Color Face Tensor Factorization and Slicing for Illumination-Robust Recognition

Yong-Deok Kim and Seungjin Choi

Department of Computer Science
Pohang University of Science and Technology
San 31 Hyoja-dong, Nam-gu, Pohang 790-784, Korea
{karma13,seungjin}@postech.ac.kr

Abstract. In this paper we present a face recognition method based on multiway analysis of color face images, which is robust to varying illumination conditions. Illumination changes cause large variations on color in face images. The main idea is to extract features with minimal color variations but with retaining image spatial information. We construct a tensor of color image ensemble, one of its coordinate reflects color mode, and employ the higher-order SVD (a multiway extension of SVD) of the tensor to extract such features. Numerical experiments show that our method outperforms existing subspace analysis methods including principal component analysis (PCA), generalized low rank approximation (GLRAM) and concurrent subspace analysis (CSA), in the task of face recognition under varying illumination conditions. The superiority is even more substantial in the case of small training sample size.

1 Introduction

Face recognition is a challenging pattern classification problem, which is encountered in many different areas such as biometrics, computer vision, and human computer interaction (HCI). Cruxes in practical face recognition systems result from varying illumination conditions, various facial expressions, pose variations, and so on. Of particular interest in this paper is the case of varying illumination conditions in color face image ensemble.

Various approaches and methods have been developed in face recognition. Subspace analysis is one of the most popular techniques, demonstrating its success in numerous visual recognition tasks such as face recognition, face detection and tracking. Exemplary subspace analysis methods include singular value decomposition (SVD), principal component analysis (PCA), independent component analysis (ICA), nonnegative matrix factorization (NMF), and Fisher linear discriminant analysis (LDA). All these methods seek a linear representation of face image ensemble such that basis images and encoding variables are learned, satisfying a certain fitting criterion with each face image represented by a vector.

Face images are formed by the interaction of multiple factors related to illuminations, color information, various poses, facial expressions, identities. Portion of such information are embedded in 2D spatial structure. Thus, face images,

intrinsically fit in multiway representation, known as *tensor*, reflecting various interactions between different modes. However, aforementioned subspace analysis methods are confined to at most 2-way representation. For example, color face images are converted to gray scale-valued vectors. This vectorization is a typical pre-processing in conventional subspace methods. It leads to high-dimensional vectors with losing some spatial structure, which results in a curse of dimensionality problem, making such methods to suffer from the small sample size problem.

Recently, there are a great deal of studies on multiway analysis in computer vision. These include 2D-PCA [1], generalized low rank approximation (GLRAM) [2], concurrent subspace analysis (CSA) [3], tensor faces [4] which employs the multilinear SVD (a.k.a HOSVD) [5,6], and multilinear ICA [7]. The basic idea of tensor analysis goes back to Tucker decomposition [8,9]. See [10] for a recent review of tensor factorization.

In the case of color face images, illumination change yields large variations on color in face images, even through they have exactly the same pose and facial expression. In this paper, we present a tensor factorization-based method for illumination-robust feature extraction in a task of color face image recognition. The method is referred to as *color face tensor factorization and slicing* (CFTFS). We form a 4-way tensor whose coordinates are associated with rows and columns of face images, color, and samples. The CFTFS employs the multilinear SVD in the 4-way tensor, simultaneously analyzing subspaces corresponding to rows, columns, and color. Then it chooses slices where the information about variations on row and column modes remain but variations on color mode are minimized. We demonstrate that CFTFS outperforms existing methods, including PCA, GLRAM, and CSA. The useful behavior of the method becomes more substantial, especially in the case of small training sample size.

The rest of this paper is organized as follows. In the next section we describe a brief overview of tensor algebra and Tucker decomposition. The proposed method, CFTFS is presented in Sec. 3. In Sec. 4, numerical experimental results are presented, showing that CFTFS is an indeed effective method for illumination-robust face recognition. Finally conclusions are drawn in Sec. 5.

2 Background: Multiway analysis

2.1 Tensor algebra

A tensor is a multiway array of data. For example, a vector is 1-way tensor and a matrix is 2-way tensor. The N-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ has N indices $(i_1, i_2, \ldots, i_N)$ and its elements are denoted by $x_{i_1 i_2 \ldots i_N}$ where $1 \leq i_n \leq I_n$. Mode-n vectors of an N-way tensor $\boldsymbol{\mathcal{X}}$ are I_n-dimensional vectors obtained from $\boldsymbol{\mathcal{X}}$ by varying index i_n while keeping the other indices fixed. In matrix, column vectors are referred to as mode-1 vectors and row vectors correspond to mode-2 vectors.

The mode-n vectors are column vectors of the matrix $\boldsymbol{X}_{(n)}$ which is the mode-n matricization (matrix unfolding) of the tensor $\boldsymbol{\mathcal{X}}$. The mode-n matriciza-

tion of $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is denoted $\boldsymbol{X}_{(n)} \in \mathbb{R}^{I_n \times I_{n+1} I_{n+2} \cdots I_N I_1 I_2 \cdots I_{n-1}}$ where $I_{n+1} I_{n+2} \cdots I_N I_1 I_2 \cdots I_{n-1}$ is the cyclic order after n. The original index i_{n+1} runs fastest and i_{n-1} slowest in the columns of the matrix $\boldsymbol{X}_{(n)}$. Pictorial illustration of the mode-n matricization of a 3-way tensor is shown in Fig. 1.
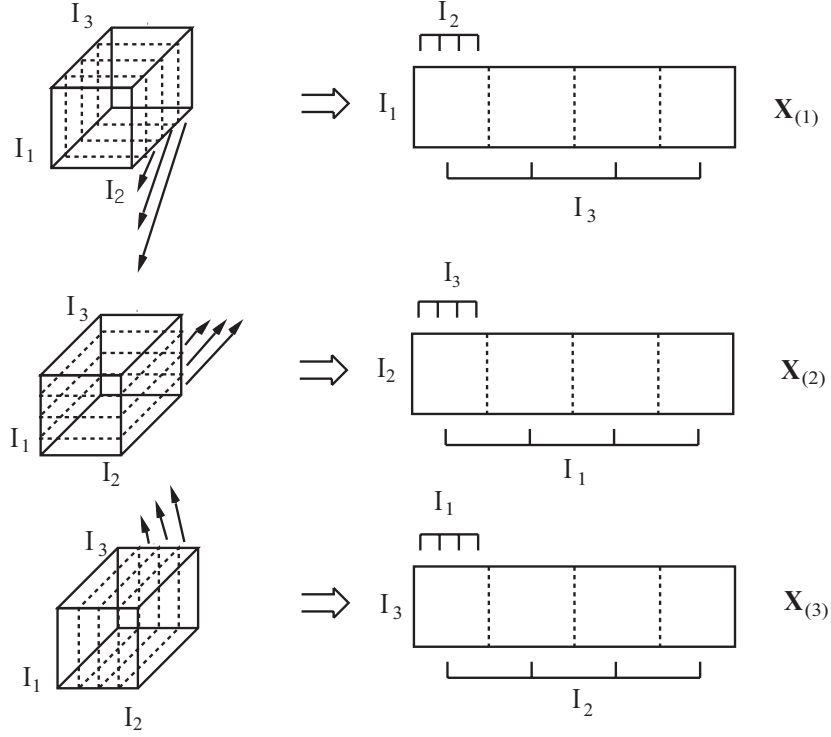

Fig. 1. Matricization of a 3-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ leads to $\boldsymbol{X}_{(1)} \in \mathbb{R}^{I_1 \times I_2 I_d}$, $\boldsymbol{X}_{(2)} \in \mathbb{R}^{I_2 \times I_3 I_1}$, and $\boldsymbol{X}_{(3)} \in \mathbb{R}^{I_3 \times I_1 I_2}$ which are constructed by the concatenation of frontal, horizontal, vertical slices, respectively.

The scalar product of two tensors $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as $\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i_1, i_2, \ldots, i_N} x_{i_1 i_2 \cdots i_N} y_{i_1 i_2 \cdots i_N}$. The Frobenius norm of a tensor $\boldsymbol{\mathcal{X}}$ is given by $\|\boldsymbol{\mathcal{X}}\| = \sqrt{\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}} \rangle}$.

The mode-n product of a tensor $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_n \times \cdots \times J_N}$ by a matrix $\boldsymbol{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ is defined by

$$\left[\boldsymbol{\mathcal{S}} \times_n \boldsymbol{A}^{(n)} \right]_{j_1 \cdots j_{n-1} i_n j_{n+1} \cdots j_N} = \sum_{j_n=1}^{J_n} s_{j_1 \cdots j_{n-1} j_n j_{n+1} \cdots j_N} a_{i_n j_n}, \qquad (1)$$

leading to a tensor $\boldsymbol{S} \times_n \boldsymbol{A}^{(n)} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times I_n \times \cdots \times J_N}$. With the mode-$n$ product, a familiar matrix factorization $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ is written as $\boldsymbol{X} = \boldsymbol{S} \times_1 \boldsymbol{U} \times_2 \boldsymbol{V}$ in the tensor framework.

2.2 Tucker decomposition

The Tucker decomposition seeks a factorization model of an N-way tensor $\boldsymbol{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ as mode products of a core tensor $\boldsymbol{S} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ and N mode matrices $\boldsymbol{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$,

$$\boldsymbol{X} \approx \boldsymbol{S} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \cdots \times_N \boldsymbol{A}^{(N)} \tag{2}$$

$$x_{i_1 i_2 \cdots i_N} \approx \sum_{j_1, j_2, \ldots, j_N} s_{j_1 j_2 \cdots j_N} a_{i_1 j_1}^{(1)} a_{i_2 j_2}^{(2)} \cdots a_{i_N j_N}^{(N)}. \tag{3}$$

Usually, mode matrices are constrained to orthogonal for an easy interpretation and there is no loss of fit. The pictorial illustration of the Tucker decomposition is shown in Fig. 2.
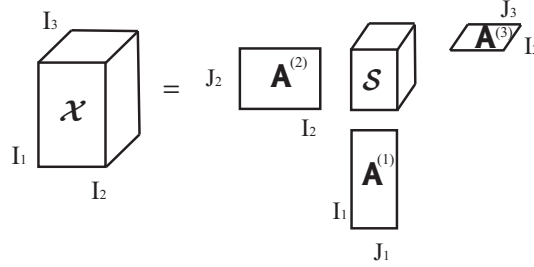

Fig. 2. The 3-way Tucker decomposition.

In Tucker decomposition, all modes don't need to be analyzed. If mode-n is not analyzed, then associated component matrix $\boldsymbol{A}^{(n)}$ becomes $I_n \times I_n$ identity matrix $\boldsymbol{I}_{I_n}$. Now, we introduce new terminology 'N-way Tucker M decomposition' where *N-way* means we incorporate N-way tensor and M is the number of analyzed modes. This new terminology explains PCA, 2D-PCA, GLRAM, CSA, and HOSVD in general framework. Table. 1 summarizes each method using this new terminology in the case of general multiway data and face image ensemble.

Suppose that $\boldsymbol{\Omega}$ is the set of analyzed modes. If all mode matrices are orthogonal, minimizing least square discrepancy between the data and model in Eq. (2) is equivalent to maximizing the function:

$$\mathcal{G}(\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \ldots, \boldsymbol{A}^{(N)}) = \|\boldsymbol{X} \times_1 \boldsymbol{A}^{(1)\top} \times_2 \boldsymbol{A}^{(2)\top} \cdots \times_N \boldsymbol{A}^{(N)\top}\|^2, \tag{4}$$

over $\boldsymbol{A}^{(n)}$ for all $n \in \boldsymbol{\Omega}$. In the Tucker decomposition, Eq. (4) has no closed form solution, except for PCA, so local solution is found iteratively with Alternative

Data type	Method	Tucker decomposition	Remark
N-way tensor	PCA	N-way Tucker 1	$\boldsymbol{A}^{(n)} = \boldsymbol{I}_{I_n}$ for $n \in \{1, \ldots, N-1\}$
	CSA	N-way Tucker $N-1$	$\boldsymbol{A}^{(N)} = \boldsymbol{I}_{I_N}$
	HOSVD	N-way Tucker N	$J_n \leq I_n$ for all n
Face image ensemble	PCA	3-way Tucker 1	$\boldsymbol{A}^{(1)} = \boldsymbol{I}_{I_1}, \boldsymbol{A}^{(2)} = \boldsymbol{I}_{I_2}$
	2D-PCA	3-way Tucker 1	$\boldsymbol{A}^{(1)} = \boldsymbol{I}_{I_1}, \boldsymbol{A}^{(3)} = \boldsymbol{I}_{I_3}$
	GLRAM	3-way Tucker 2	$\boldsymbol{A}^{(3)} = \boldsymbol{I}_{I_3}$
	CSA	4-way Tucker 3	$\boldsymbol{A}^{(4)} = \boldsymbol{I}_{I_4}$

Table 1. Tucker decomposition explains PCA, 2D-PCA, GLRAM, CSA, and HOSVD in general framework. Assume the last mode is associated with samples. The grayscale face image ensemble constructs the 3-way tensor(rows, columns, samples) and color face image ensemble does the 4-way tensor(rows, columns, color, samples).

Least Squares(ALS). In each step, only one of the component matrix is optimized, while keep others fixed. Suppose $n \in \boldsymbol{\Omega}$ and $\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(n-1)}, \boldsymbol{A}^{(n+1)}, \ldots, \boldsymbol{A}^{(N)}$ are fixed. Then Eq. (4) is reduced to a quadratic expression of $\boldsymbol{A}^{(n)}$, consisting of orthonormal columns. We have $\mathcal{G}(\boldsymbol{A}^{(n)}) = \|\mathcal{K}^{(n)} \times_n \boldsymbol{A}^{(n)\top}\|^2 = \|\boldsymbol{A}^{(n)\top} \boldsymbol{K}_{(n)}\|^2$, where

$$\mathcal{K}^{(n)} = \boldsymbol{\mathcal{X}} \times_1 \boldsymbol{A}^{(1)\top} \cdots \times_{n-1} \boldsymbol{A}^{(n-1)\top} \times_{n+1} \boldsymbol{A}^{(n+1)\top} \cdots \times_N \boldsymbol{A}^{N\top}, \quad (5)$$

and $\boldsymbol{K}_{(n)}$ is the mode-n matricization of $\mathcal{K}^{(n)}$. Hence the columns of $\boldsymbol{A}^{(n)}$ can be found as an orthonormal basis for the dominant subspace of the column space of $\boldsymbol{K}_{(n)}$. The resulting algorithm is presented in Table 2.

3 Color Face Tensor Factorization and Slicing

In the case of color face images, illumination change yields large variations on color in face images, even through they have exactly the same pose and facial expression. Conventional methods, such as PCA and GLRAM, convert color image into grayscale image and then reduced the dimension. As the illumination changes, the intensity value also varies extremely. However there is no way to reduce this fluctuation in garyscale image based method, since they already thrown away the information on color. If there are large illumination changes but only small number of samples are available, then these methods can't prevent a sharp decline on face recognition performance. Our propose method, *color face tensor factorization and slicing* (CFTFS), solves these problem by conserving the color information and the spatial structure of original color face image ensemble.

With the help of multiway analysis, the CFTFS simultaneously analyzes the subspace of rows, columns, and color. Then CFTFS slices a feature tensor where information about variations on rows and columns modes are retained but on color mode are minimized. Basic idea of CFTFS is illustrated in Fig. 3.

Table 2. ALS algorithm for Tucker decomposition.

Input: $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Omega}$, J_n for all $n \in \boldsymbol{\Omega}$.

Output: $\mathcal{S}$, $\boldsymbol{A}^{(n)}$ for all $n \in \{1, 2, \ldots, N\}$.

1. Initialize
· $\boldsymbol{A}^{(n)} \leftarrow \boldsymbol{I}_{I_n}$ for all $n \notin \boldsymbol{\Omega}$.
· $\boldsymbol{A}^{(n)} \leftarrow \mathrm{SVDS}(\boldsymbol{X}_{(n)}, J_n)$ for all $n \in \boldsymbol{\Omega}$.
· $\mathcal{S} \leftarrow \boldsymbol{\mathcal{X}} \times_1 \boldsymbol{A}^{(1)\top} \times_2 \boldsymbol{A}^{(2)\top} \cdots \times_N \boldsymbol{A}^{(N)\top}$.

2. Repeat until converges
for all $n \in \boldsymbol{\Omega}$
· $\boldsymbol{\mathcal{K}}^{(n)} \leftarrow \boldsymbol{\mathcal{X}} \times_1 \boldsymbol{A}^{(1)\top} \cdots \times_{n-1} \boldsymbol{A}^{(n-1)\top} \times_{n+1} \boldsymbol{A}^{(n+1)\top} \cdots \times_N \boldsymbol{A}^{N\top}$.
· $\boldsymbol{K}_{(n)} \leftarrow$ mode-n matricization of $\boldsymbol{\mathcal{K}}^{(n)}$.
· $\boldsymbol{A}^{(n)} \leftarrow \mathrm{SVDS}(\boldsymbol{K}_{(n)}, J_n)$.

3. $\mathcal{S} \leftarrow \boldsymbol{\mathcal{X}} \times_1 \boldsymbol{A}^{(1)\top} \times_2 \boldsymbol{A}^{(2)\top} \cdots \times_N \boldsymbol{A}^{(N)\top}$.

CFTFS uses the 4-way tensor $\boldsymbol{\mathcal{X}}$ in which (I_1, I_2) is the size of face image, $I_3 = 3$ is the number of color coordinates(RGB), and I_4 is the number samples. The face image data is centered so that it has zero mean:

$$\boldsymbol{\mathcal{X}}_{:,:,:,i_4} \leftarrow \boldsymbol{\mathcal{X}}_{:,:,:,i_4} - \boldsymbol{\mathcal{M}} \text{ for all } 1 \leq i_4 \leq I_4, \tag{6}$$

where $\boldsymbol{\mathcal{M}} = \frac{1}{I_4} \sum_{i_4}^{I_4} \boldsymbol{\mathcal{X}}_{:,:,:,i_4}$.

As its name hinted, CFTFS consists two stage: dimension reduction and slicing. At dimension reduction stage, it play the 4-way Tucker 3 decomposition, where mode-4 is not analyzed and $J_1 < I_1$, $J_2 < I_2$, and $J_3 = I_3 = 3$. In fact, it is equivalent to CSA which minimizes

$$\sum_{i_4=1}^{I_4} \|\boldsymbol{\mathcal{X}}_{:,:,:,i_4} - \mathcal{S}_{:,:,:,i_4} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \times_3 \boldsymbol{A}^{(3)}\|^2 \tag{7}$$

over $\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{A}^{(3)}$, and $\boldsymbol{\mathcal{S}}_{:,:,:,i_4}$ for all $1 \leq i_4 \leq I_4$. The color face tensor $\boldsymbol{\mathcal{X}}_{:,:,:,i_4}$ is projected to an intermediate feature tensor $\boldsymbol{\mathcal{S}}_{:,:,:,i_4} = \boldsymbol{\mathcal{X}}_{:,:,:,i_4} \times_1 \boldsymbol{A}^{(1)\top} \times_2 \boldsymbol{A}^{(2)\top} \times_3 \boldsymbol{A}^{(3)\top}$. Hence the dimension is reduced from $I_1 \times I_2 \times 3$ to $J_1 \times J_2 \times 3$.

Since we use SVDS in our algorithm, the first columns in mode matrices represent the most dominant subspace, the second columns do the second most subspace orthogonal to the first one, and so on. Thus the third slice $\boldsymbol{\mathcal{S}}_{:,:,3,i_4}$ of an intermediate feature tensor is a final illumination-robust feature matrix. In the final feature matrix, information of $\boldsymbol{\mathcal{X}}_{:,:,:,i_4}$ about variations on rows and columns are retained but on color are minimized.

The illumination-robust feature extraction for a test color face tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times I_2 \times 3}$ is summarized to

$$\boldsymbol{\mathcal{T}} = (\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{M}}) \times_1 \boldsymbol{A}^{(1)\top} \times_2 \boldsymbol{A}^{(1)\top} \times_3 \boldsymbol{a}_3^{(3)\top}, \tag{8}$$

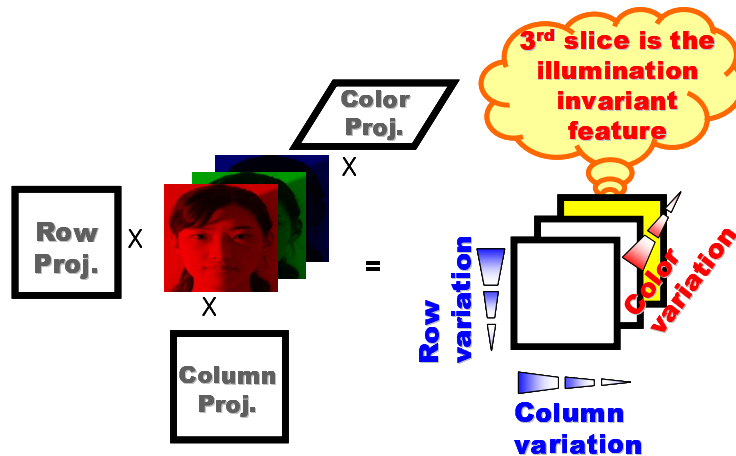where $\boldsymbol{a}_3^{(3)}$ is the third column vector of $\boldsymbol{A}^{(3)}$.

Fig. 3. CFTFS finds a transformation which reduces the size of image and minimizes the variation on color.

4 Numerical Experiments

Our MATLAB implementation of the CFTFS partly uses the tensor toolbox [11]. We show the effectiveness of our proposed method for the illumination-robust face recognition, with CMU PIE face database [12], comparing it PCA, GLRAM, and CSA. CMU PIE database contains 41,368 face images of 68 peoples. In our experiment, we use the sub-database contains 1,632 face images. It has 24 varying illumination condition with exactly same pose(C27) and facial expression(neutral) for each person. We fix the location of the two eyes, crop the face, and resize to 30×30 pixel. Sample face images are shown in Fig. 4.

Fig. 4. Sample face images are shown. Illumination change yields large variations on color in face images, even through they have exactly the same pose and facial expression.

Fig. 5 and Table. 3 show the recognition result for 4 methods. The (x, y, z) axis represent the number of features, training sample ratio, and recognition accuracy. The number of features are $\{9, 16, 25, 36, 49, 64\}$ and training sample ratio are $\{0.1, 0.2, \ldots, 0.9\}$. The experiments are carried out 20 times independently for each case and the mean accuracy are used.
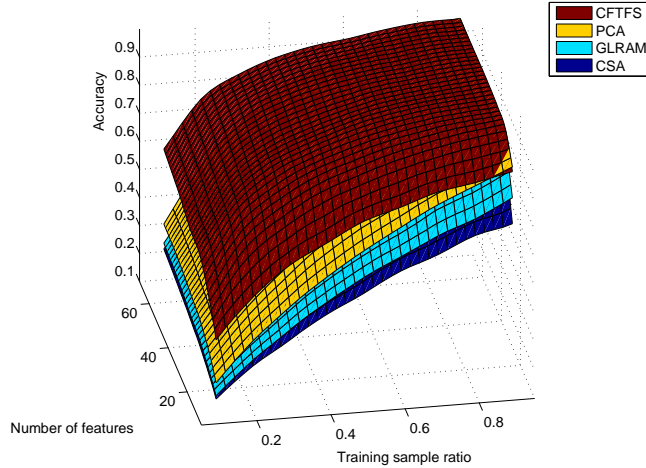
Fig. 5. From top to bottom, face recognition accuracy of CFTFS, PCA, GLRAM, and CSA. CFTFS has the highest accuracy and robust to lack of training samples.

It is known that GLRAM and CSA are has more image compression ability than PCA. However, our experiment results show that they are not suitable for the face recognition under varying illumination conditions. Especially, CSA shows the poorest result since it captures the feature where dominant variations on color remain. The slicing, difference between CFTFS and CSA, dramatically increases the recognition performance. As Fig. 5 and Table. 3 show, CFTFS has the higher recognition accuracy than others in the most of the time. Moreover superiority of our methods becomes more substantial, especially in the case of small training sample size.

5   Conclusions

In this paper, we have presented a method of *color face tensor factorization and slicing* which extracts an illumination-robust feature. Revisiting the Tucker decomposition, we have explained our algorithm in general framework with PCA, 2D PCA, GLRAM, CSA, and HOSVD. Using the 4-way Tucker 3 decomposition, subspaces of rows, columns, and color are simultaneously analyzed and then feature, in which information about variations on rows and columns are retained but on color are minimized, is extracted by slicing. Numerical experiments have confirmed that our method indeed effective for face recognition under a condition in which a large illumination change exists and only small number of training samples are available.

References

1. Yang, J., Zhang, D., Frangi, A.F., Yang, J.Y.: Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Analysis and Machine Intelligence **26** (2004) 131–137
2. Ye, J.: Generalized low rank approximations of matrices. In: Proceedings of International Conference on Machine Learning, Banff, Canada (2004) 887–894
3. Xu, D., Yan, S., Zhang, L., Zhang, H.J., Liu, Z., Shum, H.Y.: Concurrent subspace analysis. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, San Diego, CA (2005) 203–208
4. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear subsapce analysis of image ensembles. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin (2003)
5. de Lathauwer, L., de Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21** (2000) 1253–1278
6. de Lathauwer, L., de Moor, B., Vandewalle, J.: One the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensros. SIAM J. Matrix Anal. Appl. **21** (2000) 1324–1342
7. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear independent component analysis. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, San Diego, California (2005)
8. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31** (1966) 279–311
9. Kroonenberg, P.M., de Leeuw, J.: Principal component analysis of three-mode data by means of alternating least squares algorithms. Psychometrika **45** (1980) 69–97
10. Kolda, T.G.: Multilinear operators for higher-order decompositions. Technical Report SAND2006-2081, Sandia National Laboratories (2006)
11. Bader, B.W., Kolda, T.G.: Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. ACM Trans. Mathematical Software **32** (2006) 635–653
12. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Trans. Pattern Analysis and Machine Intelligence **25** (2003) 1615–1618

	# of features	training sample ratio(%)								
		10	20	30	40	50	60	70	80	90
CFTFS	9	36.67	50.85	61.27	68.04	74.75	78.74	83.06	85.32	88.19
	16	45.32	63.63	73.45	80.81	85.63	89.81	93.01	95.54	97.45
	25	52.99	70.82	80.50	86.46	90.88	94.19	96.56	98.10	99.14
	36	54.62	72.90	82.42	87.87	91.84	95.17	97.03	98.51	99.69
	49	58.45	75.62	85.07	90.20	93.67	95.97	97.43	99.13	99.63
	64	61.41	78.12	85.80	90.74	93.79	95.58	97.75	99.23	99.45
PCA	9	21.37	35.31	46.41	55.95	64.27	71.62	78.59	84.22	89.85
	16	26.44	41.71	54.17	65.26	74.04	80.54	87.89	91.89	96.93
	25	30.14	47.25	59.86	70.21	78.52	85.65	90.48	94.43	97.73
	36	32.10	50.62	63.77	73.47	81.42	87.89	92.63	96.12	98.50
	49	34.06	52.32	65.92	76.39	84.32	90.12	93.99	97.70	98.99
	64	34.72	54.47	67.35	78.28	85.28	90.38	94.49	97.58	99.60
GLRAM	9	16.74	28.44	38.18	46.97	54.76	61.10	67.46	72.42	78.90
	16	19.68	32.87	44.53	54.93	63.31	70.74	79.06	84.46	91.44
	25	22.76	37.89	50.01	60.25	69.30	78.08	84.58	90.05	95.09
	36	24.38	40.69	53.62	63.68	72.94	80.82	87.02	92.67	96.84
	49	26.65	43.19	55.74	67.46	76.53	84.03	89.67	94.54	97.30
	64	27.90	45.50	58.09	69.45	78.13	84.62	90.49	95.38	98.83
CSA	9	16.74	28.44	38.18	46.97	54.76	61.10	67.46	72.42	78.90
	16	19.68	32.87	44.53	54.93	63.31	70.74	79.06	84.46	91.44
	25	22.72	37.99	49.67	60.59	68.82	77.02	84.06	90.37	95.06
	36	20.85	34.78	46.62	57.43	66.02	73.84	80.92	87.16	92.55
	49	18.34	31.11	41.50	51.35	59.67	67.19	74.55	79.57	86.53
	64	15.59	26.33	34.61	42.67	49.14	56.00	60.76	66.53	69.63

Table 3. Recognition accuracy over various number of features and training sample ratio. CFTFS has the higher recognition accuracy than others in the most of the time. Moreover superiority becomes more substantial, especially in the case of small training sample size.