Surf points based Moving Target Detection and Long-term Tracking in Aerial Videos

Juan-juan Zhu, Wei Sun, Bao-long Guo and Cheng Li
[1] School of Aerospace science and technology, Xidian University
Xi'an 710071 China
[e-mail: zhujoo@126.com]
*Corresponding author: Juan-juan Zhu

Received January 23, 2016; revised April 27, 2016; accepted October 13, 2016; published November 30, 2016

Abstract

A novel method based on Surf points is proposed to detect and lock-track single ground target in aerial videos. Videos captured by moving cameras contain complex motions, which bring difficulty in moving object detection. Our approach contains three parts: moving target template detection, search area estimation and target tracking. Global motion estimation and compensation are first made by grids-sampling Surf points selecting and matching. And then, the single ground target is detected by joint spatial-temporal information processing. The temporal process is made by calculating difference between compensated reference and current image and the spatial process is implementing morphological operations and adaptive binarization. The second part improves KALMAN filter with surf points scale information to predict target position and search area adaptively. Lastly, the local Surf points of target template are matched in this search region to realize target tracking. The long-term tracking is updated following target scaling, occlusion and large deformation. Experimental results show that the algorithm can correctly detect small moving target in dynamic scenes with complex motions. It is robust to vehicle dithering and target scale changing, rotation, especially partial occlusion or temporal complete occlusion. Comparing with traditional algorithms, our method enables real time operation, processing 520×390 frames at around 15fps.

Keywords: Moving target detection and tracking, Background compensation, Surf feature points, KALMAN filter

This work is supported by the National Science Foundation of China (61201290) and Fundamental Research Funds for the Central Universities (JB141301, JB141307).

http://dx.doi.org/10.3837/tiis.2016.11.023 **ISSN : 1976-7277**

1. Introduction

Nowadays aerial videos processing becomes more important with the Unmanned Aerial Vehicle (UAV) developing, for example intelligent aerial traffic surveillance [1] and aerial video registration [2]. In actual environment, the complex factors including large scenes, random vehicle or target motion and target occlusion bring great challenges to stable long-term target tracking [3][4].

The aerial camera aims to follow the target actively from multi-angles and multi-directions. In target tracking applications, the existing algorithms build target models of edges, features, modeling, or their combination and then track them with optical flow[5], mean shift[6], Kalman filter or particle filter. In [6], the color texture and contour are tracked with mean shift method, but it is difficult to follow single color and small object. The KLT features [7] can track objects well in controlled environments, but they usually fail due to occlusion, large scaling, and illumination variation. The Canny edges [8] are detected with dynamic Bayesian network, which shows good vehicle detection if vehicle colors are unchanged. Miss detection and false detection are caused by low color contrast and similar rectangular structures as vehicles. The feature points including Harris [9], Susan [10] or Sift [11] are now widely used. However, the matching of Harris points or Susan points will fail due to large rotation or scale, and the computation cost of Sift points is too large to enable real time operation. The surf points demonstrate robust tracking, which is a fast growing research top.

To improve surf points matching accuracy, Miao [12] enhances repeatability by a classifier-based on-line boosting. The matching process is complex and it does not consider object occlusion. Ta [13] gives efficient surf detection inside the 3D image pyramid without computing traditional descriptor. It has limitations that the object should be initialized tracking and the tracker fails in the outdoor environment with few points.

During the long-term tracking, target scale can vary greatly as it moves toward or away from the camera. Vijay [14] combines a focus of attention mechanism. It guides tracking by visual attention with complex computation. The scale of the mean-shift kernel [15] is combined by using projective geometry of the object. The mean-shift [16] algorithm is modified by the Hellinger distance to estimate scale. Ning also proposed weighting histogram [17]. They are based on the entire features such as target contour line and color histogram, which are not robust to image occlusion. However, the temporary target occlusion and large deformation easily bring tracking failure. The local sift features [18] are chosen to represent the whole target and are particle filtered, which brings great computation cost. The new method of graph matching [19] is proposed to separate target with Markov random filed and combine the weighted graph. While many trackers [20-22] have considered occlusion and deformation, they are dedicated to recognize the current target by matching the very beginning target model. Actually, the inter-frame target deformation is not obvious, and we can discard the original template and update the newly-detected target.

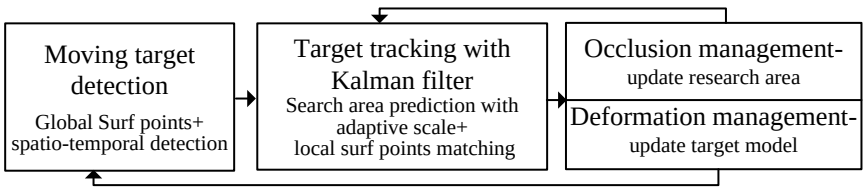


Fig. 1. Block diagram of our aerial video tracking system

Related to these studies, the key challenges of surf points tracking are to reduce time cost of object detection and improve robustness to object scaling and occlusion even when the object is very small with low image quality. Our approach uses global and local surf points in objects detection and tracking respectively, which is shown in **Fig. 1**. It aims to realize robust tracking along with unsteady background motion, object occlusion and scale changing. It includes two parts: (1) Moving target detection based on global Surf points. The proposed grid sampling surf point can prevent clustering points. The proposed distance criterion can initially delete mismatching points and classic RANSAC (Random Sample Consensus) further obtains global points to compute global motion of background. After background correction, the moving target template is detected by proposed joint spatial-temporal processing including morphological operations and adaptive binarization. (2) Moving target tracking based on local Surf points. The proposed search area prediction is realized by estimating central position of target and adjusting scale ratio. The local Surf points detected in the search area are matched with that in the target template. For the points mostly locate in the target, it can implement fast matching. In order to keep robust long-term tracking in the presence of scale changing and object occlusion, the two-layered update mechanism is proposed. When the occluded target reappears, the search area is updated as the whole image to find reappearing target. When target has significant changes with no matching points for some frames, the new target is updated by background correction to detect target.

The paper is organized as follows. Section 2 will discuss background compensation based on global Surf points for detecting single ground target. Section 3 will present moving target stable tracking based on local Surf points matching in search area. Experimental results are analyzed in section 4 and conclusion is given in section 5.

2. Moving Target Detection based on Global SURF Points

Our challenge in aerial video is how to reliably detect small moving target from complex scenes with large camera scan or jitter. Considering camera jitter, we use background compensation to warp platform motion and then detect target area roughly using temporal image difference. The target template is then obtained by spatial fine detection. The flow chart is shown in **Fig. 2**. In two successive images, global Surf points are selected in reference frame and then matched in current frame to compute motion. The reference frame is then compensated to make difference with current frame. The morphological operations are used to eliminate noise and the adaptive binarization is adopted from gray histograms. Detecting the extreme pixels of target's margins, the target template is built as the rectangle.

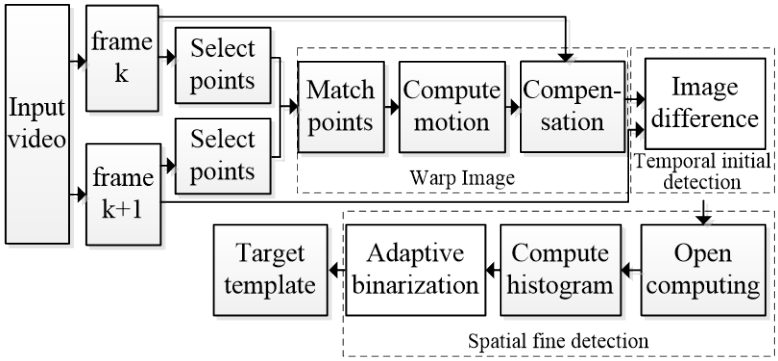


Fig. 2. Flow chart of single target detecting method

2.1 Global Surf Points Selection by Grid Sampling

The commonly used Sift feature point is robust to translation, rotation, scaling [23], but its high computational complexity decreases execution speed. Surf [24] takes advantage of Sift detector and reduces computational cost by cutting down point descriptor dimensions. Traditionally, surf points are directly detected in the whole image by checking largest Hessian values, as shown in **Fig. 3 (a)**. The number of selected points is too large, which can be reduced by modifying the Surf point contrast parameters. However, the points with high similarity still locate closely as shown in **Fig. 3 (b)**, resulting in points' redundancy.

In order to realize fast and accurate background compensation, we need to reduce points' number and extract points evenly in the background area. Therefore, the grid sampling method is proposed to delete those points that have low contrast or are cluster localized. The steps are detailed as follows.

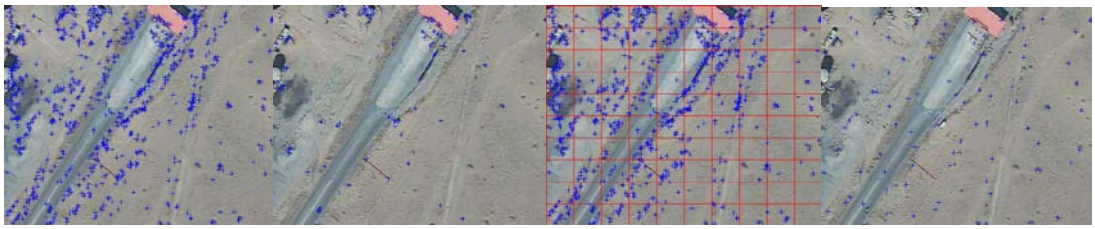
Step1: All the Surf points are initially detected in the whole image.

Step2: The image is divided into $M_1 \times M_2$ non-overlapping blocks as grids.

Step3: According to the nearest distance criterion, each Surf point is assigned to its grid.

Step4: In each grid, we get one global Surf point having the largest Hessian value.

The grid sampling method can improve points' significance and reduce amount. As shown in **Fig. 3**, **Fig. 3 (c)** shows 553 points of initial direct selection and **Fig. 3 (d)** shows 82 global points of grid sampling selection. The global points are distinctive and distribute uniformly in the background area. They are sparse but their even distribution in the whole image guarantees global motion for accurate background compensation.



(a) Direct selected points (b) Points by modification (c) Image grids (d) Evenly selected points

Fig. 3. Comparison of point selection

2.2 Global Surf Points Matching

Since the aerial video is long-range captured from air, the scene can be regarded as a plane and therefore the affine motion model [10] can describe global motion. Given the matching point P in reference frame and P' in current frame, their motion satisfies equation (1).

$$P' = \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = AP \tag{1}$$

In order to match points from coarse to fine, the distance criterion is proposed to make initial matching. And the Ransac [24] method is then used to eliminate false matches. In order to further improve matching speed, the trace of Hessian matrix is used. According to **Fig. 4**, the bright point has positive Hessian trace and the low light point has negative trace. One pair of positive and negative traces stands for mismatched points, which can be deleted

immediately. The following three steps are presented to realize points matching in complex motions, as shown in **Fig. 5**.

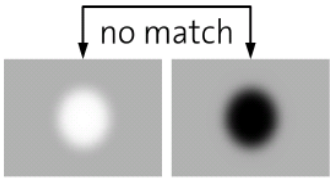


Fig. 4. Points with opposite brightness

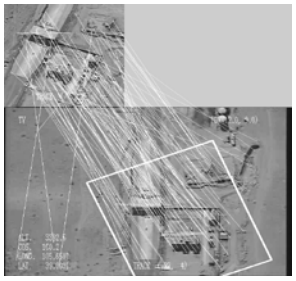


Fig. 5. Points matching results

Step1: At each Surf point in reference frame, the kd-tree storage structure is built.

Step2: In the kd-tree of point $P_i = (p_{ij})$, find the minimum Euclidean distance P_{m1} and the next nearest distance P_{m2} with point $P_i' = (p_{ij}')$ in the current frame. The pair of P_{m1} and P_i' is the approximately correct matching if their distances satisfy the following criterion:

$$\frac{d(P_{m1}, P_i')}{d(P_{m2}, P_i')} = \frac{\left[\sum_{j=1}^{64}\left(p_{ij}' - p_{m1j}\right)^2\right]^{1/2}}{\left[\sum_{j=1}^{64}\left(p_{ij}' - p_{m2j}\right)^2\right]^{1/2}} < Threshold \tag{2}$$

Step3: To further eliminate mismatching, the Ransac algorithm based on global constraint is adopted to improve matching accuracy and compute the affine matrix A.

2.3 Background Correction

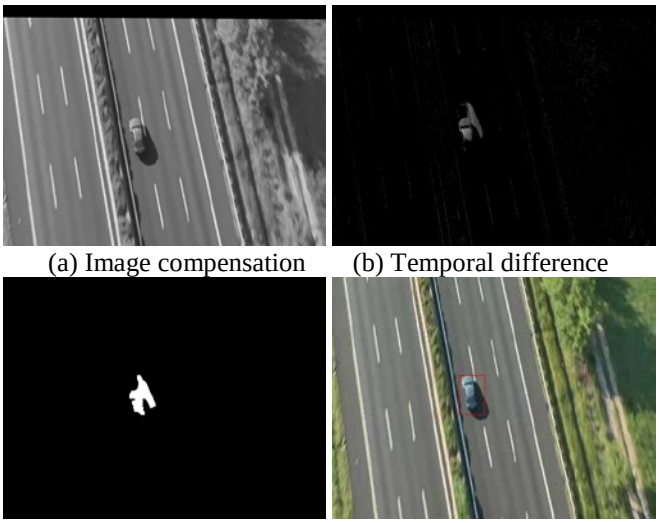
The affine matrix A represents the global motion caused by aerial platform. The background compensation is to remove camera scan or jitter. So, the matrix A is taken into affine transformation model (1) to compute new coordinates of each pixel at reference image. In real applications, we use the bilinear interpolation to determine the gray at non-integer pixels. After background correction, the foreground moving target is significantly enhanced. Although this correction is not a sophisticated process, we have found it vital for detecting small ground target. In **Fig. 6**, the target size is only few pixels and enhanced by compensated difference comparing with the direct difference.

(a)Small target (b)Direct difference (c)Compensated difference

Fig. 6. Illustration of image difference after background compensation

2.4 Joint spatial-temporal target detection

The joint spatial-temporal processing method is used to detect single ground target. After background compensation, the temporal difference is first calculated between compensated reference and current image, which can give initial result. **Fig. 7 (b)** is the temporal difference image obtained between compensated **Fig. 7 (a)** and current frame. It can be clearly seen that background compensation can remove the disturbance of camera scanning and keep the integrity of moving object. However, due to illumination variance, texture repetition or noise, the pixel difference still has noise and would result in false detection. In order to get accurate and integral target, the spatial process implements morphological operations and adaptive binarization. Fragments are removed using standard morphological dilation and erosion to capture the rough target. Despite this noise removal, small misdetections due to sensor artifacts, residual image misalignment still exist. Then the gray histogram is built to select threshold adaptively to acquire binary image **Fig. 7 (c)**. By detecting the extreme point of the outline, the rectangle in **Fig. 7 (d)** is obtained as the moving target template.



(a) Image compensation (b) Temporal difference

(c) Spatial detection (d) Moving target template

Fig. 7. Process of detecting moving target template

3. Moving Target Tracking based on Local SURF Points

Using features of color, texture, contour or shape for target modeling, the object is always tracked with Mean Shift and particle filter. For the model depicts the entire moving object, we can track it frame by frame accurately. However when the object occlusion, large rotation and scaling occur, it is difficult to achieve model matching and tracking.

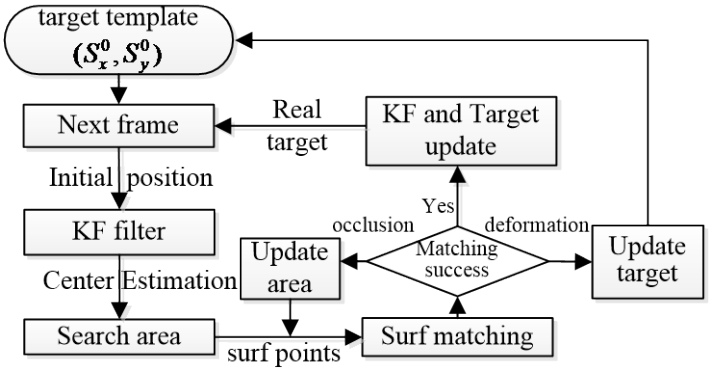


Fig. 8. Flow chart of target tracking

Considering stable points in target, local Surf points are tracked. It first gives the estimated mass center of target and the search area in next frame using KF (KALMAN filter). Then the Surf points are extracted in this search area and matched with points in target template. According to the corresponding points, the position of target center is corrected and the true target size is updated with scale information. The above process is illustrated in Fig.8. It helps increase speed by predicting new position of target center and matching local Surf points in local search area. Furthermore, if target is blocked partly, we can still track target by no less than 3 matched Surf points. When complete occlusion ends, we update search area as the whole image to track Surf points. When target has significant changes, the points matching fails in continuous frames, the background is corrected to update the new detected target.

3.1 Target Center Estimation by KF

In the long-term active tracking, the target motion is continuous and uniform. Therefore, it is reasonable that KALMAN filter can be applied to predict its path. Setting the mass center of target template as initial value, the KALMAN filter predicts the central position in next frame. The state function is described by equation (3) and (4).

$$R_k = \begin{bmatrix} x_k \\ y_k \\ \dot{x}_k \\ \dot{y}_k \end{bmatrix} = F * R_{k-1} + w = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{k-1} \\ y_{k-1} \\ \dot{x}_{k-1} \\ \dot{y}_{k-1} \end{bmatrix} + w \tag{3}$$

$$Z_k = T * R_k + v = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} R_k + v \tag{4}$$

where $R_k = [x_k, y_k, \dot{x}_k, \dot{y}_k]^T$ is the position and velocity of target center in frame k. F and T is state transition matrix and observation matrix, respectively. Z_k is the measuring position of target center, and t is time interval of two consecutive frames. w and v is process noise and measurement noise respectively.

3.2 Target search area prediction

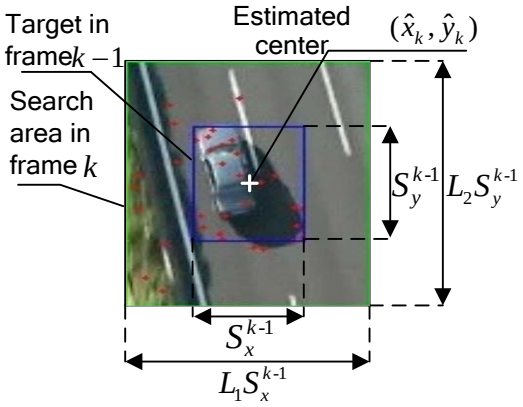


Fig. 9. Target search area prediction

KF gives the estimated center position $(\hat{x}_k, \hat{y}_k)$ of target in frame k. We define (S_x^{k-1}, S_y^{k-1}) as the scale in x and y axis of target in frame $k-1$. Supposing L_1 and L_2 is scale ratio, the search area in frame k is computed as $(L_1 S_x^{k-1}, L_2 S_y^{k-1})$, as shown in **Fig. 9**. We only match the local Surf points in this target search area rather than in the traditional whole image. Therefore, the speed improves due to fewer points in search area. Meanwhile, we also improved KALMAN filter with surf points scale information to predict search area adaptively. Two surf points nearest to the center point in the target model are matched in the current window around the predicted center point. The matched surf points are brought into equation (1), and then the affine motion matrix is computed to get the scale ratio $L_1 = L_2 = \sqrt{h_{11}h_{22} - h_{12}h_{21}}$. The width of the window proves changeable with the same height ratio. Therefore, the scale ratio is adjusted to ensure that the new search area includes moving target.

3.3 Target Tracking and KALMAN Update

The target is tracked by matching Surf points in the initial target template with those in search area frame by frame. The Surf points set is stable to describe target template and robust to transformation. The matched points can offer scaling information to predict new search area. Even if the target is partially occluded, it can still be tracked with no less than three matching points. Based on matching points, the real position of target and its center is updated for Kalman filter to locate next search area.

If tracking fails due to temporary occlusion, the two-layered update mechanism is activated. The first layer update is to extend search area to the whole image to find matched points. The second layer update is to match background points for compensated difference to detect a new target. The above long-term tracking process is illustrated as follows.

Step1: The Surf points in target template (S_x^0, S_y^0) are extracted as local features P_{target}.

Step2: According to the estimated center position $(\hat{x}_k, \hat{y}_k)$ by Kalman filter, the adaptive search area in frame k is determined.

Step3: The Surf points in search area are extracted and matched with Ransac scheme to obtain N corresponding pairs. N is compared with threshold N_{thres} to solve occlusion problem.

(3a) If $N \geq N_{thres}$, tracking is considered successful and the scale factor λ between matched target in frame k and template target is determined by equation (5). The real target size is $(S_x^k, S_y^k) = (\lambda S_x^0, \lambda S_y^0)$ and its mass center position is computed by $(x_k, y_k)=(\overline{x}_k, \overline{y}_k)$ for updating the measuring center position at frame k of Kalman filter. Return to step 2 and keep target tracking in next frame.

$$\lambda = \frac{\frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_{ki}-\overline{x}_k)^2+(y_{ki}-\overline{y}_k)^2}}{\frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_{0i}-\overline{x}_0)^2+(y_{0i}-\overline{y}_0)^2}}, \overline{x}_{0,k} = \frac{\sum_{i=1}^{N} x_{0i,ki}}{N}, \overline{y}_{0,k} = \frac{\sum_{i=1}^{N} y_{0i,ki}}{N} \tag{5}$$

(3b)If $N < N_{thres}$, points matching in frame k fails and the amendment is as follows: The estimated target central position $(\hat{x}_k, \hat{y}_k)$ is kept as measuring central position of Kalman filter and the target in frame $k-1$ is saved as the target $(S_x^k, S_y^k) = (S_x^{k-1}, S_y^{k-1})$ in frame k. Return to step2 and keep target tracking in frame $k+1$.

(3c)If $N < N_{thres}$ for some consecutive frames, target tracking fails due to temporal complete occlusion. The search area extends to the entire image in case of target reappears at a different position. We select Surf points in the whole image and match them with P_{target} to find N matched points. If $N \geq N_{thres}$, return to step (3a); if $N < N_{thres}$, the current frame might be corrupted by significant object changes, go to step (3d).

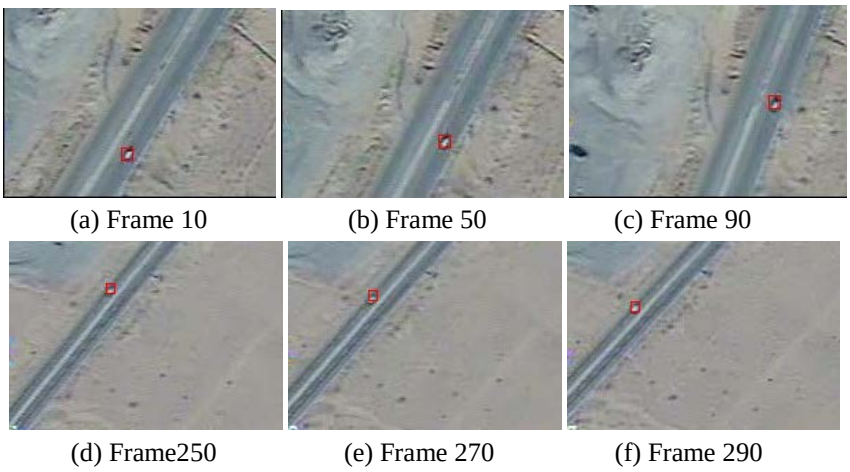
(3d) We randomly choose some Surf points around the last tracked target in frame k-1 as background points. They are matched in current frame to compute global motion and compensate background. The previously mentioned target detection in section 2.4 is employed to label a new target as model. Goto step1 to continue tracking .

4. Experimental Results and Analysis

The algorithm is tested on various aerial videos (520×390 pixels), containing translation, rotation, scaling and occlusion. Every target is identified with a rectangle. We evaluate the tracking performance in a qualitative way and perform a quantitative comparison with some classic methods.

4.1 Tracking Result of Small Object

When the target is small, accurate global motion compensation is vital to warp background for object detection. In **Fig. 10**, the tests are made to find that the minimum target size is 6×8 pixels. Comparing with the minimum size as 10×10pixels in COCOA [25] system, we can detect smaller target. The images in Fig.10 are cut from original images and enlarged locally to show small target. The small moving car is tracked correctly in sequence car1 with camera translation, rotation and zoom.



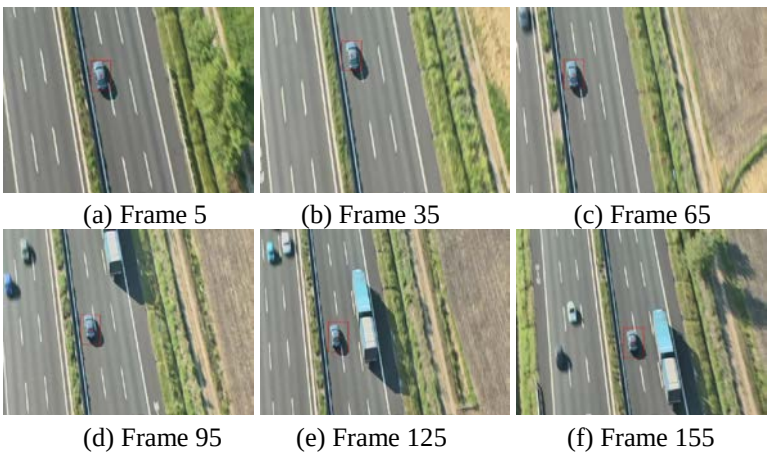
(a) Frame 10 (b) Frame 50 (c) Frame 90

(d) Frame250 (e) Frame 270 (f) Frame 290

Fig. 10. Small target tracking result in video car1

4.2 Results of Target Lock Tracking in Multi Objects

Fig. 11 shows target lock tracking results of frame 5, 35, 65, 95, 125 and 155 in video car2 from a camera mounted on an airplane. The relative motion between airplane carrier and tracked car is not stable. The background compensation removes inter-frame background jitter to obtain accurate target template. Moreover, the search area predicted by KALMAN filter ensures reliable target scope regardless of multiple targets interference.



(a) Frame 5 (b) Frame 35 (c) Frame 65

(d) Frame 95 (e) Frame 125 (f) Frame 155

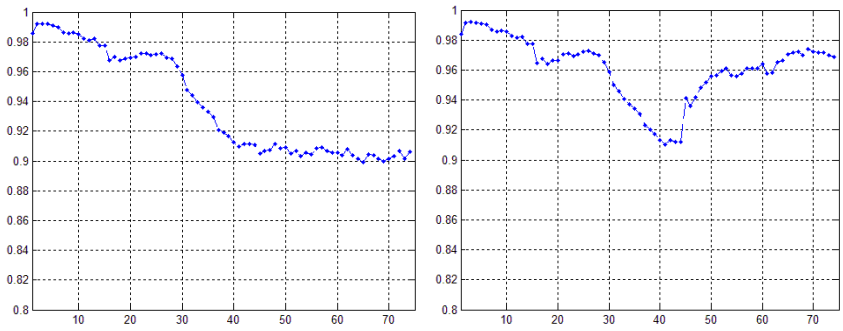
Fig. 11. Target tracking result in video car2

4.3 Tracking result of object occlusion and deformation

(a) Tracking results of Frames 70, 90,110 in video car3

(b) Tracking results of Frames 782, 813, 834 in video redcar

Fig. 12. Target tracking results in video car3 and video redcar



(a) Traditional Kalman tracking (b) Updated Kalman tracking

Fig. 13. Comparison results of Bhattacharya coefficient

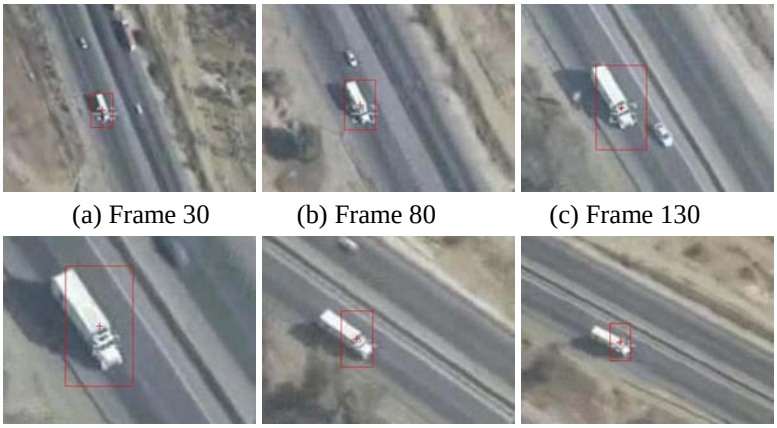
In **Fig. 12 (a)**, there exists large angle rotation and target occlusion in video car3. The property of rotation invariant Surf points ensures correct matching in the event of rotational scene. From the results, we can also see that moving object can be successfully tracked by extending search area to the whole image after the occluded target appears again. The Bhattacharyya [26] coefficient between tracked target and the target model is compared in **Fig. 13**. The curve drops when target is occluded gradually and the lowest point in the curve stands for complete occlusion. Comparing with continuous low curve in **Fig. 13 (a)** of the traditional KALMAN filter, the curve rises up from the bottom in **Fig. 13 (b)**. It verifies that the update mechanism can find target again when complete occlusion ends.

In **Fig. 12 (b)**, there exists deformation in video redcar. When the redcar turns around, almost all the points can not be matched. The target is newly detected to update target template. The Surf points are selected as features to be tracked in the sequence. This update mechanism is easy and effective, which prevents complex matching with deformed target.

4.4 Tracking Result of Object Scaling

Fig. 14 exhibits results of target tracking in video car4 with object scaling. According to the position and scale of the matched Surf points, the state of KALMAN filter is updated to give a size-scaling search area. The correct matching of Surf points is achieved by its scaling

invariance, which provides basis for target size changing. It can be seen that the real matched target is tracked with adaptive scale when the camera zooms in or out.

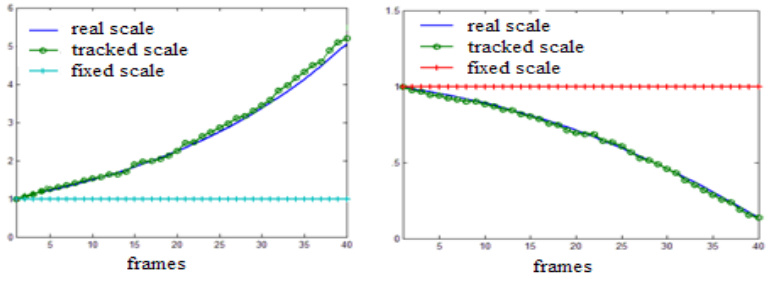


(a) Frame 30 (b) Frame 80 (c) Frame 130

(d) Frame 180 (e) Frame 230 (f) Frame 280

Fig. 14. Target tracking result in video car4

We choose two sequences with target scaling to testify the adaptation. **Fig. 15** shows the comparisons between tracked target scale and the real scale increasing and decreasing. We label the target frame by frame to give the real size and the tracked target size is approximate with the real size, which is realized by predicting search area adaptively according to motion ratio. The traditional KALMAN filter uses the fixed scale, which is not applicable to target scaling videos.



(a)Scale increasing (b) Scale decreasing

Fig. 15. Result of tracking target scaling

4.5 Analysis of Average Time and Quantitative Performance

Table 1 shows the analysis of computation time of the proposed algorithm. The target detection methods based on Sift points, traditional direct Surf points and grid sampling Surf points are compared. The time cost of Surf points detection and matching reduces 77% comparing with Sift points, and the grid sampling method shows 60% improvement in speed by contrast with direct method. In target tracking process, the proposed local points in predicted search area by KALMAN filter can realize target tracking at 15fps. This is because KALMAN filter gives an approximate area for target match and local Surf points are stable to be tracked with lower computation complexity.

Table 1. Comparison of average time

Processes	Methods	Detection(ms)	Description (ms)	Match(ms)	Sum (ms)
Target detection	Sift[11]	248	312	536	1096
	Direct surf[12]	133	197	281	621
	Grid Surf	142	65	40	247
Target tracking	Contour[6]	114	157	103	374
	KLT[7]	77	94	82	253
	Whole search	132	199	78	409
	Local search	32	22	16	70

Table 2 shows the quantitative evaluation of the proposed algorithm. Probability of Detection (PD) measure provides vital insights on the performance of the detection module. Tests are made in all the above videos car1 to car4. PD is computed as the rate between the numbers of correct detection with real target number. If the detected target is not complete or too large, the detection is wrong. False Tracking Rate (FTR) is presented for each of these sequences to measure the performance of the tracking module. If false target occurs or target is missing due to partial occlusion and reappearance, the track result is false. As can been seen from table 2, the proposed method achieves better PD especially in car1 because global motion compensation corrects background and the joint spatial-temporal processing can detect small target. The search area prediction and KALMAN update reduce rate of false or missing target, which shows the lowest FTR value in car3 with occlusion.

Table 2. Comparison of tracking reliability

Methods	PD				FTR			
	Car1	Car2	Car3	Car4	Car1	Car2	Car3	Car4
Sift[11]	0.83	0.88	0.86	0.92	0.23	0.10	0.18	0.09
Surf[12]	0.81	0.87	0.87	0.92	0.23	0.09	0.19	0.08
Contour[6]	0.75	0.76	0.72	0.73	0.26	0.12	0.25	0.18
KLT[7]	0.74	0.78	0.81	0.83	0.24	0.08	0.22	0.17
Proposed	0.92	0.89	0.91	0.90	0.11	0.09	0.06	0.10

5. Conclusion

This paper presents a new aerial-to-ground target detection and tracking algorithm based on Surf point. It selects evenly distributed global Surf points to improve global estimation, compensation and target template detection. The local Surf points in the target template are then tracked in KALMAN estimated search area to achieve accurate target tracking. Experimental results show that the algorithm is robust to aerial video jitter, large rotation and scale changing, irregular movement and occlusion of target. In the future, we will study the multi-target [27] tracking, and image mosaic to demonstrate the trajectory of the whole scene image.

References

[1] Xianbin Cao, Jinhe Lan, Pingkun Yan, Xuelong Li, "KLT Feature Based Vehicle Detection and Tracking in Airborne Videos," in *Proc. of Sixth International Conference on Image and Graphics*, pp.673-678, 2011. Article (CrossRef Link).

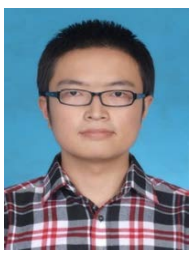
[2] E. Molina, Z. Zhu, "Persistent Aerial Video Registration and Fast Multi-View Mosaicing," *IEEE Trans. on Image Processing*, vol.23, no.5, pp.2184-2192, 2014.Article (CrossRef Link).
[3] Peter Reinartz, Marie Lachaise and Elisabeth Schmeer, Thomas Krauss, Hartmut Runge, "Traffic monitoring with serial images from airborne cameras," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol.61,pp.149-158, 2006. Article (CrossRef Link).
[4] Subhabrata Bhattacharya, Haroon Idrees, Imran Saleemi, Saad Ali, "Moving Object Detection and Tracking in Forward Looking Infra-Red Aerial Imagery," *Machine Vision Beyond Visible Spectrum Augmented Vision and Reality*, Vol.21, pp.221-252, 2011. Article (CrossRef Link).
[5] Yunfei Wang, Zhaoxiang Zhang, Yunhong Wang, "Moving Object Detection in Aerial Video," in *Proc. of International Conf. on Machine Learning and Applications*, vol.2, pp.446-450, 2012. Article (CrossRef Link).
[6] Ido Leichter, Michael Lindenbaum, Ehud Rivlin, "Mean Shift tracking with multiple reference color histograms," *Computer Vision and Image Understanding*, Vol.114, pp.400-408, 2010. Article (CrossRef Link).
[7] Myung Hwangbo, Jun-Sik Kim, and Takeo Kanade, "Inertial-Aided KLT Feature Tracking for a Moving Camera," *Intelligent Robots and Systems*, pp.909-916, 2009. Article (CrossRef Link).
[8] Hsu-Yung Cheng, Chih-Chia Weng, Yi-Ying Chen, "Vehicle detection in aerial surveillance using dynamic Bayesian networks," *IEEE Trans. on Image Processing*, vol.21, no.4, pp.2152-2159, 2012. Article (CrossRef Link).
[9] Yingqian YANG, Fuqiang LIU, "Vehicle detection methods from an unmanned aerial vehicle platform," in *Proc. of IEEE International Conference on Vehicular Electronics and Safety*, pp.411-416, 2012. Article (CrossRef Link).
[10] Xu Zhang, Baolong Guo, Yunyi Yan, Wei Sun, Meng Yi, "Image Retrieval Method Based on IPDSH and SRIP," *KSII Transactions on Internet and Information Systems*, vol.8, no.5, pp.1676-1689, 2014. Article (CrossRef Link).
[11] Szottka, I. and Butenuth, M, "Tracking multiple vehicles in airborne image sequences of complex urban environments," *Urban Remote Sensing Event*, pp. 13-16, 2011. Article (CrossRef Link).
[12] Q. Miao, G. Wang, C. Shi, X Lin, Z Ruan, "A new framework for on-line object tracking based on SURF," *Pattern Recognition Letters*, vol.32, pp.1564-1571, 2011. Article (CrossRef Link).
[13] D.N. Ta, W.C. Chen, N. Gelfand, and K. Pulli, "SURF Trac: Efficient tracking and continuous object recognition using local feature descriptors," *CVPR*, pp. 2937-2944, 2009. Article (CrossRef Link).
[14] Vijay Mahadevan and Nuno Vasconcelos, "Biologically Inspired Object Tracking Using Center-Surround Saliency Mechanisms," *IEEE Transactions on Pattern Analysisand Machine Intelligence*, vol.35, no.3, pp.541-554, 2013. Article (CrossRef Link).
[15] Zhongyu Lou, Guang Jiang, ChengkeWu, "2D scale-adaptive tracking based on projective geometry," *Multimed Tools Appl*, vol.72, pp.905–924, 2014. Article (CrossRef Link).
[16] Tomas Vojir , Jana Noskova, Jiri Matas, "Robust scale-adaptive mean-shift fortracking," *Pattern Recognition Letters*, vol.49, pp.250–258, 2014. Article (CrossRef Link).
[17] J. Ning, L. Zhang, D. Zhang, "Robust mean-shift tracking with corrected background-weighted histogram," *IET Computer Vision*, vol.6, pp.62–69, 2012. Article (CrossRef Link).
[18] Li Sun and Guizhong Liu, "Visual Object Tracking Based on Combination of Local Description and Global Representation," *IEEE Transactionson Circuitsand Systems for Video Technology*, vol.21, no.4, pp.408-420, April 2011. Article (CrossRef Link).
[19] Zhaowei Cai, Longyin Wen, Zhen Lei, Nuno Vasconcelos, and Stan Z. Li, "Robust Deformable and Occluded Object Tracking With Dynamic Graph," *IEEE Transactions on Image Processing*, vol.23, no.12, pp.5497-5509, December, 2014. Article (CrossRef Link).
[20] Bouachir, W., Bilodeau, G.-A, "Structure-aware keypoint tracking for partial occlusion handling," *IEEE Conf. on Applications of Computer Vision (WACV)*, pp.877–884, 2014. Article (CrossRef Link).
[21] Zarezade, A., Rabiee, H.R., Soltani-Farani, A., Khajenezhad, A., "Patchwise Joint Sparse Tracking With Occlusion Detection," *IEEE Transactions on Image Processing*, vol.23, no.10, pp.4496–4510, 2014. Article (CrossRef Link).

[22] Bing-Fei Wu, Chih-Chung Kao, Cheng-Lung Jen, Yen-Feng Li, Ying-Han Chen, Jhy-Hong Juang, "A Relative-Discriminative-Histogram-of-Oriented-Gradients-Based Particle Filter Approach to Vehicle Occlusion Handling and Tracking," *IEEE Transactions on Industrial Electronics*, vol.61, no.8, pp.4228–4237, 2014. Article (CrossRef Link).
[23] D.G. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, Vol.60, pp.91-110, 2004. Article (CrossRef Link).
[24] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *European Conference on Computer Vision*, pp.404-417, 2006. Article (CrossRef Link).
[25] Saad Ali, Mubarak Shah, "COCOA-Tracking in Aerial Imagery," *SPIE Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications*, Orlando, 2006. Article (CrossRef Link).
[26] D. Comaniciu, V. Ramesh, P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," in *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, South Carolina, Vol. 2, pp.142-149, 2000. Article (CrossRef Link).
[27] Y Chai, H Hong, T Kim, "Disjoint Particle Filter to Track Multiple Objects in Real-time," *KSII Transactions on Internet and Information Systems (TIIS)*, Vol.8, No.5, pp.1711-1725, May 2014. Article (CrossRef Link).

Juanjuan Zhu received her doctor degree in Electronics and Systems from Xidian University, China in 2009. She now is working as an associate professor in School of Aerospace science and technology at Xidian University. Her research interests are computer vision, visual information processing, analysis and identification.

Wei Sun is an associate professor in School of Aerospace science and technology at Xidian University. His research interests are pattern recognition and intelligent information processing.

Baolong Guo is a professor in School of Aerospace science and technology at Xidian University. He received his B.S., M.S. and Ph.D degrees from Xidian University in 1984, 1988 and 1995, respectively, all in communication and electronic system. From 1998 to 1999, he was a visiting scientist at Doshisha University in Japan. His research interests include pattern recognition, intelligent information processing, image processing and video communication.



Cheng Li is currently studying in the Intelligent Control and Image Engineering Institute for his master degree. He focuses on researching in moving object tracking and video surveillance.