

3D HAND POSE RETRIEVAL FROM A SINGLE 2D IMAGE

Haiying Guan Chin-Seng Chua Yeong-Khing Ho

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798
ehyguan@ntu.edu.sg

ABSTRACT

Current research efforts about passive sensing of 3D human hand pose have been hampered by the computational complexity caused by inverse kinematics and 3D reconstruction. In this paper, we focus on 3D hand posture estimation based on a single 2D image. We analyze some of its constraints to reduce the 27 DOFs to 12 DOFs without any significant degradation of performance. A novel algorithm to estimate the 3D hand posture from eight 2D projected feature points is proposed. Experimental results confirm that our algorithm gives good estimates of the 3D hand pose.

Keywords: 3D hand posture estimation, model-based approach, gesture recognition.

1. INTRODUCTION

Hand posture analysis is an interesting research field and has received much attention in recent years. General solutions for posture analysis are divided into two categories: glove-based techniques and vision-based techniques. Vision is a non-invasive way of sensing and more suitable for passive sensing such as remote robot control. Vision-based techniques can be classified into two types: appearance-based approach [1] and 3D model-based approach.

Conventional model-based methods are mainly used in two areas: 3D hand tracking [2] and 3D hand posture estimation. By analysis of static and dynamic motions of the human hand, Lee *et. al* [3] present some constraints on the joints and use them to simulate the human hand in real images. Shimada *et. al* [4] present a method to track the pose (joint angles) of a moving hand and refine the 3D shape (widths and lengths) of the given hand model from a monocular image sequence.

Although the above algorithms have promising results, posture estimation is not yet advanced enough to provide a flexible and reliable performance for potential applications. Model-based methods always involve finding the inverse kinematics which are in general ill-posed. It is obviously a task of computational complexity to estimate these kinematic parameters from the detected features. In addition,

previous methods on 3D require the use of multiple cameras, which not only is resource-consuming, but also needs some form of 3D reconstruction which itself is computationally intense.

In this paper, the goal of our work is to avoid the complex computation of inverse kinematics and 3D reconstruction without the motion information; that is, without using 3D information, we propose a new approach to estimate the 3D hand posture from a single 2D image. Preliminary results can be found in [5] [6], which deals only with finger posture. This paper extends the idea further to compute the 3D posture for the entire hand. First, we analyze the human hand model with 27 DOFs and its constraints. Second, we develop an algorithm to estimate the 3D hand posture by using eight feature points to retrieve the 3D hand posture.

This paper is organized as follows: Section 2 discusses the hand model and its constraints. Section 3 presents the methodology to estimate the hand posture. Two test cases involving various degrees of finger-extension are investigated in Section 4.

2. HAND MODEL AND ITS CONSTRAINTS

2.1. 27 DOFs hand model

Lee *et. al* [3] defined a hand model with 27 DOFs. The joints of the human hand are classified into three kinds: flexion, directive or spherical joints which consist of 1 DOFs (extension/flexion), 2 DOFs (one for extension/flexion and one for adduction/abduction) and 3 DOFs (rotation) respectively (see Figure 1). For each finger, there are four DOFs described by $\theta_1 - \theta_4$. The thumb has five DOFs described by $\theta_1 - \theta_5$. Including the 6 DOFs for the translation and rotation of the wrist, the model has 27 DOFs.

2.2. Model constraints

Conventional models of the human hand are lacking in constraints. It limits their usefulness in computer vision and animation. The lack of constraints leads to unnatural model behavior. On the other hand, because the movements of the

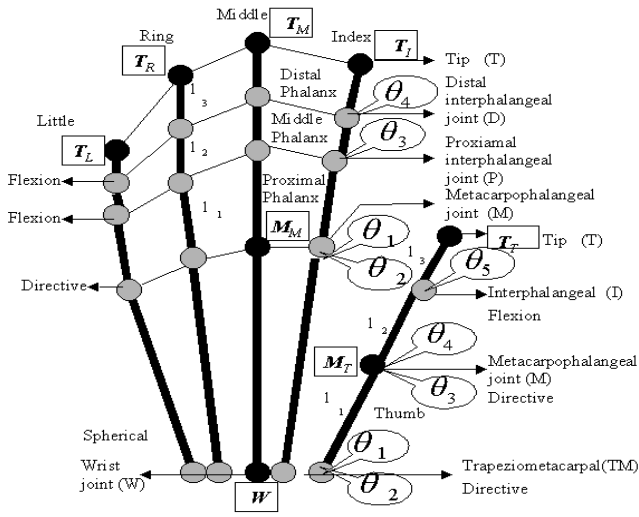


Fig. 1. 27 DOFs hand model

fingers are inter-dependent in the human hand and in order to reduce search space of matching, constraints are essential to further realize the hand motion. By careful analysis of the constraints proposed by Rijpkema, Kuch, Lee and the hand model we have established, we adopt some other static constraints.

The following constraints is proposed by Rijpkema [7]:

Constraint 1

The angles of *D* (Distal) joints and *P* (Proximal) joints are dependent:

$$\theta_4 = \frac{2}{3}\theta_3 \quad (1)$$

where θ_3 and θ_4 represent the extension/flexion DOF for the *P* and *D* joints of each finger (see Figure 1).

Constraint 2

The thumb is a more complicated manipulator, because a large part of the thumb is part of the palm and the joints are moving along non-trivial axes. However, it is known that the kinematics of the thumb can be calculated almost uniquely by experimental observations:

$$\theta_1 = 2 * (\theta_3 - \frac{1}{6} * \pi) \quad (2)$$

$$\theta_2 = \theta_4 * \frac{7}{5} \quad (3)$$

The following constraint is proposed by Kuch and Huang [8]:

Constraint 3

The joint angles of *P* and *M* joints of finger have a dependency represented by the following equation:

$$\theta_1 = k\theta_3 \quad 0 \leq k \leq 1/2 \quad (4)$$

where θ_1 and θ_3 represent extension/flexion DOF for the *M* and *P* joints of the finger respectively.

The following constraints are proposed by Lee and Kuniti [3]:

Constraint 4

There is little adduction (*add.*) / abduction (*abd.*) of the *M* (metacarpophalangeal) joint of the middle finger; that is:

$$\theta_2 = 0 \quad (5)$$

According to the hand model we have established, we propose the following constraints:

Constraint 5

Five points (Wrist joint, Metacarpalangeal joint, Proximal joint, Distal joint and Tip represented by *W*, *M*, *P*, *D* and *T* respectively) of each finger are coplanar. We define this plane as the “finger plane”.

According to Constraint 4, we omit the adduction / abduction DOF for the *M* joint, so that *M*, *P*, *D* joints of four fingers are all extension/flexion joints. For accuracy, we may define a different *W* point for each finger (see Figure 1).

Constraint 6

The joint angles of *I* and *M* of the thumb have a dependency represented by the following equation:

$$\theta_5 = a\theta_4 \quad a \geq 0 \quad (6)$$

Constraint 7

Four points (*TM*, *M*, *I* and *T*) of the thumb are coplanar. We define this plane as the “thumb plane”. In the experiments, the *TM* joint is approximated by the *W_T* point (see Figure 1).

Constraint 8

We assume that the *M* joint of each finger are located in the plane called “palm plane” and this plane is perpendicular to the “finger plane” of the middle finger.

Using Constraint 1, 3, 4 and 5, we omit 3 DOFs for each finger. Using Constraint 2 and 6, we omit 3 DOFs for the thumb. Over all, our simplified model is reduced from 27 DOFs to 12 DOFs.

3. HAND POSTURE ESTIMATION

3.1. Problem description

The purpose of this part is to analyze the geometric characteristics of the hand and replicate its pose based on 2D projected information. Without loss of generality, we assume that the world coordinate frame is aligned with the camera frame; in other words, the image plane coincides with the *X – Y* frame of the world. In our experiments, we adopt color markers to identify the feature points of the joints and tips.

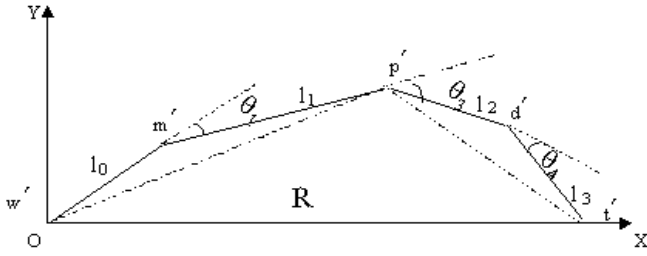


Fig. 2. Pentagon geometry

3.2. Finger posture estimation

3.2.1. Solution for five points in the 2D “finger plane”

For a particular finger, we define the 3D distance between the tip of finger T to the wrist joint W as R . We denote the positions of W , M , P , D and T in the “finger plane”¹ The coordinate system is oriented such as is shown in Figure 2. In the pentagon (see Figure 2) which consists of the five points W , M , P , D and T , the lengths of l_0 , l_1 , l_2 and l_3 are constant parameters. The relationship between θ_3 and R can be approximated by a cubic polynomial. Using a regression method, for a given R , we can find the shape of the pentagon. Hence we can conclude that if the distance R between the W (W joint) and T (Tip) is given, the shape of the pentagon is determined. The positions of W' , M' , P' , D' and T' are obtained at the same time.

3.2.2. 3D solution for the five points

The task at hand is to find the 3D positions (relative to the camera frame) of the finger joints and tip (defined as W , M , P , D and T) given any three 2D image positions of these points (defined as w , m , p , d and t). For ease of description, we define the 3D coordinates of the detected finger points as P_1 , P_2 , and P_3 , which are any three points of W , M , P , D and T . Their orthographic projection coordinates in the image plane are: p_1 , p_2 , and p_3 . Taking P_1 as the reference, the positions of P_2 and P_3 are defined by P_{21} and P_{31} . Rotating them to the finger plane, their new 2D coordinates are p'_1 , p'_2 , p'_3 respectively.

In order to find the solution for this problem, the following steps are proposed (See [5] for detail):

First, obtain the 3D hand model parameters (the lengths of the finger links or l_1 , l_2 and l_3) from the 2D image information by using the feature extraction method. For simplicity of computation, we treat Y_{31} as a variable and R as known. For each possible value of R , we calculate the positions of w' , m' , p' , d' , t' in 2D plane by using the algorithm described in Section 3.2.1. We obtain the distances among

¹Positions with “'” are used to denote oriented finger plane positions while those without “'”, which will be introduced in Section 3.2.2, are used to denote positions referenced from the image plane.

P_1 , P_2 and P_3 and let them be l_{21} , l_{23} and l_{31} (the distance between P_2P_1 , P_2P_3 and P_3P_1 , respectively). Z_{21} is calculated by Euclidean distance.

Second, obtain Y_{31} and Z_{31} using the Euclidean distance. Calculating all possible values of R , we obtain the relationship between Y_{31} and R . We approximate the relationship between Y_{31} and R as the quadratic curve. Using the regression method, the curve function can be approximated. For the actual Y_{31} , we obtain R .

Using the computed R , we find the 2D solutions of w' , m' , p' , d' and t' with the method described in Section 3.2.1. The equivalent 3D coordinates of W' , M' , P' , D' and T' in the “finger plane” are extracted.

Finally, compute the rotation matrix mapping points P'_1 , P'_2 and P'_3 to points P_1 , P_2 and P_3 . Obtain the 3D position of the other points with the rotation matrix.

3.3. Thumb posture estimation

Combined with the Constraint 4 (Eq. (2) and (3)) and Constraint 6 (Eq. (6)), we use a similar approach as above to obtain the four-point solution for the thumb. In the algorithm, we use the three 2D feature points of the thumb (that is, TM , MP and T) to calculate the 3D positions of the four points: TM , MP , IP and T (see Figure 1).

3.4. Hand posture estimation

Based on the algorithms for finger posture estimation and thumb posture estimation which is described in Section 3.2 and Section 3.3. We present a hand posture estimation algorithm by using eight feature points of the hand; that is, the wrist point (W), the metacarpal joint of the middle finger (M_M), the metacarpophalangeal joint of the thumb (M_T), the four tips of the fingers (T_I , T_M , T_R and T_L) and the tip of the thumb (T_T).

From the feature extraction stage, we obtain the W point of the middle finger. For the 2D positions of the three feature points of middle finger (that is, W , M_M and T_M), we use the finger posture solution algorithm (Section 3.2) to retrieve the middle finger posture.

Once the 3D posture of the middle finger is determined, the “finger plane” for the middle finger in 3D space is known. According to Constraint 8 (see Section 2.2 for details), the “palm plane” is perpendicular to the “finger plane” of the middle finger. This constraint, together with the knowledge of two points (M_M and W joints) on the palm, constraints the palm plane. From this plane, and using the predetermined hand model, the positions of the metacarpal (M) joint of the index, ring and little finger can be identified. To obtain better finger plane estimations for the other fingers (index, ring and little finger), different wrist points for each finger (W_I , W_R and W_L) are approximated using the hand model.

For the three feature points (W , M and T) of each of index, ring and little fingers, we use the finger posture algorithm (see Section 3.2 for details) to retrieve the 3D posture. In the same way, for the three feature points (W_T , M_T and T_T) of the thumb, the thumb posture is retrieved by the thumb posture algorithm. Finally, we synthesize the hand posture in 3D space.

4. EXPERIMENTS

For simplicity of feature extraction, color markers are commonly used by other research groups to identify the main feature points. We use color markers to detect the eight main points in the real image. Eight points (W , M_T , M_M , T_T , T_I , T_M , T_R and T_L) are identified from the centroids of the detected color regions (see Figure 1 for definition)².

The tests (Figure 3) show hand postures with different degrees of extension for each finger. For each posture, we show the original images (left image), the estimated 3D pose overlaid on the original image (center image) and the 3D pose from a different viewpoint (right image). We show in Figure 3 (c) that the posture for non-frontal views of the hand can be retrieved as well. The constraint that we place on our experiments is that all eight color markers (identifying the 8 positions of W_T , M_T , M_M , T_T , T_I , T_M , T_R , T_L) be visible. Occlusion of one or several markers is out of the scope of this paper and is currently under investigation.

5. CONCLUSION

The new algorithm is promising because of the following reasons: First, the algorithm uses 2D positions of the feature points and avoids the computational complexity caused by the 3D reconstruction. Second, the algorithm does not involve the computation of inverse kinematics. Third, the algorithm uses a only single 2D image to retrieve the 3D hand posture. There is no need to know the motion information of fingers or thumb which can only be provided by a sequence of images. Experimental results have shown that our 3D hand posture estimation works well even with real images.

6. REFERENCES

7. REFERENCES

[1] T.E. Starner, *Unencumbered Virtual Environments*, Ph.D. thesis, MIT Media Arts and Sciences Section, USA, 1993.
 [2] J. M. Rehg and T. Kanade, "Visual tracking of high DOF articulated structures: An application to human hand tracking," in *ECCV94*, 1994, pp. B:35–46.

²The subscripts represent different fingers and thumb.

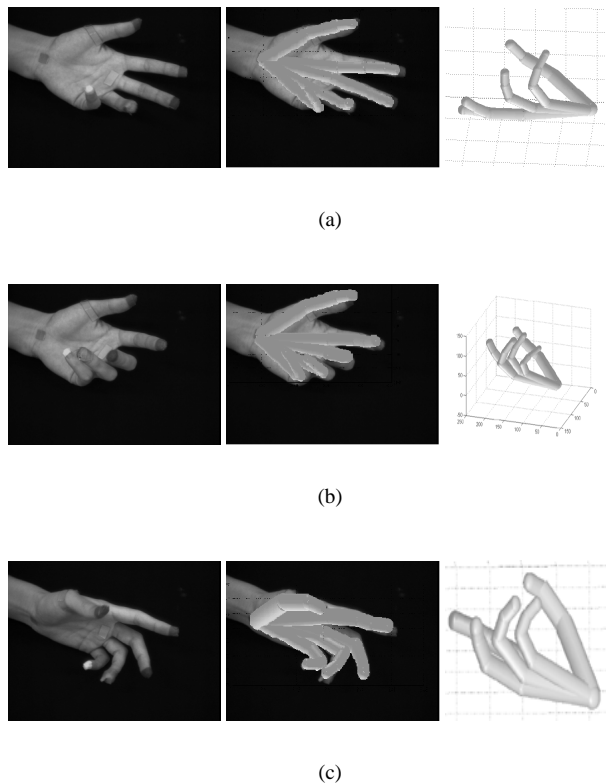


Fig. 3. Hand postures with different degrees of extension for each finger (using color markers)

[3] J. Lee and T. L. Knuii, "Model-based analysis of hand posture," *IEEE Computer Graphics and Application*, pp. 77–86, Sept. 1995.
 [4] N. Shimada, Y. Shirai, Y., and J. Miura, "Hand gesture estimation and model refinement using monocular camera - ambiguity limitation by inequality constraints," in *The Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, Los Alamitos, Calif, 1998, pp. 268–273, IEEE computer Society Press.
 [5] C.S. Chua, H. Y. Guan, and Y. K. Ho, "Model-based finger posture estimation," in *Fourth Asian Conference on Computer Vision*, Taipei, Taiwan, 2000, pp. 43–48.
 [6] H. Y. Guan, C.S. Chua, and Y. K. Ho, "Hand posture estimation from 2d monocular image," in *Second International Conference on 3-D Digital Imaging and Modeling*, Ottawa, Canada, 1999, pp. 424–429.
 [7] H. Rijkema and M. Girard, "Computer animation of knowledge-based human grasping," *IEEE Computer Graphics and Application*, vol. 25, no. 4, pp. 339–348, July 1991.
 [8] J. M. Kuch and T. S. Huang, "Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration," in *ICCV95*, 1995, pp. 666–671.