The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Robust Real-Time 3D Head Tracking Based on Online Illumination Modeling and Its Application to Face Recognition

Kwang Ho An and Myung Jin Chung, *Senior Member, IEEE*

Abstract—This paper investigates the estimation of 3D head poses and its identity authentication using a simple ellipsoid model. To achieve robust motion estimation even under time-varying lighting conditions, we incorporate illumination correction into the conventional 3D model-based tracking framework with a single camera. In addition, by computing the illumination bases online from the registered face images, after estimating the 3D head poses, user-specific illumination bases can be obtained, and therefore illumination-robust tracking without a prior learning process can be possible. Furthermore, our unified tracking is approximated as a linear least-squares problem; a closed-form solution is then provided. Therefore, it can be executed in real-time at 20 frames per second. After recovering full motion of the head, we can register face images with pose variations into stabilized (frontal) view images which are suitable for pose-robust face recognition. To verify the feasibility and applicability of our proposed 3D head-tracking framework, we performed extensive experiments with three sets of challenging image sequences.

I. Introduction

SMART environments of the future will interact with us more like humans. A key components of that interaction will be their abilities to detect, track, and recognize our faces and even our expressions. A human face provides a variety of different communicative functions such as identification and the perception of emotional expression. For this reason, face analysis has been carried out by many researchers over the past two decades in terms of detection, tracking, and recognition [1]-[6].

An accurate estimation of 3D head position and orientation is important in many applications. 3D head pose information can be used in human-computer interfaces (HCI), active telecommunication, virtual reality, and visual surveillance. In addition, a face image aligned in terms of the recovered head motion would facilitate face recognition and facial expression analysis.

Face recognition, as one of the primary biometric technologies, became more and more important owing to rapid advances in technologies such as digital cameras, the Internet and mobile devices, and increased demands on security. Face recognition has several advantages over other biometric technologies: It is natural, nonintrusive, and easy to use.

Manuscript received March 1, 2009.

K. H. An and M. J. Chung are with the Electrical Engineering and Computer Science Department, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea (e-mail: akh@cheonji.kaist.ac.kr, mjchung@ee.kaist.ac.kr).

Although significant work has been done in face recognition, the current systems are still not close to the human perceptual system. Traditionally, face recognition research has been limited to recognizing faces from still images. Most of these approaches discount the inherent 3D structure of the face and therefore deteriorate with changes in pose, illumination, and other disturbing factors, among which pose variation is the most difficult one to deal with [7]. Therefore, face registration (alignment) is the key of robust face recognition. If we can register face images into frontal views, the recognition task would be much easier. To align a face image into a frontal view, we need to know 3D pose information of a human head.

With consideration of all of these issues, in this paper, a unified 3D head pose estimation method using a simple ellipsoidal head model [8] is presented. Our unified tracking is considering online illumination correction, and thus provides a stable 3D motion estimation even under time-varying lighting conditions. After recovering full motion of the head, face images with pose variations can be registered into stabilized view images, which are suitable for frontal face recognition.

The remainder of the paper is organized as follows. Section II presents a unified 3D head pose estimation method including online illumination correction. Section III explains how to generate stabilized and mirrored texture maps, which are suitable for frontal face recognition, by using the unified tracking framework proposed in Section II. In Section IV, we provide extensive experimental results with three sets of challenging image sequences. Section V presents conclusions and discussions.

II. Unified 3D Head Pose Estimation

Generally, image-based tracking is based on the brightness change constraint equation (BCCE). The BCCE for image velocity estimation arises from the assumption that image intensity does not change from one frame to the next. However, this assumption does not hold true under conditions of varying illumination. Tracking based on the minimization of the sum of squared differences between the input and reference images is inherently susceptible to changes in illumination. Hence, we need to consider the effect of ambient illumination changes for stable tracking even under such circumstances.

$$\mathbf{I}_t \approx \mathbf{I}_{m,t} + \mathbf{I}_{i,t}. \quad (1)$$

We assume that image intensity changes arise from both

motion and illumination variations as shown in (1). $\mathbf{I}_t$ is image gradient with respect to time t, and both $\mathbf{I}_{m,t}$ and $\mathbf{I}_{i,t}$ are the instantaneous image intensity changes due to motion and illumination variations respectively.

A. Motion

First, we assume static ambient illumination and thus that instantaneous image intensity changes arise from variations in motion only. If then, the following BCCE holds true.

$$\begin{aligned} I(x,y,t) &= I(x+\Delta x, y+\Delta y, t+\Delta t) \\ &\approx I(x,y,t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t. \end{aligned} \tag{2}$$

$$\frac{\partial I}{\partial x}v_x + \frac{\partial I}{\partial y}v_y = -\frac{\partial I_m}{\partial t}, \tag{3}$$

where $v_x = dx/dt$ and $v_y = dy/dt$ are the x- and y-components of the 2D image velocity of object motion after projection onto the image plane. In addition, we replace $\partial I/\partial t$ with $\partial I_m/\partial t$ to denote that the intensity changes are due to motion variations.

$$\begin{bmatrix} I_x & I_y \end{bmatrix}\begin{bmatrix} v_x \\ v_y \end{bmatrix} = -I_{m,t}, \tag{4}$$

where I_x, I_y, and $I_{m,t}$ are the spatial and temporal derivatives of the image intensity computed at location $\mathbf{p} = \begin{bmatrix} x & y \end{bmatrix}^T$ respectively, where $I_{m,t}$ arise from the motion changes. However, we are interested in solving for 3D velocities of object points, which are related to 3D motion parameter estimation. Under the perspective projection camera model with focal length f, 2D image velocities can be related to 3D object velocities by the following equations.

$$\begin{aligned} v_x &= \frac{dx}{dt} = \frac{d}{dt}\left(f\frac{X}{Z}\right) = \frac{f}{Z}V_X - \frac{x}{Z}V_Z, \\ v_y &= \frac{dy}{dt} = \frac{d}{dt}\left(f\frac{Y}{Z}\right) = \frac{f}{Z}V_Y - \frac{y}{Z}V_Z, \end{aligned} \tag{5}$$

where $\mathbf{V} = \begin{bmatrix} V_X & V_Y & V_Z \end{bmatrix}^T$ is the 3D velocity of a point $\mathbf{P} = \begin{bmatrix} X & Y & Z \end{bmatrix}^T$, corresponding to the image pixel $\mathbf{p}$, in the camera coordinate frame. The relationship between the two corresponding velocities can be expressed in compact matrix form as shown below.

$$\mathbf{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \frac{1}{Z}\begin{bmatrix} f & 0 & -x \\ 0 & f & -y \end{bmatrix}\mathbf{V}. \tag{6}$$

Any rigid body motion can be expressed in terms of the instantaneous rotations and translation of the object. For small inter-frame rotations, the rotation matrix can be linearly approximated as ($\Delta\mathbf{R} \approx \mathbf{I} + \left[\Delta\mathbf{r}\right]_\times$). $\mathbf{I}$ is an 3×3 identity matrix, and $[]_\times$ denotes a skew-symmetric matrix. Also, assuming that time interval Δt is unity, temporal derivatives of rotation and translation vectors can be approximated by finite differences $\Delta\mathbf{r}$, $\Delta\mathbf{t}$ respectively.

$$\mathbf{V} \approx \mathbf{R}\begin{bmatrix} \mathbf{I} & -\left[\mathbf{P}_o\right]_\times \end{bmatrix}\begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{r} \end{bmatrix}, \tag{7}$$

where $\mathbf{P}_o$ is a 3D sampled model point in the object coordinate frame corresponding to the point $\mathbf{P}$ in the camera reference frame. $\mathbf{R}$ is the rotation matrix computed in the previous frame between the camera and object coordinate frames. $\Delta\mathbf{r}$ and $\Delta\mathbf{t}$ are the inter-frame rotation and translation vectors expressed in the object coordinate frame, respectively. The above equation describes the relationship between the 3D object velocity in the camera coordinate frame and inter-frame rigid body motion parameters in the object coordinate frame. Substituting (6) and (7) into (4), we obtain a simple linear equation as shown below.

$$\frac{1}{Z}\begin{bmatrix} fI_x & fI_y & -\left(xI_x + yI_y\right) \end{bmatrix}\mathbf{R}\begin{bmatrix} \mathbf{I} & -\left[\mathbf{P}_o\right]_\times \end{bmatrix}\begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{r} \end{bmatrix} = -I_{m,t}. \tag{8}$$

The above single linear equation relates the spatial and temporal image intensity derivatives to rigid body motion parameters under the perspective projection model at a single pixel. Because (8) is linear with respect to motion parameters, we can combine it across n pixels by stacking the equations in matrix form. n is the number of model points that can be seen from the camera under the current estimated head pose.

$$\begin{pmatrix} \frac{1}{Z_1}\begin{bmatrix} fI_{x,1} & fI_{y,1} & -\left(x_1I_{x,1} + y_1I_{y,1}\right) \end{bmatrix}\mathbf{R}\begin{bmatrix} \mathbf{I} & -\left[\mathbf{P}_{o,1}\right]_\times \end{bmatrix} \\ \vdots \\ \frac{1}{Z_n}\begin{bmatrix} fI_{x,n} & fI_{y,n} & -\left(x_nI_{x,n} + y_nI_{y,n}\right) \end{bmatrix}\mathbf{R}\begin{bmatrix} \mathbf{I} & -\left[\mathbf{P}_{o,n}\right]_\times \end{bmatrix} \end{pmatrix}\begin{pmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{r} \end{pmatrix} = \begin{pmatrix} -I_{m,t,1} \\ \vdots \\ -I_{m,t,n} \end{pmatrix}. \tag{9}$$

Let the left-hand side of (9) be $\mathbf{M}$ and the right-hand side be $\mathbf{I}_{m,t}$. Then, (9) can be represented in compact matrix form as shown below.

$$\mathbf{M}\boldsymbol{\alpha} = \mathbf{I}_{m,t},\ \boldsymbol{\alpha} = \begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{r} \end{bmatrix}. \tag{10}$$

B. Illumination

As mentioned in the beginning of Section II, BCCE does not hold true under time-varying illumination conditions. To handle face image variations due to changes in lighting conditions, many methods have been proposed in the field of face recognition thus far. Among them, for modeling illumination variations, subspace-based methods have often been used [9]. These kinds of methods model the face image variations due to illumination changes with a low-dimensional linear subspace. They approximate the intensity changes due to illumination variations as a linear combination of illumination bases that are obtained from the training samples of different people taken under a wide variety of lighting conditions.

However, these kinds of subspace-based methods construct an illumination subspace from training images for different people, which includes not only illumination conditions but also face identities. This subspace is not capable of representing the lighting conditions uniquely,

because the intrinsic (facial geometry and albedo) and the extrinsic (illumination conditions) information is mixed. Otherwise, extremely large training sets would be needed. Furthermore, these methods need a prior training process and thus suffer from the cost of training data acquisition and processing.

Hence, in this paper, by computing these illumination bases online from the registered face images, after estimating the head poses, user-specific illumination bases can be obtained, and therefore illumination-robust tracking without a prior learning process can be possible as shown in Fig. 1. Therefore, we can approximate the intensity changes due to illumination variations as a linear combination of illumination bases obtained through online illumination modeling based on principal component analysis (PCA) as shown below.

$$\mathbf{I}_{i,t} = \mathbf{L}\boldsymbol{\beta}, \tag{11}$$

where $\mathbf{I}_{i,t}$ is the instantaneous image intensity changes due to illumination variations. The columns of the matrix $\mathbf{L} = [\mathbf{l}_1, \cdots, \mathbf{l}_k]$ are the illumination bases obtained by PCA, and $\boldsymbol{\beta}$ is the illumination coefficient vector. k is the number of principal components.

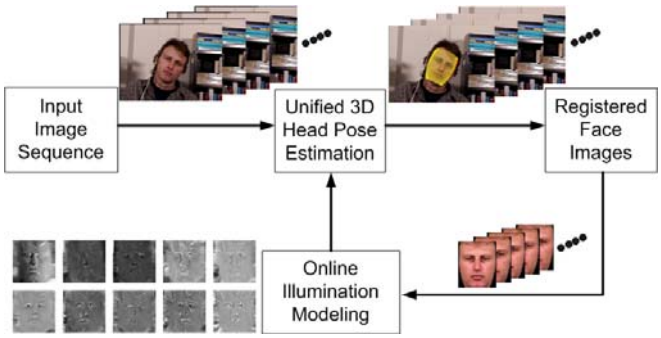

Fig. 1. Online illumination modeling.

C. Combined into a Unified Framework

Because we assumed (1) in the beginning of Section II, and because (10) and (11) are linear with respect to motion parameter $\boldsymbol{\alpha}$ and illumination coefficient vector $\boldsymbol{\beta}$ respectively, we can combine them into a unified framework as shown below.

$$\begin{bmatrix}\mathbf{M} & \mathbf{L}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha} \\ \boldsymbol{\beta}\end{bmatrix} = \mathbf{I}_t. \tag{12}$$

Let the left-hand side of (12) be $\mathbf{A}$ and the right-hand side be $\mathbf{b}$. Then, the least-squares solution of (12) can be easily obtained as shown below.

$$\mathbf{s}^* = \arg\min_{\mathbf{s}} \|\mathbf{A}\mathbf{s} - \mathbf{b}\|^2 = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}. \tag{13}$$

Due to the presence of noise, non-rigid motion, occlusion, and projection density, some pixels in the face image may contribute less to motion estimation than others may. To account for these errors, the pixel should be weighted by their contributions. If then, a weighted least-squares solution can be obtained as shown below.

$$\mathbf{s}^* = \left((\mathbf{W}\mathbf{A})^T(\mathbf{W}\mathbf{A})\right)^{-1}(\mathbf{W}\mathbf{A})^T(\mathbf{W}\mathbf{b}), \tag{14}$$

where $\mathbf{W}$ is a diagonal matrix whose components are pixel weights assigned according to their projection densities as in [6]. Finally, motion parameters between the object and camera coordinate frames are updated by (15) and iterated until the estimates of the parameters converge. Initial motion parameters are assumed to be known.

$$\mathbf{R} \leftarrow \mathbf{R}\Delta\mathbf{R},\ \mathbf{T} \leftarrow \mathbf{R}\Delta\mathbf{t} + \mathbf{T}. \tag{15}$$

III. Face Recognition

As mentioned in Section I, if we can align the face images with pose variations into canonical frontal views, the recognition task would be much easier, and higher recognition rate can be achieved.

Fig. 2 presents how to obtain a pose-compensated face image when given 3D pose information of the head under perspective projection. The general idea of stabilization is as follows. First, a pixel in the stabilized texture map corresponds to one point on the surface of a 3D ellipsoid in the object coordinate frame. Second, we can estimate the current pose of the ellipsoid corresponding to a human head using the proposed unified motion estimation technique. If so, then we can project all surface points of the ellipsoid onto the input image plane under the perspective projection model. Following this procedure, we can find out the complete relationship between an arbitrary input face image and its corresponding stabilized texture map.

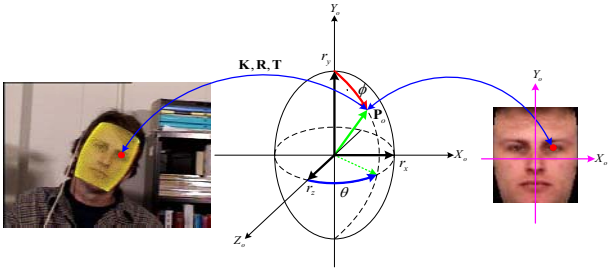

Fig. 2. Geometrical mapping from an input face image to its stabilized texture map under the estimated head pose and perspective projection. $\mathbf{K}$ is the camera calibration matrix (represents intrinsic parameters) and is assumed to be known. $\mathbf{R}$ and $\mathbf{T}$ are the estimated rotation matrix and the translation vector of current head pose, respectively.

However, there might be missing pixels in the stabilized texture map, which correspond to invisible regions from the camera due to self-occlusion and camera's viewing direction. Therefore, such invisible regions are considered as missing pixels, and their intensities are set to be zeros in the stabilized texture map. This may deteriorate the recognition performance. However, fortunately, we know that a human face has a symmetric property around its vertical axis. Therefore, we can also generate a mirrored texture map through a simple mirror operation on the stabilized texture map around its vertical axis as shown in Fig. 3. By doing so,

we can make up for the missing pixels and improve the recognition performance.

Finally, simple and efficient frontal face recognition can be easily carried out in the stabilized (or mirrored) texture map space, which is nearly linear-separable, instead of the original input image space that is highly nonlinear and complex.

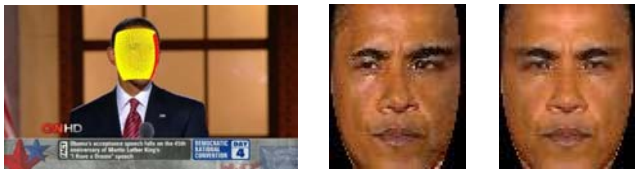

Fig. 3. Mirrored texture map generation. An input image (taken from CNN news) and its stabilized and mirrored texture maps are shown from left to right, respectively.

IV. Experimental Results

To verify the feasibility and applicability of our proposed 3D head-tracking framework, we performed extensive experiments with three sets of challenging image sequences. All the three experiment sets of image sequences were collected with a vision module named "Bumblebee". All the image sequences were digitized at 30 frames per second at a resolution of 320×240. Ground truth data for the first and second sets was simultaneously collected via a 3D magnetic sensor named "Flock of Birds". The magnetic sensor has a positional accuracy of 2.54*mm* and rotational accuracy of 0.5°. The first set consists of 20 image sequences (two sequences for each of 10 subjects) taken under near-uniform illumination conditions. The second set consists of 20 image sequences (two sequences for each of 10 subjects) taken under time-varying illumination. All the sequences in the first and second sets are 300 frames long and are including free and large head motions. The third set was collected for face recognition test and consists of 17 image sequences (16 males + 1 female) taken under near-uniform lighting conditions. All the sequences in this set are 200 frames long and are also including free and large head motions.

Note that all the measured ground truth and the estimates of the visual tracking are expressed with respect to the initial object (head) coordinate frame for the comparison of estimation errors.

A. Experiment 1: Near-Uniform Illumination

The first experiment was designed to compare the performance of the proposed tracker including online illumination correction with that of a conventional optical flow-based tracker and also intended to evaluate the effects of online illumination correction. In this experiment, for modeling the illumination changes in face images, we used 10 illumination bases. They were obtained through online illumination modeling based on PCA from the registered face images that had been stored until the previous frame.

Fig. 4 presents typical tracking results on one of the test sequences from the first experiment set. The estimations for 3D motion on this sequence are displayed in Fig. 5. This sequence involves large pitch, yaw, and roll motions up to 26°, 60°, and 33° respectively. On the average, our proposed tracker shows a translational error of 6.91*mm* and rotational error of 2.23° on this sequence. On the other hand, conventional tracking has a translational error of 49.03*mm* and rotational error of 7.89° on this image sequence.

Fig. 4. Typical tracking results on one of the sequences taken under near-uniform illumination. Frames 1, 55, 112, 164, 193, and 300 are shown (left to right, top to bottom). Rows 1~2: conventional method without illumination correction; Rows 3~4: our proposed unified method including illumination correction.

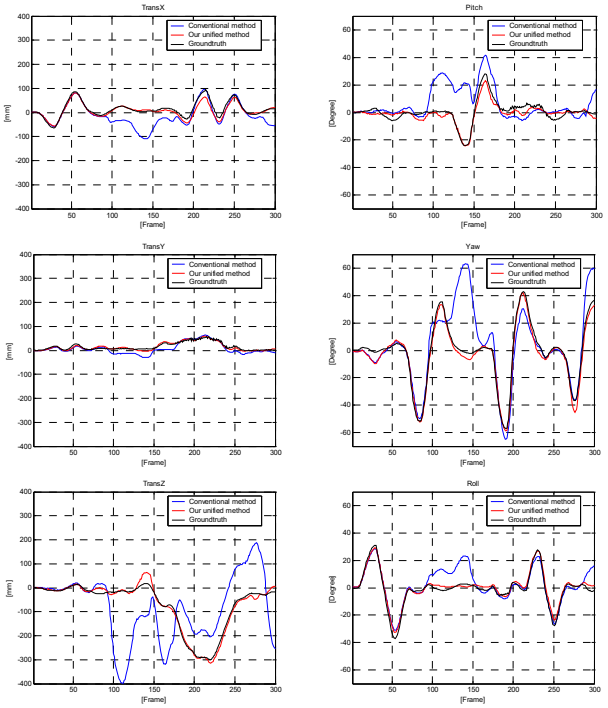

Fig. 5. Comparison between the ground truth and the estimated head poses on the sequence corresponding to Fig. 4. Red line: our proposed method; Blue line: conventional method; Black line: the ground truth.

TABLE I
MOTION ESTIMATION ERRORS ON 20 IMAGE SEQUENCES TAKEN UNDER NEAR-UNIFORM ILLUMINATION CONDITIONS

	Existing method	Our method
Trans. X [mm]	11.27	5.98
Trans. Y [mm]	9.65	4.36
Trans. Z [mm]	66.61	19.49
Pitch [degree]	5.46	2.49
Yaw [degree]	6.08	4.17
Roll [degree]	2.54	1.76

Average errors of 3D motion estimation on 20 image sequences are shown in Table I. As can be seen in this table, conventional single camera-based tracking is not robust to large out-of-plane rotations (especially for pitch and yaw) and translation in depth direction. On the other hand, even though there are no changes in ambient illumination, motion estimation is greatly improved through the proposed unified tracking including online illumination correction. This is because self-shading is likely to occur in face images even under uniform illumination, depending on the current head pose. Hence, our proposed unified tracking can provide robust motion estimation by reducing the negative effects of self-shading thanks to the illumination correction term.

B. *Experiment 2: Time-Varying Illumination*

The second experiment was set up to evaluate the performance of the proposed tracker under time-varying illumination conditions. In this experiment, we also used 10 illumination bases obtained through online illumination modeling based on PCA.

Fig. 6 presents typical tracking results on one of the test sequences from the second experiment set. The estimations for 3D head motion on this sequence are displayed in Fig. 7. This sequence involves large pitch, yaw, and roll motions up to 21°, 60°, and 35° respectively. Whenever there are changes in illumination, significant tracking errors occur in conventional single camera-based tracking. On the other hand, our method shows stable tracking even under time-varying illumination. On the average, our proposed tracking shows a translational error of $14.31mm$ and rotational error of 3.25° on this sequence. On the other hand, conventional tracking has a translational error of $64.30mm$ and rotational error of 11.62° on this image sequence.

TABLE II
MOTION ESTIMATION ERRORS ON 20 IMAGE SEQUENCES TAKEN UNDER TIME-VARYING ILLUMINATION CONDITIONS

	Existing method	Our method
Trans. X [mm]	18.85	6.92
Trans. Y [mm]	16.02	5.51
Trans. Z [mm]	112.37	30.10
Pitch [degree]	9.91	3.84
Yaw [degree]	18.89	4.30
Roll [degree]	6.86	2.12

Average errors of 3D motion estimation on 20 image sequences are shown in Table II. As can be seen in this table, there exist significant tracking errors in conventional tracking, because it cannot cope with illumination changes. On the other hand, our method shows slightly deteriorated but almost similar performance of motion estimation to that evaluated in Experiment 1 even under time-varying lighting conditions, thanks to the illumination correction term.

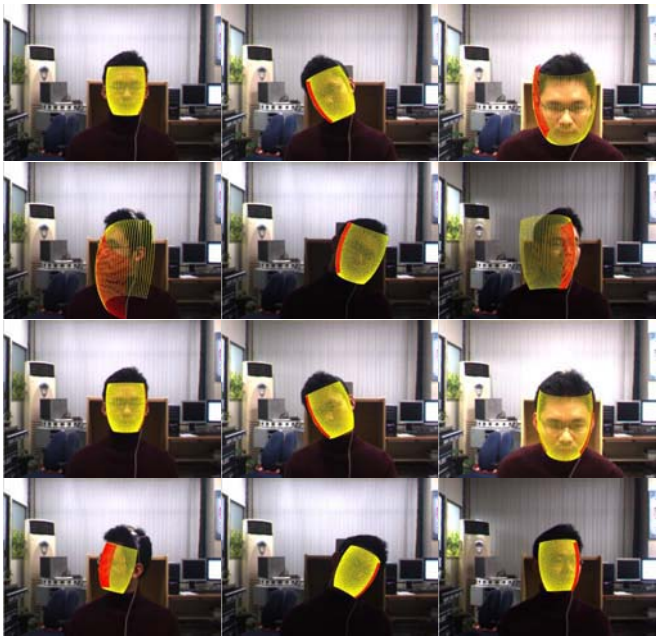

Fig. 6. Typical tracking results on one of the sequences taken under time-varying illumination. Frames 1, 77, 127, 172, 197, and 285 are shown (left to right, top to bottom). Rows 1~2: conventional method without illumination correction; Rows 3~4: our proposed unified method including illumination correction.

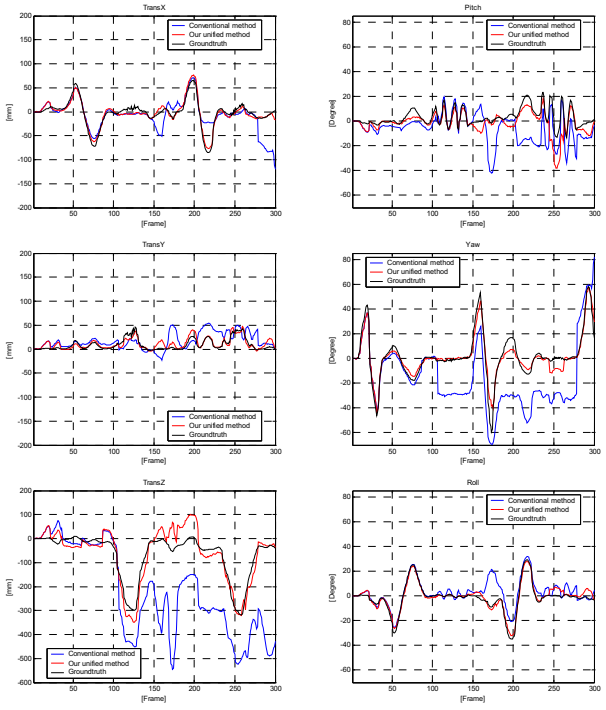

Fig. 7. Comparison between the ground truth and the estimated head poses on the sequence corresponding to Fig. 6. Red line: our proposed method; Blue line: conventional method; Black line: the ground truth.

C. *Experiment 3: Face Recognition*

The third experiment was intended to verify that our

proposed head tracking method is helpful to improve the performance of face recognition. In this experiment, we constructed three test sets such as unregistered, stabilized, and mirrored sets. For the unregistered test set, we manually cropped 200 face images from the input image sequence for each of 17 classes. For the stabilized set, we obtained 200 face images, registered into frontal views by the proposed unified tracker, for each class. For the mirrored set, we made a mirror operation on the stabilized test set. For face recognition on each test set, we used only a single frontal face image for the training and 200 face images, obtained by the aforementioned methods, for the test from each class. We performed PCA-based face recognition for each test set [10].

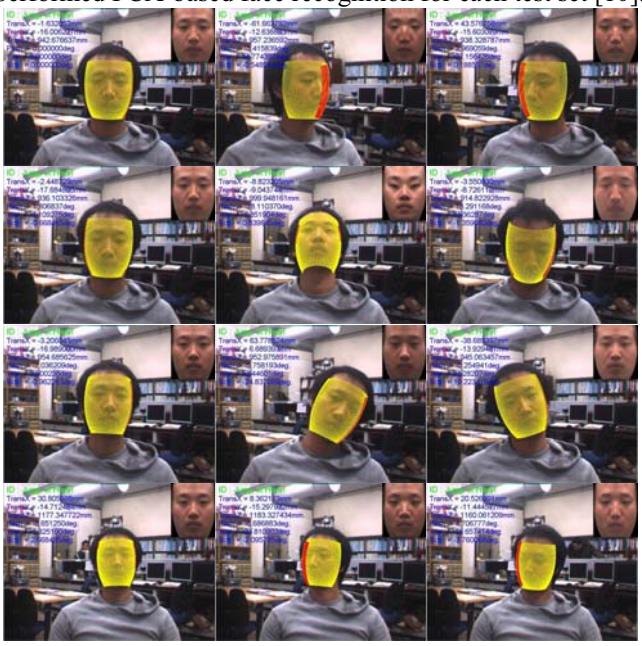

Fig. 8. 3D head pose estimation and 2D texture map-based face recognition on one of the image sequences from the third experiment set. Frames 1, 23, 47, 55, 66, 80, 90, 107, 116, 145, 152, and 186 are shown (left to right, top to bottom).

TABLE III
PERFORMANCE OF PCA-BASED FACE RECOGNITION ON OUR LABORATORY TEST SETS

	Unregistered	Stabilized	Mirrored
Recognition rate	55.15%	83.74%	90.41%

Fig. 8 shows the typical results of 3D head pose estimation and 2D texture map-based face recognition on one of the image sequences from the third experiment set. Our proposed unified tracking including online illumination correction was used for 3D head pose estimation, and PCA-based face recognition was performed on the mirrored texture maps. 3D head pose estimates and face recognition results are displayed in the top-left corner of result images. Mirrored view images generated by pose estimation, texture mapping, and mirror operation are displayed in the top-right corner of result images. They were used for face recognition, and this sequence showed 100% recognition rate. Table III shows the recognition rates with PCA-based classification for each test set. As can be seen in the recognition rates, we can verify that face registration is helpful to improve the recognition performance, and also the recognition rate using mirrored texture maps is much better.

V. CONCLUSION

In this paper, we presented a long-term stable and robust technique for 3D head tracking even in the presence of varying illumination conditions. We incorporated the online illumination correction term into conventional 3D tracking with a single camera. This enables us to cope with large out-of-plane rotations and depth-motion even under time-varying illumination conditions. We approximated the intensity changes due to illumination variations as a linear combination of illumination bases. In addition, by computing these illumination bases online from the registered face images, after estimating the head pose, user-specific illumination bases can be obtained, and finally illumination-robust tracking, without a prior learning process that needs a great cost of training data acquisition and processing, can be possible.

This paper has shown the feasibility and applicability of the proposed approach by carrying out various challenging experiments. First, it was verified that the proposed unified tracking method is able to cope with fast and large out-of-plane rotations and depth-motion. This is true even under time-varying illumination conditions. Second, it was proved that our proposed unified head tracking method is helpful to improve the performance of face recognition (over 90% recognition rate when using mirrored texture maps). Finally, our proposed method has shown computational efficiency and has achieved 20 frames/sec. processing speed.

REFERENCES

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Int'l Conf. Computer Vision and Pattern Recognition*, 2001.

[2] G. D. Hager and P. N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 10, pp. 1025-1039, 1998.

[3] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminant common vectors for face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 1, 2005.

[4] X. Lu, A. K. Jain, and D. Colby, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 1, 2006.

[5] M. L. Cascia and S. Sclaroff, "Fast, reliable head tracking under varying illumination," *Int'l Conf. Computer Vision and Pattern Recognition*, 1999.

[6] J. Xiao, T. Kanade, and J. F. Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," *Int'l Conf. Automatic Face and Gesture Recognition*, 2002.

[7] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, "FRVT 2002: Evaluation Report," Mar. 2003.

[8] K. H. An and M. J. Chung, "3D head tracking and pose-robust 2D texture map-based face recognition using a simple ellipsoid model," *Int'l Conf. Intelligent Robots and Systems*, pp. 307-312, 2008.

[9] P. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions," *Int'l Conf. Computer Vision and Pattern Recognition*, pp. 995-999, 1994.

[10] M. Turk and A. P. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.