applied sciences

MDPI

Article

Improved Single Sample Per Person Face Recognition via Enriching Intra-Variation and Invariant Features

Huan Tu [1], Gesang Duoji [2,*], Qijun Zhao [1,2,*] and Shuang Wu [1]

[1] College of Computer Science, Sichuan University, Chengdu 610065, China; tuhuan722@outlook.com (H.T.); ws981117@gmail.com (S.W.)

[2] School of Information Science and Technology, Tibet University, Lhasa 850000, China

* Correspondence: gsdj@utibet.edu.cn (G.D.); qjzhao@scu.edu.cn (Q.Z.)

Received date: 8 December 2019; Accepted date: 3 January 2020; Published: 14 January 2020

Abstract: Face recognition using a single sample per person is a challenging problem in computer vision. In this scenario, due to the lack of training samples, it is difficult to distinguish between inter-class variations caused by identity and intra-class variations caused by external factors such as illumination, pose, etc. To address this problem, we propose a scheme to improve the recognition rate by both generating additional samples to enrich the intra-variation and eliminating external factors to extract invariant features. Firstly, a 3D face modeling module is proposed to recover the intrinsic properties of the input image, i.e., 3D face shape and albedo. To obtain the complete albedo, we come up with an end-to-end network to estimate the full albedo UV map from incomplete textures. The obtained albedo UV map not only eliminates the influence of the illumination, pose, and expression, but also retains the identity information. With the help of the recovered intrinsic properties, we then generate images under various illuminations, expressions, and poses. Finally, the albedo and the generated images are used to assist single sample per person face recognition. The experimental results on Face Recognition Technology (FERET), Labeled Faces in the Wild (LFW), Celebrities in Frontal-Profile (CFP) and other face databases demonstrate the effectiveness of the proposed method.

Keywords: face recognition; single sample per person; sample enriching; intrinsic decomposition

1. Introduction

Face recognition has been an active topic and attracted extensive attention due to its wide potential applications in many areas [1–3]. There are multiple modalities of face data that can be used in face recognition, such as near infrared images, depth images, Red Green Blue (RGB) images, etc. Compared with near infrared and depth images [4], RGB images include more information and have broader application scenarios. In the past decades, many RGB-based face recognition methods have been proposed and great progress has been made, especially with the development of deep learning [5–8]. However, there are still many problems to be solved. Face recognition with single sample per person, i.e., SSPP FR, proposed in 1995 by Beymer and Poggio [9], is one of the most important issues. In SSPP FR, there is only one training sample per person but various testing samples with appearance different from training samples. This situation could appear in many actual scenarios such as criminal tracing, ID card identification, video surveillance, etc. In SSPP FR, the limited training samples provide insufficient information of intra-class variations, which significantly decreases the performance of most existing face recognition methods. Tan et al. [10] showed that the performance of face recognition drops with the decreasing number of training samples per person and a 30% drop of recognition rate happens when only one training sample per person is available. In recent years, many

methods have been suggested to solve the SSPP FR problem. These methods can be roughly divided into three categories: robust feature extraction, generic learning, and synthetic face generation.

Algorithms in the first category extract features that are robust to various variations. Some of them extract more discriminative features from single samples based on variants of improved principal component analysis (PCA) [11–13]. Others focus on capturing multiple face representations [14–16], mostly by dividing face image into a set of patches and applying various feature extraction techniques to get face representations. For instance, Lu et al. [14] proposed a novel discriminative multi-manifold analysis (DMMA) method to learn features from patches. They constructed a manifold from patches for every individual and formulated face recognition as a manifold–manifold matching problem to identify the unlabeled subjects. Dadi et al. [17] proposed to represent human faces by Histogram of Oriented Gradients (HOG), which captures edge or gradient structure and is invariant to local geometric and photometric transformations [18]. Local Binary Pattern (LBP) texture features extractor proposed by Ahonen et al. [19] has also been explored for face recognition thanks to its computational efficiency and invariance to monotonic gray-level changes. With the development of deep learning, there are many other methods that utilize the deep learning ability to extract more robust features, such as VGGNet [20], GoogleNet [21], FaceNet [22], ResNet [23], and SENet [24].

Generic learning attempts to utilize a generic set, in which each person has more than one training samples, to enhance the generalization ability of model. An implicit assumption of this kind of algorithms is that the intra-class variations for different datasets are similar and can be employed to share useful information to learn more robust model. The idea of sharing information has been widely used in [25–29] and achieved promising results. Sparse-representation-based classification (SRC) [30] is often used for face recognition, but its performance depends on adequate samples for each subject. Deng et al. [25] extended SRC framework by constructing an intra-class variation dictionary from generic training set together with gallery dictionary to recognize query samples. A sparse variation dictionary learning (SVDL) technique was introduced by Yang et al. [27], which learns a projection from both gallery and generic set and rebuilds a sparse dictionary to perform SSPP FR.

For the last category, some researchers synthesize some samples for each individual from the single sample to compensate the limited intra-class variations [31–37]. Mohammadzade and Hatzinakos [32] constructed expression subspaces and used them to synthesize new expression images. Cuculo et al. [36] extracted features from images that are augmented by standard augmentation techniques, such as cropping, translation, and filtering, and then applied sparsity-driven sub-dictionary learning and k-LIMAPS for face identification. To solve the lighting effect, Choi et al. [37] proposed a coupled bilinear model that generates virtual images under various illuminations using a single input image, and learned feature space based on these synthesized images to recognize a face image. Zeng et al. [33] proposed an expanding sample method based on traditional approach and used the expanded training samples to fine-tune a well-trained deep convolutional neural network model. 3D face morphable model (3DMM) is widely applied to face modeling and face image synthesis [35,38–40]. Zhu et al. [38] fitted 3DMM to face images via cascaded convolutional neural networks (CNN) and generated new images across large poses, which compose a new augmented database, namely 300W-LP. Feng et al. [39] presented a supervised cascaded collaborative regression (CCR) algorithm that exploits 3DMM-based synthesized faces for robust 2D facial landmark detection. SSPP-DAN introduced in [35] combines face synthesis and domain-adversarial network. It first generates synthetic images with varying poses using 3DMM and then eliminates the gap between source domain (synthetic data) and target domain (real data) by domain-adversarial network. Song et al. [40] explored the use of 3DMM in generating virtual training samples for pose-invariant CRC-based face classification.

The core idea of all the aforementioned methods is to train a model that can extract the identity features of face images, and ensure that the features are discriminative enough to find a suitable classification hyperplane that can accurately divide the features of different individuals. Unfortunately, many external factors, such as pose, facial expression, illumination, resolution, etc., heavily affect the appearance of facial images, and the lack of samples with different external factors leads to insufficient

learning of feature space and inaccurate classification hyperplane. Being aware of these problems with existing methods, we propose a method to improve SSPP FR from two perspectives: *Normalization* and *Diversification*. *Normalization* is to eliminate the external factors so as to extract robust and invariant features, which are helpful for defining more accurate classification hyperplane. *Diversification* means to enrich intra-variation through generating additional samples with various external factors. More diverse samples also enable the model to learn more discriminative features for distinguishing different individuals. To achieve this goal, a 3D face modeling module including 3D shape recovery and albedo recovery is presented at first. For the albedo recovery particularly, we make full use of the physical imaging principle and face symmetry to complete the invisible areas caused by self-occlusion while reserving the identity information. Since we represent albedo in the form of UV map, which is theoretically invariant to pose, illumination and expression (PIE) variations, we can alleviate the influence of these external factors. Based on the recovered shape and albedo, additional face images with varying pose, illumination, and expression are generated to increase intra-variation. Finally, we are able to improve the SSPP face recognition accuracy thanks to the enriched intra-variation and invariant features.

The remaining parts of this paper are organized as follows. Section 2 reviews face recognition with single sample per person and inverse rendering. Section 3 presents the detail of the proposed method. Section 4 reports our experiments and results. Section 5 provides the conclusion of this paper.

2. Related Work

2.1. Face Recognition with Single Sample Per Person

With the unremitting efforts of scholars, face recognition has made great progress. However, the task becomes much more challenging when only one sample per person is available for training the face recognition model. Dadi et al. [17] extracted histogram of oriented gradients (HOG) features and employ support vector machine (SVM) for classification. Li et al. [41] combined Gabor wavelets and feature space transformation (FST) based on fusion feature matrix. They projected the combined features to a low-dimensional subspace and used nearest neighbor classifier (NNc) to complete classification. Pan et al. [42] proposed a locality preserving projection (LPP) feature transfer based algorithm to learn a feature transfer matrix to map source faces and target faces into a common subspace.

In addition to the traditional methods introduced above, there are many other methods that utilize the learning ability of deep learning to extract features. To make up for the lack of data in SSPP FP, some algorithms combine deep learning and sample expanding. In [34], a generalized deep autoencoder (GDA) is firstly trained to generate intra-class variations, and is then separately fine-tuned by the single sample of each subject to learn class-specific DA (CDA). The new samples to be recognized are reconstructed by corresponding CDA so as to complete classification task. Similarly, Zeng et al. [33] used a traditional approach to learn an intra-class variation set and added the variation to single sample to expand the dataset. Then, they fine-tuned a well-trained network using the extended samples. Sample expanding can be done not only in the image space but also in the feature space. Min et al. [43] proposed a sample expansion algorithm in feature space called k class feature transfer (KCFT). Inspired by the fact that similar faces have similar intra-class variations, they trained a deep convolutional neural network on a common multi-sample face dataset at first and extracted features for the training set and a generic set. Then, k classes with similar features in the generic set are selected for each training sample, and the intra-variation of the selected generic data is transferred to the training sample in the feature space. Finally, the expanded features are used to train the last layer of SoftMax classifier.

Unlike these existing methods, this paper simultaneously recovers intrinsic attributes and generates diversified samples. Compared with sample expanding methods mentioned above, our method uses face modeling to decompose intrinsic property and generates images with richer intra-variation via simulating the face image formation process rather than following the idea of intra-variation migration. Our method also takes full advantage of intrinsic properties that can more robustly represent identity information. Deep learning is used as a feature extractor in our method due to its superiority demonstrated in many existing studies.

2.2. Inverse Rendering

The formation of face images is mainly affected by intrinsic face properties and external factors. Intrinsic properties consist of shape (geometry) and albedo (skin properties), while external factors include pose, illumination, expression, camera setting, etc. Inverse rendering refers to reversely decomposing internal and external properties in facial images. Many inverse rendering methods have been proposed. CNN3DMM [44] represents shape and albedo, respectively, as a linear combination of PCA bases and uses a CNN to regress the combination coefficients. SfSNet [45] mimics the process of imaging faces based on physical models and estimates the albedo, light coefficients, and normal of the input face image.

As one of the intrinsic properties, albedo has natural advantage for face recognition owning to its robustness to variations in view angle and illumination. However, most inverse rendering algorithms pay more attention to recovering a more accurate and detailed 3D face shape, and treat the albedo as an ancillary result. As one of the few algorithms using albedo to assist face recognition, Blanz and Vetter [46] captured the personal specific shape and albedo properties by fitting a morphable model of 3D faces to 2D images. The obtained model coefficients that are supposed to be independent of external factors can be used for face recognition. However, due to the limited representation ability of the statistical model, the recovered albedo would lose its discrimination to some extent. To solve this problem, Tu et al. [47] proposed to generate albedo images with frontal pose and neutral expression from face images of arbitrary view, expression, and illumination, and extract robust identity features from the obtained albedo images. They experimentally showed that albedo is beneficial to improving face recognition. However, they only realize the synthesis of normalized albedo images in two-dimensional image space, lacking the exploration on the principle of physical imaging, which leads to a poor performance on a cross-database.

3. Proposed Method

3.1. Overview

3.1.1. Preliminary

In this paper, densely aligned 3D face shapes are used, each containing n vertices. Generally, we denote an n-vertex 3D face shape as point cloud $S \in \mathbb{R}^{3\times n}$, where each column represents the coordinates of a point. The face normal, represented as $N \in \mathbb{R}^{3\times n}$, is calculated from the 3D face shape. The texture and albedo are denoted as $T, A \in \mathbb{R}^{3\times n}$, where each column represents the color and reflectivity of a point on the face.

However, using only a collection of attributes of each point to represent S, N, T, and A misses information about the spatial adjacency between points. Inspired by position maps in [48], we denote albedo as a UV map: $UV_A \in \mathbb{R}^{256\times256\times3}$ (see Figure 1). Each point in A can find a unique corresponding pixel on UV_A. Different from the traditional UV unwrapping method, each pixel in our UV map will not correspond to multiple points in A. In addition, we also use UV_T and UV_N to represent facial texture and facial normal as UV maps.

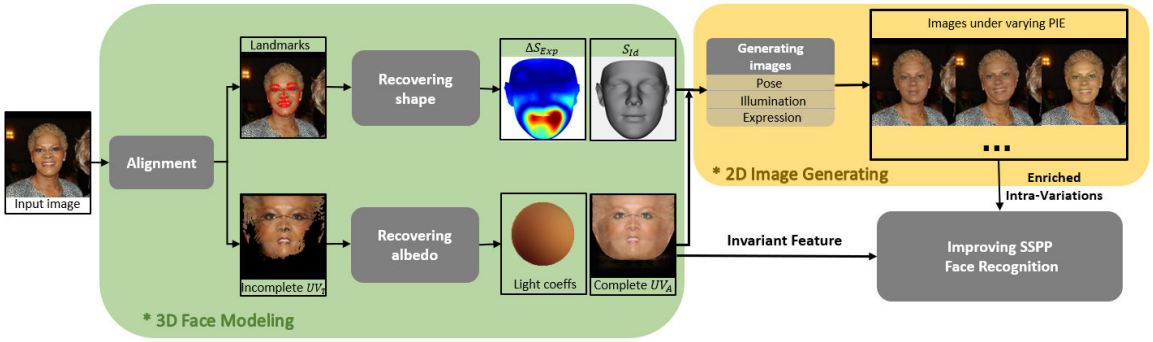

Figure 1. Pipeline of proposed method.

3.1.2. Pipeline

Figure 1 shows the framework of the proposed method for single sample per person face recognition. The method consists of three modules: 3D face modeling, 2D image generation, and improved SSPP FR. Given a face image of a person, we detect 68 landmarks, U, and generate the incomplete UV map of texture (Incomplete UV_T) using PRNet algorithm in [48] at first. We then recover its 3D face shape and complete UV map of albedo (Complete UV_A), respectively, from landmarks and Incomplete UV_T. With the recovered properties, images under varying pose, illumination, and expression are generated in the 2D image generation module. Finally, in the improved SSPP FR module, the reconstructed Complete UV_A and generated images are used to assist SSPP face recognition. Next, we detail: (i) albedo recovery; (ii) shape recovery; (iii) data enrichment; and (iv) SSPP FR.

3.2. Albedo Recovery

3.2.1. Network Structure

We assume that the face is Lambertian and illuminated from the distant. Under the Lambertian assumption, we represent the lighting and reflectance model as second-order Spherical Harmonics (SH) [49,50], which is a natural extension of the Fourier representation to spherical function. In SH, the irradiance at a surface point with normal (n_x, n_y, n_z) is given by

$$B(n_x, n_y, n_z|\Theta^{sh}) = \sum_{k=1}^{b^2} \Theta_k^{sh} H_k(n_x, n_y, n_z), \tag{1}$$

where H_k are the $b^2 = 3^2 = 9$ SH basis functions, and Θ_k^{sh} is the corresponding kth illumination coefficient. Since we consider colored illumination, there are totally $3 \times 9 = 27$ illumination coefficients with nine coefficients for each of the R, G, and B channels. The texture of a surface point can be calculated by multiplying the irradiance and albedo of the point. To sum up, the texture under certain illumination is a function of normal, albedo, and illumination, and can be expressed as

$$\begin{aligned} T(p) &= f_{sh}(A(p), N(p), SHL), \\ UV_T(p) &= f_{sh}(UV_A(p), UV_N(p), SHL), \end{aligned} \tag{2}$$

where p represents a pixel (2D) or point (3D), and SHL denotes the SH illumination coefficients.

Inspired by Sengupta et al. [45], we propose an end-to-end network that can recover the missing part in the incomplete UV_T and generate its complete version UV_A. As can be seen in Figure 2, we concatenate the incomplete UV_T with its horizontally flipped image as input of the network. The proposed network follows an encoder–decoder structure, in which the encoder module extracts a common feature from input image, and the albedo decoder and the normal decoder decode the complete albedo UV_A and the complete normal UV_N from the common feature, respectively, and the light decoder computes the spherical harmonics illumination coefficients SHL from the concatenation

of common feature, albedo feature, and normal feature. At last, following Equation (2), a rendering layer is used to recover the texture based on the above decoded attributes.

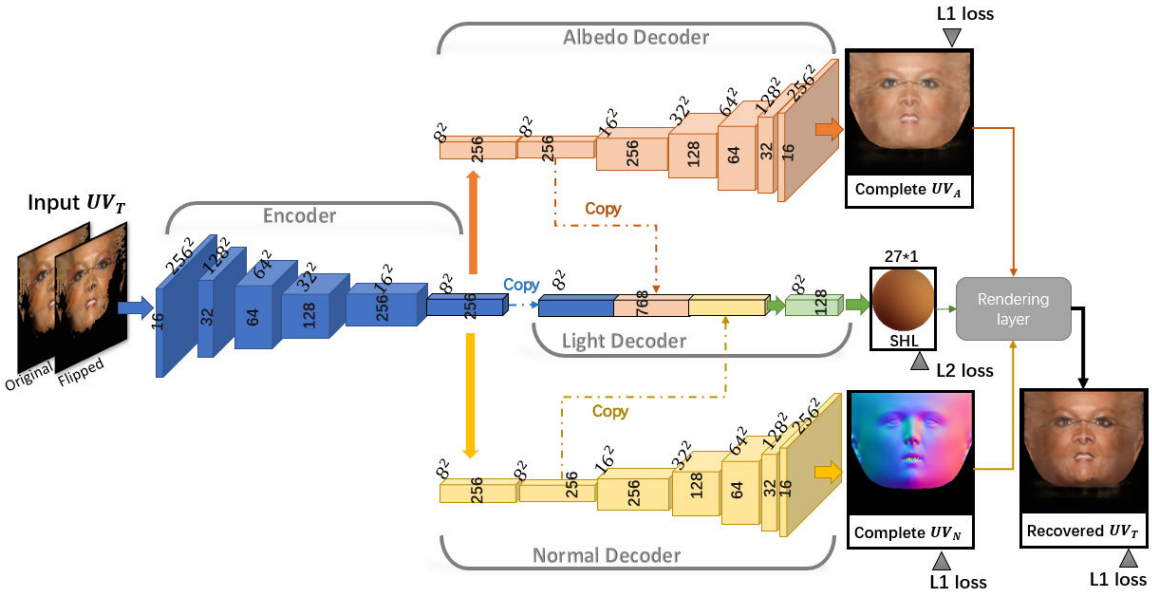

Figure 2. Pipeline of albedo recovery.

3.2.2. Loss Functions

To train the albedo recovery model, we minimize the error between the reconstructed value and the ground truth. However, the ground truth of unseen regions in real face is unavailable. To address the issue, we flip the reconstructed texture horizontally and make it as similar as possible to the input texture image. The loss function for reconstructed texture is defined as

$$\begin{aligned} L_{recon} &= \frac{1}{t}\sum_{p}\left(|UV_M[p]\left(UV_T^*[p]-\hat{UV}_T[p]\right)\right) \\ &+\frac{\lambda_f}{t}\sum_{p}\left(|UV_M[p]\left(UV_T^*[p]-\hat{UV}_{T_{flip}}[p]\right)\right), \end{aligned} \tag{3}$$

where UV_M is the visibility mask, $[p]$ denotes the pixel spatial location, t is the number of visible pixels, $\hat{UV}_{T_{flip}}$ means horizontally flipping the reconstructed texture $\hat{UV}_T$, λ_f denotes the weight of the reconstruction loss component associated with the horizontally-flipped reconstructed texture with respect to that associated with the original reconstructed texture, and the symbols "*" and "$\hat{}$" indicate the ground truth and reconstructed values, respectively.

The loss functions for SH illumination coefficients, albedo, and normal are formulated, respectively, as

$$L_l = \| SHL^* - S\hat{H}L \|_2^2, \tag{4}$$

$$L_a = |UV_A^* - \hat{UV}_A|, \tag{5}$$

$$L_n = |UV_N^* - \hat{UV}_N|. \tag{6}$$

The total loss is defined as

$$L_{total} = \lambda_{recon}L_{recon} + \lambda_l L_l + \lambda_a L_a + \lambda_n L_n, \tag{7}$$

where λ_{recon}, λ_l, λ_a, and λ_n are the weights to balance different losses.

3.2.3. Implementation Details

Since the ground truth of albedo, normal, and light coefficients of real facial images is not available, synthetic data are firstly used in this paper. Following the definition of 3DMM [46], the shape and the albedo of a face can be represented as

$$S = \overline{S} + \alpha_{id}\, b^s_{id} + \alpha_{exp}\, b^s_{exp}, \tag{8}$$

$$A = \overline{A} + \beta_{id}\, b^a_{id}, \tag{9}$$

where S and A are 3D face shape and albedo; $\overline{S}$ is the mean shape; $\overline{A}$ is the mean albedo; b^s_{id} and b^a_{id} are, respectively, the identity-related shape bases and albedo bases; b^s_{id} is the expression-related shape bases; and α_{id}, α_{exp}, and β_{id} are corresponding shape parameters, expression parameters, and albedo parameters.

Bar et al. [51] provided a method to flatten a 3D shape into 2D embedding and re-sample a 3DMM (including $\overline{S}$, $\overline{A}$, b^s_{id}, b^s_{exp}, and b^a_{id}) from the initial 3DMM over a uniform grid of size $H \times W$ in this flattened space. The re-sampled 3DMM has $n = HW$ vertices and each vertex has an associated UV coordinate. In our work, b^s_{id} and b^a_{id} of initial 3DMM come from BFM 2009 [52] and b^s_{exp} comes from FaceWarehouse [53]. Using the method in [51], we obtain a re-sampled 3DMM with $n = 256^2$ = 65,536 vertices. Then, we remove the vertices in the neck area resulting in a 3D face of 43,867 vertices and UV map of 256×256 with 43,867 valid pixels. We use the re-sampled 3DMM to synthesize 2330 subjects with a total of 150,704 facial images.

In view of the difference in data distribution between synthetic images and real images, real training data are necessary. We used real data from the CelebA database [54], including a total of 202,599 facial images of 10,177 subjects, and adopted a two-step training strategy to solve the problem of unavailability of real data labels. (i) We first trained the network on synthetic data. The pre-trained network was applied to real data from CelebA to obtain their albedo, normal, and SH illumination coefficients, which were taken as "pseudo ground truth" of the real data. (ii) We then fine-tuned the network with the combined set of synthetic data with "golden ground truth" and real data with "pseudo ground truth". For both steps, we applied Equation (7) to supervise the learning process and set λ_{recon}, λ_l, λ_a, and λ_n as 1.0 and λ_f as 0.5. Optimization was done by Adaptive Moment Estimation (Adam) with mini-batch of 32 samples, and the exponential decay rate for the first-moment estimates and second-moment estimates were set as 0.5 and 0.999, respectively. When we trained the network with only synthetic data, the learning rate was initialized as 10^{-3} and then decreased by factor of 10 after every four epochs. When training the network with combination of synthetic and real data, we set the learning rate to 10^{-4}.

3.3. Shape Recovery

To recover the 3D face shape, motivated by the method in [55], we train a cascade coupled-regressor $\{R^k_{Id}\}_{k=1}^K$, $\{R^k_{Exp}\}_{k=1}^K$ to reconstruct the 3D shape based on the detected 2D landmarks U^* on the input image. Here, 3D face shape is disentangled as

$$S = \overline{S} + \Delta S_{Id} + \Delta S_{Exp}, \tag{10}$$

where $\overline{S}$ is the mean 3D shape of frontal pose and neutral expression, termed pose-and-express-normalized (PEN) 3D face shape, ΔS_{Id} is the difference between the subject's PEN 3D shape, and $\overline{S}$, ΔS_{Exp} is the deformation in S caused by expression.

Given a 3D face shape and the corresponding 3D landmark vertices D, its 2D landmarks U at the same pose as the input face is obtained through weak perspective projection M. Note that the 3D-to-2D projection matrix is computed via least squares such that the projection of the D is as close as possible to the ground truth landmarks U^*, i.e.,

$$M = ((D)^T D)^{-1}(D)^T U^*. \tag{11}$$

The core idea of this method is to assume that there is a "relationship" between the difference on two 3D face shapes and the difference on their corresponding 2D landmarks, which can be learned from training samples. Thus, the reconstruction can be described as:

$$\begin{aligned} S^k = S^{k-1} \quad & + R^k_{Id}(U^* - U^{k-1})) \\ & + R^k_{Exp}(U^* - U^{k-1})), \end{aligned} \tag{12}$$

where S^k is the reconstructed 3D face shape after k iteration, $S^0 = \overline{S}$.

To train the coupled-regressors, we synthesized 1000 face images under various poses and expressions based on the resampled 3DMM. The ground truth ΔS_{Id} and ΔS_{Exp} of these images were recorded during the synthesis. In this study, we cascaded five pairs of identity and expression shape regressors, i.e., $K = 5$.

3.4. Data Enrichment

With the obtained identity shape $\overline{S} + \Delta S_{Id}$, in short S_{Id}, and albedo UV_A of a subject, face images of the subject with arbitrary poses, illuminations, and expressions could be generated. Firstly, we obtain shape S with the arbitrary expression via Equation (10), in which random ΔS_{Exp} is generated through the resampled 3DMM. Secondly, we add pose to S and obtained S_{pose}. The process of adding a pose can be expressed as:

$$S_{pose} = f * R * S + t_{3d}, \tag{13}$$

where R is the 3×3 rotation matrix calculated from the pitch, yaw, and roll angles; f is the scale factor; and t_{3d} is the 3×1 translation vector. Thirdly, we generate texture T with random illumination through Equation (2), where N is calculated by the obtained S_{pose}, A is constructed from UV_A, and SHL is randomly generated. Finally, S_{pose} and T are used to render the final image by orthographic projection matrix and z-buffer.

3.5. SSPP FR

In this section, we utilize the decomposed intrinsic facial properties (i.e., albedo) and the generated arbitrary face images to solve SSPP FR problem. This is done from two perspectives: (i) enriching the intra-variation of training samples with the generated arbitrary face images; and (ii) exploring additional invariant features extracted from albedo maps and fusing the match scores of different features.

3.5.1. Face Recognition via Enriching Intra-Variation

In SSPP FR, performance of face matcher drops due to the deficiency of intra-variation. One implementation to improve SSPP face recognition is to enrich intra-variation in training set via 3D-modeling-based image generation. As shown in Figure 3a, 3D face modeling module is used to recover shapes and albedos of SSPP training set, based on which additional face images with varying pose, illumination, and expression are generated. Finally, the enlarged dataset of original and generated images is utilized to train a more discriminative face matcher.

3.5.2. Face Recognition via Enriching both Intra-Variation and Invariant Features

The 3D face modeling proposed in this paper not only makes it possible to enrich intra-variation, but also provides invariant features from its decomposed intrinsic attribute (i.e., albedo). Thus, we come up with another implementation of SSPP face recognition, which is to enrich both intra-variation and invariant features during test. Figure 3b shows the details. As can be seen, this implementation includes three matchers: original face matcher based on features of original input

images in gallery and probe, albedo matcher based on invariant features extracted from decomposed albedo of gallery and probe, and enriched face matcher based on features of enriched gallery and the recovered probe. Note that the recovered probe refers to synthetic images with the same PIE and background. The purpose of the enriched face matcher is to exploit the gallery of enriched intra-variation to suppress false rejections. Each matcher computes match scores by calculating cosine similarity of features. The match scores of the three matchers are fused together by a weighted sum rule to obtain the final match score. The identity of a probe image is finally decided as the subject whose gallery sample has the highest final match score with the probe.

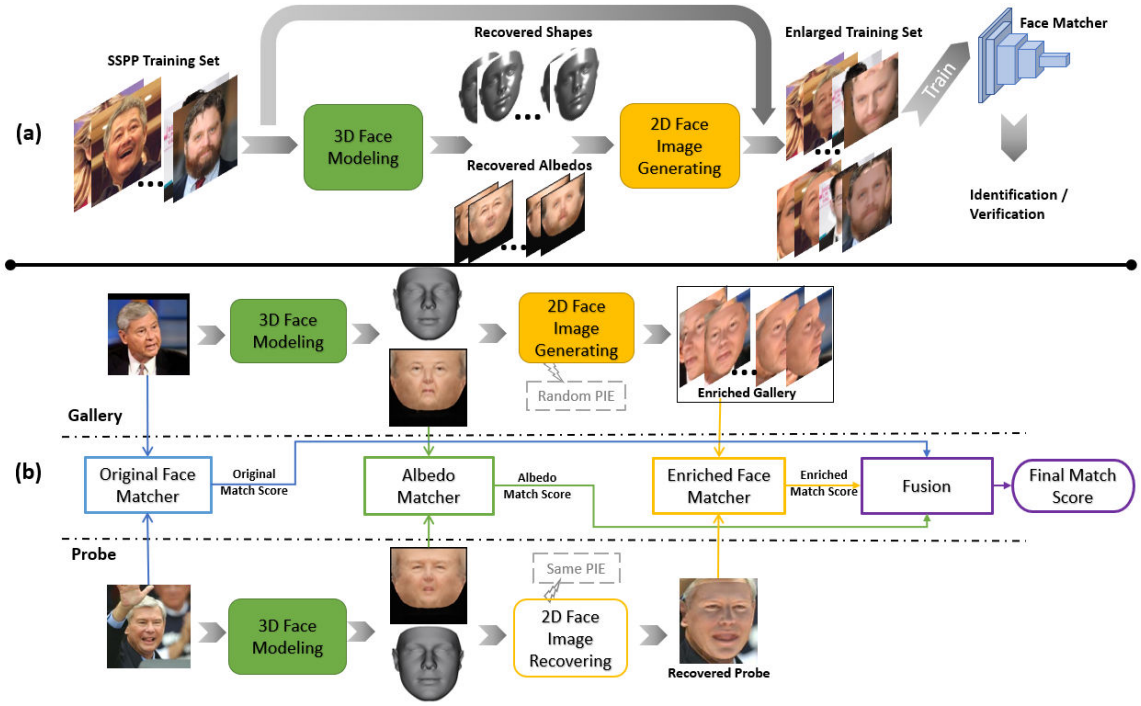

Figure 3. (**a**) Pipeline of SSPP face recognition via enriching intra-variation. (**b**) Pipeline of SSPP face recognition via enriching both intra-variation and invariant features.

3.5.3. Implementation Details

SE-Inception network [24] was employed in all the aforementioned matchers, and Stochastic Gradient Descent (SGD) with momentum coefficient of 0.9 and batch size of 64 was used to train the networks. The learning rate was initialized as 0.1, and gradually decreased by factor of 10 every 25 epochs. All face images were first aligned to the predefined template based on the five landmarks (i.e., left eye, right eye, tip of nose, left corner of mouth, and right corner of mouth) on them, and then cropped and resized to 112×112.

4. Experiments

We evaluated the effectiveness of the proposed method from two aspects: (i) visual inspection of the reconstructed albedo and shape, and the generated facial images accuracy as well; and (ii) single sample per person face recognition based on enriching intra-variation and invariant features.

4.1. Datasets and Protocols

Seven datasets were used in our experiments. Below are the details and evaluation protocols of the datasets.

FERET-b: FERET-b database [56], collected in a laboratory environment, contains 1400 images of 200 subjects with different pose, expression, and illumination. The SSPP FR experiments on this database were conducted in both verification and identification modes. In identification, we used the neutral and frontal image of each person as gallery and the remaining images as probe. In verification, we divided the database into 10 non-overlapping subsets. Each subset contained 120 positive pairs (i.e., pairs of images from the same person) and 120 negative pairs (i.e., pairs of images from different people).

LFW: Labeled Faces in the Wild (LFW) database [57] includes more than 13,000 images of 5749 different individuals taken under an unconstrained environment. The evaluation protocol suggests dividing the database into 10 non-overlapping subsets. Each subset contains 300 positive pairs and 300 negative pairs. LFW-a is a version of LFW after alignment using commercial software. For the sake of fair comparison, we selected the first 50 people who have more than 10 samples for evaluation according to the experimental setup described in [43]. We randomly selected one sample of each of the 50 subjects as gallery, and the remaining images were used as probe.

CPLFW: CPLFW [58], a renovation of LFW, constructs a cross-pose LFW database to evaluate the influence of pose variation in face recognition. It provides 10 disjoint subsets of image pairs for face verification. Each subset contains 300 positive pairs and 300 negative pairs.

CALFW: Cross-Age LFW (CALFW) database [59], similar to the CPLFW database, is a renovation of LFW. It emphasizes aging effect in addition to other variations (pose, illumination, etc.) in face recognition. The dataset is separated into 10 non-repeating subsets of image pairs for face verification, each subset containing 300 positive pairs and 300 negative pairs.

AgeDB: AgeDB [60] contains 16,488 images of various famous people, such as actors/actresses, writers, scientists, politicians, etc. There are 568 distinct subjects in AgeDB. AgeDB provides 10 folds of image pairs, with each fold consisting of 300 positive pairs and 300 negative pairs.

CFP: CFP database [61], a challenging dataset to examine the problem of frontal to profile face verification, collects 7000 images of 500 subjects with each subject having 10 frontal and 4 profile face images. Its evaluation protocol defines two separate experiments of Frontal–Profile (CFP_FP) and Frontal–Frontal (CFP_FF) face verification, and divides in each experiment the whole dataset into 10 folds each containing 350 positive pairs and 350 negative pairs.

VGGFace2-FP: The VGGFacce2 database [62] contains 3.31 million images of 9131 subjects. The dataset is divided into training set and evaluation set, in which the training set contains 8631 subjects and the evaluation set contains 500 subjects. Evaluation scenarios can be divided into two categories by pose and age. We considered the scenario of face matching across different poses, i.e., VGGFace2-FP. The evaluation data are divided into 10 folds of image pairs, with each fold consisting of 250 positive pairs and 250 negative pairs.

4.2. Visualization of Reconstructed Image Components and Generated Images

Figure 4a shows recovered 3D face shapes and complete UV_A for four input face images in LFW. It can be found that our method can not only reconstruct the albedo of the visible area but also recover the albedo of missing part. We also compared our 3D face modeling method with two existing methods [44,45]. Figure 4b shows the comparison results for two input images. As can be seen, the 3D face shape and albedo map reconstructed by Sengupta et al. [45] are incomplete and cannot be used for data augmentation, and the results recovered by Tran et al. [44] are poor in reserving identity-related appearance feature due to the limited representation capacity of PCA bases. Figure 4c shows some generated face images of varying illumination (Columns 3 and 8), pose (Columns 4 and 9), and expression (Columns 5 and 10) of four different subjects.

We also show the ability of the 3D modeling approach to address illumination and pose factors in Figure 5. As can be seen, although expressions, illuminations, and poses have influence on the appearance of input images, our method can effectively eliminate the external influence due to the

robustness of the albedo image to lighting as well as to the rigorous alignment of the face in the UV diagram representation.

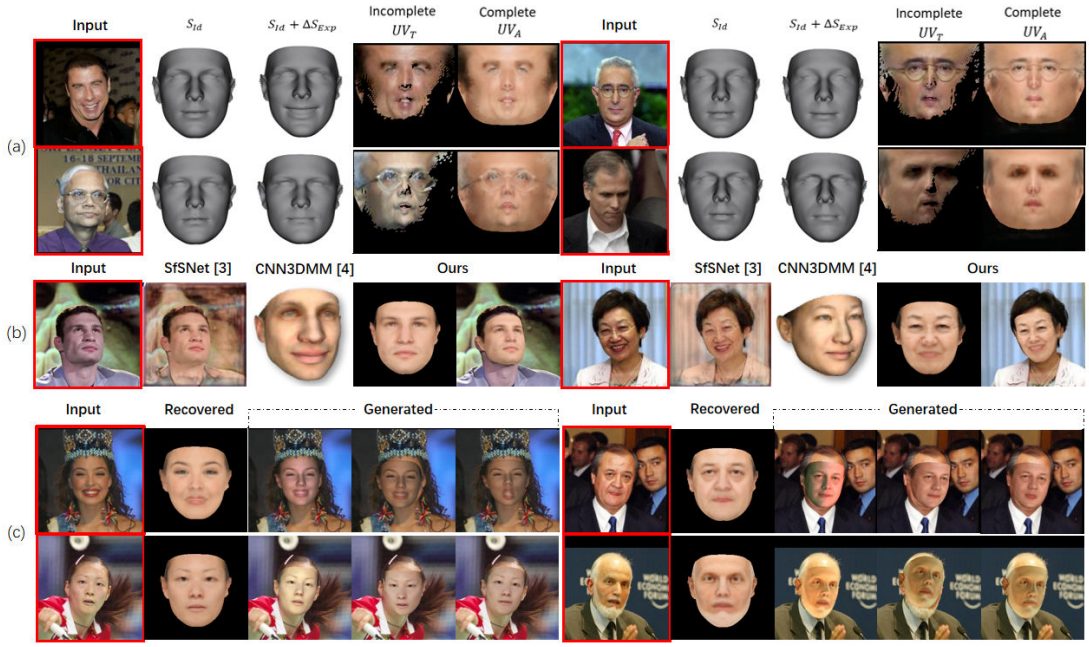

Figure 4. (**a**) Recovered 3D face shapes and albedo UV maps for two input face images by our 3D face modeling module. (**b**) Comparison between the face images reconstructed by our method and two existing methods. (**c**) Example enriched face images of varying illumination (Columns 3 and 8), pose (Columns 4 and 9), and expression (Columns 5 and 10) of four different subjects by our method.

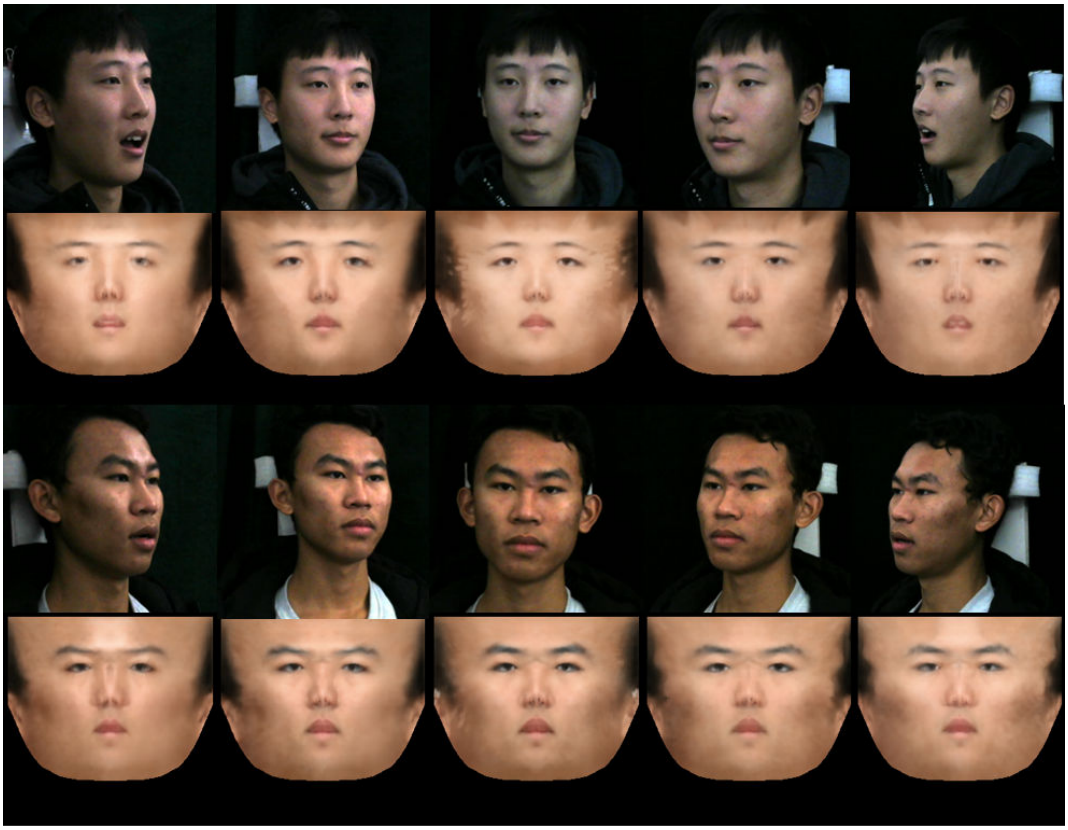

Figure 5. Albedo UV maps recovered by our 3D face modeling module for two subjects in the database constructed by ourselves with different expressions, illuminations, and poses.

4.3. Effectiveness of Enriching Intra-Variation

To validate the effectiveness of our proposed method of enriching intra-variation via 3D-modeling-based image generation, we constructed twelve groups of training data as follows.

- We randomly chose one sample per person in CelebA dataset [54] to form the original SSPP training data (denoted as *Original*). Note that face images in CelebA are in the wild images with varying PIE as well. However, their pose angles are relatively small (mostly within 30 degrees).
- The 3D modeling module was used to recover the intrinsic properties (albedo and shape) and external factors (i.e., PIE) of the face images in *Original*. We then synthesized images with the same PIE and background as the images in *Original* to form another training set, named *Synthetic*.
- With the recovered properties of the images in *Original*, we further generated another seven face images for each person with varying poses, varying illuminations, varying expressions, varying poses and illuminations, or varying poses, illuminations, and expressions. We call the respective databases of generated images as *AugP*, *AugI*, *AugE*, *AugPI*, and *AugPIE*, respectively.
- By combining the above-generated images with the original images, we obtained another five training sets, denoted as *Ori + AugP*, *Ori + AugI*, *Ori + AugE*, *Ori + AugPI*, and *Ori + AugPIE*, respectively.

We trained the implementation of our proposed method with enriching intra-variation (see Figure 3a) by using the above training sets and evaluated its face verification performance on seven of the benchmark datasets, i.e., LFW [57], CPLFW [58], CALFW [59], AgeDB [60], CFP [61], VGGFace2-FP [62], and FERET-b [56]. During the test, the features extracted from the images by the network were used to judge whether two images are from same persons based on the cosine similarity between their features. Here, ten-fold cross validation was employed. Specifically, we divided the dataset into ten subsets, nine of which were used as the training set and the remaining one used as the test set.

The results are reported in Table 1. By comparing the recognition rate obtained by training the network on *Original* and that obtained by training the network on *Synthetic*, it can be found that they are comparable. This indicates that the shape and albedo obtained by using our proposed 3D modeling method can effectively capture the identity information. When the network was trained with the generated data only, the recognition accuracies on most of the datasets were improved. Moreover, the face recognition accuracy was consistently improved when training the network with the combined set of original and generated images. It is worth mentioning that our proposed method of enriching intra-variation improved the recognition accuracy by 8.07% $(= 58.82\% - 50.72\%)$ and 13.02% $(= 82.69\% - 69.67\%)$ on the challenging AgeDB and CFP-FF test sets, respectively. It can be found that pose factor is an important factor that affects the verification rate by comparing the results of AugP, AugI, and AugE or the results of *Ori+AugP*, *Ori+AugI*, and *Ori+AugE*. While the results of using merely the augmented data (i.e., *AugP*, *AugI*, *AugE*, *AugPI*, and *AugPIE*) show consistent improvement when more variations are taken into consideration, combining the original and augmented data of expression variations, however, obtained only marginal improvement in some of the test datasets. This is probably because the original data from the CelebA database already have some expression variations and play a major role in the training, and consequently the contribution of augmented expression data becomes marginal.

Table 1. Face verification rates (%) of the first implementation of our proposed method on different benchmark datasets when different enriched face images from the CelebA database were used for training. The best result on each dataset is shown in bold.

TrainData\TestData	AgeDB	LFW	CALFW	CPLFW	CFP_FF	CFP_FP	VGGFACE2_FP	FERET-b
Original	50.75	77.33	63.03	58.7	69.67	60.77	61.46	84.33
Synthetic	53.35	73.77	59.05	57.52	69.09	59.74	60.28	79.96
AugP	52.12	75.95	58.82	60.08	71.46	61.89	61.66	87.54
AugI	50.28	75.45	59.27	57.58	71.7	60.87	61.46	80.17
AugE	50.43	75.05	57.47	58.10	71.09	62.49	61.44	80.04
AugPI	52.47	78.17	60.67	60.83	73.57	62.70	62.80	90.46
AugPIE	55.18	80.08	62.78	61.90	76.59	64.37	62.94	92.08
Ori+AugP	56.62	82.73	69.05	65.68	78.13	66.64	68.54	90.71
Ori+AugI	52.63	77.48	60.93	58.63	70.83	62.07	62.84	81.17
Ori+AugE	50.93	74.85	59.02	57.53	68.36	60.51	60.54	80.20
Ori+AugPI	**58.82**	**86.08**	**72.03**	**67.97**	81.96	**70.44**	**71.02**	94.54
Ori+AugPIE	58.65	85.98	71.02	67.15	**82.69**	70.23	70.22	**95.54**

4.4. Effectiveness of Enriching Both Intra-Variation and Invariant Features

In this experiment, we compared another implementation of our proposed method with both enriched intra-variation and enriched invariant features (see Figure 3b) with the following methods: (i) traditional methods, namely HOG+SVM [17], G-FST [41], and FT-LPP [42]; and (ii) deep learning methods, namely FaceNet [22], CDA [34], TDL [33], and KCFT [43]. For fair comparison, we used multiple samples per person database, CelebA, as the generic data to pretrain three of the matchers. To train albedo matcher and enriched face matcher, we generated the albedo image and the recovered image (synthetic image with same PIE) of CelebA by the proposed 3D face modeling module and 2D face generation module. For our proposed method, the scores from three face matchers were fused by a weighted sum rule. We set the weights for different matchers with respect to their respective recognition accuracy. Specifically, the weights were 0.45, 0.45, and 0.1 for the original matcher (O), albedo matcher (A), and enriched matcher (E), respectively, when fusing the three matchers. In the ablation study, the weights for O and A were both 0.5 when they were fused, while the weights were 0.8 and 0.2 for O and E when O and E were fused.

4.4.1. Results on FERET-b Dataset

We generated six additional images with varying pose, illumination, and expression for each subject in the gallery. Table 2 shows the recognition rates of the proposed method and the counterpart methods. Obviously, our method achieved the best recognition rates.

Table 2. Rank-1 identification rates (%) of different methods on FERET-b and LFW-a. The best result on each database is shown in bold.

Database\Method	HOG+SVM	CDA	G-FST	FT-LPP	TDL	FaceNet	KCFT	Proposed
FERET-b	50.17	77.25	82.08	86.08	89.33	91.42	93.17	**96.41**
LFW-a	40.98	51.12	57.95	62.49	74.01	89.04	**98.83**	97.25

4.4.2. Results on LFW-a Dataset

We generated 20 facial images for each person in the galley of LFW-a. The recognition rates of different methods are shown in Table 2. As can be seen, our method overwhelmedmost of the counterpart methods, and was among the top two methods.

4.4.3. Ablation Study

We further evaluated the contribution of the enriched invariant features and the enriched images on improving SSPP face recognition performance. The comparison results on FERET-b and LFW-a are shown in Table 3. It can be found that the recognition accuracy improved after we fused the match scores of the original images (denoted by *O*) with the match scores of enriched invariant features extracted from albedo (denoted by *A*) or the match scores of features extracted from the enriched images (denoted by *E*), even though the original images already set up relatively high baselines. Figure 6 shows the false rejection rates at different thresholds when fusing different match scores on the FERET-b and LFW-a databases. As can be seen, our proposed matchers *A* and *E* significantly reduced the FRR of the original matcher *O*. It is worth mentioning that both *A* and *E* helped in reducing FRR. However, according to the experimental results, *A* (using albedo-based features) was more effective than *E* (using augmented images).

Table 3. Rank-1 identification rates (%) of the proposed method when fusing the match scores of the original images (denoted by *O*) and the match scores of the enriched invariant features (denoted by *A*) or the match scores of the enriched images (denoted by *E*) on the FERET-b and LFW-a databases.

O	*A*	*E*	FERET-b	LFW-a
✓			92.25	92.46
✓	✓		96.33	97.15
✓		✓	93.17	93.53
✓	✓	✓	**96.41**	**97.25**

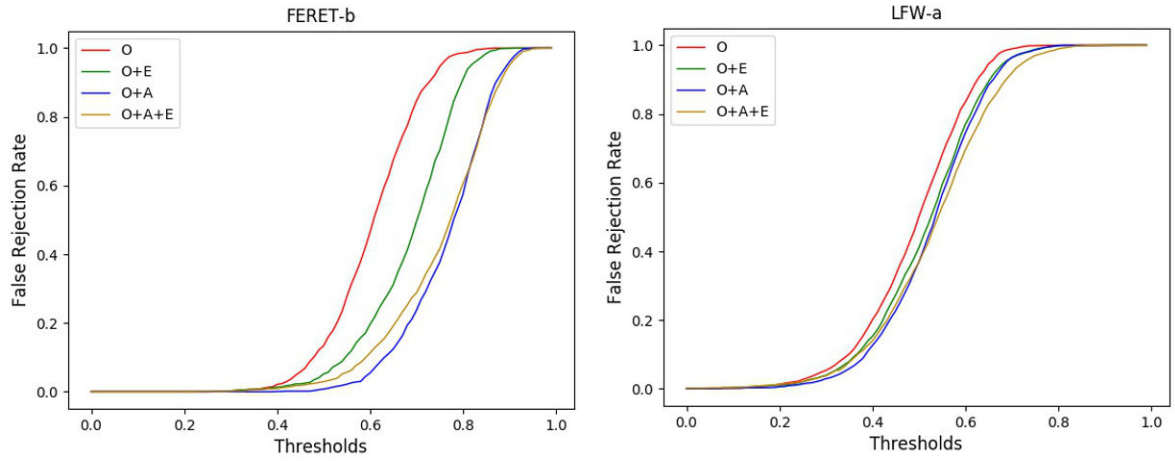

Figure 6. False rejection rates at different thresholds when fusing different match scores on the FERET-b and LFW-a databases.

5. Conclusions

A novel single sample per person face recognition algorithm based on enriching data intra-variation and invariant features is proposed in this paper. The method consists of three modules: 3D face modeling, 2D image generating, and improved SSPP FR. A novel end-to-end network is employed to recover complete albedo from the input image, which not only provides additional invariant identity features, but also can be used with the restored 3D shape to generate images containing richer intra-class variations. While existing SSPP FR methods focus either on generating synthetic images (CDA, KCFT, etc.) or on learning robust features (HOG+SVM, FaceNet, etc.), we improve the SSPP FR accuracy from the perspective of enriching both intra-variation and invariant features. Experiments were performed on multiple databases. The results show that, by using the

synthetic images generated by our proposed method to augment the train data, the SSPP FR accuracy is improved significantly by up to 13%. Moreover, using our proposed enriched invariant features boosts the rank 1 identification rate from 92.25% to 96.33% on FERET-b and from 92.46% to 97.15% on LFW-a. In the future, we are going to further improve the modeling and synthesis procedures, for example, by considering more elaborate losses and by integrating the modeling, synthesis, and identification modules in a more unified manner.

Author Contributions: Conceptualization, H.T., G.D., Q.Z., and S.W.; Data curation, H.T.; Formal analysis, H.T.; Funding acquisition, Q.Z.; Investigation, H.T.; Methodology, H.T.; Resources, Q.Z.; Supervision, G.D. and Q.Z.; Validation, H.T. and S.W.; Visualization, H.T.; Writing—original draft, H.T.; and Writing—review and editing, Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (No. 2017YFB0802300) and the National Natural Science Foundation of China (Nos. 61773270, 61971005 and 61961038).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abate, A.F.; Nappi, M.; Riccio, D.; Sabatino, G. 2D and 3D face recognition: A survey. *Pattern Recognit. Lett.* **2007**, *28*, 1885–1906. [CrossRef]
2. Hassaballah, M.; Aly, S. Face recognition: Challenges, achievements and future directions. *IET Comput. Vis.* **2015**, *9*, 614–626. [CrossRef]
3. Wang, M.; Deng, W. Deep Face Recognition: A Survey. *arXiv* **2018**, arXiv:1804.06655.
4. Hu, Z.; Gui, P.; Feng, Z.; Zhao, Q.; Fu, K.; Liu, F.; Liu, Z. Boosting Depth-Based Face Recognition from a Quality Perspective. *Sensors* **2019**, *19*, 4124. [CrossRef] [PubMed]
5. Yang, Y.; Wen, C.; Xie, K.; Wen, F.; Sheng, G.; Tang, X. Face Recognition Using the SR-CNN Model. *Sensors* **2018**, *18*, 4237. [CrossRef]
6. Koo, J.H.; Cho, S.W.; Baek, N.R.; Kim, M.; Park, K.R. CNN-Based Multimodal Human Recognition in Surveillance Environments. *Sensors* **2018**, *18*, 3040. [CrossRef]
7. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, 16–20 June 2019; pp. 4690–4699.
8. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5265–5274. [CrossRef]
9. Beymer, D.; Poggio, T.A. Face Recognition from One Example View. In Procedings of the Fifth International Conference on Computer Vision (ICCV 95), Massachusetts Institute of Technology, Cambridge, MA, USA, 20–23 June 1995; pp. 500–507. [CrossRef]
10. Tan, X.; Chen, S.; Zhou, Z.; Zhang, F. Face recognition from a single image per person: A survey. *Pattern Recognit.* **2006**, *39*, 1725–1745. [CrossRef]
11. Wu, J.; Zhou, Z. Face recognition with one training image per person. *Pattern Recognit. Lett.* **2002**, *23*, 1711–1719. [CrossRef]
12. Yin, H.; Fu, P.; Meng, S. Sampled Two-Dimensional LDA for Face Recognition with One Training Image per Person. In Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC 2006), Beijing, China, 30 August–1 September 2006; pp. 113–116. [CrossRef]
13. Lee, S.; Jung, H.; Hwang, B.; Lee, S. Authenticating corrupted photo images based on noise parameter estimation. *Pattern Recognit.* **2006**, *39*, 910–920. [CrossRef]
14. Lu, J.; Tan, Y.; Wang, G. Discriminative multi-manifold analysis for face recognition from a single training sample per person. In Proceedings of the IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, 6–13 November 2011; pp. 1943–1950. [CrossRef]
15. Abd-Almageed, W.; Wu, Y.; Rawls, S.; Harel, S.; Hassner, T.; Masi, I.; Choi, J.; Leksut, J.T.; Kim, J.; Natarajan, P.; et al. Face recognition using deep multi-pose representations. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, Lake Placid, NY, USA, 7–10 March 2016; pp. 1–9. [CrossRef]

16. Bashbaghi, S.; Granger, E.; Sabourin, R.; Bilodeau, G. Robust watch-list screening using dynamic ensembles of SVMs based on multiple face representations. *Mach. Vis. Appl.* **2017**, *28*, 219–241. [CrossRef]
17. Dadi, H.S.; Pillutla, G.K.M.; Makkena, M.L. Face Recognition and Human Tracking Using GMM, HOG and SVM in Surveillance Videos. *Ann. Data Sci.* **2018**, *5*, 157–179. [CrossRef]
18. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA, 20–26 June 2005; pp. 886–893. [CrossRef]
19. Ahonen, T.; Hadid, A.; Pietikäinen, M. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 2037–2041. [CrossRef]
20. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
21. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.E.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]
22. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015; pp. 815–823. [CrossRef]
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141. [CrossRef]
25. Deng, W.; Hu, J.; Guo, J. Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1864–1870. [CrossRef]
26. Deng, W.; Hu, J.; Zhou, X.; Guo, J. Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning. *Pattern Recognit.* **2014**, *47*, 3738–3749. [CrossRef]
27. Yang, M.; Gool, L.V.; Zhang, L. Sparse Variation Dictionary Learning for Face Recognition with a Single Training Sample per Person. In Proceedings of the IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, 1–8 December 2013; pp. 689–696. [CrossRef]
28. Su, Y.; Shan, S.; Chen, X.; Gao, W. Adaptive generic learning for face recognition from a single sample per person. In Proceedings of the Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13–18 June 2010; pp. 2699–2706. [CrossRef]
29. Cai, J.; Chen, J.; Liang, X. Single-Sample Face Recognition Based on Intra-Class Differences in a Variation Model. *Sensors* **2015**, *15*, 1071–1087. [CrossRef] [PubMed]
30. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [CrossRef] [PubMed]
31. Shao, C.; Song, X.; Feng, Z.; Wu, X.; Zheng, Y. Dynamic dictionary optimization for sparse-representation-based face classification using local difference images. *Inf. Sci.* **2017**, *393*, 1–14. [CrossRef]
32. Mohammadzade, H.; Hatzinakos, D. Projection into Expression Subspaces for Face Recognition from Single Sample per Person. *IEEE Trans. Affect. Comput.* **2013**, *4*, 69–82. [CrossRef]
33. Zeng, J.; Zhao, X.; Gan, J.; Mai, C.; Zhai, Y.; Wang, F. Deep Convolutional Neural Network Used in Single Sample per Person Face Recognition. *Comput. Intell. Neurosci.* **2018**, *2018*, 3803627:1–3803627:11. [CrossRef]
34. Zhang, Y.; Peng, H. Sample reconstruction with deep autoencoder for one sample per person face recognition. *IET Comput. Vis.* **2017**, *11*, 471–478. [CrossRef]
35. Hong, S.; Im, W.; Ryu, J.; Yang, H.S. SSPP-DAN: Deep domain adaptation network for face recognition with single sample per person. In Proceedings of the 2017 IEEE International Conference on Image Processing, ICIP 2017, Beijing, China, 17–20 September 2017; pp. 825–829. [CrossRef]
36. Cuculo, V.; D'Amelio, A.; Grossi, G.; Lanzarotti, R.; Lin, J. Robust Single-Sample Face Recognition by Sparsity-Driven Sub-Dictionary Learning Using Deep Features. *Sensors* **2019**, *19*, 146. [CrossRef] [PubMed]
37. Choi, S.; Lee, Y.; Lee, M. Face Recognition in SSPP Problem Using Face Relighting Based on Coupled Bilinear Model. *Sensors* **2019**, *19*, 43. [CrossRef] [PubMed]

38. Zhu, X.; Lei, Z.; Liu, X.; Shi, H.; Li, S.Z. Face Alignment Across Large Poses: A 3D Solution. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 146–155. [CrossRef]
39. Feng, Z.; Hu, G.; Kittler, J.; Christmas, W.J.; Wu, X. Cascaded Collaborative Regression for Robust Facial Landmark Detection Trained Using a Mixture of Synthetic and Real Images With Dynamic Weighting. *IEEE Trans. Image Process.* **2015**, *24*, 3425–3440. [CrossRef] [PubMed]
40. Song, X.; Feng, Z.; Hu, G.; Kittler, J.; Wu, X. Dictionary Integration Using 3D Morphable Face Models for Pose-Invariant Collaborative-Representation-Based Classification. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2734–2745. [CrossRef]
41. Li, L.; Ge, H.; Tong, Y.; Zhang, Y. Face Recognition Using Gabor-Based Feature Extraction and Feature Space Transformation Fusion Method for Single Image per Person Problem. *Neural Process. Lett.* **2018**, *47*, 1197–1217. [CrossRef]
42. Pan, J.; Wang, X.; Cheng, Y. Single-Sample Face Recognition Based on LPP Feature Transfer. *IEEE Access* **2016**, *4*, 2873–2884. [CrossRef]
43. Min, R.; Xu, S.; Cui, Z. Single-Sample Face Recognition Based on Feature Expansion. *IEEE Access* **2019**, *7*, 45219–45229. [CrossRef]
44. Tran, A.T.; Hassner, T.; Masi, I.; Medioni, G.G. Regressing Robust and Discriminative 3D Morphable Models with a Very Deep Neural Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 1493–1502. [CrossRef]
45. Sengupta, S.; Kanazawa, A.; Castillo, C.D.; Jacobs, D.W. SfSNet: Learning Shape, Reflectance and Illuminance of Faces 'in the Wild'. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6296–6305. [CrossRef]
46. Blanz, V.; Vetter, T. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1063–1074. [CrossRef]
47. Tu, H.; Li, K.; Zhao, Q. Robust Face Recognition with Assistance of Pose and Expression Normalized Albedo Images. In Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence, ICCAI 2019, Bali, Indonesia, 19–22 April 2019; pp. 93–99. [CrossRef]
48. Feng, Y.; Wu, F.; Shao, X.; Wang, Y.; Zhou, X. Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network. In Proceedings of the Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, 8–14 September 2018; pp. 557–574. [CrossRef]
49. Ramamoorthi, R. Modeling illumination variation with spherical harmonics—Chapter 12. *Face Process. Adv. Model. Methods* **2006**, 385–424.
50. Ramamoorthi, R.; Hanrahan, P. A signal-processing framework for reflection. *ACM Trans. Graph.* **2004**, *23*, 1004–1042. [CrossRef]
51. Bas, A.; Huber, P.; Smith, W.A.P.; Awais, M.; Kittler, J. 3D Morphable Models as Spatial Transformer Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2017, Venice, Italy, 22–29 October 2017; pp. 895–903. [CrossRef]
52. Paysan, P.; Knothe, R.; Amberg, B.; Romdhani, S.; Vetter, T. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In Proceedings of the Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009, Genova, Italy, 2–4 September 2009; pp. 296–301. [CrossRef]
53. Cao, C.; Weng, Y.; Zhou, S.; Tong, Y.; Zhou, K. FaceWarehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 413–425. [CrossRef] [PubMed]
54. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Face Attributes in the Wild. In Proceedings of the 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, 7–13 December 2015; pp. 3730–3738. [CrossRef]
55. Liu, F.; Zhao, Q.; Liu, X.; Zeng, D. Joint Face Alignment and 3D Face Reconstruction with Application to Face Recognition. *arXiv* **2017**, arXiv:1708.02734.
56. Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1090–1104. [CrossRef]
57. Huang, G.B.; Mattar, M.; Berg, T.L.; Learned-Miller, E.G. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; University of Massachusetts: Amherst, MA, USA, 2007.

58. Zheng, T.; Deng, W. *Cross-pose LFW: A Database for Studying Cross-pose Face Recognition in Unconstrained Environments*; Technical Report 18-01; Beijing University of Posts and Telecommunications: Beijing, China, 2018.
59. Zheng, T.; Deng, W.; Hu, J. Cross-Age LFW: A Database for Studying Cross-Age Face Recognition in Unconstrained Environments. *arXiv* **2017**, arXiv:1708.08197.
60. Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; Zafeiriou, S. AgeDB: The First Manually Collected, In-the-Wild Age Database. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 1997–2005. [CrossRef]
61. Sengupta, S.; Chen, J.; Castillo, C.D.; Patel, V.M.; Chellappa, R.; Jacobs, D.W. Frontal to profile face verification in the wild. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, Lake Placid, NY, USA, 7–10 March 2016; pp. 1–9. [CrossRef]
62. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. VGGFace2: A Dataset for Recognising Faces across Pose and Age. In Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2018, Xi'an, China, 15–19 May 2018; pp. 67–74. [CrossRef]