Real-time Model-based Hand Shape Estimation with Stereo Vision

Fujiki, Ryuji
Department of Intelligent Systems, Kyushu University

Yoshimoto, Hiromasa
Department of Intelligent Systems, Kyushu University

Arita, Daisaku
Department of Intelligent Systems, Kyushu University

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

出版情報：Proceedings of the 11th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp.225-230, 2005-01
バージョン：
権利関係：

Real-time Model-based Hand Shape Estimation with Stereo Vision

Ryuji Fujiki, Hiromasa Yoshimito, Daisaku Arita and Rin-ichiro Taniguchi

Department of Imtelligent Systems, Kyushu University
E-mail:{fujiki,yosimoto,arita,rin}@limu.is.kyushu-u.ac.jp

Abstract We are researching for real-time hand shape estimation for natural user interface. We have employed a computer vision approach, since unwired sensing provides restriction-free observation unlike a dataglove. Since a human's 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. On the other hand, a human's hand has several constraints at the same time, i.e., each joint has a movable range and interdependence. This paper proposes a novel method to estimate hand shapes in real-time by using both shape features from camera images and 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 unlike a dataglove. There are two approaches for hand shape estimation. One is classification of hand shape into predefined categories. The other is measurement of arbitrary hand shape. The latter approach is possible to use hand shape estimation for a lot of applications. It can be applied not only interface using recognition of hand shape predefined categories, but also interface using hand shape change: people can operate a hand of a robot or a CG character directly referring to the estimation result. Therefore we have adopted the latter approach.

This paper proposes a novel method to estimate 3D hand shapes in real-time by using both shape features from camera images and hand constraints heuristically introduced. This paper consists of the followings. Section 2 mentions the related researches. Section 3 describes the representation of a 3D hand model used our system. Section 4 and 5 describes the details of hand shape estimation: extraction of hand features is explained in section 4; hand shape estimation by Inverse Kinematics with hand constraints is explained in section 5. Section 6 show some experimental results in order to verify the effectiveness of our proposed method. Finally, we present the conclusion and the future work.

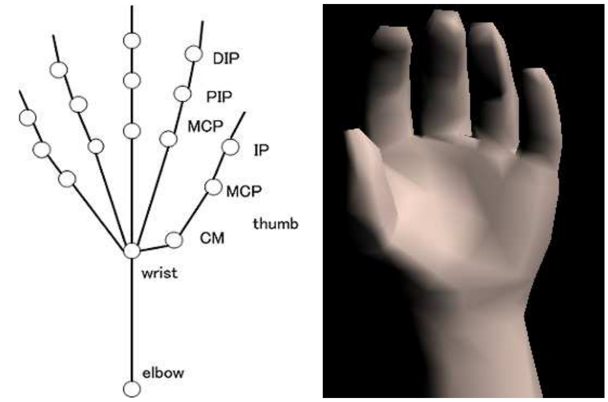
2 Related Works

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

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

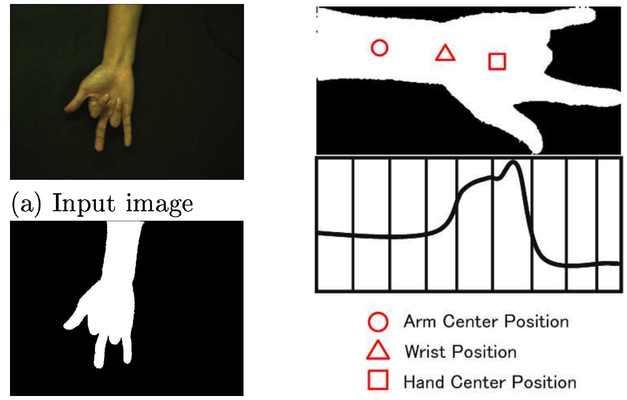
The former is based on appearance of hands in the 2D images, and consists essentially in a kind of template matching. Shimada et al.[1] represented the variations of possible shape appearances as the Locally-Compressed Feature Manifold (LCFM) in an appearance feature space. It is effective to prevent the system from tracking failures and reduce the search area. Stenger[2] used a tree-based estimator based on Bayesian filter. This approach achieves coarse to fine search by approximating image likelihood as piecewise constant at a coarse discretization, 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



(a) Skelton Model (b) Skin Model

Fig. 1: Hand Model



(a) Input image (b) Skin color image (c) Feature Points

Fig. 2: Image analysis

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 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, the former approach has the problem a large amount of computation. the latter has the problem of image feature missing by self-occlusion, since a human's hand has many joints.

Our approach is to extract the robust hand features from the images as much as possible even if it is not sufficient to know an exact hand shape. We estimate 3D hand shape using hand constraints and geometric parameters of the hand model. In addition, we can know hand shape information in more detail from the result of hand pose estimation, and we have a chance to acquire effective image features by limiting a search range of image feature detection. In this paper, we use arcs on the contour obtained by image analysis, and we identify which of the finger (or fingertip) arcs. Then we calculate hand shapes by Inverse Kinematics with hand constraints. We use multiple cameras, and we select two cameras which face to the front of a hand, since it is easy to extract feature points from the frontal image robustly. Here, the front of a hand is the direction perpendicular to a palm in the 3D space. Then, we estimate accurate 3D positions of feature points with stereo vision. We have made preliminary experiments under uncomplicated background.

3 Hand Model

In principle a human hand is a non-rigid object. In this paper, a human hand is approximated by 3D rigid articulated objects. This 3D hand model consists of a skelton model and a skin model(see Fig. 1). The skelton model consists of joint linked others, and has parent-child relation on each joint. Skin model is not currently used for hand shape estimation. In each finger, DIP (Distal Interphalangeal) and PIP (Proximal Interphalangeal) joint has 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. However, in this paper, the elbow is not estimated.

4 Feature Extraction of Hand

We extract shape features using image analysis. As preparation, we extract skin color region. First we convert a color input image into a hue image(see fig.2(a)). The image is smoothed for removing noise and binarized.(see fig.2(b)) We retrieve contours from the binary image and get the longest contour. Then we extract shape features as follows. (see fig.2(c))

4.1 Non-finger Feature Points

1. Wrist Positon
   We find a minimal-sized rectangle for extracted contour. This rectangle is normalized for rotation. Generally, a hand is wider than an arm. Therefore we find the wrist position $(x_{wrist}, y_{wrist})$by counting the number of skin color region 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.

$$x_{arm} = \frac{M_{10}}{M_{00}}, \quad y_{arm} = \frac{M_{01}}{M_{00}} \tag{1}$$

   where

$$M_{ii} \quad = \quad \sum_x \sum_y x^i y^i \tag{2}$$

   (x, y) is coordinates of skin color pixel.

3. Hand Center Position
   The centroid usually dose not coincide with hand center, because it is strongly influenced by finger'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 Arcs on the Contour

We detect arcs on the contour. These arcs are used as the top joint positions of IK mentioned later.

1. Detection of Arcs on the Contour
   We detect arcs on the contour by curvature information. The arcs correspond to joints projected outer-most in 2D image space. Here, we have to consider two problems: the correspondences between the arcs and fingers and between two cameras. This problems is complex since we can make a lot of combinations. Therefore, we use hand constraints, the finger order and intervals between fingers, to reduce combination.

2. Correspondence between arcs and fingers
   We decide which of the finger an arc corresponds to. If we can 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 estimate 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 $C_j(c_{j,1}, \cdots, c_{j,k})$:($j = 0, \cdots, \binom{5}{k}$) is a set which is selected k pieces from F_i.

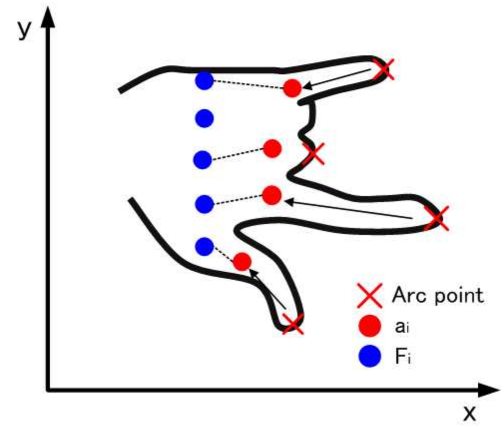


Fig. 3: Correspondence between arcs and figners

   - The best combination C_j is determined as $argmin_j \sum_{m=1}^{k} ||c_{j,m} - a_m||$ (m=0,$\cdots$,k), when a_m is the position of the base of protrusion [1] whose top is the m-th arc. when we calculate the distance $||c_{j,m} - a_m||$, only y coordinates of the positions are used. This complicated combination 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 maximam value of curvature whose y coodinate is similar to F_i(the missed finger).

4.3 Pitch and Yaw of Hand

We estimate pitch and yaw of a hand influenced by rotation of the wrist. We define the 3D positions of arm center, wrist, hand center as $P_{palm}, P_{wrist}, P_{hand}$ respectively.
$\mathrm{V_{arm}}, \mathrm{V_{hand}}$ is

$$\mathrm{V_{arm}} \quad = \quad P_{wirst} - P_{arm} \tag{3}$$
$$\mathrm{V_{hand}} \quad = \quad P_{hand} - P_{wrist} \tag{4}$$

We estimate yaw and pitch rotation of a wrist by calculating rotation of $\mathrm{V_{hand}}$ to $\mathrm{V_{arm}}$. When each vector are projected to x-y plane, angle between x axis and each vector are $\theta_{hand}, \theta_{arm}$. When each vector are projected to x-z plane, angle between x axis and each vector are ϕ_{hand}, ϕ_{arm}.

$$\theta_{yaw} \quad = \quad \theta_{hand} - \theta_{arm} \tag{5}$$

[1]The base of protrusion is detected by contour tracing, starting from an arc point.

Table 1: Movable range of joint

	DIP	PIP	MCP	abduction
Little	0∼70	0∼90	-30∼90	-20∼20
Ring	0∼70	0∼90	-20∼90	-20∼20
Middle	0∼70	0∼90	-20∼90	-20∼20
Index	0∼70	0∼90	-40∼90	-20∼20
Thumb	-20 ∼80	-20∼40	-20∼70	-20∼35

unit: degree

$$\phi_{pitch} = \phi_{hand} - \phi_{arm} \tag{6}$$

5 Inverse Kinematics with Hand Constraints

5.1 Hand Constraints

A human's hand has many constraints[5, 6, 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.

2. Limitation of abduction
   As we flex our finger, adduction happens. Inversely, we extend our finger, abduction happen. On the basic of this knowledge, we limit abduction of finger linearly with MCP joint.

3. Interdependence of finger joints
   Each joint has a interdependence.

   - On a grasping,
     DIP joint has a linearity relationship with PIP joint. ($\theta_{DIP} = \frac{2}{3}\theta_{PIP}$)
   - As MCP joint bend, PIP and DIP joint bend slowly at first. And then their joint bend suddenly in the middle of MCP joint flexion. Finally their joints bend again slowly. In brief, MCP joint has correlation of ess curve with PIP and DIP joints.

     However a human hand can extend MCP joint under PIP and DIP joints bended. Therefore we set a constant angle as PIP joint, when MCP joint angle is smaller than 10 degrees.

   On the basis of the above observation, we heuristically use relational expressions shown Fig.4.

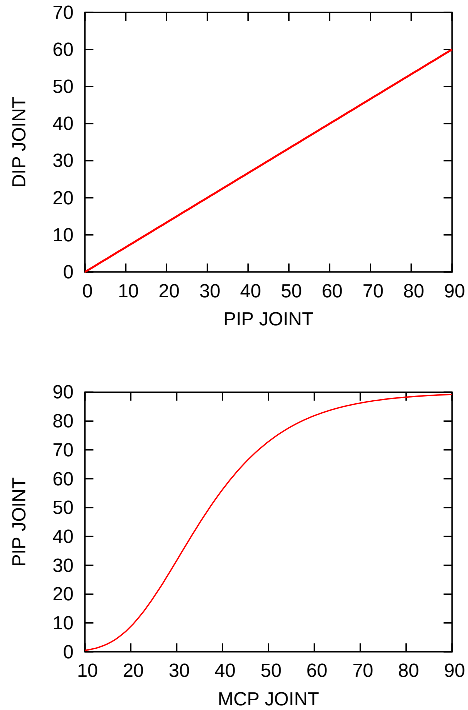


Fig. 4: Interdependence of Finger Joints: The figure above shows relation between PIP and DIP joint, the bottom shows relation between MCP and PIP joint

5.2 Hand Shape Estimation using IK

We estimate joint angles of fingers by Inverse Kinematics with the hand constraints mentioned above. Inverse Kinematics is a method to calculate the angles of the joints in order to put the target at the goal location. Here, the goal is given by an arc position detected and the target is fingertip/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.

5.2.1 Identification of Arcs

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

1. We set an arc as the goal, fingertip as the target of IK.

2. We solve Inverse Kinematics by CCD method. Then, we calculate the error between the goal and the target. 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 hypothesize +that the target is the parent joint of target, and go to 2. If the target is MCP joint, go to 4.

4. We select the joint which is the least error.

5.2.2 Estimation of joint angles

Even if we can identify arcs position correctly, the joint angle of its child can not be estimated by Inverse Kinematics. Except for the case that the arc's position is a fingertip. We solve this problem by using interdependence of joint angles, and estimate those joint angles.

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 know that the correspondences between the arcs and fingers or between tow cameras is false. Then we estimate finger angles again by using a new combination. In addition, we can know hand shape information in more detail 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. 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. PC has Pentium *IV* (2GHz) running Linux. Experimental results is shown Fig.5. The processing time is shown Table. 2

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 for real-time estimation and hand constraints in order to complement hand shape information not to be obtained image features. We have experimented

Table 2: processing time

Convert image to hue color space	12
Detection skin color region	12
Extract hand features	20
IK calculation	14
Total	58

unit: msec

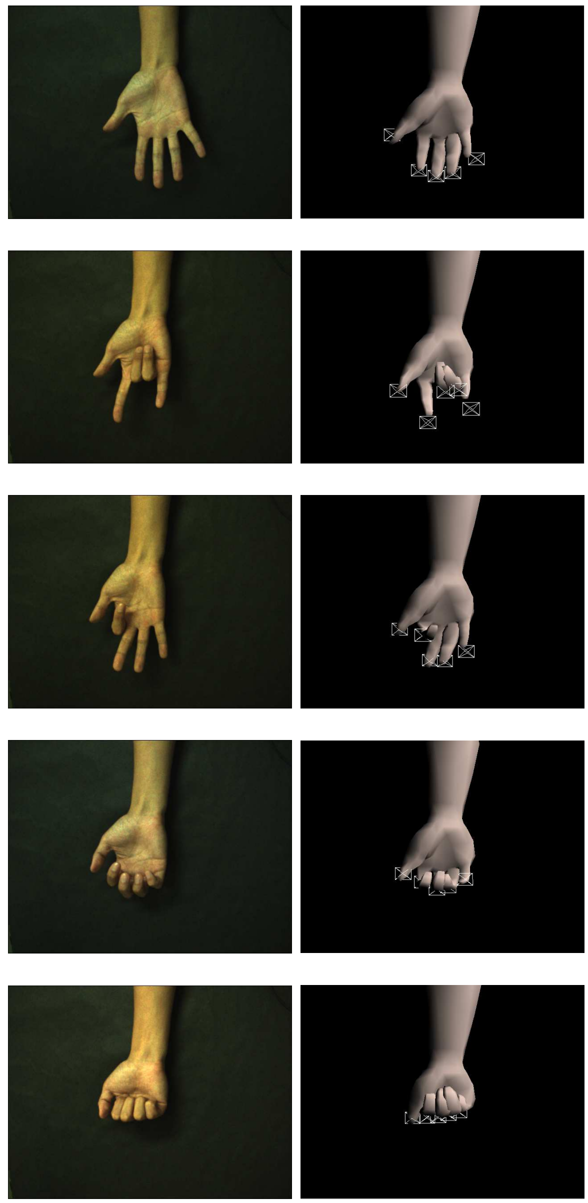


Fig. 5: The result of estimation

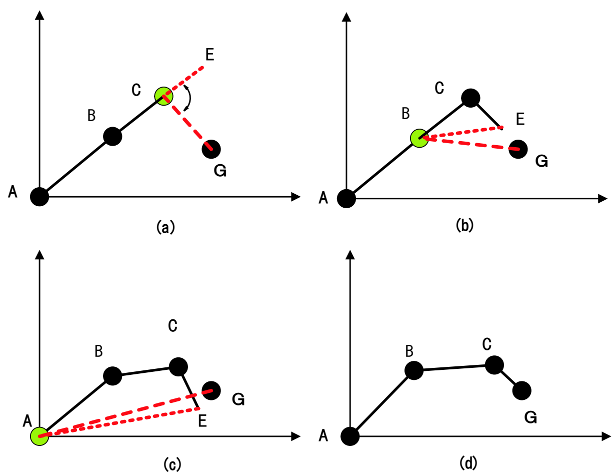


Fig. 6: Cyclic Coordinadate Descent Method

by our proposed method under non-complex background.

The next goal is to build the 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 from the first pose is also an important issue.

Appendix: Cyclic Coordinate Descent Method

The CCD method minimizes the error on the position of the end effecter by adjusting the orientation of each bone one by one. Concretely we introduce an estimated method in the joint angle of the finger using this technique. A, B, C is joint in figure. 6, Each joint has relationship of child-parent in this turn of arrival. It calculates by the following procedures to bring E (the target) to the position of G (the goal).

step1 It calculates a vector from C to E or G ($\overrightarrow{E}$, $\overrightarrow{G}$).

step2 It normalizes the vector which was calculated in step1, and calculates the angle of rotation to turn around $\overrightarrow{E'}$ to $\overrightarrow{G'}$.

step3 It adds the turn corner which was computed in step2 to joint C. If angle limitation is set to the joint, It compares with the joint angle which calculated the maximum and minimum of the movable range. If the joint corner exceeds a movable range, The maximum (or minimum) of the movable range is set up as a new joint angle. Fig. 6(a)

step4 When the distance between the goal and the target when using a new joint angle becomes bigger than the last time, it returns to the original joint angle.

step5 It repeats step2-4 with B.(B is the parent joint of C) It repeats to the specified joint. Fig. 6(b)(c)

step6 It will finish if repeating step2-5 at the decided number of times or the distance between the goal and the target becomes smaller than the threshold value. . Fig. 6(d)

References

[1] N. Shimada and Y. Shirai, "Hand Posture Estimation based on 2D Appearance Retrieval Using Monocular Camera," In 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," in ICCV, pp.1063-1070, 2003.

[3] Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai and Tsukasa Ogasawara. "Hand Pose Estimation for Vision-based Human Interface", IEEE Transactions on Industrial Electronics. Vol.50, No.4, pp.676-684, August, 2003.

[4] Shan Lu, Dimitris Metacas, Dimitirs Samaras, John Oliensis, "Using multiple cues for hand tracking and model refinement," IEEE Conf. pp443-450, 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, pp8543-8550, 2000.

[6] Kamper, DG, EG Cruz, and MP Siegel. "Stereotypical fingertip trajectories during grasp," J Neurophysiol, pp3702-3710, 2003.

[7] George ElKoura, Karan Singh "Handrix: Animating the Human Hand," SIGGRAPH Symposium on Computer Animation, 2003.

[8] Wang, L. T. and Chen, C. C, "A combined optimization method for solving the inverse kinematics problem of mechanical manipulators," IEEE J. Robotics and Automations. pp489-499, 1991