MULTIPLE OBJECT TRACKING WITH BACKGROUND ESTIMATION IN HYPERSPECTRAL VIDEO SEQUENCES

Z. Kandylakis[1,2], *K. Karantzalos*[2]†, *A. Doulamis*[2], *N. Doulamis*[2]

[1] i-SENSE Research Group, Institute of Communication and Computer Systems (ICCS), Greece
[2] National Technical University of Athens (NTUA), Greece
† karank@central.ntua.gr http://users.ntua.gr/karank/

ABSTRACT

Although, cutting-edge frame hyperspectral sensors can currently acquire hypercubes at video rates with low spatial resolution, they offer enhanced discrimination capabilities for the characterization of subtle spectral features and important object reflectance properties. To this end, a generic framework was designed, developed and validated for multiple object tracking in hyperspectral video sequences. The background estimation was efficiently addressed through advanced scale space filtering and dimensionality reduction. The detection of moving objects was performed on the reduced representation for low computational complexity. The object recognition task was based on certain spectral and geometric features which were associated with a rule-based classification. The experimental results appear promising and indicate the efficiency of the developed approach.

***Index Terms*—** Object detection, recognition, dimensionality reduction, registration, scale space filtering

1. INTRODUCTION

Recent advances in optics, photonics and nanotechnology have enabled the development of frame (snapshot) hyperspectral imaging sensors [1]. In contrast to push-broom (line-scanning) ones, they don't require the constant movement of the camera or the movement of the observed object. Frame hyperspectral sensors are currently *(i)* based on the multiple orders and air gaps of an interferometer (e.g., a Fabry-Perot one) placed in front of a standard RGB/monochromatic sensor [2], [3] or either *(ii)* based on optical filters monolithically integrated on top of CMOS image sensors [1], [4]. Still, certain distortions and vignetting can be observed between the acquired spectral bands of both solutions requiring radiometric and geometric corrections ([2], [3] and Figure 1). Moreover, the spatial and spectral resolution is relative lower compared to push-broom sensors.

Such frame hyperspectral imaging systems, which can acquire hypercubes at video rates, have already been employed for several applications in medical imaging [5], precision agriculture [2], gas plume detection [6] and moving object detection and tracking [7], [8]. In particular, in [7] and [8] the tracking of moving object was addressed through the calculation of object reflectance. Chemical plume detection has also been addressed through a two-stage detection and refinement approach using binary partition trees [6].

Hyperspectral video sequences with high temporal resolution combine the advantages of both video and hyperspectral imagery. Despite the high spatial resolution of modern RGB or monochromatic video cameras, their low spectral resolution limits their ability to classify or identify objects robustly. Multispectral and hyperspectral sensors offer repetitive, consistent and comprehensive datasets with enhanced discrimination capabilities, useful for the characterization of subtle spectral features and important chemical and physical properties of the observed terrain features/objects.

To this end, an automated framework was developed and validated for multiple moving object recognition and tracking in hyperspectral video sequences. The goal was to design a generic approach able to detect moving objects like people and crowds in interior and exterior public areas for any security purposes like surveillance or evacuation. The background estimation was efficiently addressed through advanced scale space image representations and dimensionality reduction techniques. The detection of the moving objects was performed on the reduced dimensionality, while the object recognition task was based on object's spectral and geometric features through a rule-based approach. The experimental results and the performed qualitative and quantitative evaluation indicate the efficiency of the developed approach.

2. METHODOLOGY

A flowchart of the developed algorithm is presented in Figure 2. The different algorithmic modules shown, contribute to processing the initial raw hypercubes and delivering the detected multiple objects at each frame of the video sequence.

2.1. Co-registration of spectral bands

The first step is recovering the geometry of the acquired spectral bands for every frame. This is accomplished through

The research leading to these results has received funding from the European Union's FP7 under grant agreement n.313161, eVACUATE Project www.evacuate.eu

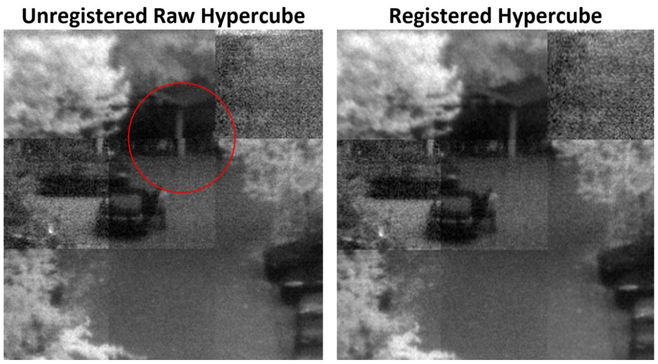

Fig. 1. The spectral bands of the raw hypercube (left, frame *#525*) require radiometric and geometric corrections. The checkerboard visualization demonstrates the effectiveness of the registration procedure.

an automatic descriptor-based registration procedure [3]. In short, the algorithm is based on feature descriptors and the problem is decomposed into three processing steps. The operated wavelength is divided into an appropriate number of spectral groups. Then the co-registration of all the spectral bands of each group is performed. In particular, after sorting the spectral bands in terms of spectral variance and proximity to the key changing spectral regions, the appropriate bands are selected for the co-registration inside and in-between the groups. Under this approach the search space of solutions is effectively narrowed through the selection, in an unsupervised manner, of the spectral bands which lead to the geometric recovery quickly. Even with important distortions and vignetting, large illumination changes, rotation and translation effects, the image descriptors can find enough correspondences and result to an acceptable geometric accuracy.

2.2. Hyperspectral Scale Space Filtering and Dimensionality Reduction

The hypercube $\mathcal{I}$, that the registered spectral bands are forming in every frame, is then simplified through anisotropic morphological levelings [9]. Briefly, by employing advanced nonlinear scale space filtering, noise is removed while the data are simplied efciently. Multiscale levelings can be constructed when the initial (reference) hypercube $\mathcal{I}$ is associated with a series of marker functions $\{h_1, h_2, ..., h_n\}$ -all h are increasingly smoother hypercubes in $\mathcal{R}^3$. The constructed levelings are respectively

$$\begin{aligned} g_1 = \mathcal{I},\ \ g_2 = \Lambda(g_1, h_1),\ \ g_3 = \Lambda(g_2, h_2), \\ g_4 = \Lambda(g_3, h_3)\ , ...,\ g_n = \Lambda(g_{n-1}, h_{n-1}) \end{aligned} \tag{1}$$

A series g_n of simpler and simpler hypercubes, with fewer and fewer smooth zones are produced forming a 4D scale space with $g : \Omega \subset \mathcal{R}^4$ and $g(x, y, z, n) = g_n(x, y, z)$. With these fewer zones, the simplified hypercubes describe, in a more distinct way, the spectral and spatial signatures of the

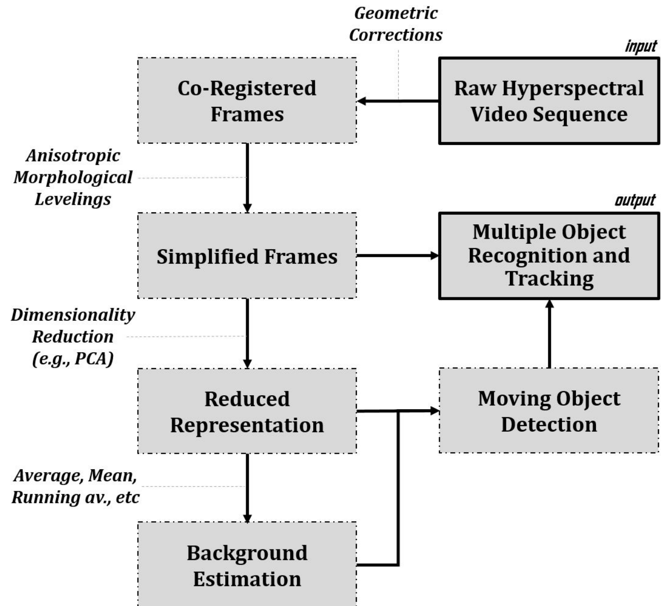

Fig. 2. A flowchart of the developed algorithm for the automated multiple object tracking in hyperspectral video sequences

observations, forming appropriate versions for adequately estimating the intrinsic dimensionality of the registered data.

In addition, the computation of complex lower dimensional manifolds was more efficient, since the variance of the embedded hyperspace was more compact [10]. Having estimated the intrinsic dimensionality of the available datasets, their transformation into a meaningful representation of reduced dimensionality followed. Ideally, the reduced representation has a dimensionality that corresponds to its intrinsic one which is the minimum number of parameters needed to account for the properties of the observations. In all the experiments conducted, the standard Principal Component Analysis was employed for calculating the reduced representation.

2.3. Background Subtraction, Moving Object Detection and Recognition

In order to keep the computation complexity as low as possible for allowing near-real time performance, the background estimation and the initial moving object detection were computed over the first principal component. In all experimental results the background estimation was performed based on the average of 50 frames. The initial moving object detection was based on the background subtraction delivering the initial seeds/blobs which feed the multiple object detection and recognition module.

At every frame the geometric and spectral properties of the detected moving objects were calculated. A hierarchical knowledge-based scheme with specific rules was then applied, for the classification of every detected object to the appropriate class. Certain rules were associated with the calcu-

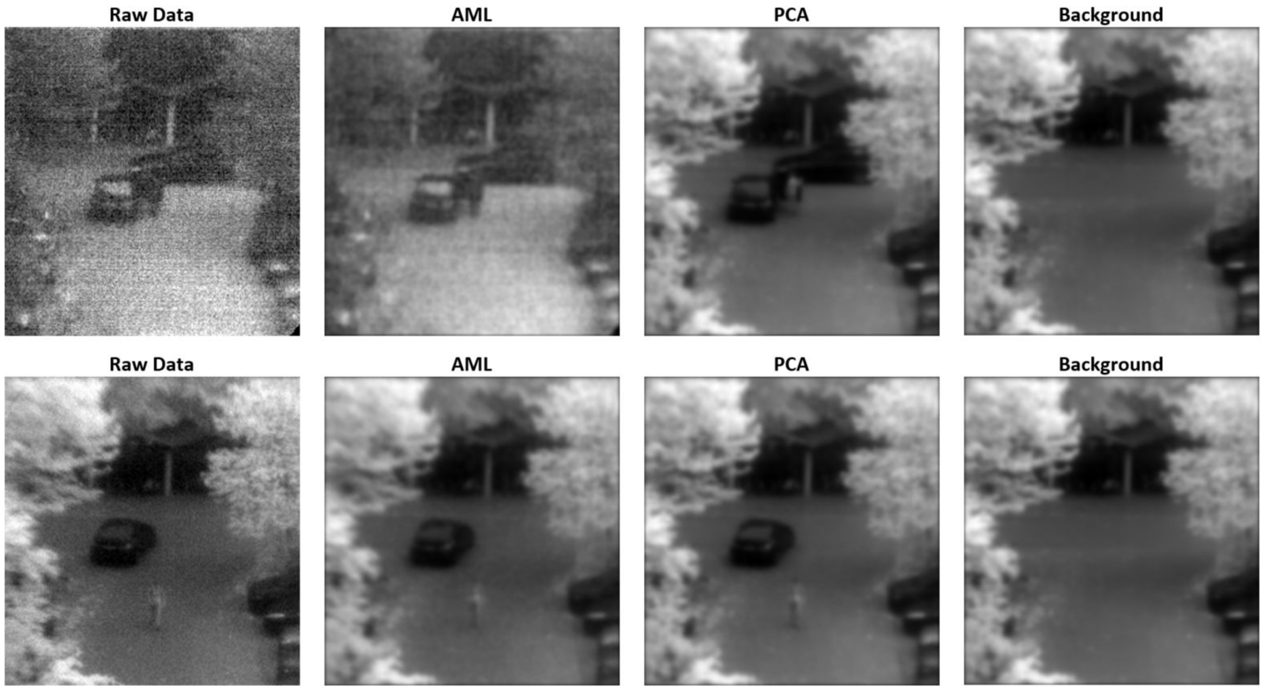

Fig. 3. The raw, anisotropically smoothed (AML), the principal component of the lower dimensional representation (PCA) along with the estimated background are shown for *(i)* Top: frame *#510* at *644nm* and *(ii)* Bottom: frame *#800* at *865nm*.

lated spectral (e.g., band ratios, indices) and geometric (e.g., area, compactness) features. These features were used as the key information for defining the inclusion-or-exclusion parameters during the classification of moving objects.

3. EXPERIMENTAL RESULTS AND VALIDATION

The developed method was applied on video datasets acquired by a frame hyperspectral sensor with integrated tiled filters and optical duplication from imec[1]. The main parts of the imaging system are the tiled filter layout and an optical subsystem which simultaneously duplicates the scene onto each filter tile. Through the use of monolithically integrated optical filters it retains the qualities of compactness, low cost and high acquisition speed. The camera can acquire 32 spectral bands in the spectral range of 600-1000nm at a speed of about 30 cubes per second at daylight conditions up to 340 cubes per second at higher illumination levels as typically used in machine vision applications. The spatial resolution of each spectral band is 256x256 pixels. In its current version the optical duplicator conveys certain distortions and vignetting.

In Figure 1 data from the frame #525 of the *ParkingFills*[2] hyperspectral video dataset are shown. With a checkerboard visualization, containing randomly 9 spectral bands in 9 sub-image blocks, one can observe that scene's geometry has been effectively recovered after the application of the employed co-registration procedure.

In Figure 3 results from the different algorithmic components (Figure 2) are shown for *(i)* Top: frame *#510* at *644nm* and *(ii)* Bottom: *#800* at *865nm*. In particular, the corresponding raw, anisotropically smoothed (AML) and the principal component of the lower dimensional representation (PCA) are shown. One can observed that anisotropic morphological levelings (AML) managed to both denoise and smooth the raw data. Furthermore, the 1st principal component further simplified the data while representing and constrained more from the near infrared bands. It should be, also, noted that in both frames the background has been efficiently estimated and is almost identical across the sequence.

In Figure 4 results after the application of the developed multiple object detection and tracking algorithm are shown. The detected and labelled moving objects are shown for certain frames. Experimental results for the entire *ParkingFills* hyperspectral video sequence can be found here[3]. In frames *162* and *163* the algorithm managed to detect and recognize both the person (colored green) and the moving bike (colored blue). In frames *384* and *385* two people and one car (colored red) were detected and labelled correctly. Despite the low spatial resolutio,n the spectral features computed across the simplified hypercube, resulted to a successful classification. From the frame *500* to *540* two moving cars and one person were detected and labelled correctly. In particular, for frames *520* and *523* the initial moving object detection resulted to a large initial blob and the rule-based classification procedure based on the calculated geometric and spectral features man-

[1] http://www2.imec.be
[2] http://vimeo.com/77218620
[3] http://users.ntua.gr/karank/Demos/TrackHyperVideo.html

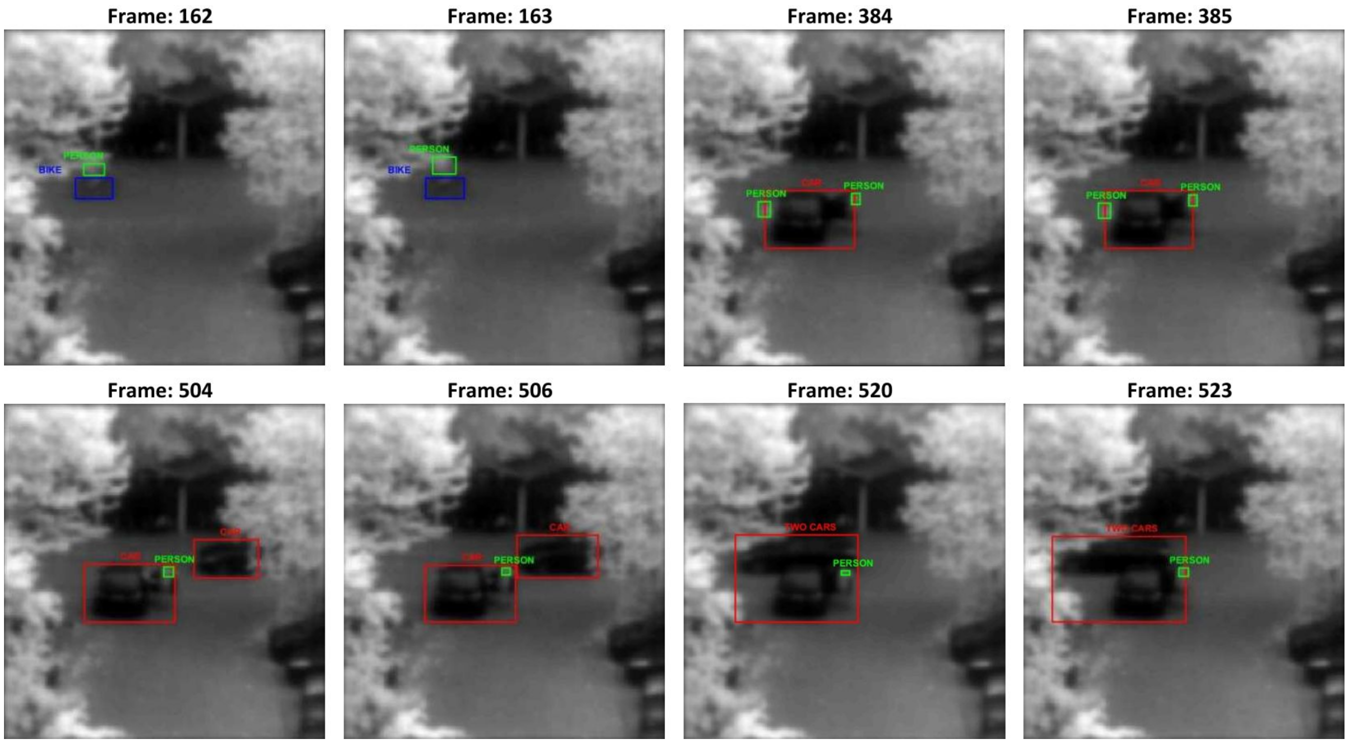

Fig. 4. Experimental results from the *ParkingFills* hyperspectral video dataset. After the application of the developed methodology the detected multiple objects have been classified and labelled correctly.

aged to label it correctly.

4. CONCLUSION

A generic framework for multiple object detection and tracking in hyperspectral video sequences was developed and validated. The background estimation was efficiently addressed through advanced scale space image representations and dimensionality reduction. The detection of the moving objects was performed on the reduced dimensionality, while the object recognition task was accomplished based on geometric and spectral features through a rule-based classification. The experimental results appear promising and indicate the eficiency of the developed approach.

5. REFERENCES

[1] Nathan Hagen and Michael W. Kudenov, “Review of snapshot spectral imaging technologies,” *Optical Engineering*, vol. 52, no. 9, pp. 090901–090901, 2013.

[2] E. Honkavaara, H. Saari, J. Kaivosoja, I. Polonen, T. Hakala, P. Litkey, J. Makynen, and L. Pesonen, “Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight uav spectral camera for precision agriculture,” *Remote Sensing*, vol. 5, no. 10, pp. 5006–5039, 2013.

[3] Maria Vakalopoulou and Konstantinos Karantzalos, “Automatic Descriptor-Based Co-Registration of Frame Hyperspectral Data,” *Remote Sensing*, vol. 6, no. 4, pp. 3409–3426, 2014.

[4] Bert Geelen, Murali Jayapala, Nicolaas Tack, and Andy Lambrechts, “Low-complexity image processing for a high-throughput low-latency snapshot multispectral imager with integrated tiled filters,” in *Proc. SPIE 8743, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX, 87431E*, 2013, vol. 8743.

[5] T. Arnold, M. de Biasio, and R. Leitner, “Hyperspectral video endoscopy system for intra-surgery tissue classification,” in *Sensing Technology (ICST), 2013 Seventh International Conference on*, Dec 2013, pp. 145–150.

[6] Guillaume Tochon, Jocelyn Chanussot, Jérôme Gilles, Mauro Dalla Mura, Jen-Mei Chang, and Andrea Bertozzi, “Gas Plume Detection and Tracking in Hyperspectral Video Sequences using Binary Partition Trees,” in *IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014)*, Lausanne, Switzerland, 2014.

[7] A. Banerjee, P. Burlina, and J. Broadwater, “Hyperspectral video for illumination-invariant tracking,” in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS ’09. First Workshop on*, Aug 2009, pp. 1–4.

[8] Hien Van Nguyen, A. Banerjee, and R. Chellappa, “Tracking via object reflectance using a hyperspectral video camera,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, June 2010, pp. 44–51.

[9] K. Karantzalos, “A 4D Morphological Scale Space Representation for Hyperspectral Imagery,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, pp. 127–132, 2008, ISSN 1682-1750.

[10] K. Karantzalos, “Intrinsic dimensionality estimation and dimensionality reduction through scale space filtering,” in *IEEE International Conference on Digital Signal Processing*, July 2009, pp. 1–6.