# Thermal infrared object tracking using convolutional shallow features and kernelized correlation filter

Suryo Adhi Wibowo[1], Hansoo Lee[1], Eun Kyeong Kim[1], and Sungshin Kim[1]

Department of Electrical and Computer Engineering, Pusan National University,
Busan, South Korea,
{suryo, hansoo, kimeunkyeong, sskim}@pusan.ac.kr

**Abstract.** Object tracking is one of important research field in computer vision area. Especially for the thermal infrared object tracking, it can be applied for surveillance systems, military field, and etc. Unfortunately, there are several challenging problems in this research such as camera motion, dynamics change, motion change, occlusion, and also size change. In order to solve these problems, we proposed kernelized correlation filter using convolutional shallow features. Further, to handle the appearance change, the update parameters are also proposed. Furthermore, to validate the proposed method, extensive experiments are conducted by using VOT-TIR2016 benchmark dataset. The results from the experiment show that the proposed method performs favorably against state-of-the-art tracking algorithms.

**Keywords:** object tracking, visual tracking, thermal infrared, convolutional features, shallow layer

## 1 Introduction

Object tracking still become hot research topic in this year. It is because there are many applications related with this topic such as surveillance systems, robotics, virtual reality, augmented reality, and etc, where these applications will grow rapidly for the future. The purpose of object tracking is how to estimate the state of the target object when the initial information of the target object (e.g. centroid of the location of the target object, scale of the target object, and etc) is provided. Further, compared to the common cameras, thermal cameras have some advantages such as robust for changes in illumination, well perform in the darkness, and also robust to the shadow effects.

However, to make robust and accurate object tracking algorithm, one of several important factors that should be considered is how to represent the target object by using a feature. The color histogram features can be categorized as a simple features. It is because it can be obtained from the pixel values of the images directly, or it can be obtained by using a simple equation which represents color histogram. One of example of the utilizing color histogram features for object tracking is provided in [1]. The authors proposed multi-scale color features based on correlation filter to developed object tracking algorithm. Based on their results, this representation shows good for changes in motion and illumination problems. For an occlusions and change in size, this representation shown less robust than the others tracking algorithm. Points representation are used to represent the target object for object tracking in [2, 3]. Unfortunately, this representation does not robust when the tracker algorithm faced some problems such as change in illumination and motion, as well as change in size problems.

In [4], the authors used sparse coefficient vectors to represent the target object. This representation combined with particle filter to modelling the motion. Because particle filter is used to modelling the motion, the accuracy of their proposed method will be influenced with the number of particle that they used. Then, it also influences to the computation time. To address the computation time problem, [5] proposed fast generative approach based on sparse coefficient vector for visual tracking. Although well perform in occlusion problem, this representation has some drawbacks for changes in motion and size. Remembering the drawback from the color histogram representation, recently, to address this problem [6] proposed combination features between color histogram and histogram of oriented gradients (HOG) in the correlation filter framework. Further, [7] proposed distractor handling since this combination features failed when faced the distractor which has similar representation with the target object.

Beside that, currently, Convolutional Neural Networks (CNNs) shows excellent performance in several computer vision applications such as object recognition [8] and object detection [9]. Inspiring from this fact, in this paper, we proposed convolutional features to represent the target object, where these features are generated from the shallow layer of pre-trained CNN. Further, correlation filter is proposed to estimate the state of the target object. Finally, since this research is focused on thermal infrared object tracking, we perform extensive experiment on a visual object tracking-thermal infrared 2016 (VOT-TIR2016) benchmark dataset. Where this benchmark dataset consists of 25 videos. Based on the results, the proposed method performs favorably against state-of-the-art tracking algorithms.

The organization of this paper as follow: Section 2 gives explanations about the proposed method includes the kernelized correlation filter, convolutional shallow features and update parameters of the proposed method. Section 3 provides experiments results. Section 4 concludes this paper.

## 2 Proposed method

In this section, our proposed method which consists of kernelized correlation filter, convolution shallow features, and update parameters are described. Figure 1 shows the framework of the proposed method.

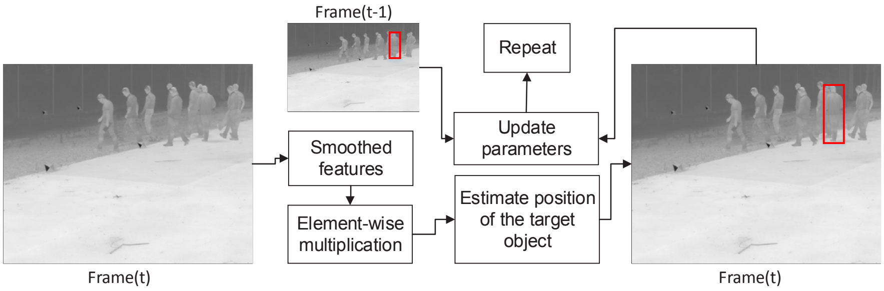

**Fig. 1.** Framework of the proposed method.

### 2.1 Kernelized correlation filter

In the first frame, correlation filter is trained by using image patch which is cropped from a given position of the target object. For detection in next frame, many types of features can be extracted from raw input data. In this paper, we used convolutional features which is generated from shallow layer. Then, a response map can be obtained by using an inverse fast Fourier transform (FFT) after performing element-wise multiplication by using FFT between smoothed features and the correlation filter. Further, location of the target object can be estimated using maximum value of the response map. Mathematically, it can be represented by

$$y = F^{-1}(\hat{x} \odot \hat{h}^{*}), \tag{1}$$

where $\hat{x}$ is the input in the Fourier domain, and $\hat{h}^{*}$ is the correlation filter in the Fourier domain. Symbol $*$ denotes complex conjugate and $\odot$ represents element-wise multiplication.

### 2.2 Convolutional shallow features

Currently, research about CNN has been growth rapidly. The features which generate from CNN have many advantages especially in computer vision applications. This network usually consists of several layers such as convolutional layers, normalization layers, and pooling layers. In this paper, we used pre-trained CNN proposed by [10] and the features to represent the target object is generated from shallow layer.

### 2.3 Update parameters

During tracking, usually the target object change their appearance. In order to handle this problem, the algorithm should update the parameter. Because we used kernelized correlation filter, the correlation filter can be updated by using

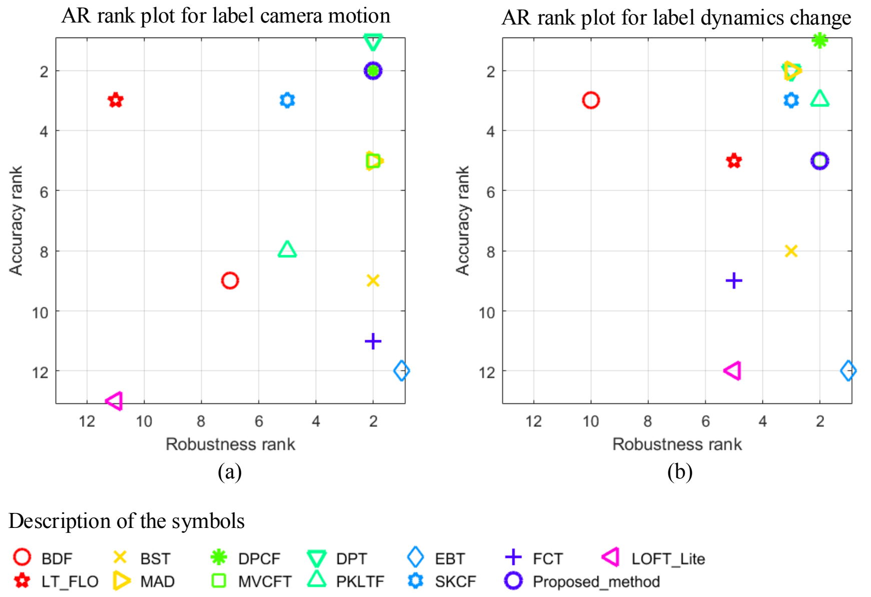

**Fig. 2.** AR rank plots based on experiments in VOTIR-2016 benchmark dataset: (a) for label camera motion, (b) for label dynamics change.

$$\min_{\hat{h}} \sum_{i} \gamma \left|\left|\hat{h} \cdot x_i - y_i\right|\right|_2^2 + \lambda \left|\left|A\right|\right|_2^2 , \tag{2}$$

where $i$ is the number of training image patch, $y$ is the desired output, and A is element-wise multiplication between weight function and correlation filters. Symbols $\gamma$ and $\lambda$ denote weight control and regularization parameter to handle over fitting. Further, parameter $\gamma$ can be updated by using

$$\gamma_{updated} = \frac{\gamma_{previous}}{1 - \alpha}, \tag{3}$$

where $\alpha$ is the learning rate.

## 3 Experimental results

In this section, the experimental results are described. Parameters $\lambda$ and $\alpha$ are equal to 1 and 0.0075, respectively. And to evaluate the proposed method, VOT-TIR2016 benchmark dataset [11] is used. This benchmark dataset consists of 25 videos, where each video has challenging problems such as camera motion,

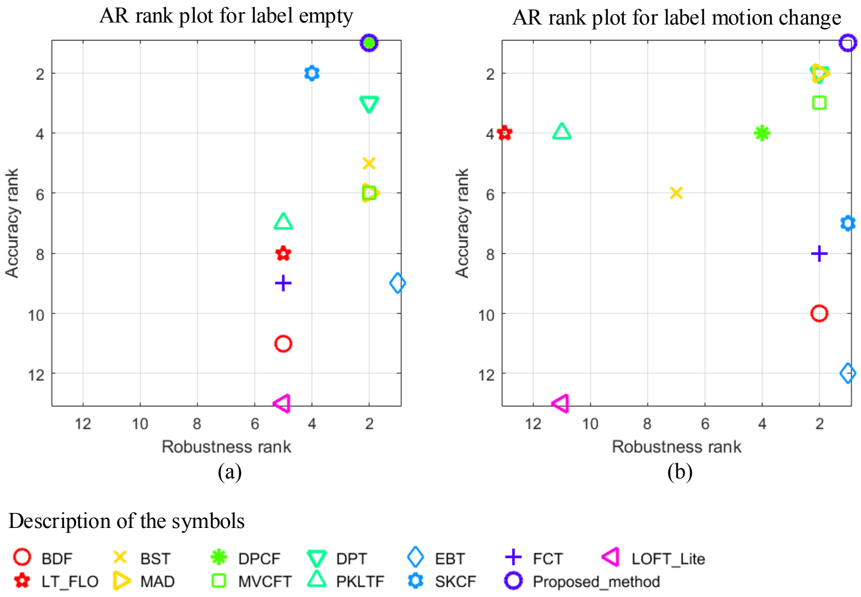

**Fig. 3.** AR rank plots based on experiments in VOTIR-2016 benchmark dataset: (a) for label empty, (b) for label motion change.

dynamics change, motion change, occlusion, and also size change. Further, the experimental results will be served in the accuracy-robustness (AR) rank. The accuracy is measured based on area under curve (AUC), where $AUC = \frac{|B_r \cap B_{gt}|}{|B_r \cup B_{gt}|}$. $B_r$ is the bounding box of the result and $B_{gt}$ is the bounding box of the ground truth. For robustness, it is measured based on the average failure rate over 3 runs. Our proposed method is implemented in MATLAB and runs approximately 0.4 fps on a 3.3 GHz i5-4590 with 4 GB memory.

The AR rank plots are shown in Figure 2, Figure 3, and Figure 4. The proposed method is compared with 12 state-of-the-art tracking algorithm such as BDF [12], BST [11], DPCF [13], DPT [14], EBT [15], FCT [11], LOFT_lite [16], LT_FLO [17], MAD [18], MVCFT [11], PKLTF [19], and SKCF [20]. For the camera motion problem, DPT tracker was ranked first, while the proposed method and DPCF tracker have same rank in the second rank. For the dynamics change problem, DPCF tracker was ranked first in the AR rank plot, while the proposed method was ranked fifth. The proposed method and DPCF tracker were ranked first for label empty problem. Further, the proposed method performs excellent and achieved ranked first for the motion change problem. For the occlusion problem, the proposed method shows more robust than DPT tracker, SKCF tracker,

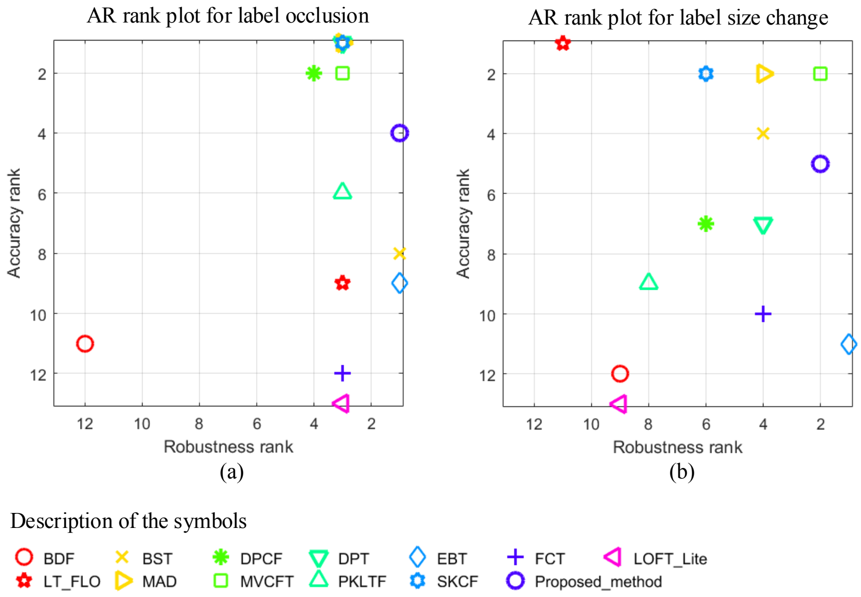

**Fig. 4.** AR rank plots based on experiments in VOTIR-2016 benchmark dataset: (a) for label occlusion, (b) for label size change.

and MVCFT tracker. Furthermore, the proposed method and MVCFT tracker were ranked second for the robustness rank in the size change problem.

## 4 Conclusion

In this paper, convolutional shallow features with kernelized correlation filter for thermal infrared object tracking is proposed. The convolutional features is generated from shallow layer of pre-trained CNN. Then, to estimate the location of the target object, maximum value from response maps of the correlation filter is used. Further, update parameters are used to handle appearance change of the target object during tracking. And based on extensive experimental results by using VOT-TIR2016 benchmark dataset, our proposed method performs favorably against state-of-the-art tracker algorithms.

## Acknowledgment

The authors would like to thank BK21PLUS, Creative Human Resource Development Program for IT Convergence for supporting this work.

## References

1. S. A. Wibowo, H. Lee, E. K. Kim, and S. Kim, "Multi-scale color features based on correlation filter for visual tracking," in *1st International Conference on Signals and Systems*, May 2017, pp. 272–277.
2. S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *Int. J. Comput. Vision*, vol. 56, no. 3, pp. 221–255, Feb. 2004.
3. Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, July 2012.
4. X. Mei, H. Ling, Y. Wu, E. P. Blasch, and L. Bai, "Efficient minimum error bounded particle resampling l1 tracker with occlusion detection," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2661–2675, July 2013.
5. S. A. Wibowo, H. Lee, E. K. Kim, and S. Kim, "Fast generative approach based on sparse representation for visual tracking," in *2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS)*, Aug 2016, pp. 778–783.
6. L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: Complementary learners for real-time tracking," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 1401–1409.
7. S. A. Wibowo, H. Lee, E. K. Kim, and S. Kim, "Visual tracking based on complementary learners with distractor handling," *Mathematical Problems in Engineering*, pp. 1–13, 2017.
8. M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1717–1724.
9. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 580–587.
10. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in *British Machine Vision Conference*, 2014.
11. M. Felsberg, M. Kristan, J. Matas, A. Leonardis, R. Pflugfelder, G. Häger, and et. al., *The Thermal Infrared Visual Object Tracking VOT-TIR2016 Challenge Results*. Cham: Springer International Publishing, 2016, pp. 824–849.
12. M. E. Maresca and A. Petrosino, *Clustering Local Motion Estimates for Robust and Efficient Object Tracking*. Cham: Springer International Publishing, 2015, pp. 244–253.
13. O. Akin, E. Erdem, A. Erdem, and K. Mikolajczyk, "Deformable part-based tracking by coupled global and local correlation filters," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 763–774, 2016.
14. A. Lukezic, L. Cehovin, and M. Kristan, "Deformable parts correlation filters for robust visual tracking," *CoRR*, vol. abs/1605.03720, 2016.
15. G. Zhu, F. Porikli, and H. Li, "Beyond local search: Tracking objects everywhere with instance-specific proposals," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 943–951.
16. R. Pelapur, S. Candemir, F. Bunyak, M. Poostchi, G. Seetharaman, and K. Palaniappan, "Persistent target tracking using likelihood fusion in wide-area and full motion video sequences," in *2012 15th International Conference on Information Fusion*, July 2012, pp. 2420–2427.

17. K. Lebeda, S. Hadfield, J. Matas, and R. Bowden, "Texture-independent long-term tracking using virtual corners," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 359–371, Jan 2016.
18. S. Becker, S. B. Krah, W. Hübner, and M. Arens, "Mad for visual tracker fusion," in *SPIE Security+ Defence.* International Society for Optics and Photonics, 2016, pp. 99 950K–99 950K.
19. A. González, R. Martín-Nieto, J. Bescós, and J. M. Martínez, "Single object long-term tracker for smart control of a ptz camera," in *Proceedings of the International Conference on Distributed Smart Cameras*, ser. ICDSC '14. New York, NY, USA: ACM, 2014, pp. 39:1–39:6.
20. A. S. Montero, J. Lang, and R. Laganire, "Scalable kernel correlation filter with sparse feature integration," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Dec 2015, pp. 587–594.