A Tracking Algorithm Based on SIFT and Kalman Filter

Song Dan,Zhao Baojun,Tang Linbo
School of Information Science and Technology
Beijing Institute of Technology

***Abstract*-This paper presents a method of target tracking based on SIFT and Kalman filter. SIFT algorithm has the ability to detect the invariant feature points which used in tracking and Kalman filter has the ability to predict the target location. Firstly, this paper uses SIFT to compute the location of target. Secondly, this paper uses Kalman filter to optimize the target location in order to correct the error of SIFT algorithm precisely. Lastly, this paper uses 2 groups of videos to test this algorithm. The results show that this is an effective tracking method.**

Keywords-Tracking, SIFT, feature point, Kalman filter

I. INTRODUCTION

Target tracking is an important branch of computer vision, and it is also one of the most popular topics in image processing [1]. The typical algorithms of tracking are Mean-Shift, frame difference and particle filter [2]. Mean-Shift is a kind of kernel probability density estimation, and it can converge the maximum of the kernel probability density estimation quickly through out iteration [3]. However, this algorithm will disturb by illumination and block easily. Frame difference is a kind of tracking algorithm which used in static background tracking [4]. However, this algorithm is not suit for the tracking of dynamic background. Particle filter can effectively deal with the estimation of nonlinear dynamic systems but the large computation and the serious degradation are the disadvantages of particle filter [5].

At the present stage, target invariant feature extraction is an effective method to identify the key positions of targets, the best algorithm is SIFT. SIFT is called Scale Invariant Feature Transform [6]. It has the abilities to fit the changes of image scale, illumination and partial occlusion and it has been widely used in the target feature extraction [7]. However, because of the large number of feature points, the location computing of target is not accurate after many frames. This paper uses Kalman filter to deal with this problem. After accumulating some prior information, we use Kalman filter to predict the target location.

II. TARGET TRACKING BASED ON SIFT

There are two parts in the SIFT algorithm which called the feature points extraction and feature points matching, and the first part also contains feature points detection and feature points description.

(1) Scale-space extrema detection: The first stage of computation searches all the scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation [8]. It has been shown by Koenderink and Lindeberg that under a variety of reasonable assumptions the only possible scale-space kernel is the Gaussian function [9]. Therefore, the scale space of an image is defined as $L(x,y,\sigma)$, which is produced from the convolution of the variable-scale Gaussian function $G(x,y,\sigma)$ and the input image.

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{1}$$

In order to efficiently detect stable feature points in scale space, we use scale-space extrema in difference-of-Gaussian images, which can be compute from the difference of two nearby scales.

$$D(x,y,\sigma)=\left[G(x,y,k\sigma)-G(x,y,\sigma)\right]*I(x,y) \tag{2}$$

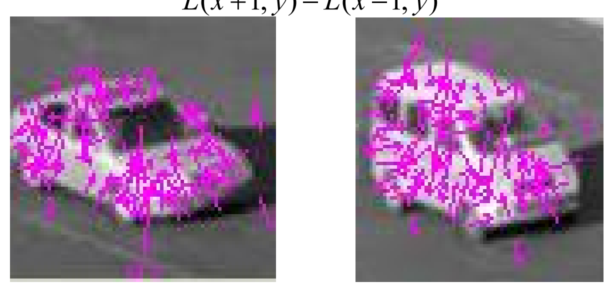
In order to detect the maxima of $D(x,y,\sigma)$, each sample point is compared with its eight neighbors in the current image and 9 neighbors in the scale above and below, then we can find the maxima in different scales and regard them as candidate points.

(2) Feature points localization: At each candidate location, a detailed model is used to determine the location and scale [10].

(3) Orientation assignment: One or more orientations are assigned to each feature point which based on local image gradient directions. The formula is shown as follows.

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \tag{3}$$

$$\theta(x,y)=\arctan\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)} \tag{4}$$



(a) (b)

Figure 2. The Orientation assignment of feature points

(4) Feature points descriptor: The local image gradients are measured at the selected scale in the region around each feature point. In order to enhance the robust of matching, each feature point is described with 4×4 descriptors [11]. There are eight orientations in each descriptor, so we use

$4 \times 4 \times 8 = 128$ elements feature vector to describe each feature point.

(5) Feature points matching: After we detect feature points of the target, we need to match them in order to make correspondence with the same target in two images. We use Euclidean distance to measure the similarity of two images, then we can use the ratio of nearest distant of the feature points and second nearest distant of the feature points in two images to match the feature points.

In the stage of feature point detection, we extracted a large number of feature points. However, in the process of target tracking, we only need a few feature points. In order to reduce the computation and determine the target location accurately, we have to select the main feature points. In the stage of orientation assignment, we calculate the gradient of each feature point. Firstly, we find the maximum of the gradient, and then select the gradients which are larger than 50% of the maximum. As a result, we find a group of feature points which have strong contrast.

Figure 3. Feature points after selected

After selecting feature points, we use kernel weighting function to calculate the target location. The role of kernel weighting function is to increase the weight of center pixels, decrease the weight of surrounding pixels. The location calculating formula is shown as follows.

$$x_{ave} = \frac{\sum_{i=1}^{n} k\left(\left\| \frac{x_{center} - x_i}{h} \right\|^2\right) x_i}{\sum_{i=1}^{n} k\left(\left\| \frac{x_{center} - x_i}{h} \right\|^2\right)} \quad (5)$$

The formula of Y coordinate is the same as formula (5). h is the length of tracking window, (x_{center}, y_{center}) is the initial target location of the first frame.

In the process of tracking, with the error accumulation, the location will not accurate. We can use Kalman filter to solve this problem.

III. MOVING PREDICTION BASED ON KALMAN FILTER

Kalman filter is proposed by Rudolf E. Kalman in 1960 [12]. The Kalman filter is a recursive estimator. This means that only the estimated state from the previous time and the current measurement are needed to compute the estimate for the current state [13]. The basic idea of Kalman filter is to build equations and observation matrix, and using the estimate of previous signal and the observation of current signal to make optimal estimation to the current signal which based on linear unbiased minimum variance estimation.

The formulas of Kalman filter are shown as follows.

(1) The predicted state estimate:

$$\hat{X}_{k|k-1} = A\hat{X}_{k-1|k-1} + BU_k \quad (6)$$

(2) The predicted estimate covariance:

$$P_{k|k-1} = AP_{k-1|k-1}A' + Q \quad (7)$$

Q is the covariance of system process.

(3) The updated state estimate:

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + Kg_k(Z_k - HX_{k|k-1}) \quad (8)$$

Kg is the Kalman gain. Z_k is the observation of time k.

(4) The optimal Kalman gain:

$$Kg_k = P_{k|k-1}H'/(HP_{k|k-1}H' + R) \quad (9)$$

(5) The updated estimate covariance:

$$P_{k|k} = (I - kg_k H)P_{k|k-1} \quad (10)$$

After a few frames of tracking, we accumulated a group of coordinates of target location, so we can use them to estimate the trajectory of target.

The target movement equation is shown as follows.

$$\begin{cases} x_{k+1} = x_k + v_k T \\ v_{k+1} = v_k + a_k T \end{cases} \quad (11)$$

x is the X coordinate of target location at time T, v is the speed of target at time T, a is the acceleration.

The state equation of system is shown as follows.

$$\begin{pmatrix} x_{k+1} \\ v_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_k \\ v_k \end{pmatrix} + \begin{pmatrix} 0 \\ a_k T \end{pmatrix} \quad (12)$$

The state transition matrix is shown as follows.

$$H(k) = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \quad (13)$$

The observation equation of system is shown as follows.

$$x_{k+1} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_k \\ v_k \end{pmatrix} \quad (14)$$

Now we can predict the location of current frame, the predict location is (x_{pre}, y_{pre}). The tracking location based on SIFT is (x_{new}, y_{new}).

$$\begin{cases} |x_{new} - x_{pre}| > 30 & x_{new} = x_{pre} \\ |x_{new} - x_{pre}| < 30 & x_{new} = (x_{pre} + x_{new})/2 \end{cases} \quad (15)$$

IV. EXPERIMENTAL RESULTS

In this paper, we use 2 groups of videos to test this algorithm.

Firstly, we use a group of video to test the robustness of SIFT. The result is shown as follows.

Figure 4 shows the invariant feature points extracted by SIFT. Figure 4(a) shows the feature points of normal target and Figure 4(b) shows the feature points and tracking result of the target which has been blocked. The result shows that SIFT algorithm has the ability to overcome the block of the target.

Secondly, we use a group of video to test prediction of Kalman filter. Figure 5 and Figure 6 compare the results of tracking a 54-frame traffic video. The results are shown as follows.

Figure 5 shows the tracking result without Kalman filter. After feature points detection by SIFT of 54 frames, the location of target has great offset. Figure 6 shows the tracking result by using Kalman filter. After 54 frames, the tracking result is still stable and accurate. In this article, we use similarity measurement (Bhattacharyya coefficient) to measure the stability of tracking. This result shows that the tracking algorithm with Kalman filter is more stable.

V. SUMMARIES

This paper presents a tracking algorithm based on SIFT and Kalman filter. Firstly, this paper explains the principle and stages of SIFT and we use kernel function to calculate the target location of tracking process. Secondly, this paper presents the principle of Kalman filter and uses it to predict the target location. Lastly, this paper shows the experimental results.

This paper uses 2 groups of videos to test this algorithm. Firstly, we test the performance of SIFT. We can see from the result that the SIFT algorithm can extract invariant feature points and can effectively overcome the block of target. Secondly, we test the tracking prediction results of Kalman filter. In this paper, we use Bhattacharyya coefficient to measure the similarity measurement of tracking. The result shows that the performance of tracking by using Kalman filter is much better than without using it.

REFERENCES

[1] Changfeng Niu, Dengfeng Chen, Yushu Liu. An Object Tracking Method Based on SIFT Feature and Particle Filter [J]. Robot, 2010, 32(2): 241~243.

[2] Fanglin Wang. Research on Important Issue of Robust Visual Tracking [D]. Shanghai Jiao Tong University, 2009: 25~27.

[3] Daniel Freedman, Pavel Kisilev. Fast Mean Shift by Compact Density Representation [J]. IEEE Trans. On Pattern Analysis and Machine Intelligence, 2009: 1818~1823.

[4] Qiu Dao-yin. Application of Frame Difference Methods in Real-time Moving Target Tracking [J]. Journal of North China Institute of Water Conservancy and Hydroelectric Power, 2009: 379-383.

[5] Yang Degui. Infrared Target Tracking based on Circular Projection and Particle Filter [J]. Signal Processing, 2009: 1702~1710.

[6] Yongbin Deng, Xinsheng Huang, Songjiang Feng. An Image Matching Method Combined with SIFT and LBP [J]. Computer Aided Design and Computer Graphics, 2010, 22(2): 286~288.

[7] Lowe D G. Object Recognition from Local Scale-Invariant Features[C]. 7th International conference on Computer Vision, 1999: 1150~1157.

[8] Lowe D G. Distinctive image features from scale-invariant feature points [J]. International Journal of Computer Vision, 2004, 60(2): 91~100.

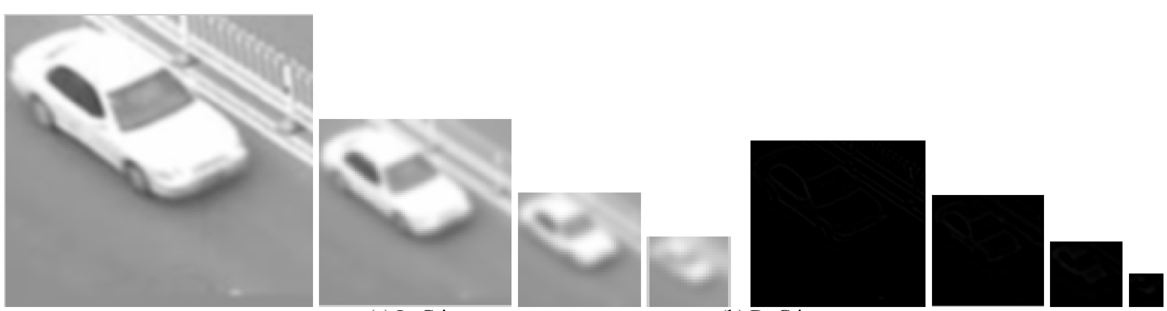
[9] Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales [J]. Journal of Applied Statics, 1991, 21(2):224~270.

[10] Zhongmin HuangFu, Xuemei Liu. Extraction of Rotational Surface Based on RANSAC [J]. Computer Engineering and Design, 2009,30(5):1295~1297.

[11] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington D C, 2004, 2: 506~513.

[12] Simon D. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. IET Control Theory and Applications, 2009:1303~1308.

[13] Julier S, Laviola J. On Kalman filtering with nonlinear equality constraints. IEEE Trans. Signal Process, 2007, 55(6):2774~2784.



(a) LoG images (b) DoG images

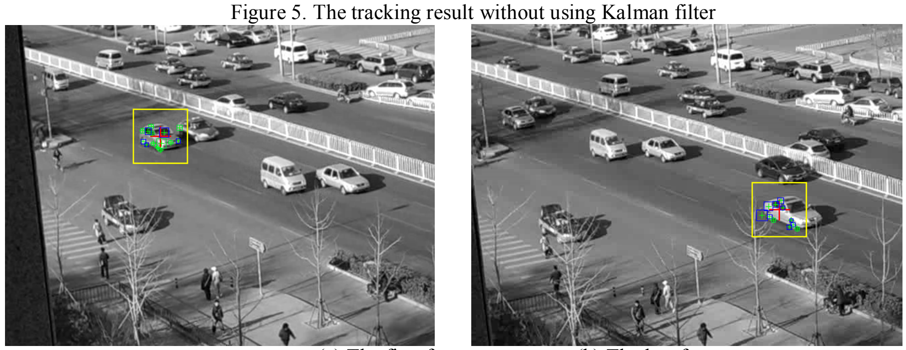
Figure 1. The LoG and DoG images of different scales

(a) The target without blocking (b) The target has been blocked

Figure 4 .The invariant feature points extracted by SIFT

(a) The first frame (b) The last frame

Figure 5. The tracking result without using Kalman filter



(a) The first frame (b) The last frame

Figure 6. The tracking result by using Kalman filter

Published by Atlantis Press, Paris, France.