Facial Expression Decomposition

Hongcheng Wang, Narendra Ahuja

Beckman Institute, University of Illinois at Urbana-Champaign

{wanghc,ahuja}@vision.ai.uiuc.edu

Abstract

In this paper, we propose a novel approach for facial expression decomposition - Higher-Order Singular Value Decomposition (HOSVD), a natural generalization of matrix SVD. We learn the expression subspace and person subspace from a corpus of images showing seven basic facial expressions, rather than resort to expert-coded facial expression parameters as in [3]. We propose a simultaneous face and facial expression recognition algorithm, which can classify the given image into one of the seven basic facial expression categories, and then other facial expressions of the new person can be synthesized using the learned expression subspace model. The contributions of this work lie mainly in two aspects. First, we propose a new HOSVD based approach to model the mapping between persons and expressions, used for facial expression synthesis for a new person. Second, we realize simultaneous face and facial expression recognition as a result of facial expression decomposition. Experimental results are presented that illustrate the capability of the person subspace and expression subspace in both synthesis and recognition tasks. As a quantitative measure of the quality of synthesis, we propose using Gradient Minimum Square Error (GMSE) which measures the gradient difference between the original and synthesized images.

1. Introduction

Facial expression plays an important role in cognition of human emotions. Basic facial expressions typically recognized by psychologists are happiness, sadness, fear, anger, disgust and surprise [15]. In this paper, we consider 'neutral' as a seventh basic facial expression for convenience. Although the appearance of these expressions may vary between individuals, humans can still recognize different expressions. For example, even if we are not familiar with someone, we can recognize the person's facial expression due to the universality of expressions [20]. Also, we can recognize a familiar person regardless of the person's facial expression.

However, it is a challenging task for computer vision to recognize an individual across different expressions or to classify the basic facial expressions across different persons. In spite of the rapid progress in recent years in automatic face and facial expression recognition [17, 18], there are still numerous problems that are not fully investigated. For example, if a database is to be searched for all the photos of a specific person, and the search is based on a mug shot of the person in which the person appears to be serious, previous work does not consider how to recognize the person in those images where he has other expressions. The following questions enumerate the issues that arise from different combinations of persons and expressions from the viewpoints of both analysis and synthesis:

- Given an unknown person with a known expression, can we recognize this person? Can we synthesize his other expressions?
- Given a known person with an unknown expression, can we classify this expression into one of the seven basic facial expressions?
- Given an unknown person with an unknown expression, can we recognize the person and his expression simultaneously?
- Given a person with a lower intensity expression than previously seen or an expression which is a blend of the seven basic expressions, can we still recognize this person?

These issues involve three major variables: a person, his facial features, and the particular expression he has. Traditional Principle Component Analysis (PCA) can model only the principle axis of variation across images, i.e. when person identity is the only factor ([16]). This paper proposes a generalized framework for facial expression decomposition, which can establish relationships among expressions, persons and facial features. Our approach uses a multilinear generalization of Singular Value Decomposition (SVD), known as Higher-Order Singular Value Decomposition (HOSVD) [12], which is a natural exten-

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)
0-7695-1950-4/03 $ 17.00 © 2003 IEEE

sion of the Tucker model in psychometrics [11]. We learn the expression subspace and person subspace from a corpus of images with seven basic facial expressions. Thus, we can synthesize novel facial expressions for a new person using the learned subspace model. The person subspace and expression subspace can also be used for simultaneous face and facial expression recognition. An important use of the simultaneous face and facial expression recognition algorithm is its application in facial expression synthesis for a new person. When the expert coded expression parameters and even human perception are unavailable, we can classify the facial expression into one of the seven basic categories using our algorithm, and synthesize this person's other facial expressions.

2. Related Work

There has been much work on face recognition and facial expression recognition in computer vision. Comprehensive literature surveys are provided in [17] and [18] on face recognition and facial expression recognition, respectively. Most facial expression recognition techniques use the feature deformation model based on optical flow [21] or local parameters [22] to realize person-independent recognition. Most face recognition techniques usually focus on specific parts of the face to deal with expression-independent face recognition [19]. However, to the knowledge of the authors, none of the previous work has investigated simultaneous face and facial expression recognition.

Facial expression mapping has been widely studied in the field of computer vision and graphics. Traditional methods are warping-based approaches [5, 7], which only capture the feature motions of the face but ignore illumination change, or morphing-based approaches [6], which relates pixels and cannot be used to generate expressions for a new face. A recent algorithm based on ratio images [4] can map facial expression to a new face and captures the illumination change and facial feature motion. However, all these facial expression mapping algorithms are designed purely from the point of view of computer graphics and do not explore how the appearance of an expression changes from person to person. The work most closely related to ours is by Du and Lin [3] who propose learning a linear mapping between emotional parameters and appearance parameters for facial expression synthesis. The emotional parameters come from semantic ratings provided by psychologists.

Some work with goals similar to our work has used a factorization model. Chuang et al. [1] propose using a bilinear model [2] to separate person-specific video data into expressive features and speech content, and synthesize new sequences which maintain the original speech while using a different expression. HOSVD subsumes the bilinear model of [2] since the bilinear model is essentially a 2-factor analysis method, while HOSVD is a general n-factor analysis method. HOSVD has been applied to human face recognition and human motion analysis [8, 9, 10]. Vasilescu et al. consider different factors such as viewpoint, illumination, and facial expression in their facial image analysis, and obtain better face recognition results than PCA [8, 10]. They also use HOSVD to obtain motion signature which is used to animate a rendered character [9]. Using modified HOSVD, this paper investigates decomposing facial expression images from different persons into separate subspaces. As a consequence of this is that any expression can be mapped onto a new person's face. Most importantly, our decomposition approach leads to expression-independent face recognition and person-independent facial expression recognition simultaneously.

The paper is organized as follows. In Section 3, we give a brief overview of the proposed algorithm, including the modified HOSVD, and the methods used for facial expression synthesis, and face and facial expression recognition. Section 4 presents the experiments we have conducted and evaluates the performance of the approach presented. Finally, we present conclusions and future research directions in Section 5.

3. Overview of the Algorithm

Given a corpus of facial expression images of different persons, we would like to decompose them into two separate subspaces, the expression subspace and the person subspace. We use a third-order tensor $\mathcal{A} \in \mathbb{R}^{I \times J \times K}$ to represent the facial expression configuration, where I is the number of persons, J is the number of facial expressions for each person and K is the dimension of the facial feature vector which describes the distribution of salient features and gray-level appearance variations.

Our goal is to decompose tensor $\mathcal{A}$ using HOSVD to extract separate subspaces along person mode and expression mode. The details of HOSVD are described in Appendix. Vasilescu et al. [8] describe an implementation of HOSVD based on a natural extension of matrix SVD proposed in [12]. This procedure is very inefficient in dealing with high-dimensional image data since very large matrices are involved in the decomposition. For an $M \times N$ image, if the third mode represents pixel gray-level, it will have the dimension of MN, which is usually very large (e.g. even for a small image size of 161×142 that we use, the dimension would be $22, 862$). The HOSVD of a tensor for facial images of different persons showing various expressions will have low computational speed and a high memory complexity. To alleviate this problem, we propose a PCA-based HOSVD, in which we first find the principle compo-

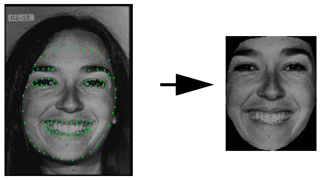


Figure 1. Facial feature extraction using deformable face model. The left figure illustrates 60 feature points around eye, mouth, etc, and the right figure illustrates the gray-level appearance.

nents for the mode with the largest dimension, so that HOSVD only needs to deal with the resulting lower dimension tensor. The dimensionality reduction in the third mode is realized by a deformable face model (Section 3.1).

3.1. Facial Feature Extraction

Facial features in the facial expression configuration are described by the spatial distribution of salient feature points (called shape parameters) and gray-level appearance. Shape is represented by a vector of feature locations on the face, and gray-level appearance is represented by pixel intensity values of the image. By using a deformable face model similar to the active appearance model employed by Cootes et al. [13], we can reduce the dimensionality of facial feature vectors using principle component analysis.

Face shape is represented by a vector of the locations of a number of manually marked feature points, denoted by $\boldsymbol{s}$. We apply PCA to the training image set for dimensionality reduction and obtain a linear model, $\boldsymbol{s} = \bar{\boldsymbol{s}} + \mathbf{P}_s \boldsymbol{b}_s$, where $\bar{\boldsymbol{s}}$ is the mean shape, $\boldsymbol{b}_s$ is a set of shape parameters and $\mathbf{P}_s$ is a set of orthogonal modes of variation.

To build a statistical gray-level appearance model, we first warp each training image so that its control points match the mean shape using a triangulation algorithm. This removes spurious texture variation due to shape differences. We then sample the intensity information from the shape-normalized image over the region covered by the mean shape to form a texture vector, $\boldsymbol{g}$. By applying PCA to the gray-level appearance model, we obtain a linear model, $\boldsymbol{g} = \bar{\boldsymbol{g}} + \mathbf{P}_g \boldsymbol{b}_g$, where $\bar{\boldsymbol{g}}$ is the mean gray-level appearance vector, $\boldsymbol{b}_g$ is a set of gray-level parameters and $\mathbf{P}_g$ is a set of orthogonal modes of variation.

The facial feature vector, $\boldsymbol{f}$, is now defined as the combination of shape vector, $\boldsymbol{b}_s$, and gray-level appearance vector, $\boldsymbol{b}_g$, i.e. $\boldsymbol{f} = [\boldsymbol{b}_g, \boldsymbol{b}_s]^T$. An example of facial feature extraction is illustrated in Figure 1.

3.2. Facial Expression Decomposition

By applying HOSVD to the tensor resulting from dimensionality reduction, we can decompose the tensor into different subspaces,

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{person} \times_2 \mathbf{U}^{expression} \times_3 \mathbf{U}^{feature} \qquad \text{(EQ 1)}$$

In Equation 1, $\mathcal{S}$ is the core tensor which represents the interactions of the person, expression and feature subspaces. Tensor $\mathcal{S}$ is analogous to the diagonal singular value matrix in matrix SVD, except that it is not diagonal. Matrices $\mathbf{U}^{person}$, $\mathbf{U}^{expression}$ and $\mathbf{U}^{feature}$ represent the person subspace, expression subspace and facial feature subspace, respectively. These matrices are all orthogonal. Each row vector in each subspace matrix represents a specific vector in this mode. For example, the row vector $\boldsymbol{u}_n^p$ in person subspace $\mathbf{U}^{person} = [\boldsymbol{u}_1^p, \boldsymbol{u}_2^p, \ldots, \boldsymbol{u}_n^p, \ldots \boldsymbol{u}_I^p]^T$ represents the characteristics of the nth person, while the row vector $\boldsymbol{u}_n^e$ in expression subspace $\mathbf{U}^{expression} = [\boldsymbol{u}_1^e, \boldsymbol{u}_2^e, \ldots, \boldsymbol{u}_n^e, \ldots \boldsymbol{u}_J^e]^T$ represents the characteristics of the nth expression. Each column vector in each subspace matrix represents the contributions of other modes. For example, the column vector in person subspace matrix represents the contributions of eigenexpressions and eigenfeatures. Such facial expression decomposition is illustrated in Figure 2. Different subspaces associated with the person and expression modes obtained from facial expression decomposition can be used for facial expression synthesis and face and facial expression recognition.

3.3. Facial Expression Synthesis

Using the facial expression decomposition given in Equation 1, we have obtained the person subspace $\mathbf{U}^{person}$, expression subspace $\mathbf{U}^{expression}$, facial feature subspace $\mathbf{U}^{feature}$ and their interaction tensor $\mathcal{S}$. These define a generative model that can observe facial features of a new person with a given expression and synthesize the remaining expressions. Using a simple transformation, we define two tensors related to the person and expression modes, and we call them the expression tensor, $\mathcal{T}^{expression}$, and the person tensor, $\mathcal{T}^{person}$, respectively, given by,

$$\mathcal{T}^{expression} = \mathcal{S} \times_2 \mathbf{U}^{expression} \times_3 \mathbf{U}^{feature} \qquad \text{(EQ 2)}$$

$$\mathcal{T}^{person} = \mathcal{S} \times_1 \mathbf{U}^{person} \times_3 \mathbf{U}^{feature} \qquad \text{(EQ 3)}$$

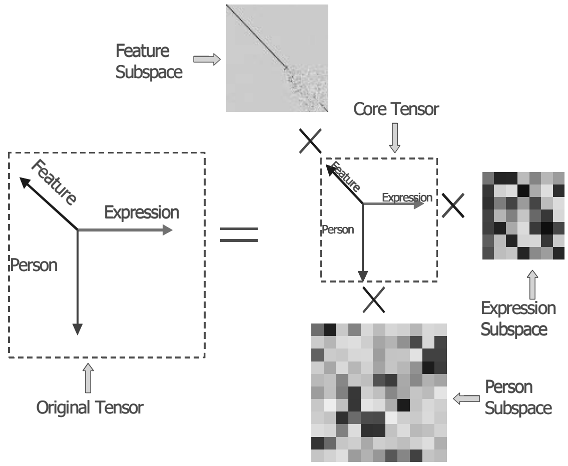


Figure 2. Facial expression decomposition to person subspace, expression subspace, feature subspace

The expression tensor defines a set of basis matrices for all the facial features associated with all expressions. The person tensor defines a set of basis matrices for all the facial features associated with all persons. $\mathcal{T}^{expression}$ and $\mathcal{T}^{person}$ have the same dimension as $\mathcal{A}$ and S.

Given an expression of a new person, we now describe synthesis of his other expressions. The idea is first to find the person vector corresponding to this person from the known facial feature vector, $\boldsymbol{f}^{test}$, and known expression type, denote by identity id variable *i*. In our case, $1 \le i \le 7$ since we code seven basic facial expressions with id varying from 1 to 7: happiness(1), sadness(2), fear(3), anger(4), surprise(5), disgust(6) and neutral(7). We apply the person vector to Equation 2 to find all the expressions associated with this person.

The input test tensor $\mathcal{T}_{test}$ is a $1 \times 1 \times I_3$ tensor using $\boldsymbol{f}^{test}$ in the third mode. The person vector $\boldsymbol{u}^p$ is represented as

$$\boldsymbol{u}^p = uf(\mathcal{T}_{test}, 1)^T \cdot (uf(\mathcal{T}^{expression}(i), 1))^{-1} \qquad \text{(EQ 4)}$$

where $uf(\mathcal{T}, 1)$ means unfolding tensor $\mathcal{T}$ in the first (person) mode. Then the synthetic expression tensor of the new person is:

$$\mathcal{T}_{syn} = (S \times_2 \mathbf{U}^{expression} \times_3 \mathbf{U}^{feature}) \times_1 \boldsymbol{u}^p \qquad \text{(EQ 5)}$$

Similarly, given a known person with an unknown expression, we can synthesize all the persons in the database with the same expressions. The expression vector $\boldsymbol{u}^e$ is represented as

$$\boldsymbol{u}^e = uf(\mathcal{T}_{test}, 2)^T \cdot (uf(\mathcal{T}^{person}(i), 2))^{-1} \qquad \text{(EQ 6)}$$

where $uf(\mathcal{T}, 2)$ means unfolding tensor $\mathcal{T}$ in the second (expression) mode. Then the synthetic tensor for all the persons with the same expression is:

$$\mathcal{T}_{syn} = (S \times_1 \mathbf{U}^{person} \times_3 \mathbf{U}^{feature}) \times_2 \boldsymbol{u}^e \qquad \text{(EQ 7)}$$

Facial feature vector $[\boldsymbol{b}_g, \boldsymbol{b}_s]^T$ can be obtained by reshaping $\mathcal{T}_{syn}$ to a column vector. The synthesized gray-level image and shape can be obtained as:

$$\begin{bmatrix} \boldsymbol{s}_{syn} \\ \boldsymbol{g}_{syn} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{s}} + \mathbf{P}_s \boldsymbol{b}_s \\ \bar{\boldsymbol{g}} + \mathbf{P}_g \boldsymbol{b}_g \end{bmatrix} \qquad \text{(EQ 8)}$$

3.4. Simultaneous Face and Facial Expression Recognition

The person subspace and facial expression subspace obtained from facial expression decomposition can not only be used for facial expression synthesis, they can also be used for person and facial expression recognition. We use cosine distance measurements to compute the distances between the measurements of an unknown person or unknown expression to the learned person vectors or expression vectors in the subspace model.

The cosine of the angle between two vectors, ***a*** and ***b***, is defined as

$$cos\_dist(\boldsymbol{a}, \boldsymbol{b}) = \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\|\boldsymbol{a}\| \|\mathbf{b}\|} = \frac{tr(\boldsymbol{a}^T, \boldsymbol{b})}{\|\boldsymbol{a}\| \|\boldsymbol{b}\|} \qquad \text{(EQ 9)}$$

Face recognition with known expression is based on the cosine distance between test person vector $\boldsymbol{u}^p$ from Equation 4 and each row of the person subspace $\mathbf{U}^{person}$. The person corresponding to the largest distance, $Max_i[cos\_dist(\boldsymbol{u}^p, \mathbf{U}^{person}(i))]$, is chosen.

Facial expression recognition is based on the cosine distance between test expression vector $\boldsymbol{u}^e$ from Equation 6 and each row of the expression subspace $\mathbf{U}^{expression}$. The expression corresponding to the largest distance, $Max_i[cos\_dist(\boldsymbol{u}^e, \mathbf{U}^{expression}(i))]$, is chosen.

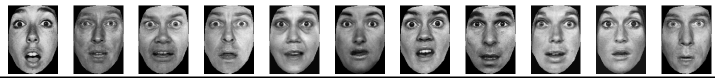
tasks	given	synthesized images
unknown person, known expression, synthesize expressions;		
unknown person, low intensity expression, synthesize expressions;		
unknown expression, known person, synthesize persons with the same expression		
synthesize expressions of a new person who is not in the training set;		
synthesize expressions of a new person (there is a similar person in the training set).		

Figure 3. Facial Expression Synthesis Experiments. Facial expression from left to right: happiness, sadness, fear, anger, surprise, disgust and neutral.

We achieve simultaneous face and facial expression recognition algorithm as follows: for the classification of expressions, we first assume the given person has one of the basic facial expressions, and then find all the person vectors associated with the expression, and find the maximum of the distances between person vectors. By repeating this procedure for all expressions, the expression sought can be identified as that corresponding to the maximum distance. Person recognition can be done in a similar way.

4. Experiments and Evaluation

We used Ekman's 'Pictures of Facial Affect' database [14] in our experiments. The database contains 110 facial expression images showing different emotions of six male and eight female subjects. We chose 77 images of 11 persons with seven basic facial expressions displayed by each person as our training set*. In the deformable face model, we chose 50 modes to account for 93.8% appearance variation, and 40 modes to account for 98.3% shape variation. Then the facial feature is a 90 dimensional vector. Thus we defined a third order tensor of size $11 \times 7 \times 90$ to describe the facial expression configuration of this dataset.

4.1. Facial Expression Synthesis

The purpose of the facial expression synthesis experiment is to validate the learned expression subspace model. By mapping the expression subspace to a new person, and synthesizing images, we can see that the model captures the expression signatures. We describe the facial expression synthesis tasks and experimental results in Figure 3.

*. The remaining three persons have images of only five or six basic facial expressions

The facial expressions are synthesized for an unknown person with 'happiness' expression, as illustrated in the first row in Figure 3. The third row in Figure 3 illustrates synthesized expressions for different persons with the same expression as the given person has. These two cases essentially are about image reconstruction using the facial expression decomposition.

If the test person is not in the training database, we can still synthesize his facial expressions under the assumption that similar persons have similar facial expression appearance and shape. We did some experiments on low intensity images achieved by warping the original, more intensely expressive training image. The synthesized images (the second row in Figure 3) have a little bit lower intensity compared to those in the first row. This is because the synthesized images are obtained by mapping the person vector of the input image to the expression subspace while the person vector of the low intensity expression image is changed only a little (as seen from the first row in Figure 5). The fourth row and the fifth row in Figure 3 are the synthesized expressions for new persons who are not in the training set. The training database contains images of a person who looks very similar to the person presented in the fifth, but not the fourth row. The synthesized images in the fifth row are sharp, but there are some artifacts and blurring in those of the fourth row. For these two examples, we do not need the expert coded expression parameters for the given images; the expressions can be identified without such coding (as shown in the second and third rows in Figure 5), which are then used to find the person vectors for the given images for expression synthesis.

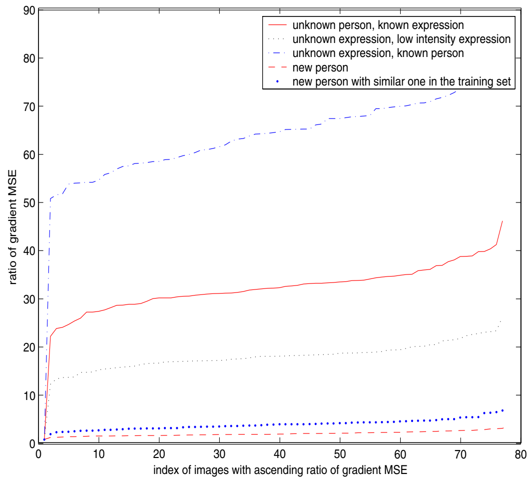


Figure 4. Ratio of GMSE for Different Facial Expression Synthesis Tasks

The synthesized images are a little brighter compared to the ground truth images because of the limit on the number of components we used. Therefore, we can restore the mean brightness of the synthesized image for comparison with the original image. How can we evaluate the quality of the synthesis? One way is to do this through experiments involving ranking by people. To obtain more objective quantitative measure, we here compared the image intensity characteristics, by comparing the intensity gradients of the original and synthesized images. We use image Gradient Mean Square Error (GMSE) instead of pixel MSE, defined as:

$$GMSE = \frac{1}{MN}\sum_{i=1}^{MN}\left\| \begin{bmatrix}\nabla I_x(i)\\ \nabla I_y(i)\end{bmatrix}_{org} - \begin{bmatrix}\nabla I_x(i)\\ \nabla I_y(i)\end{bmatrix}_{syn}\right\|^2 \quad \text{(EQ 10)}$$

for an image of size $M \times N$. $\left[\nabla I_x(i)\ \nabla I_y(i)\right]^T_{org}$ and $\left[\nabla I_x(i)\ \nabla I_y(i)\right]^T_{syn}$ are gradients of pixel i in the original and synthesized images.

By computing GMSE of the synthesized image with all the images in the training set and the given image, we found that the minimum GMSE is always the one between the given image and the synthesized image with the same expression. This means that the synthesized image is the closest to the given image. We also obtained for the five different cases the ratio of GMSE, given by,

$$ratio = \frac{GMSE}{min(GMSE)}$$

to evaluate the similarity between the synthesized image, and the set of given image and training images. Figure 4 gives the plot of the ratio of GMSE. The slope of the line between the first two points with minimum GMSE illustrates the similarity. The smaller the slope, the smaller the similarity, and the larger the GMSE, though the ratio of GMSE is smaller. Therefore, the given new person in the fourth case (dashed line) has the minimum similarity from Figure 4. The image in the fifth case has a larger GMSE ratio than that in the fourth case. The image with high intensity expression has a larger GMSE ratio than that with low intensity expression. The evaluation coincides with our visual perception and discussion before.

We also conducted an informal evaluation of the approach by showing the results to a psychological and facial expression expert and three labmates, and they could all correctly classify the different expressions. This shows the validity of our facial expression subspace model, which can capture the signatures of expressions. The artifacts and blurring in the synthesized images of a new person are reasonable considering we have only 11 persons in the training set. When a large data set is available, the probability of finding a face in the database which is close to that of the test person is large. Therefore, our approach should be able to synthesize very good facial expressions using a large training set.

4.2. Face and Facial Expression Recognition

Face recognition and facial expression classification is another way to validate our facial expression decomposition approach. Figure 5 describes the experimental results of face and facial expression recognition using the same test images as Figure 3.

In the first row in Figure 5, the high intensity expression 'happiness' (left), the low intensity expression 'happiness' (middle) and the person (right) are correctly recognized. Because there is not much difference between the person vector distance for high and low intensity 'happiness', the synthesized expressions in the first and second rows in Figure 3 do not change much. We also did experiments for the two new persons who are not in the training set. A threshold for the person vector distance is used to decide if the person is in the database or not. For example, the expression in the second row in Figure 5 is recognized as 'surprise', but the person is not recognized according to our thresholds (person threshold = 0.9, expression threshold = 0.85). But, the person in the third row is similar to the eighth person in the database from the recognition result; therefore the synthesized expressions are close to those of the eighth person. The expression is correctly recognized as 'neutral'. We did experiments with the 33 remaining images of 'Pictures of Facial Affect' as our test images on simultaneous face and facial expression recognition. These images

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)
0-7695-1950-4/03 $ 17.00 © 2003 IEEE

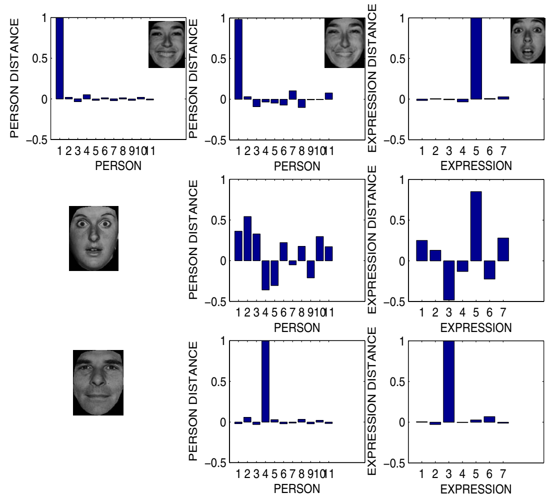


Figure 5. Face and Facial Expression Recognition. First row: recognize the person or the expression only; Second row: recognize person and expression for a new person simultaneously; Third row: recognize person and expression simultaneously for a new person with similar one in the training set. Facial expression: 1-happiness, 2-sadness, 3-fear, 4-anger, 5-surprise, 6-disgust and 7-neutral.

have different expressions of different persons. Although some persons are in the training set, their facial expressions are different in intensity. We tested our simultaneous face and facial expression recognition algorithm, and obtained an accuracy of 84.85% in recognizing the expressions, and 96.97% in recognizing the persons.

5. Conclusion and Future Extensions

We have demonstrated that the facial expression space can be decomposed into three subspaces using a generalized singular value decomposition method - HOSVD. By using a deformable face model to reduce the dimensionality of the facial images, the HOSVD of large image data is tractable. The resulting subspaces can be used for facial expression synthesis and simultaneous face and facial expression recognition. Experiments show that HOSVD is effective in learning facial expression configuration. However, it can not synthesize expressions of people with unforeseen characteristics, such as a beard if we have no similar images in the training set.

A natural extension of current work is to add more factors in the data, such as different viewpoints and illumination conditions. Then we can synthesize new face images from different viewpoints and under different illuminations. We plan to explore these cases, as well as the dynamic facial expression recognition from video sequences using HOSVD.

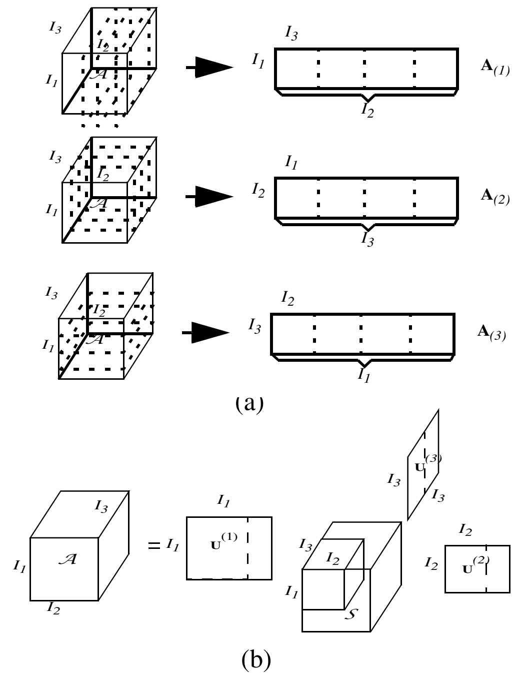


Figure 6. (a) Unfolding; (b) HOSVD of a third-order tensor

Acknowledgements

The support of the Office of Naval Research under grant N00014-03-1-0107 is gratefully acknowledged.

Appendix

Given an Nth-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. The matrix unfolding $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 I_2 \ldots I_{n-1} I_{n+1} I_{n+2} \ldots I_N)}$ contains the element $a_{i_1 \ldots i_N}$ at row number i_n and column number

$$(i_{n+1} - 1) I_1 I_2 \ldots I_{n-1} I_{n+2} I_{n+3} \ldots I_N + \ldots +$$
$$(i_N - 1) I_1 I_2 \ldots I_{n-1} + (i_1 - 1) I_2 I_3 \ldots I_{n-1} + \ldots + i_{n-1}$$

Unfolding a tensor $\mathcal{A}$ along the nth mode is denoted as $uf(\mathcal{A}, n)$. Unfolding of a third-order tensor is illustrated in Figure 6 (a).

The n-mode product of a tensor $\mathcal{A}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is defined by a tensor with entries,

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} a_{i_1 \ldots i_N} u_{j_n i_n}$$

The Nth-order SVD is proposed in the following theorem.

THEOREM: *Every* $I_1 \times I_2 \times \ldots \times I_N$ *tensor* $\mathcal{A}$ *can be written as the product:*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times \ldots \times_N \mathbf{U}^{(N)} \qquad \text{(EQ 11)}$$

where,

- $\mathbf{U}^{(n)} = (\boldsymbol{u}_1^{(n)} \boldsymbol{u}_2^{(n)} \ldots \boldsymbol{u}_{I_n}^{(n)})$ is a unitary $I_n \times I_n$ matrix;
- $\mathcal{S}$ is a $I_1 \times I_2 \times \ldots \times I_N$ tensor of which the subtensors $\mathcal{S}_{i_n = \alpha}$ has the property of all-orthogonality and ordering based on the Frobenius-norms $\|\mathcal{S}_{i_n}\|$.

$\mathcal{S}$ is in general a full matrix (not pseudodiagonal), and also known as core tensor; $\mathbf{U}^{(i)}$ provides directions of maximal oriented energy along the *i*th mode. The HOSVD of a third-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is illustrated in Figure 6 (b).

It is shown in [12] that the *n*-mode singular matrix $\mathbf{U}^{(n)}$ can directly be found at the left singular matrix of an *n*-mode matrix unfolding of $\mathcal{A}$. Therefore computing the HOSVD of an *N*th-order tensor leads to the computation of N different matrix SVDs of matrices with size $I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N$.

References

[1] E. Chuang, H. Deshpande, and C. Bregler, “Facial Expression Space Learning”, *Pacific Graphics*, 2002.

[2] J. B. Tenenbaum and W. Freeman, “Separating Style and Content with Bilinear Models”, *Neural Computation Journal*, Vol. 12, January 1999, pp. 1247-1283.

[3] Y. Du and X. Lin, “Mapping Emotional Status to Facial Expressions”, *Proceedings of the 16th International Conference on Pattern Recognition (ICPR 2002)*, August, 2002, pp. 524-527.

[4] Z. Liu, Y. Shan, and Z. Zhang, “Expressive Expression Mapping with Ratio Images”, *Proceedings of ACM SIGGRAPH 2001*, Los Angeles CA, August 2001, pp. 271-276.

[5] J.Y. Noh and U. Neumann, “Expression Cloning”, *Proceedings of ACM SIGGRAPH 2001*, Los Angeles CA, August 2001, pp. 277-288.

[6] S. M. Seize and C. R. Dyer, “View Morphing”, *SIGGRAPH* 1996, pp. 21-30.

[7] F. I. Parke and K. Waters, *Computer Facial Animation*, AK Peters, Massachusetts, 1996.

[8] M. A. O. Vasilescu and D. Terzopoulos, “Multilinear Analysis of Image Ensembles: TensorFaces”, *ECCV* 2002, pp. 447-460.

[9] M. A. O. Vasilescu and D. Terzopoulos, “Human Motion Signatures: Analysis, Synthesis, Recognition”, *Proceedings of the International Conference on Pattern Recognition (ICPR’02)*, Quebec City, Canada, Aug. 2002.

[10] M. A. O. Vasilescu and D. Terzopoulos, “Multilinear Image Analysis for Facial Recognition”, *Proceedings of the International Conference on Pattern Recognition (ICPR’02)*, Quebec City, Canada, Aug. 2002.

[11] L. R. Tucker, “Some Mathematical Notes on Three-Mode Factor Analysis”, *Psychometrika*, Vol. 31, 1966, pp. 279-311.

[12] L. De Lathauwer, B. De Moor, and J. Vandewalle, “A Multilinear Singular Value Decomposition”, *SIAM Journal of Matrix Analysis and Applications*, 21(4), 2000.

[13] T. F. Cootes, G. J. Edwards and C. J. Taylor, “Active Appearance Models”, *Proceedings European Conference on Computer Vision (ECCV)*, Vol. 2, 1998, pp. 484-498.

[14] P. Ekman and W.V. Friesen, *Pictures of Facial Affect*, Palo Alto, CA: Consulting Psychologists Press, 1978.

[15] P. Ekman, “Facial Expressions of Emotion: an Old Controversy and New Findings”, *Philosophical Transactions of the Royal Society*, London, B335:63--69, 1992.

[16] M. Turk and A. Pentland, “Eigenfaces for Recognition”, *Journal of Cognitive Neuroscience*, Vol. 3, 1991, pp. 72-86.

[17] W. Y. Zhao, R. Chellappa, **A**. Rosenfeld, and P. J. Phillips, “Face Recognition: A Literature Survey”, *UMD CfAR Technical Report CAR-TR-948*, 2000.

[18] B. Fasel and J. Luettin, “Automatic Facial Expression Analysis: A Survey”, *Pattern Recognition*, Vol. 36(1), 2003, pp. 259-275.

[19] A.M. Martinez, “Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24(6), 2002, pp. 748-763.

[20] Ekman, P. *Emotion in the Human Face* New York: Cambridge University Press, 1982.

[21] Y. Yacoob and L. S. Davis, “Recognizing Human Facial Expressions from Long Image Sequences Using Optical Flow”, *IEEE Transitions on Pattern Analysis and Machine Intelligence*, Vol. 18, June 1996, pp. 636-642.

[22] M. J. Black and Y. Yacoob, “Tracking and Recognizing Rigid and Non-rigid Facial Motions Using Local Parametric Models of Image Motion”, *ICCV*, 1995.