

Individual Stable Space: An Approach to Face Recognition Under Uncontrolled Conditions

Xin Geng, Zhi-Hua Zhou, *Senior Member, IEEE*, and Kate Smith-Miles, *Senior Member, IEEE*

Abstract—There usually exist many kinds of variations in face images taken under uncontrolled conditions, such as changes of pose, illumination, expression, etc. Most previous works on face recognition (FR) focus on particular variations and usually assume the absence of others. Instead of such a “divide and conquer” strategy, this paper attempts to directly address *face recognition under uncontrolled conditions*. The key is the individual stable space (ISS), which only expresses personal characteristics. A neural network named ISNN is proposed to map a raw face image into the ISS. After that, three ISS-based algorithms are designed for FR under uncontrolled conditions. There are no restrictions for the images fed into these algorithms. Moreover, unlike many other FR techniques, they do not require any extra training information, such as the view angle. These advantages make them practical to implement under uncontrolled conditions. The proposed algorithms are tested on three large face databases with vast variations and achieve superior performance compared with other 12 existing FR techniques.

Index Terms—Face recognition (FR), individual stable space (ISS), machine learning, neural networks, pattern recognition.

I. INTRODUCTION

DURING recent years, face recognition (FR) techniques have been making steps toward practicality. A number of FR systems achieved good performance in the latest report of Face Recognition Vendor Test (FRVT 2006) [4], yet many issues still remain to be addressed. Among those issues, perhaps the most prominent one is that most systems require the face images fed to them to satisfy certain “rules,” such as within a particular range of view angle, under homogeneous illumination, or without any occlusions. Such systems are called face recognition under controlled conditions (FRC). In fact, these rules greatly restrict the commercialization of the FR techniques because most real applications cannot satisfy such strict rules. What the real world needs are systems that can recognize any face images recognizable by human beings. Such systems are called face recognition under uncontrolled conditions (FRU). A formal definition is given as follows.

Manuscript received December 12, 2006; revised August 17, 2007 and December 21, 2007; accepted January 10, 2008. First published May 16, 2008; last published August 6, 2008 (projected). Part of this work was done when X. Geng was at the LAMDA Group, Nanjing University, Nanjing, China. The work of Z.-H. Zhou was supported by the National Science Foundation of China under Grants 60325207, 60635030, and 60721002 and the National High Technology Research and Development Program of China under Grant 2007AA01Z169.

X. Geng and K. Smith-Miles are with the School of Engineering and Information Technology, Deakin University, Melbourne, VIC 3125, Australia (e-mail: xge@deakin.edu.au; katesm@deakin.edu.au).

Z.-H. Zhou is with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: zhouzh@nju.edu.cn).

Digital Object Identifier 10.1109/TNN.2008.2000275

Definition 1 (FRU): Given still or video face images with arbitrary variation but enough information to be recognized by human beings, identify or verify one or more persons, whose features have been stored in a database, without using any information other than the images themselves.

While a wide range of applications are covered by this definition, this paper focuses on scenarios including identification, still image, and one person per image. Further work could extend to other cases like video sequences, verification, and multiple persons per image.

The development of FR techniques corresponds to a march from strictly controlled conditions to more and more uncontrolled conditions. Most early algorithms [2], [3], [21], [33] can handle expression variation well but suffer in the presence of other variations. Subsequently, many methods [9], [10], [13], [14], [16]–[20], [26], [29] were proposed to tackle view angle and illumination variations. Recently, a few works have been emerging to remove occlusion [34] and simulate aging effects [7]. Although the treatable variations are more and more complex, most of these ingenious methods yet have to assume the absence of other possible variations. The methodology adopted by the existing work appears to be “divide and conquer,” i.e., gradually reduce the restrictions through tackling possible variations one by one. However, in practice, a number of variations are often complicatedly interlaced and cannot actually be divided. Thus, the combination of several algorithms each of which handles a particular variation well will not necessarily result in a robust system against all variations. Instead of “divide and conquer,” this paper presents one of the first attempts directly targeting to FRU. Since the variations under uncontrolled conditions might be too complex to be well handled, we avoid explicitly modeling them. Instead, we focus on the information useful for recognition and try to filter out all other information. This is achieved by projecting the face images into a subspace called individual stable space (ISS), where only the useful information is reserved.

The rest of this paper is organized as follows. In Section II, the four different kinds of information in the face images are analyzed and the concept of ISS is introduced. In Section III, a neural network named individual stable neural network (ISNN) is proposed to map a face image into the ISS. In Section IV, three ISS-based algorithms for FRU are proposed. In Section V, the experimental results are reported. Finally, in Section VI, conclusions are drawn and several issues for future work are indicated.

II. EXTRACTION OF PERSONAL CHARACTERISTICS

To find a solution for FRU, we start from analyzing the components of the face images obtained under uncontrolled con-

ditions. The information conveyed by an arbitrary face image¹ might be categorized into four kinds:

- 1) *personal characteristics* (denoted by Φ_{personal}), i.e., the characteristics that make one person look different from others;
- 2) *common facial characteristics* (denoted by Φ_{facial}), i.e., the characteristics shared by all faces;
- 3) *face status* (denoted by Φ_{status}), i.e., any changes a particular face may undergo, such as expressions, aging effects, glasses, scars, etc.;
- 4) *imaging configuration* (denoted by Φ_{imaging}), i.e., the conditions under which the face is imaged, such as illumination, view angle, imaging device, etc.

Information here refers to the semantic meanings, rather than the data in the image space or any subspace. Each kind of information has a certain number (up to infinite) of states (continuous or discrete values). Consequently, the information conveyed by the image is the combination of the states of Φ_{personal} , Φ_{facial} , Φ_{status} , and Φ_{imaging} , and each kind of information can be extracted from the image. Although face images are believed to distribute on nonlinear manifolds [28], linear methods have been successfully used to extract various facial information [33], [3], [9]. Fig. 1 shows an example of using linear subspaces to extract Φ_{personal} , Φ_{facial} , Φ_{status} , and Φ_{imaging} , respectively. The data used in this example is a subset of the CMU PIE database [30]. As illustrated by Fig. 1(a), there are three states (1, 2, 3) for each of Φ_{personal} (identity), Φ_{status} (expression), and Φ_{imaging} (illumination) in the data set. Note that in order to show what each variation might look like, each line of Fig. 1(a) only shows one variation, but in fact, the three kinds of variations might happen simultaneously in one face image, just as in real applications. The face images are first projected into the linear subspace spanned by the “eigenfaces” [33]. Fig. 1(b) shows the first ten eigenfaces. As can be seen, all the eigenfaces look like faces. Thus, the subspace spanned by them is called “face space” [33]. As the name implies, projection into the face space can be viewed as extraction of Φ_{facial} . This is also evidenced by the numerous applications of coding/decoding face images with eigenfaces. The projections on the first two eigenfaces are plotted with respect to Φ_{personal} , Φ_{status} , and Φ_{imaging} in Fig. 1(c), (d), and (e), respectively. In each case, the three states cannot be well separated. Then, linear discriminant analysis (LDA) is used to find a linear subspace for each kind of information that can best separate the three states. The results are shown in Fig. 1 (f), (g), and (h), respectively. As can be seen, in each LDA subspace, the three states can be clearly separated, which means that different kinds of information contained in the images are decomposed by the linear subspaces. Of course, this is a simplified illustrative example in sense that there is only one variation in each of Φ_{personal} , Φ_{status} , and Φ_{imaging} , while in real cases, the variations are much more complex. Among the four kinds of information, Φ_{personal} is the only useful one for recognizing identity. Thus, the key step of any FR method is the extraction of Φ_{personal} , explicitly or implicitly. The problem is that, in practice, the three other kinds of information are usually complicatedly interlaced and could not be clearly separated as

shown in Fig. 1. In most cases, even the states of the information (class labels used in LDA) are not available.

Consider a set of face images obtained under uncontrolled conditions. The four kinds of information can be divided into two groups: *unstable* and *stable*, which are defined as the information with major and minor variance in the set, respectively. Note that the stable information is not defined as with zero variance because of noise and the imperfection of algorithms. The membership of each group depends on the arrangement of the image set. For example, if the face images come from different persons, then Φ_{personal} , Φ_{status} , and Φ_{imaging} are unstable, and only Φ_{facial} is stable. If the face images all belong to the same person, then Φ_{status} and Φ_{imaging} are unstable while Φ_{personal} and Φ_{facial} are stable.

Most existing FR approaches focus on the unstable information in a set of face images from different persons. In this case, Φ_{facial} is out of consideration first. The goal is set as distinguishing the variation of Φ_{personal} from that of Φ_{status} and Φ_{imaging} . Of course, the most straightforward way is to artificially freeze the states of Φ_{status} and Φ_{imaging} . However, that will result in a trivial system which can only “recognize” exactly the same face images in the database. Nontrivial work naturally starts from the relatively easier case when the variations of Φ_{status} and Φ_{imaging} are partially restricted, i.e., the case of FRC. Directly targeting FRU is challenging because the numerous variations of Φ_{status} and Φ_{imaging} might be too complex to be efficiently modeled. However, what should be kept in mind is that the goal is Φ_{personal} . Thus, if Φ_{status} and Φ_{imaging} cannot be directly modeled, why not try to “filter out” these information? Here, “filter out” or “remove” a particular information Φ_x means making Φ_x no longer *variable* in some subspace Γ so that Φ_x will no longer affect the classification in Γ . Special attention should be paid to distinguish “*variable* in a space” and “*unstable* in a data set.” For example, Φ_{facial} is variable in the image space because nonface images are also expressible, but it is stable in a face image set because all images in the set are faces.

Through properly assembling the data set, the problem of filtering out a particular kind of information can be transformed into an easier one: removing the stable/unstable information in the image set. Since Φ_{facial} is always stable in a set of face images, it can be removed first. Formally, suppose the information of a face image I in a multipersonal face image set A is denoted by $\Phi_{I \in A}(I)$, then

$$\Phi_{I \in A}(I) = \underbrace{\Phi_{\text{status}}(I)} + \underbrace{\Phi_{\text{imaging}}(I)} + \underbrace{\Phi_{\text{personal}}(I)} + \underbrace{\Phi_{\text{facial}}(I)} \quad (1)$$

where the double-underlined terms are the unstable information, and those single-underlined ones are the stable information. Note that “+” in the equation is not the arithmetic operation “add,” but means “combination” of the states of different kinds of information. Suppose a subspace Γ_u can be constructed to filter out the stable information and meanwhile preserve the unstable information in the image set (Γ_u is an ideal subspace for this purpose, while the technique described later

¹Here, the face image refers to the normalized face image, i.e., only the face region is contained in the image.

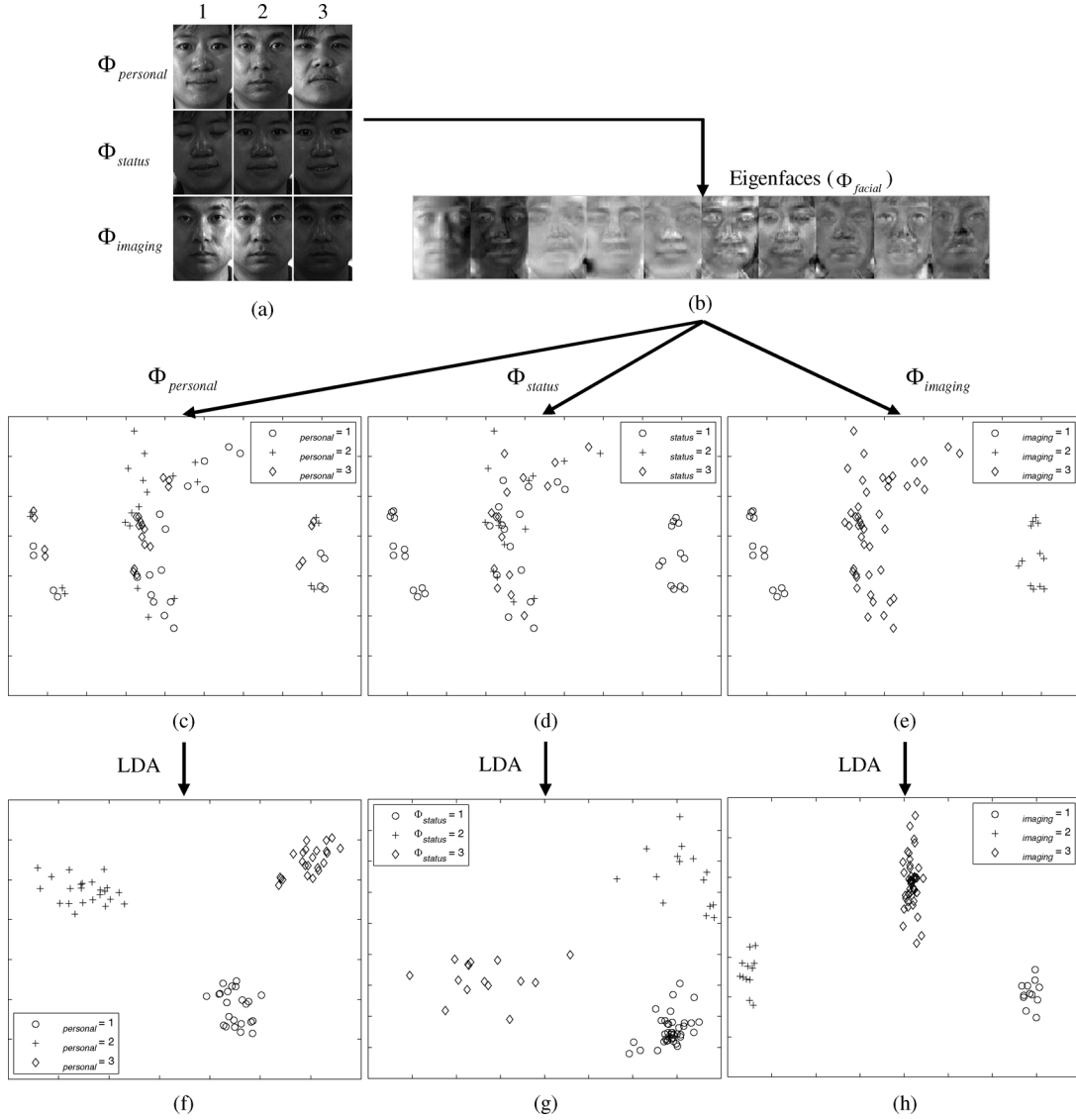


Fig. 1. Extraction of Φ_{personal} , Φ_{facial} , Φ_{status} , and Φ_{imaging} by linear subspaces.

in Section III-A is an approximation to it), then the information contained in the projection $\Gamma_u(I)$ will be

$$\begin{aligned} \Phi_{I \in A}(\Gamma_u(I)) &= \underbrace{\Phi_{\text{status}}(\Gamma_u(I)) + \Phi_{\text{imaging}}(\Gamma_u(I))}_{+ \Phi_{\text{personal}}(\Gamma_u(I))} \\ &= \underbrace{\Phi_{\text{status}}(I)} + \underbrace{\Phi_{\text{imaging}}(I)} + \underbrace{\Phi_{\text{personal}}(I)}. \end{aligned} \quad (2)$$

Since only the stable information is removed and all unstable information is preserved, the states of Φ_{status} , Φ_{imaging} , and Φ_{personal} in $\Gamma_u(I)$ are equal to those in I . Suppose A is subsequently divided into N (the number of persons in A) subsets A^i , $i = 1 \dots N$, each of which is a single-personal set, then the information of the projections in A^i will be

$$\Phi_{I \in A^i}(\Gamma_u(I)) = \underbrace{\Phi_{\text{status}}(I)} + \underbrace{\Phi_{\text{imaging}}(I)} + \underbrace{\Phi_{\text{personal}}(I)}. \quad (3)$$

Now, only Φ_{personal} is the stable information. If a second subspace Γ_i that filters out the unstable information and meanwhile preserves the stable information (Γ_i is also an ideal subspace for

this purpose, while the technique described in Section III-B is an approximation to it) is constructed on the subset of person i , then the information contained in the projections $\Gamma_i(\Gamma_u(I))$ will be only Φ_{personal}

$$\Phi_{I \in A^i}(\Gamma_i(\Gamma_u(I))) = \underbrace{\Phi_{\text{personal}}(\Gamma_i(\Gamma_u(I)))}_{= \Phi_{\text{personal}}(I)}. \quad (4)$$

Through two subspaces and one data splitting, Φ_{personal} is finally extracted from the face images. The subspace Γ_i is called ISS of the person i for two reasons. First, all face images of that person are expected to be stable in Γ_i because all unstable information has been filtered out. Second, since the commonly stable information Φ_{facial} has also been removed and only the personally stable information Φ_{personal} is left, if the face images from other persons are projected into Γ_i , the projections are expected to be unstable. Thus, ISS can be used to design an FRU system.

Different from the existing FR techniques, ISS is the first FR method without any restrictions or preassumptions about the face images. It actually represents a family of algorithms. The

implementation of ISS depends on the algorithms used to construct the subspaces Γ_u and Γ_i . Section III will describe one of the many possible solutions, which maps a face image into ISS by the ISNN.

III. INDIVIDUAL STABLE SPACE

The extraction of personal characteristics described in Section II involves two kinds of subspace. The first is the subspace that can filter out the information stable in the training set [the projection from (1) to (2)]. The second is the subspace that can filter out the information unstable in the training set [the projection from (3) to (4)]. Obviously, these two kinds of subspace have opposite properties. In this section, a method is proposed to map face images into the ISS based on a pair of neural networks with opposite learning rules, namely, the stochastic gradient ascent (SGA) network [25] and the anti-Hebbian version of SGA (ASGA) network [36].

A. SGA Network

The training of the SGA network is an unsupervised learning procedure, i.e., no personal ID is needed at this stage. The SGA network [25] was proposed to recursively learn the principal components of the input data stream. The network structure is shown in Fig. 2. It has p parallel neurons, each of which corresponds to one principal component. For each neuron, the n -dimensional input vector passes from the left-hand side to the right-hand side after being subtracted by the product of the connection weights and the internal feedback from the output end. The output signal from the right-hand side is used as the input vector to the next neuron. The learning rule of the SGA network is given by

$$y_j(t) = \mathbf{w}_j^T(t-1)\mathbf{x}(t) \quad (5)$$

$$\Delta \mathbf{w}_j(t-1) = \alpha_1 y_j(t) \left[\mathbf{x}(t) - y_j(t) \mathbf{w}_j(t-1) - 2 \sum_{i < j} y_i(t) \mathbf{w}_i(t-1) \right] \quad (6)$$

$$\mathbf{w}_j(t) = \mathbf{w}_j(t-1) + \Delta \mathbf{w}_j(t-1) \quad (7)$$

where $0 < \alpha_1 < 1$ is the learning rate. As proved by Oja [24], for $t \rightarrow \infty$, the vectors $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$ will converge to the principal components $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p$ of the input data stream.

Principal component analysis (PCA) [11] is not new in the literature of FR. Much previous work on FRC, including the famous Eigenface method [33] and its variants [22], [3], uses PCA to extract features from the face images. The initial idea of using PCA in FR is an information theory approach of coding and decoding face images. PCA can effectively extract the main variations in a collection of face images and thus can code a face image as a vector of linear weights of the most variable “eigenfaces.” In strictly controlled FR, such as the case that only moderate expression variation exists in the face images, the Eigenface method can work very well because in such cases, the main variation in the image set is just the variation of Φ_{personal} , thus the coding of main variation is approximately the coding of Φ_{personal} . However, under uncontrolled conditions, the main variation in the image set might

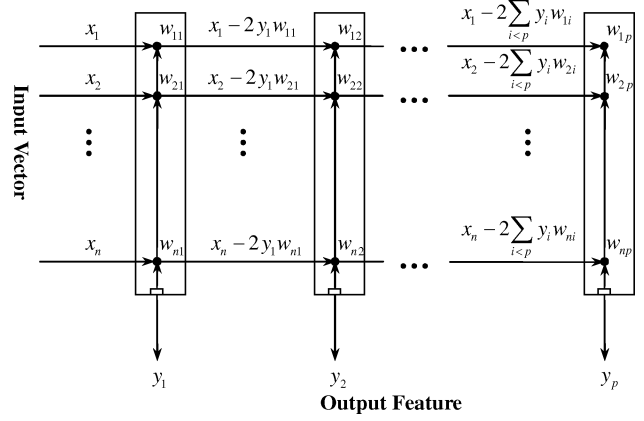


Fig. 2. SGA network. The input to the network is an n -dimensional vector $\mathbf{x} = [x_1, x_2, \dots, x_n]$, and the output is a p -dimensional ($p < n$) feature $\mathbf{y} = [y_1, y_2, \dots, y_p]$.

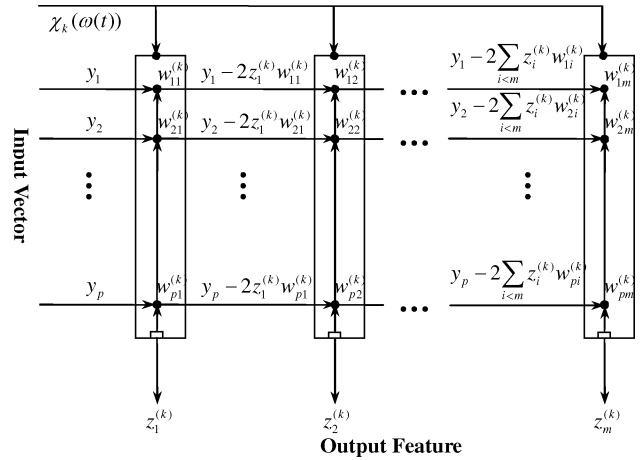


Fig. 3. Supervised ASGA network for person k . The input to the network is a p -dimensional vector $\mathbf{y} = [y_1, y_2, \dots, y_p]$ and the supervisory signal $\chi_k(\omega(t))$. The output is an m -dimensional ($m < p$) feature $\mathbf{z}^{(k)} = [z_1^{(k)}, z_2^{(k)}, \dots, z_m^{(k)}]$.

include the variations of Φ_{personal} , Φ_{status} , and Φ_{imaging} . In fact, some misleading variations are even more significant than the variation of Φ_{personal} . For instance, it has been shown that the changes caused by illumination could be even larger than the differences between individuals [1]. Thus, it is hard for the Eigenface-based methods to achieve good performance under uncontrolled conditions.

Instead of face image coding, the utility of the SGA network in our approach relies on PCA's ability to filter out the stable component (insignificant variance) of the training set, i.e., the common facial characteristics Φ_{facial} . This can also be evidenced by the fact that the first several eigenfaces look like faces. An arbitrary linear combination of these face-like eigenfaces should also be face-like images (share some common facial characteristics). Thus, any point in the subspace spanned by these eigenfaces corresponds to a face-like image. Consequently, Φ_{facial} is no longer variable in such subspace.

Conventionally, the principal components are calculated by eigendecomposing the covariance matrix of the image set. Suppose the face images are m_1 by m_2 , then the dimensionality of

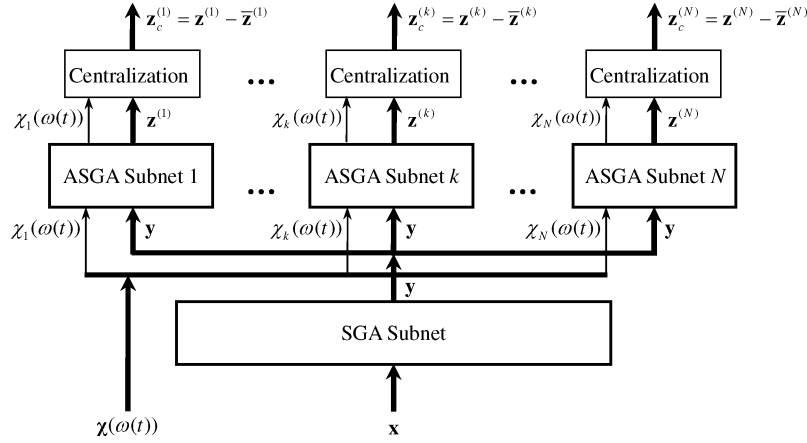


Fig. 4. Architecture of ISNN. The thick lines represent vector signals, and the thin lines represent scalar signals.

the image vector will be $n = m_1 \times m_2$, and the covariance matrix Σ will be $n \times n$. Directly determining the n -dimensional eigenvectors and eigenvalues is an intractable task for typical image size. Normally, in previous work, this problem was solved by singular value decomposition (SVD), i.e., first calculating an $M \times M$ matrix problem, where M is the number of face images in the training set, and then taking linear combination of the resulting eigenvectors [33]. Under controlled conditions, the possible variations in the face images are limited, so a small number (from tens to hundreds) of face images are enough to compose a representative training set. Thus, usually $M \ll n$, and the eigendecomposition of an $M \times M$ matrix is computationally feasible on most machines. However, under uncontrolled conditions, the vast possible variations consequentially require a large training set, so M is usually comparative to n , if not larger. As a result, the eigendecomposition of the $M \times M$ matrix is almost as hard as that of Σ . Fortunately, with the learning rule of the SGA network, the principal components can be recursively approximated, no matter how large M is (in fact, the larger, the better).

B. Supervised ASGA Network

The ASGA network [36] uses the opposite learning rule of the SGA network. It is an anti-Hebbian version of SGA. Since in the second stage of the extraction of Φ_{personal} , the projections in the face space need to be divided into subsets according to personal IDs [from (2) to (3)], a supervisory signal χ_k should be integrated into the learning rule of the ASGA network trained on the subset with the ID k . Suppose there are N different persons in the training set, the personal ID of a particular projection $y(t)$ is $\omega(t)$, where $\omega(t)$ can be any of the labels from 1 to N , then $\chi_k(\omega(t))$ is defined by

$$\chi_k(\omega(t)) = \begin{cases} 1, & \text{when } \omega(t) = k \\ 0, & \text{when } \omega(t) \neq k. \end{cases} \quad (8)$$

The structure of the supervised ASGA network is shown in Fig. 3. The output y of the SGA network becomes the input vector of the ASGA network. The structure in Fig. 3 is similar

to that in Fig. 2, except for the additional supervisory signal. The learning rule of the supervised ASGA network is given by

$$z_j^{(k)}(t) = w_j^{(k)T}(t-1)y(t) \quad (9)$$

$$\Delta w_j^{(k)}(t-1) = -\alpha_2 \chi_k(\omega(t)) z_j^{(k)}(t) \times \left[y(t) - \frac{z_j^{(k)}(t) w_j^{(k)}(t-1)}{\|w_j^{(k)}(t-1)\|^2} - 2 \sum_{i < j} z_i^{(k)}(t) w_j^{(k)}(t-1) \right] \quad (10)$$

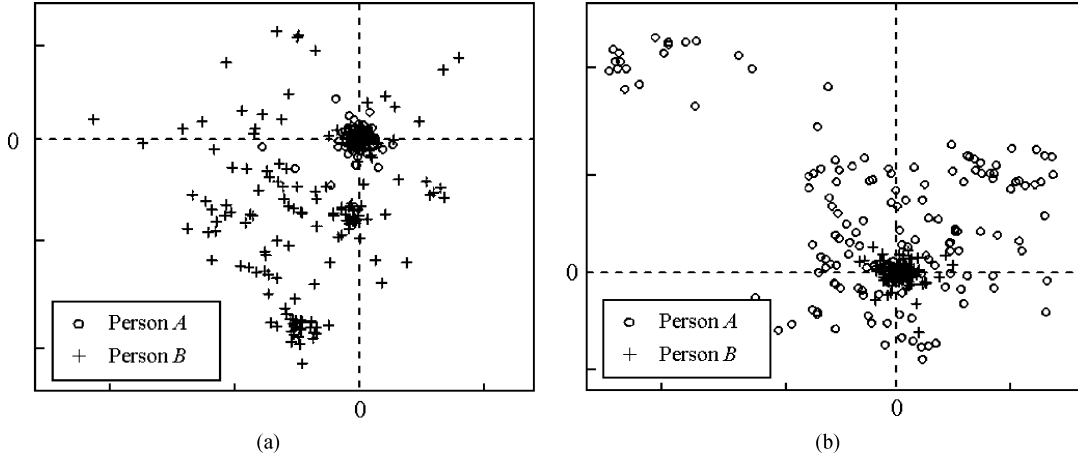
$$w_j^{(k)}(t) = w_j^{(k)}(t-1) + \Delta w_j^{(k)}(t-1) \quad (11)$$

where $0 < \alpha_2 < 1$ is the learning rate. Equation (10) is different from (6) at three points. The first is the additional supervisory signal $\chi_k(\omega(t))$. If the input projection $y(t)$ does not belong to person k , then no update will take place. The second is that the learning in (6) is changed into the opposite direction by adding a minus sign in front. The third is that an explicit normalization term $\|w_j^{(k)}(t-1)\|^2$ is inserted to guarantee that the magnitude of $w_j^{(k)}(t)$ is of unit length. It was proved [37] that for $t \rightarrow \infty$, $w_j^{(k)}(t)$ will converge to the least variable components of the input data. Such components are called minor components [36].

As principal components filtering out the stable information, minor components filter out the unstable information in the data set. In (3), both Φ_{status} and Φ_{imaging} are unstable, which are responsible for the main variance in the data set. Thus, the subspace spanned by the minor components only reserves Φ_{personal} . Similar to principal components, the minor components can also be calculated through eigendecomposition, but ASGA is more efficient. Moreover, since we have used SGA network in the previous step, using ASGA in this step is helpful to result in an integrated framework of solution.

C. Individual Stable Neural Network

The architecture of the ISNN is shown in Fig. 4. The first subnet of the ISNN is an SGA network. The raw face image x is first input into this SGA subnet to get its projection y in the


 Fig. 5. Projections in ISS. (a) Projections in the A -stable-space. (b) Projections in the B -stable-space.

face space. Then, \mathbf{y} is input into the N (the number of different individuals) ASGA subnets together with the supervisory signal $\chi(\omega(t)) = [\chi_1(\omega(t)), \chi_2(\omega(t)), \dots, \chi_N(\omega(t))]$. The output of the ASGA subnet k will be the projection $\mathbf{z}^{(k)}$ in the ISS of person k . Then, each ISS is centralized to the mean of the projections of the training set $\bar{\mathbf{z}}^{(k)}$. Such centralized ISS is called k -stable-space. $\bar{\mathbf{z}}^{(k)}$ is calculated after the convergence of the SGA and ASGA subnets by

$$\bar{\mathbf{z}}^{(k)} = \frac{\sum_t \chi_k(\omega(t)) \mathbf{z}^{(k)}(t)}{\sum_t \chi_k(\omega(t))}. \quad (12)$$

Fig. 5 gives a real example. Persons A and B are two individuals randomly selected from the CMU PIE database [30]. The face images of each person are randomly divided into two parts, one (training set) is used to train the ISNN, and the other (test set) is projected into the ISS through the trained ISNN. Fig. 5(a) shows the projections in the A -stable-space. It can be seen that the projections of person A (circles) converge around the origin while those of person B (plus signs) scatter in the space. Fig. 5(b) shows what happens in the B -stable-space. This time, the projections of person A scatter while those of person B converge around the origin. The two projected classes are not completely disjoint because the ISS is only 2-D for visualization purpose. In fact, such a low-dimensional subspace is not sufficient for the task of FRU. In the experiments described later, the ISS usually has tens of dimensions. Table I shows the average distance from the origin of the ISS to the projections of the same (\bar{d}_{same}) or different ($\bar{d}_{\text{different}}$) persons on the three data sets (PIE, FERET, and FRU) used in the experiment section. The dimensionality of the ISS in the experiments is set to 20. As can be seen, in all cases, \bar{d}_{same} is much smaller than $\bar{d}_{\text{different}}$, which indicates the key property of ISS that the images from the same person appear stable while those from different persons appear unstable. Thus, ISNN is an effective tool to map face images into ISS.

 TABLE I
AVERAGE DISTANCE FROM THE ISS ORIGIN TO THE PROJECTIONS OF THE SAME/DIFFERENT PERSONS

Avg. Distance	PIE	FERET	FRU
\bar{d}_{same}	0.0105	0.0360	0.0103
$\bar{d}_{\text{different}}$	0.1139	0.1560	0.1255
$\bar{d}_{\text{same}}/\bar{d}_{\text{different}}$	9.22%	23.08%	8.24%

IV. ISS-BASED FACE RECOGNITION

A. Algorithms

The ISNN described in Section III-C can map a face image \mathbf{x} into the centralized ISS of each individual. The output of the ISNN is N projections $\mathbf{z}_c^{(1)}, \mathbf{z}_c^{(2)}, \dots, \mathbf{z}_c^{(N)}$ in the ISS of N different individuals, respectively. As mentioned in Section II, the face images of person k will be stable only in the k -stable-space. Contrarily, the most stable projection will indicate the ID of the face image. If a function $S(\mathbf{z})$ is defined to measure the stability of the projection \mathbf{z} (larger value of $S(\mathbf{z})$ means more stable), then the personal ID of \mathbf{x} is given by

$$r = \arg \max_{j=1}^N [S(\mathbf{z}_c^{(j)})]. \quad (13)$$

Different definitions of $S(\mathbf{z})$ will result in different recognition algorithms. In this paper, three definitions of $S(\mathbf{z})$ are studied

$$S_1(\mathbf{z}) = -\|\mathbf{z}\| \quad (14)$$

$$S_2(\mathbf{z}) = -\sqrt{\mathbf{z}' \mathbf{C}^{-1} \mathbf{z}} \quad (15)$$

$$S_3(\mathbf{z}) = -\frac{1}{M'} \sum_{i=1}^{M'} \|\mathbf{z} - \mathbf{z}_i\|. \quad (16)$$

S_1 is the negative Euclidean distance from \mathbf{z} to the mean of the training set (the origin of the centralized ISS). S_2 is the negative Mahalanobis distance from \mathbf{z} to the mean of the training

set, where the covariance matrix \mathbf{C} can be estimated from the training set. S_3 is the negative average Euclidean distance from \mathbf{z} to the projections of the training images \mathbf{z}_i , where M' is the number of images from a particular person. The algorithms corresponding to the three definitions are denoted by ISS1, ISS2, and ISS3, respectively. The whole solution is summarized in Algorithm 1.

Algorithm 1: ISS-Based Face Recognition

// Training Process:

Data: Training images with IDs

Result: ISNN

- Train the SGA subnet using all training images [(5)–(7)];
- Train an ASGA subnet for each individual using the output of the SGA subnet and the corresponding IDs [(9)–(11)];
- Centralize the ISS for each individual [(12)];

// Test Process:

Input: Test image \mathbf{x} , trained ISNN

Output: ID of the test image r

- \mathbf{x} goes through the ISNN, getting $\mathbf{z}_c^{(j)}$, $j = 1, \dots, N$;
 - Calculate the stability of $\mathbf{z}_c^{(j)}$, $S(\mathbf{z}_c^{(j)})$ [(14)–(16)];
 - $r \leftarrow \arg \max_{j=1}^N [S(\mathbf{z}_c^{(j)})]$.
-

It might be argued that ISNN requires training of one network for each person, thus it is inefficient for large databases. For this, we will note that ISNN adopts the so-called one-class-one-network (OCON) structure, which has certain advantages over the all-class-one-network (ACON) structure, such as less hidden units, faster convergence, and better generalization [38]. Moreover, such architecture can easily benefit from distributed computing. It is also suitable for incremental learning. If a new individual needs to be enrolled into the system, the existing ISNN does not need to be retrained. This is because the SGA subnet describes the general concept of face, which has already been well trained when the ISNN is first set up. The only thing to do when enrolling a new individual is to add a new ASGA subnet and train it using the new face images. Moreover, since uncontrolled conditions require a relatively large training set, the efficiency of computing principal/minor components could be further improved, such as using adaptive approaches [12]. Nevertheless, ISNN is proposed as one of the many possible implementations of ISS. The main purpose is to illustrate the effectiveness of ISS in a straightforward way. Further consideration on efficiency could be carried out within the framework of ISS in the future. Finally, all the efficiency issues only exist in the training process. Once the ISNN is built, the recognition of a new face image (going through the ISNN) will be very fast.

B. Relationship to Existing Work

As mentioned before, more or less, existing FR methods usually put some restrictions on the face images (most of them are designed for one or several variations, but not for all possible variations). Thus, under uncontrolled conditions, their performance appears unstable. On the other hand, the ISS-based method concentrates on Φ_{personal} and tries to filter out all other information without any preassumptions of the face images. Thus, it is more suitable for uncontrolled conditions. Existing FR techniques dealing with large facial variations roughly follow four streams: 1) algorithms based on local facial features [6], [26]; 2) algorithms based on discriminant analysis [3], [39]; 3) algorithms based on nonlinear subspace/manifold [20], [5]; and 4) algorithms based on multiple classifiers/subspaces [15], [21], [29].

The first category is based on the assumption that some local facial features might be invariant when certain kinds of Φ_{status} or Φ_{imaging} change. One typical method in this category is Eigenfeature [26], which constructs eigenspaces for eyes, nose, and mouth, respectively. The combination of Eigenfeature and Eigenface (Eigenfeature+Eigenface) can be viewed as a layered representation of a face, where a coarse description of the whole face is augmented by additional details in terms of salient facial features, and therefore, better performance was observed [26]. The main problem of such kinds of methods is that the relationship between local features and Φ_{status} or Φ_{imaging} is very complex. Thus, key issues, such as how many features, which features, and the extension and position of each feature region, are usually empirically determined. Under uncontrolled conditions, this problem becomes more serious: intuitively, no local features can keep invariant with uncontrolled changes of Φ_{status} and Φ_{imaging} .

Fisherface [3] is a representative of the second category. It tries to find a global feature space that maximizes the ratio of the extrapersonal difference and the intrapersonal difference by applying Fisher's linear discriminant (FLD). Fisherface was proposed to tackle illumination and expression variations and it does work well supposing no other variations are present. However, under uncontrolled conditions, one global linear subspace might not be powerful enough to clearly separate different persons. There might be two possible solutions. One is nonlinear global subspace (the third category), and the other is multiple subspaces (the fourth category).

It has become a recent trend to apply nonlinear subspace [20] or manifold [5] techniques to robust FR. Face images are believed to be distributed on nonlinear manifolds in the observation space [28]. Thus, better performance could be expected by using nonlinear techniques, such as kernel PCA (KPCA) [20]. However, most existing nonlinear subspace/manifold methods have one or more of the following problems: 1) vulnerable to overfitting; 2) computationally expensive; 3) many of them cannot be directly applied to classification tasks (such as the famous Isomap [32]); 4) under uncontrolled conditions, variations of Φ_{status} and Φ_{imaging} might be more significant than that of Φ_{personal} , but most nonlinear subspace/manifold methods cannot distinguish them.

Since ISS constructs one subspace for each individual, it belongs to the fourth category. Typical algorithms in this category include probabilistic decision-based neural network (PDBNN)

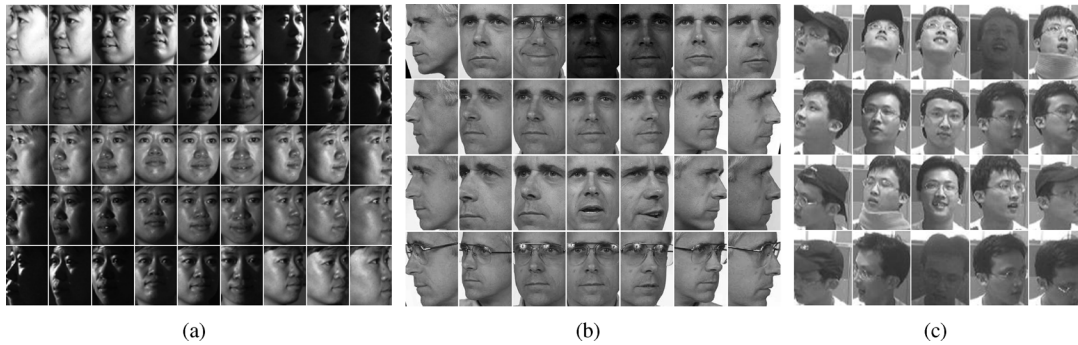


Fig. 6. Typical face images from (a) the CMU PIE database, (b) the FERET database, and (c) the FRU database.

[15], face specific subspace (FSS) [29], and Bayesian FR [21]. Both PDBNN and ISNN adopt the OCON structure. However, the effectiveness of PDBNN relies on the ability of the mixture of Gaussians to approximate any data distribution, while the effectiveness of ISNN relies on the extraction of the only useful information for recognition Φ_{personal} . FSS also utilizes the idea of personalized subspace. This method directly trains an eigenspace on the face images of each person, and then uses the reconstruction error in the eigenspace as a similarity measure. Since the space spanned by the principal components is orthogonal to that spanned by the minor components, the reconstruction error in the former space equals the Euclidean distance from the projection to the origin in the latter space. Thus, FSS is actually similar to the second step of ISS1. There are two main advantages of the ISS-based methods over FSS. The first is that FSS only filters out Φ_{status} and Φ_{imaging} . When both Φ_{facial} and Φ_{personal} are stable in the subspace, it is not reliable to expect the projection to be unstable just because Φ_{personal} is different. The second advantage is that the ISS-based method explicitly calculates the projections in the ISS ($\mathbf{z}_c^{(k)}$ in Fig. 4). This provides more room for further improvements, such as the Mahalanobis distance used in ISS2 and the average Euclidean distance used in ISS3. Bayesian FR models the extrapersonal and intrapersonal differences with two subspaces. However, under uncontrolled conditions, the possible cases of both kinds of differences will exponentially grow to an unmanageable size. On the other hand, the ISS-based methods avoid directly modeling different kinds of information, but tries to filter out all useless information and only retain what is useful for recognition.

V. EXPERIMENTS

A. Methodology

In the experiments, first, ISS1 is compared with those related methods described in Section IV-B and some of their variants. Then, the properties of the ISS-based algorithm are analyzed, including the comparison of ISS1, ISS2, and ISS3.

Three data sets are used in the experiments. The first is the CMU PIE database [30]. The face images are greatly different in pose, illumination, and expression. The completely dark (without any illumination) images are removed from the database, which leaves 38 707 images from 68 individuals used in our experiments. The faces are normalized by fixing the positions of the two eyes (for those profile faces, the positions

of one eye and the nose tip are used). The normalized face image has 66×46 pixels. Some typical normalized faces are shown in Fig. 6(a). The second data set is a subset of the grayscale FERET database [27]. Most subjects in the FERET database just have several face images with limited variations. To simulate the uncontrolled conditions, those subjects with at least 30 different face images are selected from the FERET database. In total, there are 56 persons selected. Their face images remarkably vary in pose, illumination, expression, occlusion (glasses), and image-taken date. Compared to the possible variations, the number of face images per person is relatively small, which greatly increases the difficulty of accurate recognition. The face images are normalized by the same method used on the PIE database, resulting in the 66×46 images. Some typical images are shown in Fig. 6(b). The third data set is used to test the algorithms in another case: fewer individuals, sufficient face images per person, but with more variations. We have collected 23 978 images from 14 individuals through a web camera to compose the FRU face database. This database attempts to simulate most possible variations in real FR applications. The variations include pose, illumination, expression, talking, occlusion (with/without cap, scarf, and glasses), image-taken date, and imaging noise. Moreover, in the experiments, the faces are only roughly cropped from the background (without face detection, rotation, and scale) to test the tolerance of the algorithms to inaccurate face detections. The cropped face image has 55×42 pixels (different from the PIE database and the FERET database because of different normalization method). Some typical faces from this database are shown in Fig. 6(c). Note that some faces are partly out of the cropped image, which can be regarded as a special case of occlusion. For all three data sets, after the geometric normalization, the images are histogram equalized and then vectorized to a zero-mean, one-variance vector.

The possible intrapersonal variations, the number of different subjects, and the average number of images per person in the three face image sets are summarized in Table II. The numbers in the table indicate how many possible cases for each item. “N/A” means no such variation. “Uncontrolled” means no limitation on the possible cases. It can be seen that the three data sets correspond to the following three cases:

Case 1) *PIE*: many different subjects, many possible intrapersonal variations, and sufficient face images per person;

TABLE II
VARIATIONS IN THE THREE FACE IMAGE SETS

Data Set	Pose	Illumination	Expression	Talking	Occlusion
PIE	13	45	3	2	2
FERET	20	2	2	N/A	2
FRU	Uncontrolled	Uncontrolled	3	2	4

Data Set	Date Taken	Noise	Face Detection	Subjects	Images/Person (Avg.)
PIE	N/A	N/A	Manual	68	569
FERET	Within 3 Years	N/A	Manual	56	39
FRU	Within 1 Month	Web Camera	Inaccurate	14	1712

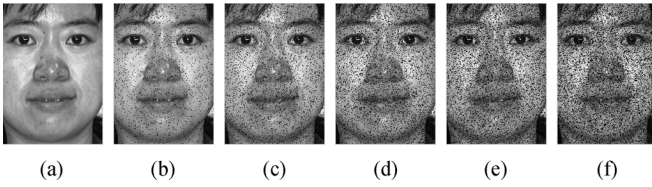


Fig. 7. Face images with (a) 0% (original), (b) 5%, (c) 10%, (d) 15%, (e) 20%, and (f) 25% missing pixels.

Case 2) *FERET*: many different subjects, many possible intrapersonal variations, and insufficient face images per person;

Case 3) *FRU*: fewer subjects, uncontrolled intrapersonal variations, and sufficient face images per person.

Note that although most face images in PIE and FERET are obtained in controlled environments, the control information (such as the pose and lighting labels) is not used to train ISS in the experiments. Moreover, these two databases are with perhaps the most diversified variations among the publicly available face databases. Thus, they can be viewed as good simulations of uncontrolled conditions. Another issue is that the number of images per person may seem too large to be practical. However, taking into consideration that these images do not need to be obtained in a controlled environment, it is not difficult to obtain a large amount of images from each person. For instance, using an ordinary web camera with frame rate 30 fps, 1712 face images can be obtained in 57 s.

There are remarkable illumination and pose variations in both databases. Among the methods mentioned in Section IV-B, only Eigenface is not specially designed to deal with illumination variation. It has been reported that discarding the first few eigenfaces will endow Eigenface with certain ability to handle illumination variation [3]. This is tested in the experiments by discarding the first three eigenfaces, which is denoted by Eigenface-3. As for the pose variation, except for ISS, PDBNN, FSS, Eigenfeature+Eigenface, and KPCA, none of the other methods is designed for the multiview case, so a multiview version is extended for each of them in a way similar to the View-based Eigenface [26] (abbreviated as V-Eigenface). The multiview algorithms are denoted by V-Bayes, V-Fisherface, and V-(Eigenface-3), respectively.

In the experiments, the parameters of PDBNN are empirically determined through several trials. When the best performance is observed, the number of Gaussians for each individual is set to six, the learning rate for the Gaussian centers is set to 10^{-6} , the learning rate for the variance is set to 10^{-4} , the learning rate for the threshold is set to 0.05, and the penalty function for the threshold is the sigmoid function. As for the Bayes method, similar to [21], 1000 intrapersonal difference images and 4000 extrapersonal difference images are randomly sampled as the training set. The class conditional density is estimated by the subspace density estimation technique [23]. For Fisherface, as described in [3], the face images are first projected into an $(M - N)$ -dimensional subspace using PCA, and then using FLD to reduce the dimensionality to $N - 1$, where M is the number of training images, and N is the number of persons. For Eigenfeature+Eigenface, only left eye, right eye, and nose are used because most expressions are related to mouth. KPCA uses the RBF kernel with the bias 1. If not explicitly stated, the number of principal components p is set to 50 for all other algorithms using PCA. Usually, the first 50 principal components will explain about 90% variance in the data set for most methods compared in this experiment. The number 50 is not deliberately tuned because different methods might favor different subspace dimensionality. For a fair comparison, all methods should use the same subspace dimensionality. However, since ISS needs to calculate m minor components based on the 50 principal components, m must be less than 50 (in fact, this can be considered “unfair” to ISS since it uses lower dimensional features than other methods). In the comparative experiments, m is set to 20 (explains about 5% variance), which is also not specially chosen in favor of ISS. The value of m will be further investigated in Section V-B2. In fact, as can be seen from Fig. 9, $m = 20$ is not the best choice for ISS. The initial weight vectors in the SGA or ASGA subnet are set to random orthogonal unit vectors. The learning rates in (6) and (10) are set as $\alpha_1 = \alpha_2 = 0.01$.

All algorithms are tested by threefold cross validation. That is, the original face images are randomly divided into three equal-sized subsets while keeping the proportion of each individual. Then, in each fold, one subset is used as test set (probe faces) and the union of the remaining two is used as training set (gallery faces). For each algorithm, the average result of these three folds is recorded as the recognition performance. Conforming to the FERET testing protocol [27], both “is the top

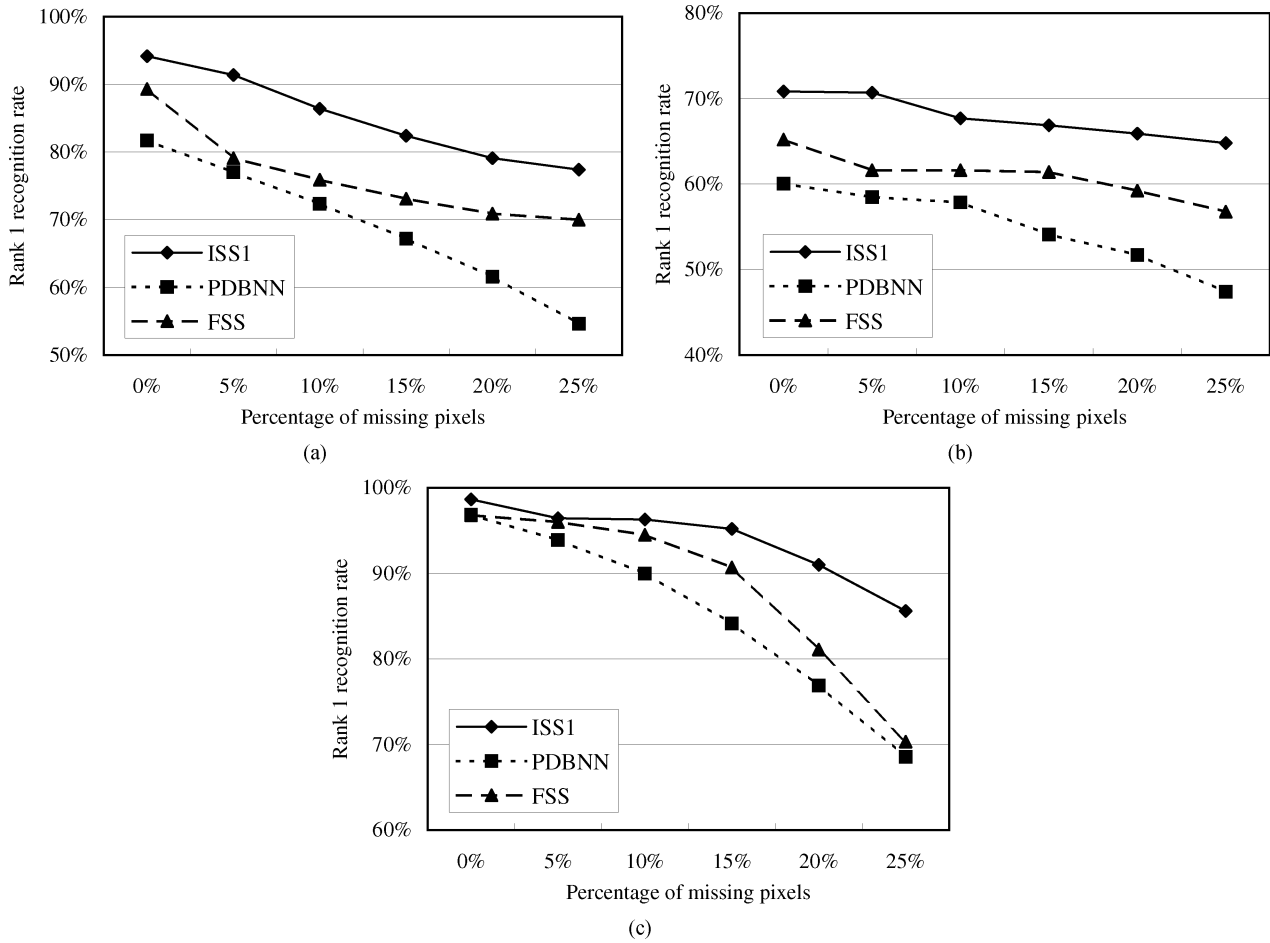


Fig. 8. Rank 1 recognition rates of ISS1, PDBNN, and FSS with respect to the percentage of missing pixels on (a) the PIE database, (b) the FERET subset, and (c) the FRU database.

match correct?” (rank 1 match) and “is the correct answer in the top k matches?” (rank k match) are considered. Moreover, the pairwise one-tailed t -test of ISS1 paired with other algorithms at the significance level 0.025 is performed.

B. Results

1) Comparative Experiments: The recognition rates from rank 1 to rank 3 on the PIE database are tabulated in Table III. The best performance in each case is highlighted in bold. The t -test results of ISS1 paired with other algorithms are listed in the parentheses by the corresponding recognition rates. The 1, 0, and -1 represent that ISS1 is significantly better, not significantly different, and significantly worse, respectively. Note that the algorithms below the dashed line require additional pose information during training.

Above the dashed line, the best performance is achieved by ISS1, which is significantly better than all the other algorithms. The superiority of ISS1 over FSS mainly comes from the SGA subnet of the ISNN, which removes the common facial characteristics in the face images. It is also worth mentioning that FSS uses a 50-dimensional subspace while ISS1 only uses a 20-dimensional subspace to describe each person. Thus, ISS1 is much faster and requires less storage than FSS. PDBNN performs worse than both ISS1 and FSS. This might be due to the

fact that under uncontrolled conditions, the distribution of the face images from each person is so complicated that the gradient-descent learning of PDBNN will tend to fall into local optima. The Bayes method results in poor performance, which is not surprising since the sampled difference images are only a small portion of all possible differences in the training set (under uncontrolled conditions, there are huge number of possible difference images due to combination explosion). As reported by previous work [3], Fisherface performs better than Eigenface and its variants because of the utilization of class information. The recognition rate of Eigenface-3 is much higher than that of Eigenface, which is consistent with the statement in [3]. Making use of local facial features also makes Eigenfeature+Eigenface perform better than Eigenface. KPCA is only slightly better than Eigenface. The reason is that KPCA shares the same shortcoming with Eigenface, i.e., under uncontrolled conditions, they cannot distinguish Φ_{personal} from Φ_{status} and Φ_{imaging} . It can also be found that there is a remarkable gap between the recognition rates of the best three methods (ISS1, FSS, and PDBNN) and those of the others, which indicates that the personalized approach might be a suitable solution to FR under uncontrolled conditions. Considering the view-based algorithms below the dashed line, ISS1 still performs the best. This is impressive because it does not use the additional information. With certain

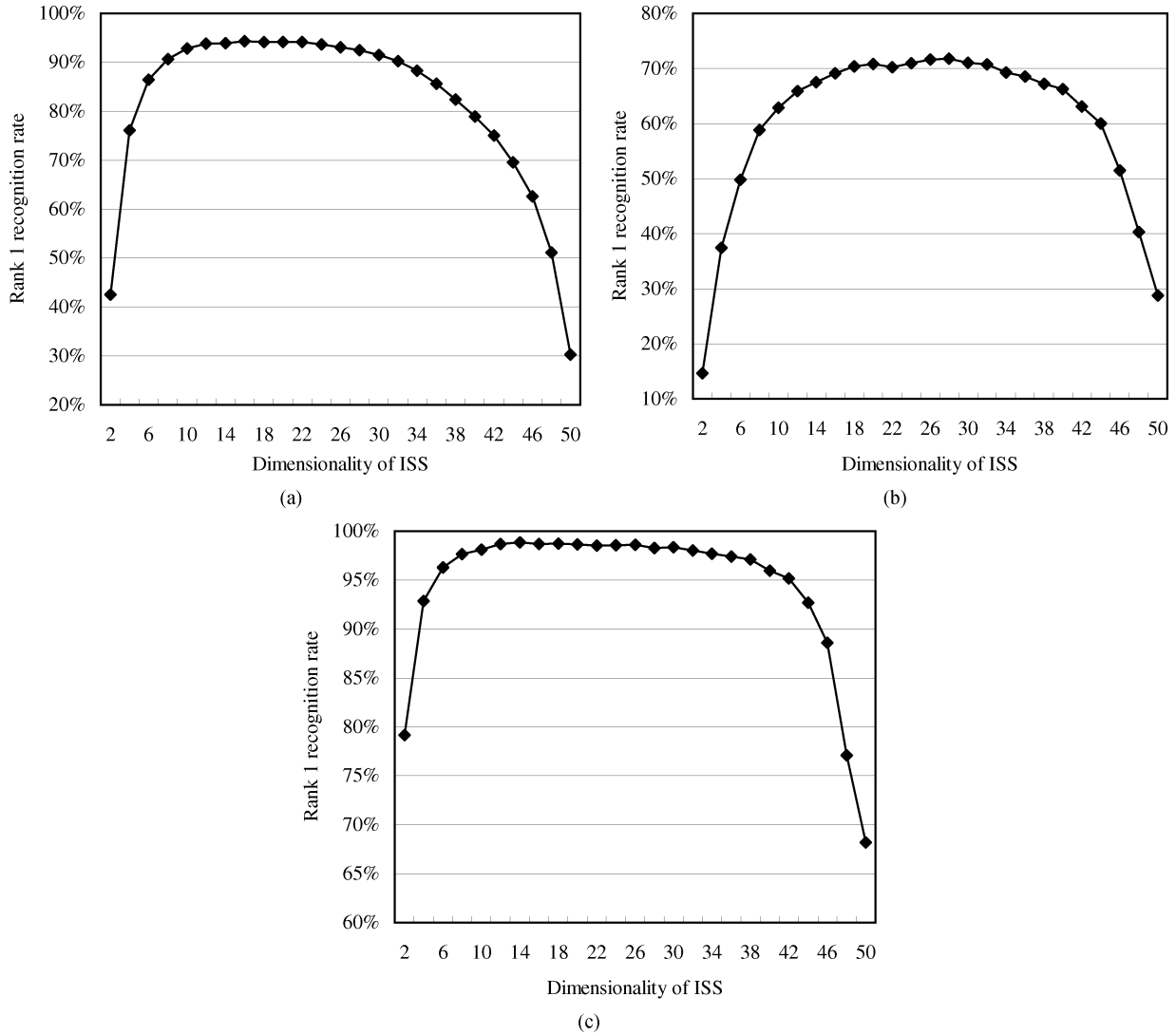


Fig. 9. Rank 1 recognition rates of ISS1 as a function of the ISS's dimensionality on (a) the PIE database, (b) the FERET subset, and (b) the FRU database.

ability to handle the pose variation, all the multiview variants make remarkable improvements over the corresponding original algorithms. Among them, V-Fisherface achieves the highest recognition rate, which slightly exceeds that of FSS, but is still worse than that of ISS1. In practice, the pose information is hard to obtain, especially under uncontrolled conditions. This greatly enlarges the superiority of ISS1 over V-Fisherface.

The recognition rates from rank 1 to rank 3 on the FERET subset are tabulated in Table IV. The number of subjects and possible variations in this data set is similar to that in the PIE database, but the number of images per person is much smaller. Relative to the vast possible variations, the average 39 images per person is insufficient. This can be evidenced by the remarkable accuracy degradation of all algorithms compared with their performance on the PIE database. Note that the high accuracy on FERET database reported in some previous literatures was usually obtained on a subset that restricts some variations, such as the most frequently used FA/FB subset which only contains the frontal face images. However, in this test, there are no limitations on the variations.

The relative performance of the 13 algorithms is similar to that on the PIE database. Whether the pose information is available or not, ISS1 performs significantly better than all the other algorithms. The runner-up is FSS, followed by V-Fisherface, and then PDBNN. Among the algorithms above the dashed line, there is also an apparent gap between the accuracy of the best three algorithms (ISS1, PDBNN, and FSS) and that of the others, which again suggests that the personalized approach might be a suitable solution to FR under uncontrolled conditions. It is noteworthy that KPCA performs worse than Eigenface. The reason might be that KPCA is relatively easier to overfit the training data. Also, it is always difficult to select the best kernel of KPCA for a particular data set. Since the number of images per person is relatively small, the training data can hardly represent all possibilities well. Thus, the overfitting problem will be more serious.

The recognition rates from rank 1 to rank 3 on the FRU database are tabulated in Table V. Since there are only 14 individuals in this database, and each of them has plenty of face images, although more variations are possible, almost all algorithms achieve better performances than those on the PIE data-

TABLE III
RECOGNITION RATES (IN PERCENT) ON THE PIE DATABASE

Methods	Rank 1	Rank 2	Rank 3
ISS1	94.16	96.59	97.32
PDBNN	81.70 (1)	88.44 (1)	91.45 (1)
FSS	89.30 (1)	92.62 (1)	94.04 (1)
Bayes	18.58 (1)	25.23 (1)	31.36 (1)
Fisherface	57.96 (1)	67.25 (1)	72.89 (1)
Eigenface	30.27 (1)	39.54 (1)	45.84 (1)
Eigenface-3	45.31 (1)	55.24 (1)	61.10 (1)
Eigenfeature+Eigenface	51.66 (1)	60.55 (1)	66.76 (1)
KPCA	30.42 (1)	40.73 (1)	47.48 (1)
-----	-----	-----	-----
V-Bayes	48.86 (1)	61.33 (1)	68.56 (1)
V-Fisherface	89.88 (1)	93.36 (1)	94.77 (1)
V-Eigenface	65.92 (1)	72.18 (1)	75.58 (1)
V-(Eigenface-3)	85.11 (1)	88.85 (1)	90.60 (1)

TABLE IV
RECOGNITION RATES (IN PERCENT) ON THE FERET SUBSET

Methods	Rank 1	Rank 2	Rank 3
ISS1	70.83	78.83	83.88
PDBNN	60.04 (1)	67.69 (1)	71.72 (1)
FSS	65.20 (1)	72.31 (1)	75.93 (1)
Bayes	10.38 (1)	18.78 (1)	24.93 (1)
Fisherface	37.74 (1)	44.61 (1)	50.40 (1)
Eigenface	28.76 (1)	41.94 (1)	50.34 (1)
Eigenface-3	35.31 (1)	46.04 (1)	52.25 (1)
Eigenfeature+Eigenface	42.49 (1)	53.83 (1)	61.20 (1)
KPCA	21.37 (1)	30.53 (1)	37.16 (1)
-----	-----	-----	-----
V-Bayes	16.63 (1)	22.91 (1)	27.76 (1)
V-Fisherface	64.28 (1)	70.42 (1)	74.39 (1)
V-Eigenface	43.92 (1)	54.24 (1)	60.31 (1)
V-(Eigenface-3)	51.02 (1)	59.43 (1)	64.55 (1)

base. Note that the main purpose of the experiment on the FRU database is not to compare with those on the PIE database and the FERET subset (it is predictable to get better result on such a database with much fewer persons), but to compare the relative performance of the algorithms under more variable conditions.

The comparative results on the FRU database are similar to those on the PIE database and the FERET subset. Above the dashed line, ISS1 is significantly better than all the other algorithms (with the only exception on the rank 3 recognition rate of PDBNN, where the two algorithms are not significantly different). PDBNN also achieves a good performance just next to that of ISS1, and better than that of FSS. This might be because with fewer classes, the mixture of Gaussians learned by PDBNN is enough to separate different classes. Different from the previous results, Eigenfeature+Eigenface performs worse than Eigenface, possibly due to the fact that the faces in the FRU

TABLE V
RECOGNITION RATES (IN PERCENT) ON THE FRU DATABASE

Methods	Rank 1	Rank 2	Rank 3
ISS1	98.65	99.51	99.79
PDBNN	96.84 (1)	99.04 (1)	99.54 (0)
FSS	96.79 (1)	98.52 (1)	98.99 (1)
Bayes	39.52 (1)	59.96 (1)	73.80 (1)
Fisherface	91.32 (1)	96.56 (1)	98.10 (1)
Eigenface	68.18 (1)	80.63 (1)	86.62 (1)
Eigenface-3	75.24 (1)	86.15 (1)	90.06 (1)
Eigenfeature+Eigenface	50.82 (1)	66.60 (1)	76.12 (1)
KPCA	68.59 (1)	81.40 (1)	86.94 (1)
-----	-----	-----	-----
V-Bayes	53.71 (1)	74.76 (1)	85.78 (1)
V-Fisherface	96.41 (1)	98.56 (1)	99.26 (1)
V-Eigenface	76.01 (1)	86.05 (1)	90.26 (1)
V-(Eigenface-3)	83.56 (1)	90.60 (1)	93.06 (1)

database are only roughly cropped from the background while the performance of Eigenfeature heavily depends on the accuracy of the feature locations. Thus, the inaccurate face detection causes poor performance of Eigenfeature+Eigenface. Below the dashed line, there is still no other algorithm that achieves accuracy higher than that of ISS1, but with extra pose information, all of the view-based algorithms are better than the original ones.

2) *Properties of the ISS-Based Algorithm:* Although the data sets used in the experiments include almost all facial variations ever mentioned in the previous literatures, some insight into how to deal with unseen situations in the future is still required to support the claim of “working under uncontrolled conditions.” Since it is impossible to collect an image set that includes all possible facial variations, we abstract the variations as information loss and use missing pixels to simulate it. In detail, the gray scales of a certain proportion of pixels in the face image are randomly set to zeros. Fig. 7 gives the example images with up to 25% missing pixels. The best three algorithms (ISS1, PDBNN, and FSS) in the comparative experiments are tested again with the same configuration except that there are from 5% to 25% missing pixels in the test images. The rank 1 recognition rates with respect to the percentage of missing pixels are plotted in Fig. 8. As expected, all algorithms degrade with the increase of missing pixels, but ISS1 is more robust against missing pixels than the other two. It performs the best in all cases, and its superiority is more significant at higher missing percentages. This reveals a positive future of ISS applying to uncontrolled conditions.

To study the impact of ISS’s dimensionality on the performance of the ISS-based FR, different values of m from 2 to 50, with 2 as the interval, are tested. The rank 1 recognition rates with different values of m of ISS1 on the PIE database, the FERET subset, and the FRU database are shown in Fig. 9(a)–(c) respectively.

It can be seen from Fig. 9(a) that the recognition rate of ISS1 increases rapidly to a high level above 90% with the dimensionality increases from 2 to 8. Referring back to Table III, it can be

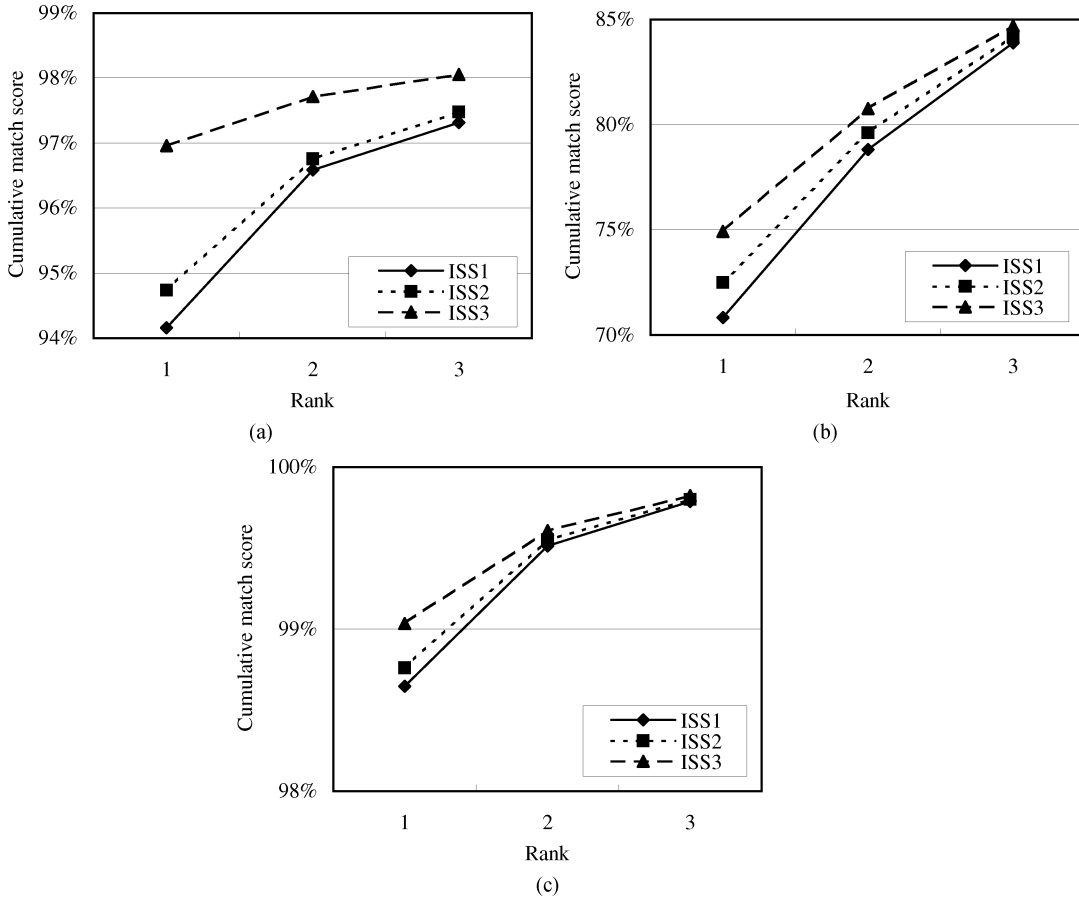


Fig. 10. Comparison of ISS1, ISS2, and ISS3 on (a) the PIE database, (b) the FERET subset, and (c) the FRU database.

found that without pose information, ISS1 is able to exceed all the other algorithms by using as few as 8 dimensions. Compared with the original image space of $66 \times 46 = 3036$ dimensions, the efficiency of the extremely low-dimensional ISS exhibits its ability to extract the only useful information Φ_{personal} . When the dimensionality exceeds 8, the performance of ISS1 remains relatively steady above 90% until m rises to 32. The highest recognition rate (94.29%) is achieved when $m = 16$, which is slightly higher than that listed in Table III (when $m = 20$). After that, the recognition rate gradually decreases to as low as about 30%. Note that when m increases to 50, all the components are used and ISS1 is equivalent to the standard Eigenface method. Fig. 9(b) is similar to Fig. 9(a). With the growth of m , the accuracy of ISS1 increases rapidly, exceeds those of all other methods when $m = 12$, achieves the best performance (71.79%) when $m = 28$, stays relatively steady until m increases to 40, and then drops to the level of standard Eigenface. Note that the dimensionality required to exceed other methods and achieve the best performance is higher than that on the PIE database. This might be due to the relatively small number of images per person, which makes the ASGA subnets in the ISNN hard to converge to the real minor component, and thus more dimensions are needed to describe Φ_{personal} . The situation in Fig. 9(c) is also similar. The recognition rate of ISS1 starts from 79.14%, surpasses all the baseline algorithms when the dimensionality increases to 8, achieves the best performance when the

dimensionality equals to 14, and then gradually decreases to the level of standard Eigenface (68.18%).

Fig. 9 reveals that too low or too high dimensionality of ISS may both lead to performance degradation. Too low-dimensional ISS may not be able to capture all information from Φ_{personal} , and too high-dimensional ISS may also contain the information from Φ_{status} and Φ_{imaging} . However, the relatively wide scale of m that corresponds to steady performance means not much effort on parameter tuning is needed for the ISS-based method to obtain a good result.

The performances of ISS1, ISS2, and ISS3 on the PIE database, the FERET subset, and the FRU database are compared in Fig. 10(a)–(c), respectively. It can be seen that on all data sets the ISS-based algorithms can be further improved. The performances of the three algorithms can be sorted as $\text{ISS3} > \text{ISS2} > \text{ISS1}$. Note that the only difference among the three algorithms is the definition of the stability measurement function $S(z)$. Thus, the improvement of ISS3 and ISS2 over ISS1 comes from the more sophisticated measurements of stability. Of course, special domain knowledge could be considered to design other stability measurements for even better performance.

VI. CONCLUSION

This paper presents one of the first attempts toward FR under uncontrolled conditions (FRU), which extends our preliminary work [8]. The proposed method is based on a subspace called

ISS, which only expresses the information of personal characteristics. The concept of ISS is derived from the analysis of the four different kinds of information contained in the face images. Instead of the "divide and conquer" strategy, i.e., modeling different kinds of information one by one, this paper focuses on what is useful for the task of FR and tries to filter out all other information. A neural network structure named ISNN is proposed to realize ISS. On this basis, three ISS-based algorithms are designed for FRU. The ISS-based method is compared with 12 existing FR techniques on three large databases with vast variations and achieves the best performance. Moreover, unlike many other FR algorithms, the ISS-based method does not require any extra information, such as face poses, which makes it more practical and reliable.

The ISS-based approach can be viewed as a general framework for FRU. Other novel subspace methods, including both linear and nonlinear ones, may be utilized to realize ISS to further improve the effectiveness or efficiency of ISNN. This will be one of the major future works following this paper.

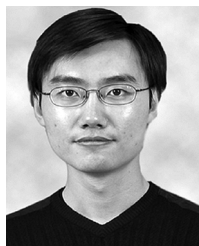
Currently, the training data for the ISS-based algorithms must contain many kinds of variations, which requires at least tens of images per person. Since the images do not require a controlled environment, creating a face database with many images per person can be easily done by an ordinary video camera. However, for those existing face databases without so many images per person, it is usually too costly to rebuild the whole database. In such a case, the problem might be partly solved by simulating different variations from single face image, such as the methods designed for addressing the problem of FR with one training image per person [31], [35], and the illumination simulation method used in [29]. Other more complex variations, such as viewing angle [9], aging effect [7], etc., might be simulated from single face image as well. This is another important future work.

Although the ISS-based algorithm is designed for uncontrolled FR, it might have many other applications in pattern recognition. Generally speaking, the first subspace is trained on all data to filter out common characteristics. Then, the projections in the subspace are divided into subsets according to the class labels. After that, a subspace that filters out the variable information is trained on each subset. These subspaces are expected to express only the information related to the class labels. Thus, ISS can be viewed as a model of complex patterns and might be further explored in other tasks of pattern recognition in the future.

REFERENCES

- [1] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problems of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721–732, Jul. 1997.
- [2] M. S. Bartlett, H. M. Lades, and T. Sejnowski, "Independent component representation for face recognition," in *Proc. SPIE Conf. Human Vis. Electron. Imag. III*, San Jose, CA, 1998, vol. 3299, pp. 528–539.
- [3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [4] "Face recognition vendor test (FRVT)," 2006 [Online]. Available: <http://www.frvt.org/FRVT2006/>
- [5] X. Geng, D.-C. Zhan, and Z.-H. Zhou, "Supervised nonlinear dimensionality reduction for visualization and classification," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 35, no. 6, pp. 1098–1107, Dec. 2005.
- [6] X. Geng and Z.-H. Zhou, "Image region selection and ensemble for face recognition," *J. Comput. Sci. Technol.*, vol. 21, no. 1, pp. 116–125, 2006.
- [7] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.
- [8] X. Geng, Z.-H. Zhou, and H. Dai, "Uncontrolled face recognition by individual stable neural network," in *Proc. 9th Pacific Rim Int. Conf. Artif. Intell.*, Guilin, China, 2006, pp. 553–562.
- [9] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 449–465, Apr. 2004.
- [10] F. J. Huang, T. Chen, Z.-H. Zhou, and H.-J. Zhang, "Pose invariant face recognition," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Grenoble, France, 2000, pp. 245–250.
- [11] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- [12] R. Kothari and S. Swaminathan, "On minor component and active learning," in *Intelligent Engineering Systems Through Artificial Neural Networks*, C. Dagli, M. Akay, Ersoy, B. Fernandez, and A. Smith, Eds. New York: ASME, 1997, vol. 7, Smart Engineering Systems: Neural Networks, Fuzzy Logic, Data Mining, and Evolutionary Programming, pp. 93–98.
- [13] K.-C. Kwak and W. Pedrycz, "Face recognition using an enhanced independent component analysis approach," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 530–541, Mar. 2007.
- [14] Q. Li, J. Ye, and C. Kambhampettu, "Linear projection methods in face recognition under unconstrained illuminations: A comparative study," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, 2004, pp. 474–481.
- [15] S. H. Lin, S. Y. Kung, and L. J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 114–132, Jan. 1997.
- [16] Q. Liu, X. Tang, H. Lu, and S. Ma, "Face recognition using kernel scatter-difference-based discriminant analysis," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 1081–1085, Jul. 2006.
- [17] Y.-H. Liu and Y.-T. Chen, "Face recognition using total margin-based adaptive fuzzy support vector machines," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 178–192, Jan. 2007.
- [18] J. Lu, K. Plataniotis, A. Venetsanopoulos, and S. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 166–178, Jan. 2006.
- [19] J. Lu, X. Yuan, and T. Yahagi, "A method of face recognition based on fuzzy c-means clustering and associated sub-NNs," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 150–160, Jan. 2007.
- [20] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 780–788, Jun. 2002.
- [21] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognit.*, vol. 33, no. 11, pp. 1771–1782, 2000.
- [22] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, Jul. 1997.
- [23] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object detection," in *Proc. 5th Int. Conf. Comput. Vis.*, Boston, MA, 1995, pp. 786–793.
- [24] E. Oja, *Subspace Methods of Pattern Recognition*. London, U.K.: Research Studies/Wiley, 1983.
- [25] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Netw.*, vol. 5, no. 6, pp. 927–935, 1992.
- [26] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. CVPR*, Seattle, WA, 1994, pp. 84–91.
- [27] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
- [28] H. S. Seung and D. D. Lee, "Cognition – The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.
- [29] S. Shan, W. Gao, and D. Zhao, "Face identification based on face-specific subspace," *Int. J. Imag. Syst. Technol.*, vol. 13, Special Issue on Face Processing, Analysis and Synthesis, no. 1, pp. 23–32, 2003.
- [30] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.
- [31] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, 2006.

- [32] J. B. Tenenbaum, V. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [33] M. A. Turk and A. Pentland, "Eigenface for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
- [34] C. Y. Wu, C. Liu, H. Y. Shum, Y. Q. Xu, and Z. Y. Zhang, "Automatic eyeglasses removal from face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 322–336, Mar. 2004.
- [35] J. Wu and Z.-H. Zhou, "Face recognition with one training image per person," *Pattern Recognit. Lett.*, vol. 23, no. 14, pp. 1711–1719, 2002.
- [36] L. Xu, A. Krzyzak, and E. Oja, "Neural nets for dual subspace pattern recognition method," *Int. J. Neural Syst.*, vol. 2, no. 3, pp. 169–184, 1991.
- [37] L. Xu, E. Oja, and C. Suen, "Modified Hebbian learning for curve and surface fitting," *Neural Netw.*, vol. 5, no. 3, pp. 441–457, 1992.
- [38] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–459, 2003.
- [39] W. Zheng, J.-H. Lai, and P. C. Yuen, "GA-fisher: A new LDA-based face recognition algorithm with selection of principal components," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 35, no. 5, pp. 1065–1078, Oct. 2005.



Xin Geng received the B.Sc. and M.Sc. degrees in computer science from Nanjing University, Nanjing, China, in 2001 and 2004, respectively.

Since submitting his Ph.D. dissertation for examination at Deakin University, Melbourne, Australia in early 2008, he has been a Research Fellow at the School of Engineering and Information Technology, Deakin University, Melbourne, VIC, Australia. His research interests include computer vision, pattern recognition, and machine learning. He has published 16 papers in these areas, including those published

in prestigious journals and top international conferences.

Mr. Geng was a Guest Editor of several international journals including PRL and IJPRAI. He has served as a Program Committee Member for several international conferences including PRICAI and CIT. He is also a frequent reviewer for various international journals and conferences.



Zhi-Hua Zhou (S'00–M'01–SM'06) received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Nanjing University, Nanjing, China, in 1996, 1998, and 2000, respectively, all with the highest honors.

He joined the Department of Computer Science and Technology, Nanjing University, as a Lecturer in 2001, and is currently Cheung Kong Professor and Head of the LAMDA group. His research interests are in artificial intelligence, machine learning, data mining, information retrieval, pattern recognition,

evolutionary computation, and neural computation. In these areas, he has published over 50 papers in leading international journals or conference proceedings.

Dr. Zhou has won various awards/honors including the National Science & Technology Award for Young Scholars of China (2006), the Award of National Science Fund for Distinguished Young Scholars of China (2003), the National Excellent Doctoral Dissertation Award of China (2003), the Microsoft Young Professorship Award (2006), etc. He is an Associate Editor of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING. He is on the editorial boards of *Artificial Intelligence in Medicine*, *Intelligent Data Analysis*, *Knowledge and Information Systems*, etc. He is/was a PAKDD Steering Committee Member, Program Committee Chair/Co-Chair of PAKDD'07 and PRICAI'08, Vice Chair/Area Chair of several conferences including ICDM'06, ICDM'08, etc., Program Committee Member of various international conferences including AAAI, ICML, ECML, SIGKDD, ICDM, ACM Multimedia, etc., and General Chair/Co-Chair or Program Committee Chair/Co-Chair of a dozen of native conferences. He is a Senior Member of the China Computer Federation (CCF), the Vice Chair of the CCF Artificial Intelligence & Pattern Recognition Society, an Executive Committee Member of the Chinese Association of Artificial Intelligence (CAAI), the Chair of the CAAI Machine Learning Society, and the Chair of the IEEE Computer Society Nanjing Chapter.



Kate Smith-Miles (S'93–M'96–SM'01) received the B.Sc. degree (with honors) in mathematics and the Ph.D. degree in electrical engineering both from the University of Melbourne, Melbourne, VIC, Australia.

Currently, she is a Professor and Head of the School of Engineering and Information Technology, Deakin University, Melbourne, VIC, Australia. Before joining Deakin University in 2006, she was a Professor in the Faculty of Information Technology at Monash University, Australia, where she held a

variety of leadership roles over the previous ten years including Deputy Head and Director of Research for the School of Business Systems, and Co-Director of the Monash Data Mining Centre. She has published two books on neural networks and data mining applications, and over 175 refereed journal and international conference papers in the areas of neural networks, combinatorial optimization, intelligent systems, and data mining. She has supervised to completion 16 Ph.D. students, and has been awarded over AUD\$1.5 million in competitive grants, including seven Australian Research Council grants and industry awards. In addition to her academic activities, she also regularly acts as a consultant to industry in the areas of data mining and intelligent systems.

Dr. Smith-Miles is on the editorial board of several international journals, and has been a member of the organizing committee for over 50 international data mining and neural network conferences, including several as chair. She is a frequent reviewer of international research activities including grant applications in Canada, U.K., Finland, Singapore, and Australia, refereeing for international research journals, and Ph.D. examinations. In 2007, she was appointed Chair of the IEEE Technical Committee on Data Mining (IEEE Computational Intelligence Society).