Development of Realistic and Real-Time Virtual Mirror Application for Electronic Cosmetic System

Hayato Kinoshita[*a] *,Takao Makino*[a], *Norimichi Tsumura*[a], *Toshiya Nakaguchi*[a], *and Yoichi Miyake*[b]

[a]*Department of Information and Image Sciences, Chiba University,*

Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan

[b]*Research Center for Frontier Medical Engineering, Chiba University,*

Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan

Abstract

In this paper, we present a virtual mirror system tracking a 3D position and orientation of a user's head in real-time. This system can exhibit a high-realistic color change of the human face. The "virtual mirror" is a system that can reproduce a mirror image of user's face onto the display by using a video camera and a half mirror. Unlike a normal mirror, virtual mirror has an advantage that various image processing can be performed to the video stream. Therefore, it is thought that the virtual mirror can be applied to various applications, for example, video chatting and electronic cosmetic and so on. Our goal is to create the electronic cosmetic system with the virtual mirror. The 3D head position and orientation are tracked to control the skin color on appropriate position in the face.

1 Introduction

Human face gives various kinds of information such as emotion, age, health condition and so on. Therefore it is important to acquire and reconstruct the facial pattern precisely for various kinds of applications. Our goal is to create the electronic cosmetics systems. A virtual mirror system for electronic cosmetics has been used with a half mirror as shown in Figure 1[1]. A camera of this system captures the facial image of the user. A half mirror is set in front of the display, and the image is captured as the image reflected by the half mirror. The image is displayed on the monitor from behind the half mirror. In this system, the user can look into his or her eyes like a real mirror. This is a high advantage compared to the camera attached on the display, since the user can not look into his or her eyes in this conventional camera system. Comparing to a real mirror, the virtual mirror is capable of adding many effects on the image[1].

In this virtual mirror system, high-precision face tracking technique is necessary to perform the image processing on the face. Various methods are proposed for tracking user's face in video image, such as methods of using templates[3], color information[4], face's feature points[5], 3D face model[6], and so on. Recent advanced tracking method using 3D face model is based on particle filter[7](one of time-line filters). This method is robust against noise and partial occlusion in the input image. Oka et al. or Dornaika et al.[8][10] proposed the method which is adaptive to the changing of facial expression by using the particle filter. In addition, the method of Oka et al.[9][10] is adaptive to the user's quick movement. However, their techniques have following problems. The problem of Oka et al.[9][10] is that their research obtained a 3D face model by two cameras. Therefore, it is necessary to perform the calibration between two cameras. The other problem is that the method by Oka at el. does not consider the situation that the user

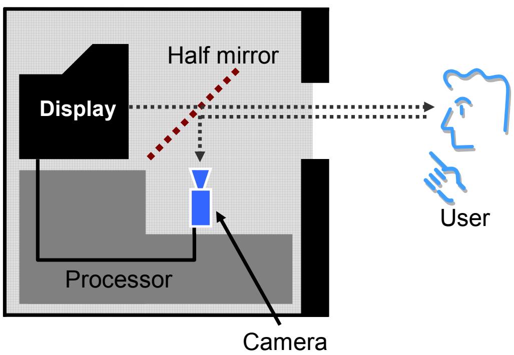

Figure 1. Typical virtual mirror interface

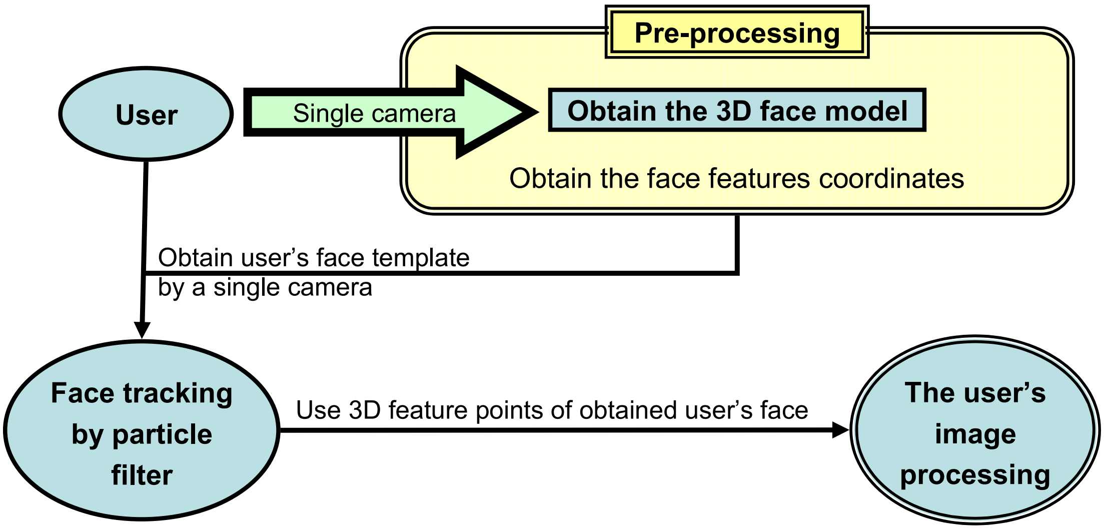

Figure 2. Block diagram of virtual mirror interface

changes the facial pose and expression at the same time. The problem of Dornaika's method[8] is that it must measure the 3D face model with respect to each user previously.

In this paper, we use a single camera to track the face movement and make the virtual mirror system for applying electronic cosmetic on the face. Since, we just use a single camera in the system, it is not necessary to perform calibration between two cameras. The block diagram of this interface is shown in Figure 2. A standard facial 3D model is used instead of the 3D model for each user. The particle filter is used for the face tracking in this system. In our work, tracking process is divided into three steps. In this process, we consider that the user move the facial pose and expression at the same time because the user's movement is not considered in previous method. Additionally, the image processing on the face is performed by 3D feature points of the model used in tracking. Using these processes, we make more practical virtual mirror system for electronic cosmetics. This paper describes the details of our proposed system and demonstrates its performance improvement in the above processes.

We describe particle filter in Section 2, and in Section 3 describe face tracking with particle filter, and experiment in Section 4, conclusion in Section 5.

2 Particle filter

The particle filter represents the probability density function (PDF) of a state as a set of many discrete samples. Each sample has the corresponding

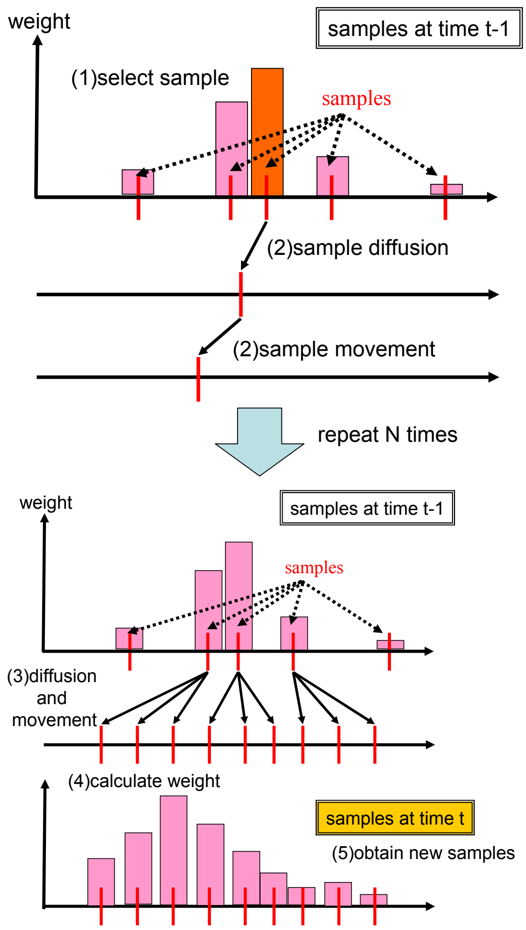

Figure 3. Update procedure of particle filter

weight. Therefore, this sample set can approximate an arbitrary PDF including non-Gaussian PDF. The particle filter is one of time-line filters. The state at frame t is obtained by the state at frame t-1.

Denote the state vector is $\boldsymbol{x}_t$, the observation vector obtained from the input image is $\boldsymbol{y}_t$, and the history of $\boldsymbol{y}_t$ is $Y_t=\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t\}$. The state at the time t

is obtained by estimating PDF p($\boldsymbol{x}_t$|Y$_t$). In the case of particle filter, PDF p($\boldsymbol{x}_t$|Y$_t$) is expressed as N samples $\{(\boldsymbol{s}_t^{(i)};\boldsymbol{\pi}_t^{(i)})\}(i=0,\ldots,N)$. Where, $\boldsymbol{s}_t^{(i)}$ is N discrete samples within state space S, and $\boldsymbol{\pi}_t^{(i)}$ is the weight of $\boldsymbol{s}_t^{(i)}$. Based on these samples, p($\boldsymbol{x}_t$|Y$_t$) can be approximated arbitrary discrete non-Gaussian PDF.

The flow of update procedure in particle filter is shown in Figure 3. Movement model using in the filtering is consisted of "movement" and "diffusion". "Movement" is to move samples by preconcerted method. "Diffusion" is to add random noise to moved samples. Update procedure of particle filter is described as below. A sample $\boldsymbol{s'}$ is selected by weight $\boldsymbol{\pi}_{t-1}^{(i)}$ in all samples $\boldsymbol{s}_{t-1}^{(i)}$. After selecting the sample $\boldsymbol{s'}$ based on PDF, the movement model determined in advance is applied to $\boldsymbol{s'}$. After repeating these processes for N times, the new samples $\boldsymbol{s}_t^{(i)}$ are generated. Then the weight $\boldsymbol{\pi}_t^{(i)}$ is calculated for each sample. In this work, template matching is used to determine each weight $\boldsymbol{\pi}_t^{(i)}$. This method is applied to all samples and is capable of obtaining the new samples sets $\{(\boldsymbol{s}_t^{(i)};\boldsymbol{\pi}_t^{(i)})\}$. These processes are performed at every frame t.

The problem about particle filter is the calculation cost due to the number of samples enough to keep accuracy. As the dimension of the state vector increases, the number of samples enough to keep calculation accuracy grows exponentially. Therefore, in using particle filter, we need algorithm to reduce the number of samples without reducing calculation accuracy. In our method, the next frame estimation in the tracking algorithm is modified to by dividing the all process into 3 steps. The process will be described in detail in Section 3.3.

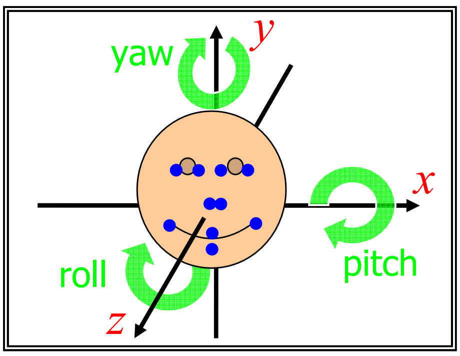

Figure 4. Coordinate system face model

3 Face Tracking

3.1 Face model

Particle filter requires a 3D face model to perform high accuracy tracking. In our method, the face model has K feature points. The face model coordinate system is shown in Figure 4. Let M be the shape vector that consists of 3D coordinates of K feature points in the model coordinate system. T defined as the image template sets. In this paper, K is set to 10 to represent these feature points: the inner and the outer corners of both eyes, both corners of the mouth, upper lip, lower lip, and both nostrils. Each feature point is also shown as blue dots in Figure 4.

We can obtain 3D coordinates of K feature points by using a standard facial 3D model. The standard 3D facial model has the 3D coordinate of a replica of the face. We apply the z-coordinate of the standard facial 3D model as the z-coordinate of K feature points. x and y-coordinates of K feature points are obtained manually from the input face image. It is thought that because the z-coordinates of feature points have low personal difference, this 3D coordinate is available to the generally persons. By this method, we are able to obtain z-coordinate of the feature points with a single camera. When we appoint feature points to acquire the z-coordinate, the template images of feature points are obtained at the same time. The size of images is 16x16.

3.2 Flow of process in the face tracking

In the tracking process, we estimate 3 dimension face pose and expression changing with particle filter by the face model and the input image. Oka et al. or Dornaika et al. have reported that face tracking with particle filter[9][8]. We aim to improve their method. The flow improved tracking process is shown in Figure 5. During tracking we produce successive estimation of a (6+B) dimensional state vector for each image frame t.

$$x_t = \begin{pmatrix} p_t \\ a_t \end{pmatrix} \tag{3.1}$$

In this paper, $\boldsymbol{p}_t$ is the translation and the rotation from the world coordinate system to the model coordinate system. Such state vector $\boldsymbol{p_t}$ shows by following equation:

$$p_t = (x_t, y_t, z_t, roll_t, pitch_t, yaw_t)^T \quad (3.2)$$

It is defined that $\boldsymbol{a_t}$ is a B-dimensional vector of face expression changing, B is set to 5 without loss of generality. We have chosen the five vectors: such as both lip corner depressor, lower lip raiser or drop.

The main flow of our estimation method is shown in Figure.5. First, we generate N new samples $\boldsymbol{s}_t^{(i)}(i=1\ldots N)$ based on the sample group $\{(\boldsymbol{s}_{t-1}^{(i)};\boldsymbol{\pi}_{t-1}^{(i)})\}$. The movement model is obtained by repeating the following process N times. A base sample $\boldsymbol{s'}_{t-1}$ we choose from $\{(\boldsymbol{s}_{t-1}^{(i)};\boldsymbol{\pi}_{t-1}^{(i)})\}$ is based on the probability proportional to the weight $\boldsymbol{\pi}_{t-1}$. Next, the chosen sample $\boldsymbol{s'}_{t-1}$ is drifted into $\boldsymbol{s}_t^{(i)}$ by assuming fixed motion of a user's face between two successive image frames as:

$$\mathbf{s}_t = \mathbf{s}'_{t-1} + \tau\mathbf{v}_{t-1} + \omega \quad (3.3)$$

where τ is the time interval between frames, $\boldsymbol{v}_{t-1}$ is the previous velocity of the face movement, and $\boldsymbol{\omega}$ is a

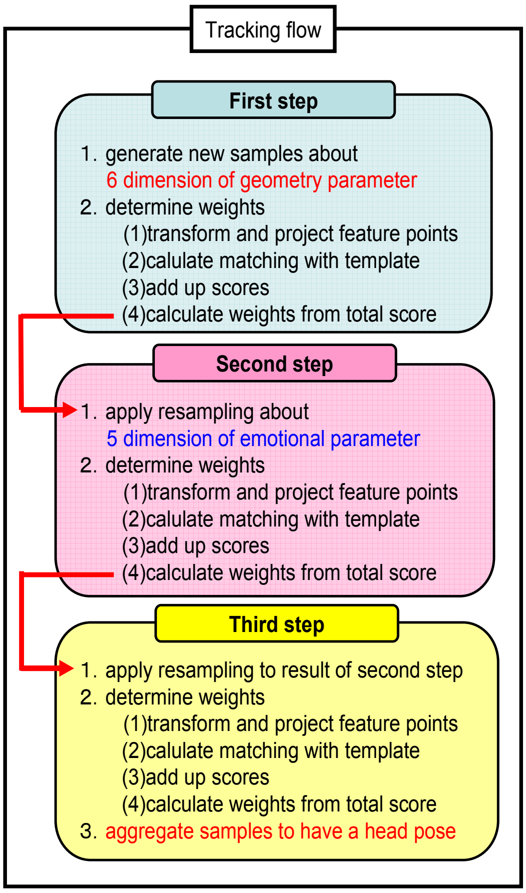

Figure 5. Tracking flow

system noise added to $\boldsymbol{s'}_{t-1}$. $\boldsymbol{v}_{t-1}$ is calculated at the end of this tracking step.

After obtaining new samples $\boldsymbol{s}_t^{(i)}$, we compute the weight $\boldsymbol{\pi}_t^{(i)}$ for $\boldsymbol{s}_t^{(i)}$. The evaluation is based on the current input image. The method to determine $\boldsymbol{\pi}_t^{(i)}$ is four levels from (1) to (4) as shown in Figure 5. First, using a obtained sample $\boldsymbol{s}_t^{(i)}$, we can change the face model depending on the pose represented by $\boldsymbol{s}_t^{(i)}$. The feature points M of the transformed face model is projected to the points $\boldsymbol{m}$ on the image plane by the projection function. Then, a matching score is computed between the neighboring region of $\boldsymbol{m}$. We calculate corresponding template T by normalized correlation-based function N(T, $\boldsymbol{m}$). For each sample $\boldsymbol{s}_t^{(i)}$, we apply the normalized correlation function N to K projected image points $\boldsymbol{m}$, and those results are added up to obtain a total score $c_t^{(i)}$. Finally, the weight $\boldsymbol{\pi}_t^{(i)}$ is calculated from the total score $c_t^{(i)}$ by using Gaussian function.

$$c_t^{(i)} = \sum_{k=1}^{K} \mathbf{N}\left(\mathbf{T}'_k, m_{t,k}^{(i)}\right) \quad (3.4)$$

$$\pi_t^{(i)} \propto \exp\left(-\frac{(K - c_t^{(i)})}{\sigma^2}\right) \quad (3.5)$$

where σ is the standard deviation of Gaussian function. In this paper, σ is set to 3.0 experimentally. All weights $\boldsymbol{\pi}_t^{(i)}$ are normalized to become the sum of $\boldsymbol{\pi}_t^{(i)}$ is 1.

3.3 Tracking by three steps

Previous works performed tracking of the facial pose and expression. If the facial pose and expression are changing at same time, previously proposed methods are not responded to these changing. Therefore, seriously problem in tracking is occurred by this situation. For this purpose, we propose the three steps method to respond to this changing of facial pose and expression. Using 3 steps, the facial pose and expression can be considered separately. In this section, as Figure.5, we mention about the three steps of the tracking method. Particulars about the tracking method are described in section 3.2. Each step is explained as below.

In the first step, the tracking is carried out in 6-dimension of the facial position and direction parameter. The estimation from frame *t-1* to frame *t* is carried out as shown in section 3.2. Especially, in terms of eq.(3.3) used in this step, noise vector $\boldsymbol{\omega}$ is controlled for each parameter. Because it is highly possible that each value of x, pitch and yaw change suddenly, the value of $\boldsymbol{\omega}$ relatively becomes high. On the other hand, about y, z and roll, the value of $\boldsymbol{\omega}$ relatively become low number.

In the second step, the tracking is carried out in B-dimension of expression parameter. The result of the first step is applied to resampling of the samples. The estimation is calculated as shown section 3.2. In the same procedure, in terms of eq.(3.3) used in this step, noise vector $\boldsymbol{\omega}$ is controlled for each parameter. Because it is highly possible that each value of lower lip change suddenly, the value of $\boldsymbol{\omega}$ relatively becomes high number. On the other hand, in terms of both corners of the mouth, the value of $\boldsymbol{\omega}$ relatively becomes low number.

In the third step, the tracking is carried out in (6+*B*) dimension of all parameter. In the same procedure, the result of the second step is applied to resampling of the samples by the input image at frame *t*. When the samples are resampled with Eq(3.3), the previous velocity of the pose $\boldsymbol{v}_{t-1}$ is set to 0 since the samples are adjusted in this step. Therefore, the new samples become ones weighted in the neighborhood with a high degree of accuracy. Then, we estimate the face model from weighted samples at frame *t*.

In addition, since the user's face is tracked by the feature points, the points of 3D coordinate information are used in face image processing. By this processing, we make effects on the face for electronic cosmetic system.

4 Experiments

We performed experiments using our proposed tracking method, and showed the effects for electronic cosmetic. Our PC consists of a Windows XP based OS with Intel PentiumD 2.8GHz processor and 1GB memory. Figure 6 shows our proposed virtual mirror system. To avoid incoming extra lights in this system and reflecting environmental background, the system is surrounded by black boards and black curtain. A single camera we used is USB2.0camera (Lumenera Lu170C) and the transmission factor of the half mirror is 30%. The result of face tracking is shown in Figure 7. The size of images was captured at a resolution of 640x480. The number of samples is set to 1000 of particle filter. Our method ran at about 30 frames per second. Moreover, using 3D coordinate information of the tracking points, we can perform image processing on the user's image. These effects can be applied to electronic cosmetic system. In this work, we have composed effects for electronic cosmetic.

In Figure 7, blue circles on the image show tracked feature points. We can see that face tracking is successful exactly. Face tracking in the proposed method can be stable by a single camera. Even when the user moved suddenly, this tracking can respond to such movement. In addition, it is easy for users to use this system since using a standard 3D face model to

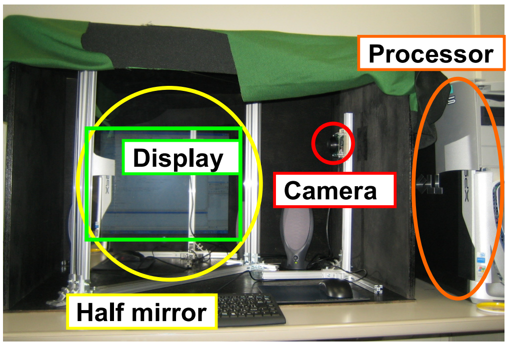

Figure 6.Our virtual mirror system

Figure 7. Result of tracking

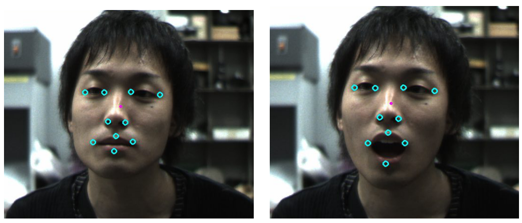

Figure 8 application the system to other user

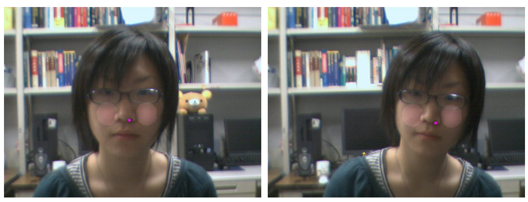

Figure 9. User's image put on rouge

obtain 3D coordinate information. Figure 8 shows the resulting images of other users by our method. The our system can process effects on user's image for electronic cosmetic. One of these effects is shown as Figure 9. The effect is to put on rouge to user's cheeks.

5 Conclusion

In this paper, we have proposed the virtual mirror system for electronic cosmetic. This system consists of a single camera and a half mirror. This system is able to approximate user's face depth of the feature points with a single camera. For tracking the face, we have used the standard 3D face model. It is able to obtain 3D information of the face without measuring 3D shape for each user by the face model. In addition, we have proposed three steps method for tracking user's face with particle filter. In conclusion, the facial pose and expression are treated separately, and the tracking is applied to changing of the pose and expression at the same time. Moreover, using obtained 3D face information, image processing is able to be applied to user's face. This processing is expected to be used as the real-time electronic cosmetics processing.

For a future work, we consider that this system can also be applied to the electronic cosmetics systems to simulate make-up with more reality. To make the system, it requires more realistic effects. In addition, it is also necessary to improve tracking accuracy itself. For instance, on the face tracking, feature points used are increased for the expression changing without decreasing tracking accuracy. This system is also required to automate at all points of the procedure obtaining feature points of the user.

References

1. T.Darrell, G.Gordon, J.Woodfill, M.Harville, "A Virtual Mirror Interface using Real-time Robust Face Tracking", In Proceedings of the Third International Conference on Face and Gesture Recognition, pp.14-16, 1998.
2. T.Makino, T.Nakaguchi, N.Tsumura, Y.Miyake, "Virtual mirror based on 3D shape reconstruction and real-time face tracking", Internet Imaging VI. Edited by Santini, Simone; Schettini, Raimondo; Gevers, Theo. Proceedings of the SPIE, Volume 5670, pp.252-259 , 2004.
3. M. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion", In International Journal of Computer Vision, vol. 25, no. 1, pp. 23-48, 1997.
4. V. Vezhnevets. "Face and Facial Feature Tracking for Natural Human-Computer Interface." Graphicon 2002. Nizhny Novgorod, Russia, pp.16-21, September 2002.
5. A. Azarbayejani, B. Horowitz, "Recursive estimation of structure and motion using the relative orientation constraint", Proc. IEEE CVPR, pp.70-75, 1993.
6. M.L. Cascia and S. Sclaroff, "Fast, Reliable Head Tracking under Varying Illumination", In CVPR99, pp. 604-610, 1999.
7. M. Isard and A. Blake, "Condensation–conditional density propagation for visual tracking", Int. J. Computer Vision, vol.29, no.1, pp.5–28, 1998.
8. F. Dornaika and F. Davoine, "Head and facial animation tracking using appearance-adaptive models and particle filters", Proc. Workshop on Real-Time Vision for Human-Computer Interaction, pp.153-153, 2004.
9. Kenji Oka and Yoichi Sato, "Real-time modeling of face deformation for 3D head pose estimation", Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2005), pp. 308-320, October 2005.
10. Kenji Oka and Yoichi Sato, "Head pose estimation system based on particle filtering with adaptive diffusion control", Proc. IAPR MVA 2005, pp.586-589, May 2005.

