

A review of visual moving target tracking

Zheng Pan¹ · Shuai Liu¹ · Weina Fu^{1,2}

Received: 5 April 2016 / Revised: 17 May 2016 / Accepted: 26 May 2016 /
Published online: 16 June 2016
© Springer Science+Business Media New York 2016

Abstract Recently, computer vision and multimedia understanding become important research domains in computer science. Meanwhile, visual tracking of moving target, one of most important application in computer vision, becomes a highlight today. So, this paper reviews research and technology in this domain. First, background and application of visual tracking is introduced. Then, visual tracking methods are classified by different thinking and technologies. Their positiveness, negativeness and improvement are analyzed deeply. Finally, difficulty in this domain is summarized and future prospect of related fields is presented.

Keywords Computer vision · Visual tracking · Meanshift · Particle filter · Active Contour · Scale-invariant

1 Background

Vision is the main form to achieve outside information for being. According to the scientific investigation, it is found that more than 80% of total information is achieved by vision. Today, computer vision, pattern recognition, computer graphics, artificial intelligence etc., are developed swiftly. So, it is time to study computer vision that makes computers achieve information as human. Also, computer vision contains all domains that makes computers recognize, analyze, understand things as human.

Based on object detection, visual tracking is presented which is applied to detect, extract, recognize and track moving objects in image sequence [40, 93]. With the extraction and tracking, parameters (especially moving parameters, such as position, speed, acceleration, etc.) are achieved. Next, these parameters are processed and analyzed to understand behavior of moving target. Today, visual tracking is expected to apply in multi-domains such as radar

✉ Shuai Liu
cs_liushuai@imu.edu.cn

¹ College of Computer Science, Inner Mongolia University, Hohhot 130012, China

² College of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 130012, China

guidance, aerospace, medical diagnosis, intelligent robot, video compression, human computer interaction, education and entertainment by fused technologies of image processing, pattern recognition, artificial intelligence, automatic control and so on.

For visual tracking, its application mainly summarized as follows.

1) **Intelligent video surveillance (IVS)** [57]

IVS is an automatic survey system which supervises and inspects behavior of target real-time in specific area, such as safety department, private residence and public occasion. It prevents the occurrence of criminal acts by recognized suspicious behaviors. In IVS, main functions are target detecting, recognition and tracking in video. Especially, behavior recognition is most difficulty in IVS.

2) **Video compression/encoding** [2]

In video compression based on objects (H.265, HEVC etc.), encoding computation mainly spends on two aspects which are block matching and filtering. So it is necessary to apply target tracking technology into encoding method. In this way, object in image sequence can be track easily and fast. This enhances matching speed and accuracy. Further, it raises encoding effectiveness, peak signal to noise ratio (PSNR), and reduces bit error rate (BER).

3) **Intelligent traffic system (ITS)** [19, 64, 73]

Target detection and visual tracking are widely applied in modern ITS. Target detection is applied to position vehicle in surveillance video or monitor. Then, visual tracking is applied to track detected vehicle to compute vehicle flowrate, condition and abnormal behavior. A real-time computer vision system for vehicle tracking and traffic surveillance is useful. A novel method for tracking and counting pedestrians real-time can compute density of pedestrians. Detection and tracking method of pedestrians in crossroad can help vehicle move safely on crossroad.

4) **Other application** [34, 80, 94]

Visual tracking can be applied in many other domains. On one hand, it is applied as main technology in military, such as missile guidance, radar detection, unmanned aerial vehicle (UAV) flight control, individual combat system. On the other hand, it is also applied in modern medicine. In fact, medical image technology becomes a novel technology in medical clinical diagnostics and therapeutics. This can be applied to improve health status and exercise habits by a healthcare PC. Also, it can be applied to track trajectory of protein stress granule in cells and analyze dynamical features of cell structure.

Development of visual tracking was based on development of computer vision. Before 1980s, processing with static image was main research in image processing [75]. This is mainly due to the limited development of computer vision technology. So, research and analysis of dynamical image sequence was similar to apply technology of static image research. Research on dynamical image sequence became a highlight after optical flow method was presented [39, 63]. The study upsurge of optical flow continued to 1990s [8]. Recently, there were papers that reviewed L-K method.

Since 1980s, there were lots of visual tracking method appeared, in which some of them are classic methods today. This review introduces and analyzes these methods. In order to introduce and compare these methods well enough, this review classifies them to several classes.

1) **Classification of visual tracking problem**

Based on different application, there is a classification of visual tracking problem.

i) **Classification by scene**

There are two scenes in real application of visual tracking. One is tracking in single scene, another is tracking in multi-scenes. Target tracking in single scene tracks target in a static scene. In other words, it tracks one designated target in one video (scene), which is taken by one static camera. Target tracking in multi-scenes tracks multi-targets in one monitoring networks constructed by multi-cameras. This method establishes unique identity for every targets and tracks every targets continuously in this networks by fused all monitors [41].

ii) **Classification by number of moving targets**

According to number of moving targets, visual tracking can be classified to two classes, which are single target tracking and multi-targets tracking. Single target tracking is relatively easier than multi-target tracking. But they are same essentially. Both of them need to detect and extract tracking target from scene completely and accurately. The detection and extraction are influenced by multiple factors, such as transformation of background, illumination and noise. Multi-target tracking must consider occlusion, merger and separation of these targets. But when consider one target as the target and other targets as background, multi-targets tracking is same to single target tracking.

iii) **Classification by movement of camera**

In this classification, visual tracking is classified to tracking with static camera and moving camera. In the case of visual tracking with static camera, the background is fixed. This means foreground and background can be relatively separated easily. There are many methods can separate foreground and background, such as Gaussian filter and differential method. In the case of visual tracking with moving camera, it is difficult to separate foreground and background because both of them are moving. A common practice is to restore every image (frames in video) by applied a sensor in camera to measure speed of camera.

2) **Classification of visual tracking method**

Based on different extracting features, there is a classification of visual tracking method.

i) **Tracking method based on target geometrical similarity**

Target tracking method with geometrical similarity must apply an image block which is called target template. This method assumes that target can be expressed by simple geometrical

form which is stored as target template. Target template can be extracted by target detection method or artificial designation. Then, similarity is computed between target form and template. A threshold is applied to determine if there is a target in tracking scene. This kind of method is widely applied to track target in image sequence, such as meanshift. This kind of method has accurate tracking result because it usually applies global information of target, such as color and texture. But the tracking accurate dims when there is deformation or large occlusion of moving target.

ii) **Tracking method based on local feature**

Target tracking method with local feature extracts local features of target, then recognizes and tracks target in image sequence by fused these features. Common local features contain edge which is extracted by Canny operator, angular point which is extracted by Susan operator and so on. Tracking with local features can be applied more positive when there is partly occlusion of target. This is because these local features are extracted all over the whole target, which means that recognition efficiency is less influenced than geometrical features when there is partly occlusion of target. The difficulty of this kind of method is to recognize and track when target changes its feature by movement. The local features can be changed when there is zooming or rotation of target.

iii) **Tracking method based on contour profile**

Target tracking method with contour profile is a newer research than previous two kinds of methods. This kind of method needs a general position of target, which is provided by human. Then recursion by differential equation is solved to converge contour profile to local minimum value by energy function method. An energy function is usually constructed according to feature of image and smoothness of contour, such as edge and curvature. Snake algorithm is an example of this kind of method. It adopts combined action of image, internal and external constraining force to control movement of contour, in which image force pushes snake curve to edge of the image, internal force constrains local smoothness of contour, external force pushes snake curve to local minimum position of expectation (can be provided by human).

iv) **Tracking method based on forecast**

Target tracking method with forecast transforms a target tracking problem to a mode estimation problem. Mode of target contains all movement features concerned. In this kind of method, key point of target tracking is to deduce and express posterior probability of target mode accurately effectively by exist observed data. So forecast tracking is usually based on filtering method. Two well-known filtering methods are Kalman filtering and particle filtering. Recent years, there are newer tracking methods based on fusion of one filtering method and some other kind of methods.

So in this review, we put forward and discuss mainstream algorithm of target tracking today which includes meanshift, particle filter, active contour and SIFT method. In second section, we discuss meanshift method and its improvement. Then we study particle filter method and its improvement in third section. In fourth and fifth sections, active contour method and SIFT method are researched. Finally, we conclude these methods and present our mind.

2 Meanshift and its extension

2.1 Meanshift algorithm

Meanshift algorithm, known as shift of mean vector, was first provided by Fukunaga et al. [33] in year 1975. They presented this algorithm from estimation of gradient function in probability density. Later, Cheng [14] extended meanshift algorithm into domain of computer vision in year 1995. Then, Comaniciu et al. [23] applied meanshift algorithm to solve target tracking problem.

Meanshift is such a pattern recognition algorithm that it provides a fast search method to computer local extremum in distribution of data probability density. Meanshift algorithm computes similarity of object model and candidate image data. It moves searching window to search object by applied climbing algorithm of similarity gradient direction. It reaches target tracking when meanshift algorithm converges to local extremum.

Meanshift changes by extension of its theory. Earlier, it was a mathematical definition, which expressed a mean vector of shift. Today, meanshift denotes an iterative process. This iterative process first computes shift of current data mean, and moves current data along with direction of its shift mean. Then, applying moved data as current data, meanshift continues iteratively.

A basic meanshift vector is defined in Eq. 1, where S_k is a sphere in d-dimensional space with radius h, see Eq. 2, x is a chosen point in S_k , x_i is a sample point in S_k , k denotes there are k points in S_k in total n points x_i ($i = 1 \dots n$).

$$M_h = \frac{1}{k} \sum_{x_i \in S_k} (x_i - x) \tag{1}$$

$$S_h(x) = \left\{ y : (y - x_i)^T (y - x_i) < h^2 \right\} \tag{2}$$

Then, to insert a kernel function $k()$ into meanshift, we have Eq. 3.

$$\tilde{f}_{h,k}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) \tag{3}$$

To reach its derivation, we have meanshift equation in Eq. 4.

$$\nabla \tilde{f}_{h,k}(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x - x_i) k' \left(\left\| \frac{x - x_i}{h} \right\| \right) \tag{4}$$

Then, to apply $g(x) = -k'(x)$ and Gaussian kernel function, Eq. 4 transforms to Eq. 5 where meanshift vector $m(x)$ of x is presented in Eq. 6.

$$\tilde{\nabla} f_{h,k}(x) = \tilde{f}_{h,G}(x) \cdot \frac{2c_{k,d}}{h^2 c_{g,d}} m_{h,G}(x) \tag{5}$$

$$m_{h,G}(x) = \frac{\sum_{i=1}^n x_i g\left(\left\| \frac{x - x_i}{h} \right\|\right)}{\sum_{i=1}^n g\left(\left\| \frac{x - x_i}{h} \right\|\right)} - x \tag{6}$$

So, we have Eq. 7 to computer convergent direction of meanshift vector when let $m_{h,G}(x)=0$.

$$x = \frac{\sum_{i=1}^n x_i g\left(\left|\frac{x-x_i}{h}\right|\right)}{\sum_{i=1}^n g\left(\left|\frac{x-x_i}{h}\right|\right)} \quad (7)$$

Detail of meanshift is presented in following Alg.1.

Alg.1 Meanshift Algorithm

Step 1.

Let x is the chosen point in C^d space; d -dimensional sphere S is constructed.

Step 2.

For all x_i in S do

Alg.1 Meanshift Algorithm

Step 1.

Let x is the chosen point in C^d space; d -dimensional sphere S is constructed.

Step 2.

For all x_i in S do

If $m_{h,G}(x) > \varepsilon$ Then // $m_{h,G}(x)$ is computed by applied Eq.6

{

Apply Eq.7 to compute x ;

Goto Step 1.

}

Alg.1 finished.

Alg.1 finished.

2.2 Algorithm analysis and improvement

Meanshift algorithm has many advantages. First, computation of the algorithm is relatively not large, so that it can track target real time and efficiently with known target area. Second, acting as a density estimation algorithm with no parameter, meanshift can easily become a module and integrate other algorithms to increase effectiveness of target tracking. Third, meanshift constructs its model by applied kernel function. This means meanshift algorithm has robustness that is not sensitive to edge occlusion, target rotation, deformation and background change.

Meanshift algorithm has some defects in its tracking process. First, feature representation model of target lacks essential update ability. Tracking process can not adapt searching window with different sizes. So tracking process has high failure ratio when there is transformation of scale or form in tracking target. Second, color histogram is a weak feature as the only description of target. Indeed, this feature does not work well when color distributions are similar between background and target. Third, meanshift algorithm relies on relation between



Fig. 1 An instance of meanshift algorithm in target tracking

current and next frames. So tracking result of target with fast velocity usually converges to an object in background, which color distribution of the object is similar to tracking target. This is because there is no overlapped area between two adjacent frames when target moves quickly. So tracking method loses efficacy because it can not find target in the later frame when it searches area near target’s position in earlier frame.

Figure 1 shows one instance for meanshift algorithm. Videos we applied are from Ref [81].

In Fig. 1, we find that meanshift algorithm successful tracks moving target with red rectangle in Fig. 1a, but it fails in Fig. 1b. The failure occurs because another target occludes the tracking target. In this case, meanshift algorithm can not recognize tracking target and occurs error.

Today, there are many improved methods of meanshift. Comanicui presented Camshift (Continuously Adaptive Meanshift) method based on meanshift [21]. They applied probability density estimation of kernel function in visual target tracking. Camshift method applied histogram to describe target, adopted Bhattacharya parameter to measure distance between model and current object, and tracked target by meanshift method. It has higher tracking rate and effectiveness by used hill climbing method by gradient instead of global search algorithm.

To consider an image as a matrix, probability density of pixel x is ruled when to draw a circle with original point x and radius h , and to define two pattern rules for all x_i in the circle.

- 1) Probability density is larger when color is more similar between pixel x and x_i .
- 2) Probability density is larger when position is nearer between pixel x and x_i .

So probability density of pixel x can be constructed in following Eq. 8 where $C/h_s^2h_r^2$ denotes density of unit measure, $K(x)$ is a kernel function, s denotes space distance between x and x_i , r denotes color difference between x and x_i .

$$K_{h_s, h_r}(x) = \frac{C}{h_s^2 h_r^2} K\left(\left\|\frac{x^s - x_i^s}{h_s}\right\|\right) K\left(\left\|\frac{x^r - x_i^r}{h_r}\right\|\right) \tag{8}$$

Then, Alg.2 is presented for detail Camshift.

Alg.2 Continuously Adaptive Meanshift Algorithm

Step 1. Construct feature representation model of target.

Begin

Apply color distribution of moving target in presented area as feature.

Compute color distributional status by applied distance between pixels and barycenter of the target.

Give weights for every pixel in moving target by color distribution and distance. The distance is inversely proportional to the weight.

End

Step 2. Construct description model for potential area that may contains the target.

Specify searching areas around target of previous result in advance. Extract feature in the detection area and generate feature representation model for potential areas.

Step 3. Compute similarity between feature representation models of target and potential area.

Usually, meanshift applies Bhattacharyya coefficient to measure similarity between color histograms of target and potential area. Bhattacharyya coefficient is computed as Eq. 9 where a , b denote two histograms, n denotes number of color in each histogram. $\sum a_i$ and $\sum b_i$ denote sum of pixels with same color i . Also, other methods of similarity measure are also applied in the comparison [24].

$$B(a, b) = \sum_{i=1}^n \sqrt{\sum a_i \cdot \sum b_i} \quad (9)$$

Step 4. Extract moving target.

While length of vector is larger than given threshold

Choose area with largest Bhattacharyya coefficient as the target.

Goto Step 1.

Alg.2 finished.

Camshift based method applies meanshift method in continuous frames. It tracks target by applied color distribution graph and k -order moment. It has higher tracking robustness in indoor environment. But Camshift confuses target and other objects when color of object is similar to target. This will cause expansion of tracking window.

Figure 2 shows one instance for Camshift algorithm.

In Fig. 2, we find that camshift algorithm successful tracks moving target with red rectangle in Fig. 2a, but it fails in Fig. 2b. The failure occurs because another target occludes the tracking target and confuses Camshift method.

Later, Collins studied blob tracking in scale space of meanshift [20]. Peng et al. presented an automatic selecting method of window in kernel function [74]. Li et al. studied convergent threshold and environment of meanshift [60]. Birchfield et al. improved meanshift algorithm by spatiograms versus histograms [9]. Collins et al. first lead online learning method to meanshift algorithm [21]. Jeyakar et al. improved meanshift by background-weighted local kernels [45]. Their work effective improve meanshift when the target is partially occluded. Leichter et al. applied integral histogram to conquer model migration of meanshift [56]. Vojir et al. improved adaptation of scale change for meanshift by hellinger distance [82]. This method was compared with TLD (Tracking Learning Detection) and shows better robustness [48].

Recently, Li et al. combines spatial context information to meanshift [59]. This algorithm is robustness when target moves fast. Yang et al. applied HSI color feature histogram into meanshift

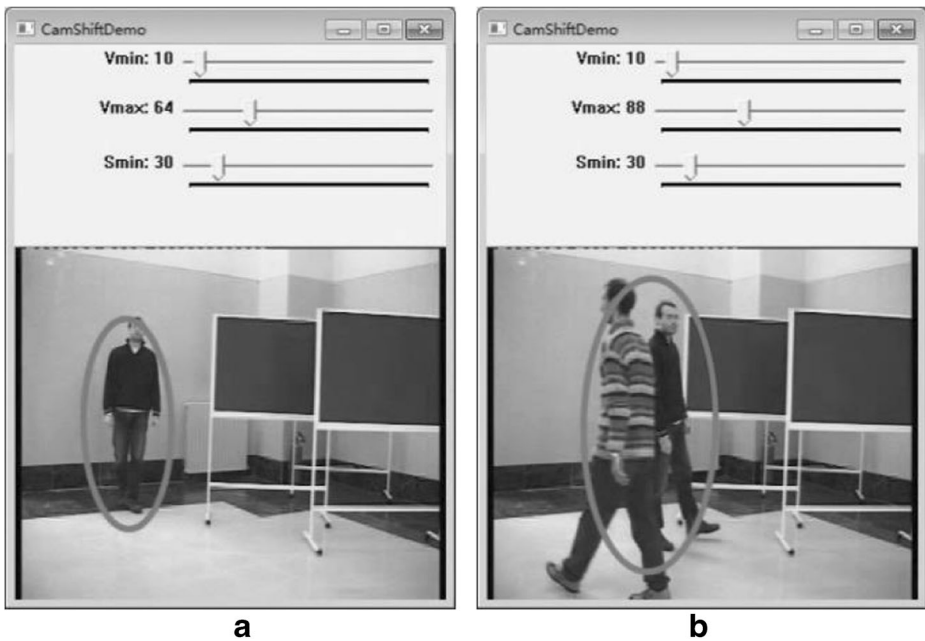


Fig. 2 An instance of camshift algorithm in target tracking

[91]. They classify superpixels and showed good performance for moving target tracking. Jia et al. improves kernel function of meanshift in their method, which shows higher robustness and more applicable scenarios [47]. Jatoh et al. analyzes performance of many tracking methods, and provide a novel method combined α - β filter, Kalman filter and meanshift for target tracking in frame sequences [44]. Guo et al. improves meanshift to non-rigid target tracking method [35]. Their algorithm further promotes accuracy and robustness of tracking.

3 Particle filter and its extension

3.1 Particle filter algorithm

Particle filter is extended from Kalman filter. Different from Kalman filter, particle filter can be applied to track target in nonlinear and non-Gaussian condition. It disseminates weighted discrete random variables in status space. These variables can be approach to probability of every formation when number of particles extends to infinite. Particle filter computes integration of Bayesian estimation by applied large number theorem and statistical testing method. It first samples particles by defined sample function. Then it sets weights of particles according to similarity of particles by likelihood function. Finally, it computes posterior probability by applied particles and their weights. Since UKF (Unscented Kalman Filter) is a local optimization method and is not suitable to apply in nonlinear and non-Gaussian condition, particle filter algorithm is applied in this condition instead of UKF.

Equation 10 is applied to compute probability density function of color distribution $p(y) = \{p_u(y)\}_{u=1 \dots m}$ at position y , where m denotes number of features in feature space, $p_u(y)$ denotes the u^{th} feature of y , n is number of pixels in chosen area, x_0 is center position,

μ is normalization factor that expresses in Eq. 11, c is diagonal length of display area, $s()$ is applied to get serial number of feature, $\delta(0)=1$ and $\delta(\text{other})=0$ is a flag function.

$$p_u(y) = \mu \cdot \sum_{i=1}^n \left(k \left(\frac{\|y_i - x_0\|}{c} \right) \cdot \delta(s(y_i) - u) \right) \quad (10)$$

$$\mu = \frac{1}{\sum_{i=1}^n k \left(\frac{\|y_i - x_0\|}{c} \right)} \quad (11)$$

Then, sample of every color distribution is defined as Eq. 12, where p_y is positions of particle y , \vec{v}_y is velocity vector of y , θ is deflection angle in horizontal direction.

$$y = \{p_y, \vec{v}_y, \theta, c\} \quad (12)$$

Sample set of particle is updated and broadcasted through the transition function of system status in Eq. 13, where A is status transition matrix and w_{t-1} is Gaussian noise at time $t-1$.

$$y_t = Ay_{t-1} + w_{t-1} \quad (13)$$

In detail, Particle filter algorithm is presented in following Alg.3.

Alg.3 Particle Filter Algorithm

Step 1. Construct feature representation model of target.

Assuming $\{y_{t-1}^i\}_{i=1}^m$ as sample set of posterior probability distribution of target status, applying Eq. 13 to compute current particle set y_t and positions of current particles.

Step 2. Compute weight for all particles.

For all samples of particle set in status set y_t , to compute its color histogram by applied Eq. 14, which is from Eq. 10.

$$p_u(y_t) = \mu \cdot \sum_{i=1}^n \left(k \left(\frac{\|y_t^i - x_0\|}{c} \right) \cdot \delta(s(y_t^i) - u) \right) \quad (14)$$

Then, according to current status, to compute weight of samples by applied Eq. 16 (Bayesian Theorem) when consider target model is equal to prior probability distribution as Eq. 15.

$$P(y_t | p_u(y_t)) = P(y_t | p_u(y_{t-1})) \quad (15)$$

$$\omega_t^i = \frac{P(y_t | y_t^i)}{\sum_{i=1}^n P(y_t | y_t^i)} \quad (16)$$

Step 3. Estimate target status.

Finally, target status is estimated as Eq. 17.

$$\bar{y}_t = \sum_{i=1}^n (\omega_t^i \cdot y_t^i) \quad (17)$$

Alg.3 finished.

3.2 Algorithm analysis and improvement

Particle filter algorithm has its advantage. First, it can be applied in nonlinear and non-Gaussian condition. So compared to the Kalman filter, which can only applied to solve problem of linear system, particle filter algorithm is applied widely. Another advantage is that computational complexity of particle filter algorithm can be simplified by Monte Carlo method.

Disadvantage of particle filter algorithm also exists. First, there is particle degeneracy phenomenon in this algorithm. Particle degeneracy means most weights of particles changed very small after several iterations. This is because generated random sample changed to invalid sample which are predicted by state variables. Invalid sample is not much significance for the estimation. However, along with increment of invalid sample, number of valid samples is relative reduced. This makes large deviation in estimation. Meantime, particle filter method slows because it spends a large amount of computing resources in sample with small weight.

Second, there is poor sample phenomenon in particle filter algorithm. To avoid particle degeneracy, a variety of resampling algorithms is applied. Though resampling can eliminate particle degeneracy phenomenon, it introduces poor sample phenomenon. When system noise is relatively smaller, this problem becomes more serious. This is because excessive replication of particles with large weight in resampling process. This makes the particle diversity decreases. Poor diversity of particle can not reflect status of the system with their statistical characteristics.

Figure 3 shows one instance for particle filter algorithm.

In Fig. 3, we find that particle filter algorithm successful tracks moving target with particles in Fig. 3a, but it fails in Fig. 3b. The failure occurs because of particle degeneracy.

In order to improve the weakness of particle filter algorithm, and to make up for weakness by particle degeneracy and poor sample, many researchers provide different methods to improve the efficiency of particle filter algorithm.

Resampling method is applied to reduce particle degeneracy. A frequently used resampling method is multiterm sampling algorithm. Sampling variance of this method is presented in Eq. 18, where N_i is sampling times and $\sum_{i=1}^n N_i = N$.

$$\text{var}(N_i) = N\omega_k^i(1-\omega_k^i) \quad (18)$$



Fig. 3 An instance of particle filter algorithm in target tracking

To reduce resampling variance and improve resampling effectiveness, many resampling methods are presented today. For example, residual sampling, systematic sampling and adaptive resampling algorithm are all resampling method [37, 43, 49].

To avoid poor sample, researchers also provide many solutions, regularized particle filter and Markov chain Monte Carlo method etc. [38, 90]. There are also other particle filter algorithms today. Gaussian Particle Filter (GPF) applied Gaussian distribution to construct importance function [52]. It generate next particle by computed Gaussian distribution of all current particles in particle set. GPF is an asymptotically optimal method when number of particle tends to infinity and particle distribution obeys Gaussian distribution. Chen et al. applied prior distribution under process of SA (annealing simulated) as importance function [13]. SA uses Metropolis method to reach a well particle distribution. Real-time Particle Filter (RTPF) improves real-time performance by applied reduction of particle number [54]. It constructs optimal importance function to ensure accuracy of particle filter method.

Particle filter also can be combined other models and generated more real-time and robustness tracking methods. Ross et al. presented incremental visual tracking by PCA of difference, which was generated by image space of frames before current frame [65]. Kwon et al. improved Ref.44, and presented two dimensional affine transformation matrix to describe object [55]. Bao et al. combined L1 minimum tracking to particle filter to solve negativeness rotation, scale transformation, occlusion and illumination change [6]. Jia et al. applied structural local sparse appearance model to describe target [46]. Zhong et al. presented their tracking method by collaborative model with holistic templates and local representations [95]. Their method showed robustness when background changed. Oron et al. provided locally orderless tracking method with locally orderless matching of target [72, 76].

Recently, there are also many researches in this area. Yi et al. applies robust combination of particle filter and sparse representation in single target tracking [92]. Nieto et al. presents a real-time tracking system which has well tracking speed and accuracy [70]. Zhou et al. presents hybrid particle filter and sparse representation to avoid defect of classical particle filter method [96]. Their method solves problem of target offset in traditional tracking algorithm.

4 Image processing method for target tracking

4.1 Active contour model

Active contour model is a recognition method in image recognition. It needs a general contour of tracking target first. Then it solves the differential equation recursively to converge to local minimum of the energy function. Energy function is usually constructed according to the smoothness of the features and contour profile of image, such as edges and curvature. Target tracking algorithm based on contour has no restriction on target shape and movement.

Snake model is a representative method of active contour model [50]. It can also be applied in visual moving target tracking. Snake method applies a deformation contour line that moves by internal and external force of image. Contour of target is defined by constructed suitable deformation energy $E_s(v)$, where s denotes unit variable domain, $v(x(s), y(s))$ denotes a mapping from $s \in [0, 1]$ to image surface. To consider external force of target contour as a differential of potential energy $P(v(s))$, total energy of target contour can be defined in Eq. 19,

where $E_s(v)$ can be computed in Eq. 20. In Eq. 20, v is rewritten to $v(s)$ because v is a function of s . So $E_s(v)$ also can be rewritten to $E_s(v(s))$.

$$E(v) = E_s(v) + P(v) \quad (19)$$

$$E_s(v(s)) = \int_0^1 \left(k_1 \left| \frac{\partial}{\partial s} v(s) \right|^2 + k_2 \left| \frac{\partial^2}{\partial s^2} v(s) \right|^2 \right) ds \quad (20)$$

Equation 20 defines a internal deformation energy of an extensible and flexible $v(s)$. $E_s(v(s))$ contains two variables which are $k_1(s)$ and $k_2(s)$. In Eq. 20, $k_1(s)$ controls stress of contour and $k_2(s)$ controls rigidity of contour. Then, Eq. 21 is presented to describe $P(v)$. $P(v)$ is a scalar function that its definition domain is whole image surface $I(x, y)$.

$$P(v) = \int_0^1 p(v(s)) ds \quad (21)$$

In $p(v)$, Eq. 22 is construct to attract Snake to edge (contour) of target, where k_3 controls attracting strength of potential energy, G_e means Gaussian smoothing filter with characteristics density e , symbol “*” means convolution.

$$p(x, y) = \pm k_3 \cdot |G_e * I(x, y)| \quad (22)$$

Active contour model has many advantages. First, in snake method, both features of image data, initial estimation, target contour and constraint based on knowledge unify to one process of feature extraction. Second, the method converges to status of minimum energy automatically based on suitable initialization. Third, extracting area extends and computational complexity decreases in snake because energy is minimized from large size to small size in scale space.

There are two main defects for active contour model. One is that active contour model is sensitized with its initial position. It works well when snake is first put near features of tracking target. But it can not track correct when initial position is far from useful features. The other defect is that active contour model is a non-convex function. This means the method may converge to local extreme value, even divergent.

Figure 4 shows one instance for active contour model.

In Fig. 4, we find that snake algorithm successful tracks moving target with particles in Fig. 4a, but it fails in Fig. 4b. The failure occurs because of contour is changed when two persons overlapping. This makes previous features and deformation contour line changed to features and line for both the two persons. In other words, snake method consider the two persons in Fig. 4b as one target.

There are many researchers improve active contour model. Menet et al. provide B-snake model. Target contour is expressed by B- spline, which can express angular points impliedly [66]. This method can be applied to edge detection and contour extraction without priori knowledge of target contour. Initial snake method will converge to one point or one line without image force. Cohen et al. presented a balloon model based on classical active contour model [18]. Their model depresses sensitiveness of snake model, and can cross some pseudo edge points (local extreme value) in image. Cohen applied the method to medical image



Fig. 4 An instance of snake in target tracking

processing, and extract ventricular contour from magnetic resonance and ultrasound images. Allili et al. applied adaptive mixture model to active contour [4]. Brox et al. provided a level set-based segmentation method by combined spatial and temporal variations to active contour [11]. Chiverton et al. presented an automatic tracking and segmentation method by motion-based bootstrapping and shape-based active contour [15]. Aitfares et al. presented a mixed target tracking method by region intensities and motion vector [3]. Ning et al. combined registration and active contour segmentation [71].

Recently, Cai et al. present DGT (dynamic graph-based tracker) to solve deformation and illumination change [12]. Sun et al. provide non-rigid target tracking method by supervised level set model and online boosting [79]. Choi et al. combine optical flow to active contour, in which optical flow solves deformation and speed change and active contour tracks contour of moving target [16]. Mondal et al. combine local neighborhood information to fuzzy k-nearest neighbor classifier [67]. Huang et al. improve snake method and apply it to segmentation and tracking of lymphocytes effectively [42]. Their result has a greater improvement with previous algorithms. Cuenca et al. construct new computational mode of constraining force that adopt GPU-CUDA to compute convolution between image and Gaussian smoothing filter [26]. Their method is now adopted into computerized tomographic scanning (CT).

4.2 Scale-invariant feature transform method

Scale-invariant feature transform (SIFT) is a kind of machine vision algorithm. It detects and describes local features of image. SIFT searches extreme points in the scale of space, and extracts position, scale, rotation invariant of extreme points as features. This algorithm is first presented by David Lowe in 1999 [61], and summarized in 2004 [62]. There are many researchers improve SIFT method these years. PCA-SIFT method applies principal components analysis method to reduce dimension of descriptor. It reduces usage of internal memory and speed up the matching speed [51]. SURF method approximates many processes in SIFT method to accelerate process of recognition [7].

First, applying Gaussian convolution kernel as linear kernel in scale transformation, an scale space in R^2 can be defined in Eq. 23 where $G()$ is a Gauss's function with variable scale and described in Eq. 24, (x, y) denotes scale position, σ controls image smoothness. Then, in order to search stable key points in scale space, difference of Gaussian scale space (DOG) is

presented to instead of Laplacian of Gaussian operator (LOG) in Eq. 25 by different Gaussian difference kernel.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (23)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (24)$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (25)$$

SIFT algorithm can be divided into 4 steps in Alg.4.

Alg.4 SIFT Algorithm

Step 1. Extreme value detection in scale space.

DOG operator is applied to construct difference scale space. Then, every sample point is compared with all neighbors, which conclude 8 points in same scale and 18 points in higher and lower scales. A point is an extreme value if it is extreme value of its scale in all these 27 points.

Step 2. Key point positioning.

Positioning key points and their scales need function fitting, which is presented in Eqs. 26 and 27, where Eq. 27 is presented when let $D'(z) = 0$.

$$D(z) = D + \frac{\partial D^T}{\partial z} z + \frac{1}{2} z^T \frac{\partial^2 D}{\partial z^2} z \quad (26)$$

$$\dot{z} = -\frac{\partial^2 D^{-1}}{\partial z^2} \frac{\partial D}{\partial z} \quad (27)$$

Then, feature points with low contrast and instable edge points are dropped by Eqs. 28 and 29 where ε denotes dropped threshold, H is curvature and presented in Eq. 30, $r > 1$ is ratio between two eigenvalue of H .

$$|D(\dot{z})| = \left| D + \frac{1}{2} \cdot \frac{\partial D^T}{\partial z} \dot{z} \right| \geq \varepsilon \quad (28)$$

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r+1)^2}{r} \quad (29)$$

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (30)$$

Step 3. Construct SIFT feature vectors.

First, applying the gradient direction distribution of neighborhood of key points, direction parameters of each key point are assigned to keep rotation invariance of operators. Then, primary direction of key points is computed by its modulus and direction of gradient. It is computed by statistical maximum of all directions of gradients in neighborhood of key position.

After that, according to primary direction of key point, one window is extracted with size c . Gradients of all pixels in this window is weighted computed by Gaussian window. Then, with size $2c$ and $c/2$, same computation is processed for all blocks in window of one key point. So, there are 16n features to describe one key point if we define n directions.

Finally, illumination effect is removed by normalization of this feature vector.

Step 4. Applying SIFT in extraction.

To generate feature vectors for two images A and B. Then, Euclid distances between every key points b_i of B and a_j of A are computed as d_{ij} . To find the smallest two $d_{i'j}$ and $d_{i''j}$ for j , $b_{i'}$ and a_j is a pattern pair if $d_{i'j}/d_{i''j} < \varepsilon$, where ε controls stability of SIFT.

SIFT has many advantages. First, SIFT feature extracts local feature of image. It is robust with rotation, zooming scale, illumination changes. SIFT also shows good stability in perspective changes, affine transforms and noise. Second, Features of SIFT contain much information, that suit for fast and accurate recognition in big data features. Third, lots of SIFT features are constructed even if only a few targets.

SIFT also has some defects. First, traditional SIFT shows bad performance for nonrigid target. Recognition occurs error when target is deformed. Second, computational complexity of SIFT is high because each key point uses many dimensional vectors to represent. Time costs large when features of image increase. Third, in SIFT, image feature extraction is not sufficient and recognition effectiveness is not well enough.

Figure 5 shows one instance for SIFT.

In Fig. 5, we find that SIFT algorithm successful tracks moving target with quadrilateral in Fig. 5a, but it fails in Fig. 5b. The failure occurs because many features of SIFT present wrong direction. So, in Fig. 5b, SIFT is degeneration and recognizes error.

Today, there are many improvements for SIFT. First, Ke and Sukthankar presented PCA-SIFT in 2004 [51]. They combined principle component analysis (PCA) into SIFT to reduce dimensions of points. Their method reduced computational time of SIFT. Dalal et al. presented HOG to divide image into small connecting unit, which were combined to block to compute gradient [27]. Bosch et al. changed feature selection space and computed SIFT operators in HSV space [10]. Abdel-Hakim and Farag applied RGB color information to construct CSIFT method [1]. They extended SIFT from grey image recognition to color. Their method matched features by three colors and applied three dimensions in recognition. Morel et al. presented



Fig. 5 An instance of SIFT in target tracking

affine SIFT method to improve affine performance of SIFT [68]. Their method added two angles which cause affine transformation.

SIFT and its derivative algorithm, such as SURF, HOG etc. are important in visual target tracking and studied by many researchers. Zhu et al. first applied HOG into human detection [97]. Wang et al. combined multi-features into target tracking [85]. Xu et al. combined HOG to SVM [89]. Chu et al. extracted features by RGB-based SURF which was better than gray-based SURF [17]. Babenko et al. applied SURF features into multi-instance learning [5]. Sun combined HOG features to online boosting framework and achieve good performance [78]. Xia applied Haar feature and HOG features into online boosting method to classify features [87].

SIFT can also be combined to other method. Wang et al. combined SIFT with Camshift [83]. They use Bhattacharyya coefficient to determine matching of SIFT features. Their method has well robustness when tracking target is deformed, revolved and occluded. Xie et al. applied features of SIFT in support vector machine (SVM) [88]. They extended SIFT to multi-targets (rigid) tracking domain. Their method is applied in multiple aircrafts tracking.

5 Experimental results and analysis

5.1 Dataset

Recently, though benchmark of visual target tracking has been made a lot of progress, it is a very necessary work to construct a video data set which has objective evaluation of the tracker performance with a variety of complex changes. It is a key to collect a representative data set for comprehensive performance evaluation of trackers. There are some visual tracking data set for surveillance, such as CAVIAR [29] and VIVID [22]. But targets in both of them moves simple with static background. So we need more complex data set to evaluate algorithms.

Recently, there are many researchers study to construct video data sets and evaluation frameworks. Ref.87 constructs a video test data set, which includes 50 labeled image sequences to evaluate tracking algorithms. It also has many open trackers (29 algorithms can be found now) to evaluate algorithms. Ref. 77 selects 315 sequences with different kinds, and evaluate 19 open trackers in this data set. Positiveness of Ref. 77 is that this reference provides online evaluation at website <http://www.alov300.org>, which only needs uploaded tracking results with standard data set. Then, the website can present comparison result of test algorithm and existing algorithms by F-score and accuracy. Then, in Ref. 58, they provide a data set including 356 video sequences of 12 kinds. This data set focuses on rigid targets and pedestrian. In its website http://www.lv-nus.org/pro/nus_pro.html, there is test data from 20 good tracking algorithms.

In 2015, Wang et al. provided an evaluation method to evaluate trackers by motion model, feature extractor, observation model, model updater, and ensemble post-processor [84]. Further, In VOTC (Visual Object Tracking Challenge), data sets and evaluation criteria is uniform [53].

5.2 Experimental parameters

5.2.1 Tracker

This paper select 9 algorithms to compare each other. These algorithms are selected because they show effective results in our experiment. The comparison is only to show positiveness

Table 1 Evaluated Tracking Algorithms

Method	Category	Representation	Search	MU	Code
PF	Particle Filter	H, PCA, GM	PF	Y	MC
ASLA [46]	Particle Filter Extension	L, SR, GM	PF	Y	MC
SCM [95]	Particle Filter Extension	L, SR, GM + DM	PF	Y	MC
L1APG [6]	Particle Filter Extension	H, SR, GM	PF	Y	MC
SIFT	Based On SIFT	L, T	LOS	Y	M
HAH [87]	SURF Extension	L, T, SR	LOS	Y	M
KMS [23]	Mean Shift	H, IH	LOS	N	MC
CAM [25]	Mean Shift Extension	H, IH	LOS	N	MC
SNA	Based On Active Contour	–	–	N	M

and negativeness of all kinds of introduced tracking algorithms in this paper. Experimental results do not imply overall performance of these algorithms. Table 1 presents evaluated 9 algorithms, where ‘MU’ implies model update in table head; ‘L’ implies local, ‘H’ implies holistic, ‘T’ implies template, ‘IH’ implies intensity histogram, ‘PCA’ implies principal component analysis, ‘SR’ implies sparse representation, ‘DM’ implies discriminative model, ‘GM’ implies generative model in ‘Representation’ column; ‘PF’ implies particle filter, ‘LOS’ implies local optimum search in ‘Search’ column; ‘N’ implies no, ‘Y’ implies yes in ‘MU’ column; ‘M’ implies Matlab code, ‘C’ implies C/C++ code, ‘MC’ implies the algorithm combines Matlab and C/C++ in ‘Code’ column.

5.2.2 Test set

This paper applies test platform in Ref.87, which contains 50 video sequences for performance test of trackers. In Fig. 6, we can find 1st frame of all test sequences, which is labeled by standardized bounding box to compare (Red rectangle in each image). In this paper, we choose 5 representative video sequences to evaluate these 9 algorithms in Table 1.

Ref.87 provides 11 different attributes of data set. So in this paper, we analyze tracking accuracy in IV, SV, OCC, DEF, MB, FM, IPR, OPR, OV, BC and LR, which were presented in

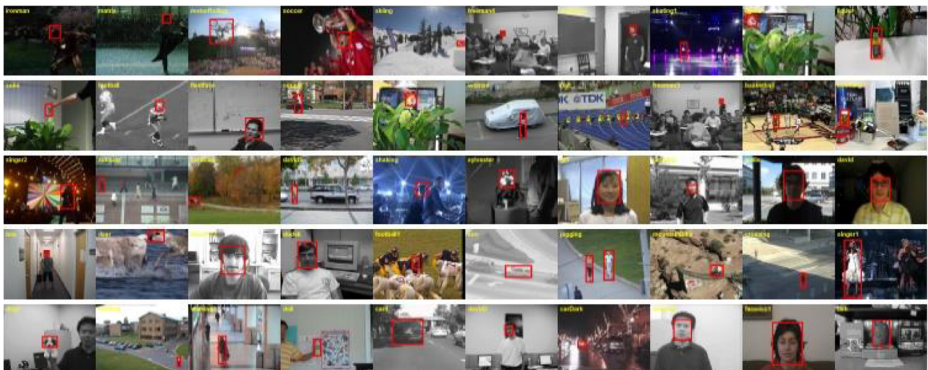


Fig. 6 1st frame with label of 50 video sequences

Table 2 List of the attributes annotated to test sequences [86]

Attr	Description
IV	Illumination Variation - the illumination in the target region is significantly changed
SV	Scale Variation - the ratio of the bounding boxes of the first frame and the current frame is out of the range $[1/ts, ts]$, $ts > 1$ ($ts = 2$)
OCC	Occlusion - the target is partially or fully occluded
DEF	Deformation - non-rigid object deformation
MB	Motion Blur - the target region is blurred due to the motion of target or camera
FM	Fast Motion - the motion of the ground truth is larger than tm pixels ($tm = 20$)
IPR	In-Plane Rotation - the target rotates in the image plane
OPR	Out-of-Plane Rotation - the target rotates out of the image plane
OV	Out-of-View - some portion of the target leaves the view
BC	Background Clutters - the background near the target has the similar color or texture as the target
LR	Low Resolution - the number of pixels inside the ground-truth bounding box is less than tr ($tr = 400$)

Table 2, with same test data. Then, we apply experiments to illustrate application environment and their advantages, defects in next section.

5.2.3 Evaluation criterion

This paper applies benchmark in Ref.87 as evaluation platform. So we apply evaluation criterion of Ref.87 to evaluate these algorithms, which is same to Refs.87-88.

This evaluation criterion is bounding box overlap score, which is defined as $S = |\gamma_t \cap \gamma_a| / |\gamma_t \cup \gamma_a|$, where T and S represent the intersection and union of two regions, respectively, $|\cdot|$ denotes the number of pixels in the region, γ_t is tracked bounding box and γ_a is ground truth bounding box. Then. Because applying one success rate value at a specific threshold (e.g. $to = 0.5$) for tracker evaluation may not be fair, AUC (area under curve) is also applied of each success plot to rank the tracking algorithms

Then, we have evaluation of robustness. Classic evaluation method is OPE (one-pass evaluation), which is first initialized by accurate position in 1st frame, then processed in text sequence, and last reported as a result of average accuracy. OPE is sensitive to initialization, and changes its performance by different initial frame. So other two evaluation methods are applied



Fig. 7 Test with video sequence 1 (Walking 1)

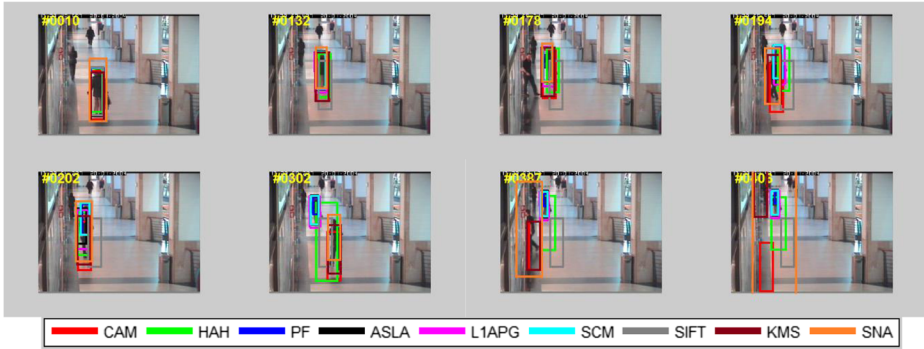


Fig. 8 Test with video sequence 2 (Walking 2)

to evaluate robustness of initialization which disturbs initialization for time (tracking from different frame) and space (tracking from different bounding box). These two methods are TRE (temporal robustness evaluation) and SRE (spatial robustness evaluation).

In this paper, each tracker is tested with more than 29,000 frames in OPE, is evaluated by 12 times in each sequence with more than 350,000 bounding box in SRE, is tested with about 310,000 frames (each sequence is divided into 20 segments) in TRE.

5.3 Experiment result and analysis

5.3.1 Tracking results

This sub-section mainly consists with two parts. In first part, we choose 5 representative video sequences from the data set of Ref.87 to test selected 9 trackers. Test results are provided and analyzed by combined 11 annotated attributes. In second part, we provide tracking result and analyze overall performance of 9 trackers in 50 video sequences. Also, we analyze tracking results for 9 trackers in 11 annotated attributes. So experiments in these two parts can validate our conclusions of these 4 kinds of algorithms.

In video sequence ‘Walking’, there are only three attribute changes, which are ‘SV’, ‘OCC’ and ‘DEF’. Indeed, it is a video sequence with less change and lower difficulty. Some tracking results are provided in following Fig. 7. In Fig. 7, we find that SNA misses its tracking target

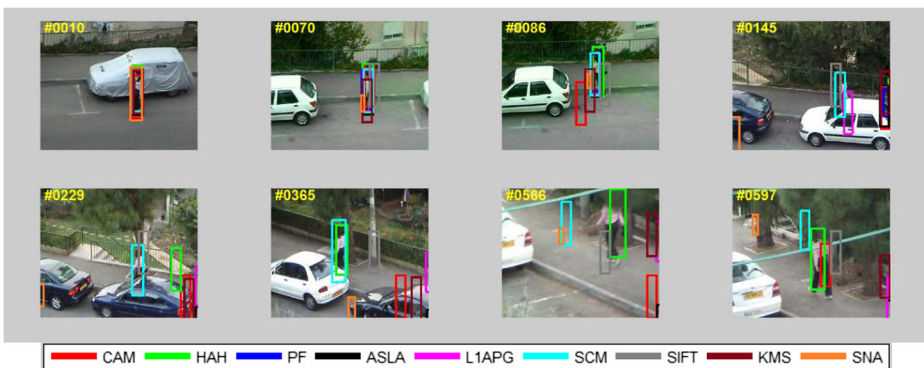


Fig. 9 Test with video sequence 3 (Woman)

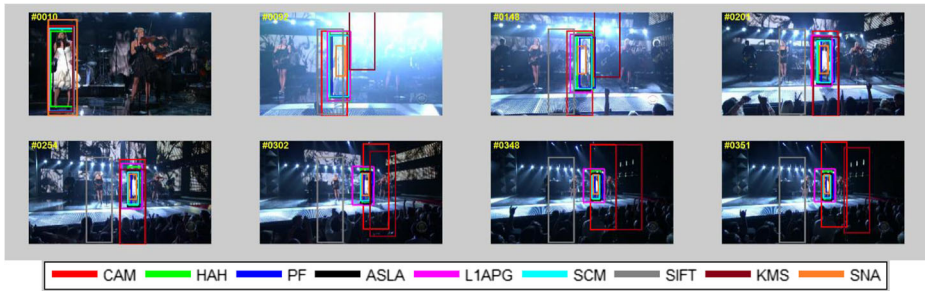


Fig. 10 Test with video sequence 4 (Singer)

partly because occlusion appears by telegraph pole in 88th frame. In 203rd frame, change of background leads to failure of HAH. CAM fails in 265th frame because of partly deformation of target.

Then, we test algorithms by video sequence Walking 2, which is also a relative simple test data and has three attribute changes ‘SV’, ‘OCC’ and ‘LR’. Partly tracking results are provided in following Fig. 8. In Fig. 8, we can find that dislocation between tracking box and target appears in SNA and KMS because background is changed in 132nd frames. Then, in 194th frame, another moving object appears and disturb tracking by occlusion. Tracking error occurs in both KMS, SIFT, SNA and CAM. Then, both SNA, KMS, CAM, SIFT and HAH fail to track target because of disturbance of the other moving object in 302nd frame. Finally, in 387th–403rd frames, after disappearance of disturbed object, both of these 5 algorithms can not track target accurately.

Then, we test algorithms with video sequence ‘Woman’, which contains more changed attributes ‘IV’, ‘SV’, ‘OCC’, ‘DEF’, ‘MB’, ‘FM’ and ‘OPR’. Experimental results are provided in following Fig. 9. In Fig. 9, we find SNA converges to local target and shows bad performance by scale change in 70th frame. In 86th frame, deformation appears in target, which leads to failure of KMS and CAM. Then, both HAH, PF and ASLA fails to track target because target occludes partly by car in 145th frame. Further, in 229th frame, partly occlusion appears by complex background, which makes tracking error for all 9 algorithms. When target blurs by its quicken movement and partly occluded by tree, only HAH partly tracks target in 566th frame.

Then, video sequence ‘Singer’ with attributes ‘IV’, ‘SV’, ‘OCC’ and ‘OPR’ is chosen because it has complex illumination changes, which can compare robustness of tracking



Fig. 11 Test with video sequence 5 (Skating)

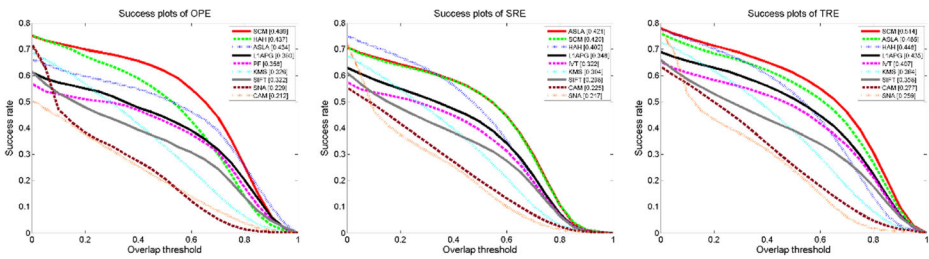


Fig. 12 Overall performance evaluation

algorithms with illumination changes. Experimental results are provided in Fig. 10. In Fig. 10, KMS fails to track and SNA converges local target when illumination enhances in 92nd frame. Then KMS and SIFT fails their tracking when target rotates out of the image plane (OPR) in 148th frame. Also, tracking box of CAM enlarges. In 302nd frame, target keeps rotation with illumination changes, which leads to failure of CAM.

Final video sequence ‘Skating’ has attributes ‘IV’, ‘SV’, ‘OCC’, ‘DEF’, ‘OPR’ and ‘BC’. This video sequence is chosen because it has many negative attributes which can evaluate algorithms overall. Also, tracking failure is easy to be appeared because of these attributes. Experimental results are provided in Fig. 11. In Fig. 11, SNA fails to track when target quickly deforms in 23rd frame. In 88th frame, only SCM tracks target when quick scale variation of target appears. Between 165th and 268th frames, all algorithms can not track effectively because of quick movement, rotation, and occlusion. Between 329th and 393rd frame, target moves quickly with large illumination change, which leads to failure of all tracking algorithms.

5.3.2 Evaluation results

After experimental analysis of 9 algorithms in 5 video sequences, we apply benchmark in Ref.87 to analyze tracking results of tracking in 50 video sequences with 9 algorithms. This can evaluate 4 kinds of algorithms in this paper. We first evaluate overall performance of 9 algorithms. Then, performances of 5 reprehensive attributes in all 11 attributes are analyzed.

Figure 12 is presented to analyze overall performance. AUC value is applied to conclude and rank performances of algorithms. In Fig. 12, we find that average performance of TRE is higher than OPE. It is because that frames of OPE is less than sum of frames in TRE. Average performance of TRE is higher because performances of tracking algorithms are better when video sequences are shorter. Similarly, average performance of SRE is lower than OPE. Overall performance shows that algorithms based on particle filter and its extension have

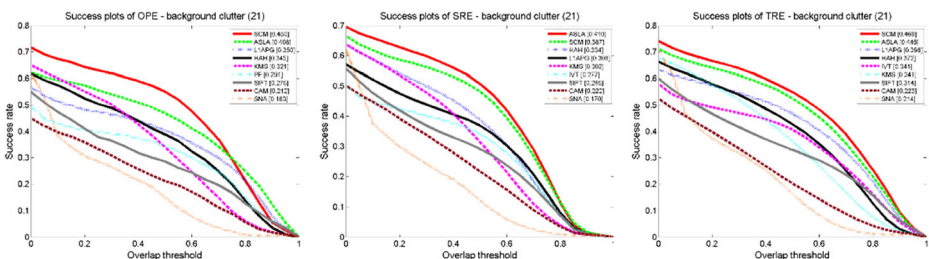


Fig. 13 performance evaluation with background clutter

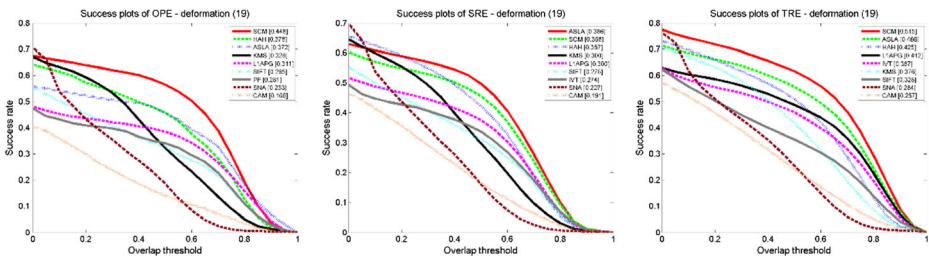


Fig. 14 performance evaluation with deformation

better performances. One reason is algorithms based on particle filter are more than other kinds of algorithms. Also, newer algorithms show better performances than older ones.

Figure 13 is applied to show performances of algorithms with background clutter. There are 21 video sequences has attribute ‘background clutter’. SNA has lowest average performance in all algorithms. This is because SNA needs clearer background to converges contour of target. Also, we find SCM has best robustness of complex background.

Figure 14 is applied to show performances of algorithms with deformation. There are 19 video sequences has attribute ‘deformation’. CAM has lowest average performance in all algorithms, which is according to our conclusion with it. Also, we find both SCM and ALSA have best performance when target deforms.

Figure 15 is applied to show performances of algorithms with illumination change. SNA does not achieve good average performance because illumination change makes weak degree to divide background and target. Also, we find both SCM and ALSA have good performance. Then, we find HAH has good performance because it is based on SIFT, which has Illumination invariance.

Figure 16 is applied to show performances of algorithms with scale variation. CAM does not achieve good average performance because its adaptability of scale is poor. Both SCM and ALSA have good performance with this attribute.

Figure 17 is applied to show performances of algorithms with occlusion. Only HAH has good robustness with partly occlusion in all 9 algorithms. This is because HAH is based on feature extraction which can extract some features to track target in occlusion.

6 Discussion and conclusion

After decades of technological development, visual target tracking technology is still a highlight in computer vision domain. This paper first briefly introduced visual moving target

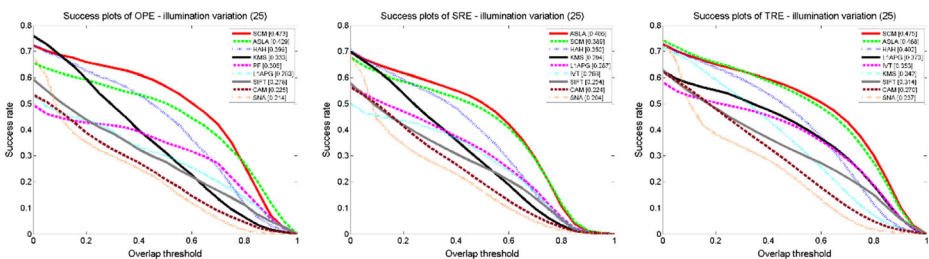


Fig. 15 performance evaluation with illumination change

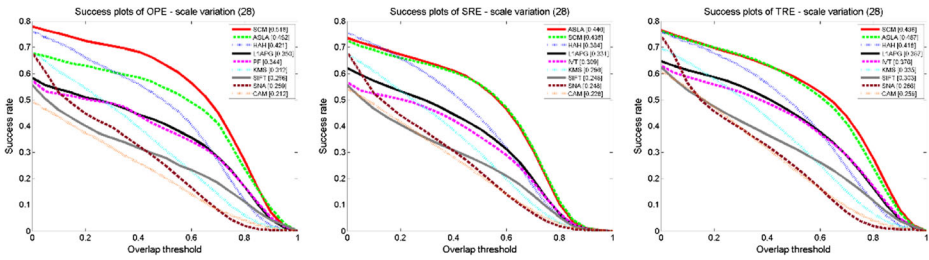


Fig. 16 performance evaluation with scale variation

tracking technology and its application. Then target tracking methods is classified by their thinking and type of extracted features. Further, every kind of tracking method are elaborated with related algorithm, that include advantages and disadvantages as well as improvement of these algorithms. Moreover, algorithms were also compared with advantages, disadvantages and application scenarios. In summary, this paper provides a comprehensive and objective review in this research domain.

Now there are also many problems in related research domains. They are mainly in the following aspects.

1) Occlusion

Though there are many researches in recognition with occlusion, most of the target tracking system can not completely solve this problem. Occlusion appears in several conditions, which are occlusion between different targets, occlusion due to foreground or background, and object occlusion itself.

Target is only partly visible when occlusion appears. So tracking results always extract tracking mistakes by classical tracking method, soever overall or local features, static or dynamical model.

Our team constructs a new target tracking method based on movement feature [30–32]. This method can partly solve short time occlusion. But it still needs a long way to go.

2) Feature selection

Target feature selection and measurement is an important problem in target tracking method. Appropriate features can improve the tracking efficiency, accuracy. So how to select features, how to combine different features, how to weight features are still valued to be research today.

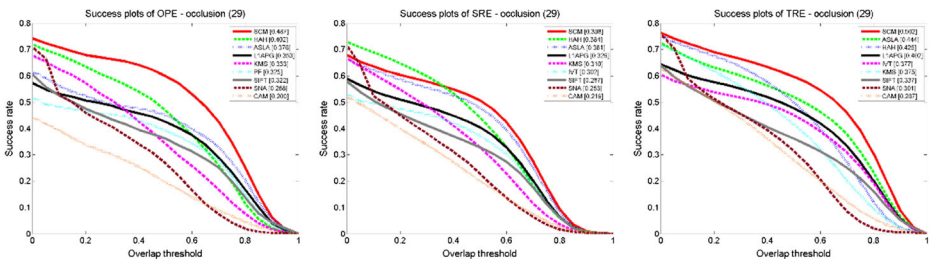


Fig. 17 performance evaluation with occlusion

Adaboost classifier is a widely applied classifier to combined different features. Also, learning system is an effective method in weight value. Recently, deep learning shows its potential in image recognition, and it will be applied in multimedia tracking soon.

3) Tracking of multi-targets

Earlier, target tracking systems are mainly applied to track single target. Today, requirement of multi-targets tracking becomes large. The main difference between tracking of single target and multi-targets is complexity. Compared to single target tracking, multi-targets tracking is uncertainty in the existence of the corresponding relationship between observation and target. The tracking mistake may occur when multiple targets are close enough.

Multi-vision method is widely used recently to recognize multi tracking targets. It applies multiple cameras in same background. Then it applies difference between cameras to detect and track targets.

At last, with rise of natural network again from year 2006, deep learning has been widely applied in many research domains, including computer vision [36]. Also, deep learning is applied in visual target tracking recent years. In 2015 VOTC (Visual Object Tracking Challenge), the best algorithm MDNet applied CNN (Convolutional Neural Network) to learn objects [69]. Meantime, Deep-SRDCF also achieve a top 3 result in this challenge [28]. So we can see that deep learning will be another hotspot in this research domain.

However, target tracking method in video is now an important research domain, and already has a wide range of research value. This research field has lots of directions to explore and worth exploring.

Acknowledgments This research is supported by following grants:

National Natural Science Foundation of China [61502254] and National Natural Science Foundation of Inner Mongolia [2014BS0606].

References

1. Abdel-Hakim AE, Farag AA (2006) CSIFT: A SIFT descriptor with color invariant characteristics// Computer Vision and Pattern Recognition, 2006 I.E. Computer Society Conference on. IEEE 2:1978–1983
2. Agrafiotis D, Davies SJC, Canagarajah N et al (2007) Towards efficient context-specific video coding based on gaze-tracking analysis. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 3(4):4
3. Aitfares W, Bouyakhf EH, Herbulot A et al (2013) Hybrid region and interest points-based active contour for object tracking. *Appl Math Sci* 7(118):5879–5899
4. Allili MS, Ziou D (2008) Object tracking in videos using adaptive mixture models and active contours. *Neurocomputing* 71(10):2001–2011
5. Babenko B, Yang MH, Belongie S (2011) Robust object tracking with online multiple instance learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33(8):1619–1632
6. Bao C, Wu Y, Ling H et al (2012) Real time robust l1 tracker using accelerated proximal gradient approach//Computer Vision and Pattern Recognition (CVPR), 2012 I.E. Conference on. IEEE: 1830–1837
7. Bay H, Tuytelaars T, Gool LV (2006) SURF: speeded up robust features. *Computer Vision & Image Understanding* 110(3):404–417
8. Beauchemin SS, Barron JL (1995) The computation of optical flow. *ACM Computing Surveys (CSUR)* 27(3):433–466
9. Birchfield ST, Rangarajan S (2005) Spatiograms versus histograms for region-based tracking//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE 2:1158–1163

10. Bosch A, Zisserman A, Muoz X (2008) Scene classification using a hybrid generative/discriminative approach. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 30(4):712–727
11. Brox T, Rousson M, Deriche R et al (2010) Colour, texture, and motion in level set based segmentation and tracking. *Image and Vision Computing* 28(3):376–390
12. Cai Z, Wen L, Lei Z et al (2014) Robust deformable and occluded object tracking with dynamic graph. *Image Processing*, IEEE Transactions on 23(12):5497–5509
13. Chen Z (2003) Bayesian filtering: fKalman filters to particle filters, and beyond. *Statistics* 182(1):1–69
14. Cheng Y (1995) Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 17(8):790–799
15. Chiverton J, Xie X, Mirmehdi M (2012) Automatic bootstrapping and tracking of object contours. *Image Processing*, IEEE Transactions on 21(3):1231–1245
16. Choi JW, Whangbo TK, Kim CG (2015) A contour tracking method of large motion object using optical flow and active contour model. *Multimedia Tools and Applications* 74(1):199–210
17. Chu DM, Smeulders AWM (2010) Color invariant surf in discriminative object tracking,” in Proc. IEEE ECCV, Heraklion
18. Cohen LD (1991) On active contour models and balloons. *CVGIP: Image understanding* 53(2):211–218
19. Coifman B, Beymer D, McLauchlan P et al (1998) A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C: Emerging Technologies* 6(4):271–288
20. Collins RT (2003) Mean-shift blob tracking through scale space//Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. IEEE, 2: II-234–40 vol. 2
21. Collins RT, Liu Y, Leordeanu M (2005) Online selection of discriminative tracking features. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 27(10):1631–1643
22. Collins R, Zhou X, Seng KT (2005) An open source tracking testbed and evaluation web site. *Perf.eval.track. & Surveillance*
23. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 24(5):603–619
24. Comaniciu D, Ramesh V, Meer P (2000) Real-time tracking of non-rigid objects using mean shift, BEST PAPER AWARD, IEEE Conf. Computer Vision and Pattern Recognition (CVPR’00), Hilton Head Island, South Carolina 2:142–149
25. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 25(5):564–577
26. Cuenca C, González E, Trujillo A et al (2015) Fast and accurate circle tracking using active contour models. *J Real-Time Image Process*:1–10
27. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition* 1:886–893
28. Danelljan M, Hager G, Shahbaz Khan F et al (2015) Learning spatially regularized correlation filters for visual tracking//Proceedings of the IEEE International Conference on Computer Vision:4310–4318
29. Fisher RB (2004) The PETS04 surveillance ground-truth data sets. Proc. Sixth IEEE Int. Work. on Performance Evaluation of Tracking and Surveillance (PETS04)
30. Fu W, Zhou J, Liu S et al (2014) Differential trajectory tracking with automatic learning of background reconstruction. *Multimedia Tools & Applications*. doi:10.1007/s11042-014-2391-6
31. Fu W, Zhou J, An C (2015) Distributed dynamic target tracking method by block diagonalization of topological matrix. *Journal of Supercomputing*. doi:10.1007/s11227-015-1499-4
32. Fu W, Zhou J, Ma Y (2015) Moving tracking with approximate topological isomorphism. *Multimedia Tools & Applications*. doi:10.1007/s11042-015-2519-3
33. Fukunaga K, Hostetler LD (1975) The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory*, IEEE Trans 21(1):32–40
34. Gress O, Posch S (2012) Trajectory retrieval from Monte Carlo data association samples for tracking in fluorescence microscopy images//Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on. IEEE:374–377
35. Guo S, Shi X, Wang Y et al (2016) Non-rigid object tracking using modified mean-shift method//Information Science and Applications (ICISA) 2016. Springer Singapore:451–458
36. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
37. Hong S, Bolić M, Djuric PM (2012) An efficient fixed-point implementation of residual resampling scheme for high-speed particle filters. *IEEE Signal Processing Letters* 11(5):482–485
38. Hong S, Shi Z, Wang L et al (2013) Adaptive regularized particle filter for synchronization of chaotic colpitts circuits in an AWGN Channel. *Circuits Systems & Signal Processing* 32(2):825–841
39. Horn B K, Schunck B G (1981) Determining optical flow//1981 Technical symposium east. Int Soc Optics Photon:319–331

40. Hou Z, Han C (2006) A survey of visual tracking. *Acta Automatica Sinica* 32(4):603–617
41. Huang K, Chen X, Kang Y et al (2015) Intelligent visual surveillance: a review. *Chinese Journal of Computers* 38(6):1093–1118
42. Huang Y, Liu Z (2015) Segmentation and tracking of lymphocytes based on modified active contour models in phase contrast microscopy images. *Comput Math Methods Med* 2015
43. Hwang SS, Speyer JL (2011) Particle filters with adaptive resampling technique applied to relative positioning using GPS carrier-phase measurements. *IEEE Transactions on Control Systems Technology* 19(19):1384–1396
44. Jatoth RK, Gopisetty S, Hussain M (2015) Performance analysis of alpha beta filter, kalman filter and mean shift for object tracking in video sequences. *International Journal of Image, Graphics and Signal Processing (IJIGSP)* 7(3):24
45. Jeyakar J, Babu RV, Ramakrishnan KR (2008) Robust object tracking with background-weighted local kernels. *Computer Vision and Image Understanding* 112(3):296–309
46. Jia X, Lu H, Yang MH (2012) Visual tracking via adaptive structural local sparse appearance model// *Computer vision and pattern recognition (CVPR)*, 2012 I.E. Conference on. IEEE:1822–1829
47. Jia D, Zhang L, Li C (2015) The improvement of mean-shift algorithm in target tracking. *International Journal of Security and Its Applications* 9(2):21–28
48. Kalal Z, Mikolajczyk K, Matas J (2012) Tracking-learning-detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(7):1409–1422
49. Kalton G (2014) Systematic sampling//*Wiley StatsRef: Statistics Reference Online*. Wiley
50. Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. *International journal of computer vision* 1(4):321–331
51. Ke Y, Sukthankar R (2004) PCA-SIFT: A more distinctive representation for local image descriptors// *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 I.E. Computer Society Conference on. IEEE 2: II-506-II-513 Vol. 2*
52. Kotecha JH, Djurić PM (2003) Gaussian particle filtering. *Signal Processing, IEEE Transactions on* 51(10): 2592–2601
53. Kristan M, Matas J, Leonardis A et al (2015) The Visual Object Tracking VOT2015 challenge results// *Proceedings of the IEEE International Conference on Computer Vision Workshops:1–23*.
54. Kwok C, Fox D, Meila M (2004) Real-time particle filters. *Proceedings of the IEEE* 92(3):469–484
55. Kwon J, Lee KM, Park FC (2009) Visual tracking via geometric particle filtering on the affine group with optimal importance functions//*Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE:991–998*
56. Leichter I (2012) Mean shift trackers with cross-bin metrics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(4):695–706
57. Li X, Hu W, Shen C et al (2013) A survey of appearance models in visual object tracking. *ACM transactions on Intelligent Systems and Technology (TIST)* 4(4):58
58. Li A, Lin M, Wu Y et al (2016) NUS-PRO: a new visual tracking challenge. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 38(2):335–349
59. Li S, Wu O, Zhu C et al (2014) Visual object tracking using spatial context information and global tracking skills. *Computer Vision and Image Understanding* 125:1–15
60. Li X, Wu F, Hu Z (2005) Convergence of a mean shift algorithm. *Journal of Software* 16(3):365–374
61. Lowe DG (1999) Object recognition from local scale-invariant features//*Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Ieee 2:1150–1157*
62. Lowe DG (2004) Distinctive image features from scale-invariant key points. *International journal of computer vision* 60(2):91–110
63. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. *IJCAI* 81:674–679
64. Masoud O, Papanikolopoulos NP (2001) A novel method for tracking and counting pedestrians in real-time using a single camera. *Vehicular Technol, IEEE Trans* 50(5):1267–1278
65. Mei X, Ling H (2009) Robust visual tracking using L1 minimization//*Computer Vision, 2009 I.E. 12th International Conference on. IEEE:1436–1443*
66. Menet S, Saint-Marc P, Medioni G (1990) B-snakes: implementation and application to stereo//*proceedings DARPA* 720:726
67. Mondal A, Ghosh S, Ghosh A (2016) Efficient silhouette-based contour tracking using local information. *Soft Computing* 20(2):785–805
68. Morel JM, Yu G (2009) ASIFT: a new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences. SIAM J Imaging Sci* 2(2):438–469
69. Nam H, Han B (2015) Learning multi-domain convolutional neural networks for visual tracking. *Comput Sci*

70. Nieto M, Cortés A, Otaegui O et al (2016) Real-time lane tracking using Rao-Blackwellized particle filter. *Journal of Real-Time Image Processing* 11(1):179–191
71. Ning J, Zhang L, Zhang D et al (2013) Joint registration and active contour segmentation for object tracking. *Circuits and Systems for Video Technology, IEEE Transactions on* 23(9):1589–1597
72. Oron S, Bar-Hillel A, Levi D et al (2015) Locally orderless tracking. *International Journal of Computer Vision* 111(2):213–228
73. Pai CJ, Tyan HR, Liang YM et al (2004) Pedestrian detection and tracking at crossroads. *Pattern Recognition* 37(5):1025–1034
74. Peng N, Yang J, Liu Z et al (2005) Automatic selection of kernel-bandwidth for mean-shift object tracking. *Journal of Software* 16(9):1542–1550
75. Rosenfeld A (2001) From image analysis to computer vision: an annotated bibliography, 1955–1979. *Computer Vision and Image Understanding* 84(2):298–324
76. Shaul Oron DL, Aharon B-H, Avidan S (2012) Locally orderless tracking,” in Proc. IEEE CVPR, Providence
77. Smeulders AWM, Chu DM, Cucchiara R et al (2014) Visual tracking: an experimental survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36(7):1442–1468
78. Sun S, Guo Q, Dong F et al (2013) On-line boosting based real-time tracking with efficient hog//Acoustics, Speech and Signal Processing (ICASSP), 2013 I.E. International Conference on. IEEE:2297–2301
79. Sun X, Yao H, Zhang S et al (2015) Non-rigid object contour tracking via a novel supervised level set model. *Image Processing, IEEE Transactions on* 24(11):3386–3399
80. Tao X, Shaogang G (2006) Beyond tracking: modeling activity and understanding behavior. *Int J Comput Vis*
81. Vezzani R, Cucchiara R (2010) Video surveillance online repository (visor): an integrated framework. *Multimedia Tools and Applications* 50(2):359–380
82. Vojir T, Noskova J, Matas J (2014) Robust scale-adaptive mean-shift for tracking. *Pattern Recognition Letters* 49:250–258
83. Wang Z, Hong K (2012) A new method for robust object tracking system based on scale invariant feature transform and camshaft //Proceedings of the 2012 ACM Research in Applied Computation Symposium. ACM:132–136
84. Wang N, Shi J, Yeung D Y et al (2015) Understanding and diagnosing visual tracking systems//Proceedings of the IEEE International Conference on Computer Vision:3101–3109
85. Wan H, Wang J, Ren M et al (2009) A new robust object tracking algorithm by fusing multi-features. *Journal of Image and Graphics* 14(3):489–498
86. Wu Y, Lim J, Yang MH (2013) Online object tracking: a benchmark//IEEE Conference on Computer Vision & Pattern Recognition:2411–2418
87. Xia C, Sun S F, Chen P et al (2014) Haar-like and HOG fusion based object tracking//Advances in Multimedia Information Processing–PCM 2014. Springer International Publishing:173–182
88. Xie Z, Wei Z, Bai C (2015) Multi-aircrafts tracking using spatial–temporal constraints-based intra-frame scale-invariant feature transform feature matching. *Computer Vision, IET* 9(6):831–840
89. Xu F, Gao M (2010) Human detection and tracking based on HOG and particle filter//Image and Signal Processing (CISP), 2010 3rd International Congress on. IEEE 3:1503–1507
90. Yan H, Dechant CM, Moradkhani H (2015) Improving soil moisture profile prediction with the particle filter-markov chain Monte Carlo method. *IEEE Transactions on Geoscience & Remote Sensing* 53(11):6134–6147
91. Yang F, Lu H, Yang MH (2014) Robust superpixel tracking. *Image Processing, IEEE Transactions on* 23(4):1639–1651
92. Yi S, He Z, You X et al (2015) Single object tracking via robust combination of particle filter and sparse representation. *Signal Processing* 110:178–187
93. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. *Acm computing surveys (CSUR)* 38(4):13
94. Youm S, Liu S (2015) Development healthcare PC and multimedia software for improvement of health status and exercise habits. *Multimedia Tools and Applications*. doi:10.1007/10.1007/s11042-015-2998-2
95. Zhong W, Lu H, Yang MH (2012) Robust object tracking via sparsity-based collaborative model//Computer vision and pattern recognition (CVPR), 2012 I.E. Conference on. IEEE:1838–1845
96. Zhou Z, Zhou M, Li J (2016) Object tracking method based on hybrid particle filter and sparse representation. *Multimed Tools Appl*:1–15
97. Zhu Q, Yeh MC, Cheng KT et al (2006) Fast human detection using a cascade of histograms of oriented gradients//Computer Vision and Pattern Recognition, 2006 I.E. Computer Society Conference on. IEEE 2:1491–1498



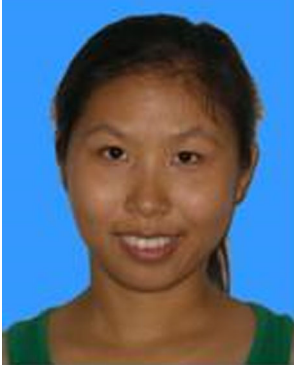
Mr Zheng Pan postgraduate, male, born in 1992. He studies in moving target tracking and recognition.



Associate Professor Shuai Liu master's supervisor, male, born in 1982, His interesting area contains applied mathematics and complex computation.

Following are his main 5 papers.

- [1] Liu S, Cheng X, LAN C, et al. Fractal property of generalized M-set with rational number exponent, *Applied Mathematics and Computation*, 2013, 220, 668-675.
- [2] S Liu, W Fu, H Deng, et al. Distributional Fractal Creating Algorithm in Parallel Environment, *International Journal of Distributed Sensor Networks*, 2013 doi:[10.1155/2013/281707](https://doi.org/10.1155/2013/281707).
- [3] Liu S, Fu W, He L, et al. Distribution of primary additional errors in fractal encoding method. *Multimedia Tools & Applications*, 2014:1-16. doi: [10.1007/s11042-014-2408-1](https://doi.org/10.1007/s11042-014-2408-1).
- [4] G Yang, S Liu. Distributed Cooperative Algorithm for k-M Set with Negative Integer k by Fractal Symmetrical Property, *International Journal of Distributed Sensor Networks*, 2014, doi: [10.1155/2014/398583](https://doi.org/10.1155/2014/398583).
- [5] S Liu, X Cheng*, W Fu, et al. Numeric characteristics of generalized M-set with its asymptote. *Applied Mathematics and Computation*, 2014, 243:767–774.



Dr. Weina Fu female, born in 1982, her interesting area contains moving recognition and multimedia encoding. Following are her main 5 papers.

- [1] Fu W, Zhou J, Liu S, et al. Differential trajectory tracking with automatic learning of background reconstruction. *Multimedia Tools & Applications*, 2014:1-13. doi:[10.1007/s11042-014-2391-6](https://doi.org/10.1007/s11042-014-2391-6)
- [2] Fu W, Zhou J, Ma Y. Moving tracking with approximate topological isomorphism. *Multimedia Tools & Applications*, 2015:1-18. doi:[10.1007/s11042-015-2519-3](https://doi.org/10.1007/s11042-015-2519-3).
- [3] Fu W, Zhou J, An C. Distributed dynamic target tracking method by block diagonalization of topological matrix. *Journal of Supercomputing*, 2015:1-18. doi:[10.1007/s11227-015-1499-4](https://doi.org/10.1007/s11227-015-1499-4)
- [4] Liu S, Fu W, Deng H, et al. Distributional Fractal Creating Algorithm in Parallel Environment. *International Journal of Distributed Sensor Networks*, 2013, doi:[10.1155/2013/281707](https://doi.org/10.1155/2013/281707).
- [5] S Liu*, W Fu, W Zhao. A Novel Fusion Method by Static and Moving Facial Capture, *Mathematical Problems in Engineering*, 2013. doi:[10.1155/2013/503924](https://doi.org/10.1155/2013/503924)