Zoom-Based Head Tracker in Complex Environment

*Teng-Kai Kuo¹   Li-Chen Fu^{1,2}   Jong-Hann Jean³
Pei-Ying Chen¹   Yu-Ming Chan¹*

Department of Electrical Engineering ¹

Department of Computer Science and Information Engineering ²

National Taiwan University, Taipei, Taiwan, R.O.C.

Department of Electrical Engineering ³

St. John's & St. Mary's Institute of Technology

E-mail: lichen@ccms.ntu.edu.tw

Abstract

In this paper, we establish a nearly real-time surveillance tracking system, which is able to detect any person who intrudes to a prohibited area and lock his head image at the scene center. The entire underlying algorithm consists of a motion detector, an elliptical head tracking algorithm, auto-zooming ability, and a VPDA filter which is modified by Probabilistic Data Association filter (PDA filter). The system operates about 35ms and actively controls the camera platform pan and tilt motion to track a person in real environment.

1 Introduction

Although visual tracking of a person in an unmodified environment has greatly progressed in recent years for computer vision research, there are still problems for visual tracking with the active servo control. Because for the stationary camera or a video, there exist some technologies, like background differencing [5], which can achieve the tracking objective; however, it does not allow for the active camera. The reason is that when camera moves, the background relatively moves, too.

Owing to that the human head is an obviously

distinguishable feature of the body, we focus on the tracking of the human head. Nevertheless, the common face detection methods on stationary images are time-consuming with respect to the real-time requirement and hence the active tracking of the tracking system would easily fail in such a situation. Based on the elliptical head tracker [1], this paper proposes a real-time algorithm for tracking human head with a moving camera, which turns out to be suitable for the underlying problem. Therefore, we here compose this ellipse head tracker with our motion detector, auto-zooming ability and modified VPDA filter to establish a robust surveillance tracking system. Each component will be described in Section 2 and 3.

Fig. 1 The overview of overall system architecture.

2 Head Tracking System

The architecture of our visual tracking system can be easily shown by the block diagram in Fig. 1. The real-time image sequence is provided from a camera, which is mounted on a pan-tilt servo platform, and is then passed to the target detection unit.

2.1 Motion Detector

A natural question to be revised here is that: "Why don't we directly use the ellipse method to find the moving human's head with stationary camera, but use the motion detector?" This is due to the fact that the image size is too large to do the complete-search in each frame, or it may meet the problem of time-consuming. In other words, we need to locate the human's position by using other methods.

As shown in Fig. 2, two consecutive image frames $I(k)$ and $I(k-1)$ are subtracted pixel by pixel, and the results are then binarized. Later in the sequel, we will call it subtracted image. The number of bright pixels are summed up and the sum is compared with the threshold. If this number is less than the motion detection threshold, the motion detection unit examines the next two consecutive frames; if the number exceeds the motion detection threshold, we declare that a target is detected.

Fig. 2 Block diagram of the motion detection unit.

Fig. 4 The overall flowchart of moving edge image.

We include the concept of "moving edges" [2]. This can be achieved by the following steps. First, we obtain an edge image of the current frame at time t . Then, we incorporate this information with the subtracted image as mentioned previously by taking a logic AND operation between the two images. This highlights the edges within the moved pixel region to obtain the moving edges in the latest image, as shown in Fig. 4. Now, we can finally feed this moving edge image to the ellipse method.

2.2 Ellipse Algorithm

In [1], to emphasize the contour of head in an image, we use the edge image for head tracking (see Fig. 5). And just like the template matching, we need to sum the total gradient values pixel by pixel in the edge image along a pre-constructed ellipse model. When finishing the sum-up action for an ellipse, we will shift the center of ellipse to the next pixel and compute the total sum along the perimeter of the ellipse again. After going through

all the search area, we can get the largest value corresponding to the distribution of edge pixels which resembles the ellipse model the most, that is, mostly like the head contour.

The process in each searching loop can be summarized by the following normalized sum equation

$$s^* = \max \left\{ \frac{1}{N_\sigma} \sum_{i=1}^{N_\sigma} |g_i| \right\}, \quad (1)$$

where $s = (x, y, sz)$ is the ellipse's state, in which (x, y) is the position of the center and sz means the size of the ellipse; N_σ is the total numbers of all pixels along the perimeter of ellipse.

Fig. 5 Search the head shape in edge image.

2.3 Auto Zooming

Although ellipse algorithm derives good performance of tracking, it still has some defects. Due to the limitation of fixed zoom of camera, a human being to be tracked cannot stand too far or too close. That is because when human being comes closer, his head would become larger and even bigger than our maximum size of the head model. Similarly, when human being goes far away, his head will shrink to a small bright ball, not an ellipse any more. If the environment unfortunately is cluttered enough, the noisy edges of a group can form the same bright ball, too. Hence, we can't let the size of the ellipse model be too small to track in image. It may lose the head very soon.

For this reason, we want to control the zoom of

Fig. 6 Similar triangles of optical lens.

the camera. In other words, we want to keep the size of the head image at a fixed value. We not only can track target more stable, but also can enlarge the tracking field due to the zoom capability.

From Fig. 6, let z_i' be the desired fixed size, and z_i be the real head size in image. By using principle of similar triangles, we can get the following equality

$$\frac{u}{z_o} = \frac{v}{z_i} = \frac{v'}{z_i'} \quad (2)$$

After slight modification on this equation, we can obtain

$$v' = \frac{v}{z_i} \times z_i' = \frac{v}{z_i} \times 16. \quad (3)$$

Based on the above equation, we can always compute the camera constant v' corresponding to the fixed size z_i' , which is defined to be 16, as desired. The reason of choosing 16 to be the fixed size is not only for clarity but also for suitability for ellipse algorithm. Because if the ellipse size is too small, the edge detection will more easily converge to the facial edge inside the head image, whereas if the size is too large, the computational cost will increase dramatically and the tracking will be lost more easily. The relationship between focal length (camera constant) and zoom value is calculated in advance, as shown in Fig. 7.

Fig. 7 Focal length (camera constant) vs. zoom.

3 Trajectory Design

The VPDA filter, in fact, is the original probabilistic data association (PDA) filter ([3][4]) integrated with the visual information, which is introduced by this work. The original PDA filter has been proposed in radar tracking for many years. By the radar or sonar sensors in realistic environments, there many extraneous incorrect data, such as electronic system noise, false alarms, and other targets, result in the degradation of tracking performance, or even result in loss of tracking. Therefore, someone brings up the PDA theory.

The concept of PDA filter is that since the best-match may be wrong due to the other incorrect data, why don't we consider all of the possible targets? If the best-match target is not the actual target, it still can be the second-match or else. Therefore, the PDA filter obtains just an averaged data which includes all of the possible target's data. The method it applied is the probability. It computes the posterior association probabilities for all current possible candidate measurements and uses them to form a weighted sum of innovations for updating the target's state in a suitably modified version of the Kalman Filter.

3.1 The Stochastic Model

The dynamics of the target of interest are modeled by the equation

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{v}(k) \quad (4)$$

where the processing noise $\mathbf{v}(k)$ is assumed to be white Gaussian with zero mean. The measurement system is modelled as follows,

$$\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{w}(k) \quad (5)$$

and the measurement noise $\mathbf{w}(k)$ is also taken to be white Gaussian with zero mean.

3.2 PDA Filter

The measurements are reduced to a set of validated measurements by defining the validation region [4]. Each measurement $\mathbf{z}_i$ that lies within this region is considered validated.

The procedure that yields the probabilities is called Probabilistic Data Association. Assuming a nonparametric diffuse prior [4], the association probabilities are given by

$$\beta_i(k) = \frac{e_i}{\sum_{i=0}^{m_k} e_i}, \quad i = 0, 1, \dots, m_k \quad (6)$$

where

$$e_i = \exp\left(-\frac{1}{2} \mathbf{v}_i^T(k) \mathbf{S}^{-1}(k) \mathbf{v}_i(k)\right), \quad i = 1, \dots, m_k \quad (7)$$

$$e_0 = (2\pi/\gamma)^{n/2} m_k c_n (1 - P_D P_G) / P_D \quad (8)$$

By the total probability theorem [4], the conditional mean of the state can be written as

$$\begin{aligned} \hat{\mathbf{x}}(k|k) &= E\{\mathbf{x}(k) | Z^k\} \\ &= \sum_{i=0}^{m_k} E\{\mathbf{x}(k) | \theta_i(k), Z^k\} P\{\theta_i(k) | Z^k\} \\ &= \sum_{i=0}^{m_k} \hat{\mathbf{x}}_i(k|k) \beta_i(k) \end{aligned} \quad (9)$$

where $\hat{\mathbf{x}}_i(k|k)$ is the updated state estimate conditioned on the event that the i^{th} validated measurement is correct. This is given by the standard Kalman Filter as

$$\hat{\mathbf{x}}_i(k|k) = \hat{\mathbf{x}}_i(k|k-1) + \mathbf{W}(k) \mathbf{v}_i(k) \quad (10)$$

where $v_i(k) = z_i(k) - \hat{z}(k|k-1)$ is the corresponding innovation, and $W(k)$ is the standard Kalman gain. The error covariance associated with the updated state estimate is defined in [4].

3.3 VPDA Filter

Given the result in the research work presented in [6], the most obvious difference between our VPDA filter and theirs is the image features taken in the similarity association term. Generally, larger Ellipse score means higher image feature similarity between the current measurement and the target. Therefore, we define these similarity values and convert them into *likelihood ratios* like,

$$\lambda_i = \frac{\text{SCORE}_i}{\max_{j \in 1, \dots, m_k} \text{SCORE}_j}, \quad i = 1, \dots, m_k. \quad (11)$$

This likelihood ratio can be incorporated into the standard PDA Filter's association probabilities as follows:

$$\beta_i(k) = \frac{1}{c} \cdot \frac{e_i \exp(\phi \lambda_i)}{b + \sum_{i=1}^{m_k} e_i \exp(\phi \lambda_i)}, \quad i = 1, \dots, m_k \quad (12)$$

$$\beta_0(k) = \frac{1}{c} \cdot \frac{b}{b + \sum_{i=1}^{m_k} e_i \exp(\phi \lambda_i)} \quad (13)$$

The other details can be referred to [6].

4 Experimental Results

To demonstrate the tracking performance of the surveillance system under different backgrounds, we choose three scenarios, as shown in Fig. 8 (a)(b)(d). Our image sequence consists of frames with size 320×240 pixels at the time intervals 35 to 40 ms. However, the image figures in this paper are much smaller due to limited paper space and hence become somewhat blurred. Advantages of the ellipse method, such as partial occlusion, rotation, etc., have all been revealed in our experiments. Fig.

8(a) illustrates successful tracking in cluttered environment. The following images are the edge images of the original ones. Though it looks clear to most of the people, however, due to the usage of ellipse method, its edge image (Fig.8(c)) will appear in a cluttered background considered in our algorithm. In Fig. 8(b), we try to track the head in front of the blackboard, but there is low contrast between the hair on the head and the background. However, we still can find out that the tracking remains satisfactory, though it may not be smooth enough. In the final Fig. 8(d), we augment the tracking system with auto-zooming capability and show the zoom values at the bottom-left image. Owing to the mechanical zooming, the speed of zooming is a little bit slow and can not respond in real-time. Nevertheless, this would not hurt the performance of the tracking.

5 Conclusion

We successfully develop a real-time head tracking system. It mainly applies the theory of the VPDA filter to upgrade the performance of original elliptic head tracker in cluttered environment. And due to the auto-zooming capability, our tracking system can increase the tracking range and enlarge the tracking field. As a consequence, we are confident that the developed system can be used in many surveillance applications.

Acknowledgements

This research is sponsored by Ministry of Education, R.O.C. The authors would also like to send thanks to Prof. M. C. Tsai for his instructive comments and encouragement.

References

- [1] Birchfield, Stan, "An Elliptical Head Tracker," *31st Asilomar Conference on Signals, Systems, and Computers*, November 1997.
- [2] Murray, D. and A. Basu, "Motion Tracking with an Active Camera," *IEEE Trans. Pattern Analysis and Intelligence*, vol.16, no.5, pp. 449-459, May 1994.
- [3] Bar-Shalom, Y., E. Tse, "Tracking in a Cluttered Environment with Probabilistic Data Association," *Automatica*, Vol. 11, pp. 451-460, 1975.
- [4] Bar-Shalom, Y., T. E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.
- [5] Yalamanchili, S., W. N. Martin, and J. K. Aggarwal, "Extraction of moving object description via differencing," *CGIP*, Vol. 18, pp. 188-201, 1982.
- [6] Liu, D. and L. C. Fu, "Target Tracking in an Environment of Nearly Stationary and Biased Clutter," *IEEE Int. Conf. on Intelligent Robots and Systems*, Vol. 3, pp. 1358-1363, 2001.

(a) Partial occlusion in textured background.

(b) Rotation in a low contrast background.

(c) Edge images of (a) and (b).

(d) Overall tracking with auto-zooming.

Fig. 8 Three demonstrations of different backgrounds.