results on the number of kernels. Given K_s as 1.5, the number of kernels varied from 5 to 60. We can see that the error rate is lowest when L_K is around 32. Basically, more kernels mean more computing time. As long as we have a low error rate, using fewer kernels is often preferable. It is also observed that compared with the baseline *kLDA*, MK-RDA has attained marginally better accuracy.

We then ran an experiment on K_s. As shown in Eq.(17), K_s=0 means no bias. The bigger K_s is, the more biased it is toward the salient features. Fig.8-c) shows the experimental results. It can be seen that the error rate is lowest when K_s is around 2.5. It is also observed that biased sampling with higher K_s simply worsens the accuracy because it means some non-salient features may be abandoned in the random process even though they may contribute to the recognition process.

V. EXPERIMENTS

To validate our algorithm, we implemented our face recognition method in Matlab, and ran on a PC with 2.5GHz dual-core Intel CPU. Before running the benchmark on face datasets, all images in the datasets were scrambled using the (triple) Arnold transform [7~8]. Fig.11 shows selected face images from the three datasets: ORL, PIE and PUBFIG.

The ORL database has 40 subjects, each with 10 faces at different poses. In total, 400 faces are used for the test. The CMU PIE database [40] has 41,368 faces, comprising 67 classes with about 170 faces per class, including a wide spectrum of variations in terms of pose, illumination, expression and appearance. In our tests, we use 50 faces per subject, similar to [30] and [31].

The PUBFIG database [42] contains wild faces selected from the internet. It is very similar to LFW [43] but it provides standard cropped faces. As has been shown [43], background textures in LFW can help attain a higher accuracy. Since we consider face recognition only, PUBFIG fits better with our purpose.

In many previous reports [9], the leave-one-out test scheme

has been widely used. However, this test is too simple because it leaves one image out as the test image and keeps all the rest in the training set. In our test scheme, we try to increase the challenge and adopt a test scheme called *leaving-k-out*, where in each test *k* samples per category are left out as test samples. For example, we have *N* samples, and then we choose all faces of (*N*-k) samples as the training dataset, and use *k* samples for the test.

For a leaving *k* out scheme, there are usually C_N^k choices. In our experiment, we just chose 3 sets of consecutive faces from *N* samples, starting at *N*/4, *N*/2 and 3*N*/4. As a result, we have 3 sets of tests in turn for a *leave-k-out* experiment. The final accuracy is given by the average of all three tests. It is noted that the consecutive splitting will usually bring out the large difference between test and training datasets, because the datasets have faces varied consecutively and the first *k* faces are usually very different from the last (*N*-*k*) faces.

Our benchmark tests aim to verify whether or not the proposed MK-RDA can enhance the accuracy on scrambled face recognition. Our approach is a pure data-driven face classification method. Hence, similar to Ref.[15], we compared our approach with a number of typical data-driven methods, including Eigenface [25], Fisherface [25], kPCA[26], kLDA[26], and Laplacianface (LPP) [31], each applied to facial images in the scrambled domain. In the evaluation of the proposed scheme, we simply use the nearest neighbor classifier because any involvement of any other methods may blur the comparison and we then cannot easily assert if the enhancement comes from our MK-RDA scheme or any other underlying more complicated classifiers.

A. Tests on the ORL Dataset

The ORL database has 10 faces per subject. In our *leave-k-out* test, *k* varies from 1 to 6. In total, each *k*-test has 3 subtests, with different selections of query faces from 10 faces. The final accuracy is the average on all subtests.

Fig.10-a) shows all *leave-k-out* tests, where *k* varies from 1 to 6. We can see that the proposed MK-RDA attained the best accuracy in all five *k* tests.

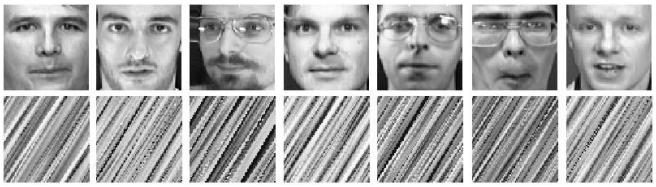
Fig.10-b) lists out the overall accuracy by averaging all *k* tests. Here, we included PCA, LDA, kPCA, kLDA and LPP for comparison because they are typical data-driven face recognition methods based on dimensionality reduction. We can see that our MK-RDA attained the best accuracy over all *k*-tests of around 95.7%. In comparison, LPP attained 91.5%, kLDA 93.3%, LDA 93.6%, and kPCA and PCA attained87.5%.

B. Tests on the PIE Dataset

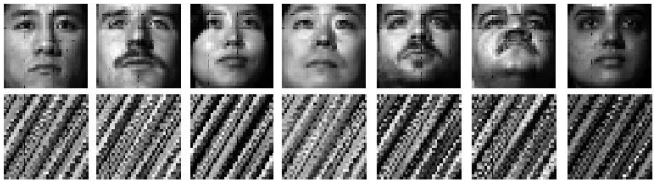
In our experiment, we used 50 faces per subject and in total 3350 faces were used in our *leave-k-out* experiment. In this test scheme, *k* faces from *N* samples per subject are selected as test samples, and the rest are used as training samples.

Fig.11 gives the test results on the PIE dataset. Fig.11-a) shows all *leave-k-out* tests, where *k* varies from 5 to 25. We can see that the proposed MK-RDA attained the best accuracy in all *k* tests. However, when *k* is increased, fewer samples are left for training and as a result the accuracy drops in all methods.

Fig.11-b) lists out the overall accuracy by averaging all *k* tests. PCA and kPCA attained an average accuracy of around 76.0%,



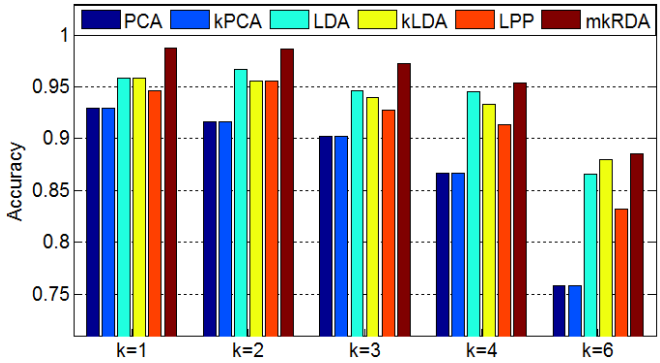
a) Samples in the ORL database and their scrambled images



b) Samples in the PIE database and their scrambled images

c) Wild faces in the PubFig dataset and their scrambled images

Fig.9. Facial images in the ORL, PIE and PUBFIG datasets.

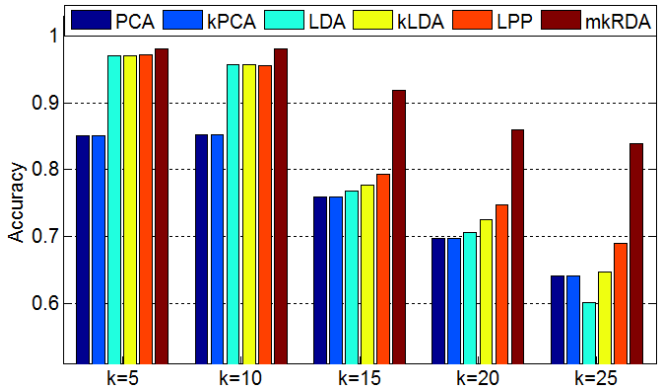


a) *Leave-k-out* tests

Method	PCA	kPCA	LDA	kLDA	LPP	MK-RDA
Accuracy	87.5	87.5	93.6	93.3	91.5	95.7

b) Over all accuracy of all *k tests*

Fig.10. *Leave-k-out* tests on ORL dataset.

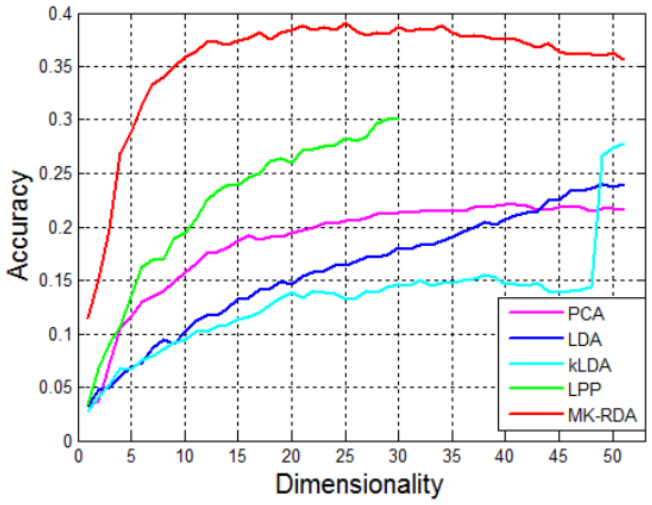


a) *Leave-k-out* tests

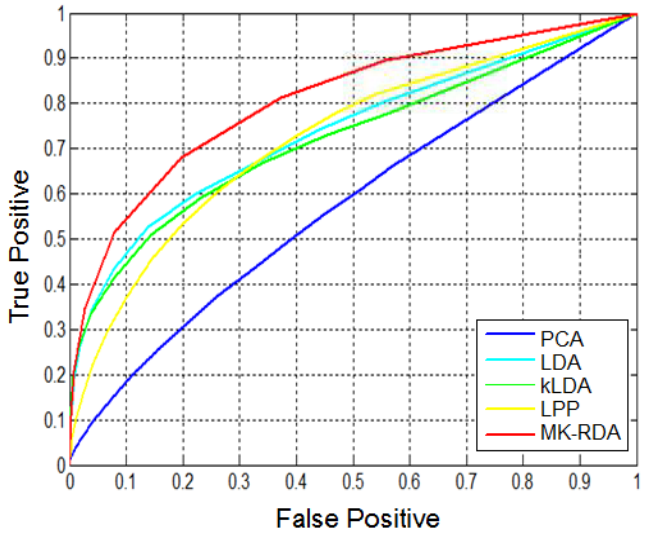
Method	PCA	kPCA	LDA	kLDA	LPP	MK-RDA
Accuracy	76.0	76.0	80.0	81.5	83.1	91.5

b) Over all accuracy of all *k tests*

Fig.11. *Leave-k-out* tests on PIE dataset.



a) Rank-1 accuracy versus dimensionality



b) TP-FP curves

Fig.12. Experimental results on PubFig wild faces.

LDA attained 80.0%, kLDA got a better score of 81.5%, and LPP has the second best accuracy of 83.1%. In comparison, our MK-RDA attained the best accuracy of 91.5, clearly better than the other data-driven approaches.

C. Tests on PUBFIG Dataset

The PUBFIG dataset is designed to compare various algorithms against the human vision system. Its typical benchmark test can have as many as 20,000 pairs of faces for comparison. However, in IoT-targeted scrambled domain, human perception can barely recognize any scrambled faces, making it meaningless to carry out this human-compared test. On the other hand, in the scenarios of IoT applications, usually we have training datasets on the server side, making it most likely as a *leave-k-out* experiment. For this reason, we need to design a new evaluation scheme.

In our experiment, we selected 52 subjects with 60 faces each, and split it randomly into test and training datasets, with each having 30×52=1560 faces. We then test all data-driven methods by comparing each test face against all training faces. In total, we have 1560×1560=2.4 million pairs for testing. Here we use two criteria to evaluate our experiment. One is the rank-1 accuracy versus dimensionality. The other is the true positive (TP) versus the false positive (FP).

Fig.12-a) shows the accuracy versus dimensionality. It is shown that the proposed MK-RDA attained marginally better accuracy-dimensionality performance, consistently corroborating the underlying conjecture that the proposed many kernels method may help capture the intrinsic multiple manifolds lying under the given dataset, as discussed in Section III.

Fig.12-b) gives the results on TP-FP curves. Here, we obtained a likelihood matrix of 1560×1560 elements by comparing each test sample against all training samples. Then we applied varying thresholds on the likelihood matrix, and counted how many pairs classified as positive are false positive and true positive pairs. From the results shown in Fig.12-b), it is observed that PCA has the worst performance, nearly no different from random guessing. From the comparison, we can clearly see that the proposed MK-RDA has clearly better performance on the true/false positive tests, with consistently better true/positive rates (TPR) over other data-driven face recognition methods.

VI. Conclusion

In conclusion, we have identified a new challenge in scrambled face recognition originated from the need for biometric verification in emerging IoT applications, and developed a salience-aware face recognition scheme that can work with chaotic patterns in the scrambled domain. In our method, we conjectured that scrambled facial recognition could generate a new problem in which “many manifolds” need to be discovered for discriminating these chaotic signals, and we proposed a new ensemble approach – Many-Kernel Random Discriminant Analysis (MK-RDA) for scrambled face recognition. We also incorporated a salience-aware strategy into the proposed ensemble method to handle chaotic facial patterns in the scrambled domain, where random selection of features is biased towards semantic components via salience modelling. In our experiments, the proposed MK-RDA was tested rigorously on three standard human face datasets. The experimental results successfully validated that the proposed scheme can effectively handle chaotic signals and drastically improve the recognition accuracy, making our method a promising candidate for emerging IoT applications.

References

[1] Singh, A. ; Karanam, S. ; Kumar, D. "Constructive Learning for Human-Robot Interaction", IEEE Potentials, Vol 32, Issue 4, 2013, Page(s): 13 – 19.

[2] Jayatilake, D. ; Isezaki, T. ; Teramoto, Y. ; Eguchi, K. ; Suzuki, K. "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] McDuff, D. ; Kaliouby, R.E. ; Picard, R.W. "Crowdsourcing Facial Responses to Online Videos", IEEE Trans Affective Computing, Vol 3, Issue 4, 2012 , Page(s): 456 – 468
[4] Fleck, S.; Strasser, W. "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
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] F. Dufaux, T. Ebrahimi, "Scrambling for Video Surveillance with Privacy," Proc. 2006 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 2006, pp.106-110.
[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] 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.
[16] 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.
[17] C. Shan, S. Gong, and P. McOwan. Facial expression recognition based on Local Binary Patterns: A comprehensive study. J. IVC, 27(6):803–816, 2009.
[18] Ping Liu, Shizhong Han, Zibo Meng, Yan Tong, "Facial Expression Recognition via a Boosted Deep Belief Network", CVPR 2014.
[19] 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.
[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, Page(s): 1552 – 1564.
[21] 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.
[22] R. Hsu, A. Jain, "Semantic face matching", IEEE International Conference on Multimedia and Expo, 2002, pp.145.
[23] T. F. Cootes, G. J. Edwards, C. J. Taylor, "Active appearance models", IEEE T rans. Pattern Analysis & Machine Intelligence, Vol. 23, No. 6, 2001, pp.681.
[24] 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.
[25] A.M. Martinez, A.C. Kak, "PCA versus LDA", IEEE Trans. Pattern Analysis & Machine Intelligence, Vol. 23, No. 2, Feb. 2001, pp. 228-233.
[26] M. H. Yang, "Kernel Eigenfaces vs. kernel Fisherface: face recognition using kernel methods", International Conference on Automatic Face and Gesture Recognition, 2002, pp.215.
[27] B. Tenenbaum, V. Silva, J. Langford, "A global geometric framework for nonlinear dimensionality", Science, Vol.290, No.5500, 2000, pp.2319.
[28] S. Roweis, L. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science, Vol.290, No.5500, pp.2323.
[29] M. Belkin, P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering", Advances in Neural Information Processing Systems, 2001, pp.14.
[30] 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.
[31] Xiaofei He, Deng Cai and Partha Niyogi, "Tensor Subspace Analysis", Advances in Neural Information Processing Systems 18 (NIPS), Vancouver, Canada, Dec. 2005.
[32] D. Foster, S. Kakade, T. Zhang, Multi-view dimensionality reduction via canonical correlation analysis, TTI-C Technical Report, TTI-TR-2008-4, 2008.
[33] Hedge et al., Random Projections for Manifold Learning, NIPS 2008
[34] Hou C, Zhang C, Wu Y, et al. Multiple view

semi-supervised dimensionality reduction. Pattern Recognition, 2010, 43(3): 720-730.
[35] 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.
[36] 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.
[37] Lin Y Y, Liu T L, Fuh C S. Multiple kernel learning for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(6): 1147-1160.
[38] A. Zien and C. S. Ong. Multiclass multiple kernel learning.In International Conference on Machine Learning, 2007.
[39] R. Jiang, D. Crookes, "Deep Salience: Visual Salience Modelling via Deep Belief Propagation", AAAI 2014, , Quebec, Canada, July 2014.
[40] T. Sim, S. Baker, M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database," Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition, May 2002.
[41] 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.
[42] Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar,"Attribute and Simile Classifiers for Face Verification," International Conference on Computer Vision (ICCV), 2009.
[43] Gary B. Huang, Vidit Jain, and Erik Learned-Miller. "Unsupervised joint alignment of complex images," International Conference on Computer Vision (ICCV), 2007.
[44] Ping Liu, Shizhong Han, Zibo Meng, Yan Tong, "Facial Expression Recognition via a Boosted Deep Belief Network", CVPR 2014.

Richard Jiang is currently a Lecturer in the department of Computer Science and Digital Technologies, Northumbria University, Newcastle upon Tyne, United Kingdom. He received his PhD in Computer Science from Queen's University Belfast, Belfast, UK, in July 2008.

After his PhD study, he has been working in Brunel Univ., Loughborough Univ., Swansea Univ., Univ. of Bath and Univ. of Sheffield. He joined Northumbria in May 2013. His research interests mainly reside in the fields of Artificial Intelligence, Man-Machine Interaction, Visual Forensics, and Biomedical Image Analysis. His research has been funded by EPSRC, BBSRC, TSB, EU FP, and industry funds, and he has authored and coauthored more than 40 publications.

Somaya Almaadeed received the Ph.D. degree in computer science from the University of Nottingham, Nottingham, U.K., in 2004. She has been a Visiting Fellow with Northumbria University, Newcastle upon Tyne, U.K., since 2012.

She is currently with the Department of Computer Science and Engineering, Qatar University, Doha, Qatar, as an Assistant Professor, where she is involved in research on biometrics, writer identification, image processing, and document analysis. She has been awarded a number of grants, and has authored around 40 papers. Dr. Almaadeed is a member of different international computer science committees. Her team received the Best Performance Award in the 2011 International Conference on Document Analysis and Recognition's Signature Verification Competition and Music Scores Competition.

Ahmed Bouridane received the "Ingenieur d'État" degree in electronics from Ecole Nationale Polytechnque of Algiers (ENPA), Algeria, in 1982, the M.Phil. degree in electrical engineering (VLSI design for signal processing) from the University of Newcastle-Upon-Tyne, U.K., in 1988, and the Ph.D. degree in electrical engineering (computer vision) from the University of Nottingham, U.K., in 1992.

From 1992 to 1994, he worked as a Research Developer in telesurveillance and access control applications. In 1994, he joined Queen's University Belfast, Belfast, U.K., initially as Lecturer in computer architecture and image processing and then as a Reader in computer science. He became a Professor in Image Engineering and Security at Northumbria University at Newcastle (U.K.) in 2009. His research interests are in imaging for forensics and security, biometrics, homeland security, image/video watermarking and cryptography. He has authored and co-authored more than 200 publications.

Danny Crookes received the B.Sc. degree in Mathematics and Computer Science in 1977, and the Ph.D. degree in Computer Science in 1980, both from Queen's University Belfast.

He became Professor of Computer Engineering in 1993 at Queen's University Belfast, Belfast, U.K., and was Head of Computer Science from 1993-2002. He is currently Director of Research for Speech, Image and Vision Systems at the Institute of Electronics, Communications and Information Technology, Queen's University Belfast. His current research interests include the use of novel architectures (GPUs and FPGAs) for high performance speech and image processing. Professor Crookes is currently involved in projects in automatic shoeprint recognition, speech separation and enhancement, and medical imaging. Professor Crookes has some 220 scientific papers in journals and international conferences.

M. Emre Celebi received the B.Sc. degree in computer engineering from the Middle East Technical University, Ankara, Turkey, in 2002 and the M.Sc. and Ph.D. degrees in computer science and engineering from The University of Texas at Arlington, Arlington, TX, USA, in 2003 and 2006, respectively.

He is currently a Professor with the Department of Computer Science, University of Central Arkansas, Conway, AR, USA. He has pursued research in the field of image processing and analysis. He has published more than 130 articles in journals and conference proceedings. His recent research is funded by grants from the National Science Foundation.