Tensor-based Factor Decomposition for Relighting

[1]Dahua Lin, [2]Yingqing Xu, [1,2]Xiaoou Tang and [1]Shuicheng Yan

[1]Department of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong SAR
Email: {dhlin4, scyan}@ie.cuhk.edu.hk

[2]Visual Computing Group
Microsoft Research Asia
Beijing, China
Email: {yqxu, xitang}@microsoft.com

***Abstract*— Lighting condition is an important factor in face analysis and synthesis, which has received extensive study in both computer vision and computer graphics. Motivated by the work on multilinear model, we propose a learning-based algorithm for relighting based on tensor framework, which explicitly accounts for the interaction of the identity factor and the lighting factor. The major contribution of our work is that we develop a novel algorithm based on a two-stage decomposition scheme to simultaneously and robustly solve for the identity parameter and the lighting parameter which are both unknown. Equipped with the decomposition algorithm, the capability of the tensor model is significantly extended. Experiment results illustrate the effectiveness of our algorithm.**

I. Introduction

Lighting variation is one of the most challenging problems in computer vision and computer graphics. Recent years, a variety of models and algorithms have been proposed to address the problem [1] [2] [3]. However, in many methods, the attention is focused on the effect of illumination itself and the interaction between identity factor and lighting factor is neglected. Actually, the effect of the same lighting condition varies from person to person. To effectively tackle the lighting variation problem, both the identity factors and the lighting factors should be analyzed under a unified framework.

Tensor Algebra [4] is a generalization of linear algebra, which offers a powerful mathematical framework to model the communication of multiple factors. Vasilescu et. al introduced the multilinear analysis to face modelling, which is called tensor face [5] [6] and demonstrated the promising application of the model. In this model, each face image is explicitly formulated as the compound effect of multiple factors including identity, illumination, pose and expression in an elegant mathematical form.

Our analysis reveals that in the multilinear analysis model, two problems have been addressed: (1)given a set of training samples, the factor parameters associated with each sample are obtained in the training process; (2)with all factor parameters of a sample known, we can synthesize the sample by tensor product. However, there remains an unsolved problem which seriously limits the application of the multilinear model: when a new image is given with all its factor parameters unknown, how to solve the factor parameters? The multilinear model in itself does not tell us.

To solve the problem of extracting factor parameters for a new sample, we develop a two-stage tensor-based factor decomposition algorithm based on careful analysis of the multilinear model. In the initialization stage, the K-nearest Equation Construction strategy and the Rank-1 approximation of matrix are integrated to give a robust estimation of initial parameter values. In addition, Constrained Alternate Least Square is employed to further optimize the estimation in the second stage. Encouraging results acquired in a set of experiments on relighting convincingly demonstrate the effectiveness of our framework.

II. Analysis on Multilinear Model

A. Tensor Algebra

To make the discussion clear, we first briefly introduce the basic concepts and notations in tensor algebra [4]. A *tensor*, also known as a n-mode matrix, is a higher order generalization of matrix. The *order* of tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is N. The *mode-n product* of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and a matrix $\mathbf{M} \in \mathbb{R}^{J_n \times I_n}$ is a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \ldots I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N}$, denoted as $\mathcal{A} \times_n \mathbf{M}$, can be calculated as

$$b_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} m_{j_n i_n}, \quad (1)$$

where $a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N}$ is the entries of $\mathcal{A}$, $m_{j_n i_n}$ is the entry of $\mathbf{M}$, and $b_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N}$ is the entries of $\mathcal{B}$. The tensor matrix product satisfies commutability:

$$\mathcal{A} \times_m \mathbf{U} \times_n \mathbf{V} = \mathcal{A} \times_n \mathbf{V} \times_m \mathbf{U} \qquad m \neq n. \quad (2)$$

In a tensor, a *mode-n vector* is an I_n-dimensional vector formed by varying index i_n with all other indices fixed. We can arrange all the mode-n vectors in order of indices into an $I_n \times (I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N)$ matrix, which is called *mode-n flatten matrix* of $\mathcal{A}$, denoted as $\mathbf{A}_n$. It has been

0-7803-9134-9/05/$20.00

shown that [4] the flatten matrices and tensor product take on an important relation:

$$\mathcal{B} = \mathcal{A} \times_n \mathbf{M} \Leftrightarrow \mathbf{B}_{(n)} = \mathbf{M}\mathbf{A}_{(n)}. \quad (3)$$

Singular Value Decomposition (SVD) also has a higher-order generalization: *N-Mode SVD* [4], which expresses a tensor $\mathcal{D}$ as follows:

$$\mathcal{D} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \ldots \times_N \mathbf{U}_N, \quad (4)$$

where $\mathcal{C}$ is a *core tensor* coordinating the interaction of *mode matrices*: $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_N$. Core tensor assumes some special properties including all-orthogonality and subtensor- ordering. Mode matrices are all orthonormal.

B. Analysis of Multilinear Face Model

In this paper, we adopt a multilinear model explicitly accounting for the identity factor and the lighting factor. Training samples are arranged into a data tensor $\mathcal{D}$, which is a $d \times I_{lig} \times I_{id}$ tensor. Here, d is the sample vector dimension, I_{lig} is number of lighting conditions, I_{id} is number of persons. Total $I_{lig} \times I_{id}$ samples are grouped as follows: each mode-1 vector is a sample, mode-2 index indicates the type of lighting condition, mode-3 index indicates the person identity.

The multilinear analysis model is denoted as

$$\mathcal{D} = \mathcal{C} \times_1 \mathbf{U}_{fea} \times_2 \mathbf{U}_{lig} \times_3 \mathbf{U}_{id}. \quad (5)$$

Here $\mathcal{C}$ is the core tensor that coordinates the interaction of identity parameters and lighting parameters. $\mathbf{U}_{fea}$ is the mode matrix spanning the sample vector space, $\mathbf{U}_{lig}$ is the mode matrix spanning the lighting parameter space, $\mathbf{U}_{id}$ is the mode matrix spanning the identity parameter space.

To acquire a deeper insight into the mechanism of the multilinear model, we delve into the mathematical representation. First, we can simplify (5) to

$$\mathcal{D} = \mathcal{S} \times_2 \mathbf{U}_{lig} \times_3 \mathbf{U}_{id}. \quad (6)$$

Here, we call $\mathcal{S} = \mathcal{C} \times_1 \mathbf{U}_{fea}$ *Bases Tensor*. Denote the mode-1 vector of $\mathcal{S}$ with mode-2 index p and mode-3 index q as $\mathbf{s}_{pq}$, and the mode-1 vector from $\mathcal{D}$ with mode-2 index i and mode-3 index j as $\mathbf{d}_{ij}$. According to the formation of $\mathcal{D}$, $\mathbf{d}_{ij}$ is also the sample vector in i-th lighting condition from j-th person. Then we have:

$$\mathbf{d}_{ij} = \sum_{p=1}^{I_{lig}} \sum_{q=1}^{I_{id}} a_{ip} b_{jq} \mathbf{s}_{pq}. \quad (7)$$

From (7), we have the following observations:

1) Each sample is a linear combination of a set of bases with the weights determined by both identity parameters and lighting parameters.
2) The combination weights for the sample in the i-th lighting condition from the j-th person depends only on the i-th row of $\mathbf{U}_{lig}$ and the j-th row of $\mathbf{U}_{id}$. Denote the i-th row vector of $\mathbf{U}_{lig}$ as $\mathbf{v}_i^{(lig)T}$, and the j-th row vector of $\mathbf{U}_{id}$ as $\mathbf{v}_j^{(id)T}$. We have

$$\mathbf{d}_{ij} = \mathcal{S} \times_2 \mathbf{v}_i^{(lig)T} \times_3 \mathbf{v}_j^{(id)T}. \quad (8)$$

With $\mathbf{v}_i^{(lig)}$ fixed, the lighting condition of the sample will also be fixed, and with $\mathbf{v}_j^{(id)}$ fixed, the identity of the sample will not change either. In this sense, we can say that for sample $\mathbf{d}_{ij}$, $\mathbf{v}_i^{(lig)T}$ is its lighting parameter, and $\mathbf{v}_j^{(id)T}$ is its identity parameter.

Formula (8) can be extended to the samples not in the training set to synthesis a sample with lighting parameter $\mathbf{v}^{(lig)}$ and identity parameter $\mathbf{v}^{(id)}$ given

$$\mathbf{d} = \mathcal{S} \times_2 \mathbf{v}^{(lig)T} \times_3 \mathbf{v}^{(id)T}. \quad (9)$$

Though we have a simple way for synthesis, however, how to solve $\mathbf{v}^{(lig)}$ and $\mathbf{v}^{(id)}$ with a new sample $\mathbf{d}$ given remains a challenging problem.

III. Two-Stage Tensor-based Decomposition

A. Formulation

In mathematics, the problem of solving factor parameters can be formulated as the following optimization problem:

$$(\hat{\mathbf{v}}^{(lig)}, \hat{\mathbf{v}}^{(id)}) = \operatorname*{argmin}_{\mathbf{v}^{(lig)}, \mathbf{v}^{(id)}} ||\mathbf{d} - \mathcal{S} \times_2 \mathbf{v}^{(lig)T} \times_3 \mathbf{v}^{(id)T}||^2. \quad (10)$$

The objective function is biquadratic, so there is no general closed form solution for the problem. In addition, $\mathbf{v}^{(lig)}$ and $\mathbf{v}^{(id)}$ are multidimensional vector, hence the optimization procedure is sensitive to initial estimation and vulnerable to local minima. To address the optimization problem, we develop a novel algorithm: *Two-Stage Tensor-based Factor Decomposition*, which effectively exploits the mathematical form of tensor product.

B. First Stage: Initialization

Because $\mathbf{U}_{lig}$ and $\mathbf{U}_{id}$ are both orthonormal matrices, we can introduce $\mathbf{c}_{lig} = \mathbf{U}_{lig}\mathbf{v}_{lig}$ and $\mathbf{c}_{id} = \mathbf{U}_{id}\mathbf{v}_{id}$. This can also be written as $\mathbf{v}_{lig}^T = \mathbf{c}_{lig}^T \mathbf{U}_{lig}$ and $\mathbf{v}_{id}^T = \mathbf{c}_{lig}^T \mathbf{U}_{id}$. Here we regard the factor parameter vectors as linear combination of factor parameter vectors of training samples, $\mathbf{c}_{lig}$ and $\mathbf{c}_{id}$ serve as combination coefficients. Then the synthesis formula can be written in terms of $\mathbf{c}_{lig}$ and $\mathbf{c}_{id}$ as

$$\mathbf{d} = \mathcal{S} \times_2 (\mathbf{c}_{lig}^T \mathbf{U}_{lig}) \times_3 (\mathbf{c}_{id}^T \mathbf{U}_{id}), \quad (11)$$

$$\mathbf{d} = (\mathcal{S} \times_2 \mathbf{U}_{lig} \times_3 \mathbf{U}_{id}) \times_2 \mathbf{c}_{lig}^T \times_3 \mathbf{c}_{id}^T. \quad (12)$$

According to (5), (6) and (12) we have

$$\mathbf{d} = \mathcal{D} \times_2 \mathbf{c}_{lig}^T \times_3 \mathbf{c}_{id}^T. \quad (13)$$

Denote the i-th entry of $\mathbf{u}_{lig}$ as $c_{lig}(i)$, and that of $\mathbf{u}_{id}$ as $c_{id}(i)$, then Eq.13 can be expanded to

$$\mathbf{d} = \sum_{i=1}^{I_{lig}} \sum_{j=1}^{I_{id}} c_{lig}(i) c_{id}(j) \mathbf{d}_{ij}. \quad (14)$$

We denote $c_{ij} = c_{lig}(i) c_{id}(j)$, arrange all c_{ij} values to an $I_{lig} \times I_{id}$ dimensional vector $\mathbf{c}$ and flatten $\mathcal{D}$ along mode-1 to a $d \times (I_{lig} \times I_{id})$ matrix $\mathbf{D}$. Then Eq(14) can be written in matrix form as

$$\mathbf{d} = \mathbf{D}\mathbf{c}. \quad (15)$$

However, in a large training set, it is possible that $d < I_{lig} \times I_{id}$. In such cases, Eq(15) is underdetermined equations, and $\mathbf{c}$ cannot be uniquely solved. To address the problem, we propose a strategy called *K-Nearest Equation Construction*, which states that each sample should be constructed by only K most relevant samples with relevance measured in terms of Euclidean distance in sample space. In detail, we select the K column vector from $\mathbf{D}$ that are nearest to $\mathbf{d}$ to form a $d \times K$ matrix $\mathbf{D}_K$. Then the K weights of these vectors can be obtained by solving the overdetermined equations in least square error

$$\mathbf{d} = \mathbf{D}_K \mathbf{c}_K, \tag{16}$$

while the weights of other vectors are set to zeros. Here we impose a constraint to $\mathbf{c}$ and thus $\mathbf{c}_K$: all coefficients are positive and their sum equals to 1. We call it *Unitary Positive Constraint* Intuitively speaking it is reasonable constraint when the coefficients are viewed as ratio of contributions to construction.

With K-Nearest Equation Construction and Unitary Positive Constraint, we can easily solve $\mathbf{c}$ by linear least square methods. Then we arrange the long vector $\mathbf{c}$ into an $I_{lig} \times I_{id}$ matrix $\mathbf{C}$, with entries $c_{ij} = c_{lig}(i)c_{id}(j)$. This can be written in outer product form as

$$\mathbf{C} = \mathbf{c}_{lig}\mathbf{c}_{id}^T. \tag{17}$$

Then $\mathbf{c}_{lig}$ and $\mathbf{c}_{id}$ can be solved under following formulation

$$(\mathbf{c}_{lig}, \mathbf{c}_{id}) = \underset{\mathbf{c}_{lig},\mathbf{c}_{id}}{\operatorname{argmin}} ||\mathbf{C} - \mathbf{c}_{lig}\mathbf{c}_{id}^T||^2. \tag{18}$$

It is also a biquadratic problem. However, this problem can be solved based on rank-approximation. The theory of singular value decomposition states that: any matrix can be rewritten as a linear combination of outer product terms

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i. \tag{19}$$

Here, r is the rank of matrix $\mathbf{A}$, $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th left-singular-vector and right-singular-vector respectively. In addition, it has been shown in matrix theory that the best rank-1 approximation of a matrix, i.e. least-square-error outer product approximation, is $\lambda_1\mathbf{u}_1\mathbf{v}_1$, here λ_1 is the largest singular value.

Hence, $\mathbf{c}_{lig}$ and $\mathbf{c}_{id}$ can be solved by SVD. However, SVD can only determine the vectors up to scale. Here we impose the constraint $\mathbf{c}_{id}^T\mathbf{1} = 1$ to $\mathbf{c}_{id}$, then the two vectors can be uniquely solved as

$$\mathbf{c}_{lig} = (\mathbf{v}_1^T\mathbf{1})(\lambda_1\mathbf{u}_1), \tag{20}$$

$$\mathbf{c}_{id} = (\mathbf{v}_1^T\mathbf{1})^{-1}(\mathbf{v}_1). \tag{21}$$

C. Second Stage: Constrained ALS Optimization

Reviewing the process of initialization, we can find that it is in essence a relax-by-impose procedure: relax the rank-1 constraint on $\mathbf{C}$ in the first step and re-impose the rank-1 constraint in the second step. Though in each step, the result is optimum, however, we cannot guarantee that the final result obtained by cascading the two steps are also optimum.

To further optimize the coefficients, we develop *Constrained Alternate Least Square (CALS)*as follows:

We denote $\mathbf{c}_{lig}$ and $\mathbf{c}_{id}$ obtained in the t-th iteration as $\mathbf{c}_{lig}^{(t)}$ and $\mathbf{c}_{id}^{(t)}$, and those obtained in initialization stage as $\mathbf{c}_{lig}^{(0)}$ and $\mathbf{c}_{id}^{(0)}$. The following three steps are repeated until the reconstruction error becomes stable:

1) Keep $\mathbf{c}_{id}^{(t-1)}$ fixed, then the optimization objective function is converted to $||\mathbf{d} - \mathbf{B}_{lig}^{(t)}\mathbf{c}_{lig}^{(t)}||$, where $\mathbf{B}_{lig}^{(t)} = (\mathcal{S} \times_3 \mathbf{c}_{id}^{(t-1)})_{(1)}$. Then update the lighting parameters as

$$\mathbf{c}_{lig}^{(t)} = \operatorname{argmin} ||\mathbf{d} - \mathbf{B}_{lig}^{(t)}\mathbf{c}_{lig}^{(t)}||^2 = (\mathbf{B}_{lig}^{(t)T}\mathbf{B}_{lig}^{(t)})^{-1}\mathbf{B}_{lig}^{(t)T}\mathbf{d}. \tag{22}$$

2) Keep $\mathbf{c}_{lig}^{(t)}$ fixed, then update the identity parameters

$$\mathbf{c}_{id}^{(t)} = \underset{\mathbf{c}_{id}^{(t)T}\mathbf{1}=1,\mathbf{c}_{id}>0}{\operatorname{argmin}} ||\mathbf{d} - \mathbf{B}_{id}^{(t)}\mathbf{c}_{id}^{(t)}||^2, \tag{23}$$

where $\mathbf{B}_{id}^{(t)} = (\mathcal{S} \times_2 \mathbf{c}_{lig}^{(t)})_{(1)}$, The unitary positive constraint $\mathbf{c}_{id}^{(t)T}\mathbf{1} = 1, \mathbf{c}_{id} > 0$ is imposed based on the rationale that the coefficients for combining persons reveals the grade of membership.

3) After the final solutions of coefficients $\mathbf{c}_{lig}$ and $\mathbf{c}_{id}$ are obtained, and the corresponding parameters $\mathbf{v}_{lig}$ and $\mathbf{v}_{id}$ can be computed as

$$\mathbf{v}_{lig} = \mathbf{U}_{lig}^T\mathbf{c}_{lig} \qquad \mathbf{v}_{id} = \mathbf{U}_{id}^T\mathbf{c}_{lig}. \tag{24}$$

IV. Applications and Experiments

In this section, we discuss the application of the multilinear model and our Tensor-based Factor Decomposition algorithm to the problem of lighting effect processing. For each application, a set of experiments are designed to test the performance of our framework.

All these experiments are conducted on the PIE Database [7] and share the following configuration: In all 69 subjects with each subject having images captured under 24 different lighting conditions, samples in 12 lighting conditions of all subjects are used for training the multilinear model, while the other 12 lighting conditions are used for testing.

A. Lighting Cloning

Lighting cloning is to "clone" the effect of lighting condition from one face image to another face image. With our framework, this can be accomplished as follows: for two input images, their first 300 PCA coefficients are extracted, and denoted as $\mathbf{x}_1$ and $\mathbf{x}_2$ respectively. With a trained multilinear model, the lighting parameters and identity parameters of these two samples are solved by Two-Stage Tensor-based

Factor Decomposition, denoted as $\mathbf{v}_{lig1}$, $\mathbf{v}_{id1}$, $\mathbf{v}_{lig2}$ and $\mathbf{v}_{id2}$ respectively.

To transplant the lighting condition in the second face image to the face image of the first person, we can synthesize a new image as

$$\mathbf{x}_{clone} = \mathcal{C} \times_1 \mathbf{U}_{pix} \times_2 \mathbf{v}_{lig2}^T \times_3 \mathbf{v}_{id1}^T. \quad (25)$$

The results obtained with different methods are compared in Fig.1. We can see that without proper initialization, the final results are erroneous. It is illustrated in column (c) and column(d) that the our initialization scheme can give robust initial estimations and the optimization stage based on CALS further enhances the quality of cloned images.

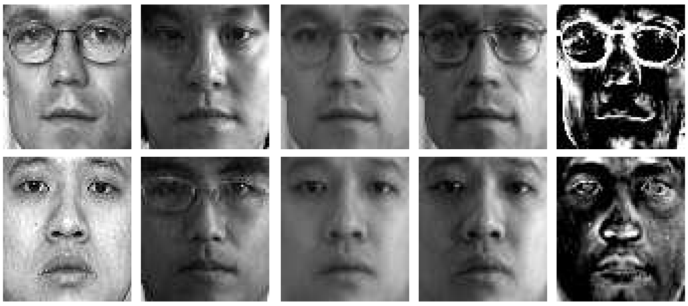

Fig. 1. Results of Light Cloning: Column 1: 1st image, Column 2: 2nd image, Column 3: Results obtained after initialization, Column 4: Results obtained by the two-stage algorithm, Column 5: Results obtained with random initialization

B. Recognition under Lighting Variation

As mentioned above, the identity parameter vector $\mathbf{v}_{id}$ reflects the intrinsic characteristics of a person which is invariant with lighting change. In our method, the identity parameter vector for each sample is extracted and compared. The similarity of two samples are measured in the identity parameter space in Euclidean distance.

The following table compare the recognition performance of our method with some representative methods. The comparative results convincingly demonstrate the superiority of our algorithm over others on face recognition.

method	recognition accuracy
PCA	74.2%
LDA	93.3%
Tensor-based Decomposition	100%

TABLE I

THE COMPARATIVE RESULTS

V. CONCLUSION

In this paper, a new algorithm for extracting identity parameters and lighting parameters are proposed. Its effectiveness and robustness are sufficiently reflected by a set of experiments. Moreover the significance of our work is not confined to modelling the relation of the identity and the lighting condition, analysis of other factors such as pose and expression can also use our framework.

ACKNOWLEDGEMENTS

The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong SAR.

REFERENCES

[1] P. N. Belhumeur, J. P. Hespanda, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class-specific linear projection," *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[2] A. S. Georghiades and P. N. Belhumeur, "Illumination cone models for faces recognition under variable lighting," *Proceedings of CVPR'98*, 1998.

[3] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based rerendering and recognition with varying illuminations," *IEEE Trans. PAMI*, vol. 23, no. 2, pp. 129–139, 2001.

[4] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix ANAL. APPL.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[5] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," *Proceedings of ECCV'02*, 2002.

[6] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," *Proceedings of ICPR'02*, vol. 2, pp. 511–514, 2002.

[7] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," .