Reconstructing a Whole Face Image from a Partially Damaged or Occluded Image by Multiple Matching

Bon-Woo Hwang[1] and Seong-Whan Lee[2]

[1] The Robotics Institute, Carnegie Mellon University,
5000 Forbes Ave., Pittsburgh, PA 15213, USA
bwhwang@cs.cmu.edu
[2] Center for Artificial Vision Research, Korea University,
Anam-dong, Seongbuk-gu, Seoul 136-713, Korea
swlee@image.korea.ac.kr

Abstract. The problem we address in this paper is, given a facial image that is partially occluded or damaged by noise, to reconstruct a whole face. A key process for the reconstruction is to obtain the correspondences between the input image and the reference face. We present a method that matches an input image with multiple example images that are generated from a morphable face model. From the matched feature points, shape and texture of the full face are inferred by the non-iterative data completion algorithm. Compared with single matching with the particular "reference face", this multiple matching method increases the robustness of the matching. The experimental results of applying the algorithm to face images that are contaminated by Gaussian noise and those which are partially occluded show that the reconstructed faces are plausible and similar to the original ones.

Keywords: Face reconstruction, morphable face model, SIFT feature, data completion.

1 Introduction

Reconstructing a whole face from partially damaged facial image due to occlusion or sensor noise can improve the performance of face recognition and authentication applications. Most of previous methods proposed for the purpose take advantage of the fact that face images have a certain statistical structure of shape and texture representable by a lower dimensional subspace, and that therefore once the subspace coefficients are recovered from the input (even partially impaired) the whole face can be reconstructed. The two most popular representations are the Eigenface [4][12] and a morphable model [2][3][5][6][7]; the major difference between the two is whether shape and texture are represented jointly or separately.

The most critical process of the reconstruction is to precisely align the input image to the reference image coordinates so that the subspace coefficients can be computed. The task is not trivial since the input is assumed to be partially damaged. Everson and Sirovich [4], Jones and Poggio [7] established the correspondence of the input face with the iteration of the stochastic gradient procedure. Hwang et al. proposed a

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 692–701, 2007.

method for reconstructing 2D shape and texture from correspondence of a set of 2D point without iteration procedure [5][6]. Blanz et al. presented a method for inferring missing coordinates from sparse 2D or 3D feature coordinates [3]. These matching methods have limitations in that reconstruction of damaged region was obtained as only a side-effect in iterative optimization process [2][7] or in that feature points required for reconstruction should be labeled by hand [3][5][6].

In order to obtain the correspondence between the input facial image and the reference face image without human intervention, the stable and effective algorithms for extracting, describing and matching feature points are required. Matching should be robust to illumination change, noise and deformation within a single object category. One can consider use of Scale Invariant Feature Transform (SIFT) [10]. The SIFT features are invariant to image scale and rotation, and robust to affine distortion, illumination, and additive image noise. In addition to object recognition [8][9][11] in general, SIFT feature matching has been also used for face authentication [1]. However, this algorithm is applied to matching between facial images for two different persons, the number of matched feature points is typically 5 to 15 even if pose, facial expression and illumination conditions are almost identical. For the purpose of face reconstruction, the input and the reference must be matched densely.

In this paper, we present an algorithm to precisely and densely align the input partially damaged image with the reference face coordinates by extending the number of matched feature points between them. We use the morphable face model [2] for shape and texture representation of faces. An input image and the reference images are matched by using a set of example images that are generated from a morphable face model. From the matched feature points, full shape and texture for the input face are inferred from by the non-iterative data completion algorithm. We can reconstruct a facial image similar to original one even from damaged one by Gaussian noise and occluded one by an object.

2 Face Reconstruction

2.1 Face Reconstruction Procedure

If the input face matches with the single reference face, due to the large difference of appearances between the input face and the reference face, only a small number of keypoints can be matched. We prepare multiple example images that are generated from a morphable face model and store the correspondence between the generated example images and the reference face image. If the input facial image is given, the multiple matching results between the input facial image and all the example images are combined into a dense matching between the input facial image and the reference face image.

The entire procedure for reconstructing a facial image consists of 6 steps, categorized into on-line and off-line processes(Fig. 1). In the reconstruction procedure, forward and backward warping mean deformation of a texture onto each face with a shape and deformation of input face onto the defined reference face with a shape [13].

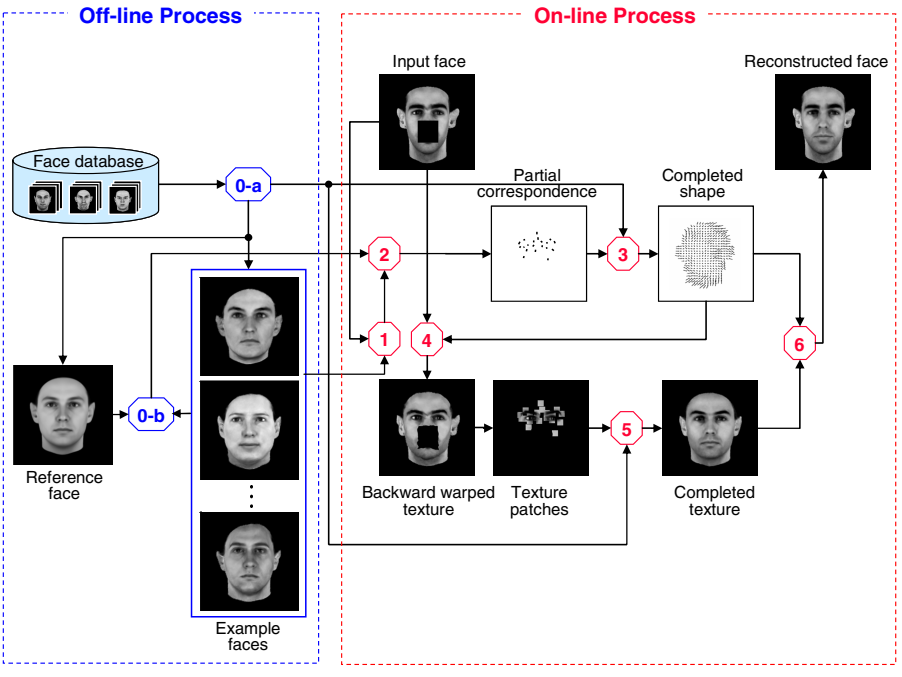

Fig. 1. Face reconstruction procedure: off-line(steps 0-a and 0-b) and on-line(steps 1 to 6) processes

- **Step 0-a:** a) Develop a morphable face model from a face database in which all facial images have dense correspondence with a defined reference face; and b) Synthesize multiple example faces by forward warping textures with shapes. Shape and texture are obtained by generating random coefficients with multivariate Gaussian distribution in the morphable face model.
- **Step 0-b:** a) Obtain SIFT descriptors at keypoints from the example faces by the SIFT algorithm; and b) Compute the corresponding points on the reference face to keypoints on the example face by triangle mesh interpolation from the shape of example face.
- **Step 1:** a) Obtain SIFT descriptors at keypoints from an input face by using the SIFT algorithm; and b) Match the SIFT descriptor of the input face with those of multiple example faces.
- **Step 2:** Obtain the correspondence between the input face and the reference face only in an internal face mask using correspondences among the input face, example faces and the reference face in Step 0-b. The internal face mask includes only main facial components such as eyebrows, eyes, a nose and a mouth and is defined on the reference face.

- **Step 3:** Complete shape from partial correspondence at the matched keypoints by using the data completion algorithm which will be described in section 2.3. The shape eigenvectors $\mathbf{S}$ in the morphable face model are exploited in this step.
- **Step 4:** a) Warp the input face to texture with the completed shape; and b) Extract $n \times n$ texture patch at each keypoint point of reference face from the backward warped texture. Other regions of the backward warped texture are masked.
- **Step 5:** Complete texture from texture patches at the matched keypoints by using the data completion algorithm which is used for shape completion. The texture eigenvectors $\mathbf{T}$ in the morphable face model are also exploited in this step.
- **Step 6:** a) Warp the completed texture with the completed shape only for the inside region defined by the internal face mask. This step results in a reconstructed facial region containing eyebrows, eyes, a nose and a mouth; b) Overlay the reconstructed facial region on the input face to evaluate the reconstruction results.

Steps 0-a and 0-b are performed as an off-line process to prepare: (1) a morphable face model for completing shape and texture; (2) SIFT descriptors of example faces for multiple matching with those of an input face; and (3) correspondence between reference face and the example face. Contrarily, steps 1 to 6 are performed as an on-line process after an input face is given.

2.2 Preparing Multiple Reference Image

By employing Gaussian random coefficients with multivariate normal distribution on the morphable face model (Step 0-a), we can generate facial images as many as required for getting enough number of keypoints matched with those for an input face. In this study, 1,000 example faces are synthesized to match with input face. SIFT descriptors of each example face are obtained at keypoints by using the SIFT algorithm and stored separately (Step 0-b). Next, the correspondence between point on the reference and a keypoint on the example face is computed by using the known shape for the example face and triangle mesh interpolation algorithm. This correspondence is saved into a "keypoint lookup table".

2.3 Matching

If an input face is given for the reconstruction, the SIFT descriptors at keypoints are obtained and matched with those of multiple example faces by using SIFT matching algorithm (Step 1). This on-line multiple matching allows getting point on the reference face corresponding to a keypoint on the input face using the "keypoint lookup table". If a keypoint on the input face corresponds to that on the example face, the counter for corresponding point of reference increases. The counter for each keypoint on the reference face is accumulated for all example face to select the best matched point on the reference face. A single point on the reference face may have several corresponding points on the input face because the input face is matched with multiple example faces. By selecting the point with the largest value, accurate correspondence can be obtained. If the second largest value for the keypoint is greater than the given threshold multiplied by the largest one, this keypoint is rejected similar to Lowe's SIFT matching strategy for the stability of the matched keypoints(Step 2) [11].

It is assumed that the input faces are roughly aligned and normalized by translation, rotation and scale by a face detector. Therefore, if the distance of a matched keypoint in common image coordinate is larger than the given threshold, this is excluded from the list of the matched keypoints. After eliminating the incorrect keypoints, we estimate three parameters, translation, rotation and scale, for the input face by minimizing L2-norm of difference vector of matched points' coordinates. The input face is normalized by using the estimated parameters.

Fig. 2 shows matching between an input face and the reference face by matching the input face with the synthesized multiple example faces. The blue and red dots in Fig. 2 represent the position of SIFT keypoints. The correspondences from keypoints(blue dots) on the example face to points on the reference face are represented by blue dashed arrows and the correspondence from keypoints(red dots) on the input face to those on the example face are represented by red solid arrows. In our experiments, more than 30 keypoints are finally matched between the input face and the reference face even if an input face is damaged by a virtual object(about 10% of face region) or Gaussian noise(standard deviation of 20). These are enough to reconstruct a quite plausible face (Fig. 3)

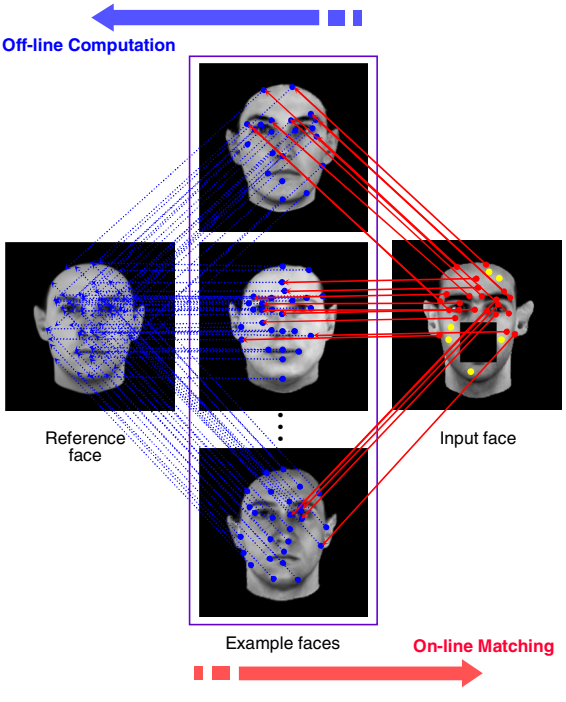

Fig. 2. Indirect matching between the reference face and an input face by using example faces

2.4 Reconstruction

By using defined reference face and pixelwise correspondences between given facial images and the reference face, facial images can be separated into shape and

texture[13]. With the shape S and the texture T separated from the facial image, we fit a multivariate normal distribution to a data set of faces. This is based on the mean of shape, $\overline{S}$ and that of texture $\overline{T}$, covariance matrices $\sum_S$ and $\sum_T$ computed over the differences of the shape and texture:

$$X^S = S - \overline{S},\ X^T = T - \overline{T} \tag{1}$$

By Principal Component Analysis(PCA), a basis transformation is performed to an orthogonal coordinate system formed by eigenvectors S_i and T_i of the covariance matrices, $\sum_S$ and $\sum_T$ on our data set of m faces.

$$S = \overline{S} + \sum_{i=1}^{m-1} c_i^S S_i,\ T = \overline{T} + \sum_{i=1}^{m-1} c_i^T T_i, \tag{2}$$

where $C = (c_1, c_1, \cdots, c_{m-1}) \in \Re^{m-1}$. σ_i^S and σ_i^T are standard deviations within the shape and texture along the eigenvectors S_i and T_i. The dimension of the space spanned by S_i and T_i is at most $m-1$.

To reconstruct a facial image from the obtained correspondence at SIFT keypoints, we can apply two methods: (1) shape and texture reconstruction method using a set of 2D point coordinates[5] and (2) the 3D extended and regularized method[3] of the method (1). In this paper, the latter method is selected due to the robustness and the stability of reconstruction.

The barycentric shape vector X^S is defined by the set of the scaled eigenvector $\sigma_i S_i$ and coefficients c_i :

$$X^S = \sum_{i=1}^{m-1} c_i \sigma_i S_i = \mathbf{S} \cdot diag(\sigma_i) C \tag{3}$$

The data selection matrix, $\mathbf{P}: \Re^n \to \Re^p$ is also defined for representing shape corresponding to the keypoints on an input face. p is the number of matched keypoints. $\mathbf{P}$ is a linear mapping that select a subset of components from an entire vector. Using the data selection matrix $\mathbf{P}$, partial shape vector, F corresponding to the keypoints is represented by

$$F = \mathbf{P}S - \mathbf{P}\overline{S} = \mathbf{P}X \tag{4}$$

The reduced version of the scaled eigenvectors, $\mathbf{Q}$is also defined as:

$$\mathbf{Q} = \mathbf{PS}diag(\sigma_i) \in \Re^{p \times (m-1)} \tag{5}$$

According to the number of keypoints, p and the number of eigenvectors, solution of Equation (4) may be not unique. In order to find optimal coefficients for linear combination of the basis vectors, we minimize the cost function, E, which is given as:

$$\begin{aligned} C^* &= \arg\min_C E(C), \\ E(C) &= \|\mathbf{Q}C - F\|^2 + \kappa \cdot \|C\|^2 \end{aligned} \tag{6}$$

where $\kappa \geq 0$ is a regularization factor. This regularization factor derived from a statistical approach provides the stability and robustness to noise by controlling the tradeoff between fitting accuracy and plausibility[3].

The optimal solution C^* is obtained by Singular Value Decomposition $\mathbf{Q} = \mathbf{UWV}^T$ with diagonal matrix $\mathbf{W} = diag(w_i)$ [3].

$$C^* = \mathbf{V} diag\left(\frac{w_i}{w_i^2 + \kappa}\right)\mathbf{U}^T F \tag{7}$$

The completed shape can be obtained from (1) and (3):

$$S = \overline{S} + S \cdot diag\ (\sigma_i)\mathbf{V}\, diag\left(\frac{w_i}{w_i^2 + \kappa}\right)\mathbf{U}^T F \tag{8}$$

From Equation (8), we can get the complete correspondence for all pixels. Similarly, we can reconstruct complete texture T . From the complete shape and texture, facial image can be synthesized by forward warping.

2.5 Summary of the Process

If an input facial image is given, the partial correspondences between the input face and the reference face are obtained by using multiple matching and a "keypoint lookup table". The partial correspondences are completed to full shape by the data complete algorithm. From the input face image and the completed shape, the backward warped texture is generated and extract texture patch at each keypoint from the texture. The completed texture from texture patches is obtained by using the same data completion algorithm which is used for shape completion. The completed texture is warped with the completed shape to the reconstructed facial image.

3 Experimental Results

3.1 Face Database

Two hundred of 2D faces were used to test and validate the proposed method. These images were rendered with only ambient light using a database of three-dimensional head models recorded with a laser scanner (CyberwareTM) [2][13]. The resolution was 256 by 256 pixels and the color images were converted to 8-bit gray level images. One hundred of facial images in the database were randomly selected to generate a morphable face model by PCA. The other 100 images were used to test our reconstruction algorithm. The test sets were strictly separated from training tests in our experiments.

3.2 Face Reconstruction

If a facial image is given, facial image are reconstructed by the on-line procedure described in section 2.1. Fig. 3 shows the reconstructed examples for two persons.

The facial images for each person are reconstructed from an original face, a damaged face by Gaussian noise and an occluded face by a virtual object, respectively. The damaged faces by Gaussian noise are generated by adding the Gaussian noise with standard deviation of 20 to the original image. The size of the virtual object is 60×60 pixels(about 10% of internal face region) and its position is restricted to the inside of the face in order to occlude facial components such as eyes, a nose and a mouth. Although input faces are damaged by Gaussian noise(left , center in each face group) or occluded by virtual objects(left, bottom), the reconstruction results(right, middle and right, bottom) from them are very similar to original faces(left, top) or the reconstructed faces(right, top) from original faces.

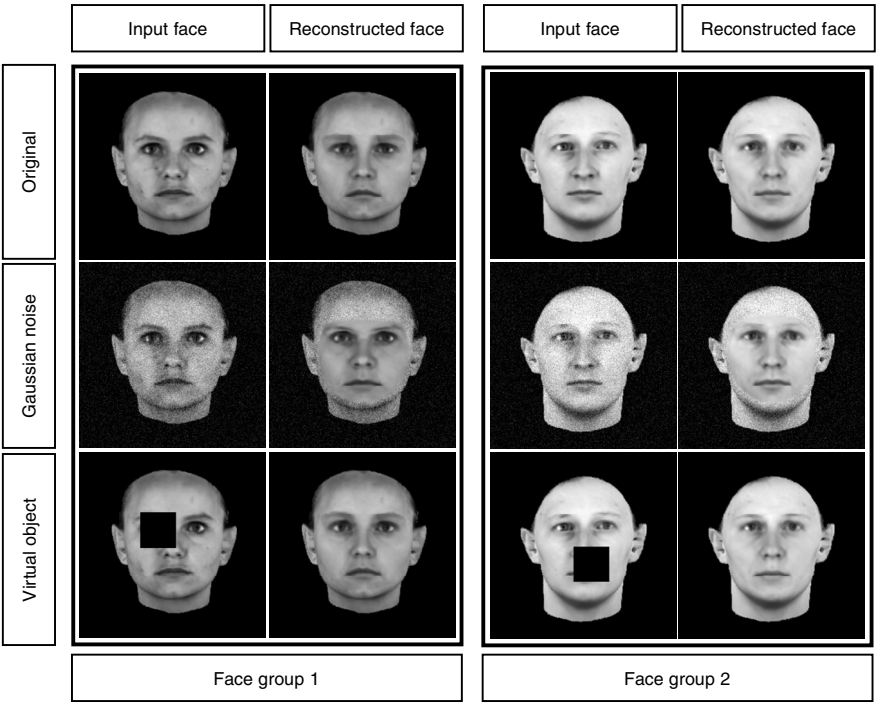

Fig. 3. Examples of input faces and reconstructed faces

Fig. 4 shows the mean reconstruction errors for shapes, textures and synthesized images. Horizontal axis of each graph indicates the type of input for reconstruction face except PCA projection. In the case of PCA projection, we get the projected shape and texture for test faces to the shape and texture eigenvector space, $\mathbf{S}$ and $\mathbf{T}$, and the synthesized image from the projected shape and texture by forward warping. Vertical axes of the graphs are the mean displacement error per pixel and the mean intensity error per pixel (for an 8 bit gray scale image), respectively. Err_S_x and Err_S_y in Fig. 4a imply the x-directional and the y-directional mean displacement error for shape, respectively. Err_T and Err_I in Fig. 4b imply the mean intensity errors for texture and image, respectively.

The case of PCA projection naturally shows the minimum errors in the graphs because they contain the errors occurred by only PCA projection. The reconstruction errors in a damaged face by Gaussian noise do not increase much due to the stability of SIFT descriptor and the robustness of our indirect matching algorithm using the SIFT descriptors. The differences in the reconstruction errors from an original face and a damaged face by Gaussian noise are 0.03(x-direction) and -0.02(y-direction) for shape, 2.25 for texture, and 1.48 for image while the reconstructions from an original face and an occluded face by virtual object causes the differences in the error by 0.08(x-direction) and 0.05(y-direction) for shape, 3.26 for texture and 3.60 for image. This reveals that the errors of occluded faces by a virtual object tend to be relatively higher than the errors of damaged faces by Gaussian noise. The possible reasons for this result are that the principal facial components are occluded by a virtual object and that keypoints on occluded region can not be extracted. The damaged region can be just statistically estimated by the data completion algorithm. Nevertheless, we can verify that the occluded face regions are plausibly reconstructed by the proposed method(Fig. 3).

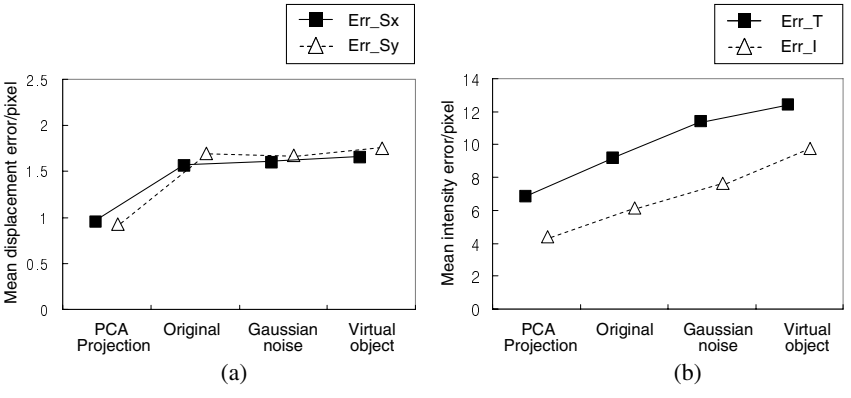

Fig. 4. Mean reconstruction errors for shape, texture and synthesized images

4 Conclusion

This paper present an automatic algorithm to align the input partially damaged image with the reference face by extending the number of matched feature points between them. It has been a challenge for many researchers to obtain the correspondences between two faces when one face of them is damaged by an objects or sensor noise. The proposed matching technique provides the enough number of the matched feature points and the accurate correspondence between an input face and the defined reference face by matching the input face with multiple example faces synthesized by the morphable face model. From the matched feature points, full shape and texture for the input face are inferred by the non-iterative data completion algorithm. The proposed methods were tested and evaluated in three types of test sets: original faces, damaged faces by Gaussian noise and occluded faces by a virtual object. The experimental results showed that the reconstructed faces are plausible and similar to original faces.

Our approach will be tested for facial images including real objects such as sunglasses, a gauze mask and hands, and camera sensor noise. In addition, face reconstruction in various pose, facial expression and illumination condition is also considered as future works. We expect that the proposed method be applied to various practical applications such as face recognition and authentication system.

Acknowledgments. This work was supported by the Korea Research Foundation Grant(KRF-2005-000-10384-0). We would like to thank Prof. Takeo Kanade for his helpful discussions and advices. In addition, we also thank the Max-Planck-Institute for providing the MPI Face Database.

References

1. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the Use of SIFT Features for Face Authentication. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops, June 2006, p. 35 (2006)
2. Blanz, V., Vetter, T.: Morphable Model for the Synthesis of 3D Faces. In: Proc. of SIGGRAPH '99, Los Angeles, USA, August 1999, pp. 187–194 (1999)
3. Blanz, V., Mehl, A., Vetter, T., Seidel, H.-P.: A Statistical Method for Robust 3D Surface Reconstruction from Sparse Data. In: Proc. of International Symposium on 3D Data Processing, Visualization and Transmission, Thessaloniki, Greece, September 2004, pp. 293–300 (2004)
4. Everson, R., Sirovich, L.: The Karhunen-Loeve Transform for Incomplete Data. Journal of the Optical Society of America A. 12(8), 1657–1664 (1995)
5. Hwang, B.-W., Blanz, V., Vetter, T., Song, H.-H., Lee, S.-W.: Face Reconstruction from a Small Number of Feature Points. In: Proc. of International Conference on Pattern Recognition, Barcelona, Spain, September 2000, vol. 2, pp. 842–845 (2000)
6. Hwang, B.-W., Lee, S.-W.: Reconstruction of Partially Damaged Face Images Based on a Morphable Face Model. IEEE Transaction on Pattern Analysis and Machine Intelligence 25(3), 365–372 (2003)
7. Jones, M.J., Poggio, T.: Multidimensional Morphable Models: A Framework for Representing and Matching Object Classes. International Journal of Computer Vision 29(2), 107–131 (1998)
8. Lowe, D.G.: Object Recognition from Local Scale-Invariant Features. In: Proc. of International Conference on Computer Vision, Corfu, Greece, September 1999, pp. 1150–1157 (1999)
9. Lowe, D.G.: Local Feature View Clustering for 3D Object Recognition. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, December 2001, pp. 682–688 (2001)
10. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
11. Mikolajczyk, K., Schmid, C.: A Performance Evaluation of Local Descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
12. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
13. Vetter, T., Troje, N.E.: Separation of Texture and Shape in Images of Faces for Image Coding and Synthesis. Journal of the Optical Society of America A 14(9), 2152–2161 (1997)