A Novel Object Tracking Algorithm Combined with SIFT Feature Points

Xinfeng Yang[*1], Chunguang Guo[2]

[1.] *School of Computer and Information Engineering*, Nanyang Institute of Technology, Nanyang, Henan 473004, China
[2.] *Department of Information Engineering*, Henan College of Industry & Information Technology, Jiaozuo, Henan 454000, China

***Abstract* — In this paper a SIFT (Scale Invariant Feature Transform) is used for object tracking to overcome the deterioration in tracking accuracy, by using feature extraction and object position. Combined with the movement and space relations, an adaptive Mean Shift filter is proposed to obtain the exact location of object borders, and Bhattacharyya trust theory is used for the correlation coefficient to deal with the emergence and disappearance of the object. A linked list queue data association is used to record the relation between moving objects to improve the detection accuracy and reduce computing complexity.**

Keywords - motion tracking; particle filter; Mean Shift; SIFT feature points

I. INTRODUCTION

Moving target tracking means to find the tracked target in each frame image of the video image sequence. In order to effectively portray tracked target, we need to select the appropriate feature space F. Agreement with X representing an image coordinate space, called the configuration space; with tracking windows representing the region which may be the target in the image; with target model representing the tracked target. The core image tracking research is how to effectively use this information to modeling the tracked target, but it is a topic which has not yet been solved.

Tracking moving targets, not only can monitor its operation, but also provides a reference basis for movement act analysis understanding for vehicle target in traffic scene; it has been widely used in traffic junction surveillance. Tracking is equivalent to create the corresponding feature matching based on position, velocity, shape, texture, and color between successive image frames, the technical tools commonly used have particle filter algorithm and Dynamic Bayesian networks. Currently, for tracked target, there is pedestrian tracking or vehicle tracking; for track perspective, there is only perspective based on a single camera, multiple perspectives or comprehensive based on multi-camera; also include classify by spatial dimension, different tracking environment, the number of tracking and the motion status of the camera.

In recent years, researchers have proposed many algorithm strategies for the moving target tracking, but there are shortcomings of slow processing speed and low accuracy. To this, this paper proposes particle filter tracking algorithm combined with SIFT feature points to tracking the moving vehicles target in traffic scene, and improve the traffic movement target tracking accuracy and effect.

II. SIFT PARTICLE FILTER TRACKING

A. *Particle Filter*

Sequential importance sampling is a Monte Carlo sampling method [1], the key idea is using a group random sample with corresponding weights to say the posterior probability density function needed, and calculate the estimated value based on these samples and weights. Set the probability density function $p(x)\infty\pi(x)$, and $\pi(x)$ is a probability density function, and $\hat{x}^{(i)} \sim q(x), i=1,...,N$ is a sample obtained by a probability density function $q(\square)$ sampling, called sample particle, and $q(\square)$ became importance density function. The weighted approximation for probability density function $p(\square)$ can be expressed as [2]:

$$p(x) \approx \sum_{i=1}^{N} \lambda^{(i)} \delta(x - \hat{x}^{(i)}) \qquad (1)$$

Wherein, $\lambda^{(i)}\infty\frac{\pi(\hat{x}^{(i)})}{q(\hat{x}^{(i)})}$ is regular weight of the i-th particle, meet $\sum_{i=1}^{N}\lambda^{(i)} = 1$.

Suppose $X^k = \{x_1, x_2, ..., x_k\}$, $Z^k = \{z_1, z_2, ..., z_k\}$, $k \in N$ in nonlinear dynamic system, respectively said vector set all the states constituted until k time and vector set all measurements constituted, then the posterior probability density function of k time can be approximated as:

$$p(X^k \mid Z^k) \approx \sum_{i=1}^{N} \lambda^{(i)} \delta(X^k - X^k_{(i)}) \quad (2)$$

Wherein, $X^k_{(i)} = \left\{x_0^{(i)}, x_1^{(i)}, ..., x_k^{(i)}\right\}$. Regular weight may also be expressed as:

$$\lambda_k^{(i)} \propto \frac{p(\hat{x}_{0:k}^{(i)} \mid Z^k)}{q(\hat{x}_{0:k}^{(i)} \mid Z^k)} \quad (3)$$

If at the moment k, there has the approximate reconstruction of the sample to $p(X^{k-1} \mid Z^{k-1})$, and after the new measurement Z^k obtained, need a new set to approximate reconstruct $p(X^k \mid Z^k)$. Suppose importance density function can be decomposed into:

$$q(X^k \mid Z^k) = q(x_k \mid X^{k-1}, Z^k) q(X^{k-1} \mid Z^{k-1}) \quad (4)$$

We can use the existing sample $\hat{x}_{0:k-1}^{(i)} \sim q(X^{k-1} \mid Z^{k-1})$ and the new state sample $\hat{x}_{0:k}^{(i)} \sim q(x_k \mid X^{k-1}, Z^k)$ gets the sample $\hat{x}_{0:k}^{(i)} \sim q(X^k \mid Z^k)$.

If $q(x_k \mid X^{k-1}, Z^k) = q(x_k \mid X_{k-1}, Z_k)$, then according to the Bayesian formula can derive:

$$\begin{aligned} p(X^k \mid Z^k) &= \frac{p(z_k \mid X^k, Z^{k-1}) p(X^k \mid Z^{k-1})}{p(z_k \mid Z^{k-1})} \\ &= \frac{p(z_k \mid X^k, Z^{k-1}) p(x_k \mid X^{k-1}, Z^{k-1}) p(X^{k-1} \mid Z^{k-1})}{p(z_k \mid Z^{k-1})} \\ &= p(X^{k-1} \mid Z^{k-1}) \frac{p(z_k \mid x_k) p(x_k \mid x_{k-1})}{p(z_k \mid Z^{k-1})} \\ &\propto p(z_k \mid x_k) p(x_k \mid x_{k-1}) p(X^{k-1} \mid Z^{k-1}) \end{aligned} \quad (5)$$

Corresponding weights are available:

$$\lambda_k^{(i)} \propto \frac{p(\hat{x}_{0:k}^{(i)} \mid Z^k)}{q(\hat{x}_{0:k}^{(i)} \mid Z^k)} = \frac{p(z_k \mid \hat{x}_k^{(i)}) p(\hat{x}_k^{(i)} \mid \hat{x}_{k-1}^{(i)}) p(\hat{x}_{k-1}^{(i)} \mid Z^{k-1})}{q(\hat{x}_k^{(i)} \mid \hat{x}_{k-1}^{(i)}, z_k) q(\hat{x}_{0:k-1}^{(i)} \mid Z^{k-1})} \quad (6)$$

So the revised weight is:

$$\lambda_k^{(i)} \propto \lambda_{k-1}^{(i)} \frac{p(z_k \mid \hat{x}_k^{(i)}) p(\hat{x}_k^{(i)} \mid \hat{x}_{k-1}^{(i)})}{q(\hat{x}_k^{(i)} \mid \hat{x}_{k-1}^{(i)}, z_k)} \quad (7)$$

So the posterior filtering approximate density function is:

$$p(x_k \mid Z^k) \approx \sum_{i=1}^{N} \lambda_k^{(i)} \delta(x_k - \hat{x}_k^{(i)}) \quad (8)$$

Particle Filter has a defect problem that is degeneration phenomenon, after several iterations, weights of almost all particles are very small or even close to zero, so the large number of calculations is used to updating the particles, but the contribution for the approximation posterior probability $p(X^k \mid Z^k)$ [3], the effective sample size is generally defined as follows:

$$N_{eff} = \frac{N}{1 + \mathrm{var}(\overline{\lambda}_k^{(i)})} \quad (9)$$

Its estimate is often used in practice:

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} (\lambda_k^{(i)})^2} \quad (10)$$

Resampling method reserves high-weight particles through the elimination of low-weight particles, it can curb the degradation to a certain extent. But after resampling, sample particles are no longer independent, high-weight particles is copied multiple times, Low-weight particles gradually disappear, after several iterations, the particles are concentrated to a point, depletion occurs [4]. In this paper, through SIFT feature point added as part particles to the iterative process, through retaining the good characteristics and low weight of the particles, solve the particle degradation depletion impacts to a certain extent.

B. *SIFT algorithm*

Scale Invariant Feature Transform (SIFT) algorithm is a local feature description operator based on scale space which were proposed by Lowe based on the previous characteristics detection method summarized in 1999 [5]. The operator can maintain invariance for image rotation, zooming and affine transformation, which has the most robust in the same class operator [6]. For traffic image, SIFT algorithm has a very good matching characteristics and robustness in the target occlusion each other, scale change, the effects of noise and other aspects, so it can be used to track the movement of vehicles and other targets.

SIFT algorithm includes the detection of the key points and characterization generation, firstly detect feature point in scale space, describe the gray gradient distribution of inside the motion target area, and determine the location and dimensions of the feature point, by describing the direction of the point feature, achieve the scale and direction of independence. SIFT algorithm includes four steps: feature point detection, eliminating the unstable point, feature point description and feature matching.

1) Generate Scale Space

Gaussian convolution kernel is the only linear convolution kernel for space invariance [7]. Define a two-dimensional Gaussian filter function as formula (11), wherein, σ represents the variance of the Gaussian function, generally take 0.5.

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \tag{11}$$

Set an $N\times N$ image $I(x,y)$, images of different scale space can be obtained by the corresponding Gaussian convolution kernel:

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{12}$$

Wherein, σ is also called scale space factor, the size of the value is said how much an image be smoothed. Since the Difference of Gaussians (DoG) has a fast algorithm and good robustness, so SIFT use DoG to form multi-scale space:

$$\begin{aligned}D(x,y,\sigma)&=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)\\&=L(x,y,k\sigma)-L(x,y,\sigma)\end{aligned} \tag{13}$$

$$G(x,y,k\sigma)-G(x,y,\sigma)\approx(k-1)\sigma^2\nabla^2 G \tag{14}$$

Wherein, I is input image, D is DoG difference kernel, L is the convolution of I with gaussian kernel, k is the ratio of adjacent scales. Laplacian scale invariance can be approximately normalized in the DoG scale space, the local extremum feature points with good uniqueness and stability can be easily detected, Formula (14) instructions that DoG kernel is the linear approximation of scale invariant kernel $\sigma^2\nabla^2 G$.

2) Space extreme point detection

Compare each sample point with all its adjacent points to determine whether the point is both extreme point of two-dimensional image domain and the scale space domain. As shown in Fig. 1, the middle detection points need compare with 26 points, including 8 neighbors with same scale and all 18 points the adjacent scale corresponded to.

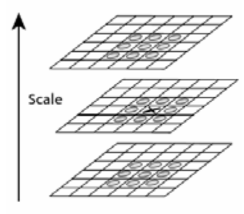


Figure 1. Choosing all extreme

3) Feature points direction distribution

According to the gradient direction distribution of feature point neighbor pixels, determine direction of feature points to ensure describe rotational invariance of the operator. For a scale space image L with measure of σ, calculate the gradient magnitude and direction of each sampling point $L(x,y)$:

$$\begin{aligned}X&=L(x+1,y)-L(x-1,y)\\Y&=L(x,y+1)-L(x,y-1)\\m(x,y)&=\sqrt{X^2+Y^2}\\\theta(x,y)&=\tan^{-1}(Y/X)\end{aligned} \tag{15}$$

Wherein, m is gradient mode of the pixel (x,y), θ is the gradient direction of the pixel (x,y).For each feature point, set a gradient histogram, which range is from 0 to 360 degrees, divide into 36 grids, the sampling range is the neighborhood window of the point as the center, use a histogram to statistics the gradient direction distribution of each pixel neighborhood, the direction of the histogram peak is the main direction of the key points, as shown in Fig. 2.

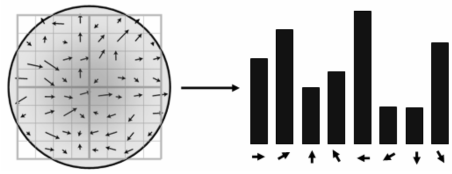


Figure 2. Histograms of oriented gradients

4) Generate feature descriptor

To ensure rotation invariance of descriptors, coordinate is rotated as the direction of the feature point. In order to enhance the robustness of matching, this paper is with the construction process for 128-dimensional SIFT descriptor as an example, for any one key point, in their scale space, take 16×16 window of the key point as the center, as shown in Fig. 3, the gradient histogram is obtained from gradient information of 8 directions on one small pieces of 4×4, one feature point composite by 16 seed points with 8 direction vector information, and then, gradient histograms of 8 direction for 4×4 sub-regions is sequentially ordered according to location, constitute 128-dimensional SIFT feature vector of 4×4×8.

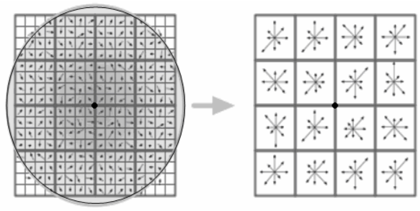


Figure 3. Feature vectors according to local gradients

Fig. 4 is a schematic for SIFT feature vectors. The starting point of the arrow indicates the position of the key

DOI 10.5013/IJSSST.a.17.49.42

ISSN: 1473-804x online, 1473-8031 print

point, the length indicates the scale of the point, the pointing indicates the direction of the SIFT feature vector.

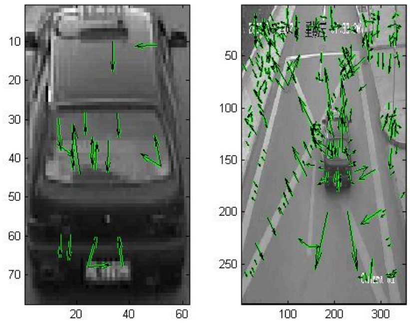


Figure 4. SIFT Key point feature vector

5) SIFT descriptor Matching

After characteristic descriptors of N_1 and N_2 feature points in two images were calculated, take a feature descriptor of 1×128 dimension in standard image, with N_2 feature descriptors of the matching image to calculate distance, then affinity matching can be obtained. In practical application, mainly use the nearest neighbor method to match the feature points. For the extracted descriptor, firstly need to calculate the Euclidean distance of the matching point descriptors and the feature point descriptors which needs to match and number is greater, secondly follow the actual need to select the appropriate matching criteria. Typically, when the ratio r of distance from the nearest point and next nearest point is less than a certain threshold value R, consider matching successfully [8].

C. SI-P algorithm

Because the SIFT feature point represents the movement robustness characteristics, so it is combined with particle filter algorithm, form Scale Invariant-Particle (SI-P) filtering algorithm. In the iterative process selection of particles combine SIFT feature points, as a particle update values into filtering process, so improve the accuracy of particle description of target, by keeping the independence of a certain number of particle with strong characteristic, alleviate the particle degradation phenomenon, thus improve the accuracy and performance of motion tracking.

Particle filter tracking combined SIFT feature points can effectively get the center of the target, but for the traffic vehicle tracking, often also need to know accurate boundary position of the vehicle, so this paper uses the mean shift algorithm to obtain the boundary of the target through tracking window, and by the Mean Shift analysis for each particle, after the Mean Shift iterations, all the particles are concentrated in the local area of the observation vectors, the particle will get big weight in the process of concentration, so as to further improve the tracking accuracy.

III. MEAN FILTER TRACKING

Mean shift algorithm (Mean Shift) is a non-parametric probability density gradient estimation algorithm [9], and its original meaning is the shift of the mean vector, then define a family kernel in the literature [10], so that the contribution of the offset (different distances samples with shifted points) to the mean shift vector is not the same. It also set a weighting factor, so that different sample points have different importance, thus expanding the application fields of mean shift method. Since then mean shift is widely applied to many other related areas, such as pattern classification, image segmentation and target tracking.

A. Mean Shift object tracking

Mean Shift tracking field is applied [11] is based on the external characteristics of the tracking algorithm, it is capable of moving targets better real-time tracking. Suppose $f(x)$ is the probability density function, $x_i (i = 1, 2, ..., n)$ is a d-dimensional space of n sampling points, then uses the kernel function approximation $f(x)$ can be expressed as:

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right) w(x_i)}{h^d \sum_{i=1}^{n} w(x_i)} \tag{16}$$

In the above formula weight of sample points xi represented as $w(x_i)(w(x_i) \geq 0)$, kernel function with $K(x)$, and meet $\int K(x)dx = 1$, then define $k(x)$ is the corresponding cross section function of $K(x)$, namely: $K(x) = k(\|x\|^2)$, at the same time, $g(x)(g(x) = -k'(x))$ and $G(x) = g(\|x\|^2)$ is defined as the corresponding kernel function, then the estimation of gradient $\nabla f(x)$ for $f(x)$ is may be expressed as:

$$\hat{\nabla} f(x) = \nabla \hat{f}(x) = \frac{2\sum_{i=1}^{n} (x - x_i) k'\left(\left\|\frac{x_i - x}{h}\right\|^2\right) w(x_i)}{h^{d+2} \sum_{i=1}^{n} w(x_i)} \tag{17}$$

Because $G(x) = g(\|x\|^2)$, $g(x) = -k'(x)$, so:

$$\hat{\nabla} f(x) = \frac{2\sum_{i=1}^{n}(x_i - x)G\left(\left\|\frac{x_i - x}{h}\right\|^2\right)w(x_i)}{h^{d+2}\sum_{i=1}^{n} w(x_i)}$$
$$= \frac{2}{h^2}\left[\frac{\sum_{i=1}^{n} G\left(\frac{x_i - x}{h}\right)w(x_i)}{h^{d+2}\sum_{i=1}^{n} w(x_i)}\right]\left[\frac{\sum_{i=1}^{n}(x_i - x)G\left(\left\|\frac{x_i - x}{h}\right\|^2\right)w(x_i)}{\sum_{i=1}^{n} G\left(\frac{x_i - x}{h}\right)w(x_i)}\right] \tag{18}$$

In formula (18), in the first bracket is the estimate for the probability density function $f(x)$ defined by $G(x)$ as kernel function, write for $\hat{f}_G(x)$, in the second bracket is Mean Shift vector, and $\hat{f}(x)$ defined in formula (16) is rewritten as $\hat{f}_K(x)$, so formula (16) can be rewritten as:

$$\hat{\nabla} f(x) = \nabla \hat{f}_K(x) = \frac{2}{h^2}\hat{f}_G(x) M_h(x) \tag{19}$$

So can get:

$$M_h(x) = \frac{1}{2}h^2 \frac{\nabla \hat{f}_K(x)}{\hat{f}_G(x)} \tag{20}$$

The formula (20) shows, use probability density of point x estimated by kernel function G can generate the normalized factor, and the mean shift vector $M_h(x)$ calculated by function G in point x and the gradient of the normalized probability density function $\hat{f}_K(x)$ estimated by kernel function K is proportional.

For real-time tracking moving target, set the center start position obtained from SI-P filter as $\hat{y}_0$, calculate distribution probability $p_u(\hat{y}_0)$ to get Bhattacharyya correlation coefficient: $\rho[\hat{p}_u(\hat{y}_0), \hat{q}_u] = \sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u}$, then calculate weight $\{w_i\}$, Among them, $w_i = \sum_{u=1}^{m}\delta[b(x_i - u)]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}}$, Then the new position of the moving target center can be estimated as: $\hat{y}_1 = \frac{\sum_{u=1}^{n_k} x_i w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}{\sum_{u=1}^{n_k} w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}$, at the same time calculate the new distribution probability and Bhattacharyya correlation coefficient: $\rho[\hat{p}_u(\hat{y}_1), \hat{q}_u] = \sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{y}_1)\hat{q}_u}$. When $\rho[\hat{p}_u(\hat{y}_1), \hat{q}_u] < \rho[\hat{p}_u(\hat{y}_0), \hat{q}_u]$, order $\hat{y}_1 = \frac{1}{2}(\hat{y}_1 + \hat{y}_0)$. If $\|\hat{y}_1 - \hat{y}\|_0 < \varepsilon$ (ε generally take a pixel value), then the algorithm is over, otherwise order $\hat{y}_0 = \hat{y}_1$, iterate again. For real time consideration, the largest number of iterations actual use may be taken in actual use, this paper is for 5 times.

In the tracking process, the tracking window (kernel width) of mean shift method generally remains unchanged, which cannot meet the campaign objectives (especially vehicles) scaling changes. Aim at the shortcomings of the kernel tracking algorithms exist, currently some improved algorithms are proposed, such as the kernel tracking algorithm based on the gradient direction histogram feature, tracking algorithm of multi-color distribution model and subspace model learning method and so on [12]. But the gradient histogram which ignores the gradient color information cannot better represent the entire target information even if chunking, and it only applies to the target tracking with slower speed and simple background; the modeling number of multi-model and model selection is only artificial determined currently, and there is no optimal strategy; due to subspace model learning has the lack of necessary monitoring mechanisms in model update process, underlying model drift make these algorithms at the risk of missing the target. So this paper uses DoG kernel tracking algorithm, reference multi-scale image analysis theory, adaptively change tracking kernel width through expectation maximization algorithm, then solve the above problems.

B. Improved mean filter method

1) Analysis of kernel Function

Consider d dimensions real Euclidean space R^d, the inner product of vector $x, y \in R^d$ is defined as $\langle x, y\rangle = x^T y = \sum_i x_i y_i$, the vector module can be exported by inner product product $\|x\| = \langle x, y\rangle^{1/2}$. For a given function K: $R^d \to R$, if there is an unary function k : $[0, \infty) \to R$ makes $K(x) = k(\|x\|^2)$ establishment, Wherein, $k(r)$ is non-negative, bounded, monotone decreasing, piecewise continuous and integral $\int_0^\infty k(r)dr$ bounded in the range $[0, \infty)$, the function $K(x)$ is called the kernel function, $k(r)$ is for the corresponding cross-sectional function. Because the function $k(r)$ is piecewise continuous, and the Lebesgue measure of the point set of non-differentiable is 0, Therefore, after supplementary defined on the point set of non-differentiable, the function $k(r)$ can be definitude everywhere in its domain, that is, $k^{'}(r)$ exists.

Bandwidth of kernel function is a very important parameter for Mean Shift algorithm. The graph shown in Fig. 5 is a graph showing the Gaussian kernel, bandwidth of kernel function can be set to 17.5. Search window usually choose this area. Ideally, if the size of the object has major changes in the tracking, the target window should also be a corresponding change. However, the Mean Shift algorithm itself has not mechanism of adapting the bandwidth of kernel function, so if the target is small, but bandwidth of kernel function is constant, a lot of background pixels will be included into the window, the background noise will dilute the characteristics distribution of the target model, thus affect the track results. Conversely, if the target becomes large, bandwidth is unchanged, candidate target model only counts part of the pixels of the target, the search window can only wander within the interior area of the target, which will affect the accurate positioning of the target boundary.

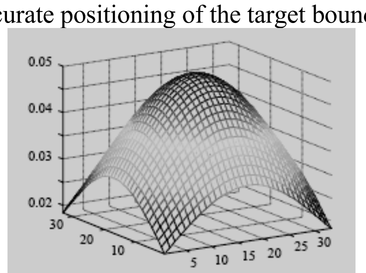


Figure 5. Gaussian kernel function

2) DoG multi-scale kernel function

Visual theory consider that [13], the number of elements in initial sketch visual proceeded can be used as a very good information metrics. Based on this idea, this paper has introduced information volume measure method of the multi-scale image into the moving target tracking; automatically choose the size of the tracking window according to changes of the information volume. Information volume measure method indicates that a kind of visual detail information of image in scale space is inversely proportional to the scale; similarly, the distance from the observer to the object is reflected thereby the information volume, corresponding information volume is greater, but smaller when close. Therefore, when track the target which scales change, the size of tracking window can be automatically expand or shrink according to the changes of information volume. Multi-scale image can be obtained by the convolution for the image and different scales Gaussian template, and then use the appropriate differential operator (such as DoG operator) to process images with different scales, robust features can be obtained when configuration space and scale space have reached local extreme. DoG function is selected as the kernel function to describe the target, define kernel function $K_N(x,\sigma) = c_1 G_{\sigma_1}(x) - c_2 G_{\sigma_2}(x)$, c_1 and c_2 are constants, set $K_N^{'}(x) = -g_N(x)$, then:

$$y = \frac{\sum_{i=1}^{n} x_i w_i g_N\left(\left\|\left(y_0 - x_i\right)/\sigma\right\|^2\right)}{\sum_{i=1}^{n} w_i g_N\left(\left\|\left(y_0 - x_i\right)/\sigma\right\|^2\right)} \tag{21}$$

Order $\sigma_1 = \sigma/\sqrt{1.6}$ and $\sigma_2 = \sigma \cdot \sqrt{1.6}$, combine with Expectation-Maximization (EM) Algorithm[14], after locating the target center, the maximum Bhattacharyya correlation coefficient $\rho : \rho\left[P(y), q\right] = \sum_{u=1}^{m}\sqrt{P_u(y) q_u}$ is obtained by changing the scale $\sigma\left(0.5\sigma \sim 1.5\sigma\right)$, and thus ultimately determine the tracking window scale. Fig. 6 shows the tracking effect of scale changing Shift Mean with a moving vehicle as an example.

Figure 6. Scale changing Mean Shift tracking (Frames: 10, 24, 39)

IV. EXPERIMENTS AND ANALYSIS

Fig. 7 and Fig. 8 compare the traditional particle filter tracking algorithm and SI-P filter tracking algorithm, the cross symbol represents the state of particles, the initial particle number is 300. The curve is the motion trajectory obtained after filtering. As can be seen in every frame, particle remains within the target area in SI-P algorithm, but in the traditional particle filter algorithm, because the target area changes, the track error is larger and thus offset moving objects lead to tracking failure.

Figure 7. Traditional particle filtering tracking (Frames: 10, 39, 60)

Figure 8. SI-P filtering tracking (Frames: 10, 39, 60)

The running time of the two were compared, the image frame with 768*576 resolution, selected 15 frame, CPU: 1.4GHz, memory: 512M, software environment is VC ++ 6.0. Fig. 9 shows that the average running time of SI-P is only more 0.15s than particle filter even adds SIFT feature point extraction process, it is within the monitoring system latency allowable range.

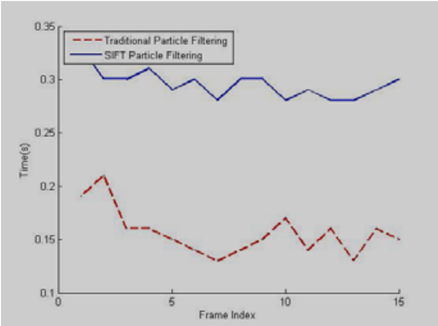


Figure 9. Comparison of runtime

Fig. 10(a) is the tracking results for Gaussian kernel function Mean Shift with classic fixed window width. We can see the tracking window remains in the movement area of the vehicle, but unable to adaptively determine the target border. (b) is the tracking results for Mean Shift with variable bandwidth the literature [15] presented. First, selected the target in frame i, got the initial tracking window T_i, and Mean Shift tracking performed in frame i to get T_{i+1}, then Mean Shift tracking performed in frame i to get $T_i^{'}$ using T_{i+1} as the initial tracking window, according the position difference of the centers of T_i and $T_i^{'}$ to move, and change its size: $r = r \cdot \varepsilon$. Finally, add the target kernel histogram dot product, and then select the maximum, ultimately determine T_{i+1}. But when the object and background colors are similar, Mean Shift has a local maximum in a smaller area, window scale shrink rapidly, thus cannot be accurately tracked. (c) is the DoG kernel function multi-scale tracking results, from which we can see along with changes in the size of moving objects, the window scale can be changed adaptively to accurately obtain the target center position and the surrounding boundary.

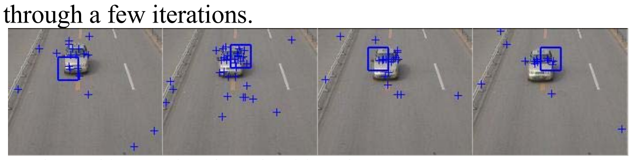
(a) Fixed window width Gaussian kernel Mean Shift with tracking

(b) The tracking results of Lecture [15]

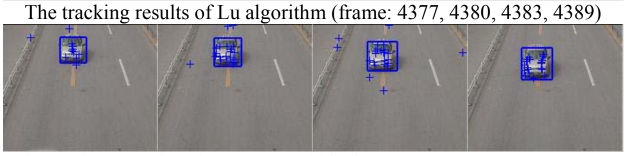
(c) DoG kernel function adaptive window scale Mean Shift tracking

Figure 10. Comparison of different Mean Shift tracking methods

In the aspect of combing with Particle Filter and Mean Shift, compared SI-P-mean-shift algorithm and Lu algorithm, test sequence is the actual traffic monitoring sequence, Lu algorithm first use Mean-shift algorithm to track, obtained the similar coefficients by calculating the detected target position, compared with the predefined threshold coefficient, used different particle filter strategy, it has a certain anti-blocking, but the algorithm does not consider the case of the target size changes evidently, the threshold selected to the particle sampling need to manually preset, it is not conducive to practical use. As shown in Fig. 11, cross symbol represents the particle; the rectangular is the target area boundary. When the target size changes, tracking window of Lu algorithm cannot adaptively change the scale, wandering around the target, the distribution of particles is relatively sparse, the tracking accuracy is low.SI-P-mean-shift algorithm detailed display the tracking performance the window scale changes when the target from small to large and from large to small, and particles focused on inside the moving target. The comparison of both runtime is shown in Fig. 12, drawn 10 frames, the initial number of particles is 150, image frames resolution is 352*288, CPU: 1.4GHz, Memory: 512M, software environment is VC ++ 6.0. The overall runtime of the tracking algorithm this paper proposed is even less than Lu algorithm, this due to base on SIFT feature matching can quickly identify the characteristics of the target area, the follow-up filter tracking achieved optimal through a few iterations.



The tracking results of Lu algorithm (frame: 4377, 4380, 4383, 4389)



The tracking results of SI-P-mean-shift (frame: 4360, 4369, 4377, 4389)

Figure 11. Comparison of tracking accuracy

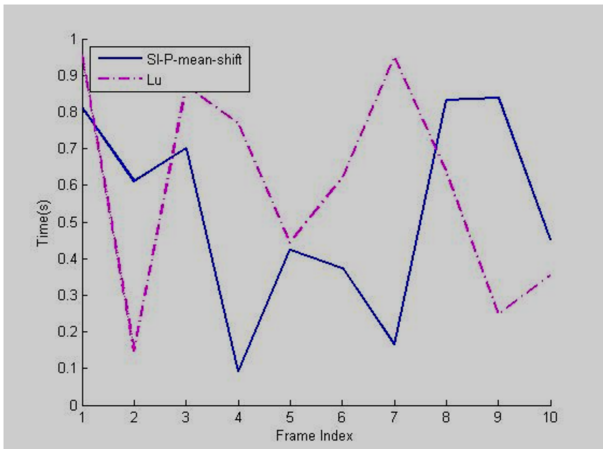
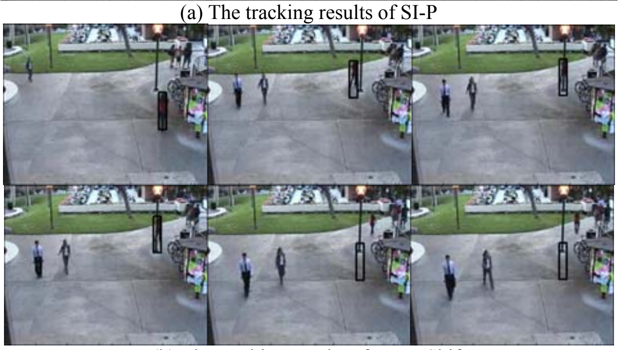


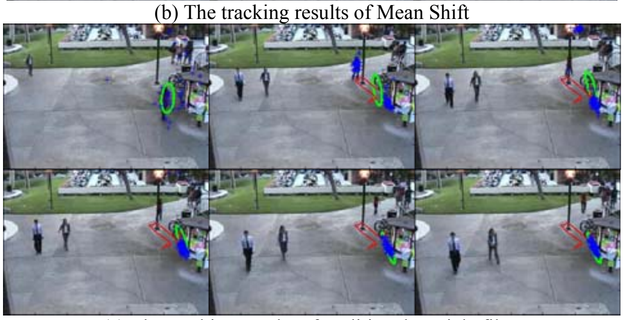
Figure 12. Comparison of runtime

In addition, in order to highlight the robustness of SI-P filter tracking in terms of anti-occlusion, this paper uses a pedestrian surveillance video with more complex background, initial detection region as a template, for a monitored pedestrian, not only the scale has changes, but also almost completely obscured by poles in motion, and then reappear. As can be seen from Fig. 13, the tracking region of SI-P filter algorithm has remained inside the moving target, Mean Shift algorithm lost accurate track on the target because mismatch when the target is obscured and appears again, due to target is smaller and scale dramatically changes, the traditional particle filter algorithm has lost accurate track on the target before the occlusion.

(a) The tracking results of SI-P



(b) The tracking results of Mean Shift



(c) The tracking results of traditional particle filter

Figure 13. Comparison of occlusion resistance (Frames: 49, 171, 196, 201, 239, 247)

V. CONCLUSIONS

Based on motion detection, aimed at that the shape and appearance (color distribution) of the target is usually very complex, and the target will be deformed, partially or completely obscured in the tracking process, as well as the scale size of target in image will changing with its distance from the camera changes, study the moving target tracking method in video surveillance image sequences through the combination of SIFT feature points and particle filter, this paper give the corresponding algorithm processes, which include feature extraction, target location, status estimation and the implementation of tracking status update. While using background subtraction results, remove the background, reconstruct foreground object image, combine with relationships of the target moving time and spatial, use window adaptive mean filtering method, obtain the exact location of the object boundary, and thereby improve the accuracy and efficiency of the moving target tracking.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was financially supported by the key scientific research project of Henan Province (15A520090) and the key science and technology project of Henan Province (122102210563 and 132102210215).

REFERENCES

[1] Satoh Y., Okatani T., Deguchi K.. A color-based tracking by Kalman particle filter, In: Proceedings of International Conference on Pattern Recognition, Cambridge, UK, 2004, 3: 502-505.

[2] Gordon N. J., Salmond D. J., Smith A. F. M.. Novel approach to nonlinear/ non-Gaussian Bayesian state estimation, In: IEE Proceedings on Radar and Signal Processing, 1993. 140(2): 107-113.

[3] Arulampalam M. S., Maskell S., Gordon N., et al.. A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking, IEEE Transactions on Signal Processing, 2002, 50(2): 174-188.

[4] Doucet A.. On sequential simulation-based methods for Bayesian filtering, University of Cambridge: Technical Report, CUED/F-INFENG/TR.310, 1998.

[5] Lowe D. G.. Distinctive image features International Journal of Computer Vision, 2004 from scale-invariant keypoints,60(2): 91-110.

[6] Mikolajczyk K., Schmid C.. A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.

[7] Lindeberg T.. Scale-space theory: a basic tool for analyzing structures at different scales, Journal of Applied Statistics, 1994, 21(2): 224-270.

[8] Beis J. S., Lowe D. G.. Shape indexing using approximate nearest-neighbor search in high-dimensional spaces, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 1997, 1000-1006.

[9] Fukunaga K., Hostetler L.. The estimation of the gradient of a density function with applications in pattern recognition, IEEE Transactions on Information Theory, 1975, 21(1): 32-40.

[10] Cheng Y.. Mean Shift, mode seeking, and clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790-799.

[11] Yang C. J., Duraiswami R., Davis L.. Efficient mean-shift tracking via a new similarity measure, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, California, USA, 2005, 1: 176-183.

[12] Ross D., Lim J., Yang M. H.. Adaptive probabilistic visual tracking with incremental subspace update, In: Proceedings of European Conference on Computer Vision, Prague, 2004, 215-227.

[13] Milan S., Vaclav H., Roger B.. Image processing, analysis, and machine vision (second edition), London: Brooks/Cole Publishing, 1999.

[14] Dempster A. P., Laird N.. Complete data via the EM Series B (Methodological), and Rubin D. B., Maximum likelihood from in-algorithm, Journal of the Royal Statistical Society, 1977, 1(39): 1-38.

[15] Collins R.. Mean-shift blob tracking through scale space, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Wisconsin, USA, 2003, 2: 234-240.