International Journal of Natural Sciences Research
2016 Vol. 4, No. 5, pp. 92-99
ISSN(e): 2311-4746
ISSN(p): 2311-7435
DOI: 10.18488/journal.63/2016.4.5/63.5.92.99

MOVING TARGET TRACKING IN INFRARED IMAGE SEQUENCES BASED ON DIFFERENTIAL KERNEL COVARIANCE DESCRIPTOR

Burak ALPTEKİN[1]† --- Hamidullah BİNOL[2] --- Hüseyin ÇUKUR[3]
[1,2,3]*Department of Electronics and Communications Engineering, Yildiz Technical University Istanbul, Turkey*

ABSTRACT

Forward looking infrared (FLIR) imaging has been used in many areas of research and everyday life, but it has been mostly employed in military and security domains. In these fields, remote infrared target tracking is a crucial element for surveillance. However, long-range captured IR image sequences generally have poor contrast, variable illumination, and high background clutter. These challenges make target tracking difficult. This paper suggests a technique for target tracking in different ranges in challenging FLIR image sequences, based on Differential Kernel Covariance Descriptor (DKCD). This new method diminishes rotation and illumination variation effects. The proposed technique calculates the differential kernel matrix of reference target by using various statistical and spatial features such as first and second derivatives, location information, and the intensity value of pixels. Later, the differential covariance matrix is constructed by using different pixel features and applying the appropriate kernel function to the matrix. Thanks to the kernel functions, the algorithm redefines the target's differential spatial features in Hilbert space. This process makes the descriptor non-linear. The predicted position of the target is calculated with the nearest neighbor algorithm in the candidate regions in the sub-frame. The performance of the suggested single target tracking system is then tested on challenging real-life video sequences.

Keywords: Differential kernel covariance descriptor, Visual tracking, Infrared image sequence.

Received: 2 November 2016/ **Revised:** 2 December 2016/ **Accepted:** 7 December 2016/ **Published:** 14 December 2016

Contribution/ Originality

This paper proposes a new nonlinear descriptor which mainly uses kernel covariance matrix based on difference of features. The technique minimizes the effect of pose variation, illumination, size, and background changes.

1. INTRODUCTION

Thanks to the advancement in thermal camera technology in recent years, it is possible to get images of better quality and resolution at a lower cost and magnitude. This has led to a rise in the use of thermal imaging, especially in the military, industry, and medical domains. For example, it is used for communication devices and navigation systems by the military; environmental control systems, astronomical spectroscopy, and atmospheric chemistry surveys in industry; and blood sugar measurement, pulse oximetry, and functional neuron imaging in the medical field.

In military and security usage, IR cameras are mainly employed for long-range surveillance and passive infrared homing missile systems. The main advantage of thermal cameras in the military domain is their ability to capture in total darkness. Missile homing systems and surveillance cameras seek the heat emanating from the target. Many objects such as people, vehicles, tanks, and aircraft can be defined as targets that generate and retain

† Corresponding author

heat. However, long-range captured IR images generally have poor quality with low contrast, variable illumination, and high background clutter. The reason for this is that water vapor and certain gases in the atmosphere absorb infrared radiation making the target description, recognition, and tracking in IR images a challenging problem. Descriptor selection is the most important step for tracking. A descriptor should define the target with discriminative and fast calculating features while making it invariant to illumination, size, rotation, and particularly, the noise.

Most techniques used for tracking images in IR presume the target is hotter than its background with an acceptable threshold. This was followed by histogram and fast histogram construction methods that were extensively used for tracking and representation. Probabilistic prediction methods such as Kalman filter [1] were then employed. Comaniciu [2-4] uses the mean shift method which consists of color histograms. Color-based particle tracker [5] is a robust algorithm but it is affected by fast motion and illumination change. Dawoud, et al. [6] utilized the weighted composite reference function for modeling the target. Tuzel, et al. [7] described the target with the covariance matrix which uses various statistical and spatial features. Covariance Descriptor (CD) depends on the linear relation of the pixel value. These cannot describe or represent non-linear dependence on a feature in an appropriate manner. Arif and Vela [8] used a combination of kernel principal component analysis and covariance region description for spatial covariance matrix while using the kernel principal component analysis for learning. The second process of this technique makes the descriptor non-linear. Wu, et al. [9] present a kernel-based region covariance descriptor for differential tracking.

The proposed technique, Differential Kernel Covariance Descriptor (DKCD), offers non-linear covariance matrix for describing the target. The algorithm first calculates the spatial and statistical properties of each pixel. Then a differential covariance matrix is built with the difference of each pixel's properties. Finally the kernel function is applied to this matrix to make the descriptor non-linear. The predicted new position of the target is found with the nearest neighbor algorithm among the candidate regions in the sub-frame.

Section 2 of this paper describes the target tracking algorithm with the features used, the covariance matrix, and DKCD. We demonstrate the performance of the algorithm followed by a detailed comparison in Section 3. Section 4 concludes this paper.

2. TARGET TRACKING

A. Spatial and Statically Feature Matrix Used for Description

The descriptor uses a specific feature that is exclusive to the target. The success of the algorithm in tracking the target, while rejecting a false alarm, is dependent on this exclusive feature. The other important subject for feature selection is fast and easy computation. In this technique, we use first and second derivative of intensity with respect to horizontal and vertical direction, location information of pixels, and intensity value. These seven features characterize the reference image. One can decrease or increase the number of features used. A rise in features leads to an increase in the time taken for calculation. Dimension of created descriptor is W x H x d.

$$F(x) = \phi(I, x, y, dx, dy, dx^2, dy^2) \qquad (1)$$

Function ϕ is an operator which maps the *d*-dimensional feature.

B. Differential Covariance Matrix

Classical CD is a covariance of the *d*-dimensional feature matrix ($F(x,y)$). Calculating thus: (quantity of features) × (quantity of features) produces the resultant covariance matrix given in Equation 2.

$$C_R = \frac{1}{n-1}\sum_{k=1}^{n}(z_k-\mu)(z_k-\mu)^T \qquad (2)$$

The selected region is R. z_k is the *d*-dimensional feature point inside R. μ represents means of features. n represent amount of pixel. The proposed technique fuses various spatial and statistical features. The diagonal elements of the matrix are variations while the other elements are covariance of features and diagonally symmetrical. Covariance matrix as a descriptor possesses some superiority. The size of the matrix is independent from the size of the selected region. The quantities of features used determine the size of the matrix so that CD consists of a low-dimensional matrix. It provides real-time tracking, but scale and rotation changes affect it in a limited manner.

The proposed technique uses a new type of covariance matrix. It is built with the feature that shows the difference of pixel value. Each pixel in the selected region has seven different values.

$$F(x) = \phi(I, x, y, dx, dy, dx^2, dy^2)$$

$$G_R = \Phi_1^m(F_{i.} - F_j) \qquad (3)$$

Function Φ is the operator which maps the *d*-dimensional differential covariance matrix (G_R). The descriptor can identify the target in various views, poses, and distance.

C. Differential Kernel Covariance Descriptor (DKCD)

CD is excellent in many tracking scenarios but covariance matrix consists only of linear relations of pixels. An improved technique is required for more complicated scenarios. In CD, the target is placed in a rectangular window. When high scale and orientation change, the structure of the matrix changes as well. The distribution of the background and target information too changes inevitably. This circumstance causes loss of important data and interfaces more background information. In this technique, we can utilize windows of various shapes depending on the shape of the target as this decreases the effect of background noise. Kernel covariance matrix offers non-linear covariance matrix for describing the target. The suggested technique minimizes the effect of pose variation, illumination, size, and background changes. You can use different types of kernels such as Laplacian, Gaussian, Cauchy kernel, etc., depending on the selected image sequences. These superiorities make DKCD a successful tracker in IR image sequences. DKCD applies kernel function to *d*-dimensional differential feature matrix which changes the disruption of feature information and builds a new matrix. Diagonal elements of the matrix are differential variations; other elements are covariance of features and a diagonally symmetrical matrix like CD. We give only the results of Gaussian kernel among the several popular kernel functions which can be seen in Eq. 4-7:

Gaussian Kernel

$$K(x_i, x_j) = (2\pi)^{-\sigma/2}\exp(\frac{1}{2}\left\|x_i - x_j\right\|^2) \qquad (4)$$

Exponential Kernel

$$K(x_i, x_j) = \exp(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma}) \qquad (5)$$

Logarithmic Kernel

$$K(x_i, x_j) = -\log(\left\|x_i - x_j\right\|^d + 1) \qquad (6)$$

Cauchy Kernel

$$K(x_i, x_j) = \frac{1}{1 + \frac{\|x_i - x_j\|^2}{\sigma^2}}) \qquad (7)$$

The value of σ*(variation)*, as is evident, changes depending on the image. σ controls the width of the kernel function. The algorithm includes selection of sub-frame, center point of target, fast factor, and the shape of RIO. *d*-dimensional feature matrix ($F(x,y)$) is created for the selected region. The differential covariance matrix is created later. The matrix is converted to DKCD by means of the Gaussian kernel function ($K(x_i, x_j)$).

D. Region Similarity Measure

Although machine learning algorithms mainly rely on Euclidean space, it is not appropriate for this technique. The nearest neighbor algorithm is used for distance calculation. The distance of features points [10] give the dissimilarity of two covariance matrices (C_1, C_2).

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^{n} In^2 \lambda_i (C_1, C_2)} \qquad (7)$$

where $\{ \lambda_i (C_1, C_2) \}_{i=1\ldots n}$ are the generalized eigen value of C_1 and C_2, computed from

$$\lambda_i C_1 x_i - C_2 x_i = 0 \quad i = 1.....d \qquad (8)$$

and x_i = 0 are the generalized eigenvectors. The instance measure ρ satisfies the metric axioms (1-3) for positive definite symmetric matrices *C1* and *C2*

1. ρ (*C1*,*C2*) $\leq$ 0 and ρ (*C1*,*C2*) = 0 only if *C1* = *C2*,
2. ρ (*C1*,*C2*) = ρ (*C2*,*C1*),
3. ρ (*C1*,*C3*) + ρ (*C1*,*C3*) $\geq$ ρ (*C2*,*C3*).

3. EXPERIMENTAL RESULTS

The performance of suggested single target tracking system is tested on challenging real-life video sequences. The data sets used in this work are Linköping Thermal IR (LTIR) data set [11] OTCBVS benchmark dataset collection [12] and our videos. The algorithm is tested on many video sequences, seven of which are shown in this paper. LTIR data set is used in the thermal infrared visual object tracking (VOT-TIR2015) challenge [11]. The different scenarios consisted object motion, background clutter, size change, aspect ratio change, target like clutter, partial occlusion object deformation, and scene complexity.

The stableman sequence consists of pose variation with legs and arms moving (Fig. 1). The background is relatively simple and the illumination is stable. The first image contains two different backgrounds that change with movement. As he goes further, the size of the man becomes smaller, fusing more background, and the head of horse interferes with the target. The tracker manages to track till the end of the video.

Fig-1. Tracking results of the stableman (The Linköping Thermal Infrared (LTIR) dataset).

A small merchant vessel is video-recorded from long range (Fig. 2). Although the background is simple, high clutter makes it look complex. In the first image, the ship covers nearly all ROI. As it goes further, the size of ship gets smaller and smaller, fusing more background. The illumination is stable. The tracker manages to track till the end of video.

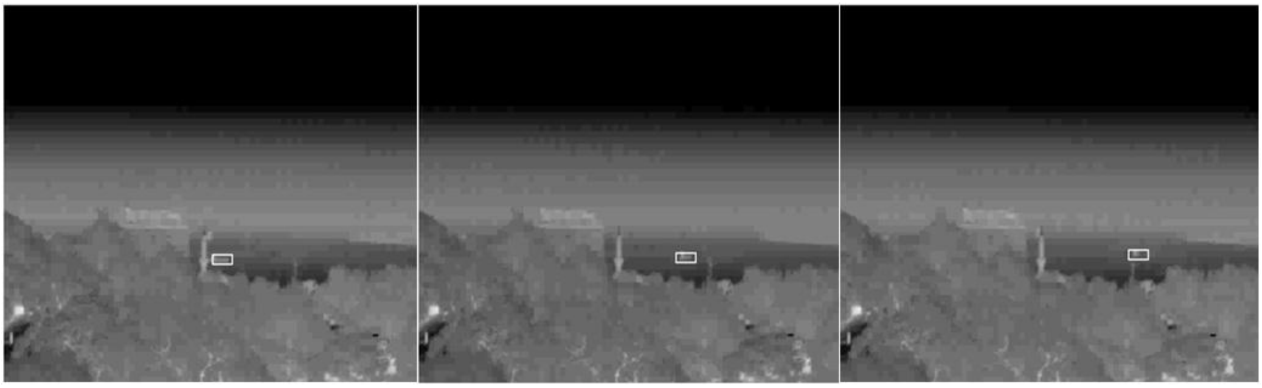

Fig-2. Tracking results of the small merchant vessel (Video is recorded by authors in Bosphorus).

The quadcopter is continuously moving and rotating in different directions (Fig. 3). It possess very little ROI. In the first image, the background is simple and plain. The quadcopter goes further towards a tree and its size gets smaller, fusing more background. When it is in the tree, the size becomes even smaller and the background interference is relatively high, as compared to the others. The background interference increases significantly since RIO is small. The tracker manages to track it till it reaches the tree.

Fig-3. Tracking results of the quadcopter (The Linköping Thermal Infrared (LTIR) dataset).

As Fig. 4 shows, as the dog is getting closer, its size is getting bigger with the ROI partly covering it, fusing less background. The scenario consists of pose variation and rotation and the background is simple and plain. The illumination is stable. The tracker manages to track till the end of video.

Fig-4. Tracking results of the dog (The Linköping Thermal Infrared (LTIR) dataset).

In this video, a man walks in a forest (Fig. 5). The background is complex. Trees interface the target and severe occlusion happens. Twice, the target is completely behind the tree. The algorithm manages to track till the end of the video impressively. The scale and illumination is stable.

Fig-5. Tracking results of the man in forest (OTCBVS Benchmark Dataset Collection)

A rhino grazes in a forest (Fig. 6). It is partly occluded along with pose variation at times. Other parameters are stable. The algorithm achieves good results.

Fig-6. Tracking results of the rhino (The Linköping Thermal Infrared (LTIR) dataset).

Target like clutter and severe occlusion happen in the crossing people video (Fig. 7).When the target is walking, another human crosses the target. Tracker tracks the target in all sequences perfectly with a high tracking accuracy. The noticeable factor giving good performance is the high contrast between the target and background.

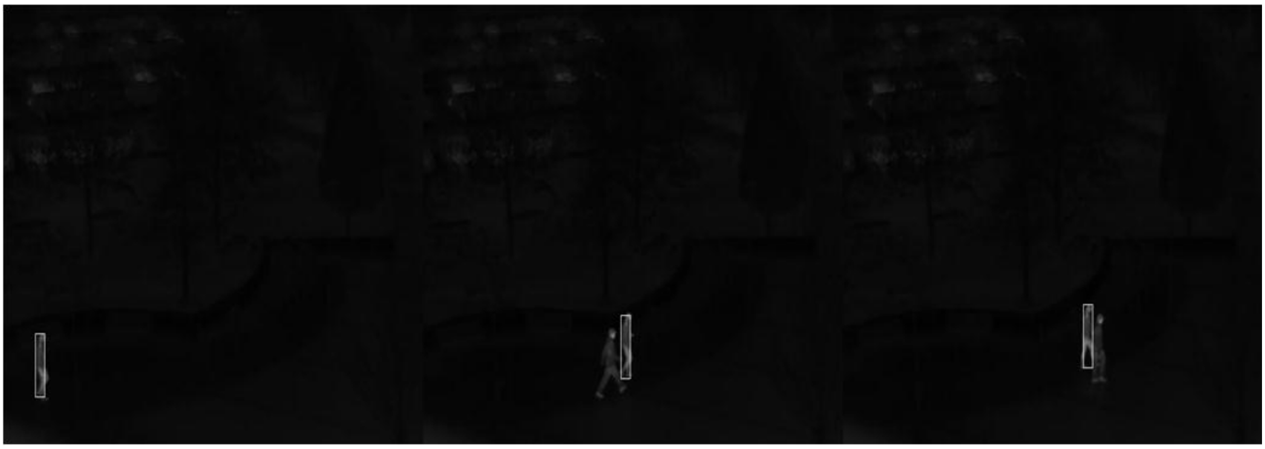

Fig-7. Tracking results of the crossing people (The Linköping Thermal Infrared (LTIR) dataset).

4. CONCLUSION

In this paper, we suggested a modified technique for single target tracking in IR image sequence based on differential kernel covariance descriptor. Differential kernel covariance matrix proposes a new non-linear image region descriptor which exhibits non-linear relation of each pixel. We tested the proposed technique in different challenging scenarios. Tracking results show that the new technique is suitable for popular problems and is far superior to most existing techniques. The disadvantage is that it's slower than classic covariance descriptor and kernel covariance descriptor.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Contributors/Acknowledgement: The authors would like to thank O. Tuzel, F. Porikli, P. Meer, and D. Comaniciu for rewarding discussions. The authors also thank the creators of Linköping Thermal IR (LTIR) data set [6] and OTCBVS benchmark dataset collection for sharing their real-life infrared video sequences.

REFERENCES

[1] P. Gutman and M. Velger, "Tracking targets using adaptive kalman filtering," *EEE Transactions on Aerospace and Electronic Systems*, vol. 26, pp. 691-699, 1990.

[2] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects usingmean shift," presented at the Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR Hilton Head Island, 2000.

[3] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603 – 619, 2002.

[4] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 564-577, 2003.

[5] K. Nummiaro, K.-M. Esther, and V. G. Luc, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, pp. 99-110, 2003.

[6] A. Dawoud, M. Alam, A. Bal, and C. Loo, "Target tracking in infrared imagery using weighted composite reference function-based decision fusion," *IEEE Transactions on Image Processing*, vol. 15, pp. 404-410, 2006.

[7] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. ECCV Conf.*, 2006, pp. 589–600.

[8] O. Arif and P. A. Vela, "Kernel covariance image region description for object tracking," in *16th IEEE International Conference on Image Processing (ICIP)*, 2009.

[9] Y. Wu, B. Ma, and P. A. Y. Jia, "Differential tracking with a kernel-based region covariance descriptor," *Pattern Analysis and Applications*, vol. 18, pp. 45–59, 2015.

[10] W. Forstner and B. Moonen, "A metric for covariance matrices," Technical Report, Dept. of Geodesy and Geoinformatics, Stuttgart University, 1999.

[11] A. Karteek, "The thermal infrared visual object tracking VOT-TIR2015 challenge results," presented at the IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, 2015.

[12] A. Berg, J. Ahlberg, and M. Felsberg, "A thermal object tracking benchmark," presented at the 12th IEEE International Conference on Advanced Video- and Signal-Based Surveillance, Karlsruhe, Germany, 2015.