Optical Engineering

SPIEDigitalLibrary.org/oe

Classifier-based offline feature selection and evaluation for visual tracking of sea-surface and aerial targets

Serdar Çakır
Tayfun Aytaç
Alper Yıldırım
Ö. Nezih Gerek

SPIE

Classifier-based offline feature selection and evaluation for visual tracking of sea-surface and aerial targets

Serdar Çakır
Tayfun Aytaç
Alper Yıldırım
TÜBİTAK BİLGEM
UEKAE/İLTAREN
Şehit Mu. Yzb. İlhan Tan Kışlası
2432. cad., 2489. sok.
Ümitköy
Ankara, TR-06800, Turkey
E-mail: serdar.cakir@iltaren.tubitak.gov.tr.

Ö. Nezih Gerek
Anadolu University
Department of Electrical and Electronics Engineering
İki, Eylül Kampusü
Eskişehir, TR-26470, Turkey

Abstract. An offline feature selection and evaluation mechanism is used in order to develop a robust visual tracking scheme for sea-surface and aerial targets. The covariance descriptors, known to constitute an efficient signature set in object detection and classification problems, are used in the feature extraction phase of the proposed scheme. The performance of feature sets are compared using support vector machines, and those resulting in the highest detection performance are used in the covariance based tracker. The tracking performance is evaluated in different scenarios using different performance measures with respect to ground truth target positions. The proposed tracking scheme is observed to track sea-surface and aerial targets with plausible accuracies, and the results show that gradient-based features, together with the pixel locations and intensity values, provide robust target tracking in both surveillance scenarios. The performance of the proposed tracking strategy is also compared with some well-known trackers including correlation, Kanade–Lucas–Tomasi feature, and scale invariant feature transform-based trackers. Experimental results and observations show that the proposed target tracking scheme outperforms other trackers in both air and sea surveillance scenarios.
[DOI: 10.1117/1.3640826]

Subject terms: target tracking; feature selection; feature evaluation; covariance tracker; target detection; surveillance systems.

Paper 110727R received Jun. 27, 2011; revised manuscript received Aug. 26, 2011; accepted for publication Aug. 29, 2011; published online Sep. 29, 2011.

1 Introduction

Extraction and selection of appropriate features for target representation are the key steps in most target detection and tracking and behavior analysis applications.[1] Features may be color, edges, displacement vectors in optic flow-based approaches, textures, and their combinations depending on the target model (appearance and motion) and imaging system. The features extracted from the target locations should have several properties including high differentiation property, scale and rotation invariance, robustness to noise, and partial invariance to affine transformation, intensity changes, and occlusion. Another important issue is the real time requirement where proper features decrease the computational cost of the detection and tracking algorithm. A comprehensive survey about object tracking is made in Refs. 1 and 2.

In the literature, raw pixel intensities or first or higher order statistics extracted from these values are commonly used in different target tracking problems.[3–5] Kernel-based methods are widely used in target tracking,[6–9] where histogram features extracted from raw pixel intensities are utilized. Features obtained by Scale invariant feature transform (SIFT) (Ref. 10) are independent of scale, rotation, and intensity change and robust against affine transformation. They are widely used in applications for target detection,[11] tracking,[12] classification,[13] image matching,[14] and constructing mosaic images.[15] When compared to other point-of-interest detectors such as Moravec[16] and Harris,[17] SIFT features are more robust to background clutter, noise, and occlusion. Harris corner detection is not applicable in most cases where the target changes its orientation rapidly or the camera angle changes in the subsequent frames in surveillance systems because it is not invariant to scale. Unfortunately, despite the distinctive properties of SIFT, the feature extraction process is time consuming and the method is hardly used in real time applications. Correlation filters for the target tracking problem are proposed in Refs. 18–20. Covariance descriptors with low computational load are defined in Ref. 21, and these descriptors are successfully used in applications such as indoor and outdoor target tracking.[22]

In this study, an offline feature selection and evaluation mechanism is developed for robust tracking of sea-surface and aerial targets in visual band videos. The covariance descriptors,[21] which cover both spatial and statistical information and their correlations, are used as features. Image features used in the computation of the covariance matrix are tested offline using a database of target and background images for different scenarios using support vector machines (SVMs).[23] Scenarios are defined for the different sea and sky backgrounds and sea-surface and aerial targets. The major contribution in this work is the development of an offline feature extraction and cross verification scheme for robust tracking of sea-surface and aerial targets.

Eventually, a complete solution is proposed as a surveillance system using a visual band camera mounted on land or sea platforms to track symmetric and

0091-3286/2011/$25.00 © 2011 SPIE

asymmetric targets which gains importance in military and security applications.[24–32] To the best of our knowledge, covariance-based trackers were not previously used for tracking of sea-surface and aerial targets in the literature, hence the proposed method imposes a novel attempt to the problem. Moreover, an offline feature selection mechanism for robust target tracking has not been visited in the literature. An adaptive appearance-based update strategy is developed by considering the motion characteristics of the targets, and, in order to test the usefulness of the proposed method, the performance of the proposed tracking system is evaluated using different performance measures in field trials considering scale, rotation, and partial occlusion cases.

The paper is organized as follows: In Sec. 2, local covariance descriptors are briefly described. Feature selection and two-class (target and background) differentiation using SVMs are explained in Sec. 3. In Sec. 4, proposed target tracking framework is given in detail. Experimental work and results with a comparison using the ground truth data with different performance measures are provided in Sec. 5. In Sec. 5, the performance of the proposed system is compared with correlation, Kanade–Lucas–Tomasi (KLT) feature, and SIFT-based trackers. Concluding remarks are made and direction for future research is provided in Sec. 6.

2 Local Covariance Descriptors

In this study, local covariance descriptors are used for target tracking purposes. They are chosen due to their low computational complexity and robustness to partial occlusion. They also enable to add or remove features in a simple manner to adapt the tracker for the different target types.

In the computation of the covariance matrix, the feature matrices $(f_i;\ i = 1, 2, \ldots, D)$ extracted from a $W \times H$ subregion of an image are stacked to form $W \times H \times D$ dimensional feature tensor $T(:, :, :)$ (Fig. 1). W and H are the width and height of the template and D is the number of features extracted from the template. In the feature tensor, the elements in each layer with the index (m, n) are sorted to construct the feature vector $(\underline{S}_t)$ [Eq. (1)]. In total, $W \times H$ feature vectors $(\underline{S}_t)$ are constructed

$$\underline{S}_t = [f_1(m, n) \quad f_2(m, n) \quad \cdots \quad f_D(m, n)], \tag{1}$$

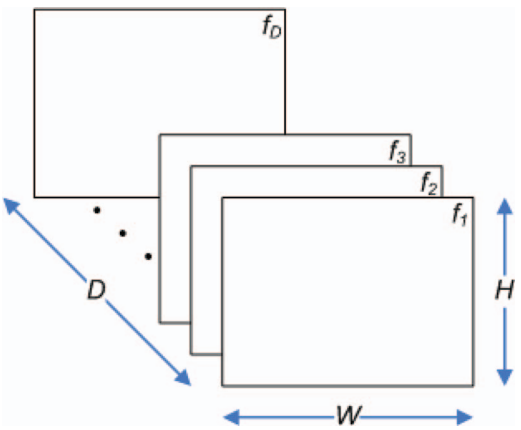

Fig. 1 The feature tensor $T(:, :, :)$ formed by placing feature matrices back to back.

where $m = 1, 2, \ldots, W$, $n = 1, 2, \ldots, H$, $t = 1, 2, \ldots, k$, and $k = W \times H$.

Local covariance matrix, M_{loc}, is computed by using the feature vectors at each index (m, n) as follows:

$$M_{\text{loc}}(p, q) = \frac{1}{k-1}\left[\sum_{t=1}^{k} \underline{S}_t(p)\underline{S}_t(q) - \frac{1}{k}\sum_{t=1}^{k} \underline{S}_t(p) \times \sum_{t=1}^{k} \underline{S}_t(q)\right], \tag{2}$$

where $p, q = 1, 2, \ldots, D$.

In the classical formulation of the covariance matrix [Eq. (2)], there are too many multiplication operations that increase the computational cost. Therefore, the "integral image" (Ref. 21) approach, known as a computationally efficient method, is adopted for the computation of the covariance matrix. The integral image (ψ) can be expressed in general as the accumulation of intensities (I) in a selected region. Considering the image shown in Fig. 2, the ψ image for the selected region is computed as follows:

$$\psi(x', y') = \sum_{x<x'} \sum_{y<y'} I(x, y). \tag{3}$$

In Ref. 21, the local covariance descriptor for a given region is computed by using the Eq. (4).

$$P(x', y', p) = \sum_{x<x'} \sum_{y<y'} T(x, y, p)$$

$$Q(x', y', p, q) = \sum_{x<x'} \sum_{y<y'} T(x, y, p)T(x, y, q)$$

$$P_{x',y'} = [P(x', y', 1) \cdots P(x', y', D)]^T,$$

$$Q_{x',y'} = \begin{pmatrix} Q(x', y', 1, 1) & \cdots & Q(x', y', 1, D) \\ \vdots & \ddots & \vdots \\ Q(x', y', D, 1) & \cdots & Q(x', y', D, D) \end{pmatrix}$$

$$M_{\text{loc}(1,1\ ;\ x',y')} = \frac{1}{k-1}\left[Q_{x',y'} - \frac{1}{k}P_{x',y'}P^T_{x',y'}\right]. \tag{4}$$

In this work, the local covariance descriptors are computed for the two scenarios separately: sea and air surveillance. Since the target and background characteristics of

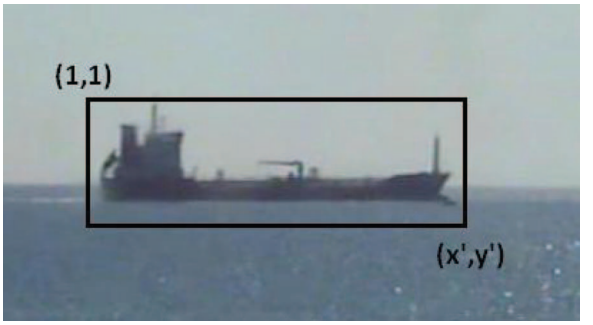

Fig. 2 The illustration of a sample region used in the computation of the integral image.

the sea and air surveillance scenarios differ significantly, different types of features are used in the computation of the covariance descriptors for both surveillance scenarios. Feature extraction and selection mechanism is explained in Sec. 3.

3 Feature Extraction and Selection

The camera used in the study captures interlaced videos. Although interlacing structures have important merits such as reducing the signal bandwidth by a factor of 2, interlacing artifacts occur when there exists rapid camera movements and vibrations. The elimination of interlacing artifacts is a crucial task in appearance-based applications. Since the developed tracker is intended to be used in real-time applications, a fast and simple deinterlacing algorithm is needed. Therefore, a "line doubling" type of approach is used, where the odd numbered (even numbered) rows of each frame are taken and the interpolation of two consecutive rows are placed between these rows. At the end, a reasonably deinterlaced video frame at the same dimension with the original video frame is obtained.

After deinterlacing, RGB to gray-scale conversion is performed as a preprocessing step to generate covariance descriptors. The gray-scale conversion is preferred since the target color does not change significantly in sea and air surveillance scenarios. The features extracted from the gray-scale image in the region of interest (ROI) are listed below:

- image intensity (I)
- image horizontal position (x)
- image vertical position (y)
- first derivative of the image in the horizontal direction ($\partial_{1,x} = \partial I/\partial x$)
- first derivative of the image in the vertical direction ($\partial_{1,y} = \partial I/\partial y$)
- second derivative of the image in the horizontal direction ($\partial_{2,x} = \partial^2 I/\partial x^2$)
- second derivative of the image in the vertical direction ($\partial_{2,y} = \partial^2 I/\partial y^2$)
- gradient magnitude ($\mathrm{GM} = \sqrt{\partial_{1,x}^2 + \partial_{1,y}^2}$)
- gradient orientation ($\mathrm{GO} = \tan^{-1}(\partial_{1,y}/\partial_{1,x})$).

Features are extracted from the target and background images within a data set. The efficiency of these features on target-background classification is observed using the SVM classifier.[23] SVM, developed by Vladimir Vapnik, is a supervised machine learning method based on the statistical learning theory. The method constructs a hyperplane or a set of hyperplanes in a high dimensional space that can be used in applications such as object, voice, and handwritten character recognition and text classification. SVM preprocesses and represents the features in a higher dimensional space where they may become linearly separable.

In a dataset (sea or air surveillance), the features or feature sets extracted from the target and background images are mutually or in combination used to construct regional covariance descriptors. The descriptors belonging to target and background images are applied to the SVM classification engine to separate target feature vectors from the background feature vectors. In the classification process, C-support vector classification (C-SVC) type of formulation of the

Fig. 3 The target images above sea-level.

SVM software package[33] is used. The radial basis function kernel [$K(\omega_i, \omega_j) = \exp(-\gamma \|\omega_i - \omega_j\|^2), \quad \gamma > 0$, ω_i and ω_j are feature vectors] is selected in the computation and parameters are determined after cross validation. For the experimental dataset, the kernel parameter (γ) and the penalty parameter of the error term (C) are selected after a cross validation process as 10,000 and 0.1, respectively.

Among the covariance descriptors, the features with higher target-background classification accuracies are selected as a candidate feature set and used in the target tracking phase.

In the Secs. 3.1 and 3.2, the features providing the best target-background classification accuracy for sea and air surveillance are given and the procedure is explained for both surveillance environments.

3.1 *Feature Selection for Sea-Surface Targets*

In order to observe the performance of different features in sea-surface target-background classification, a two-class dataset including the images of sea-surface targets and background is constructed. The dataset consists of 1000 images of several sea-surface targets and 1000 background images captured with a visual band camera at different day times, places, and scenarios. Sample images from the dataset are given for target and background in Figs. 3 and 4, respectively. As seen in Fig. 3, target images have also some portion of the background because we are interested not only in target texture but also in target to background transitions, which provides useful information. Five hundred target and 500 background images are randomly selected and used in the training of the SVM. The remaining 500 target and 500 background images are used in the test of the SVM. This procedure is repeated 10 times and the average target-background classification rate is computed.

The target-background classification performances of the feature sets are given in Table 1. Although all possible combinations of the features are used in target-background separation, several but representative ones are presented in Table 1 to preserve the readability of the results. The results indicate that individual and combined utilization of features may significantly alter the performance.[34] As an example, although the pixel positions are not discriminative features individually, higher classification performance is obtained when the pixel positions are used together with the other types of features. On the other hand, it can be said that the correlation between the pixel positions and derivative-based features has

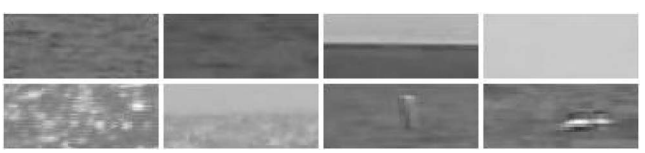

Fig. 4 The background images in sea surveillance.

Table 1 SVM-based sea-surface target-background classification accuracies for different feature sets. The candidate feature sets for sea-surface target tracking tests are separated with a line in the table. Although all possible combinations of the features are used in the experiments, several but representative ones of them are presented to preserve the readability of the results.

Feature set	Target-background classification accuracy (%)
y	55.8
x	58.5
$\partial_{2,y}$	64.4
$\partial_{1,y}$	64.8
GO	67.8
$\partial_{1,x}$	72.7
$\partial_{2,x}$	73.8
I	85.8
I, x, y	95.4
I, x, y, GO	98.8
$I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}, GM, GO$	98.9
I, x, y, GM	99.1
$I, x, y, \partial_{2,x}, \partial_{2,y}$	99.2
I, x, y, GM, GO	99.5
$I, x, y, \partial_{1,x}, \partial_{1,y}$	99.7
$I, x, y, \partial_{2,x}, \partial_{2,y}, GM$	99.8
$I, x, y, \partial_{1,x}, \partial_{1,y}, \partial_{2,x}, \partial_{2,y}$	99.8
$I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}, GM$	99.8

more discriminative power. The results presented in Table 1 show that one cannot generalize a rule that the classification performance increases by adding more features. Therefore, in order to examine the feature characteristics of the sea-surface targets in the tracking scenarios, the feature sets providing over 90% classification accuracy are selected as candidate sets and the tracking performances of these feature sets are measured. By this way, a representative feature set could be determined for a specific target type in order to develop a target tracker that is robust to sudden camera movements, scale changes, and occlusion, which is the main motivation of the paper.

3.2 *Feature Selection for Aerial Targets*

Similar to the approach in Sec. 3.1, a two-class dataset is constructed for air surveillance in order to compare the performance of different features and feature sets on aerial target-background classification. The dataset constructed for air surveillance consists of 600 images of several aerial vehicles (airplane and helicopter) and 600 background images (sky, clouds, etc.). The images are the frames of different visual-band videos captured at different day times, locations, and

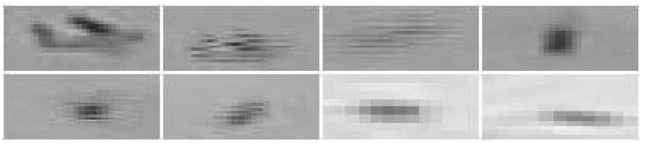

Fig. 5 The images of aerial targets.

scenarios. Sample aerial target and background images are given in Figs. 5 and 6, respectively.

Three hundred out of 600 images are randomly selected from both target and background images in order to train the SVM classifier. The remaining 300 aerial target and 300 background images are used in the test phase of the SVM. The procedure is repeated 10 times in order to obtain robust accuracies.

The target-background classification performances of the feature sets are given in Table 2. As per the results presented for sea-surface targets, all possible combinations of features are used in the experiments for aerial targets, but only representative ones are listed in Table 2 to preserve the readability of the results. Similar to the results given for the sea-surface targets, the pixel coordinate locations are not discriminative individually and provide low target-background classification accuracies. However, the classification accuracies dramatically increase if the coordinate locations are used together with other types of features.

As per the classification performance for sea surveillance presented in Table 1, adding more features does not cause an increase in target-background classification performance. Therefore, the performance of the feature sets in aerial target tracking scenarios are examined by selecting candidate feature sets that provide over 90% classification accuracy. In Sec. 4, a regional covariance descriptor-based tracker is described.

4 Covariance-Based Tracker

The flow diagram of the proposed regional covariance descriptor-based tracking scheme is given in Fig. 7. The tracking algorithm is initiated as soon as the target is selected. In general, the initial selection of the target can be automatic or manual. In this work, the initial target is selected manually in the first frame. Manual target selection must be done such that the selected gate includes the target and some proportion of the background. When the target is not centered in the selected gate (i.e., background covers much more area than the target), this may result in the degradation of the performance of the tracker. After determination of the target template (TT), feature descriptors are extracted from the template and matched with those extracted in the next frame to find the best matching region ($M_{R,\mathrm{Best}}$). During matching, a generalized eigenvalue-based similarity metric is used in order to compare the covariance descriptors extracted

Fig. 6 The background images in air surveillance.

Table 2 SVM-based aerial target-background classification accuracies for different feature sets. The candidate feature sets for aerial target tracking tests are separated with a line in the table. Although all possible combinations of the features are used in the experiments, several but representative ones of them are presented to preserve the readability of the results.

Feature set	Target-background classification accuracy (%)
x	45.5
y	54.5
$\partial_{1,y}$	64.8
$\partial_{2,y}$	67.8
GO	70.6
$\partial_{1,x}$	72.0
$\partial_{2,x}$	73.5
I, x, y, $\partial_{1,x}$, $\partial_{2,x}$, $\partial_{1,y}$, $\partial_{2,y}$, GM, GO	86.6
I, x, y, GM, GO	96.0
I, x, y, GO	96.1
I	98.3
I, x, y, $\partial_{1,x}$, $\partial_{1,y}$	98.3
I, x, y, GM	98.6
I, x, y, $\partial_{2,x}$, $\partial_{2,y}$	99.0
I, x, y, $\partial_{1,x}$, $\partial_{2,x}$, $\partial_{1,y}$, $\partial_{2,y}$, GM	99.3
I, x, y, $\partial_{2,x}$, $\partial_{2,y}$, GM	99.6
I, x, y, $\partial_{1,x}$, $\partial_{1,y}$, $\partial_{2,x}$, $\partial_{2,y}$	99.6
I, x, y	99.8

from the TT (M_{TT}) and descriptors extracted from the region that is desired to be compared (M_R):[35]

$$\rho(M_{\mathrm{TT}}, M_R) = \sqrt{\sum_{i=1}^{D} \ln^2 \lambda_i(M_{\mathrm{TT}}, M_R)}, \tag{5}$$

where $\{\lambda_i(M_{\mathrm{TT}}, M_R)\}_{i=1,\ldots,D}$ are the generalized eigenvalues of M_{TT} and M_R. In the target search, the sub-region giving the minimum ρ value with the TT is chosen as the target region.

In a video frame, the target is searched within a region [search region (SR)] that surrounds the target center in the previous frame. The SR consists of pixels that are located in the τ-pixel neighborhood of the target center in the previous frame. Depending on the ρ value in the previous frame, the target is searched in different scales or a single scale. If ρ is larger than a threshold e_0, the target is searched in different scales (meaning camera zoom or target approach/leave). In this case, each pixel of the SR is determined as the center of the candidate region and different scaled rectangles centered at each pixel of the SR are determined to be candidate regions. The dimensions of the different scaled rectangles are determined by multiplying the dimensions of the target template of the previous frame with the scale coefficient κ. The tracker contains two shrinkage ($\kappa = \{0.8, 0.9\}$) and two growth ($\kappa = \{1.1, 1.2\}$) scale coefficients. By this way, the target is searched within SR using four different scales by considering the target dimension changes in both a positive and negative manner. This approach is similar to the Monte Carlo-based target update strategy presented in Ref. 36. The candidate region resulting in the smallest ρ value with the current TT is selected as $M_{R,\mathrm{Best}}$ and the TT is updated using the $M_{R,\mathrm{Best}}$.

In the TT update, Euclidean distance rate (α) and ρ are used together. α is calculated using the Euclidean distance (ED) and defined in the Eq. (6).

$$\mathrm{ED} = \|M_{R,\mathrm{Best}} - \mathrm{TT}\|_2, \quad \alpha = \frac{\mathrm{ED}}{\text{number of pixels } (M_{R,\mathrm{Best}})}. \tag{6}$$

The TT update strategy depends on the α and ρ values and their predefined thresholds e_2 and e_3. If ρ is smaller than e_2, a match is assumed and TT is taken as $M_{R,\mathrm{Best}}$. Otherwise, the TT is updated according to the α value. In this case, template change counter (TCC), which is defined to indicate the number of similar ($\alpha < e_3$) TTs and $M_{R,\mathrm{Best}}$'s in the consecutive frames, is altered. If the α value defined in Eq. (6) is less than e_3, the TCC value is incremented by one and TT is updated according to Eq. (7).

$$\mathrm{TT}_{\mathrm{Next}} = \alpha(\mathrm{M}_{\mathrm{R,Best}}) + (1-\alpha)\mathrm{TT}. \tag{7}$$

In Eq. (7), since α has small values, the previous TT value is more emphasized in the updated TT.

When the TCC reaches a predefined value (N), existing TT is updated with the same strategy, but the $M_{R,\mathrm{Best}}$ is more emphasized in TT update. Therefore, the update in Eq. (7) is modified as follows:

$$\mathrm{TT}_{\mathrm{Next}} = (1-\alpha)\mathrm{M}_{\mathrm{R,Best}} + \alpha\mathrm{TT}. \tag{8}$$

In this case, after TT is updated, TCC is reset to zero. The same zero-resetting is also applied if the α value is larger than the threshold e_3.

In the covariance-based tracker, if TT is largely different from $M_{R,\mathrm{Best}}$, ρ metric becomes greater than its value in a normal match. In this case, the algorithm assumes that the target faced a scale change and initiates a target search with varying scales. This property enables to track targets with varying scale and shape. It also provides robustness to abrupt camera movements, camera vibrations, and sudden displacements.

In aerial target tracking, if ρ is larger than threshold e_1, the tracker assumes that there is a significant change in the target model and a target detection strategy is initiated in order to adapt the TT to the rapid changes in the target model. The target detection algorithm used in the air surveillance case is a simple intensity thresholding-based technique that takes the advantage of contrast difference between the aerial target and the sky background. The reason to use a simple target detection algorithm is to meet the real-time requirements. The detection algorithm is tested over plenty of air surveillance videos and satisfactory detection performances are achieved.

5 Experimental Work and Results

The candidate feature sets providing over 90% target-background classification accuracy are used in the

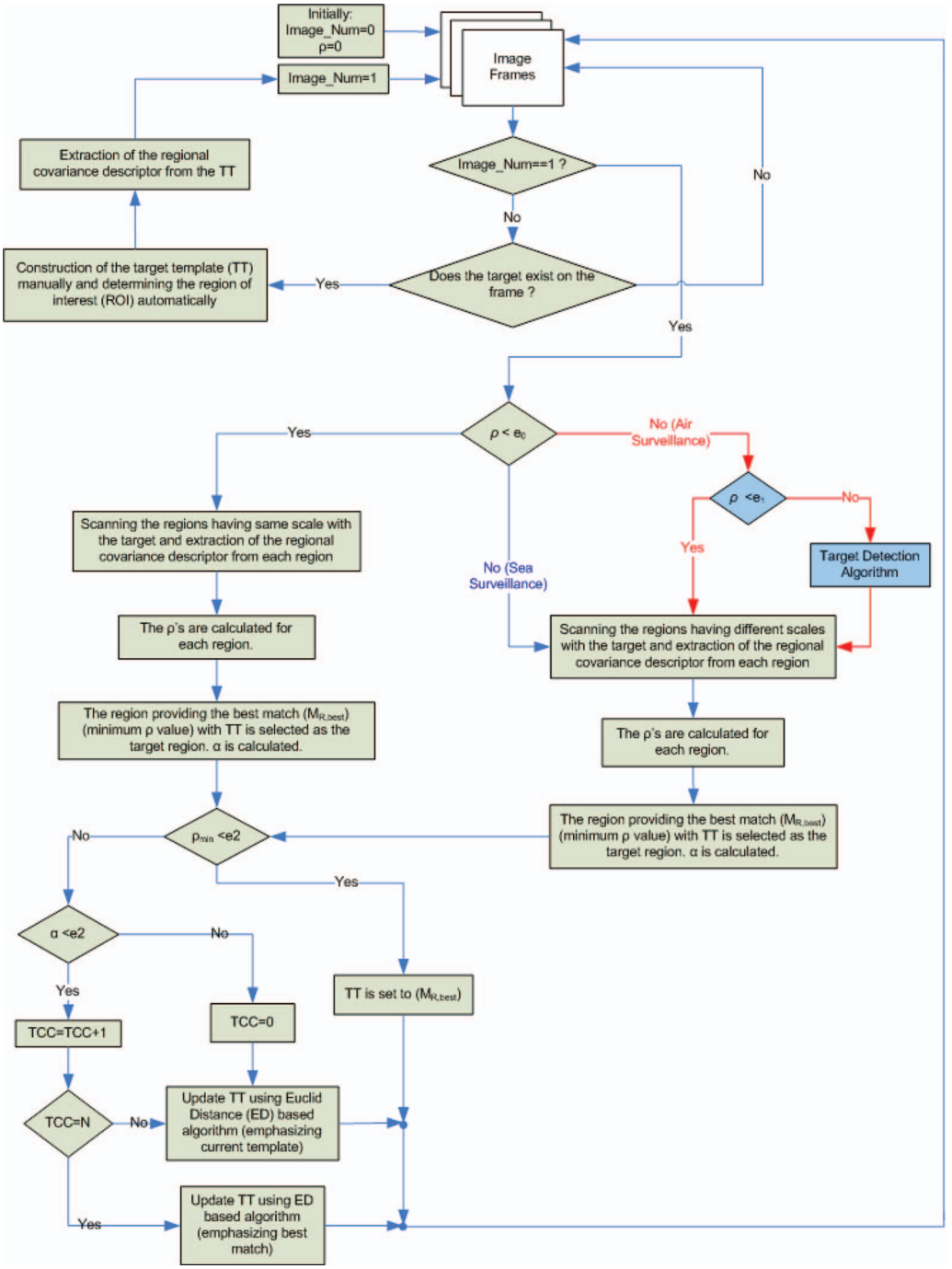

Fig. 7 The flow diagram of the regional covariance descriptor based target tracker.

covariance-based tracker, as explained in Sec. 4. The performance of each candidate set is examined for both sea and aerial surveillance scenario. In each surveillance scenario, the performance of the tracker is examined by running the tracker algorithm on two different videos with different candidate feature sets. In Sec. 5.1, the performance measures used to evaluate the tracking performance are described.

5.1 *Performance Measures*

In order to evaluate the tracking performance in a robust way, four different morphological similarity measures (PM_i, where $i = 1, 2, 3, 4$) are used. These approaches are pixel-wise area (PM_1 and PM_2) and distance-based measures (PM_3 and PM_4). Depending on the different scenario cases, a suitable performance measure can be selected for the evaluation of the track performance. For example, PM_3 and PM_4 may give erroneous results when used to compare the track performances for two different targets, one of which covers a large area in the field of view of the imaging system and the other is far away from the imaging system. A similar case is also possible in the comparison of the tracking results across imaging systems of different spatial dimensions. For the evaluation of the tracking performance, other measures relating to track quality such as number of false

tracks, duration of false tracks, successful track length, may also be used. However, in our case, surveillance videos are not long enough to carry these analyses. For example, track length becomes important where targets are tracked in long term and drift errors result in divergence from the actual trajectory.

Considering the sample illustration in Fig. 8, the first performance measure (PM_1) is defined as follows:

$$PM_1 = \frac{\text{Area}(ABGXHD)}{\text{Area}(ABCD)}, \tag{9}$$

where Area($ABGXHD$) is the area of the polygon that is formed by the difference of the actual target gate from the intersection of the target gate obtained by the tracker and the actual target gate. Area($ABCD$) is the area of the rectangle that symbolizes the actual target gate. The range of PM_1 varies as:

$$PM_1 = \begin{cases} 0, & \text{exactly the same,} \\ 1, & \text{no similarity,} \\ (0,\,1), & \text{otherwise.} \end{cases} \tag{10}$$

The second performance measure (PM_2) actually provides a measure for the similarity of the actual target gate and the target gate obtained by the tracker (as depicted in Fig. 8) and PM_2 is defined as follows:

$$PM_2 = \frac{\text{Area}(XGCH)}{\text{Area}(XYZT)}, \tag{11}$$

where Area($XGCH$) is the area of the rectangle that is formed by the intersection of the target gate obtained by the tracker and the actual target gate. Area($XYZT$) is the area of the rectangle that symbolizes the target gate obtained by the tracker. The range of the PM_2 attains the following values:

$$PM_2 = \begin{cases} 0, & \text{no similarity,} \\ 1, & \text{exactly the same,} \\ (0,\,1), & \text{otherwise.} \end{cases} \tag{12}$$

The third performance measure (PM_3) is the ED between the center of the actual target gate (O_1 in Fig. 8) and the center of the target gate obtained by the tracker (O_2 in Fig. 8).

The last metric (PM_4) is the "city block distance" between the center of the actual target gate (O_1 in Fig. 8) and the center of the target gate obtained by tracker (O_2 in Fig. 8).

The performance measures defined in this section are calculated for each video frame and the tracking statistics for each measure (PM_i) are obtained by averaging the frame-wise measures.

After obtaining four performance measures, these values are parametrically fused as in Eq. (13) for the final evaluation of the tracking performance.

$$\text{Track\_Score} = (1 - \text{PM}_1)\text{PM}_2 - \zeta\left(\frac{\text{PM}_3 + \text{PM}_4}{2}\right). \tag{13}$$

The PM_i's are optimized to form the overall expression in Eq. (13) in the combined case that PM_1 converges to 0, PM_2 converges to 1, and PM_3 and PM_4 also converge to 0. The parameter ζ in Eq. (13) provides the normalization factor for the significance of the first and second summands, and is experimentally selected as 0.01.

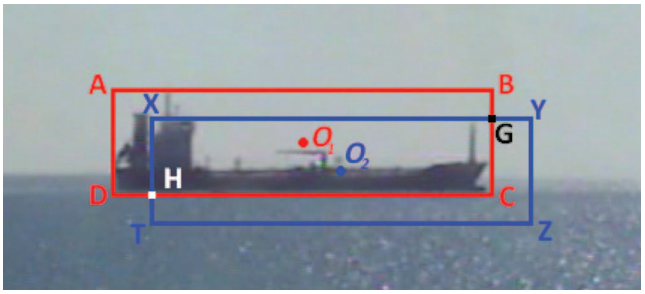

Fig. 8 The illustration of actual target gate and the target gate obtained by the tracker. rectangle ABCD: Actual target gate. rectangle XYZT: The target gate obtained by the tracker.

The tracking results for both surveillance scenarios are presented in the Secs. 5.2–5.4.

5.2 *Tracking of Sea-Surface Targets*

As stated in Sec. 3, the feature sets providing over 90% accuracy are selected as candidate features using Table 1. In this section, the covariance-based tracker is used to track sea-surface targets for each candidate feature set. The tracking experiments are carried out using two visual band videos captured at different daytimes and at 640×480 dimension. The first video (Sea_Video_Seq_1) contains 1000 frames of a moving sea-surface target that is exposed to occlusion in certain frames. The second video (Sea_Video_Seq_2) used for sea surveillance consists of 500 frames of a moving fishing boat toward a transport ship. The tracking performance of the tracker in these videos are measured by comparing the target gate locations obtained by the tracker with the ground truth target gate locations using the measures defined in Sec. 5.1.

The tracker parameters τ, e_0, e_2, e_3, and N for sea surveillance scenario are selected as 5, 0.4, 0.1, 0.0019, and 10, respectively. These values are obtained experimentally by considering different cases. The evaluation of the tracking performance for each candidate feature set is given in Table 3. The tracking results presented in Table 3 are the results obtained by averaging each frame-wise performance measures. The frame-wise PM_i's for the video Sea_Video_Seq_2 are provided in Figs. 9 and 10 as a more detailed analysis. The gradient-based features are used in this analysis.

Table 3 shows that the gradient-based feature sets $\{I, x, y, \text{GM}\}$ and $\{I, x, y, \text{GM}, \text{GO}\}$ provide satisfactory tracking performances in both videos. Since GM and GO contain the magnitude of the intensity changes in both directions as well as their angular orientations, they are expected to have better target representation in sea-surveillance scenarios.

By looking at the results presented in Table 3, one can say that there exists some feature sets that provide satisfactory performance only on one video. Apparently, these feature sets are case-dependent features. Therefore, we can conclude that it is better not to rely on such a case-dependent feature set for sea-surface target tracking.

The results presented in Table 3 also show that adding more features do not necessarily increase the tracking performance.[34] As an example, although the feature

Table 3 The performance of candidate feature sets in sea-surface target tracking.

	Sea_Video_Seq_1				Sea_Video_Seq_2			
Candidate Feature Sets	PM_1	PM_2	PM_3	PM_4	PM_1	PM_2	PM_3	PM_4
I, x, y	0.099	0.513	4.98	6.35	0.184	0.807	2.49	2.99
I, x, y, GO	0.245	0.541	5.00	6.28	0.208	0.783	1.97	2.53
$I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}, GM, GO$	0.692	0.816	4.21	5.78	0.746	0.120	25.93	31.63
I, x, y, GM	0.158	0.834	1.45	1.83	0.127	0.863	1.77	2.03
$I, x, y, \partial_{2,x}, \partial_{2,y}$	0.376	0.638	2.93	3.30	0.127	0.710	1.48	1.74
I, x, y, GM, GO	0.066	0.908	0.99	1.12	0.120	0.890	1.61	1.81
$I, x, y, \partial_{1,x}, \partial_{1,y}$	0.001	0.676	2.06	2.74	0.131	0.859	2.13	2.40
$I, x, y, \partial_{2,x}, \partial_{2,y}, GM$	0.524	0.791	8.39	9.87	0.265	0.843	1.51	1.87
$I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}$	0.359	0.690	3.37	4.27	0.115	0.894	2.03	2.24
$I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}, GM$	0.563	0.975	7.10	7.48	0.513	0.258	14.06	17.29

set $\{I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}, \mathrm{GM}, \mathrm{GO}\}$ contains more features, it does not have a positive effect on the tracking performance. Nevertheless, the set of gradient-based features was found to be useful on sea-surface target tracking scenarios. In Figs. 11 and 12, the sea-surface target tracking results obtained by using gradient-based features are presented for Sea_Video_Seq_1 and Sea_Video_Seq_2, respectively. In Figs. 11 and 12, the sea-surface targets are marked with a rectangular marker and the line stands for the trajectory of the target from the initial frame to the current frame.

Finally, the tracking scenario of a scale-varying sea-surface target is presented in Fig. 13. In the video, there is a slow moving cargo ship on the horizon. The operator zoomed into the cargo ship, causing significant scale changes within a few hundred frames. The tracking results presented in Fig. 13 show that the tracker can successfully handle sudden scale changes due to zooming.

5.3 *Tracking of Aerial Targets*

In this section, the candidate feature sets for aerial surveillance scenarios determined in Sec. 3 are used in the covariance-based tracker. The performance of the tracker is evaluated by using the performance measures on two visual band videos. The first video used for aerial surveillance (Air_Video_Seq_1) consists of 187 frames of dimension 640×480. The video contains a helicopter moving away from the capture device. The video was captured on a windy day and the capture device was not properly stabilized. Therefore, there are some vibrations and sudden movements that reduce the quality of the captured video and make the target tracking task more complicated. The second video used for aerial surveillance (Air_Video_Seq_2) consists of 100 frames. The video contains a trainer aircraft moving fast to the right of the capture device. The video was captured on a windy day and the capture device was not properly fixed to the ground.

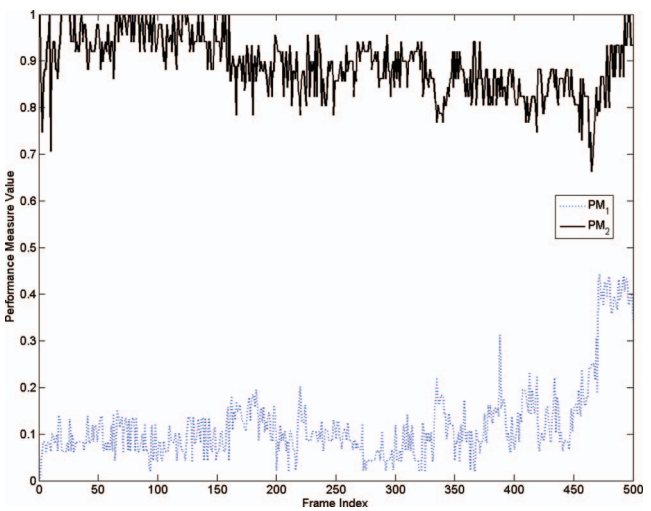

Fig. 9 The frame-wise PM_1's and PM_2's of the video Sea_Video_Seq_2.

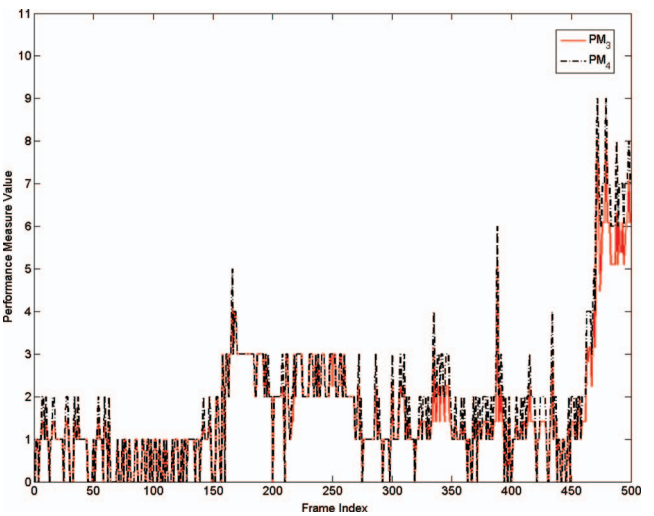

Fig. 10 The frame-wise PM_3's and PM_4's of the video Sea_Video_Seq_2.

Fig. 11 The illustration of the sea-surface target tracking in certain frames of Sea_Video_Seq_1.

Fig. 13 The illustration of the sea-surface target tracking when the target scale changes drastically.

Consequently, there are abrupt movements, vibrations, and flicker effects in both of the video sequences.

The performance of each feature set used in the tracker is examined by running the tracker on the videos and comparing the target locations obtained by the tracker with the ground truth target locations. The tracker parameters τ, e_0, e_1, e_2, e_3, and N for air surveillance scenarios are selected as 10, 0.4, 0.5, 0.1, 0.0019, and 3, respectively. As in the sea surveillance case, these values are obtained experimentally by considering different cases. The performances of the candidate feature sets on aerial surveillance videos are given in Table 4. The results indicate that, similar to the sea surveillance scenario, the gradient-based feature sets $\{I, x, y, \mathrm{GM}, \mathrm{GO}\}$, $\{I, x, y, \mathrm{GO}\}$, and $\{I, x, y, \mathrm{GM}\}$ provide satisfactory tracking performance on both videos. In addition, the feature set $\{I, x, y, \partial_{1,x}, \partial_{1,y}\}$ outperforms most of the feature sets in the aerial surveillance scenarios.

Since the aerial surveillance scenarios do not contain complex backgrounds, i.e., the gray colored aircraft and low-contrast sky background, the availability of the feature sets that enable robust tracking increases. As an example, the feature set $\{I, x, y\}$, which seemed to be one of the most simple candidate sets, provides an acceptable performance in aerial target tracking.

Fig. 12 The illustration of the sea-surface target tracking in certain frames of Sea_Video_Seq_2.

Parallel to the results given for the sea surveillance scenario, adding more features may not guarantee an increase in the tracking performance. As an example, the feature set $\{I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}, \mathrm{GM}\}$ does not provide better tracking results on both videos. Therefore, it is reasonable to use the feature set that has the lowest computational cost on aerial target tracking scenarios. But, as the background complexity increases (existence of target colored clouds or target-like structures), it is better to use gradient-based feature sets that may provide more discriminative power. In Figs. 14 and 15, the aerial target tracking results obtained by using gradient-based features are presented for Air_Video_Seq_1 and Air_Video_Seq_2, respectively. In Figs. 14 and 15, the aerial targets are marked with a rectangular marker and the line stands for the trajectory of the aerial target from the initial frame to the current frame. In Figs. 14 and 15, there exists some camera movements to locate the target to the center of the screen. These movements affect the appearance of the target trajectory and cause the trajectory lines to intersect at certain locations. For example, in Fig. 14, the target moves out of the SR due to an abrupt camera movement at a certain frame. The tracker handles the situation with the SR and catches the target again in the next frame. This illustration is a good example of how the tracker handles the instantaneous changes in target's position and shape.

Finally, the tracking scenario of a scale-varying aerial target is presented in Fig. 16. The scenario shows that the tracker can easily handle gradual scale changes such as an aircraft approaching the capture device. The tracker can also compensate sudden scale changes such as zooming.

5.4 *Performance Comparison Using Different Target Trackers*

As an additional experiment, the performance of the proposed target tracker scheme is compared with the standard target tracker techniques including correlation,[18] KLT feature,[37–39] and SIFT[10] based trackers. The performance of these trackers in both sea-surface and aerial videos are presented in Table 5.

Table 4 The performance of candidate feature sets in aerial target tracking.

	Air_Video_Seq_1				Air_Video_Seq_2			
Candidate Feature Sets	PM_1	PM_2	PM_3	PM_4	PM_1	PM_2	PM_3	PM_4
I, x, y, GM, GO	0.085	0.666	0.87	1.05	0.230	0.785	1.84	2.13
I, x, y, GO	0.085	0.653	0.95	1.16	0.244	0.805	1.69	1.96
I	0.585	0.407	2.36	2.95	0.432	0.550	3.61	4.66
$I, x, y, \partial_{1,x}, \partial_{1,y}$	0.091	0.687	0.88	1.02	0.266	0.807	1.80	2.09
I, x, y, GM	0.072	0.647	0.91	1.07	0.270	0.779	2.27	2.58
$I, x, y, \partial_{2,x}, \partial_{2,y}$	0.069	0.651	0.89	1.04	0.293	0.727	2.92	3.29
$I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}, GM$	0.096	0.659	0.94	1.15	0.284	0.757	2.46	2.81
$I, x, y, \partial_{2,x}, \partial_{2,y}, GM$	0.100	0.656	0.90	1.10	0.303	0.747	2.51	2.91
$I, x, y, \partial_{1,x}, \partial_{2,x}, \partial_{1,y}, \partial_{2,y}$	0.082	0.664	0.85	1.03	0.324	0.772	2.91	3.27
I, x, y	0.091	0.681	0.90	1.09	0.247	0.749	2.37	2.73

In the literature, correlation-based tracking approaches are widely used due to their naive structures and low computational cost. Although many of the correlation-based approaches provide low computational cost and perform at real-time, they are too sensitive to the occlusion and rapid changes in the target appearance. For performance comparison purposes, an image correlation-based tracker is implemented. The correlation-based tracker is initiated by selecting the target manually. It uses the normalized cross-correlation function in order to obtain a measure that indicates the quality of the match between the target template and the candidate target region in the ROI. The candidate region providing the largest matching measure is determined as the best matched region. If the largest matching measure is greater than a predefined threshold (selected as 0.9 in the experiments), the target template is updated using the Euclidean distance-based update strategy defined in Eq. (7). Otherwise, the target template is not updated. In the experiments, the parameter α

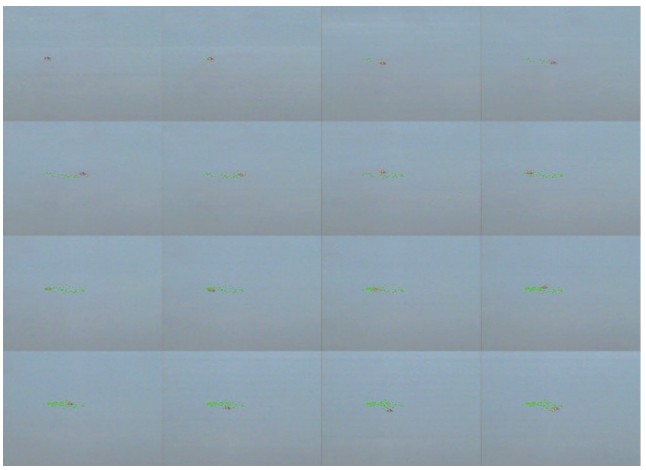

Fig. 15 The illustration of the aerial target tracking in certain frames of Air_Video_Seq_2.

Fig. 14 The illustration of the aerial target tracking in certain frames of Air_Video_Seq_1.

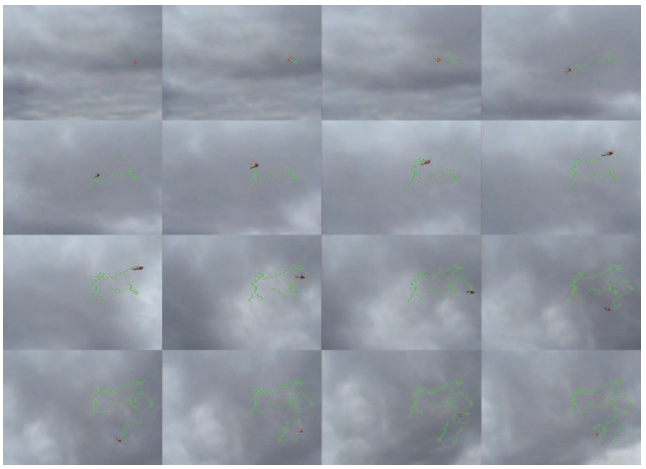

Fig. 16 The illustration of the aerial target tracking when the target scale changes gradually.

Table 5 The performance of reference trackers in both sea-surface and aerial target tracking.

	Sea_Video_Seq_1				Sea_Video_Seq_2				Air_Video_Seq_1				Air_Video_Seq_2			
Reference Trackers	PM_1	PM_2	PM_3	PM_4	PM_1	PM_2	PM_3	PM_4	PM_1	PM_2	PM_3	PM_4	PM_1	PM_2	PM_3	PM_4
Correlation	0.195	0.701	3.52	4.06	0.021	0.855	1.24	1.56	0.464	0.524	2.08	2.47	0.672	0.372	5.37	7.10
KLT feature	0.918	0.123	29.43	32.29	0.697	0.571	3.47	4.76	0.534	0.367	5.73	7.09	0.142	0.142	6.49	8.48
SIFT	0.305	0.612	4.77	5.82	0.167	0.234	10.92	14.06	0.217	0.292	3.39	4.47	0.556	0.564	5.49	6.82

of the Euclidean-based update strategy is selected as 0.05. The implemented correlation tracker performs at real-time indeed, the track is lost when there is an occlusion or sudden changes in the target appearance. The performance of the correlation tracker on sea-surface and aerial scenarios is provided in Table 5. Especially on Sea_Video_Seq_2, the correlation tracker provides satisfactory results. However, the video sequence Sea_Video_Seq_2 contains neither any rapid changes nor partial occlusions in the target model. The correlation tracker starts to perform poorly as in the scenarios Sea_Video_Seq_1, Air_Video_Seq_1, and Air_Video_Seq_2, if the target model is exposed to rapid changes and partial occlusion. Additionally, an example set of tracking result images that illustrate the failure of the correlation tracker in the case of rapid target appearance and scale changes is presented in Fig. 17. Although the covariance tracker still tracks the target under sudden changes in the target model due to its scale invariance property, the correlation tracker loses the track under the same circumstances.

The second tracker used in the performance comparison is the KLT feature tracker. In the literature, this method is used to extract interest points and match these features in the subsequent frames. KLT feature tracker is a sparse optical flow method based on three assumptions: constant brightness, small movements in time, and coherent motion at neighboring elements. A sparse and iterative version of the KLT feature tracker in a pyramidal structure, described in Ref. 40, is implemented for performance comparison in this study. The target is selected manually in order to initiate the KLT feature tracker. A 10×10 search window and 5 level pyramid is used for tracking purposes in the implementation. Similar to the correlation tracker, the KLT feature tracker performs beyond real-time but it is even much more sensitive to the occlusion than the correlation tracker. The performance of the KLT feature tracker on both sea-surface and aerial scenarios are provided in Table 5. According to tracking results provided in Table 5, the KLT feature tracker performs poorly on both sea-surface and aerial surveillance scenarios. One reason for this poor performance is the KLT feature tracker's structure. Since the KLT tracker tracks a number of strong features, it cannot localize the target gate properly. Therefore the PM_i's used in this paper are not eligible for the KLT feature tracker. Another reason for the poor performance of the KLT feature tracker is its inability to handle any level of occlusion. In addition to the tracking results given in Table 5, an example set of tracking result images that illustrate the failure of the KLT feature tracker in the case of partial occlusion is presented in Fig. 18. By looking at Fig. 18, one can say that the KLT feature tracker is not robust to any level of occlusion. Since the KLT feature tracker is not a target-model–based tracker, it does not provide any information about the target shape. Due to these disadvantages, the proposed covariance descriptor-based tracker framework outperforms the KLT feature tracker in both sea and air surveillance scenarios.

The last tracker implemented for performance comparison is the SIFT-based tracker.[10] In the initialization of the tracker, the target model is selected manually. Then, the scale-invariant features proposed by Lowe are extracted from the target template and from the search region, i.e., ROI. These extracted features are invariant to scale changes due to the pyramidal structure of the feature extraction scheme. The features, i.e., SIFT descriptors, extracted from both target and search region are matched with an Euclidean distance-based measure. If the measure is greater than a threshold, the descriptors are considered as matched. However, there may be false matches between the descriptors that cause information losses in the target model and track loss. Therefore, a match pruning mechanism is implemented as a post-processing operation using a gradient clustering type of approach. By this way, the outliers of the gradient of the matching lines are removed. Although SIFT features are known as powerful features, the computational complexity of the implementation is high due to the pyramidal feature extraction process with additional post-processing operations. Moreover, the SIFT-based tracker is not suitable to track point/point-like targets because the discriminative SIFT-based features cannot be extracted from these types of targets. In our case, the SIFT-based tracker cannot extract important interest points from the aerial targets that have a tendency to be point-like structures. The SIFT-based tracker performs poorly on both sea and aerial surveillance videos as given in Table 5, since it

Fig. 17 The correlation tracker fails to track the target when there exists a rapid change in the target model.

Fig. 18 The KLT feature tracker is far too sensitive to the partial occlusion. It cannot track the desired target.

cannot locate the target position properly. The SIFT features extracted from the target template may not span the whole target model that provides information and localization losses in the target model. Therefore, the proposed covariance tracking scheme outperforms the SIFT-based structure, especially in aerial target tracking.

The experimental studies show that the proposed tracking system outperforms classical target tracking algorithms. By comparing the results in Tables 3 and 4 with Table 5, one can generalize that PM_1, PM_3, and PM_4 obtained with the reference trackers is greater than the results obtained using the proposed tracking scheme. Similarly, PM_2 values face a harsh decrease when the reference trackers are used instead of the proposed tracking scheme. Although there exists track losses in the reference trackers, we computed performance measures for the completeness of the comparisons. The track losses cause a harsh decrease in the tracker performance and may not be suitable for a fair comparison between the proposed tracking scheme and baseline tracking techniques.

6 Conclusion

In this paper, an offline feature selection system is proposed for the regional covariance descriptor-based target tracking framework to track sea-surface and aerial targets in visual band videos in a robust way. Due to the low computational complexity of the covariance descriptors, the proposed target tracking scheme can be easily adapted to the real-time applications. The contribution of the paper is to obtain feature characteristics for both sea-surface and aerial targets by performing an offline feature selection followed by feature evaluation based on tracking using several performance measures. In order to extract the feature characteristics of different platforms, an offline target-background classification mechanism followed by a tracking performance test is proposed. Here, the successful sets of features for classification were determined as candidates for promising tracking performance. The experimental results show that the background complexity is the major factor that affects the feature selection. As the background complexity increases, more complex features that have more discriminative power are needed. In our case, the background in sea-surveillance is more complex than the background in aerial surveillance. Therefore, the gradient-based features that contain the magnitude and the orientation of the intensity changes in both directions provide better tracking results in sea surveillance scenarios. Based on our experimental observations, the addition of more features was observed not to necessarily guarantee a better target representation in both surveillance scenarios, providing a sample case of "curse of dimensionality". The target search strategy does not use the kinematic information of the target in the successive frames and therefore is robust to abrupt target changes. The tracking performance of the proposed algorithm is compared with the classical tracking methods including correlation, KLT feature, and SIFT-based trackers. Experimental results show that the proposed tracking scheme outperforms these classical baseline methods.

Despite the difficulty of obtaining real-life experimental video sequences, great effort was paid in achieving a reasonably large data set with different backgrounds, weather conditions, and platforms for the sake of robust offline feature selection and evaluation. Nevertheless, it is clear that the proposed approach can be readily extended to various different scenarios and other imaging systems, including infrared imaging, by the same offline feature analysis idea (proposed herein).

As a future work, an online feature selection technique with real-time processing capability is aimed to be developed. Additionally, a target motion estimator-based framework is intended to be implemented and used together with the proposed tracking scheme in order to provide a compact solution for target tracking applications.

Acknowledgments

This study is supported by Project No. 109A001 in the framework of TÜBİTAK 1007 Program. The authors would like to thank Dr. S. Gökhun Tanyer for his support in this study, A. Onur Karalı for his efforts in video capture and helpful discussions, and the anonymous reviewer for his/her constructive critiques and suggestions.

References

1. A. Yılmaz, J. Omar, and S. Mubarak, "Object tracking: A survey," *ACM Comput. Surv.* **38**(4), 1–45 (2006).
2. F. Porikli, "Achieving real-time object detection and tracking under extreme conditions," *Real-Time Image Process.* **1**(1), 33–40 (2006).
3. K. Nishida, T. Kurita, Y. Ogiuchi, and M. Higashikubo, "Visual tracking algorithm using pixel-pair feature," in *Proceedings of the IEEE International Conference on Pattern Recognition* pp. 1808–1811 (2010).
4. Y. Zhou, A. Mayyas, A. Qattawi, and M. Omar, "Feature-level and pixel-level fusion routines when coupled to infrared night-vision tracking scheme," *Infrared Phys. Technol.* **53**, 43–49 (2010).
5. J. Czyz, B. Ristic, and B. Macq, "A particle filter for joint detection and tracking of color objects," *Image Vis. Comput.* **25**, 1271–1281 (2007).
6. D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 142–149 (2000).
7. P. Brasnett, L. Mihaylova, D. Bull, and N. Canagarajah, "Sequential Monte Carlo tracking by fusing multiple cues in video sequences," *Image Vis. Comput.* **25** 1217–1227 (2007).
8. S.-X. Li, H.-X. Chang, and C.-F. Zhu, "Adaptive pyramid mean shift for global real-time visual tracking," *Image Vis. Comput.* **25**, 1271–1281 (2007).
9. R. V. Babu, P. Perez, and P. Bouthemy, "Robust tracking with motion estimation and local Kernel-based color modeling," *Image Vis. Comput.* **25**, 1206–1216 (2007).
10. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comp. Vis.* **60**, 91–110 (2004).
11. C. Park, K. Baea, and J.-H. Jung, "Object recognition in infrared image sequences using scale invariant feature transform," in *Proceedings of the SPIE Signal Processing, Sensor Fusion, and Target Recognition XVII*, I. Kadar, Ed., Vol. 6968, pp. 1–9 (2008).
12. H. Lee, P. G. Heo, J.-Y. Suk, B.-Y. Yeou, and H. Park, "Scale-invariant object tracking method using strong corners in the scale domain," *Opt. Eng.* **48**, 1–9 (2010).
13. P. B. W. Schwering, H. A. Lensen, S. P. Broek, R. J. M. Hollander, W. Mark, H. Bouma, and R. A. W. Kemp, "Application of heterogeneous multiple camera system with panoramic capabilities in a harbor environment," in *Proceedings of the SPIE Electro-Optical and Infrared Systems: Technology and Applications VI*, D. A. Huckridge and R. R. Ebert, Eds., pp. 1–11 (2009).
14. L. Jing-zheng, Y. Xu-chu, G. Ling-qing, and Y. Wen-shuai, "Automatic matching of infrared image sequences based on rotation invariant," in *Proceedings of the IEEE International Conference on Environmental Science and Information Application Technology*, pp. 365–368 (2009).
15. Y. Wang, A. Camargo, R. Fevig, F. Martel, and R. .R. Schultz, "Image mosaicking from uncooled thermal ir video captured by a small uav," in *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 161–164 (2008).
16. H. P. Moravec, "Visual mapping by a robot rover," in *International Joint Conference on Artificial Intelligence*, Tokyo, Japan, pp. 598–600 (1979).
17. C. Harris and M. J. Stephens, "A combined corner and edge detector," in *Alvey Vision Conference*, Manchester, UK, pp. 147–152 (1988).
18. S. M. A. Bhuiyan, M. S. Alam, and M. Alkanhal, "New two-stage correlation-based approach for target detection and tracking in forward-looking infrared imagery using filters based on extended maximum average correlation height and polynomial distance classifier correlation," *Opt. Eng.* **46**(8), 1–14 (2007).
19. D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proceedings of*

the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2544–2550 (2010).
20. A. Mahalanobis, R. R. Muise, and S. R. Stanfill, "Quadratic correlation filter design methodology for target detection and surveillance applications," *App. Opt.* **43**(27) 5198–5204 (2004).
21. O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proceedings of the IEEE European Conference on Computer Vision*, pp. 589–600 (2006).
22. F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Proceedings of the IEEE International Conference on Pattern Recognition*, pp. 728–725 (2006).
23. V. N. Vapnik, *Estimation of Dependences Based on Empirical Data*. [in Russian], Moscow: Nauka (1979). (English translation: New York: Springer Verlag, 1982).
24. M. Diani, A. Baldacci, and G. Corsini, "Novel background removal algorithm for navy infrared search and track systems," *Opt. Eng.* **40**(8), 1729–1734 (2001).
25. Y. Xiong, J. Peng, and M. Ding, "An extended track-before-detect algorithm for infrared target detection," *IEEE Trans. Aerosp. Electron. Syst.* **33**(3), 1087–1092 (1997).
26. M. de Visser, P. B. W. Schwering, J. G. De Groot, and E. A. Hendriks, "Passive ranging using an infrared search and track sensor," *Opt. Eng.* **45**, 1–14 (2006).
27. C. Bibby and I. Reid, "Visual tracking at sea," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1802–1805 (2005).
28. H. Wei, H. Nguyen, P. Ramu, C. Raju, X. Liu, and J. Yadegar, "Automated intelligent video surveillance system for ships," in *Proceedings of the SPIE Conference on Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI*, Vol. 7306, pp. 1–9 (2009).
29. F. Robert-Inacio, A. Raybaud, and E. Clement, "Multispectral target detection and tracking for seaport video surveillance," in *Proceedings of the IVS Image and Vision Computing New Zealand*, pp. 169–174 (2007).
30. S. Fefilatyev, "Detection of marine vehicles in images and video of open sea", PhD thesis, University of South Florida, Florida (2008).
31. F. D. Lapierre, A. Borghgraef, and M. Vandewal, "Statistical real-time model for performance prediction of ship detection from microsatellite electro-optical imagers," *EURASIP J. on Appl. Signal Process.* **2010**, 1–16 (2010).
32. R. Wijnhoven, K. Rens, E. G. T. Jaspers, and P. H. N. With, "Online learning for ship detection in maritime surveillance," in *Proceedings of the WIC Thirty-first Symposium on Information Theory*, pp. 73–80 (2010).
33. C.-C. Chang and C.-J. Lin, www.csie.ntu.edu.tw/cjlin/libsvm, LIBSVM: *a library for support vector machines* (2001).
34. S. Günal, Ö. N. Gerek, D. G. Ece, and R. Edizkan, "The search for optimal feature set in power quality event classification," *Expert Syst. Applic.* **36**(7), 10266–10273 (2009).
35. W. Forstner and B. Moonen, "A metric for covariance matrices," Technical Report, Department of Geodesy and Geoinformatics, Stuttgart University (1999).
36. X. Ding, C. Huang, and F. Huang, "Region covariance based object tracking using Monte Carlo method," in *Proceedings of the IEEE International Conference on Control and Automation*, pp. 1802–1805 (2010).
37. C. Tomasi and T. Kanade, "Detection and tracking of point features," Technical Report, Carnegie Mellon University Technical Report (1991).
38. B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence*, Vancouver, Canada, pp. 674–679 (1981).
39. J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–10 (1994).
40. J.-Y. Bouguet, "Pyramidal implementation of the Lucas-Kanade feature tracker," Technical Report, Intel Corp., Microelectronics Research Laboratory (2000).

Serdar Çakır received his BSc degree from Eskişehir Osmangazi University in 2008. Immediately after graduation he joined Bilkent University, and got his MSc degree in electrical and electronics engineering in 2010. He joined TÜBİTAK BİLGEM UEKAE/İLTAREN Research and Development Group in 2010, where he is currently a research scientist. He also continues his graduate study at Bilkent University as a PhD candidate under the supervision of Professor A. Enis Çetin. His main research interests are image/video processing, computer vision, and pattern recognition.

Tayfun Aytaç received his BSc degree in electrical engineering from Gazi University, Ankara, Turkey in 2000 and MSc and PhD degrees in electrical and electronics engineering from Bilkent University, Ankara, Turkey in 2002 and 2006, respectively. During his graduate studies, he was a research and teaching assistant at Bilkent University, Department of Electrical and Electronics Engineering. He joined TÜBİTAK BİLGEM UEKAE/İLTAREN Research and Development Group in 2006, where he is currently a senior research scientist. His current research interests include infrared imaging systems, automatic target recognition, target tracking and classification, and sensor-based robotics.

Alper Yıldırım received a BSc degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 1996, an MSc degree in digital and computer systems from Tampere University of Technology, Tampere, Finland, in 2001, and a PhD degree in electronics engineering from Ankara University, Ankara, in 2007. He was a design engineer with Nokia Mobile Phones, Tampere. He is currently a chief research scientist with the Scientific and Technological Research Council of Turkey, Ankara. His research interests include digital signal processing, optimization, and radar systems.

Ö. Nezih Gerek received BSc, MSc, and PhD degrees in electrical engineering from Bilkent University, Ankara, Turkey, in 1991, 1993, and 1998, respectively. During his PhD studies, he spent a semester in the University of Minnesota as an exchange researcher in an NSF project. Following his PhD degree, he spent 1 year as a research associate at EPFL, Lausanne, Switzerland. Currently, he is a full professor of electrical engineering at Anadolu University, Eskişehir. He is also a member of the Electrical, Electronics and Informatics Research Fund Group of the Scientific and Technological Research Council of Turkey. He is in the Editorial Board of Turkish Journal of Electrical Engineering and Computer Science and Elsevier: Digital Signal Processing. His research areas include signal analysis, image processing, and signal coding.