HAL open science

3D Face Verification across pose based on Euler rotation and tensors.

Ammar Chouchane, Abdelmalik Ouamane, Elhocine Boutellaa, Mebarka Belahcene, Salah Bourennane

To cite this version:

Ammar Chouchane, Abdelmalik Ouamane, Elhocine Boutellaa, Mebarka Belahcene, Salah Bourennane. 3D Face Verification across pose based on Euler rotation and tensors.. Multimedia Tools and Applications, 2017, 77 (16), pp.20697-20714. 10.1007/s11042-017-5478-z . hal-02475988

HAL Id: hal-02475988

https://hal.science/hal-02475988v1

Submitted on 9 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

3D Face Verification Across Pose Based on Euler Rotation and Tensors

Ammar Chouchane, Abdelmalik Ouamane, Elhocine Boutellaa,
Mebarka Belahcene, and Salah Bourennane

Abstract

In this paper, we propose a new approach for 3D face verification based on tensor representation. Face challenges, such as illumination, expression and pose, are modeled as a multilinear algebra problem where facial images are represented as high order tensors. Particularly, to account for head pose variations, several pose scans are generated from a single depth image using the Euler transformation. Multi-bloc local phase quantization (MB-LPQ) histogram features are extracted from depth face images and arranged as a third order tensor. The dimensionality of the tensor is reduced based on the higher-order singular value decomposition (HOSVD). HOSVD projects the input tensor in a new subspace in which the dimension of each tensor mode is reduced. To discriminate faces from different persons, we utilize the Enhanced Fisher Model (EFM). Experimental evaluations on CASIA-3D database, which contains large head pose variations, demonstrate the effectiveness of the proposed approach. A verification rate of 98.60% is obtained.

Index Terms

3D face verification, tensor analysis, multilinear dimensionality reduction, Euler angles.

I. Introduction

In real world applications, face recognition is confronted to several challenges, in terms of, for e.g., illuminations, expressions, occlusions and poses variations. This situation has highly motivated researchers from the fields of computer vision and pattern recognition to consider face recognition challenge under so called 'in the wild' conditions, referring to face images acquired

A. Chouchane, A. Ouamane and M. Belahcene are with Université de Biskra, Algeria.

E. Boutellaa is with Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland.

S. Bourennane is with Institut Fresnel, Université de Marseille, France.

in uncontrolled environments. The possibility to mitigate the influence of lighting, expression and poses variations makes 3D facial scans more effective and robust to these challenges compared to usual RGB images. This explains the recent rising interest in exploiting 3D sensing technologies for enhancing the face recognition systems.

Head pose variation and camera viewpoint have been identified as the major unsolved problems of face recognition, which attracts a great interest [1], [2]. In many face recognition applications, there is always a difference in the head inclination angles between the training images and the test image. Usually, the training images consist of a frontal face, while the test image are rotated faces with arbitrary angles in the 3D space, i.e., around X-axis Y-axis and Z-axis (roll, yaw and pitch, respectively). In the present work, we tackle the face viewpoint problem by generating several faces at different rotations using Euler transformation of 3D face scans.

In face recognition systems, facial representation plays a main role in the choice of methods and mathematical tools. Linear transformations, which are based on matrices and vectors representation of data, have been largely investigated due to their conceptual and computational simplicity [3]. Among the most famous algorithms are Principal Components Analysis (PCA) [4], Linear Discriminant Analysis (LDA) [5] and Independent Components Analysis (ICA) [6]. These linear transformations are based on matrix representation of data. However, they are not a natural way for representation of the facial data [7], since the image vectorization process cause the loss of pixels location [8]. On the other hand, multilinear transformations which are extensions of linear transformations to high order tensors representation, offer a powerful and natural way of data representation. An example a tensor representation of face images is depicted in Figure 1.

This paper adopts a 3rd-order face tensor representation that takes into account head pose variations. The Euler transform is used the generate face rotations in the range $\pm 60^{\circ}$ around the Y-axis, which form a mode of the face tensor.

The remainder of our paper is organized as follows. Section II discusses the related work in two directions, multilinear transformations and non-frontal face recognition. Section III presents the our proposed algorithm for tensor dimensionality reduction and discrimination. Section IV reminds the Euler transform we use for rotating frontal faces. Our complete face verification system is presented in Section V. Section VI introduces the experimental data and test protocol, and discusses the results. Finally, Section VII concludes the paper and outlines the future work.

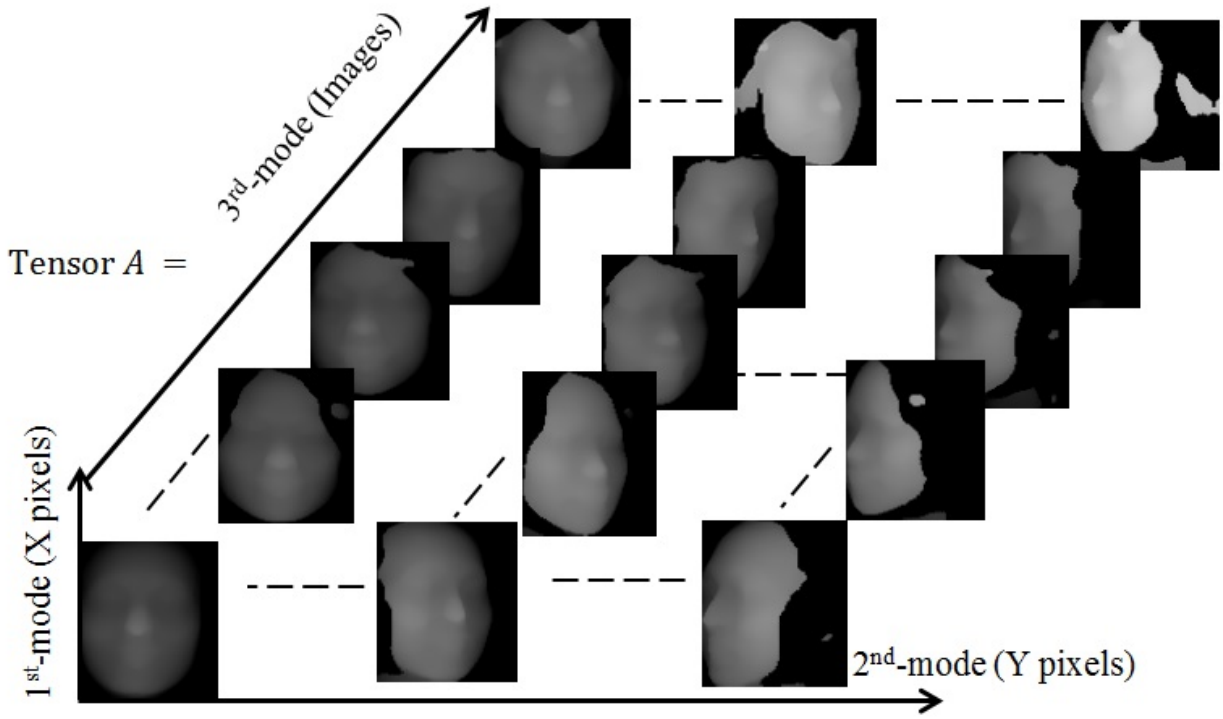

Fig. 1. Example of high order tensor representation of face images.

II. Related Work

This section reviews the related work focusing on two main directions. First, the multilinear dimensionality reduction transformations are presented. Then, the non-frontal facial recognition approaches, which deal with the pose variations, are summarized.

Recently, many multilinear subspace transformations, based on multilinear algebra, are proposed for tensor reduction and classification. Multilinear Principal Component Analysis (MPCA) [7] is a multilinear extension of PCA. MPCA projects the high order tensor to a new subspace with lower dimension by applying an orthogonal transformation on each tensor mode. Since MPCA is an unsupervised method, that do not take into consideration the class labels, it does not consider the discrimination between classes in the new tensor subspace. To deal this isue, LDA is extended to Multilinear Discriminant Analysis (MDA) [8]. MDA deal very well with the small sample size problem. However, MDA algorithm is very sensitive to parameter settings causing a converge problem [9]. In order to enhance the discrimination between classes, Lu et al. [10] proposed the uncorrelated MDA by minimizing the redundancy of features. In this

paper, we propose to use a tensor-based approach to model local face features and head pose variation. We first project the tensor to a subspace via HOSVD then employ EFM to enhance the discrimination.

Numerous algorithms have been already developed in terms of frontal face recognition when the person is looking directly the acquisition device. Recently, the problem of non-frontal face recognition has been studied. As indicated in a recent review [1], the protocols for testing facial recognition systems in the presence of poses are not yet unified; it means that we still have a long way to build an efficient face recognition system infected with pose variations. Various approaches address this problem using 2D and 3D methods. In this regard, 3D facial information has attracted more attention due to its advantages compared to 2D images [11], [12], [13], [14]. Here we interest to the methods based on 3D face scans, that we summarize in the following.

Passalis et al. [15] have manually selected face landmarks to estimate the head poses and to detect the occluded areas. A face model is applied on the scan, during assembly, facial symmetry is used to identify the missing data resulting from different head poses. The authors make the comparisons between inter-poses scans using a biometric signature based on wavelets. A limitation of this approach is that the face landmarks extraction, during the training, is done manually. A solution to solve this kind of problems is introduced by Dibeklioglu et al. [16]. The authors proposed an automatic method to detect face landmarks on 3D depth images, using curvatures and heuristic analysis. The head oose is the corrected allowing the system to identify persons with head pose variations. This method is applied to the facial images with only rotations around the Y-axis and less than $\pm 45°$. Blanz et al. [17] proposed a holistic method based on facial reconstruction by adapting the 3D Morphable Model (3DMM) to the 3D face scans. The 3DMM is constructed using PCA on 3D face shapes and textures, obtained from a laser scanner. The face recognition tests, in this paper, are performed with faces under poses less than $\pm 40°$ along the Y-axis. In [13] the pose of textured 3D face is automatically corrected based on the nose tip point using the Hotelling transformation. The approach is based on the PCA algorithm, where a rotation matrix aligns the point cloud on the principal axes. This method is evaluated on facial data with small poses variations (less than $\pm 15°$ along the Y-axis). Another interesting idea, named Rotated Profile Signatures (RPS), is presented in [18]. The RPS is used to detect the nose as well as the categorization of different poses. 3D face is turned from 0 to $180°$ with a step of $5°$. At each step, the rightmost profile points are extracted and adapted at a large variety of nose models and the result is taken as a similarity score. The minimum score indicates the

correct location of the nose. The weakness of this method is that it requires a large variety of nose models to compare and categorize different types of profiles. A similar method proposed by Rajwade et al. [19] for detection and automatic pose correction using Support Vector Regression (SVR) on the sub-band wavelet. This technique is able to accurately classify subjects with head poses of $\pm 9^{\circ}$ in both X and Y-axes.

In general, the presented approaches are based on step of pose correction where the rotation angles are estimated then an inverse rotation is applied to bring to face to a frontal pose. This process requires complex mathematical calculations on the three-dimensional coordinates of faces. Additionally, for training, the above approaches require face databases that contain different variations of poses for each individual. To deal with pose variation challenge, Euler transformation is used in our work to generate the 3D poses of a frontal face. This avoids requirement of training data with per-subject head pose variation

III. Tensor Reduction

This section introduces the dimensionality reduction methods that our approach is based on. After, giving the mathematical notations and definitions, we present the tensor projection techniques.

A. Notations and Definitions

In this paper, we adopt the following notations: tensors are denoted by italic uppercase letters, e.g., T, B; scalars by lowercase and uppercase letters, e.g., i, j, K, L; matrices by uppercase bold letters, e.g., **B, M**; and vectors by lowercase bold letters, e.g., **e**, **v**.

Tensors are a high order extension of vectors and matrices. A tensor can be thought of as a multidimensional array with N inputs, called modes. The tensor T of Nth-order is denoted by $T \in \Re^{\mathrm{I}_1 \times \cdots \times \mathrm{I}_\mathrm{N}}$. The elements of the tensor T are noted by $\mathrm{t}_{\mathrm{i}_1 \cdots \mathrm{i}_\mathrm{n} \cdots \mathrm{i}_\mathrm{N}}$, where I_n represents the dimension of mode n and $1 \leq \mathrm{i}_\mathrm{n} \leq \mathrm{I}_\mathrm{N},\ 1 \leq \mathrm{I}_\mathrm{N} \leq \mathrm{N}$.

The space of Nth-order tensor consists of N spaces for each mode. Therefore, a data set comprising N parameters may be modeled as an array of N inputs, wherein each tensor mode is associated with a different parameter. For example, the RGB image can be seen as a 3rd order tensor in which the first two modes are associated with the spatial information and the third mode represents the three color channels R, G and B.

Let A, B and $T \in \Re^{\mathrm{I}_1 \times \ldots \mathrm{I}_\mathrm{N}}$ be three tensors. We consider the following definitions:

The n-mode matrix $\mathbf{T}_{(\mathrm{n})} \in \Re^{\mathrm{I_n}(\mathrm{I_n}\times\ldots\times\mathrm{I_{n-1}}.\mathrm{I_{n+1}}\ldots\times\mathrm{I_N})}$ is formed by arranging all the n-mode vectors sequentially as well as its columns. The procedure of forming the n-mode matrices from a tensor of order $\succ 2$ is called tensor **unfolding**. The unfolding matrices in n-mode $\mathbf{T}_{(\mathrm{n})} \in \Re^{\mathrm{I_n}(\mathrm{I_n}\times\ldots\times\mathrm{I_{n-1}}.\mathrm{I_{n+1}}\ldots\times\mathrm{I_N})}$ are used for the manipulation of the tensor in this mode.

The **scalar product** of A with B is given by $\langle A|B\rangle = \sum_{\mathrm{i_1},\ldots,\mathrm{i_N}} \mathrm{a_{i_1\ldots i_N}}.\mathrm{b_{i_1\ldots i_N}}$

The Frobenius norm of B is $\|B\|^2 = \sum_{\mathrm{i_1},\ldots,\mathrm{i_N}} \mathrm{t}^2_{\mathrm{i_1\ldots i_N}} = \mathrm{tr}(\mathbf{B_n}\mathbf{B_n^T})$, where $\mathrm{tr}(.)$ is the trace operation of a matrix.

The **outer product** of vectors $\mathbf{e}^{(1)},\ldots,\mathbf{e}^{(\mathrm{N})}$ (of dimensions $\mathrm{I_1},\ldots,\mathrm{I_N}$) belonging to the vector spaces $\mathrm{E}^{(1)},\ldots,\mathrm{E}^{(\mathrm{N})}$)), respectively defines the tensor $T \in \Re^{\mathrm{I_1}\times\ldots\times\mathrm{I_N}}$, where

$$T = \mathbf{e}^{(1)}\mathbf{e}^{(2)}\ldots\mathbf{e}^{(\mathrm{N})} \tag{1}$$

The element $\mathrm{t_{i_1\ldots i_N}}$ of T is defined by the product $\mathrm{t_{i_1\ldots i_N}} = \prod_{\mathrm{n=1}}^{\mathrm{N}} \mathbf{e}^{(\mathrm{N})}_{\mathrm{i_n}}$, where $\mathbf{e}^{(\mathrm{N})}_{\mathrm{i_n}}$ is the $i^t_n h$ component of the vector $\mathbf{e}^{(\mathrm{N})}$.

The **n-mode product** of a tensor T with a matrix $\mathbf{M} \in \Re^{\mathrm{J_n}\times\mathrm{I_n}}$, denoted $T \times_n \mathbf{M}$, is $A = T \times_n \mathbf{M} \in \Re^{\mathrm{I_1}\times\ldots\times\mathrm{I_{n-1}}\times\mathrm{J_n}\times\mathrm{J_{n+1}}\times\ldots\times\mathrm{J_N}}$. This product can be written in the matrix form using the unfolding matrices $\mathbf{A_n} = \mathbf{M}.\mathbf{T_n}$, where $\mathbf{T_n}$ and $\mathbf{A_n}$ are the n-mode unfolding matrices of the tensors T and A, respectively.

B. Proposed HOSVD+EFM algorithm

Tensor decomposition is a crucial operation for the tensor representation of data and its manipulation. The main idea of tensor decomposition is to find the best approximation of a tensor $T\in \Re^{\mathrm{I_1}\times\ldots\times\mathrm{I_N}}$ by another tensor $\hat{T}$ based on optimizing a specific criterion. In the present work, the optimization criterion of the tensor decomposition is the minimization of the dimension of different tensor modes. The best approximation in the sense of least squares of a matrix **M** is obtained by truncating its singular value decomposition (SVD) [20], i.e. keeping only the k first singular vectors associated with the k largest singular values of **M**. The tensor decomposition by high order SVD (HOSVD) is a multilinear extension of SVD proposed by Lathauwer [21]. HOSVD finds the orthogonal matrices $\mathbf{U}^{(\mathrm{n})}, \mathrm{n} = 1,\ldots,\mathrm{N}$ of the input tensor data, where each matrix contains the $\mathrm{K}^1_{\mathrm{n}}$ unfolding matrix $\mathrm{T_n}$ in the n-mode of the tensor T. Using multilinear algebra, one can express any tensor T by the following decomposition:

$$K = T\times_1\mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \times_4 \ldots \times_{\mathrm{N}} \mathbf{U}^{(\mathrm{N})} \tag{2}$$

where, K is the new core tensor and $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(\mathrm{N})}$ are the orthogonal projection matrices.

HOSVD ensures the projection of the tensor to lower multilinear subspace. However, HOSVD is an unsupervised method, which does not take the data classes into account. In our case (face verification), the discrimination etween data belonging to different classes is the ultimate goal. To handle this issue, we integrate a discrimination algorithm to the original HOSVD process. We employ the Enhanced Fisher Method (EFM) [] to maximize the discrimination between the classes while reducing the tensor dimensionality. The EFM algorithm ensures a better separation of the targeted classes, by the diagonalization of the two within-class and between-class scatter matrices. In the tensor case, we unfold the tensor to each mode and label the vectors of the mode unfolding matrix with their original class label then we apply EFM to this vectors. The proposed algorithm HOSVD+EFM is presented in Algorithm 1.

IV. Face rotation by Euler transformation

To deal with head pose challenge in face recognition problem, we propose to increment the training data by generating new rotated faces from each frontal face of the subjects in the database. To this end, we utilize the Euler transform. In the 3D space, the Euler transform is given by three orthogonal matrices with a unit determinant, given by:

$$\mathbf{R}_{\mathrm{x}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) \\ 0 & sin(\alpha) & cos(\alpha) \end{pmatrix} \tag{3}$$

$$\mathbf{R}_{\mathrm{y}} = \begin{pmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{pmatrix} \tag{4}$$

$$\mathbf{R}_{\mathrm{z}} = \begin{pmatrix} cos(\gamma) & -sin(\gamma) & 0 \\ sin(\gamma) & cos(\gamma) & 0) \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

where $\mathbf{R}_{\mathrm{x}}$, $\mathbf{R}_{\mathrm{y}}$ and $\mathbf{R}_{\mathrm{z}}$ are the rotations along the axes X, Y and Z, respectively. The rotation of 3D face around X, Y and Z-axis are controlled by angles α ,β and γ, respectively. The 3D rotation matrix (3DRM) is given by:

$$\mathbf{3DRM} = (\mathbf{R}_{\mathrm{x}} \times \mathbf{R}_{\mathrm{y}} \times \mathbf{R}_{\mathrm{z}}) \tag{6}$$

Algorithm 1 HOSVD+EFM

Input:

- The tensor $T \in \Re^{\mathrm{I}_1 \times \ldots \times \mathrm{I}_\mathrm{N}}$ of $N^t h$ order: contains L classes and each class $\mathrm{C_j}$ contains $\mathrm{n_j}$ samples.
- The number $\mathrm{n_k}$ of selected singular vectors for each mode k.
- The feature dimension of EFM ($\mathrm{f_n}$).

Output:

- The projection matrices of HOSVD $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(\mathrm{N}-1)}$.
- The projection matrix $\mathbf{W}_\mathrm{E}\mathrm{FM}$ of EFM.

Algorithm:

1) For $\mathrm{n} = 1, \ldots, \mathrm{N}-1$ do
   a) Unfold the tensor T in the N-mode to obtain the matrix $\mathbf{T}_\mathrm{N}$;
   b) Perform the SVD of the matrix $\mathbf{T}_\mathrm{N}$;
   c) Extract the $\mathbf{U}^{(\mathrm{n})}$ matrices formed by the $\mathrm{k_n}$ singular vectors associated with the largest singular values of $\mathbf{T}_\mathrm{N}$;
2) Return: the reduced tensor B and the orthogonal projector matrices $B = T \times_1 \mathbf{U}^{(1)^\mathrm{T}} \times_2 \ldots \times_{\mathrm{N}-1} \mathbf{U}^{(\mathrm{N}-1)^\mathrm{T}}$.
3) Unfold the tensor B in the N-mode to obtain the matrix $\mathbf{B}_\mathrm{N}$.
4) Calculate the intra-class ($\mathbf{S}_\mathrm{W}$) and the inter-class ($\mathbf{S}_\mathrm{b}$) scatter matrices of $\mathbf{B}_\mathrm{N}$:
   a) $\mathbf{S}_\mathrm{W} = \sum_{\mathrm{j}=1}^{\mathrm{L}} \sum_{\mathrm{i}=1}^{\mathrm{n_j}} (\mathbf{B}_\mathrm{N}^{\mathrm{ij}} - \overline{\mathbf{B}_\mathrm{N}^{\mathrm{j}}})(\mathbf{B}_\mathrm{N}^{\mathrm{ij}} - \overline{\mathbf{B}_\mathrm{N}^{\mathrm{j}}})^\mathrm{T}$.
   b) $\mathbf{S}_\mathrm{b} = \sum_{\mathrm{j}=1}^{\mathrm{L}} \mathrm{n_j} (\overline{\mathbf{B}_\mathrm{N}^{\mathrm{j}}} - \overline{\mathbf{B}_\mathrm{N}})(\overline{\mathbf{B}_\mathrm{N}^{\mathrm{j}}} - \overline{\mathbf{B}_\mathrm{N}})^\mathrm{T}$
5) Calculate the new diagonalized inter-class ($\mathbf{K}_\mathrm{b}$) scatter matrices:
   a) $\mathbf{S}_\mathrm{b} = \mathbf{Y}^\mathrm{T}\mathbf{E}\mathbf{Y}$
   b) $\mathbf{S}_\mathrm{b} = \mathbf{Y}^{-1/2}\mathbf{E}^\mathrm{T}\mathbf{S}_\mathrm{b}\mathbf{E}\mathbf{Y}^{-1/2}$
6) Calculate the final EFM projection matrix ($\mathbf{W}_\mathrm{E}\mathrm{FM}$)
   a) $\mathbf{K}_\mathrm{b} = \mathbf{H}^\mathrm{T}\mathbf{O}\ \mathbf{E}$
   b) $\mathbf{K}_\mathrm{b} = \mathbf{E}\mathbf{Y}^{-1/2}\mathbf{H}$

In our case, the 3D faces are only rotated by β angles around Y-axis. Therefore, we set α and γ equal to $0°$. Thus, the **3DRM** become:

$$\textbf{3DMR} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{pmatrix} \tag{7}$$

Each point (x_i, y_i, z_i) in the 3D face is rotated by:

$$\begin{pmatrix} \mathrm{x_i^\beta} \\ \mathrm{y_i^\beta} \\ \mathrm{z_i^\beta} \end{pmatrix} = \begin{pmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{pmatrix} \begin{pmatrix} \mathrm{x_i} \\ \mathrm{y_i} \\ \mathrm{z_i} \end{pmatrix} \tag{8}$$

V. Face verification pipeline

In this section, we detail the proposed pipeline for 3D face verification based on tensor representation. An overview of the proposed system is illustrated in Figure 2. Depth face images are preprocessed and described by LPQ features, which are used to form the tensor representation. Afterward, the tensor data is projected into a lower discriminative subspace by the HOSVD+EFM algorithm. In the test phase, face matching is performed in the new subspace using cosine distance. In the following, we present the detailed steps of our method.

A. Preprocessing

The first step of the prepressing is the generation of different face poses. For each subject in the database, we use a 3D frontal face scan to generate new face poses by applying Euler transformation (see Section **??**). The 3D face scan consists in a point cloud. Each point, $\mathrm{P_i(x_i, y_i, z_i)}$, in this cloud is transformed to $\mathrm{P_i^\beta(x_i^\beta, y_i^\beta, z_i^\beta)}$ corresponding to a rotation with a specific angle. The new rotated faces cover angles around Y-axes in the rage of $-60°$ to $+60°$ with a step of $10°$. In total, the 12 rotated faces are generated for each subject. Figure 3 illustrates an example of the rotated faces.

After generating the new poses, the 3D point cloud of each face is projected into the 2D plane (depth image). We use an elliptical mask to crop the facial area. This mask is centered at the

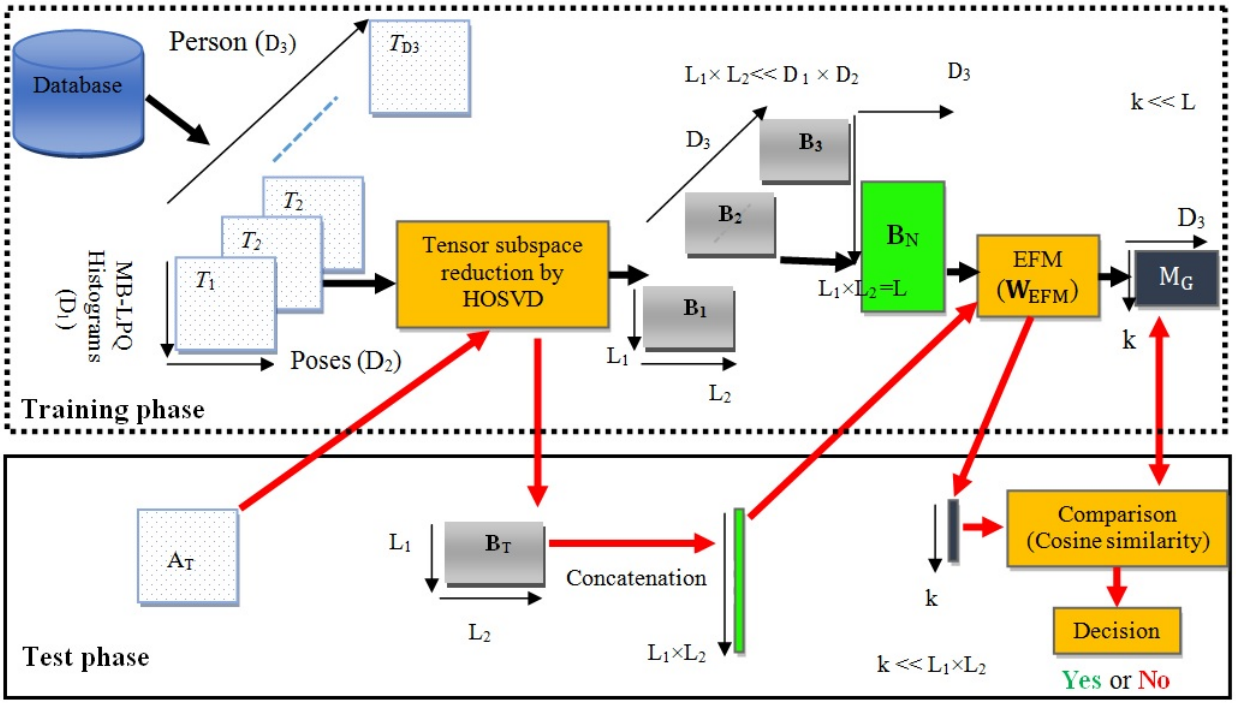

Fig. 2. Overview of the proposed approach.

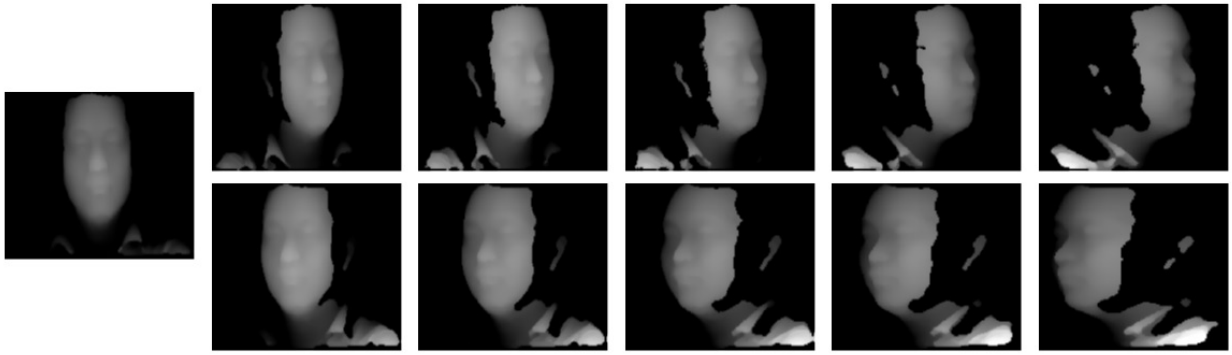

Fig. 3. Example of a frontal face scan and the generated rotated faces.

nose tip point, which is located using the maximum intensity method [11], [12]. usually, the generated face depth images contain holes and spikes. To recover the missing data, holes are filled using cubic interpolation. To eliminate the noise, in the form of spikes, we apply a median filter of 9×9 on the input depth image.

B. Feature extraction

In our work, depth face images are described using local phase quantization (LPQ) [22]. LPQ is originally proposed as a blur invariant operator that describe texture images. LPQ has been shown to perform well for face depth images [12]. This motivates our choice of this descriptor. In LPQ, the 2D Short-Term Fourier Transform (STFT) is used to extract the local phase information. STFT is calculated in a square of size M by M centered at each pixel of the depth image. Then, only four low frequency Fourier coefficients are kept, and a scalar quantizer is applied to encode the phase information of the four frequencies. The quantizer is based on the sings of the real and imaginary values of each coeficient. Thus, the eight binary digits encoding the signs are taken as integer values between 0 and 255 forming the code of the pixel. The face image described by LPQ by generating the code of each pixel. For mathematical details, we refer to [22].

In our work, we use the multi-blocks LPQ in which the histograms of the small rectangular blocks of the LPQ labelled image are used as discriminative features to represent the local face information. LPQ labelled image is subdivided into k rectangular blocks and the histograms of the blocks are concatenated $(\mathrm{h} = [\mathrm{h}_1, \mathrm{h}_2, \ldots, \mathrm{h}_\mathrm{n}])$ building a feature vector v of size n =256 $\times$ k per face image. The feature vectors of different faces are finally assembled to create the $3^r d$ order tensor data. In our case, the three tensor modes (D_1, D_2, D_3) are defined as follows:

- D_1: The face image feature vector (MB-LPQ concatenated histograms).
- D_2: Pose angles, 13 angles in the range $[-60°, +60°]$ with a step of $10°$.
- D_3: Persons, different persons with different variations of expressions and illuminations.

The feature extraction and tenor design step of our system is illustrated in Figure 4.

In our work, the input of the proposed algorithm is a 3rd order tensor $T \in \Re^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3}$ which is reduced by the multilinear method HOSVD followed by the linear method EFM.

After the detection of facial area based on nose tip and the preprocessing of the 3D depth images (see Section 4), the features of MB-LPQ histograms are extracted and stored as a $3^r d$ order tensor $T \in \Re^{\mathrm{D}_1 \times \mathrm{D}_2 \times \mathrm{D}_3}$. Hence, the poses angles for each scan in the database are generated using Euler transformation. Based on the HOSVD of T, keeping in D_1-mode (MB-LPQ histograms) the K_n singular vectors associated with the largest singular values of the unfolding matrix $\mathbf{T}_\mathrm{N}$. We get a reduced core tensor $B \in \Re^{\mathrm{L}_1 \times \mathrm{L}_2 \times \mathrm{D}_3}$, in which, $B = T \times_1 \mathbf{U}^{(1)^\mathrm{T}} \times_2 \mathbf{U}^{(2)^\mathrm{T}} \times_3 \mathbf{U}^{(3)^\mathrm{T}}$ where $(\mathrm{L}_1 \times \mathrm{L}_2 \prec\prec \mathrm{D}_1 \times \mathrm{D}_2)$. The tensor B is unfolded according to the 3rd-mode (N=3) to get the unfolding matrix $\mathbf{B}_\mathrm{N}$ of size $\mathrm{L} \times \mathrm{D}_3$. The column vectors of $\mathbf{B}_\mathrm{N}$ represents the persons and its

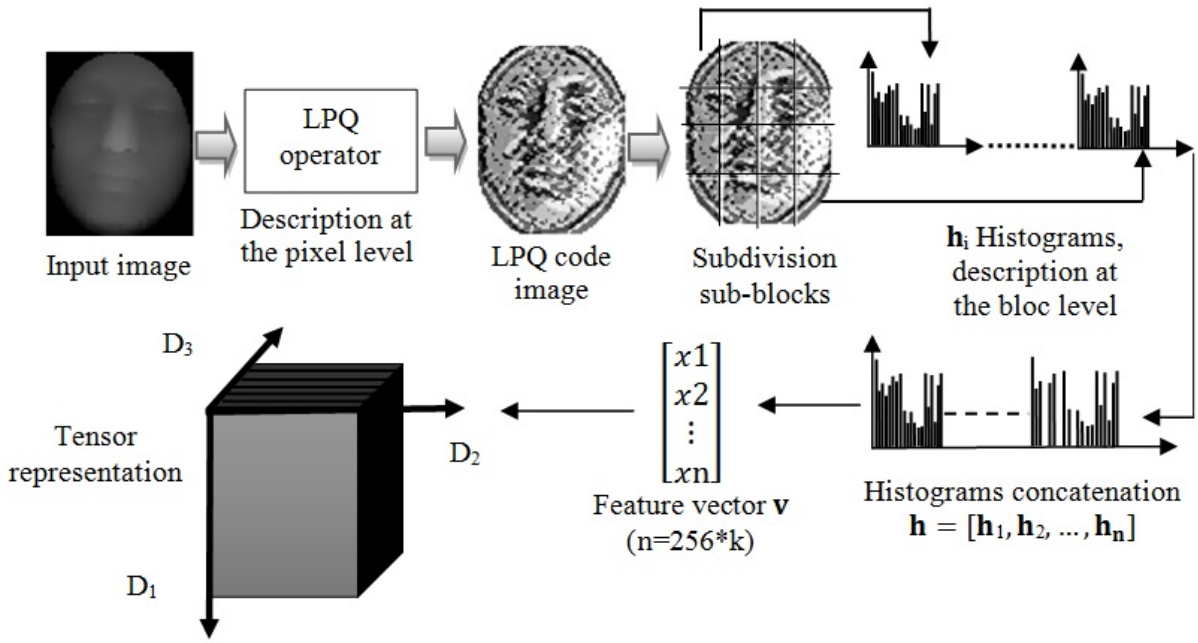

Fig. 4. Feature extraction and higher-order tensor face design.

row vectors are the features of these persons under variations of illuminations, expressions and poses.

Then, this feature space ($\mathbf{B}_{\mathrm{N}}$) is projected using EFM in order to increase the efficiency of our method through the class appearance provided by the EFM algorithm. The intra-class ($\mathbf{S}_{\mathrm{W}}$) and the inter-class ($\mathbf{S}_{\mathrm{b}}$) scatter matrices are calculated and a new inter-class scatter matrices ($\mathbf{k}_{\mathrm{b}}$) is obtained from $\mathbf{S}_{\mathrm{b}}$ (step 5 of Table 1). Subsequently, $\mathbf{k}_{\mathrm{b}}$ is diagonalized in which $\mathbf{H}$ and $\mathbf{E}$ are their eigenvectors and eigenvalues, respectively. Finally, the projection matrix EFM ($\mathbf{W}_{\mathrm{E}}\mathrm{FM}$) is calculated as presented in the step 6 (b) of Table 1. After that, a reduced training matrix $\mathbf{M}_{\mathrm{G}}$ of size $\mathrm{k} \times \mathrm{D}_3$ ($\mathrm{k} \prec\prec \mathrm{L}$) is obtained by the projection of $\mathbf{B}_{\mathrm{N}}$ in the EFM subspace using $\mathbf{W}_{\mathrm{E}}\mathrm{FM}$ matrix. This $\mathbf{M}_{\mathrm{G}}$ matrix contains significant feature vectors for each subject in the database.

In the test phase, the test image $\mathbf{A}_{\mathrm{T}}$ (candidate) is processed in the same manner as the training images and projected into HOSVD subspace, then concatenated as a vector of size $\mathrm{L}_1 \times \mathrm{L}_2$. This vector is projected into EFM subspace using $\mathbf{M}_{\mathrm{G}}$ matrix. Then, we get a reduced feature vector of size k, where $\mathrm{k} \prec\prec \mathrm{L}_1 \times \mathrm{D}_2$, for the test candidate. This test vector is matched with the training features of size $\mathrm{k} \times \mathrm{D}_3$ using the cosine similarity, where D_3 is the number of persons.

VI. Experiments

In this section, we perform a number of experiments to evaluate the proposed approach for the 3D face verification based on face representation using 3rd-order tensors. First, we describe the used CASIA-3D face database and the followed test protocol as well as the parameter settings. Then, we assess three variants of the 3D face verification system: pose concatenation using EFM, the multilinear approach using HOSVD and the proposed approach (HOSVD+EFM).

A. CASIA-3D database

We use the CASIA-3D face database **??** to evaluate the performance of the proposed approach. This database consists of 123 persons with 4624 scans captured under variations in terms of illumination, expressions and poses, using 3D Minolta VIVID 910 scanner. Additionally, for each frontal pose of a person (images from 1 to 15), 12 new depth images with different poses are generated using the Euler transformation. The test protocol subdivides the database into three parts: training, evaluation and test. The subjects are used as follows: 100 clients, 13 evaluation impostors and 10 test impostors. Table I summarizes the test protocol of our experiments on CASIA-3D face database.

TABLE I

Test protocol used for the experiments on CASIA-3D face database.

Subset	Client	Impostor
Training	6500 (65 Images/ Person)	N/A
Evaluation	6500 (65 Images/ Person)	2535 (195 Images/ Person)
Test	6500 (65 Images/ Person)	1950 (195 Images/ Person)

B. Parameter Settings

In the experiments, depth face images are resized to 90×70 pixels. Each face image is subdivided into 30 blocks, where the size of each bloc is 16×16 pixels. The histograms of these blocks are extracted and concatenated into one feature vector of dimension 7680 (30×256). For the MBLPQ descriptor, the window size is set to 5. These parameters have been empirically tuned.

The training data is used to estimate the subspace projection matrices. For each experiment, the dimension of feature vectors is varied until obtaining the best performances. In the proposed

approach, ”HOSVD + EFM”, PCA is used before the EFM to reduce the size of the histogram features after multilinear projection by HOSVD. This step is needed to make the within scatter matrix $\mathbf{S}_{\mathrm{w}}$ be non singular and the EFM is more stable in this case [12] [20]. PCA deals also with the SSS problem. The PCA subspace is defined by retaining 99% energy of the eigenvalues.

C. Results and discussion

For multilinear HOSVD approach, the face data is represented as 3rd-order tensors. Where, the dimension D_1 represents the feature vector (MB-LPQ histograms), the dimension D_2 represents different poses and the dimension D_3 represents the persons' samples with different illuminations variations and facial expressions.

The dimension of the first mode D_1 of the tensor varies between 0 and 600, the best results of multilinear approach HOSVD can be achieved with the dimensions from 400 to 600. For the linear variant, EFM, and the combined variant, HOSVD+EFM, the best performance is achieved with a features dimension under 100.

The variations of the verification rates and the equal error rates against changes in the dimension of the tensor for HOSVD method are shown in Figure 5

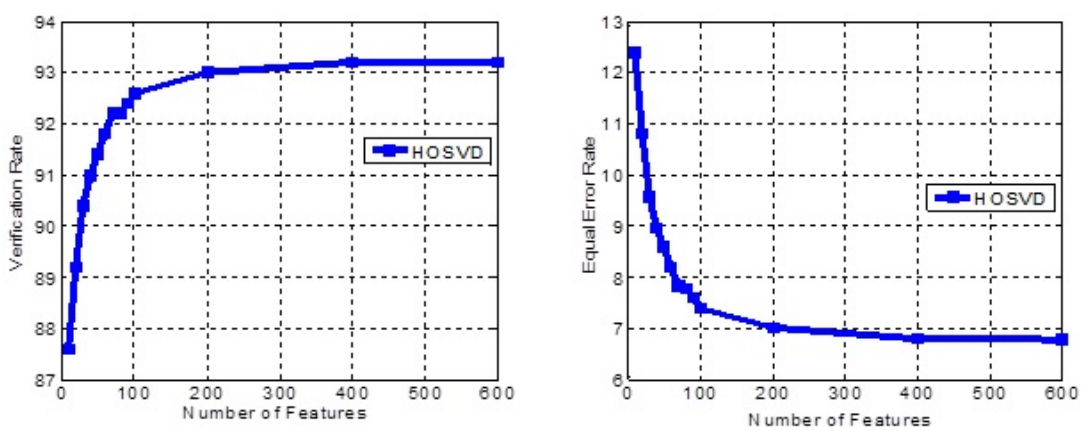

Fig. 5. VR and EER of HOSVD in function of the tensor dimension.

The variations of verification rates and equal error rates against the dimension of features vectors for EFM and HOSVD+EFM are shown in Figure 6.

According to the experimental results, it is evident that the multilinear methods based on tensors generalizes better between the evaluation (EER) and test (VR). The linear method EFM

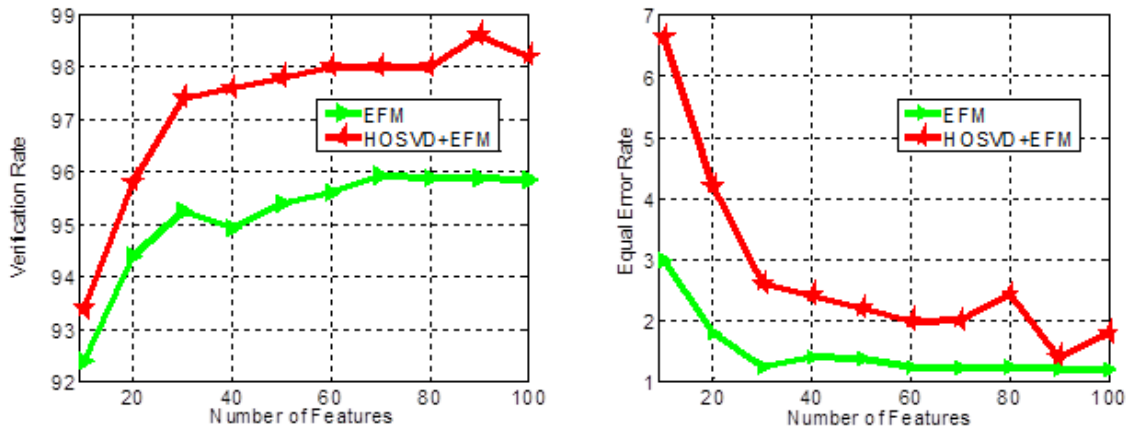

Fig. 6. VR and EER of EFM and HOSVD+EFM in function of the features dimension.

gives good results during the evaluation set (EER between 3.0% and 1.2%). Despite this, the VR is not improved in the test set (TV around 95%). Unlike for the multilinear approach, when we use the HOSVD, the results between the evaluation and test sets are more compatible. For example, HOSVD+EFM gives an EER of 1.4% in the evaluation set and a VR of 98.6% in the test set.

As can be seen in Figure 6, HOSVD+EFM gives the best performance with an equal error rate (EER) of 1.4% and a verification rate (VR) of 98.6%. These results demonstrate the robustness of the proposed method to different variability factors (illuminations, expressions and poses variations 60).

The stability of our approach HOSVD+EFM (based on the tensor representation) compared to EFM (based on the matrix representation) is obvious in Figure 6.

Table II summarizes the best performances and the computation time (CT, for the verification of one person) for each method. We emphasize the fact that the computation time is considerably reduced, which makes our method practical. When we compare the results in terms of computation time, EFM achieves a VR of 95.9% with the best CT (0.5 s). Our method HOSVD+EFM gives a VR of 98.6% with a CT of 0.6 s. Therefore, HOSVD+EFM reaches a gain of 2.4% in the verification rate with a loss of 0.1 s in terms of CT. , this is another advantage for our method.

The comparison of the three methods in terms of ROCs curves is shown in Figure 7. The ROC depicts the probability of a correct match to the false acceptance rate. The ROC curve of

TABLE II

PERFORMANCE OF PROPOSED APPROACH VARIANTS.

Method	Evaluation	Test		
	EER (%)	HTER (%)	VR (%)	CT (s)
Multilinear: HOSVD	6.8	6.0	93.2	0.6
Linear: EFM	1.2	3.5	95.9	**0.5**
Combination: HOSVD+EFM	**1.4**	**1.0**	**98.6**	0.6

our approach HOSVD+EFM has better performance and it is the more stable, which means that the verification rate of our algorithm is at each false acceptance rate greater than the verification rate of the other methods.

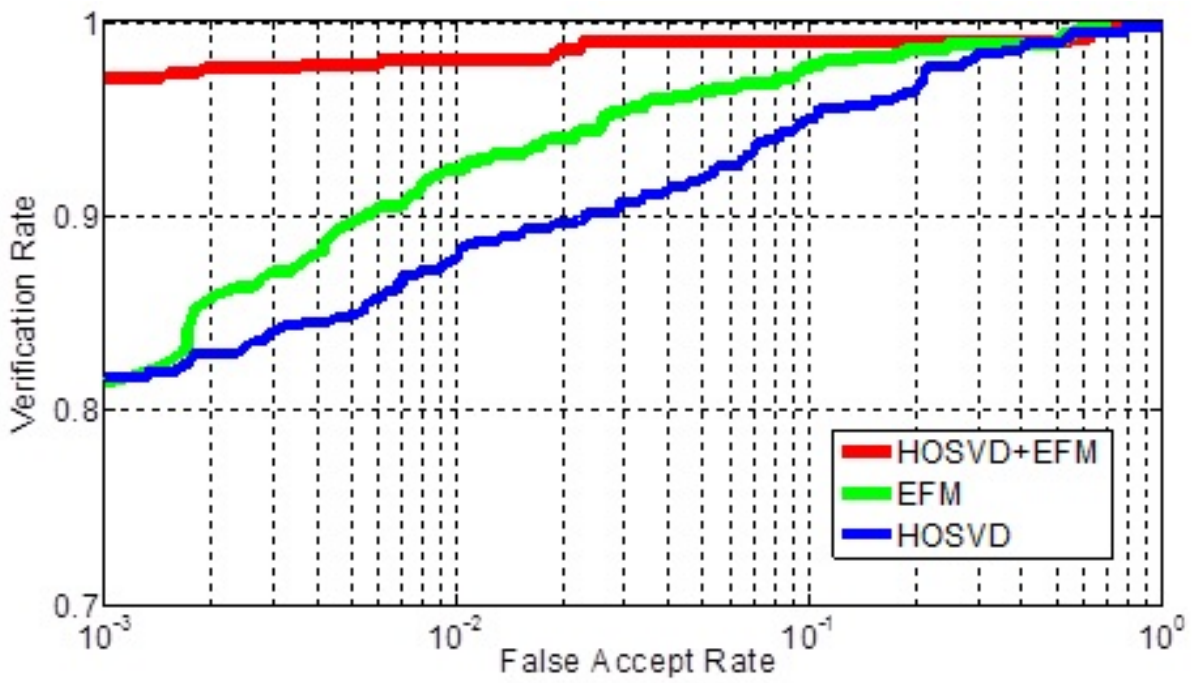

Fig. 7. ROC curves of the three methods.

The experiments demonstrates that combining two methods of dimension reduction, multilinear using HOSVD followed by linear method based on EFM, gives the best performance with a good generalization, compared to using of each method independently.

Table III compares the verification performance attained by our method HOSVD+EFM based on tensor representation of face data with the state-of-the-art methods on CASIA database.

HOSVD+EFM outperforms the literature methods based on matrix representation of face data despite the presence of large pose variations.

TABLE III

COMPARISON TO STATE OF THE ART.

Authors	Data representation	Method	VR (in %)
Xu et al. [23]	Matrix	LDA	91.0
Wang et al. [24]	Matrix	PCA	91.7
Ming [14]	Matrix	OSR	96.2
Ouamane et al. [12]	Matrix	PCA+EFM	97.2
Our method	**3rd order tensor**	**HOSVD+EFM**	**98.6**

VII. CONCLUSION

In this paper, we have proposed a new 3D face verification approach based on tensor analysis. In the training stage, Euler transformation is used to generate faces with different pose angles from each frontal 3D face to account for pose challenges at test stage. These face images are described by MB-LPQ features. Different face variations are modeled as a multilinear algebra problem, where face images with these challenges are represented as a 3rd-order tensor. This latter is decomposed using HOSVD and projected into a discriminate subspace by EFM. Promising results are obtained on CASIA-3D face database, demonstrating the effectiveness of the proposed approach, HOSVD+EFM.

Face pose generation give the possibility to train and evaluate a 3D face verification system under pose challenge using the databases that does not contain the face scans with poses variations. As future work, the generalization of our approach on other 3D databases and the investigation of alternative face tensor design are planed.

ACKNOWLEDGMENT

The authors would like to thank the Algerian Ministry of Higher Education and Scientific Research (MESRS) for supporting this work.

REFERENCES

[1] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876 – 2896, 2009.

[2] X. Liu and T. Chen, "Pose-robust face recognition using geometry assisted probabilistic modeling," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 502–509.

[3] M. A. O. Vasilescu, "A multilinear (tensor) algebraic framework for computer graphics, computer vision, and machine learning," Ph.D. dissertation, Citeseer, 2009.

[4] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991, pp. 586–591.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[6] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on neural networks*, vol. 13, no. 6, pp. 1450–1464, 2002.

[7] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.

[8] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 212–220, 2007.

[9] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H.-J. Zhang, "Human gait recognition with matrix representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 896–903, 2006.

[10] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 103–123, 2009.

[11] C. Ammar, B. Mebarka, O. Abdelmalik, and B. Salah, "Evaluation of histograms local features and dimensionality reduction for 3d face verification," *J Inf Process Syst*, vol. 12, no. 3, pp. 468–488, 2016.

[12] A. Ouamane, M. Belahcene, A. Benakcha, S. Bourennane, and A. Taleb-Ahmed, "Robust multimodal 2D and 3D face authentication using local feature fusion," *Signal, Image and Video Processing*, vol. 10, no. 1, pp. 129–137, 2016.

[13] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.

[14] Y. Ming, "Rigid-area orthogonal spectral regression for efficient 3d face recognition," *Neurocomputing*, vol. 129, pp. 445–457, 2014.

[15] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3d face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1938–1951, 2011.

[16] H. Dibeklioglu, A. A. Salah, and L. Akarun, "3d facial landmarking under expression, pose, and occlusion variations," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*. IEEE, 2008, pp. 1–6.

[17] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[18] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Rotated profile signatures for robust 3d feature detection," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–7.

[19] A. Rajwade and M. D. Levine, "Facial pose from 3d data," *Image and Vision Computing*, vol. 24, no. 8, pp. 849–856, 2006.

[20] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[21] D. Letexier, "Filtrages tensoriels adaptatifs pour la restauration d'images multidimensionnelles," Ph.D. dissertation, Aix-Marseille 3, 2009.

[22] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *International conference on image and signal processing*. Springer, 2008, pp. 236–243.

[23] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3d face recognition from depth and intensity gabor features," *Pattern recognition*, vol. 42, no. 9, pp. 1895–1905, 2009.

[24] X. Wang, Q. Ruan, and Y. Ming, "3d face recognition using corresponding point direction measure and depth local features," in *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 86–89.