Vision–Based Dynamic Tracking of Motion Trajectories of Human Fingertips

D.W. Ren and J.T. Li

Robotics Institute, Beihang University, Beijing 100083, P.R. China
rendawei636@gmail.com

Abstract. Dynamic tracking of motion trajectories of human fingertips is a very challenging job considering the requirement of accuracy as well as speed. A binocular vision is adopted to dynamically measure the fingertip positions of the human hand. Based on Kalman filter, combined with the historical motion data of human hand, a dynamic tracking model is presented which can fast and accurately track the fingertip positions. The experimental result shows that when human fingers move in a natural speed, the dynamic fingertip positions can be tracked successfully.

1 Introduction

The information of human hand motion, especially the fingers, is very useful when designing dexterous robotic hands or planning their grasping and manipulation. At present, dataglove is often used as the interface to measure the human hand motion [1-5]. While it is convenient for dataglove to measure the joint angles, however it cannot satisfy the expected precision when the fingertip positions are accurately needed. As for computer vision, it can achieve higher accuracy than dataglove but slower speed because of processing a large amount of information. When tracking the moving trajectories of human fingertips, it is involved in dealing with the multi-dynamic targets within a relative small space, where mismatching easily happens. Therefore how to use computer vision effectively and efficiently is a very challenging research. In tele-manipulation system for DLR Hand, computer vision is used to calibrate the dataglove [2]. Choosing the appropriate interface is obviously relevant to the motion parameters which to be measured.

The research of this paper focuses on the dynamic tracking of fingertip motion of human hand. As stated above, dataglove cannot satisfy the precision requirement of the positions. Therefore the binocular vision is adopted in our system to measure the positions of the human fingertips. The research goal is to increase the recognition rate, which is not quite satisfied in our previous research [6]. We think that the unique characters of the human finger motion can provide important clues to deal with such matters, thus should not be neglected. Therefore based on Kalman filter [7-8], combined with the full use of the just passed motion data of human hand, a dynamic tracking model is presented which can fast and accurately match the currently measured fingertip positions to their corresponding past ones. The experiment shows that when human fingers move in a natural speed, the recognition rate is almost 100 percent. This is critical to our master-slave grasp system for the dexterous hand, which is developed in our previous research [6].

T.-J. Tarn et al. (Eds.): Robot. Weld., Intellige. & Automation, LNCIS 362, pp. 429–435, 2007.
springerlink.com

2 Targets to Be Tracked

When binocular vision is used to track the fingertips, the basic issue to be solved is to match the current dynamic targets to their corresponding ancestors on both images. The targets in the images are 2 dimensional, of which the coordinates vary in different frames of the images. Therefore the tracking variables in the images are the 2-D coordinates (x, y) of targets. To make the tracking easy and fast, the coordinates x and y are tracked independently. The presented method in the following sections is suitable for both x and y coordinate.3 Target tracking based on kalman filter.

The following model is based on the assumption that human fingers move in constant speeds, which can be roughly controlled by the operator.

In order to shorten the tracking time, every target is searched only in a relatively small area rather than in the whole image one. The searching area, of which scope is predetermined by the experiment, is located by estimation according to the last frame position and speed of the target. Considering the possible noises, the target is recursively modified and be searched out at last.

2.1 Discrete Time Model for Position and Speed

At time t_k , assume that the position and velocity, along one of the coordinate axes, of the target is s_k and v_k , respectively. With the constant speed assumption, we have

$$s_k = s_{k-1} + v_{k-1}\Delta t, v_k = v_{k-1} \tag{1}$$

Where the subscripts k and $k-1$ mean at time t_k and t_{k-1} . Δt is the sampling interval.

Equation (1) can be rewritten into the following expression

$$x_k = Ax_{k\text{-}1} \tag{2}$$

Where $x_k = \begin{bmatrix} s_k \\ v_k \end{bmatrix}$ is called state vector and $A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$ is called transition matrix.

Considering the uncertainty that is brought by the assumption of constant velocity model, Equation (2) is further estimated as follows.

$$x_k = Ax_{k-1} + \begin{bmatrix} 0.5\Delta t^2 \\ \Delta t \end{bmatrix} w_k = Ax_{k-1} + \begin{bmatrix} w_{ks} \\ w_{kv} \end{bmatrix} \tag{3}$$

Where the sequence $\{w_k\}$ is acceleration which is supposed to be a white Gaussian noise. The variance of w_k is q_k^2 . $\begin{bmatrix} w_{ks} \\ w_{kv} \end{bmatrix}$ is called process noise. The w_{ks} is position noise and the w_{kv} is velocity noise. $Q_k = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} = q_k^2 \begin{bmatrix} \Delta t^4/4 & \Delta t^3/2 \\ \Delta t^3/2 & \Delta t^2 \end{bmatrix}$ is covariance matrix of process noise[7]. When the Δt in the system is short enough, q_{11}, q_{12}, q_{21} are small which can be considered as zero. So the Q_k can be expressed as $\begin{bmatrix} 0 & 0 \\ 0 & q_{22} \end{bmatrix}$, where $q_{22} = \sigma_{vk}^2$, σ_{vk}^2 is variance of random speed changes.

The measurement model of target position takes the following form:

$$z_k = Cx_k + n_k \tag{4}$$

Where z_k is the measured position (x or y coordinate) of target from the image plane of camera. $C = [1 \quad 0]$ is called the observation matrix. n_k is the uncertainty of the camera measurement. Sequence $\{n_k\}$ is also supposed to be a white Gaussian noise which is called measurement noise and independent with the process noise sequence $\{w_k\}$.

2.2 Recursive Algorithm

The recursive algorithm is as follows. The estimative state vector $x_{k|k-1}$:

$$x_{k|k-1} = Ax_{k-1|k-1} \tag{5}$$

Where the subscription $k \mid k-1$ denotes the estimated value at time k from the actual value at time $k-1$, and the subscription $k-1 \mid k-1$ means the actual value at time $k-1$.

The estimative system error covariance matrix $P_{k|k-1}$:

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q_{k-1} \tag{6}$$

Where $P_{k|k\text{-}1} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$, p_{11} is uncertainty(variance) in position; p_{22} is uncertainty(variance) in speed; p_{12}, p_{21} are uncertainty(covariance) in position and speed. $P_{k|k-1}$ is measure of accuracy of $x_{k|k-1}$.

The system gain matrix K_k :

$$K_k = P_{k|k-1}C^T[CP_{k|k-1}C^T + R_k]^{-1} \tag{7}$$

K_k is optimal correction factor used to weight the error between the estimate position $s_{k|k-1}$ and the measurement position z_k.

The updating state vector $x_{k|k}$:

$$x_{k|k} = x_{k|k-1} + K_k[z_k - Cx_{k|k-1}] \tag{8}$$

The updating system error covariance matrix $P_{k|k}$:

$$P_{k|k} = P_{k|k-1} - K_k C P_{k|k-1} \tag{9}$$

By the above recursive procedure, the actual target position $s_{k|k}$ can be determined.

2.3 Measurement Noise Covariance R_k

When determining the measurement noise covariance R_k , affect of the image distortion is ignored, so all the pixels on the image have measurement white noise with the same covariance. So R_k is a constant in this system which can be expressed as $R_k = \lfloor \sigma_n^2 \rfloor$. R_k can be got by the following method. The target is placed at an arbitrary position in the view field of camera and kept stationary. Record the image coordinates s_i of the target at time i . The number of recorded samples is m which should be adequate. Then the actual image coordinate s is approximated as:

$$s = \sum_{i=1}^{m} s_i / m \tag{10}$$

So the covariance of measurement noise matrix R_k is

$$\mathrm{R_k} = [\sigma_n^2] \tag{11}$$

Where $\sigma_n^2 = \sum_{i=1}^{m} (s_i - s)^2 / (m-1) \cdot$

2.4 Process Noise Covariance Q_{k+1}

The process noise is caused by the constant velocity model assumption, which is actually difficult to control precisely for a human operator. Because of difficulty to calculate the real-time process noise covariance matrix, the approximated method is used to calculate it.

First, the speed noise at time $k+1$ approximately calculated as follows:

$$w_{v\,k+1} \cong v_{k|k} - v_{k|k-1} \tag{12}$$

In our system, the sampling period is 0.04 second, so the process noise covariance matrix can be expressed as:

$$\mathrm{Q}_{k+1} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{v\,k+1}^2 \end{bmatrix} \tag{13}$$

Where $\sigma_{v\,k+1}^2 = (\sigma_{v\,k}^2 \times (k-1) + w_{v\,k+1}^2) / k$.

2.5 Initial Variables Selection

It is very important to select suitable initial state vector $x_{0|0}$ and initial system error covariance matrix $P_{0|0}$ to make the tracking model stable and convergent in the initial period[9]. The initial state vector $\mathbf{x}_{0|0}$ and the initial system error covariance matrix $P_{0|0}$ are calculated by the following method.

$$\mathbf{x}_{0|0}=\begin{bmatrix} z_0 \\ 0 \end{bmatrix}, P_{0|0}=\begin{bmatrix} \sigma_n^2 & 0 \\ 0 & \sigma_{v0}^2 \end{bmatrix} \tag{14}$$

Where σ_{v0}^2 is determined by the experience.

3 Experiment

As shown in Fig.1, the subject wears a white glove with four small black balls stuck on the fingertips to make the targets remarkable.

Two white and black CCD cameras are used to track the positions of these balls which are considered to be the human fingertips, when the human hand moves. The resolution of CCD is 768×576. The sampling frequency of the image collection card is 25 frames per second. The code is written in VC++ 6.0 and run on Windows 2000 operating system. And the speed of CPU is Pentium IV 2.0G. with 512M memory. $\sigma_{v0}^2 = 5000(pixels/s)^2$ and $\sigma_n^2 = 0.005(pixels)^2$ in this system.

Fig. 1. The targets to be tracked

Fig. 2. Experiment system

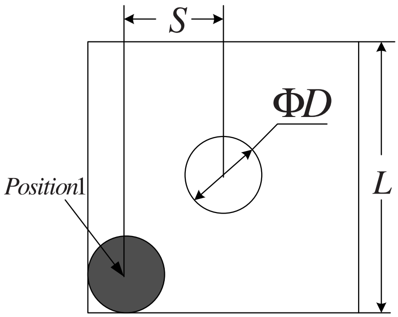

Fig. 3. Position constraint

3.1 Acceleration Constraint of the Tracking Algorithm

To ensure there is no failing tracking, the constraint of the speed of the human fingertips is estimated. As shown in Fig.3, when the target ball is located within the predefined searching area (demonstrated by the square), the tracking is successful. The extreme position of the ball is occurred when the circle is tangent to the edges of the square, such as position 1 (the dark colored circles). The position constraint can be described by the inequation $|\Delta S| < (L-D)/2$. ΔS is the distance error between the estimated position (the light colored circle) and actual position(the deeply colored circle) along each of the coordinate axes. L is the edge length of square and D is the image diameter of the black balls. Because ΔS is resulted from the speed changes that is materially acceleration, we have the equations $|\Delta S| = a_{\text{limit}}\Delta t^2/2$. a_{limit} is the limit acceleration along the coordinate axis which can be calculated by the inequation $|a_{\text{limit}}|\Delta t^2 < L-D$.

In the experiment, the diameter of target ball is 8mm. Distance of camera lens and the human hand is about 600mm. The diameter of target ball is about 23.2 pixels on the image and the edge length of the square is 36 pixels. At last the acceleration constraint is about $\text{a}_{\text{limit}} < 7500\, pixel/\text{s}^2$ along the coordinate axis which is equal to $2.5 m/s^2$.

3.2 Coping with Outranged Acceleration

During the experiment, the acceleration occasionally breaks the constraint, which results in failure of tracking. Two examples are demonstrated in Fig.4. Case 1 represents occurrence of abrupt acceleration and case 2 represents that of abrupt stop.

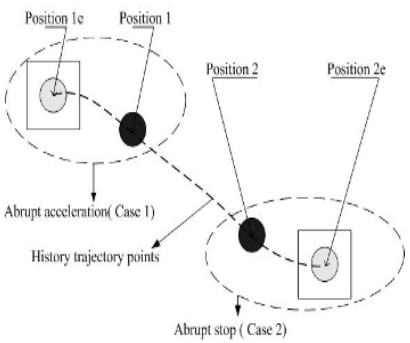

Fig. 4. Tracking failure situation

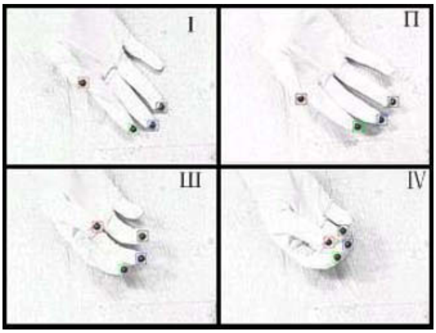

Fig. 5. Tracking procedure

In order to coping with this situation, certain positions on the trajectory the target passed are recorded. If the target is out of the range of the square, the previous recorded positions will be searched until one of the positions is within the searching square and this position will be assumed to be the current tracked target. The recorded process is been done offline, just before the real time tracking of the human fingertips, which can ensure the trajectories are similar enough. The distance between adjacent two recorded points must be less than the edge length of square. To balance the reliability and the searching efficiency, 20 pixels distance is selected.

3.3 Experimental Results

With the method described above, the human fingertips are successfully tracked in the experiment when human fingers move no more than 55 times per minute. It costs 6 ms to simultaneously track 4 targets for each CCD image, altogether 8 targets for 2 CCD. Some of the snapshots of the tracking procedure are shown in the Fig.5.

4 Conclusion

A vision-based approach is investigated to dynamically track the motion of human fingertips. The proposed tracking model is validated by the experiment to be feasible and efficient when human fingers move in a natural speed. Owning to the high-efficiency of algorithm, it is possible to get more human hand motion information timely, which we will investigate next.

The research is also helpful to establish the accurate kinematical model of human hand, which will benefit the design and control of humanoid dexterous hand.

References

1. Sing Bing Kang and Katsushi Ikeuchi(1997) Toward Automatic Robot Instruction from Perception－Mapping Human Grasps to Manipulator Grasps, IEEE Trans. on Robotics and Automation, pp. 81-95, 13(1).
2. M. Fischer, P. van der Smagt, and G.Hirzinger (1998) Learning Techniques in a Dataglove Based Telemanipulation System for the DLR Hand, Proc. 1998 IEEE Intl. Conf. on Robotics and Automation, pp. 1603-1608,Leuven, Belgium.
3. Haruhisa Kawasaki, Kanji Nakayama, Tetsuya Mouri, and Satoshi Ito (2001) Virtual Teaching Based on Hand Manipulability for Multi-Fingered Robots, Proc. 2001 IEEE Intl. Conf. on Robotics and Automation, pp. 1388-1393, Korea.
4. Bruno M. Jau (1995) Dexterous Telemanipulation with Four Fingered Hand System, Proc. 1995 IEEE Intl. Conf. on Robotics and Automation, pp. 338-343.
5. Michael L. Turner, et al. (2000) Development and Testing of a Telemanipulation System with Arm and Hand Motion, Proc. of ASME IMECE DSC-Symposium on Haptic Interfaces, pp. 1-8.
6. Jiting Li, Wenkui Su, Yuru Zhang, Weidong Guo (2004) Vision-Based Grasp Planning System for Dexterous Hands, Proc. of Intl. Conf. on Intellligent manipulation and Grasping, July.
7. Salmond, D. (2001) Target Tracking: Introduction and kalman tracking filter, Algorithms and Applications, IEE Volume Workshop, vol.2:1/1 - 1/16, Oct.
8. Rong Li, X.; Jilkov, V.P. (2003) Survey of maneuvering target tracking. Part I. Dynamic models, IEEE Transactions on Aerospace and Electronic Systems, Vol.39, Issue 4:1333 - 1364 ,Oct.
9. LIU Gang, PAN Quan, ZHANG Hong-cai (2000) A Study of Tracking Filter Initiation Algorithm Based on 2-Step Estimation, Journal of system simulation, Vol. 14 No. 7, pp.849~851, July.