Pattern Recognition 46 (2013) 45–56

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A complete and fully automated face verification system on mobile devices

Jianfeng Ren [a,*], Xudong Jiang [b], Junsong Yuan [b]

[a] BeingThere Centre, Institute for Media Innovation, Nanyang Technological University, 50 Nanyang Drive, Singapore 637553, Singapore
[b] Electrical & Electronic Engineering, Nanyang Technological University, Nanyang Link, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 8 November 2011
Received in revised form
16 April 2012
Accepted 20 June 2012
Available online 28 June 2012

Keywords:
Face verification system
Face detection
Eye detection
Subspace approach
Attentional cascade
Class-specific threshold

ABSTRACT

Mobile devices have been widely used not only as a communication tool, but also a digital assistance to our daily life, which imposes high security concern on mobile devices. In this paper we present a natural and non-intrusive way to secure mobile devices, i.e. a complete and fully automated face verification system. It consists of three sub-systems: face detection, alignment and verification. The proposed subspace face/eye detector locates the eyes at a much higher precision than Adaboost face/eye detector. By utilizing attentional cascade strategy, the proposed face/eye detector achieves a comparable speed to Adaboost face/eye detector in this "close-range" application. The proposed approach that determines the class-specific threshold without sacrificing the training data for the validation data further boosts the performance. The proposed system is systematically evaluated on O2FN, AR and CAS-PEAL databases, and compared with many different approaches. Compared to the best competitive system, which is built upon Adaboost face/eye detector and ERE approach for face recognition, the proposed system reduces the overall equal error rate from 8.49% to 3.88% on the O2FN database, from 7.64% to 1.90% on the AR database and from 9.30% to 5.60% on the CAS-PEAL database. The proposed system is implemented on O2 XDA Flame and on average it takes 1.03 s for the whole process, including face detection, eye detection and face verification.

1. Introduction

Mobile devices have been widely utilized in our daily life. A large amount of private information are stored on mobile devices, e.g. personal videos, pictures, documents or even financial transaction records. Thus, it imposes high security concerns on mobile devices. Traditionally, mobile devices are secured by password. Short password is easily attacked and long password is difficult to remember. In contrast, biometrics cannot be lost, stolen, shared, forgotten or easily duplicated. Particularly, face recognition has several advantages over others, e.g. it is natural, non-intrusive and easy to use. Fingerprint authentication system has been successfully implemented on mobile devices. However, it may be uncomfortable to the user to release fingerprint due to its usage in law force. Face recognition has less such concern. In addition, an embedded sensor is required to capture the fingerprint, whereas no additional hardware is required to capture the face image since currently most mobile phones are equipped with a camera.

The proposed face recognition system consists of three subsystems: face detection, alignment and verification. It has been shown that the performance of a fully automated face recognition system highly depends on face alignment accuracy [1]. Subspace approaches are widely used in face recognition and superior performance has been demonstrated in [2–12]. However, high precision of face alignment is required for subspace face recognition. Although we can provide guidance for the user to position the camera to get the face aligned, it is not an easy task to get the face precisely aligned even for an experienced user. For a complete face recognition system, it is not trivial to build a fast and accurate face/eye detector.

Mobile face detection for verification introduces both challenges and opportunities. On one side, mobile phones have limited computational power and memory compared to PC, and hence it imposes higher requirement on the detection speed. On the other side, as users are trying to pass the verification system, they intend to capture a large face in the self-taken photo. It results in a "close-range" face detection, where the face occupies a significant portion of the image. Different from general face detection, where millions of scanning windows are required, only a limited number of scanning windows are required for mobile face detection. Thus, it imposes lower requirement on the detection speed. The number of scanning windows for eye detection is also limited as eyes are detected within the face window. In fact, many face detection tasks belong to "close-range", e.g. in face tracking, after the first detection, the face can be detected within a small neighborhood in the subsequent frames.

* Corresponding author. Tel.: +65 6790 5018; fax: +65 6793 3318.
E-mail addresses: jfren@ntu.edu.sg (J. Ren), exdjiang@ntu.edu.sg (X. Jiang), jsyuan@ntu.edu.sg (J. Yuan).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.06.013

In this paper, a subspace face/eye detector is proposed. Subspace approaches have been successfully applied in object classification and have demonstrated impressive classification performance [2,13]. However, it is not widely used in object detection due to the speed limitation, e.g. it requires large matrix multiplication to derive subspace features. Inspired by Viola & Jones' face detector [14], an attentional cascade structure is proposed for subspace approach to speed up the detection process. Classifiers built from a few subspace features are utilized in the early stages to fast filter out most of the negative samples and only those tough samples are evaluated by classifiers with more subspace features in the latter stages. On O2 XDA Flame, it only takes 0.26 s to detect both eyes for the proposed face/eye detector. The proposed subspace face/eye detector is optimized for mobile face verification, which aims to achieve higher detection accuracy at a comparable speed with respect to Adaboost face/eye detector.

For face verification sub-system, a novel approach to determine the class-specific threshold (CST) is proposed, which greatly improves the performance of the system. Naturally, some faces are easy to be recognized and some are difficult. Therefore, a CST in general provides better performance than a global threshold (GT). Traditionally, a separate validation dataset is needed to determine the CST. However, separating a portion of the dataset as the validation dataset results in less training data, and hence the performance may degrade. The proposed approach determines the CST based on the training data only. The experimental results on O2FN, AR [15] and CAS-PEAL [16] databases demonstrate the superior performance of the proposed approach.

In order to evaluate the performance of the proposed system on mobile device, the O2FN database[1] is built. It contains 2000 images from 50 persons, each with 40 images. The variations in facial expressions are minimized. It mainly contains illumination variations and in-plane rotations. The database will be released to other researchers as a common evaluation dataset for face recognition system on mobile devices.

The proposed system is implemented on O2 XDA Flame, on which it takes 1.03 s to recognize a face. Comprehensive experimental results on O2FN, AR [15] and CAS-PEAL [16] databases demonstrate the superior performance of the proposed system compared to the system built upon Adaboost face/eye detector.

The rest of the paper is organized as follows: the related work is discussed in Section 2; the proposed algorithms are described in Section 3; the creation of the O2FN mobile face database is given in Section 4; experimental results are in Section 5; finally, conclusion and future work are in Section 6.

2. Related work

The related work on each sub-system, e.g. face detection, alignment and verification is discussed first, followed by a review of several complete systems.

In [17–21] Adaboost face detector [14] was built on mobile devices. The main advantage of Adaboost face detector is its speed, which is critical in general face detection. Due to the limited discrimination power of rectangular features, a huge amount of samples are required to train a Viola & Jones' face detector, and hence training is time consuming. The proposed face detector detects faces more precisely at a comparable speed with respect to Adaboost face detector on "close-range" face images. In addition, the proposed face detector has the advantages of fast training and easy implementation.

Face alignment is critical to subspace face recognition. The face is often aligned through eye locations. In [17,20] Adaboost eye detector was built. Tao & Veldhuis extended it to detect more fiducial points [19]. Eyes are detected within the detected face window, and hence the location and scale of eyes are roughly known. Only a limited number of scanning windows are needed. Thus, Adaboost eye detector may not be the best choice due to its limited accuracy.

Although tons of face verification algorithms are proposed in literature, very few are suitable for mobile applications. In [22], superior performance is achieved for SIFT features, but the recognition speed is extremely slow. Sparse representation [3] demonstrated impressive performance on recognizing partially occluded faces. However, due to the huge computational cost on iterative optimization, it is not suitable for mobile applications. Simple features, e.g. local binary pattern (LBP) [19,20], Haar wavelet features [23], DCT coefficients [17], Fourier coefficients [18] and LRP features [24] are extracted for classification on the resource-constraint devices. Correlation filter [21,25] is another option due to its simplicity and tolerance to noise. However, those approaches do not yield optimal recognition performance. Subspace face recognition approaches have demonstrated superior performance in [2–12]. Yang et al. [26] ported direct linear discriminant analysis [27] into mobile devices. In this work, various subspace approaches are implemented and tested on the O2FN database. The best performed one is built into our system.

Several complete face verification systems on mobile devices have been proposed in literature. Ng et al. implemented a real-time face verification system on O2 XDA IIi [21], using Viola & Jones' face detector and unconstrained minimum average correlation energy (UMACE) filter for face verification. It was tested on a database of 25 subjects and the overall equal error rate (EER) is 8.49%. Schneider et al. implemented FaceScry on Nokia 6630 [28], but they only performed a sampling check on their system. Hadid et al. implemented a system on Nokia N90 [20], using Viola & Jones' face/eye detector and LBP features for face verification. Tao & Veldhuis implemented a system on Eten M500 Pocket PC [19], using Viola & Jones' detector for face/facial landmark detection and a simplified LBP filter for face verification. It was tested on a database of 20 subjects, and the EER was 2% by fusing the results from 450 frames. However, it is not practical to take a video of 30 s for authentication. Adaboost face detector [19–21] and Adaboost eye detector [19,20] are popular because of its fast speed. UMACE filter [21], geometric features [28] and LBP features [19,20] are used for face verification because of their simplicity. However, they do not yield optimal recognition performance. In the following sections, we will show that the proposed system achieves much higher recognition accuracy at a comparable speed.

3. The proposed face verification system on mobile devices

The proposed face verification system consists of three subsystems: face detection, eye detection and face verification as shown in Fig. 1. In view of the asymmetric nature of face/eye detection, by studying the eigen-spectrum of the covariance matrices of face/non-face class and eye/non-eye class, a subspace face/eye detector is proposed. An attentional cascade structure is proposed to speed up the detector. Various subspace face recognition algorithms are studied and the one most suitable for mobile devices is built into our system with the proposed mechanism of class-specific decision.

3.1. Face detection

Asymmetric PCA (APCA): Face detection can be formulated as a two-class classification problem: face class and non-face class.

[1] "O2" stands for "O2" XDA Flame; "F" stands for "front" camera; "N" stands for "native" setting, e.g. no advance image processing enabled during capturing.

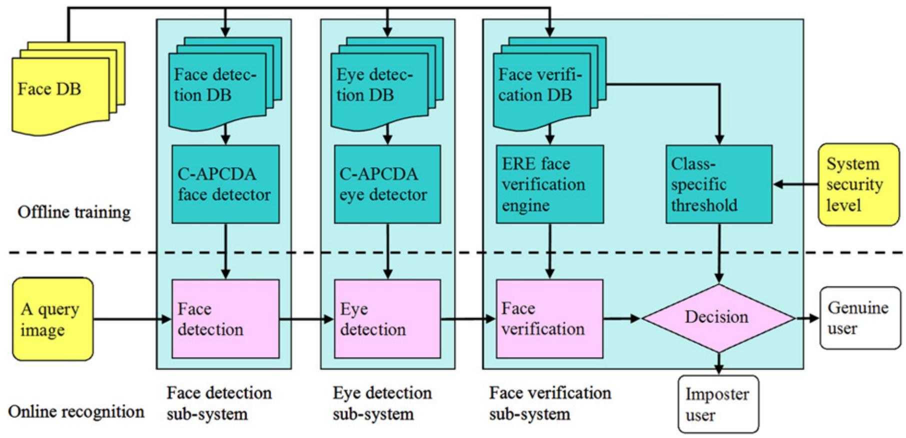

Fig. 1. Block diagram of the proposed face verification system on mobile devices.

Traditionally, PCA treats two classes equally. As pointed in [2,13], the role of PCA in pattern classification is to remove the unreliable dimensions that are harmful to accurate classification. Since the training samples of two classes may have different reliability, we should weight them according to the class reliability when applying PCA. The covariance mixture $\mathbf{\Sigma}_\alpha$ is defined as

$$\mathbf{\Sigma}_\alpha = \alpha\mathbf{\Sigma}_p + (1-\alpha)\mathbf{\Sigma}_q + \mathbf{\Sigma}_m \quad (1)$$

where $\mathbf{\Sigma}_p$, $\mathbf{\Sigma}_q$ are the covariance matrices of face/non-face class; $\mathbf{\Sigma}_m$ is the between-class scatter matrix; α is the weighting factor of $\mathbf{\Sigma}_p$ and $0 \le \alpha \le 1$.

Face detection is highly asymmetric in nature. In general face detection it is relatively easy to collect representative face samples, but not easy to collect representative negative samples, since the non-face is "everything else except face". However, in "close-range" face detection the non-face samples are well defined, e.g. mainly the cropped partial faces. As a result, $\mathbf{\Sigma}_q$ can be estimated as reliable as $\mathbf{\Sigma}_p$. Furthermore, much more non-face samples than face samples can be collected, and hence $\mathbf{\Sigma}_q$ can be more reliably estimated. Dimensions corresponding to the small eigenvalues of the covariance matrix of unreliable class should be removed. Therefore, we assign a large weight α to $\mathbf{\Sigma}_p$. PCA is applied on $\mathbf{\Sigma}_\alpha$, and the eigenvectors $\mathbf{\Phi}_m$ corresponding to first m largest eigenvalues are extracted. By assigning a large weight to $\mathbf{\Sigma}_p$, more dimensions corresponding to the unreliable small eigenvalues of $\mathbf{\Sigma}_p$ are removed, and hence better generalization performance is achieved.

Asymmetric discriminant analysis (*ADA*): APCA is effective to alleviate over-fitting problem by removing the unreliable dimensions of covariance mixture matrix. For a fast classification, discriminant analysis is necessary to extract a compact feature set from the reliable APCA subspace. Linear discriminant analysis (LDA) is commonly used to extract discriminant features. However, for two-class classification problem, only one feature can be extracted. Covariance discriminant analysis (CDA) [29] can provide more dimensions, but the asymmetric nature of training data reduces the effectiveness of CDA. In fact, these two approaches can be combined in one discriminant evaluation. In asymmetric principal and discriminant analysis (APCDA) algorithm [13], ADA is applied in the APCA subspace, which aims to maximize the Bhattacharyya distance [30] defined in the following equation:

$$D = \frac{1}{8}(\hat{M}_p - \hat{M}_q)^T \left(\frac{\hat{\mathbf{\Sigma}}_p + \hat{\mathbf{\Sigma}}_q}{2}\right)^{-1} (\hat{M}_p - \hat{M}_q) + \frac{1}{2} \ln \frac{|(\hat{\mathbf{\Sigma}}_p + \hat{\mathbf{\Sigma}}_q)/2|}{\sqrt{|\hat{\mathbf{\Sigma}}_p||\hat{\mathbf{\Sigma}}_q|}} \quad (2)$$

where $\hat{M}_p$, $\hat{M}_q$, $\hat{\mathbf{\Sigma}}_p$, $\hat{\mathbf{\Sigma}}_q$ are the mean vectors of the face/non-face class and the covariance matrices of the face/non-face class in APCA subspace, respectively.

The first term of Bhattacharyya distance is maximized by LDA in Eq. (3), where $\hat{\mathbf{\Sigma}}_m$ is between-class covariance matrix in APCA subspace:

$$\hat{\mathbf{\Sigma}}_m \mathbf{\Phi} = (\hat{\mathbf{\Sigma}}_p + \hat{\mathbf{\Sigma}}_q)\mathbf{\Phi}\mathbf{\Lambda} \quad (3)$$

It is proven in [31] that the second term of Bhattacharyya distance is maximized in the subspace spanned by the generalized eigenvectors corresponding to the largest $\lambda_k + 1/\lambda_k$, where λ_k is the generalized eigenvalue of matrix pair either $(\hat{\mathbf{\Sigma}}_p, \hat{\mathbf{\Sigma}}_q)$ or $(\hat{\mathbf{\Sigma}}_q, \hat{\mathbf{\Sigma}}_p)$, or equivalently $(\hat{\mathbf{\Sigma}}_p, \hat{\mathbf{\Sigma}}_p + \hat{\mathbf{\Sigma}}_q)$. However, covariance matrix $\hat{\mathbf{\Sigma}}_p, \hat{\mathbf{\Sigma}}_q$ are still biased even after APCA, thus the generalized eigenvalue problem is modified as in Eq. (4), where $0.5 < \beta \le 1$ is determined by prior knowledge about the asymmetry of the two classes:

$$\hat{\mathbf{\Sigma}}_p \mathbf{\Phi} = (\hat{\mathbf{\Sigma}}_p + \beta\hat{\mathbf{\Sigma}}_q)\mathbf{\Phi}\mathbf{\Lambda} \quad (4)$$

Then, Eqs. (3) and (4) are combined, e.g.

$$(\hat{\mathbf{\Sigma}}_p + \gamma\hat{\mathbf{\Sigma}}_m)\mathbf{\Phi} = (\hat{\mathbf{\Sigma}}_p + \beta\hat{\mathbf{\Sigma}}_q)\mathbf{\Phi}\mathbf{\Lambda} \quad (5)$$

where γ is a constant that weights the discriminatory information of the class mean against the covariance. ADA is applied in the APCA subspace and extracts d eigenvectors $\tilde{\mathbf{\Phi}}_d$ corresponding to the d largest $\max(\lambda_k, 1-\lambda_k)$, where λ_k is the generalized eigenvalue of Eq. (5).

Cascade APCDA (*C-APCDA*) *face detector*: Most non-face samples are significantly different from the face samples. It is possible to filter out a large number of non-face samples and preserve most face samples by utilizing a few APCDA dimensions. Thus, cascade APCDA (C-APCDA) is proposed to boost the detection speed. Easy samples can be processed extremely fast and only those tough samples will be evaluated at the latter stages with more dimensions.

After ADA, feature dimensions are sorted according to their discriminant power in the descending order. The proposed C-APCDA face detector consists of l stages. At stage i, it is an APCDA face detector Ω_i of d_i dimensions, $i = 1, 2, \ldots, l$. We choose first d_i eigenvectors from $\tilde{\mathbf{\Phi}}_d$ to construct the projection matrix $\mathbf{U}_i = [\tilde{\Phi}_1 \tilde{\Phi}_2 \ldots \tilde{\Phi}_{d_i}]$, where $\tilde{\Phi}_i$ is i-th eigenvector of $\tilde{\mathbf{\Phi}}_d$. As the stage number increases, the number of dimensions increases, and better classification performance is expected. Here, the face detector at each stage is constructed from the first d_i eigenvectors, as these feature dimensions are un-correlated. For each scanning window, we define its score [13] as

$$s_i = (\tilde{X}_i - \tilde{M}_q)^T (\beta\tilde{\mathbf{\Sigma}}_q)^{-1} (\tilde{X}_i - \tilde{M}_q) - (\tilde{X}_i - \tilde{M}_p)^T \tilde{\mathbf{\Sigma}}_p^{-1} (\tilde{X}_i - \tilde{M}_p) \quad (6)$$

where $\tilde{X}_i$, $\tilde{M}_p$, $\tilde{M}_q$, $\tilde{\mathbf{\Sigma}}_p$, $\tilde{\mathbf{\Sigma}}_q$ are testing sample, the mean vector of face/non-face class, the covariance matrix of face/non-face class in the APCDA subspace, respectively. If $s_i \geq T_j$, where T_j is the threshold at j-stage, it is a face sample, otherwise it is a non-face sample. The optimal threshold at each stage is determined as follows: assume the target false rejection rate (FRR) is ϵ_F, and each stage has equal FRR ϵ, then ϵ can be estimated by $(1-\epsilon)^l = 1-\epsilon_F$. The threshold at each stage is adjusted accordingly to achieve desired ϵ. As there is only one face in the image, at the last stage Ω_l, the scanning window with maximum score is detected as the face window.

The procedures to build the proposed C-APCDA face detector are summarized as follows:

1. Perform APCA on $\mathbf{\Sigma}_\alpha$ as defined in Eq. (1) to remove the unreliable dimensions.
2. Perform ADA in the APCA subspace, and choose generalized eigenvectors $\tilde{\mathbf{\Phi}}_d$ of Eq. (5) corresponding to d largest max $(\lambda_k, 1-\lambda_k)$.
3. Select the first d_i APCDA dimensions to construct the face detector Ω_i at i-stage of the C-APCDA face detector.
4. At the last stage Ω_l, the scanning window with maximum score as defined in Eq. (6) is detected as the face window.

3.2. Eye detection

Although the eye detector can be built in a similar approach to the C-APCDA face detector, there are some differences between these two. Eye detection is to precisely locate two eyes, while face detection is to roughly locate the face region. Eye localization should be as accurate as possible so that the face can be aligned precisely, whereas much larger margin is allowed for face detection. For eye detection, only the cropping window at the precise eye location can be used as an eye sample. The scale of eyes is roughly known since eyes are detected within the detected face window. Therefore, only one eye sample can be generated for each eye. In order to determine the precise eye locations, the scanning windows that are very close to the eyes are included in the non-eye database. Lastly, eye detection is more sensitive to obstacles, e.g. wearing or not wearing glasses is not critical to face detection, but it is critical to eye detection. Similarly, open eyes, half-open eyes or closed eyes greatly affect the appearance of the eyes. All these increase the intra-class variations.

Eye detection is modeled as a two-class classification problem, e.g. positive eye class and negative non-eye class. The non-eye is not "everything else except eye", but the scanning windows within the face region. Thus, the covariance matrix of the non-eye class can be estimated as reliable as that of the eye class. Furthermore, much more non-eye samples can be collected, and hence the covariance matrix of the non-eye class can be estimated more reliably. A larger weight should be put on the less reliable covariance matrix of the eye class so that more unreliable dimensions from the eye class can be removed.

The face is assumed to be almost upright. Thus, the left eye is detected in the left-up quarter and the right eye is detected in the right-up quarter. Right eyes are left–right flipped and only the left eye detector is built. The procedures similar to the C-APCDA face detector are adopted to build the proposed C-APCDA eye detector.

3.3. Face verification

Many face recognition algorithms have been proposed in literature. Due to the limited resources on mobile devices, many are not suitable for mobile applications. In contrast, subspace approaches require less memory and computational power, and work well on low-resolution images. Thus, they are more suitable for mobile applications. Subspace approaches have demonstrated state-of-the-art performance in face recognition [2,5–10]. Our face verification sub-system is based on subspace approaches.

Subspace face recognition can be formulated as global formulation or class-specific formulation. In global formulation, one unified projection matrix is trained for all the classes through constructing a pooled covariance matrix [32]. In class-specific formulation, a class-specific projection matrix is trained for each class [33]. Similar to the training process, to determine whether a user is a genuine user or an imposter user, either a global threshold (GT) or a class-specific threshold is needed. Now the question is: *Which combination of formulation and threshold should be used for face verification?* In class-specific formulation, the genuine model generated from the limited number of training samples of each class may not be able to fully capture high complexity of face manifold, and may be easily over-fit to the training data. Those small eigenvalues of the class-specific covariance matrix are largely deviated from their true values [2,13]. The inverses of those small unreliable eigenvalues are used to weight the features, which causes severe over-fitting problem. In contrast, in global formulation the within-class scatter matrix is constructed by summing up the class-specific covariance matrix of each class, and hence it can be better estimated. As a result, better generalization performance is achieved for global formulation. For class-specific formulation, n^2 (n is face dimensionality) free parameters of the class-specific covariance matrix need to be estimated from very few training samples of each class, whereas only one parameter needs to be estimated for the CST. Thus, CST will not cause severe over-fitting. Naturally, some faces are easy to be recognized, and hence a small threshold will be sufficient, whereas others may need a large threshold. Thus, CST in general provides better performance than GT. Therefore, we propose to train the face verification system based on global formulation and make the decision based on CST. In the latter section, a novel approach is proposed to determine the CST without sacrificing part of the training data as the validation data, which further boosts the performance of the system.

Subspace approaches share the same framework, e.g. originated from linear discriminant analysis. They aim to find the most discriminant dimensions by maximizing the Fisher ratio. Face dimensionality n is often larger than the number of samples N. $\mathbf{S}_w \in \Re^{n \times n}$ is singular since its rank $r \leq N-c$, where c is the number of class. Many approaches are proposed to solve this singularity problem. Fisherface [9] solves the singularity problem of $\mathbf{S}_w$ by applying PCA on $\mathbf{S}_t$ to reduce the feature dimensionality to $N-c$. However, discriminant information also resides in those dimensions being thrown away. Null space approach (NLDA) [10] only utilizes the null space of $\mathbf{S}_w$, but the discriminant information in range space is ignored. In fact, discriminant information resides in both subspaces. The features in the range space can be weighted by their discriminant power, e.g. $\phi_k^w / \sqrt{\lambda_k^w}$, where ϕ_k^w, λ_k^w are k-th eigenvector and eigenvalue of $\mathbf{S}_w$. However, for the features in the null space, $\lambda_k^w = 0$. It is not feasible to directly combine the features from those two subspaces. In addition, in the range space, only the eigenvectors corresponding to leading largest eigenvalues are reliable. Those dimensions corresponding to small eigenvalues are not reliable, but a large amount of discriminant information resides in those dimensions.

In eigenfeature regularization and extraction (ERE) approach [2,5], it not only alleviates the problem of unreliable small and zero eigenvalues caused by noise, but also enables discriminant evaluation to be performed in the full dimension of the image data under one unified discriminant criteria. The subspace spanned by eigenvector ϕ_k^w of $\mathbf{S}_w$ is decomposed into three subspaces: reliable range space $\mathbf{F} = \{\phi_k^w\}_{k=1}^m$, unreliable range space $\mathbf{N} = \{\phi_k^w\}_{k=m+1}^r$ and null space $\emptyset = \{\phi_k^w\}_{k=r+1}^n$, where r is the rank of $\mathbf{S}_w$ as defined early, and dimension ϕ_{m+1}^w is the starting dimension of unreliable range space. The face component in **F** decays rapidly and can be viewed as the outliers of the whole spectrum. Median operator works well in separating outliers from a dataset. To determine the start point of unreliable range space, we first find a point near the center of unreliable range space by $\lambda^w_{med} = median\{\forall\lambda^w_k | k \le r\}$. The distance between λ^w_k and the smallest nonzero eigenvalue is $d_{m,r} = \lambda^w_{med} - \lambda^w_r$. The upper bound of the unreliable eigenvalues is estimated by $\lambda^w_{med} + d_{m,r}$. More generally, λ^w_{m+1} can be estimated by

$$\lambda^w_{m+1} = \max\{\forall\lambda^w_k | \lambda^w_k < \lambda^w_{med} + \mu(\lambda^w_{med} - \lambda^w_r)\} \tag{7}$$

where μ is a constant and $\mu \approx 1$. The optimal μ may vary for different applications. For simplicity, $\mu = 1$ is used for all the experiments.

In reliable range space **F**, the eigenvalues remain unchange. In unreliable range space **N**, the new eigenvalues are determined by fitting the eigenvalues in reliable range space **F** into a nonlinear model, e.g. $\tilde{\lambda}^w_k = \alpha/(k+\beta)$. Although not limited from other possible fitting, α and β are determined by letting $\tilde{\lambda}^w_1 = \lambda^w_1$ and $\tilde{\lambda}^w_m = \lambda^w_m$, which yields

$$\alpha = \frac{\lambda^w_1 \lambda^w_m (m-1)}{\lambda^w_1 - \lambda^w_m}, \quad \beta = \frac{m\lambda^w_m - \lambda^w_1}{\lambda^w_1 - \lambda^w_m}$$

As a result, the over-fitting problem caused by those unreliable dimensions is alleviated. In null space $\emptyset$, the new eigenvalues are chosen as the first eigenvalue of the nonlinear model in null space $\emptyset$. As a result, the singularity problem of $\mathbf{S}_w$ is alleviated. In summary, the eigen-spectrum λ^w_k is regularized by an eigen-spectrum model:

$$\tilde{\lambda}^w_k = \begin{cases} \lambda^w_k & \text{if } k \le m \\ \dfrac{\alpha}{k+\beta} & \text{if } m < k \le r \\ \dfrac{\alpha}{r+1+\beta} & \text{if } r < k \le n \end{cases} \tag{8}$$

Discriminant information is extracted in the full image dimension under one unified discriminant criteria. Now the features are weighted by $\tilde{\lambda}^w_k$ as

$$\tilde{\mathbf{\Phi}}_w = \left\{ \phi^w_1 / \sqrt{\tilde{\lambda}^w_1}, \phi^w_2 / \sqrt{\tilde{\lambda}^w_2}, \ldots, \phi^w_n / \sqrt{\tilde{\lambda}^w_n} \right\} \tag{9}$$

Different eigen-spectrum models can be built. The optimal model is application-dependent. For simplicity, the original model in [5] is used, as no free parameter selection is needed.

Then, discriminant analysis is applied to extract a compact feature set. The between-class scatter matrix $\mathbf{S}_b$ is projected as $\tilde{\mathbf{S}}_b = \tilde{\mathbf{\Phi}}^T_w \mathbf{S}_b \tilde{\mathbf{\Phi}}_w$. Then, we perform eigen-decomposition on $\tilde{\mathbf{S}}_b$, i.e. $\tilde{\mathbf{S}}_b = \tilde{\mathbf{\Phi}}_b \tilde{\mathbf{\Lambda}}_b \tilde{\mathbf{\Phi}}^T_b$. The first d eigenvectors $\tilde{\mathbf{\Phi}}^d_b$ corresponding to d largest eigenvalues of $\tilde{\mathbf{S}}_b$ are selected. Finally, the projection matrix is obtained as $\tilde{\mathbf{P}} = \tilde{\mathbf{\Phi}}_w \tilde{\mathbf{\Phi}}^d_b$.

The procedures of ERE approach are summarized in Fig. 2.

Class-specific decision based on training data: Conventionally, a validation set is needed to find the optimal threshold of each class. Separating a validation set from the training set results in less training samples, and hence the overall performance of the system may degrade. Instead, we propose to determine the CST based on the training data.

The proposed approach is based on the following observation: for the imposter classes whose training samples are very close to the claimed class center, their testing samples are also likely close to the claimed class center, and hence easily falsely accepted. In view of this, we can use the distances from the claimed class center to other class centers as the indicator of the CST, e.g. $d_{i,j} = D(M_i, M_j)$, $i \ne j$, where M_i is the claimed class center, M_j is the class center of j-th class, and D is a distance measure. We can choose the distances to k-nearest class centers and use the average as the indicator of the CST. For simplicity, we choose $k = 1$. The illustration for determining the CST based on the training data is shown in Fig. 3. Imposter samples that are close to the claimed class center are easily falsely accepted. Thus, the

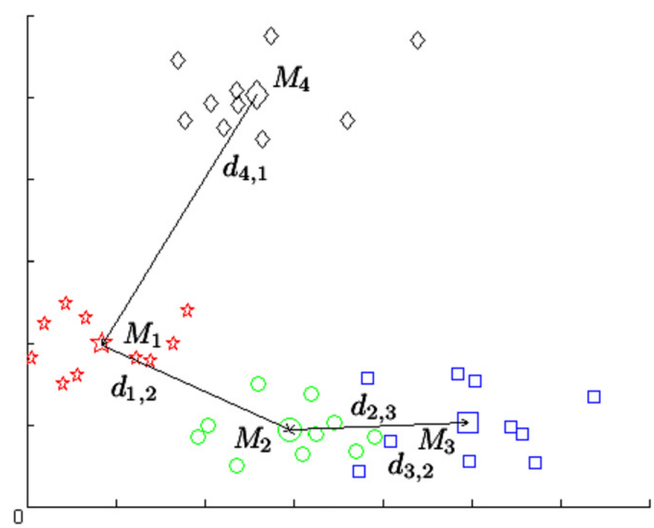

Fig. 3. Illustration for determining the class-specific threshold based on the training data. The small stars, circles, squares and diamonds represent samples of Classes 1–4, respectively. The big ones represent the class centers. The distance of a class center to its nearest class center is used as the indicator of the CST.

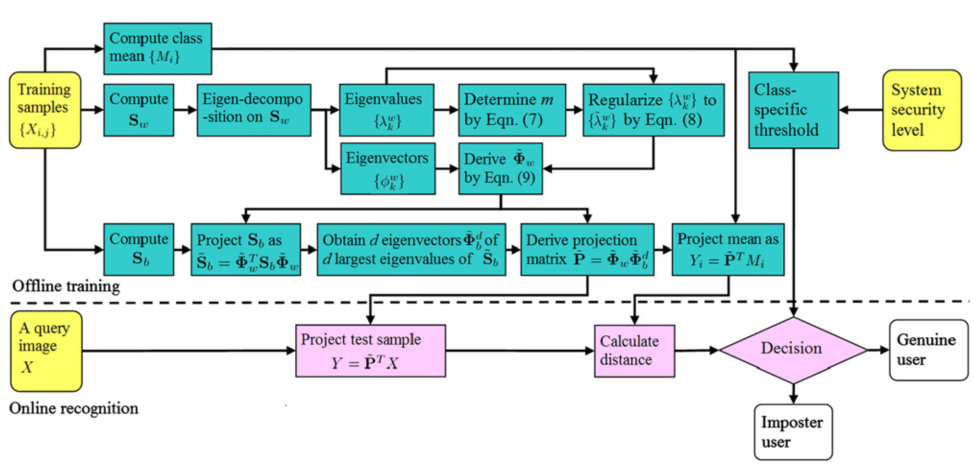

Fig. 2. Block diagram of ERE approach.

distance from the claimed class center to its nearest class center is chosen as the indicator of the CST, i.e. $d_i = \min_j(d_{i,j})$. As shown in Fig. 3, $d_{1,2}$, $d_{2,3}$, $d_{3,2}$ and $d_{4,1}$ are the indicators for Classes 1, 2, 3, 4, respectively. Based on this distance, the CST is calculated as $t_i = d_i R$, where $R \geq 0$ is used to control the security level of the system. If a high security level is required for some applications, i.e. a low false acceptable rate is desirable, user can choose a small value of R. If a low false rejection rate is desirable, user can choose a large value of R. In this way, the proposed approach determines the CST without a validation dataset. More samples can be used in training, and hence a better performance is expected.

In literature, to the best of our knowledge, when the CST is used, no receiver characteristic curve (ROC) is provided except for some discrete operational points. When a validation dataset is used to determine the CST, the number of operational points is determined by the limited number of genuine validation samples in each class. Therefore, only a limited number of operational points can be determined. For the proposed approach, ROC can be easily generated by varying the control parameter R.

In order to determine the CST, a large amount of training samples of each user are required. For general face recognition tasks, a large number of users use a public system. Thus, it is very difficult to collect such a large amount of training images per each user. However, for mobile applications, the mobile device is personalized. Thus, a user can comfortably collect a large number of his/her own photos gradually during the usage of the system, which can be used to compute a reliable CST.

4. O2FN mobile face database creation

In order to evaluate the proposed face verification system, the O2FN mobile face database is built. It contains 2000 face images of size 144×176 pixels from 50 subjects. The images are self-taken photos. The users are told to take roughly 20 indoor images and 20 outdoor images, with minimum facial expression variations and out-plane rotations. As shown in Fig. 4, the O2FN database mainly contains in-plane rotations and illumination variations.

The O2FN database is then used to generate the training and testing databases to evaluate each sub-system. For the face detection sub-system, 53 924 face samples and 477 097 non-face samples are generated and normalized to the size of 24×28 pixels. For the eye detection sub-system, the detected faces are first normalized to 96×112 pixels, and then 4000 eye samples and 207 331 non-eye samples of size 35×25 pixels are generated from the detected faces. To evaluate the face verification sub-system, face images are cropped, rotated to the same horizon and normalized to the size of 82×96 pixels based on the manually marked eyes. In the system-level evaluation, face images are normalized based on the eye locations determined by face/eye detector. An ellipse mask is applied to remove the unwanted corner, and the final face vector is of length 6634. Histogram equalization is applied to reduce illumination variations and the face vector is normalized to zero mean and unit variance.

5. Experimental results

In this section, the proposed face verification system is firstly systematically evaluated on the O2FN database. Then, it is further compared with the system built upon Adaboost face/eye detector on the AR database [15] and the CAS-PEAL database [16]. Superior performance of the proposed system is demonstrated.

The O2FN database is partitioned in a consistent way as shown in Table 1. The performance on each sub-system is evaluated in a pilot run, where 20 odd images of the first 40 subjects are used in training. Then, in the system-level evaluation, the experiment is repeated five times. For each trail, 40 subjects are randomly selected and used in training.

5.1. Face detection results on the O2FN database

In this section we show the gradual performance improvement, i.e. PCA → APCA → APCDA → C-APCDA face detector on the O2FN database. Then, the proposed C-APCDA face detector is compared with Adaboost face detector.

Table 1
Partition of the O2FN database.

Sub-system	40 subjects		10 subjects
	20 odd images	20 even images	40 images
Face detection	Training	Testing	Testing
Eye detection	Training	Testing	Testing
Face verification	Training	Genuine matching	Imposter matching

Fig. 4. Sample images of the O2FN database.

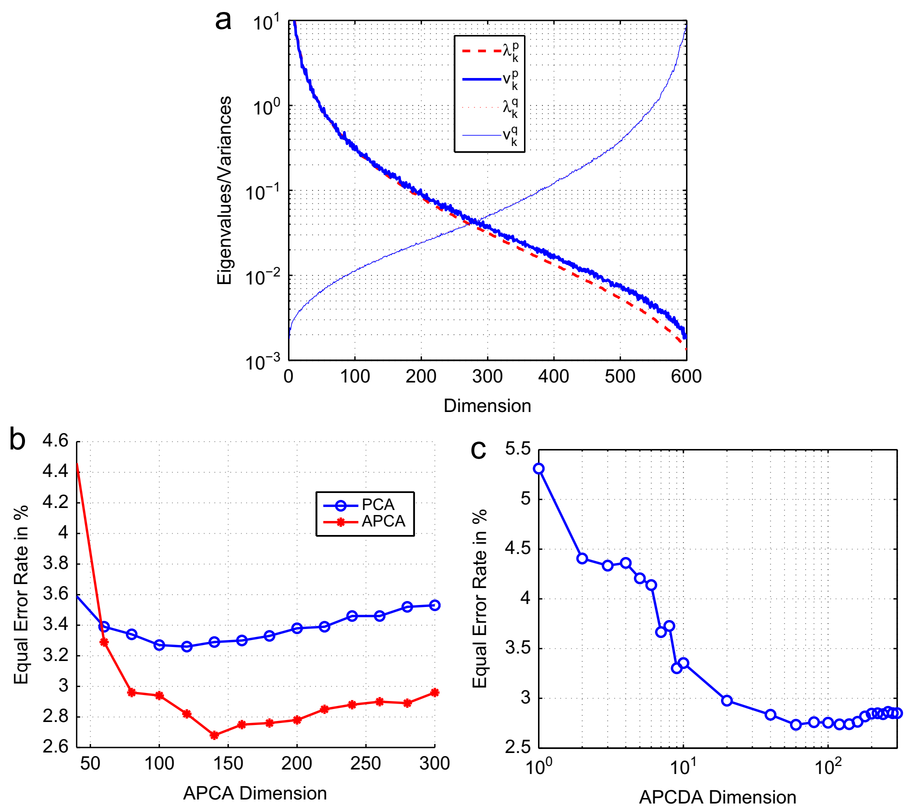

Fig. 5. Experimental results on the O2FN database. (a) Eigen-spectrum plot for the eigenvalue λ_k^p, λ_k^q of the face/non-face covariance matrix, and the corresponding variance υ_k^p, υ_k^q of the face/non-face testing samples, (b) the EER for the APCA/APCA face detector and (c) the EER for the APCDA face detector.

To study the reliability of face/non-face training samples, we analyze the eigen-spectrums of $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_q$. We perform eigen-decomposition on $\mathbf{\Sigma}_p$, $\mathbf{\Sigma}_q$, and obtain corresponding eigenvalue λ_k^p, λ_k^q and eigenvector Φ_k^p, Φ_k^q. Then, face/non-face testing samples are projected onto Φ_k^p, Φ_k^q, respectively. Corresponding variance υ_k^p, υ_k^q are calculated. We present λ_k^p, υ_k^p in the descending order, and λ_k^q, υ_k^q in the ascending order as shown in Fig. 5(a). λ_k^q and υ_k^q are almost overlapped, and hence non-face training samples can well represent testing samples. On the other hand, λ_k^p and υ_k^p differ significantly after dimension 300, and hence face training samples cannot well represent face testing samples. A large weight $\alpha = 0.9$ is assigned to $\mathbf{\Sigma}_p$ so that more unreliable dimensions from the face class can be removed by APCA. When $\alpha = n_p/(n_p+n_q) \approx 0.1$, it is a standard PCA face detector. n_p, n_q are the number of face/non-face training samples.

Then, we test the PCA/APCA face detector on extracted testing samples, and plot the EER vs. dimensions as shown in Fig. 5(b). APCA consistently outperforms PCA after dimension 60. It is consistent with our analysis in Section 3.1. By assigning a large weight to the unreliable face class, more dimensions from the unreliable covariance matrix of the face class are removed. Thus, better generalization performance is achieved.

The objective of ADA in APCDA is to find the most discriminant dimensions within the APCA subspace. The EER of the APCDA face detector of different dimensions is shown in Fig. 5(c). The EER of the APCDA face detector is comparable to that of the APCA face detector, but with much less dimensions. With 60 dimensions, the lowest EER of 2.73% is achieved for the APCDA face detector. Even with the first dimension only, the EER of 5.32% is achieved.

In view of the discriminant power of first few APCDA dimensions, a cascade APCDA face detector is built. It consists of eight stages with 1, 2, 3, 6, 12, 24, 48, 96 dimensions at each stage. A large amount of negative samples are filtered out by the classifier built from the first APCDA dimension. The proposed C-APCDA face detector is tested on 1200 face testing images. Let us define the overlap ratio R_O as the ratio of the overlapped region over the union area. 97.00% and 96.50% of faces are detected for $R_O \geq 85\%$ when $\epsilon = 0.5\%$ and $\epsilon = 1\%$, respectively. For $R_O \geq 70\%$, almost perfect detection rate is achieved: 99.67% and 99.25%, respectively. Unless otherwise stated, the face detection rate is calculated based on $R_O \geq 70\%$. In terms of time saving, the proposed C-APCDA face detector with $\epsilon = 0.5\%$ is 5.7 times faster than the APCDA face detector of 96 dimensions.

Table 2
Comparisons of APCA, APCDA, C-APCDA face detector and Adaboost face detector on the O2FN database.

Algorithm	$R_O \geq 85\%$ (%)	$R_O \geq 70\%$ (%)	Dimensions required
APCA300	97.25	99.58	300
APCDA100	97.00	99.83	100
C-APCDA	97.00	99.67	17
Adaboost face detector	96.17	99.17	–

The proposed face detector is further compared with Adaboost face detector. As it requires a huge number of samples and tremendous time to train an optimal Adaboost face detector, we utilize pre-trained face detectors in OpenCV library [34]. We test four Adaboost face detectors and the one with best performance (haarcascade_frontalface_alt_tree.xml) is reported. Table 2 summarizes the experimental results of the APCA face detector of 300 dimensions, the APCDA face detector of 100 dimensions, the proposed C-APCDA face detector and Adaboost face detector. The number of dimensions required for the C-APCDA face detector is estimated as $96/5.7 \approx 17$. Among three subspace face detectors, the C-APCDA face detector needs much less dimensions compared to the other two for roughly the same detection rate.

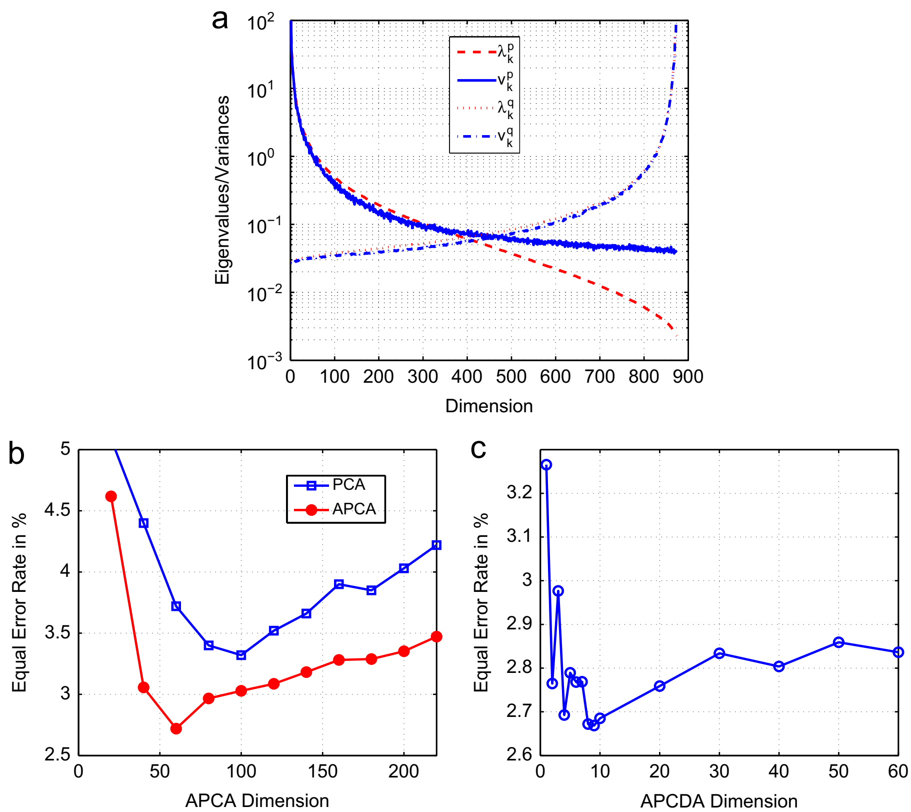

Fig. 6. Experimental results on the O2FN database. (a) Eigen-spectrum plot for the eigenvalue λ_k^p,λ_k^q of the eye/non-eye covariance matrix, and the corresponding variance $\upsilon_k^p,\upsilon_k^q$ of the eye/non-eye testing samples, (b) the comparison of the PCA/APCA eye detector in terms of EER and (c) the EER of the APCDA eye detector.

Compared to Adaboost face detector, the C-APCDA face detector achieves better face detection rate.

5.2. Eye detection results on the O2FN database

In this section, we present the experimental results of the APCA/APCDA eye detector, the proposed C-APCDA eye detector and the comparisons to other eye detectors. Eyes are detected within the detected face region. Thus, the detection rate reported here is the overall detection rate of face/eye detection.

To study the reliability of eye/non-eye training samples, we apply eigen-spectrum analysis similarly as in the face detection. For eye samples, eigenvalue λ_k^p of the covariance matrix $\mathbf{\Sigma}_p$ and corresponding variance υ_k^p of the testing samples are presented in the descending order as shown in Fig. 6(a). For non-eye samples, eigenvalue λ_k^q and corresponding variance υ_k^q are presented in the ascending order. υ_k^p differs significantly from λ_k^p for $k>350$, whereas υ_k^q in general is very close to λ_k^q. A limited number of eye training samples cannot well represent the testing samples, and hence are less reliable. A large weight is assigned to the eye class, e.g. $\alpha=0.3$. It is significantly larger than the weight in PCA, e.g. when $\alpha\approx 0.02$, APCA is deduced to standard PCA.

Then, the PCA eye detector and the APCA eye detector are tested on extracted samples. The results are shown in Fig. 6(b). The APCA eye detector consistently outperforms the PCA eye detector. By assigning a large weight to the unreliable eye class, more dimensions from the unreliable covariance matrix of the eye class are removed. Thus, better performance is achieved.

With much less dimensions, the APCDA eye detector can achieve comparable performance to the APCA eye detector. The EER of APCDA eye detector of different dimensions is shown in Fig. 6(c). With first two dimensions only, almost the same EER as APCA60 is achieved.

Table 3
Eye detection rate on the O2FN database for the C-APCDA eye detector of $\epsilon=0.5\%$, 1% and Adaboost eye detector on the detected face images under different tolerances.

Tolerance	C-APCDA eye detector $\epsilon=0.5\%$ (%)	C-APCDA eye detector $\epsilon=1\%$ (%)	Adaboost eye detector (%)
$T=1$	53.33	53.67	37.33
$T=2$	83.25	84.00	77.42
$T=3$	93.17	93.42	90.33
$T=4$	96.42	96.50	94.25
$T=5$	97.75	98.00	95.92

In view of the discriminant power of the first few APCDA features, a C-APCDA eye detector is built. It consists of six stages, with 1, 2, 3, 6, 12, 20 dimensions at each stage. It is tested on 1200 face images for $\epsilon=0.5\%$ and 1%. 97.75% and 98.00% eyes are detected within 5 pixels to the manually marked eyes, respectively. Table 3 shows the detection rates under different tolerances. A large portion of eyes are detected within 2 pixel to the manually marked eyes. In the latter sections, unless otherwise states, the eye detection rate is calculated based on 5-pixel tolerance. Compared to Adaboost eye detector [34], the proposed C-APCDA eye detector detects the eyes at much more precise locations.

For the C-APCDA eye detector, by utilizing the focus-attention strategy, most of non-eye samples are rejected at the early stages and only those tough samples are evaluated at the latter stages. Thus, the speed is greatly improved. The computational complexity of the C-APCDA eye detector is only 22.82% of the APCDA eye detector of 20 dimensions. Table 4 shows the performance comparison of the APCA eye detector of 60 dimensions, the APCDA eye detector of 20 dimensions and the proposed C-APCDA eye detector. For the C-APCDA eye detector, the number of dimensions required is calculated as $20 \times 22.82\% \approx 4.6$. Compared to the APCA/APCDA eye detector, the proposed C-APCDA eye detector achieves a comparable detection accuracy, but with much less dimensions.

The proposed C-APCDA eye detector is further compared with four feature-based eye detectors [35–38] and Adaboost eye detector [34]. Three sets of results for the proposed C-APCDA eye detector are presented: corresponding to scanning step sizes 1–3. For Adaboost eye detector, the experimental results of scale factor 1.01 and 1.1 are given. The algorithms are implemented and tested on a PC with Intel Core 2 Duo CPU @ 2.66 GHz and 2 GB RAM. The experimental results are shown in Table 5. Feature-based eye detectors are simple and fast, and hence suitable for mobile devices with limited computation power. However, as shown in Table 5, feature-based eye detectors cannot work well. The highest detection rate is achieved for the proposed C-APCDA eye detector of scanning step size 1. The processing time is 166.6 ms, which is longer than Adaboost eye detector of scale factor 1.1. As shown in the last two rows of Table 5, if we increase the scanning step size of the proposed algorithm to 2 or 3, the proposed C-APCDA eye detector consumes less processing time yet still achieves significantly higher detection rate than Adaboost eye detector. In the latter sections, the scale factor of Adaboost eye detector is set as 1.01. Some sample images of face/eye detection results for the proposed C-APCDA face/eye detector are shown in Fig. 7. The eyes are detected almost as accurate as the marked eyes.

5.3. Face verification results on the O2FN database

The proposed ERE approach [2,5] and various other subspace face recognition algorithms, e.g. Bayesian maximum likelihood (BML) [6], enhanced maximum likelihood (EML) [7], dual-space LDA (DSL) [8], FisherFace [9], null space approach [10] and graph embedding [12], are implemented and tested on the O2FN database. They are tested on the face images normalized based on the manually marked eyes ("Marked eyes") and the detected eyes ("Detected eyes"), and evaluated for CST and GT. "Detected eyes" are determined by the proposed C-APCDA face/eye detector. The experimental results at optimal parameters are reported in Table 6.

In original FisherFace approach the null space of $\boldsymbol{\Sigma}_w$ may not be removed when PCA is performed on $\boldsymbol{\Sigma}_t$. Those dimensions corresponding to the null space of $\boldsymbol{\Sigma}_w$ will have very large weights, and hence dominate the matching score. Therefore, we modify original FisherFace by performing PCA on $\boldsymbol{\Sigma}_w$ to reduce the dimension to $N-c$. Much better performance is achieved. For graph embedding, marginal Fisher analysis [12] is implemented.

Table 6 shows that the proposed ERE approach achieves the best performance for almost all the settings. The EER is as low as 1.37% for "Marked eyes" and CST. When tested for "Detected eyes", the EER is as low as 4.12% for CST. The eigen-spectrum model of ERE approach alleviates the over-fitting problem caused by the small eigenvalues in the range space of $\boldsymbol{\Sigma}_w$, and utilizes the discriminant information in both the range place and the null space. Thus, better performance is achieved.

Table 6 also shows that better performance is achieved for CST compared to GT. The CST determined by the proposed approach greatly improves the performance. In addition, we can easily obtain the ROC for different algorithms. E.g. the ROC for ERE approach, "Marked eyes", CST is shown in Fig. 8.

Finally, Table 6 shows that subspace approaches are sensitive to face alignment error. Previously, we achieve face/eye detection rate as high as 98%, but the performance based on the detected eyes still degrades significantly. Thus, an accurate face/eye detector is crucial to the performance of the complete and fully automated face verification system.

Face recognition via sparse representation [3] has demonstrated superior performance to handle occluded faces. We implement it based on l1-MAGIC [39] and compare it with ERE approach on the images normalized based on the manually marked eyes. The original normalized image of size 82×96 pixels

Table 4
Performance comparison of APCA, APCDA and C-APCDA eye detector on the O2FN database.

Eye detector	Detection rate (%)	Dimension required
APCA60	97.58	60
APCDA20	96.83	20
C-APCDA	98.00	4.6

Table 5
Comparisons of the C-APCDA eye detector with other eye detectors on the O2FN database.

Algorithm	Detection rate (%)	Time (ms)
Li et al. algorithm [35]	55.63	2.8
Lin and Yang's algorithm [36]	74.17	27.4
Ren and Jiang's algorithm [37]	83.42	51.8
Ren and Jiang's algorithm [38]	89.50	10.0
Adaboost [34]—SF 1.01	95.92	576.9
Adaboost [34]—SF 1.1	90.58	61.0
The proposed algorithm—step size 1	**98.00**	166.6
The proposed algorithm—step size 2	97.92	44.0
The proposed algorithm—step size 3	96.50	22.6

Table 6
Comparisons of various subspace face recognition algorithms at different settings in terms of EER on the O2FN database.

Algorithm	Marked eyes		Detected eyes	
	GT (%)	CST (%)	GT (%)	CST (%)
BML [6]	12.16	13.52	16.03	17.76
EML [7]	9.63	11.27	13.07	15.23
Dual-space [8]	3.13	1.83	8.08	4.98
FisherFace [9]	32.53	35.37	34.43	37.91
FisherFace—modified	2.69	1.88	**5.87**	4.88
Null space [10]	3.16	1.90	8.24	4.98
Graph embedding [12]	5.95	3.36	11.64	10.76
ERE approach [5]	**2.42**	**1.37**	6.28	**4.12**

Fig. 7. The sample images of face/eye detection results for the C-APCDA face/eye detector on the O2FN database.

is too large for sparse representation. Thus, we resize it to 41×48 pixels. Sparse representation is originally formulated as face identification problem with a reject option. Sparsity concentration index (SCI) is used to reject the imposter samples. Then for the genuine samples, Algorithm 1 in [3] is used to find the identity. To handle the occlusion, an identity matrix is appended after the image matrix. The EER for SCI is 2.94%, and subsequent face recognition rate on the genuine samples is 98.25%. The overall EER of sparse coding will surely be larger than 2.94%. On the same image set of size 41×48 pixels, the EER of ERE approach is 2.5% when using the global threshold and 1.78% when using the CST. Clearly ERE approach outperforms sparse coding.

5.4. System-level comparison on the O2FN database

The proposed face verification system, i.e. the C-APCDA face/eye detector, ERE face verification with the CST, is compared with the system built from Adaboost face/eye detector, ERE face verification with the global threshold. The O2FN database is partitioned as shown in Table 1. The experiment is repeated five times, and each time 40 subjects are randomly selected and used in training. The average performance and standard derivation are reported in Table 7. The performance is fairly consistent over five trials. The EER of ERE face verification on the images normalized based on the manually marked eyes is $0.94\% \pm 0.27\%$. The proposed system demonstrates superior performance compared to the competitive system. An overall EER of 3.88% is achieved for the proposed system.

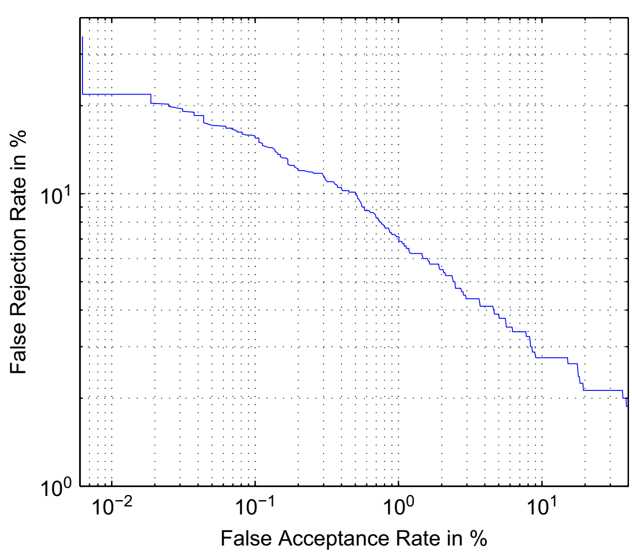

Fig. 8. The ROC for ERE approach based on the class-specific threshold.

5.5. Experimental results on the AR database

Seventy-five subjects with 14 non-occluded images per subject are selected from the AR database [15]. Those images mainly vary in facial expressions. The first seven images of first 60 subjects serve as the training images, and the remaining seven images of first 60 subjects serve as the genuine testing images. The 14 images of last 15 subjects serves as the imposter testing images. The C-APCDA face/eye detector is built based on 420 training images and then used to determine the eye locations of the testing samples. The face verification sub-system is based on ERE approach. The proposed system is compared with the system built upon the Adaboost face/eye detector as shown in Table 8. The abbreviations are the same as defined in Table 6.

The proposed C-APCDA face detector locates the faces at a much higher precision compared to the Adaboost face detector, e.g. 96.98% of faces are located with $R_O \geq 85\%$. Subsequently, the proposed C-APCDA eye detector detects the eyes at a higher precision compared to Adaboost eye detector, e.g. 98.73% of eyes are detected within 5 pixels to the manually marked eyes. As a result, compared to the system built upon Adaboost face/eye detector, the proposed system achieves much lower EER, e.g. 4.54% for the global threshold and 1.90% for the CST. In addition, it can be seen that the proposed approach to determine the CST significantly improves the performance.

5.6. Experimental results on the CAS-PEAL database

There are in total 30 863 images of 1040 subjects in the CAS-PEAL database, in which 9031 are frontal faces and 21 832 are profile faces. We choose the frontal dataset for testing. The CAS-PEAL database is more challenging than the AR database. Besides variations in facial expressions, there are variations in accessory, aging, background, distance and lighting. 2359 images of first 101 subjects are selected. Other subjects are discarded as they have much less images per subject, and hence not suitable for face verification task. The odd images of first 80 subjects serve as the training images, and the even images of first 80 subjects serve as the genuine testing images. The images of last 21 subjects serve as the imposter testing images. The face verification sub-system is

Table 7
Performance comparison of the proposed face verification system and the system built upon Adaboost face/eye detector, evaluated on 2000 images of the O2FN mobile face database for five trials.

Algorithm	Face detection rate (%)	Eye detection rate (%)	System overall EER (%)
C-APCDA face/eye detector+ERE+CST	99.33 ± 0.19	96.98 ± 0.29	3.88 ± 0.27
Adaboost face/eye detector+ERE+GT	98.73 ± 0.16	95.40 ± 0.18	8.49 ± 0.71

Table 8
Performance comparison of the proposed face verification system and the system built upon Adaboost face/eye detector, evaluated on 1050 images of 75 subjects on the AR database.

Algorithm	Face detection rate		Eye detection rate (%)	Marked eyes		Detected eyes	
	$R_O \geq 85\%$ (%)	$R_O \geq 70\%$ (%)		GT (%)	CST (%)	GT (%)	CST (%)
C-APCDA face/eye detector+ERE	**96.98**	99.84	**98.73**	2.70	**1.66**	4.54	**1.90**
Adaboost face/eye detector+ERE	91.27	99.21	95.71			7.64	3.37

Table 9
Performance comparison of the proposed face verification system and the system built upon Adaboost face/eye detector, evaluated on 2359 images of 101 subjects on the CAS-PEAL database.

Algorithm	Face detection rate		Eye detection rate (%)	Marked eyes		Detected eyes	
	$R_O \geq 85\%$ (%)	$R_O \geq 70\%$ (%)		GT (%)	CST (%)	GT (%)	CST (%)
C-APCDA face/eye detector+ERE	**94.48**	99.42	**98.55**	4.13	**3.18**	6.60	**5.60**
Adaboost face/eye detector+ERE	93.46	99.27	93.02			9.30	6.77

based on ERE approach. The proposed system is compared with the system built upon Adaboost face/eye detector as shown in Table 9.

The proposed C-APCDA face/eye detector locates the eyes at a much higher precision compared to Adaboost face/eye detector, e.g. the eye detection rate increases from 93.02% to 98.55%. As a result, the proposed system outperforms the system built upon Adaboost face/eye detector, e.g. the overall EER is reduced from 9.30% to 6.60% for the GT. The proposed approach to determine the CST further improves the performance. Compared to the GT, the overall EER of the proposed system is reduced from 6.60% to 5.60% for the CST.

5.7. Implementation on mobile devices

The proposed complete and fully automated face verification system is implemented on O2 XDA Flame, with 2 Mb camera at the rear, one VGA video camera in front, 520 MHz IntelR PXA-270 processor, 128 Mb RAM and Windows Mobile 6 operating system. Two techniques are employed to further speed up the system. Firstly, we utilize the focus-attention strategy in the spatial domain for face/eye detection, e.g. checking the face/eye at a coarse scale first and then focus on the region of interest at a fine scale. Secondly, in view of the discriminant power of the APCDA face/eye detector of first few dimensions, we build the C-APCDA face/eye detector based on down-sampled face/eye samples. The face is down-sampled from 24×28 pixels to 6×7 pixels and the eye is down-sampled from 35×25 pixels to 7×5 pixels. Scanning windows are first filtered by those down-sampled C-APCDA detector, and then only those tough samples are evaluated by the C-APCDA detector of higher resolution. In such a way, the detection speed is further improved. On average, it takes 110.5 ms for face detection, 145.1 ms for eye detection and 772.5 ms for face verification. In total, it takes 1028.1 ms to recognize a face on O2 XDA Flame mobile phone. If the system is implemented on the new mobile phones with more memory and computation power, e.g. iPhone 4S, much higher speed is expected.

The core module is implemented in native C++ (no windows API) and packaged in dll. The system can be easily ported into various applications and different platforms. Several applications have been built upon this system, e.g. mobile access control, personalized media consumption and data/folder locker. For more implementation details, readers can refer to [40].

6. Conclusion and future work

In order to secure the mobile devices, a complete and fully automated face verification system is proposed in this paper. The proposed system is systematically evaluated on O2FN, AR and CAS-PEAL databases. Comprehensive experiments demonstrate the superior performance of the proposed system compared to the state-of-the-art technologies. The best competitive system is built upon Adaboost face/eye detector and ERE approach for face recognition. Compared to this system, the proposed system reduces the overall EER from 8.49% to 3.88% on the O2FN database, from 7.64% to 1.90% on the AR database and from 9.30% to 5.60% on the CAS-PEAL database. The proposed system is implemented on O2 XDA Flame, in which it takes 1.03 s to recognize a face.

The proposed cascade APCDA face/eye detector achieves better detection rate than Adaboost face/eye detector. The eye detection rate increases from 95.40% to 96.98% on the O2FN database, from 95.71% to 98.73% on the AR database and from 93.02% to 98.55% on the CAS-PEAL database. The eyes are detected almost as accurate as the manually marked eyes. Furthermore, by utilizing the focus-attention strategy, the detection speed is greatly improved. Compared to Adaboost face/eye detector, the proposed C-APCDA eye face/eye detector detects the eyes at a much higher precision and a comparable speed.

Various subspace approaches are evaluated. In ERE approach, the eigen-spectrum of $\mathbf{\Sigma}_w$ is replaced by a spectrum model, which solves the problem of the unreliable small eigenvalues in the range space and utilizes the discriminant information in both the range space and the null space. The proposed approach to determine the CST from the training data instead of extra validation data further improves the performance.

A representative database is important to the success of the training-based face verification system. Therefore, in order to further improve the system, a larger mobile face database is necessary, which will be our future work.

Acknowledgment

The authors would like to thank National Science and Engineering Research Council, A*STAR for funding the research of the face verification system on mobile devices under the Research Grant no. 0621300056. This work is supported in part by the Nanyang Assistant Professorship (SUG M58040015), and supported in part by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

References

[1] P.J. Philips, M. Hyeonjoon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.
[2] X. Jiang, Linear subspace learning-based dimensionality reduction, IEEE Signal Processing Magazine 28 (2) (2011) 16–26.
[3] J. Wright, A. Yang, A. Ganesh, A. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.
[4] D.Q. Dai, P.C. Yuen, Regularized discriminant analysis and its application to face recognition, Pattern Recognition 36 (3) (2003) 845–847.
[5] X. Jiang, B. Mandal, A. Kot, Eigenfeature regularization and extraction in face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (3) (2008) 383–394.
[6] B. Moghaddam, T. Jebara, A. Pentland, Bayesian face recognition, Pattern Recognition 33 (11) (2000) 1771–1782.
[7] X. Jiang, B. Mandal, A. Kot, Enhanced maximum likelihood face recognition, Electronics Letters 42 (19) (2006) 1089–1090.
[8] X. Wang, X. Tang, Dual-space linear discriminant analysis for face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2004, pp. 564–569.

[9] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[10] W. Liu, Y. Wang, S. Li, T. Tan, Null space approach of Fisher discriminant analysis for face recognition, Biometric Authentication 3087 (2004) 32–44.
[11] L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, G.J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition 33 (10) (2000) 1713–1726.
[12] S. Yan, D. Xu, B. Zhang, H.J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2007) 40–51.
[13] X. Jiang, Asymmetric principal component and discriminant analyses for pattern classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (5) (2009) 931–937.
[14] P. Viola, M. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2) (2004) 137–154.
[15] A.M. Martinez, R. Benavente, The AR Face Database, CVC Technical Report #24, June 1998.
[16] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The CAS-PEAL large-scale Chinese face database and baseline evaluations, IEEE Transactions on Systems, Man, and Cybernetics (Part A) 38 (1) (2008) 149–161.
[17] S.U. Jung, Y.S. Chung, J.H. Yoo, K.Y. Moon, Real-time face verification for mobile platforms, Advances in Visual Computing, Part II 5359 (2008) 823–832.
[18] S. Paulus, N. Pohlmann, H. Reimer, P. Abeni, M. Baltatu, R. D'Alessandro, A face recognition system for mobile phones, in: Information Security Solutions Europe, 2006, pp. 211–217.
[19] Q. Tao, R. Veldhuis, Biometric authentication system on mobile personal devices, IEEE Transactions on Instrumentation and Measurement 59 (4) (2010) 763–773.
[20] A. Hadid, J.Y. Heikkila, O. Silven, M. Pietikainen, Face and eye detection for person authentication in mobile phones, in: IEEE International Conference on Distributed Smart Cameras, 2007, pp. 101–108.
[21] C.K. Ng, M. Savvides, P.K. Khosla, Real-time face verification system on a cellphone using advanced correlation filters, in: Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies, 2005, pp. 57–62.
[22] C. Geng, X.D. Jiang, Face recognition based on the multi-scale local image structures, Pattern Recognition 44 (10–11) (2011) 2565–2575.
[23] S. Jassim, H. Sellahewa, J. Ehlers, Wavelet-based face verification for mobile personal devices, in: Proceedings of the COST 275 Workshop on Biometrics on the Internet, 2005, pp. 81–84.
[24] K. Choi, K.A. Toh, H. Byun, Realtime training on mobile devices for face recognition applications, Pattern Recognition 44 (2) (2011) 386–400.
[25] K. Venkataramani, S. Qidwai, B.V.K. Vijayakumar, Face authentication from cell phone camera images with illumination and temporal variations, IEEE Transactions on Systems, Man and Cybernetics Part C—Applications and Reviews 35 (3) (2005) 411–418.
[26] J. Yang, X. Chen, W. Kunz, H. Kundra, Face as an index: knowing who is who using a PDA, International Journal of Imaging Systems and Technology 13 (1) (2003) 33–41.
[27] H. Yu, H. Yang, A direct LDA algorithm for high-dimensional data—with application to face recognition, Pattern Recognition 34 (10) (2001) 2067–2070.
[28] C. Schneider, N. Esau, L. Kleinjohann, B. Kleinjohann, Feature based face localization and recognition on mobile devices, in: Proceedings of the International Conference on Control, Automation, Robotics and Vision, 2006, pp. 1–6.
[29] P.F. Hsieh, D.S. Wang, C.W. Hsu, A linear feature extraction for multiclass classification problems based on class mean and covariance discriminant information, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2) (2006) 223–235.
[30] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bulletin of the Calcutta Mathematical Society 35 (1943) 99–109.
[31] K. Fukunaga, Introduction to Statistical Pattern Recognition, second ed., Academic Press, 1990.
[32] S. Yan, D. Xu, X. Tang, Face verification with balanced thresholds, IEEE Transactions on Image Processing 16 (1) (2006) 262–268.
[33] G. Goudelis, S. Zafeiriou, A. Tefas, I. Pitas, Class-specific kernel-discriminant analysis for face verification, IEEE Transactions on Information Forensics and Security 2 (3) (2007) 570–587.
[34] G. Bradski, The opencv library, Dr. Dobb's Journal: Software Tools for the Professional Programmer 25 (11) (2000) 120–124.
[35] G. Li, X. Cai, X. Li, Y. Liu, An efficient face normalization algorithm based on eyes detection, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 3843–3848.
[36] D.T. Lin, C.M. Yang, Real-time eye detection using face-circle fitting and dark-pixel filtering, in: IEEE International Conference on Multimedia and Expo (ICME 2004), 2004, pp. 1167-1170.
[37] J. Ren, X. Jiang, Eye detection based on rank order filter, in: International Conference on Information, Communications and Signal Processing, 2009, pp. 1–4.
[38] J. Ren, X. Jiang, Fast eye localization based on pixel differences, in: IEEE International Conference on Image Processing, 2009, pp. 2733–2736.
[39] 〈http://users.ece.gatech.edu/~justin/l1magic/〉.
[40] 〈http://www.cis.eee.ntu.edu.sg/Projects/Pages/FaceVerificationSystemonMobileDevices.aspx〉.

Jianfeng Ren received the BEng from National University of Singapore, major in Communication and MSc from Nanyang Technological University, major in Signal Processing in 2001 and 2009, respectively. He received Professional Engineers Board Gold Medal for his best academic results among all EEE MSc graduates in 2009. From 2003 to 2007, he was in industry sections. In December 2007, he joined Nanyang Technological University as a project officer, responsible for the development of the face verification system on mobile devices. Currently he is pursuing Ph.D. in Electrical and Electronic Engineering at Nanyang Technological University, Singapore. His research interests include face recognition, general pattern recognition and computer vision.

Xudong Jiang received the BEng and MEng degrees from the University of Electronic Science and Technology of China (UESTC) in 1983 and 1986, respectively, and the Ph.D. degree from Helmut Schmidt University, Hamburg, Germany, in 1997, all in Electrical and Electronic Engineering. From 1986 to 1993, he was a lecturer at UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1993 to 1997, he was with Helmut Schmidt University, Hamburg, as a scientific assistant. From 1998 to 2002, he was with Nanyang Technological University (NTU), Singapore, as a senior research fellow, where he developed a fingerprint verification algorithm that achieved the most efficient and the second most accurate fingerprint verification at the International Fingerprint Verification Competition (FVC'00). From 2002 to 2004, he was a lead scientist and the head of the Biometrics Laboratory at the Institute for Infocomm Research, Singapore. He joined NTU as a faculty member in 2004. Currently, he serves as the director of the Centre for Information Security, School of Electrical and Electronic Engineering, NTU, Singapore. His research interests include pattern recognition, signal and image processing, computer vision, and biometrics.

Junsong Yuan (M'08) received the Ph.D. and M.Eng degrees from Northwestern University, Illinois, USA, and National University of Singapore, respectively. Before that, he graduated from the special program for the gifted young in Huazhong University of Science and Technology, Wuhan, PR China, and received his B.Eng in Communication Engineering. During the summer 2008, 2007 and 2006, he was a research intern with the Communication and Collaboration Systems group, Microsoft Research, Redmond, WA, Kodak Research Laboratories, Rochester, NY, and Motorola Laboratories, Schumburg, IL, respectively. From 2003 to 2004, he was a research scholar at the Institute for Infocomm Research, Singapore.

In September 2009, he joined Nanyang Technological University as a Nanyang assistant professor. His current research interests include computer vision, image and video data mining and content analysis, multimedia search, etc. He was a recipient of the Doctoral Spotlight Award from IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'09), a recipient of the Nanyang Assistant Professorship from Nanyang Technological University, and received the Outstanding Ph.D. Thesis award from the EECS department in Northwestern University.