

Design and Implementation of UAV Vision Tracking System

Qiang Wu^{1,a}, Wenpeng Chen^{2,b}, Xin Zheng^{3,c} and Xuwen Li^{4,d}

^{1,2,3}College of Information and Communication Engineering, Faculty of Information Technology,
Beijing University of Technology, Beijing, 100124, China

⁴ College of Life Science and Bioengineering, Beijing University of Technology
Beijing, 100124, China

^aemail: wuqiang@bjut.edu.cn, ^bemail: wpchen24@163.com,

^cemail: xz@bjut.edu.cn, ^demail: lixuwen@bjut.edu.cn

Keywords: Visual Tracking, UAV, Scale Adaptive Mean Shift, Image Processing

Abstract. UAV is increasingly popular in the field of military, civil and scientific research, since its low loss, strong adaptability and high mobility. This paper introduces the design and implement of the visual tracking system based on the rotor UAV, and gives the UAV a function of continuous and stable tracking of the ground target. Based on the standard Mean Shift algorithm, we introduce the improved object tracking algorithm with the spatial histogram and scale estimation. The robustness to scale adaptive and background related interference is improved with the enhanced tracking algorithm. It is finally implemented in our image processing board with TMS320C6657 and ZYNQ 7020. The whole system is optimized and accelerated by both of the two core of DSP. We finally achieve the real-time tracking and monitoring of the ground target in the range of 100m ~ 1km.

Introduction

In recent years, UAV is more and more popular as for its god perspective. And it is more usually used in the field of computer vision applications. In this paper, we choose the rotor UAV as the experiment platform. And a image processing board with TMS320C6657 and ZYNQ 7020 is designed and produced too. That let us to give the UAV a function of continuous and stable tracking of the ground target. We use the standard Mean Shift target tracking algorithm as the visual processing basic algorithm in this system. The improved object tracking algorithm with the spatial histogram and scale estimation is proposed to make the algorithm more robust. Finally, the proposed tracker is implemented in the image processing board, and it completes the real-time solution of the target position in the image sequence. The UAV achieves the real-time tracking and monitoring of the ground target in the range of 100m ~ 1km with our achievement.

Mean shift tracking is a popular algorithm for object tracking since it is easy to implement and it is fast and robust. It was first proposed by Fukunage and Hostetler in 1975 [1]. And it was first applied in the target tracking field by Comaniciu. The Mean Shift is a non-parametric mode-seeking method for density functions with no prior knowledge. The original Mean Shift algorithm uses a color histogram as a feature description of the tracked target. And the tracking window is fixed. Once the target is zoomed in, the rotation and the target color are similar to the background, the tracking performance is degraded or even lost. For the problems of the Mean Shift algorithm, the tracking window is scaled at a fixed scale of $\pm 10\%$ was proposed by Comaniciu, and that the window can be adjusted by multiple iterations [2]. However this method cannot be better adapted to the process of getting bigger. Collins uses the Lindeberg scale space theory to adapt to the target scale change [3]. Liang et al. search for the object boundary by correlating image with the four templates [4]. This kind of method often has the problem of large computation and difficult to transplant, and it is difficult to realize real-time processing in the embedded platform. Zhao et al. exploit affine structure to recover object relative scale from feature point correspondences between consecutive frames [5]. Methods depending on feature matching are able to estimate the scale quite robustly, but they cannot be

seamlessly integrated to the Mean-Shift framework. This method is strongly dependent on the local invariant feature of the target, it is difficult to achieve continuous tracking of non-rigid targets.

In this paper, the proposed algorithm uses spatial histogram to replace the single description of the target with a color histogram. And it is more robust when the target color is similar to the background. Refer to the standard Mean Shift algorithm iterative process, we propose a scale adaptive estimation algorithm. It has been shown to have good adaptability to target scale scaling.

Standard Mean Shift Tracking Algorithm

The standard Mean Shift tracking algorithm includes two steps: target description and target location.

Let $\{x_i^*\}_{i=1,2,\dots,m}$ denote the pixel locations of the target model centered at the y^* . We choose an m -bins kernel-estimated histogram to describe the color space of the target area as follows:

$$\hat{q}_u(y^*) = C \sum_{i=1}^m k\left(\left\|\frac{y^* - x_i^*}{h}\right\|^2\right) \delta[b(x_i^*) - u], \quad u = 1, 2, \dots, m \tag{1}$$

where $b(x_i^*)$ is the pixel value of x_i^* . And h is the width of the kernel function that determines the size of the target area. C is a normalization constant. Similarly, the candidate target area is as follows:

$$\hat{p}_u(y) = C_h \sum_{i=1}^m k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \delta[b(x_i) - u], \quad u = 1, 2, \dots, m \tag{2}$$

The difference between target model and candidate target is measured by Bhattacharyya distance, which is known to be a metric:

$$\hat{\rho}(y) = \rho[\hat{p}(y), \hat{q}(y^*)] = \sum_{u=1}^m \sqrt{\hat{p}_u(y) \hat{q}_u(y^*)} \tag{3}$$

The search for the new target location in the current frame starts at the location \hat{y}_0 of the target in the previous frame using gradient ascent with a step size equivalent to the MeanShift method. And the new location is described as follows:

$$\hat{y}_1 = \frac{\sum_{i=1}^m x_i w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^m w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)} \tag{4}$$

where $w_i = \frac{\sqrt{\hat{q}_u(y^*)}}{\sqrt{\hat{p}_u(\hat{y}_0)}}$

Scale Adaptive Tracking Algorithm

The standard Mean Shift tracking algorithm includes two steps: target description and target location. In the traditional Mean Shift tracking algorithm, only the target position joins the process of iteration, while the kernel function window remains unchanged. To achieve the scale adaptation of the tracking window, it is assumed that the target is elliptical in the current frame and the scale changes in an isotropic manner. The target model and target candidate are described as follows:

$$\hat{q}_u = C \sum_{i=1}^N k\left(\frac{(x_i^1)^2}{a^2} + \frac{(x_i^2)^2}{b^2}\right) \delta[b(x_i) - u] \tag{5}$$

$$\hat{p}_u(y, h) = C_h \sum_{i=1}^N k\left(\frac{(y^1 - x_i^1)^2}{a^2 h^2} + \frac{(y^2 - x_i^2)^2}{b^2 h^2}\right) \delta[b(x_i) - u] \tag{6}$$

where $C_h = \frac{1}{\sum_{i=1}^N k\left(\frac{(y^1 - x_i^1)^2}{a^2 h^2} + \frac{(y^2 - x_i^2)^2}{b^2 h^2}\right)}$

The parameter h defines the scale of the target candidate. As known as the area ratio of the ellipse, the number of the candidate pixels is the h^2 times of the model pixels. Using the definition of Riemann integral we obtain that $C_h \approx C / h^2$ and for any two values h_0, h_1 $C_{h_1} \approx C_{h_0} \frac{h_0^2}{h_1^2}$.

Using the approximations above, the Bhattacharyya distance can be described as follows:

$$\rho[\hat{p}(y, h), \hat{q}] \approx \rho(y, h) = \sum_{u=1}^m \sqrt{C_{h_0} \frac{h_0^2}{h^2} \sum_{i=1}^N k\left(\frac{(y^1 - x_i^1)^2}{a^2 h^2} + \frac{(y^2 - x_i^2)^2}{b^2 h^2}\right) \delta[b(x_i) - u]} \hat{q}_u \quad (7)$$

Using the same gradient analysis method as the traditional Mean Shift tracking algorithm, the local maximum of the Bhattacharyya distance can be obtained. And the kernel with a scale parameter h_0 is iteratively moved from current location \hat{y}_0 in direction of $\nabla \rho(\hat{y}_0^1, \hat{y}_0^2, h_0)$ to the new location \hat{y}_1 , changing its scale to h_1 . At this time, we can get that:

$$w_i = \sum_{u=1}^m \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0, h_0)}} \delta[b(x_i - u)] \quad (8)$$

Let us denote

$$m_k(\hat{y}_0, h_0) = (m_k^1(\hat{y}_0, h_0), m_k^2(\hat{y}_0, h_0))^T = \frac{\sum_{i=1}^N x_i w_i g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)}{\sum_{i=1}^N w_i g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)} - \hat{y}_0 \quad (9)$$

Then seeking derivative of ρ at y_1, y_2 and h , and the derivative of h is:

$$\frac{\partial \hat{\rho}(y, h)}{\partial h}(\hat{y}_0, h_0) = \frac{C_{h_0}}{(h_0)^2} \sum_{i=1}^N w_i g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right) \cdot \left[\frac{1}{h_0} \frac{\sum_{i=1}^N w_i \left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right) g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)}{\sum_{i=1}^N w_i g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)} - h_0 \frac{\sum_{i=1}^N w_i k\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)}{\sum_{i=1}^N w_i g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)} \right] \quad (10)$$

Finally, we obtain that:

$$\hat{y}_1^1 = \frac{1}{a^2} m_k^1(\hat{y}_0, h_0) + \hat{y}_0^1, \hat{y}_1^2 = \frac{1}{b^2} m_k^2(\hat{y}_0, h_0) + \hat{y}_0^2 \quad (11)$$

$$h_1 = \left[1 - \frac{\sum_{i=1}^N w_i k\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)}{\sum_{i=1}^N w_i g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)} \right] h_0 + \frac{1}{h_0} \frac{\sum_{i=1}^N w_i \left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2}\right) g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)}{\sum_{i=1}^N w_i g\left(\frac{(\hat{y}_0^1 - x_i^1)^2}{a^2 h_0^2} + \frac{(\hat{y}_0^2 - x_i^2)^2}{b^2 h_0^2}\right)} \quad (12)$$

Using the iterative method consistent with the standard Mean Shift tracking algorithm, the local maximum value is iterated along the gradient direction, and the scale of the new position is replaced by the scale iterated in the tracking process.

The comparison between the standard Mean Shift tracking algorithm and the scale adaptive Mean Shift tracking algorithm is shown in Fig.1 and Fig.2.

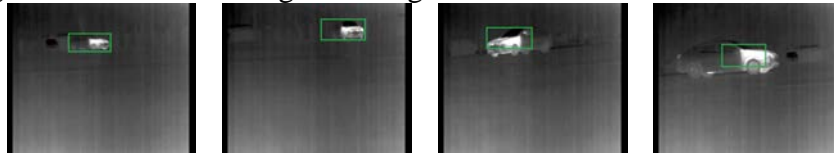


Figure 1. Standard Mean Shift tracking algorithm renderings

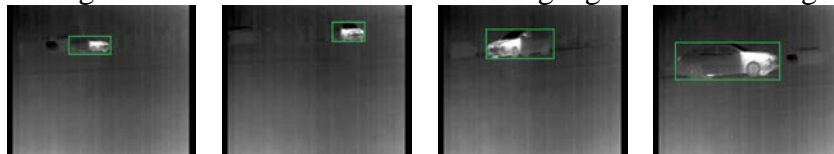


Figure 2. Scale adaptive Mean Shift tracking algorithm renderings

The color information is used as the only description of the target area as traditional algorithm does. Using the concept of second order spatial histogram, we add the color space mean and covariance matrix to the description here.

$$\mu(y) = \frac{1}{\sum_{i=1}^N \delta_i} \sum_{i=1}^N (x_i - y) \delta_i \quad (13)$$

$$\Sigma(y) = \frac{1}{\sum_{i=1}^N \delta_i} \sum_{i=1}^N (x_i - \mu(y))^T (x_i - \mu(y)) \delta_i \quad (14)$$

The spatial histogram describes the spatial information of the regional color distribution while acquiring the color information. Now the Bhattacharyya distance can be described as follows:

$$\rho(y) = \rho(h(y), h') = \sum_{i=1}^N \psi(y) \sqrt{n(y) \cdot n'} \quad (15)$$

where $\psi = \eta \exp\left(-\frac{1}{2}(\mu - \mu')^T \cdot \sum_{i=1}^n (\mu - \mu')^{-1} \cdot \sum_{i=1}^n (\mu - \mu')^{-1}\right)$, and n is the pixel of the target area.

Also use the iterative principle of the standard Mean Shift algorithm, we can obtain that:

$$y_1 = \frac{\sum_{i=1}^N \alpha_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right) x_i - \sum_{i=1}^n v}{\sum_{i=1}^N \alpha_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)} \tag{16}$$

where $\alpha_i = \frac{C_h}{h^2} \sum_{i=1}^n \varphi(y_0) \sqrt{\frac{n}{n(y_0)}} \delta_i$, $v = \varphi(y_0) \sqrt{n n(y_0)} \sum^{-1}(y_0) (\mu^i - \mu(y_0))$.

With spatial characteristics of the regional color distribution integrated into the iterative process, the tracking effect is better robust when the target is similar to the background color. The Mean Shift tracking algorithm is compared with the Mean Shift tracking algorithm based on spatial histogram. The result is shown in Fig.3 and Fig.4.



Figure 3. Standard Mean Shift tracking algorithm renderings



Figure 4. Mean Shift tracking algorithm based on spatial histogram renderings

We compare the algorithm above with the traditional Mean Shift algorithm and the $\pm 10\%$ scale adaptation algorithm mentioned in the article [2] in the accuracy and time complexity. Performance of the algorithms was assessed by the recall (the number of correctly tracked frames divided by number of frames where the target is visible). Let the tracking algorithm window area be denoted by T, and the target actual window area is denoted as G. The frame was considered tracked correctly if the overlap with the ground truth was higher than 0.5 measured as $(T \cap G)/(T \cup G)$. The result is shown in Fig.5.

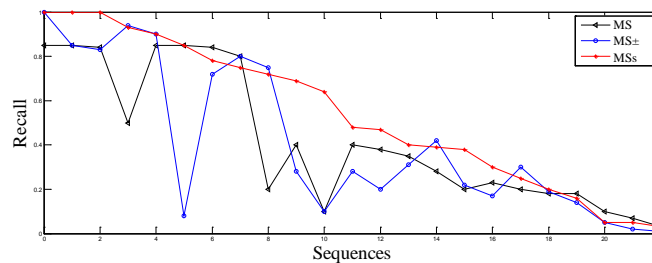


Figure 5. Comparison of different algorithms

We know that the execution time of the MeanShift tracking algorithm is related to the size of the trace window. In the comparison experiment, we specify the same initial window position and size, the comparison results are shown in Table 1.

Table 1. Algorithm time complexity comparison result (Unit: milliseconds)

	MSs	MS±	MS
Max	28	125	16
Min	1	3	1
Mean	8	14	3

System Test

The UAV system selected in this paper includes the basic components such as unit platform, ground station equipment, Futaba multimode remote control and wireless data transmission. We chose the eight-rotor UAV as an experimental platform, which is better stability, maneuverability and better load performance than the more common four-rotor UAV. In order to obtain better shooting results, we use the three-axis stabilization platform to correct the mechanical vibration, and the control accuracy is better than 0.01° .

Here are the system design block diagram and physical map:

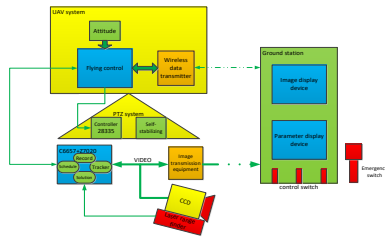


Figure 6. System design block diagram



Figure 7. System design physical map

The image processing board is designed with TMS320C6657 DSP + ZYNQ7020. Among them, the ADV7180 acquires the PAL@25Hz analog video signal from the CCD and translates it to YUV422 format video frame to the ZYNQ7020 PL (FPGA). Then PL sends them to C6657 DSP through the uPP data bus. In order to realize the real-time processing of video frames, we use both the two core of DSP to accelerate the proposed algorithm. We let the core0 to be the main control core, which is responsible for data transceiver and target description. And let the core1 to be responsible for iterative operations. With the acceleration of DSP, the proposed algorithm has reached the real-time requirements. The result of the algorithm is sent back to the ZYNQ7020 PS part in the form of pixel coordinates. PS (ARM) is responsible for summarizing the coordinate information, laser ranging information and CCD parameters. Then PS calculates them to the real spatial coordinates and completes the PTZ angle control. The video and coordinates are transmitted to the ground station at the same time. With above, we give the UVA a function of sustained and stable tracking of ground targets.

The actual experimental renderings are as follows:

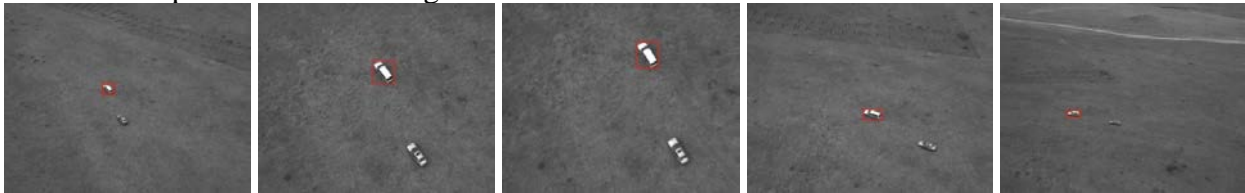


Figure 8. The actual experimental renderings

Conclusions

In this paper, we design a scale-adaptive Mean Shift tracking algorithm based on spatial histogram. Compared with the standard Mean Shift algorithm, the robustness to scale adaptive and background related interference is improved. The proposed algorithm is implemented on the image processing board consists of TMS320C6657 DSP and ZYNQ7020. After serial and parallel optimization, the tracker can achieve the 25fbps video signal real-time processing. The whole system is mounted on the rotor UAV platform. After actual experiments, we do have achieved the stability tracking of the UAV system to the ground target.

References

- [1] Fukunaga K, Hostetler L D. The Estimation of the Gradient of a Density Function [J].IEEE Trans. Information Theory, 1975, 21: 32-40
- [2] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid object using mean shift[C], Proc of IEEE Conf on Computer Vision and Pattern Recognition. Hilton Head island:[s.n.],2000:142-149.
- [3] Collins, R.T.: Mean-shift blob tracking through scale space. In: Computer Vision and Pattern Recognition. pp. 234–240. IEEE Computer Society (2003)
- [4] Liang, D., Huang, Q., Jiang, S., Yao, H., Gao, W.: Mean-shift blob tracking with adaptive feature selection and scale adaptation. In: International Conference Image Processing (2007)
- [5] Zhao, C., Knight, A., Reid, I.: Target tracking using mean-shift and affine structure. In: ICPR. pp. 1–5 (2008)