Real-time 3D hand shape estimation based on inverse kinematics and physical constraints

Fujiki, Ryuji
Department of Intelligent Systems, Kyushu University

Arita, Daisaku
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

https://hdl.handle.net/2324/5910

出版情報：Proceedings of International Conference on Image Analysis and Processing || || p850-858, pp.850-858, 2005-09

Real-time 3D Hand Shape Estimation based on Inverse Kinematics and Physical Constraints

Ryuji Fujiki, Daisaku Arita and Rin-ichiro Taniguchi

Department of Intelligent Systems, Kyushu University
6-1, Kasuga-koen, Kasuga, Fukuoka, 816-8580, JAPAN
{fujiki,arita,rin}@limu.is.kyushu-u.ac.jp

Abstract. We are researching for real-time hand shape estimation, which we are going to apply to user interface and interactive applications. We have employed a computer vision approach, since unwired sensing provides restriction-free observation, or a natural way of sensing. The problem is that since a human hand has many joints, it has geometrically high degrees of freedom, which makes hand shape estimation difficult. For example, we have to deal with a self-occlusion problem and a large amount of computation. At the same time, a human hand has several physical constraints, i.e., each joint has a movable range and interdependence, which can potentially reduce the search space of hand shape estimation. This paper proposes a novel method to estimate 3D hand shapes in real-time by using shape features acquired from camera images and physical hand constraints heuristically introduced. We have made preliminary experiments using multiple cameras under uncomplicated background. We show experimental results in order to verify the effectiveness of our proposed method.

1 Introduction

Human hands, expressing our intension, are often used for communication. Therefore, hand shape recognition can be used in various interactive applications and user interface. We have developed hand shape estimation based on a computer vision approach, since unwired sensing provides restriction-free observation, or a natural way of sensing. There are, in principle, two approaches for hand shape estimation. One is classification of hand shapes into predefined categories based on pattern recognition techniques. The other is measurement of arbitrary hand shapes in 3D space. Though, the former approach can be used to indicate symbolic information, such as command labels of interaction, it can not present continuous information such as changes of hand shapes. On the other hand, the latter approach can acquire hand shape parameters in 3D space, or continuous information about 3D hand shape, and, therefore, it can be also applied to hand-manipulation-based control, such as control of robot hands, real-time 3D animation, etc.

Considering the applicability, we have adopted the latter approach. In this paper, we propose a novel method to estimate 3D hand shapes in real-time by using shape features acquired from camera images and physical hand constraints heuristically introduced. At first, we present the representation of a 3D hand model used our system. Then, we describe the details of hand shape estimation: extraction of hand features and 3D hand shape estimation by Inverse Kinematics with hand constraints. Finally, we show some experimental results in order to verify the effectiveness of our proposed method.

2 Related Works

The aim of our research is real-time estimation of 3D hand shape. Basically, there are two approaches proposed for 3D hand shape estimation.

- 2D appearance-based approach[1, 2]
- 3D model-based approach[3, 4]

The former is based on appearances of hands in 2D images, and essentially consists in a kind of template matching. Shimada et al.[1] represented variations of possible shape appearances as Locally-Compressed Feature Manifold (LCFM) in an appearance feature space. It is effective to prevent the system from tracking failures and to reduce the search area. Stenger[2] used a tree-based estimator based on Bayesian filter. This approach achieves coarse to fine search by approximating the posterior distribution at multiple resolutions, and hopeless sub-trees of the search space are not further evaluated.

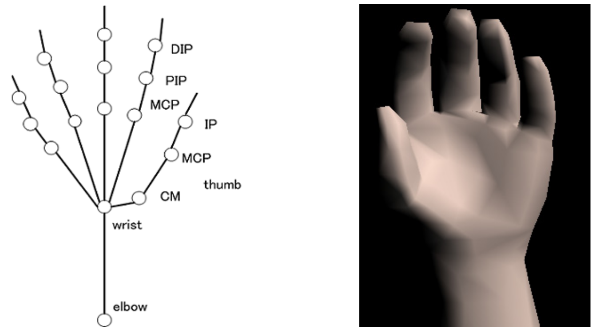
The latter is a method of extracting local hand features from images and estimating hand shapes, fitting a 3D hand model to the features. Ueda et al.[3] demonstrated the following method. Voxel representation is reconstructed from silhouette images by a multi-viewpoint camera system. Then a 3D hand shape is estimated using model fitting between a 3D hand model and the voxel model. Lu et al.[4] used a dynamic model and image force is calculated from image edges, optical flows and shading information. Then they perform fitting the dynamic model to image data applying image forces to the hand model.

Simply speaking, the former approach has the problem to deal with a large amount of templates. The latter has the problem of image feature missing by self-occlusion, since a human hand has many joints. Our approach is to extract the robust hand shape features from the images as much as possible even if it is not sufficient to know an exact hand shape. We estimate 3D hand shapes from the extracted features using hand constraints and geometric parameters of the hand model. In our system, based on the estimation result, we can re-extract effective image features by limiting a search range of image feature detection, and improve the estimation result.

Here, we use arcs on image contours obtained by image analysis, and we identify which finger (or fingertip) arc is detected. Then, we calculate the 3D positions of the arcs using multi-view analysis. Finally, we estimate hand shapes from the features by Inverse Kinematics with hand constraints.

3 Hand Model

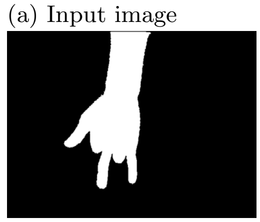
In principle a human hand is a non-rigid object. In this paper, a human hand is approximated by a 3D rigid articulated object. This 3D hand model consists of a skeleton model and a skin model(see Fig.1). The skeleton model consists of joints linked others, and has parent-child relation among the joints. Skin model is not currently used for hand shape estimation, except for visualization of results of hand shape estimation. In each finger, DIP (Distal Interphalangeal) and PIP



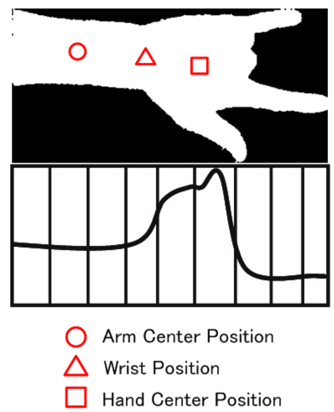
(a) Skeleton Model (b) Skin Model

Fig. 1. Hand Model

(a) Input image



(b) Skin color image



(c) Feature Points

Fig. 2. Image analysis

(Proximal Interphalangeal) joints have 1 DOF (degree of freedom), and MCP (Metcarprophalangeal) has 2 DOFs. The wrist joint has 2 DOFs of yaw and pitch rotations. The elbow joint has 4 DOFs of translation and roll rotation, though, in this paper, the elbow is not fully estimated.

4 Feature Extraction of Hand

We extract shape features using image analysis. At first, we extract skin color region: we convert a color input image into a hue image (see Fig.2(a)) and, then, the image is smoothed for removing noises and binarized (see Fig.2(b).

4.1 Non-finger Feature Points

First, we extract non-finger shape features as follows (see Fig.2(c)).

1. Wrist Position
   We find a minimal-sized rectangle for extracted contour. This rectangle is normalized for rotation: x-axis is set to coincide with the arm direction. Generally, a hand is wider than an arm. Therefore we find the wrist position by searching for the minimum number of skin-color pixels projected on x-axis.

2. Arm Center Position
   The arm center position is computed as the centroid of skin-color region which is at the left part of the wrist position.
3. Hand Center Position
   The centroid usually dose not coincide with hand center, because it is strongly influenced by hand's opening and closing. We calculate an approximated distance from every binary image pixel to the nearest contour pixel. Then the position of a pixel with the maximum value is judged to be the hand center position(x_{hand}, y_{hand})

4.2 Detection of Arcs on the Contour

As major features of 3D hand shape estimation, we detect arcs on the contour. These arcs are used as the positions of end effectors in inverse kinematics mentioned later.

1. Detection of Arcs on the Contour
   We detect arcs on the contour by curvature information, i.e., we detect contour points with large curvatures. The arcs correspond to joints projected outer-most in 2D image space. Here, we have to consider two problems: the correspondence between the arcs and fingers and correspondence between arcs extracted in two cameras. This problems is complex since we can make a lot of combinations. Therefore, we use physical hand constraints, the finger order and intervals between fingers, to reduce the combination.
2. Correspondence between arcs and fingers
   We decide which finger corresponds to each of arcs detected. When we detect five arcs, we can relate their order with the finger order. In other cases, i.e, when we can only detect k(< 5) arcs, we heuristically decide finger positions as follows.
   - We calculate possible finger position F_i (see Fig.3) based on the width of a hand detected and the hand model.
   - We define $C_j(c_{j,1}, \cdots, c_{j,k})$:($j = 0, \cdots$,$\binom{5}{k}$) as a finger set which consists of selected k pieces from F_i.
   - The best combination C_j is determined as $argmin_j \sum_{m=1}^{k} ||c_{j,m} - a_m||$ (m=0,$\cdots$,k), where a_m is the position of *base of protrusion* corresponding to the m-th arc. The base of protrusion is detected by contour tracing as follows:
     - Beginning from an arc point, contours of its both sides are traced at the same speed until a concave point is found in one of the two contours (see Fig.3).
     - a_m is the average of the positions of two points which are reached by the contour tracing.

     When we calculate the distance $||c_{j,m} - a_m||$, only y coordinates of the positions are used. This procedure is introduced to reduce the influence of finger abduction.

   As we decided the best combination, we can estimate which finger could not be detected. In case that a finger is not detected, we estimate its positions by finding a point with local maximum value of curvature whose y coordinate is similar to F_i(the missed finger).

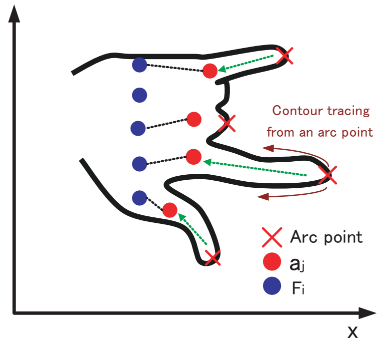


Fig. 3. Correspondence between arcs and fingers

4.3 Pitch and Yaw of Hand

We estimate pitch and yaw of a hand influenced by rotation of the wrist. They are easily calculated from the 3D positions of the arm center, the wrist position, and the hand center, which are defined in Fig.2.

5 Inverse Kinematics with Hand Constraints

5.1 Hand Constraints

A human's hand has many constraints[5–7]. Here, we consider hand constraints are combined with Inverse Kinematics.

1. Movable range of joints
   The movement of each finger is limited by movable ranges of joints. Movable range of each joint is shown Table 1.

Table 1. Movable range of joint

	DIP	PIP	MCP	abduction
Little	$0° \sim 70°$	$0° \sim 90°$	$-30° \sim 90°$	$-20° \sim 20°$
Ring	$0° \sim 70°$	$0° \sim 90°$	$-20° \sim 90°$	$-20° \sim 20°$
Middle	$0° \sim 70°$	$0° \sim 90°$	$-20° \sim 90°$	$-20° \sim 20°$
Index	$0° \sim 70°$	$0° \sim 90°$	$-40° \sim 90°$	$-20° \sim 20°$
Thumb	$-20° \sim 80°$	$-20° \sim 40°$	$-20° \sim 70°$	$-20° \sim 35°$

2. Limitation of abduction
   As we flex our fingers, adduction happens. Inversely, we extend our fingers, abduction happens. Considering this characteristics, we limit abduction of finger linearly with an MCP joint.
3. Interdependence of finger joints
   Each joint has a interdependence as follows:
   – On a grasping, a DIP joint has a linearity relationship with a PIP joint. ($\theta_{DIP} = \frac{2}{3}\theta_{PIP}$)

- As an MCP joint bends, PIP and DIP joints bend slowly at first. Then their joints bend suddenly in the middle of the MCP joint flexion. Finally their joints bend again slowly. In brief, an MCP joint has correlation of an ess curve with PIP and DIP joints. However, a human hand can extend an MCP joint under PIP and DIP joints bended. Therefore we assume a PIP joint makes a constant angle, when an MCP joint angle is smaller than 10 degrees.

On the basis of the above observation, we have heuristically introduced relational expressions shown in Fig.4.

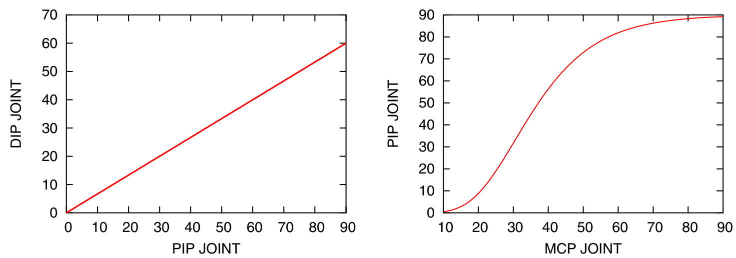


Fig. 4. Interdependence of Finger Joints: The figure above shows relation between PIP and DIP joint, the bottom shows relation between MCP and PIP joint

5.2 Hand Shape Estimation using IK

We estimate joint angles of fingers by Inverse Kinematics based on the hand constraints mentioned above. In general, IK is to determine joint angles $\theta_{\mathbf{i}}$ of a manipulator so that the position of an end effector, or a final point, $\mathbf{P}_n$, coincides with a given goal point $\mathbf{G}$: $\mathbf{P}_n(\theta_1, ..., \theta_n) = \mathbf{G}$: where the manipulator has n segments. Here, the goal is given by an arc position detected and the target is a fingertip or a finger joint of the hand model.

We use Cyclic Coordinate Descent (CCD) method to solve IK[8]. The merit of the CCD method is that it estimates joint angles based on a posture in previous frame and it is a fast algorithm. This is effective for real-time processing.

Identification of Arcs We decide correspondence between fingers (fingertip) and arc points. The algorithm is summarized as follows.
On a finger,

1. We set an arc as the goal, and fingertip as the target of IK.
2. If the distance between wrist and target is longer than the distance between wrist and parent joint, the target is refused. The joint must be on the outside on the wrist, because the arc detected is on the contour.
3. We solve Inverse Kinematics by CCD method, satisfying the hand constraints. Then, we calculate the error between the goal and the target.

4. Unless the target is an MCP joint, make the next target be the parent joint of the current target, and go to 2.
5. We select a joint which has the least error as the correct joint corresponding to the given arc.

Estimation of joint angles Even if we can identify an arc position correctly using the method above, angles of its child joints can not be estimated by Inverse Kinematics, except for the case that the arc position corresponds to a fingertip. We solve this problem by using interdependence of joint angles, and estimate those joint angles.

Table 2. Processing time

Algorithm	Time (msec)
Convert image to hue color space	12
Detection skin color region	12
Extract hand features	20
IK calculation	14
Total	58

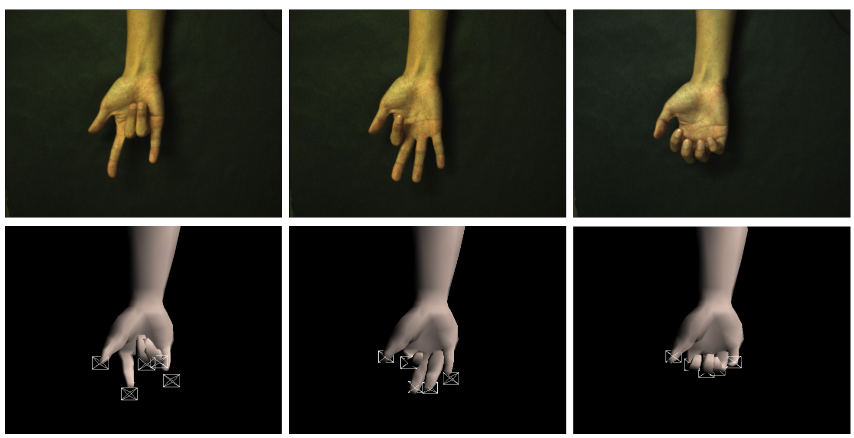


Fig. 5. The result of estimation

5.3 Post processing

We can add the following processing so that we estimate more accurate hand shape. From the error of Inverse Kinematics, we can evaluate false estimation in image analysis. When the error of Inverse Kinematics estimation is pretty large, we suppose that the correspondences between the arc and the finger or the correspondence between two views for depth measurement is not correct. Then we estimate finger angles again by using another correspondence. In addition, we can know hand shape information from the result of Inverse Kinematics, and we have a chance to acquire effective image features such as fingertip position by limiting a search range of image feature detection. We can use those features for more accurate hand shape estimation.

6 Preliminary Experiment

We experimented the 3D hand shape estimation by our proposed method except for the post processing. In this experiment, we have used IEEE-1394-based color cameras (Point Grey Research Inc; Flea) with f:8 mm lenses, which are geometrically calibrated in advance. The images are captured with the size of 640×480 pixels. Several experimental results are shown in Fig.5, which indicates the effectiveness of our method. The processing time is shown Table. 2 using PC with Pentium *IV* (2GHz). It shows that our algorithm can be used for real-time applications.

7 Conclusion

In this paper, we have shown a real-time 3D hand shape estimation without special marker-sensors. The key point is that we use Inverse Kinematics with physical hand constraints in order to complement hand shape information not to be directly obtained image features. We have experimented our proposed method under non-complex background.

The next goal is to build a system which can handle the turn of the palm of the hand, by selecting two cameras which face to the front of a hand from multiple cameras. In other words, we estimate the roll rotation of elbow joint. We also have to extract the shape features which are effective to estimate more accurate 3D hand shape from cluttered background. Acquisition of the geometrical parameters of 3D hand model, such as lengths of finger bones, from the first pose is also an important issue.

References

1. N.Shimada and Y.Shirai, "Hand Posture Estimation based on 2D Appearance Retrieval Using Monocular Camera," Proc. Int, Workshop on RATFG-RTS, pp.23-30, 2001.
2. B.Stenger, A.Thayananthan, P.H.S.Torr, and R.Cipolla, "Filtering using a tree-based estimater," Proc. ICCV, pp.1063-1070, 2003.
3. E.Ueda, Y.Matsumoto, M.Imai and T.Ogasawara. "Hand Pose Estimation for Vision-based Human Interface", IEEE Trans. on Industrial Electronics. Vol.50, No.4, pp.676-684, 2003.
4. S.Lu, D.Metaxas, D.Samaras and J.Oliensis, "Using multiple cues for hand tracking and model refinement," Proc. CVPR, pp.443-450, 2003.
5. C.Häger-Ross and M.H.Schieber, "Quantifying the Independence of Hand Finger Movements: Comparisons of Digit, Hands, nad Movement Frequencies," The Journal of Neuroscience, Vol.20, No.22, pp.8542-8550, 2000.
6. D.G.Kamper, E.G.Cruz, and M.P.Siegel, "Stereotypical fingertip trajectories during grasp," Journal of Neurophysiology, Vol.90, No.6, pp.3702-3710, 2003.
7. G.ElKoura, K.Singh "Handrix: Animating the Human Hand," Proc. SIGGRAPH, pp.110-119, 2003.
8. L.T.Wang and C.C.Chen, "A combined optimization method for solving the inverse kinematics problem of mechanical manipulators," IEEE Trans. on Robotics and Automations, Vol.17, No.4, pp.489-499, 1991.