The application of tensor representation and HoSVD for detecting 3D-facial shape recovery

Hamid Reza Yazdani (hamidreza.yazdani@gmail.com)
Imam Hossein Comprehensive University
Ali Reza Shojaerfard
Imam Hossein Comprehensive University

Research Article

Keywords: Facial shape recovery, HoSVD, Tensor, Tensor representation

Posted Date: April 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1524787/v1

The application of tensor representation and HoSVD for detecting 3D-facial shape recovery

Hamid Reza Yazdani[1,2*] and Ali Reza Shojaerfard[2,1†]

[1*]Department of mathematics and statistics, Imam Hossein Comprehensive University, Babaei Highway, Tehran, 1698715861, Tehran, Iran.

[2]Department of mathematics and statistics, Imam Hossein Comprehensive University, Babaei Highway, Tehran, 1698715861, Tehran, Iran.

*Corresponding author(s). E-mail(s): hamidreza.yazdani@gmail.com;
Contributing authors: ashojaeifard@ihu.ac.ir;
†These authors contributed equally to this work.

Abstract

This paper proposes a fast 3D-facial shape recovery algorithm from a single image with general, Unknown lighting. To derive the algorithm, we formulate a nonlinear least-square problem with two-parameter vectors related to personal identity and light conditions. We then combine the spherical harmonics for the surface normal of a human face with tensor algebra and show that in a particular condition, the dimensionality of the least-square problem can be further reduced to one-tenth of the regular subspace-based model by using tensor decomposition (N-mode SVD), which speeds up the computations. To enhance the shape recovery performance, we have incorporated prior information in updating the parameters. The proposed algorithm takes less than 0.4 s to reconstruct a face in the experiment and shows a significant performance improvement over other reported scheme.

Keywords: Facial shape recovery, HoSVD, Tensor, Tensor representation.

MSC Classification: 65F99 , 15A69 , 58C05

1 Introduction

Face recognition is one of the most popular applications of image analysis. In the present scenario, face recognition plays a significant role in security, personal information access, improved human-machine interaction, and personalized advertising. Hence a recognition system that, is inexpensive to use at any location, performs quicker matching, handles extensive database and do recognition in a varying environment is the need of the hour [1]. Natural images are formed by interacting multiple factors related to scene structure, illumination, and imaging. The human perception remains robust despite significant variation of these factors. Although humans are quite nice at identifying known faces, it is tough to deal with many unknown faces. This human limitation is overcome by supercomputers using strong and solid algorithms. For many applications, the performance of face recognition systems in controlled environments has now reached a satisfactory level; however, there are still many challenges posed by uncontrolled environments. Some of these challenges are posed by the problems caused by variations in illumination, face pose, expression, etc [4]. Matrix and tensor completion methods have many applications in big data analysis, prediction based on collected data, image processing, and computer vision. Incomplete, distorted, and noisy data has always been a significant challenge on the big data analysis, especially image processing [9].

A tensor is a multidimensional array. The concept of tensors was introduced by Gauss, Riemann, and Christoffel in the 19th century in the study of differential geometry. Operations with tensors have become increasingly prevalent in recent years. There are many application domains of tensors, such as data analysis, data mining, information science, image processing, and computational biology, etc. A first-order tensor is a vector, a second-order tensor is a matrix, and a tensor of order three or higher is a higher-order tensor. In many cases, it is suited to store data in higher-order multidimensional arrays, rather than as a metricized equivalent form. Multilinear algebra, the algebra of higher-order tensors, offers a powerful mathematical framework for analyzing ensembles of images resulting from the interaction of any number of underlying factors [13]. Consider the example of storing photos of different individuals taken under several lighting situations with distinct gestures and viewpoints. The use of a higher-order tensor would allow for storage of the images into a fifth-order tensor with modes of people, lighting conditions, gestures, viewpoints, and pixels [12].

Tensors (i.e., multiway arrays) provide an eective and faithful representation of structural properties of the data, especially for multidimensional data or data ensembles aected by multiple factors [7]. For instance, a video sequence can be represented by a third-order tensor with a dimensionality of $(Height \times Width \times Time)$. An image ensemble measured under multiple conditions can be represented by a higher-order tensor with a dimensionality of $(Pixel \times Pose \times Illumination)$ [11]. A face recognition system can be developed as a three-step process. Face localization/ detection is the process of extracting a specific image region as a face. This procedure has many applications like face tracking, pose estimation, or compression. Face normalization is one of the most critical issues in using a vector of geometrical features. The extracted features must be somehow normalized to be independent of position, scale, and rotation of the face in the image plane. The next step of feature extraction involves acquiring relevant facial features from the data. The feature extraction process must be efficient in terms of computing time and memory usage. The output should also be optimized for the classification step [4].

The goal of this paper is to provide a practical method, which can be applied to a single picture taken by a regular camera and which achieves good accuracy in the recovery of facial shape in a short time. When taking a picture using a typical camera in a general environment, management of the pose of a face is easy, but control of the light conditions is not. Hence, we assume that the face in an input image is in frontal pose under general light conditions, which are unknown [8]. We then formulate a non-linear least-square problem of two-parameter vectors by using the spherical harmonics for the surface normal of a human face to handle general light conditions, based on the Lambertian assumption. In order to speed up the calculation, we introduce tensor algebra and show how to reduce the dimensionality of the least-square problem [5].

In this paper, after introducing tensor completion problem, we state the tensorial face shape recovery method for image recovery, and implement it on some examples. The remainder of the paper is organized as follows. We introduce some preliminaries briey in Section 2, explain the proposed algorithm in section 3. The performance evaluation and experimental results is shown in Section 4, and nally we conclude the paper in Section 5.

2 Preliminaries

Here, we use tensor algebra and to explain for standard operations such as inner product, F-norm, and *singular value decomposition (SVD)*. Therefore, we briefly state some preliminaries for tensor calculus and tensor completion. For more details and information, please read [7], [10].

Definition 1 A tensor is a multidimensional array, with a dimensionality that is referred to as its order. X stands for a Nth-order tensor (i.e. an N-way array) which is identified as N-dimensional or N-mode tensor, too. Here, the word "order" is used for referring to the dimensionality of a tensor (like Nth-order tensor), and the word "mode" is employed for describing operations on a particular dimension (like mode-n product)[4]. We denote the set of all n-dimensional tensors of order m by $T_{m,n}$. The tensor A is called symmetric, if all $a_{i_1,\ldots,i_n}$ are invariant under any permutation of indices. The set of all real n-dimensional symmetric tensors of order m is shown with $S_{m,n}$.

Definition 2 A tensor fiber is defined as a one-dimensional section or vector obtained by fixing all indices but one. For instance, the (i, j) tube fiber is denoted by $X(i, j, :)$. Observe that a third-order tensor is a matrix of tubal fibers [3]. Fibers are generalizations of matrix columns and rows. $Mode - n$ fibers are obtained by fixing all indices but n^{th}.

Definition 3 *Mode-n matricization (unfolding)* of tensor X, denoted as $X_{(n)}$, is obtained by arranging all mode-n fibers as columns of a matrix. The precise order in which fibers are stacked as columns is not important as long as it is consistent. Figure 1 shows the fibers of 3-tensor. *The folding* is the inverse operation of matricization/unfolding. A generalization of the product of two matrices is the product of a tensor and a matrix [11].

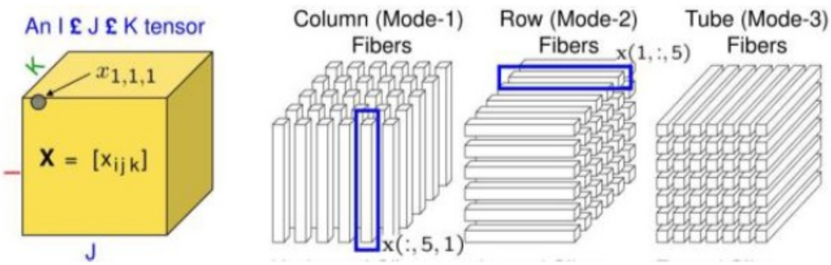


Fig. 1 Fibers of a tensor from rank 3.

Definition 4 $Mode - n$ product of tensor X and matrix A is denoted by $X \times_n A$, and defined by

$$Y = X \times_n A \Longleftrightarrow Y_{(n)} = AX_{(n)}. \tag{1}$$

This product is commutative (when applied in distinct modes), i.e.

$$(X \times_n A) \times_m B = (X \times_m B) \times_n A. \tag{2}$$

for $m \neq n$.

Definition 5 The inner product of two tensors X and Y of same size is defined as $< X, Y >$. Unless otherwise specified, we treat it as dot product defined as follows [10]:

$$< X, Y >= \sum_{i_1 \cdots i_m = 1}^{n} x_{i_1 \cdots i_m} y_{i_1 \cdots i_m}. \tag{3}$$

The F-norm of a tensor X (Generalized from matrix Frobenius norm), is defined as [10]:

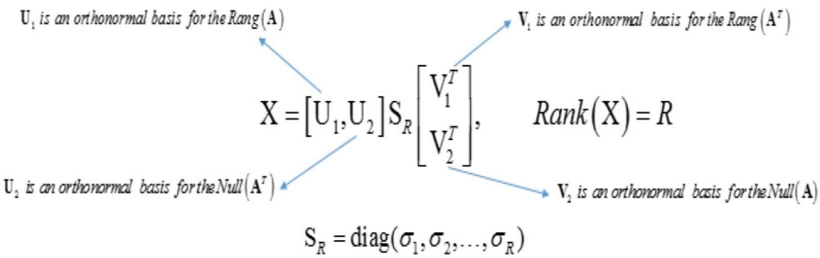
$$X_F := \sqrt{< X, X >} = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} X^2_{i_1, i_2, \cdots, i_N}}. \tag{4}$$

Definition 6 Suppose X is a symmetric tensor of $S_{m,n}$, r is a positive integer number and $u^{(k)} \in \mathbb{R}^n$ for $k \in \{1, \cdots, r\}$ exist such that

$$X = \sum_{k=1}^{r} (u^{(k)})^m. \tag{5}$$

Therefore, X is called a *completely positive tensor (CP)*, and *Eq.*(3) is a *CP-decomposition* of X (For example, see Figure 2). In the CP-decomposition of *Eq.*(5), the minimum of r is called *CP-rank* of X [5].

Definition 7 In the general, The *Singular Value Decomposition (SVD)* is a factorization of a real or complex matrix that generalizes the eigen decomposition, which only exists for square normal matrices to any $m \times n$ matrix via an extension or the polar decomposition. In the tensor calculus, similar concepts proposed as follows:



Where U_i (for $i = 1, 2$) are orthonormal and can be extended to orthonormal basis, S_R is a diagonal and positive definite of dimension R, with R as the number of non-zero eigenvalues of tensor X^*X and $V = [V_1, V_2]$ is a unitary tensor of $rank(X)$ [13]. The representation of SVD shown in figure 2.

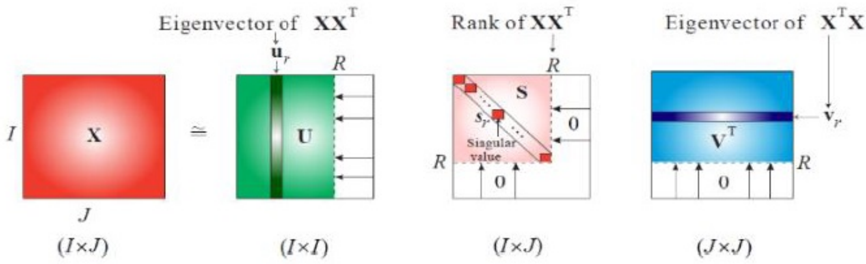


Fig. 2 The representation of SVD.

Definition 8 Given a low-rank (either CP rank or other ranks) tensor T with missing entries, the goal of completing it can be formulated as the following optimization problem:

$$Min_X rank_*(X),$$
$$Subject \quad to \quad X_\Omega = T_\Omega.$$

Where $rank_*(X)$ denotes a specific type of tensor rank based on the rank assumption of given tensor T, X represents the completed low rank tensor of T and Ω is an index set of observations. For this paper, the specific rank is completed positive (CP) rank [6]. Figure 3 compares matrix completion with tensor completion.

Tensor completion is a problem of filling the missing or unobserved entries of partially observed tensors [1]. Due to the multidimensional character of tensors in describing complex datasets, tensor completion algorithms and their applications have received wide attention and achievements in areas like data mining, computer vision, signal processing, and neuroscience. On the other hand, in practice, due to the multiway property of modern datasets, tensor completion naturally arises in data-driven applications such as image completion, image processing, and video compression.

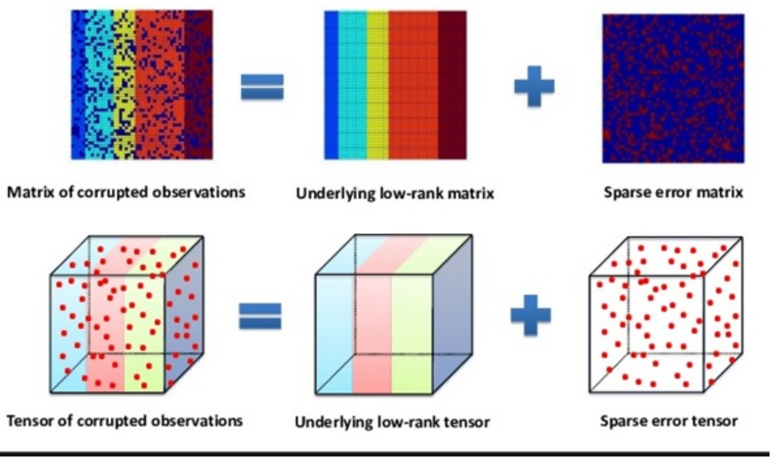


Fig. 3 The comparison scheme between matrix and tensor completion.

Definition 9 The generalized of SVD or N-mode SVD (also named Higher-order SVD) is defined as [3]:

$$D = G \times_1 U_1 \times_2 \cdots \times_N U_N. \tag{6}$$

Where G is the core tensor and U_k is derived from SVD of

$$D_{(k)} = U_k \sum_k V_k^T. \tag{7}$$

and G is defined as

$$D = G \times_1 U_1^T \times_2 \cdots \times_N U_N^T. \tag{8}$$

HoSVD of 3-tensor shown in figure 4.

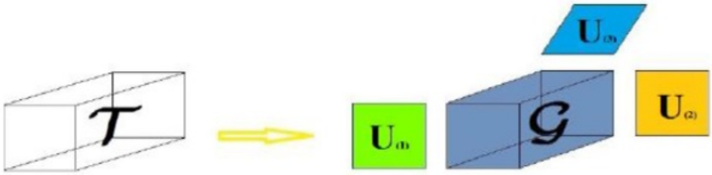


Fig. 4 HoSVD decomposition of 3-tensor.

3 Algorithm Implementation

Multilinear algebra offers a natural approach to the analysis of the multifactor structure of image ensembles and to addressing the complex problem of disentangling the constituent factors or modes [12]. An image of a human face depends on various parameters, such as its 3D structure, head pose, light and exposure, surface reflection property, etc. This picture can be approximated as linear equation i.e.,

$$I(x,y) \approx f(x,y)^T S. \tag{9}$$

Where $I(x,y)$ is the brightness of the pixel $(x,y), s \in \mathbb{R}^{n_l}$ is the light condition vector and $f(x,y)$ is a n_l-dimensional vector which is related to the surface characteristics and is either the scaled normal ($n_l = 3$) or the spherical harmonic representation ($n_l = 4$ or 9) at the pixel (x,y). Figure 5 shows the primary process of the algorithm.

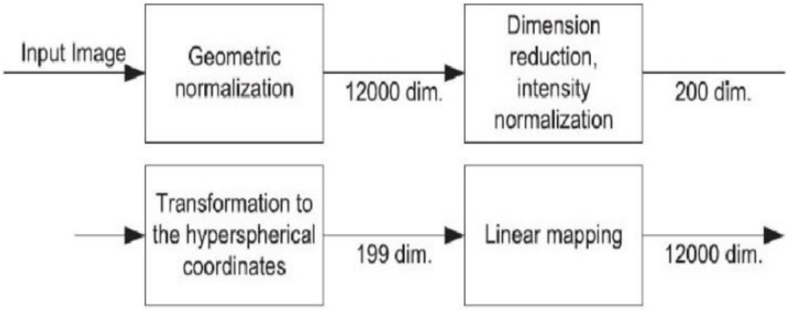


Fig. 5 The schematic view of algorithm.

To reduce the dimension of I, we apply tensor decomposition (N-mode SVD). N-mode SVD is known to be suitable for dimensional reduction when the data is in the form of a multidimensional array (e.g., an image in matrix form) rather than a vector. By applying HoSVD, we obtain a dimension reduction matrix, followed by an orthogonal process, to obtain an orthogonal matrix and the transformed basis [6]. Because geometric normalization of images and depths is essential for the accuracy of the proposed scheme, a straightforward afne transform or active appearance model can be used for this purpose [3]. Before taking any further steps, we apply an afne transform to the images and depths of all the subjects, so that the eyes and the center of the mouth are located at the designated locations in 2-D coordinates [9]. Figure 6 shows a 3D-face reconstruction. The coordinate plane part of the face is cut in 3D and represented by the tensor representation methods (with various tensor analyzes).

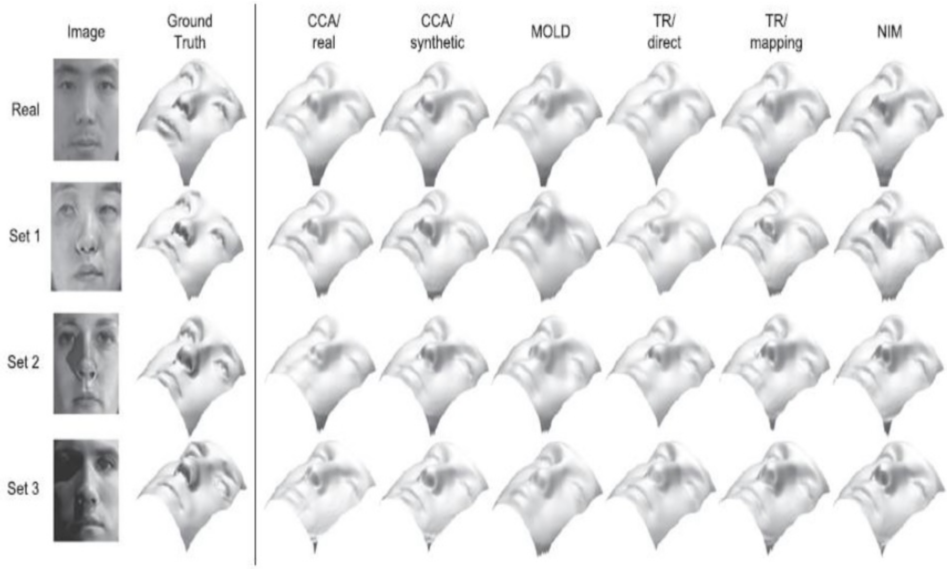


Fig. 6 The samples of 3D face reconstructions by tensorial representation methods.

The overall procedure for the proposed method has two following steps:

1. Approval: a statement which confirms that all experimental protocols were approved by a named institutional and/or licensing committee. Please identify the approving body in the methods section
2. In the first step, we apply the affine transform to the harmonic images of each training sample, so that the centers of the eyes and mouth of all pieces are located at the same positions. After that, calculate the mean tensor of the flat face (F) and apply N-mode SVD to it (Q). Now, the train image is stored by HoSVD version of the mean tensor.

3. In the second step, we apply the affine transform to a test image; The resultant image is denoted as I'. Now, calculate

$$L = U_Y^T I' U_X \tag{10}$$

In the reduced dimensional space and indicate its tensorized version as L. After that, if the norm distance between Q and L is minimized, the best result is obtained; Otherwise, search the best Q for this purpose.

In fact, the main problem is minimizing the problem of norm distance between L, Q.

4 Experimental results

For our experiments, we run **MATLAB R2020a** on the laptop system *(Asus A53sv)* with configuration as shown in table 1.

Table 1 The configuration of experiment system.

Laptop	**Asus A53sv**
CPU	Ci7-2670QM (6MB Smart Cache)
No of Cores	4 Physical/8 Thread
Frequency	2.2-3.1 GHz
RAM	16 GB (DDR3/1600 MHz)
H.D.D.	750 GB (7200 rpm)
GPU	GeForce Gt 630m (2GB/96 Cuda Cores)
Performance (FP32)	367 Gflops
O.S.	Win 10 Pro 64 bit

After importing the original picture and running the algorithm, the results in figure 7 are obtained.

For another example, we apply the algorithm on the other images by the different poses of head, background, and exposure. The result is shown in figure 8.

5 Conclusion

Prior information was used to the parameter update procedure to enhance the performance, and the experimental results showed that the proposed algorithm reconstructs a facial shape pretty accurately in less than 0.02 s under various unknown light conditions and also gives a signicant performance improvement over the other schemes. Practical experiments show that the proposed algorithm takes only a few hundredths to a few tenths of a second to reconstruct a human face, and improving performance dramatically. Therefore, based on our experiments, the efficiency, accuracy, and speed of this method for recovering the shape of the face in different exposure conditions are very high. The

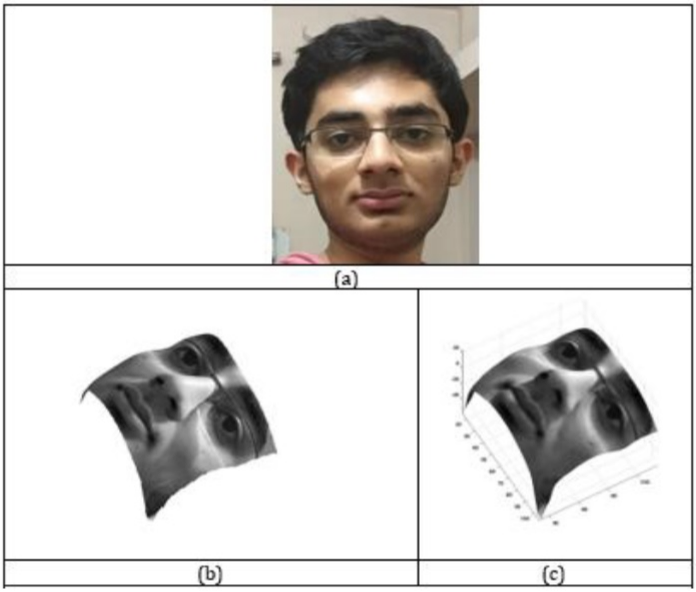


Fig. 7 The implementation of Facial Shape Recovery by Tensorial methods. (a): original photo, (b): RR-tensor representation output and (c): SIFR-tensor representation output.

image space is created by selecting a suitable nonlinear function and approximating a set of multiplicative transformed samples. The final image is the result of minimizing the sum of all errors at multiple vertices. We use various tensor analyzes such as CP and Tucker to reduce unwanted changes and improve efficiency. The experimental results show that the proposed algorithm has high reconstruction accuracy even in the presence of shadows in different exposure conditions; Also processing speed is high enough for instantaneous (real-time) applications. By using the tensor representation of the spherical harmonics for the surface standard of a human face, we have shown that the problem becomes a nonlinear least-square problem with two vectors, which are personal identity and light condition. Fast facial shape recovery can be applied in many other applications, such as 3D-face recognition.

References

[1] M. P. Beham, and S. M. M. Roomi, A REVIEW OF FACE RECOGNITION METHODS, International Journal of Pattern Recognition, 27(4) (2013), World Scientific.

[2] E. J. Candes, and B. Recht, Exact Matrix Completion via Convex Optimization, Found Comput Math, 9 (717) (2009), Springer.

[3] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (6) (2001)

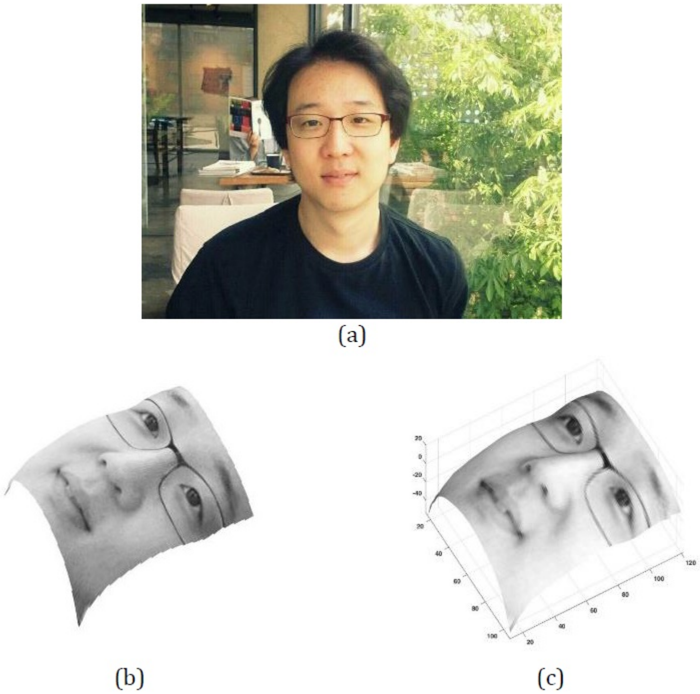


Fig. 8 The implementation of Facial Shape Recovery by Tensorial methods. (a): original photo, (b): RR-tensor representation output and (c): SIFR-tensor representation output.

681685.

[4] N. Hao, M. E. Kilmer, K. Braman, and C. Hoover, Facial Recognition using tensor-tensor decompositions, Siam journal on imaging sciences, 4 (2013), SIAM.

[5] M. Lee, C-H. Choi, Fast facial shape recovery from a single image with general, unknown lighting by using tensor representation, Pattern Recognition, 44(2011)14871496, Elsevier.

[6] M. Lee, C-H. Choi, A robust real-time algorithm for facial shape recovery from a single image containing cast shadow under general, unknown lighting, Pattern Recognition, 46 (2013) 3844, Elsevier.

[7] M. Lee, C-H. Choi, Real-time facial shape recovery from a single image under general, unknown lighting by rank relaxation, Computer Vision and Image Understanding, 120 (2014) 5969, Elsevier.

[8] J. Liu, P. Musialski , P. Wonka, and Ye J., Tensor Completion for Estimating Missing Values in Visual Data, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1 (35) (2013), 208-220.

[9] L. Qi, The Spectral Theory of Tensors, arXiv (2012): 1201.3424v1.

[10] A. Rovi, Analysis of 222 Tensors, MAI mathematics: Master thesis (2010), Linkopings University.

[11] Q. Song, H. Ge, J. Caverlee, and X. Hu, Tensor Completion Algorithms in Big Data Analytics, ACM Trans- actions on Knowledge Discovery from Data (TKDD), 6 (13) (2019), 148.

[12] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," 2002 International Conference on Pattern Recognition, 2 (2002), doi: 10.1109/ICPR.2002.1048350.

[13] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings, 2003, doi: 10.1109/CVPR.2003.1211457.