Object Tracking with a Novel Visual-Thermal Sensor Fusion Method in Template Matching

Satbir Singh
Dr.B R Ambedkar National Institute of Technology, Jalandhar, India
Email: satbir2008@ymail.com

Arun Khosla
Dr.B R Ambedkar National Institute of Technology, Jalandhar, India
Email: khoslaak@nitj.ac.in

Rajiv Kapoor
Delhi Technological University, Delhi, India
Email: rajivkapoor.dtu@gmail.com

Abstract—Recently there has been an increase in the use of thermal-visible conjunction technique in the field of surveillance applications due to complementary advantages of both. An amalgamation of these for tracking requires a reasonable scientific procedure that can efficiently make decisions with sound accuracy and excellent precision. The proposed research presents a unique idea for obtaining a robust track estimate with the thermo-visual fusion in the context of fundamental template matching. This method firstly introduces a haphazard transporting control mechanism for individual modality tracking that avoids unexpected estimates. Then it brings together an efficient computation procedure for providing the weighted output using minimal information from the individual trackers. Experiments performed on publically available datasets mark the usefulness of the proposed idea in the context of accuracy, precision and process time in comparison with the state of art methods.

Index Terms—Sensor Fusion, Object Tracking, Template matching, Thermal Imaging

I. INTRODUCTION

The combined use of thermal and visible sensors in the tracking problem is motivated by the fact that each imagery has its limitations and advantages. The visible imagery though provides a foremost hint of colour of the object, but thermal imagery has got its advantages in case of night or improper vision conditions, illumination variations like the shadow, etc. and in case of the object being camouflaged.

Earlier, many efforts have been formulated to utilize the advantages of these imaging modalities in different surveillance applications. These techniques are discussed in detail in the section 2. The proposed method advances the field by working for the betterment of tracking related issues like time constraints, accuracy, and reliability.

Our method carves out a statistical correlation value based template match framework into an effective fusion approach. Initially, individual imagery track is obtained by template matching, and it is modified with the track object's dimension based haphazard transporting controlling (HTC) process in each frame. Then, the outputs retrieved from individual imagery tracker are fed into the proposed merger algorithm. This algorithm firstly makes adaption in the weight factors to be provided to each imagery and later adjusts the positional coordinates of the final track estimate to remove the offset caused by least weighing modality. The rest organization of the paper is detailed in the following paragraph.

The earlier work done is presented in Section 2. Section 3 describes the details of the work methodology adopted for the proposed method. In this section, first tracking using improved template matching for each sensing domain is presented. It follows a detail of the fusion steps used in this template matching practice. The results obtained after performing experiments are presented in Section 4. The discussion on the results is also provided in this section. The last section consists of conclusion remarks & the future scope of the paper.

II. RELATED WORK

Due to exciting advantages of thermal-visual fusion, various uses of visual-thermal fusion are found in the literature.

Face recognition inhibiting thermal and visible sensor fusion using background subtraction had been proposed earlier in the literature [1, 2, 3, and 4]. It has also been actively utilized in robotics-based vision applications [5, 6, and 7]. Besides this, the use of visible and thermal

imagery in fusion has also been hosted in spacecraft proximity operations [8] to overcome the effect of different lighting conditions in the orbital environment. [9, 10] have advanced the field of image fusion of multisensory data for the improved vision applications.

Notably, the use of said bi-modal information has been increased for object tracking under challenging conditions of vision in recent times. The significance of the field has even made researchers incorporate new intelligent techniques like deep learning and convolutional neural networks [11] in thermo-visual tracking. A two-stream convolution net was formed independently for each domain, and the fusion was achieved by forming a fusion net. Despite a new approach, it is to be noted that it is dependent on a rigorous training process and requires complex hardware processing systems.

It is interesting to note that several methods like [15, 16, and 17] make use of the particle filter [18] based approach to aggregate the information from each sensing domain and formulate a sound decision of track. However, in general, the particle filter based approaches have demerit of particle degeneration and an associated unpredictability in its output estimate, and it generally fails to track in severe visual conditions.

To address this issue, [15] proposed a basic fusion that involved multiplication of the individual specialists that used cues from thermal as well as visible imagery to form their opinions. However, simple multiplication only serves the purpose of video sequences with no veiling conditions. In case of false tracking in one mode of imagery, it may lead to overall false result due to the absence of a procedure to verify this.

Further, the techniques employing fusion to overcome demerits of particle filter [16, 17] make use of fusion step per single frame of the track as many times as the no. of particle chosen. This mechanism takes many computational efforts since most of times background calculations are involved in performing a single step fusion. In addition, the computation cost rises as if no. of the particle is increased to increase the accuracy.

More insight into the advances in this field can be reviewed in [12, 13, and 14]. Table 1 lists the description of some methods that incorporated the visible sensor and thermal sensor conjunctively. The fusion strategy adopted, along with the applicative aim of these, are covered under this table.

III. Proposed Work Methodology

The algorithm is formulated into a three-step approach. First is the original track estimate using correlation-based template matching, second the proposed HTC procedure and third is the said thermo-visual fusion process overview. Fig. 1 presents the overall illustration of the workflow of the complete process. The complete section is divided into the following subsections:

A. Template matching based Tracking

For tracking in a particular domain, a fast template matching process was incorporated. The template matching process used the value of the correlation coefficient between the template region and current frame image regions. The formulae for this is given by:

$$c(x,y) = \frac{\sum_i \sum_j (I(x+i,y+j)-\bar{I})(T(i,j)-\bar{T})}{\sqrt{\{\sum_i \sum_j (I(x+i,y+j)-\bar{I})^2\}\{\sum_i \sum_j (T(i,j)-\bar{T})^2\}}} \quad (1)$$

In (1), I stands for the image and T for the template and resembles the pixel position in the template region of the object and the image frame, respectively. Values, and represents the mean values of pixels in the template region and the corresponding image region with a size equivalent to the template size. The output c(x, y) represents the template matching result of a rectangular region in the image with (x, y) as the starting pixel location.

Table 1. Type Sizes for Camera-Ready Papers

Reference	Fusion Strategy	Application
C.LI [11]	Fusion net	Tracking in public RGB-Thermal datasets
M. Talha et al. [16]	Adaptive Foreground-Background relation	Tracking in camouflaging conditions
O'Conaire et al. [15]	Product of Expert Beliefs	Object tracking for publically available datasets
G. Xiao et al. [19]	Weighted Fusion Rules	Object tracking
Stolkin et al. [20]	Bayesian Fusion	Human Detection and Tracking
E. Fendri [21]	Logical AND between the binary mask regions in separate imageries	Moving object detection
S.R. Schnelle [22]	Different fusion methods based on spatial and pyramidal approaches	Object tracking under a surveillance system
Y. Niu [23]	Fusion based on target region segmentation and discrete wavelet transforms	Target perception for Unmanned Aerial Vehicles (UAV)
Y. Wu [24]	Fusion via sparse representation	2D target object tracking
C. Li [25]	Collaborative Sparse representation	Tracking in Grayscale-Thermal framework
C. Li [26]	Laplacian sparse representation	Tracking in Grayscale-Thermal framework
H. P. Liu [27]	Using Min operation on Sparse representation	OCTVBS dataset
Palmerini et al. [8]	Extended Kalman Filter	Spacecraft Proximity Tracking under light varying conditions

For visible domain tracking, separate measures of R, G, and B were calculated, and the combined correlation measure was calculated as the mean of these.

$$c^V(x,y) = (c_r^V(x,y) + c_g^V(x,y) + c_b^V(x,y))/3 \quad (2)$$

In the above equation, $c_r^V(x, y)$, $c_g^V(x, y)$, $c_b^V(x, y)$ represent the individual color channel based correlation similarity score, and $c^V(x, y)$ denotes the combined value for visible imagery based output.

Since the information from the thermal sensor contains the temperature gradient information only, it was converted to grayscale, and the corresponding correlation coefficients are stored as $c^T(x, y)$.

B. Controlling Haphazard Transport of Track Estimate

Moreover, an additional check treatment was introduced to avoid superfluous hoping of state transition. This step helps in sustaining the constancy of track estimation. This factor used for this was named as the haphazard transporting factor (HTF). It was found as:

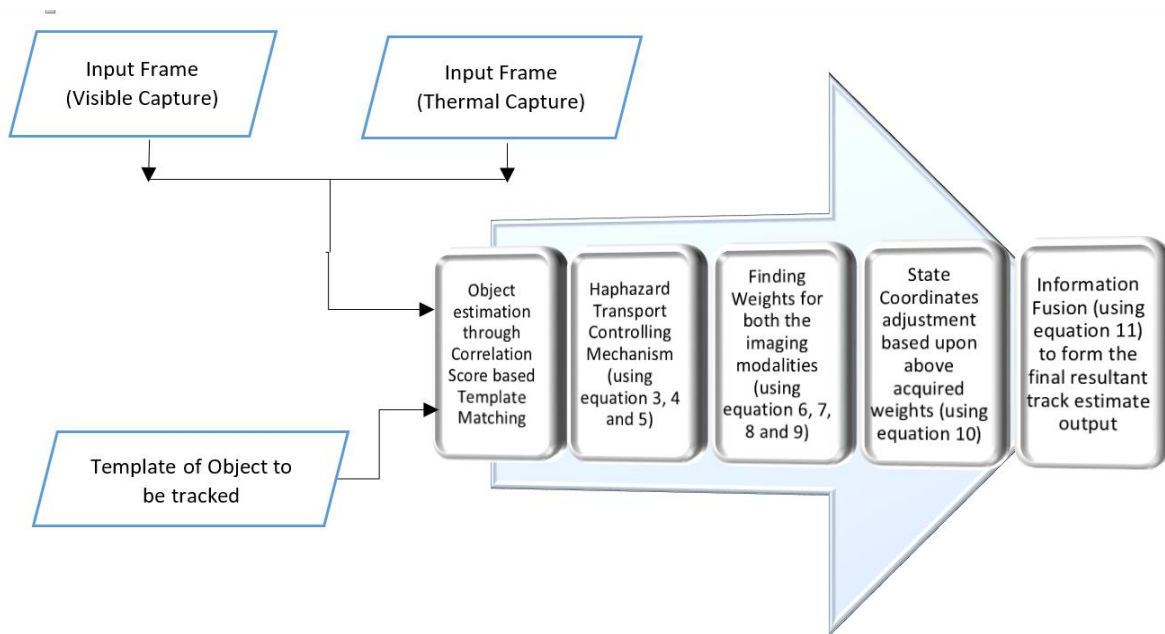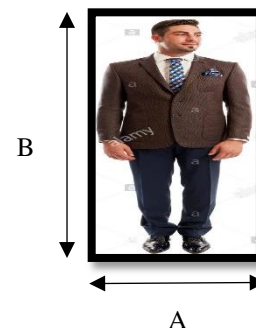$$HTF_t = \sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2} \qquad (3)$$

Fig.1. Block Diagram to illustrate the workflow of the proposed syste

Here, (X_t, Y_t) stands for initial coordinates of rectangular track estimate from the input tracker at image frame at a time interval t. Besides, the corresponding control action was taken was:

$$(X_t, Y_t) = \begin{cases} (X_{t-1}, Y_{t-1}), & if \ HTF_t \ > d \\ (X_t, Y_t), & otherwise \end{cases} \qquad (4)$$

The value of d was strictly dependent directly on the dimensions of the object being tracked. Its value was estimated using the following relation:

$$d = \sqrt[2]{(A^2 + B^2)} \qquad (5)$$

Where A and B represent the width and height of the target selected initially. Fig. 2 shows an example of this.

The maximum value of the correlation coefficient from the above process and the region corresponding to this was chosen as the track estimate from the particular modality. The maximum value of correlation score for individual imaging was found and saved as C_t^V, C_t^T and the region corresponding to this was saved as state estimates S_t^V, S_t^T respectively.

Fig.2. Illustration for the object dimensions

C. Fusion Scheme

Now to decide the individual imagery inputs, a weighting technique is adopted precisely depending upon the fact that the modality tracking should have a less overall average correlation with the background whereas the image showing more considerable correlation measure with the background can be treated as less competent. For this, the average correlation of the background leaving the region of tack estimate of a tracker is found, and the weight adjustment is made accordingly.

$$\alpha_t = B_t^T/(B_t^V + B_t^T) \qquad (6)$$

$$\beta_t = 1 - \alpha_t \qquad (7)$$

$$W_t^V = \frac{\alpha_t \ c_t^V}{(\alpha_t \ c_t^V + \beta_t \ c_t^T)} \qquad (8)$$

$$W_t^T = 1 - W_t^V \qquad (9)$$

In the context of fusion, the adaptive weight factor for visible imagery is represented by α_t and the corresponding adjusted weight by W_t^V. The thermal counterpart parameters are represented by β_t and W_t^T respectively. Here B_t^V and B_t^T represent the average correlation values of the background regions in visible imaging and thermal imaging, respectively.

Fig. 3 helps to illustrate the idea of the fusion process adopted. As seen that in this particular scene, the region of the object tracked by individual imagery is labelled as region R in both visible (Fig. 3(a)) and thermal (Fig. 3(b)). Also, the rest region except R is selected as the region for determining the average correlation for the background in the current image frame. The final correlation weight of each imaging domain is adjusted depending upon the average correlation of the template with the average correlation of the region formed by leaving out R from the whole image.

Fig.3. Information from the section of the image except for region R in both imageries is utilized for fusion.
a Visible imagery scene from OCTBVS dataset [18] sequence 2
b Thermal Imagery counterpart

Besides, an adjustment in state coordinates is proposed using the above-said concept. The check is made for the least weighting measure, so that if its track location is very far from that of the highest weighing domain, then this should not produce too much offset and a significant deviation in track estimate. A limit is set depending on the tracked object dimensions and the weights obtained after fusion.

$$l = \frac{(AB)(\min(W_t^V, W_t^T))}{2(A+B)(\max(W_t^V, W_t^T))} \qquad (10)$$

If, the less weighted state's coordinates differ from the higher weighted state's coordinates by more than l pixels than shrink this difference to at most pixels, else keep the individual tracker's estimated state locations unchanged.

Finally, after the adaptations made, the final estimate can be found using the following fusion rule

$$F_t = W_t^V S_t^V + W_t^T S_t^T \qquad (11)$$

Here, F_t represents the final output after completion of the fusion process.

IV. EXPERIMENTAL RESULTS & DISCUSSION

The experiments were conducted on 6 different sequences. These sequences were taken from publically known thermo-visual datasets: Ohio State University (OSU) Dataset [28], Bristol Eden Project Multi-Sensor Dataset (BES) [29] and Video Analytics dataset [30]. These datasets contained thermal and visible signatures simultaneously without the need for calibration. The sequences picked had many challenging problems for vision tracking such as camouflaging, the presence of similar background, the presence of similar texture objects, shadow effects, illumination variation, object rotation, and occlusions, etc. Illustration of different dataset sequence used in experimentation is presented in Fig. 4.

Results obtained from our algorithm were compared with five different states of the art trackers viz. ordinary visible colour particle filter (VPF) [18], individual visible template matching (VTM), Track before fusion in

(a)

(b)

(c)

(d)
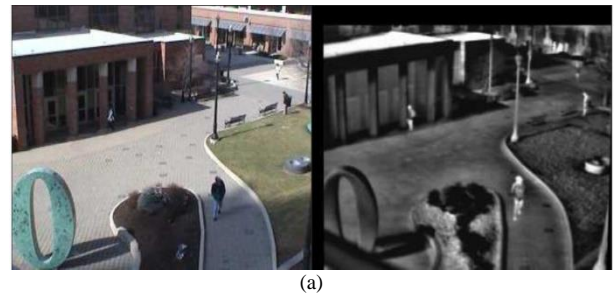
Fig.4. Illustrations of different dataset sequences used
a Visible and Thermal Captures of Frame number 1315 of OSU3
b Visible and Thermal Captures of Frame number 25 of OSU6
c Visible and Thermal Captures of Frame number 782 of SPS
d Visible and Thermal Captures of Frame number 20 of BES

Thermal and Visible (TBF) [19] , Adaptive Data Fusion of Thermal and Visible (ADF) [16], and Adaptive Data fusion via color and gradient (ACGF) [17]. All the computations were performed with MATLAB on the I3 processor system.

A. Parameters Used for Comparitive Performance Analysis

The quantitative comparative analysis was performed using statistical parameters. The parameters opted for comparing the efficiency of the proposed algorithm are Positional Root Square Positional Error (PRSTE), Precision value (PV), Recall Measure Value (RMV) and finally the F1-Measure Value. The comparisons were obtained by taking the object's ground truth positions in the video sequence. The RSPEV resembles the closeness in the position of the track estimate with the accrual position of the target, whereas PV and RV are the resultant of bounding box overlap efficiencies. PV is found as a ratio of intersection area of ground-truth bounding and estimated track bounding box to the area covered by the later. Whereas for RV, the numerator in the ratio remains as it is but the denominator is exchanged by the area of the ground truth-bounding box. F1-measure is the combined measure that is equal to the Harmonic mean of PV and RV. F-score can be criticized in particular circumstances due to its bias as an evaluation metric. The following four equations provide the mathematical calculations of these parameters.

$$PRSTE_t = \sqrt{\left(G_t(x) - P_t^F(x)\right)^2 + \left(G_t(y) - P_t^F(y)\right)^2} \qquad (12)$$

$$PV_t = G_t(area) \cap P_t(area)/P_t(area) \qquad (13)$$

$$GV_t = G_t(area) \cap P_t(area)/G_t(area) \qquad (14)$$

$$FM_t = 2 \times P_t \times R_t/(P_t + R_t) \qquad (15)$$

In above equations, $G_t(x)$ and $G_t(y)$ are the X and Y coordinate of starting point of the actual target state respectively, whereas, $P_t^F(x)$ and $P_t^F(y)$ represent the X and Y coordinate of the starting point of the final track

estimate found through applied fusion algorithm.

B. Detailed Analusis of the performance of proposed Algorithm on the OSU2 Sequence

The OSU2 sequence [28] sequence inhibits different challenging conditions for single visual sensor based tracking. Notably, there are dark spaces between the pillars of the building and a tree present on the right side of the view that is making difficult for target judgment in the visible domain. Whereas the thermal signatures robustly mark the presence of the object as compared to the visible captures due to the temperature difference between the human object and the background.

The aim chosen was to track the human with pure black clothing walking around the street of Ohio State University. It is interesting to note that for this target, the background is very much similar in colour appearance in the visible frame.

As shown in Fig. 5 and Fig. 6, the black rectangle indicates the outcome of VTM, straightway after (1), the yellow rectangle indicates the outcome of thermal correlation-based tracking after application of (3) and (4). The red circle indicates the outcome of the fusion algorithm after (11).

In Fig. 5, it is clear that even if the visible tracker estimates the output due to the colour similarity found at the background void space, but due to clarity provided by the thermal tracker, the proposed algorithm favours more towards the thermal capture information and hence adjusts accordingly to the situation.

Further, in Fig. 6, it is shown that although the estimate of the VTM indicated by the black rectangle is very far away from the actual position of the target object, but visible tracker output does not follow it blindly.
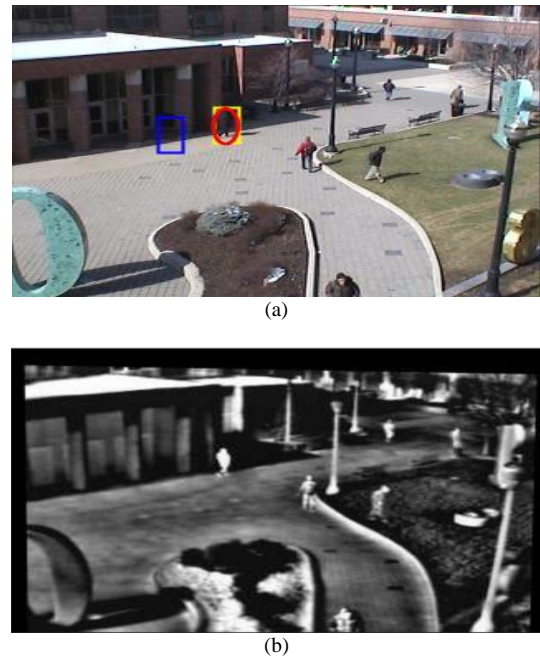

(a)

(b)

Fig.5. Frame number 60 of OSU2 Sequence [7]
a Visible Capture
b Thermal Capture

(a)

(b)

Fig.6. Frame number 180 of OSU2 Sequence [7]
a Visible Capture
b Thermal Capture

Rather, Instead, it makes use of the proposed HTF mechanism and makes a closer judgment, as shown by a trace of the blue rectangle. Finally, after the application of (10) and (11), the outcome is obtained and is enclosed by the red circle. The plots are shown in Fig. 7(a) and 7(b), which provide a comparison of the performance of the different algorithm for the OSU2 sequence. The value of parameters PRSTE and F1- Measure were calculated with the help of (12) and (15), respectively.

It is evident from Fig. 7 that only two algorithms viz. the proposed algorithm (PA) and ADF were able to track the chosen object in OSU2 sequence with reasonable accuracy. The black and red coloured curves in Fig. 7 represented the statistical parameter's obtained values. The F-Measure value for these is mainly lying between 0.8 – 0.9 consistently.

On the other hand, algorithms performing the tracking operation without any input from the thermal imaging (VTM and VPF) provided an inaccurate estimate during the sequence. This fact is supported by high values of RSTE and low F-Measure values in Fig. 7(a) and Fig. 7(b), respectively. The TBF algorithm put up a better performance than alone vision-based methods, but due to improper fusion rules and deputation of fixed value to parameters irrespective to tracking conditions, its performance was not up to mark. The ACGF, although used more than one cue (intensity and gradient) to formulate the fusion basis, its performance deteriorates considerably after 80 starting frames, because of false predictions of visible imaging. The improvement in tracking results by ADF, once again confirms the case of the effectiveness of thermal imaging assisted visual tracking. However, the method is complicated and time-consuming, so the proposed method provides a better alternative to it.

(a)
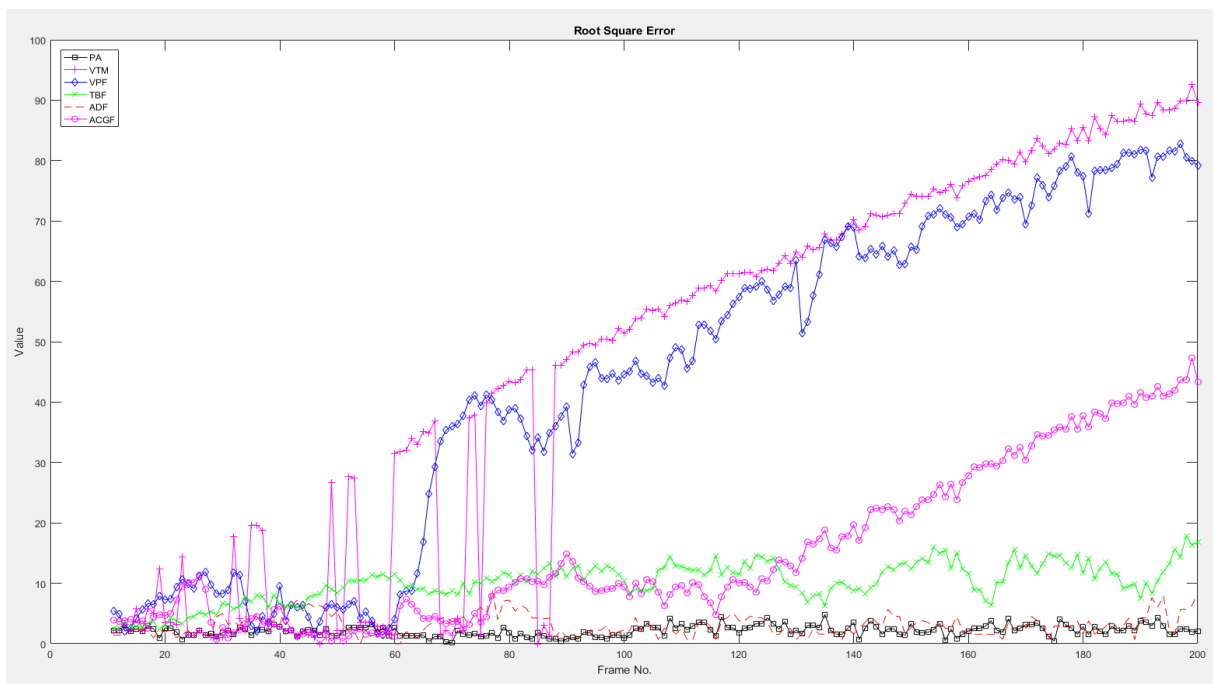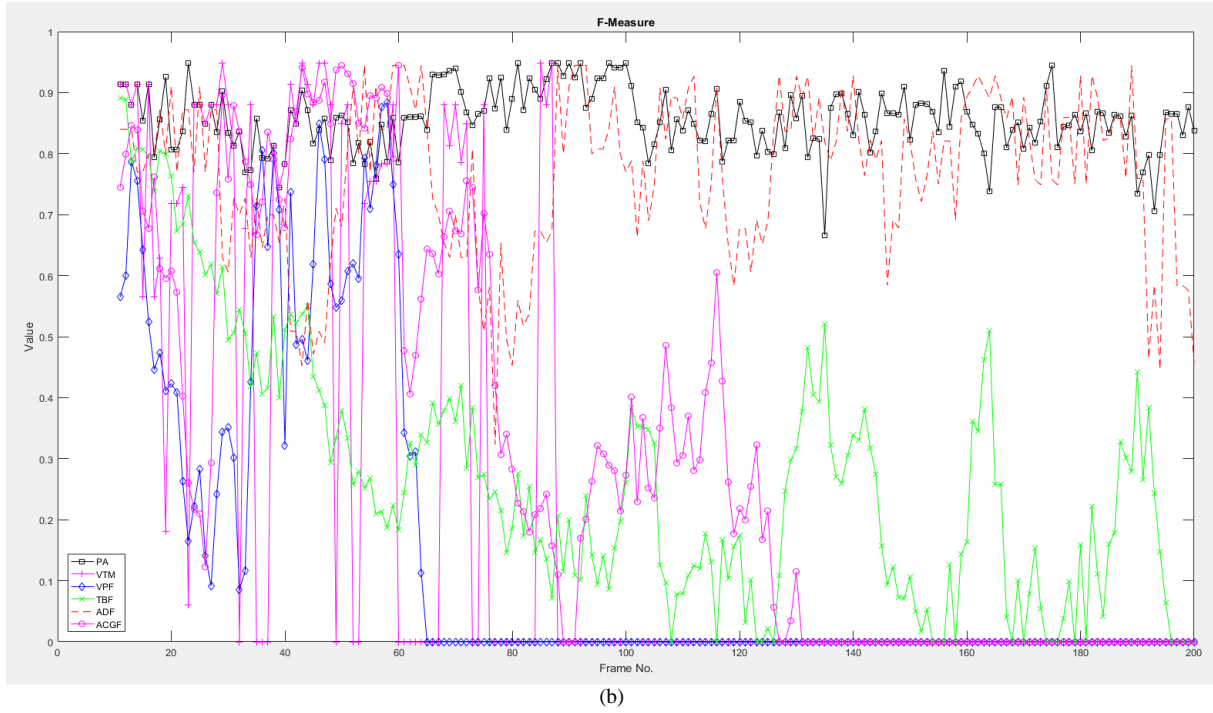
(b)

Fig.7. Plots of Statistical Parameters used for Quantitative Comparisons of the different algorithm for OSU2 sequence

C. Overall Performance Comparison

The overall results obtained are comprehensively presented in this subsection with the help of Fig. 8, Table 2 and Table 3. The computation time analysis in Fig. 8 depicts that the proposed method only lags behind the single sensor imaging frameworks, i.e., VTM and VPF which themselves have very dejected track accuracy as indicated by data from Table 2 and Table 3. Nevertheless, the computation time is least among the compared thermos-visual fusion based state of art trackers.
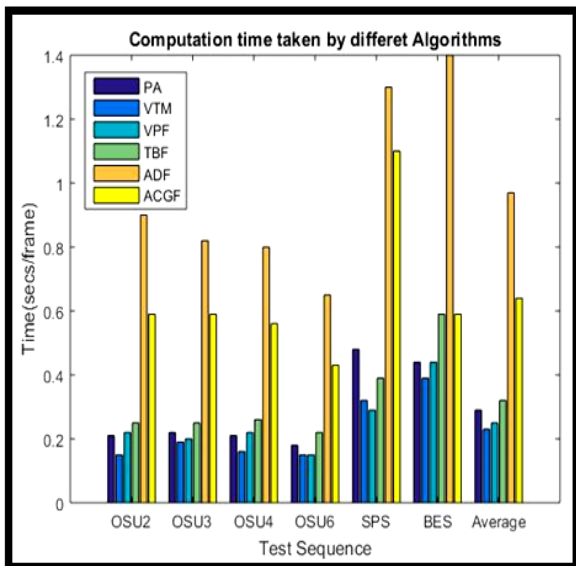

Fig.8. Computation time comparisons under similar experimental conditions.

Table 2. PRSTE comparisons

Test Seq.	Algorithm Used					
	PA	VTM	VPF [18]	TBF [19]	ADF [16]	ACGF [17]
OSU2 [28]	**2.82**	47.15	43.42	8.67	*3.46*	13.06
OSU3 [28]	**3.30**	*4.20*	14.22	28.35	13.27	7.17
OSU4 [28]	*4.27*	52.67	5.81	**3.72**	7.35	5.78
OSU6 [28]	*3.26*	40.52	60.57	47.78	**3.22**	40.45
SPS [30]	**10.67**	84.89	20.4	*10.7*	11.7	23.1
BES [29]	**6.10**	91.81	18.74	9.96	*6.90*	13.60
Average	**5.07**	53.54	27.19	18.20	*7.65*	17.19

Table 3. F1-Measure comparisons

Test Seq.	Algorithm Used					
	PA	VTM	VPF [18]	TBF [19]	ADF [16]	ACGF [17]
OSU2 [28]	**0.86**	24	0.16	0.3	*0.77*	0.44
OSU3 [28]	**0.80**	*0.78*	0.44	0.24	0.54	0.61
OSU4 [28]	*0.70*	0.65	0.69	**0.72**	0.66	0.70
OSU6 [28]	*0.76*	0.10	0.13	0.17	**0.77**	0.24
SPS [30]	**0.64**	0.24	0.51	0.61	*0.63*	0.44
BES [29]	**0.85**	0.32	0.55	0.73	*0.83*	0.58
Average	**0.77**	0.38	0.41	0.46	*0.70*	0.50

The tables present the performance of algorithms in terms of the positional error term, and overlap area based F-Measure statistics computations. The computed parameter values represent the average value computed over the total no. of frames chosen for experimentation for an individual sequence, respectively. The top-performing algorithm is marked in bold letters and the second in italics letters.

It was observed that the proposed method outperforms other trackers by sustaining a consistent performance throughout all sequences. Out of the six sequences chosen for experimentation, it performs best in four cases viz. OSU2, OSU3, SPS and BES. Along with this, the proposed algorithm secures runner up position for the remaining two sequences.

V. Conclusion

The results obtained in the proposed work give the impression that the presented idea of thermo-visual fusion tracking in the template-matching framework delivers accurate and robust track estimates. A novel fusion strategy was developed, which was found to perform better in comparison to the renowned methods in the literature. Further, lesser time to process indicates the computational ease for tracking by this method. Along with the proposal of a fusion strategy, the accuracy in standard template matching was increased by the introduction of an HTC process. The algorithm acquired the leading position in accuracy and F0measures for four different data sequence and attained a second-best position for the rest of datasets. The overall average was the best among the various state of the art tracking systems in the thermo-visual fusion-tracking field. Future emphasis concentrates on the design of a real-time system in the template matching system that adds to this work by associating multiple features/cues from multiple sources to further increase affectivity of a tracking system under various challenged vision conditions.

References

[1] U. Ali and M. Hanif, "Optimized Visual and Thermal Image Fusion for Efficient Face Recognition," in *IEEE International Conference on Information Fusion*, 2006.

[2] G. Bebis, A. Gyaourova and I. Pavlidis, "Face Recognition by Fusing Thermal Infrared and Visible Imagery," *Image and Vision Computing*, vol. 24, no. 7, pp. 727–742, 2006.

[3] J. Heo, S. G. Kong, B. R. Abidi, and M. A. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2004, pp. 122–122.

[4] T. Wilhelm, H. J. B¨ohme, and H. M. Gross, "A multi-modal system for tracking and analyzing faces on a mobile robot," *Robotics and Autonomous Systems*, vol. 48, no. 1, pp. 31–40, 2004.

[5] D. R. Perrott, J. Cisneros, R. L. McKinley, and W. R. D'Angelo, "Aurally aided visual search under virtual and free-field listening conditions." *Human Factors*, vol. 38, no.4, pp. 702-715, 1996.

[6] G. Cielniak, and T. Duckett, "Active People Recognition using Thermal and Grey Images on a Mobile Security Robot," *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2005, pp. 3610–3615.

[7] G. Cielniak, T. Duckett, and A. J. Lilienthal, "Improved data association and occlusion handling for vision-based people tracking by mobile robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 3436–3441.

[8] G. B. Palmerini and S. Universit`a, "Combining Thermal and Visual Imaging in Spacecraft Proximity Operations," in *International Conference on Control Automation Robotics Vision*, 2014, pp. 383–388.

[9] Y. Tong, L. Liu, M. Zhao, J. Chen, and H. Li, "Adaptive fusion algorithm of heterogeneous sensor networks under different illumination conditions," *Signal Processing*, vol. 126, pp. 149–158, 2016.

[10] Z. Zhou, B. Wang, S. Li, and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters," *Information Fusion*, vol. 30, pp. 15–26, 2016.

[11] C. Li, X. Wu, N. Zhao, X. Cao, and J. Tang, "Fusing two stream convolutional neural networks for RGB-T object tracking," *Neurocomputing*, vol. 281, pp. 78–85, 2018.

[12] G. S. Walia and R. Kapoor, "Recent advances on multicue object tracking: a survey," *Artificial Intelligence Review*, vol. 46, no. 1, pp. 821–847, 2016.

[13] S. Singh, R. Kapoor, and A. Khosla, *Cross-Domain Usage in Real Time Video-Based Tracking*. U.S.A: IGI Global, 2017, pp. 105–129.

[14] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Information Fusion*, vol. 45, pp. 153–178, 2018.

[15] C. O. Conaire, N. E. O. Connor, and A. Smeaton, "Thermo-visual feature fusion for object tracking using multiple spatiogram trackers," *Machine Vision and Applications,* vol. 19, no. 5-6, pp. 483–494, 2008.

[16] M. Talha and R. Stolkin, "Particle filter tracking of camouflaged targets by adaptive fusion of thermal and visible spectra camera data," *IEEE Sensors Journal*, vol. 14, no. 1, pp. 159–166, 2014.

[17] J. Xiao, R. Stolkin, M. Oussalah, and A. Leonardis, "Continuously Adaptive Data Fusion and Model Relearning for Particle Filter Tracking With Multiple Features," *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2639– 2649, 2016.

[18] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive colorbased particle filter," *Image and Vision Computing*, vol. 21, no. 1, pp. 99–110, 2003.

[19] G. Xiao, X. Yun, and J. Wu, "A new tracking approach for visible and infrared sequences based on tracking-before-fusion," *International Journal of Dynamics and Control*, vol. 4, no. 1, pp. 40-51, 2016.

[20] R. Stolkin, D. Rees, M. Talha, and I. Florescu, "Bayesian fusion of thermal and visible spectra camera data for region based tracking with rapid background adaptation," in *IEEE Int. Conf. Multisens. Fusion Integr. Intell. Syst.*, 2012, pp. 192–199.

[21] E. Fendri, R. R. Boukhriss, M. Hammami, "Fusion of thermal infrared and visible spectra for robust moving object detection," *Pattern Anal. Appl.*, vol. 20, no. 4, pp. 907–926, 2017.

[22] S. R. Schnelle, and A.L. Chan, "Enhanced target tracking through infrared-visible image fusion," in *14th Int. Conf. Inf. Fusion*, 2011, pp. 1–8.

[23] Y. Niu, S. Xu, L. Wu, and W. Hu, "Airborne infrared and visible image fusion for target perception based on target

region segmentation and discrete wavelet transform," *Math. Probl. Eng.*, 2012.

[24] Y. Wu, E. Blasch, G. Chen, L. Bai, L. Ling, "Multiple source data fusion via sparse representation for robust visual tracking," in *14th International Conference on Information Fusion,* 2011, pp. 1–8.

[25] C. Li, H. Cheng, S. Hu, X. Liu, J. Tang, and L. Lin, "Learning Collaborative Sparse Representation for Grayscale-Thermal Tracking," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5743–5756. 2016.

[26] C. Li, S. Hu, S. Gao, and J. Tang, "Real-time grayscale-thermal tracking via laplacian sparse representation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, pp. 54–65.

[27] H. P. Liu and F.C. Sun, "Fusion tracking in color and infrared images using joint sparse representation," *Sci. China-Information Sci.*, vol. 55, no. 3, pp. 590–599, 2012.

[28] J. Davis and V. Sharma, "Background-subtraction using contour based fusion of thermal and visible imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2-3, pp. 162–182, 2007.

[29] Bristol Eden Project Multi-Sensor Data Set, http://www.cis.rit.edu/pelz/scanpaths/data/bristol-eden.htm/.

[30] Video Analytics Dataset, https://www.ino.ca/en/video-analytics-dataset/.

Arun Khosla Arun Khosla received his Ph.D. degree from Indraprastha University, Delhi in the field of Information Technology. He is presently working as Professor in the Department of Electronics and Communication Engineering, National Institute of Technology, Jalandhar, India. Dr. Khosla has been reviewer for various IEEE and other National and International conferences and also serves on the editorial board of International Journal of Swarm Intelligence Research. He is a life member of Indian Society of Technical Education.

Rajiv Kapoor Rajiv Kapoor received M.E. and Ph.D. degree in Electronics & Communication Engineering from Delhi College of Engineering, Delhi University. Dr. Kapoor presently working as Professor in Electronics & Communication Engineering Department, AIACT&R (Govt. of NCT of Delhi). He has authored over 90 research papers in various renowned international journal and conferences. His primary research interests are machine learning, computer vision, signal and image processing.

Authors' Profiles

Satbir Singh Satbir Singh received M.E in Electronics and Communication Engineering from Thapar University, Patiala, India. He is presently working with Central scientific Instruments Organization, Chandigarh, India. Previously, he has a working experience as Senior Scientific Officer with Electronics and Communication Engineering Department of Delhi Technological University, Delhi and as a Project Engineer with Centre of Advanced Computing, Mohali, India. His research interests include signal processing, computer vision and IoT. Currently, he is pursuing Ph.D. from National Institute of Technology, Jalandhar, India.