Object Tracking Algorithm based on Bilateral Structure Tensor under Lie Group Manifold

Yinghong Xie[1,a*], Qiang Guo[2,b], and Chunhe Shi[1]

[1]School of Information Engineering, Shenyang University, Shenyang, China,

[2]Library, National Police University of China, Shenyang, 110854, China,

[a]xieyinghong@163.com; [b]royinchina@163.com

Keywords: Object tracking, Bilateral structure tensor, Particle filtering, Lie group, Riemann geometry

Abstract: Considering of object modeling problem with the complex background, the paper proposes a new object tracking algorithm based on bilateral structure tensor using particle filter under Lie group. Bilateral structure tensor is adopted to describe the appearance of the object, which can maintain the information of image edge better. At the same time, for the bilateral structure obeys Lie group structure, Riemann geometry which is much more accurate is applied to calculate the mean value and design particle filtering algorithm. The experimental results show that the proposed algorithm can achieve stable tracking for the visual and the variable targets under water.

1. Introduction

In the field of video tracking, the establishment of accurate and stable target apparent model and dynamic model for particle filter is the key to improve the performance of the algorithm. Because particle filter is nonlinear and does not limit the noise to Gaussian noise, using particle filter online learning method to construct tracking framework[1-6]has become a hot spot of various tracking algorithms.

In reference [1], by using the characteristic that affine deformation parameters are distributed on low dimensional manifolds, the state model of particle filter is established by affine lie group, and the tracking of deformed target is realized.

The paper[2,3] proposed a method for tracking the increasing covariance tensor on the Riemannian's flow. By using the affine transformation to describe the deformation process of the target, the particle filter frame is fused, and the background interference can be better removed. And real-time tracking can be realized.

The paper[4]proposes to construct a dual-particle filter by using the Riman flow geometry and the geodetic line method in Bayesian frame.At present, many algorithms [5] introduce structure tensor into the field of image tracking or matching.

These algorithms construct a positive definite matrix for certain properties of the image, and then pre-process Gaussian filtering. However, Gaussian filter is linear filter, which has the characteristics of isotropicity, which can easily filter out weak corner information. By combining the gray scale information and spatial information in the image, the bilateral filtering method makes the output image save the edge information well while filtering background noise.

Paper [6]applies an improved bilateral filtering method to the cost fusion algorithm of images. Paper [7]applies bilateral filtering theory to target detection and improves detection performance.

Considering the large non-planar attitude change of video target, the target tracking method based on the traditional Euclidean space basic framework often leads to the target tracking offset or failure. The apparent characteristic data of the deformed target exist on the smooth surface. Lie groups contain both differential manifolds and related theories of groups, where differential manifolds include topological manifolds and differential structures. It has become a feasible and effective method to realize target tracking based on the basic theory of lie group manifolds. In this

  DOI: 10.25236/meici.2019.061

paper, based on the characteristics of bilateral structure tensor, the bilateral structure tensor matrix is constructed as the apparent model of the target to be traced, the mean value of the lie algebra corresponding to the lie group space is used to calculate the mean value, and the particle filter is used to realize the stable tracking of the visual and infrared targets in the lie group space.

2. Bilateral Filtering Theory

The bilateral filtering method is based on the Gaussian filter, and is an improvement to the Gaussian filter. The Gaussian filter directly uses the Gaussian coefficient as the weight, and performs image filtering. Only the spatial information of the image is taken into account, and when the bilateral filtering method calculates the weight value, the gray level information of the image is integrated, and the average value of the pixel values adjacent to the space and the gray scale of each pixel of the image is replaced with the original value. By the non-linear combination of the gray level information and the spatial information in the image, the edge detail part of the image can be better preserved.

Let I be the original image, and I (x0, y0) denotes the gray value at the pixel (x0, y0). If the pixel point (x, y) is any point in the neighborhood *w* of (x0, y0), it is the same as the Gaussian filter. The spatial distance between the pixel point (x, y) and (x0, y0) can be expressed as follows:

$$d_I = \sqrt{(x\text{-}x_0)^2+(y-y_0)^2} \tag{1}$$

Let $d_{I'}$ represents the gradient distance between two points, defined as follows:

$$d_{I'} = \sqrt{(I_x-I_{x_0})^2+(I_y-I_{y_0})^2} \tag{2}$$

The weight value merges the spatial distance information and gradient distance information between two pixel points. for any (x, y) $\in$ *w* weight, the definition of $N_{\rho,\delta}$ is as follows:

$$N_{\rho,\delta}=\frac{1}{C_{\rho,\delta}}\exp(-\frac{d_I^2}{2\rho^2})\exp(-\frac{d_{I'}^2}{2\delta^2}) \tag{3}$$

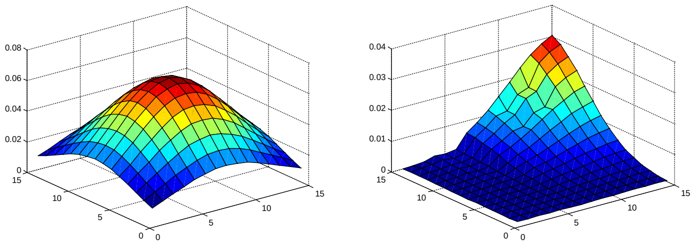
Which the parameters ρ , δ control the decay rate of space and gradient distance respectively.

$$C_{\rho,\delta}=\sum_{(x,y)\in w}\exp(-\frac{d_I^2}{2\rho^2})\exp(-\frac{d_{I'}^2}{2\delta^2}) \tag{4}$$

the bilateral filtering formula is as follows:

$$\tilde{I}(x_0,y_0) = \sum_{(x,y)\in w} N_{\rho,\delta}*I(x,y) \tag{5}$$

Fig.1 is a comparison of the weight distribution of Gaussian kernel function G_σ and bilateral tensor filter kernel function $N_{\rho,\delta}$ in a given image centered on a pixel point.



(a) Gauss Kernel function G_σ (b) bilateral filer function $N_{\rho,\delta}$

Figure 1 Comparison of the Weight Distribution of the Gaussian Kernel Function G_σ and the bilateral Filtering Function $N_{\rho,\delta}$

It can be seen that G_σ has the characteristics of isotropism and does not depend on the local features of the image. However, the $N_{\rho,\delta}$ has the characteristics of anisotropy, and the distribution of the point depends on the local characteristics of the point.

3. Establishment of Bilateral Structural Tensor Matrix

First, for the point (x,y) in a given window W, the weighted structure information of the point is defined as[4,5]:

$$T_{(x,y)} = \begin{pmatrix} N_{\rho,\delta} * I_x^2 & N_{\rho,\delta} * I_x I_y \\ N_{\rho,\delta} * I_x I_y & N_{\rho,\delta} * I_y^2 \end{pmatrix} \tag{6}$$

Among them, the definition of $N_{\rho,\delta}$ is the same as formula (5.2). The bilateral structure tensor for the point (x_0, y_0) is defined as the algebra sum of $T_{(x,y)}$ of all points in a search window centered on that point:

$$M_{(x_0,y_0)} = \sum_{(x,y)\in w} T_{(x,y)} \tag{7}$$

However, due to the $T_{(x,y)}$ disobedience of vector space, the structure tensor obtained directly by arithmetic addition may lead to inaccurate operation results. Because symmetric positive definite manifolds have lie group structure, their intrinsic mean values can be calculated by Riemann geometry.

For matrix lie groups, lie group exponential mapping is an algebra matrix index in the general sense. If given matrix x ∈ g, there is exponential mapping:

$$\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!} \tag{8}$$

When the matrix has no negative real eigenvalues, the matrix lie group exponential mapping has inverse mapping, which is called lie group logarithmic mapping.

$$\log(X) = \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i}(I - X)^i \tag{9}$$

For $T_{(x,y)}$ of any point in the search window, there is a lie group structure. Therefore, this paper defines the bilateral structure tensor of (x_0, y_0) based on the lie group structure as follows:

$$M'_{(x_0,y_0)} = \exp(m'_{(x_0,y_0)}) \quad M'_{(x_0,y_0)} \in G \tag{10}$$

$$\text{where,} \quad m'_{(x_0,y_0)} = \frac{1}{n}\sum_{(x,y)\in w} \log(T_{(x,y)}) \quad m'_{(x_0,y_0)} \in g \tag{11}$$

4. Particle Filter Algorithm

Let the initial probability density be as follows:

$$p\left(X_0 \mid Y_0\right) = p\left(X_0\right) \tag{12}$$

In which X_t is system state variables. Y_t is an observation variable of the system. The system is divided into two stages: a prediction stage and an update stage.

(1) the prediction stage: according to the state transition model of the system, the observation value that the system is likely to be obtained at the time t of the future is computed, The prior probability $p\left(X_t \mid Y_{1:t-1}\right)$ is predicted by the prior probability $p\left(X_{t-1} \mid Y_{1:t-1}\right)$:

$$p\left(X_t \mid Y_{1:t-1}\right) = \int p\left(X_t \mid X_{t-1}\right) p\left(X_{t-1} \mid Y_{1:t-1}\right) dX_{t-1} \tag{13}$$

(2) the updating stage: on the basis of obtaining the observed values at time *t*, according to the observation model of the system, the posterior probability is obtained:

$$p\left(X_t \mid Y_{1:t}\right) = \frac{p\left(Y_t \mid X_t\right) p\left(X_t \mid Y_{1:t-1}\right)}{p\left(Y_t \mid Y_{1:t-1}\right)} \tag{14}$$

Among them, $p\left(Y_t \mid X_t\right)$ is the likelihood probability, which represents the degree of similarity between the system state and the observed value after the transition from time t-1 to time t. Formula (3.2) is called prediction equation, which represents the probability model of target state transition, and formula (3.3) is called update equation. The process is described as follows: if the probability density of the target state at t - 1 is known as $p\left(X_t \mid Y_{1:t-1}\right)$, the posterior probability density of the target state $p\left(X_t \mid Y_{1:t-1}\right)$ at *t* is calculated recursively according to the probability model of state transition $p\left(S_t \mid S_{t-1}\right)$ and observation Y_t.

During the tracking process, the observed data are used to correct the latest predicted state at each moment. The probability of the sample is estimated by measuring the similarity between the observed data and the model. Given $p\left(I_t \mid S_t\right)$ is the observation of I_t under the state S_t, the state model can be constructed as follows:

$$p\left(I_t \mid S_t\right) \propto \exp(-\lambda \left\| d^2\left(M_*, M_{S_t}\right)\right\|^2) \tag{15}$$

Where, C_* is represented by the bilateral structure tensor of the image template, and C_{S_t} is represented by the sampling bilateral structure tensor at time *t*.

Define the weight of each particle as follows:

$$w_t^j = \exp(-\lambda \left\| d^2\left(M_*, M_{S_t}^j\right)\right\|^2) \tag{16}$$

The normalization weight is:

$$w_t^j = w_t^j / \sum_{j=1}^{M} w_t^j$$

5. Tracking Algorithm Design and Template Update Strategy

5.1 Tracking Algorithm Design

The target apparent model is estimated by using the bilateral structure tensor of the target region M_t .The complete tracking algorithm design is as follows:

input:. M_{t-1}

Output: the bilateral structure tensor M_t of the region at time *t*, *t* = 2, 3,. . .

Step 1. Initialize particle weight $w_1^j = 1/N$

step 2. Produces sample particles.

Step 3. Calculate the weight value $w_t^i, i = 1, 2, ..., N$

Step 4. The weighted mean value is calculated according to formula (10).

Step6. According to formula (9), output the target state M_t, let $t = t + 1$ turn to step 2.

5.2 Template Update Strategy

Because of the non-rigidity, the shape, scale and appearance of the deformed target may change at any time. In order to achieve stable target tracking, it is necessary to establish the corresponding template updating strategy to adapt to these changes. The algorithm updates the template every other *m* frames. Let {T1, T2,. , Tm} represents the bilateral structure tensor matrix of the nearest *m* frame tracking target. The value of *m* is determined by the practical application. The Riemann mean of the nearest m frame covariance matrix is calculated, which is used as a new tracking template. According to formula (10), the logarithmic Euclidean Riemann mean algorithm is used to calculate the mean value of bilateral structural tensor matrix. The algorithm is shown in Table 1.

Table 1 Riemann mean algorithm for bilateral structural tensor matrices

Algorithm 1

Input: sample matrix (m, n*n), m is the number of bilateral structural tensor matrices, and n is the dimension of covariance matrix.
Initialization: Sample mean Mean=zeros（n,n）;
for i=1:m
    Mean=sample(:,i);
    mean=reshape(Mean,n,n);
    meanlog=meanlog+logm(mean);
end
meanlog=1/m*meanlog;
meanlog=expm(meanlog);
output: meanlog

Then meanlog is the mean value of the bilateral structure tensor matrix, that is, the updated template. The time complexity of the algorithm is O (n).

6. Experimental Results and Analysis

In order to verify the effectiveness of the algorithm, the proposed algorithm (BPA) is compared with the tracking algorithm (CPA) using covariance matrix particle filter. Simulation experiments are carried out on the image sequence of rigid body and light transformation, and the tracking effects of the two algorithms are compared.

frame 1 frame 37 frame 196 frame 228

A. CPA algorithm tracking results

frame 1 frame 37 frame 196 frame 228

B. BPA algorithm tracking results

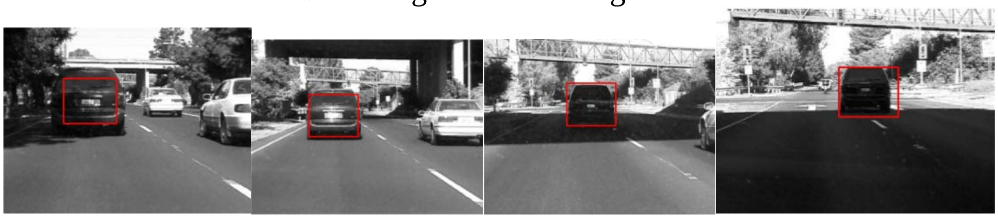
Figure 2. tracking results of video sequence 1

In the first group of experiments, two algorithms are used to track the rigid deformation motion in video sequence 1. Figure 2 (a) is the tracking effect of CPA algorithm, and figure 2 (b) is the tracking effect of BPA algorithm. In this case, in pixels, the neighborhood w=1*1 is taken. The template update frequency of both algorithms is 10 frames per time. The experimental results show that both algorithms can achieve stable tracking of rigid body motion, but for each frame effect, the tracking effect of BPA algorithm is more accurate.

In the second group of experiments, two algorithms are used to track the rigid body motion in video sequence 2. Figure 3 (a) is the tracking effect of CPA algorithm, and figure 3 (b) is the tracking effect of BPA algorithm. In this case, in pixels, the neighborhood w=1*1is taken. The template update frequency of both algorithms is 10 frames per time. It can be seen that the tracking algorithm proposed in this paper can achieve stable tracking of video targets with obvious light changes. The experimental results show BPA algorithms can achieve stable tracking of rigid body motion, while CPA algorithm tracks deviation.

frame 1 frame 131 frame 196 frame 229

A. CPA algorithm tracking results

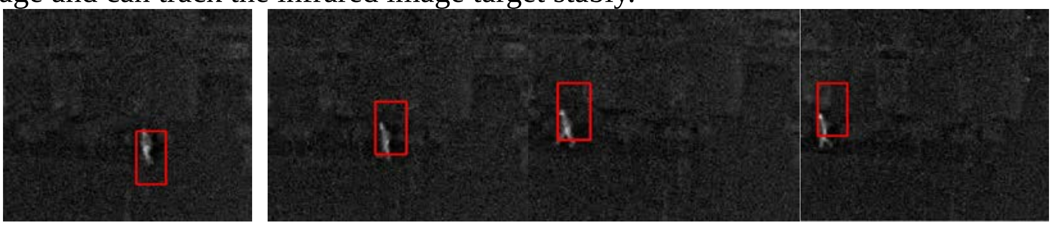


frame 1 frame 131 frame 196 frame 229

B. BPA algorithm tracking results

Figure 3. tracking results of video sequence 2

In the third group of experiments, we track a group of infrared video targets. Fig. 4 shows the tracking results of the two algorithms, when the neighborhood size is 2*2 (pixels) and the smoothing scale is 0.8. The template is updated every 10 frames. The experimental results show that the tracking results of CPA become worse due to the cumulative error. Because the bilateral structure tensor is used as the image feature data, the BPA algorithm enhances the edge information of the image and can track the infrared image target stably.



frame 218 frame 245 frame 308 frame 338

A. CPA algorithm tracking results

frame 218 frame 245 frame 308 frame 338

B. BPA algorithm tracking results

Figure 4. tracking results of video sequence 3

Table 2 lists the average coordinate difference between the traced center point and the real target center point in the x axis and y axis direction. It can be seen that the algorithm proposed in this paper has little deviation and the tracking results are more accurate.

Table 2 distance between tracking results and the center point of the real target

	First video sequence	**Second video sequence**	**Third video sequence**
CPA	8.13	16.23	13.43
BPA	3.11	5.66	2.19

In the fourth group of experiments, we test the effect of the tracking algorithm on underwater target tracking. As can be seen from figure 5, the tracking algorithm can track underwater targets very well.

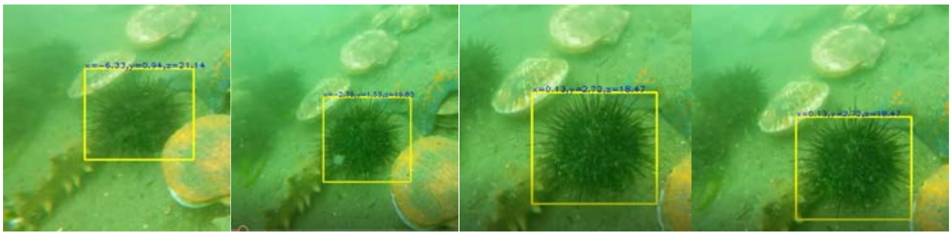


Figure 5. Underwater Target Tracking Results

7. Conclusions

In the complex background, the apparent model of the target is the key to obtain a stable and accurate tracking effect. In this paper, we apply the bilateral filtering theory to the target tracking field and the target tracking algorithm based on the tensor particle filtering of the bilateral structure. Because the bilateral filtering has the anisotropic characteristics, the filtered image can better preserve the edge information, the surface of the image area is characterized by the bilateral structure tensor, the particle filtering algorithm is implemented, and the target tracking is finished. The experimental results of various conditions show that the algorithm can realize stable tracking on the target of deformation and the target undergoing large-scale illumination change, and can accurately track the infrared target, and can accurately track the underwater target.

Acknowledgements

This work was supported by the National Science Foundation of China under Grant [No.61503274, No. 61603415, No.61703285], Science and Technology Program Project of Shenyang city under Grant [No.18013015] and the National Science Foundation of Liao Ning Province [No.20170540644]

References

[1] Kwon J H, Lee K M, Par, F C. Visual Tracking via Geometric Particle Filtering on the Affine Group with Optimal Importance Functions[A]. IEEE Conference on Computer Vision and Pattern Recognition[C].. Seoul, 2009: 991 - 998.

[2] Liu Y P, Li G W, Shi Z L. Covariance Tracking via Geometric Particle Filtering[J]. Journal on Advances in Signal Processing.2010,(0):1-9.

[3] Wu Y, CHENG J, WANG J Q. Real-time Probabilistic Covariance Tracking with Efficient Model Update [J]. IEEE Transactions on Image Processing, 2012, 21(5): 2824-2837.

[4] Khan Z H, Gu I Y H. Tracking Visual and Infrared Objects using Joint Riemannian Manifold Appearance and Affine Shape Modeling[A]. IEEE Interational Conference on Computer Vision Workshops[C]. Gothenburg, 2011:1847-1854.

[5] DONOSER M, KLUCKNER S, BISCHOF H. Object tracking by structure tensor analysis[C]. 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010:2600 -2603.

[6] LI L, YAN H. Cost Aggregation Strategy with Bilateral Filter Based on Multi-scale Nonlinear Structure Tensor[J]. Journal of Networks, 2011, 6(7): 958-965.

[7] ZHANG L, ZHANG L, ZHANG D. A multi-scale bilateral structure tensor based corner detector[C]. 9th Asian Conference on Computer Vision, Xi'an, China, 2009: 618-627.