A Robust Visual Servo Based Headtracker with Auto-Zooming in Cluttered Environment

Teng-Kai Kuo[1]   *Cheng-Ming Huang[1]*   *Li-Chen Fu[1,2]*   *Pei-Ying Chen[1]*

Department of Electrical Engineering [1]
Department of Computer Science and Information Engineering [2]
National Taiwan University, Taipei, Taiwan, R.O.C.

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, a modified ellipse 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. Finally, a number of experiments are conducted to validate the effective functionality of the head tracking system.

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.
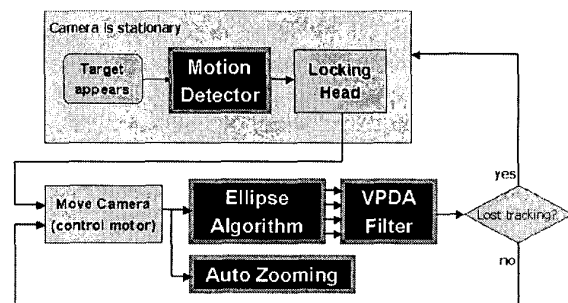

Fig. 1 The overview of overall system architecture.

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 raised here is that: "Why don't we directly use the ellipse method to find the moving human's head with a 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.
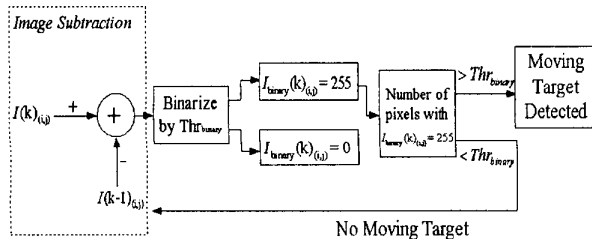
Fig. 2 Block diagram of the motion detection unit.

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. 3. Now, we can finally feed this moving edge image to the ellipse method.
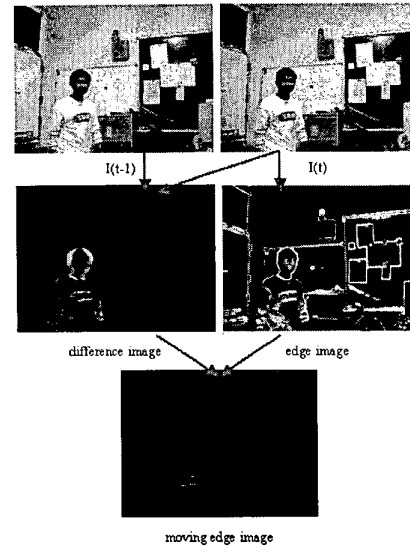
2.2 Ellipse Algorithm

Fig. 3 The overall flowchart of moving edge image.

In [1], to emphasize the contour of head in an image, we use the edge image for head tracking (see Fig. 4). 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

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

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; g_i is the gradient at perimeter pixel i and N_σ is the total numbers of all pixels along the perimeter of ellipse.

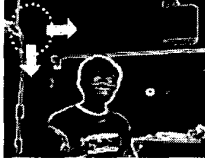## 2.3 Ellipse + SAD algorithm

Fig. 4   Search the head shape in edge image.

From the past work [8], though the author combines the ellipse algorithm with use of the color histogram, the performance is not improved considerably. The potential problem is that the color model does not perfectly match the head's color distribution at different poses. For example, if we only use the facial skin color for matching, it will fail when one turns back with his head covered by hairs. Therefore, a better approach is taken from [8], which includes both the hair and skin colors. Nevertheless, the problem still remains. Since the ratio of the hair and skin color is heuristic and situation dependent, the color model for perfect matching is hardly built. As a result of the ambiguity of the color model, the tracking performance has limited improvement and hence we decide to take the template matching method instead.

To integrate these two methods, we take the same center for both the ellipse model and the template which is chosen to be of size 16×16. By this, not only the similarity information can be fused together, but also the 16×16 template will always be guaranteed to lie inside the ellipse regardless of the size of the ellipse model. Thus, it is impossible for the template to include the background image besides head. The best head position is now modified as

$$s^* = \arg\max_{s_i \in S} \left\{ \overline{\phi}_g(s_i) + \overline{\phi}_{SAD}(s_i) \right\}, \qquad (2)$$

where S is the center position of the ellipse model and also of the template, S is the search area, $\overline{\phi}_g$ is the score from the original gradient value of the ellipse method, and $\overline{\phi}_{SAD}$ is the score from the template matching. Both are normalized as the following

$$\overline{\phi}_g(s) = \frac{\phi_g(s) - \min_{s_i \in S} \phi_g(s_i)}{\max_{s_i \in S} \phi_g(s_i) - \min_{s_i \in S} \phi_g(s_i)}$$

and

$$\overline{\phi}_{SAD}(s) = \frac{\max_{s_i \in S} \phi_{SAD}(s_i) - \phi_{SAD}(s)}{\max_{s_i \in S} \phi_{SAD}(s_i) - \min_{s_i \in S} \phi_{SAD}(s_i)}$$

Another problem is the weighting factor. When one of $\overline{\phi}_g$ and $\overline{\phi}_{SAD}$ fails, the representation of the weighted match-score may no longer be valid. Therefore, if the weighted match-score can be adaptively chosen by assigning different weighting ratios, as shown below:

$$s^* = \arg\max_{s_i \in S} \left\{ \frac{w_g}{10} \overline{\phi}_g(s_i) + \frac{w_{SAD}}{10} \overline{\phi}_{SAD}(s_i) \right\}, \qquad (3)$$

via evaluating the confidence measure of each matching method, the representation of the weighted match-score should be much more effective. Motivated by the current prevalent use of fuzzy membership functions, we construct a look-up table that can assign the suitable weighting ratios for the weighted match-score formula, as shown in Fig. 5.

SAD values

	1200	1000	800	600	400	200	0
135	5:5	4:6	3:7	2:8	1:9	0:10	
155	6:4	5:5	4:6	3:7	2:8	1:9	
175	7:3	6:4	5:5	4:6	3:7	2:8	
195	8:2	7:3	6:4	5:5	4:6	3:7	
215	9:1	8:2	7:3	6:4	5:5	4:6	
235	10:0	9:1	8:2	7:3	6:4	5:5	
255							

gradient edge scores

Fig. 5   The look-up table of weighting factors

$$w_g : w_{SAD}.$$

2.4   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 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.
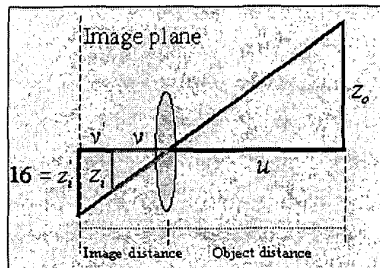

Fig. 6   Similar triangles of optical lens.

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

$$\frac{u}{z_o} = \frac{v}{z_i} = \frac{v'}{z_i'}. \tag{4}$$

After slight modification on this equation, we can obtain

$$v' = \frac{v}{z_i} \times z_i' = \frac{v}{z_i} \times 16. \tag{5}$$

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.
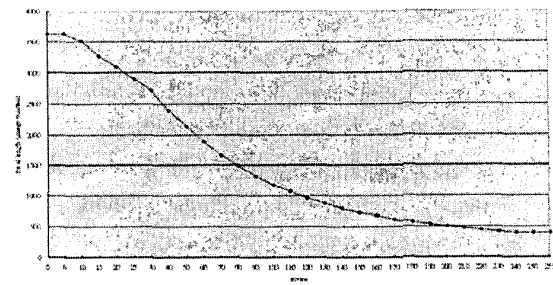
3   Trajectory Design

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

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 filer 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. These details can be referred to [7].

4   Experimental Results

In the past work [7], we have demonstrated many successful experimental results. However, tracking in the low-contrast background is somewhat weak. Hence, we especially provide a modified ellipse algorithm, which is combined with the SAD template

matching, to improve the tracking performance in front of the low-contrast background. Figure 8 successfully demonstrates the better tracking performance in the low-contrast environment of the modified ellipse method, which is combined with the SAD template matching. Figure 8(a) shows the 360° rotation test and Fig. 8(b) demonstrates the partial occlusion test.

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.

References

[1] Birchfield, Stan, "An Elliptical Head Tracker," *31st Asilomar Conference on Signals, Systems, and Computers*, pp.1710-1714,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.

[7] Kuo, T. K., L. C. Fu, J. H. Jean, P. Y. Chen, and Y. M. Chan, "Zoom-Based Head Tracker in Complex Environment," *IEEE Conference on Control Applications, Sep.,* 2002, at U.K.

[8] Birchfield, Stan, "Elliptical Head Tracking Using Intensity Gradiants and Color Histograms," *IEEE Conf. on Computer Vision and Pattern Recognition,* Santa Barbara, California, July 1998

(1)　　　　(2)　　　　(3)　　　　(4)

(5)　　　　(6)　　　　(7)　　　　(8)
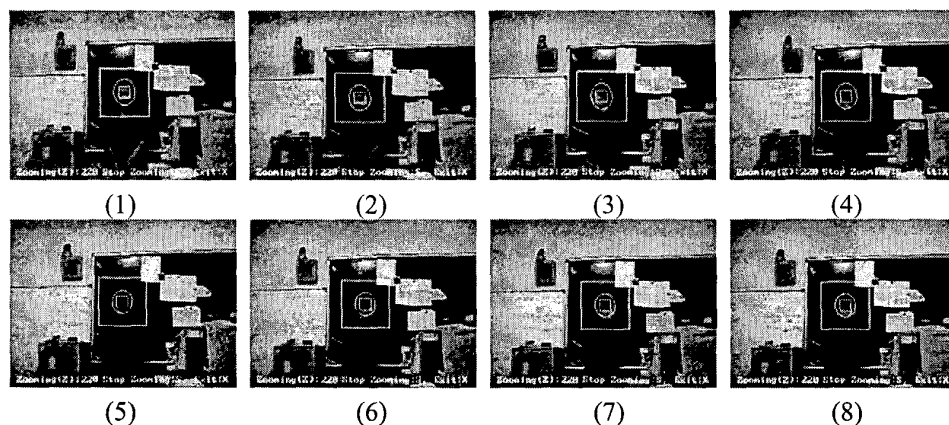
(a) The 360° rotation test in low-contrast background for modified ellipse method, which is combined with SAD template matching.

(1)　　　　　　　(2)　　　　　　　(3)　　　　　　　(4)

(5)　　　　　　　(6)　　　　　　　(7)　　　　　　　(8)
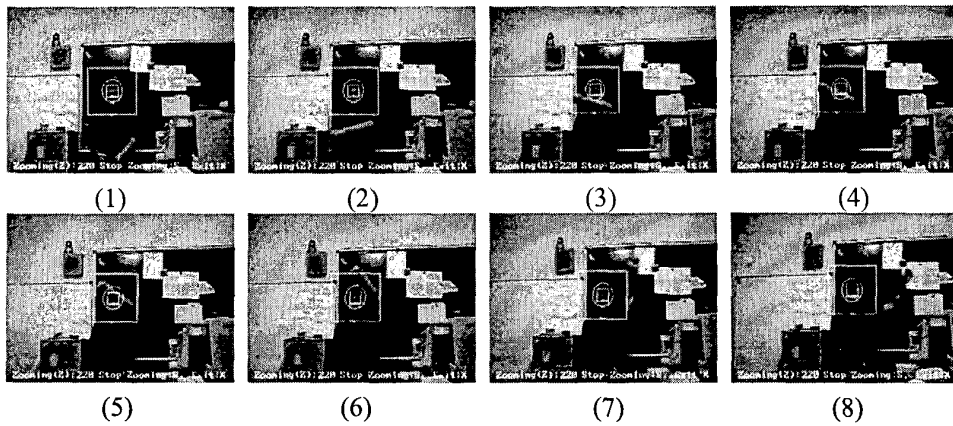
(b) The partial occlusion test in low-contrast background for modified ellipse method,
which is combined with SAD template matching.

Fig. 8　The demonstrations of the ellipse + SAD template matching method.