A performance evaluation method for infrared tracker

Haichao Zheng
Institute of Integrated Electronics and Information
China Academy of Electronics and Information Technology
Beijing, China
zhctbt@163.com

Jie Yang
Schulich School of Engineering
University of Calgary
Calgary, Canada
jie.yang2@ucalgary.ca

Jianjun Chen
Institute of Integrated Electronics and Information
China Academy of Electronics and Information Technology
Beijing, China
chenjianjun512@126.com

***Abstract*—In recent years, significant progress has been made in the evaluation of tracking performance. However, there is still one issue that remains unsolved. All existing works evaluate tracking performance based on empirical comparison using limited video sequences. Performance evaluation based on empirical comparison leads to the following problems: (1) we don't know that how robust a tracker is for all possible scenarios; (2) we will never know what the performance of a tracker is for untested video sequences until we execute it on those sequences. To address these problems, we propose a performance evaluation method for infrared tracker, which can predict the tracking performance of a tracker for untested infrared video sequences. In the proposed method, an image sequence metric is introduced first to quantify tracking difficulty of the infrared video sequence. Afterwards, we establish a tracking dataset including infrared video sequences with gradually increasing tracking difficulties. Then, the tracker need to be evaluated is executed on the tracking dataset. Meanwhile, the tracking performance is quantitatively evaluated by a measure. Thereafter, we identify the relationship between tracking difficulties of infrared video sequences and corresponding quantitative tracking results of that evaluated tracker. Based on that relationship, when tracking difficulty of an untested video sequence is determined, we can predict the tracking performance of that evaluated tracker for that untested infrared video sequence. Experimental results prove that the proposed performance evaluation method can effectively predict the tracking performance of a tracker for untested infrared video sequences.**

Keywords—infrared video, object tracking, tracking evaluation, performance evaluation method

I. INTRODUCTION

Object tracking is one of the most important problems in computer vision with applications ranging from surveillance, human computer interaction, to medical imaging. Although object tracking has been studied for several decades and much progress has been made in recent years, it remains a very challenging problem. Numerous factors affect the performance of a tracker and there exists no single tracking approach that can successfully handle all scenarios. Therefore, it is crucial to evaluate the performance of state-of-the-art trackers and identify their tracking performance under different scenarios.

Numerous efforts have been made to assess the performance of different trackers. Wang et al. [1] compare several trackers using center error and overlap measures. Their research is focused primarily on investigating strengths and weaknesses of a few trackers. Wu et al. [2] propose an approach utilizing a time-reversed Markov chain to evaluate trackers in the absence of annotations. A number of tracking approaches based on sparse representation are compared in [3]. In order to determine the properties and relationships of different performance evaluation measures, L. Čehovin et al. [4] perform an experimental evaluation on a diverse set of 13 well established or recently presented trackers. For a deeper insight into trackers' performance, Nawaz et al. [5] propose a diagnosis of the performance of multi-target trackers, which analyzes key contributory factors (false positives, false negatives, ID changes) that lead to the final performance. Nawaz and Cavallaro [6] present a system for evaluation of video trackers that aims at addressing the real-world conditions. The system can simulate several real-world sources of noisy input, such as initialization noise and image noise. An interesting idea has been suggested by Pang and Ling [7], who aggregate existing experiments, published in various works, in a page-rank fashion to form a less biased ranking of trackers. Smeulders et al. [8] provide an experimental survey of 19 trackers on 315 videos together with an analysis of several performance measures. They search for multiple measures that describe different aspects of tracking performance.

In recent years, several performance evaluation methodologies have been established in order to assess and understand the advancements made by large number of publications. One of the pioneers for building a common ground in tracking performance evaluation is PETS [9], followed-up more recently by the Visual Object Tracking (VOT) challenges [10] [11] [12] and the Object Tracking Benchmarks [13] [14]. VOT challenges aim at comparing short-term single-object visual trackers that do not apply pre-learned models of object appearance. Organizers of VOT challenges provide an evaluation kit and a dataset for automatic evaluation of the trackers. Authors attending VOT challenges are required to integrate their tracker into the evaluation kit, which automatically performed a standardized experiment. The results are analysed by the VOT evaluation methodology. In the most recent VOT2016 [12], results of 70 trackers are presented, with a large number of trackers being published at major computer vision conferences and journals in the recent years. The number of tested trackers makes the VOT 2016 the largest and most challenging benchmark on short-term tracking to date. In [13] and [14], Wu et al. extensively evaluate the performance of 31 trackers on 100 videos with different initialization settings. By analyzing quantitative results, they identify effective approaches for robust visual tracking and provide potential research directions in this field.

All the above-mentioned researches focus on visual tracking, but far fewer researches exist for infrared tracking. The essential difference between infrared tracking and visual tracking is that infrared image can only provide gray

978-1-5386-4276-4/17/$31.00 ©2017 IEEE

information, while color image can provide color information that is usually used as feature for visual tracking. Visual Object Tracking thermal infrared (VOT-TIR2015) challenge [15] is the first infrared (TIR), short-term tracking challenge. Like the VOT challenge, the VOT-TIR challenge considers single-camera, single-target, model-free, causal trackers, applied to short-term tracking. It has been featured as a sub-challenge to VOT2015. VOT-TIR2015 evaluates 24 trackers on the Linkoping TIR dataset. VOT-TIR2016 [16] is the second benchmark on short-term tracking in TIR sequences. Results of 24 trackers are presented. The VOT-TIR2016 challenge is similar to the 2015 challenge, the main difference is the introduction of new, more difficult sequences into the dataset. Furthermore, VOT-TIR2016 evaluation adopted the improvements regarding overlap calculation in VOT2016. In addition, to demonstrate the proposed tracker, Lamberti et al. [17] [18] address performance comparisons of 33 trackers on several public infrared databases and implicitly identify some effective approaches for pedestrian tracking in infrared videos.

Regardless of the fact that numerous efforts have been made to assess the performance of different trackers, there is still one issue that remains unsolved. All the existing works evaluate trackers based on empirical comparison using limited video sequences and demonstrate corresponding tracking success rates or tracking precisions. Performance evaluation based on empirical comparison leads to the following problems: (1) we don't know that how robust the tracker is for all possible scenarios; (2) we will never know what the performance of a tracker is for untested video sequences until we execute it on those sequences. One way to address these problems is to correlate the video sequence characteristics with the tracking performance and measure the video sequence using quantitative metrics, which is usually called "image sequence metric" [19] [20] [21]. It should be noticed that the concept of "image sequence metric" in the field of object tracking is different from that for an ordinary video: they are conceived to describe the factors interfering with the performance of trackers [20] [21]. Thereafter, when characteristics of an untested video sequence are determined by image sequence metric, the tracking performance can be predicted correspondingly. Inspired by the above ideas, we propose a performance evaluation method for infrared tracker, which can predict the tracking performance of a tracker for untested infrared video sequences.

The goal of this study is not to evaluate many trackers and compare their tracking abilities in different aspects, though the proposed performance evaluation method can offer a quantitative comparison. This work mainly focuses on the performance evaluation for infrared trackers for a single target.

II. The Proposed Performance Evaluation Method

The block diagram of the proposed performance evaluation method is illustrated in Fig. 1.

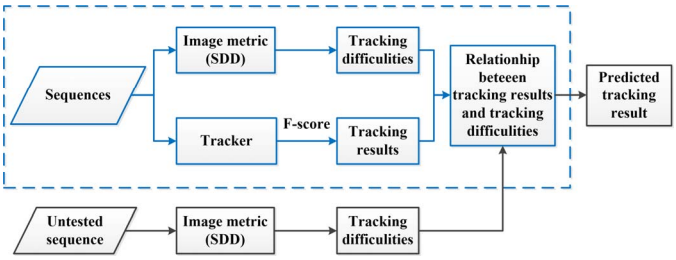

Fig. 1 The block diagram of the proposed performance evaluation method.

We first introduce the sequence difficulty degree (SDD) metric [20] to quantify tracking difficulty of the infrared video sequence. Research on quantifying tracking difficulty of the infrared video sequence is still in the initial stage. To our best knowledge, there are only two infrared sequence metrics that have been suggested in previous works: inter-frame change degree (IFCD) metric [21] and SDD metric. The SDD metric is developed from IFCD metric and has been proved that its performance is far more superior to that of IFCD metric. Thus, we choose the SDD metric to quantify tracking difficulty of the infrared video sequence in this study. Thereafter, we establish a tracking dataset, which includes infrared video sequences with gradually increasing tracking difficulties (increasing SDD values). Afterwards, the evaluated tracker is executed on each infrared video sequence of the tracking dataset. Tracking performance is quantitatively evaluated by the widely used measure: F-score [22] [23]. Then, we identify the relationship between the F-scores of the evaluated tracker and corresponding tracking difficulties of infrared video sequences. With the help of that relationship, when the specific SDD value of an untested video sequence is determined, we can predict the F-score of that evaluated tracker for that untested infrared video sequence.

III. Sequence Difficulty Degree (SDD)

The SDD metric is utilized to quantify tracking difficulty of the infrared video sequence. Due to the space limitations of this paper, we only briefly review SDD for self-completeness here. The interested reader can find more details in work [21].

The work [21] summarizes the factors affecting infrared tracking into five categories: intra-frame occlusion, intra-frame confusion, inter-frame target texture variation, inter-frame target size variation and inter-frame target location variation. Five corresponding image metrics are proposed by quantitatively describing the above five factors. These five metrics are intra-frame degree of occlusion (IFDO), intra-frame degree of confusion (IFDC), inter-frame variation degree of target texture (IFVDTT), inter-frame variation degree of target size (IFVDTS) and inter-frame variation degree of target location (IFVDTL). Infrared tracking is affected by all these five factors mentioned above at the same time. It is meaningless to describe the relationship between any individual metric and the tracking result. Thus, the SDD metric is constructed by combining these five metrics as an integrated indicator to intuitively represent the sequence-level tracking difficulty of the infrared video sequence as follows:

$$SDD = \sqrt{IFDO^2 + IFDC^2 + IFVDTT^2 + ITVDTS^2 + IFVDTL^2}. \quad (1)$$

The larger SDD value means the higher tracking difficulty of the given infrared video sequence.

IV. Experiments and Analysis

To prove the effectiveness of the proposed performance evaluation method, we conduct experiments in this section. We will show how to utilize the proposed performance evaluation method to provide quantitative comparison for different trackers and predict the tracking performance of a tracker for untested infrared video sequences. To that end, specific experimental conditions are given below.

A. Tracking Dataset

Our tracking dataset includes 69 real infrared video sequences. Part of these infrared video sequences is collected from the public datasets Army Missile Command (AMCOM) Dataset [24] [25] and the OTCBVS Dataset 05: Terravic Motion IR Database [17] [18]; the rest infrared video sequences are collected by ourselves. These infrared video sequences include both the long-wave infrared and medium-wave infrared video sequences. The infrared camera is moving to shoot thirty-nine video sequences, while it is fixed to shoot the rest of these video sequences. The lengths of these infrared video sequences vary from 39 frames to 150 frames. For these infrared video sequences, resolutions range from 128×128 to 571×356 pixels. All the sequences in the tracking dataset are from outdoor environments and were recorded in different weather conditions. The following gallery in Fig. 2 gives an overview of our tracking dataset (several snapshots from the video sequences in our tracking dataset).

Fig. 2 Several snapshots from our tracking dataset

This work focuses on the performance evaluation for infrared trackers for single targets. There is no case of multiple targets in our experiments. When there are multiple objects in an infrared video sequence, we only choose an object of interest as the target. The targets appear in the tracking dataset including pedestrian, truck, car, aircraft and speedboat. The target presents throughout the video sequence. We utilize SDD to quantify the tracking difficulty of each infrared video sequence. For the 69 infrared video sequences, their SDD values range from 0.2875 to 0.8522. Computing SDD metrics in practice will require the ground truth of the target. We create the ground truth of the target by manually selecting the smallest rectangle that contains the target. There are no rotating bounding boxes, all of them are axis-aligned.

B. Evaluated Trackers

The goal of this study is not to evaluate many state-of-the-art trackers and compare their tracking abilities in different aspects. Instead, we mainly focus on offering a method to predict the tracking performance of a tracker for untested infrared video sequences. We select three state-of-the-art online trackers that can be adapted for infrared tracking, appearing in TPAMI and CVPR in the recent six years. They are fast compressive tracking (FCT) [26] [27], distribution fields for tracking (DF) [28] and least soft-thresold squares tracking (LSST) [29] [30]. These three trackers are utilized to conduct tracking experiments on our tracking dataset.

Due to the space limitations of this paper, we only briefly introduce these three trackers here. The interested reader can find more details in works [26] [27] [28] [29] [30]. FCT is a simple yet effective and efficient tracker with an appearance model based on features extracted from a multiscale image feature space with data-independent basis. The tracking task is formulated as a binary classification via a naive Bayes classifier with online update in the compressed domain. DF builds an image descriptor using distribution fields (DFs), a representation that allows smoothing the objective function without destroying information about pixel values. DFs provide a convenient way to aggregate the observations of the object through time and maintain an updated model. LSST is a generative tracking method based on a robust linear regression algorithm. Based on maximum joint likelihood of parameters, a least soft-thresold squares (LSS) distance is derived to measure the difference between an observation sample and the dictionary. Compared with the distance derived from ordinary least squares methods, the LSS metric is more effective in dealing with outliers. We use the default parameters as the authors have given in their works for all of our experiments.

C. Tracking Performance Measure

Tracking performance is quantitatively measured by the widely used measure: F-score [22] [23]. F-score is given by

$$\text{F-Score} = \frac{2\times TP}{2\times TP \ + \ FP \ + \ FN}. \tag{2}$$

where TP is true positive; FP is false positive; FN is false negative. As shown in Fig. 3, all TP, FN and FP values are calculated based on pixel count. In Fig. 3, the bounding box drawn in solid lines represents the ground truth and the one drawn in dashed lines is the bounding box found by the tracker.

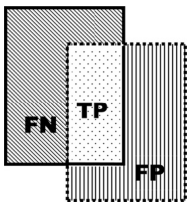

Fig. 3 Calculation of TP, FP and FN.

The value range of F-score is [0, 1]. The smaller the F-score value is, the worse the tracking performance is. To intuitively measure the performance of a tracker in an infrared video sequence, the average F-score is used in this study, which is the mean of the sum of each frame's F-score. If the obtained average F-score is smaller than 0.2, we believe that tracker fails to track the target and set the average F-score to 0.

D. Experimental Results

Under above experimental conditions, we get the experimental results of FCT, DF and LSST. The 69 infrared video sequences in the tracking dataset are divided into two parts. We first use 66 infrared video sequences to conduct experiments. Relationship plots between SDD values and F-score values of these three trackers are shown in Fig. 4, Fig. 5 and Fig. 6, respectively. Then, we show how to utilize the proposed performance evaluation method to provide quantitative comparison for different trackers. Thereafter, we show how to utilize the proposed performance evaluation method to predict the tracking performance of a tracker for untested infrared video sequences. Afterwards, we use the left 3 untested infrared video sequences in the tracking dataset to verify the predicted tracking results. To verify the tracking results of a wide range of sequences, these three sequences are selected with small, middle and high SDD value, respectively.

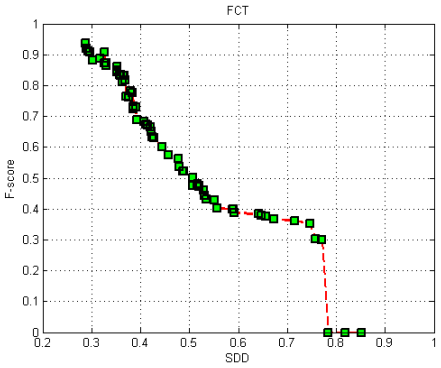

Fig. 4 Relationships between SDD values and F-score values of FCT.

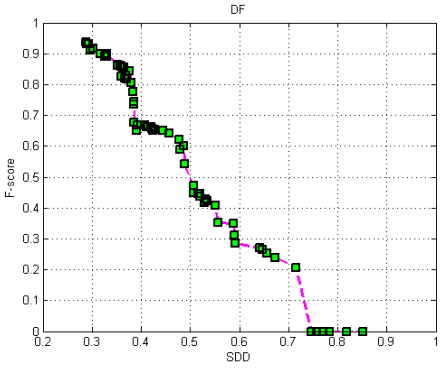

Fig. 5 Relationships between SDD values and F-score values of DF.

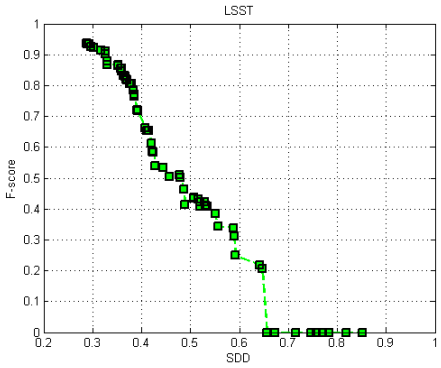

Fig. 6 Relationships between SDD values and F-score values of LSST.

In Fig. 4, Fig. 5 and Fig. 6, the horizontal axis represents SDD values of infrared video sequences, while the vertical axis represents the F-scores of a specific tracker. Thus, each point on the curve is the tracking performance of a tracker corresponding to an infrared video sequence with specific tracking difficulty.

Take Fig. 4 as an example at first to show how these three relationship figures in Fig. 4, Fig. 5 and Fig. 6 are obtained. For all sixty-six tested infrared video sequences, we calculate the SDD value of every sequence. Then, we utilize FCT to execute tracking on every sequence and then calculate its corresponding F-score value. Thus, specific SDD value corresponds to a specific F-score value for each sequence. We use the horizontal axis and the vertical axis to represent the SDD and F-score, respectively. Thus, we have sixty-six scatter points. We arrange these sixty-six scatter points in ascending order of SDD values. Then, we get the relationship between SDD and F-score of FCT as shown in Fig. 4. Similarly, Fig. 5 and Fig. 6 show the relationship between SDD and F-Score of DF and LSST, respectively.

E. Quantitative Comparison for Different Trackers

Before we show how to provide quantitative comparison for different trackers, we need to give one definition.

Definition 1 SDD failure threshold: when a tracker is executed on many infrared video sequences with gradually increasing SDD values, the first sequence in which the tracker loses the target is denoted as s_f . The specific SDD value of s_f is the SDD failure threshold corresponding to the tracker.

Quantitative comparison results of FCT, DF and LSST are given in Table I. For the sixty-six infrared video sequences, their SDD values range from 0.2875 to 0.8522. As shown in Table I, FCT only fails to track the target for 3 sequences; while the failed sequence number of DF and LSST is 6 and 9, respectively. Meanwhile, the average F-score of FCT is 0.6148, which is higher than that of DF (0.5941) and LSST (0.5648). Consequently, these evaluation results show that FCT overall performs better than DF and LSST. Although LSST successfully tracks the target on most of these sequences, its average F-score is only 0.5648. It reflects that it is difficult for LSST to accurately track the target. However, as shown in

Table I, the maximum F-score of LSST can reach 0.9559. Whereas, the maximum F-score of FCT and DF is only 0.9365 and 0.9372, respectively. These evaluation results indicate that LSST can achieve more accurate tracking than FCT and DF on the sequence with a low tracking difficulty.

TABLE I. Quantitative Comparison Results of FCT, DF and LSST

Tracker	Failed sequence number	SDD failure threshold	F-score value	
			Average	Maximum
FCT	3	0.7837	0.6148	0.9365
DF	6	0.7459	0.5941	0.9372
LSST	9	0.6563	0.5648	0.9559

With the help of relationship plots between SDD values and F-score values, we can obtain some other useful information. As mentioned above, the higher SDD value indicates the higher tracking difficulty of the given infrared video sequence. Meanwhile, the smaller F-score value indicates the worse the tracking performance. As shown in Fig. 4, Fig. 5 and Fig. 6, these three relationships show that the higher the SDD value is, the worse the tracking performance is. A tracker will lose the target when the SDD value is high enough, but that specific SDD value may vary for different trackers. For example, FCT loses the target when the SDD value is higher than or equal to 0.7837; whereas the SDD value is only 0.6563 when LSST begins to lose the target. Thus, the SDD failure threshold of FCT, DF and LSST is 0.7837, 0.7459 and 0.6563, respectively. FCT has the highest SDD failure threshold. It indicates that FCT has a more widespread applicability and can adapt to the infrared video sequence with high tracking difficulty. In contrast, LSST has the lowest SDD failure threshold. Thus, LSST has the most narrow tracking applicability and cannot deal with those infrared video sequences with high tracking difficulty.

F. Predicting Tracker's Tracking Performance

Before we show how to predict the tracking performance of a tracker for untested infrared video sequences, we need to give another definition.

Definition 2 Minimum SDD range: Denote the SDD value of an untested infrared video sequence S_x as SDD_x. Arrange SDD_x with the SDD values of the tested sequences together in ascending order. The SDD value on the left side of SDD_x is denoted as SDD_l, while the SDD value on the right side of SDD_x is denoted as SDD_r. Then the minimum SDD range of S_x is $[SDD_l, SDD_r]$.

Generally speaking, evaluation results of image metric should be associated with tracking performance. As pointed out by the work [20], a monotonously decreasing relationship can be expected between SDD values and the F-score values. It means that the tracking performance gradually deteriorates with increasing tracking difficulty. Since we have obtained the relationship plots between SDD values and F-score values of FCT, DF and LSST. We can utilize these relationship plots to predict the tracking performance of each tracker for untested infrared video sequences. We used the left three infrared video sequences (S1, S2 and S3) in our tracking dataset to verify the predicted tracking results. To verify the tracking results of a wide range of sequences, these three predicted sequences are selected with small, middle and high SDD value (0.3552, 0.6941 and 0.8260), respectively. The experimental results of these three predicted sequences are given in Table II.

TABLE II. Details of Three Predicted Sequences and Corresponding Experimental Results

Sequence	SDD value	Minimum SDD range	Predicted F-score range			Actual F-score value		
			FCT	DF	LSST	FCT	DF	LSST
S1	0.3552	[0.3529, 0.3574]	[0.8361, 0.8429]	[0.8599, 0.8620]	[0.8564, 0.8629]	0.8393	0.8600	0.8609
S2	0.6941	[0.6725, 0.7145]	[0.3623, 0.3668]	[0.2070, 0.2396]	[0, 0]	0.3666	0.2087	0
S3	0.8260	[0.8191 0.8522]	[0, 0]	[0, 0]	[0, 0]	0	0	0

Let us take FCT for an example to predict its tracking performance for untested sequences S1, S2 and S3. According to the relationship plot between SDD values and F-score values of FCT in Fig. 4, we first find the minimum SDD range containing the SDD value of sequence S1 (0.3552). That minimum SDD range is [0.3529, 0.3574]. The SDD value 0.3529 and 0.3574 correspond to two specific sequences in the relationship plot. It indicates tracking difficulty of sequence S1 should be between that of these two sequences. Correspondingly, the tracking performance of S1 should be between that of these two sequences. According to the relationship plot in Fig. 4 we can find the F-score value corresponds to 0.3529 and 0.3574 is 0.8429 and 0.8361, respectively. Consequently, we can predict that the F-score of FCT for sequence S1 should range from 0.8361 to 0.8429. Similarly, minimum SDD range containing the SDD value of S2 (0.6941) can be determined as [0.6725, 0.7145]. The corresponding F-score value range is [0.3623, 0.3668]. Thus, we can predict that the F-score of FCT for sequence S2 should range from 0.3623 to 0.3668. When it comes to the sequence S3, we can notice that the SDD value of S3 (0.8260) exceeds the SDD failure threshold of FCT (0.7837). Consequently, we can predict that FCT will definitely lose the target when it is executed to track the target on S3 and the corresponding F-score value should be 0.

To prove the effectiveness of the proposed performance evaluation method, we execute FCT, DF and LSST on sequence S1, S2 and S3, respectively. Then, we obtain the actual F-score values of these three sequences as shown in Table II. Let us first verify the predicted tracking results of FCT. From Table II, we can see that the actual F-score value of FCT for S1 is 0.8393. It meets with the predicted F-score range [0.8361, 0.8429]. Whereas the actual F-score value of FCT for S2 is 0.3666, which also meets with the predicted F-score range [0.3623, 0.3668]. When it comes to the sequence S3, FCT fails to track the target and the actual F-score value of FCT is 0. The actual F-score value meets with the predicted F-score value. Similarly, we can utilize the relationship plot in

Fig. 5 and Fig. 6 to predict the tracking performance of DF and LSST. As shown in Table II, for sequence S1, the predicted F-score ranges of DF and LSST are [0.8599, 0.8620] and [0.8564, 0.8629], respectively. Whereas the predicted F-score ranges of DF and LSST for sequence S2 are [0.2070, 0.2396] and [0, 0], respectively. The predicted results indicate that DF can successfully track the target on sequence S2; whereas LSST will fail to track the target. When it comes to the sequence S3, the predicted F-scores of DF and LSST both are 0. It indicates that both DF and LSST will lose the target on sequence S3. As shown in Table II, the actual F-score values of DF and LSST for S1 are 0.8600 and 0.8609, which meet with their corresponding predicted F-score range, respectively. For sequence S2, the predicted F-score value of DF (0.2087) meets with its predicted F-score range; whereas LSST fails to track the target as predicted. When it comes to the sequence S3, both DF and LSST fail to track the target as predicted.

V. CONCLUSION

In this study, we propose a performance evaluation method for infrared tracker, which can predict the tracking performance of a tracker for untested infrared video sequences. The SDD metric is introduced first to quantify the tracking difficulty of the infrared video sequence. Afterwards, we establish a tracking dataset including 69 infrared video sequences with gradually increasing tracking difficulties. Then, three trackers are executed on the tracking dataset. Meanwhile, the tracking performance is quantitatively evaluated by F-score. Thereafter, we identify the relationship between SDD values and F-score values of a tracker. Based on that relationship, we can predict the tracking performance of that tracker for untested infrared video sequences. Experimental results prove that the proposed method can effectively predict the tracking performance of a tracker for untested infrared video sequences.

ACKNOWLEDGMENT

This work is funded by China Postdoctoral Science Foundation and the Natural Science Foundation of China (No.61601422).

REFERENCES

[1] Q. Wang, F. Chen, W. Xu, et al., "An experimental comparison of online object tracking algorithms," SPIE Conference on Image and Signal Processing, pp. 81381A-81381A (2011).

[2] H. Wu, A. C. Sankaranarayanan, and R. Chellappa, "Online empirical evaluation of tracking algorithms," IEEE Trans. Pattern Anal. Mach. Intell. 32(8), pp.1443–1458 (2010).

[3] S. Zhang, H. Yao, X. Sun, et al., "Sparse coding based visual tracking: Review and experimental comparison," Pattern Recognit. 46(7), pp.1772–1788 (2013).

[4] L. Čehovin, M. Kristan, and A. Leonardis, "Is my new tracker really better than yours?" IEEE Winter Conference on Applications of Computer Vision, pp. 540-547 (2014).

[5] T. Nawaz, and A. Ellis, "A method for performance diagnosis and evaluation of video trackers," SIViP. 11(7), pp. 1287-1295 (2017).

[6] T. Nawaz, and A. Cavallaro, "A protocol for evaluating video trackers under real-world conditions," IEEE Trans. Image Process. 22(4), pp. 1354-1361 (2012).

[7] Y. Pang and H. Ling, "Finding the best from the second bests-inhibiting subjective bias in evaluation of visual tracking algorithms," IEEE Conference on Computer Vision, pp. 2784-2791 (2013).

[8] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, et al., "Visual tracking: An experimental survey," IEEE Trans. Pattern Anal. Mach. Intell. 36(7), pp. 1442-1468 (2014).

[9] D. P. Young, and J. M. Ferryman, "Pets metrics: Online performance evaluation service," IEEE 14th International Conference on Computer Communications and Networks, pp. 317-324 (2005).

[10] L. Agapito, M. Bronstein, C. Rother, et al., "The visual object tracking VOT2014 challenge results," European Conference on Computer Vision Workshops, pp. 191-217, 2014.

[11] M. Kristan, J. Matas, A. Leonardis, et al., "The visual object tracking VOT2015 challenge results," IEEE International Conference on Computer Vision Workshops, pp. 1-23 (2015).

[12] M. Kristan, A. Leonardis, J. Matas, et al., "The visual object tracking VOT2016 challenge results," European Conference on Computer Vision Workshops, pp. 197-216 (2016).

[13] Y. Wu, J. Lim, and M.H. Yang, "Online object tracking: A benchmark," IEEE Conference on Computer Vision, pp. 2411-2418 (2013).

[14] Y. Wu, J. Lim, and M.H. Yang, "Object tracking benchmark," IEEE Trans. Pattern Anal. Mach. Intell. PP Issue: 99, pp. 1-14 (2015).

[15] M. Felsberg, A. Berg, G. Hager et al., "The thermal infrared visual object tracking VOT-TIR2015 challenge results," IEEE International Conference on Computer Vision Workshops, pp. 76-88 (2015).

[16] M. Felsberg, M. Kristan, J. Matas, et al., "The thermal infrared visual object tracking VOT-TIR2016 challenge results," European Conference on Computer Vision Workshops, pp. 639-651 (2016).

[17] F. Lamberti, R. Santomo, A. Sanna, "Intensity variation function and template matching-based pedestrian tracking in infrared imagery with occlusion detection and recovery," Opt. Eng. 54(3), pp. 033106 (2015).

[18] F. Lamberti, A. Sanna, G. Paravati, et al., "IVF3: exploiting intensity variation function for high-performance pedestrian tracking in forward-looking infrared imagery," Opt. Eng. 53(2), pp. 023105 (2014).

[19] Y. Chen, G. Chen, R.S. Blum, et al., "Image quality measures for predicting automatic target recognition performance," IEEE Conference on Aerospace, pp. 1-9 (2008).

[20] H. Zheng, X. Mao, L. Chen, "A novel method for quantifying target tracking difficulty of the infrared image sequence," Infrared Phys. Technol. 72, pp. 8-18 (2015).

[21] W. Diao, X. Mao, H. Zheng, "Image sequence measures for automatic target tracking," Prog. Electromagn. RES. 130, pp. 447-472 (2012).

[22] H. Zheng, X. Mao, L. Chen, "Adaptive edge-based meanshift for drastic change gray target tracking," Optik. 126(23), pp. 3859-3867 (2015).

[23] L. Wang, H. Yan, K. Lv, et al., "Visual tracking via kernel sparse representation with multi-kernel fusion," IEEE Trans. Circuits Syst. Video Technol. 24, pp. 1132-1141 (2014).

[24] A. Bal, and M. S. Alam, "Automatic target tracking in FLIR image sequences using intensity variation function and template modeling," IEEE Trans. Instrum. Meas. 54(5), pp. 1846–1852 (2005)

[25] G. Paravati, and S. Esposito, "Relevance-based template matching for tracking targets in FLIR imagery," Sensors. 14(8), pp. 14106-14130 (2014).

[26] K. Zhang, L. Zhang, and M.H. Yang, "Fast compressive tracking," IEEE Trans. Pattern Anal. Mach. Intell. 36, pp. 2002-2015 (2014).

[27] K. Zhang, L. Zhang, and M.H. Yang, "Real-time compressive tracking," European Conference on Computer Vision, pp. 864-877 (2012).

[28] L.S. Lara, and E.L. Miller, "Distribution fields for tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1910-1917 (2012).

[29] D. Wang, H. Lu, and M.H. Yang, "Least soft-threshold squares tracking," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2371-2378 (2013).

[30] D. Wang, H. Lu, and M.H. Yang, "Robust visual tracking via least soft-threshold squares," IEEE Trans. Circuits Syst. Video Technol. 99, pp. 1-12 (2015).