Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Subspace transform induced robust similarity measure for facial images*

Jian ZHANG[†1], Heng ZHANG[†‡2], Li-ling BO[2], Hong-ran LI[1], Shuai XU[1], Dong-qing YUAN[2]

[1]*Department of Computer Engineering, Jiangsu Ocean University, Lianyungang 222005, China*

[2]*Department of Mathematics, Jiangsu Ocean University, Lianyungang 222005, China*

[†]E-mail: zhangjian@jou.edu.cn; zhangheng@jou.edu.cn

Received Oct. 11, 2019; Revision accepted Mar. 23, 2020; Crosschecked July 23, 2020

Abstract: Similarity measure has long played a critical role and attracted great interest in various areas such as pattern recognition and machine perception. Nevertheless, there remains the issue of developing an efficient two-dimensional (2D) robust similarity measure method for images. Inspired by the properties of subspace, we develop an effective 2D image similarity measure technique, named transformation similarity measure (TSM), for robust face recognition. Specifically, the TSM method robustly determines the similarity between two well-aligned frontal facial images while weakening interference in the face recognition by linear transformation and singular value decomposition. We present the mathematical features and some odds to reveal the feasible and robust measure mechanism of TSM. The performance of the TSM method, combined with the nearest neighbor rule, is evaluated in face recognition under different challenges. Experimental results clearly show the advantages of the TSM method in terms of accuracy and robustness.

Key words: Subspace analysis; Image similarity measure; Face recognition; Pattern recognition
https://doi.org/10.1631/FITEE.1900552 **CLC number:** TP391

1 Introduction

Similarity measure has long played a critical role and attracted great interest in face recognition and other pattern recognition tasks (Cover and Hart, 1967; Bowyer and Phillips, 1998; Sebe et al., 2000; Perlibakas, 2004; Paredes and Vidal, 2006; Wang H, 2006; Liu and Jin, 2006). Many new similarity measure methods have been proposed and reviewed (Sebe et al., 2000; Perlibakas, 2004; Paredes and Vidal, 2006; Zou and Yuen, 2010; Wu J et al., 2013; Liu CJ, 2014; Chen et al., 2015; Wen et al., 2016; Zhang YM et al., 2016; Peng et al., 2017, 2018; Sun et al., 2018; Zhou et al., 2018; He et al., 2019; Wu MC et al., 2019). Among the proposed methods for face recognition, the most widely used metric is the vector 2-norm, i.e., Euclidean distance (Chen et al., 2015). There are two key advantages of the Euclidean distance: it is invariant to translation and rotation, and it

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 61873106), the Natural Science Foundation of Jiangsu Province, China (No. BK20171264), the Jiangsu Qing Lan Project to Cultivate Middle-Aged and Young Science Leaders, China, the Jiangsu Six Talent Peak Project, China (Nos. XYDXX-047 and XYDXX-140), the University Science Research General Research General Project of Jiangsu Province, China (Nos. 18KJB520005 and 19KJB520004), the Innovation Fund Project for Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, China (No. JYB201609), the Lianyungang Hai Yan Plan, China (Nos. 2018-ZD-003, 2018-QD-001, and 2018-QD-012), the Science and Technology Project of Lianyungang Hightech Zone, China (Nos. ZD201910 and ZD201912), and the Natural Science Foundation Project of Huaihai Institute of Technology, China (No. Z2017005)

ORCID: Jian ZHANG, https://orcid.org/0000-0001-5764-9351; Heng ZHANG, https://orcid.org/0000-0002-4201-3892

is equal to the Frobenius norm of an image matrix. Nonetheless, the Euclidean distance may fail to offer a discriminative and robust metric for some face recognition challenges. For example, variations between faces of the same person due to illumination or occlusion are almost always larger than image variations of the change of identities (Adini et al., 1997; Georghiades, 2001).

To address this problem, discriminative metric learning approaches fusing prior information by a learning process with training datasets (e.g., local adaptive distance metric learning, whitened cosine distance, and feature extraction methods) have sprung up in the pattern recognition and face recognition (Paredes and Vidal, 2006; Wang H, 2006). Generally, a discriminative metric adopts training examples to learn a global or local model of the sample distributions, such as a linear subspace or a manifold model generalized to new samples by strategies like the weighted method. Nevertheless, the recognition performance depends largely on the diversity of training datasets, which is difficult to obtain. For example, there must be a variety of face images containing illumination changes to ensure that the learning model can acquire illumination variation information.

The linear reconstruction measure (LRM), which determines the similarity between the query sample and all the other known training samples by sorting the minimum L_2-norm error, has significant potential in solving illumination variation and block occlusion problems in face recognition (Zhang J and Yang, 2014). LRM-L_1 is a sparse representation classifier (Wright et al., 2009) and LRM-L_2 is a linear regression classifier (Naseem et al., 2010), which were both expressed as similarity measurement mechanisms by Zhang J and Yang (2014). Although achieving impressive results, compared with many well-known face recognition methods, LRM is not robust enough to illumination variation since it is difficult for the known samples to span the real illumination variation space in practical applications. Deng et al. (2012) applied an auxiliary intra-class variant dictionary to boost the recognition performance of LRM under illumination changes and block occlusion. Inspired by this idea, Zhuang et al. (2013) proposed a sparse illumination learning and transfer (SILT) technique. Illumination in SILT is learned by fitting illumination examples of auxiliary face images from one or more additional subjects with a sparsely used illumination dictionary. Wagner et al. (2012) proposed an illumination dictionary construction method by simulating a realistic lighting scene to enhance the ability of sparse representation classification under illumination variations. Generally, the above-mentioned methods show better performance than LRM with the generation of virtual samples. However, illumination dictionary construction method will inevitably lead to a large-scale training dataset, which not only has high computational cost but also increases the difficulty in acquiring a large number of images. In addition, it is still an open problem to guarantee the quality and reality of the generated virtual lighting samples.

We propose a fundamental similarity measure technique called the transformation similarity measure (TSM) without the learning process. The motivation of this study is to decrease the gap between the Euclidean distance (Frobenius metric) and some face recognition challenges by removing the related transformation factors that are independent of the intrinsic structural changes of human faces. To achieve this objective, we model the similarity of two facial images as a linear regression with the minimum Frobenius norm loss. Then, singular value decomposition is adopted to remove some interference related transformations such as illumination changes and block occlusion. Finally, the Frobenius norm distance between the transformation matrix and the unit matrix is employed as the similarity of two face images based on Assumption 1 in Section 2.

To illustrate the effectiveness of our approach, we present some mathematical features and examples to explain the rationale and robust mechanism of TSM for several face recognition challenges, such as illumination changes, block occlusion, sketch-photo matching, and inverse image recognition. While the proposed similarity measure is significant in face recognition and other image recognition tasks, in this study, we will focus on face image classification, which determines the identities of face images in a large-scale face image gallery. We perform experiments on several benchmark face databases to verify the TSM performance, and combine TSM with the nearest neighbor rule under different challenges. Experimental results show the advantages of the TSM method in terms of accuracy and robustness.

Note that we have proposed a face image representation method, named nearest orthogonal matrix representation (NOMR), with singular value decomposition (SVD) for face recognition in our previous work (Zhang J et al., 2015). NOMR is quite different from the proposed TSM. Essentially, NOMR is an image representation method which processes each image by SVD separately. However, TSM is an image similarity measure approach, where SVD is used to analyze the similarity between two images.

2 Image transformation similarity measure

Given two $m\times n$-dimensional gray facial image matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, the similarity between them can be formulated by the following steps:

First, linear transformation $\boldsymbol{T}$ from $\boldsymbol{B}$ to $\boldsymbol{A}$ can be calculated by

$$\boldsymbol{T}=\arg\min_{\boldsymbol{T}} \|\boldsymbol{A}-\boldsymbol{T}\boldsymbol{B}\|_{\mathrm{F}}^2. \tag{1}$$

$\boldsymbol{T}$ represents the row-direction linear transformation, incarnating the process of turning image $\boldsymbol{B}$ into image $\boldsymbol{A}$ with the minimum Frobenius norm error. By transposing $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ and $\boldsymbol{B}\in\mathbb{R}^{m\times n}$, the column-direction transformation could also be achieved with Eq. (1). The solution to Eq. (1) is as follows: if $m=n$ and $\boldsymbol{B}$ is invertible, then $\boldsymbol{T}=\boldsymbol{A}\boldsymbol{B}^{-1}\in\mathbb{R}^{m\times m}$; otherwise, the least-squares solution $\boldsymbol{T}=\boldsymbol{A}\boldsymbol{B}^{+}\in\mathbb{R}^{m\times m}$ will be obtained, where $\boldsymbol{B}^{+}$ denotes the Moore-Penrose generalized inverse matrix.

Secondly, transformation $\boldsymbol{T}$ will be decomposed as follows by SVD:

$$\boldsymbol{T}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}, \tag{2}$$

where $\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m)$ is a diagonal matrix of size $m\times m$ with nonnegative real numbers at the diagonal (σ_i is the singular value of $\boldsymbol{T}$ by convention arranged in a non-increasing order $\sigma_1\geq\sigma_2\geq\ldots\geq\sigma_m\geq 0$), and the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are termed left-singular vectors and right-singular vectors of $\boldsymbol{T}$, respectively.

SVD provides a convenient way to break a matrix into simpler, meaningful pieces. From the perspective of linear transformation, these three pieces (i.e., $\boldsymbol{\Sigma}$, $\boldsymbol{U}$, and $\boldsymbol{V}$) have clear physical meanings; $\boldsymbol{U}$ and $\boldsymbol{V}$ represent the rotational transformations and $\boldsymbol{\Sigma}$ denotes the stretching transformation (Demirel et al., 2008). Generally, the stretching transformation is independent of the identity variation information. For instance, the stretching transformation often corresponds to the intra-class variations (such as illumination variations and mode changes), which are obstacles for some special face recognition tasks, such as alternating illumination recognition and heterogeneous pattern recognition (Zhang J et al., 2015). Specifically, in Zhang DQ et al. (2005), more similar images were obtained by singular value perturbation to solve the single training sample recognition problem. Thus, we can remove the stretching transformation $\boldsymbol{\Sigma}$ from $\boldsymbol{T}$ and obtain transformation $\boldsymbol{T}^*$ as

$$\boldsymbol{T}^*=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}, \tag{3}$$

where $\boldsymbol{T}^*\in\mathbb{R}^{m\times m}$ represents the transform process from images $\boldsymbol{B}$ to $\boldsymbol{A}$ excluding some intra-class variations.

$\boldsymbol{T}^*$ is the solution to the following model:

$$\boldsymbol{T}^*=\arg\min_{\boldsymbol{T}} \|\boldsymbol{A}-\boldsymbol{T}\boldsymbol{B}\|_{\mathrm{F}}^2 \ \text{s.t.}\ \boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}=\boldsymbol{I}. \tag{4}$$

Eq. (4) is a model of the classical orthogonal Procrustes problem (OPP). The OPP solves the problem of how closely matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ can approximate a given matrix $\boldsymbol{B}\in\mathbb{R}^{m\times n}$, which is multiplied by matrix $\boldsymbol{T}\in\mathbb{R}^{m\times m}$ with orthogonal columns in the sense of the Frobenius norm. Hence, $\boldsymbol{T}^*$ is a unique and orthogonal matrix.

In fact, with images $\boldsymbol{A}$ and $\boldsymbol{B}$ taken as two vector subspaces by row or column, $\boldsymbol{T}^*$ and $\boldsymbol{\Sigma}$ will represent the direction change and length change between the bases of these two vector subspaces, respectively.

Based on the above discussion, we have the following assumption:

Assumption 1 For similar face images $\boldsymbol{A}$ and $\boldsymbol{B}$, the solution $\boldsymbol{T}^*$ to Eq. (4) tends to be the unit matrix $\boldsymbol{I}$ in the sense of the Frobenius norm.

If $\boldsymbol{A}=\boldsymbol{B}$ (i.e., $\boldsymbol{A}$ and $\boldsymbol{B}$ are the same image of one object class), it is easy to obtain $\boldsymbol{T}^*=\boldsymbol{I}$. In other cases, an example is established to verify Assumption 1. In this example, a face image is selected from the

Extended Yale B database as $\boldsymbol{B}$ and three homogeneous and heterogeneous images of $\boldsymbol{B}$ are randomly selected as $\boldsymbol{A}$. Then, we calculate each $\boldsymbol{T}^*$ by the steps mentioned above and present these transformations as visual images. Illustration of this example is shown in Fig. 1. $\boldsymbol{T}^*$ from $\boldsymbol{B}$ to its homogeneous $\boldsymbol{A}$ is close to the unit matrix $\boldsymbol{I}$, but its heterogeneous image does not possess this feature.

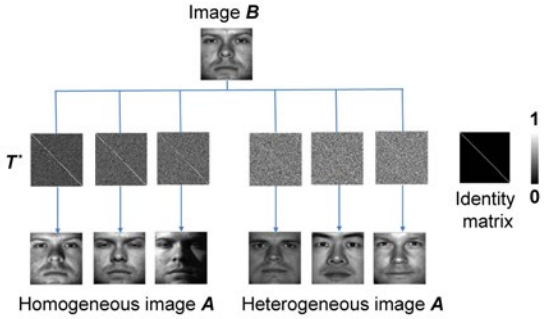

Fig. 1 Illustration of Assumption 1

The second row shows the images of transformation $\boldsymbol{T}^*$

Inspired by Assumption 1, similarity between $\boldsymbol{A}$ and $\boldsymbol{B}$ can be converted into the similarity between $\boldsymbol{T}^*$ and $\boldsymbol{I}$. Thus, we take the Frobenius norm (it is equal to the Euclidean distance) between $\boldsymbol{T}^*$ and $\boldsymbol{I}$ as the similarity measure of face images $\boldsymbol{A}$ and $\boldsymbol{B}$. TSM between $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as

$$\mathrm{TSM}(\boldsymbol{A},\boldsymbol{B})=1-f(\|\boldsymbol{T}^*-\boldsymbol{I}\|_{\mathrm{F}}^2), \tag{5}$$

where $f(\cdot)$ represents the normalization function normalizing $\|\boldsymbol{T}^*-\boldsymbol{I}\|_{\mathrm{F}}^2$ into [0, 1]. Considering human habits and the sigmoid function (which can imitate inputs and outputs of the human brain), we set $f(t)=2/(1+\mathrm{e}^{-t})-1$, $t\in[0,+\infty)$. Thus, the concrete TSM can be calculated by

$$\mathrm{TSM}(\boldsymbol{A},\boldsymbol{B})=2-\frac{2}{1+\exp(-\|\boldsymbol{T}^*-\boldsymbol{I}\|_{\mathrm{F}}^2)}. \tag{6}$$

Obviously, the larger the TSM($\boldsymbol{A}$, $\boldsymbol{B}$), the more similar the image matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The complete algorithm is outlined in Algorithm 1.

Fig. 2 shows the TSM obtained from one image and the images from various classes. Here, we randomly select one facial image from a certain class of the Extended Yale B face database as image $\boldsymbol{B}$ and one good-condition facial image from each class of the Extended Yale B face database as image $\boldsymbol{A}$ (totally 38 individual facial images), and calculate the TSM between $\boldsymbol{B}$ and $\boldsymbol{A}$. Fig. 2 demonstrates that the maximum TSM belongs to the homogeneous image of $\boldsymbol{B}$ and is significantly larger than other TSMs between image $\boldsymbol{B}$ and its heterogeneous images. This result further verifies the validity of the proposed TSM. Additionally, Fig. 2 shows that the TSMs between image $\boldsymbol{B}$ and its heterogeneous images are not zeros. This indicates that all face images have a certain similarity.

Algorithm 1 Transformation similarity measure

Input: two image matrices $\boldsymbol{A}$ and $\boldsymbol{B}\in\mathbb{R}^{m\times n}$

Output: TSM($\boldsymbol{A}$, $\boldsymbol{B}$)

Main procedure:

1. Solve Eq. (1) and obtain linear transformation $\boldsymbol{T}\in\mathbb{R}^{m\times m}$
2. Perform singular value decomposition of $\boldsymbol{T}$
3. Remove singular values by Eq. (3) and obtain $\boldsymbol{T}^*$
4. Obtain TSM($\boldsymbol{A}$, $\boldsymbol{B}$) by Eq. (6)

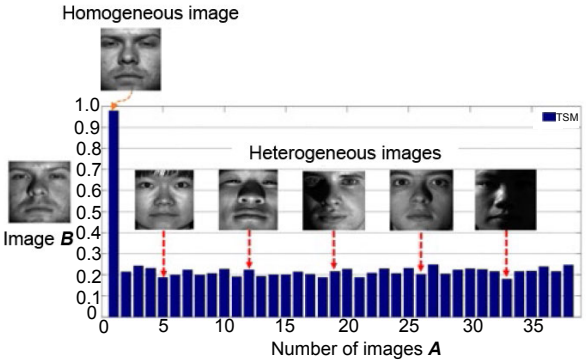

Fig. 2 Illustration of the proposed transformation similarity measure (TSM)

3 Model analysis

3.1 Mathematical features

In this subsection, we analyze the mathematical features of TSM. In general, the definition of the similarity measure function $S(\boldsymbol{A},\boldsymbol{B})$ should satisfy the following rules:

1. $S(\boldsymbol{A},\boldsymbol{A})=1$ (self-similarity).

Obviously, if $\boldsymbol{B}=\boldsymbol{A}$ then $\boldsymbol{T}=\boldsymbol{I}$ and $\boldsymbol{T}^*=\boldsymbol{I}$. Then TSM($\boldsymbol{A}$, $\boldsymbol{A}$)=1. So, the proposed TSM satisfies the self-similarity rule.

2. $S(\boldsymbol{A},\boldsymbol{B})\geq 0$ (non-negativity).

By the definition of TSM, we have 0<

TSM($\boldsymbol{A}$, $\boldsymbol{B}$)≤1. Thus, the proposed TSM satisfies the non-negativity rule.

3. $S(\boldsymbol{A}, \boldsymbol{B})=S(\boldsymbol{B}, \boldsymbol{A})$ (symmetry).

For TSM($\boldsymbol{A}$, $\boldsymbol{B}$), setting $\boldsymbol{T}_{AB}=\boldsymbol{A}\boldsymbol{B}^{-1}$ and $\boldsymbol{A}\boldsymbol{B}^{-1}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ with SVD, we have $\boldsymbol{T}^{*}=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$. Likewise, $\boldsymbol{T}_{BA}=\boldsymbol{B}\boldsymbol{A}^{-1}=(\boldsymbol{A}\boldsymbol{B}^{-1})^{-1}=\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}$ for TSM($\boldsymbol{B}$, $\boldsymbol{A}$). Performing SVD on $\boldsymbol{T}_{BA}$, we obtain $\boldsymbol{B}\boldsymbol{A}^{-1}=\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}$ and $\boldsymbol{T}_{BA}^{*}=\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}$. Knowing $\|\boldsymbol{T}_{AB}^{*}-\boldsymbol{I}\|_{\mathrm{F}}^{2}=\|\boldsymbol{T}_{BA}^{*}-\boldsymbol{I}\|_{\mathrm{F}}^{2}$, we have TSM($\boldsymbol{A}$, $\boldsymbol{B}$)=TSM($\boldsymbol{B}$, $\boldsymbol{A}$). So, the proposed TSM satisfies the symmetry rule.

In summary, from the mathematical point of view, the proposed TSM can be seen as a reasonable similarity measure model.

3.2 Advantage features for face recognition

Now we further analyze TSM in face recognition tasks. In general, the proposed TSM provides good features and innovative viewpoints for face image similarity measure and recognition as follows:

1. TSM is a two-dimensional (2D) holistic similarity measure method without parameters, and can be easily achieved by SVD. Without the parameter tuning and learning processes, TSM is time-saving and, more importantly, can avoid the embarrassment that happens in many distance learning methods. Compared with traditional distance-based methods (e.g., Euclidean distance, correlation coefficient, and angle-based distance), 2D TSM contains structural similarity of the images and is more suitable for image classification.

2. Like the cosine distance, TSM is not sensitive to the scaling changes, i.e.,

$$\mathrm{TSM}(\boldsymbol{A}, \boldsymbol{B})=\mathrm{TSM}(a\boldsymbol{A}, b\boldsymbol{B}), \tag{7}$$

where a and b are the scaling factors for all pixels of one image.

Fig. 3 shows the appearance of one image with the scaling factor changing. In this case, traditional methods, such as Euclidean distance and Mahalanobis distance, cannot measure the similarity, while TSM still works well.

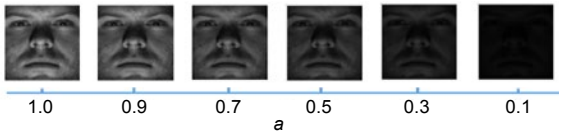

Fig. 3 Appearance of one image with the scaling factor changing

3. TSM can weaken the influence of illumination variation because the singular values associated with illumination variation are discarded during its calculation process. Here, we take an example to reveal this feature. In this example, we select two facial images with different illumination conditions from a certain class of the Extended Yale B face database as image $\boldsymbol{B}$, randomly select one good-condition facial image from each class of the Extended Yale B face database as image $\boldsymbol{A}$ (totally 38 individual facial images), and calculate the TSM between $\boldsymbol{B}$ and $\boldsymbol{A}$. Fig. 4 shows that the maximum TSM belongs to the homogeneous image $\boldsymbol{B}$ and is significantly larger than other TSMs between image $\boldsymbol{B}$ and its heterogeneous images. This verifies the validity of the proposed TSM in coping with the illumination variation problem. In particular, Fig. 4b shows that TSM handles extreme illumination variation conditions well.

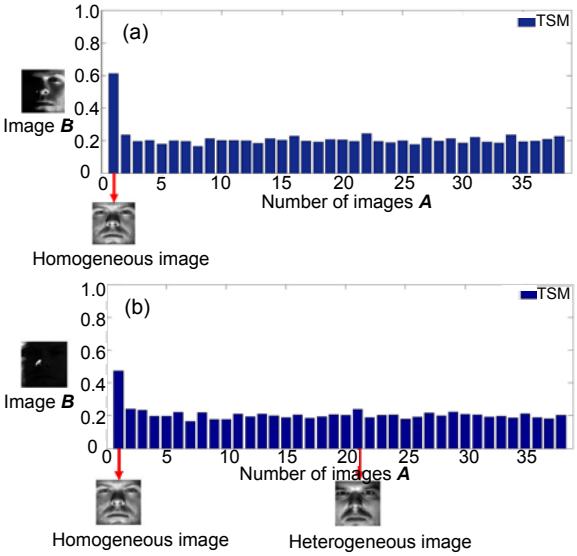

Fig. 4 An intuitive illustration of the insensitivity of TSM to the illumination variation conditions: (a) extreme illumination condition 1; (b) extreme illumination condition 2

4. TSM is insensitive to non-face block occlusion conditions. Eq. (1) can be rewritten as

$$\boldsymbol{T}=\arg\min_{\boldsymbol{T}}\sum_{i=1}^{n}\|\boldsymbol{a}_i-\boldsymbol{T}\boldsymbol{b}_i\|_2^2. \tag{8}$$

Suppose that image $\boldsymbol{B}$ is blocked by $\boldsymbol{C}\in\mathbb{R}^{m\times t}$ in the column direction and that $t\leq n$. Then the blocked

image can be denoted as $\bar{\boldsymbol{B}}=[\boldsymbol{b}_1,...,\boldsymbol{b}_s,\boldsymbol{c}_1,...,\boldsymbol{c}_t,\boldsymbol{b}_{s+t+1},...,\boldsymbol{b}_n]$ and the linear transformation from $\bar{\boldsymbol{B}}$ to $\boldsymbol{A}$ can be calculated by

$$\boldsymbol{T}=\arg\min_{\boldsymbol{T}}\left(\sum_{i=1}^{s}\|\boldsymbol{a}_i-\boldsymbol{T}\boldsymbol{b}_i\|_2^2+\sum_{j=1}^{t}\|\boldsymbol{a}_{s+j}-\boldsymbol{T}\boldsymbol{c}_j\|_2^2+\sum_{k=s+t+1}^{n}\|\boldsymbol{a}_k-\boldsymbol{T}\boldsymbol{b}_k\|_2^2\right). \tag{9}$$

From Eq. (9), it can be seen that the transformation $\boldsymbol{T}$ should balance the errors between the remaining and blocked parts. Actually, the transformation tends to be constant when $\boldsymbol{C}$ is not a face image since it can be treated as the transformation from the non-face subspace to the face subspace. Thus, we can obtain an effective similarity measure for the recognition task according to the remaining part of $\boldsymbol{B}$.

We show an example of this feature. Here, we select one image from a certain class of the Extended Yale B face database as image $\boldsymbol{B}$ and randomly select one good-condition facial image from each class of the Extended Yale B face database as image $\boldsymbol{A}$ (totally 38 individual facial images). Then image $\boldsymbol{B}$ is blocked by a baboon image of different proportions from the Internet. We calculate the TSMs and classify the test image with the nearest neighbor rule. Fig. 5 shows the ratios between the largest and second-largest TSMs as well as the classification results with the variations of the occlusion proportion. From Fig. 5 we can see that image $\boldsymbol{B}$ is still correctly classified and that the maximum TSM is significantly larger than that when the occlusion proportion is 80%. This verifies the validity of the proposed TSM in dealing with the block occlusion problem.

5. One more specific problem is that whether the inverse image (Fig. 6) can be used to describe identity. Here, the inverse image refers to the image obtained by the inverse operation of the original image matrix. This has potential application in the field of image encryption. Generally, it is hard to believe that the inverse image contains identity information. Indeed, almost all the existing methods have poor performances with the inverse image. However, by derivation, we can obtain

$$\mathrm{TSM}(\boldsymbol{A}^{\mathrm{T}},\boldsymbol{B}^{\mathrm{T}})=\mathrm{TSM}(\boldsymbol{A}^{-1},\boldsymbol{B}^{-1}). \tag{10}$$

Eq. (10) means that the similarity between the inverse images is equal to the similarity of the original images. In Section 4, experimental results show that the proposed TSM performs well on the inverse image recognition problem.

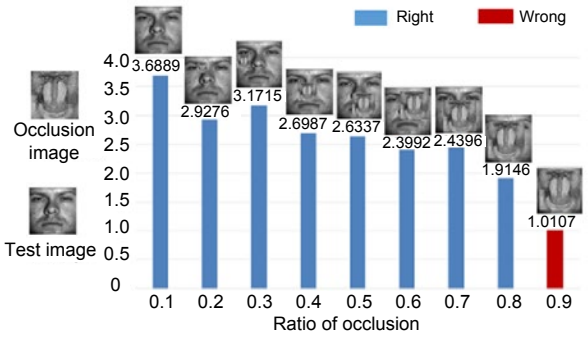

Fig. 5 Intuitive illustration of the insensitivity of TSM to the block occlusion condition

Blue bars mean that the test image is correctly classified and red bars mean misclassification. Numbers above the bars indicate the ratios between the largest and the second-largest TSMs. References to color refer to the online version of this figure

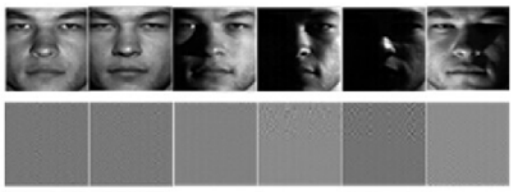

Fig. 6 Images of one person from the Extended Yale B database and their inverse images

The top row shows the original images and the bottom row shows the corresponding inverse images

3.3 Computational complexity of face recognition

Computational complexity is commonly used to measure the pros and cons of a model. The computational speed of TSM is often slow with the involvement of the SVD process. In theory, SVD can be achieved by eigen-decomposition. The computational complexity of the general SVD algorithm is about $O(n^3)$, where n is the rank of one matrix. The parallel SVD is more efficient in practical calculation. For specific algorithms, readers can refer to Demmel and Kahan (1990). In addition, the rank of the practical face recognition image is generally so small that we can quickly calculate it on a personal computer. We will compare the time consumptions of CPU of several similarity measure methods in Section 4.2, and

the results show that the proposed TSM can still meet the practical requirements on larger known datasets.

4 Experiments

In this section we present various experiments on public available databases for face recognition, which demonstrate the efficacy of the proposed algorithm.

Since the proposed TSM does not include a learning process, we compare the proposed TSM just with some classical and fundamental distance-based similarity measure methods in face recognition: Euclidean distance, Manhattan distance, angle distance (cosine distance), Mahalanobis distance, nuclear norm, Hausdorff distance, and LRM. Details of these methods can be found in Yambor et al. (2002), Vivek and Sudha (2007), Gu et al. (2012), and Zhang J and Yang (2014). Here, the nearest neighbor (NN) rule is used for classification (Cover and Hart, 1967). For LRM, we take the coefficient as the similarity indicator, whose performance is equal to that of the sparse representation based classifier (SRC), which was reported as one of the best methods in the Extended Yale B and AR databases (Zhang J and Yang, 2014). The solution tools and parameter settings follow Zhang J and Yang (2014)'s suggestions. Note that all experiments are done on the original face images without any image preprocessing or feature extraction step. In our experiments, data is randomly permuted 20 times, and thus all the results are reported as the average. Experiments are carried out on a personal computer (CPU: Intel® Core™ i7-4790 2.66 GHz; RAM: 16 GB).

4.1 Datasets

Four databases are involved. The details are given in the following:

1. Extended Yale B database

The Extended Yale B database contains about 2414 frontal face images of 38 individuals (Lee et al., 2005). We use the cropped and normalized face images (marked with P00) of size 80×80, which were captured under various laboratory-controlled lighting conditions with only small changes in the head pose and facial expression. Example images of one person are shown in Fig. 7.

Fig. 7 Samples of a person under different illumination conditions in the Extended Yale B face database

2. AR face database

The AR face database consists of over 3000 face images of 126 individuals (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions, and occlusions (Martínez and Benavente, 1998). There are 26 images of each individual, taken on two different occasions (i.e., two sessions separated by two weeks). We randomly select 120 individuals for our experiments. We manually crop the face portion of the image and then normalize it to 45×45 pixels. The normalized images of one person are shown in Fig. 8.

3. CMU PIE face database

The CMU PIE face database consists of 337 different subjects taken in four sessions with simultaneous variations in pose, expression, and illumination (Gross et al., 2010). In our experiments, the subset contains images of pose C27 (a nearly frontal pose) of 68 persons, each with 21 different direction illuminated images (Fig. 9). All the images are manually aligned, cropped, and resized to 64×64 pixels.

4. CUHK sketch face database

The CUHK sketch face database contains 188 persons (376 facial images) from the Chinese

Fig. 8 Samples of a person in the AR face database

The top row illustrates samples from session 1, and the bottom row illustrates samples from session 2

University of Hong Kong student database (Wang XG and Tang, 2009). For each person, there is a sketch drawn by an artist based on a photo taken in a frontal pose, under the normal lighting condition, and with a neutral expression (Fig. 10).

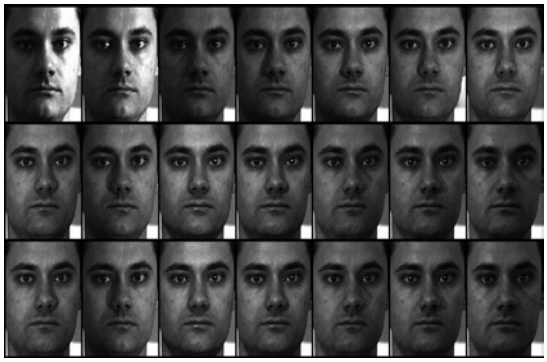

Fig. 9 Illustration of one person from the subset pose C27 in the CMU PIE face database

Fig. 10 Illustration of some photos and the corresponding sketch face images from the Chinese University of Hong Kong

4.2 Recognition with general conditions

In the first experiment, we use the Extended Yale B database to test the performance of TSM with the known variation of the sample number. The randomly selected k (k=1, 4, 8, 16, and 32) images of each subject are used as the known sample set and the remaining images for testing. Table 1 shows the average recognition rates and the test sample running time with the known sample sizes of TSM compared with the Euclidean distance, Manhattan distance, angle distance, Mahalanobis distance, nuclear norm, Hausdorff distance, and LRMs.

As shown in Table 1, the proposed TSM always obtains the highest recognition rate irrespective of the variation of the sample set size. Specifically, with the small sample size, such as k=1 and 4, TSM achieves recognition rates of over 70% and 90%, respectively, which are about 52% and 25% higher than those of LRM-L_2, clearly showing the advantages of TSM in face recognition. In terms of computational complexity, even though the performance of TSM is not the worst, we must admit that there exists a gap between the proposed TSM and practical applications.

4.3 Recognition with the inverse of face image

In theory, the inverse of image can be used to describe the identification performance of our proposed TSM model. In this experiment, we use the inverse images (Fig. 6) with an experiment setting similar to that in the first experiment in Section 3.2 to test the performance of the proposed algorithm.

Table 2 shows the recognition rates of Euclidean distance, Manhattan distance, angle distance, Mahalanobis distance, nuclear norm, Hausdorff distance, LRMs, and our proposed TSM. From Table 2 we can see that the proposed algorithm shows excellent performance irrespective of the variation of the known sample set size, but all other methods fail. As we have demonstrated in Section 3.1, the TSM between inverse images is equal to that of the original images. Although the appearance of the inverse image has

Table 1 Average recognition rate and test sample running time on the Extended Yale B database with known sample size variation

Similarity measure	Average recognition rate (%)					Test sample running time (ms)				
	k=1	k=4	k=8	k=16	k=32	k=1	k=4	k=8	k=16	k=32
Euclidean distance	9.67	35.21	61.42	66.93	69.73	3	11	27	45	99
Manhattan distance	9.58	33.92	57.31	62.71	69.14	6	17	41	73	137
Nuclear norm	14.82	57.48	80.00	84.26	86.24	68	162	402	786	1342
Angle distance	13.18	44.36	79.75	86.41	90.12	17	28	47	88	157
Hausdorff distance	6.78	18.30	35.58	52.17	79.86	257	694	1456	2238	3967
Mahalanobis distance	14.36	45.16	82.51	87.65	89.18	637	234	4783	8876	15 736
LRM-L1 (SRC)	19.18	63.17	84.68	92.03	95.06	105	113	122	340	253
LRM-L2 (LRC)	21.46	66.23	86.43	92.67	95.51	0.37	0.56	0.72	1.19	1.32
Proposed TSM	73.72	91.30	96.80	98.41	98.95	139	447	834	1521	2986

obscured its actual identity and is disorganized, TSM can still measure its similarity. However, other methods, no matter based on the image structure or pixel value, lose their measurement function. For example, obviously, $\|\boldsymbol{A}-\boldsymbol{B}\|_{\mathrm{F}}$ is not equal to $\|\boldsymbol{A}^{-1}-\boldsymbol{B}^{-1}\|_{\mathrm{F}}$. This feature suggests that TSM may have application in the field of image encryption.

Table 2 Average recognition rate on the Extended Yale B database with inverse images

Similarity measure	Average recognition rate (%)				
	k=1	k=4	k=8	k=16	k=32
Euclidean distance	2.43	2.51	2.68	2.34	2.71
Manhattan distance	2.54	2.86	3.02	3.21	3.24
Nuclear norm	1.26	2.38	2.28	2.74	2.62
Angle distance	2.43	2.28	2.54	2.67	2.68
Hausdorff distance	1.12	2.28	2.18	2.24	2.37
Mahalanobis distance	3.15	3.26	3.46	4.25	5.38
LRM-L_1 (SRC)	3.47	3.61	3.65	4.17	4.28
LRM-L_2 (LRC)	3.48	3.62	3.65	4.12	4.32
Proposed TSM	71.96	92.91	97.13	97.96	98.16

4.4 Recognition under different illumination conditions

In this subsection, we test the proposed method under various lighting conditions. In the first experiment, the Extended Yale B database is divided into five subsets of different illumination conditions. We use subset 1 consisting of 266 images (seven images per subject) under the nominal lighting condition as the known sample set, and all others for testing. Subsets 2 and 3 each contain 12 images per subject and are characterized by slight to moderate luminance variations. Subset 4 (14 images per subject) and subset 5 (19 images per subject) depict severe illumination variations. Results are shown in Fig. 11.

Fig. 11 shows that the proposed method achieves excellent performance for either moderate or severe lighting variations and obtains the highest recognition rate for almost all subsets. In particular, for both subsets 4 and 5 with extreme lighting conditions, our proposed TSM achieves the highest rates of 81.3% and 75.4%, respectively. Some robust methods like LRM-L_1 and LRM-L_2 do not seem to be robust to extreme illumination changes.

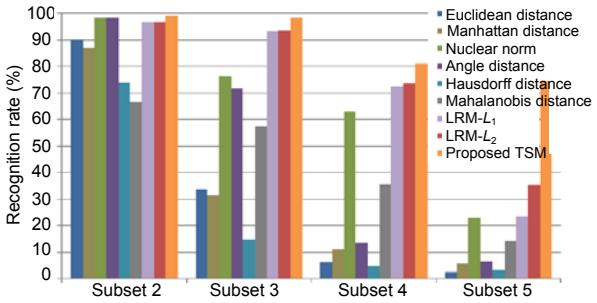

Fig. 11 Recognition rate on the Extended Yale B database under various lighting conditions

We conduct the second experiment on the subset of the CMU PIE face database containing images of pose C27 (a nearly frontal pose) of 68 persons, each with 21 different direction illuminated images (Fig. 9). In our experiment, some images of each subject are randomly selected for the known sample set, and the remaining images for testing.

Fig. 12 presents the average recognition rates on the CMU PIE face database. Clearly, in all cases, the proposed method achieves the best results. Specifically, for the single sample problem, TSM can still obtain an 83.21% recognition rate, about 20% higher than that of the second-highest LRM-L_1. Other methods like LRMs and nuclear norm achieve competitive results. This is possibly because the CMU PIE face database has more moderate lighting conditions than Extended Yale B and these reconstruction-based similarity measure methods are insensitive to relatively slight illumination changes.

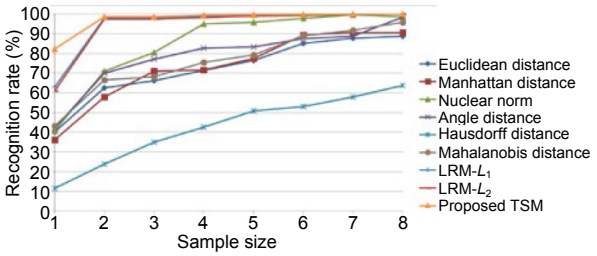

Fig. 12 Average recognition rate on the CMU PIE face database

4.5 Recognition with occlusions

In the first experiment, we use an experimental setting similar to that in Zhang and Yang (2014) to test the performance of the proposed TSM. Subsets 1 and 2 of Extended Yale B are used as the known sample sets and subset 3 for testing (these subsets

have been described in Section 4.1). Each test image is corrupted by the randomly located square block of a "baboon" image. The block size determines the occlusion level of an image. Fig. 13 shows these images with the occlusion level varying from 10% to 90%. Fig. 14 shows the recognition rates under different occlusion levels.

Fig. 14 shows that the proposed method significantly outperforms other similarity measure methods when the occlusion level is equal to or larger than 10%. The recognition rate of TSM with the nearest neighbor classifier drops slowly with the increase of the occlusion level, and thus the proposed method is insensitive to the level of structural noise. Note that the proposed method can still achieve an over 90% recognition rate when the occlusion level is 70%.

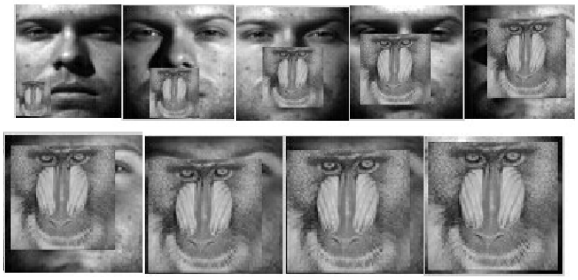

Fig. 13 Illustration of some images with the occlusion level varying from 10% to 90%

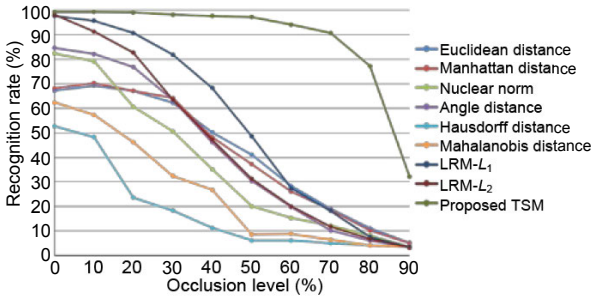

Fig. 14 Recognition rate under different occlusion levels

Now we test the TSM performance in real face disguise on the AR database. Different from Chen et al. (2015), we test these methods with more challenging conditions. For each subject, we randomly select seven images from 14 images with only illumination change and expressions as known samples. Others are divided into two separate subsets (with sunglasses and scarves, six samples per subject per session) for testing.

From Fig. 15, we observe that the proposed method can achieve the highest recognition rates for both tests with a scarf and sunglasses. For the test with sunglasses, all methods can achieve good results because the occlusion level is relatively low. There is no significant performance difference between the proposed method and others. For the test with a scarf, the proposed method significantly outperforms others, but does not achieve the same prime performance as in the first experiment. We believe that the mediocre performance of the proposed method is caused by the irregular nature of the occlusions in this experiment.

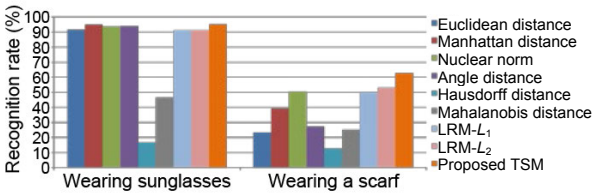

Fig. 15 Average recognition rate on the AR database with real face disguise

4.6 Experiments on the CUHK sketch face database

We test the performance of the proposed TSM on the CUHK sketch face database (Wang XG and Tang, 2009) compared with Euclidean distance, Manhattan distance, angle distance, Mahalanobis distance, nuclear norm, Hausdorff distance, LRMs, and the sketch transform method (STM) proposed by Wang XG and Tang (2009). In this experiment, we adopt the same experimental setting as in Wagner et al. (2012). We compute the recognition rates with the feature space dimension of 50×50 obtained by the down-sample method. The recognition rates are shown in Fig. 16.

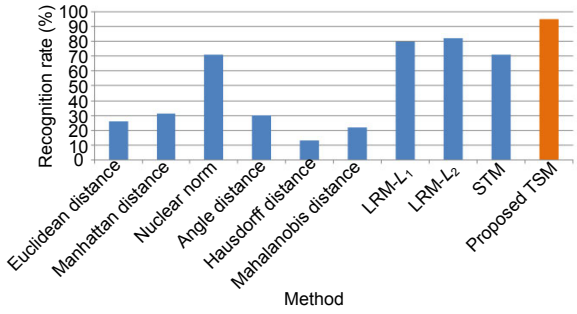

Fig. 16 Recognition rate on the CUHK sketch face database

As shown in Fig. 16, it is surprising that our proposed TSM achieves the highest 95% recognition rate versus 26% of Euclidean distance, 31% of

Manhattan distance, 30% of angle distance, 22% of Mahalanobis distance, 71% of nuclear norm, 13% of Hausdorff distance, 80% of LRM-L_1, 82% of LRM-L_2, and 71% of STM. The results show that TSM has the potential to deal with the heterogeneous face image recognition problem.

5 Conclusions and future work

It is important to improve the robustness of similarity measure in face recognition. We have studied the similarity between face image identities from the perspective of image matrix transformation. We believe that some transformations are not related to the identity information portrayed by the image. Based on this assumption and the nature of singular value decomposition, we have proposed the transformation similarity measure (TSM) to enhance the robustness of the metric method under lighting and occlusion conditions. The proposed method is both simple and practical since it does not involve learning process. Experimental results indicated that the proposed TSM achieves encouraging performance compared with the general similarity measure methods with respect to some robustness issues, such as alternating illumination, structural noise caused by occlusion, and heterogeneous pattern.

Our future work includes mainly applying TSM to other image-based applications, such as nature image detection and recognition tasks. In addition, our proposed TSM model uses linear transformation to capture the intrinsic similarity of two face images. Whether this model is effective against more complex conditions or how to extend the model for nonlinear cases such as pose variations needs further investigation.

Contributors

Jiang ZHANG and Heng ZHANG developed the idea of this study. Jian ZHANG designed the research. Jian ZHANG, Li-ling BO, Hong-ran LI, and Shuai XU processed the data. Jian ZHANG drafted the manuscript. Heng ZHANG and Dong-qing Yuan helped organize the manuscript. Jian ZHANG and Heng ZHANG revised and finalized the paper.

Compliance with ethics guidelines

Jian ZHANG, Heng ZHANG, Li-ling BO, Hong-ran LI, Shuai XU, and Dong-qing YUAN declare that they have no conflict of interest.

References

Adini Y, Moses Y, Ullman S, 1997. Face recognition: the problem of compensating for changes in illumination direction. *IEEE Trans Patt Anal Mach Intell*, 19(7):721-732. https://doi.org/10.1109/34.598229

Bowyer KW, Phillips PJ, 1998. Empirical Evaluation Techniques in Computer Vision. IEEE Computer Society Press, Washington, USA.

Chen SW, Dai XL, Pan BB, et al., 2015. A novel discriminant criterion based on feature fusion strategy for face recognition. *Neurocomputing*, 159:67-77. https://doi.org/10.1016/j.neucom.2015.02.019

Cover T, Hart P, 1967. Nearest neighbor pattern classification. *IEEE Trans Inform Theory*, 13(1):21-27. https://doi.org/10.1109/TIT.1967.1053964

Demirel H, Anbarjafari G, Jahromi MNS, 2008. Image equalization based on singular value decomposition. Proc 23rd Int Symp on Computer and Information Sciences, p.1-5. https://doi.org/10.1109/ISCIS.2008.4717878

Demmel J, Kahan W, 1990. Accurate singular values of bidiagonal matrices. *SIAM J Sci Stat Comput*, 11(5):873-912. https://doi.org/10.1137/0911052

Deng WH, Hu JN, Guo J, 2012. Extended SRC: undersampled face recognition via intraclass variant dictionary. *IEEE Trans Patt Anal Mach Intell*, 34(9):1864-1870. https://doi.org/10.1109/TPAMI.2012.30

Georghiades AS, Belhumeur PN, Kriegman DJ, 2001. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans Patt Anal Mach Intell*, 23(6):643-660. https://doi.org/10.1109/34.927464

Gross R, Matthews I, Cohn J, et al., 2010. Multi-PIE. *Image Vis Comput*, 28(5):807-813. https://doi.org/10.1016/j.imavis.2009.08.002

Gu ZH, Shao M, Li LY, et al., 2012. Discriminative metric: Schatten norm vs. vector norm. Proc 21st Int Conf on Pattern Recognition, p.1213-1216.

He XT, Peng YX, Zhao JJ, 2019. Fast fine-grained image classification via weakly supervised discriminative localization. *IEEE Trans Circ Syst Video Technol*, 29(5):1394-1407. https://doi.org/10.1109/TCSVT.2018.2834480

Lee KC, Ho J, Kriegman DJ, 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans Patt Anal Mach Intell*, 27(5):684-698. https://doi.org/10.1109/TPAMI.2005.92

Liu CJ, 2014. Discriminant analysis and similarity measure. *Patt Recogn*, 47(1):359-367. https://doi.org/10.1016/j.patcog.2013.06.023

Liu Y, Jin R, 2006. Distance Metric Learning: a Comprehensive Survey. Michigan State University, p.4.

Martínez A, Benavente R, 1998. The AR Face Database. 24 CVC Technical Report.

Naseem I, Togneri R, Bennamoun M, 2010. Linear regression for face recognition. *IEEE Trans Patt Anal Mach Intell*, 32(11):2106-2112.

https://doi.org/10.1109/TPAMI.2010.128

Paredes R, Vidal E, 2006. Learning weighted metrics to minimize nearest-neighbor classification error. *IEEE Trans Patt Anal Mach Intell*, 28(7):1100-1110. https://doi.org/10.1109/TPAMI.2006.145

Peng YX, He XT, Zhao JJ, 2017. Object-part attention model for fine-grained image classification. *IEEE Trans Image Process*, 27(3):1487-1500. https://doi.org/10.1109/TIP.2017.2774041

Peng YX, Qi JW, Yuan YX, 2018. Modality-specific cross-modal similarity measurement with recurrent attention network. *IEEE Trans Image Process*, 27(11):5585-5599. https://doi.org/10.1109/TIP.2018.2852503

Perlibakas V, 2004. Distance measures for PCA-based face recognition. *Patt Recogn Lett*, 25(6):711-724. https://doi.org/10.1016/j.patrec.2004.01.011

Sebe N, Lew MS, Huijsmans DP, 2000. Toward improved ranking metrics. *IEEE Trans Patt Anal Mach Intell*, 22(10):1132-1143. https://doi.org/10.1109/34.879793

Sun YY, Tong F, Zhang ZK, et al., 2018. Throughput modeling and analysis of random access in narrowband Internet of Things. *IEEE Int Things J*, 5(3):1485-1493. https://doi.org/10.1109/JIOT.2017.2782318

Vivek EP, Sudha N, 2007. Robust Hausdorff distance measure for face recognition. *Patt Recogn*, 40(2):431-442. https://doi.org/10.1016/j.patcog.2006.04.019

Wagner A, Wright J, Ganesh A, et al., 2012. Toward a practical face recognition system: robust alignment and illumination by sparse representation. *IEEE Trans Patt Anal Mach Intell*, 34(2):372-386. https://doi.org/10.1109/TPAMI.2011.112

Wang H, 2006. Nearest neighbors by neighborhood counting. *IEEE Trans Patt Anal Mach Intell*, 28(6):942-953. https://doi.org/10.1109/TPAMI.2006.126

Wang XG, Tang XO, 2009. Face photo-sketch synthesis and recognition. *IEEE Trans Patt Anal Mach Intell*, 31(11): 1955-1967. https://doi.org/10.1109/TPAMI.2008.222

Wen Y, Zhang L, von Deneen KM, et al., 2016. Face recognition using discriminative locality preserving vectors. *Dig Signal Process*, 50:103-113. https://doi.org/10.1016/j.dsp.2015.11.001

Wright J, Yang AY, Ganesh A, 2009. Robust face recognition via sparse representation. *IEEE Trans Patt Anal Mach Intell*, 31(2):210-227. https://doi.org/10.1109/TPAMI.2008.79

Wu J, Shen H, Li YD, et al., 2013. Learning a hybrid similarity measure for image retrieval. *Patt Recogn*, 46(11):2927-2939. https://doi.org/10.1016/j.patcog.2013.04.008

Wu MC, He SB, Zhang YT, et al., 2019. A tensor-based framework for studying eigenvector multicentrality in multilayer networks. *PNAS*, 116(31):15407-15413. https://doi.org/10.1073/pnas.1801378116

Yambor WS, Draper BA, Beveridge JR, 2002. Analyzing PCA-based face recognition algorithms: eigenvector selection and distance measures. Empirical Evaluation Methods in Computer Vision, p.39-60. https://doi.org/10.1142/9789812777423_0003

Zhang DQ, Chen SC, Zhou ZH, 2005. A new face recognition method based on SVD perturbation for single example image per person. *Appl Math Comput*, 163(2):895-907. https://doi.org/10.1016/j.amc.2004.04.016

Zhang J, Yang J, 2014. Linear reconstruction measure steered nearest neighbor classification framework. *Patt Recogn*, 47(4):1709-1720. https://doi.org/10.1016/j.patcog.2013.10.018

Zhang J, Yang J, Qian JJ, 2015. Nearest orthogonal matrix representation for face recognition. *Neurocomputing*, 151: 471-480. https://doi.org/10.1016/j.neucom.2014.09.019

Zhang YM, He SB, Chen JM, 2016. Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks. *IEEE/ACM Trans Netw*, 24(3):1632-1646. https://doi.org/10.1109/TNET.2015.2425146

Zhou CW, Gu YJ, He SB, et al., 2018. A robust and efficient algorithm for coprime array adaptive beamforming. *IEEE Trans Veh Technol*, 67(2):1099-1112. https://doi.org/10.1109/TVT.2017.2704610

Zhuang LS, Yang AY, Zhou ZH, et al., 2013. Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.3546-3553. https://doi.org/10.1109/CVPR.2013.455

Zou WW, Yuen PC, 2010. Discriminability and reliability indexes: two new measures to enhance multi-image face recognition. *Patt Recogn*, 43(10):3483-3493. https://doi.org/10.1016/j.patcog.2010.05.024