OBJECT TRACKING USING IMPROVED SPATIO-TEMPORAL CONTEXT WITH KALMAN FILTER

Zhou Zhao[1], *Panfeng Huang[2], Lu Chen[3] and Zhenyu Lu[4]

[1]*Research Center for Intelligent Robotics and National Key Laboratory of Aerospace Flight Dynamics, School of Astronautics, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an Shaanxi, P.R.China, E-mail:* zhaozhou@mail.nwpu.edu.cn

[2]*Research Center for Intelligent Robotics and National Key Laboratory of Aerospace Flight Dynamics, School of Astronautics, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an Shaanxi, P.R.China, E-mail:* pfhuang@nwpu.edu.cn

[3]*Research Center for Intelligent Robotics and National Key Laboratory of Aerospace Flight Dynamics, School of Astronautics, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an Shaanxi, P.R.China, E-mail:* chenlu11@mail.nwpu.edu.cn

[4]*Research Center for Intelligent Robotics and National Key Laboratory of Aerospace Flight Dynamics, School of Astronautics, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an Shaanxi, P.R.China, E-mail:* luzhenyu@mail.nwpu.edu.cn

ABSTRACT

To increase the robustness of satellites on-orbit self-servicing and solve the severe occlusion, pose change and illumination variation problems of the moving target in space, a method based on the combination of improved spatio-temporal context (ISTC) and Kalman filter is proposed. In this method, ISTC is used to track the interested object accurately and steadily. But this algorithm will lose the tracking target once the Euclidean distance between the image intensity of the object in two consecutive frames is greater than the threshold. Thus we apply Kalman filter to predict and estimate the possible location of the target under severe occlusion. These characteristics make this method suitable for object tracking, and make up for the deficiency of STC method that the object will lose under severe occlusion and high speed. The experimental results show that the ISTC method with Kalman filter is feasible and effective when the tracking process exists some challenges of occlusion, pose change and illumination variation. In image sequences, the information of the moving object in the current frame is obtained through the tracking method, which is better beneficial to the surveillance of the object in the process of satellites on-orbit self-servicing.

*Corresponding author

1 INTRODUCTION

In recent years, the development of space robot technology shows that its great potential in the satellites self-servicing. Computer vision plays a very important role in space robot technology with numerous applications (e.g., measurement, positioning, auxiliary operation, object tracking, human-computer interaction and surveillance), and our main mission in this paper is to solve the problem of the tracking process of the moving object. If the satellite is able to maintain on-orbit or be serviced by itself, not only its performance can be enhanced in space systems and the life of the satellite can be extended, but also the substantial costs and the growing problem of space junk can be reduced. Automatic object tracking in image sequences is still a challenge in computer vision due to many factors that include occlusion, pose change, illumination variation, complex motion, and background clutter, although some progress have been completed in the past decades, developing a robust tracking algorithm is indispensable so that the object tracking is better applied to space robot.

To tackle these problems in the tracking process, numerous algorithms have been presented by researcher from all over the world for object tacking, which rely on template matching [1]-[2], small patch tracking [3]-[4], particle filters [5]-[6], sparse representations [7], contour modeling [8] or image segmentation [9]. For example, an algorithm using Kanade- Lucas-Tomasi (KLT) and Kalman filter is proposed in [10], and this method applies KLT to track object and then predicts and estimates the best tracked patch of KLT tracking results through Kalman filter. Wang [11] proposes a method which combines a texture-based tracking method and the Kalman filter, and this method first estimates the object pose in the current frame and predicts the target pose in the next frame through Kalman filter. In order to improve the stability of the Kalman filter, [12] proposes a method based on centroid weighted Kalman filter (CWKF) for object tracking, and this method firstly uses background subtraction method to search for the region of the moving target, and then combines centroid weighted method and Kalman filter to optimize the predictive state value. [13] presents an improved visual saliency model and combines it with the particle filter to solve a problem of abrupt motion.

This paper presents a tracking method that combines improved spatio-temporal context and Kalman filter. The main contributions are as follows. Firstly, we give an improved spatio-temporal context that achieves better results in the terms of accuracy, robustness and speediness than spatio -temporal context (STC), and make up for the deficiency of STC method that the object will lose under severe occlusion and high speed. Secondly, we give a condition to judge severe occlusion in the tracking process and solve the problem of severe occlusion. If the Euclidean distance between the image intensity of the tracking object in two consecutive frames is greater than the threshold that is varied for different tracking objects, and the center position of the tracking object in the current frame and former frame remains unchanged simultaneously, we affirm that the object exists severe occlusion. The problem of severe occlusion in the tracking process is solved by Kalman filter that is used to predict and estimate the possible location of the target under severe occlusion, and object's position is updated in the progress of prediction.

The rest of the paper is organized as follow. Section 2 and 3 simply introduce the spatio-temporal context method for tracking and the theory of Kalman filter respectively. In section 4, we explain our method in details. Section 5 gives experimental results and analysis. Finally, we conclude in section 6.

2 SPATIO-TEMPORAL CONTEXT (STC) FOR TRACKING

The STC algorithm [14] formulates the spatio-temporal relationships between the object and its local context based on a Bayesian framework by

$$\begin{aligned} m(\boldsymbol{x}) &= P(\boldsymbol{x} \mid o) \\ &= \textstyle\sum_{\boldsymbol{m}(\boldsymbol{r})\in X^r} P(\boldsymbol{x},\boldsymbol{m}(\boldsymbol{r}) \mid o) \\ &= \textstyle\sum_{\boldsymbol{m}(\boldsymbol{r})\in X^r} P(\boldsymbol{x} \mid \boldsymbol{m}(\boldsymbol{r}),o)P(\boldsymbol{m}(\boldsymbol{r}) \mid o) \end{aligned} \tag{1}$$

where $X^r = \left\{\boldsymbol{m}(\boldsymbol{r}) = \left(I(\boldsymbol{r}),\boldsymbol{r}\right) \mid \boldsymbol{r} \in \Omega_m(\boldsymbol{x}^*)\right\}$ denotes the context feature set, $m(\bullet)$ denotes the object location likelihood function, o denotes the object present in the current frame and $\boldsymbol{x} \in R^2$ is a target position, $I(\boldsymbol{r})$ is image intensity at position $\boldsymbol{r}$, $\boldsymbol{x}^*$ denotes the center position of the tracking object, $\Omega_m(\boldsymbol{x}^*)$ is the region around the center location $\boldsymbol{x}^*$, $P(\boldsymbol{x} \mid \boldsymbol{m}(\boldsymbol{r}),o)$ is the conditional probability which models the spatial relationship between the local surrounding background of the target and the target position, and $P(\boldsymbol{m}(\boldsymbol{r}) \mid o)$ denotes a context prior probability.

This method solves the tracking problem through a way that the best object location is obtained by computing a confidence map and maximizing a target position likelihood function. Although the STC algorithm is robust to appearance variations introduced by partial occlusion, pose and illumination variations, this method has a fatal problem that the object will lose under severe occlusion, high speed and high maneuver.

3 KALMAN FILTER FOR TRACKING

Kalman filter [15] is used as a target state estimator, and is also acquired as a predictor of the next target' position. In this case, Kalman filter can predict the next object' position under severe occlusion. The standard Kalman filter can be represented by transition and observation model as shown in the following equations:

$$\boldsymbol{x}_{t+1} = \boldsymbol{\phi}\boldsymbol{x}_t + \boldsymbol{w}_t \tag{2}$$

and

$$\boldsymbol{Y}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{V}_t \tag{3}$$

where $\boldsymbol{x}_t$ denotes state vector at time t^{th} sample, $\boldsymbol{\phi}$ is a state transition matrix , $\boldsymbol{w}_t$ denotes process noise taking into account the perturbations to the system, $\boldsymbol{H}$ is an observation matrix, $\boldsymbol{V}_t$ denotes measurement noise, and $\boldsymbol{Y}_t$ is a measurement vector obtained from the sensor.

4 IMPROVED SPATIO-TEMPORAL CONTEXT (ISTC) ALGORITHM WITH KALMAN FILTER

The STC tracker has a smooth tracking on the object, but this method will lose object under severe occlusion, high speed and high maneuver. These problems can be solved by ISTC method with Kalman filter as a predictor.

Figure 1 illustrates the basic flow of our method. First, the tracking window is given by manual label at the first frame and the target is tracked by ISTC method under no occlusion and partial occlusion. Next, if the tracking process of object exists severe occlusion that the object position remains unchanged at the t-th and (t-1)-th frames and the Euclidean distance between the image intensity of the tracking object of the (t-1)-th and (t-2)-th frames more than the threshold, Kalman filter is applied to predict the object location at the next frame, otherwise the object is still tracked by ISTC.

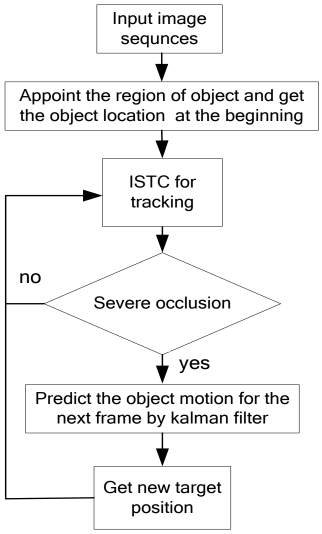


Figure 1: Flow of ISTC with Kalman filter

4.1 Improved Spatio-temporal Context (ISTC) Algorithm

We redefine the conditional probability which models the spatial relationship between the local surrounding background of the target and the target position. In order to better suitable for the scale change of the tracking object in the tracking process, Gaussian function is used to substitute for the weighted function in STC method. And the context prior probability is simply modeled by

$$P\left(\boldsymbol{m}(\boldsymbol{r}) \mid o\right) = I(\boldsymbol{r}) \cdot \frac{1}{2\pi\beta^2} \cdot e^{-\frac{\left|\boldsymbol{r}-\boldsymbol{x}^*\right|^2}{\beta^2}} \tag{4}$$

where β is a scale parameter of the tracking target.

To solve the losing problem of high speed and high maneuvering target in the tracking process, the hunting zone of the position of the tracking object will be broadened. Figure 2 illustrates the method of obtaining $\boldsymbol{x}_t^*$ .The black rectangle denotes the position of the tracking window in the (t-1)-th frame, and $\boldsymbol{x}_{t\text{-}1}^*$ has been obtained. In the t-th frame, the center location of the tracking window will be moved to nine positions, and the confidence map of the nine positions can be calculated respectively. Then compared the confidence map of the nine positions, we can obtain the maximum of confidence map at nine positions. And the center location of the target in the t-th frame can be obtained with:

$$\boldsymbol{x}_t^{*} = \underset{\boldsymbol{x} \in \Omega_m(\boldsymbol{x}_{t-1}^{*})}{\operatorname{argmax}} \; m_t^i(\boldsymbol{x}), i = 1, \cdots, 9 \tag{5}$$

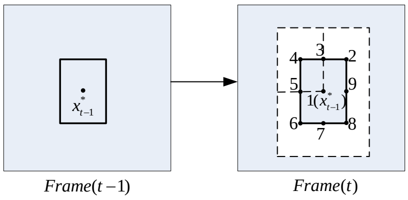


Figure 2: the method of obtaining the center position of the tracking object

4.2 Solve Severe Occlusion

The Euclidean distance between the image intensity of the tracking object of the (t-1)-th frame and (t-2)-th frame is given by

$$d_E(I_{t-1}(\boldsymbol{r}), I_{t-2}(\boldsymbol{r})) = \sqrt{\sum_{r_i \in \boldsymbol{r}} (I_{t-1}(r_i) - I_{t-2}(r_i))^2} \tag{6}$$

If $d_E(\bullet) > threshold$ and $\boldsymbol{x}_t^{*}$ remains unchanged, the tracking object may undergoes a severe occlusion, hence let $\boldsymbol{x}_t^{*}$ substitute for $\hat{\boldsymbol{x}}_{t|t}$ in (7).

$$\hat{\boldsymbol{x}}_{t+1|t} = \boldsymbol{\phi} \hat{\boldsymbol{x}}_{t|t} \tag{7}$$

The $\boldsymbol{P}_{t+1|t}$ can be derived by

$$\boldsymbol{P}_{t+1|t} = \boldsymbol{\phi} \boldsymbol{P}_{t|t} \boldsymbol{\phi}^T + \boldsymbol{Q} \tag{8}$$

where $\boldsymbol{P}_{t|t}$ denotes a covariance matrix in the t-th frame, and $\boldsymbol{Q}$ is a process noise covariance matrix, and here we assume $\boldsymbol{Q}$ is a constant.

The values of $\boldsymbol{P}_{t+1|t+1}$ and $\hat{\boldsymbol{x}}_{t+1|t+1}$ can be obtained with

$$\boldsymbol{P}_{t+1|t+1} = \boldsymbol{P}_{t+1|t} - \boldsymbol{K}_{t+1} \boldsymbol{H} \boldsymbol{P}_{t+1|t} \tag{9}$$

and

$$\hat{\boldsymbol{x}}_{t+1|t+1} = \hat{\boldsymbol{x}}_{t+1|t} + \boldsymbol{K}_{t+1}(\boldsymbol{Y}_{t+1} - \boldsymbol{H} \hat{\boldsymbol{x}}_{t+1|t}) \tag{10}$$

where $\boldsymbol{K}_{t+1}$, the so called Kalman gain matrix, is determined with

$$\boldsymbol{K}_{t+1} = \boldsymbol{P}_{t+1|t} \boldsymbol{H}^T (\boldsymbol{H} \boldsymbol{P}_{t+1|t} \boldsymbol{H}^T + \boldsymbol{R})^{-1} \tag{11}$$

where $\boldsymbol{R}$ denotes an observation noise covariance matrix, and here we assume it is a constant.

Here $\hat{\boldsymbol{x}}_{t+1|t+1}$ is the new location, and substitutes for the old position $\boldsymbol{x}_t^{*}$:

$$\underset{\boldsymbol{x} \in \Omega_m(\boldsymbol{x}_t^{*})}{\operatorname{argmax}} \; m_{t+1}(\boldsymbol{x}) = \hat{\boldsymbol{x}}_{t+1|t+1} \tag{12}$$

Then $\hat{\boldsymbol{x}}_{t+1|t+1}$ is applied to calculate the center location of the object $\boldsymbol{x}_{t+2}^{*}$ at the next frame. If $\hat{\boldsymbol{x}}_{t+1|t+1} = \boldsymbol{x}_{t+2}^{*}$, this method continues to go (7)~(12) for predicting the center location of the object at the next frame, otherwise this method goes the process of ISTC.

5 EXPERIMENTAL RESULTS

In order to verify this method's feasibility and effectiveness, this paper has executed a tracking experiment. In the experiment, the computer is configured to 2.31GHz machine with 1GB RAM. The experimental location is different environment. The experiments are carried out in MATLAB (2012b). We set this system' state transition matrix and observation matrix, i.e.

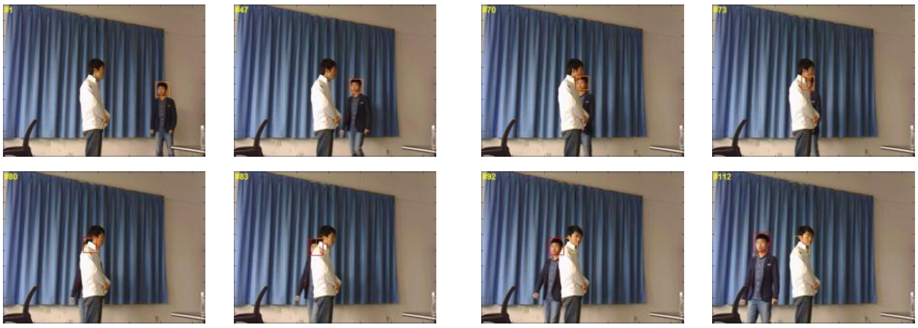
$$\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The tracking target in the *severe occlusion* sequence in fig.3a undergoes severe occlusion. Overall, our method performs well on this sequence. Frames #70, #73, #80 and #83 show examples where the object is partially and fully occluded. Our method solves the problem of severe occlusion, and makes up for the deficiency of STC method.

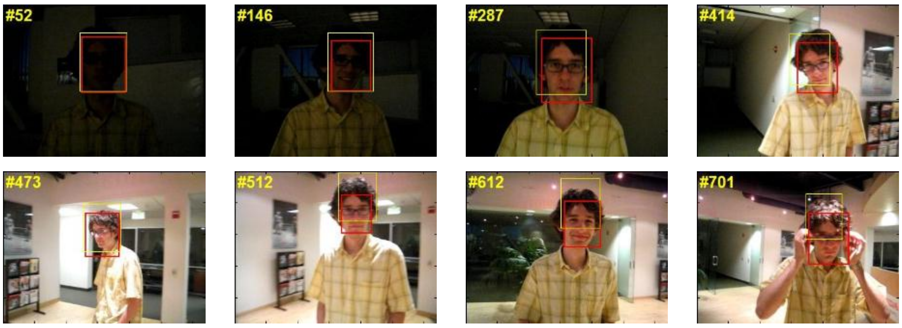
For the *David indoor* sequence shown in Fig.3b, the appearance changes gradually due to pose and illumination variations when the person walks on the meeting room. The STC method gradually drifts away from the target after frame #287, but our method is able to track the object reliably. We also evaluate our method in terms of handling scale change. The object undergoes scale change due to the man movements, and our method can suitable for scale change in the experiment. Overall, the proposed tracker is robust to pose and illumination changes.

The target in the *high speed and high maneuvering target* sequence in fig.3c undergoes abrupt movements with 360 degree out-of-plane rotation and high speed. Our method performs well on this sequence, but the STC method cannot handle high speed and high maneuvering object well as illustrated by frames after frame #5. Especially in the frames #58 and #63, the scale of the tracking window of the STC method suddenly becomes bigger and bigger. Finally, the size of the tracking window is greater than the size of the background image, and the STC method fails to track the target.

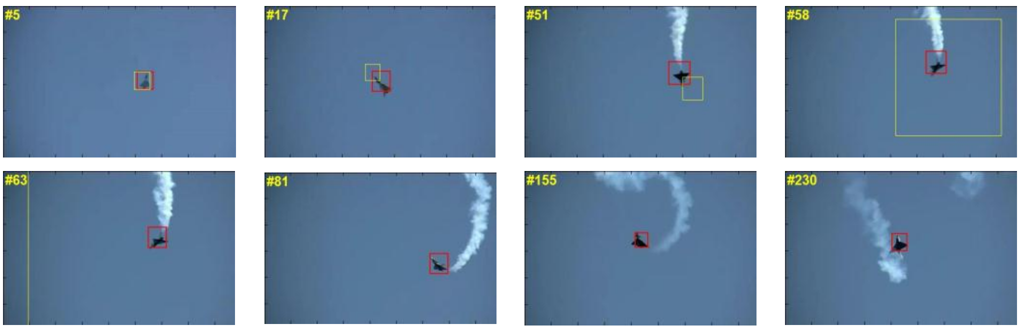
These experimental results show that the ISTC method with Kalman filter is feasible and effective when the tracking process undergoes some challenges of occlusion, pose change and illumination variation.



(a) Tracking result of the *severe occlusion* sequence



(b) Tracking result of the *David indoor* sequence



(c) Tracking result of the *high speed and high maneuvering target* sequence

STC ISTC with Kalman filter

Figure 3: Screenshots of some sample tracking results when there are severe occlusion, pose variations, severe illumination changes, high speed and high maneuver

6 CONCLUSION

In this paper, a novel framework for object tracking based on ISTC trackers and Kalman filter has been proposed. ISTC is acquired to track object of interest and Kalman filter is used to predict and estimate the possible location of the target under severe occlusion. In the experiment, this method shows real-time tracking stability when the object undergoes severe occlusion, pose change and illumination variation.

Acknowledgement

This research is sponsored by the National Natural Science Foundation of China (Grant No: 11272256, 61005062 and 60805034) and the Fundamental Research Funds for the Central Universities (Grant No:3102015BJ006).

References

[1] Zimmermann Karel, Matas Jirí (2009) Svoboda Thomás. Tracking by an Optimal Sequence of Linear Predictors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 31(4):677-692.

[2] Ellis Liam, Dowson Nicholas, Matas Jiri, et al (2011) Linear Regression and Adaptive Appearance Models for Fast Simultaneous Modelling and Tracking. *International Journal of Computer Vision* 95(2):154-179.

[3] Kalal Zdenek, Mikolajczyk Krystian, Matas Jiri (2010) Forward-backward Error: Automatic Detection of Tracking Failures. In: *2010 20th International Conference on Pattern Recognition (ICPR)*. Istanbul, pp.2756-2759.

[4] Zhang Lu, van der Maaten, Laurens JP (2014) Preserving Structure in Model-free Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 36(4):756-769.

[5] Tzimiropoulos Georgios, Zafeiriou Stefanos, Pantic Maja (2011) Sparse Representations of Image Gradient Orientations for Visual Recognition and Tracking. In: *2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Colorado Springs, pp.26-33.

[6] Liwicki Stephan, Zafeiriou, Stefanos, Tzimiropoulos, Georgios, et al (2012) Efficient Online Subspace Learning with an Indefinite Kernel for Visual Tracking and Recognition. *IEEE Transactions on Neural Networks and Learning Systems* 23(10):1624-1636.

[7] Mei Xue, Ling Haibin (2011) Robust Visual Tracking and Vehicle Classification via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 33(11):2259-2272.

[8] Horbert Esther, Rematas Konstantinos, Leibe Bastian (2011) Level-set Person Segmentation and Tracking with Multi-region Appearance Models and Top-down Shape Information. In: *2011 IEEE International Conference on Computer Vision (ICCV)*. Barcelona, pp.1871-1878.

[9] Nicolas Papadakis, Aurélie Bugeau (2011) Tracking with Occlusions via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 33(1):144-157.

[10] Pouya Bagherpour, Seyed Ali Cheraghi, Musa bin Mohd Mokji (2012) Upper Body Tracking Using KLT and Kalman Filter. *Procedia Computer Science* 13:185-191.

[11] Wang Yu, Liu Gang (2011) Head Pose Estimation Based on Head Tracking and the Kalman Filter. *Physics Procedia* 22:420-427.

[12] Fu Zhaoxia, Han Yan (2012) Centroid Weighted Kalman Filter for Visual Object Tracking. *Measurement* 45(4):650-655.

[13] Su Yingya, Zhao Qingjie, Zhao Liujun, et al (2014) Abrupt Motion Tracking Using a Visual Saliency Embedded Particle Filter. *Pattern Recognition* 47(5):1826-1834.

[14] Zhang Kaihua, Zhang Lei, Liu Qingshan, et al (2014) Fast Tracking via Dense Spatio-temporal Context Learning. In: *European Conference on Computer Vision*, Zurich, pp.127-141.

[15] Chen S. Y. (2012) Kalman Filter for Robot Vision: a Survey. *IEEE Transactions on Industrial Electronics* 59(11): 4409-4420.