Hand Posture Estimation
in Complex Backgrounds
by Considering Mis-match of Model

Akihiro Imai[1], Nobutaka Shimada[2], and Yoshiaki Shirai[2]

[1] Dept.of Computer-Controlled Mechanical Systems,
Osaka University Yamadaoka, Suita, Osaka 565-0871, Japan
[2] Dept.of Human and Computer Intelligence, Ritumeikan University,
Nojihigashi, Kusatsu, Shiga 525-8577, Japan

Abstract. This paper proposes a novel method of estimating 3-D hand posture from images observed in complex backgrounds. Conventional methods often cause mistakes by mis-matches of local image features. Our method considers possibility of the mis-match between each posture model appearance and the other model appearances in a Baysian stochastic estimation form by introducing a novel likelihood concept "Mistakenly Matching Likelihood (MML)". The correct posture model is discriminated from mis-matches by MML-based posture candidate evaluation. The method is applied to hand tracking problem in complex backgrounds and its effectiveness is shown.

1 Introduction

Precise hand-finger shape estimation methods using visual cues have been developed [1][2][3][4][5][6] in order to implement the gestural interfaces in a touch-less manner which are utilized in interaction with virtual environments and automatic sign-language translation.

One of the difficulties of implementing the interfaces based on the hand shape estimation exists in its situations where the interfaces are needed: its complex backgrounds like colorful and textured clothes, skin-colored region as human face and some desktops on which various tools and objects are scattered. Since the hand shape even in simple backgrounds is a tough problem due to its great varieties of posture, shape estimation with simultaneous segmentation is still a challenging problem.

To solve the problem with feasible computing resources, some trials were reported from the following two viewpoints: 1) how to reduce the number of posture candidates to consider (i.e. how to predict the posture), 2) how to evaluate the matching degree between the posture candidate and observed image features.

From the viewpoint 1), Active Shape Model [7] is proposed, which learns acceptable shape deformations and tracks the region contour or texture assuming smooth deformation and motion. Non-smooth deformation can be treated by introducing Switching Linear Dynamics [8]. 3-D model-based shape prediction and tracking, not based on appearance learning, is also proposed [5]. Most of
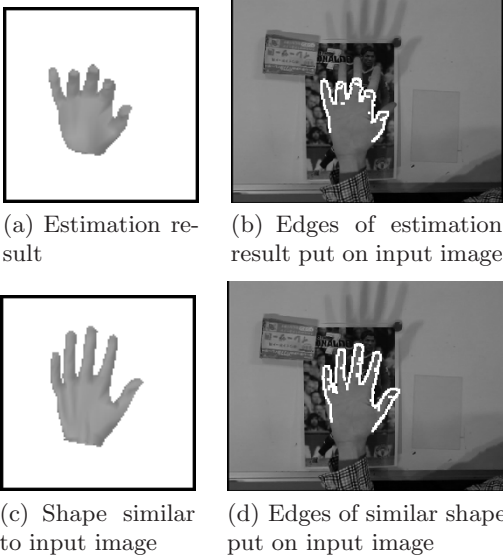
(a) Estimation re-
sult

(b) Edges of estimation
result put on input image

(c) Shape similar
to input image

(d) Edges of similar shape
put on input image

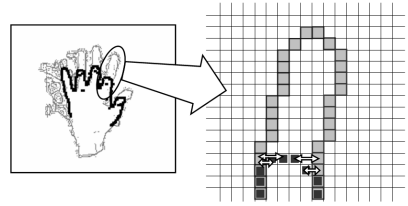**Fig. 1.** Mistake of conventional method

Fig. 2. Corresponding of edges

those methods employ a parallel search scheme in tracking, like beam-search or
particle filter for robustness against temporal mis-estimation and tracking failure
[5][9][10][11][12][13][14].

While many improvements from the first viewpoint are reported, those from
the second viewpoint, concerned with the evaluation of matching degree, are
comparatively few and most implementations employ a simple feature corre-
sponding and evaluating method: *chamfer matching* [15][16]. Chamfer matching
makes correspondences between the features with the least distance in the im-
age, and evaluates the matching degree by the sum of the distances (chamfer
distance). This simple matching scheme, of course, causes often a wrong shape
estimate on the complicated backgrounds.

Fig. 1 is an example of wrong estimate of hand posture caused by chamfer
matching. Because many edge textures are observed in the hand region, the fin-
ger tips of the posture model are mistakenly corresponding to the inner edges
(see Fig. 2) and as a result its chamfer distance is evaluated too small. While
this problem is hard to avoid as long as using chamfer matching, no more ap-
propriate matching method is found other than chamfer matching. Therefore
the matching degree should be evaluated under the consideration of that such a
wrong matching often happens.

Nevertheless, the existence of the mismatch caused by chamfer matching does
not directly mean its uselessness. Embedding approach [15] evaluates the match-
ing degree between an input image feature and not only one posture model
but also several other reference models. For example of Fig. 1, in addition to
the correct posture candidate (c), the candidate (a) also has so high matching

degree that (a) is picked up as the estimate. However, if only (a) has the high matching degree when (a) should be the correct matching, these two cases can be discriminated by evaluating the matching degrees with both reference models (a) and (c). Since Embedding approach only uses an ad-hoc way by evaluating the squared sum of the matching degrees of all reference models, its estimate is not the optimum in Bayesian point of view.

This paper mathematically derives the Baysian form of Embedding approach. In its derivation, a novel concept of likelihood is introduced: *Mistakenly Matching Likelihood (MML)*, which predicts the high evaluation caused by wrong matching and gives the ability to discriminate the true estimate from the false match in a stochastic way. The derived MML-based candidate evaluation is applied to hand tracking problem in complex backgrounds and its effectiveness is experimentally shown.

2   Acquisition of Typical Hand Posture Images

3-D Hand model used in our research is originally wireframe model. The model is modified into a shaded model. The joint bending angles are denoted by $\theta_{b,t,1}, \theta_{b,t,2}, \theta_{b,t,3}, \theta_{b,i,1}, \theta_{b,i,2}, \theta_{b,i,3}, \ldots$ and opening angles at a base joint of the fingers are denoted by $\theta_{o,t}, \theta_{o,i}, \ldots$ (shown in Fig.3). The posture of the whole hand model is represented by translation t_x, t_y, t_z and Ritalin $\theta_{r,x}, \theta_{r,y}, \theta_{r,z}$. As a whole, the shape of the hand model has 26 degrees of freedom

$$\boldsymbol{\theta} = (\theta_{b,t,1}, \ldots, \theta_{r,z}). \tag{1}$$

CG images of typical hand models are shown in Fig. 4. The finger joints dependently moves in natural actions [5]. In index, middle, ring, pinky fingers, adjacent joint angles are usually similar. This kind of joint constraints reduce
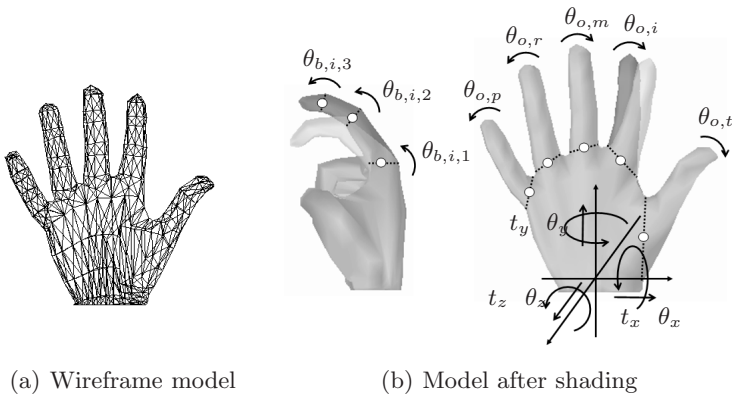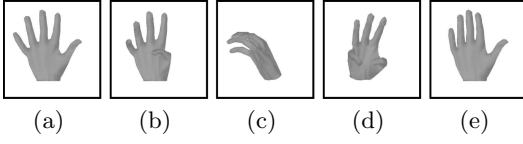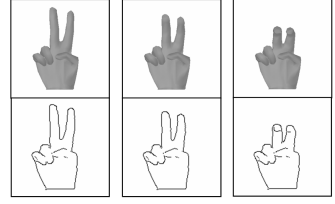

(a) Wireframe model   (b) Model after shading

Fig. 3. Hand model

Fig. 4. CG images of typical hand postures

Fig. 5. Edge images of hand model

Table 1. Quantization width from search center

$\Delta\theta^c_{bend}$ [°]	$-\zeta$	0	ζ
$\Delta\theta^c_{open}$ [°]	-6	0	6
$\Delta\theta^c_{o,m}$ [°]	-	0	-
$\Delta\theta^c_{r,x}$ [°]	-30	0	30
$\Delta\theta^c_{r,y}$ [°]	-15	0	15
$\Delta\theta^c_{r,z}$ [°]	-6	0	6
Δt^c_x, Δt^c_y [mm]	-8	0	8
Δt^c_z [mm]	-40	0	40

the number of possible postures. Fig.5 shows the edge images generated from the CG images of the typical hand postures under the constraints.

The posture parameters θ to be estimated are quantized. The change is shown in Table 1 (The change of θ_b is represented by $\Delta\theta^c_{bend}$, that of θ_o other than middle is represented by $\Delta\theta^c_{open}$). Each parameter of θ_b is assumed to change 0 or $\pm\zeta$. Each parameter has own ζ between 9 deg to 15 deg.

3   Matching Method

The system has several hand models with various dimensions: i.e. lengths and widths of the palm and fingers. Since input image sequences are assumed to start from a predefined simple shape and an initial position, the dimensions are easily initialized at the first frame. The posture parameters to be estimated are around the posture estimate at the previous image frame.

When each input image is obtained, the best-matched model is determined by the maximum likelihood criterion. Let I denote edges and skin regions extracted from an input image and Θ_j denote a quantized parameter vector of the j th model. The criterion is

$$\hat{\Theta} = \arg\max_{\Theta_j} P(I|\Theta_j) \tag{2}$$

where $P(I|\Theta_j)$ denotes the likelihood of the j th model for the input. The likelihood is defined based on the difference of the image I and the the appearance of the shape model.

3.1   Difference of Image and Appearance of Shape Model

In this paper, edge image $I^{(e)}$ and skin-color region image $I^{(s)}$ are used as the image features of the input I. The difference between the silhouette of a typical hand model and that of an image is computed. Let $A_{\boldsymbol{\theta}}^{(s)}$ be the silhouette generated from a typical hand model $\boldsymbol{\theta}$. The Difference of the silhouette, $f_{skin}(A_{\boldsymbol{\theta}}^{(s)}; I^{(s)})$, is defined as the area of $A_{\boldsymbol{\theta}}^{(s)}$ which does not overlap with $I^{(s)}$.

The difference between the $I^{(e)}$ and the edge appearance $A_{\boldsymbol{\theta}}^{(e)}$, $f_{dist}(A_{\boldsymbol{\theta}}^{(e)}; I^{(e)})$, is computed by a modified chamfer matching, in which the edge points are classified by gradient direction and the edges is matched by the original chamfer matching in each direction class[13][18]. The distance is weighted by the edge contrast and as a result f_{dist} is defined as follows:

$$f_{dist}(A_{\boldsymbol{\theta}}^{(e)}; I^{(e)}) = \sum_j w_{\boldsymbol{\theta},j} \min_k \left(||\boldsymbol{x}_{\boldsymbol{\theta},j} - \boldsymbol{x}_{I^{(e)},k}|| + f_{I^{(e)},k} + g(j,k) \right) \qquad (3)$$

where $\boldsymbol{x}_{\boldsymbol{\theta},j}$ and $\boldsymbol{x}_{I^{(e)},k}$ denotes the j th edge point of the model and the k th edge point of an input edge image. $|| \cdot ||$ is 2-dimensional Euclidean norm, $w_{\boldsymbol{\theta},j}$ is a weight constant, $f_{I^{(e)},k}$ is a penalty for an edge with a low contrast,

$$w_{\boldsymbol{\theta},j} = \frac{d_{\boldsymbol{\theta},j}}{\sum_l d_{\boldsymbol{\theta},l}}. \qquad (4)$$

$$f_{I^{(e)},k} = -w_d d_{I^{(e)},k} \qquad (5)$$

where w_d represents weight constant. The difference of gradient direction $g(j,k)$ is defined in terms of gradient direction $\phi_{\boldsymbol{\theta},j}$ of the model edge and gradient direction $\phi_{I^{(e)},k}$ as

$$g(j,k) = w_\phi ||\phi_{\boldsymbol{\theta},j} - \phi_{I^{(e)},k}||. \qquad (6)$$

where w_ϕ represents weight constant. All weights are experimentally determined. The modified chamfer matching is computed by using distance transformation as fast as the original chamfer matching.

3.2   Discrimination Principle

The example of mis-matching by a conventional method has been shown as Fig. 1 in section 1. Let Θ_a and Θ_c respectively denote the posture parameter of model (a) (fist shape) and (c) (flat shape). Suppose one case that the true hand posture should be Θ_c. In this section, we describe the principle to discriminate the true match from the wrong matches due to the complicated skin textures and the backgrounds and finally introduce its stochastic forms giving the discriminating criterion.

Then the probability of that the appearance A_{Θ_c} is matched to the input image, $p(A_{\Theta_c}|\Theta_c)$, should be large enough because the true posture is the same as the one which generates the appearance, Θ_c.

On the other hand, the probability of A_{Θ_a} (small fist shape), $p(A_{\Theta_a}|\Theta_c)$ can be also large in spite of the posture difference between Θ_a and Θ_c because
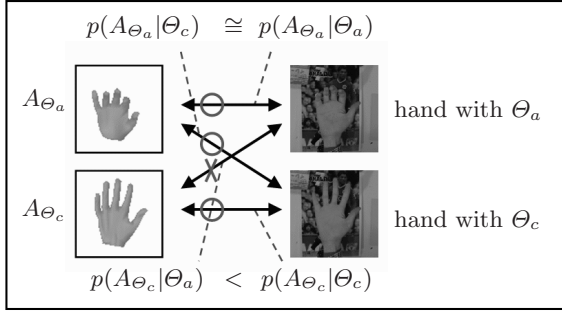
Fig. 6. Likelihood for edge images

almost all of the area of A_{Θ_a} is included and the inner texture edges can be wrongly matched to the finger contours. Therefore the conventional likelihood maximization can often choose Θ_a mistakenly for flat hand shapes like Θ_c due to the image capture noise, inaccuracy of the 3-D shape model and quantization errors of the posture parameters.

In order to solve the mis-matches, we carefully analyze the behaviours of two more probabilities: $p(A_{\Theta_a}|\Theta_a)$ and $p(A_{\Theta_c}|\Theta_a)$. $p(A_{\Theta_a}|\Theta_a)$ should be large and $p(A_{\Theta_c}|\Theta_a)$ should be small because A_{Θ_c} protrudes from the area of Θ_a. The four probabilities take the behaviours as follows: while $p(A_{\Theta_a}|\Theta_c)$, $p(A_{\Theta_c}|\Theta_c)$ are large together for the posture Θ_c, $p(A_{\Theta_a}|\Theta_a)$ is large and $p(A_{\Theta_c}|\Theta_a)$ is small for the posture Θ_a (see Fig. 6). Therefore when the appearance A_{Θ_a} and A_{Θ_c} are observed together, that posture should be estimated as Θ_c. When A_{Θ_a} is observed alone, that posture should be Θ_a.

When the likelihood of an appearance A_{Θ_k} to a model Θ_j, $p(A_{\Theta_k}|\Theta_j)$, is obtained for each possible combination of k and j in advance, the appropriate model can be chosen by taking the all $p(A_{\Theta_k}|\Theta_j)$ values into account like the above discussion.

When k and j are identical, $p(A_{\Theta_k}|\Theta_j)$ is equivalent to the conventional likelihood function. Otherwise, it means the likelihood of that an appearance A_{Θ_k} comes from a mistakenly chosen model Θ_j. We call the likelihood as "mistakenly matching likelihood" (MML).

3.3   Model Selection using Mistakenly Matching Likelihood

We introduce the stochastic discrimination criterion from the principle described in the previous section by employing Bayesian estimation framework. Let A_{Θ_1}, $A_{\Theta_2}, \ldots$ denote appearances of typical hand models. Assuming $A_{\Theta_1}, A_{\Theta_2}, \ldots$ are exclusive under each Θ_j, the likelihood of Θ_j for I can be expanded as below:

$$p(I|\Theta_j) = \sum_k p(I, A_{\Theta_k}|\Theta_j) = \sum_k p(I|A_{\Theta_k}, \Theta_j)p(A_{\Theta_k}|\Theta_j). \tag{7}$$

Assuming the appearance A_{Θ_k} has all information to generate the observed image I, condition Θ_j can be removed,

$$p(I|\Theta_j) = \sum_k p(I|A_{\Theta_k})p(A_{\Theta_k}|\Theta_j). \qquad (8)$$

In the conventional maximum likelihood estimation method, the likelihood for the case of $k = j$ is considered alone. In contrast, we additionally consider the MML for the case of $k \neq j$.

Assuming that $I^{(e)}$ and $I^{(s)}$ are independent when a certain Θ_j is specified,

$$p(I|\Theta_j) = p(I^{(e)}|\Theta_j)p(I^{(s)}|\Theta_j) \qquad (9)$$

is derived as the discrimination criterion in stochastic form.

The likelihood $p(I^{(e)}|\Theta_j)$ and $p(I^{(s)}|\Theta_j)$ are respectively derived from the following equations:

$$p(I^{(e)}|\Theta_j) = \sum_k p(I^{(e)}|A_{\Theta_k}^{(e)})p(A_{\Theta_k}^{(e)}|\Theta_j) \qquad (10)$$

$$p(I^{(s)}|\Theta_j) = \sum_k p(I^{(s)}|A_{\Theta_k}^{(s)})p(A_{\Theta_k}^{(s)}|\Theta_j) \qquad (11)$$

The probabilistic distributions $p(A_{\Theta_k}^{(e)}|\Theta_j)$ and $p(I^{(e)}|A_{\Theta_k}^{(e)})$ for edge images is introduced the following sections. Those for skin color silhouette, $p(A_{\Theta_k}^{(s)}|\Theta_j)$ and $p(I^{(s)}|A_{\Theta_k}^{(s)})$ can be introduced in the same manner of those for edge image.

3.4   Likelihood of Typical Hand Models for Appearances

The likelihood of typical hand model is obtained as the following form because of quantization errors of Θ_j.

$$
\begin{aligned}
p(A_{\Theta_k}^{(e)}|\Theta_j) &= \int_{\Theta_j} p(A_{\Theta_k}^{(e)}, \boldsymbol{\theta_j}|\Theta_j)d\boldsymbol{\theta_j} \\
&= \int_{\Theta_j} p(A_{\Theta_k}^{(e)}|\boldsymbol{\theta_j}, \Theta_j)p(\boldsymbol{\theta_j}|\Theta_j)d\boldsymbol{\theta_j} \\
&= \int_{\Theta_j} p(A_{\Theta_k}^{(e)}|\boldsymbol{\theta_j})p(\boldsymbol{\theta_j}|\Theta_j)d\boldsymbol{\theta_j}.
\end{aligned}
\qquad (12)
$$

The sampling distribution of $p(\theta_j|\Theta_j)$ can be assumed as a uniform distribution under each Θ_j. Assuming $p(A_{\Theta_k}^{(e)}|\boldsymbol{\theta_j})$ is constant for each j since the the quantization interval of Θ_j is small enough, $p(A_{\Theta_k}^{(e)}|\boldsymbol{\theta_j})$ is reduced to $p(A_{\Theta_k}^{(e)}|\boldsymbol{\theta_j^*})$, where $\boldsymbol{\theta_j^*}$ is the mean value of the interval Θ_j. $p(A_{\Theta_k}^{(e)}|\boldsymbol{\theta_j^*})$ is derived as follows from the definition of f_{dist} in sec.3.1 and assuming that f_{dist} obeys a gaussian distribution f_{dist}[12][13][17]:

$$p(A_{\Theta_k}^{(e)}|\boldsymbol{\theta_j^*}) = \alpha_{\boldsymbol{\theta_j^*}}^{(e)}exp(-(d_M^{(e)}(k,j))^2) \qquad (13)$$

where $I_r^{(e)}(\boldsymbol{\theta})$ is the edge image rendered from the posture $\boldsymbol{\theta}$, and

$$d_M^{(e)}(k,j) = \frac{f_{dist}(A_{\Theta_k}^{(e)};I_r^{(e)}(\boldsymbol{\theta_j^*}))}{\sigma_M^{(e)}}. \qquad (14)$$

$\sigma_M^{(e)2}$ is the variance of the value of $f_{dist}(A_{\Theta_k}^{(e)}; I_r^{(e)}(\boldsymbol{\theta_j^*}))$. $\sigma_M^{(e)}$ is experimentally determined. $\alpha_{\boldsymbol{\theta_j^*}}^{(e)}$ is normalization constant,

$$\alpha_{\boldsymbol{\theta_j^*}}^{(e)} = \left(\textstyle\sum_k exp(-(d_M^{(e)}(k,j))^2) \right)^{-1}. \tag{15}$$

In the same manner as the above,

$$p(A_{\Theta_k}^{(s)}|\boldsymbol{\theta_j^*}) = \alpha_{\boldsymbol{\theta_j^*}}^{(s)} exp(-(d_M^{(s)}(k,j))^2) \tag{16}$$

where $I_r^{(s)}(\boldsymbol{\theta})$ is the silhouette generated from $\boldsymbol{\theta}$, and

$$d_M^{(s)}(k,j) = \frac{f_{skin}(A_{\Theta_k}^{(s)}; I_r^{(s)}(\boldsymbol{\theta_j^*}))}{\sigma_M^{(s)}}. \tag{17}$$

$\sigma_M^{(s)2}$ is the variance of the value of $f_{skin}(A_{\Theta_k}^{(s)}; I_r^{(s)}(\boldsymbol{\theta_j^*}))$. $\alpha_{\boldsymbol{\theta_j^*}}^{(s)}$ is normalization constant,

$$\alpha_{\boldsymbol{\theta_j^*}}^{(s)} = \left(\textstyle\sum_k exp(-(d_M^{(s)}(k,j))^2) \right)^{-1}. \tag{18}$$

3.5   Likelihood of Appearance

In this section, we explain the evaluation of the likelihood of an appearance $p(I^{(e)}|A_{\Theta_k}^{(e)})$. The likelihood is defined based on the definition of f_{dist} as

$$p(I^{(e)}|A_{\Theta_k}^{(e)}) = \beta_{\Theta_k}^{(e)} exp\left(-\frac{(f_{dist}(A_{\Theta_k}^{(e)}; I^{(e)}))^2}{\sigma_I^{(e)2}} \right). \tag{19}$$

where, $\sigma_I^{(e)2}$ is the variance of $(f_{dist}(A_{\Theta_k}^{(e)}; I^{(e)}))$. It is experimentally determined. The normalization constant $\beta_{\Theta_k}^{(e)}$ is derived from the integral condition of probabilistic distributions:

$$\int p(i^{(e)}|A_{\Theta_k}^{(e)})di^{(e)} = 1 \tag{20}$$

Assuming that $p(i^{(e)}|A_{\Theta_k}^{(e)})$ can be large value only for $i^{(e)}$ of hand images and is 0 for most of other $i^{(e)}$,

$$\begin{aligned}
\int p(i^{(e)}|A_{\Theta_k}^{(e)})di^{(e)} &\approx \int p(I_r^{(e)}(\theta_l)|A_{\Theta_k}^{(e)})d\theta_l \\
&\approx \textstyle\sum_l p(I_r^{(e)}(\theta_l^*)|A_{\Theta_k}^{(e)}) \cdot \delta \\
&= \beta_{\Theta_k}^{(e)} \textstyle\sum_l exp(-(d_M^{(e)}(k,l))^2) \cdot \delta \\
&\equiv 1
\end{aligned} \tag{21}$$

where δ is the range of the quantization of Θ.

$$\beta_{\Theta_k}^{(e)} = \left(\textstyle\sum_l exp(-(d_M^{(e)}(k,l))^2) \cdot \delta \right)^{-1} \tag{22}$$

When $A_{\Theta_k}^{(e)}$ wrongly matches to many of $I_r^{(e)}(\theta_l)(l \neq k)$, $\beta_{\Theta_k}^{(e)}$ becomes small. On the other hand when a few of those, $\beta_{\Theta_k}^{(e)}$ becomes large. It means that ambiguous appearance model, which is easy to mis-match to other posture's appearances, are automatically low evaluated.

4   Estimation of More Accurate Posture Parameters

Posture parameters of the best-matched model are slightly different from that of the hand of an input image due to quantization errors of posture parameters. Thus, more accurate parameters must be estimated.

The wireframe CG model of the hand is deformed so that the model is matched to an input image, and the 3-D hand shape is reconstructed from the deformed model[19]. In this method, while the curved surface shape of the hand is reconstructed, posture parameters are not estimated. We deform the CG model so that the appearance of the model is matched to those of the input image by using this method. The accurate posture parameters are estimated from coordinates of the vertices of the triangle patches of the deformed wireframe model.

Parameters are estimated by the following steps of a procedure.

1. We make correspondences of edges of the best-matched model to those of the input image.
2. The change of the appearance is evaluated from the correspondences so that the edges of the model move toward those of the input image.
3. In order to reduce the huge search region of posture parameters due to the high DOF of human hand, available deformations of surface mesh of the CG model are learned by PCA in advance for each of typical postures, and then the best approximated mesh deformation is estimated by the projection to the PCA subspace.
4. Return to 1. We make correspondences of edges of the CG model deformed at 3. to those of the input image. CG model is deformed by the change of appearance evaluated from the correspondence, again. Repeat these processes.
5. We evaluate the 2-dimensional Euclidean norm of the vertices of the triangle patches between the deformed CG model and CG model generated by the posture parameters. The sum of the norms is minimized using steepest descent method. The posture parameters with minimized sum of the norms are the posture estimate.

5   Experiment

We did the experiment of posture tracking for 250 hand images. The resolution of the images is 320×240. The images are captured by 30 fps. In the conventional method, where $p(I|\Theta_j) = p(I|A_{\Theta_j}^{(e)})p(I|A_{\Theta_j}^{(s)})$ is used as a matching criterion, 70.4% images are correctly matched. In our method, 82.0% images are correctly matched. The success rates show effectiveness of our method.
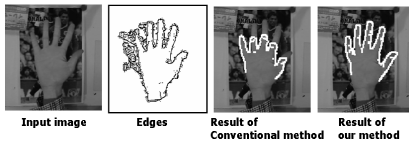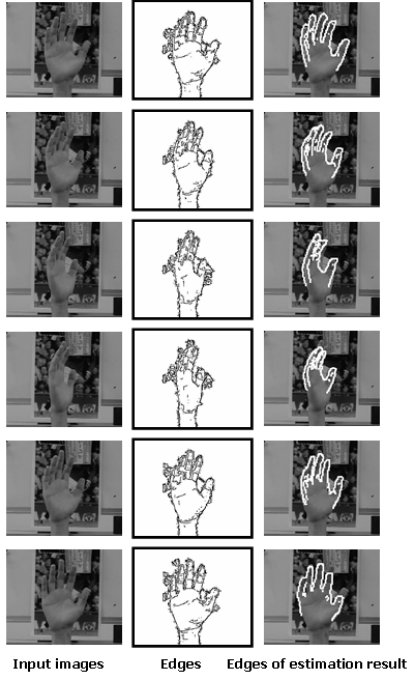
Input image   Edges   Result of   Result of
                         Conventional method   our method

Fig. 7. Experimental result

Input images   Edges   Edges of estimation result

Fig. 9. Experimental result

Input images   Edges   Edges of estimation result

Fig. 8. Experimental result

Input images   Edges   Edges of estimation result

Fig. 10. Experimental result

The example of the image which is correctly matched in our method while mis-matched in the conventional method, is shown in Fig. 7. While the wrong fist hand shape is matched in conventional method, the correct flat shape is matched in our method.

Fig. 8 shows the results of an image sequence. In the input images, fingers are partially occluded and the edges of the background are confusingly appeared near the fingers. Such cases causing mismatches are correctly matched in our method. Fig. 9 and Fig. 10 show the tracking results for other hand shapes. These images are also correctly matched in our method.

6   Conclusion and Discussion

The paper introduces a Bayesian form of evaluation of posture candidates for hand tracking in complex backgrounds. The novel concept of Mistakenly

Matching Likelihood (MML) enables to discriminate the true posture candidate from other confusing ones when the mismatch of image features frequently occurs. Experimental results for tracking of the real human hand show the effectiveness of this evaluation method. Additional image features like optical flows or range other than edges and silhouette should be considered on this framework as future work.

Acknowledgment

References

1. Liu, X., Fujimura, K.: Hand Gesture Recognition using Depth Data. In: Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition, pp. 529–534 (2004)
2. Iwai, Y., Yagi, Y., Yachida, M.: Estimation of Hand Motion and Position from Monocular Image Sequence. In: Li, S., Teoh, E.K., Mital, D., Wang, H. (eds.) ACCV1995. LNCS, vol. 1035, pp. 230–234. Springer, Heidelberg (1996)
3. Lee, S.U., Cohen, I.: 3D Hand Reconstruction from a Monocular View. In: Proc. 17th Int. Conf. on Pattern Recognition, vol. 3, pp. 310–313 (1995)
4. Kameda, Y., Minoh, M., Ikeda, K.: Three Dimensional Pose Estimation of an Articulated Object from its Silhouette Image. In: ACCV 1993, pp. 612–615 (1993)
5. Shimada, N., Kimura, K., Shirai, Y.: Real-time 3-D Hand Posture Estimation based on 2-D Appearance Retrieval Using Monocular Camera. In: Proc. Int. Workshop on RATFG-RTS, pp. 23–30 (2001)
6. Imai, A., Shimada, N., Shirai, Y.: 3-D Hand Posture Recognition by Training Contour Variation. In: Proc. of The 6th Int. Conf. on Automatic Face and Gesture Recognition, pp. 895–900 (2004)
7. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models-Their Training and Application. COMPUTER VISION AND IMAGE UNDERSTANDING 61(1), 38–59 (1995)
8. Jeong, M., Kuno, Y., Shimada, N., Shirai, Y.: Recognition of shape-changing hand gestures. IEICE Trans. inf.Syst. E85-D(10), 1678–1687 (2002)
9. Isard, M., Blake, A.: Visual tracking by stochastic propagation of conditional density. In: Proc. European Conf. Computer Vision, pp. 343–356 (1996)
10. Isard, M., Blake, A.: ICONDENSATION:Unifying low-level and high-level tracking in a stochastic framework. In: Proc. European Conf. Computer Vision, pp. 767–781 (1996)
11. Heap, T., Hogg, D.: Wormholes in Shape Space:Tracking through Discontinuous Changes in Shape. In: 6th Int. Conf. on Computer Vision, pp. 344–349 (1998)
12. Zhou, H., Huand, T.S.: Tracking Articulated Hand Motion with Eigen Dynamics Analysis. 9th Int. Conf. on Computer Vision 2, 1102–1109 (2003)

13. Stenger, B., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Model-Based Hand Track-
    ing Using a Hierarchical Bayesian Filter. IEEE Transactions on Pattern Analysis
    and Machine Intelligence 28(9), 1372–1384 (2006)
14. Wu, Y., Lin, J., Huang, T.S.: Analyzing and Capturing Articulated Hand Motion
    in Image Sequences. IEEE TRANS. ON PATTERN ANALYSIS AND MACHINE
    INTELLIGENCE 27(12), 1910–1922 (2005)
15. Athitsos, V., Sclaroff, S.: Estimating 3D Hand Pose from a Cluttered Image. In:
    IEEE Conf. on Computer Vision and Pattern Recognition, vol. II, pp. 432–439.
    IEEE Computer Society Press, Los Alamitos (2003)
16. Barrow, H.G., Tenenbaum, J.M., Bolles, R.C., Wolf, H.C.: Parametric correspon-
    dence and chamfer matching: Two new techniques for image matching. In: Proc.
    5th Int. Joint Conf. Artificial Intelligence, pp. 659–663 (1977)
17. Blake, A., Isard, M.: Active Contours. Springer, Heidelberg (1998)
18. Navaratnam, R., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Hierarchical Part-
    Based Human Body Pose Estimation. In: Proc. British machine Vision Conference
    (2005)
19. Heap, T., Hogg, D.: Towards 3D Hand Tracking using a Deformable Model. In:
    2nd Int. Conf. on Automatic Face and Gesture Recognition, pp. 140–145 (1996)