ELSEVIER

Image and Vision Computing 16 (1998) 121–134

image AND vision COMPUTING

Model-based articulated hand motion tracking for gesture recognition

Cheng-Chang Lien, Chung-Lin Huang*

Institute of Electrical Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan, ROC

Received 14 October 1996; revised 21 May 1997; accepted 22 May 1997

Abstract

Conventional model-based hand gesture analysis systems require high computation cost to solve the finger inverse kinematics that makes them very difficult for real-time implementation. In this paper, we propose a fast hand model fitting method for the tracking of hand motion. The model fitting method consists of (1) finding the closed-form inverse kinematics solution for the finger fitting process, and (2) defining the alignment measure for the wrist fitting process. In the experiments, we illustrate that our hand model fitting method is effective and real-time implementable. © 1998 Elsevier Science B.V.

Keywords: Model-based; Gesture recognition; Real-time implementable hand motion tracking; Finger inverse kinematics

1. Introduction

Recently, hand motion analysis researches have become more and more popular for applications in hand gesture recognition, intelligent human–machine interface, virtual reality, and computer animation. The hand gesture identification algorithms can be classified into the gloves-based [1] and vision-based methods [2]. All the gloves-based methods are designed to detect the hand shape and finger motion in real time. However, the mechanical gloves are expensive, uncomfortable to wear, and finger movement is limited. The vision-based methods provide a more suitable and natural human–machine interaction. However, the vision-based methods have problems with occlusion, noise, and spurious data. This paper proposes a fast model-based method for articulated hand model fitting which compares the internal model with input images.

The vision-based methods consist of contour-based [3–5] and model-based methods [6–8]. The contour-based methods have problems in that (1) the edge extraction is not reliable because of the occlusion and motion of the fingers, and (2) the silhouette or contour feature from the images cannot provide enough information to estimate 3-D hand posture. The model-based method uses a hand model to analyze the 3-D human hand posture which is an articulated object with about 27 degrees of freedom (DOF). Larger degrees of freedom leads to more computation. Therefore, constraint-based kinematics methods [7,8] have been used to reduce the search space and the number of variables. However, the computation cost is still too high.

The Digiteye [9] is a recently developed real-time gesture recognition system which recognizes the articulated motion of fingers as a 3-D mouse. Another real-time system, the Digitaldesk [10], uses fingers for human–computer interaction. However, these two systems only analyze the finger-tip motion. Badler et al. [12] did a pioneer research by proposing the concept of the reach tree method to analyze the 3-D posture of the entire hand. They employed a human model with 18 joints and 48 DOF to fit various body postures by using the reach tree data structure. Each limb of the body is represented as a node with different weight in the reach tree. In the reach tree data structure, the lwd (local weighted distance) and stwd (subtree weighted distance) are introduced for fitting the body model to the specified body posture. Lee and Kunii [7,8] modified the method of reach tree to design a two-phase fitting algorithm for the finger-fitting problem. In the first phase, a wrist-positioning process moves the hand model from an arbitrary position and orientation to the one in the vicinity of the hand image. The second phase is a finger-positioning process.

In the above algorithms, one of the most time-consuming steps is the fingers-fitting process. The inverse kinematics proposed by Hans and Girard [11] does not generate closed-form solutions. The joint angles are computed by the binary search method. In this paper, we propose a fast hand model fitting method which consists of two processes to speed up the fitting process. The first one generates a new closed-form solution for the finger kinematics problem. The second

* Corresponding author. E-mail: clhuang@ee.nthu.edu.tw

PII S0262-8856(97)00041-3

one applies the alignment measure to fit the hand model to the posture of the real hand. Furthermore, the 3-D positions of the markers in the images can be obtained by detecting the color markers on the hand and with stereo vision [7,8].

2. The inverse kinematics solution function generation

Human motion analysis has always been a challenging problem. In this paper, we propose a model-based gesture analysis by using inter-dependent finger movements (or constraints) and reducing the searching space of matches between the hand models and the real input position data. Firstly, we review the hand model. Secondly, we propose a new method to solve the finger inverse kinematics problem.

2.1. The kinematics of the hand model

Lee and Knuii [7,8] classified the joints of the hand as flexive, directive, or spherical joints. The flexive joints have 1 DOF, the directive joints have 2 DOF, and the spherical joints have 3 DOF. Fingers II–V have four DOF, while the thumb has five DOF. Therefore, there exist 27 DOF in the hand, including 3 DOF for wrist translation and 3 DOF for wrist rotation. According to the D-H (Denavit-Hartenberg) rules [12], we define the local coordinate system for each finger. Fig. 1 illustrates the hand model and the local coordinate systems with the origins on the joints.

For a joint i of a finger, there are four structural kinematics parameters: (1) a_i, the length of the link i; (2) α_i, the angle of rotation about positive x_{i-1} axis measured from the positive z_{i-1} axis to the positive z_i axis, i.e., the twist angle of link i; (3) θ_i, the angle of rotation about positive z_{i-1} axis measured from the positive x_{i-1} axis to the positive x_i axis; (4) d_i, the perpendicular distance between the x_i axis and the x_{i-1} axis.

The four structural kinematics parameters determine the local transformation matrix A. If a vector p_i is known in the ith coordinate system, it can be expressed in the $(i-1)$th coordinate system as p_{i-1}, i.e.,

$$p_{i-1} = A^i_{i-1} p_i. \tag{1}$$

Because fingers II–V have four links, a vector p_4 in the fourth coordinate system can be expressed as a vector p_0 in the base coordinate system as

$$p_0 = A^1_0 A^2_1 A^3_2 A^4_3 p_4 = A^4_0 p_4. \tag{2}$$

Based on Eq. (2), the forward kinematics method can fit the hand model to the 3-D positions of the identified markers. Since fingers II–V are planar manipulators, the most important structural kinematics parameters are θ_i.

Similar to [11], we summarize some important constraints for the joint angles and the movements of fingers: (1) the four fingers are planar manipulators with the exception of the MP (metacarpophalangeal) joints; (2) the relation between joint angles 3 and 4 for finger II–V is represented by $\theta_4 = 2/3\theta_3$; (3) joint angle 2 around the y axis of middle finger is negligible, i.e., $\theta^y_2(\mathrm{III}) = 0$; (4) the dependency of joint angles 1 and 2 of the thumb around the z axis can be represented by $\theta^z_2(\mathrm{I}) = 2\theta^z_1(\mathrm{I}) - (1/6)\pi$; (5) the dependency of joint angles 1 and 2 of thumb around the y axis can be depicted by $\theta^y_2(\mathrm{I}) = \theta^y_1(\mathrm{I})(7/5)$.

2.2. The two-separable-step method

The conventional inverse kinematics solution [14,15] is obtained by solving the matrix algebra between the transformation matrix of the end-effector and the product of local transformation matrices of every joint. The *san*-coordinate frame is used to represent the coordinate of the end-effector. The end-effector is the same as the finger tip in the hand model. Let T^n_0 be a 4×4 matrix that relates the finger tip coordinate frame to the base frame system, i.e.,

$$T^n_0 = \begin{bmatrix} n_{x_0} & s_{x_0} & a_{x_0} & \bar{p}_{x_0} \\ n_{y_0} & s_{y_0} & a_{y_0} & \bar{p}_{y_0} \\ n_{z_0} & s_{z_0} & a_{z_0} & \bar{p}_{z_0} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} n & s & a & \bar{p}_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where, $\boldsymbol{n}$, $\boldsymbol{s}$, $\boldsymbol{a}$, are the unit orientation vectors of the finger tip represented in space coordinate, and $\bar{p}_0$ is the position vector of the finger tip. The relation between the T^n_0 and the local transformation matrix A^i_{i-1} is

$$T^n_0 = A^n_0 = A^1_0 A^2_1 A^3_2 \ldots A^n_{n-1}. \tag{4}$$

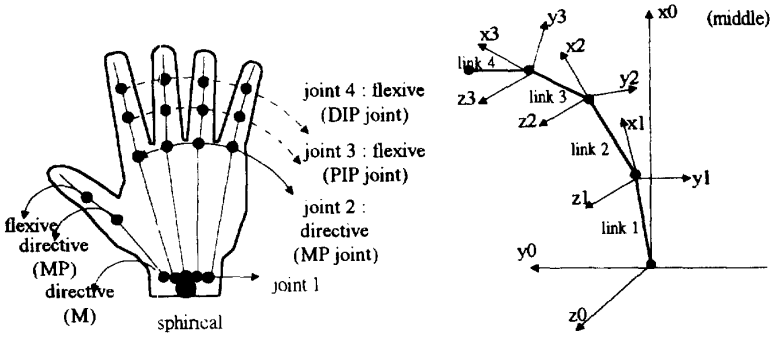

Fig. 1. (a) The hand model. (b) The local coordinate frames on the joint position for middle finger.

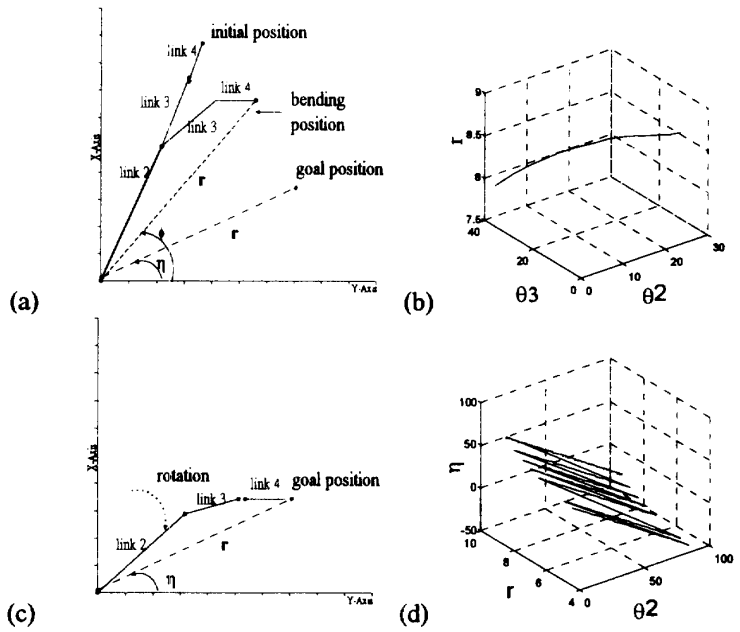

Fig. 2. (a) The first step is to move the finger tip along a certain ϕ angle to the desired distance r. (b) All the possible r, θ_2 and θ_3 values are illustrated. (c) The second step is to rotate the link 1 (MP joint) of the finger from angle ϕ to angle θ. (d) All the possible η, r, and θ_2 values are illustrated.

However, unlike the robotics manipulators [14,15], the [n, s, a] vector (the orientation vector of last link) is unknown for the vision-based gesture tracking system. Because the stereo vision can only detect the 3-D position of the hand markers, we assume that we can only acquire the 3-D position vector $[\bar{p}_{x_0}, \bar{p}_{y_0}, \bar{p}_{z_0}]$ by using stereo vision, and we do not use the conventional inverse kinematics method to find the values of joint angles. Another disadvantage of using conventional inverse kinematics is that it will lead to multiple solutions. We have to choose the correct one that satisfies the fingers' movement constraints. Based on our observations of the finger motion of ordinary people, we find that there are some constraints for bending our fingers. Therefore, given a 3-D finger tip position, there exists a unique solution for each joint angle. Instead of using the conventional inverse kinematics method, we propose a two-separable-step method to generate the closed-form solution for each joint angle.

If the 3-D position of a finger tip is known, for example the index finger, we can calculate the distance r and the slant angle η from the position of MP (metacarpophalangeal) joint to that position. The movement of index finger can be decomposed into two separable steps. The first step is to bend the finger along a certain ϕ angle to the desired r distance (see Fig. 2(a)). The second step is to rotate the link 2 (MP joint) of the finger from angle ϕ to the angle η (see Fig. 2(c)).

2.3. Solution function for finger inverse kinematics

Since the movement of fingers can be decomposed into two separable steps, we may define two solution functions to describe the relationship between the goal position of finger tip, i.e., the radius and slant angle, (r,η), and the joint angles, $(\theta_2, \theta_3, \theta_4)$. Because $\theta_4 = 2/3\theta_3$, the DOF of each finger can be reduced to 2. First, we apply an exhaustive search method to find as many solutions as possible for all reachable goal positions of the finger tips. Second, we observe that (1) all the possible r, θ_2, and θ_3 values can be depicted by an analytic curve (i.e., Fig. 2 (b)), and (2) all the possible η, r, and θ_2 values can also be described by a conical surface (i.e., Fig. 2(d)). Third, we apply the regression method to identify the coefficients of these two analytical (or solution) functions. The first function which depicts the relationship among r, θ_2 and θ_3 for a certain elevation angle ϕ is defined as

$$r = f(\theta_2, \theta_3). \tag{5}$$

The second function which relates slant angle η and variables r and θ_2 is defined as

$$\eta = g(r, \theta_2). \tag{6}$$

Once the above two functions having been developed, they can be applied to find the closed-form solution for the finger inverse kinematics. It can also be applied for the model-based hand motion tracking.

2.4. The solution functions $r = f(\theta_2,\theta_3)$ and $\eta = g(r,\theta_2)$

From the observation of all possible solutions in (r, θ_2, θ_3) space, we assume that the first solution function is can be described by as analytical curve which is expressed by

intersecting the conical surface with a vertical plane as

$$r = f(\theta_2, \theta_3) = a\theta_2^2 + b\theta_2\theta_3 + c\theta_3^2 + d\theta_2 + e\theta_3 + f, \quad \text{and}$$

$$\theta_2 = m\theta_3 + k \tag{7}$$

The first solution function can be further simplified as

$$r = A\theta_3^2 + B\theta_3 + C \tag{8}$$

where $A = am^2 + bm + c$, $B = 2amk + bk + dm + e$, and $C = ak^2 + dk + f$. Given r, θ_3 can be derived.

The same assumption can be made for the second solution function $\eta = g(r,\theta_2)$. The solution function can be described by another conical surface in (η, r, θ_2) space as

$$\eta = a\theta_2^2 + b\theta_2 r + cr^2 + d\theta_2 + er + f. \tag{9}$$

Given r, the solution function can be simplified as

$$\eta = l\theta_2^2 + m\theta^2 + n \tag{10}$$

where $l = a$, $m = br + d$, $n = cr^2 + er + f$. Given the destination slant angle η and radius r of the goal position of finger tip, we may derive the joint angle θ_3 and then the joint angle θ_2.

2.5. The regression method

Once all the possible combination of the solutions (r, θ_2, θ_3) and (η, r, θ_2) are obtained, we may use the curve fitting and surface fitting methods to find the solution functions. Assuming that the combination of all the solutions are on the same analytical contour or the same conical surface, we may apply the least-square regression method [13] to find the coefficients of Eq. (7) and Eq. (9). Although the solutions may not be precisely fitted with the analytical contour or surface, we still can make the best estimation of the coefficients by using as many solutions as possible in the solution space. The co-analytic-contour and the co-conical-surface assumptions together with the regression results can be easily verified. First, we make some movements of any finger (with known joint angles) and calculate the 3-D positions of finger-tips. Then, by using these 3-D positions as input, we may apply the solution functions to calculate the joint angles. In the experiments, we compare these two sets of joint angles, and find very little difference.

2.6. The accuracy of the new inverse kinematics solution

Fig. 3(a) shows the first solution function $r = f(\theta_2, \theta_3)$, and Fig. 3(b) illustrates the second solution function $\eta = g(r, \theta_2)$. These solution functions for finger I–V are listed in Appendix A. How to use the solution functions to calculate the joint angles θ_2, θ_3, and θ_4 is illustrated as follows: (a) apply the destination radius of finger tip to Eq. (8) for the joint angle θ_3; (b) apply the destination slant angle of the designated finger to Eq. (10) to calculate the angle θ_2.

To verify the accuracy of the regression process, we can posture our hand and then apply the solution functions to find the best hand model fitting. In the experiment, we use several gestures to verify the accuracy of regression process. From the experimental results, we find that the maximum error is 3° because of the inaccurate surface fitting of the solution function $\eta = g(r, \theta_2)$.

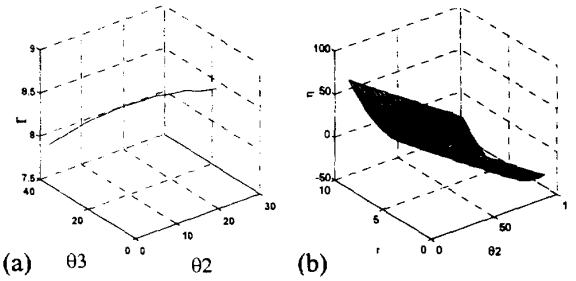

Fig. 3. The final regression of the solution functions: (a) $r = f(\theta_2,\theta_3)$ and (b) $\eta = g(r,\theta_2)$.

3. The fast model fitting method

Here, we propose a fast model fitting method with less computation time than the Lee and Kunii method [8], but with similar accurate results.

3.1. Overview of the fitting methods

The well-known hand model fitting method proposed by Lee and Kunii [8] is summarized in the following steps:

Step 1:	Determine the subtree weighted torque (stwt) for the hand tree root of the hand model.
Step 2.	Select the rotation axis that have the greatest torque component.
Step 3.	Incrementally rotate the hand model around the selected rotation axis to the position that the torque on the axis decreases to zero. At each increment, apply the inverse-kinematics subroutine to relocate all fingers to optimal positions.
Step 4.	If torque of the entire hand is smaller than the threshold value then quit.

In Step 3, the calculations of the stwt and the inverse-kinematics subroutine is the bottleneck. Once the hand root is rotated, the inverse-kinematics subroutine has to be applied for each increment in the hand model fitting process. To solve the problem, we propose the closed-form solutions for the inverse-kinematics to reduce a large amount of computation in the inverse-kinematics. Once the direction of hand root is determined, the fingers can be allocated to the optimal positions efficiently.

To increase the speed of model fitting, we propose a separable fitting method. Based on the structure of hand tree, we decompose the hand tree into three parts: wrist, fingers II–V, and thumb, then we apply the wrist fitting, finger fitting, and thumb fitting processes, respectively.

The separable fitting process is similar to the reach tree method. It firstly deals with the hand root fitting and then the leaf node fitting. The hand root fitting is the crucial operation of the entire fitting process. Here, we propose a fast and accurate hand model fitting process.

3.2. The fast fitting method

The hand model fitting method can be decomposed into three phases. In the first phase, the wrist fitting process estimates the translation of the wrist of the hand model to the real hand position and the rotation angles of the hand model about the x, y, and z axes, i.e., yaw, pitch, and roll, with respect to the directions of the real hand.

In the second phase, finger fitting process, which deals with the movements of adduction and abduction for fingers II–V, (excluding the thumb), consists of MP (metacarpophalangeal) joint fitting process and the finger fitting process. The latter uses the new inverse kinematics solutions to allocate the positions for fingers II–V as precisely as possible. Finally, the thumb fitting process is applied in the third (last) phase, that is because the thumb can move with high degree of freedom and almost independently of the motion of fingers II–V. The thumb fitting, similar to the wrist and finger fitting, fits the thumb of the hand model to the 3-D tip position of the thumb.

The three phase fitting methods are developed by taking advantages of the 3-D positions of seven markers. Fig. 4 shows the coordinates of the wrist and the seven markers on the hand. Marker ***a*** is pasted on the wrist position and marker ***b*** is pasted on the centroid of the palm. Markers ***c***–***g*** are pasted on the tips of the four fingers and thumb. The seven markers are easily identified by the image feature extraction process and their 3-D finger tip positions can be calculated by a stereo vision technique. The above two processes can be referred to in Refs. [16,17].

3.2.1. The wrist fitting phase

The wrist fitting process consists of the translation and rotation of the hand model about the x, y, and z axes of the wrist coordinate to fit the 3-D positions of a real hand. In this wrist fitting process, the main parameters to be determined are the rotation angles of the wrist about the x, y, and z axes: ϕ_x, ϕ_y and ϕ_z. In our wrist fitting algorithm, we apply three hand motion constraints to determine the angles ϕ_x, ϕ_y and ϕ_z. The first constraint indicates that fingers II–V are planar manipulators. The second constraint is with regard to the restriction of abduction motion or adduction motion for the middle finger. Actually, the middle finger may have abduction/adduction movements, especially when the other fingers are straight. However, when the other fingers are moving, the abduction/adduction movements of the middle finger are very limited. Besides, the assumption of limited abduction/adduction movements for the middle finger can provide a useful constraint for the wrist fitting process. The assumption of limited abduction/adduction movements for the middle finger is based on the considerations of physiological and model fitting efficiency.

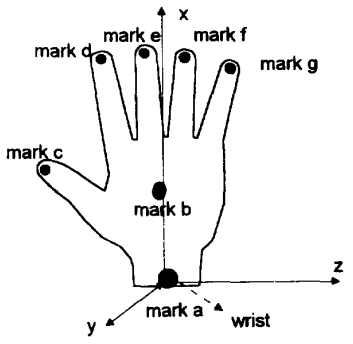

Fig. 4. The coordinate of the wrist and the seven characteristic marks on the hand.

The third constraint assumes that the motion of the palm is rigid body motion. If the palm is a non-rigid object, then the first link of fingers II–IV (the links in the palm) will have 2 DOF and there will be 35 DOF in the hand model. The fitting process will become unnecessarily complex and inefficient. Therefore, we assume that the palm is a rigid object and use 27 DOF to describe the movements of the hand model. The parameters ϕ_x and ϕ_y are calculated by using the first two constraints, and the third parameter ϕ_z is computed by using the third constraint.

The following equation describes the transformation between the space coordinate and the local coordinate of the wrist as

$$p_0 = W_0^1 W_1^2 W_2^3 p_w = W_0^3 p_w, \tag{11}$$

where

$$W_0^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi_x & -\sin\phi_x & 0 \\ 0 & \sin\phi_x & \cos\phi_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$W_1^2 = \begin{bmatrix} \cos\phi_y & 0 & \sin\phi_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\phi_y & 0 & \cos\phi_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$W_2^3 = \begin{bmatrix} \cos\phi_z & -\sin\phi_z & 0 & 0 \\ \sin\phi_z & \cos\phi_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and $\boldsymbol{p}_w$ is the vector $\boldsymbol{p}_0$ represented in wrist coordinate. Eq. (11) describes the rotation transform about the x, y,

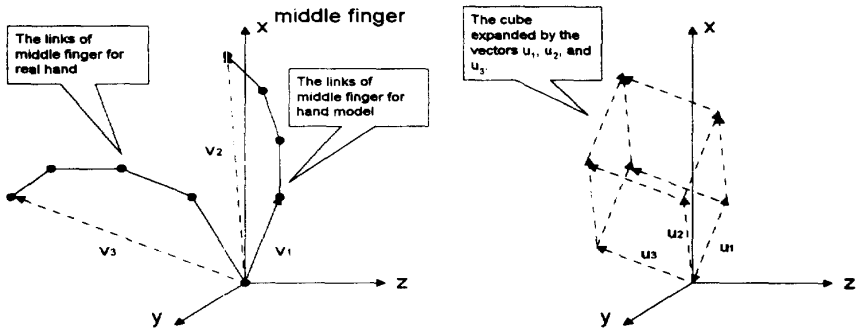

Fig. 5. The middle finger is a coplanar manipulator.

and z axes. Based on the above formula, the transformation between space coordinate and local coordinate of the wrist is used to calculate the angles ϕ_x, ϕ_y, and ϕ_z. The transform matrices W_0^1, W_1^2, and W_2^3 are assigned as the rotation matrices about x, y, and z axis, respectively, such that the rotation of the wrist can be depicted by the product of three transform matrices. Given the 3-D positions of the seven markers, we can apply the model fitting by first rotating the hand model about the x axis, then the y axis, and finally the z axis.

3.2.1.1. Calculation of ϕ_x and ϕ_y. Here, we propose the coarse finding and the fine adjustment to calculate ϕ_x and ϕ_y.

3.2.1.1.1. Coarse finding. In order to estimate the parameters ϕ_x and ϕ_y, we apply the coplanar constraint on the movement of fingers and the constraint of limited abduction and adduction movement on the middle finger. The coplanar constraint for the bending movement of the middle finger can be depicted in the Fig. 5. There are two model vectors (v_1 and v_2) originating from the position of joint 1 of the middle finger to the position of finger tip and the position of the MP (metacarpophalangeal) joint. The third vector v_3 originates from the position of joint 1 of middle finger of the hand model to the tip position of middle finger of a real hand. The unit vectors (u_1, u_2, and u_3) form a cube whose volume determines how closely the three vectors are located on the same plane. Here, the cubical volume defines the alignment measure, which can be used to determine how closely the finger tip position is located on the finger bending plane of the hand model. The alignment measure is defined as

$$V_i = |(\vec{u}_1 \times \vec{u}_2)\cdot\vec{u}_3|, \tag{12}$$

where, i is the finger index. Similarly, we can apply the alignment measurements of finger 2 (index, finger 4 (ring), and finger 5 (little), respectively, for the fitting process.

Because the movements of abduction and adduction of the index, ring and little fingers may not be used to estimate ϕ_x and ϕ_y accurately, the different weighting value is applied to the alignment measure for each finger. Therefore, the alignment measure for each finger may have different weighting factor representing different 'influence'. The rotation angles ϕ_x and ϕ_y of the wrist are determined by minimizing the hybrid measurement U defined as follows:

$$U = w_1V_1 + w_2V_2 + w_3V_3 + w_4V_4 \tag{13}$$

where w_i is the weighting for finger i, V_1, V_2, V_3, and V_4 are the alignment measures for the index finger, middle finger, ring finger, and little finger, respectively. The highest weighting value is assigned to the middle finger, i.e., $w_2 > w_i$, $i = 1, 3, 4$. The searching algorithm of the rotation angles ϕ_x and ϕ_y is described as follows: (1) rotate the wrist of the hand model incrementally with the angles $\Delta\phi_x$ and $\Delta\phi_y$, and then calculate the alignment measurements: V_i, $i = 1, 2, 3, 4$; (2) calculate the hybrid alignment measurement U, defined in Eq. (13); (3) find the angles of ϕ_x and ϕ_y that minimize the value of U.

3.2.1.1.2. Fine adjustment. The direction of the wrist can be adjusted precisely by applying the constraint that limits the abduction motion or adduction motion of the middle finger. Using the algorithm mentioned in Step 1, we can rotate the hand model around the direction of ϕ_x and ϕ_y with a small deviation $\Delta\phi_x$ and $\Delta\phi_y$ until the alignment measure for the middle finger of hand model and real hand is below a certain lower threshold. Before performing the fine alignment process, we need to transform the tip position of the middle finger to the local coordinate of the wrist. The following equation describes this transformation as

$$(W_0^3)^{-1}p_0 = p_{\text{wrist}}, \tag{14}$$

where W_0^3 is the global transform matrix of the wrist. When the tip position of the middle finger is represented by the coordinate of the wrist, the fine adjustment of angles ϕ_x and ϕ_y can be performed by rotating the hand model in the direction such that the hand model and the real hand are aligned.

3.2.1.2. Calculation of ϕ_z. Here, we explain how to determine the main parameter ϕ_z (pitch angle). Similar to Step 1 of the previous calculation, we define three vectors and apply the volume estimation method to acquire the

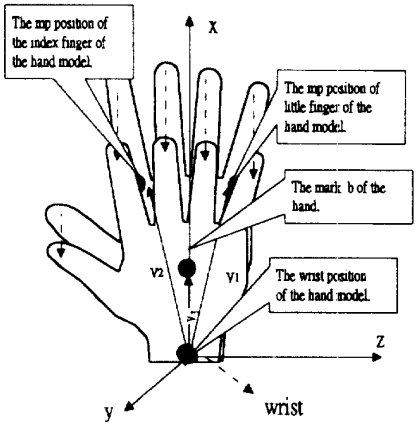

Fig. 6. The three vectors that are applied to calculate the angle ϕ_z.

angle ϕ_z. In Fig. 6, there are two vectors originating from the wrist position of the hand model to the finger tip positions of index and little finger of the hand model. The third vector originates from the wrist position of hand model to the position of marker $\boldsymbol{b}$ on the real hand. We rotate the wrist incrementally with the angles ϕ_z, to minimize the value of V (until the three vectors are coplanar) so that the palm orientation of the hand model is closely fitted to the orientation of the markers on the real hand.

3.2.2. Finger fitting

The accuracy of finger fitting depends on the correctness of the wrist fitting. The finger fitting can be decomposed into two processes. The first is the MP (metacarpophalangeal) joints fitting process which calculates the angles of adduction and abduction for fingers II–V. The second is the inverse-kinematics fitting process.

3.2.2.3. MP joint fitting. This process estimates the angles of adduction and abduction for fingers II–V. If the direction of the wrist is accurately estimated, the palm of the hand model can be accurately fitted to the position of a real palm. Here, we assume that the limited abduction motion or adduction motion for the middle finger is negligible. Therefore, we do not need to calculate abduction angular motion or adduction angular motion of the middle finger. Before calculating the abduction angle or adduction angle, we need to transform the tip positions of fingers to the local coordinate of the MP joint described as

$$(A_{\mathrm{MP}})^{-1}p_0 = p_{\mathrm{MP}}, \tag{15}$$

where A_{MP} is the global transform matrix of MP joint. When the tip positions of fingers are represented by the coordinate of MP joint, the angle of abduction motion or adduction motion can be calculated by rotating the finger about the y-axis until the alignment process is accomplished. The adjusting process is shown in Fig. 7.

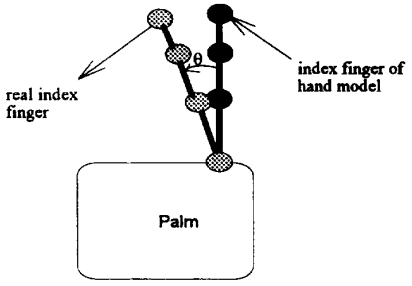

Fig. 7. By adjusting the abduction angle or adduction angle of a finger of the hand model to the tip position of real finger.

3.2.2.4. Inverse/forward-kinematics fitting. In Section 2, we developed closed-form solutions for the inverse kinematics problems. By applying the technique of stereo vision, we can obtain the 3-D positions of the markers $\{c,\ldots,g\}$. To use the closed-form solutions, we need to know the distance r and slant angle η relative to the MP joint positions. Here, we apply the method of coordinates transformation to acquire the distance r and slant angle η. In the finger fitting phase, the 3-D positions of the five markers are transformed to the coordinate of the MP joint. The transformation between the space coordinate and local coordinate of the last joint is described as

$$p_0 = W_0^3 A_0^1 A_1^2 \ldots A_{i-1}^i p_i, \tag{16}$$

where joint i is the last joint. Multiplying the inverse matrices of A_0^1, A_1^2, and W_0^3 on both sides of Eq. (16), we may have the 3-D positions of markers $\{c,\ldots,g\}$ transformed to the local coordinates of the MP joint, i.e.,

$$p_{\mathrm{MP}} = (A_1^2)^{-1}(A_0^1)^{-1}(W_0^3)^{-1}p_0 = A_2^3 \ldots A_{i-1}^i p_i. \tag{17}$$

The distance r and slant angle η are obtained from the vector p_{MP}. Because the closed-form solutions for the fingers are not exact, the forward-kinematics is applied to reduce the error. The inverse/forward-kinematics fitting is implemented as follows.

(1) Obtain the values of r and η from the vector p_{MP}, and then substitute these values into the closed-form solutions to obtain the angular values of the finger joints.

(2) Apply the method of the forward-kinematics to relocate the fingers by fine adjusting the joint angles such that the hand model may reach the real location of finger tips as closely as possible.

3.2.2.5. The thumb fitting. Because the M joint (see Fig. 1) of the thumb has two DOF with a large rotation angle and a bending direction, we use two local transformation matrices A_0^1 *and* A_1^2 *to describe the movements of the M joint. The two*

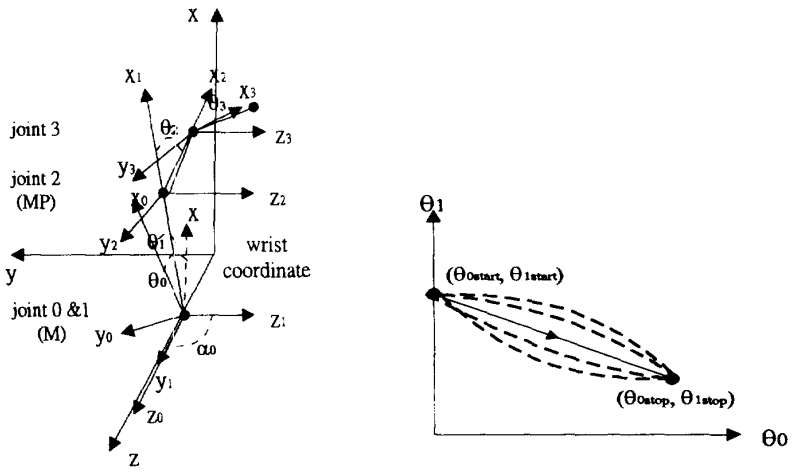

Fig. 8. (a) The kinematics for the thumb. (b) The traces of the joint angles (θ_0, θ_1) of the M joint.

transformation matrices A_0^1 and A_1^2 are defined as

$$A_0^1 = \begin{bmatrix} \cos\theta_0 & -\cos\alpha_0 \sin\theta_0 & \sin\alpha_0 \sin\theta_0 & a\cos\theta_0 \\ \sin\theta_0 & \cos\alpha_0 \cos\theta_0 & -\sin\alpha_0 \cos\theta_0 & a\sin\theta_0 \\ 0 & \sin\alpha_0 & \cos\alpha_0 & d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A_1^2 = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 & 0 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where the angles θ_0, α_0, and θ_1 are shown in Fig. 8(a). The transformation matrix A_0^1 indicates that the M joint rotates θ_0 about the z-axis and then rotates α_0 about the x-axis such that the links 1, 2, and 3 (see Fig. 8(a)) form a planar manipulator bending with angles θ_1, θ_2, and θ_3. The value of the parameter α_0 is determined by the hand structure and in our 3-D hand model, $\alpha_0 \approx 90°$. Similar to the MP joint fitting process of the finger fitting, the MP joint fitting process of the thumb fitting can be developed.

Based on our observations of the movement of the M joint of the thumb, we find that there are many ways for the joint angles, (θ_0, θ_1) (shown in Fig. 8(b)), to vary from ($\theta_{0\text{start}}$, $\theta_{1\text{start}}$) to ($\theta_{0\text{stop}}$, $\theta_{1\text{stop}}$). However, for a normal thumb movement, the relationship between joint angles θ_0 and θ_1 is almost linear and can be approximated by

$$\theta_1 = m\theta_0 + b. \tag{18}$$

The above approximation can be used to find the angle θ_1 efficiently after the value of θ_0 has been identified. In the experiments, we find that $m = -0.643$ and $b = 45$. The value of θ_0 is found by applying the method of alignment measure to the thumb. The fitting process of the thumb is implemented by the following steps: (1) apply the coplanar constraint to find the joint angle of θ_0; (2) apply Eq. (18) to calculate the joint angle of θ_1; (3) calculate the angle of abduction motion or adduction motion by rotating the MP joint of the thumb about the y-axis until the alignment process is accomplished; (4) calculate joint angles θ_2 and θ_3 by using the closed form solution function (i.e., Eq. (8) and Eq. (10)) of inverse-kinematics for the thumb fitting.

3.3. The error analysis of the fast fitting method

Among the above three fitting phases, the most important phase is the hand root fitting phase, i.e., the wrist fitting. The accuracy of finger fitting depends on the correctness of the wrist fitting. Since the 3-D positions of the seven markers cannot be measured without error, we need to develop a robust fast fitting algorithm to reduce the influence of the measurement error. The test is performed by simulating an arbitrary gesture and disturbing the 3-D positions of the seven markers with different noise levels. The noise is generated randomly on the x, y, z axes simultaneously within the ranges of 2.5, 5, and 7.5 mm, respectively. By disturbing

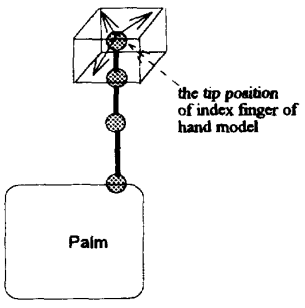

Fig. 9. The cube indicates noise distribution for the tip position of index finger.

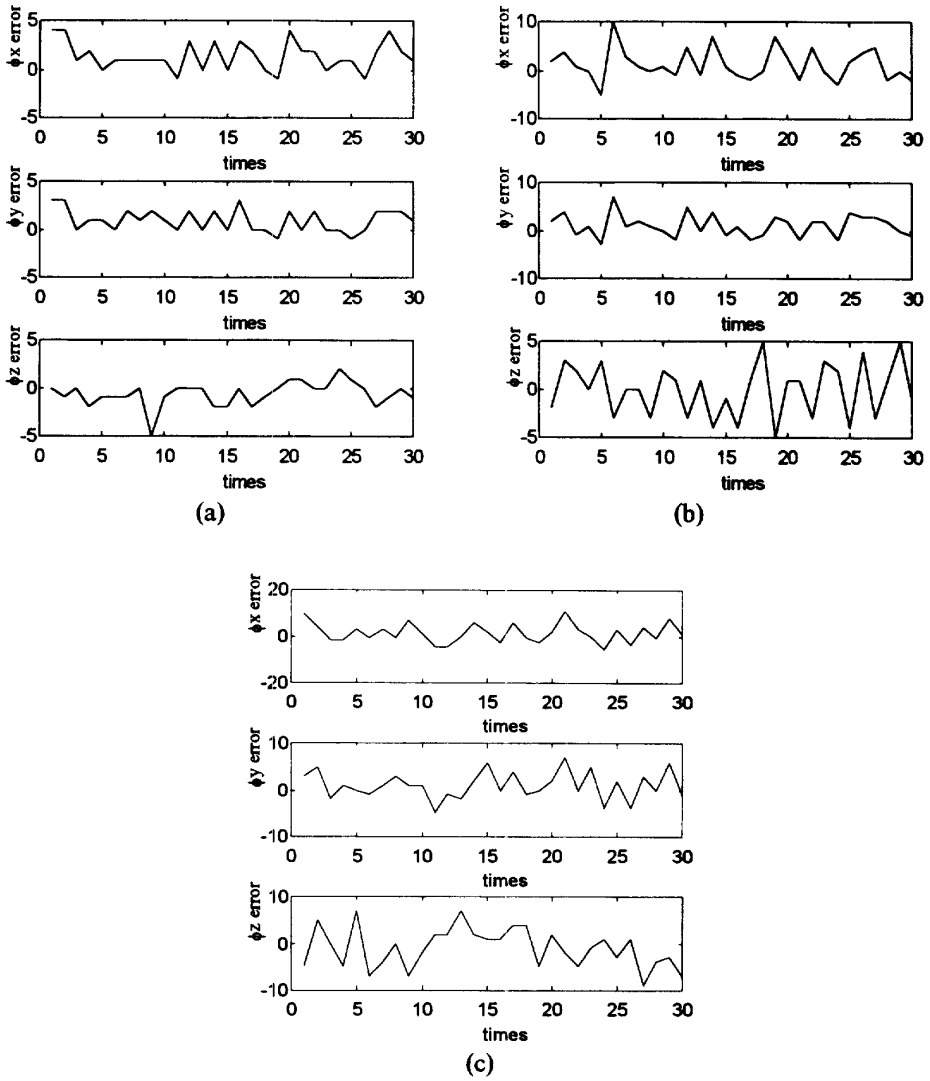

Fig. 10. The error analysis of the wrist rotation angles for the seven points disturbed by random noise with levels (a) 2.5, (b) 5.0, and (c) 7.5 mm.

the 3-D positions of the seven markers with different noise levels, the 3-D positions of the seven markers will be distributed in a small cubic space. The cube indicating the possible distribution for the tip position of the index finger is shown in Fig. 9.

The angular errors of the wrist and the joint angles of the fingers can be obtained from the difference between the input joint angles of simulated gesture and the identified joint angles of the hand model fitted with the noise-disturbed 3-D finger tip positions. Figs 10 and 11 show the angular error of the wrist and the little finger for different noise levels: 2.5, 5, and 7.5 mm. The angular error of the other fingers is similar to the little finger; here, we only illustrate the angular error analysis for the little finger. When the noise level approaches 7.5 mm, the maximum angular error of the wrist is about 12° found in the ϕ_x direction. The angular error of the wrist generates inaccurate joint angles of the little finger, and the maximum angular error is about 18°. Therefore, the angular error of the wrist fitting will deteriorate angular error of the joint angles of the finger fitting. However, for a noise level below 7.5 mm, the accuracy of identified hand posture is acceptable. In our hand motion tracing system, we limit the maximum angular error of the wrist within 10° in ϕ_x, ϕ_y and ϕ_z directions (noise level below 7.5 mm) such that the fingers are fitted accurately in the finger fitting phase. Through the stereo vision technique, the 3-D positions of the finger tips are generated, of which the maximum position sensing error is within the range of 5.0 mm.

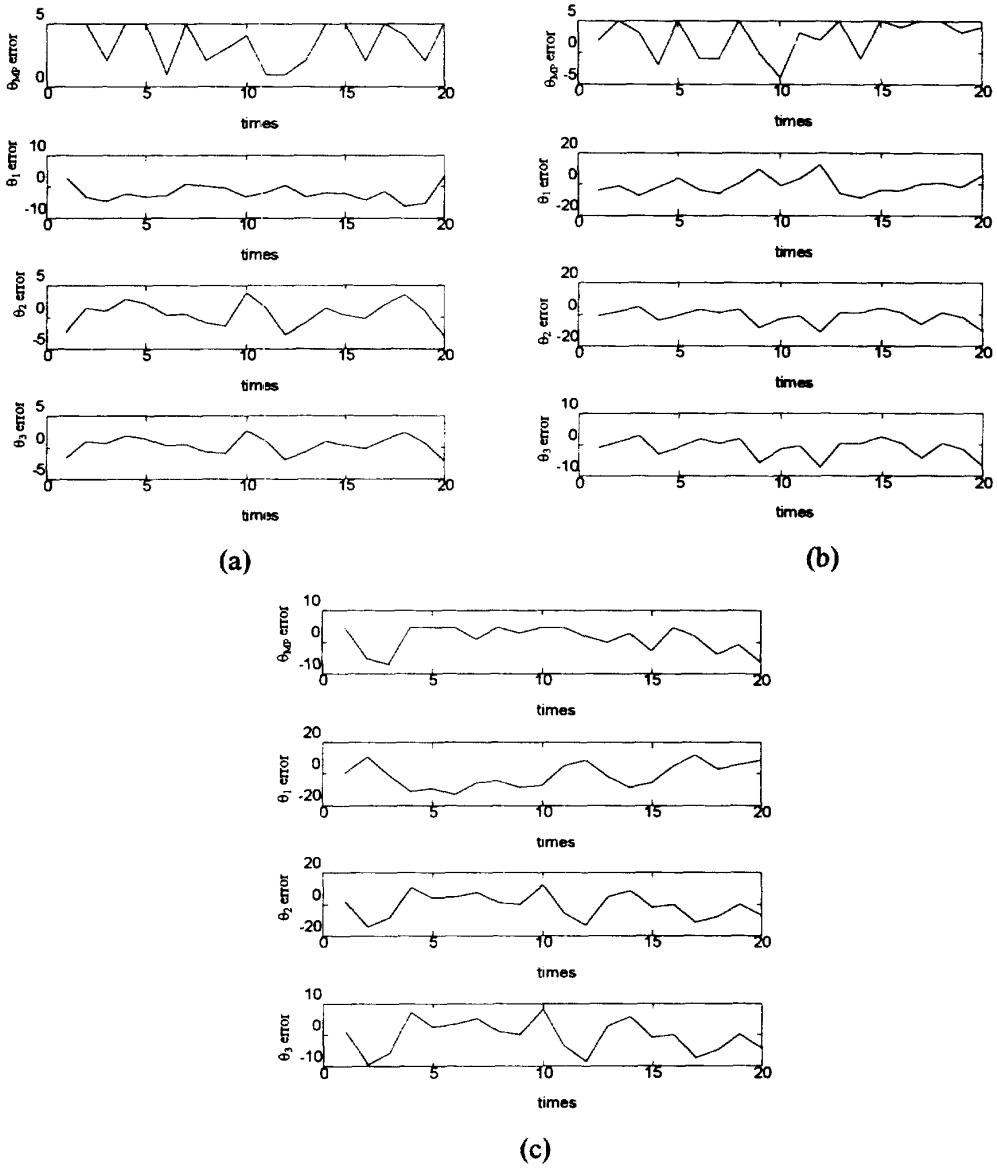

Fig. 11. The error analysis of the joint angles of the little finger for the seven points disturbed by random noise with levels (a) 2.5, (b) 5.0, and (c) 7.5 mm.

4. Experimental results

The fast hand model fitting consists of two new techniques. The first technique generates a new closed-form solution for the finger kinematics problem which will be mentioned in wSection 4.1 and Section 4.2. The second new technique applies the hand motion constraints to the hand model fitting the results of which will be mentioned in Section 4.3 and Section 4.4.

4.1. The possible ranges of r and η for the closed-form solutions

For different hand postures, the values of r and η will change. Before substituting the values of r and η into the closed-form solutions for the five fingers, we have to specify the variation range of r and η. The relationship between the values of r and η for the index finger is plotted in Fig. 12. The curves in Fig. 12(a) show the possible values of r and η. In Fig. 12(b), the bottom curve and the top curve are

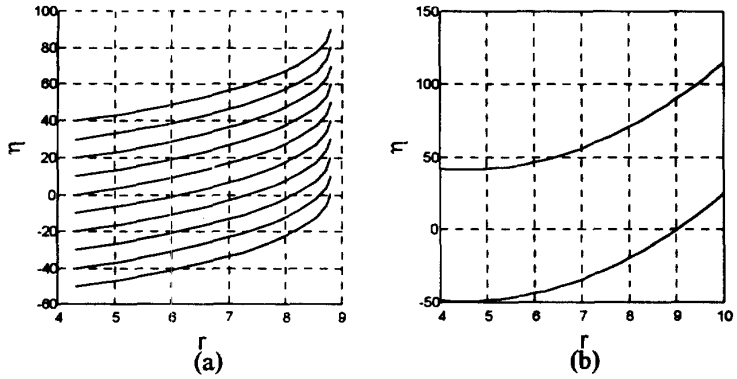

Fig. 12. (a) The values of the r and η are calculated from the tip positions of the five fingers. (b) The lower bound and upper bound of (a) are represented by the second-order polynomials.

represented by second-order polynomials. The bottom curve and the top curve are the lower and upper bounds for r and η, and these curves for the five fingers are generated.

4.2. The accuracy of the closed-form solution for the five fingers

Because the procedure for finding closed-form solutions requires the distance r and slant angle η relative to the MP joint positions, we apply the method of coordinate transformation to acquire the distance r and slant angle η. This transformation is described in Eq. (17). The distance r and slant angle η are obtained from the vector p_{MP}. The analysis accuracy of the closed-form solution is performed by simulating all the possible combinations of the θ_2, and θ_3, and then calculating the r and η values which are defined in Section 2. The range of θ_2 and θ_3 for the five fingers are described as follows:

$$\begin{cases} 0° \le \theta_2 \le 45°,\ 0° \le \theta_3 \le 90° \text{ for the thumb;} \\ 0° \le \theta_2 \le 90°,\ 0° \le \theta_3 \le 90° \text{ for the index, middle,} \\ \text{ring, and little fingers.} \end{cases}$$

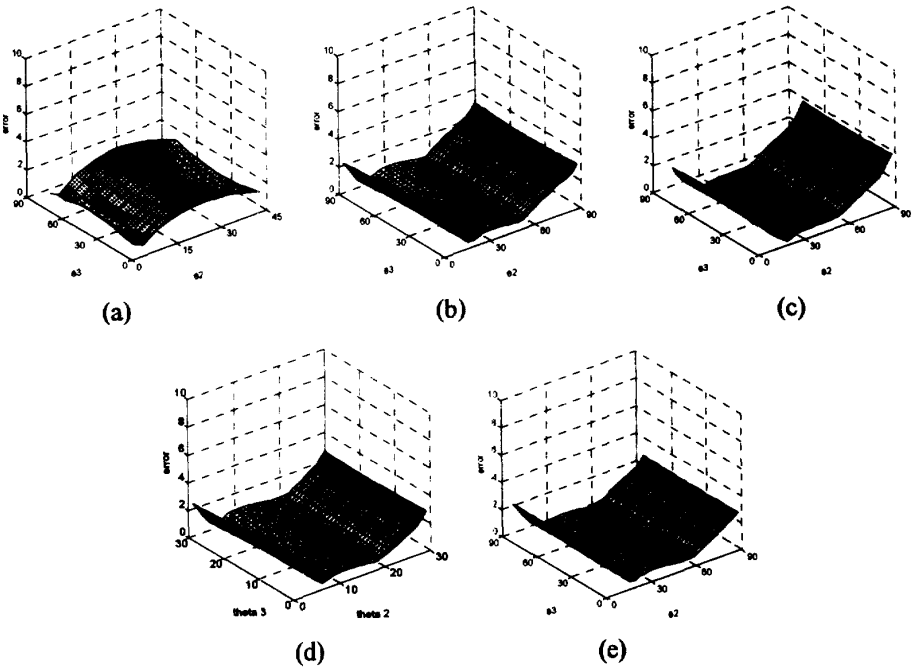

Fig. 13. The error between the simulated data and the data acquired from the closed form solutions for: (a) thumb, (b) index finger, (c) middle finger, (d) ring finger, and (e) little finger.

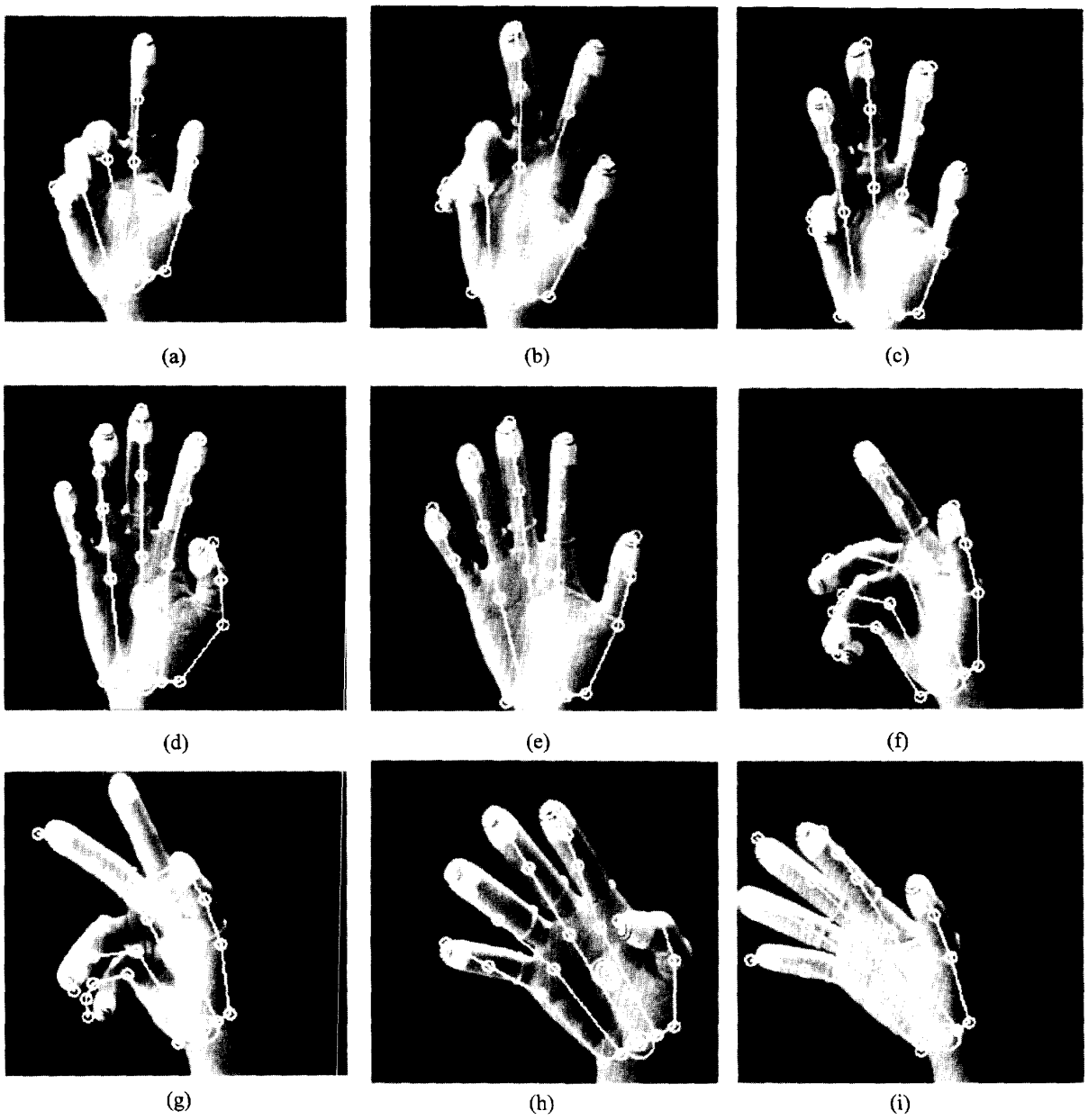

Fig. 14. The images of the nine gestures which are overlaid with their skeleton hand model.

The criterion for the error analysis is measurment of the mean absolute error defined as

$$e = \sum_{i=1}^{n} abs(e(\theta_i))/n, \tag{19}$$

where $e(\theta_i) = \theta_i - \theta_i'$, θ_i is the simulated angle for the joint i, and θ_i' is the angle acquired from the closed-form solutions. Fig. 13 shows the mean absolute error between the simulated angles and the angles acquired from the closed-form solutions for each finger. It shows that the maximum error is about 3° and the average error is about 1.5°.

4.3. The fast hand model fitting

The initial hand model is very important for the hand model fitting. Fingers II–IV are bent as the initial model, such that in the wrist fitting phase the wrist can be fitted accurately. The recognition of the gestures is performed by

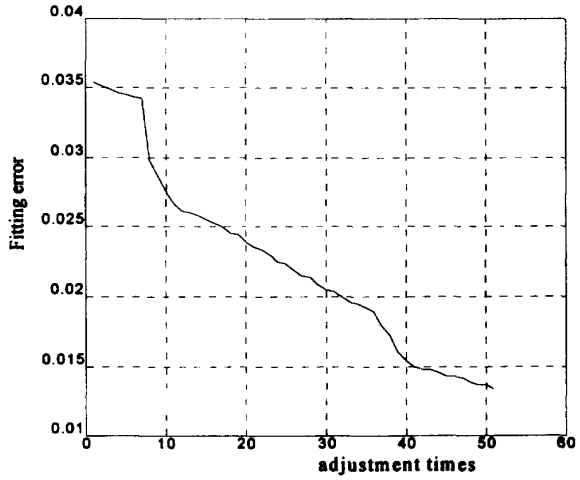

Fig. 15. The fitting error rate for the hand model fitting process.

fitting the hand model to the nine different gestures and the images of the 9 gestures are overlaid with their skeleton hand models. For each hand model fitting, the computation time of the fitting process executed by a Pentium-90 processor takes about 12 s. The fast hand model fitting can fit the hand model to the real hand efficiently for further gesture recognition, the results of which are shown in Fig. 14.

4.4. The fitting error of the fast hand model fitting process

The algorithm of the fast hand model fitting consists of three phase: the wrist fitting phase, the fingers fitting phase, and the thumb fitting phase. The direction of the palm is computed by using the method of wrist fitting, and then the fingers are allocated by the process of the inverse/forward-kinematics. The processes of the inverse-kinematics are implemented by using the closed-form solution functions, and the processes of the forward-kinematics are used to reduce the inaccuracy of the solutions. After the processes of the wrist fitting and inverse-kinematics, the error of the fitted hand model is analyzed by the following formula:

$$E(j) = \sum_{i=1}^{5} d_i(j) / \sum_{i=1}^{5} l_i \tag{20}$$

where $E(j)$ is the error rate for jth adjustment, d_i is the distance from the tip position of finger i of the fitted hand model to the tip position of the same finger i of the real hand, and l_i is the length of finger i. Fig. 15 shows the fitting error for the posture shown in Fig. 14(e). The fitting error converges quickly to the value of 0.013, and the fitting process running on a pentium-90 PC requires 12 s. In Lee and Kunii's model-fitting method, most of the computation time (99.2% of processing time) is spend on the inverse-kinematics process, and the total processing time is about 45 min on an Iris workstation. Our fast model-fitting method has greatly improved the fitting algorithm in both computation efficiency and accuracy.

5. Conclusion and future work

This paper has proposed a fast model-fitting method, which consists of two new techniques that can speed up the fitting process. The first technique generates closed-form solutions for the inverse-kinematics process. The second technique consists of fast hand fitting methods, which has greatly improved the fitting efficiency. However, this paper has not considered the size variation and marker occlusion problems. The size difference between the hand model and the real hand will introduce large fitting error which will be one of the main concerns in future studies.

Appendix A

The polynomials for fingers I–V are(a) thumb:

$$\eta = -1 \times 10^{-4}\theta_1^2 + 1.24 \times 10^{-2}\theta_1 r + 14.015 r^2 - 1.0457\theta_1 - 125.02 r + 332.78$$

$$r = -2.1947 \times 10^{-4}\theta_2^2 - 2.17 \times 10^{-4}\theta_2 + 6.0$$

(b) index finger:

$$\eta = -1 \times 10^{-3}\theta_1^2 + 7 \times 10^{-4}\theta_1 r + 1.8414 r^2 - 0.9336\theta_1 - 15.305 r + 72.2669$$

$$r = -6.15 \times 10^{-4}\theta_2^2 - 7.434 \times 10^{-4}\theta_2 + 8.8$$

(c) middle finger:

$$\eta = -7 \times 10^{-4}\theta_1^2 + 7 \times 10^{-4}\theta_1 r + 1.3781 r^2 - 0.9542\theta_1 - 14.7677 r + 84.7329$$

$$r = -7.29 \times 10^{-4}\theta_2^2 - 9.32 \times 10^{-5}\theta_2 + 10.39$$

(d) ring finger:

$$\eta = 3 \times 10^{-4}\theta_1^2 - 1.79 \times 10^{-2}\theta_1 r + 1.2592 r^2 - 0.9208\theta_1 - 8.7488 r + 54.7401$$

$$r = -6.2 \times 10^{-4}\theta_2^2 - 1.3 \times 10^{-3}\theta_2 + 9.3$$

(e) little finger:

$$\eta = -3 \times 10^{-4}\theta_1^2 - 1.9 \times 10^{-3}\theta_1 r + 2.2785 r^2 - 0.9764\theta_1 - 16.1729 r + 70.4728$$

$$r = -5.2 \times 10^{-4}\theta_2^2 - 1.11 \times 10^{-3}\theta_2 + 7.7$$

References

[1] D.J. Sturman, D. Zeltzer, A survey of glove-based input, IEEE Computer Graphics Application, 1994, pp. 30–39.
[2] J.J. Kuch, T.S. Huang, Vision based hand modeling and tracking for virtual teleconference and telecollaboration. ICCV, 1995, pp. 666–671.
[3] J. Davis, M. Shah, Visual gesture recognition. IEE Proc. Vis Image Signal Process. 141 (2) (1994) pp. 101–106.
[4] T.J. Darrell, A.P. Pentland, Recognition of space-time gestures using a distributed representation. MIT Media Laboratory Vision and Modeling Group Technical Report No. 197.
[5] S. Tamura, S. Kawasaki, Recognition of sign language motion images, Pattern Recognition 21 (4) (1988) 343–353.
[6] I.J. Mulligan, A.K. Mackworth, P.D. Lawrence, A model-based vision system for manipulator position sensing, in: IEEE Proc. Workshop on Interpretation of 3D Scenes, 1989, pp. 186–193.
[7] J. Lee, T.L. Knuii, Model-based analysis of hand posture. IEEE Computer Graphics Application, 1995, pp. 77–86.
[8] J. Lee, T.L. Knuii, Constraint-based hand animation. Models and Techniques in Computer Animation, Springer-Verlan, Tokyo: 1993, pp. 110–127.
[9] J.M. Rehg, T. Kanade, Digiteyes: vision-based hand tracking for human-computer interaction, in IEEE Workshop on Motion of Non-rigid and Articulated Objects, 1994, pp. 16–22.
[10] P. Wellner, Digitaldesk, Commun. ACM 36 (7) (1993) 87–96.
[11] H. Rijpkema, M. Girard, Computer animation of knowledge-based human grasping, Computer Graphics Proc. Siggraph 25 (4) (1991) 339–347.
[12] N.I. Badler, K.H. Manoochehri, G. Walters, Articulated figure positioning by multiple constraints, IEEE Computer Graphics & Application, 1987, pp. 28–38.
[13] P. Lancaster, K. Salkauskas, Curve and Surface Fitting, Academic Press, 1986.
[14] A.J. Koivo, Fundamentals for control of robotics manipulators, John Wiley & Sons, Inc., 1989.
[15] S.C.A. Thomopoulos, R.Y.J. Tam, An interactive solution to the inverse kinematics of robotics manipulators, Mech. Mach. Therory 26 (4) (1991) 359–373.
[16] R. Jain, R. Kasturi, B. Schunck, Machine Vision, McGraw-Hill Inc., 1995.
[17] B.K.P. Horn, Robot Vision, MIT Press, 1986.