

# Simultaneous 3D Face Pose and Person-specific Shape Estimation from a Single Image using a Holistic Approach

Fadi Dornaika  
IKERBASQUE, Basque Foundation for Science  
University of the Basque Country, San Sebastian, Spain  
fadi\_dornaika@ehu.es

Bogdan Raducanu  
Computer Vision Center  
Bellaterra, Barcelona, Spain  
bogdan@cvc.uab.es

## Abstract

*This paper presents a new approach for the simultaneous estimation of the 3D pose and specific shape of a previously unseen face from a single image. The face pose is not limited to a frontal view. We describe a holistic approach based on a deformable 3D model and a learned statistical facial texture model. Rather than obtaining a person-specific facial surface, the goal of this work is to compute person-specific 3D face shape in terms of a few control parameters that are used by many applications. The proposed holistic approach estimates the 3D pose parameters as well as the face shape control parameters by registering the warped texture to a statistical face texture, which is carried out by a stochastic and genetic optimizer.*

*The proposed approach has several features that make it very attractive: (i) it uses a single grey-scale image, (ii) it is person-independent, (iii) it is featureless (no facial feature extraction is required), and (iv) its learning stage is easy. The proposed approach lends itself nicely to 3D face tracking and face gesture recognition in monocular videos. We describe extensive experiments that show the feasibility and robustness of the proposed approach.*

**Keywords:** simultaneous 3D face shape and pose, 3D model fitting, holistic approaches, face subspace, evolutionary algorithms

## 1. Introduction

Offline or online computed 3D face shapes can be used in many applications such as face recognition [3, 6], 3D face pose tracking [10], and facial expression recognition [12]. Model-based applications exploiting monocular vision systems (the face model is given by a 3D mesh or a range model) need to personalize the face model of the person utilizing the system in order to achieve an accurate estimation. This holds true even with simple 3D models such as cylinders and ellipsoids. Recently many authors used spe-

cial sensors such as a travelling camera or a 3-D scanner in order to build personalized facial shape [2]. These shape models are then used for art production or for 3D face detection and recognition using 3D sensors. Such systems suffer from several shortcomings. Some of the shortcomings can be alleviated by using stereo vision sensors [5]. In [7], the authors propose to infer side-view shape parameters from one single frontal image using learned statistical correlation between the frontal-view parameters and the side-view parameters. The facial points (MPEG-4 points) and the frontal view parameters (relative distances) are extracted from the frontal image using some heuristics and prior knowledge.

The mainstream for face pose and shape estimation relies on extracting and matching some salient facial features such as the locations and local statistics of the eyes, nose, and mouth in one or more views. A taxonomy of head pose estimation approaches can be found in [10]. Feature-based approaches suffer from self-occlusions and drifting. A solution to overcome the drawbacks of feature-based approaches is given by holistic approaches (appearance-based approaches), which try to analyze the whole facial appearance [4, 9]. For example, Active Appearance Models (AAMs) were mainly used for 2D model fitting and tracking.

Given the huge amount of work on face modeling and tracking, we state the problem addressed in this paper as follows. We address the simultaneous estimation of two types of parameters (3D head pose and person specific shape parameters that are constant for a given subject) from just one single image using only a statistical facial texture model and a standard deformable 3D model. The face is not necessarily frontal. The proposed holistic approach estimates the 3D pose parameters as well as the face shape control parameters by registering the input texture (warped region of the image) to a statistical face texture. Compared to AAMs methods our proposal has two advantages. First, there is no need to compute a Jacobian matrix neither offline nor online. Second, while AAMs merge both the inter and intra-person shape variabilities, our method separates these

variabilities, and therefore the proposed method can be easily and efficiently used for initializing a real time 3D face tracker and facial expression recognizer in videos (both the 3D deformable model and its 3D pose are computed for the first frame in the video sequence). However, it is not clear how these tasks can be performed with AAMs.

We stress the fact that our approach does not use neither 2D AAM nor 3D AAM. The only similarity with AAMs is the use of a statistical facial texture model based on Principal Component Analysis (PCA). The remainder of the paper is organized as follows. Section 2 describes the face modeling aspects. Section 3 presents the proposed holistic approach for the simultaneous estimation of the 3D pose and shape. Section 4 presents some qualitative and quantitative evaluations of performance. Section 5 concludes the paper.

## 2. Modeling faces

### 2.1. A deformable 3D mesh

In our study, we use the *Candide* 3D face model [1]. This 3D deformable wireframe model accounts for person specific shape variation as well as for facial animation. The 3D shape of this wireframe model (triangular mesh) is directly recorded in coordinate form. It is given by the coordinates of the 3D vertices  $\mathbf{P}_i, i = 1, \dots, n$  where  $n$  is the number of vertices. Thus, the shape up to a global scale can be fully described by the  $3n$ -vector  $\mathbf{g}$ ; the concatenation of the 3D coordinates of all vertices  $\mathbf{P}_i$ . The vector  $\mathbf{g}$  is written as:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{S} \boldsymbol{\tau}_s + \mathbf{A} \boldsymbol{\tau}_a \quad (1)$$

where  $\bar{\mathbf{g}}$  is the standard shape of the model,  $\boldsymbol{\tau}_s$  and  $\boldsymbol{\tau}_a$  are shape and animation control vectors, respectively, and the columns of  $\mathbf{S}$  and  $\mathbf{A}$  are the Shape and Animation Units. A Shape Unit provides a means of deforming the 3D wireframe so as to be able to adapt eye width, head width, eye separation distance, etc (see Figure 1). Thus, the term  $\mathbf{S} \boldsymbol{\tau}_s$  accounts for shape variability (inter-person variability) while the term  $\mathbf{A} \boldsymbol{\tau}_a$  accounts for the facial animation (intra-person variability). The shape and animation variabilities can be approximated well enough for practical purposes by this linear relation. Also, we assume that the two kinds of variability are independent. With this model, the ideal neutral face configuration is represented by  $\boldsymbol{\tau}_a = \mathbf{0}$ . In this study, we assume that the images are depicting quasi-neutral faces. Using this assumption it is then possible to omit the third term in the left side of Eq.(1). Thus, the expression for the deformable mesh becomes:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{S} \boldsymbol{\tau}_s \quad (2)$$

The shape modes were created manually to accommodate the subjectively most important changes in facial shape. Even though a PCA was initially performed on manually adapted models in order to compute the shape modes,

we preferred to consider the *Candide* model with manually created shape modes with semantic signification that are easy to use by human operators who need to adapt the 3D mesh to facial images.

In the model package, the number of modes associated with facial Shape Units matrix  $\mathbf{S}$  (inter-person variability) is twelve. However, for the purpose of our study which deals with the automatic image-based extraction of the control vector  $\boldsymbol{\tau}_s$  only six components are considered as the most significant indicators of the perceived person-dependent facial shape in a given near frontal facial image. These components are: Head height, vertical position of the eye brows, vertical position of the eye, eyes separation distance, vertical position of the nose, vertical position of the mouth. The remaining components are set to nominal values.

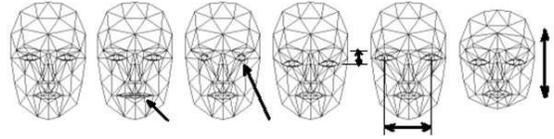


Figure 1. Effects of some facial shape control parameters on the deformable 3D model (neutral shape, mouth width, eyes width, eyes vertical position, eye separation distance, head height).



Figure 2. (a) 3D mesh with correct control shape parameters. (b) 3D mesh with another set of control shape parameters.

Figure 2 illustrates the importance of having a correct person-specific shape parameters—the control vector  $\boldsymbol{\tau}_s$ . This figure illustrates the 3D mesh overlaid on the face subject using the same 3D face pose parameters in two cases. In (a), the 3D mesh is obtained from the control vector  $\boldsymbol{\tau}_s$  associated with the displayed subject. In (b), the 3D mesh is obtained with the control vector  $\boldsymbol{\tau}_s$  associated with another subject. As can be seen, facial expression tracking and recognition based on the deformable 3D mesh become unfeasible if arbitrary control shape parameters are used.

In equation (1), the 3D shape is expressed in a local coordinate system. However, one should relate the 3D coordinates to the image coordinate system. To this end, we adopt the weak perspective projection model. We neglect the perspective effects since the depth variation of the face can be considered as small compared to its absolute depth. There-

fore, the mapping between the 3D face model and the image is given by a  $2 \times 4$  matrix,  $\mathbf{M}$ , encapsulating both the 3D face pose and the camera parameters. Thus, the state of the 3D wireframe model is given by the 3D face pose parameters (three rotations and three translations) and the shape control vector  $\boldsymbol{\tau}_s$ . This is given by the 12-dimensional vector  $\mathbf{b}$ :

$$\begin{aligned} \mathbf{b} &= [\theta_x, \theta_y, \theta_z, t_x, t_y, t_z, \boldsymbol{\tau}_s^T]^T \quad (3) \\ &= [\mathbf{h}^T, \boldsymbol{\tau}_s^T]^T \quad (4) \end{aligned}$$

where the vector  $\mathbf{h}$  represents the six degrees of freedom associated with the 3D face pose. Note that if only the aspect ratio of the camera is known, then the component  $t_z$  is replaced by a scale factor  $s \propto \frac{1}{t_z}$  having the same mapping role between 3D and 2D.

## 2.2. Shape-free facial patches

A facial patch is represented as a shape-free image (geometrically normalized rawbrightness image). The geometry of this image is obtained by projecting the standard shape  $\bar{\mathbf{g}}$  using a centered frontal 3D pose onto an image with a given resolution. The texture of this geometrically normalized image is obtained by texture mapping from the triangular 2D mesh in the input image (see figure 3) using a piece-wise affine transform,  $\mathcal{W}$ . The warping process applied to an input image  $\mathbf{y}$  is denoted by:

$$\mathbf{x}(\mathbf{b}) = \mathcal{W}(\mathbf{y}, \mathbf{b}) \quad (5)$$

where  $\mathbf{x}$  denotes the shape-free patch and  $\mathbf{b}$  denotes the geometrical parameters. Several resolution levels can be chosen for the shape-free patches. Regarding photometric transformations, a zero-mean unit-variance normalization is used to partially compensate for contrast variations.

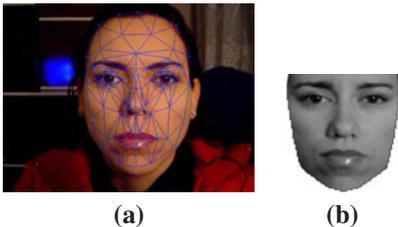


Figure 3. (a) an input image with correct adaptation. (b) the corresponding shape-free facial patch.

## 2.3. Statistical facial texture

Our statistical facial texture should describe the appearance variation of the shape-free facial patches  $\mathbf{x}$  (see figure 3.(b)). These patches are obtained from the training images (individual snapshots or video sequences) by fitting the 3D deformable model to the face. This fitting can be

manual or automatic [1]. Using these training patches one can easily build a statistical facial texture model. For this purpose we use the Principal Component Analysis (PCA)—a well-known technique used for modeling face subspaces. We assume that we have  $K$  shape-free patches. Applying a PCA on the training patches we can compute the mean and the principal modes of variation. Thus, the parameters of the facial texture model will be given by the average texture  $\bar{\mathbf{x}}$  and the principal texture modes encoded by the  $d \times K$  matrix  $\mathbf{T}$ . The columns of  $\mathbf{T}$  represent the principal modes ( $d$  is the size of the shape-free facial patch).

If the model instance,  $\mathbf{b}$ , is a good fit to the input image (i.e., the 3D mesh is aligned with the actual 3D face pose and shape), then the residual error between the shape-free patch  $\mathbf{x}$  and its projection onto the PCA space  $\hat{\mathbf{x}}$  is small since the remapped texture will be consistent with the statistical model of a face texture. Thus, a reliable measure of the goodness of any fit,  $\mathbf{b}$ , can be given by the norm of the associated residual image between the shape-free patch and its PCA approximation:

$$e(\mathbf{b}) = \|\mathbf{r}\|^2 = \|\mathbf{x}(\mathbf{b}) - \hat{\mathbf{x}}(\mathbf{b})\|^2 \quad (6)$$

The above error is called Distance From Feature Space (DFFS). The projection of the texture  $\mathbf{x}(\mathbf{b})$  onto the space spanned by the texture modes is given by:

$$\hat{\mathbf{x}}(\mathbf{b}) = \bar{\mathbf{x}} + \mathbf{T}\mathbf{T}^T(\mathbf{x}(\mathbf{b}) - \bar{\mathbf{x}}) \quad (7)$$

Thus, the basic idea is to estimate the 3D face pose and shape parameters, i.e. the vector  $\mathbf{b}$ , such that the associated shape-free patch will be as close as possible to the facial sub-space.

## 3. 3D face pose and person-specific shape parameters from a single image

The unknown 3D face pose and shape parameters encapsulated into the vector  $\mathbf{b}$  can be estimated by seeking the minimum of the cost function (6):

$$\mathbf{b} = \underset{\mathbf{b}}{\operatorname{arg\,min}} e(\mathbf{b}) \quad (8)$$

To this end, we use the Differential Evolution (DE) algorithm [11] in order to minimize the error (6) with respect to the 3D face pose and shape parameters. The DE algorithm is a practical approach to global numerical optimization that is easy to implement, reliable and fast. The crucial idea behind DE is a scheme for generating trial parameter vectors. Basically, DE adds the weighted difference between two population vectors to a third vector. In a population of potential solutions within a  $D$  dimensional search space, a fixed number of vectors are randomly initialized,

then evolved over time to explore the search space and to locate the extremum of the objective function.

In our case, the initial population is randomly selected between the lower and upper bounds defined for each variable using uniform distributions. In other words, the population of the first generation is randomly chosen around a rough solution  $\mathbf{b}^*$ . Thus, the first population is centered on a solution formed by  $\mathbf{b}^* = (0, 0, 0, t_x^*, t_y^*, s^*, \boldsymbol{\tau}_s^*)^T$ . By definition of *Candide* model, each component of  $\boldsymbol{\tau}_s$  is normalized and belongs to the interval  $[-1, 1]$ , thus the value of all initial individuals are drawn from the uniform distribution  $U(-1, 1)$ , i.e.,  $\boldsymbol{\tau}_s^* = \mathbf{0}$ .

The rough 2D translation  $(t_x^*, t_y^*)$  is set to the center of the rectangle found by Viola & Jones face detector [13]. The scale  $s^*$  is directly related to the size of the detected rectangle. A learned Look Up Table is used for this purpose.

The optimization adopted by the DE algorithm is based on a population of  $N$  solution candidates  $\mathbf{b}_{n,i}$  ( $n = 1, \dots, N$ ) at iteration (generation)  $i$ . Initially, the solution candidates are randomly generated around the solution  $\mathbf{b}^*$  within the provided intervals of the search space. The population then improves by generating new solutions iteratively for each candidate [11].

## 4. Experimental results

Experiments were conducted to evaluate the performance of the proposed fitting algorithm in image snapshots extracted from several video sequences recorded under realistic conditions. The videos depict rich person activity, covering as many head pose angles as possible. The distance of the faces from the recording camera ranged from 50 cm to one meter. The yaw and pitch angles were belonging to the interval  $[-40^\circ, +40^\circ]$ .

In this section, we report qualitative and quantitative evaluation of the proposed algorithm. For the purpose of ground-truth data associated with the 3D head pose we used home-made tools in order to compute the ground truth 3D head pose by registering dense 3D facial surfaces obtained by a stereo camera.

### 4.1. Qualitative evaluation

We have found that PCA models with 20 principal components are usually enough for representing the face space. More precisely, we found that the retained variance is above 95% of the total variance. We built a PCA model by using the shape-free templates associated with a training set of 500 images.

Figure 4 illustrates the application of the proposed scheme on a single image of an unseen person. Figure 4.(a) shows the original image together with the 2D face detection results obtained by Viola & Jones detector. Figure 4.(b) shows the projection of the 3D mesh using the obtained 3D

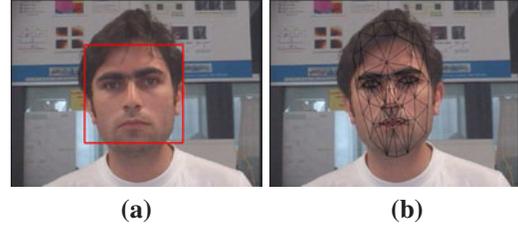


Figure 4. Full automatic estimation of 3D face pose and person-specific shape associated with an unseen person. (a) The original image and the 2D face detection results. (b) The estimated 3D face pose and shape parameters projected onto the image.

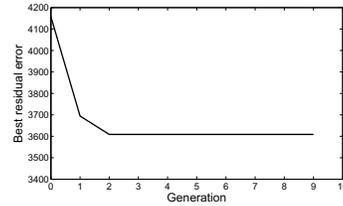


Figure 5. The evolution of the best residual error obtained by the Differential Evolution algorithm associated with the image shown in Figure 4. The population size was 500. As can be seen, the convergence was obtained in three iterations/generations.



Figure 6. Simultaneous 3D face pose and shape associated with four different snapshots.

pose and shape parameters.

Figure 5 illustrates the evolution of the best residual error obtained by the Differential Evolution algorithm associated with the image shown in Figure 4. The population size of DE algorithm was 500. As can be seen, the convergence was obtained in three iterations/generations. With a non-optimized C code, the algorithm took about 5 seconds to compute the unknown parameters.

Figure 6 illustrates the application of the proposed scheme on four snapshots.

### 4.2. Quantitative evaluation

In the previous section, the evaluation of the fitting algorithm was carried out by visual inspection. We checked

that the features of the 3-D model (deformed 3D mesh) projected onto their corresponding 2-D features in the image. In this section, we aim at a quantitative evaluation. The problem with an objective evaluation is that the absolute truth is not known. This is particularly true for the shape parameters that represent normalized intensities that deform a 3D model.

In this section, we report comparisons with a manual fitting. Then we present performance studies considering robustness to occlusions. Finally, we provide a comparison with ground-truth data obtained from a stereo-based 3D facial surfaces.

#### 4.2.1 Control shape parameters: Automatic fitting versus manual fitting

In this section we compare the automatically estimated parameters with the manually fitted ones. The manual fitting is carried out using an interactive graphical interface that displays the current image together with a 2D projection of the *Candide* model.

Table 1 depicts the average deviation between the manually obtained parameters and the automatically obtained parameters over ten different individuals. Recall that the shape parameters are normalized, i.e., each parameter belongs to the interval  $[-1, 1]$ . Thus, one can conclude that the largest deviation is associated with the vertical position of the nose 4.57%. This can be explained by the fact that the nose and its surrounding areas are somewhat featureless. Table 2 illustrates the manually and automatically fitted shape parameters for a given individual.

	eyebrow	eye	eyes separa.	nose	mouth
Ave. dev.	2.65%	3.27%	1.7%	4.57%	1.22%

Table 1. Average deviation (in %) over ten different individuals.

	eyebrow	eye	eyes separa.	nose	mouth
Manual	0.550	0.680	0.037	0.580	0.160
Automatic	0.493	0.687	0.032	0.635	0.139

Table 2. Manual fitting versus automatic fitting associated with a given subject. The first row illustrates the shape parameters obtained manually by deforming the 3D mesh using a graphical user interface. The second row depicts the same parameters estimated by the proposed algorithm.

#### 4.2.2 Evaluation in the presence of occlusions

Figure 7 shows the fitting results on a partially occluded face. The occlusions affected about half of the shape-free facial image. Despite the presence of these occlusions, the estimated pose and shape parameters do not deviate significantly from their estimated values with no occlusion. Table 3 summarizes the deviations associated with the 3D face

pose parameters (Figure 7.(a)). The first row depicts the deviations when the DFFS used the Sum of Squared Distance SSD (the classical distance). The second row depicts the deviations when the DFFS used the Trimmed Least Square distance TLS (the inlier percentage was set to 60%).



Figure 7. 3D face pose and shape estimation when the face is partially occluded.

	$t_x(\text{pixels})$	$t_y(\text{pixels})$	$s$	$\theta_x$	$\theta_y$	$\theta_z$
SSD	0.08	0.15	0.03	1.35°	2.1°	1.64°
TLS	0.47	2.4	0.03	1.30°	1.1°	0.95°

Table 3. The deviation in the 3D face pose parameters when the face is partially occluded (Figure 7.(a)). The first row corresponds to the classical DFFS (Sum of Squared Difference). The last row corresponds to the Trimmed Least Square distance.

#### 4.2.3 3D face pose accuracy using stereo-based ground-truth data

We point out that the input to our proposed fitting algorithm is a single image depicting a quasi neutral face. Thus, the proposed approach can be used for initializing 3D face trackers in videos using the 3D mesh. However, the proposed approach is not intended to be used as a 3D face tracker since the person-specific shape parameters are constant for a given person in the sense that only one single image is enough for estimating them.

In this section, we evaluate the accuracy of the proposed fitting algorithm on images extracted from three video sequences whose ground-truth 3D face pose data are known. In total, about 900 images are used in the evaluation process. Figure 8 presents some snapshots taken from two different videos. In these videos, the subjects were asked to move their face such that it performs the three out-of-plane motions (pitch, yaw and depth). The aim of using our proposed approach as a 3D face tracker was two-fold: to evaluate the accuracy of the estimated 3D head pose parameters over a large number of images, and to measure the similarity of the estimated person-specific shape parameters within a given sequence. Recall that those parameters are ideally constant for a given person.

We run our proposed fitting algorithm (Section 3) on the monocular sequence in order to retrieve the 3D face pose

parameters and the 3D shape parameters associated with every frame of the sequence. Using the above stereo sequence, the relative 3D face pose was estimated using two approaches: (i) the proposed fitting algorithm (Section 3), and (ii) the joint use of the stereo-based facial surfaces and the Iterative Closest Point (ICP) algorithm. The second kind of data can be used as ground-truth 3D face poses since the 3D data associated with the face surface are accurate and since the used registration (the Iterative Closest Point) performs a fine 3D registration.

The average errors over the whole data set were ( $4.37^\circ$ ;  $4.82^\circ$ ;  $0.73^\circ$ ) on the three rotation angles, and (0.24cm; 0.26cm; 1.33cm) on the 3D position. The average over the sequences of the standard deviation of the estimated shape-parameters the vertical position of the eye brows, the vertical position of the eye, the eyes separation distance, the vertical position of the nose, and the vertical position of the mouth were respectively 0.9%, 1.0%, 1.1%, 0.8%, and 0.9%. This shows the ability of the fitting algorithm to estimate almost the same person-shape parameters under different 3D face poses.



Figure 8. Simultaneous 3D face pose and shape parameters associated with two 300-frame video sequences.

## 5. Conclusion

This paper presented a holistic method that fits a generic deformable 3D face model to a single facial image where the face is not required to be frontal. The fitted parameters are the 3D face pose parameters as well as some shape control parameters. The proposed method has several advantages that make it attractive. These advantages are summarized in Table 4. The core of our method is an holistic approach that is based on a genetic optimizer. The proposed method is useful for the tasks of 3D face pose tracking and 3D facial expression recognition.

## Acknowledgements

B. Raducanu is supported by MITYC Grant TSI-020302-2008-38 and MEC Grant CONSOLIDER-INGENIO 2010 (CSD2007-00018), Spain. The authors thank Dr. Franck Davoine from CNRS, Compiègne, France, for providing some training video sequences.

Criterion	Yes/No
Requires continuous video data?	No
Requires an initialization?	No
Causes drifting?	No
Uses stereo vision?	No
Needs a frontal view of the face?	partially yes
Needs facial feature labelling?	No
Subject dependent?	No

Table 4. The main features of the proposed holistic approach to face pose and shape estimation according to [8].

## References

- [1] J. Ahlberg. An active model for facial feature tracking. *EURASIP Journal on Applied Signal Processing*, 2002(6):566–571, June 2002. 2, 3
- [2] M. D. Breitenstein, D. Kuettel, T. Weise, L. Gool, and H. Pfister. Real-time face pose estimation from single range images. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2008. 1
- [3] A. Bronstein, M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64:5–30, 2005. 1
- [4] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–684, 2001. 1
- [5] S. Gurbuz and N. Inoue. Real-time head pose estimation using reconstructed 3D face data from stereo image pair. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007. 1
- [6] F. Haar and R. Veltkamp. 3D face model fitting for recognition. In *European Conference on Computer Vision*, 2008. 1
- [7] C. J. Kuo, R. Huang, and T. Lin. 3D facial model estimation from single front-view facial image. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(3):183–192, 2002. 1
- [8] S. Malassiotis and M. G. Strintzis. Robust real-time 3D head pose estimation from range data. *Pattern Recognition*, 38:1153–1165, 2005. 6
- [9] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004. 1
- [10] E. Murphy-Chutorian and M. Trivedi. Head pose estimation in computer vision: a survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(4):607–626, 2009. 1
- [11] K. V. Price, J. A. Lampinen, and R. M. Storn. *Differential Evolution: A Practical Approach To Global Optimization*. Springer, 2005. 3, 4
- [12] Y. Sun and L. Yin. Facial expression recognition based on 3D dynamic range model sequences. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008*, volume 5303 of *LNCS*, 2008. 1
- [13] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. 4