MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara- ken New Public Hall, Nara, Japan

8—16

Two-handed Gesture Tracking in the Case of Occlusion of Hands

Takanao INAGUMA[†], Hitoshi SAJI[†,††], and Hiromasa NAKATANI[†,††]
[†] Graduate School of Science and Technology
[††] Department of Computer Science
Shizuoka University, Hamamatsu 432-8011, Japan.
Email: inaguma@cs.inf.shizuoka.ac.jp

Abstract

In this paper, we propose an efficient technique for tracking three-dimensional palm motion, which is captured by two cameras.

If we perform template matching independently for each image, two points in two images do not always correspond to each other. Thus we propose to set the search area in the three-dimensional space, not in image planes, so that we can find the correct correspondences. The three-dimensional coordinates of the search area are projected on each image plane. We perform template matching at the projected point in each image. We add those two similarities and use it as a similarity at the three-dimensional coordinates. We search for the position that has the maximum similarity in the search area.

When hands are crossing, we cannot track an occluded palm because the palm region is occluded by an occluding palm. Still at that situation we can track arm regions. We use the arm tracking result to solve the occlusion of hands.

1 Introduction

Computer analysis of gestures has been widely studied to provide a natural interface in human-computer interaction [1, 2, 3]. Hand gesture recognition, in particular, has many promising applications that provide new means of manipulation in artificial and virtual environments.

The methods with sensors or markers can recognize hand gesture accurately and computational cost is low. Attaching sensors or markers, however, becomes a burden for users. Therefore the methods with using image processing are widely studied. An image has only two-dimensional information but we need three-dimensional information for recognizing detailed hand gestures. Thus three-dimensional hand models or multiple cameras are used for tracking.

In this paper, we propose a palm tracking technique based on a constraint of three-dimensional continuity. The outline of the tracking procedure we propose is illustrated in Figure 1. In the proposed method we set the search area in the three-dimensional space and project that area into each image. We perform template matching on the projected point in each image, search for the position that has the maximum similarity in the search area, and we obtain the correct corresponding result.

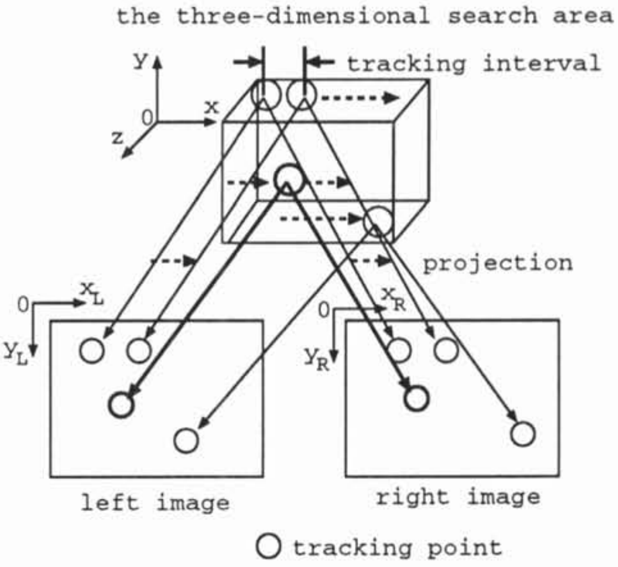


Figure 1: Search area and projected points

When hands are crossing, we cannot track an occluded palm because the palm region is occluded by an occluding palm. Still at that situation we can track arm regions. We use the arm tracking result to solve the occlusion of hands.

2 Outline of the System

2.1 System Configuration

The main elements of the experimental environment are two video cameras (SONY DCR-VX1000). The distance of the baseline between two cameras is set to 0.25m. A subject is asked to move his/her hand and keep the palm in the direction of the cameras. The subject is positioned at a distance of 2.4m from the baseline. To simplify the detection of the initial position of the hand, the subject is asked to keep the face and hands not to overlap in the image. The system tracks the palm motion in stereo from an image sequence; each image is captured at the video frame rate of 30Hz with the resolution of 320 × 240 pixels.

2.2 Outline of the Procedure

Our hand tracking procedure consists of the following steps: 1. camera calibration, 2. initialization of template, 3. computing the three-dimensional trajectory, 4. palm motion tracking, 5. arm motion tracking, and 6. return to the 4th step.

For measuring the three dimensions of an object, we need the internal and external camera calibration before three dimensional measurements. From image correspondences between two views of an object with known size, we estimate the camera parameters such as the orientation and position of each camera.

Let (x, y, z) be an object point, (x', y') be the corresponding image point. Then, the transformation between the object and image point is represented by the camera parameters s_i [4]:

$$x'(s_1x + s_2y + s_3z + 1) + s_4x + s_5y + s_6z + s_7 = 0 \quad (1)$$

$$y'(s_1x + s_2y + s_3z + 1) + s_8x + s_9y + s_{10}z + s_{11} = 0 \quad (2)$$

Thus, we can obtain the camera parameters by solving the above equations for each camera.

We set the initial template in the left image firstly, and the template is used for setting it in the right image.

The subject is asked to keep their hands from their face so that we can easily separate the hand region and the face region in the image.

The initial shape of the template is determined as follows: First, we transform the input image into YCrCb color space to extract skin color regions [5]. Next, we make a silhouette image by thresholding the subtraction image between the input image and the background image. We find face and hand regions from the skin color segmentation image and the silhouette image. For the hand region we search for such the biggest square inside that all the pixels are white, and set the square as the initial template for the hand.

After we locate the initial template in the left image, we set the initial template in the right image by considering the correspondence between the right and left images. If we set the templates independently in the two views, we might estimate inaccurate three-dimensional positions.

The initial position of the right template is determined under the epipolar constraint. The epipolar line is calculated by substituting the center coordinates of the left template into Equations (1) and (2). We locate the right template by template matching along the epipolar line.

From the initial positions in the right and left image and the camera parameters, we calculate the three-dimensional coordinates of the hand. We make use of it for the tracking procedure.

In the following sections, we will explain palm and arm tracking procedure in detail.

3 Palm Tracking Procedure

3.1 Search Area

We determine the search area not in the image plane, but in the three-dimensional space. If we determined the search area in the image plane, it would be difficult to find the appropriate size of the search area. Even if we set the size of the search area to be constant in the three-dimensional space, its appearance changes in a two-dimensional image. Therefore we set the search area in the three-dimensional space so that the size of the search area changes appropriately according to the three-dimensional position of the moving object.

We set the size of the search area by considering the constraint of the continuity of the hand location. Because the location of the hand does not change abruptly between consecutive frames that are captured at every 1/30 second, we can restrict the search area of the next location of the hand to the neighborhood of the current location. We can also limit the area by predicting the next position from the positions in the last three frames.

Then, we set the size of the search area ±3cm from the predicted position for each xyz axis.

3.2 Tracking Interval

We call the distance between the two consecutive three-dimensional points in the search area as the tracking interval (Figure 1). If we set the tracking interval short, we can obtain the accurate tracking results. However the computational costs increase and most operations are useless because the projected points in the image plane stay at the same locations while the tracking points move in the three-dimensional space (Figure 2(a)). Conversely, if we set the tracking interval long, we may skip the appropriate position (Figure 2(b)). In this study, we set the suitable tracking interval experimentally.

To determine the tracking interval, we use camera parameters and the initial position of the hand. We substitute these values into Equations(1)(2), and search for such a tracking interval that changes the locations of the projected points. As a result, 0.1cm is necessary in x and y directions, and 0.5cm in z. Thus we set the tracking interval to be 0.1cm.

3.3 Similarity

We define dissimilarity $D(x, y)$ for template matching between an image $f(x, y)$ and a template $g(x, y)$:

$$D(x, y) = \frac{\sum_i \sum_j |f(x + i, y + i) - g(i, j)|}{T_x T_y}, \quad (3)$$

where T_x and T_y are the width and height of the template, respectively. The template is given by the current image.

Now, let $p = (x, y, z)$ be the matching point, (x_R, y_R) be the projected point of p in the right image, and (x_L, y_L) in the left. The similarity of the point p is defined as

$$S(x, y, z) = 1 - \frac{D(x_L, y_L) + D(x_R, y_R)}{2}. \quad (4)$$

We determine the location of the hand by finding such location (x, y, z) that gives the maximum value of $S(x, y, z)$. We usually find many points which have the same value of $S(x, y, z)$ because there are multiple points that are projected onto the same position in an image. Thus, we determine the location of the hand by calculating the center of the points which have the maximum value.

3.4 Experimental Result

We track the palm motion with two kinds of methods and compare the results to show the effectiveness of the proposed method.

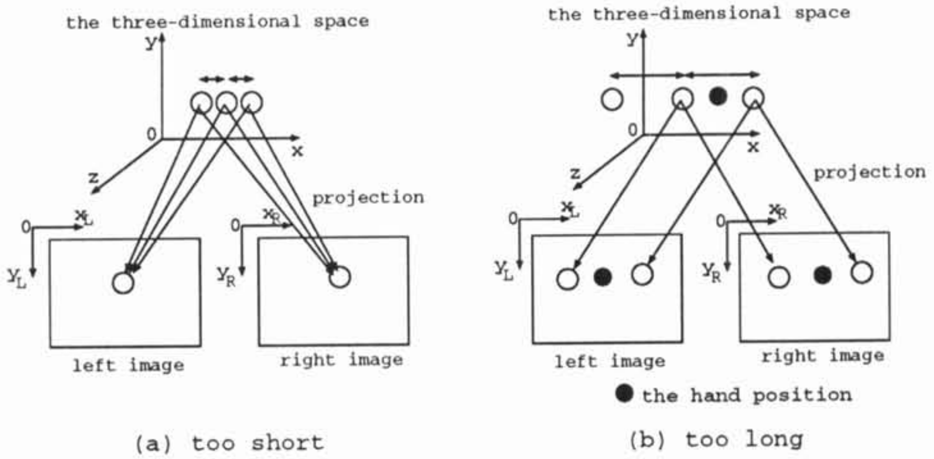
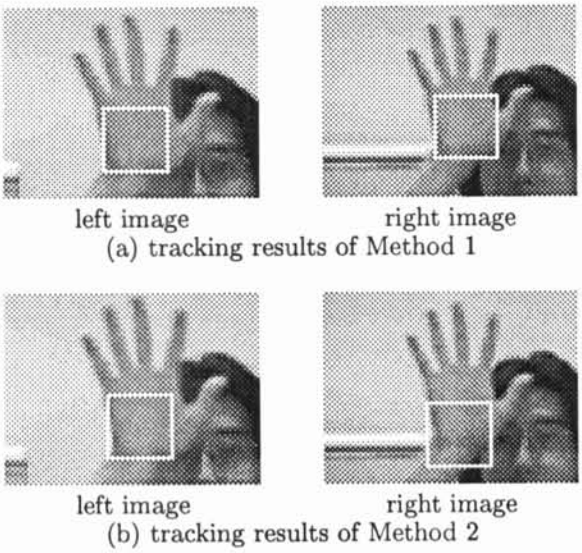


Figure 2: Tracking interval



(a) tracking results of Method 1

(b) tracking results of Method 2

Figure 3: Comparison of the tracking results

Method 1 is the proposed method.

Method 2 performs template matching independently in each image without the epipolar constraint and calculates the three-dimensional position by the tracking results in image planes.

We examine the correspondence between two images for all frames and show the results in Table 1. The ratio of the incorrect correspondence between two images is reduced from 41% to 12% by using the proposed method. The average time of the proposed method is 0.27 second per frame, and 0.91 second for Method 2. We can see that the proposed method can restrict the search area efficiently and track the three-dimensional motion accurately.

4 Arm Tracking Procedure

We track the right and left hands with the method explained in the previous section. However, when hands are crossing, we cannot track an occluded palm. Figure 4 shows a mistracking result. For solving the problem of the occlusion, we need to consider the following three points: 1. Which hand is occluded and when? 2. When is the occluded hand found again in an image? 3. How to resume tracking the hand?

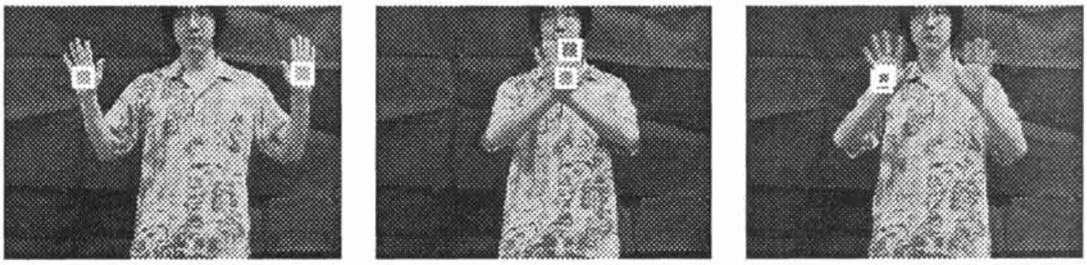
Table 1: the number of the incorrect tracking results

sequence	# of frames	# of incorrect frames	
		Method 1	Method 2
1	270	38	162
2	270	6	70
3	180	30	35
4	270	22	122
5	75	26	49
total	1,065	122	438

For solving these problems, we use the arm tracking results. At that situation we can track arm regions because most region of the arms can be found in the image.

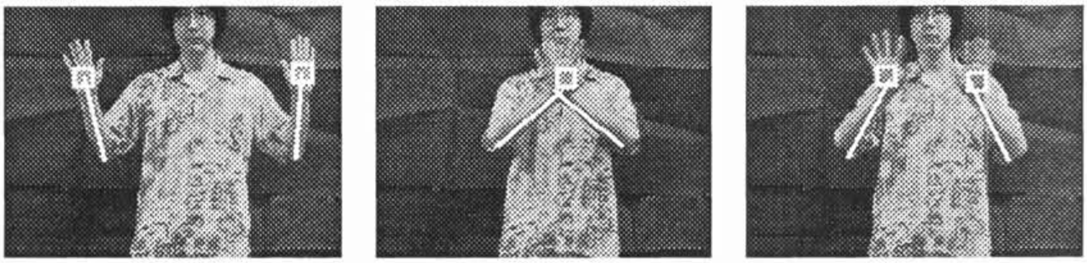
We track arm motion by template matching with Equation(3) for the left image sequence. The arm template has the following informations: the point of the elbow (x_e, y_e), the point of the palm (x_p, y_p), the length l of the template, and the slope θ of the template. We rotate the template with the center at (x_e, y_e), and calculate the dissimilarity by changing (x_e, y_e) and θ. We set the arm position by finding the minimum dissimilarity.

Next, we explain the procedure which solves the occlusion problems. When the palm templates overlap, we start this procedure. Since we calculate the three-dimensional positions of each hand, it is possible to determine which hand is in front when hands are crossing. However, it is difficult to determine that in such a case as the left and right hand have similar depths. Therefore we use the similarity of the palm tracking result. The dissimilarity of the occluded hand increases suddenly while the dissimilarity of the occluding hand does not change much. So we can determine which hand is occluded by comparing the dissimilarities. If



(a) before crossing (b) crossing (c) after crossing

Figure 4: Palm tracking results without arm tracking



(a) before crossing (b) crossing (c) after crossing

Figure 5: Palm tracking results with arm tracking

the occluded hand is determined, we stop updating the palm template of the occluded hand and store the palm template of the previous frame.

If we detect the occlusion, we focus the distance between the palm points of arm template. Let (xr_p, yr_p) be the palm point of the right arm, (xl_p, yl_p) be the left. The distance D is defined as:

$$D = |xl_p - xr_p| + |yl_p - yr_p| \qquad (5)$$

If the value of D is bigger than a threshold, we resume tracking the occluded hand. In this procedure, we restrict the search area of occluded hand to the neighborhood of the palm point of the arm template and perform template matching with the template which is stored before the hands are crossing.

As an experiment, we track the palm motion from the same image sequences as in Figure 4, and show the tracking result in Figure 5. In this figure, white lines show the arm tracking results. When hands are crossing(Figure 5(b)), we stop tracking the occluded hand(left hand). When the left hand is found in the image, we start tracking the left hand again(Figure 5(c)).

5 Conclusion

We have proposed a stereo tracking method by template matching for measuring the accurate three-dimensional positions. We set a search area in the three-dimensional space, perform the template matching at the points that are projected onto an image plane from the search area, and track a moving palm by finding the correspondence between right and left images. We also use the arm tracking result to solve the occlusion of hands. Using this method, we can obtain the accurate three-dimensional trajectories and track the palm which occluded by crossing hands.

In this study, we do not allow the rotation, expansion and reduction of the palm template. If the palm slants or moves toward the cameras, the accuracy of the trajectory becomes low since the similarity of the template matching becomes low. To solve this problem, we will try to allow a template to change its size. The update of the template size is left for our future work.

References

[1] V. I. Pavlovic, R. Sharma and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677-695, 1997.

[2] D. M. Gavrila, "The visual analysis of human movement: a survey," *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82-98, 1999.

[3] A. Pentland, "Looking at people: sensing for ubiquitous and wearable computing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 107-119, 2000.

[4] B. K. P. Horn, *Robot vision*, MIT Press, 1986.

[5] D. Chai, K. N. Ngan, "Locating facial region of a head-and-shoulders color image," *Proc. 3rd International Conference on Automatic Face and Gesture Recognition*, pp. 124-129, 1998.