Int J Comput Vis (2008) 80: 260–274
DOI 10.1007/s11263-007-0125-1

Pose Robust Face Tracking by Combining Active Appearance Models and Cylinder Head Models

Jaewon Sung · Takeo Kanade · Daijin Kim

Received: 1 December 2006 / Accepted: 28 December 2007 / Published online: 23 January 2008

Abstract The active appearance models (AAMs) provide the detailed descriptive parameters that are useful for various autonomous face analysis problems. However, they are not suitable for robust face tracking across large pose variation for the following reasons. First, they are suitable for tracking the local movements of facial features within a limited pose variation. Second, they use gradient-based optimization techniques for model fitting and the fitting performance is thus very sensitive to initial model parameters. Third, when their fitting is failed, it is difficult to obtain appropriate model parameters to re-initialize them. To alleviate these problems, we propose to combine the active appearance models and the cylinder head models (CHMs), where the global head motion parameters obtained from the CHMs are used as the cues of the AAM parameters for a good fitting or re-initialization. The good AAM parameters for robust face tracking are computed in the following manner. First, we estimate the global motion parameters by the CHM fitting algorithm. Second, we project the previously fitted 2D shape points onto the 3D cylinder surface inversely. Third, we transform the inversely projected shape points by the estimated global motion parameters. Fourth, we project the transformed 3D points onto the input image and computed the AAM parameters from them. Finally, we treat the computed AAM parameters as the initial parameters for the fitting. Experimental results showed that face tracking combining AAMs and CHMs is more pose robust than that of AAMs in terms of 170% higher tracking rate and the 115% wider pose coverage.

Keywords Face tracking · Active appearance models · 2D+3D active appearance models

Electronic supplementary material The online version of this article (http://dx.doi.org/10.1007/s11263-007-0125-1) contains supplementary material, which is available to authorized users.

J. Sung · D. Kim (✉)
Department of Computer Science and Engineering, POSTECH, San 31, Hyoja-Dong, Nam-Gu, Pohang, 790-784, Korea
e-mail: dkim@postech.ac.kr

J. Sung
e-mail: jwsung@postech.ac.kr

T. Kanade
Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: tk@cs.cmu.edu

1 Introduction

The active appearance models (AAMs) (Cootes et al. 2001; Matthews and Baker 2004; Xiao et al. 2004a) are flexible models that can explain the varying shape and appearance of a non-rigid object. They are especially useful when we are interested in describing a specific part of a 3D object and the part is always visible because the AAMs require that the topology of the shape must be consistent. The 3D morphable models (MMs) (Blanz and Vetter 1999) are another type of flexible models that are very similar to the 2D+3D AAMs (Xiao et al. 2004a) that are extension of the traditional 2D AAMs by incorporating 3D shape models into themselves. They require a large number parameters to describe the detailed variations of the shape and appearance. We may improve the descriptive power of the AAMs by taking more and more training data, but this also increases the required number of the model parameters to treat the increased variations of training data, and makes the AAM fitting more difficult.

When the AAMs are used to model human faces, the number of model parameters usually amounts to dozens or more than a hundred. This high dimensionality of the AAMs complicates the fitting surface and increases the trapping of the AAM fitting to the local minima. The AAM fitting algorithms also suffer from its sensitivity to initial parameters, which is a common problem of the iterative gradient-descent optimization algorithms. Therefore, the face tracking using only the AAMs cannot be successful especially when the face freely moves around and changes its pose largely. When the head pose is deviated from the frontal view too much, the AAMs fail to fit the input face image correctly because most part of the face image becomes invisible. Once the AAMs fail to track the face, they cannot work until it is re-initialized. Since the re-initialization can be usually done by the face detector at the font view, the AAMs can do nothing until the face pose returns to near frontal view.

The global head motion can be estimated by model-based 3D face tracking algorithms. Some of them used simple geometric head models such as a cylinder (Cascia et al. 2000; Xiao and Kanade 2002), an ellipsoid (Basu et al. 1996), or a head-like 3D shape (Malciu and Preteux 2000). They recover the global motion by minimizing the difference of texture or optical flow between observation and their model. Vacchetti et al. (2004) used multiple key frames and feature point matching to estimate the motion of their model under large pose variation. These approaches assume that the 3D shape of the object does not change during tracking, which means that they do not have shape parameters. On the other hand, some researchers tried to track the deforming shape and global motion at the same time. Strom et al. (1999) used feature point tracking result with structure from motion in Kalman filter framework. DeCarlo and Metaxas (1996) used a deforming face model whose fitting algorithm integrated optical flow and edge information. However, these algorithms are not adequate for tracking of large head pose change and the quality of the estimated 3D shapes cannot be guaranteed.

The global head motion can be represented by a rigid motion, which can be parameterized by 6 parameters; three for 3D rotation and three for 3D translation. Therefore, the number of all the model parameters are only 6. The low dimensionality of the parameter space results in robust tracking performance when compared to the high dimensionality of the AAMs. In addition, these methods do not require any learning stage, which means that they are person independent. Moreover, they are robust to a large pose change because they use the whole area of the head in the image instead of a specific part. However, the rigid head model cannot provide detailed information such as the local movement of the facial features; opening mouth, closing eyes, raising eye brows, and so on. Among three different geometric head models, we take the cylinder head model because it is the generally applicable and simplest method. The cylinder model is more appropriate for approximating the 3D shape of the generic faces than the ellipsoid model. Also, the head-like 3D shape model requires a large number of parameters and its fitting performance is very sensitive to their initialization.

We propose a new face tracker by combining the AAMs and the CHMs together. They are tightly coupled with each other in the following manner. In the very beginning of face tracking, the AAM local motion parameters provide the initial cues of the global motion parameters of the CHMs. During the face tracking, the global motion parameters of CHMs provide the initial cues of the facial feature parameters of the AAMs. The proposed face tracker has advantages by the following. From the viewpoint of AAMs, the global motion parameters obtained from the CHMs can be used to estimate a good candidate of initial parameters for the AAM fitting, and to re-initialize the face tracking when AAMs fail to fit the current face image. This makes the face tracking robust to the change of face pose and extends the range of face tracking effectively. From the view point of CHMs, the detailed information of the local movements of the facial features obtained from the AAMs enables the CHMs to recognize the facial expressions, gazes, and several facial gestures such as nodding/disapproval by head motion, blinking eyes, and opening/closing mouth, and so on.

The proposed face tracker is different from the 3DMM-based face tracker in the following way. The 3DMM-based face tracker often fails to track the face under the large pose variation because the 3DMM covers a partial area of the head. However, the proposed CHM-based face tracker tracks well the face under the large pose variation by the following reasons. First, the CHM has a very simple geometric shape while the 3DMM has a detailed 3D mesh structure. Fidaleo et al. (2005) compared the tracking performance of different geometric models with different level of details such as ellipsoid model, generic 3D shape model, true 3D shape model. According to their experimentations, the ellipsoid model, which is the most simple model, showed the better tracking performance than the generic 3D shape model. This result comes from the fact that the detailed information about the face in the generic 3D shape model imposes a wrong prior on the tracker that is apt to make an erroneous tacking result. Therefore, the CHMs are considered as the better choice for the global head motion tracking of arbitrary person. The good tracking performance using the CHMs has been demonstrated in many literatures (Cascia et al. 2000; Xiao and Kanade 2002). Second, the CHMs covers the wider area of the head than the 3DMM. Fidaleo et al. (2005) showed that the coverage of the model affects the tracking performance because the smaller coverage can miss some important feature points.

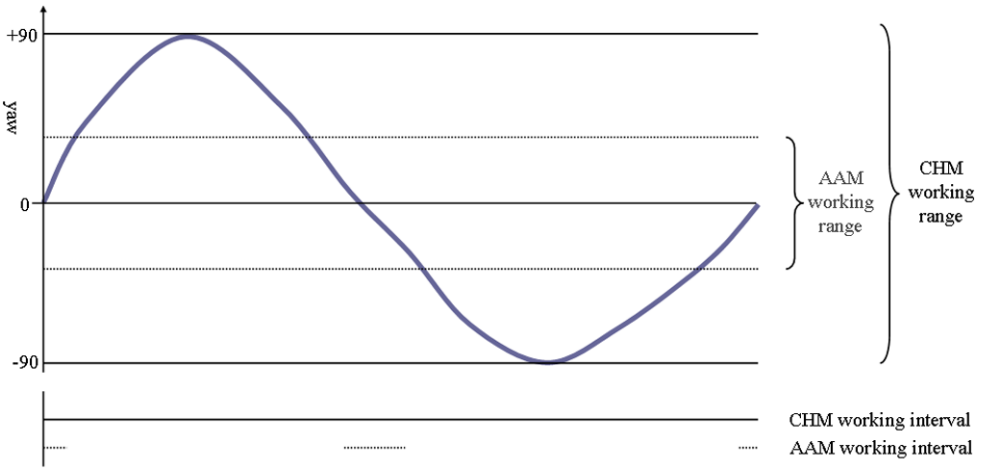


Fig. 1 Comparison of working time intervals of the AAM and the CHM

Figure 1 compares the working range of pose angles, in which an AAM and a CHM are tracking a face that is moving horizontally, where the horizontal and vertical axes represent the time and the yaw angle of the head, respectively. We assume that the CHM can follow the head motion within the range of $\pm$ 90 degree and the AAM can fit within a limited range of pose angle from the frontal view. The two lines in the bottom represent the time-interval of successful face tracking of the CHM (solid line) and the AMM (dotted line), respectively. As it is shown, the AAM can track the face within the very limited time interval while the CHM can track the face through the whole range of pose angles.

This paper is organized as follows. Section 2 describes the 2D+3D AAMs and its local motion tracking briefly.[1] Section 3 describes the CHMs and its global motion tracking.[2] Section 4 introduces a novel face tracker by combining the 2D+3D AAMs and the CHMs together. Section 5 shows the experiment results that validates the robustness of the proposed face tracking in terms of tracking rate and pose coverage. Finally, Sect. 6 draws a conclusion.

2 2D+3D AAMs and Local Motion Tracking

2.1 2D+3D AAMs

Xiao et al. (2004a) proposed the *c*ombined 2D+3D AAMs and its fitting algorithm by adding the 3D shape constraints to the 2D AAMs (Matthews and Baker 2004). In the 2D AAMs, the 2D shape is represented by a vector $\boldsymbol{s}$ that consists of 2D coordinates of the l 2D landmark points as $\boldsymbol{s} = (x^1, y^1, \ldots, x^l, y^l)^t$.[3] The appearance A is defined on the shape-normalized template to which the landmarked training images are warped. Therefore, the variations in the appearance data are due to the texture variations not the shape variations. The variations of the 2D shape and appearance are represented by two linear models

$$\boldsymbol{s} = \sum_{i=0}^{n} p_i \boldsymbol{s}_i, \qquad A = \sum_{i=0}^{m} \alpha_i A_i, \tag{1}$$

where $\boldsymbol{s}_0$, $\boldsymbol{s}_i$ $(i = 1, \ldots, n)$, and p_i $(i = 0, \ldots, n, p_0 = 1)$ are the mean 2D shape, the 2D shape bases, and the 2D shape parameters, respectively, and n is the number of 2D shape bases. Similarly, A_0, A_i $(i = 1, \ldots, m)$, and α_i $(i = 0, \ldots, m, \alpha_0 = 1)$ are the mean appearance, the appearance bases, and the appearance parameters, respectively, and m is the number of appearance bases.

In the 2D+3D AAMs, the additional 3D shape is represented by a 3D shape vector that consists of the l landmark 3D points as $\bar{\boldsymbol{s}} = (x^1, y^1, z^1, \ldots, x^l, y^l, z^l)^t$. The variation of the 3D shape is represented by the linear model

$$\bar{\boldsymbol{s}} = \sum_{i=0}^{\bar{n}} \bar{p}_i \bar{\boldsymbol{s}}_i, \tag{2}$$

where $\bar{\boldsymbol{s}}_0, \bar{\boldsymbol{s}}_i$ $(i = 1, \ldots, \bar{n})$, and $\bar{p}_i$ $(i = 0, \ldots, \bar{n}, \bar{p}_0 = 1)$ are the mean 3D shape, the 3D shape bases and the 3D shape parameters, respectively, and $\bar{n}$ is the number of 3D shape bases.

Since the 2D shape model is obtained by applying the principal component analysis (PCA) to the aligned shape data, in which the scaling, rotation, and translations are removed in the alignment process, the extra parameters must be incorporated to the AAM parameters

[1] In this paper, the local motion tracking implies the tracking of the local movements of facial features such as eyebrows, eyes, nose, and mouths, etc. in the face.

[2] In this paper, the global motion tracking implies the tracking of global head movements.

[3] The superscript ^{t} represents transpose.

to represent the removed scaling, rotation, and translation in the target image. They can be incorporated by adding four special shape bases to the 2D shape model as $\boldsymbol{s}_{n+1} = \boldsymbol{s}_0 = (x_0^1, y_0^1, \ldots, x_0^l, y_0^l)^t$, $\boldsymbol{s}_{n+2} = (-y_0^1, x_0^1, \ldots, -y_0^l, x_0^l)^t$, $\boldsymbol{s}_{n+3} = (1^1, 0^1, \ldots, 1^l, 0^l)^t$, $\boldsymbol{s}_{n+4} = (0^l, 1^l, \ldots, 0^l, 1^l)^t$, where x_0^j and y_0^j are the jth components of x and y coordinate of the mean shape $\boldsymbol{s}_0$, respectively. The extra shape parameters $\boldsymbol{q} = (p_{n+1}, \ldots, p_{n+4})^t$ represent the similarity transform, where the first two parameters (p_{n+1}, p_{n+2}) are related to the scale and rotation, and the last two parameters (p_{n+3}, p_{n+4}) are the translations along the x and y coordinates, respectively (Matthews and Baker 2004). Therefore, when a 2D shape $\boldsymbol{s}$ is given, the optimal shape parameter vector $\boldsymbol{p} = (p_1, \ldots, p_{n+4})^t$ can be easily computed as

$$\boldsymbol{p} = \Phi_s^{-1}(\boldsymbol{s} - \boldsymbol{s}_0), \tag{3}$$

where $\Phi_s = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{s}_{n+1}, \ldots, \boldsymbol{s}_{n+4}]$ is the shape basis matrix.[4]

Xiao et al. (2004a) also derived an efficient fitting algorithm of the 2D+3D AAMs that was based on the inverse compositional project out algorithm. The basic idea of their fitting algorithm is to use a set of 3D constraints that keep the 2D shape of the face to be in accordance with the projection of 3D face shape. When we use a single camera, the 3D constraint term can be represented as

$$\left\| N\left(\sum_{i=0}^{n} \boldsymbol{s}_i p_i; \boldsymbol{q}\right) - P\left(M\left(\sum_{i=0}^{\bar{n}} \bar{p}_i \bar{\mathbf{s}}_i; \bar{\mathbf{q}}\right)\right) \right\|^2 = 0, \tag{4}$$

where $N(\cdot; \boldsymbol{q})$ represents a similarity transform function by $\boldsymbol{q}$, $P()$ represents a camera projection function and $M(\cdot; \bar{\boldsymbol{q}})$ represents a rigid transformation that is parameterized by $\bar{\boldsymbol{q}}$. The overall objective function to optimize is

$$E = \sum_{\boldsymbol{x} \in \boldsymbol{s}_0} \left[\sum_{i=0}^{m} \alpha_i A_i(\boldsymbol{x}) - I(W(\boldsymbol{x}; \boldsymbol{p})) \right]^2 + K \cdot \left\| N\left(\sum_{i=0}^{n} \boldsymbol{s}_i p_i; \boldsymbol{q}\right) - P\left(M\left(\sum_{i=0}^{\bar{n}} \bar{p}_i \bar{\mathbf{s}}_i; \bar{\boldsymbol{q}}\right)\right) \right\|^2. \tag{5}$$

In this paper, we take an efficient fitting algorithm of 2D+3D AAMs introduced by Xiao et al. (2004a). We do not explain the details of the fitting algorithm because the fitting algorithm itself is not our concern. Refer to the references (Cootes et al. 2001; Matthews and Baker 2004) for more details on the fitting algorithm.

[4]From hereafter, we will use the symbol $\boldsymbol{p}$ to represent the concatenated vector of the pure 2D shape parameter vector and the similarity transform parameter vector $\boldsymbol{q}$.

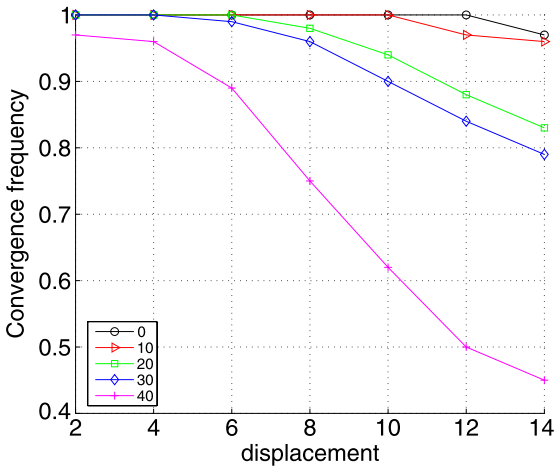
2.2 Local Motion Tracking

When the AAMs are used for the face tracking problems, the AAM parameters obtained in the previous frame are usually used as the initial parameters for the fitting in the current frame. The AAM fitting fails frequently when the head movement is large or the pose deviation from the frontal face is great. Eventually, this results in the failure of the local motion tracking. When the local motion tracking fails, we usually find the face area via face detector and eyes via eye detector to restart the AAM fitting from the initial AAM parameters estimated from these initial cues.

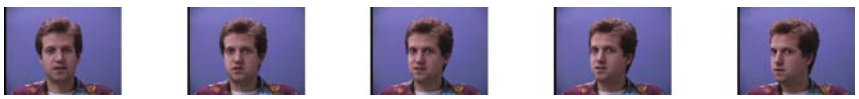
Table 1 shows a typical face tracking algorithm using the 2D+3D AAMs. It alternates two operating modes: detection and tracking, where they are different in obtaining the initial AAM parameters for fitting. In the beginning, the algorithm starts in the detection mode (step 1) and reads an input image (step 2). Then, it tries to find a face in the image. If a face is detected, then it computes the 2D similarity transform vector $\boldsymbol{q}$ so that the two eye points of the mean shape $\boldsymbol{s}_0$ match to the detected two eye points and sets the 2D shape parameters $(p_1, \ldots, p_n)$ to zero (step 3). Next, it synthesizes a 2D shape $\hat{\boldsymbol{s}}$ and computes the 3D shape parameters $\bar{\boldsymbol{p}}$ and projection parameters $\bar{\boldsymbol{q}}$ (the parameters of the affine projection model: 3D rotation, scale and 2D translation) from $\hat{\boldsymbol{s}}$ by using the algorithm proposed by Romdhani et al. (2003)

Table 1 Face tracking algorithm using the 2D+3D AAMs

	Procedure Local_motion_tracking_using_AAMs
(1)	Set *mode* = *detection* and $t = 1$.
(2)	Obtain an input image I_t.
	If *mode* = *detection*
	If a face is detected
(3)	Compute 2D similarity transform parameters $\boldsymbol{q}_{t-1} = (p_{n+1}, \ldots, p_{n+4})^t$ using the detected two eye points, and set $(p_1, \ldots, p_n)_{t-1} = \mathbf{0}$.
(4)	Synthesize a 2D shape $\hat{\boldsymbol{s}}$ using estimated AAM parameters in step (3), compute $\bar{\boldsymbol{p}}_{t-1}$ and $\bar{\boldsymbol{q}}_{t-1}$ from $\hat{\boldsymbol{s}}$, and set $\boldsymbol{\alpha}_{t-1} = \mathbf{0}$.
(5)	Set *mode* = *tracking*.
	End
	End
	If *mode* = *tracking*
(6)	Set $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1}$, $\boldsymbol{p}_t = \boldsymbol{p}_{t-1}$, $\bar{\boldsymbol{p}}_t = \bar{\boldsymbol{p}}_{t-1}$, and $\bar{\boldsymbol{q}}_t = \bar{\boldsymbol{q}}_{t-1}$.
(7)	Obtain optimal 2D+3D AAM parameters $\boldsymbol{\alpha}_t$, $\boldsymbol{p}_t$, $\bar{\boldsymbol{p}}_t$, and $\bar{\boldsymbol{q}}_t$ by fitting the 2D+3D AAM to I_t.
	If failed to converge correctly
(8)	Set *mode* = *detection*.
	End
	End
(9)	Set $t = t + 1$ and goto (2).



(a) Convergence rate vs. initial displacements



(b) Faces at five yawing angles: 0°, 10°, 20°, 30°, and 40°, respectively.

Fig. 2 The sensitivity of AAM fitting to initial model parameters

(step 4). Then, it changes the operating mode to the tracking mode (step 5). In the tracking mode, it takes the previous 2D+3D AAM parameters as the initial 2D+3D AAM parameters (step 6), and fits the 2D+3D AAM to the current image to obtain optimal 2D+3D AAM parameters (step 7). If it fails to converge to the current image correctly, then it changes the operating mode to detection mode and goes back to step (2) (step 9). Although it is difficult to determine whether the AAM's fitting is successful or not, we use the amount of difference between the model synthesized image and the current image as a measure of successful fitting in this work.

However, the AAM parameters in the previous frame are not always a good estimate for the current frame when the face is moving fast. The AAM's sensitivity to initial parameters becomes more critical problem when the face pose is deviated from the frontal view. As the face pose is approaching to the profile view, the larger area of the face becomes invisible, which means that only a small part of the texture information is available and the fitting on the little textured image is apt to fail. Figure 2(a) shows how sensitive the 2D+3D AAM fitting is to the initial model parameters at five different pose angles (0°, 10°, 20°, 30°, and 40°, see Fig. 2(b)). The fitting performance is measured by the convergence rate that is defined by the ration of the number of successful AAM fitting over the number of trials of AAM fitting (= 100). The initial model parameters for each trial of AAM fitting are computed by translating the ground truth landmark points by the amount of displacement value to a random direction. This figure shows that (1) the convergence rate decreases as the initial displacement increases for a given pose angle and (2) the convergence rate decreases abruptly as the face pose is deviated from the frontal view for a given initial displacement due to the lack of texture information; The area of the left half side of the face becomes small (15,675, 11,803, 9,387, 8,044, and 5,423) as the pose angle is deviated from the frontal view.

Another problem occurs when the most part of the face becomes invisible. In this situation, the AAMs fail to fitting, and the algorithm loses its tracing of the face (step 10). The problem is that the AAMs cannot work until they are re-initialized even though they possess the fitting capability to the images that arrives before the re-initialization.

3 CHMs and Global Head Motion Tracking

3.1 Cylinder Head Models

The cylinder head models assume that the head is shaped as a cylinder and the face is approximated by the cylinder surface. Since the global motion of the cylinder is a kine of rigid motion, the global motion can be parameterized by a rigid motion vector $\boldsymbol{\mu}$, which includes the 3D rotation angles (w_x, w_y, w_z) and the 3D translations (t_x, t_y, t_z). When the 3D coordinate of a point on the cylinder surface is $\boldsymbol{x} = (x, y, z)^t$ in the camera-centered 3D coordinate system, the new location of $\boldsymbol{x}$ transformed by the rigid motion vector $\boldsymbol{\mu}$ is

$$\boldsymbol{M}(\boldsymbol{x}; \boldsymbol{\mu}) = R\boldsymbol{x} + T, \tag{6}$$

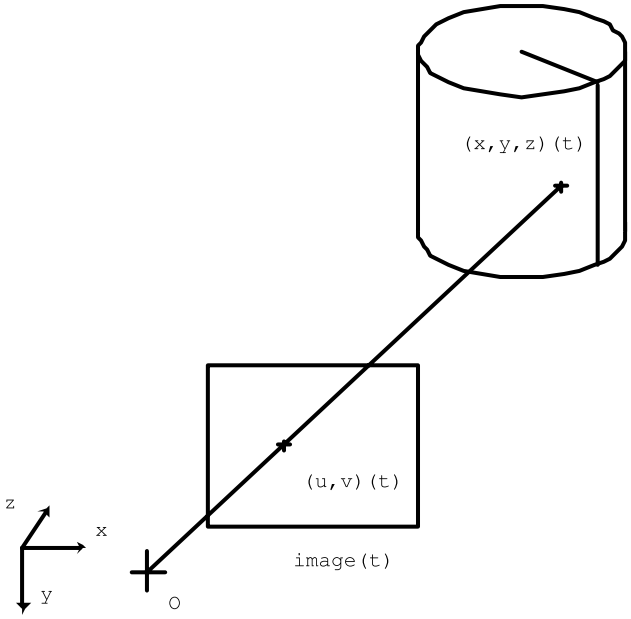


Fig. 3 A cylinder model in the camera-centered 3D coordinate system

where M is a rigid transformation function that is represented by a 3D rotation matrix $R \in \mathbb{R}^{3\times3}$ corresponding to (w_x, w_y, w_z) and a 3D translation vector $T \in \mathbb{R}^{3\times1} = (t_x, t_y, t_z)^t$. Figure 3 shows a cylinder surface point $\boldsymbol{x}_t = (x, y, z)$ in the camera-centered 3D coordinate system and its projection point $\boldsymbol{u}_t = (u, v)$ in the image plane at time t.

When the rigid head motion between the time t and $t+1$ is represented by the rigid motion vector $\Delta\boldsymbol{\mu}$, the new location of $\boldsymbol{u}_t$ at time $t+1$ can be represented as

$$\boldsymbol{u}_{t+1} = \boldsymbol{F}(\boldsymbol{u}_t; \Delta\boldsymbol{\mu}), \tag{7}$$

where $\boldsymbol{F}$ is the 2D parametric motion function of $\boldsymbol{u}_t$. If we assume that the illumination condition does not change between two image frames, then the intensity of the pixel $I_t(\boldsymbol{u}_t)$ must be consistent with that of the corresponding pixel in the next frame image as

$$I_{t+1}(\boldsymbol{F}(\boldsymbol{u}_t; \Delta\boldsymbol{\mu})) = I_t(\boldsymbol{u}_t). \tag{8}$$

The rigid motion vector $\Delta\boldsymbol{\mu}$ can be obtained by minimizing the image difference between two image frames as

$$\min E(\Delta\boldsymbol{\mu}) = \sum_{\boldsymbol{u}\in\Omega} \left\{ I_{t+1}(\boldsymbol{F}(\boldsymbol{u}_t; \Delta\boldsymbol{\mu})) - I_t(\boldsymbol{u}_t) \right\}^2, \tag{9}$$

where Ω is the region of the template I_t whose corresponding pixel at the time $t+1$ is also visible. The minimization problem can be solved by the Lucas-Kanade method (Kanade and Lucas 1981) as

$$\Delta\boldsymbol{\mu} = -\left(\sum_{\Omega} (I_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{\mu}})^t (I_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{\mu}}) \right)^{-1} \sum_{\Omega} (I^t (I_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{\mu}})^t), \tag{10}$$

where $I_{\boldsymbol{u}}$ and I_t are the spatial and temporal image gradient, and $\boldsymbol{F}_{\boldsymbol{\mu}} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\mu}}|_{\Delta\boldsymbol{\mu}=0}$ denotes the partial derivative of $\boldsymbol{F}$ with respect to the rigid motion vector.

The new location $\boldsymbol{u}_{t+1}$ is the perspective projection of a 3D point that is the rigid transformation M of $\boldsymbol{x}_t$ by the rigid motion vector $\Delta\boldsymbol{\mu} = (w_x, w_y, w_z, t_x, t_y, t_z)^t$ as

$$\boldsymbol{u}_{t+1} = \boldsymbol{p}(\boldsymbol{M}(\boldsymbol{x}_t; \Delta\boldsymbol{\mu})), \tag{11}$$

where $\boldsymbol{p}$ represents the perspective projection function. When the rotation is small, the rotation matrix can be approximated as

$$R = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_y & 1 \end{bmatrix} \tag{12}$$

using the exponential map (Bregler and Malik 1998). Then, the perspective projection function $\boldsymbol{p}$ in (11) can be represented as

$$\boldsymbol{p}(\boldsymbol{M}(\boldsymbol{x}_t; \boldsymbol{\Delta\mu})) = \begin{bmatrix} x - y\omega_z + z\omega_y + t_x \\ x\omega_z + y - z\omega_x + t_y \end{bmatrix} \times \frac{f}{-x\omega_y + y\omega_x + z + t_z}, \tag{13}$$

where f is the camera focal length.

By comparing (7) and (13), we know that the perspective projection function $\boldsymbol{p}$ is merely a parametric motion function $\boldsymbol{F}(\boldsymbol{u}_t; \Delta\boldsymbol{\mu})$. Thus, the Jacobian of the parametric motion function $\boldsymbol{F}_{\boldsymbol{\mu}}|_{\Delta\boldsymbol{\mu}=0}$ can be computed by the derivative of the perspective projection function $\boldsymbol{p}$ as

$$\boldsymbol{F}_{\boldsymbol{\mu}}|_{\Delta\boldsymbol{\mu}=0} = \begin{bmatrix} -xy & x^2+z^2 & -yz & z & 0-x \\ -(y^2+z^2) & xy & xz & 0 & z-y \end{bmatrix} \frac{f}{z^2}. \tag{14}$$

By plugging the computed Jacobian of the parametric motion function into (10), we can obtain the rigid head motion vector $\boldsymbol{\Delta\mu}$ (Xiao and Kanade 2002).

If we want to obtain the exact distance of head from the camera, we need to know the exact size of head and the focal length of camera. However, we fix the size of the cylinder model to be constant in this work. Hence, the estimated translation t_z does not give the exact distance of head from the camera. In addition, we do not know the exact focal length f when we do not calibrate the camera. However, it is known that the effect of f is usually small (Aggarwal et al. 2005).

3.2 Global Motion Tracking

Although the AAMs and the CHMs use the same gradient-based iterative optimization techniques for their fitting algorithms, fitting of the CHMs is easier than that of the AAMs

Table 2 Global motion tracking algorithm using the CHMs

	Procedure Global_motion_tracking_using_CHMs
(1)	Set $mode = detection$ and $t = 1$.
(2)	Obtain an input image I_t.
	If $mode = detection$
	If a face is detected
(3)	Initialize the global motion parameters $\boldsymbol{\mu}$ of CHMs using the face Detection result.
(4)	Set $mode = tracking$ and go to step (10).
	End
	Else
(5)	Set $\boldsymbol{\mu}_{t-1}$ as the initial global motion parameters of $\boldsymbol{\mu}_t$.
(6)	Estimate the current global motion parameters $\boldsymbol{\mu}_t$ using the fitting algorithm given in Sect. 3.1.
(7)	Perform *re-registration* given in Table 3.
(8)	Perform *dynamic template update* that takes the current input image I_t as the template image T for the next time.
	If failed to converge correctly
(9)	Set $mode = detection$.
	End
	End
(10)	Set $t = t + 1$ and goto step (2).

Table 3 Re-registration procedure

	Procedure Re-registration
(7.1)	Find a reference image I_r whose global motion parameters $\boldsymbol{\mu}_r$ are the most closest to $\boldsymbol{\mu}_t$ among the registered reference images.
	If $\|\|\boldsymbol{\mu}_r - \boldsymbol{\mu}_t\|\| < \delta_1$
(7.2)	Estimate the global motion parameters $\boldsymbol{\mu}'_t$ using the selected reference image I_r.
	If $error(I_r, T) < error(I_t, T)$
(7.3)	Set $\boldsymbol{\mu}_t = \boldsymbol{\mu}'_t$
	End
	End

because the CHMs have only 6 global motion parameters. Therefore, it is better to use the CHMs for global motion of the face. Among the various face tracking methods that use the CHMs, we take the tracking method that was proposed by Xiao and Kanade (2002) due to its robustness to the occlusion, the local motion in the face, and the gradual lighting change.

Table 2 shows the face tracking algorithm using the CHM. Initially, the algorithm operates in the detection mode (step 1) and reads an image (step 2). Then, it tries to find a face in the current image. If a face is detected, it initializes the global motion parameters $\boldsymbol{\mu}_t$ using the face detection results (step 3), and changes the operating mode into the tracking mode and proceeds to the next input image (step 10).

When it is operating in tracking mode, it takes the previous global motion parameters $\boldsymbol{\mu}_{t-1}$ as the initial values of the current global motion parameters $\boldsymbol{\mu}_t$ (step 5). Then, it estimates the global motion parameters $\boldsymbol{\mu}_t$ using the fitting algorithm given in Sect. 3.1 (step 6). Next, it performs the re-registration procedure that is given in Table 3 (step 7). Finally, it performs the dynamic template update in order to adapt the template image with the gradually changing current input image (step 8). When it fails to converge to the current image, it change the operating mode into the detection mode and goes back to step (2).

4 Combining AAMs and CHMs and Pose Robust Face Tracking

4.1 Combining AAMs and CHMs

As mentioned before, the AAMs and CHMs are appropriate for the local motion tracking of facial feature movements and global motion tracking of head movements, respectively. Hence, the face tracking using the AAMs works only when the face pose is deviated not too much from the frontal view. To overcome this limitation, we propose to combine the AAMs and the CHMs, where the global motion parameters of the CHMs provide the cues of the initial parameters that are appropriate for the AAM fitting. By combining these two models, we expect that we can obtain advantages by the following. First, the duration of successful face tracking increases because the CHMs re-initialize the AAMs effectively. Second, the pose angle of successful face tracking extends because the CHMs can provide accurate initial parameters for the AAM fitting. Third, the pose angle of the CHMs can be used to determine whether to run the AAM fitting or not. Last, the result of the AAM fitting is used to initialize the CHMs.

Although the global motion parameters of the CHMs can be initialized using the result of the face detection, its location and size may not exactly match the face in the image. This incorrect initialization of the CHMs results in the failure of the face tracking. To avoid this problem, we propose to use the AAM fitting results for the CHM initialization because the AAMs provide precise locations of the landmark points such as face boundaries and eyes. The initial global motion parameters of the CHMs are computed in two steps. First, the AAMs are initialized using the face detection result and fitted to the image as given in Sect. 2.2. Second, once the AAM is fitted to the image, we can obtain the 3D rotation angles and 3D translation parameters as follows. We took the 3D rotation angles of the 2D+3D AAM fitting (pitch, yaw, and roll that are the rotations with respect to x, y, and z axis, respectively) as the initial 3D rotation parameters of

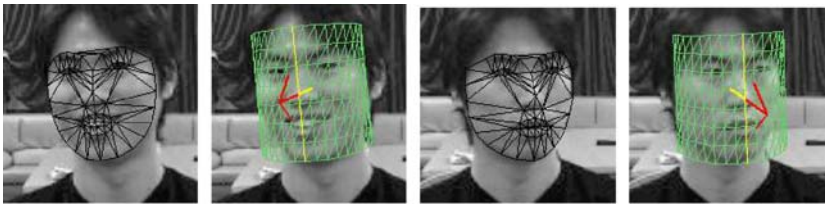
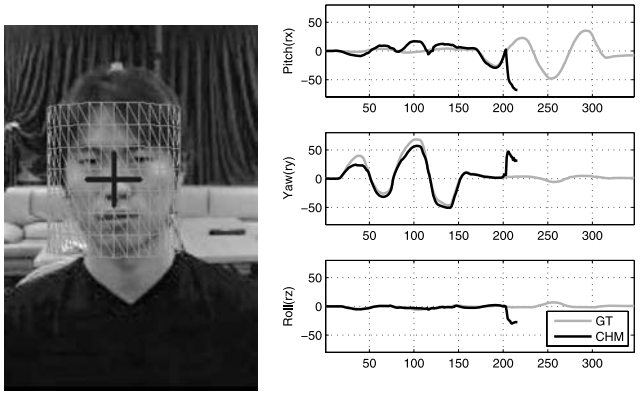


Fig. 4 Examples of automatic initializations of the CHM using the AAM fitting results

(a) (b) (c) (d)



(a) (b)

Fig. 5 Face tracking results using badly initialized CHM

the CHMs. Then, the t_z parameters are computed. Because we assumed that the physical size of the cylinder model and the focal length of the camera were fixed, the t_z parameters can be computed by comparing the width of the face in the image and the physical width of our cylinder model. Next, the t_x and t_y parameters are computed so that the bottom line of the face (jaw) matches the bottom of the projected cylinder model and the t_x is computed so that the left and right boundaries of the face match to the boundary of the projected cylinder model whose distance is t_z. Figure 4(b) and Fig. 4(d) show the two examples of the automatically initialized CHMs, using the two AAM fitting results, Fig. 4(a) and Fig. 4(c), respectively. Figure 4(b) shows that the CHM could be successfully initialized although the face was rotated: the face was not upright and not looking the front view.

Sometimes, the tracking performance of the CHM was largely affected by the correctness of the initial pose parameters. The initialization of the CHM using only the front view face detector can be bad when the face detection result was bad or the face is rotated. Figure 5(a) shows an example of badly initialized CHM, where the face area was estimated larger than the face in the image, and Fig. 5(b) shows the face tracking results, where the three plots, from top to bottom, represent pitch, yaw, and roll rotation angles obtained from the ground truth and the CHM, respectively. In this example, the CHM failed to track the face after 200th image frame.

Table 4 A procedure that estimates the initial AAM parameters from the CHM fitting result $\Delta\boldsymbol{\mu}$

Procedure Estimation_of_initial_AAM_parameters

(1) Each point of the 2D shape $\boldsymbol{s}_{t-1}$ that is obtained in the previous time is inversely projected to the cylinder surface to compute corresponding 3D coordinates $\{\boldsymbol{x}^j_{t-1}\}^v_{j=1}$ as $\boldsymbol{x}^j_{t-1} = \boldsymbol{p}^{-1}(\boldsymbol{s}^j_{t-1}; \boldsymbol{\mu}_{t-1})$, where $\boldsymbol{\mu}_{t-1}$ is the previous global motion parameters.

(2) The 3D surface points $\{\boldsymbol{x}^j_{t-1}\}^v_{j=1}$ are transformed by the global motion $\Delta\boldsymbol{\mu}$ that is estimated by the CHM fitting: $\hat{\boldsymbol{x}}^j_t = M(\boldsymbol{x}^j_{t-1}; \Delta\boldsymbol{\mu})$.

(3) Project the transformed 3D surface points to image plane to obtain the estimate of the current 2D shape $\hat{\boldsymbol{s}}_t$: $\hat{\boldsymbol{s}}^j_t = \boldsymbol{p}(\hat{\boldsymbol{x}}^j_t)$.

(4) Compute the 2D and 3D shape parameters of the 2D+3D AAM from $\hat{\boldsymbol{s}}_t$, and set the previous appearance parameters as the initial appearance parameters: $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1}$.

Table 4 shows the detailed procedure to compute the initial 2D+3D AAM parameters using the global motion parameters $\Delta\boldsymbol{\mu}$ estimated by the CHM fitting. In the procedure, the initial shape and motion parameters ($\boldsymbol{p}$, $\bar{\boldsymbol{p}}$ and $\bar{\boldsymbol{q}}$) are estimated using $\Delta\boldsymbol{\mu}$ and the appearance parameters are copied from previous appearance parameters.

Sometimes, it is impossible to compute the inverse projection when the locations of the shape points do not belong to the area that is the projection of the CHM. In that case, we used the previously computed surface coordinates corresponding to such shape points, instead of computing the surface coordinates of them in the step (step 1) of the Table 4. The points on the cylinder surface may become invisible as the face rotates. Thus, only the visible surface points are projected to image and used to compute the model parameters of the 2D+3D AAM in step (3) of Table 4.

4.2 Face Tracking Using the AAMs and CHMs

Table 5 shows the proposed face tracking algorithm using the combination of the AAMs and the CHMs. Initially, the algorithm operates in the detection mode (step 1) and reads an input image (step 2). Then, it tries to find a face in the current image. If a face is detected, it computes the initial

Table 5 The proposed face tracking algorithm using the AAMs and CHMs

Procedure Face_tracking_using_AAMs_and_CHMs

(1) Set *mode* = *detection* and $t = 1$.
(2) Obtain an input image I_t.
If *mode* = *detection*
  If a face is detected
(3) Initialize the global motion parameters $\boldsymbol{\mu}_t$ of the CHM using the result of 2D+3D AAM fitting as explained in the Sect. 4.1.
(4) Set *mode* = *tracking* and go to step (9).
  End
Else
(5) Estimate the global motion parameters $\boldsymbol{\mu}_t$ using the global motion tracking algorithm in Sect. 3.2.
  If failed to converge correctly
(6) Set *mode* = *detection* and go to step (9).
  End
  If pose angle is within the AAM working range
(7) Perform the procedure *Estimation of initial AAM parameters* given in Table 4.
(8) Obtain the optimal model parameters $\boldsymbol{\alpha}_t$, $\boldsymbol{p}_t$, $\bar{\boldsymbol{p}}_t$, and $\bar{\boldsymbol{q}}_t$ by fitting the 2D+3D AAM on I_t.
  End
End
(9) Set $t = t + 1$ and goto step (2).

2D+3D AAM parameters using the face detection result and fits the 2D+3D AAM to the current image. Then, it initializes the global motion parameters $\boldsymbol{\mu}_t$ using the 2D+3D AAM fitting result (step 3), changes the operating mode into the tracking mode, and proceeds to the next input image (step 4). When it is operating in the tracking mode, it estimates the global motion parameters $\boldsymbol{\mu}_t$ as in the face tracking algorithm given in Sect. 3.2 (step 5). If it fails to converge to the current image, then it changes the operating mode into the detection mode and proceeds to the next input image (step 6), else it performs the *estimation of initial AAM parameters from the CHM fitting result* procedure, whose details are given in Table 4 (step 7). Then, it fits the 2D+3D AAM to the current image (step 8).

5 Experiment Results and Discussion

5.1 Global Motion Tracking Performance of the CHM

The role of the global head motion tracker is crucial in the proposed algorithm because the local motion tracker is initialized using the estimated global motion parameter. To show the global motion tracking performance of the CHM, we evaluated it using the BOSTON database (Cascia et al. 2000) that contains various global head motion image sequences and their ground truth head pose angles. We applied the global head motion tracking algorithm on the 45 image sequences that are recorded under the uniform light environment.

The global head motion tracking performance is measured by the average of three pose angle errors, which are computed by summing pose angle errors over all the image sequences and averaging the sums over the entire number of frames in the tested image sequences. The average pose angle errors with respect to yaw, tilt, and roll head motions were 5.4°, 5.6°, and 3.1°, respectively. These results show that the CHM-based global head motion tracking algorithm works well under the large pose variation.

Figures 6(a), (b), and (c) show three examples of the global head motion tracking results in the cases of the tilt, yaw, and roll head motion image sequences, respectively, which are successfully tracking the head in all cases. Also, Figs. 6(d), (e), and (f) show the estimated head pose angles and the ground truth angles in the cases of the tilt, yaw, and roll head motion image sequences, respectively, where the estimated head pose angle and its ground truth are denoted by the dotted and solid lines, respectively.

5.2 Global and Local Motion Tracking Performance of the AAM+CHM

5.2.1 Data Set

We prepared two datasets: DATA-I and DATA-II. The first dataset DATA-I consists of two image sequences of a specific person with clear backgrounds, which were ideal cases that are good to validate the proposed method. One image sequence consists of 409 images with freely yawing and tilting head movements and the other image sequence consists of 839 images with freely yawing head movements. The head moves from left to right and then from right to left repeatedly increasing the magnitude of yawing angle. The second dataset DATA-II consists of 10 pose varying image sequences, one for each person, which were recorded in a normal indoor situation with cluttered backgrounds. Each image sequence of DATA-II consists of between 140 to 250 images, where the head motions are similar to that of the second image sequence of DATA-I but the speed of motions are different from person to person. They were recorded using a cheap web camera, so the image quality of DATA-II is not as good as that of the DATA-I. The ground truth pose angles were also recorded using the FASTRAK motion capture system.

5.2.2 Evaluation Methods

We compared three face tracking methods: 'AAM', 'CHM', and 'AAM+CHM', where they denote the face tracking

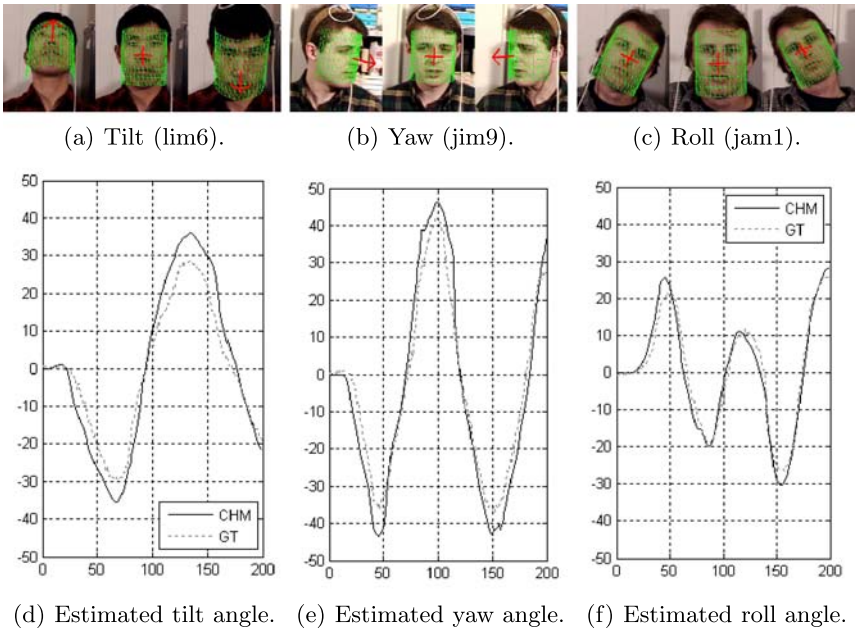


Fig. 6 Examples of global head motion tracking results using the CHM

(a) Tilt (lim6). (b) Yaw (jim9). (c) Roll (jam1).

(d) Estimated tilt angle. (e) Estimated yaw angle. (f) Estimated roll angle.

method using only the 2D+3D AAM, the face tracking method using only the CHM, and the face tracking method that combines the 2D+3D AAM and the CHM together, respectively. In this work, we evaluated the face tracking performance in terms of two measures: *tracking rate* and *pose coverage*.

The tracking rate measures how many image frames are successfully tracked as

$$\text{tracking rate} = \frac{\text{the number of successfully tracked image frames}}{\text{the number of total image frames}}, \tag{15}$$

where the face is successfully tracked when the RMS error between the fitted shape points and the ground truth landmark points is smaller than a given threshold.

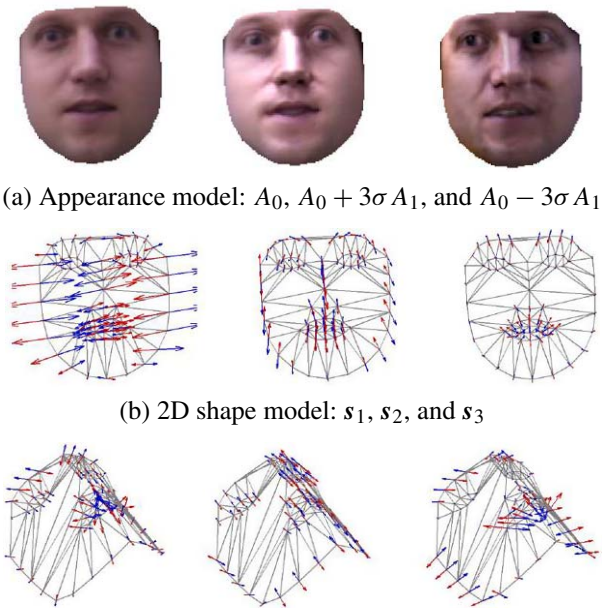
The pose coverage measures the range of pose angles where the face is successfully tracked as

$$\text{pose coverage} = (\min(\{\tilde{\theta}_k\}), \max(\{\tilde{\theta}_k\})), \quad k \in \mathcal{B}, \tag{16}$$

where $\tilde{\theta}_k$ and $\mathcal{B}$ represent the pose angles of the kth frames and a set of frame numbers in which the face is tracked successfully.

5.2.3 Face Tracking Experiments Using DATA-I

In this experiment, the first image sequence of DATA-I was used to construct the 2D+3D AAM, i.e., the sequence was manually landmarked, and gathered to build a 2D+3D AAM. The 3D shape model was also built from the landmarked shapes using the *structure from motion* algorithm (Xiao et al. 2004b). The number of bases in each model was determined to keep 95% of variations. Figure 7 shows some examples of the appearance, the 2D, and 3D shape models.



(a) Appearance model: A_0, $A_0 + 3\sigma A_1$, and $A_0 - 3\sigma A_1$

(b) 2D shape model: $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, and $\boldsymbol{s}_3$

(c) 2D shape model: $\bar{\boldsymbol{s}}_1$, $\bar{\boldsymbol{s}}_2$, and $\bar{\boldsymbol{s}}_3 e$

Fig. 7 The appearance, the 2D shape, and the 3D shape model of the 2D+3D AAM

We compared the performances of three different face tracking algorithms: 'AAM', 'CHM' and 'AAM+CHM' using the first image sequence in DATA-I, which contains the moderate yawing and tilting head motions. Figure 8 shows the measured RMS errors of three different tracking methods, where the horizontal and vertical axes represent the

frame number and the RMS error value, respectively, and the thin solid, and dashed, and thick solid lines denote the RMS errors of the 'AAM', 'CHM', and 'AAM+CHM' face tracking methods, respectively. To measure the RMS error of the 'CHM' tracking method, we computed the surface points of the CHM that corresponded to the ground truth landmark points by the inverse projection when the CHM was initialized. In the next frames, we applied the 'CHM' tracking method and computed the 2D coordinates of the surface points that were moved along the estimated global motion. The 2D coordinates were treated as the fitted shape points of the 'CHM' tracking method to measure the RMS error. Figure 8 shows that the 'AAM+CHM' tracking method tracks the whole image sequence successfully, while the 'AAM' tracking method fails to track the face in many frames, especially at the end of the image sequence even though the AAM was built from the training image sequence. The 'CHM' tracking method fails to track the face because it was badly initialized due to the bad face detection result.

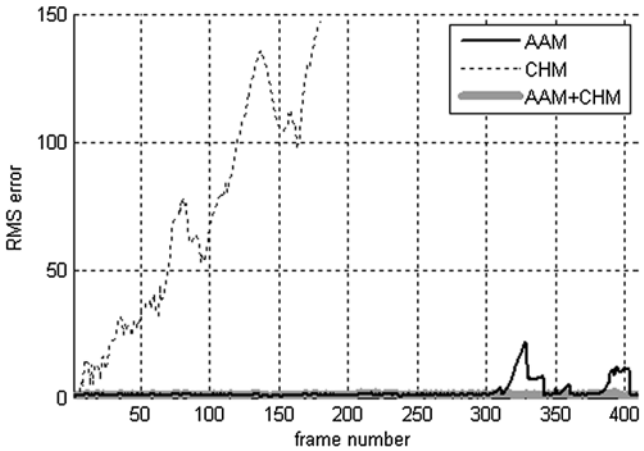


Fig. 8 Face tracking results in a training image sequence

Figure 9 compares the tracking results of the three different tracking algorithms at some specific image frames, where the first, second and third rows correspond to tracking result of the 'AAM', 'CHM' and 'AAM+CHM' tracking methods and the face turns from the upper left side (310th frame), through the front (330th frame), to the upper right side (350th frame). In the case of the 'AAM' tracking method, the face tracking begin to fail near the 320th frame because there is a large head motion between the 310th and 320th frame. Then, it continues to fail because the model parameters at the previous frame are not appropriate for the initial parameters for the face image at the current frame. The success of face tracking at the 350th frame occurred unexpectedly because the fitting result at the 340th frame was as good as the initial parameters for the face image at the 340th frame. In the case of 'CHM' tracking method, the tracking begin to fail soon after an initialization, where the CHM was initialized at the 250th frame such that the cylinder is larger than the face in the image due to the bad face detection result. In the case of the 'AAM+CHM' tracking method, it tracks the face successfully during all image frames.

Table 6 compares three different tracking methods in terms of the tracking rate and the pose coverage, where the pose coverage of the 'CHM' is not determined due to the failure of tracking. We know that (1) the pose coverages of 'AAM' and 'AAM+CHM' tracking methods are almost the

Table 6 Comparison of the tracking performances in the training image sequence

Tracking rate	AAM	CHM	AAM+CHM
	0.87	0.18	1.00
Pose coverage (yaw)	$(-37.2, 40.4)°$	–	$(-37.2, 40.4)°$
Pose coverage (tilt)	$(-31.9, 10.9)°$	–	$(-31.9, 10.9)°$
Pose coverage (roll)	$(-4.9, 13.2)°$	–	$(-6.0, 13.8)°$

Fig. 9 Some tracking results of three different tracking methods

(a) Results of AAM tracking method.

(b) Results of CHM tracking method.

(c) Results of AAM+CHM tracking method.

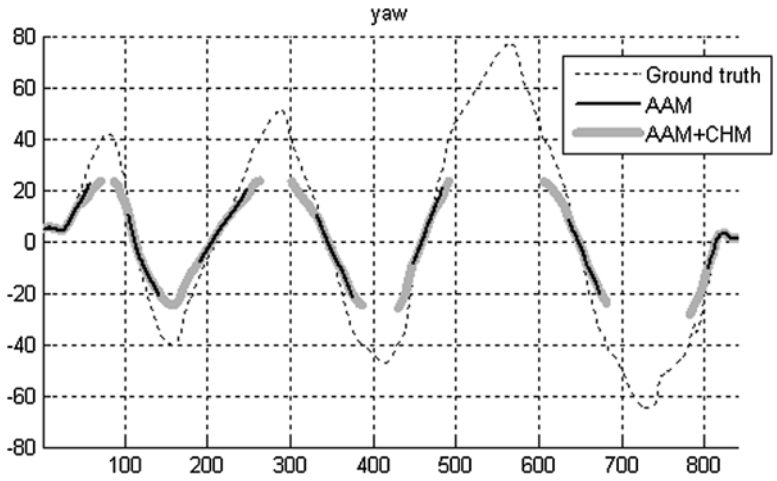


Fig. 10 Face tracking result in a test image sequence

same because we compare the tracking performance of two tracking methods by using the training image sequence, (2) the maximum angles of yawing and tilting are about $\pm 40°$, (3) the tracking rate of the 'CHM' tracking method can be very low when it is badly initialized; In this experiment, the tracking rate was 0.18, and (4) the 'AAM+CHM' tracking method greatly improves the face tracking performance; the tracking rate increased from 0.87 to 1.0. Since the head motion in the training image sequence is within the range of the AAM fitting, the re-initialization did not occur. Therefore, the improvement of the tracking rate of the 'AAM+CHM' tracking method was obtained by the proposed initialization method of combining the AAMs and the CHMs.

Second, we compared the face tracking performance of the 'AAM' and 'AAM+CHM' tracking methods in terms of the tracking rate and the pose coverage using the second image sequence, which contains a large amount of the yawing head motions. When we employed the 'AAM+CHM' tracking method, we used the tracking result of the CHM for the AAM re-initialization, where the CHM was precisely initialized using the fitting result of the AAM at the first frame. As mentioned before, the 'AAM+CHM' tracking method can designate the working range of the AAMs. In this work, we took the working range $\pm 40°$ of yawing angles because the previous experiment showed that the yawing angle coverage was $(-37.2, 40.4)$. When we employed the 'AAM' tracking method, we used the front view face detector for the AAM re-initialization. For this experiment, we used the fitted face shapes obtained from by the 'AAM+CHM' tracking method as the ground truths. This implies that the tracking rate of the 'AAM+CHM' tracking method is 1. Then, in the case of 'AAM' tracking method, an image frame was judged to be successfully tracked if the RMS error between the fitted shape and the ground truth is less than 2.5. Figure 10 compares the estimated yawing angles of two different tracking methods: 'AAM' and 'AAM+CHM', where the horizontal and vertical axes represent the frame number and the estimated yawing angles, respectively, and the thin and solid line denote the estimated yawing angles obtained by the 'AAM' and 'AAM+CHM' tracking method, respectively, and the dotted line denotes the ground truth yawing angle. This figure shows that the 'AAM+CHM' tracking method outperforms the 'AAM' tracking method in terms of the tracking rate and the pose coverage.

Table 7 Comparison of the tracking performances in the test image sequence

Tracking rate	AAM	AAM+CHM
	0.37	0.63
Angle coverage (yaw)	$(-34.4, 34.8)°$	$(-39.8, 40.0)°$
Angle coverage (tilt)	$(-17.2, 3.3)°$	$(-17.2, 4.6)°$
Angle coverage (roll)	$(-7.4, 5.0)°$	$(-7.4, 5.2)°$

Table 7 compares two different tracking methods in terms of the tracking rate and the pose coverage, where the second and third columns correspond to the 'AAM' and 'AAM+CHM' tracking method, respectively. This table shows that (1) the tracking rates of the 'AAM' and 'AAM+CHM' tracking method are 0.37 and 0.63, respectively, which implies the improvement of tracking rate by 170%, and (2) the pose coverage of the 'AAM' and 'AAM+CHM' tracking method are $(-34.4, 34.8)$ and $(-39.8, 40.0)$, respectively, which implies the improvement of pose coverage by 115%.

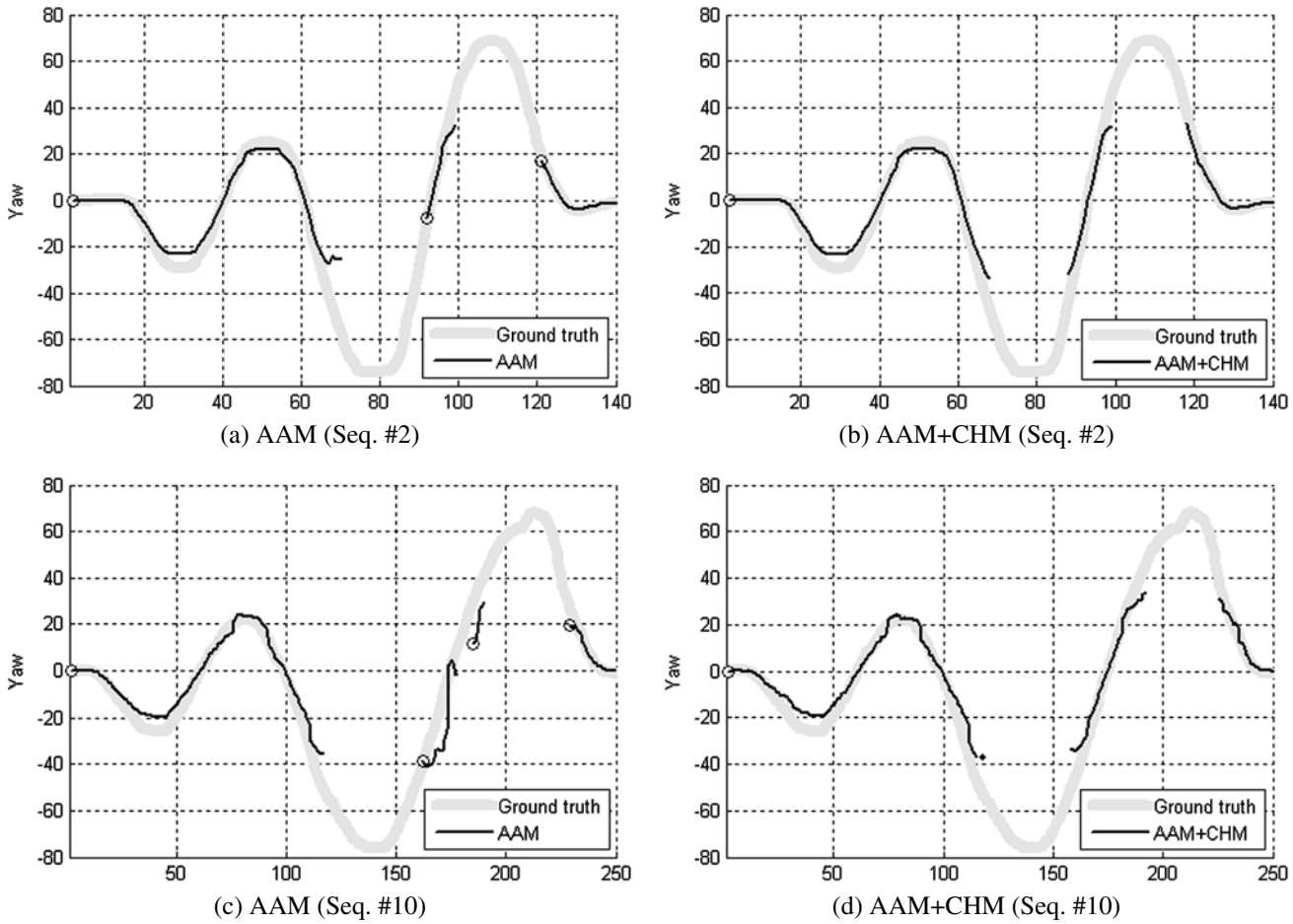
The improvement of tracking performance in the 'AAM+CHM' tracking method is due to the following facts. First, the improved tracking rate was obtained by the proper re-initialization of the AAM by the estimated global motion parameters of the CHM when the head pose moves inside from the outside of the AAM's working range. Second, the improved pose coverage was obtained by the good initialization of the AAM by the estimated global motion parameters of the CHM near the boundary between the working and non-working area of the AAM fitting.

5.2.4 Face Tracking Experiments Using DATA-II

In these experiments, we used a generic 2D+3D AAM instead of using the person specific AAM, which was used in the face tracking experiments using DATA-I (Gross et al. 2004a). To build a generic 2D+3D AAM, we gathered the training images from the image sequences of 10 people in the DATA-II. For each image sequence, we picked about 30 images, which approximately correspond to every 5th image in the sequence, and gathered total 298 images. Next, the gathered images were manually landmarked and used to construct a generic 2D+3D AAM. The *structure from motion* algorithm (Xiao et al. 2004b) was used again and the number of bases in the 2D+3D AAM was determined to keep 95% of variations of the training data.

We applied two tracking methods, 'AAM' and 'AAM+CHM', to the 10 image sequences, which contain gradually increasing yawing head motions. We set the AAM working range $\pm 40°$ of yawing angles because the training images had $\pm 30°$ variations of yawing angles and we used robust AAM fitting algorithm (Gross et al. 2004b) that enables the AAM to fit to largely rotated face images. The success of fitting was judges by the RMS error, where the ground truth landmark positions were obtained by from the fitting result of the 'AAM+CHM' and the threshold value was 2.5. Because some fitting results of the 'AAM+CHM' was not perfect in these experiments, we examined all the fitting results and manually corrected the position of landmark points when the fitting result was unsatisfactory.

Figure 11 shows some examples of the tracking results, where the vertical and horizontal axis represent the yawing angle and the image frame number, respectively, where the top and bottom row correspond to the tracking results of the second person and the 10th person, respectively, and the left and right column correspond to the tracking results of the 'AAM' and 'AAM+CHM' tracking methods, respectively. For the 2nd image sequence, we can see that both tracking methods track the face successfully when the yawing angle is small and fail at the almost same time (70th frame) but the 'AAM+CHM' tracking method is re-initialized earlier than the 'AAM' tracking method. For the 10th image sequence, we can see that the 'AAM+CHM' tracking method is far more stable than the 'AAM' tracking method; the 'AAM' tracking method failed at the 175th frame again because the



(a) AAM (Seq. #2) (b) AAM+CHM (Seq. #2)

(c) AAM (Seq. #10) (d) AAM+CHM (Seq. #10)

Fig. 11 Face tracking results for the 2nd and 10th image sequences of DATA-II

Table 8 Comparisons of the tracking rates

Seq. No.	Tracking rate		Pose coverage (yaw)	
	AAM	AAM+CHM	AAM	AAM+CHM
1	0.67	0.90	(−42, +25)	(−45, +44)
2	0.92	0.93	(−41, +38)	(−41, +38)
3	0.89	0.99	(−42, +32)	(−45, +45)
4	0.78	0.93	(−30, +20)	(−44, +45)
5	0.92	0.97	(−43, +38)	(−43, +44)
6	0.80	0.89	(−44, +22)	(−44, +32)
7	0.88	0.89	(−40, +28)	(−30, +28)
8	0.74	0.97	(−44, +44)	(−44, +44)
9	0.88	0.98	(−45, +45)	(−43, +44)
10	0.94	0.96	(−45, +43)	(−45, +43)
Average	0.84	0.94	avg. range: 75°	avg. range: 83°

face detector detects the face at 160th image frame but the pose angle is about 40° and the face detection result provides bad initial parameters for further AAM fittings. However, the 'AAM+CHM' tracking method shows a stable face tracking result within the AAM working range.

Table 8 compares the 'AAM' and 'AAM+CHM' tracking methods in terms of the tracking rate and pose coverage, where each row corresponds to each image sequence in DATA-II. In this table, we present only the coverage of yaw angle because the largest pose variation in the image sequences is the yawing head motion. Because we have the ground truth pose angles of all the images in DATA-II, the tracking rates are measured over a subset of images whose pose angles belong to the AAM working range. We know that (1) the tracking rates of the 'AAM+CHM' tracking method are higher than those of the 'AAM' tracking method, where the average tracking rates of the 'AAM+CHM' and 'AAM' tracking method are 0.94 and 0.84, respectively, and (2) the pose coverages of the 'AAM+CHM' tracking method are wider than those of the 'AAM' tracking method, where the average yaw angle ranges of the 'AAM+CHM' and 'AAM' tracking method are 83°, and 75°, respectively.

6 Conclusion

The human face has both rigid and non-rigid natures. The rigid nature has been adapted in the CHM-based face tracking algorithms, which are interested in extracting 6 rigid motion parameters and have worked robustly to large pose changes. The non-rigid nature has been modeled in the AAM-based face fitting algorithms, which usually have many parameters to represent the variations of the shape and appearance, and have suffered from the problem of estimating good initial AAM parameters.

We proposed a face tracking algorithm that uses two face models internally; the AAM and CHM, which focus on the rigid and non-rigid natures, separately. The proposed 'AAM+CHM' tracking method increased the tracking rate and pose coverage by 170% and 115%, respectively, when compared to those of the 'AAM' tracking method when we used the database DATA-I and a person specific AAM. The improvements of the tracking rate and pose coverage were 112%, and 111%, respectively, when we used the database DATA-II and a generic AAM. The improvements were obtained by combining the CHM and AAM, where the CHM re-initialized the AAM effectively during the tracking process by providing a good estimate of the initial AAM parameters, and the AAM initialized the CHM precisely by providing estimated 3D rotation angles and the positions of facial feature points.

In the future, we will conduct more experiments considering multi-person data (Gross et al. 2004a), occlusion handling (Gross et al. 2004b) and so on. In addition, we expect that the proposed algorithm can be improved further by making the CHM and AAM interact in a bidirectional way during the tracking process. Currently, the proposed algorithm is designed for the CHM to help the AAM during the tracking process.

Acknowledgements It was financially supported by the Ministry of Education and Human Resources Development (MOE), the Ministry of Commerce, Industry and Energy (MOCIE) and the Ministry of Labor (MOLAB) through the fostering project of the Lab of Excellency. Also, it was partially supported by the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Commerce, Industry and Energy of Korea.

References

Aggarwal, G., Veeraraghavan, A., & Chellappa, R. (2005). 3D facial pose tracking in uncalibrated videos. In *International conference on pattern recognition and machine intelligence*.

Basu, S., Essa, I., & Pentland, A. (1996). Motion regularization for model-based head tracking. In *International conference on pattern recognition*.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Siggraph* (pp. 187–194).

Bregler, C., & Malik, J. (1998). Tracking people with twists and exponential maps. In *Proc. of IEEE conference on computer vision and pattern recognition*.

Cascia, M., Sclaroff, S., & Athitsos, V. (2000). Fast, reliable head tracking under varying illumination: an approach based on registration of texture-mapped 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(4), 322–336.

Cootes, T., Edwards, G., & Taylor, C. (2001). Active appearance models. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, *23*(6), 681–685.

DeCarlo, D., & Metaxas, D. (1996). The integration of optical flow and deformable models with applications to human face shape and motion estimation. In *Proc. of IEEE conference on computer vision and pattern recognition*.

Fidaleo, D., Medioni, G., Fua, P., & Lepetit, V. (2005). An investigation of model bias in 3D face tracking. In *IEEE analysis and modeling of faces and gestures* (pp. 125–139).

Gross, R., Matthews, I., & Baker, S. (2004a). Generic vs. person specific active appearance models. In *Proc. of the British machine vision conference*.

Gross, R., Matthews, I., & Baker, S. (2004b). Constructing and fitting active appearance models with occlusion. In *Proc. of IEE workshop on face processing in video*.

Kanade, T., & Lucas, B. (1981). An iterative image registration technique with an application to stereo vision. In *International joint conference on artificial intelligence* (Vol. 1, pp. 674–679).

Malciu, M., & Preteux, F. (2000). A robust model-based approach for 3D head tracking in video sequences. In *IEEE international conference on automatic face and gesture recognition*.

Matthews, I., & Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision*, *60*(2), 135–164.

Romdhani, S., Canterakis, N., & Vetter, T. (2003). *Selective vs. global recovery of rigid and non-rigid motion* (Technical Report). CS Dept., University of Basel.

Strom, J., Jebara, T., Basu, S., & Pentland, A. (1999). Real time tracking and modeling of faces: an EKF-based analysis by synthesis approach. In *Proc. of the modelling people workshop at ICCV*.

Vacchetti, L., Lepetit, V., & Fua, P. (2004). Stable real-time 3D tracking using online and offline information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(10), 1385–1391.

Xiao, J., & Kanade, T. (2002). Robust full-motion recovery of head by dynamic templates and registration techniques. In *Automatic face and gesture recognition*.

Xiao, J., Baker, S., Matthews, I., & Kanade, T. (2004a). Real-time combined 2D+3D active appearance models. In *International conference on computer vision and pattern recognition*.

Xiao, J., Chai, J., & Kanade, T. (2004b). A closed-form solution to non-rigid shape and motion recovery. In *European conference on computer vision*.