The research on the mean shift algorithm for target tracking

Honghong CAO

Software Engineering College (Chongqing University of Posts and Telecommunications), Chongqing 400065, China

Abstract. The traditional mean shift algorithm for target tracking is effective and high real-time, but there still are some shortcomings. The traditional mean shift algorithm is easy to fall into local optimum in the tracking process, the effectiveness of the method is weak when the object is moving fast. And the size of the tracking window never changes, the method will fail when the size of the moving object changes, as a result, we come up with a new method. We use particle swarm optimization algorithm to optimize the mean shift algorithm for target tracking, Meanwhile, SIFT (scale-invariant feature transform) and affine transformation make the size of tracking window adaptive. At last, we evaluate the method by comparing experiments. Experimental result indicates that the proposed method can effectively track the object and the size of the tracking window changes.

1 Introduction

Moving target tracking in video or image sequence is aimed at marking the region of interest in each frame image of the video. With the development of science and technology, moving target tracking has been widely used in such fields as motion analysis, behavior recognition, intelligent video surveillance, and human-computer interaction[1]. Moving target tracking has gradually become one of the important research subject in the field of computer vision. Although under the efforts of researchers moving target tracking methods have got mature, there are also some problem to solve[2]: occlusion,illumination, complex background disturbances, scale changes, appearance changes. Many tracking algorithms have been proposed, the existing target tracking algorithm can be roughly divided into four classes: (1) target tracking algorithm based on feature matching;(2) target tracking algorithm based on learning[3]; (3) target tracking algorithm based on active contour[4]; (4) target tracking algorithm based on model. Mean shift tracking algorithm which was come up by foreign researchers belongs to target tracking algorithm based on feature matching. Mean shift tracking algorithm also have shortcomings, which has many advantages such as good real-time performance and small amount of calculation.

Mean shift target tracking algorithm make color histogram feature by calculate the gray value of the pixels in the target region and the candidate region, then in order to make the highest degree of similarity between the target area and he target candidate region, we should iterative the mean shift vector constantly to make sure the target center converge to the true location of the target, eventually the algorithm achieve the purpose of tracking. There are some disadvantages in the traditional algorithm: (1) the size of the tracking window won't change while the scale of the target changing; (2) the mean shift algorithm is easy to trap into local optima. In order to overcome the shortcomings, we come up with a method which makes use of the particle swarm optimization to improve the mean shift algorithm to make the local optima into global optima, meanwhile extracts and matches the SIFT features of the for the previous frame and the current frame, then changes the size of the tracking window by figuring out the affine transformation matrix.

2 Mean shift algorithm for target tracking

Mean shift was come up by Fukunaga and Hostetler[5] in 1975, but people did not expect to apply this theory to target tracking. In 2003, Comaniciu published a paper[6] which successfully applied the mean shift theory to target tracking.The principle of the mean shift algorithm for target tracking: (1) manually mark the moving target region in the first frame and extract the feature of the moving target region; (2) extract the feature of the candidate moving target region and calculate the degree of the similarity between the candidate model and the target model using some algorithm. (3) figure out the mean shift vector making sure the similarity is the highest.

2.1 Feature extraction

At first, we mark the target region of interest using rectangular box,assume that the center of the target area is x_0,there are n pixels which are indicated by $\{x_i\}_{i=\dots\dots n}$. There are m bins,u indicates the corresponding bin, the color histogram is

$$\widehat{q}_u = c\sum_{i=1}^{n} k\left(\left\|\frac{x_0 - x_i}{h}\right\|^2\right)\delta[b(x_i)-u] \qquad (1)$$

$$C = \frac{1}{\sum_{i=1}^{n} k\left(\left\|\frac{x_0 - x_i}{h}\right\|^2\right)} \qquad (2)$$

$$\sum_{u=1}^{m} \widehat{q}_u = 1 \qquad (3)$$

where C is a constant,for each pixel,if the pixel belongs to the u bin,the value of $\delta(x)$ is 1,or else the value of $\delta(x)$ is 0. The same as the target model ,the color histogram of the candidate target model is

$$\widehat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)\delta[b(x_i)-u] \qquad (4)$$

2.2 Similarity measure

After obtaining the features of the target model and the target candidate model, we need to calculate the similarity between them to judge whether the current region matches the target model. We use Bhattacharyya coefficient[8] to measure, the formula of Bhattacharyya coefficient is

$$\widehat{\rho}(y) \equiv \rho(\widehat{p}(y),\widehat{q}) = \sum_{u=1}^{m} \sqrt{\widehat{p}_u(y)\widehat{q}_u} \qquad (5)$$

the lager the value of $\widehat{\rho}(y)$ is,the higher of the similarity.And the scale is [0,1].

2.3 Mean shift vector

The location of the target center that maximize value of the Bhattacharyya coefficient is the target center of the current frame. Assume that the position of the target center of the previous image is y_0, we Taylor expand $\hat{\rho}(y)$ at y_0, the Bhattacharyya coefficient can be approximately expressed as

$$\rho\big(\hat{p}(y),\hat{q}\big)=\frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(y_o)\hat{q}_u}+\frac{C_h}{2}\sum_{i=1}^{n_h}w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \qquad (6)$$

Where

$$w_i=\sum_{u=1}^{m}\sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}}\delta\big[b(x_i)-u\big] \qquad (7)$$

Obviously ρ is determined by the second part in the formula (6), we define

$$f_{n,k}=\sum_{i=1}^{n_k}\frac{C_h}{2}w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \qquad (8)$$

To maximize the Bhattacharyya coefficient is to make sure the $f_{n,k}$ is the largest. By calculating the derivative of $f_{n,k}$ and making it equal to zero, the mean shift vector will be obtained, consequently we can get the location of the target center in the current frame.

3   Using particle swarm optimization algorithm to optimize the mean shift tracking algorithm

3.1 The principle of the particle swarm optimization algorithm

Particle swarm optimization[9,10] is a biological simulation algorithm, which simulates the foraging behavior of birds. As shown in Figure 1, the basic principle of the algorithm[11]: (1) randomly initialize a group of particles and set the initial speed; (2) according to fitness function, calculate the fitness of each particle and record the historical optimal value pbest and global optimal value of each particle gbest; (3) calculate and update particle position; (4) repeat from step (1) to step (3) to meet the stop condition and calculate the latest global optimal value according to the position updating formula.
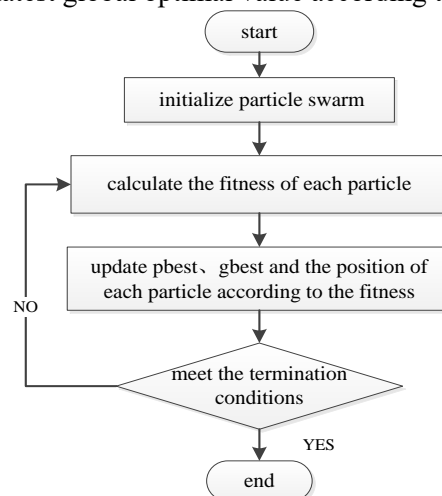

Figure 1:   The principle of particle swarm optimization

3.2 Application of particle swarm optimization algorithm in moving target tracking

The basic principle of the mean shift target tracking algorithm is to use the gradient ascent of the probability density to find the local optimum, as a consequence when the object is moving fast, the tracking effect is poor. In order to solve this problem, the traditional mean shift method need to be improved to search the global optimal solution as far as possible. In this paper, particle swarm optimization algorithm is applied to moving object tracking: initialize a set of random particles in a circular region with a radius of r around the center of the tracking window of the previous frame, the best particle position is obtained by using the Bhattacharyya coefficient as the fitness function. And then through the mean shift iteration to find the optimal location.

4   Scale adaptive

In the tracking process using the traditional mean shift target tracking algorithm, the size of the tracking window always remain unchanged. In fact, the target may be enlarged or reduced in motion, so it is necessary that the size of the tracking window changes while the target enlarging or shrinking. At the same time, the scale of the kernel function will make the same change, and the adaptive scale will improve the tracking accuracy.

SIFT was proposed by David Lowe[12] in 1999. The algorithm is widely used in target recognition, image mosaic, 3D modeling, gesture recognition, video tracking. SIFT[13,14] can be widely used because it is robust to rotation, scaling and affine transformation. This feature provides a theoretical basis for our next use of affine transform to achieve scale adaptation. After affine transformation of a two-dimensional graphics, it is still a two-dimensional graphics, it can be completed by a series of atomic operations, such as scaling, translation, wrong cut, flip, rotate. Since the SIFT feature points are stable for rotation, scale and illumination, it is feasible that extract SIFT feature of the previous and current frames, then   match the SIFT feature points and calculate the affine transformation matrix between matching points, and finally achieve the purpose of adjusting the size of the tracking window.

5   Experimental results and analysis

In order to verify the effectiveness of the proposed method, the experimental results obtained by the proposed method and the classical mean shift method are compared from two aspects of scale adaptation and accuracy. The software and hardware equipment used in this experiment are: DELL series laptop, Processor is Intel Core i5-4210, CPU clocked at 1.7 GHZ, memory (RAM) is 4GB, the installation of windows10 64 operating system, equipped with MATLAB r2013b software. The data used in this experiment is image sequences on the website of Visual Tracker Benchmark. The advantage of using this data set is that the data set comes with ground truth data, which saves a lot of work.

Figure 2:   The experimental results after improvement

Figure 3:   Target tracking using mean shift

Figure 2 and 3 are experimental results on the image sequence Car, figure 2 shows the experimental results obtained using the proposed method, figure 3 is the experimental results obtained using the traditional mean shift method. In order to facilitate the observation , we pick five discontinuous images. It is obvious that the size of the tracking window changes using the proposed method, however, the traditional mean shift target tracking method always keeps the size of the tracking window in the tracking process.
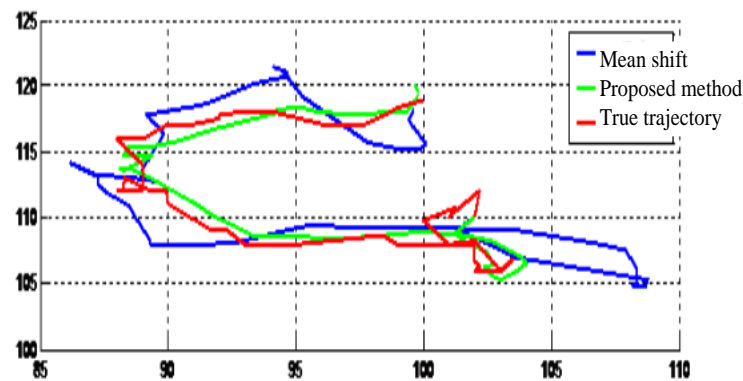
Figure 4:   The trajectory of moving target

Figure 4 depicts the position coordinates of the target trajectory that is the center of the target, where the red curve represents the true position coordinates, the blue curve represents the position coordinates using the traditional mean shift method, the green curve represents the position coordinates using the proposed method. It can be seen from the figure that the fitting degree of the green curve and the red curve is higher than that of the blue curve and the red curve. Table 1 compares the accuracy of the two methods. The evaluation criteria[15] adopted in this paper is: the accuracy C is equal to the ratio of the intersection of the tracking window and the true window and their union. The higher the value of C, the higher the accuracy. The comparison experiment shows that the method is superior to the traditional method in terms of accuracy.

Accuracy C	Dog	Girl	Car
mean shift algorithm	0.8016	0.8592	0.8651
Proposed method	0.8481	0.9323	0.8997

Table 1: Accuracy

6 Conclusion

This paper first analyzes the shortcomings of the traditional mean shift target tracking algorithm: it is easy to fall into local optimum, the accuracy is not high enough; and the size of tracking window will not change with the change of target size. As a result, a new method based on mean shift algorithm is proposed to solve these problems. To adjust the position of the target center, which avoids the deviation caused by the case that the mean shift algorithm is easy to fall into local optimum in the iterative process, the proposed method takes advantage of the particle swarm optimization algorithm to improve the accuracy of target tracking. At the same time, the affine transformation is used to obtain the transformation matrix of the SIFT feature points to realize the scale adaptation. The experimental results show that compared with the traditional mean shift target tracking method, the accuracy of this method is improved. There are still some shortcomings of this method, further work is necessary: optimize the code to reduce processing time; the parameters may not be optimal, so we need to repeat the experiment to find the optimal value; collect different types of data sets and the evaluate the performance on different data sets.

References

[1] Parekh, Himani S., Darshak G. Thakore, and Udesang K. Jaliya. "A survey on object detection and tracking methods." International Journal of Innovative Research in Computer and Communication Engineering 2.2   pp. 2970-2979, (2014).
[2] Kristan M, Pflugfelder R, Leonardis A, et al. The Visual Object Tracking VOT2014 challenge results[J]. Lecture Notes in Computer Science,   8926:pp. 191-217, (2015).
[3] Kalal, Zdenek, Krystian Mikolajczyk, and Jiri Matas. "Tracking-learning-detection." IEEE transactions on pattern analysis and machine intelligence 34.7 pp. 1409-1422, (2012).

[4] Kass, Michael, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active contour models." International journal of computer vision 1.4 pp. 321-331, (1988).

[5] Fukunaga, Keinosuke, and Larry Hostetler. "The estimation of the gradient of a density function, with applications in pattern recognition." IEEE Transactions on information theory 21.1 pp. 32-40, (1975).

[6] Comaniciu, Dorin, Visvanathan Ramesh, and Peter Meer. "Kernel-based object tracking." IEEE Transactions on pattern analysis and machine intelligence 25.5 pp. 564-577, (2003).

[7] Shengli Zhu. "Research on video tracking based on Mean Shift and correlation algorithm." zhejiang university:[PhD dissertation] (2006).

[8] Yang, Changjiang, Ramani Duraiswami, and Larry Davis. "Similarity measure for nonparametric kernel density based object tracking." Eighteenth Annual Conference on Neural Information Processing Systems, Vancouver. (2004).

[9] Kennedy, J., and R. Eberhart. "Particle swarm optimization [A] Proceedings of the IEEE International Conference on Neural Networks [C] Piscataway." NJ, USA: IEEE (1995).

[10] Eberhart, Russell, and James Kennedy. "A new optimizer using particle swarm theory." Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on. IEEE, (1995).

[11] Chen, Guo-chu, and Jin-shou Yu. "Particle swarm optimization algorithm." INFORMATION AND CONTROL-SHENYANG- 34.3 pp. 318, (2005).

[12] Lowe, David G. "Object recognition from local scale-invariant features." Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, 1999.

[13] Juan, Luo, and Oubong Gwun. "A comparison of SIFT, pca-SIFT and surf." International Journal of Image Processing (IJIP) 3.4 pp. 143-152, (2009).

[14] Valgren, Christoffer, and Achim J. Lilienthal. "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments." Robotics and Autonomous Systems 58.2   pp. 149-156, (2010).

[15] Zhang, Kaihua, et al. "Fast tracking via spatio-temporal context learning." arXiv preprint arXiv:pp. 1311-1939, (2013).