Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

3D Face Model Estimation based on Similarity Transform using Differential Evolution Optimization

K. Punnam Chandar\textsuperscript{a,}\textsuperscript{*} and T. Satya Savithri\textsuperscript{b}

\textsuperscript{a}Kakatiya University, Warangal, India
\textsuperscript{b}Jawaharlal Nehru Technological University, Hyderabad, India

Abstract

3D Face model reconstruction from a single 2D image is fundamentally important for face recognition because the 3D model is invariant to changes of viewpoint, illumination, background clutter and occlusions. In this paper, an efficient 3D Face Reconstruction algorithm is proposed based on multi-view 2D images of human face based on similarity transform measurements. In this algorithm, the pose and depth estimation from 2D feature points of the respective face images is considered as an optimization problem and solved using Differential Evolution optimization. Further different strategies of differential evolution are used to optimize and the results are compared. Simple model integration method is proposed to improve the estimation accuracy of the 3D structure of face, when more than one non-frontal-view face images are available. Furthermore, Pearson Linear Correlation is computed to show the efficiency of the proposed approach. Experimental results on 2D Head Pose and 3D Bosphorus databases demonstrate the feasibility and efficiency of the proposed methods.

1. Introduction

Many practical 2D-face recognition systems have been developed with more than three decades of research. However, the algorithms used in these systems are based on 2D-images and assume that input face images are Clear, Frontal, with good illumination, without expressions and have large face region with sufficient details. Under these controlled conditions the 2D-face recognition systems achieve reasonable performance level\textsuperscript{1}. With the increased security requirements the installation of surveillance cameras increased, ranging from small-scale standalone applications in Financial Institutions and Supermarkets to large-scale multiple networked Closed-Circuit Televisions (CCTVs) in law enforcement for public streets. In these applications, wide angle cameras are normally used and installed in a way that the viewing area is maximized. In turn controlled conditions are hard to meet. In real time with the query image having arbitrary pose with very small face region when the person is far away from the camera the performance of 2D face recognition systems suffers dramatically.

One sensible way to improve the Face recognition performance under arbitrary pose is to use multiple training images under different poses. However, face images under different poses may not be available and use of multiple

\textsuperscript{*}Corresponding author. Tel.: +91-970-440-0064. Fax: +91-874-425-7125.
\textit{E-mail address:} k..punnam@yahoo.co.in
faces will greatly increase both the size of a database and the computation time required for recognition. A potential method to improve the recognition performance in low quality input video data is to employ super resolution algorithms prior to face recognition. For these methods to give satisfactory performance, the low quality images should comply data constraints like visual quality and face similarity and will be computationally burden. Therefore 3D face models have been developed to overcome the deficiencies of classical 2D based algorithms, and finds useful in offline applications of face recognition, face tracking and face animation, etc.

Presently, there are two main stream approaches usually adopted to create human facial 3D models. One of the approach is to use specialized 3D capture systems like 3D scanners to capture texture in addition to depth i.e., 3D shape. The high cost and speed limitations of present 3D scanning devices are the obvious short comings to acquiring sufficient and useful data. The second approach is to develop algorithms to reconstruct 3D face models from 2D images, such as video sequences or multi-view photographs. This approach, call 3D reconstruction algorithm is an important tool that can be used in surveillance and in various multimedia applications. During the past decade many 3D reconstruction algorithms have been developed and can be classified in to four groups, shape-from-shading (SFS), the 3D Morphable Model (3DMM), structure from motion (SFM) and Learning.

By exploring the shading information in an image, e.g., the intensity and its derivatives, SFS deals with the recovery of shape based on some reflectance models, such as the Lambertian model, specular reflectance model, and hybrid model, etc. For the SFS algorithm, although various constraints, including brightness, smoothness, and integrability, are explored in sequence, obtaining a unique correct solution is still an intractable problem. A 3D morphable model is generally built from a set of 3D point cloud of human head. As a crucial step, the 3D point cloud scans are first registered in a dense point-by-point correspondence, using an optical-flow algorithm to reduce artifacts. In these methods, statistical signal-processing techniques, such as principal component analysis (PCA), usually play an important role in obtaining feature subspace from shape and texture features of training samples. The feature subspace can be regarded as a generic 3D face model.

Learning based algorithms exploit the common information shared by the 2D image subspaces and 3-D shape to recover the 3-D shape. Thus the algorithms in this category require a coupled training set comprising of 2D and corresponding 3D faces. However, the reconstruction performance is affected by the illumination variation as these algorithms assume that the 2D and 3D faces are embedded in the corresponding linear subspaces.

Structure from motion (SFM) is a popular approach to recover the 3D shape of an object when multiple frames of an image sequence are available. In SFM, the 3D information about a collection of discrete structures, such as lines, curves and points, is recovered from a 2D collection of lines, curves and points. 2D images are formed by projections from the 3D world. As from the literature two well-known projection models are available, the perspective and the orthographic. Much research has been conducted on determining the motion and structure of moving objects under orthographic projection. Ullman proved that four point correspondences over three views yield a unique solution to motion and structure. Tomasi proved the rank-3 theorem, i.e., the rank of the observation matrix is 3 under an orthographic projection, and proposed a robust factorization algorithm to factor the observation matrix into a shape matrix and a camera motion matrix using the singular value decomposition (SVD) technique. Xioufakis and Delopoulus extracted the motion and shape parameters of a rigid 3D object by computing the rotation matrices via the eigenvalues and eigenvectors of appropriate defined 2 × 2 matrices, where the eigenvalues are the expression of four motion vectors in two successive transitions. Bregler assumed a 3D object to be non-rigid, and the observed shapes are represented as a linear combination of a few basis shapes. Further, a Gaussian prior is assumed for the shape coefficients, and the optimization is solved using the expectation-maximization (EM) algorithm. Koo and Lam proposed Similarity-transform-based (SM) method to derive the 3D structure of a human face from multi-view photographs. Unfortunately, the genetic algorithm (GA) is used in SM algorithm to estimate the depth and is usually encounters a heavy computational burden. Moreover, how to design a reasonable chromosome, how to make a feasible gene operation scheme and how to adjust the parameters remain difficult problems. To reduce the computation of the SM method, the Differential Evolution (DE) optimization based methods are proposed in this paper to estimate the depth values of facial feature points. In the DE optimization, not only the pose parameters, but also the depth values of the facial feature points, are considered as the variables to be optimized.

The CANDIDE 3D face model is a parameterized face mask specifically developed for the model based coding of human faces. During the past several decades, candied has been a popular face model used in different face-related
applications, because of its simplicity and public availability among all of the existing methods. The third version of the CANDIDE model, called CANDIDE-3, is composed of 113 vertices and 168 triangular surfaces, as shown in Fig. 1a. Each vertex is represented by its 3-D coordinates \((x, y, z(\text{Depth}))\). Further, to show the efficiency of the proposed depth estimation algorithm using DE optimization, Pearsons Linear Correlation coefficient is computed. As linear correlation is a common criteria to measure similarity between the true signal and estimated signal.

The work in this paper summarized as follows: First, Differential Evolution optimization is used to estimate the depth values of some important feature points of face images by means of the similarity transform. Further, six DE optimization strategies are investigated with regard to the accuracy levels and the training time required. Second, a model integration method is developed to further improve the depth-estimation accuracy when multiple non-frontal-view face images are available. Experimental results on 2D and 3D databases demonstrate the feasibility and efficiency of the proposed methods. The remaining part of the paper is organized as follows; in section 2, reconstruction of 3D face using DE optimization is explained. Experimental results are given in section 3, and concluding remarks are presented in section 4.

2. Reconstruction of 3D Face using Differential Evolution

2.1 Differential evolution

In the last two decades, research on global optimization has been very active, and many different deterministic and stochastic algorithms for continuous optimization have been developed. Among the stochastic approaches, Evolutionary Algorithms (EAs) offer a number of advantages that make them attractive: implicit parallelism, robust and reliable performance, global search capability, no need of specific information about the problem to solve, easy implementation, good insensitivity to noise, and no requirements for differentiable or continuous objective functions. Differential Evolution (DE), first introduced by Storn and Price, has recently been one of the most successful evolutionary algorithms with very few control parameters: Crossover rate (Cr), Scale FACTOR (F), and Population Size (NP). Unlike traditional EAs, DE perturbs the current population members with the scaled differences of randomly selected and distinct individuals. This way, in the first iterations, the elements are widely scattered in the search space and have great exploration ability. During optimization, the individuals tend to concentrate in the regions of the search space with the best fitness values.

In DE, new individuals that will be part of the next generation are created by combining members of the current population. Every individual acts as parent and is associated to a donor vector. In the basic version of DE, the donor vector \(V_i\) for the \(i^{th}\) parent \((X_i)\) is generated, by combining three random and distinct population members \(X_{r1}, X_{r2}\) and \(X_{r3}\), as follows:

\[
V_i = X_{r1} + F \cdot (X_{r2}, X_{r3})
\]

where \(F\) is scale factor and also a real-valued parameter that strongly influences DEs performance and typically lies in the interval \([0.4, 1]\). Other mutation strategies have been applied to DE, experimenting with different base vectors and different numbers of vectors for perturbation. After mutation, every parent-donor pair generates an offspring \(O_i\) by means of crossover operation. The newly-generated offspring is evaluated and its fitness and its parents are compared. The better survives and will be part of the next generation.
2.2 Differential evolution based 3D face reconstruction

In proposed algorithm the \( n \) facial features which are represented by the coordinates \( (x_i, y_i)_{i=1,2,3,...,n} \) are used to estimate the corresponding depth values, i.e., the \( z \) coordinates. These feature points are manually marked to avoid any potential marking errors and is shown in Fig. 1b. \( (M_{xi}, M_{yi}, M_{zi}) \) represent the \( i^{th} \) feature point of a frontal-view 3-D face model \( M \), and \( (q_{xi}, q_{yi}) \) the \( i^{th} \) feature point of a non-frontal-view 2D face \( q \). Then the rotation matrix \( R \) for \( q \) is given as follows:

\[
R = \begin{bmatrix}
\cos \phi & \sin \phi & 0 \\
-\sin \phi & \cos \phi & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\cos \psi & 0 & -\sin \psi \\
0 & 1 & 0 \\
\sin \psi & 0 & \cos \psi
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \theta & \sin \theta \\
0 & -\sin \theta & \cos \theta
\end{bmatrix}
= \begin{bmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23} \\
r_{31} & r_{32} & r_{33}
\end{bmatrix}
\tag{2}
\]

where the pose parameters \( \phi, \psi, \) and \( \theta \) are the rotation angles around the \( x, y, \) and \( z \) axes, respectively. Then the rotation and translation process for mapping the frontal-view face image to the non-frontal-view face image can be given by similarity transform:

\[
\begin{bmatrix}
q_{x1} & q_{x2} & \cdots & q_{xn} \\
q_{y1} & q_{y2} & \cdots & q_{yn}
\end{bmatrix}
= k \cdot
\begin{bmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23}
\end{bmatrix}
\begin{bmatrix}
M_{x1} & M_{x2} & M_{x3} & \cdots & M_{xn} \\
M_{y1} & M_{y2} & M_{y3} & \cdots & M_{yn} \\
M_{z1} & M_{z2} & M_{z3} & \cdots & M_{zn}
\end{bmatrix}
+ \begin{bmatrix}
t_{x1} & t_{x2} & \cdots & t_{xn} \\
t_{y1} & t_{y2} & \cdots & t_{yn}
\end{bmatrix}
\tag{3}
\]

where \( k \) is the scale factor and \( (t_x, t_y) \) are the translations along \( x \) and \( y \) axes. In matrix form equation above can be written as follows:

\[
q = k \cdot R_{2 \times 3} \cdot M + T
\tag{4}
\]

where \( q \) is a \( 2 \times n \) matrix such that each column represents the \((x, y)\) coordinates \((q_{xi}, q_{yi})^T\) of one feature point, \( M \) is a matrix such that each column represents \((x, y, z)\) the coordinates \((M_{xi}, M_{yi}, M_{zi})^T\) of one feature point, and \( T \) is a \( 2 \times n \) matrix such that all columns are \((t_{xi}, t_{yi})^T\).

In terms of the shape-alignment approach in\(^3\), the translation term \( T \) can be eliminated if both \( q \) and \( M \) are centered at the origin, then

\[
q = k \cdot R_{2 \times 3} \cdot M
\tag{5}
\]

Denote \( A = k \cdot R_{2 \times 3} \) then Eq. 5 can be written as

\[
q = A \cdot M
\tag{6}
\]

If the estimated pose of the non-frontal-view face and corresponding depths fit the projection model then distance between the feature points, \( q \), of the 2D face image concerned and the corresponding projection points, \( M \), of the 3D face model can be given by

\[
d = \|q - A \cdot M\|_2^2
\tag{7}
\]

Denoting \( f(x) = q - A \cdot M \), where \( x = (\phi, \psi, \theta, M_{z1}, M_{z2}, \ldots, M_{zn}) \) as the parameter vector, including the pose parameters and the depth values of the feature points. The required parameters in the similarity measurement \( \phi, \psi, \theta \), \( k \) and the depth values \( M_{zi} \) can be obtained by minimizing the distance \( d \) in Eq. 7 using DE Optimization.

Therefore, the DE optimization is employed to search for the optimal solution i.e., optimal pose and estimate the accurate depth of important features. DE is used as it is much simpler and straight forward to implement compared to genetic algorithm used in\(^3\) and most other Evolutionary algorithms. Despite simplicity, DE exhibits much better performance and finds the approximate solution in reasonable amount of computational time. On the other hand the numbers of control parameters [Crossover rate (Cr), Scale Factor (F), and population size (NP)] are very few when compared to Genetic algorithm.

2.3 Proposed DE optimization

The DE optimization can be carried in two steps, first the initial depth values are substituted with the candied depth values \( z_c \) and the pose parameters are estimated by

\[
\min_{M_{z1}, M_{z2}, \ldots, M_{zn}} \| f(\phi, \psi, \theta) \|_2^2
\tag{8}
\]
Secondly, given the estimated pose parameters, the depth values can be estimated by

$$\min_{\phi, \psi, \theta} \| f(z_1, z_2, \ldots, z_n) \|^2_2$$

(9)

The above method is termed as two step DE optimization and coined in this work as DE2. The iteration of DE2 is repeated until a predefined maximum number of iterations is reached i.e., iter < niter, and the other way to stop the iteration is when the difference between two successive iterations is small enough i.e., given a small positive real number $\epsilon$, if the condition $d_{k+1} - d_k < \epsilon$ is satisfied.

2.4 Model integration for multiple non-frontal-view face images

Note that one non-frontal view face image is sufficient to estimate the depth $M_{z_i}, i = 1, 2, \ldots, n$ values using DE optimization. And for cases when multiple non-frontal-view face images are available, a model integration method is proposed in this paper to further improve the depth-estimation accuracy (denoted as DE2-MI). Figure 2 shows the flow chart of the model-integration method. For each non-frontal-view image $q_i$ on set of depth values $z_i$ are computed with the initial candidate depths $z_c$ using the proposed DE2 optimization. The mean of the $z_i$ in Eq. 10 is substituted with $z_c$. The iteration is repeated a predefined no. of iterations.

$$\bar{z} = \frac{1}{N} \sum_{i=1}^{N} z_i$$

(10)

![Flow chart of the model integration method.](image-url)
Table 1. Differential optimization strategies & parameters.

| DE optimization schemes | Mutation Parameters | Parameters |
|-------------------------|---------------------|------------|
| DE2−S1                  | DE/rand/1           | Cr = 0.2; F = 0.4 |
| DE2−S2                  | DE/Local-to-best/1  | Population Size |
| DE2−S3                  | DE/best/1 with jitter | Step 1: 30 |
| DE2−S4                  | DE/rand/1 with per-vector-dither | Step 2: 220 |
| DE2−S5                  | DE/rand/1 with per-generation-dither | Crossover: Uniform |
| DE2−S6                  | DE/rand/1 either-or-algorithm | Iterations: 100 |

3. Experimental Results

We evaluated the performance of the proposed two step optimization method on two different databases. The first one is the Head Pose database\textsuperscript{25}. This is a standard 2D database for evaluating 2D pose estimation algorithms, and contains images under various poses (tilt, pan & roll). Although the true depth values of the feature points are not available, the pose estimation accuracy can be well evaluated. For this database, 15 feature points at the positions shown in Fig. 1b are used in experiments. In our experiments the facial feature points are manually located to avoid potential marking errors. The second database is the Bosphorus database\textsuperscript{26}, which is a relatively new 3D face database that includes a rich set of expressions, systematic variations of poses, and different types of occlusions. One prominent advantage of this database is that both 2D and 3D facial feature points are available. Therefore, a fair experimental comparison of this database can be performed for any algorithm without considering the marking errors. Moreover, the estimated depth values can be compared with the true depth values provided in this database. Each frontal-view face image in this database has 24 facial feature points marked and is shown in Fig. 1c. As there are no corresponding points in the CANDIDE model for the feature points 23 and 24, only the feature points 1-22 are used in experiments. All simulations were conducted using MATLAB. The parameters set for the Differential Evolution optimization scheme initially have been chosen based on the most commonly used, and then refined during the experimentation. The values chosen for control parameters and different DE strategies are shown in Table 1.

3.1 Experimental results on head pose database

Head Pose database contains 2790 monocular face images of 15 persons with variations in Pan (Horizontal variation) and Tilt (Vertical variation) angles from $-90$ to $+90$ degrees, sample images of one subject are shown in Fig. 3. A subset of the Head pose database is derived comprising of total 65 images, 15 frontal and 40 non frontal images consisting of Head pose variation limited to $[-45, 45]$.

Taking the subject shown in Fig. 4 having head pose variation in $Y$ and $Z$ direction as an example, the first series of experiments were carried out on this subject to serve as an illustration of the method described in this paper. In the experiments, the elements of the pose parameters $\phi$, $\psi$, and $\theta$ are initially set to be zeros. The scale parameter $k$ is set to be 1 in the first iteration. The scaled depth values of the CANDIDE model are used as the initial values of the model. The best estimated poses using DE2−S1 are shown in Fig. 5. These results show that the poses are satisfactorily estimated.
Fig. 4. Five sample poses of person 5 of head pose database: different pan and tilt angles.

Fig. 5. Head pose database images with adapted candide model.

Fig. 6. (a) PR−D; (b) PR−SD; (c) PR−SU; (d) PR−U (e) YR−R10; (f) Front.

Table 2. Correlation coefficients of the true and estimated depth values obtained with different training images for GA and the various DE2.

| Optimization | PR−D | PR−SD | PR−SU | PR−U | YR−R10 | Average ± Std |
|--------------|------|-------|-------|------|--------|---------------|
| GA           | 0.1962 | 0.2787  | 0.5568  | 0.7283  | 0.5128  | 0.4545 ± 0.2159 |
| DE2−S1       | 0.8939 | 0.8315  | 0.5409  | 0.5258  | 0.8113  | 0.7207 ± 0.1737 |
| DE2−S2       | 0.9002 | 0.8201  | 0.5201  | 0.4115  | 0.8779  | 0.7073 ± 0.2253 |
| DE2−S3       | 0.9049 | 0.8148  | 0.4861  | 0.4614  | 0.8931  | 0.7120 ± 0.2204 |
| DE2−S4       | 0.8962 | 0.8325  | 0.6649  | 0.5320  | 0.8884  | 0.7628 ± 0.1591 |
| DE2−S5       | 0.9024 | 0.8509  | 0.5610  | 0.4739  | 0.7698  | 0.7116 ± 0.1859 |
| DE2−S6       | 0.9104 | 0.5392  | 0.6590  | 0.6211  | 0.8804  | 0.7250 ± 0.1688 |

3.2 Experimental results on Bosphorus database

The first 30 subjects from the Bosphorus database were used in the experiments. Figure 6 shows the face images of one subject in this database. Note that images with unseen feature points cannot be selected as training images, as the corresponding depth values cannot be estimated. As a result, only five non-frontal view face images, PR−D, PR−SD, PR−SU, PR−U and YR−R10 can be used to train the model in the experiments. We can obtain one set of depth values for the facial-feature points when each non-frontal-view face image is combined with its corresponding frontal-view face image for the optimization. The correlation coefficients of the true depth values and the estimated depth values are given in Table 2. It can be seen that the different DE2 Optimization strategies have higher depth-estimation accuracy than the GA-based method. Further from the Table 2 it can be observed that when PR−SU and PR−U is used as the training sample the correlation is low for all DE2 Strategies. From this it can be mentioned that the DE2 optimization is sensitive to the samples taken for training.

Figure 7 shows the true depth values and the estimated depth values of the facial feature points obtained by DE2−MI. From Fig. 7, it can be seen that the depth values of most facial feature points are estimated correctly. It should be stressed that no prior information about the true depth is utilized in the whole optimization procedure. The achieved highest correlation coefficient by DE2−MI is 0.92, which is higher than the mean values and the best result obtained by other methods when only one nonfrontal-view face image is used. Therefore, the depth estimation accuracy level can be improved efficiently by the proposed model integration method when more non-frontal-view face images are available.
Figure 8 shows the correlation coefficients $c(M_{zb}, M_{zc})$ of the true depth values and the estimated depth values obtained by the DE2-S1-MI and DE2-S2-MI algorithm, for 30 subjects and the correlation coefficients $c(M_{zb}, M_{z})$ of the true depth values and the depth values of the candied model. We can see that the estimated depth values are closer to the true depth values for most subjects, as compared to the candied model. Therefore, proposed method can accurately estimate the depth values of face images.

3.3 Experimental comparisons of depth estimation for some similar methods

Since the proposed method belongs to the structure from motion algorithms, as a comparison, we present here the depth estimation results on the same face-image database with some rigid and nonrigid SFM algorithms, including the
Table 3. Mean and standard deviation of the correlation coefficients of depth estimation results obtained using different SFM methods.

| SFM Methods | Mean (μ) | STD (σ) |
|-------------|----------|---------|
| DE2−S1−MI  | 0.845    | 0.052   |
| DE2−S2−MI  | 0.848    | 0.050   |
| SM          | 0.454    | 0.216   |
| cICA        | 0.839    | 0.063   |
| NLS1        | 0.639    | 0.279   |
| c(Mzb, Zc)  | 0.837    | 0.055   |
| FAC−SMF     | 0.909    | 0.032   |
| EM−SMF      | 0.842    | 0.166   |

Table 4. Computational time (sec.) of DE optimization.

| Genetic algorithm DE2 all strategies |
|-------------------------------------|
| ≈50                                 |
| ≈10                                 |

popular rank-3 factorization method (denoted as FAC−SMF), the expectation-maximization algorithm (denoted as EM−SMF). Depth Estimation based on Constrained Independent Component Analysis cICA, NLS-based methods utilizing symmetrical information without regularization term (denoted as NLS2, NLS2−S, NLS1, NLS1−S). The mean and standard deviations of the correlation coefficients of the depth estimation using the different SFM methods are given in Table 3. As the candied model depth values are initially chosen for the optimization algorithm, the μ and σ of c(Mzb, Mzc) for all subjects are also given in Table 3. Further, the Computation time required for GA and DE2 is given in Table 4. The proposed depth-estimation methods provide a frame work to utilize the prior information about human faces and a generic 3D face model. Further, more prior information can be explored and utilized in the proposed model to improve the depth estimation accuracy of face images. The proposed algorithm is a SFM method that is specially designed for the depth value estimation of important face features for offline applications.

4. Conclusion

In this paper, we have proposed the Differential Evolution Optimization based algorithm to estimate the pose and reconstruct the 3D facial shape from 2D face images. The proposed algorithm requires one non-frontal-view and one frontal view face image to reconstruct the sparse 3D face model. The optimization is carried out in two steps. Further simple model integration is proposed for the cases when more than one non-frontal view image is available to improve the accuracy of the 3D model estimation. The computed Pearson correlation coefficients verify the proposed 3D face model estimation algorithm efficiency. Compared to the genetic algorithm-based SM method, and constrained independent component analysis, the proposed DE based algorithm have comparable reconstruction accuracy, while the training times required decrease significantly compared to genetic algorithm. Experiment results on Head Pose Database and 3D Bosphorus database have demonstrated the pose and depth estimation efficiency of the proposed methods. For future work, several limitations of the proposed approach deserve improvement. First, the other correlations need to be investigated to measure the similarity between estimated depth and true depth. Second, the proposed approach is sensitive to the training sample which requires further research to alleviate the sensitivity and make the algorithm robust.

References

[1] R. Chellappa, C. L. Wilson and S. Sirohey, Human and Machine Recognition of Faces: A Survey, *Proceedings of the IEEE*, vol. 83(5), pp. 705–741, (1995).
[2] A. R. Chowdhury and R. Chellappa, Statistical Error Propagation in 3D Modeling from Monocular Video, in *Conference on Computer Vision and Pattern Recognition Workshop, 2003, CVPRW’03*, IEEE, vol. 8, pp. 89–89, (2003).
[3] H.-S. Koo and K.-M. Lam, Recovering the 3D Shape and Poses of Face Images based on the Similarity Transform, *Pattern Recognition Letters*, vol. 29(6), pp. 712–723, (2008).
[4] A. Thelen, S. Frey, S. Hirsch and P. Hering, Improvements in Shape-from-Focus for Holographic Reconstructions with Regard to Focus Operators, Neighborhood-size, and Height Value Interpolation, *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 18(1), pp. 151–157, (2009).

[5] M. Castelán and E. R. Hancock, Acquiring Height Data from a Single Image of a Face using Local Shape Indicators, *Computer Vision and Image Understanding*, vol. 103(1), pp. 64–79, (2006).

[6] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang and W. Gao, Efficient 3D Reconstruction for Face Recognition, *Pattern Recognition*, vol. 38(6), pp. 787–798, (2005).

[7] S. Romdhani and T. Vetter, Efficient, Robust and Accurate Fitting of a 3D Morphable Model, In *Ninth IEEE International Conference on Computer Vision, 2003 Proceedings, IEEE*, pp. 59–66, (2003).

[8] C. Tomasi and T. Kanade, Shape and Motion from Image Streams Under Orthography: A Factorization Method, *International Journal of Computer Vision*, vol. 9(2), pp. 137–154, (1992).

[9] J. Fortuna and A. M. Martinez, Rigid Structure from Motion from a Blind Source Separation Perspective, *International Journal of Computer Vision*, vol. 88(3), pp. 404–424, (2010).

[10] M. Castelán and E. R. Hancock, A Simple Coupled Statistical Model for 3D Face Shape Recovery, In *IEEE 18th International Conference on Pattern Recognition, ICPR 2006*, vol. 1, pp. 231–234, (2006).

[11] M. Song, D. Tao, X. Huang, C. Chen and J. Bu, Three-Dimensional Face Reconstruction from a Single Image by a Coupled RBF Network, *IEEE Transactions on Image Processing*, vol. 21(5), pp. 2887–2897, (2012).

[12] R. Zhang, P.-S. Tsai, J. E. Cryer and M. Shah, Shape-from-Shading: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21(8), pp. 690–706, (1999).

[13] A. Li, S. Shan, X. Chen, X. Chai and W. Gao, Recovering 3D Facial Shape Via Coupled 2D/3D Space Learning, In *8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008, FG’08, IEEE*, pp. 1–6, (2008).

[14] S. Ullman, The Interpretation of Visual Motion, (2007).

[15] Y. Xirouhakis and A. Delopoulos, Least Squares Estimation of 3d Shape and Motion of Rigid Objects from their Orthographic Projections, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22(4), pp. 393–399, (2000).

[16] C. Bregler, A. Hertzmann and H. Biermann, Recovering Non-Rigid 3D Shape from Image Streams, In *IEEE Conference on Computer Vision and Pattern Recognition, 2000 Proceedings*, vol. 2, pp. 690–696, (2000).

[17] L. Torresani, A. Hertzmann and C. Bregler, Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30(5), pp. 878–892, (2008).

[18] J. Ahlberg, An Active Model for Facial Feature Tracking, *EURASIP Journal on Applied Signal Processing*, 2002(1), pp. 566–571, (2002).

[19] Z. Sun, An Extension of Misep for PostNonlinearLinear Mixture Separation, *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 56(8), pp. 654–658, (2009).

[20] N. Hansen and A. Ostermeier, Completely derandomized Self-Adaptation in Evolution Strategies, *Evolutionary Computation*, vol. 9, pp. 159–195, (2001).

[21] K. Deb, A. Anand and D. Joshi, A Computationally Efficient Evolutionary Algorithm for Real-Parameter Optimization, *Evol. Comput.*, vol. 10(4), pp. 371–395, (2002). doi:10.1162/106365602760972767. URL http://dx.doi.org/10.1162/106365602760972767

[22] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Springer Verlag, (2003).

[23] R. Storn and K. Price, Differential Evolution &ndash; A Simple and Efficient Heuristic for Global Optimization Over Continuous Spaces, *J. of Global Optimization*, vol. 11(4), pp. 341–359, (1997). doi:10.1023/A:100820821328. URL http://dx.doi.org/10.1023/A:100820821328

[24] F. Neri and V. Tirronen, Recent Advances in Differential Evolution: A survey and Experimental Analysis, *Artificial Intelligence Review*, vol. 33(1–2), pp. 61–106, (2010). doi:10.1007/s10462-009-9137-2. URL http://dx.doi.org/10.1007/s10462-009-9137-2

[25] N. Gourier, D. Hall and J. L. Crowley, Estimating Face Orientation from Robust Detection of Salient Facial Structures, In *FG Net Workshop on Visual Observation of Deictic Gestures*, (2004).

[26] A. Savran, N. Alyuz, H. Dibeklioğlu, O. Çelikkutan, B. Gökbër, B. Sankur and L. Akarun, Bosphorus Database for 3d Face Analysis, *Biosystems and Identity Management*, pp. 47–56, (2008).

[27] S. Zhan-Li and L. Kin-Man, Depth Estimation of Face Images based on the Constrained Ica Model, *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 6, pp. 360–370, (2011).

[28] Z.-L. Sun, K.-M. Lam and Q.-W. Gao, Depth Estimation of Face Images using the Nonlinear Least-Squares Model, *IEEE Transactions on Image Processing*, vol. 22(1), pp. 17–30, (2013).