This version is available at https://strathprints.strath.ac.uk/66379/

# Fusion of Thermal and Visible Imagery for Effective Detection and Tracking of Salient Objects in Videos

Yijun Yan[1], Jinchang Ren[1], Huimin Zhao4, Jiangbin Zheng2, Ezrinda Mohd Zaihidee[1,3] and John Soraghan[1]

[1] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K
[2] School of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, China
[3] Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Malaysia
[4] School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou, China

**Abstract.** In this paper, we present an efficient approach to detect and track salient objects from videos. In general, colored visible image in red-green-blue (RGB) has better distinguishability in human visual perception, yet it suffers from the effect of illumination noise and shadows. On the contrary, thermal image is less sensitive to these noise effects though its distinguishability varies according to environmental settings. To this end, fusion of these two modalities provides an effective solution to tackle this problem. First, a background model is extracted followed by background-subtraction for foreground detection in visible images. Meanwhile, adaptively thresholding is applied for foreground detection in thermal domain as human objects tend to be of higher temperature thus brighter than the background. To deal with cases of occlusion, prediction based forward tracking and backward tracking are employed to identify separate objects even the foreground detection fails. The proposed method is evaluated on OTCBVS, a publicly available color-thermal benchmark dataset. Promising results have shown that the proposed fusion based approach can successfully detect and track multiple human objects.

**Keywords:** Video salient objects, Pedestrian detection/tracking, Image fusion, Visible image, Thermal image.

## 1    Introduction

In the past decades, detection and tracking of video objects has always been a major task in the computer vision field [1, 2]. As one subset of video object tracking, pedestrian detection and tracking has drawn massive research attention and been applied on many applications such as visual surveillance [3, 4], driver-assistance systems [5, 6], human activity recognition[7, 8], etc. For pedestrian detection and tracking, visible camera and thermal imagery are

two popularly used sources of image modalities, though not necessarily in a combined solution [9-11]. However, either visible image or thermal image has their advantages and disadvantages. Visible image can show detailed color information, however it really suffer from lighting variations, cluttered backgrounds, artificial appearances i.e. shadows and etc. Since the object is detected by its temperature and radiated heat, thermal image can eliminate the influence of color and illumination changes on the objects' appearance [12] in any weather conditions and at both day and night time. However, in some cases e.g. occlusions, thermal camera may fail to detect the object properly. In Fig.1, there are three pedestrian templates, for the one with yellow rectangle; both visible and thermal image can detect it very well since it has high contrast to the background in visible and human temperature in thermal domain. For the template in red rectangle, it has a compact shape in the thermal image; however in visible image, we can just identify it coarsely due to similar appearance in color of the background and the person's cloth. The one in green rectangle can be seen in visible image but hardly observed in the corresponding thermal image. This is because thermography is only able to directly detect surface temperatures, and it can't work well when the object is (partially) occluded. Moreover, it will detect any objects (e.g. windows and car in Fig.1) with surface temperature.
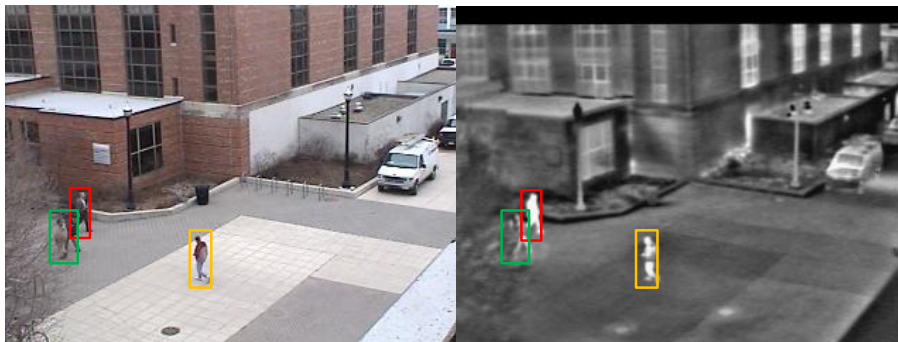


Fig. 1. Visible image of a scene (left) and thermal image of the same scene (right).

In this paper, we present a pedestrian detection and tracking algorithm that fuse the information from both visible and thermal images. To detect the region of pedestrian in visible domain, we firstly apply a statistical method to model the background before applying background subtraction to extract the

foreground regions. In thermal domain, we apply a histogram-based adaptive threshold method to detect all the objects with temperature. Then, the salient pedestrian regions can be obtained by the fusion of both visible and thermal binary maps. In addition, we also proposed a backward tracking method to automatically identify the individual pedestrian template from pedestrian group (detailed in Section 3).

The rest of the paper is organized as follows: Section 2 describes the foreground detection approach. Section 3 elaborates the backward tracking method. Experimental results are presented and discussed in Section 4. Finally, some concluding remarks and future work are summarized in Section 5.

## 2    Foreground detection

To capture the region of pedestrians in visible image, we first estimate the background image by computing a median map (Fig.2(a)) of N frames randomly selected from the video sequence. And background subtraction (Fig.2(b)) for each visible frame is defined as :

$$BS_{vis}(x, y) = |I(x, y) - I_{med}(x, y)|$$

After that, we binarize the $BS_{vis}$ with an empirical threshold (Fig.2(c)) to get a binary image $I_{vis}$ with coarse human body region. Then a binary mask $I_{mask}$ with compact ROI is generated by applying morphology and filtering the insignificant region (Fig. 2(d)) on.

For thermal image, we introduce a histogram based threshold selection method inspired by [13] to segment the objects with salient temperature and get a binary image $I_{thm}$ (Fig.2(e)). Given a thermal map with L grey levels, the probability of each grey level is denoted by:

$$P_i = \frac{n_i}{N}, \text{and } P_i \geq 0, \sum_{i=1}^{L} P_i = 1$$

Furthermore, the probability distribution $P_i$ is normalized within [0,1] and regarded as probability occurrence $\lambda(k)$ at each gray level.

$$\lambda(\mathrm{k}) = 1 - \sum_{i=1}^{L} P_i\,(k)$$

Let $\mu$, $\gamma$, $\sigma$ be the mean, median and variance of the probability occurrence value in $\lambda$, and let the threshold T be defined as a weighted combination of these parameters.

$$\mathrm{T} = \omega_\mu \mu + \omega_\gamma \gamma + \sigma$$

where $\omega_\mu, \omega_\gamma > 0 \ and \ \omega_\mu + \ \omega_\gamma = 1$.
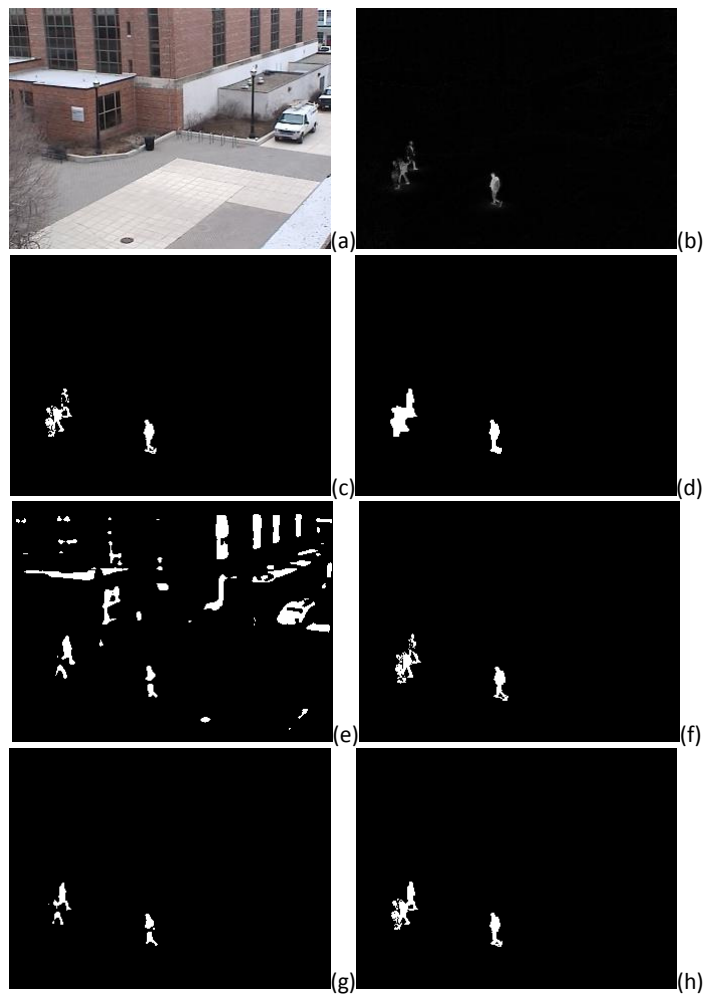
**Fig. 2.** (a) Median map from N frames, (b) grayscale background subtraction result, (c) binary visible image from Fig.1, (d) filtered mask, (e) binary thermal image from Fig.1, (f) intersection of (c) and (d), (g) intersection of (d) and (e), (h) fusion of binary visible and thermal image.

Then we can calculate the intersection of the binary visible image and filtered mask (Fig.2(f)), and intersection of the binary thermal image and mask (Fig.2(g)), respectively. Finally, the salient pedestrian segmentation (Fig.2(h)) can be done by the fusion of visible and thermal binary results as follows:

$$I_{final} = (I_{mask} \cap I_{vis}) \cup (I_{mask} \cap I_{thm})$$

## 3    Object tracking

After object detection, we can identify most individual objects very well. However, in some frames, the occlusion problem is inevitable. Fig.3 (right and middle) shows the detection detail of two adjacent frames where there should be three pedestrian patterns detected in both frames, but for the frame in the middle, the object detection method considers the left two patterns as one object because they are close to each other. Therefore, we use mean-shift method [14] to detect the individual objects from the objects group. As can be seen from the right image in Fig. 3, the identification result for second frame is updated.
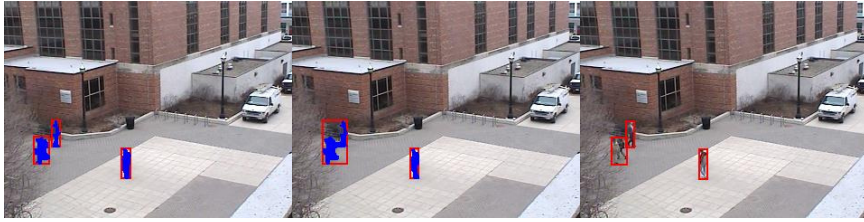


**Fig. 3.** Initial detection result of frame 1 (left) and frame 2 (middle), and updated detection result of frame 2 after mean-shift tracking progress.

## 4    Experimental results

To evaluate the performance of our method, we present our method on sequence 4, 03 OSU Color-Thermal Database from the popular benchmark da-

tabase OTCBVS. Thermal sequences are captured by Raytheon PalmIR 250D thermal sensor and color sequence are captured by Sony TRV87 Handycam color sensor. All the frames in both sequences have a spatial resolution of 320*240 pixels.
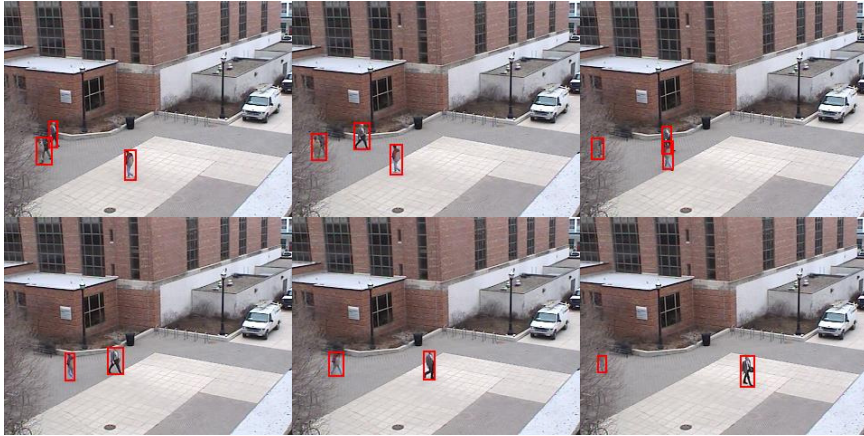


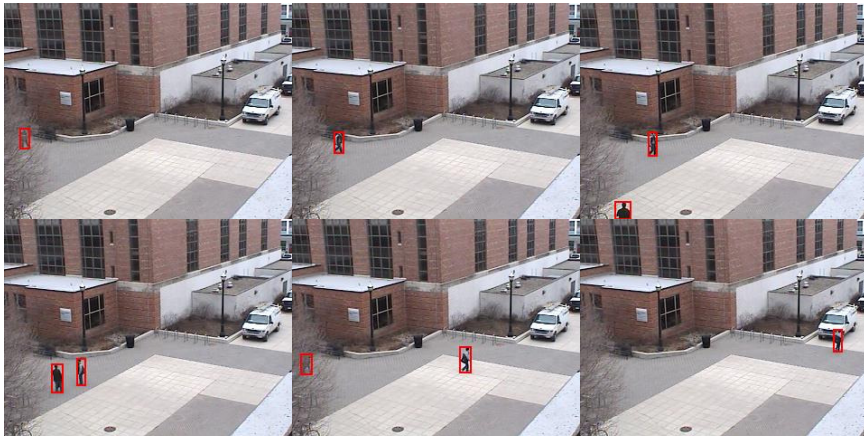**Fig. 4. Detection and tracking results for sequence 4.**



**Fig. 5. Detection and tracking results for sequence 5.**

To validate the performance of the proposed approach, three different sequences are used in our experiments. In Figs. 4-6, detection and tracking results from these sequences are given to illustrate the extracted/tracked objects using their bounding boxes. As can be seen, the proposed method

can give reliable pedestrian detection and tracking results under various conditions, including occlusion and changes in terms of illumination and scale. When the pedestrians are independent, we can detect them very well with proper scale bounding box. We can also identify the people even they are overlapped e.g. the images in first row in Fig. 4. In addition, when there is some occlusion appears like tree or wall, or the object is getting out of the screen (e.g. the image in Fig. 5 and 6), we can still locate the objects and track their motion.
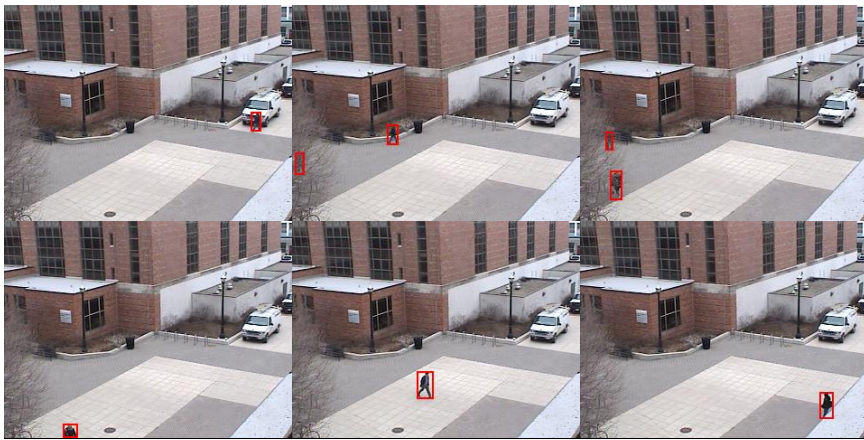


**Fig. 6. Detection and tracking results for sequence 6.**

## 5    Conclusion

In this paper, we proposed an efficient approach to fuse visible and thermal images for salient object detection and tracking in videos. By estimating the background model followed by background subtraction, foreground objects can be successfully detected with adaptive thresholding applied to the salient temperature maps of the thermal image.  Prediction based forward and backward tracking is found particularly useful to separate overlapped or occluded objects. In future work, we will further improve the method and test the performance on other more challenging datasets, including those with human shadow and large scale illumination changes.

# 6    References

1.  M. Li, C. Wei, Y. Yuan, and Z. Cai, "A Survey of Video Object Tracking," *International Journal of Control and Automation,* vol. 8, pp. 303-312, 2015.

2.  A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm computing surveys (CSUR),* vol. 38, p. 13, 2006.

3.  B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 3457-3464.

4.  O. Sidla, Y. Lypetskyy, N. Brandle, and S. Seer, "Pedestrian detection and tracking for counting applications in crowded situations," in *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on*, 2006, pp. 70-70.

5.  J. Ge, Y. Luo, and G. Tei, "Real-time pedestrian detection and tracking at nighttime for driver-assistance systems," *Intelligent Transportation Systems, IEEE Transactions on,* vol. 10, pp. 283-298, 2009.

6.  D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* pp. 1239-1258, 2009.

7.  R. Bodor, B. Jackson, and N. Papanikolopoulos, "Vision-based human tracking and activity recognition," in *Proc. of the 11th Mediterranean Conf. on Control and Automation*, 2003.

8.  R. Poppe, "A survey on vision-based human action recognition," *Image and vision computing,* vol. 28, pp. 976-990, 2010.

9.  J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *null*, 2005, pp. 364-369.

10. J. W. Davis and V. Sharma, "Robust background-subtraction for person detection in thermal imagery," *IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum,* 2004.

11. J. W. Davis and V. Sharma, "Robust detection of people in thermal imagery," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. 713-716.

12. D.-E. Kim and D.-S. Kwon, "Pedestrian detection and tracking in thermal images using shape features," in *Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on*, 2015, pp. 22-25.

13. J. Ren and T. Vlachos, "Segmentation-assisted detection of dirt impairments in archived film sequences," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 37, pp. 463-470, 2007.

14. D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, pp. 142-149.