FACIAL AGE ESTIMATION BY MULTILINEAR SUBSPACE ANALYSIS

Xin Geng and Kate Smith-Miles

School of Engineering and Information Technology, Deakin University, Australia

ABSTRACT

Automatic estimation of human facial age is an interesting yet challenging topic appearing in recent years. Since different people might age in different ways, solving the problem of age estimation involves two semantic labels: identity and age. In this paper, aging face images are organized in a third-order tensor according to both identity and age. Due to the difficulty in data collection, the aging pattern for each person in the training set is always incomplete. Therefore, the tensor contains a large amount of missing values. Through a series of multilinear subspace analysis algorithms operating on tensor with missing values, the aging pattern contained in the training aging images can be iteratively learned and be used to predict the age of a given test image. In the experiment, the proposed method not only outperforms the existing algorithms, but also exceeds the human ability in age estimation.

Index Terms— Facial age estimation, Tensor analysis, Machine vision, Pattern recognition

1. INTRODUCTION

The way people prefer to interact with technology, whether it be considerations regarding linguistics, aesthetics, or consumption habits, varies widely according to their age. Automatic age estimation techniques thus have vast potential applications from age-specific service systems to protecting minors from adult web sites and venues. One of the most salient and unobtrusive traits of human age is the face. Consequently, facial age estimation becomes the most practical approach to automatic human age estimation.

Despite the high demand of society's needs, in reality not much work has been done on automatic facial age estimation up to the present. The first true age estimation algorithm was proposed by Lanitis et al. [1] [2] based on function regression, where the aging pattern was represented by a quadratic function called *aging function*. Our latest work on automatic facial age estimation is the AGES algorithm [3] [4]. The basic idea of AGES is to model the aging pattern by constructing a representative linear subspace. The proper aging pattern for a previously unseen face image is determined by the projection in the subspace that can reconstruct the face image with minimum reconstruction error, while the position of the face image in that aging pattern will then indicate its age.

In AGES, the training face images are organized in a data format called an *aging pattern*. Each aging pattern is a sequence of a particular individual's face images sorted in the time order. The aging pattern can be vectorized through concatenating features extracted from each component image within the aging pattern. The vectorized aging patterns are regarded as basic data samples by the learning algorithm of AGES. The final training data set thus can be represented by a matrix with each row being a vectorized aging pattern. But in fact, there exists a more natural way to organize the aging face images. Instead of a matrix, the face images can be assembled in a higher-order tensor so that different modes (dimensions) of the tensor correspond to different semantic meanings of the images, such as personal identity and age. Thus the inherent relationship between the images can be better preserved. Through multilinear subspace analysis on such a tensor, this paper proposes a new approach to facial age estimation.

The rest of this paper is organized as follows. Section 2 introduces tensor fundamentals. Section 3 proposes a multilinear subspace analysis method for facial age estimation. The experimental results are reported and analyzed in Section 4. Finally in Section 5, conclusions are drawn.

2. TENSOR FUNDAMENTALS

Tensors are multilinear mappings over a set of vector spaces. They are higher-order generalization of scalar (zero-order tensor), vector (first-order tensor), and matrix (second-order tensor). In this paper, lowercase italic letters ($a, b, \dots$) denote scalars, bold lowercase letters ($\boldsymbol{a}, \boldsymbol{b}, \dots$) denote vectors, bold uppercase letters ($\mathbf{A}, \mathbf{B}, \dots$) denote matrices, and calligraphic uppercase letters ($\mathcal{A}, \mathcal{B}, \dots$) denote tensors. The *order* of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is N. An element of $\mathcal{A}$ is denoted by $\mathcal{A}_{i_1 i_2 \dots i_N}$ or $a_{i_1 i_2 \dots i_N}$, where $1 \le i_n \le I_n, n = 1, 2, \dots, N$. The *mode-$n$ vectors* of $\mathcal{A}$ are the I_n-dimensional vectors obtained from $\mathcal{A}$ by varying index i_n while keeping other indices fixed to certain values. A tensor $\mathcal{A}$ can be *flattened* into matrices in different ways. The *mode-n flattened matrix* of $\mathcal{A}$, denoted by $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)}$, is obtained by parallel concatenating all the mode-n vectors of $\mathcal{A}$. The *mode-n rank* of $\mathcal{A}$, denoted by R_n, is defined as the dimensionality of the vector space generated by the mode-n vectors: $R_n = rank_n(\mathcal{A}) = rank(\mathbf{A}_{(n)})$.

A tensor can be multiplied by a matrix. The *mode-n prod-*

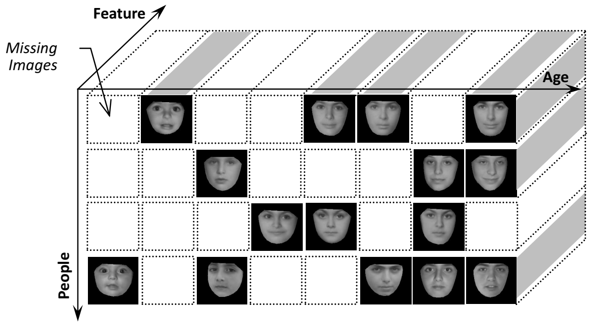

Fig. 1. Organizing aging face images in a third-order tensor $\mathcal{D}$ with missing values. Dotted squares indicate the positions of missing face images.

uct of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ and a matrix $\mathbf{M} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{B} = \mathcal{A} \times_n \mathbf{M}$, is a tensor of dimensionality $\mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$, whose entries are $\mathcal{A} \times_n \mathbf{M})_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} a_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} m_{j_n i_n}$. Alternatively, $\mathcal{B} = \mathcal{A} \times_n \mathbf{M}$ can also be calculated by re-tensorizing the matrix $\mathbf{B}_{(n)} = \mathbf{M}\mathbf{A}_{(n)}$.

3. MULTILINEAR SUBSPACE ANALYSIS FOR FACIAL AGE ESTIMATION

Compared with other facial variations, such as expression, gender and identity, aging effects display three unique characteristics:

1. *The aging progress is uncontrollable.* No one can advance or delay aging at will. The procedure of aging is slow and irreversible.

2. *Personalized aging patterns.* Different people age in different ways. The aging pattern of each person is determined by his/her genes as well as many external factors, such as health, lifestyle, weather conditions, *etc*.

3. *The aging patterns are temporal data.* The aging progress must obey the order of time. The face status at a particular age will affect all older faces, but will not affect those younger ones.

The first characteristic means the collection of sufficient training data for age estimation is extremely laborious. The available data are always 'incomplete' in the view of aging pattern, i.e., for each person, only images at some ages can be obtained, others are missing. The last two characteristics indicate the two significant labels associated with each face image, i.e., the identity (to determine the personalized aging pattern) and the age (the target). Considering these factors, a set of aging face images can be naturally organized in a third-order tensor $\mathcal{D}$ with missing values. As illustrated in Fig. 1, in $\mathcal{D}$, the first dimension corresponds to age, the second corresponds to people, and the third corresponds to features extracted from the images. If the image of a certain person at a certain age is not available, then the corresponding positions are marked as missing values.

Algorithm 1: The N-mode SVD algorithm

1. For $n = 1, \ldots, N$, compute the matrix SVD of the mode-n flattened matrix $\mathbf{D}_{(n)}$ of $\mathcal{D}$, and then set $\mathbf{U}_n$ to be the left matrix of the SVD.

2. Solve for the core tensor:
$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_n \mathbf{U}_n^T \cdots \times_N \mathbf{U}_N^T$.

There are two kinds of semantic labels contained in $\mathcal{D}$, i.e., age and people. Classification in either the 'age space' (for age estimation) or the 'people space' (for age-invariant face recognition) will be made much easier if $\mathcal{D}$ can be decomposed to separate these two constituent factors complicatedly hidden in the image features. Toward this end, $\mathcal{D}$ is decomposed as the mode-n product of two orthogonal spaces:

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{age} \times_2 \mathbf{U}_{people}, \tag{1}$$

where $\mathcal{Z}$ is called the *core tensor*, $\mathbf{U}_{age}$ and $\mathbf{U}_{people}$ contain the orthonormal vectors spanning the 'age space' and the 'people space' respectively. This can be done through the *N-mode SVD* algorithm [5], which is a generalization of matrix SVD to higher-order tensor SVD. The algorithm is summarized in Algorithm 1.

In order to get a compact representation of the constituent factors, the dimensionality of the decomposed orthogonal spaces can be reduced. Unfortunately, the optimal dimensionality reduction in multilinear analysis (operating on tensor) is not as simple as that in PCA (operating on matrix) by directly removing those eigenvectors associated with the smallest eigenvalues. The *N-mode dimensionality reduction* algorithm [5] is summarized in Algorithm 2. The target is to find a best rank-$(R_1, R_2, \ldots, R_N)$ approximation $\hat{\mathcal{D}} = \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \times_2 \hat{\mathbf{U}}_2 \cdots \times_N \hat{\mathbf{U}}_N$, with orthonormal mode matrices $\hat{\mathbf{U}}_n$ of lower rank R_n, $n = 1, 2, \ldots, N$.

Unfortunately, due to the existence of large amount of missing values in the tensor $\mathcal{D}$ consist of aging face images, the existing multilinear analysis algorithms (Algorithm 1 and Algorithm 2) cannot be directly applied to $\mathcal{D}$. In order to deal with this problem, here we propose an iterative algorithm for multilinear analysis on tensors with missing values. Instead of finding a best approximation for $\mathcal{D}$, the target is changed into finding a best approximation for the available values in $\mathcal{D}$, i.e., $\|\mathcal{D}_a - \hat{\mathcal{D}}_a\| < \varepsilon$, where $\mathcal{D}_a$ and $\hat{\mathcal{D}}_a$ are the parts in $\mathcal{D}$ and $\hat{\mathcal{D}}$ corresponding to those available values in the initial training data. The algorithm is summarized in Algorithm 3.

The most attractive property of Eq. (1) is that it provides a way to represent each age, regardless of which person, with the same coefficient vector. In detail, suppose all images at the m-th age in $\mathcal{D}$ (one slice of $\mathcal{D}$ along the 'age' dimension)

Algorithm 2: The N-mode dimensionality reduction algorithm

1. Apply step 1 of the N-mode SVD algorithm to $\mathcal{D}$, truncate each mode matrix $\mathbf{U}_n$ to R_n columns, thus obtaining the initial ($k = 0$) mode matrices $\mathbf{U}_1^0, \mathbf{U}_2^0, \ldots, \mathbf{U}_N^0$.
2. Iterate for $k = 1, 2, \ldots$ until convergence, i.e., $\|(\mathbf{U}_n^k)^T \mathbf{U}_n^{k-1}\| > (1 - \varepsilon)R_n$, for $1 \le n \le N$:

   For $n = 1, 2, \ldots, N$:

   (a) Set $\tilde{\mathcal{U}}_n^k = \mathcal{D} \times_1 (\mathbf{U}_1^k)^T \cdots \times_{n-1} (\mathbf{U}_{n-1}^k)^T \times_{n+1} (\mathbf{U}_{n+1}^{k-1})^T \cdots \times_N (\mathbf{U}_N^{k-1})^T$.

   (b) Mode-n flatten tensor $\tilde{\mathcal{U}}_n^k$ to obtain $\tilde{\mathbf{U}}_n^k$.

   (c) Set $\mathbf{U}_n^k$ to the first R_n columns of the left matrix of the SVD of $\tilde{\mathbf{U}}_n^k$

3. Set $\hat{\mathbf{U}}_n$ to the converged mode-n matrix $\mathbf{U}_n^k$. Compute the core tensor $\hat{\mathcal{Z}} = \tilde{\mathcal{U}}_N^k \times_N \hat{\mathbf{U}}_N^T$
4. The rank-reduced approximation of $\mathcal{D}$ is $\hat{\mathcal{D}} = \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \times_2 \hat{\mathbf{U}}_2 \cdots \times_N \hat{\mathbf{U}}_N$.

is denoted by $\mathcal{D}_m$, then

$$\mathcal{D}_m = \mathcal{Z} \times_1 \mathbf{U}_{age}^{(m)} \times_2 \mathbf{U}_{people} = \mathcal{B} \times_1 \boldsymbol{c}_m^T, \qquad (2)$$

where $\mathcal{B} = \mathcal{Z} \times_2 \mathbf{U}_{people}$ is a constant tensor. $\mathcal{D}_m$ is totally determined by the m-th row vector of $\mathbf{U}_{age}$, $\boldsymbol{c}_m^T = \mathbf{U}_{age}^{(m)}$. Therefore, each row vector in $\mathbf{U}_{age}$ uniquely represents one age, no matter whose age. This is a particularly useful property for age estimation since the personalized aging pattern is one of the major difficulties in this problem.

Given a previously unseen test image, its feature vector $\boldsymbol{b}$ is first extracted. Then, $\boldsymbol{b}$ can be represented by the mode-n product of a set of coefficient vectors:

$$\boldsymbol{b}^T = \mathcal{Z} \times_1 \boldsymbol{c}_a^T \times_2 \boldsymbol{c}_p^T, \qquad (3)$$

where $\boldsymbol{c}_a$ is the coefficient vector corresponding to $\boldsymbol{b}$'s age, and $\boldsymbol{c}_p$ corresponds to the person's identity. Since the person in the test image is supposed to be excluded from the training set, $\boldsymbol{c}_p$ is assumed to be a linear combination of the row vectors in $\mathbf{U}_{people}$, i.e., $\boldsymbol{c}_p^T = \boldsymbol{\alpha}^T \mathbf{U}_{people}$, where $\boldsymbol{\alpha}$ is a combination weight vector. Based on this, the age estimation method based on multilinear subspace analysis is summarized in Algorithm 4.

4. EXPERIMENT

The FG-NET Aging Database [1] is used in the experiment. There are $1,002$ face images from 82 subjects in this database.

Algorithm 3: N-mode dimensionality reduction with missing values

1. Initialize the missing values in the tensor $\mathcal{D}$ with the mean features of the available images at the same age to obtain the initialized tensor $\mathcal{D}^0$.
2. Apply the N-mode dimensionality reduction algorithm to $\mathcal{D}^0$ to get the initial low-rank approximation $\hat{\mathcal{D}}^0 = \hat{\mathcal{Z}}^0 \times_1 \hat{\mathbf{U}}_1^0 \times_2 \hat{\mathbf{U}}_2^0 \cdots \times_N \hat{\mathbf{U}}_N^0$.
3. Iterate for $k = 1, 2, \ldots$ until $\|\mathcal{D}_a - \hat{\mathcal{D}}_a^k\| < \varepsilon$:

   (a) Replace the missing part in $\mathcal{D}$ with the corresponding part in $\hat{\mathcal{D}}^{k-1}$ to obtain the updated tensor $\mathcal{D}^k$.

   (b) Apply the N-mode dimensionality reduction algorithm to $\mathcal{D}^k$ to get the low-rank approximation $\hat{\mathcal{D}}^k$.

4. Set $\hat{\mathcal{D}} = \hat{\mathcal{Z}} \times_1 \hat{\mathbf{U}}_1 \times_2 \hat{\mathbf{U}}_2 \cdots \times_N \hat{\mathbf{U}}_N$ to $\hat{\mathcal{D}}^k = \hat{\mathcal{Z}}^k \times_1 \hat{\mathbf{U}}_1^k \times_2 \hat{\mathbf{U}}_2^k \cdots \times_N \hat{\mathbf{U}}_N^k$

Fig. 2. Typical aging faces in the FG-NET Aging Database.

Each subject has 6-18 face images at different ages. The ages are distributed in a wide range from 0 to 69. Besides age variation, most of the age-progressive image sequences display other types of facial variations, such as significant changes in 3D pose, illumination, expression, *etc*. Some typical aging face sequences are shown in Fig. 2.

In this experiment, age estimation based on multilinear subspace analysis (MSA) is compared with AGES [3] [4], WAS [1], AAS [2], as well as some conventional classification methods including k-Nearest Neighbors (kNN), Back Propagation neural network (BP), C4.5 decision tree (C4.5), and Support Vector Machines (SVM). The algorithms are tested through the Leave-One-Person-Out (LOPO) mode, *i.e.*, in each fold, the images of one person are used as the test set and those of the others are used as the training set. After 82 folds, each subject has been used as test set once, and the final results are calculated from all the estimates.

The face feature extractor for all the compared methods is the Appearance Model [6]. The extracted feature requires 200 model parameters to retain about 95% of the variability in the training data. For MSA, the dimensionality R_n of the mode-n subspace is set to $2/3$ of the original space. For AGES, the

Table 1. Mean Absolute Error (in Years) of Age Estimation on the FG-NET Aging Database

Method	MSA	AGES	WAS	AAS	kNN	BP	C4.5	SVM	HumanA	HumanB
MAE	**5.36**	**6.77**	**8.06**	14.83	8.24	11.85	9.34	**7.25**	8.13	6.23

Algorithm 4: Age estimation based on multilinear subspace analysis

Given a previously unseen test image feature vector $\boldsymbol{b}$:

1. For $m = 1, 2, \ldots, M$ (M is the number of ages):
   (a) $\mathcal{C}_m = \mathcal{Z} \times_1 \mathbf{U}_{age}^{(m)} \times_2 \mathbf{U}_{people}$.
   (b) Mode-2 flatten $\mathcal{C}_m$ to obtain $\mathbf{C}_m$.
   (c) Compute the weight vector $\boldsymbol{\alpha}_m^T = \boldsymbol{b}^T \mathbf{C}_m^+$, where $\mathbf{C}_m^+$ is the pseudo-inverse of $\mathbf{C}_m$.
   (d) Compute the approximation of $\boldsymbol{b}$, $\hat{\boldsymbol{b}}_m = \mathcal{C}_m \times_2 \boldsymbol{\alpha}_m$.
   (e) Compute the reconstruction error $\Delta_m = \|\boldsymbol{b} - \hat{\boldsymbol{b}}_m\|$.
2. Output the age estimation as the age $\tilde{m}$ with the minimum reconstruction error $\Delta_{\tilde{m}}$

aging pattern subspace dimensionality is set to 20. In AAS, the error threshold in the appearance cluster training step is set to 3, and the age ranges for the age specific classification are set as 0-9, 10-19, 20-39 and 40-69. The k in kNN is set to 30. The BP neural network has a single hidden layer of 100 neurons and the same number of output neurons as the number of classes. The parameters of C4.5 are set to the default values of the J4.8 implementation. SVM uses the RBF kernel with the inverse width of 1.

As an important baseline, the human ability in age perception is also tested. 51 face images are randomly selected and presented to 29 human observers. There are two stages in the experiment. In each stage, the 51 face images are randomly shown to the observers, and the observers are asked to choose an age from 0 to 69 for each image. The difference is that in the first stage (HumanA), only the gray-scale face regions are shown, while in the second stage (HumanB), the whole color images are shown. HumanA intends to test age estimation purely based on face, while HumanB intends to test age estimation based on multiple cues including face, hair, skin color, clothes, and background. Note that the information provided in HumanA is the same as that provided to the algorithms, while in HumanB, additional information is available to the observers.

The age estimation performance is evaluated by the Mean Absolute Errors (MAE), i.e., the average absolute difference between the estimated age and the real age. The MAE of the all the compared methods, including the human tests, on the FG-NET Aging Database are tabulated in Table 1. The algorithms performing better than HumanA are highlighted in boldface and those better than HumanB are underlined. As can be seen, MSA is significantly better than all the other algorithms. It is interesting to note that the MAE of MSA is even lower than that of HumanB, where the human observers are provided with more information than that input into the algorithms. Thus at least under this experimental setting, MSA outperforms the human observers in the ability of facial age estimation.

5. CONCLUSIONS

This paper proposes a novel facial age estimation method based on multilinear subspace analysis. The aging face images are naturally organized in a third-order tensor without any pre-assumptions. In the experiment, the proposed method outperforms all the compared algorithms. Most interestingly, under the experimental conditions, the proposed method even exceeds the human ability in facial age estimation.

6. REFERENCES

[1] A. Lanitis, C. J. Taylor, and T.F. Cootes, “Toward automatic simulation of aging effects on face images,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 4, pp. 442–455, 2002.

[2] A. Lanitis, C. Draganova, and C. Christodoulou, “Comparing different classifiers for automatic age estimation,” *IEEE Trans. Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 34, no. 1, pp. 621–628, 2004.

[3] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, “Learning from facial aging patterns for automatic age estimation,” in *Proc. the 14th ACM Int’l Conf. Multimedia*, Santa Barbara, CA, 2006, pp. 307–316.

[4] X. Geng, Z.-H. Zhou, and K. Smith-Miles, “Automatic age estimation based on facial aging patterns,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 12, pp. 2234–2240, 2007.

[5] M. Alex O. Vasilescu and Demetri Terzopoulos, “Multilinear subspace analysis of image ensembles,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Madison, WI, 2003, pp. 93–99.

[6] G. J. Edwards, A. Lanitis, and C. J. Cootes, “Statistical face models: Improving specificity,” *Image Vision Comput.*, vol. 16, no. 3, pp. 203–211, 1998.