Face Mosaicing for Pose Robust Video-Based Recognition *

Xiaoming Liu[1] Tsuhan Chen[2]

Visualization and Computer Vision Lab,
General Electric Global Research, Schenectady, NY, 12309[1]
Advanced Multimedia Processing Lab,
Carnegie Mellon University, Pittsburgh, PA, 15213[2]

Abstract. This paper proposes a novel face mosaicing approach to modeling human facial appearance and geometry in a unified framework. The human head geometry is approximated with a 3D ellipsoid model. Multi-view face images are back projected onto the surface of the ellipsoid, and the surface texture map is decomposed into an array of local patches, which are allowed to move locally in order to achieve better correspondences among multiple views. Finally the corresponding patches are trained to model facial appearance. And a deviation model obtained from patch movements is used to model the face geometry. Our approach is applied to pose robust face recognition. Using the CMU PIE database, we show experimentally that the proposed algorithm provides better performance than the baseline algorithms. We also extend our approach to video-based face recognition and test it on the Face In Action database.

1 Introduction

Face recognition is an active topic in the vision community. Although many approaches have been proposed for face recognition [1], it is still considered as a hard and unsolved research problem. The key of a face recognition system is to handle all kinds of *variations* through *modeling*. There are different kinds of variations, such as pose, illumination, expression, among which, pose variation is the hardest, and contributes more recognition errors than others [2]. In the past decade, researchers mainly model each variation separately. For example, by assuming constant illumination and the frontal pose, expression invariant face recognition approaches are proposed [1]. However, although most of these approaches perform well for specific variation, the performance degrades quickly when multiple variations present, which is the case in real-world applications [3].

Thus, a good recognition approach should be able to model different kinds of variations in an efficient way. For human faces, most prior modeling work target at facial appearance using various pattern recognition tools, such as Principal Component Analysis (PCA) [4], Linear Discriminate Analysis [5], Support

* The work presented in this paper is performed in Advanced Multimedia Processing Lab, Carnegie Mellon University.

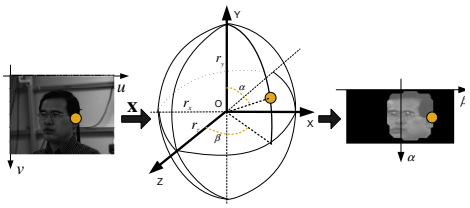

Fig. 1. Geometric mapping

Fig. 2. Up to 25 labeled facial features.

Vector Machine [5]. On the other hand, except for the 3D face recognition, the human face geometry/shape is mostly overlooked in face recognition. We believe that, similar to the facial appearance, the face geometry is also a unique characteristic of human being. Face recognition can benefit if we can properly model the face geometry, especially when pose variation is presented.

This paper proposes a face mosaicing approach to modeling both the facial appearance and geometry, and applies it to face recognition. This paper extends the idea introduced in [6, 7] by approximating the human head with a 3D ellipsoid. As shown in Fig. 1, an arbitrary view face image can be back projected onto the surface of the 3D ellipsoid, and results in a texture map. In multi-view facial images based modeling, combining multiple texture maps is conducted, where the same facial feature, such as the mouth's corner, from multiple maps might not correspond to the same coordinate on the texture map. Hence the blurring effect, which is normally not a good property for modeling, is observed.

To reduce such blurring, the texture map is decomposed into a set of local patches. Patches from multi-view images are allowed to move locally for achieving better correspondences. Since the amount of movement indicates how much the actual head geometry deviates from the ellipsoid, a deviation model trained from patch movements models the face geometry. Also the corresponding patches are trained to model facial appearance. Our mosaic model is composed of both models together with a probabilistic model $\mathbf{P}_d$ that learns the statistical distribution of the distance measure between the test patch and the patch model [8].

Our face mosaicing approach makes a number of contributions. First, as the hardest variation, pose variation is handled naturally by mapping images from different view-angles to form the mosaic model, whose mean image can be treated as a compact representation of faces under various view-angles. Second, all other variations that could not be modeled by the mean image, for example, illumination and expression, are taken care of by a number of eigenvectors. Therefore, instead of modeling only one type of variation, as done in conventional methods, our method models all possible appearance variations under one framework. Third, a simple geometrical assumption has the problem since the head geometry is not truly an ellipsoid. This is taken care of by training a geometric deviation model, which results in better correspondences across multiple views.

There are many prior work on face modeling [9, 10]. Among them, Blanz and Vetter's approach [9] is one of the most sophisticated that applied to face recognition as well, where two subspace models are trained for facial texture and

shape respectively. Given a test image, they fit the new image with two models by tuning the models' coefficients, which are eventually used for recognition. Intuitively better modeling leads to better recognition performance. However, a more sophisticated modeling also makes model fitting to be too difficult. For example, both training and test images are manually labeled with 6 to 8 feature points [9]. On the other hand, we believe that, unlike the rendering applications in computer graphics, we might not need a very sophisticated geometric model for recognition applications. The benefit with a simpler face model is that model fitting tends to be easier and automatic, which is the goal of our approach.

2 Modeling the Geometric Deviation

To reduce the blurring issue in combining multiple texture maps, we obtain a better facial feature alignment by relying on the landmark points. For the model training, it is reasonable to manually label such landmark points.

Given K multi-view training facial images, $\{\mathbf{f}_k\}$, firstly we label the position of facial feature points. As shown in Fig. 2, 25 facial feature points are labeled. For each training image, only a subset of the 25 points is labeled according to their visibility. We call these points as *key points*.

Second, we generate the texture map $\mathbf{s}^k$ from each training image, and compute key points' corresponding coordinates $\mathbf{b}_k^i (1 \leq i \leq 25)$ in the texture map $\mathbf{s}^k$, as shown in Fig. 3. Furthermore, we would like to find the coordinate on the mosaic model where all corresponding key points deviate to. Ideally if the human head is a perfect 3D ellipsoid, the same key point $\mathbf{b}_k^i (1 \leq k \leq K)$ from multiple training texture maps should exactly correspond to the same coordinate. However, due to the fact that the human head is not a perfect ellipsoid, these key points deviate from each other. The amount of deviation is an indication of the geometrical difference between the actual head geometry and the ellipsoid.

Third, we compute the averaged positions $\mathbf{b}_k^i (1 \leq k \leq K)$ of all visible key points $\mathbf{b}^i$ that correspond to the same facial feature. We treat this averaging, shown in the 3^{rd} row of Fig. 3, as the target position in the final mosaic model where all corresponding key points should move toward. Since our resulting mosaic model is composed of an array of local patches, each one of the 25 averaged key points falls into one particular patch, namely *key patch*.

Fourth, for each texture map, we take the difference between the positions of key point $\mathbf{b}_k^i$ and that of the averaged key point $\mathbf{b}^i$ as the key patch's deviation flow (DF) that describes which patch from each texture map should move toward that key patch in the mosaic model. However, there are also non-key patches in the mosaic model. As shown in Fig. 4, we represent the mosaic model as a set of triangles, whose vertexes are the key patches. Since each non-key patch falls into at least one triangle, its DF is interpolated by the key patch's DF.

For each training texture map, its geometric deviation is a 2D vector map $\mathbf{v}^k$, whose dimension equals to the number of patches in vertical and horizontal directions, and each element is one patch's DF. Note that for any training texture map, some elements in $\mathbf{v}^k$ are considered missing. Finally the deviation model

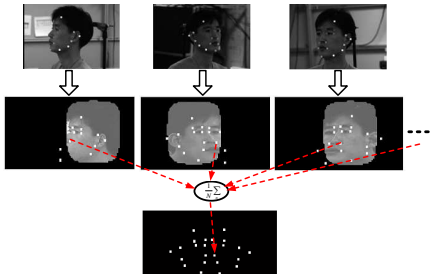

Fig. 3. Averaging key points: the position of key points in the training texture maps (2^{nd} row), which correspond to the same facial feature are averaged and result in the position in the final model (3^{rd} row).

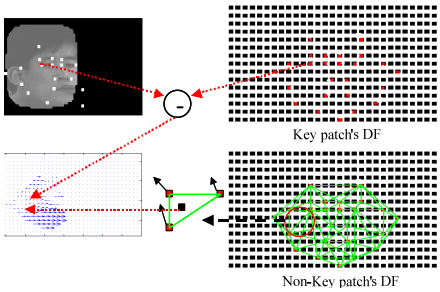

Fig. 4. Computation of patch's DF: each non-key patch falls into at least one triangle; the deviation of a non-key patch is interpolated by the key patch deviation of one triangle.

$\theta = \{\mathbf{g}, \mathbf{u}\}$ is learned from the geometric deviation $\{\mathbf{v}^k\}$ of all training texture maps using the robust PCA [11], where $\mathbf{g}$ and $\mathbf{u}$ are the mean and eigenvectors respectively. Essentially this linear model describes all possible geometric deviation of any view angle for this particular subject's face.

3 Modeling the Appearance

After modeling the geometric deviation, we need to build an appearance model, which describes the facial appearance for all poses. On the left hand side of Fig. 5, there are two pairs of training texture maps $\mathbf{s}^k$ and their corresponding geometric deviation $\mathbf{v}^k$. The resulting appearance model $\mathit{\Pi} = \{\mathbf{m}, \mathbf{V}\}$ with one mean and two eigenvectors are shown on the right hand side. This appearance model is composed of an array of eigenspaces, where each is devoted to modeling the appearance of the local patch indexed by (i, j). In order to train one eigenspace for one particular patch, the key issue is to collect one corresponding patch from each training texture map $\mathbf{s}^k$, where the correspondence is specified by the geometric deviation $\mathbf{v}^k_{i,j}$. For example, the summation of $\mathbf{v}^1_{i,j}$ and (40,83) determines the center of corresponding patch, $\mathbf{v}^1_{i,j}$, in the texture map $\mathbf{s}^1$. Using the same procedure, we find the corresponding patches $\mathbf{s}^k_{i,j} (2 \leq k \leq K)$ from all other texture maps. Note some of $\mathbf{s}^k_{i,j}$ might be considered as missing patches. Finally the set of corresponding patches are used to train a statistical model $\mathit{\Pi}_{i,j}$ via PCA. We call the array of PCA models as the *patch-PCA mosaic*. Modeling via PCA is popular when the number of training samples is large.

However, when the number of training samples is small, such as the training of an individual mosaic model with only a few samples, it might not be suitable to train one PCA model for each patch. Instead we would rather train a universal PCA model based on all corresponding patches of all training texture maps, and keep the coefficient of these patches in the universal PCA model as well. This is

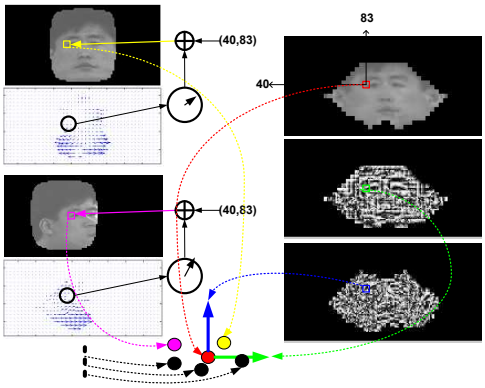

Fig. 5. Appearance modeling: the deviation indicates the corresponding patch for each of training texture maps; all corresponding patches are treated as samples for PCA.

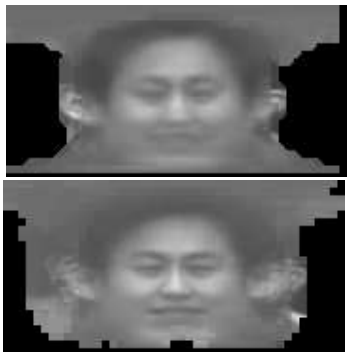

Fig. 6. The mean images of two mosaic models without geometric deviation (top) and with geometric deviation (bottom).

called the *global-PCA mosaic*. Note that the patch-PCA mosaic and the global-PCA mosaic only differ in how the corresponding patches across training texture maps are utilized to form a model, depending on the availability of training data in different application scenarios.

Eventually the statistical mosaic model includes the appearance model Π, the geometric deviation model θ and the probabilistic model $\mathbf{P}_d$. We consider that the geometric deviation model plays a key role in training the mosaic model. For example, Fig. 6 shows the mean images of two mosaic models trained with the same set of images from 10 subjects. It is obvious that the mean image on the bottom is much less blurring and captures more useful information about facial appearance. Note that this mean image covers much larger facial area comparing to the up-right illustration of Fig. 5 since extrapolation is performed while computing the geometric deviations of non-key patches.

4 Face Recognition using the Statistical Mosaic Model

Given L subjects with K training images per subject, an individual statistical mosaic model is trained for each subject. For simplicity, let us assume we have enough training samples and obtain the patch-PCA mosaic for each subject. We will discuss the case of the global-PCA mosaic in the end of this section. We now introduce how to utilize this model for pose robust face recognition.

As shown in Fig. 7, given one test image, we generate its texture map by using the universal mosaic model, which is trained from multi-view images of many subjects. Then we measure the distance between the test texture map and each of the trained individual mosaic model, namely the *map-to-model* distance. Note that the appearance model is composed of an array of patch models, which is called the *reference patch*. Hence, the map-to-model distance equals to the

summation of the map-to-patch distances. That is, for each reference patch, we find its corresponding patch from the test texture map, and compute its distance to the reference patch.

Since we have been deviating corresponding patches during the training stage, we should do the same while looking for the corresponding patch in the test stage. One simple approach is to search for the best corresponding patch for the reference patch within a search window. However, this does not impose any constraint on the deviation of neighboring reference patches. To solve this issue, we make use of the deviation model that was trained before.

As shown in Fig. 7, if we randomly sample one coefficient in the deviation model, the linear combination of this coefficient describes the geometric deviation for all reference patches. Hence, the key is to find the coefficients that provide the optimal matching between the test texture map and the model. In this paper, we adopt a simple sequential searching scheme to achieve this. That is, in a K-dimensional deviation model, uniformly sample multiple coefficients along the 1^{st} dimension while the coefficients for other dimensions are zero, and determine one of them which results in the maximal similarity between this test texture map and the model. The range of sampling is bounded by the coefficients of training geometric deviations. Then we perform the same searching along the 2^{nd} dimension while fixing the optimal value for the 1^{st} dimension and zero for all other dimensions. The searching is finished until the K^{th} dimension. Basically our approach enforces the geometric deviation of neighboring patches to follow certain constraint, which is described by the bases of the deviation model.

For each sampled coefficient, the reconstructed 2D geometric deviation (in the bottom-left of Fig. 7) indicates where to find the corresponding patches in the test texture map. Then the residue between the corresponding patch and the reference patch model is computed, which is further feed into the probabilistic model [8]. Finally the probabilistic measurement tells how likely this corresponding patch belongs to the same subject as the reference patch. By doing the same operation for all other reference patches and averaging all patch-based probabilistic measurements, we obtain the similarity between this test texture map and the model based on the current sampled coefficient. Finally the test image is recognized as the subject who provides the largest similarity.

Depending on the type of the mosaic model (the patch-PCA mosaic or the global-PCA mosaic), there are different ways of calculating the distance between the corresponding patch and the reference patch model. For the patch-PCA mosaic, the residue with respect to the reference patch model is used as the distance measure. For the global-PCA mosaic, since one reference patch model is represented by a number of coefficients, the distance measure is defined as the nearest neighbor of the corresponding patch among all these coefficients.

5 Video-based Face Recognition

There are two schemes for recognizing faces from video sequences: image-based recognition and video-based recognition. In image-based recognition, usually the

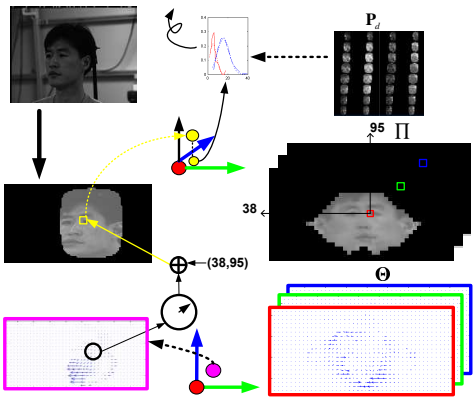

Fig. 7. The map-to-patch distance: the geometric deviation indicates the patch correspondence between the model and the texture map; the distance of corresponding patches are feed into the Bayesian framework to generate a probabilistic measurement.

face area is cropped before feeding to a recognition system. Thus image-based face recognition involves two *separate* tasks: face tracking and face recognition.

In our face mosaicing algorithm, given one video frame, the most important task is to generate a texture map and compare it with the mosaic model. Since the mapping parameter $\mathbf{x}$, which is a 6-dimensional vector describing the 3D head location and orientation [7], contains all the information for generating the texture map, the face tracking is equivalent to estimating $\mathbf{x}$, which can result in the maximal similarity between the texture map and the mosaic model. We use the condensation method [12] to estimate the mapping parameter $\mathbf{x}$.

In image-based recognition, for a face database with L subjects, we build the individualized model for each subject, based on one or multiple training images. Given a test sequence and one specific model, a distance measurement is calculated for each frame by face tracking. Averaging of the distances over all frames provides the distance between the test sequence and one specific model. After the distances between the sequence and all models are calculated, comparing these distances provides the recognition result of this sequence.

In video-based face recognition, two tasks, face tracking and recognition, are usually performed simultaneously. Zhou *et al.* [13] propose a framework to combine the face tracking and recognition using the condensation method. They propagate a set of samples governed by two parameters: the mapping parameter and the subject ID. We adopt this framework in our experiments.

6 Experimental Results

We evaluate our algorithm on pose robust face recognition using the CMU PIE database [14]. We use half of the subjects (34 subjects) in PIE for training the

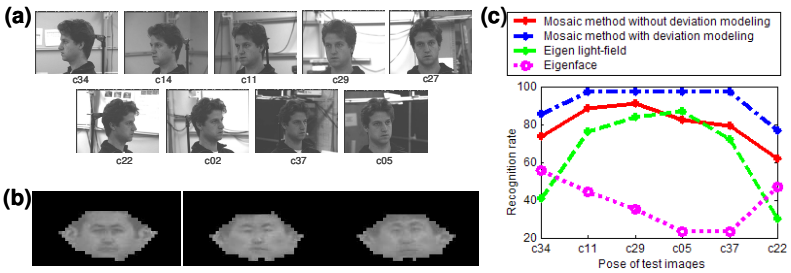

Fig. 8. (a) Sample Images of one subject from the PIE database. (b) Mean images of three individual mosaic models. (c) Recognition performances of four algorithms on the CMU PIE database based on three training images.

probabilistic model. The 9 pose images per subject from remaining 34 subjects are used for the recognition experiments.

Sample images and the pose labels from one subject in PIE are shown in Fig. 8(a). Three poses (c27, c14, c02) are used for the training, and the remaining 6 poses (c34, c11, c29, c05, c37, c22) are used for test using four algorithms. The first is the traditional eigenface approach [4]. We perform the manual cropping and normalization for both training and test images. We test with different number of eigenvectors and plot the one with the best recognition performance. The second is the eigen light-field algorithm [15] (one frontal training image per subject). The third algorithm is our face mosaic method without the modeling of geometric deviation, which essentially sets the mean and all eigenvectors of $\theta = \{\mathbf{g}, \mathbf{u}\}$ to be zero. The fourth algorithm is the face mosaic method with the modeling of geometric deviation. Since the number of training images is small, we train the global-PCA mosaic for each subject. Three eigenvectors are used in building the global-PCA subspace. Thus each reference patch from the training stage is represented as a 3-dimentional vector. For the face mosaic method, the patch size is 4×4 pixels and the size of the texture map is 90×180 pixels. For illustration purpose, we show the mean images of three subjects in Fig. 8(b). Fig. 8(c) shows the recognition rate of four algorithms for each specific pose.

Comparing among these four algorithms, both of our algorithms works better than the baseline algorithms. Obviously the mosaic approach provides a better way of registering multi-view images for an enhanced modeling, unlike the naive training procedure of the traditional eigenface approach. For our algorithms, the one with deviation modeling performs better than the one without deviation modeling. There are at least two benefits for the former. One is that a geometric model can be used in the test stage. The other is that as a result of deviation modeling, the patch-based appearance model also better captures the personal characteristic of the multi-view facial appearance in a non-blurred manner.

We perform video-based face recognition experiments based on the Face In Action (FIA) database [16], which mimics the "passport checking" scenario. Multiple cameras capture the whole process of a subject walking toward the desk, standing in front of the desk, making simple conversation and head motion,

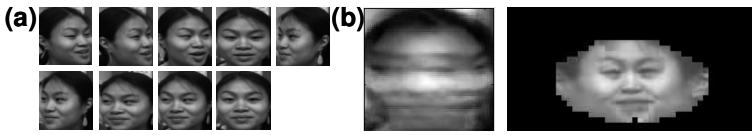

Fig. 9. (a) 9 training images from one subject in the FIA database. (b) The mean images of the individual models in two methods (left: Individual PCA, right: mosaicing).

	PCA	Mosaic
image-based method	17.24%	6.90%
video-based method	8.97%	4.14%

Table 1. Recognition error rate of different algorithms.

and finally walking away from the desk. Six video sequences are captured from six calibrated cameras simultaneously for 20 seconds with 30 frames per second.

We use a subset of the FIA database containing 29 subjects, with 10 sequences per subject as the test sequences. Each sequence has 50 frames, and the first frame is labeled with the ground truth data. We use the individual PCA algorithm [17] with image-based recognition and the individual PCA with video-based recognition as the baseline algorithms. For both algorithms, 9 images per subject are used for training and the best performance is reported by trying different number of eigenvectors. Fig. 9(a) shows the 9 training images for one subject in the FIA database. The face location of training images is labeled manually, while that of the test images is based on the tracking results using our mosaic model. Face images are cropped to be 64×64 pixels from video frames.

We test two options for our algorithms based on the same training set (9 images per subject). The first is to use the individual patch-PCA mosaic with image-based recognition, which uses the averaged distance between the frames to the mosaic model as the final distance measure. The second is to use the individual patch-PCA mosaic with video-based recognition, which uses the 2D condensation method to perform tracking and recognition. Fig. 9(b) illustrates the mean images in two methods. We can observe significant blurring effect in the mean image of the individual PCA model. On the other hand, the mean image of our individual patch-PCA mosaic model covers larger pose variation while keeping enough individual facial characteristics.

The comparison of recognition performance is shown in Table 1. Two observations can be made. First, given the same model, such as the PCA model or the mosaic model, video-based face recognition is better than image-based recognition. Second, the mosaic model works much better than the PCA model for pose-robust recognition.

7 Conclusions

This paper presents an approach to building a statistical mosaic model by combining multi-view face images, and applies it to face recognition. Multi-view face

images are back projected onto the surface of an ellipsoid, and the surface texture map is decomposed into an array of local patches, which are allowed to move locally in order to achieve better correspondences among multiple views. We show the improved performance for pose robust face recognition by using this new method and extend our approach to video-based face recognition.

References

1. Zhao, W.Y., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Survey **35**(4) (2003) 399–458
2. Phillips, P., Grother, P., Micheals, R., Blackburn, D., Tabassi, E., Bone, J.: Face recognition vendor test (FRVT) 2002: Evaluation report. (2003)
3. Sim, T., Kanade, T.: Combining models and exemplars for face recognition: An illuminating example. In: Proc. of the CVPR 2001 Workshop on Models versus Exemplars in Computer Vision. (2001)
4. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3**(1) (1991) 71–86
5. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edition. John Wiley & Sons. Inc., New York (2001)
6. Liu, X., Chen, T.: Geometry-assisted statistical modeling for face mosaicing. In: Proc. 2003 International Conference on Image Processing (ICIP 2003), Barcelona, Catalonia, Spain. Volume 2. (2003) 883–886
7. Liu, X., Chen, T.: Pose-robust face recognition using geometry assisted probabilistic modeling. In: Proc. IEEE Computer Vision and Pattern Recognition, San Diego, California. Volume 1. (2005) 502–509
8. Kanade, T., Yamada, A.: Multi-subregion based probabilistic approach toward pose-invariant face recognition. In: IEEE Int. Symp. on Computational Intelligence in Robotics Automation, Kobe, Japan. Volume 2. (2003) 954–959
9. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. IEEE Trans. on Pattern Analysis and Machine Intelligence **25**(9) (2003) 1063–1074
10. Dimitrijevic, M., Ilic, S., Fua, P.: Accurate face models from uncalibrated and ill-lit video sequences. In: Proc. IEEE Computer Vision and Pattern Recognition, Washington, DC. Volume 2. (2004) 1034–1041
11. De la Torre, F., Black, M.J.: Robust principal component analysis for computer vision. In: Proc. 8th Int. Conf. on Computer Vision, Vancouver, BC. Volume 1. (2001) 362–369
12. Isard, M., Blake, A.: Active Contours. Springer-Verlag (1998)
13. Zhou, S., Krueger, V., Chellappa, R.: Probabilistic recognition of human faces from video. Computer Vision and Image Understanding **91** (2003) 214–245
14. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Trans. on Pattern Analysis and Machine Intelligence **25**(12) (2003) 1615–1618
15. Gross, R., Matthews, I., Baker, S.: Appearance-based face recognition and light-fields. IEEE Trans. on Pattern Analysis and Machine Intelligence **26**(4) (2004) 449–465
16. Goh, R., Liu, L., Liu, X., Chen, T.: The CMU Face In Action (FIA) database. In: Proc. of IEEE ICCV 2005 Workshop on Analysis and Modeling of Faces and Gestures, Beijing, China. (2005)
17. Liu, X., Chen, T., Kumar, B.V.K.V.: Face authentication for multiple subjects using eigenflow. Pattern Recognition **36**(2) (2003) 313–328