

MULTIMODAL DATA FUSION FOR EFFECTIVE SURVEILLANCE OF CRITICAL INFRASTRUCTURE

Z. Kandylakis^{a,b*}, K. Karantzas^{a,b}, A. Doulamis^a, L. Karagiannidis^b

^aRemote Sensing Lab, National Technical University of Athens, (kandylakis, karank)@central.ntua.gr

^bInstitute of Communication and Computer Systems, lkaragiannidis@iccs.gr

Commission III, WGIII/4 Hyperspectral Image Processing

KEY WORDS: Hyperspectral, SWIR, thermal, video, multisensor, detection, tracking, moving object

ABSTRACT: Monitoring critical infrastructures, especially those that are covering wide-zones, is of fundamental importance and priority for modern surveillance systems. The concurrent exploitation of multisensor systems, can offer additional capabilities, on day and night acquisitions and different environmental/illumination conditions. Towards this direction, we have designed a multi-sensor system based on thermal, shortwave infrared and hyperspectral video sensors. Based on advanced registration, dynamic background modelling and data association techniques, possible moving targets are detected on the thermal and shortwave infrared modalities. In order to avoid the computational intensive co-registration with the hyperspectral video streams, the detected targets are projected through a local coordinate system on the hypercube image plane. The final detected and verified targets are extracted through fusion and data association, based on temporal spectral signatures and target/background statistics. The developed multisensor system for the surveillance of critical infrastructure has been validated for monitoring wide-zones against different conditions showcasing abilities for detecting and tracking moving targets through fog and smoke.

1. INTRODUCTION

During the last decade, research and development in optics, photonics and nanotechnology permitted the introduction of new innovative video sensors which can cover a wide range of the UV, visible as well as near, shortwave and longwave infrared spectrum. Multispectral and hyperspectral video sensors based mainly on filterwheels, on micro-patterned coatings on individual pixels, on optical filters monolithically integrated on top of CMOS image sensors have been developed and are gradually becoming available from the industry.

Hyperspectral video technology has been employed for the detection and tracking of moving objects for engineering, security and environmental monitoring applications. Several detection algorithms have been studied for different applications with moderate to sufficient effectiveness (Manolakis et al., 2014), (Pieper et al., 2015). In particular, hyperspectral video systems have been employed for developing object tracking methodologies through hierarchical decomposition for chemical gas plume tracking (Tochon et al., 2017). Multiple object tracking based on background estimation in hyperspectral video sequences has been also addressed (Kandylakis et al., 2015). Certain processing pipelines have been also proposed to address the changing environmental illumination conditions (Pieper et al., 2015).

These detection capabilities are gradually starting to be integrated with other video modalities like e.g., standard optical (RGB), thermal and other sensors towards the effective automation of the recognition modules. For security applications, the integration of multisensor information has been recently proposed towards the efficient fusion of the

heterogeneous information for developing robust large-scale video surveillance system (Fan et al., 2017).

Multiple target detection, recognition and tracking and security event recognition is an important application of computer vision with significant attention on human motion/activity recognition and abnormal event detection (Liao et al., 2017). Most algorithms are based on learning robust background model however, estimating a foreground/background model is very sensitive to illumination changes, extracting the foreground objects as well as recognition its class/label is not always trivial.

Novel approaches in multiple target tracking algorithms include automated segmentation and tracking modules based on CRF models (Milan et al., 2015). Moreover, the simultaneous addressing of data association and trajectory reconstruction tasks has been proposed through the use of energy minimization functions, signifying a shift from the traditional tracking-by-detection paradigm (Milan et al., 2016). The use, furthermore, of socio-topology models for the minimization of the topology-energy-variation function, has shown promising results for multiple person tracking in crowd scenes (Gao et al. 2017).

Towards the same direction, a greedy batch-based minimum-cost flow approach has been proposed, employing a generalized minimum-cost flows (MCF) algorithm on each batch to generate a set of trajectories with different probabilities (Wang et al., 2017). In addition, a hybrid data association framework has been proposed which utilizes global data association, taking multiple video frames into account to alleviate

* Corresponding author

irrecoverable errors caused by the local data association between adjacent frames (Yang et al., 2017).

Moreover, to address mutual occlusions and imprecise image based observations, a new predictive model on the basis of Gaussian Process Regression has been proposed, which utilizes generic object detection, as well as instance-specific classification, for refined localisation (Klinger et al., 2017). For the re-detection of the target in the case of long-term tracking drifts, a feature integration object tracker named correlation filters and online learning (CFOL) has shown promising results as well (Zhang et al., 2017).

In this paper, we build upon recent developments (Kandylakis et al., 2015) on multiple object tracking from a single hyperspectral sensor and have integrated another two thermal and shortwave (SWIR) video ones. The monitoring system can perform both during the night and daytime by exploiting, through multimodal data fusion, the spectral observations of every sensor. Moreover, the developed system has been validated against different conditions, showcasing abilities for detecting moving targets through fog/smoke, delivering approximation and/or intrusion alerts effectively.

2. METHODOLOGY

2.1 The multisensor video system

The developed multisensor system consists of a thermal camera, a hyperspectral camera, a SWIR camera as well as an RGB sensor for cross-reference and validation (Table 1). The thermal sensor is FLIR’s TAU2, with the capability of recording one band in the range of 8 to 13 μm , at a spatial resolution of 620×480 , and at a rate of 9 Hz. The SWIR sensor is Xeneth’s Bobcat 640, which covers the range of 900 to 1700 nm, at a resolution of 640×512 and a recording rate of 100 Hz. The hyperspectral sensor is based on an imec snapshot mosaic CMOS. It acquires 41 bands, in the range of 400 to 950 nm, at a resolution of 500×270 per band and has a frame rate of 24 fps.

Sensor	Thermal	Hyperspectral	SWIR
Spectral Range	8-13 μm	400 – 950nm	900 – 1700nm
Spectral Resolution	1 band	41 bands	1 band
Spatial Resolution	620×480	500×270	640×512
Frame Rates	9 Hz	24 fps	max 100 Hz
Lens	19mm	35mm	25mm
Interface	USB3	GiGE	GiGE

Table 1. Sensor Specifications

The hyperspectral sensor is accompanied with an FPGA that handles the frame acquisition. All the sensors are then connected to a mini-ATX local processing unit which handles the rest of the processing.

During data acquisition, the sensors are mounted on a relatively high fixed platform or tripod, acquiring oblique views of the Region of Interest (ROI). Although fixed, the sensors and the video sequence is affected by the changing wind and sudden abrupt bursts. The sensors and their corresponding field of view (FOV) are presented in Figure 1.

Due to their lens configuration, the thermal sensor has a wider field of view, followed by the SWIR sensor which observes a relatively smaller area. The hyperspectral sensor has the relatively smaller FOV, while all are covering the ROI. The ROI plane is associated with a Local Coordinate System (LCS).

2.2 Multi-modal data processing

The developed system integrates a number of software modules like dimensionality reduction algorithms (Karantzas 2009), registration (Karantzas et al., 2014), (Vakalopoulou and Karantzas, 2014), background subtraction, moving objection detection and calculation of optical flows, velocity/ density/ direction (Kandylakis et al., 2015).

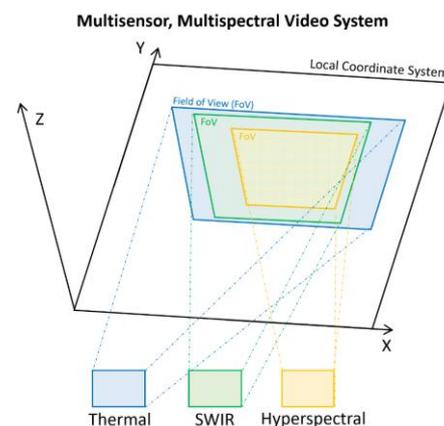


Figure 1: The main three sensors of the multisensor system and their corresponding field of view (FOV) on the region of interest (ROI). The ROI plane is associated with an arbitrarily defined, local coordinate system (LCS).

Moreover, certain software modules are responsible for performing scene classification tasks based on recent approaches like in (Makantasis et al., 2015). The monitoring of activity inside a desired ROI, and the projection of all frames in the same coordinate system for geo-referencing have also been addressed.

The first step before the main processing pipeline is presented in Figure 2, which is the calculation of the 3×3 transformation matrices for the projection of all three image planes on an arbitrarily defined Local Coordinate System (LCS). Also, all the inverse transformation matrices were computed for the inverse perspective transformations from the LCS to the image planes. A significant advantage of this approach is that, the actual projection of the entire image (hypercube) is not required, omitting a computationally expensive step. Instead, only the coordinates of the possible moving objects are converted between reference systems.

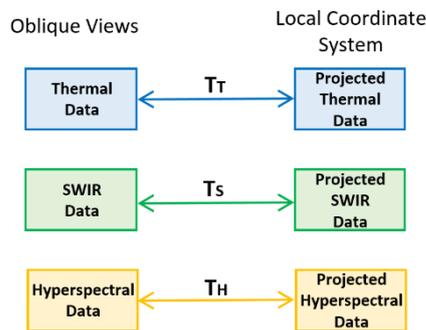


Figure 2: Establishing correspondences among the Field of Views. The perspective transformations and inverse perspective transformations are estimated and employed to convert coordinates to the Local Coordinate System (LCS) from the oblique views of all image planes and vice versa.

The main processing pipeline is summarized in Figure 3. In order to keep the computation complexity as low as possible while allowing near-real time performance, the possible moving targets are detected on the SWIR and/or the thermal sensor (covering both day and night acquisitions). On these modalities data registration, dynamic background estimation and data association are executed towards the detection of the possible moving objects/targets (PMT). The background estimation is based on an adaptive procedure during which the background was dynamically estimated based on the mean intensity value of approximately 50 frames. The registration is performed per frame in order to address the slightly moving FOV due to abrupt winds.

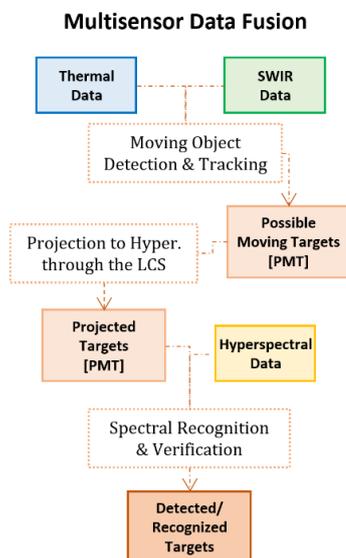


Figure 3: The processing pipeline for the multisensor data fusion

The possible detected targets, in a binary form on the SWIR or thermal image plane, are then projected into the LCS (Figure 2). In particular, the bounding box or polyline coordinates are projected to the LCS through the use of the transformation matrix T_s . The resulting coordinates are then projected to the hyperspectral image plane using the inverse transformation matrix T_H^{-1} . These projected targets are then directly fused on the hyperspectral image plane, avoiding the hypercube co-

registration with the other modalities. The final detected targets are extracted (Figure 4) after their spectral verification and recognition with through smoke/fog capabilities based on data association modules that exploit their temporal spectral signatures (Figure 5).

3. EXPERIMENTAL RESULTS AND VALIDATION

Several experiments have been performed in order to develop and validate the performance of the different hardware and

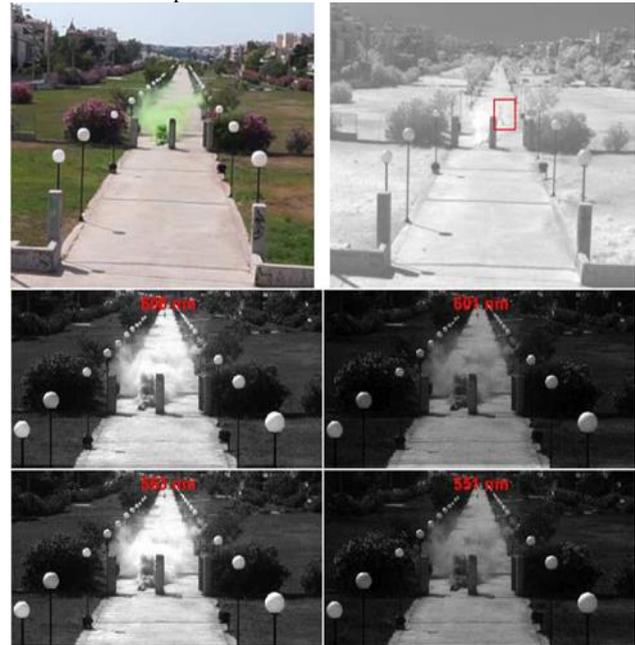


Figure 4: The multimodal data fusion module can detect the moving target even through smoke/fog conditions based on the best available modality, data association through temporal spectral signatures and efficient fusion modules on a given region of interest.

software modules. A number of experiments have taken place in the framework of the ZONeSEC FP7 EU project (<https://www.zonsec.eu/>). During all our experiments, although the sensors were mounted on a single platform and carefully fixed, due to the changing wind and abrupt wind bursts, slight movements on each FOV were encountered, which were addressed by the co-registration software modules in near-real time.

In Figure 6, experimental results after the application of the developed hardware and software systems are presented. In particular, the final detection moving targets (their polygons) for four indicative frames are presented. These polygons are overlaid on the respective acquired SWIR image that was employed in the detection process, as well as on three hyperspectral bands centered around 476, 539, and 630nm, respectively. The bounding boxes are outlined in red, and zoomed in views are provided for better distinction. It can be observed that the projection of the quadrilateral bounding box on the hyperspectral image plane, distorts it slightly into a more general polygon shape.

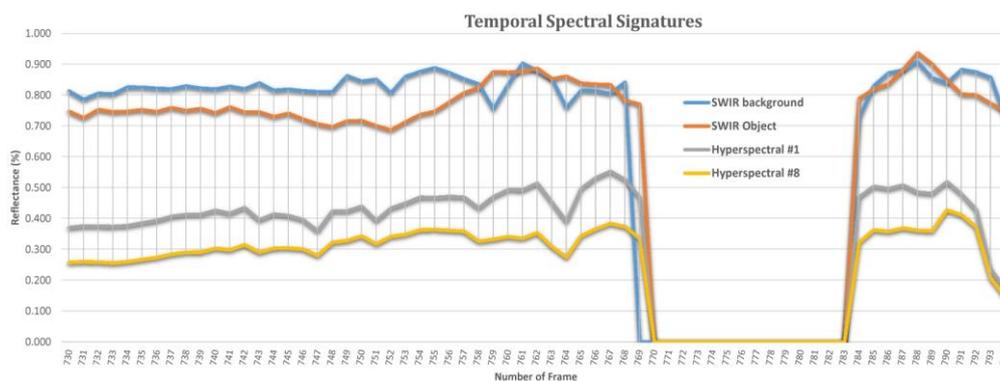


Figure 5: The temporal spectral signatures of both the targets as well as the background are calculated and employed during the data association and fusion recognition and verification step.

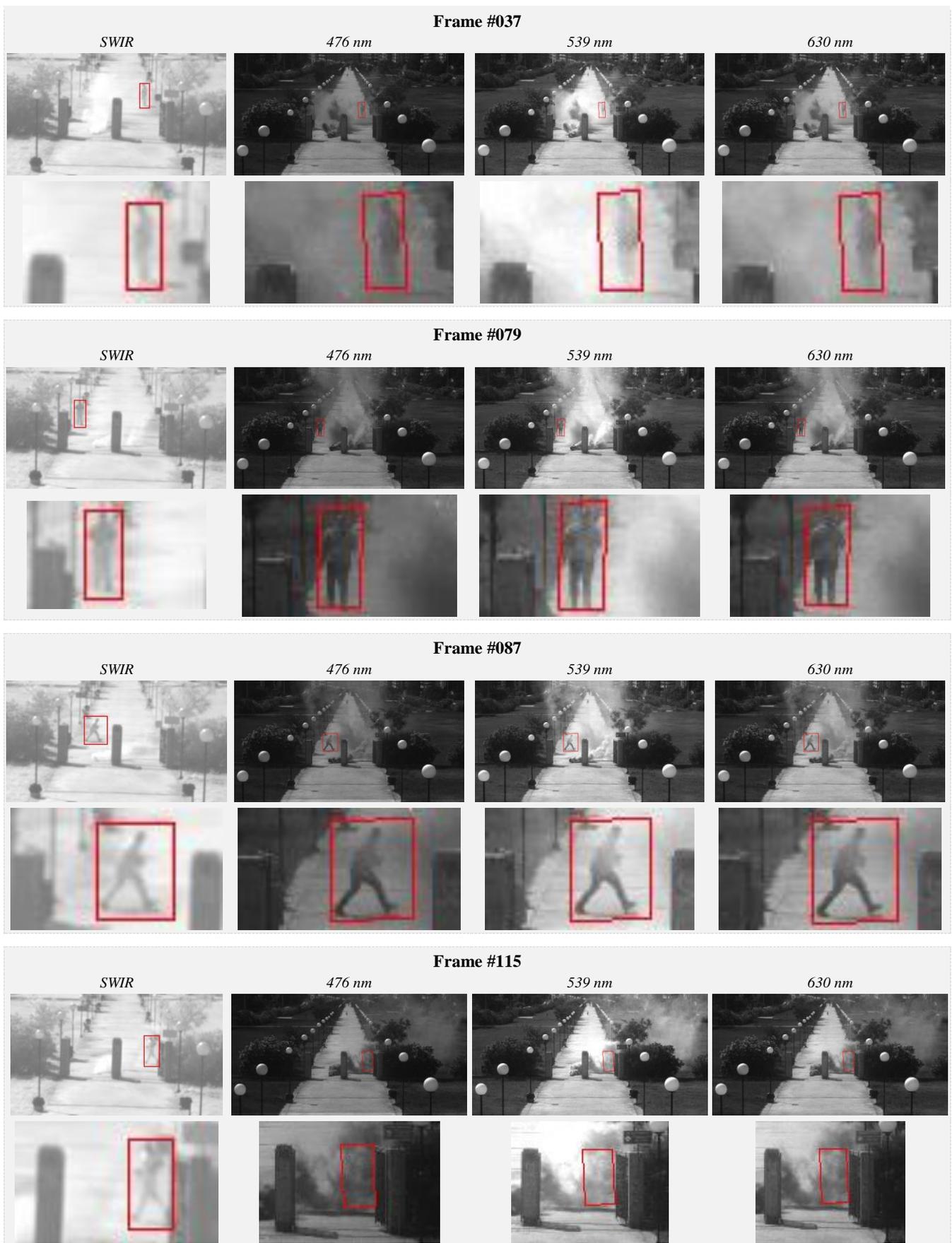


Figure 6: Experimental results after the application of the developed system. The indicative frames number #037, #079, #087 and #115 are shown. For each frame the SWIR and three hyperspectral bands (476, 539, 630 nm) are presented. The detected targets are annotated with a red color onto the SWIR. Their projections are also shown onto the hyperspectral images. Zoom-in views are, also, provided.

In Figure 6 (Frame #037), a relative difficult detection case is presented. The moving target is barely discernible from the background on the SWIR imagery, however the detection works correctly, and its projection on the hyperspectral image

plane is also accurate, as the same part of the background and moving object are represented in both.

In Figure 6 (Frame #079), the detection is working smoothly, with the bounding box containing 100% of the moving target. The projection on the hyperspectral image is also accurate in this case. The target has moved in closer proximity to the spectral system sensors, allowing for slightly increased geolocation accuracy.

The detection and projection modules also seem to perform for the next presented case in Figure 6 (Frame #087). Again, the developed detection and tracking modules managed to extract correctly the location both on the SWIR and hyperspectral modalities. This was a relative different case since the target was running.

Finally, a significant challenging case is presented in Figure 6 (Frame #115). The detected moving target on Frame #115 is slightly visible in the SWIR which possess certain through fog/smoke capabilities. The algorithm managed to detect and successfully continue tracking the target. However, during the fusion process the hyperspectral cue indicated relative high reflectance values for this projected object which was not matching with the actual target. The data association term indicated high confidence levels in the initial SWIR detection and tracking steps and therefore the final decision was positive and correctly verified.

In particular, the data association term is based on spectral statistics (mean, standard deviation, etc) which are computed both for the object (possible moving objects: top left Figure 7) as well as the background. For the spectral statistics, as background is considered the rest of the area that surrounds each possible moving object inside its bounding box. These statistics are captured and calculated at every frame and feed the fusion and data association modules.

4. CONCLUSIONS

In this paper, we propose the use of a multisensory system, which can address a range of critical environmental and illumination conditions like smoke, fog, day and night acquisitions, etc. These conditions have been proved challenging for conventional RGB sensors, or any system based on a single sensor, in general. Multimodality may answers a direct need of the security industry, for round the clock, precise monitoring in any weather or emergency condition.

We have developed the required hardware and software modules in order to perform near real-time video analysis for detecting and tracking moving objects/targets. The software modules and algorithms developed are of low-complexity, to achieve near-real time processing of the multimodal data, and timely provision of events/alerts. These preliminary experimental results demonstrate the capabilities of the proposed system to monitor critical infrastructure in challenging conditions. The system is able to detect possible moving targets as well as to track and recognise them in time and through smoke, fog, etc.

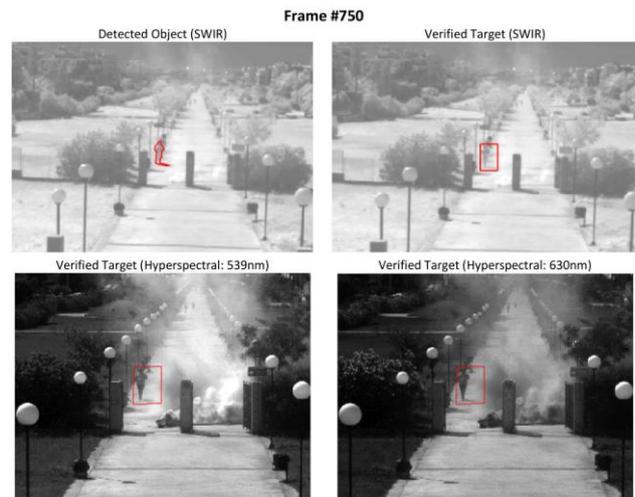


Figure 7: The spectral signature of the possible detected target (top left) from all sensors is stored and analyzed during the fusion and data association module. The same information is also computed regarding the object background which is the rest of the surrounding area inside its bounding box.

5. ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union's FP7 under grant agreement n. 607292, ZONeSEC Project <https://www.zonsec.eu/>

REFERENCES

- Gao, S., Ye, Q., Xing, J., Kuijper, A., Han, Z., Jiao, J. and Ji, X., 2017. Beyond Group: Multiple Person Tracking via Minimal Topology-Energy-Variation. *IEEE Transactions on Image Processing*, 26(12), pp. 5575-5589.
- Fan, C. T., Wang, Y. K. and Huang, C. R., 2017. Heterogeneous Information Fusion and Visualization for a Large-Scale Intelligent Video Surveillance System. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47(4), pp. 593-604.
- Kandylakis, Z., Karantzas, K., Doulamis, A. and Doulamis, N., 2015. Multiple Object Tracking with Background Estimation in Hyperspectral Video Sequences. In: *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*.
- Karantzas, K., 2009. Intrinsic dimensionality estimation and dimensionality reduction through scale space filtering. In: *16th International Conference on Digital Signal Processing, Santorini-Hellas on*, pp. 1-6.
- Karantzas, K., Sotiras, A. and Paragios, N., 2014. Efficient and Automated Multi-Modal Satellite Data Registration through MRFs and Linear Programming. In: *IEEE Computer Vision and Pattern Recognition Workshops on*, pp.1-8.
- Klinger, T., Rottensteiner, F. and Heipke, C., 2017. Probabilistic multi-person localisation and tracking in image

sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 127, pp. 73-88, ISSN 0924-2716.

Makantasis, K., Karantzas, A., Doulamis and N., Doulamis, 2015. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4959-4962.

Manolakis, D., Truslow, E., Pieper, M., Cooley, T. and Brueggeman, M., 2014. Detection Algorithms in Hyperspectral Imaging Systems: An Overview of Practical Algorithms. *IEEE Signal Processing Magazine* 31(1), pp. 24-33.

Milan, A., Leal-Taixé, L., Schindler, K. and Reid, I., 2015. Joint Tracking and Segmentation of Multiple Targets In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Milan A., Schindler, K. and Roth, S., 2016. Multi-Target Tracking by Discrete-Continuous Energy Minimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10), pp 2054-2068.

Liao, W., Yang, C., Ying Yang, M. and Rosenhahn B., 2017. Security Event Recognition for Visual Surveillance, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume IV-1/W1.

Pieper, M., Manolakis, D., Cooley, T., Brueggeman, M., Weisner, A. and Jacobson, J., 2015. New insights and practical considerations in hyperspectral change detection. In: *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) on*, pp. 4161-4164.

Tochon, G., J. Chanussot, J., Dalla Mura, M. and Bertozzi, A. L., 2017. Object Tracking by Hierarchical Decomposition of Hyperspectral Video Sequences: Application to Chemical Gas Plume Tracking, *IEEE Transactions on Geoscience and Remote Sensing*, PP(99), pp.1-19.

Vakalopoulou, M. and Karantzas, K., 2014. Automatic Descriptor-based Co-registration of Frame Hyperspectral Data, *Remote Sensing of Environment*, Vol.6(4).

Wang, X., Fan, B., Chang, S., Wang, Z., Liu, X., Tao, D. and Huang, T.S., 2017. Greedy Batch-Based Minimum-Cost Flows for Tracking Multiple Objects. *IEEE Transactions on Image Processing*, 26(10), pp. 4765-4776.

Yang, M., Wu, Y. and Jia, Y., 2017. A Hybrid Data Association Framework for Robust Online Multi-Object Tracking. *IEEE Transactions on Image Processing*, PP(99), pp. 1-1.

Zhang, X., Xia, G.S., Lu, Q., Shen, W. and Zhang, L., 2017. Visual object tracking by correlation filters and online learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, ISSN 0924-2716.