TELKOMNIKA, Vol.11, No.11, November 2013, pp. 6611~6617
e-ISSN: 2087-278X

Real-time Moving Object Tracking System Using Cam-shift Algorithm

Wei Song
School of Information Engineering, Minzu University of China, No. 27, Zhongguancun South Street, Haidian District, Beijing, 100081, P.R. CHINA.
Research Institutes of Information Technology, Tsinghua University, Information Science & Technology Building, Beijing 100084, P.R. CHINA.
e-mail: sw_muc@126.com

Abstract

A real-time tracking system with the ability of tracking the moving target using cam-shift algorithm is designed. The basic technique of the stereo vision system is discussed. The architecture, the hardware component and the interface of the tracking system is analyzed. Two digital cameras are fit on the cradle head, and the electric pan/tilt/zoom controller has two independent degrees of freedom: horizontal rotation and pitch. The video signals are captured by two digital cameras, and then the position of the target is calculated. The electric pan/tilt/zoom controller is moved in order to make the target always in the surveillance window. This active system extends vision coverage over wide areas to get more detail information.

Keywords: *tracking system, cam-shift, position based, computer vision*

1. Introduction

With the development of the information, the stereo vision technique has been widely researched. It can be used in the navigation of robots, or the navigation of aircraft. The stereo vision system [1-10] used in sur-veillance can get the distance information of the target. The basic architecture of the binocular vision system has different parts in Figure 1. Firstly, the system can auto-matically detect the target object, such as the car in the park, or the obstacle to tell the robotic to change the way. In some special area, for example, in the surveillance system, the task is to track the terrorists, so we should analyze the detail information of the target object. The target object tracking algorithm is used to show the posi-tion and the shape of the target in the screen. Then the moving target object match algorithm is developed to decide that the target object in two cameras is the same or not. Then motion control algorithm is used to control the PTZ to make the target object always in the vision, so the target object can be tracked continually. In the end, the depth information of the target object can be calculated, and we can know the exact distance between the target and the system.

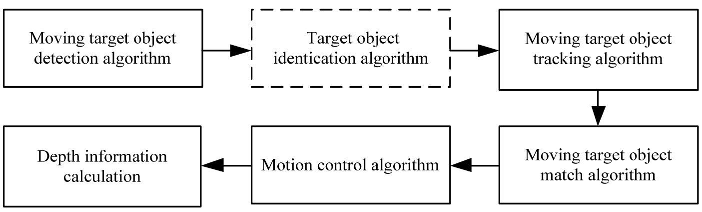

Figure 1. Basic Algorithm in the Stereo Vision System

The moving target tracking algorithm has been widely researched in computer vision area for many years. The exiting algorithms of moving target tracking can be di-vided into three types [11]: kernel tracking based, point tracking based and silhouette tracking based. And the

Received April 8, 2013; Revised June 21, 2013; Accepted July 22, 2013

system can be classified as static camera system [1-10] and active camera system [11, 14]. In the static camera system, the two cameras are fixed, when the target object appears in the vision of the camera, the tracking algorithm works. In the active camera system, the camera can move due to the position of the target object. So the vision range of the active camera system is wider than the static camera system

The remainder of this paper is organized as follows. The architecture, the hardware component and the inter-face of the tracking system is discussed in Section 2, and we will give the detail components of this system. In using algorithms about Cam-shift algorithm and position based pose adjusted algorithm are described in Section 3. Some experimental results are given to show the validity of our tracking system in section 4. Finally, we draw our con-clusions in section 5.

2. Architecture of this Tracking System

In this part, the basic architecture of this tracking sys-tem is analyzed, and then the basic components and interface are discussed.

2.1. Basic Architecture of the Tracking System

From Figure 2, we can see that the system is an active camera tracking system. The PTZ (electronic pan/tilt/zoom controller) and the computer are connected by the controller. When the location or the velocity of the target is get, the computer set comment to controller. The electronic PTZ controller is moved to make the target object always in the vision. The two cameras (left camera and the right camera) are fixed on the cradle head to get video signal. The active motion of two digital cameras is completed by the cradle head. If the camera is the analog device, the image capture card is used. Otherwise, the digital video can be get directed by the computer through digital cameras. From this architecture, we can know that, the system has three parts: they are digital video getting unit, moving controller unit and image processing unit. So, through study visual servo control method, target recog-nition algorithm and tracking theory, the real-time track-ing system can be constructed.

Table 1. Hardware Components List

Device	Amount	Type and Parameters
Digital Camear	2	Dynamic pixel:500MP; Dynamic Resolution:1024×768; Frame rate 30f/s Size 1024×768 800×600 640×480 320×240
Cradle Head	1	Hjy PT1030;2 degrees of freedom, real-time feedback of rotation angles
Host Computer	1	FounerPC:Memory>2G
Cradle Head Controller	1	Hjy PT10302:cradle head compatible

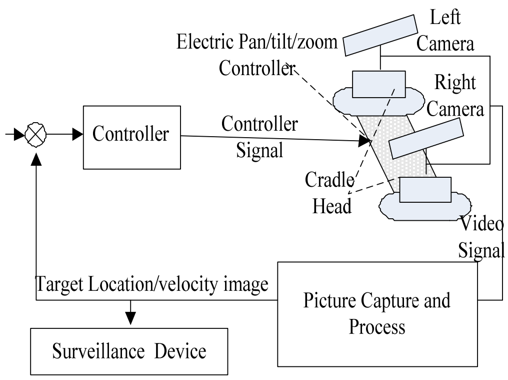

Figure 2. Tracking System Architeture

2.1. Basic Components and the Interface of the Tracking System

According to the tracking system architecture, Table 1 give the hardware component list. The digital video is captured by two digital cameras. The frame rate of the two digital cameras is 30f/s, and this capture image can be in different size. The two digital cameras are fixed on the cradle head, so the motion is decided by the cradle head, with two independent degrees of freedom: pitch and horizontal rotation. And the cradle head can send the real-time feedback of rotation angle information to the cradle head controller. When the two digital cameras are fixed on two dependent cradle heads, the feedback in-formation is used to coordinate the two cradle head to construct the binocular vision system. Host computer is used to surveillance the target object and send commands to the controllers. The pictures of the selected hardware component are shown in Figure 3, and they are selected by the consideration of interface consistency and compati-bility [11]. Figure 4 give the whole integrated binocular system.

We use visual C++ to build a visual surveillance environment (Figure 4), including system initialization, the surveillance windows, and some controlling bars. The system initialization is used to initialize the port used by controller, and when the system is connected with the computer, the system move horizontally and vertically to make sure that the system can work well. The four fine tuning bars are used to modify the position of the cradle head, and order the cradle head scanning in a setting angle. The surveillance information of two digital cameras is showed in two windows. And the positions of the current two cameras' pose are shown under the windows by the feedback information of the cradle head. In the future, the improved system will use two cradle head to make the two digital cameras independent. The current position information will become so important that the host camera will call the salve one to build a new system like the system our discussed. The system is often in the surveillance state without using tracking bar. The tracking bar is used to open the function to track the target object. When the target object appears, the target object detection algorithm and tracking algorithm are combined to calculate the position of the target, and then the computer send commends to cradle head to move.

Figure 3. Components of this System: (a) Cradle Head (b) Motion Controller (c) Digital Camera

Figure 4. Binocular Vision System

3. Active Object Tracking

In part 2, we have demonstrated that the system has three important units, video capture, tracking algorithm and visual servo control method. The video capture is easy for the use of digital camera, in this part, the tracking algorithm named Cam-shift and the visual servo control method based position are discussed in detail.

3.1. Cam-shift Algorithm

Mean shift [13] is a procedure for locating the maxima of a density function, which is given discrete data sampled from that function. It is useful for detecting the modes of density. And the procedure of mean shift is an iterative process. If give an initial estimate x, and a kernel function . This function determines the weight of nearby points for re-estimation of the mean. Usually, the Gaussian kernel in Equation (1) is used to estimate the distance.

$$K(x_i - x) = e^{c\|x_i - x\|} \quad (1)$$

The weighted mean of the density in the window determined by K is:

$$m(x)=\frac{\sum_{x_i\in N(x)}K(x_i-x)x_i}{\sum_{x_i\in N(x)}K(x_i-x)} \quad (2)$$

Where $N(x)$ is the neighborhood of x, a set of points for which $K(x)\neq 0$. The mean-shift algorithm nowsets:

$$x \leftarrow m(x) \quad (3)$$

and repeats the estimationuntil $m(x)$ converges.

Mean-shift [11] is a common used tracking method, and it has three features. They are: (1) the mean-shift approach is suitable for tracking non-grid moving target; (2) the convergence speed of the mean-shift algorithm is fast; (3) the kernel of the algorithm assigns smaller weight value to the edge area, and this character make the mean-shift method is insensitive to the edge noises. However, mean-shift algorithm is not robust to track the high speed target for its searching method.

And the Cam-Shift algorithm [11] is the improved algorithm of mean-shift algorithm, which has higher real-time processing ability and stronger robustness. The basic procedure of Cam-shift algorithm is the following 5 steps [11]:

(1) The region of interest (ROI) of the probability distribution image is set to the capture picture.

(2) An initial location of the Mean Shift search window is selected. The selected location is the target distribution to be tracked, for example, we can select the human face as the initial location.

(3) The color of image is used to track the target, and then the selected color's probability distribution of the region centered at the Mean Shift search window is calculated.

(4) Mean Shift algorithm is iterated to find the centroid of the probability image. The zeroth moment (distribution area) and centroid location should be storied.

(5) For the following frame, center the search window at the mean location found in Step 4 and set the window size to a function of the zeroth moment. Then go to Step 3.

3.2. Position based Pose Adjusted Algortihem

When the target object is detected, we should track the target object continually, which is very useful. For ex-ample, if the target moving fast, the multi-cameras can continually track the moving target. The use of Cradle Head can move according to the target object's moving direction. The detail of the moving strategy is as follows in Figure 5.

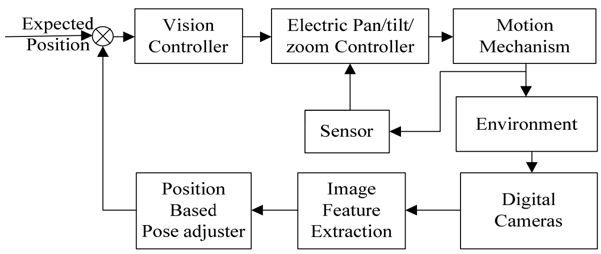

Figure 5. Position based Pose Adjusted Mechanism

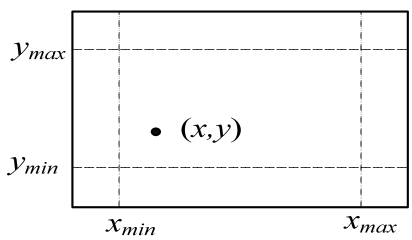

Figure 6. Moving Mechanism

Firstly, the system has an expected position due to the training information. The vision controller makes sure that the moving target is always in the scope of the vision. When the moving target is out of the defined scope, the feature of the image in the camera is firstly extracted, and then the system computes the position of the target feature. We define the point in Figure 6 is the computer feature of the moving target, and (x,y) is its coordinate. There are four defined coefficients, and they are xmin, xmax, ymin, ymax, which are the margin of the horizontal position and the vertical position, respectively. The moving mechanism commands the sensor to adjust the position of the electric Pan/tilt/ zoom controller.

The detail strategy of the cradle head moving is as follows:

If $x > x_{\max}$, move right;

If $x < x_{\min}$,move left;

If $y > y_{\max}$, move upward;

If $y < y_{\min}$, move upward;

If $x > x_{\max} \& y > y_{\max}$, move right firstly, then move upward;

If $x > x_{\max} \& y < y_{\min}$, move right firstly, then move downward;

If $x < x_{\min} \& y > y_{\max}$, move left firstly, then move upward;

If $x > x_{\max} \& y < y_{\min}$, move right firstly, then move downward;

When the cameras detect the target is out of the scope both in the horizontal direction and vertical direction, we firstly make the cradle head move horizontally in order to let the moving target in [x_{min}, x_{max}], and then make the cradle head move upward or downward in order to let the moving target always in [y_{min}, y_{max}].

4. Realization of this System

A stereo vision system that attempts to achieve ro-bustness with respect to scene characteristics is realized. This system focuses the continually tracking. We use Cam-shift algorithm to track the moving target. In Figure 7 and 8, we make the person and the colored paper is the targets. The scope of horizontal degree is -150 degree to +150 degree. And the vertical scope is -30degree and +30 degree. That is to say, the pitch angle is -30degree and +30 degree.

Figure 7. Person Tracking

Figure 8. Moving Target

From the experimental results, we can get the conclusion that the algorithm is robust, and the system can continually track the moving target. The static tracking system has many research results. The cameras are not moving, so the detection of the moving target is so easy, that the different of neighbor frame can be recognized as the moving target. For active cameras, how to track the moving target in real-time should be deeply researched. We can combine the active cameras and static cameras together to track the moving target continually. For example, in Figure 9, central area is the building; and we can arrange the two type cameras. The static cameras are fixed on the flat wall. The active cameras are fixed on the corner in order to expand the scope of the surveillance. And all of these cameras can feedback the target’s information to the controller center. And this system can be used in many fields, such as 3-D reconstruction, surveillance and forth. So object fusion is crucial. Snidaro [7] presents a same system. They employ multiple cameras to track the object; every sensor (camera) gets the position of the center of the bounding box enclosing the object searched by the ensemble of sensors and their projection on a common map as a means to perform heterogeneous sensor fusion. And it provides a detailed treatment of the novel concept of likelihood projection by both full likelihood projection and Gaussian approximation. Before fusing the position of the object detected by multiple sensors, a common coordinate frame is required. This literature provides a meaningful direction for our follow-up work, the fusion of multiple sensors will further complete this binocular vision system.

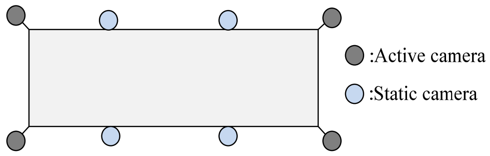

Figure 9. Tracking System with Active Cameras and Static Cameras

4. Conclusion

In this paper, we discuss the basic procedure of the Cam-shift algorithm and the algorithm of position based pose adjusted. The basic architecture of the tracking system including the hardware components, the structure is given. From the experimental result, we can see the useful of this system.

There are also some improved aspects in the next research. The tracking algorithm named cam-shift is a semi-automatic tracking algorithm, and the automatic tracking accuracy is not very high. If we want to improve the accuracy, we should select the target object manually. Secondly, this tracking method is based color, when color of the target object and the background are similar, the tracking accuracy is low. The shape information can be used to make the robustness stronger. In the future, we will improve the system architecture, and the two digital cameras will be replaced with cameras with high resolution. And we will make the two digital cameras fit on two independent cradle heads. The vision field of this system will be extended. One camera will be selected as the host camera, and the other one will be the slave one. When one of the cameras finds the moving target, the cradle head will send the feedback information to alert the other one to move until that the two cameras construct a new system like the system in this paper. Then the target match algo-rithm is to be used to make sure that the target in the two cameras is the same one. And the depth information can be got. This system will be used in the public security area, for example, this system can be arranged in the railway station, an so on. And the criminal information, such as the face images, are stored into the database, we can use this system to track the criminal's moving behavior. At the same time, the binocular vision system can construct the three-dimension structure of the target after getting the depth information, which will help us to recognize the target more clearly.

Acknowledgments

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant(No.2011CB302601) and Grant (No.2012CB315800) and in part by the N ational Natural Science Foundation of China (NSFC) Grants(No. 61222213, No. 61170290, and No. 60736012), and supported by the Fundamental Research Funds for the Central Universities (No. 1112KYQN40, No.1112KYZY49, No.0910KYZY55), and Cultural heritage protection science and technology project of Cultural relics bureau (No. 20110135). Our gratitude is extended to the anonymous reviewers for their valuable comments and professional contributions to the improvement of this paper.

References

[1] Allen JG, Xu Richard YD, Jin JS. *Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces*. In Proceedings of the Pan-Sydney area workshop on Visual information processing. Sydney. 2004; 3-7.

[2] Haritaoglu I, Harwood D, Davis L. W4: Real-time Surveillance of People and Their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(8): 809-830.

[3] Pavlidis I, Morellas V, Tsiamyrtzie P, Harp. S, Urban Surveillance system: From the Laboratory to The Commercial World.*Proceedings of the IEEE*. 2001; 89(10):1478-1497.
[4] Li J, Zhang H, Zhou JZ, GuoHW, Wang B, Zhao QJ. *Object Tracking Using Improved Camshift with SURF Method*. 2011 International Workshop on Open-Source Software for Scientific Computation. Beijing. 2011; 136-141.
[5] Comaniciu D, Meer P. *Robust Analysis of Feature Spaces: Color image segmentation*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence. 1997; 750-755.
[6] Qiu ZM, Li J, Ge J, Zhou QB. The Reserch of Automatical Object Recognizing Algorithms Based on Hausdorff Distance. *Infrared technology*. 2006; 8(4): 199-202.
[7] Snidaro L, Visentini I, Foresti GL. Fusing Multiple Video sensors for surveillance. ACM Trans. *Multimedia Comput. Commun*. 2012; 8(1). doi:10.1145/2071396.2071403
[8] Cheng PG, Yong JH. Study on Tracking of Moving Object in Intelligent Video Surveillance System. *Advanced Materials Research*. 2012. 433:6383-6585.
[9] LiX, WuHF, Qiu TS. *Real-time Object Tracking via CamShift-based Robust Framework*. International Conference on Information Science and Technology. Wuhan. 2012; 527-530.
[10] Yin JQ, Han YB, Li JP, Cao AZ. *Research on Real-Time Object Tracking by Improved Camshift*. International Symposium on Computer Network and Multimedia Technology. Wuhan. 2009; 1-4.
[11] Tian LF, Wu LC, Wang Y, Yang GS. *Binocular Vision System Design and Its Active Object Tracking*. Proceedings International Symposium on Computational Intelligence and Design. Hangzhou. 2011; 278-281.
[12] Lin LL, Ju CL, Zhou WH, Wang XP. Moving Objects Detection Based on Gaussian Mixture Model, LBP Texture and Saliency Map. *Applied Mechanics and Materials*. 2012; 63: 350-354.
[13] Li ZZ, Liu M, Wang HG, Yang Y, Chen J, Jin G. Gray-scale Edge Detection and Image Segmentation Algorithm Based on Mean Shift. *TELKOMNIKA*. 2013; DOI: 10.11591/telkomnika.v11i3.2223
[14] Firdausy K, Mirza MH. The Prototype of Real-time Object Detection System Based on SMS *TELKOMNIKA*. 2010; 08(2): 107-114.