World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:5, No:11, 2011

A Novel Tracking Method Using Filtering and Geometry

Sang Hoon Lee, Jong Sue Bae, Taewan Kim, Jin Mo Song, Jong Ju Kim

Abstract—Image target detection and tracking methods based on target information such as intensity, shape model, histogram and target dynamics have been proven to be robust to target model variations and background clutters as shown by recent researches. However, no definitive answer has been given to occluded target by counter measure or limited field of view(FOV). In this paper, we will present a novel tracking method using filtering and computational geometry. This paper has two central goals: 1) to deal with vulnerable target measurements; and 2) to maintain target tracking out of FOV using non-target-originated information. The experimental results, obtained with airborne images, show a robust tracking ability with respect to the existing approaches. In exploring the questions of target tracking, this paper will be limited to consideration of airborne image.

Keywords—Tracking, Computational geometry, Homography, Filter

I. Introduction

IMAGE target detection and tracking have been studied extensively during the past several decades. It can be used for various application such as unmanned surveillance system, target tracking in precision guided missile, face tracking, etc. Tracking methods can be classified into two main groups[1]. The first group is *filtering and data association*[2][3] as a top-down approach. In this method, objects detected in each image are used as measurements in the filter framework. Targets are modeled as dynamic system using state vector composed of their position and velocity $\mathbf{x}_k$. In general, state evolution at time k is specified by dynamic equation $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_k)$. Measured values(*i.e.* position and velocity) represented by $\mathbf{z}_k$ are used to reduce the uncertainty of dynamic system modeled as posterior probability density function $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. The objective of tracking is to estimate the state $\mathbf{x}_k$ using given all measurements $\mathbf{z}_{1:k}$. The theoretical optimal solution is given by recursive Bayesian filter. These algorithms are suited for very small objects that can not be described as a feature.

The second group is *target representation and localization* as a bottom-up approach. Targets are represented by their shape or appearance. Local descriptor or histogram can be used for finding out their distinctive features. D.G. Lowe[4] presented a remarkable masterpiece using local descriptor. Scale Invariant Feature Transform(SIFT) successfully describes distinctive regions around image keypoints detected by Difference Of Gaussian(DOG). SIFT descriptors are invariant to image scaling and rotation, also partially invariant to change in illumination and 3D camera viewpoint. On the other hand, kernel based tracking method models the target by using histogram.

S. H. Lee, J. S. Bae, T. Kim, J. M. Song and J. J. Kim are with the Department of PGM Technology, R&D Center, Hanwha Corporation, Daejeon, Korea. e-mail: shleebot, bbrother, ktw8203, jmsong81, jjkim2261@hanwha.co.kr

Previous major researches about tracking algorithms have tended to center around the question of image intensity and shape from a target. The present research topic emerged from the question: "What if the measurements from a target are not available?". In two methodologies described so far, it would not be an overstatement to say that the target tracking would be failed. It is because consecutive *prediction* step without *update* step may cause intolerable uncertainty to tracking filter in *filtering and data association*. Similarly, kernel or feature based tracking may fail with serious occlusion in *target representation and localization*. There should be further arguments to be considered.

In this paper, we will attempt to track a target with non-target originated information when target-originated information is not available. In what follows, we will explore projective geometry and algorithm scheme to overcome the situations when the targets are seriously occluded by counter measures or limited FOV.

This paper organized as follows. Section 1 dedicated to introductory overview. Section 2 provides details of projective geometry. Section 3 presents algorithm scheme to track seriously occluded targets. Then tracking results are shown in section 4 to illustrate the robustness which can be achieved by using the proposed approach. The last section presents the concluding remarks and future researches.

II. Projective geometry

In this section, we will discuss extraordinary concept, projective geometry. Traditionally, a camera can be modeled as a perspective pin-hole camera[5] as shown in Fig. 1. A real 3-D point $\mathbf{X}$, whose coordinates are $(X, Y, Z)^T$ in the frame associated to the first image, is projected into camera center C through the image plane in perspective camera model. A 2-D point is defined as $\mathbf{x}$ in the image plane. In the transformed coordinate $(X', Y', Z')^T$ by rotation R and translation t, if the same 3-D point can be projected onto the image plane, that point is defined as $\mathbf{x}'$. We aim to find out the relationship called homography $\mathbf{H}$ between two image points, $\mathbf{x} \leftrightarrow \mathbf{x}'$. Matrix $\mathbf{H}$ has 8 degrees of freedom excluding scale factor from 3×3 matrix. Since each point correspondence gives two equations, at least 4 pairs of point correspondences are required to compute $\mathbf{H}$ up to scale.

A. Homography

As described above, we can determine the 2D homography matrix $\mathbf{H}$ by four correspondences between two images, such

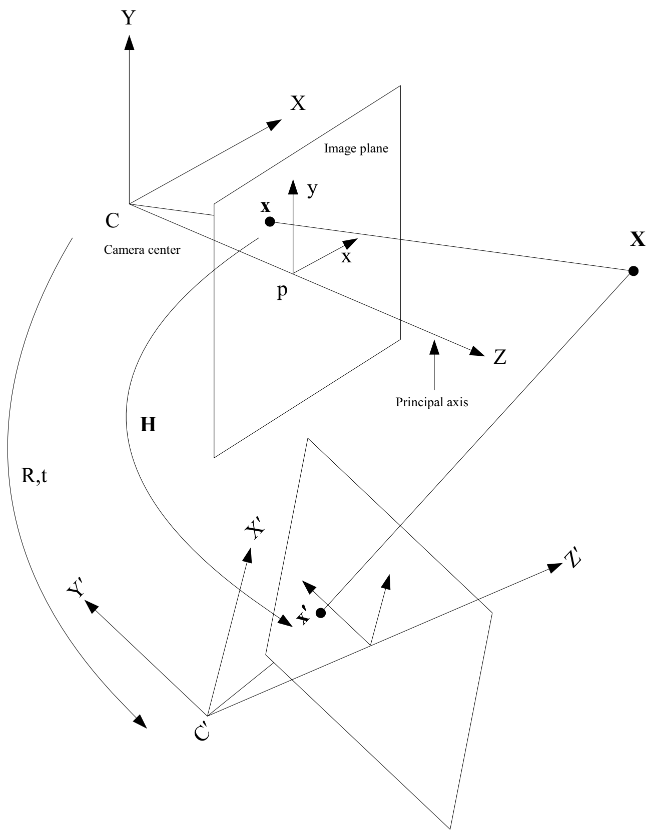


Fig. 1. Perspective camera model

that $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$. Eq. (1) gives us homogeneous linear equation.

$$\mathbf{A}_i\mathbf{h} = \mathbf{0} \tag{1}$$

Solving h from four or more point correspondences can be resolved by minimizing either the algebraic distance error using linear method such as SVD(Singular Value Decomposition)[6] or the geometric error by reprojection. We call this kind of techniques as DLT(Direct Linear Transformation), in other words, the relationship $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$

$$\mathbf{H}\mathbf{x}_i = \begin{bmatrix} \mathbf{h}^{1T}\mathbf{x}_i \\ \mathbf{h}^{2T}\mathbf{x}_i \\ \mathbf{h}^{3T}\mathbf{x}_i \end{bmatrix} \tag{2}$$

where $\mathbf{h}^{jT}$ are jth row of $\mathbf{H}$. Writing $\mathbf{x}'_i = (x'_i, y'_i, w'_i)^T$, the cross product may then be given as

$$\mathbf{x}'_i \times \mathbf{H}\mathbf{x}_i = \begin{bmatrix} y'_i\mathbf{h}^{3T}\mathbf{x}_i - w'_i\mathbf{h}^{2T}\mathbf{x}_i \\ w'_i\mathbf{h}^{1T}\mathbf{x}_i - x'_i\mathbf{h}^{3T}\mathbf{x}_i \\ x'_i\mathbf{h}^{2T}\mathbf{x}_i - y'_i\mathbf{h}^{1T}\mathbf{x}_i \end{bmatrix}. \tag{3}$$

Since $\mathbf{h}^{jT}\mathbf{x}_i = \mathbf{x}_i\mathbf{h}^j$, H can be written as

$$\begin{bmatrix} \mathbf{0}^T & -w'_i\mathbf{x}_i^T & y'_i\mathbf{x}_i^T \\ w'_i\mathbf{x}_i^T & \mathbf{0}^T & -x'_i\mathbf{x}_i^T \\ -y'_i\mathbf{x}_i^T & x'_i\mathbf{x}_i^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{bmatrix} = 0 \tag{4}$$

where

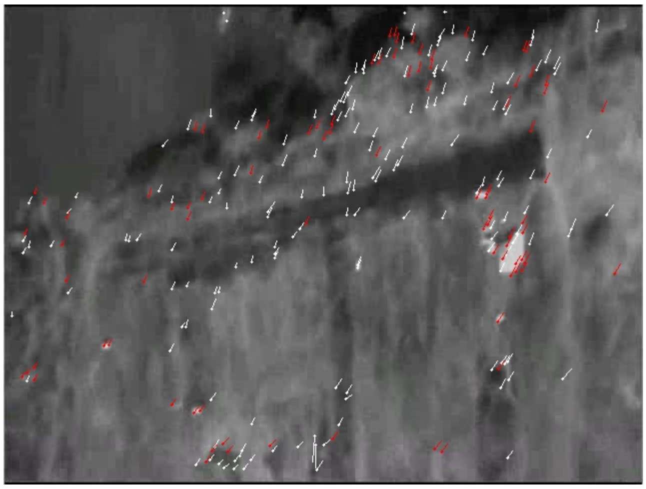


Fig. 2. Feature matching between frame 0 and 1. White and red dots are the feature points, while lines connect correspondences between consecutive two images. White dots and lines are outliers which are filtered out by RANSAC.

TABLE I
ESTIMATION OF HOMOGRAPHY MATRIX

1. Detect four or more corners from two consecutive images by using harris-corner detector. 2. Find four or more matching correspondences from corners by using NCC matching. 3. Estimate homography matrix SVD and RANSAC[8].

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}. \tag{5}$$

Only two equations are linearly independent in Eq. (4). That is the reason why we need four or more pairs of correspondences to solve Eq. (1).

We apply traditional Harris corner[7] and NCC (Normalized Cross Correlation) to consecutive images for taking pairs of point correspondences. The whole procedure for estimating homography matrix can be summarized as Table I. Fig. 2 shows the result of finding correspondences across the consecutive images.

B. Extension to target tracking

Extension to target tracking is intuitively simple. In our application, the aerial vehicle used for grabbing images is flying at a higher altitude than that of any other objects on the ground. Therefore the earth curvature can be neglected. Obviously, the error of estimated relative motion will be accumulated with the increase in number of frames. It should be taken into consideration how to deal with error accumulation problem.

III. PROPOSED ALGORITHM

We described homography matrix in projective geometry to extract the relationship between two images so far. In this section, we will focus on derivation of algorithm scheme

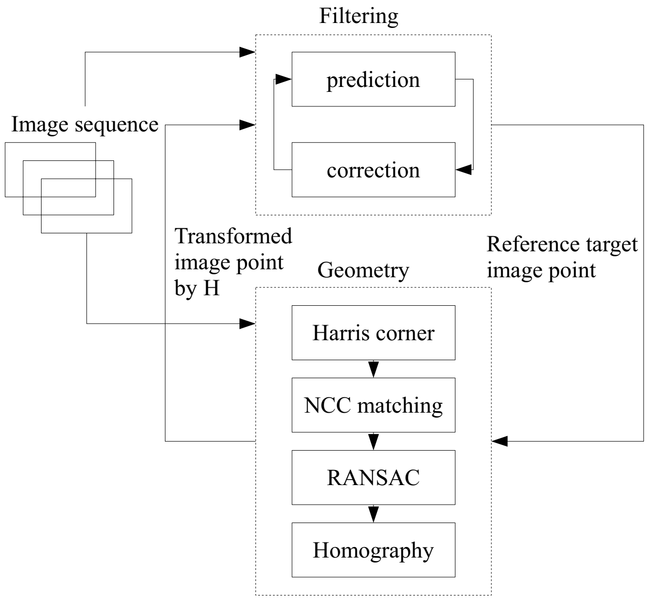


Fig. 3. Proposed algorithm scheme

along with filtering and data association method shown in our previous work[9]. The advantages of each methods (*i.e.* filtering and geometry) are effectively fused together.

A. Tracking strategy

We begin with traditional Kalman filter. Although there exists a very large literature about data association(*ex.* gating technique[10][11]), they require postulations that there exists target-originated information. Thus, these techniques cannot be applied directly to our case. Our proposed algorithm scheme is shown in Fig. 3.

Normalized Distance Square(NDS) and reprojection error could be simply applied to our strategy. NDS is defined as:

$$D_{nds} = (z_k - \bar{z}_k)^T S^{-1} (z_k - \bar{z}_k) \tag{6}$$

where S is predicted measurement covariance with measurement function H, covariance P and noise R, computed as $HPH^T + R$.

Reprojection error is defined as:

$$D_{rep} = \sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 \text{ subject to } \hat{\mathbf{x}}'_i = \hat{\mathbf{H}}\hat{\mathbf{x}}_i \ \forall i. \tag{7}$$

where $d(\mathbf{x}, \mathbf{y})$ is Euclidean image distance between $\mathbf{x}$ and $\mathbf{y}$. Data association, which is the essential problem, is effectively resolved by Eq. (6) and (7). It seems reasonable to assume that these distance measure functions provide valuable information to track or track-termination. If a certain measurement meets both conditions, obviously, we could think of the tracking filter as reliable. If two conditions are both met, the tracking filter is stable. Let us define these two cases as:

$$\begin{cases} \text{target measurement is stable} = D_{nds} \text{ and } D_{rep} \text{ are small.} \\ \text{target measurement is perturbed} = D_{nds} \text{ or } D_{rep} \text{ is large.} \end{cases}$$

In these two cases, Kalman filter works as Table II with state vector $x_k = [x, y, \dot{x}, \dot{y}]^T$.

TABLE II
KALMAN FILTERING

Prediction step

$$\bar{x}_k = \Phi_k \hat{x}_{k-1}$$
$$\bar{P}_k = \Phi_k \hat{P}_{k-1} \Phi_k^T + Q_k$$
$$\Phi_k = \begin{bmatrix} 1 & 0 & k-(k-1) & 0 \\ 0 & 1 & 0 & k-(k-1) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$Q_k = \begin{bmatrix} 20 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Update step

(1) For the case of $D_{nds} > th_{nds}$ or $D_{rep} > th_{rep}$
Wait for that region of interest transformed by homography comes into FOV. Then, re-initialize region of interest.

(2) For the case of $D_{nds} \leq th_{nds}$ and $D_{rep} \leq th_{rep}$

$$\hat{x}_k = \bar{x} + K_k(z - \bar{z})$$
$$\hat{P}_k = \bar{P}_k - K_k S_k K_k^T$$

B. Reset procedure

Error accumulation problem is an inherent limitation in relative position estimation method due to the fact that estimated ego-motion is accumulated to the reference position. Fundamentally, our proposed algorithm scheme is identical with relative position estimation problem. Estimating pixel position in the image using homography, the initial homography matrix is identity and the initial pixel position is given by Kalman filter error covariance matrix. The volume of a search region is inverse proportional to covariance. Four image points are injected and transformed by *accumulated* homography matrix. The accumulation of homography matrix can be represented as follows:

$$\mathbf{x}_2 = \mathbf{H}_{1,2}\mathbf{x}_1 \tag{8}$$

where $\mathbf{x}_1$ is pixel position measured in first image and $\mathbf{H}_{1,2}$ is homography matrix that describes the projective transformation from first image to second image. Therefore, the inverse representation can be evaluated directly according to the following relations:

$$\mathbf{x}_1 = \mathbf{H}_{1,2}^{-1}\mathbf{x}_2 \tag{9}$$

without loss of generality,

$$\mathbf{x}_1 = \prod_{i=2}^{n} \mathbf{H}_{i-1,i}^{-1}\mathbf{x}_n, \tag{10}$$

$$\mathbf{x}_n = \prod_{i=n}^{2} \mathbf{H}_{i-1,i}\mathbf{x}_1. \tag{11}$$

To clear accumulated error on homography matrix $\prod_{i=2}^{n} \mathbf{H}_{i-1,i}^{-1}$, we introduce the procedure for error clearance. It can be summarized as Table III.

TABLE III
RESET ACCUMULATION ERROR

1. Compute homography matrix between two consecutive images using Table I.
2. Detect target candidate from interest region given by homography transformation by using morphological operation and clustering; we described these procedure in our previous work.
3. Compute D_{nds} and determinant of S by using Eq. (6); if these are reasonably small, set homography matrix to identity at frame m. Simultaneously, accumulated homography matrix in Eq. (11) may becomes $\mathbf{x}_n = \prod_{i=n}^{m+1} \mathbf{H}_{i-1,i}\mathbf{x}_m$.
4. Repeat 1∼3.

Fig. 4. Aerial vehicle

IV. EXPERIMENTAL RESULTS

In this section, we illustrate the robustness and the efficiency of our proposed method against those of other methods which only use *filtering and data association* or *target representation and localization* for real applications. First, we show experimental setup to capture images using aerial vehicle and ground station. We then describe and analyze performance of proposed method compared to other methods.

A. Experimental setup

We develop Captive Flight Test(CFT) unit to capture the images for experiments as shown in Fig. 4. IIR and CCD images were taken from the unit which is equipped with wireless image transmission system. The ground station receives the image data from the aerial vehicle during the manually controlled flight, as shown in Fig. 5. Unfortunately, IIR and CCD images cannot be captured at the same time due to the fact that our wireless communication unit provides only one video channel.

B. Performance analysis and discussion

In our experiments, there are four types of test results: filtering only, target representation only, homography based, and proposed method. Each type of tests uses HPDAF proposed in our previous work, template matching, transformation using accumulated homography described in section II-A, and proposed fusing method respectively. SIFT descriptor is not included in our experiments due to pixel localization of DOG detector as shown in Fig. 7. Although DOG detector detect interest points, which is repeatable and scale-rotation invariant, it is not appropriate for finding keypoints of small target in our application. This is the main reason why SIFT is not taken into consideration. In Fig. 10, the results of four

Fig. 5. Ground station

Fig. 6. Captured CCD and IIR images

types of test are shown. Although captive camera is moving agilely, all of the results seem to be reliable for a short period of time. For a relatively long period of time, however, inherent limitations are revealed as show in Fig. 11. While two methods(homography only and proposed method) keep tracking target, the others(template matching and HPDAF) lost targets because of limited FOV; the result images of formers are shown with outside of boundaries as much as 100 pixels to show what happens outside of FOV - (horizontal: 740 pixels, vetical: 580 pixels). The enlarged template images are shown in Fig. 8, which result in failure to track by using template matching based tracking method. As the target is approaching the boundary of an image, template image are changed. Finally, template image does not describe the target any more.

Fig. 12 shows the effectiveness of error clearance strategy which is one of the main contribution of our proposed method. Because of error clearance, the estimation of the area and the position of the target using our approach is far more precise than that of homography based method.

Whole trails of the tracking is shown in Fig. 9, to aid comparability. The trail of ground truth is manually chosen by hand.

V. CONCLUSION AND FUTURE WORKS

We proposed a novel method which is robust against serious occlusion or *long-term* nonexistence of the target information. Accumulation error problem, which is an inherent limitation of relative position estimation approach, is effectively resolved

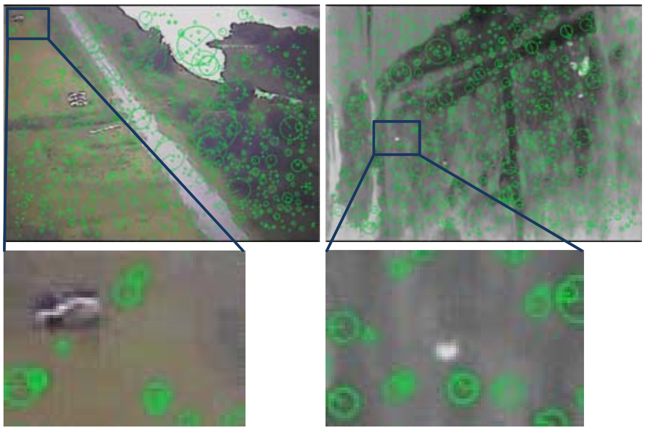
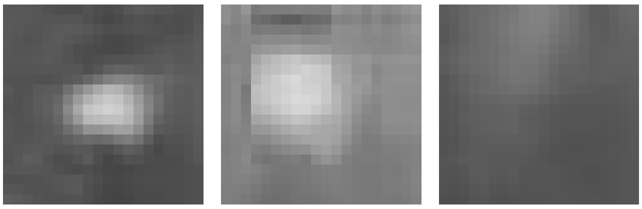


Fig. 7. SIFT descriptors on CCD and IIR images; DOG detector can not detect keypoints of small target.



(a) frame 10 (b) frame 93 (c) frame 130

Fig. 8. Enlarged template images for frame 10, 93 and 130.

by fusing filtering and geometry. The essence of proposed method is that non-target-originated information is also used to estimate the target position. This is major difference from previous methods which estimate validation gate by predicting dynamic model of the target, or search target by investigating best matching with target model.

In the future, we plan to focus on followings: 1) to deal with agilely moving target and moving camera; and 2) to detect target in CCD image(*ex. saliency).*

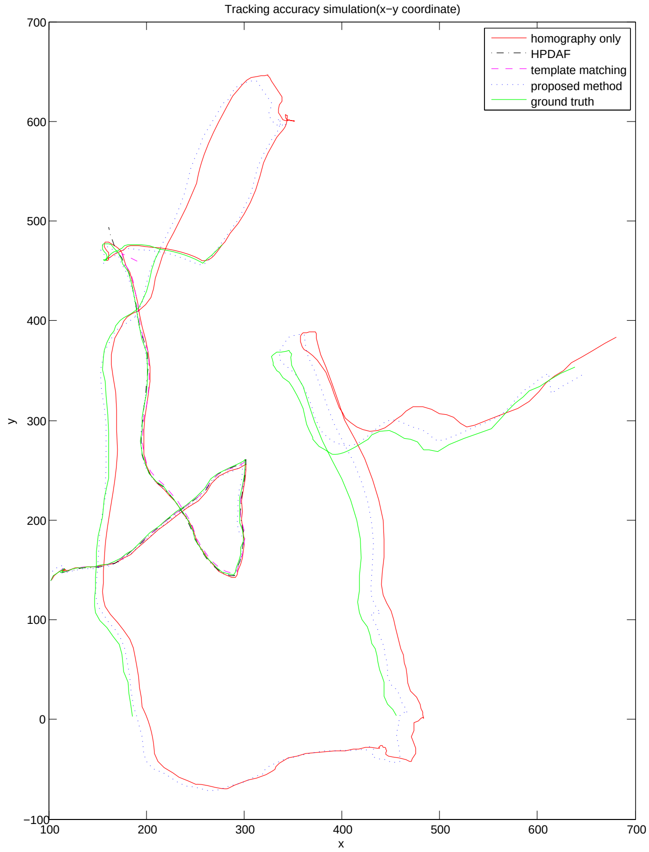


Fig. 9. The trails of tracking.

REFERENCES

[1] *Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking.* IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–575, 2003.*

[2] *X.R. Li and Y. Bar-Shalom. Tracking in clutter with nearest neighbor filters: analysis and performance.* IEEE transactions on aerospace and electronic systems*, 32(3):995–1010, 1996.*

[3] *X.R. Li and X. Zhi. PSNF: A refined strongest neighbor filter for tracking in clutter. In* IEEE CONFERENCE ON DECISION AND CONTROL*, volume 3, pages 2557–2562. INSTITUTE OF ELECTRICAL ENGINEERS INC (IEE), 1996.*

[4] *D.G. Lowe. Distinctive image features from scale-invariant keypoints.* International journal of computer vision*, 60(2):91–110, 2004.*

[5] *R. Hartley and A. Zisserman.* Multiple view geometry. *Cambridge university press, 2000.*

[6] *G.H. Golub and C. Reinsch. Singular value decomposition and least squares solutions.* Numerische Mathematik*, 14(5):403–420, 1970.*

[7] *C. Harris and M. Stephens. A combined corner and edge detector. In* Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.*

[8] *M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.* Communications of the ACM*, 24(6):381–395, 1981.*

[9] *J. S. Bae, S. H. Lee, Y. Kim, and Y. S. Jung. An Imaging Target Tracking Software for a Precision Guided Missile Application. In* Proc. Thirteenth International Conference on Information Fusion*, 2010.*

[10] *K.J. Rhee, D.G. Lee, and T.L. Song. A Probabilistic Strongest Neighbor Filter Algorithm for m Validated Measurements. In* Fusion 2004: Seventh International Conference on Information Fusion; Stockholm. *International Society of Information Fusion, ONERA-DTIM, BP 72, 29 Av. de la Division Leclerc, Chatillon, 92320, France,, 2004.*

[11] *T.L. Song and D.S. Kim. Highest Probability Data Association for Active Sonar Tracking. In* Information Fusion, 2006 9th International Conference on*, pages 1–8, 2006.*

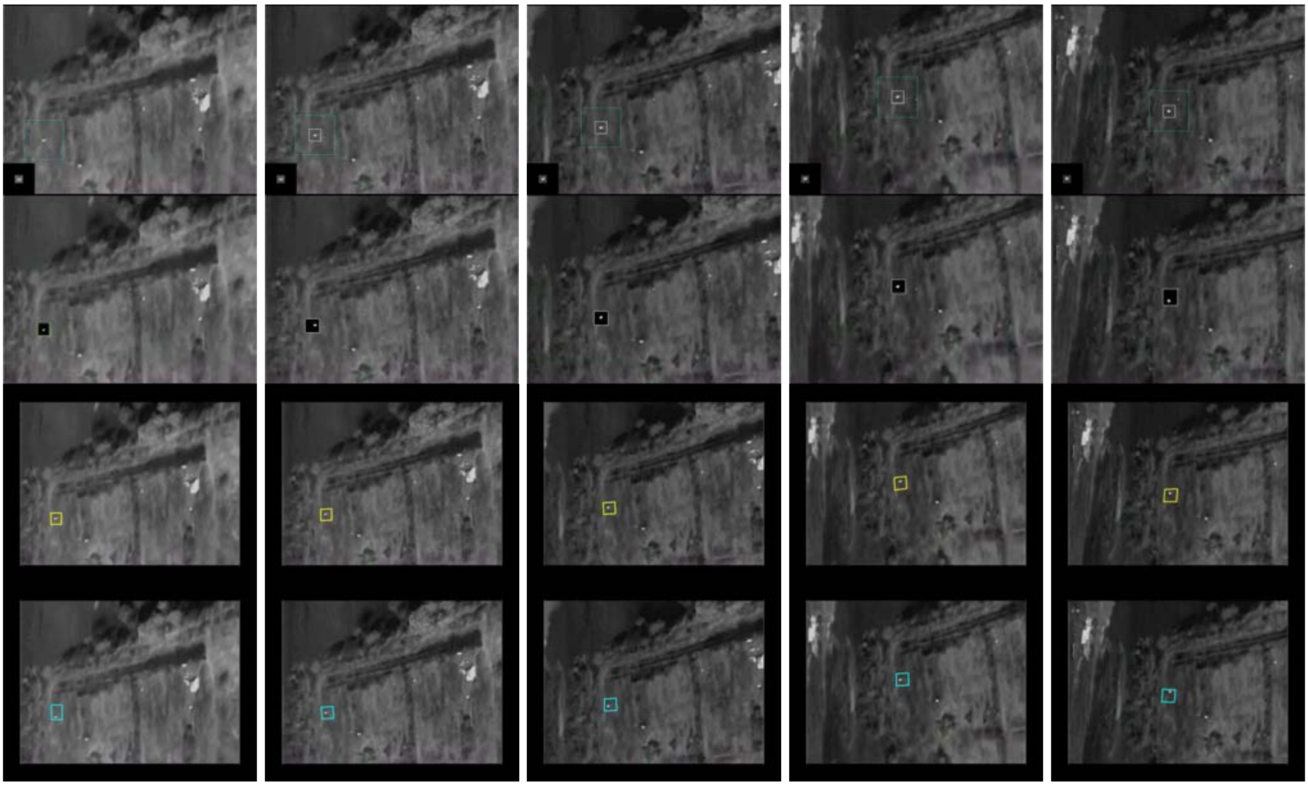
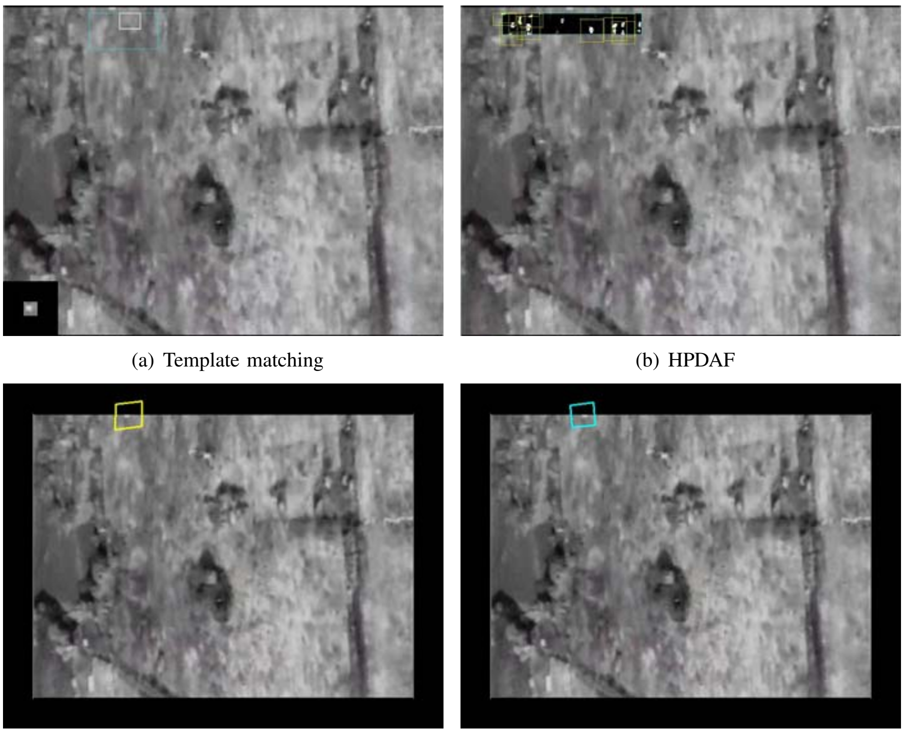


Fig. 10. The tracking results for a short period of time: frame 0, 10, 20, 30, and 40 are shown. Tracking methods, from top to bottom rows, are template matching, HPDAF, homography only, and proposed method.



(a) Template matching

(b) HPDAF

(c) Homography only

(d) Proposed method

Fig. 11. The tracking result; tracking is failed at frame 93 in two methods(template matching and HPDAF).

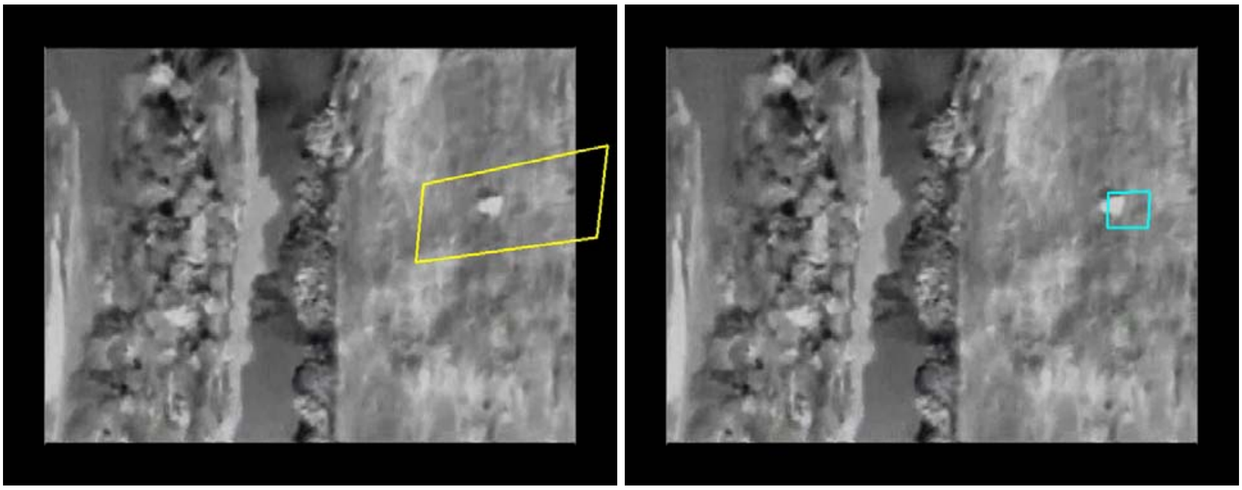


Fig. 12. Clearance of accumulation error: the tracking methods, from left to right, are homography only and proposed method.