Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Tracking facial features with occlusions

MARKIN Evgeny, PRAKASH Edmond C.

(*School of Computer Engineering, Nanyang Technological University, Singapore*)

E-mail: {evge0001; asprakash}@ntu.edu.sg

Received Apr. 14, 2006; revision accepted Apr. 29, 2006

Abstract: Facial expression recognition consists of determining what kind of emotional content is presented in a human face. The problem presents a complex area for exploration, since it encompasses face acquisition, facial feature tracking, facial expression classification. Facial feature tracking is of the most interest. Active Appearance Model (AAM) enables accurate tracking of facial features in real-time, but lacks occlusions and self-occlusions. In this paper we propose a solution to improve the accuracy of fitting technique. The idea is to include occluded images into AAM training data. We demonstrate the results by running experiments using gradient descent algorithm for fitting the AAM. Our experiments show that using fitting algorithm with occluded training data improves the fitting quality of the algorithm.

Key words: Active Appearance Model (AAM), Facial feature tracking, Computer vision

doi:10.1631/jzus.2006.A1282 **Document code:** A **CLC number:** TP39

INTRODUCTION

As one of the most interesting applications of image analysis and processing, facial expressions tracking has received great attention in the last few years. It can be explained by some reasons, such as increasing interest to human-computer-interaction (HCI), video conferencing systems, entertaining applications, and automatic 3D face modelling. Expression tracking is the process of retrieving facial expression information (i.e. location or relationship of facial features like: mouth corners, eyebrows, eyelids, wrinkles, chin). These visible elements further can be used in other high-level application such as expression recognition, or face-to-face communications. The problem presents a complex area for exploration, since it encompasses face acquisition, facial feature tracking, facial expression classification. In fact even the face localization step consisting of detecting and locating the person's face in an image, is not completely solved yet. Thus, anything that increases the complexity of the above-stated tasks will increase the difficulty of the problem, and probably increase the number of errors involved.

Even though current techniques have yielded significant results, their success is limited by the conditions imposed by many real applications and methods used. The major difficulty lies in tackling occlusions and self-occlusions. Current results show that Active Appearance Model (AAM) gives promising results. AAM enables accurate tracking of facial features in real-time, but lacks occlusions and self-occlusions. Any object placed in front of the face or 3D pose variation caused by occlusions. Hence, an appropriate solution is required to tackle this problem.

The remainder of the paper is structured as follows. In Section 2 we present a review of advances in facial expression recognition. An overview of techniques for modelling of human face is given in Section 3. In Section 4 we present our approach to the problem and experimental results. Section 5 concludes with a summary of the presented work and discusses future work.

BACKGROUND

Generic facial expression analysis system

Facial expression recognition is a complex task of image processing. In each stage we have to deal with a number of challenges occlusions, facial expressions, lighting variations, and pose variations.

Any facial expression recognition systems can be divided into three major parts, which are independent and slightly overleaping tasks:

Face acquisition or face detection is the first step to automatically locate any face region for the input image or sequence. Face acquisition task can be applied in two ways: on each frame, or only on the first frame. Face analysis is further complicated by appearance changes like illumination conditions, and pose orientation. Therefore it is a good idea to normalize input data prior to their analysis. More detailed information on face detection can be found in (Yang *et al*., 2002).

After face location is estimated, we use this area for further analysis. It is necessary, as it reduces the area of the image to be processed. Two types of features can be extracted: geometric and appearance features.

Geometric features present the shape and locations of facial components (including mouth, eyes, brows, nose). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. The appearance features present the appearance (skin texture) changes of the face, such as wrinkles and furrows. The appearance features can be extracted from either the whole-face or specific part of the face.

To recognize facial expressions, a facial expression analysis system can use (1) geometric features only (Cohen *et al*., 2003; Pantic and Rothkrantz, 2000), (2) appearance features only (Bartlett *et al*., 2001; Ford, 2002; Lyons *et al*., 1998) or (3) hybrid features (both geometric and appearance features) (Tian *et al*., 2001; 2002; Wen and Huang, 2003). Current results show that using hybrid features can achieve better results for some expressions.

The basic idea of geometric feature-based methods is to place automatically or manually unique markers on the face region for further tracking process. In this method we specify which markers respond for which facial changes. Mainly, we try to track only particular facial changes that are vital in the classification stage. Usually, we place most markers near mouth, nose, eyes, and eyebrows.

Facial expression classification is the last step. The facial changes can be represented either by existing coding system: FACS, MPEG-4. Depending on the temporal information used, we can classify facial expression classification methods as frame-based or sequence-based.

Ideal system

The problem space for facial expression includes multiple areas. An ideal facial expression analysis system has to address all these areas and outputs accurate recognition results. In addition, the ideal facial expression analysis system must perform fully automatically and in real-time for all stages. The properties of an ideal facial expression analysis system are summarized in Table 1.

Table 1 Properties of an ideal facial expression analysis system

Robustness	1	Handle lighting changes
	2	Handle large head motion
	3	Handle occlusions
	4	Handle different image resolution
Expressions	1	Recognize all possible expressions
	2	Recognize spontaneous expressions
Automatic process	1	Automatic face acquisition
	2	Automatic facial feature extraction
	3	Automatic expression recognition
Real-time process	1	Real-time face acquisition
	2	Real-time facial feature extraction
	3	Real-time expression recognition

Recent advances

An interesting approach was proposed by Lin and Ouhyoung (2005) to track facial expressions. The idea is to track dense facial motions by manually placing fluorescent markers on a human's face. In order to simplify the distinction of markers in video clip, they used fluorescent phenomenon covering markers with fluorescent pigments. When illuminated by BLB lamps, the pigments are excited and emit fluorescence. To avoid ambiguity in the markers' tracking, they employed markers of four different colors and keep them as far as possible from those of the same color. The unique feature of their approach is the using of two mirrors in front of a digital video camcorder. These mirrors are used for tracking markers on side-view of a subject's face. Mirrors are for getting stereovision picture and 3D points' reconstruction.

Wen and Huang (2003) proposed a method to accurately track subtle expression under such problems as lighting variations, different facial albedo, and skin color. To tackle this task they proposed a ratio-image based method to extract appearance features which are independent of a person's face albedo, and lighting variations. The 3D model is fit to the first frame by manually selecting landmark facial features such as corners of the eyes and mouth. The inter-frame motion is evaluated by using template-based optical flow. The results present a robust system for tracking subtle facial changes, but it is not automatic and does not operate well with a wide range of head's motions.

Cohen *et al*.(2003) used in their system an approach proposed by Tao and Huang (1998). The process of evaluating head motions and facial changes is broken into two steps. First, the 2D image motion is tracked using template matching between frames at different resolutions. From the 2D motions of many points on the model, the 3D motion can be evaluated by solving an overdetermined system of equations of the projective motions in the least-squared sense.

In the system presented by Bartlett *et al*.(2001), 3D pose and face geometry are estimated from hand-labeled feature points by using a canonical wire-mesh face model. To estimate the head's position, 8 features on these images are labelled by hand. Then, a scattered data interpolation technique is used to modify the canonical face model to fit the 8 known 3D points and to interpolate the unknown positions of all the other vertices in the face model. In the case when landmark positions are unknown, they applied a stochastic filtering approach.

Xiao *et al*.(2002) used a cylindrical model to automatically evaluate and recover 6 degrees of freedom of the head's motions. They used an AAM (Cootes *et al.*, 1998) method to map the cylindrical head model to the face region, which is detected by face detector. The 3D head's motions are processed by using three main techniques. First, they use the iteratively reweighted least squares (IRLS) technique to deal with nonrigid motion and occlusion. Second, they updated the template dynamically in order to decrease the effects of gradual changes in lighting and self-occlusion. Third, as using template model may cause error accumulation, they solved this problem by re-registering images to a reference frame when head pose is close to a reference pose. This approach prevents the system from error accumulation and enables the system to recover head pose when the head reappears after occlusion. Because head poses are recovered using templates that are constantly updated and the pose estimated for the current frame is used in estimating the pose in the next frame, errors would accumulate unless otherwise prevented. To solve this problem, the system automatically selects and stores one or more frames and associated head poses from the tracked images in the sequence (usually including the initial frame and pose) as references.

To summarize the aforementioned, we must say that the most common lack is initialization task (Lin and Ouhyoung, 2005; Wen and Huang, 2003; Bartlett *et al*., 2001; Cohen *et al*., 2003). Nevertheless each system has its own distinctive feature such as subtle expressions (Wen and Huang, 2003), dense facial expressions tracking (Lin and Ouhyoung, 2005), automatic initialization (Xiao *et al*., 2002).

MODELLING OF HUMAN FACE

Any human face is composed of the 3D face shape, the facial appearance, and the 3D face pose with respect to the camera. The process of model construction depends on parameters of interest (e.g., accuracy of model, efficiency of fitting algorithm). Fitting the model to the image gives us the face shape, the facial appearance, and the face pose. Then the recovered information can be used in applications such as face tracking (Lee *et al*., 2003; Romdhani and Vetter, 2003; Gokturk *et al*., 2001), or facial expression recognition (Romdhani and Vetter, 2003; Baker *et al*., 2004; Moghaddam *et al*., 2003). Most methods construct the model of 3D human face using either 3D range-scanned face shapes (Brand and Bhotika, 2001; Blanz and Vetter, 1999) or synchronized multi-view face images (Gokturk *et al*., 2001; Moghaddam *et al*., 2003). Since the inner depth of a face is usually small relative to the distance from the camera, a weak-perspective can be applied.

Next we present three landmark approaches: 2D AAMs (Cootes *et al*., 1998), 3D Morphable Models (MMs) (Blanz and Vetter, 1999), and Combined 2D+3D model (Xiao *et al*., 2004).

Active appearance model

An AAM (Cootes *et al*., 1998) is a nonlinear, generative, and parametric model of a certain visual phenomenon. The shape of the 2D AAM is defined by a triangulated mesh and the vertex locations of the mesh. Mathematically, the shape $\boldsymbol{s}$ is defined as the 2D coordinates of the n vertices that make up the mesh:

$$\boldsymbol{s}=\begin{pmatrix} u_1 & u_2 & \dots & u_n \\ v_1 & v_2 & \dots & v_n \end{pmatrix}. \quad (1)$$

The 2D AAM allows linear shape variation, namely, the shape matrix $\boldsymbol{s}$ can be expressed as a base shape $\boldsymbol{s}_0$ plus a linear combination of m shape matrices $\boldsymbol{s}_i$:

$$\boldsymbol{s}=\boldsymbol{s}_0+\sum_{i=1}^{n} p_i \boldsymbol{s}_i, \quad (2)$$

where the coefficients p_i are the shape parameters.

The 2D AAM is constructed from a set of images with the shape mesh marked on it. Mostly, the shapes in images are placed in different locations and at different poses, scales, and angles. The next stage is to align shapes. After that Principal Component Analysis (PCA) is utilized on the registered shape meshes to compute the deformable shape model. The base shape $\boldsymbol{s}_0$ is the mean shape and the matrices $\boldsymbol{s}_i$ are the (reshaped) eigenvectors corresponding to the m largest eigenvalues.

The appearance of the 2D AAM is defined within the base mesh $\boldsymbol{s}_0$. Let us denote $\boldsymbol{s}_0$ to be a set of pixels $\boldsymbol{u}=(u; v)^{\mathrm{T}}$ that lie inside the base mesh $\boldsymbol{s}_0$, a convenient abuse of terminology. The appearance of the 2D AAM is then an image $A(\boldsymbol{u})$ defined over the pixels $\boldsymbol{u}\in\boldsymbol{s}_0$. 2D AAMs allow linear appearance variation. This means that the appearance $A(\boldsymbol{u})$ can be expressed as a base appearance $A_0(\boldsymbol{u})$ plus a linear combination of l appearance images $A_i(\boldsymbol{u})$:

$$A(\boldsymbol{u})=A_0(\boldsymbol{u})+\sum_{i=1}^{l}\lambda_i A_i(\boldsymbol{u}), \quad (3)$$

where the coefficients λ_i are the appearance parameters. As with the shape, the base appearance A_0 and appearance images A_i are usually computed by applying PCA to the (shape normalized) face images.

There is another approach to present the 2D AAM (Cootes *et al*., 2001). While independent AAMs have separate shape p and appearance λ parameters, the combined AAMs just use a single set of parameters $\boldsymbol{c}=(c_1, c_2, \dots, c_k)^{\mathrm{T}}$.

3D morphable models

To construct a 3D morphable model (3D MM) proposed by Blanz and Vetter (1999), a set of example 3D faces (e.g., laser scans) is required. In their work the authors used over 200 laser scans acquired by a "Cyberware 3030PS" laser scanner. Generally speaking, all the work on 3D MM was carried out by Blanz and Vetter (1999; 2003).

The construction was performed in three stages. First, the laser scans are preprocessed. This semi automatic step aims to remove the scanning artifacts and to select the part of the head that is to be modelled (from one ear to the other and from the neck to the forehead). In the second stage, the correspondences are computed between one scan chosen as the reference scan, and each of the other scans. Then a principal components analysis is performed to estimate the statistics of the 3D shape and texture of the faces.

The 3D shape of the 3D MM is defined by a 3D triangulated mesh and in particular the vertex locations of the mesh. Mathematically, the shape $\tilde{\boldsymbol{S}}$ of the 3D MM is defined as the 3D coordinates of the N vertices that make up the mesh:

$$\tilde{\boldsymbol{S}}=\begin{pmatrix} x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \\ z_1 & z_2 & \dots & z_N \end{pmatrix}. \quad (4)$$

The 3D MM allows linear shape variation, namely, the mean shape matrix $\bar{\boldsymbol{S}}$ can be expressed as a linear combination of the shapes $\tilde{\boldsymbol{S}}_i$

$$\tilde{\boldsymbol{S}}=\bar{\boldsymbol{S}}+\sum_{i=1}^{N_s}\alpha_i \boldsymbol{S}_i, \quad (5)$$

where the coefficients α_i are the shape parameters.

The 3D MM is constructed from a set of images with the shape mesh marked on it. As in the construction of 2D AAM, the shapes are placed in different locations and at different poses, scales, and angles. Extracting the model requires registration of the marked shapes.

The appearance (texture) of the 3D MM is defined within a 2D triangulated mesh that has the same topology (vertex connectivity) as the base mesh $\bar{\boldsymbol{S}}$. Let $\bar{\boldsymbol{S}}$ denote the set of pixels $\boldsymbol{u}=(u; v)^{\mathrm{T}}$ that lie inside this 2D mesh. The appearance is then an image $\bar{\boldsymbol{T}}$ defined over $\boldsymbol{u}$, where $\boldsymbol{u}\in\bar{\boldsymbol{S}}$. The appearance $\tilde{\boldsymbol{T}}$ can be expressed as a base appearance $\bar{\boldsymbol{T}}$ plus a linear combination of N_T appearance images $\tilde{\boldsymbol{T}}_i$:

$$\tilde{\boldsymbol{T}}=\begin{pmatrix} r_1 & r_2 & \cdots & r_N \\ g_1 & g_2 & \cdots & g_N \\ b_1 & b_3 & \cdots & b_N \end{pmatrix}, \tag{6}$$

$$\tilde{\boldsymbol{T}}=\bar{\boldsymbol{T}}+\sum_{i=1}^{N_T}\beta_i\boldsymbol{T}_i, \tag{7}$$

where the coefficients β_i are the appearance parameters. As with the shape, the base appearance $\bar{\boldsymbol{T}}$ and the appearance images $\tilde{\boldsymbol{T}}_i$ are usually computed by applying PCA to the texture components of the training range images, appropriately warped onto the 2D triangulated mesh $\tilde{\boldsymbol{S}}$.

Combined 2D AAM+3D MM

Xiao *et al.*(2004) proposed a new method called "Combined 2D+3D AAM". The experimental results of exploiting 3D MM showed that using 3D MM is computationally expensive. With these objectives in mind they proposed a method to replace the expensive 3D image warping with the 2D image operations as in the 2D AAM fitting. Thereby they constrained a 2D AAM with the equivalent 3D shape. The authors showed that, any 2D projection of a 3D shape in the span of the 3D MM can also be instantiated by the 2D AAM, but at the expense of using more parameters. The 2D AAM requires up to 6 times more parameters than the 3D MM to model the same phenomenon.

Summary

As a summary of the aforesaid methods, we emphasize their good and weak points.

1. 2D AAM

The 2D AAM does not explicitly model the 3D face motions and facial deformations as the 3D MM does. Thus fitting the 2D AAM to images cannot recover the 3D face shapes and poses. Since the shape model of the 2D AAM is computed from the projections of the 3D face shapes, it implicitly represents the 3D information. Hence, the 2D AAM requires a larger number of parameters to represent the same phenomena as the corresponding 3D MM. Fitting the 2D AAM requires only 2D image operations and is thus very efficient. Currently the fastest 2D AAM fitting algorithm runs at over 200 frames per second (Matthews and Baker, 2004). Fitting of the 2D AAM also lacks occlusions and self-occlussions.

2. 3D MM

Fitting the 3D MM is computationally expensive because it requires 3D warping of the face appearance. The 3D MM fitting algorithm takes approximately 30 seconds per frame (Romdhani and Vetter, 2003). The 3D MM naturally represent a face model and much denser (76000 vertices are modelled by the 3D MM in (Blanz and Vetter, 1999)).

3. 2D AAM+3D MM

Needs 6 times more parameters, but is more accurate than the 2D AAM. The fitting algorithm takes approximately 60 seconds per frame (Xiao *et al.*, 2004) but lacks occlusions, and is less detailed than the 3D MM, and consists of a few tens of vertices around eye, mouth, nose.

OUR IMPROVEMENTS

In the previous section we showed that fitting of the AAM lacks occlusions and self-occlusions. Occlusions mean that an object is fully or partially overlapped by any other object. Self-occlusions imply that some regions of the object become invisible due to its rotations or displacements. In our case we deal with the human face. Occlusions may occur in the training data used to construct the AAM, or in the input videos to which the AAM is fit. The most common cause of self-occlusion is pose variation.

Generally in any AAM all of the mesh vertices $\boldsymbol{s}$ are marked manually in every training image. In the presence of occlusion the situation is sophisticated by the fact that not all of the mesh vertices can be visible in every training image.

To solve the problem of occlusions, we introduce into training data a set of images where the human face is partially occluded (Fig.1). The appearance and the shape are computed in the same way as in the general AAM.

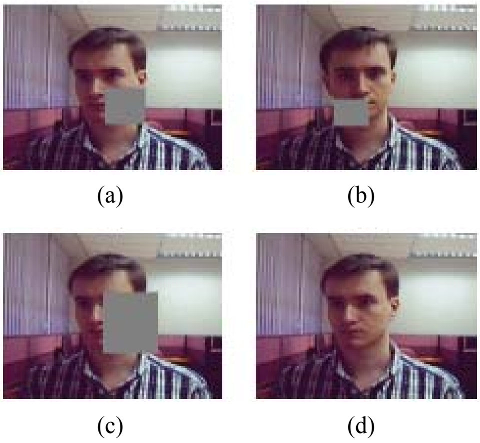

Fig.1 Examples of occluded data. (a) and (b) Images with 25% of the face region occluded; (c) Image with 50% of the face region occluded; (d) Original image

Experiments

In order to evaluate the AAM appearance with occlusion, we constructed marked image sequences of one person in which we randomly selected occluded regions. First we constructed the combined AAM for the 7 images with 50% of the face region occluded. Then for the 7 images with 25% of the face region occluded. After that, we used all occluded and non-occluded data (Fig.2).

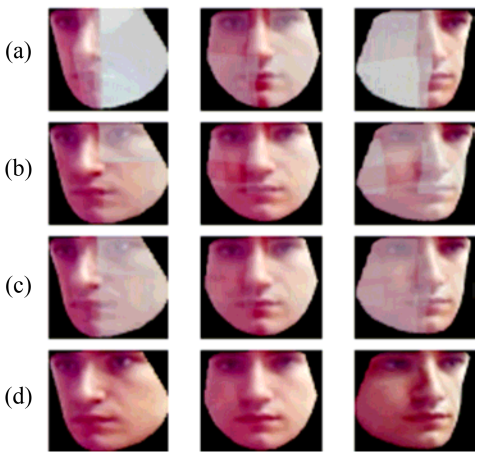

Fig.2 Mean appearance A_0 and A variations. (a) 50% occluded data; (b) 25% occluded data; (c) Appearance from the combination of occluded and original images; (d) Appearance images from non-occluded data

To evaluate visually the effect of occluded training data, we conducted a set of experiments to show the fitting results. For ground truth data we used 2 images (Fig.3). One without any occlusions, another with 25% facial region occluded. As a parameter to be minimized we used error function ε_A. Next, we employed Gradient Descent method to minimize ε_A:

$$\varepsilon_A = \left\| A - \left(A_0 + \sum_{i=1}^{l} c_i A_i \right) \right\|. \tag{8}$$

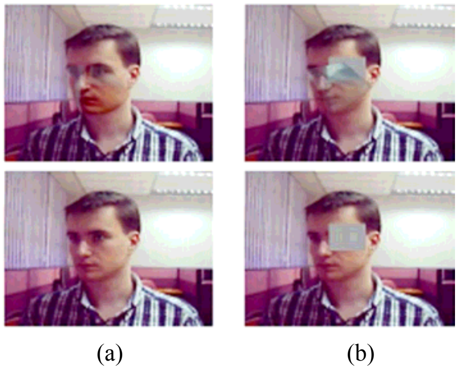

Fig.3 Fitting results. (a) Without occlusions; (b) with occlusions

The results (Fig.4) showed that using occluded data in the AAM improves the quality of the fitting algorithm. Numerically, it gives 24% precision growth.

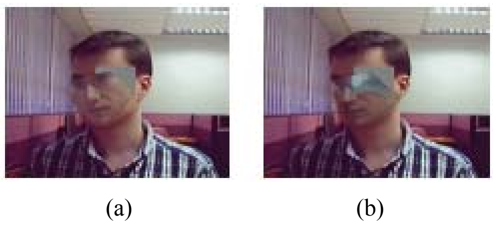

Fig.4 Fitting on occluded image without (a) and with (b) using occluded data in the AAM

CONCLUSION

In this paper we demonstrated that using occluded training data to construct the AAM improves the AAM's fitting. Even though the approach turns out to be unnatural, it refines the quality of the fitting result. Numerically, it gives 24% precision growth. In the future, this approach can be extended to deal with self-occlusions caused by 3D head pose variations.

References

Baker, S., Gross, R., Matthews, I., 2004. Lucas-Kanade 20 Years on: A Unifying Framework: Part 4. Technical Report CMURI-TR-04-14, Robotics Institute, Carnegie Mellon University.

Bartlett, M., Braathen, B., Littlewort-Ford, G., Hershey, J., Fasel, I., Marks, T., Smith, E., Sejnowski, T., Movellan, J.R., 2001. Automatic Analysis of Spontaneous Facial Behavior: A Final Project Report. Technical Report INC-MPLab-TR-2001.08, Machine Perception Lab, Institute for Neural Computation, University of California.

Blanz, V., Vetter, T., 1999. A Morphable Model for the Synthesis of 3D Faces. Proceedings of the International Conference on Computer Graphics and Interactive Techniques, p.187-194.

Blanz, V., Vetter, T., 2003. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**(9):1063-1074. [doi:10.1109/TPAMI.2003.1227983]

Brand, M., Bhotika, R., 2001. Flexible Flow for 3D Nonrigid Tracking and Shape Recovery. Proceedings of the International Conference on Computer Vision and Pattern Recognition, **1**:315-322.

Cohen, I., Sebe, N., Garg, A., Chen, L.S., Huang, T.S., 2003. Facial expression recognition from video sequences: temporal and static modelling. *Computer Vision and Image Understanding*, **91**(1-2):160-187. [doi:10.1016/S1077-3142(03)00081-X]

Cootes, T.F., Edwards, G.J., Taylor, C.J., 1998. Active Appearance Models. Proceedings of the European Conference on Computer Vision, **2**:484-498.

Cootes, T.F., Edwards, G.J., Taylor, C.J., 2001. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(6):681-685. [doi:10.1109/34.927467]

Ford, G.L., 2002. Fully Automatic Coding of Basic Expressions from Video. Technical Report INC-MPLab-TR-2002.03, Machine Perception Lab, Institute for Neural Computation, University of California.

Gokturk, S.B., Bouguet, J.Y., Grzeszczuk, R., 2001. A Data-Driven Model for Monocular Face Tracking. Proceedings of the International Conference on Computer Vision, **2**:701-708.

Lee, J., Moghaddam, B., Pfister, H., Machiraju, R., 2003. Silhouette-based 3D face shape recovery. Graphics Interface.

Lin, I.C., Ouhyoung, M., 2005. Mirror mocap: Automatic and efficient capture of dense 3D facial motion parameters from video. *The Visual Computer*, **21**(6):355-372. [doi:10.1007/s00371-005-0291-5]

Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J., 1998. Coding Facial Expressions with Gabor Wavelets. Proceedings of the International Conference on Automatic Face and Gesture Recognition, p.200-205. [doi:10.1109/AFGR.1998.670949]

Matthews, I., Baker, S., 2004. Active appearance models revisited. *International J. of Computer Vision*, **60**(2):135-164. [doi:10.1023/B:VISI.0000029666.37597.d3]

Moghaddam, B., Lee, J., Pfister, H., Machiraju, R., 2003. Model-Based 3D Face Capture with Shape-from-Silhouettes. Proceedings of the International Workshop on Analysis and Modelling of Faces and Gestures, p.20-27.

Pantic, M., Rothkrantz, L.J.M., 2000. Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, **18**(11):881-905. [doi:10.1016/S0262-8856(00)00034-2]

Romdhani, S., Vetter, T., 2003. Efficient, Robust and Accurate Fitting of a 3D Morphable Model. Proceedings of the International Conference on Computer Vision, **1**:59-66. [doi:10.1109/ICCV.2003.1238314]

Tao, H., Huang, T.S., 1998. Connected Vibrations: A Modal Analysis Approach for Non-Rigid Motion Tracking. Proceedings of the International Conference on Computer Vision and Pattern Recognition, p.735-740.

Tian, Y.I., Kanade, T., Cohn, J.F., 2001. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(2):97-115. [doi:10.1109/34.908962]

Tian, Y.I., Kanade, T., Cohn, J.F., 2002. Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity. Proceedings of the International Conference on Automatic Face and Gesture Recognition, p.218-223.

Wen, Z., Huang, T.S., 2003. Capturing Subtle Facial Motions in 3D Face Tracking. Proceedings of the International Conference on Computer Vision, p.1343-1350. [doi:10.1109/ICCV.2003.1238646]

Xiao, J., Kanade, T., Cohn, J.F., 2002. Robust Full-Motion Recovery of Head by Dynamic Templates and Re-Registration Techniques. Proceedings of the International Conference on Automatic Face and Gesture Recognition, p.156-162.

Xiao, J., Baker, S., Matthews, I., Kanade, T., 2004. Real-Time Combined 2D+3D Active Appearance Models. Proceedings of the International Conference on Computer Vision and Pattern Recognition, **2**:535-542.

Yang, M.H., Kriegman, D.J., Ahuja, N., 2002. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(1):34-58. [doi:10.1109/34.982883]