

# Task Specific Factors for Video Characterization

Ranjeeth Kumar, S. Manikandan, and C.V. Jawahar

Center for Visual Information Technology  
International Institute of Information Technology  
Gachibowli, Hyderabad - 500032, India  
jawahar@iiit.ac.in

**Abstract.** Factorization methods are used extensively in computer vision for a wide variety of tasks. Existing factorization techniques extract factors that meet requirements such as compact representation, interpretability, efficiency, dimensionality reduction *etc.* However, when the extracted factors lack interpretability and are large in number, identification of factors that cause the data to exhibit certain properties of interest is useful in solving a variety of problems. Identification of such factors or *factor selection* has interesting applications in data synthesis and recognition. In this paper simple and efficient methods are proposed, for identification of factors of interest from a pool of factors obtained by decomposing videos represented as tensors into their constituent low rank factors. The method is used to select factors that enable appearance based facial expression transfer and facial expression recognition. Experimental results demonstrate that the factor selection facilitates efficient solutions to these problems with promising results.

## 1 Introduction

Factorization methods are popular in computer vision [1,2,3,4,5]. Applications of factorization methods include recovery of structure from motion (*SfM*) [1], separation of style and content [3], decomposition of facial expressions [4] and local parts decomposition [2,5]. The use of factorization techniques differs in aspects like the manner in which factors are extracted, interpreted and modeled. Tomasi and Kanade [1] recover the scene structure and the camera motion from a sequence of images taken with a moving camera by factorizing a measurement matrix into shape and motion factors using Singular Value Decomposition (*SVD*). The factors are interpretable, and they characterize the generative process. The model of image formation assumed therein is valid. The factors extracted i.e. the scene structure and the camera motion are useful for several other tasks. Tenenbaum and Freeman [3] use a bilinear model to characterize the interaction between style and content parameters. They use *SVD* and Expectation Maximization (*EM*) algorithms to carry out the tasks of classification, extrapolation and translation which require inferring the model, and one or both of the factors. The style and content factors in [3] are interchangeable and the generative model assumed (a simple bilinear model) is rich enough to capture the interaction between the factors although it lacks formal evidence. Moreover, their

goal is not the extraction of the factors alone but to solve related tasks that require simultaneous learning of one or both of the factors and the factor model. Both these algorithms fall into the class of algorithms where the factors are fixed in number, interpretable and the factor model corresponds to the underlying generative model. Another class of algorithms attempt to identify sets of latent variables that aid in elimination of redundancies in the data. The small set of variables along with the factor model explain the structure in the data and result in compact representation. Ghahramani and Hinton [6] model the covariance structure of data assuming a simple factor model that is linear in the factors. A mixture of such models has been used for simultaneous dimensionality reduction and clustering. Techniques like Principal Component Analysis and Positive Tensor Factorization [7] extract factors that enable dimensionality reduction or compression. Tensor representation based methods are gaining wide attention in computer vision [8,9] as tensor representation is more suitable for image collections or videos. The observation that positive factorization of matrices results in local parts decomposition of objects [2] resulted in similar factorization methods for tensors [7,10]. Such factorization results in sparse encoding [5] of the data. The sparse and separable factors obtained using positive factorization are used for a variety of tasks like image encoding, model selection and face recognition [10]. The factors obtained by this latter class of techniques usually lack meaningful interpretations and do not necessarily correspond to the properties of the generative process but aid in reducing the bulkiness of data.



**Fig. 1.** The facial expression transfer problem : Given the videos of a person with different expressions the goal is to synthesize the videos of a second person with the same expressions (marked by ?'s in the left table). The right table (second row) shows the results using our method.

Orthogonal to the requirements of compact representations and interpretable factors, is the need for identification of factors (interpretable or otherwise) that cause the data to exhibit certain properties of interest. The performance of computer vision algorithms at tasks such as object classification, object detection is significantly enhanced by using a set of features that are relevant to the task rather than a full set of features representing the data. For instance, detecting a face needs features that characterize the holistic appearance of face while discriminating between two face classes would require filters capturing finer variations

in the appearance of the faces. Similarly, identification of factors that give rise to properties of interest in the observed data has interesting applications such as data synthesis and recognition. For instance, given two sets of data with different desirable properties, identifying the relevant constituent factors and the factor model enables synthesis of a third collection with both of these properties. In this paper simple and efficient methods are proposed to perform such factor selection from a pool of factors obtained by decomposing a collection of images or a video represented as a tensor. The factors are selected so as to enable efficient appearance based solutions for two challenging tasks: facial expression transfer (explained pictorially in Figure 1), and facial expression recognition. The non-negative tensor factorization proposed in [10] is used to obtain the factors. Experimental results demonstrate that the selected factors enable both of the tasks with satisfactory results despite the use of appearance information alone.

## 2 Factor Selection for Videos

The expression transfer problem will be our running example. The solution proposed in the current work is based on decomposing the appearance of the face video represented as tensor using the non-negative tensor factorization (*NTF*) [10] technique and then selecting appropriate factors to be used in the synthesis. An  $N$ -valent tensor  $G$  of dimensions  $[d_1] \times \dots \times [d_N]$  has rank  $k$  if  $k$  is the smallest number such that  $G$  can be expressed as sum of  $k$  rank-1 tensors i.e.,

$$G = \sum_{j=1}^k \otimes_{i=1}^N \mathbf{u}_i^j \quad (1)$$

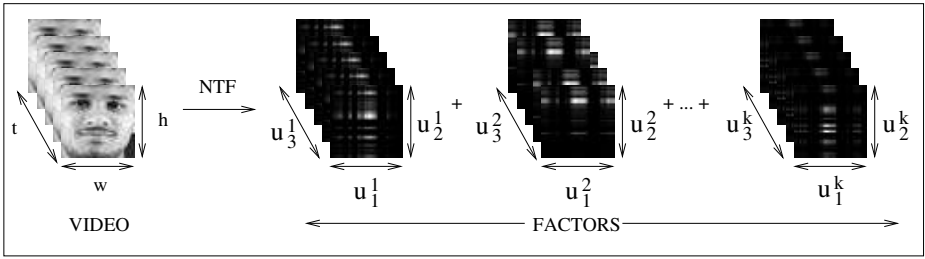
where the vectors  $\mathbf{u}_i^j \in \mathcal{R}^{d_i}$ . The *NTF* approximates the tensor  $G$  with a rank- $k$  tensor  $\hat{G} = \sum_{j=1}^k \otimes_{i=1}^N \mathbf{u}_i^j$  where  $\mathbf{u}_i^j$ 's are all positive. The  $Nk$  vectors  $\mathbf{u}_i^j$  are iteratively estimated using a gradient descent scheme such that the reconstruction error

$$\frac{1}{2} \left\| G - \sum_{j=1}^k \otimes_{i=1}^N \mathbf{u}_i^j \right\|^2 \quad (2)$$

is minimized. Given positive initial estimates of the vectors the method results in a non-negative decomposition of the tensor.

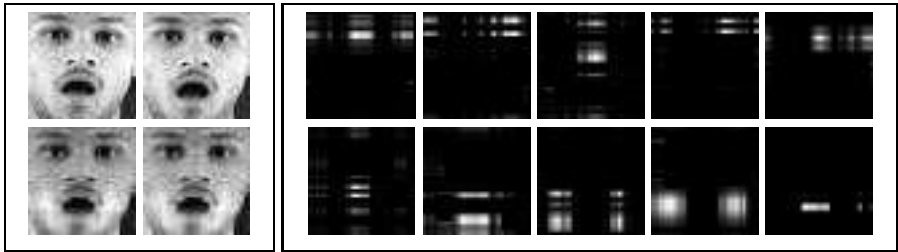
### 2.1 Video Factorization

A collection of images of an object or a video of an object can naturally be represented as a tensor by stacking the 2-D images as slices of a tensor. An alternative interpretation of the decomposition of a video provides insight into the selection of factors that best characterize the expression. Figure 2 shows the factorization of a video into constituent low-rank factors. The  $t$ -th 2D slice of the tensor  $G$  is given by  $G_t = \sum_{j=1}^k u_{3,t}^j (\mathbf{u}_1^j \otimes \mathbf{u}_2^j)$  which is a linear combination of the matrices  $\mathbf{u}_1^j \otimes \mathbf{u}_2^j$  weighted by the  $t$ -th coefficients of  $\mathbf{u}_3^j$ . The matrices  $\mathbf{u}_1^j \otimes \mathbf{u}_2^j$



**Fig. 2.** Tensor Factorization of a video. Each frame of the video is sum of the corresponding frames from the low rank factors. Each low rank factor is formed by stacking weighted versions of a basis image. Each basis image is a rank-1 matrix i.e. outer product of two vectors.

can be viewed as basis images which capture the appearance of the object. The positivity and sparseness imply that these images loosely correspond to parts of the object in the video. Figure 3 shows the factors obtained on factorization of a face video and the factors obtained. It can be seen that the energy in these images is located near the regions corresponding to various parts of the face like nose, eye etc.



**Fig. 3.** Basis images corresponding to factors obtained by tensorial factorization of a face video (with the expression *surprise*). The first table shows the representative frames of the original video (top row) and the reconstructed video (second row). The second table shows a subset of the basis-image set. It can be seen that the energy in these images is concentrated near the regions which correspond to location of parts like cheeks, mouth, nose etc.

## 2.2 Basis Image Based Method

The identification of factors that best represent an expression requires the knowledge of the appearance of a neutral face. A neutral face video of the same subject as in the expression video is used for this purpose. Identification of the relevant factors is done by factorizing both the expression video and the neutral face video. Let  $V_E$  and  $V_N$  be the expression and neutral face videos and the  $\mathbf{U}_E = \{\mathbf{u}_j^i\}$  and  $\mathbf{U}_N = \{\mathbf{v}_j^i\}$  be the factors obtained by decomposing the tensor

representation of  $V_E$  and  $V_N$ . The factors thus obtained lack alignment i.e. there is no correspondence between factors that affect the appearance of the same region in the two videos. To facilitate selection of relevant factors, the factors must be aligned. The correspondence between the factors can be established using a greedy algorithm that uses the similarity score between elements of the basis image sets  $\mathbf{B}_E$  and  $\mathbf{B}_N$  corresponding to  $\mathbf{U}_E$  and  $\mathbf{U}_N$  respectively. Since the basis images are all positive and sparse with local patches rudimentary metrics like sum of squared difference capture the similarity well. For the current work a similarity score based on the distance between the centroids of the local patches in the two images and the distribution of pixels around the centroid is used. The correspondence between the factors belonging to the two sets can be established by selecting the best matching pair of factors, eliminating them and then repeating the process for the remaining factors. Algorithm 1 gives the complete description of alignment of factors.

---

**Algorithm 1.** AlignFactors( $\mathbf{U}_E, \mathbf{U}_N, k$ )

---

```

1: Build basis image sets :  $\mathbf{B}_E \leftarrow \mathbf{U}_E, \mathbf{B}_N \leftarrow \mathbf{U}_N$ 
2: Compute similarity scores :  $S_{ij} \leftarrow \text{similarity}(\mathbf{B}_E^i, \mathbf{B}_N^j)$ 
3:  $I \leftarrow \phi$ 
4: for  $i = 1$  to  $k$  do
5:   Find  $p, q$  such that  $S_{pq}$  is maximum where  $\mathbf{B}_E^p \in \mathbf{B}_E, \mathbf{B}_N^q \in \mathbf{B}_N$ 
6:    $I = I \cup \{(p, q)\}$ 
7:    $\mathbf{B}_E \leftarrow \mathbf{B}_E - \{\mathbf{B}_E^p\}, \mathbf{B}_N \leftarrow \mathbf{B}_N - \{\mathbf{B}_N^q\}$ 
8: end for
9: return  $I$ 

```

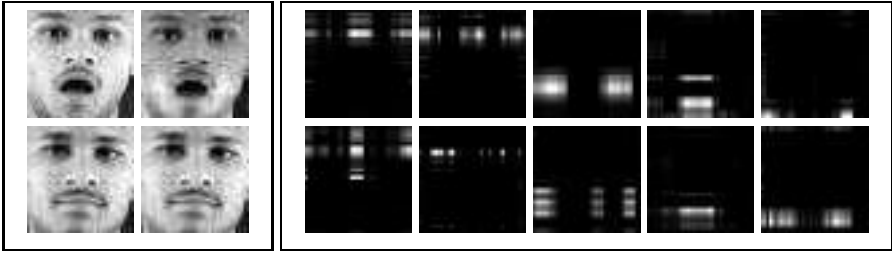
---

Figure 4 shows the results of aligning the factors using the proposed algorithm. Once the corresponding factors have been identified the factors in the expression video corresponding to the maximally dissimilar pairs of factors are more likely to cause the expression appearance. Figure 5 shows the factors identified in this manner. It is evident that the factors that are maximally dissimilar have energy centered around the mouth region where the appearance differs drastically from that of a neutral face. The factors thus identified can now be used to transfer the expression. Besides the expression transfer problem, the factors can be used as cues for recognition of expressions. Sections 3 and 4 discuss these applications.

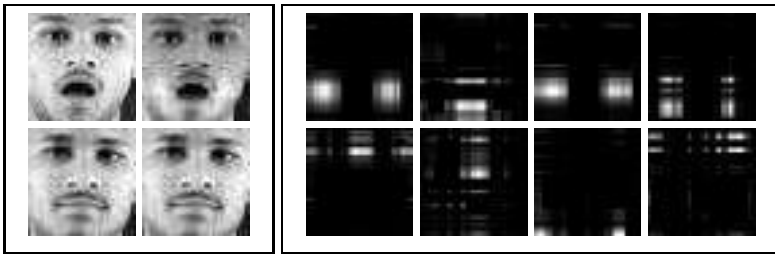
### 2.3 Factorization of the Difference Tensor

An alternative scheme for identification of factors that are useful for recognition arises from the Fisher-like criterion for discriminant analysis. Given two videos  $V_1$  and  $V_2$  represented as tensors  $\mathbf{G}_1$  and  $\mathbf{G}_2$  the objective is to find a rank- $k$  tensor  $\mathbf{W} = \sum_{j=1}^k \otimes_{i=1}^3 \mathbf{w}_i^j$  such that projections of  $\mathbf{G}_1$  and  $\mathbf{G}_2$  on to this tensor differ the most. The objective is to optimize the quantity

$$\left\langle \sum_{j=1}^k \otimes_{i=1}^3 \mathbf{w}_i^j, \mathbf{G}_1 \right\rangle - \left\langle \sum_{j=1}^k \otimes_{i=1}^3 \mathbf{w}_i^j, \mathbf{G}_2 \right\rangle \quad (3)$$

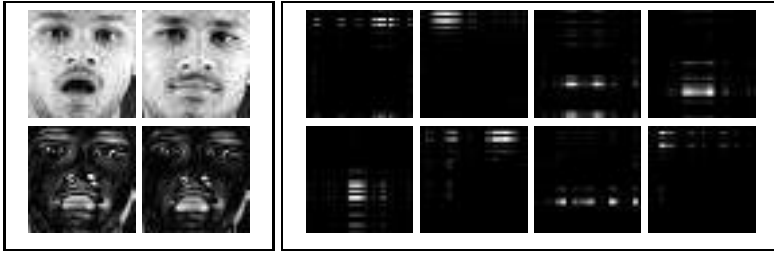


**Fig. 4.** Alignment of factors corresponding to a facial expression video and a neutral face video. The first table shows frames from both the videos and their reconstructions. The second table shows the aligned factors. The factors of the first video are shown in the top row and the corresponding factors are shown in the next row.



**Fig. 5.** Identification of expression-specific factors using basis images. The first table shows representative frames from two videos and their reconstructions from the factors : a *surprise* expression video (top row) and a neutral face video of the same subject (second row) . The second table (top row) shows the basis images corresponding to the factors chosen by the method. Note that the energy in these images is centered around the mouth region where the appearance differs significantly. The result of transferring these factors to the neutral video is shown in Figure 7. The second row shows the basis that are least preferred. These images resemble the invariant parts like nose and eyes.

which is equivalent to optimizing  $\langle \sum_{j=1}^k \otimes_{i=1}^3 \mathbf{w}_i^j, \Delta \mathbf{G} \rangle$  where  $\Delta \mathbf{G} = \mathbf{G}_1 - \mathbf{G}_2$  captures the changes in the appearance of the video. Although the tensor  $\Delta \mathbf{G}$  is not guaranteed to be positive, it can be normalized such that the elements are all non-negative and  $\mathbf{W}$  is estimated as the rank- $k$  approximation of  $\Delta \mathbf{G}$ . The factors  $\mathbf{w}_i^j$  are not useful for the problem of expression transfer but when the basis images corresponding to these factors are used as filters, the frames of the two videos give rise to markedly different responses to those filters. Thus the factors provide a bank of highly discriminative filters that can be used effectively for classification tasks. Figure 6 shows the factors obtained by factorizing the difference tensor. It can be seen that the basis images corresponding to the resulting factors are useful for discriminating between the two expressions (*neutral* and *surprise*).



**Fig. 6.** Identification of discriminative filters by factorizing the difference tensor. The first table (top row) shows representative frames from both the videos. The next row shows frames from the difference tensor and reconstructions. The next table shows the basis images corresponding to the factors chosen by the algorithm obtained by factorizing the difference tensor. The energy in the images is centered around the mouth and the eyebrows where the appearance differs from the neutral face.

### 3 Facial Expression Transfer

The problem of expression transfer has a the flavor of style and content separation problem where the expression can be thought of as style and the underlying face as content and the appearance as a result of interaction between these two. An alternative pair of factors that cause the appearance of a face are the shape and texture characteristics of the face. However, the interaction between such factors may not be amenable to simple models like linear or bilinear models. As a simple and useful alternative the appearance factors obtained by decomposition of the facial expression video are be used to transfer the expression. As observed above the knowledge of appearance of the neutral face is a prerequisite for identification of the expression-specific factors as well as for transfer of the expression. The expression transfer problem considered in the current work is posed as follows: Given a video  $V_{1E}$  of a subject  $P_1$  with certain expression  $E$ , the neutral face videos of the subject  $P_1$  and another subject  $P_2$ , synthesize the video  $V_{2E}$  of subject  $P_2$  with the same expression  $E$ .

The expression specific factors can be identified by using the algorithm described in Section 2. Once the factors are aligned and the relevant factors are identified the transfer is achieved by simply transferring the expression-specific factors in the source video to the target neutral video. Figure 7 shows the reconstructed video after the expression-specific factors are transferred to the neutral faces of the same subject and a second subject. Algorithm 2 summarizes the algorithm for expression transfer. The appearance based solution here transfers the appearance factors alone and might appear like a *cut and paste* method that results in discontinuities in the frames of synthesized videos. However experiments have shown that the synthesized videos are visually satisfactory with little or no discontinuities. However, the quality of synthesized video depends on the source video i.e. how close the shape of the source face is to the target face, differences in facial features like complexion, how well the expression is articulated in the source video etc.

**Algorithm 2.**  $\text{TransferExpression}(V_{1E}, V_{1N}, V_{2N}, k)$ 

- 
- 1:  $\mathbf{U}_{1E} \leftarrow \text{NTF}(V_{1E}); \mathbf{U}_{1N} \leftarrow \text{NTF}(V_{1N}); \mathbf{U}_{2N} \leftarrow \text{NTF}(V_{2N})$
  - 2:  $I = \text{AlignFactors}(\mathbf{U}_{1E}, \mathbf{U}_{1N})$
  - 3:  $J = \text{AlignFactors}(\mathbf{U}_{1N}, \mathbf{U}_{2N})$
  - 4: Choose  $\mathbf{U}_{1E}^* \subset \mathbf{U}_{1E}$  such that elements of  $\mathbf{U}_{1E}^*$  are maximally dissimilar to corresponding factors in  $\mathbf{U}_{1N}$ ,  $\mathbf{U}_{1N}^*$  be the corresponding factors in  $\mathbf{U}_{1N}$
  - 5: Find  $\mathbf{U}_{2N}^* \subset \mathbf{U}_{2N}$  the factors corresponding to  $\mathbf{U}_{1E}^*$  using  $I, J$  and  $\mathbf{U}_{1N}^*$
  - 6:  $\mathbf{U}_{2E} \leftarrow \mathbf{U}_{2N} \cup \mathbf{U}_{1E}^* - \mathbf{U}_{2N}^*$
  - 7:  $V_{2E} \leftarrow \text{NTFReconstruct}(\mathbf{U}_{2N})$
- 

**Experimental Results.** Experiments were conducted on a dataset that was collected in-house. An OLYMPUS C7000 camera was used to capture the videos at 30fps under controlled settings with minimal illumination or pose variation. The videos were preprocessed to segment the faces and scale them to a fixed size. The algorithms proposed here work best in presence of a good degree of alignment of faces in the videos. The dataset consisted of 13 subjects in 8 different expressions including a neutral face video for each subject. The frame count in the videos was equalized (to 60), for the ease of implementation, by deleting frames where there was no change in appearance. The frames were all scaled to fixed dimensions. Figure 7 shows representative results of the experiments on this dataset. It can be seen that the results are visually satisfactory despite appearance information alone being made use of. The minor discontinuities and artifacts are due to the erroneous transfer of some rank-1 factors which come out as horizontal/vertical lines. Further, large variations in complexion and face shape also result in discontinuities. Low-pass filtering and contrast enhancement were applied on the frames of the synthesized video as a post-processing step.

## 4 Facial Expression Recognition

The second application considered is the recognition of facial expressions. Recognition of facial expression is a challenging task as the appearance of expressions varies drastically for different subjects. We explore the possibility of solving this problem using appearance based features alone by using the factors obtained by tensorial factorization. We used the neutral face video during training for identification of expression-specific factors. The recognition is done by comparing the constituent factors of the test video with the expression specific factors of the samples in the training set. For each expression-specific factor set in the training data a maximally similar subset of the factors of the test video is found. The matching score is computed as the mean similarity score between the matched factors. The test video is assigned the label of the training sample which results in maximum matching score.

A second method for recognition of expressions uses the factors obtained by decomposition of the difference tensor. First a classifier that can discriminate between two expressions is built and the the DDAG architecture [11] is used to





**Fig. 7.** Results of expression transfer using the algorithm described in section 3. First table : The top row shows frames from the facial expression video and the neutral face video considered for expression transfer (The first neutral face corresponds to the same subject). The second row shows the frames of the synthesized video using the algorithm. The third and fourth rows show four expressions (surprise, wink, smile, giggle ) transferred to the neutral face of another subject. Second table : the expression transfer algorithm on a set of four subjects where the expression in the videos along the diagonal were transferred to other subjects (the columns show the same expression for different subjects and the rows show the same subject in different expressions). Only the diagonal videos were originally available.

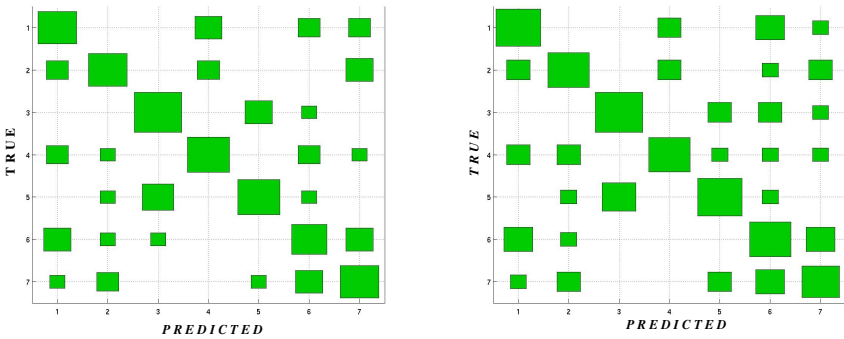
extend it to multiple expressions. The basis images corresponding to the factors of the difference tensors are used as filters and the mean response to each filter over the entire video is taken as a feature. Feature vectors built in this manner are compared to training samples using Euclidean distance as the metric.

**Experimental results.** The dataset used for expression recognition is same as the dataset used for the expression transfer experiments. Excluding the neutral video there were 7 expressions of 13 subjects. Since the sample size was small leave-one out mode testing was used for testing both the algorithms. The Hinton diagrams corresponding to the confusion matrices for both the methods table 1. The leading diagonal elements show that the recognition accuracy is quite satisfactory despite the use of appearance information alone. The accuracy improved when the images of the faces of different persons are further aligned by manually selecting the control points such the centers of eyes and tip of the nose. The overall accuracy is around 51% comparable to the state of the art methods which use feature correspondences and muscular motion models [12]. The size of the dataset that is used for the experiments precludes any definite conclusion. However, as the nature of information used by the current technique differs from that of existing ones, they are complementary and development of a hybrid solution with improved recognition rate is a promising direction.



**Fig. 8.** Expressions used for the expression recognition experiment : *Smile, Left wink, Surprise, Giggle, Mock, Right wink, Disgust*

**Table 1.** The Hinton diagram of the confusion matrix for the expression recognition task. The squares along the diagonal represent fraction of correctly recognized samples. It can be seen that the accuracy is reasonable despite the absence of feature correspondences or complex modeling schemes.



## 5 Relationship to Past Work

Expression transfer in face videos has number of applications in gaming, virtual worlds, interactive systems etc. It has been studied extensively both in computer vision and graphics. A method based on ratio images is presented in [13] which can transfer expression as well as capture illumination variation. Du and Lin [14] attempt to learn a linear mapping between parameters representing expression and appearance. Wang and Ahuja [4] use Higher-Order SVD (*HOSVD*) to decompose a collection of face expression images into two separate expression and person subspaces and use them to map expression on to a new face. Existing methods use feature point correspondences or complex motion models for expression transfer. The method that proposed in the current work is purely appearance based and is sensitive to various sources of differences in appearance such as the complexion, face shape etc. However, it does not use correspondences or complex modeling schemes for the task and is widely applicable. Moreover the current method is computationally efficient since the algorithm is dominated by the factorization step which has good convergence properties [10]. Facial Expression Recognition is a complex problem since the facial appearance corresponding to the same expression varies widely across different subjects. Previous methods use facial muscle motion models like FACS [15] and employ complex classifiers [12].

The method proposed here recognizes expressions using appearance cues alone and is complementary to the existing methods.

## 6 Conclusion

We have proposed methods for identification of task-specific factors that are useful for a range of tasks like video manipulation and recognition. To the best of our knowledge, a purely appearance based approach to expression transfer and expression recognition has not been attempted so far. The task-specific factor selection scheme provides efficient solutions for the tasks of expression transfer and recognition. The contributions of the current work are:

- Simple and efficient methods for selection of task-specific factors from the set of factors obtained by tensorial factorization of videos.
- An efficient technique to perform facial expression transfer without requiring feature correspondences or muscular motion models.
- A novel technique complementary to existing methods for recognition of expressions using expression-specific factor in facial expression videos.

The methods based on tensor representation of videos are more natural, simpler and insightful. Multilinear techniques offer new insights into analysis and processing of video information. The factors obtained by factorization of videos can be used for a number of other purposes like dynamic event analysis and background modeling. The appearance and dynamics separation achieved by tensorial factorization provides valuable cues for analysis of dynamic events in videos and we are actively pursuing this problem. In summary, task-specific factor selection makes it possible to solve a wide range of problems using tensorial factorization of videos/image-cubes and is a promising direction for future research.

## Acknowledgments

One of the authors, Ranjeeth Kumar, would like to acknowledge the financial support provided by GE (through the GE Foundation Scholar-Leaders Program 2004-06) while carrying out this work.

## References

1. Carlo Tomasi, Takeo Kanade: Shape and motion without depth. In: Proc. of the Third IEEE International Conf. on Computer Vision. (1990) 91–95
2. D. D. Lee, H. S. Seung: Learning the parts of objects by non-negative matrix factorization. *Nature* **401** (1999) 788–791
3. Joshua B. Tenenbaum, William T. Freeman: Separating style and content with bilinear models. *Neural Computation* **12** (2000) 1247–1283
4. Hongcheng Wang, Narendra Ahuja: Facial expression decomposition. In: Proc. of the Ninth IEEE International Conf. on Computer Vision. (2003) 958–965

5. Tamir Hazan, Simon Polak, Amnon Shashua: Sparse image coding using a 3d non-negative tensor factorization. Proc. of the Tenth IEEE International Conf. on Computer Vision (2005) 50–57
6. Zoubin Ghahramani, Geoffrey E. Hinton: The EM algorithm for mixtures of factor analyzers. Technical Report CRG-TR-96-1 (1996)
7. Max Welling, Markus Weber: Positive tensor factorization. Pattern Recognition Letters **22** (2001) 1255–1261
8. M. Alex, O. Vasilescu, Demetri Terzopoulos: Multilinear analysis of image ensembles: Tensorfaces. In: Proc. of the Seventh European Conf. on Computer Vision. Volume 1. (2002) 447–460
9. A. Shashua, A. Levin: Linear image coding for regression and classification using the tensor-rank principle. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. (2001) 42–49
10. Amnon Shashua, Tamir Hazan: Non-negative tensor factorization with applications to statistics and computer vision. Proc. of the International Conf. on Machine Learning (2005) 792–799
11. John Platt, Nello Cristianini, John Shawe-Taylor: Large margin dags for multi-class classification. In Solla, S., Leen, T., Mueller, K.R., eds.: Advances in Neural Information Processing Systems 12. (2000) 547–553
12. Nicu Sebe, Michael S. Lew, Ira Cohen, Ashutosh Garg, Thomas S. Huang: Emotion recognition using a cauchy naive bayes classifier. In: Proc. of the 16 th International Conf. on Pattern Recognition. Volume 1. (2002) 17–20
13. Zicheng Liu, Ying Shan, Zhengyou Zhang: Expressive expression mapping with ratio images. In: Proc. of the 28th Annual Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH). (2001) 271–276
14. Du Y., Lin X.: Mapping emotional status to facial expressions. In: Proc. of the International Conf. on Pattern Recognition. Volume II. (2002) 524–527
15. P. Ekman, W. V. Friesen: Facial Action Coding System : Investigator’s Guide. Consulting Pshycologists Press, Palo Alto, CA (1978)