

A methodology for near real-time change detection between Unmanned Aerial Vehicle and wide area satellite images

Anastasios L. Fytalis ^{a,*}, Anthony Prokos ^a, Konstantinos D. Koutroumbas ^a, Dimitrios Michail ^b, Charalambos C. Kontoes ^a

^a Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Metaxa and Vas. Pavlou Str., 15236 Athens, Greece
^b Department of Informatics and Telematics, Harokopio University of Athens, 70 El. Venizelou Av., 17671 Athens, Greece

ARTICLE INFO

Article history:

Received 11 November 2015
 Received in revised form 31 May 2016
 Accepted 2 June 2016
 Available online 18 June 2016

Keywords:

Unsupervised change detection
 Very High Resolution images
 Unmanned Aerial Vehicle imagery
 Extended descriptor
 Seeded region growing
 Misregistration

ABSTRACT

In this paper a novel integrated hybrid methodology for unsupervised change detection between Unmanned Aerial Vehicle (UAV) and satellite images, which can be utilized in various fields like security applications (e.g. border surveillance) and damage assessment, is proposed. This is a challenging problem mainly due to the difference in geographic coverage and the spatial resolution of the two images, as well as to the acquisition modes which lead to misregistration errors. The methodology consists of the following steps: (a) pre-processing, where the part of the satellite image that corresponds to the UAV image is determined and the UAV image is ortho-rectified using information provided by a Digital Terrain Model, (b) the detection of potential changes, which is based exclusively on intensity and image gradient information, (c) the generation of the region map, where homogeneous regions are produced by the previous potential changes via a seeded region growing algorithm and placed on the region map, and (d) the evaluation of the above regions, in order to characterize them as true changes or not. The methodology has been applied on demanding real datasets with very encouraging results. Finally, its robustness to the misregistration errors is assessed via extensive experimentation.

© 2016 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

Change detection (CD) is an important research problem which finds application in various remote sensing areas, including land use or land cover change (Byrne et al., 1980; Deng et al., 2008; Kontoes, 2008; Lunetta et al., 2006), forest monitoring (Collins and Woodcock, 1996; Desclee et al., 2006), damage assessment (Al-Khudhairi et al., 2005) and surveillance applications (Carlotto, 1997). The main purpose of a CD method is to produce a “change map”, i.e. a binary image that provides “change” or “no change” information between two (bi-temporal CD) or more images (multi-temporal CD). In some cases, additional information which indicates the kind of change, called “from-to” information, may also be provided (Lu et al., 2004).

The main challenge in CD stems from the fact that, as the sensed images are acquired at different times and, in some cases, by different sensors the images cannot exactly match each other; the larger the mismatch, the less accurate the change detection results are

expected to be. A pre-processing stage which includes (a) radiometric correction and (b) image registration in order to correct these differences is required (Coppin et al., 2004). Radiometric correction diminishes intensity variations between images, which are encountered due to different camera type, image acquisition conditions (height, angle of acquisition) and environmental factors (time in the day, illumination conditions, seasonal solar angle variations and atmospheric effects) (Singh, 1989). In many cases, the techniques used are simple, such as intensity normalization or linear transformations of intensity (Radke et al., 2005), while in multispectral data more advanced methods can also be applied (Song et al., 2001). Image registration is required for projecting the images on a common reference system. It is performed by extracting features from both images, matching them to acquire ground control points (GCP) and estimating a transformation model that is applied to one of the two images, while the other is used as reference (Wong and Clausi, 2007). In addition to these steps, images taken from mountainous areas also need to be resampled. In general, an average registration accuracy of 0.2–0.5 pixels (root mean square error) is considered to be acceptable (Townshend et al., 1992). However, misregistration errors may affect heavily the

* Corresponding author.

E-mail address: a.fytis@noa.gr (A.L. Fytalis).

accuracy of change detection as false changes are introduced, sometimes at the expense of detecting some true changes (Bovolo et al., 2009; Chen et al., 2014; Dai and Khorram, 1998; Townshend et al., 1992; Wang and Ellis, 2005). To avoid such a case, they should be identified and masked out (Bruzzone and Cossu, 2003; Bruzzone and Serpico, 1997; Gong et al., 1992; Marchesi et al., 2010; Shi et al., 2011; Stow, 1999).

Various CD methods have been proposed in the literature; selecting the most suitable method depends on the kind of application and on the properties of the available remote sensing data, namely spectral and ground resolution. CD methods are divided into two categories, pixel-based and object-based methods (Hussain et al., 2013). Pixel-based methods consider a single image pixel as the unit for analysis and they are the traditional methods for detecting changes. A difference image is generated between the images by performing operations on their associated pixels and a suitable threshold converts it into change map (Bruzzone and Fernandez, 2000; Rosin and Ioannidis, 2003; Singh, 1989). However, pixel-based methods exhibit some drawbacks. First, they are sensitive to "salt and pepper" noise. Furthermore, the ongoing progress in multispectral imagery has enabled the acquisition of Very High Resolution (VHR) images with rich contextual information which is not taken into account per se by the usual pixel-based methods. In addition, misregistration errors have a greater impact in fine ground resolution (Bovolo et al., 2009; Chen et al., 2014; Wang and Ellis, 2005), further supporting that pixel-based methods are unsuitable for change detection when VHR images are used. On the other hand, object-based image analysis (OBIA) methods, take advantage of the contextual information as they group neighboring pixels using spectral, textural and edge features in order to segment the images under study into objects, which form the structural units for the comparison between them. These methods are becoming more popular in remote sensing (Blaschke, 2010) and they are effectively used for CD (Desclée et al., 2006; Im et al., 2008). Hybrid CD methods, which combine image segmentation with pixel-based processing methods so that they can benefit the most from the advantages that they offer (Hussain et al., 2013), are also reported (Al-Khudairy et al., 2005). For a comparison of some of the most used CD methods, the reader is referred to Desclée et al. (2006) and Mas (1999).

1.1. Scope of the study

Utilizing VHR images acquired by Unmanned Aerial Vehicles (UAVs) for change detection without human inspection is a quite new trend. UAVs are suitable for surveillance applications as they bring many advantages, such as cost-effectiveness and quicker deployment with reduced risk compared to manned flight operations (Rango et al., 2006). In the context of surveillance applications for crisis management, land use/land cover change, illegal urbanization, border control and ecosystem disturbances, a UAV can be effectively used to acquire images of a target region in need in rush mode, meaning that a new flight plan might be designed for a new operation and that a report needs to be produced as fast as possible. Such an application would be most effective when the UAV images are sent immediately for processing to an operational Ground Control Station via a communication link, i.e. use of a co-ordinated Unmanned Aerial System is made (Colomina and Molina, 2014).

The images acquired should be compared to a reference of historical data. However, a dataset of UAV images that provides historical data is rarely available. A way out of this problem is to use satellite images (of high resolution) as reference since they are available periodically for almost all of the regions of interest in earth (Fan et al., 2010). Then the acquired UAV images can be compared to them, having almost no restriction of the spot they

are taken from. The combination of UAV and satellite images for a more complete set of data has been studied for agriculture applications (Gevaert et al., 2015). To make this CD concept possible, the following issues must be addressed; (a) the placement of the UAV image on the correct location onto the satellite image plane, i.e. the georeference of the UAV image, (b) the resulting misregistration errors and distortion effects that are evident due to the integration of images of different ground resolution, (c) the possible absence of a detailed surface model of the reference (d) the need for unsupervised CD due to the lack of ground truth of the changes of interest and (e) the need for fast processing so that any reports can be produced short after the acquisition of a set of UAV images. The latter constitutes a requirement for near real time processing that also limits the time to be spent on image pre-processing. Thus there is the need of trading off the quality of photogrammetric reconstruction of the images in return for fast overall processing.

In this paper, a methodology for detecting changes between two images taken of the same area from different kind of optical sensors, namely a UAV camera and a VHR satellite sensor, is proposed. To the best of our knowledge, this is the first time that a methodology that detects changes between low altitude VHR UAV images and a VHR satellite image that is used as a fixed reference of historical data is reported in the bibliography. The key aspect of the proposed methodology is that it consists of several processing stages, performed one after the other, which use appropriately tailored computer vision techniques with aim to diminish the effects encountered during CD. Its novelty lies mainly in the way that the combination of these stages is designed so that it brings a near real time solution to the specific CD problem as a whole in a fully integrated and unsupervised way. The methodology has been tested in challenging datasets of images with very encouraging results. Its robustness to misregistration errors is also assessed by simulating misregistration errors. This methodology has been developed to work with RGB images, but it can also be extended to multispectral images. In the latter case, one could take advantage of the richness in spectral features and add or modify some processing steps.

Few works on CD using UAV images are found in literature, most of which adopt an object-based approach for defining objects in order to detect and evaluate changes. In Coulter et al. (2011), a complete surveillance system using UAV imagery is presented. Image differencing is used with aim to achieve near real-time CD, however, misregistration errors and distortion effects are not taken into consideration. Shi et al. (2011) combine a change map with image segmentation as a refinement stage, where changes that have occurred in the strict neighborhood of matched GCPs are considered as misregistration errors and they are rejected. Qin (2014) presents a method that segments two UAV images into primitives, by using textural and geometric features, and performs object-based CD on them. Dense point clouds are extracted from Semi-Global Matching and are used to generate a Digital Surface Model, which provides the geometric features. This method is based on height information and is effectively used in urban areas. Wang et al. (2013) also perform image segmentation, a Support Vector Machine classifies the segments to known classes and finally CD is performed on the result of the classification. All of the above methods detect changes on pairs of UAV images that are taken from the same spot. It is implied (if not stated) that the images have been acquired by the same sensor and from the same height, which results in images of the same ground resolution. This is not the case for the CD problem that is examined in our paper, which becomes more challenging due to the different acquisition conditions, but this does not impose any restriction to the spot where the images are acquired to ensure sufficient image overlap. Finally, Chen et al. (2015) acquire a large number of UAV images of an urban area in different dates, create depth maps via 3D point clouds and subtract them for building change detection.

This paper is organized as follows. In Section 2, the proposed change detection methodology is presented. In Section 3, a CD experimental setup is described whose aim is (a) to assess the performance of the proposed methodology in determining true changes and (b) to evaluate the robustness of the proposed CD methodology to misregistration errors. The results of these experiments are presented and discussed in Section 4. Finally, concluding remarks are included in Section 5.

2. Proposed change detection methodology

The proposed change detection methodology consists of four major processing stages, which are summarized in Fig. 1 while the most common notations that are found throughout this paper are listed in Table 1. First of all, the UAV image is pre-processed so that to be in a form suitable for comparison with the satellite image; at this point, the part of the satellite image that represents the same area as the UAV image is automatically determined and the UAV image is ortho-rectified. A Digital Terrain Model (DTM) that has a lower ground resolution than the VHR optical images is used for the ortho-rectification, so that to represent a case where a detailed surface model of the area under study is not available. Then, a difference image between them is properly generated in a way that it remains almost unaffected by pixel misregistration problems encountered during the ortho-rectification of the UAV image; the difference image is thresholded and *potential change components* (PCC) are extracted. Independently of this process, the UAV image is also segmented into two classes, the “bare ground” and the “high saturation” class and then, a seeded region growing (SRG) algorithm based on (a) spectral features, as well as on (b) the segmentation result of the UAV image, is used to define *homogeneous regions* (Reg^{UAV}) around the locations indicated by potential changes. Finally, each homogeneous region is evaluated based on criteria concerning (a) the area of change and (b) template matching on the satellite image and it is approved as *true change* or rejected. Thus, it is considered to be a hybrid CD methodology, as it combines pixel-based image processing and thresholding with image segmentation and object extraction techniques.

In the sequel we describe in detail the processing stages involved in the proposed method. For the ease of presentation we name them as follows: (a) image pre-processing, (b) extraction of potential change components, (c) region map generation, (d) evaluation of regions.

2.1. Image pre-processing

2.1.1. Image registration

The goal of this procedure step is to generate an ortho-rectified UAV image (Im^{UAV}) scaled down to the resolution of the satellite image in order to superimpose it on the satellite image and detect

Table 1
The definition of the most common notations.

| Name | Description |
|------------------------|---|
| Im^{UAV} | The ortho-rectified UAV image that is compared to the reference |
| Im^{sat}, Im^{DTM} | The cropped images of the reference (satellite image and DTM respectively) that correspond to the area depicted in the UAV image |
| $slope^{mask}$ | The binary mask for excluding areas of large slope gradients (derived from Im^{DTM}) |
| D, D_{th} | The difference image between Im^{UAV} and Im^{sat} , and the resulting binary image after thresholding, respectively |
| W_{ij} | The search window centered on a pixel of Im^{sat} , i.e. $Im^{sat}(i,j)$ |
| PCC_k | A connected component that is extracted from D_{th} |
| Reg^{UAV} | A homogeneous region that is defined by the SRG algorithm |
| Reg_{id} | A homogeneous region that has been registered on the Region Map |
| Seg^{UAV}, Seg^{sat} | Rectangular cropped parts of Im^{UAV} and Im^{sat} for estimating edge and intensity parameters |
| $+n Seg^{UAV}$ | The number of pixels for expanding the cropped part of an image, with reference to the size of a PCC_k or a Reg_{id} that is included |

changes. In the scope of this study it is assumed that only an approximate position of each UAV image is available but its full exterior orientation is not. It is also assumed that the satellite image has been ortho-rectified using Rational Polynomial Coefficients (RPC) and a Digital Terrain Model (DTM). The method for image registration used here is the one that is described in Mendoza-Schrock et al. (2009) and Oh et al. (2011).

Full interior and exterior orientations are necessary for the generation of ortho-rectified images, which can both be calculated via photogrammetric resection. The latter is a known procedure that requires an adequate number of control points. These points are extracted automatically using a SIFT-RANSAC regime, as explained below.

SIFT (Lowe, 2004) keypoints are extracted on both UAV and satellite images. In order to assist correct match detection, a section of the satellite image is cropped given the provided approximate position of the UAV image. Initial SIFT matches are calculated between the UAV and satellite images and are then validated via RANSAC (Fischler and Bolles, 1981). The geometric model enforced by RANSAC is a projective transformation.

The matches detected by SIFT-RANSAC are then used as control points. The point coordinates detected on both UAV and satellite images are used as x, y aerial-image coordinates and X, Y world coordinates respectively, by implementing simple affine transformations. The Z world coordinate for each control point is calculated through interpolation (linear interpolation is used) from the equally spaced elevation values provided by the DTM, since the latter is expected to be of a lower resolution than that of the UAV and satellite images. The interior and exterior orientations are then

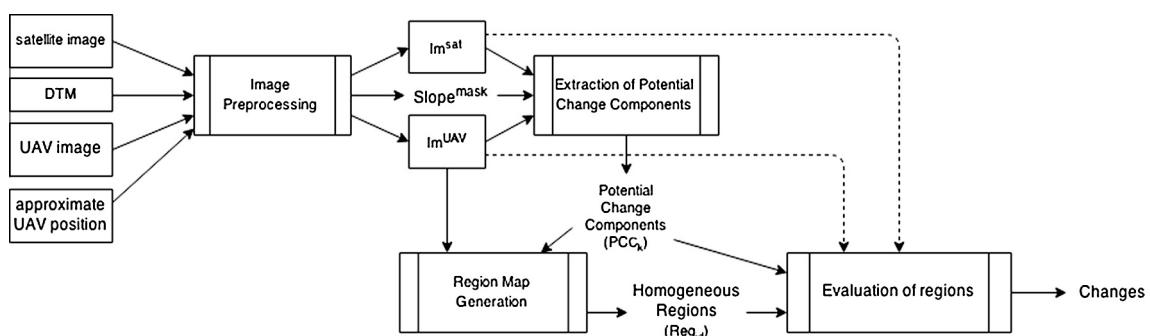


Fig. 1. The processing stages of the change detection methodology. The input, output and temporary products are also shown.

estimated via Least Squares (LS) minimization on a photogrammetric resection model.

This determines the correct position of the initial UAV image in relation to the DTM, thus allowing to superimpose color information from the UAV on the DTM. Generating an ortho-projection of the now colored DTM produces the final ortho-rectified UAV image (Im^{UAV}). This way, the UAV image falls down to the same resolution as the satellite image and the scale difference problem is solved. The last step is to crop the section from the satellite image corresponding to Im^{UAV} as to result in Im^{sat} . These two images, Im^{UAV} and Im^{sat} , are the input to the proposed CD methodology.

The registration accuracy of the ortho-rectified UAV image depends greatly on the accuracy and detail of the DTM since the latter is involved in the calculation of the control point world coordinates, the final texturing and also the ortho-rectification of the satellite image. If the elevation information is not accurate or dense enough, misregistration errors are expected to occur. Such errors are handled as explained below (Section 2.2.2).

2.1.2. Intensity normalization

In order to compensate for radiometric differences between Im^{UAV} and Im^{sat} , pixel intensity values for the channels $C^{UAV} \in \{R^{UAV}, G^{UAV}, B^{UAV}\}$ that constitute Im^{UAV} are normalized according to

$$C'^{UAV} = \frac{\sigma_C^{sat}}{\sigma_C^{UAV}} (C^{UAV} - \mu_C^{UAV}) + \mu_C^{sat} \quad (1)$$

where σ_C^{sat} , σ_C^{UAV} are the standard deviations of the intensity values of C^{sat} and C^{UAV} respectively, and μ_C^{sat} , μ_C^{UAV} their respective mean values.

2.1.3. Steep slope exclusion

As expected, extremely steep slopes were significantly distorted after the ortho-rectification of the UAV image. Consequently, change detection should exclude these areas, since the distorted areas are very likely to be incorrectly returned as changed areas while, in addition, no man-made changes are expected to be found in steep slope areas.

Using the DTM image of the region under study, denoted by Im^{DTM} , the slope of each region in the image can be estimated via slope gradients (large gradients indicate steep slopes). The approach adopted here is to calculate the horizontal and vertical elements of the elevation gradient at each pixel (i,j) using Prewitt operator (Gonzalez et al., 2002) premultiplied by 1/3. This approach smooths the difference values calculated on the interpolated Im^{DEM} image.

Speaking in more mathematical terms, the slope gradient vector $dh_{ij} = [dh_{ij}^x, dh_{ij}^y]^T$ at each pixel is calculated via the following equations

$$dh_{ij}^x = \frac{1}{3} \sum_{k=i-1}^{i+1} (Im_{k,j+1}^{DTM} - Im_{k,j-1}^{DTM}) \quad (\text{horizontal}) \quad (2)$$

$$dh_{ij}^y = \frac{1}{3} \sum_{k=j-1}^{j+1} (Im_{i+1,k}^{DTM} - Im_{i-1,k}^{DTM}) \quad (\text{vertical}) \quad (3)$$

Given that the image resolution is l meters, the two-pixel central difference is $2l$ m. If the slope gradient threshold is set at 100%, a mask that excludes the pixels with steep slope is created as

$$slope_{ij}^{mask} = \begin{cases} 1, & |dh_{ij}| \geqslant 2l \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $|dh_{ij}| = \sqrt{(dh_{ij}^x)^2 + (dh_{ij}^y)^2}$.

2.2. Extraction of potential change components

Having completed the preprocessing stage, we proceed to the first processing stage, which begins with the generation of the image of differences between Im^{UAV} and Im^{sat} . However, before that, an appropriate representation for the image pixels should be decided. Such a representation includes intensity and edge information, as well as local neighborhood information, as is described next.

2.2.1. Descriptors

A vector of extended descriptors is constructed for every pixel in both Im^{UAV} and Im^{sat} images by integrating information from the pixel itself and its 8 nearest neighbors. A similar approach which uses the neighborhood of a pixel is found in Im and Jensen (2005). The resulting vector is 36-element-long and it consists of (see also Fig. 2):

- The RGB values of the pixel itself and its eight-connected neighborhood, that is a total of 27 elements (9 pixels \times 3 intensity values).
- The intensity gradient magnitude values of the pixel's neighborhood (9 elements). These values are stored in intensity gradient images IG^{UAV} and IG^{sat} that are produced from

1. converting Im^{UAV} and Im^{sat} into grayscale luminance images Y^{UAV} and Y^{sat} , respectively, as in Poynton (2003)

$$Y_{ij}^K = 0.299 \cdot R_{ij}^K + 0.587 \cdot G_{ij}^K + 0.114 \cdot B_{ij}^K, \quad K = \{UAV, sat\} \quad (5)$$

2. filtering Y^{UAV} and Y^{sat} with Sobel operator (Sobel and Feldman, 1973) to produce IG^{UAV} and IG^{sat} , respectively
3. scaling the values of each intensity gradient image into [0, 1] range. This step is required in order to have the same range of values for all the components of the descriptor.

2.2.2. Generation of difference image

Although the algorithms used for ortho-rectification are quite accurate, they cannot guarantee total elimination of possible pixel misregistration in the two images. Thus, the homologous of a pixel in Im^{UAV} may be some pixels misplaced in Im^{sat} . Consequently, any pixel-based differencing method is very likely to produce a vast amount of unreliable results. To overcome this misregistration and to compute an accurate *difference image* D , a search is performed for each pixel in Im^{UAV} in order to locate its homologous pixel in Im^{sat} as follows.

Let Im_{ij}^{UAV} be the (i,j) UAV pixel and \underline{u}_{ij} its descriptor vector. Firstly, a search window W_{ij} of size $w \times w$ pixels, which is defined as

$$W_{ij} = \{(x,y) \in \mathbb{N}^2 : x \in [i - \lfloor w/2 \rfloor, i + \lfloor w/2 \rfloor], y \in [j - \lfloor w/2 \rfloor, j + \lfloor w/2 \rfloor]\} \quad (6)$$

where w is an odd integer and operator $\lfloor \cdot \rfloor$ denotes floor function, is centered at the (i,j) pixel of Im^{sat} , defining a square region of pixels on Im^{sat} that are possibly homologous to the Im_{ij}^{UAV} pixel. Then, the Euclidean distances between \underline{u}_{ij} and each one of the respective descriptors $\underline{u}_{x,y}$ of the pixels within W_{ij} in Im^{sat} are calculated. The position of the minimum of these distances indicates the homologous to Im_{ij}^{UAV} pixel in Im^{sat} (see Fig. 3), while this minimum distance is also registered as the difference between the two images at pixel (i,j) , i.e.,

$$D_{ij} = \min_{x,y \in W_{ij}} \|\underline{u}_{ij} - \underline{u}_{x,y}\|_2 \quad (7)$$

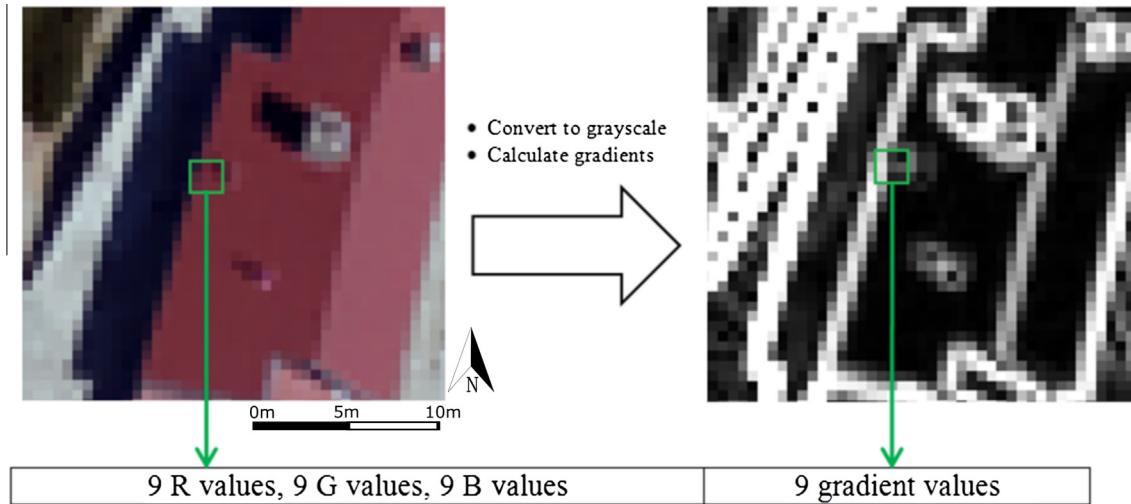


Fig. 2. Constructing the descriptor vector for a pixel.

where the operator $\|\cdot\|_2$ denotes Euclidean norm. For the sake of brevity, we will refer to values in the difference image D as *difference values*. A similar approach was successfully used and validated in [Kontoes \(2008\)](#), while a search window was also used in [Tian et al. \(2014\)](#) for producing a difference image between two Digital Surface Models in a robust way.

It is obvious that, if there is no homologous pixel within the search window, which may be due to a change at this region, the minimum distance refers to its most similar pixel. Clearly, large difference values can be used as indication of a change at that pixel. The parameter w should be adjusted to reflect the maximum misregistration error that is expected. A large window may lead to wrong matches whereas a small one restricts the search into a small area which may not contain the true homologous pixels, producing higher difference values.

2.2.3. Potential change components

The next step is to focus on high difference values in order to extract potential change components, which will be further processed and evaluated as *true changes* or no changes. *Potential changes* are defined as large compact regions in D where the pixels have high difference values. They are identified as described next.

First, D is thresholded using the non-parametric Rosin's method ([Rosin, 2001](#)), resulting in the binary image D_{th} . Then, connected components are extracted from D_{th} using 4 nearest neighbor connectivity. Connected components smaller than N_1 pixels are ignored; the rest are denoted as potential changes. The parameter N_1 is set so as to reflect the size of the smallest change component of interest, depending on the resolution of the image. For example, the N_1 value of 20 pixels on the VHR UAV image with spatial resolution of 0.5 m on the ground corresponds to the average car size. From here on, potential change components will be denoted as PCC's (the 's' notation will be used hereby as the plural number of the mentioned components).

2.3. Region map generation

Potential changes are raw clusters of pixels where notable differences have occurred. It is essential that these changes undergo processing in an object oriented way so that the natural objects that cause these differences are properly defined.

This processing is performed with the use of the seeded region growing (SRG) technique ([Gonzalez et al., 2002](#)). The PCC's designate specific spots in the image that need to undergo processing

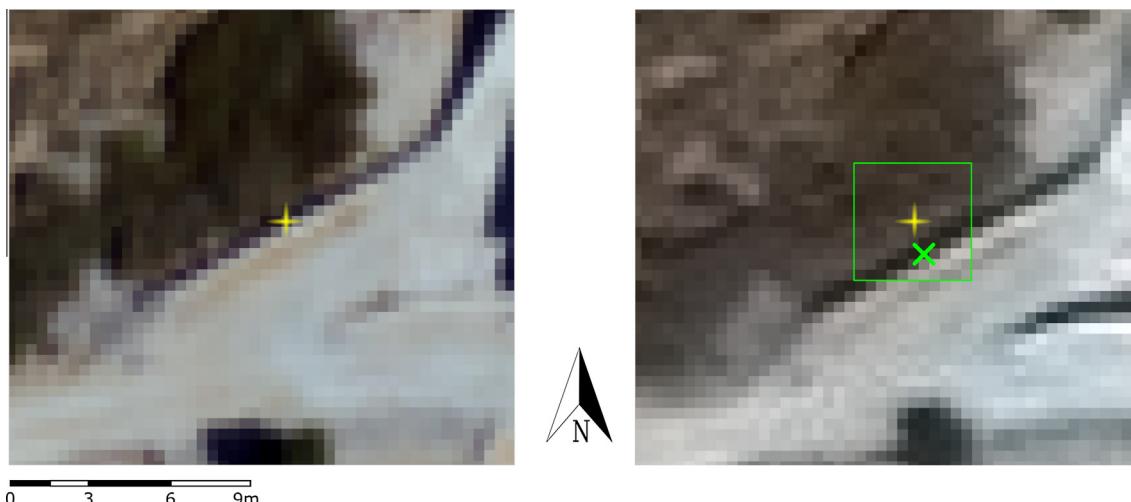


Fig. 3. Searching for homologous pixels (the \times mark) in Im^{sat} (right) for each pixel in Im^{UAV} (left) (the cross mark) using a window w_{ij} , the size of $w \times w = 11 \times 11$ pixels. A misregistration error of $\sqrt{10}$ pixels is easily noticeable for this pixel.

and a SRG algorithm is adopted to define *homogeneous regions*, denoted as Reg^{UAV} 's. Homogeneity is defined in terms of the existence of (a) vegetation, shadow components and colorful objects (high saturation class) or (b) ground surface with no vegetation (bare ground class) (Section 2.3.1). The Reg^{UAV} 's are then registered in a region map which keeps track of the defined regions (Section 2.3.2).

2.3.1. Image segmentation

This step is applied to the whole image, irrespective of the extracted PCC's. To this end, the Im^{UAV} image is transformed into the HSI space (Gonzalez et al., 2002), where each pixel is characterized by the attributes of Hue, Saturation and Intensity. The first two carry color information; hue defines pure color identity on a $[0^\circ, 360^\circ]$ range; saturation is the rate of the white light component blending with pure color. The intensity in each pixel is the received power of the visible light, it is calculated as the average of RGB components and, in fact, it is a grayscale version of the color image.

Based on these attributes, the UAV image is segmented into the following two classes:

- the “high saturation” class, which consists of vegetation, shadow components and colorful objects, and
- the “bare ground” class, which includes the ground surface that receives direct sunlight and where no vegetation is grown.

Image segmentation is carried out as follows. First, the Normalized Difference Saturation Index for each pixel is calculated as

$$NDI_{ij} = \frac{S_{ij} - I_{ij}}{S_{ij} + I_{ij}}, \quad (8)$$

where S_{ij} and I_{ij} are the pixel saturation and intensity values respectively, according to the HSI color space. The more saturated the color or the lower its intensity, the larger the ratio. This index was originally proposed in Singh et al. (2012) for shadow detection since shadow components emit light radiation of very low intensity that is also saturated with blue wavelength. However, we observed that NDI value was relatively high for vegetation pixels as well, since they are highly saturated with green (and maybe red) color and their intensity is relatively low. Then, a suitable threshold t for discriminating the two classes of interest is obtained by performing the Rosin's method (Rosin, 2001) on NDI . Thus, the generated binary image J that identifies the discrimination is defined as

$$J_{ij} = \begin{cases} 1, & NDI_{ij} \geq t \text{ (highly saturated class)} \\ 0, & \text{otherwise (bare ground class)} \end{cases} \quad (9)$$

The result of the image segmentation step is shown in Fig. 4.

Since the PCC's extraction and image segmentation steps are performed separately, a PCC may include pixels that belong to both “bare ground” and “high saturation” classes, as shown in Fig. 5. In such a case, a PCC is further divided into several *homogeneous compact components* PCC_k , $k = 1, 2, \dots$, first by class, and then by connectivity (i.e. by neighboring pixels), so that each PCC_k contains neighboring pixels that belong to *only* one of the two classes. Then, the SRG algorithm is called for each PCC_k instead.

2.3.2. Filling region map

The region map is initially a blank image, of the size of Im^{UAV} and is built in a sequential manner. Specifically, as PCC_k 's are extracted sequentially, one after the other, one or more Reg^{UAV} 's are defined via the SRG algorithm (see Section 2.3.3). Once a Reg^{UAV} region is defined, it must be registered on the region map, by acquiring a unique identification number (id) and marking this

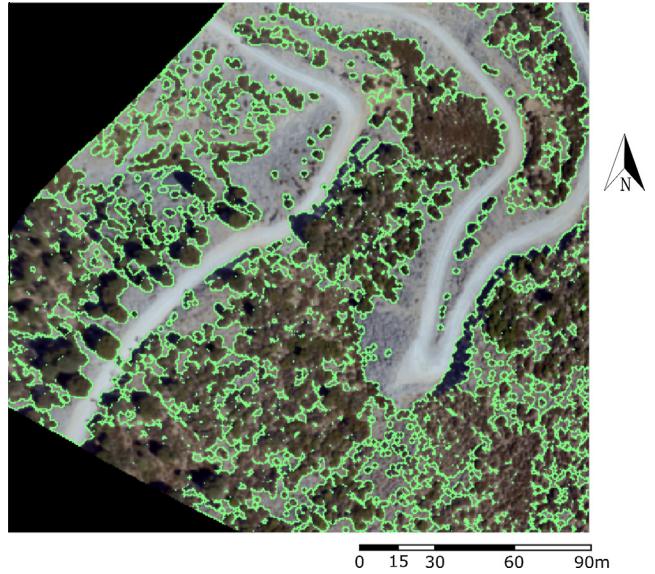


Fig. 4. The image segmentation result. The contour encircles the “high saturation” class while the rest belongs to “bare ground” class.

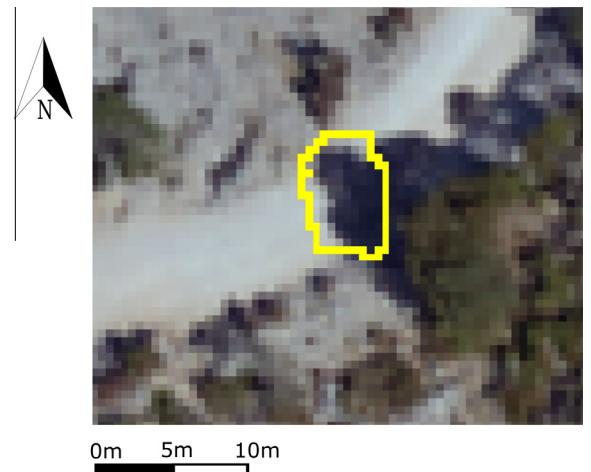


Fig. 5. The contour encircles a potential change component (PCC) that contains pixels from both classes, i.e., a large shadow component (“high saturation” class) and two small groups of neighboring road pixels (one on the top and one on the left part of the PCC – “bare ground” class). The PCC is split into three PCC_k components; then, a SRG is applied for each one of them separately.

id in the positions that correspond to the pixels that belong to Reg^{UAV} . The procedure for each PCC_k is as follows.

1. For a given PCC_k the SRG algorithm is used to define the homogeneous region (Reg^{UAV}) on Im^{UAV} . It is noted that each Reg^{UAV} is grown based only on information found in Im^{UAV} , regardless of the thresholded difference image D_{th} . As a matter of fact, pixels that belong to Reg^{UAV} may or may not belong to any PCC_k component.
2. The new region Reg^{UAV} must be registered on the region map. As the regions are grown separately, it is possible that there is an intersection between Reg^{UAV} and one or more regions that have already been registered on the region map (denoted by Reg_{id}). In this case, Reg^{UAV} may be merged with one or more of them, if they are similar with respect to a similarity criterion. If not, a screening procedure is performed. On the screening procedure,

every pixel (i,j) that belongs to both regions must be assigned to the most appropriate one. The decision is based on minimum Euclidean distance between Im_{ij}^{UAV} and the mean intensity value of each region. This action constitutes a refinement procedure, before a new Reg^{UAV} can be registered on the region map. The way the Reg^{UAV} is registered in the region map is described in full detail in [Appendix A](#). An illustrative example of the result of this procedure is shown in [Fig. 6](#).

3. Once the Reg^{UAV} has finally been registered on region map, the pixels that belong to the region are removed from PCC_k , i.e., $PCC_k = PCC_k - Reg^{UAV}$. Pixels (i,j) that have intensity gradient values higher than a specified value, i.e. $IG_{ij}^{UAV} > m \equiv p_1 \cdot m_{max}$, are also removed from PCC_k , as they contribute little to object representation and it is highly probable that they will not be included in any region. In the previous inequality:

- m_{max} is the maximum gradient value, which is computed in a region around PCC_k , denoted by Seg^{UAV} , and is defined as follows: let Box_k the bounding box that corresponds to PCC_k , of size $w_k \times h_k$. Seg^{UAV} is a rectangular frame of size $(w_k + 2 \cdot n_1) \times (h_k + 2 \cdot n_1)$ in which Box_k is centered, as shown in [Fig. 7](#). The parameter n_1 is set properly to define a rectangular region large enough to capture the edge information in a small area surrounding PCC_k . From here on, such segments will be denoted as $+n$ segments.
- $p_1 \in (0, 1)$.

In other words, m is set to a percentage p_1 of the maximum gradient value m_{max} that appears within the area around PCC_k . This way, m is not predefined, but it is adapted properly according to the local sharpness of the image, although it is still dependent on the user-defined parameter p_1 .

Finally, steps 1 and 2 are repeated until PCC_k contains no clusters of pixels of significant size (less than N_1' pixels).

2.3.3. Seeded region growing algorithm

At this point we describe the SRG algorithm, which was developed for defining a single homogeneous region Reg^{UAV} stemming from a given PCC_k . As it has been mentioned previously, the resulting region may also contain pixels that do not lie in any of the PCC_k components.

At first, the region consists of a single pixel, the “seed”. A proper selection for the seed is the centroid of a potential change component PCC_k . Then, its neighboring pixels are checked for being included in the region with respect to one or more homogeneity criteria. Those of them that satisfy these criteria are added to the region, leading to the growing of the region. On the other hand, those of them that do not satisfy these criteria are marked as

contour pixels and are not checked again. The procedure is applied iteratively and terminates when no more pixels satisfy the established criteria. At that point, the region is encircled by the contour pixels, which are given a second chance to be included in the region. This will help in defining the region in better detail. An example of the growing of a homogeneous region Reg^{UAV} is shown in [Fig. 8](#).

The adopted homogeneity criteria are:

- *The region class*, as defined from the image segmentation step. A Reg^{UAV} must grow with pixels that belong to the same class. It should be noted that both shadow components, trees and colorful objects belong to the “high saturation” class, so a region seeded in a shadow component may grow into an object-plus-shadow component.
- *The mean color intensity value*. A Reg^{UAV} must be homogeneous with respect to color intensity; that is, pixels belonging to Reg^{UAV} should not deviate much from the region’s mean intensity value. This way, for each pixel neighboring to the Reg^{UAV} as it has been formed up to now the Euclidean distance is calculated between its color intensity value and Reg^{UAV} mean color intensity value. If this distance is lower than a predefined threshold, δ_1 for a “bare ground” region and δ'_1 for a “high saturation” region, the pixel is included in the region.

Besides these criteria, the area of the region is also taken into account. It is possible that PCC_k is a part of a very large natural component, such as a road. For the sake of processing speed and interpretation simplicity, it may not be desirable to grow the whole region, but, instead, to break the iteration process when the size of the Reg^{UAV} becomes very large. For this reason, a maximum area threshold is also set.

As it was mentioned before, when the algorithm terminates, the contour pixels are given a second chance to be included in the area. The intensity gradient homogeneity criterion is used to stop the growing of Reg^{UAV} when approaching edges. However, large intensity gradient values are found in both sides of an edge. This is visible in [Fig. 7](#). This way, an extra homogeneity criterion is checked in order to include pixels which belong to an object and are located at the inner side of the edge. The mean μ^C and standard deviation σ^C for each channel $C \in \{R, G, B\}$ are calculated from Reg^{UAV} pixels. For every contour pixel, whose intensity values are represented by vector $\underline{p} = [p^R, p^G, p^B]^T$, if its difference from the region mean value is less than two standard deviations, i.e. $|p^C - \mu^C| < 2 \cdot \sigma^C$ in at least two of the $\{R, G, B\}$ channels, the pixel is included in Reg^{UAV} .

Finally, due to coarse texture of “bare ground” regions, a morphological closing is performed on Reg^{UAV} for filling small holes.



Fig. 6. The result of the merging procedure of the new region Reg^{UAV} and the registered regions Reg_1 , Reg_3 and Reg_4 .

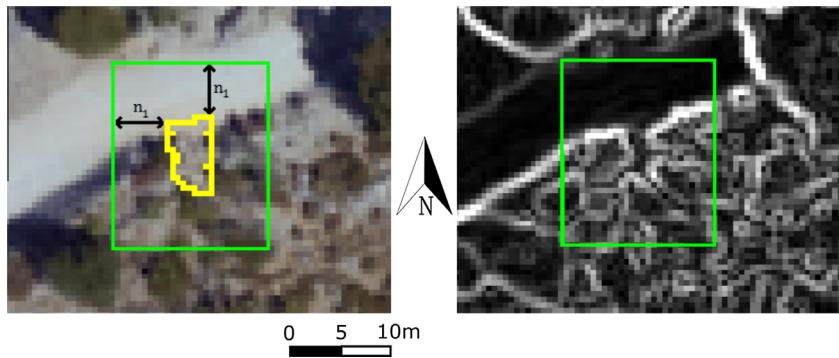


Fig. 7. Defining a $+n_1 = 10$ pixels segment Seg^{UAV} for gradient threshold generation.

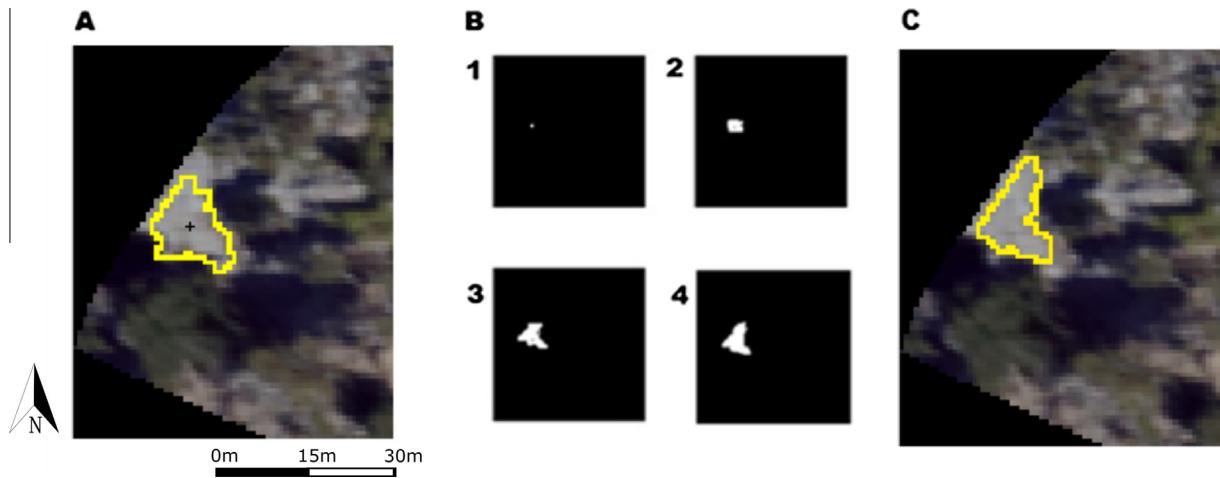


Fig. 8. The phases of seeded region growing. In (A), the contour defines a PCC_k component and the generated seed is indicated by the cross mark. In (B) some of the iterative steps of the algorithms are shown. In (C), the final Reg^{UAV} provided by the algorithm is shown.

2.4. Evaluation of regions

Homogeneous regions were defined around potential change components, having almost no restriction (other than maximum area stopping criterion) regarding the generated difference image. In this step we evaluate registered homogeneous regions, denoted as Reg_{id} 's, that are registered in the region map, either accepting them as changes or rejecting them.

A number of criteria for evaluating Reg_{id} 's are defined. The first one is a *minimum area criterion*, as regions that are not large enough are possibly noise artefacts or unimportant differences. A second criterion is the *percentage of the potential change over the region*. This derives from the concession that if a large part of a homogeneous region has produced high difference values, it should be accepted as change. If not, it should be rejected, as this slight difference may be due to noise artefacts, shadow components and differences in edge sharpness. Two examples, one for each case, are presented in Fig. 9.

A third criterion for evaluating a Reg_{id} in Im^{UAV} is to search if it appears also in Im^{sat} . This is checked via template matching (Theodoridis and Koutroumbas, 2009). The aim here is to reject regions that appear in both images but in a different way (for example, intensity and sharpness differences due to different sensor sensitivity), taking possible misplacements into account. The search is performed by the following steps.

1. Firstly, two $+n_2$ segments, Seg^{UAV} and Seg^{sat} , are cut from each image around the position of the Reg_{id} under study. Then, intensity normalization is applied on the UAV segment with respect

to the satellite segment. This local normalization aims to minimize intensity differences that could not be diminished by the global normalization during image preprocessing. Finally, both Seg^{UAV} and Seg^{sat} are converted into grayscale luminance images.

2. The search is carried out as a small $+n_3$ segment Seg^{obj} , which contains the Reg_{id} , is cut and slid over the Seg^{sat} , and correlations r are calculated for each search point (i,j) using Pearson correlation coefficient formula (Theodoridis and Koutroumbas, 2009).

The maximum correlation may indicate the position where the search object exists. Moreover, this position should be close enough to the position of the object in Im^{UAV} , which is the center of the Seg^{obj} . Therefore, the search area is limited in a disk area of ρ pixels around the center of Seg^{obj} and only pixels within this area are taken into account for finding maximum correlation. If the correlation is high enough, the object is confirmed to appear with little change in both images. The parameter n_2 can be fixed to multiples of 40–50, according to the resolution of the image, so that a large enough segment is defined to capture local color variations (we used 100 pixels for 0.5 m resolution). The parameter n_3 defines whether Seg^{obj} is expanded around Reg_{id} and its values range from 0 pixels (no expansion) to 5 pixels.

The three criteria are checked in the following order, from the less to the most computationally demanding. If one of them is not satisfied, the checking stops.

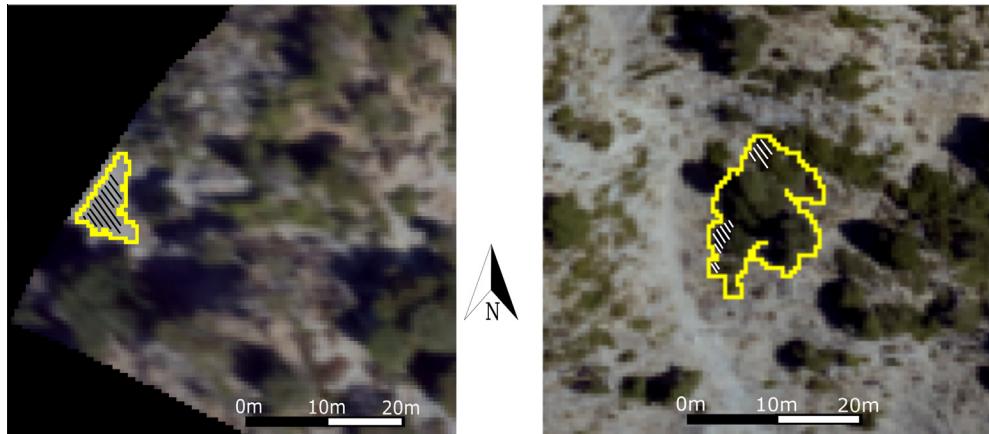


Fig. 9. The potential change percentage criterion. The contours encircle homogeneous regions, while the parts with the line pattern correspond to the intersection with PCC_k . On the left, the intersection is covering a large part of the Reg_{UAV} and it should be accepted as change. On the right, the respective area is small compared to the total area of Reg_{UAV} , it is regarded as an unimportant difference and is therefore rejected.

1. **Region area:** $A_{Reg} > N_2$ pixels, where $A_{Reg} = |\{(i,j) : (i,j) \in Reg_{id}\}|$. The threshold N_2 is set accordingly to the smallest change object of interest. Typical values range from 5 to 20 pixels for 0.5 m ground resolution.
2. **Change percentage:** $A_{Reg \cap D_{th}} / A_{Reg} > p_3$, where $A_{Reg \cap D_{th}} = |\{(i,j) : (i,j) \in Reg_{id}, D_{th|i,j} = 1\}|$.
3. **Correlation coefficient:** $r_{max} < c$, where r_{max} is the maximum correlation in a disk-shaped area with a radius of ρ pixels.

If all these criteria are true, Reg_{id} is confirmed as change and is registered on a change map.

3. Materials and methods

3.1. Case study

Datasets from two different areas are used in this study. The first study area is in Markopoulo, Attica, Greece ("Markopoulo dataset"). It is an area outside the village of Markopoulo, where the Olympic Shooting Range Hall is located. A VHR satellite image was used as reference and a Digital Terrain Model (DTM image) provided the height information. Eight UAV images that were captured on a different date from the satellite image were used for CD. On that date, several vehicles were parked outside the shooting range and detecting them was the main task of the case under study.

The second study area is located in Samaria, Crete ("Samaria dataset"). It is an uninhabited mountainous area that is renowned for the Samaria gorge (where steep slopes are encountered), a national park inside the mountain range of Lefka Ori. This is a challenging area, which was selected due to its large soil and object diversity and steep gradients. Similar to Markopoulo dataset, a satellite image and a DTM constituted the reference, while two UAV images taken from two spots inside this area were used for CD; the first one depicts a relatively land area with a rocky texture, while the second one depicts a refuge located near the edge of the gorge. A wide variety of natural objects exist in the two images, including clusters of trees, bushes, dirt road segments, rocks, and a building, thus bringing a sufficient set for testing the proposed methodology.

In both datasets, the DTM was provided by the National Cadastre & Mapping Agency S.A. and it is the best available for the Greek territory. It is worth noting that the landscape conditions of the two study areas are significantly different in terms of topography,

(e.g. in the Samaria dataset there are several steep slopes compared to the Markopoulo dataset) and land use–land cover specificity (e.g. in the Samaria dataset we have almost exclusively natural components while in the Markopoulo dataset there are also man-made constructions). In addition, the UAV images from Markopoulo were acquired from a lower flying altitude, which provides images of higher ground resolution and smaller footprint. The wide area satellite images, as well as some small cropped segments from the UAV and the satellite images, are shown in Fig. 10. The large difference in the ground resolution of the UAV and satellite images is easily observable. In addition, the properties of these datasets are listed in Table 2.

3.2. Methods

3.2.1. Detecting changes

The proposed methodology was used to register the UAV images on the reference satellite image and detect changes in an automatic way. The values of the user defined parameters and thresholds were set empirically and they are listed in Table 3. The resulting ortho-rectified images (Im^{UAV}) were also manually inspected for changes, compared to their respective part in the satellite image (Im^{sat}), and a ground truth binary change map was obtained. This was done to compare the automatic change detection methodology to the manual inspection performed by an image analyst, in the context of surveillance applications.

The manual inspection revealed the following kinds of change. Regarding the Markopoulo dataset, a total number of 33 cars were identified near the shooting range, as well as 5 motorbikes, a bus and the remnants of two buses that had been disassembled before. Concerning the Samaria dataset, three changes were highlighted for the first set of images and 15 for the second one. The main change of interest is a couple of cars that were parked outside the refuge. Small land components were also highlighted as change, whose texture or spectral appearance seemed to be different between the two images and implied land conversion. On the other hand, two common kinds of change, which were not of interest, were also observed: (a) small ground areas occluded by shadow and (b) distortion effects due to steep slope. Concerning the latter, the steep slope exclusion processing stage (Section 2.1.3) is designed to exclude steep slope areas from detecting changes. To evaluate the effectiveness of this stage in rejecting false changes, the proposed change detection methodology was repeated without

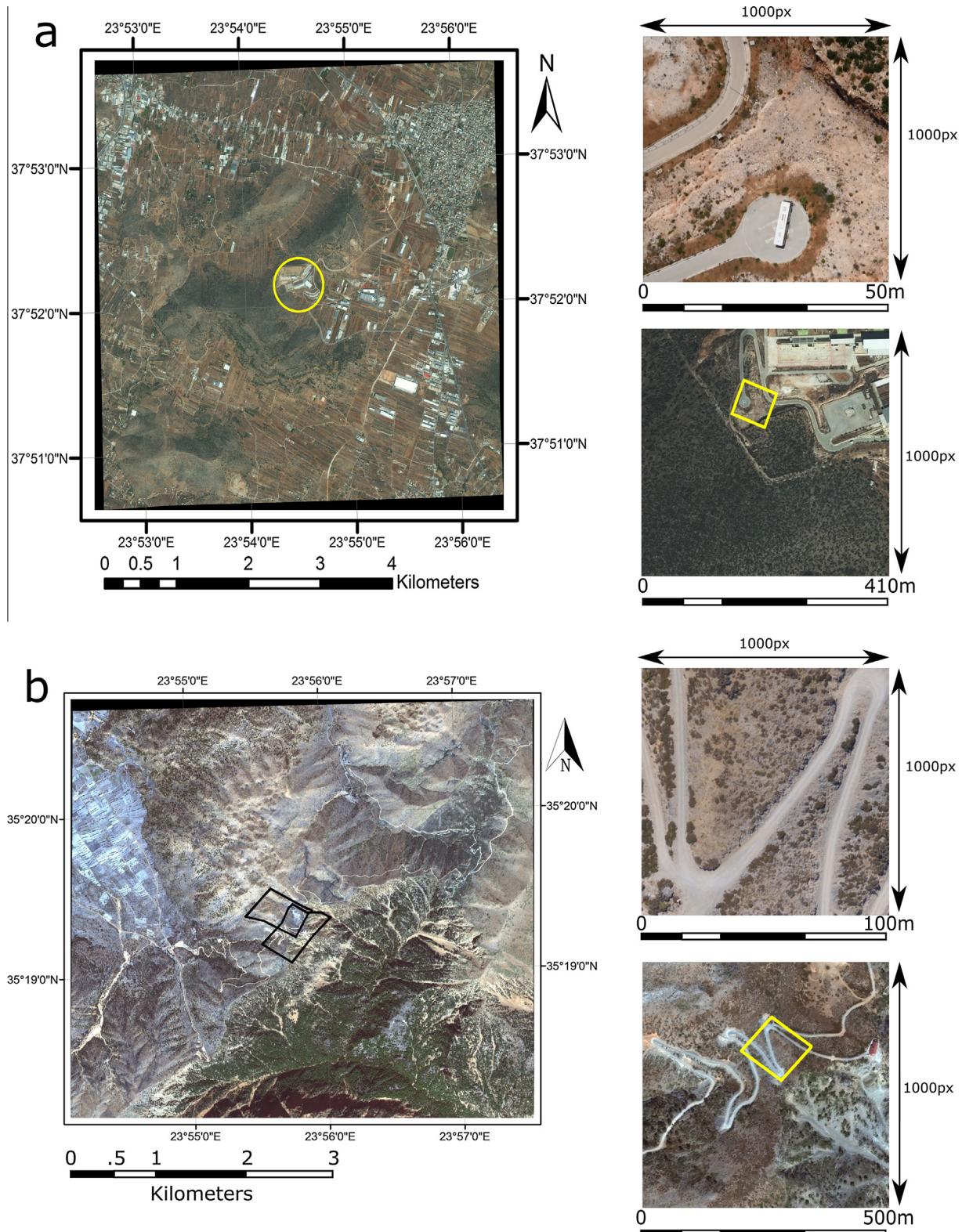


Fig. 10. The area of the case study of (a) Markopoulo and (b) Samaria datasets. In the satellite images shown on the left, in (a) the circle indicates the spots where the UAV images were acquired, while in (b) the polygons define the footprint of the two UAV images. On the right, couples of 1000 × 1000 pixel cropped image segments from one of the UAV images (top) and the satellite image (bottom) are shown.

steep slope exclusion for the second image of the Samaria dataset where this effect was evident in a large part of it.

The proposed CD methodology was implemented in MATLAB. The experiments were conducted using a PC of 64-bit i5 3.10 GHz processor and of 16 GB RAM.

3.2.2. Evaluating robustness to the misregistration errors

To assess the robustness of our methodology to misregistration errors, we simulated misregistration errors in a way similar to previous works (Bovolo et al., 2009; Chen et al., 2014; Dai and Khorram, 1998; Townshend et al., 1992; Wang and Ellis, 2005)

Table 2

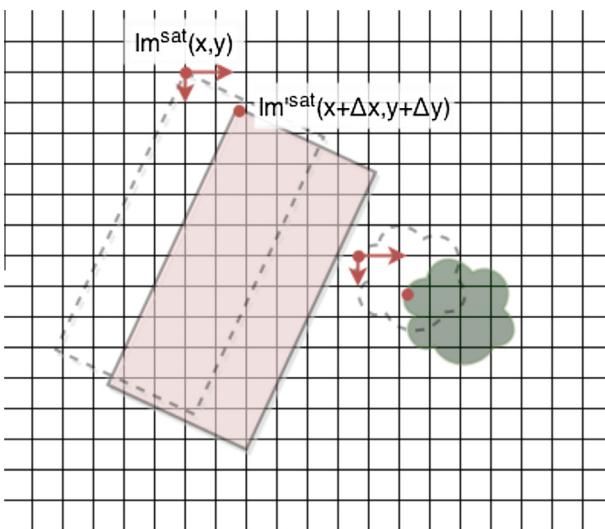
The properties of the images used in the experiment.

| Site | Type of image | Image resolution (pixels) | Ground sampling distance (m) | Area of coverage (km^2) | Date and time of acquisition | Height of acquisition (approx.) (m) |
|------------|---------------|---------------------------|------------------------------|------------------------------------|------------------------------|-------------------------------------|
| Markopoulo | Satellite | $14,031 \times 13,891$ | 0.41 | 32.76 | 18/6/2015 midday | – |
| | DTM | 1770×2160 | 5 | 3.82 | – | – |
| | UAV | 1920×1080 | 0.03–0.05 | 0.005184 | 26/5/2015 morning | 100–130 |
| Samaria | Satellite | $10,443 \times 9667$ | 0.5 | 25.24 | 14/9/2013 morning | – |
| | DTM | 2562×1398 | 5 | 89.54 | – | – |
| | UAV | 6000×4000 | 0.1 | 0.24 | 14/8/2013 morning | 650 |

Table 3

List of user defined parameters and their values used in the experiment.

| Description | Variable | Value |
|--|-----------------------|---|
| <i>Extracting potential change components</i> | | |
| Search window edge length (in pixels) | w | 11 |
| Minimum PCC size (pixels) | $N_1 (N'_1)$ | 20 (5) |
| <i>Generating region map</i> | | |
| Region growing: color criterion for “bare ground” and “high saturation” classes respectively | δ_1, δ'_1 | 0.1, 0.15 (Markopoulo) 0.1, 0.25 (Samaria) |
| Region growing: percentage of maximum gradient magnitude in Seg^{UAV} | p_1 | 0.7 |
| Region growing: additive size for Seg^{UAV} (pixels) | n_1 | 10 |
| Region merging: region similarity (see Appendix A) | δ_2, δ'_2 | 0.1, 0.15 |
| Region merging: region overlay percentage | p_3 | 0.7 |
| <i>Evaluating regions</i> | | |
| Additive size for Seg^{sat} (pixels) | n_2 | 100 |
| Additive size for Seg^{obj} (pixels) | n_3 | 4 |
| Minimum change size (pixels) | N_2 | 5 |
| Minimum change percentage | p_2 | 0.33 |
| Maximum correlation | c | 0.75 |
| Object centroid to maximum correlation point distance | ρ | 10 |

**Fig. 11.** The same translation is applied to every pixel as misregistration is simulated.

using the dataset from site Samaria; after control points have been acquired and the UAV image has been ortho-rectified (see Section 2.1.1), the satellite image is slid by increments of 0.2 pixels in both axes, before the proper part is cut. In other words, a new Im^{sat} image is produced, where (x', y') pixel coordinates correspond to $(x + \Delta x, y + \Delta y)$ pixel coordinates in Im^{sat} , where $\Delta x = k_x \cdot 0.2$, $\Delta y = k_y \cdot 0.2$ and $k_x, k_y = 0, 1, \dots, 35$ (see also Fig. 11). When either offset is not an integer number, the intensity values in each channel of the satellite image are calculated via

bilinear interpolation. Then, Im^{sat} is provided as input to the change detection method instead of Im^{sat} . At this point, we should note that the interpolation increases correlation between neighboring pixels, which was regarded as an additional factor of noise between the two images.

This process is repeated for descriptor search windows W_{ij} of different sizes by setting window sizes to 3×3 (± 1 pixel), 7×7 (± 3 pixels), 11×11 (± 5 pixels) and 15×15 (± 7 pixels) (see Section 2.2.2). This way, we observe how the choice of the window size affects the quality of difference image and the consistency in detecting changes.

3.2.3. Accuracy assessment

Let D' be the generated difference image and D'_{th} the thresholded binary image at misregistration offset $(\Delta x, \Delta y)$, and D and D_{th} the respective images at zero misregistration (these two will be referred to as baseline images). In order to assess the effect of misregistration, the following measures concerning the difference images are calculated.

i. Normalized mean square error:

$$\text{NMSE} = \frac{1}{|M|} \sum_{(x,y) \in M} \frac{(D'_{x,y} - D_{x,y})^2}{D_{x,y}^2} \quad (10)$$

where M is the set of all pixels (x, y) where differences are evaluated (pixels indicated by the slope mask are excluded, see Section 2.1.3).

ii. Correlation coefficient distance:

$$CC = 1 - \frac{1}{\sigma_D \sigma_{D'}} \cdot \frac{1}{|M| - 1} \sum_{(x,y) \in M} (D_{x,y} - \mu_D)(D'_{x,y} - \mu_{D'}) \quad (11)$$

where $\sigma_D, \sigma_{D'}$ are the standard deviations of difference values in D and D' and $\mu_D, \mu_{D'}$ their respective mean values.

In addition, we evaluate the binary images from which PCC_k are extracted. This is done via the following steps:

1. Generating the intersection image as $D_{\cap} = D'_{th} \cap D_{th}$. This image contains the true positive components that exist in both $\Delta x, \Delta y$ and zero misregistration offsets.
2. Performing morphological opening by reconstruction (Gonzalez et al., 2002) of D_{\cap} under the D'_{th} image. This is done to recover any component that is not formed exactly in the same shape as in the baseline binary image. In this way, a single baseline component is reconstructed, which may have been split into two components at $(\Delta x, \Delta y)$ misregistration. The same way, the merging of two baseline components is corrected.
3. Extracting components in D_{th}, D'_{th} and D_{\cap} and computing their cardinalities, as $|D_{th}|, |D'_{th}|$ and $|D_{\cap}|$ respectively.

Then, the following measures are calculated.

iii. Precision/False Discovery Rate (*FDR*):

$$\text{Precision} = \frac{|D_{\cap}|}{|D'_{th}|}, \quad FDR = \frac{|D'_{th}| - |D_{\cap}|}{|D'_{th}|} = 1 - \text{Precision} \quad (12)$$

iv. Recall/False Negative Rate (*FNR*):

$$\text{Recall} = \frac{|D_{\cap}|}{|D_{th}|}, \quad FNR = \frac{|D_{th}| - |D_{\cap}|}{|D_{th}|} = 1 - \text{Recall} \quad (13)$$

v. Overall Increase Percentage (*OIP*):

$$OIP = \frac{|D'_{th}| - |D_{th}|}{|D_{th}|} \quad (14)$$

In words, *Precision* measures how many of the detected changes correspond to true changes, *Recall* measures how many true changes have been detected, while their complementary measures *FDR* and *FNR*, respectively, are self-explained. Finally, *OIP* measures the relative increase in detected changes.

4. Results and discussion

In the sequel, first, the influence of the various stages of the proposed methodology in the formation of the final outcome, when it is applied on the specific datasets described before, is shown. Then, the immunity of the methodology to misregistration is discussed in detail. Also, ways of modification of the proposed methodology, in order to work with multispectral data are proposed. Finally, computational complexity issues are considered, showing that the proposed methodology is an effective algorithm for surveillance applications.

4.1. Effectiveness of the methodology

First, the processing stage of image registration is discussed. The SIFT algorithm found keypoints in the cropped reference satellite image and in each of the original UAV images. A small percentage of them initially matched in pairs and the RANSAC method then discarded some of them, providing the final matches that are used as ground control points (GCPs) for the images. The numbers of extracted and matched keypoints are shown in Table 4. It is easily noticed that the number of GCPs for the Markopoulo dataset is by far smaller than the respective for the Samaria dataset, which is mainly due to the small area that is covered in the UAV images. The disadvantage of a small area of coverage is that it provides less

Table 4

The number of the keypoints that were extracted and matched between the satellite and the UAV images. For the Samaria dataset, a rounded approximation is provided since these numbers were not much different for the two images.

| | Satellite keypoints | UAV keypoints | Initial matches | GCPs |
|---------------------------|---------------------|---------------|-----------------|-----------|
| <i>Markopoulo dataset</i> | | | | |
| Mean (std) | 1092.3 (481.5) | 462 (152.6) | 54.1 (26) | 26.8 (10) |
| Range | 307–1678 | 271–711 | 16–93 | 15–46 |
| <i>Samaria dataset</i> | | | | |
| | 12,800 | 3500 | 1200 | 780 |

keypoints, making it less likely to extract GCPs that are evenly distributed on the image and thus not bringing enough statistical information to generate the ortho-projection of the UAV image. In addition, large homogeneous segments, which produce no keypoints, are expected to cover a large portion of the image and have a larger effect on the registration. At this point we note that due to these reasons, the registration failed in one of the UAV images in Markopoulo dataset as GCPs could not be extracted, and no CD could be performed for it. Also, for the Markopoulo dataset the Least Squares model only contained exterior orientation parameters; when including the interior parameters the model was unstable due to the fact that a small number of GCPs was extracted. Therefore, the UAV should fly so that to acquire images of significant footprint size (Fan et al., 2010). The Markopoulo dataset is presented here as a marginal case, meaning that images of smaller footprint might not be automatically registered. We should note finally that for greater accuracy, Direct Linear Transformation (Abdel-Aziz and Karara, 1971) could be enforced. However, this was not necessary for our methodology since it makes up for any misregistration errors as it has been described previously.

In Table 5, the results from the photogrammetric resection are presented. We note that the resulting sigma are higher than the expected sigma from a strict photogrammetric ortho-image, which results in misregistration between the satellite image and the generated ortho-image. However, this kind of errors was expected given the available dataset but they are not a restriction for the purpose of this application and are treated properly by the steep slope exclusion, the pixel descriptor vectors and the change evaluation stages.

In the following, the steps concerning the CD are discussed. An example of the difference image D is shown in Fig. 12. Bright pixels represent high difference values, conversely for dark pixels. The integration of neighboring pixels correlated their difference values and suppressed the “salt and pepper” noise. In fact, the difference images are relatively smooth, even in textured areas; the difference values in these areas are clearly higher, but significant differences are found only in groups of several pixels. The stage of extracting potential change components finally discarded the noise in D from further processing. The thresholding method filtered out the

Table 5

Photogrammetric resection results for each UAV image.

| | σ_0 (pix) | X_0 (m) | Y_0 (m) | Z_0 (m) | ω (deg) | φ (deg) | κ (deg) |
|--------------------|------------------|-----------|-----------|-----------|----------------|-----------------|----------------|
| Markopoulo dataset | ±1.44 | ±2.33 | ±1.33 | ±0.34 | ±0.74 | ±1.33 | ±0.15 |
| | ±1.73 | ±2.97 | ±1.5 | ±0.47 | ±0.9 | ±1.86 | ±0.37 |
| | ±1.27 | ±2.74 | ±2.19 | ±0.94 | ±0.85 | ±1.06 | ±0.22 |
| | ±3.57 | ±11.8 | ±5.93 | ±1.74 | ±1.71 | ±3.43 | ±0.74 |
| | ±2.24 | ±11.12 | ±4.48 | ±1.42 | ±1.51 | ±3.83 | ±0.32 |
| | ±1.99 | ±3.69 | ±4.06 | ±3.28 | ±1.57 | ±1.57 | ±0.52 |
| | ±1.65 | ±2 | ±1.57 | ±0.73 | ±0.53 | ±0.68 | ±0.15 |
| | | | | | | | |
| Samaria dataset | ±0.8 | ±0.88 | ±0.94 | ±4.94 | ±0.05 | ±0.04 | ±0.01 |
| | ±1.3 | ±0.71 | ±1.14 | ±4.12 | ±0.11 | ±0.09 | ±0.02 |

majority of pixels and the extraction of connected components highlighted large groups of pixels, which are of great importance. The resulting PCC's are also shown in Fig. 12.

The regions that were defined by the SRG algorithm with the help of the initial global segmentation step correspond to objects (such as the cars, trees and the roof of the refuge), road segments, various bare ground areas, and shadow components. The vast majority of the regions in both datasets are properly defined, showing the robustness of the Region Map Generation stage. Moreover, the evaluation criteria rejected many regions as false changes. Many of them were rejected by the area criterion, including a large number of shadows that are cast by trees; a shadow component grew into a shadow-plus-tree region, which was rejected since the shadow part of the region is small compared to the tree part of the region. A number of changes were also rejected by template matching. An illustrative example of this case is shown in Fig. 13, which is taken from Samaria dataset. The bushes are clearly depicted in Im^{UAV} , but not in Im^{sat} . The spectral and gradient values are different for the pixels that belong to the bushes and they are signified as potential change components. However, the correlations indicated that the image segments were similar and, despite the misleading difference, there exists no true change. Another example, which belongs to the Markopoulo dataset, is shown in Fig. 14, in the top right pair of images. In this example, a set of metal sheets are clearly depicted in both images, but there is a

slight color difference due to the sensitivity of the satellite sensor, as well as evident misregistration errors. Even so, the template matching in a short search range produced high correlation and this object was finally rejected.

However, there were a few cases of objects undersegmentation that were observed in the Markopoulo dataset that corresponded mainly to objects whose spectral intensity values were not much different from their surrounding elements, such as white cars and the bus compared to the tarmac, or dark colored cars that were parked near trees. To avoid this, a lower threshold for the color distance parameters of the SRG algorithm, namely δ_1, δ'_1 , is used. A few cases of oversegmentation were also observed in white cars, because their windshield is seen as black and it was grown as a different region, but most of them (7 out of 9) were still detected as change with the help of their shadow.

On the whole, the proposed methodology succeeded in detecting most of the cars in the Markopoulo dataset (31 out of 33), the bus and the bus remnants, but only one of the motorcycles was detected because they were in general not large enough to be accepted by the evaluation criteria. In addition, it revealed some additional changes outside the road and the parking area which belong to trees and bare ground segments. A total of 82 changes were detected in the seven images that were used for CD, 53 of which were verified as changes via manual inspection. The rest of the changes (false positives) were in fact natural changes but

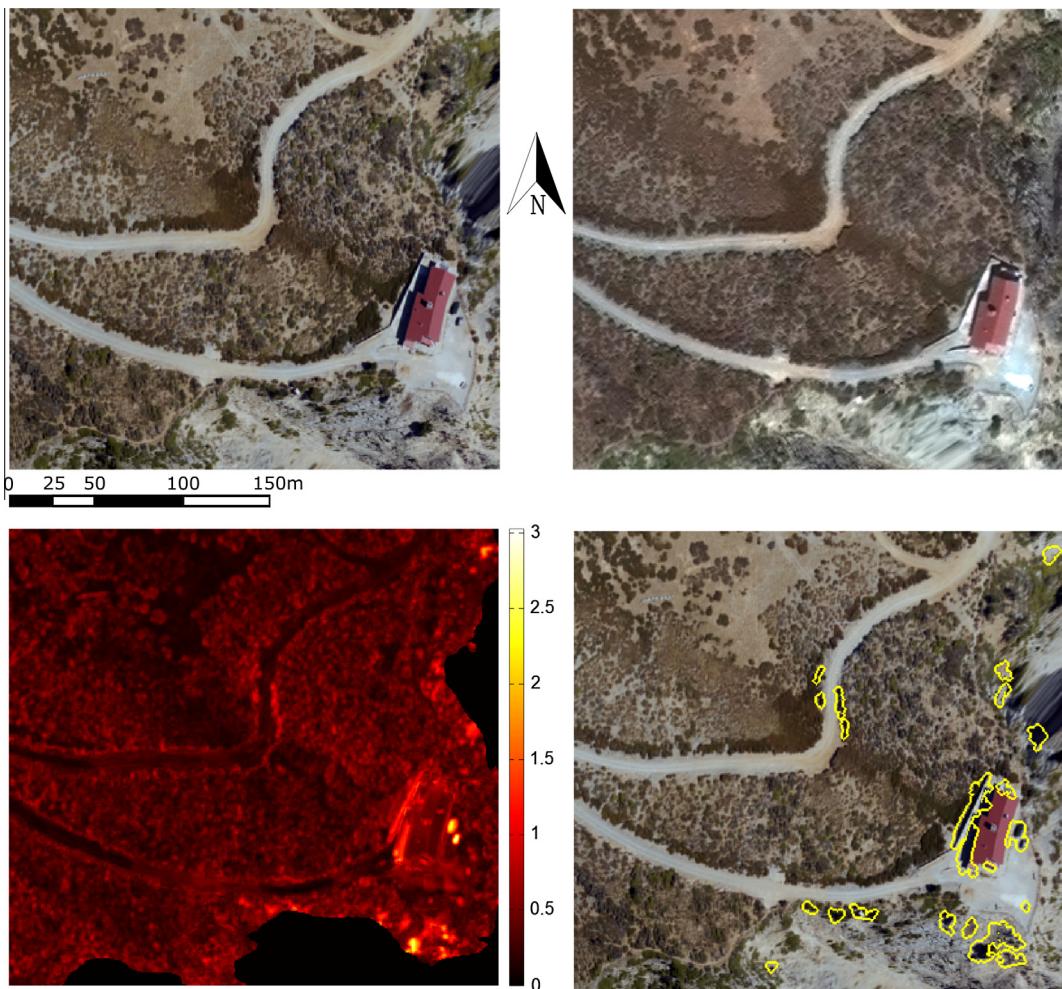


Fig. 12. The difference image D (bottom left) is calculated for the two images at the top. D is thresholded and potential change components are extracted from the binary image D_{th} (bottom right). The images belong to the Samaria dataset.



Fig. 13. The effectiveness of the evaluation criteria in rejecting false changes which arise due to the difference in image sharpness between Im^{UAV} (left) and Im^{sat} (right). The PCC's (marked on the left image) were refined via the SRG algorithm into Reg^{UAV} 's. The evaluation criteria help in rejecting many homogeneous regions as no change; finally, the rest of them are confirmed as changes (marked on the right image).

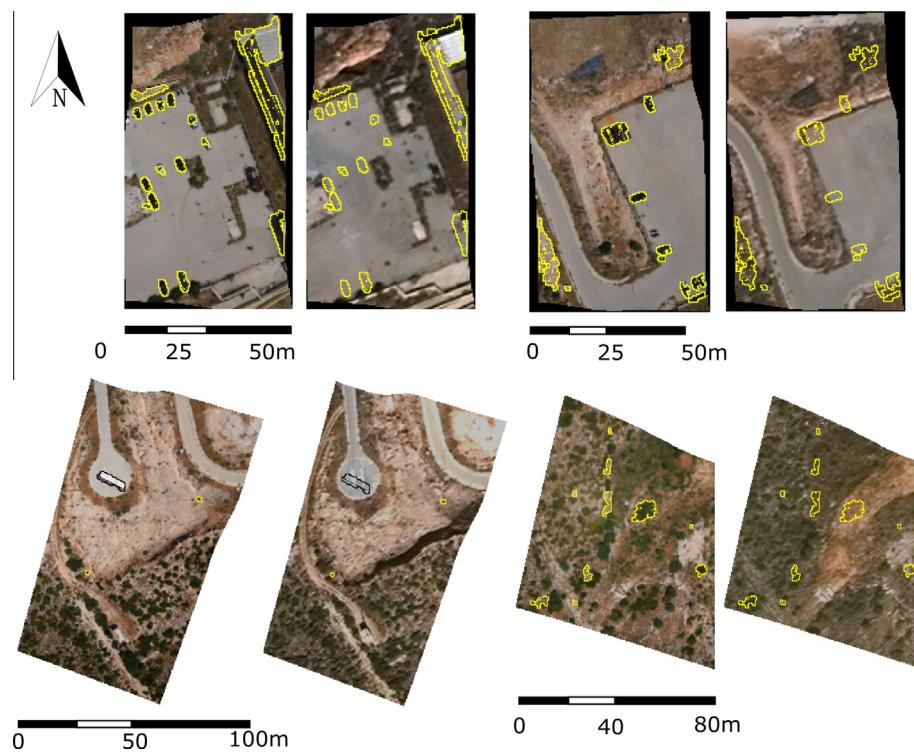


Fig. 14. Some of the changes that were detected in the Markopoulos dataset are drawn in Im^{UAV} (the image on the left of each pair) while their contours are plotted in the respective places in Im^{sat} (on the right). Dark contours are used when necessary to enhance contrast.

they were not of interest and they would only be rejected via manual inspection. Some examples of the changes that are reported are shown in Fig. 14.

As for the Samaria dataset, the manual inspection of the images revealed that the factors that are responsible for false changes prevail over the true changes that exist in the images. Shadow components appear in many occasions and they are cast in different sides of objects because the UAV and satellite images have been acquired at different times in the day. In addition, some parts of the second UAV image (at the edge of the gorge) have been severely distorted. This is the reason for choosing these images in first place, i.e. to test the proposed methodology on extremely difficult cases. The number of PCC's that were extracted is still high, as 119 and 182

components were extracted from the two images, but this number was reduced to 21 and 87 regions respectively that were accepted as changes, after the Region Map Generation and Change Evaluation processing stages.

The proposed methodology succeeded in detecting almost all of the manually highlighted changes, 2/3 in the first set of images and 13/15 in the second one. The rest of the changes which were highlighted by the proposed methodology were mainly regions which were partially occluded by shadow in either Im^{UAV} or Im^{sat} . Concerning the second image of the dataset, almost half of the changes detected, including the missed ones (false negatives), were found in the gorge area (see also Fig. 15). They corresponded to regions which were either distorted due to the viewing angle in relation

to the low resolution of the DTM or occluded by shadow. We should note that the slope gradient in these distorted regions was not high enough and they survived the threshold that was set to construct the slope mask. However, the largest part of the distorted areas was masked out successfully. When the change detection methodology was performed without the slope exclusion stage, the number of detected changes rose to 106 for the second set of images while the processing time quadrupled. At the same time, the true changes percentage decreased to 8/15. This is due to the fact that the distorted regions were responsible for great difference values and they were taken into account at the expense of other changes.

4.2. Robustness against misregistration errors

Fig. 16 shows the way NMSE grows as larger misregistration offsets ($\Delta x, \Delta y$) are simulated, for different sizes of W_{ij} search windows. Clearly, when larger windows are used, NMSE grows in a much slower rate. This is because, as the search is performed for a larger area, the homologous pixel may still be found even if it is misplaced by many pixels, until it is finally out of reach. In large window sizes (11×11 and 15×15), we also observe how the interpolation affects the quality of the satellite image, as it was previously mentioned; since pixel intensity values are interpolated, correlation between neighboring pixels is increased, thus the misregistered image is a smoothed version of the original satellite image. This way, additional errors are introduced in the descriptor elements, producing slightly greater difference values. This is not the case when integer values are used for misregistration offsets. As a consequence, in this case, the difference image D' is more similar to the baseline image, resulting thus in small “holes” in the NMSE surface. Finally, we observe that the NMSE values are larger along Δx axis, which is probably due to the vertical orientation of Im^{UAV} . In the same way, **Fig. 17** shows the CC distance, supporting further that, when large windows are used, the difference image D' is similar to D even for large ($\Delta x, \Delta y$) offsets.

The similarity measures concerning the binary images evaluate the robustness of the overall procedure of extracting potential

change components in the presence of misregistration. This evaluation is performed in a way similar as performed by [Dai and Khorram \(1998\)](#), where a pixel-based change detection method is used in medium ground resolution images. **Figs. 18 and 19** show that, when a large W_{ij} window is used, the methodology becomes more robust to misregistration errors, considering the components that are extracted. *FDR* increases in a slow rate, indicating that a small number of false changes is introduced progressively. In the same way, *FNR* increases in a slow rate as well, indicating that most of the true PCC components at zero misregistration are still identified, despite the increase in misregistration length. In **Fig. 20**, their respective complementary measures, *Precision* and *Recall*, are plotted as a function of the misregistration length regardless of direction, by averaging the values of the offsets of the same length (for example, all of $(5, 0), (3, 4), (4, 3)$ and $(0, 5)$ offsets introduce a total misregistration length of 5 pixels). In these figures, the rate of decrease is easily observed and can be characterized as linear, excluding the abrupt drop within the first pixel of misregistration. As a consequence, the accuracy in detecting true changes remains high even for large misregistration errors, as indicated by an average *Precision* value of nearly 90% and 80% for 2 and 4 pixels respectively for both large windows. In the same manner, the average *Recall* is nearly 85% and 80% for 2 and 4 pixels respectively in the case of a 11×11 window and 90% and 80% for a 15×15 window. This is a significant improvement compared to the percentages reported in [Dai and Khorram \(1998\)](#) where, even at a single pixel misregistration, the *Precision* (true positive) percentage dropped sharply. The medium sized window (7×7) seems to produce similar results ($Precision \approx 85\%$ at 2 pixels, $>70\%$ at 4 pixels and $FNR \approx 20\%$ at 2 pixels, $<30\%$ at 4 pixels). However, when the small window is used, *Precision* drops in a faster rate within the first 5 pixels of misregistration and stabilizes at 40% (5 pixels) to 30% (10 pixels), while the *FNR* increases in a smoother way and reaches 25% at 2 pixels and 38% at 4 pixels. All things considered, the potential change extraction approach does not require the image registration procedure to be perfectly accurate, as it can compensate for spurious differences.

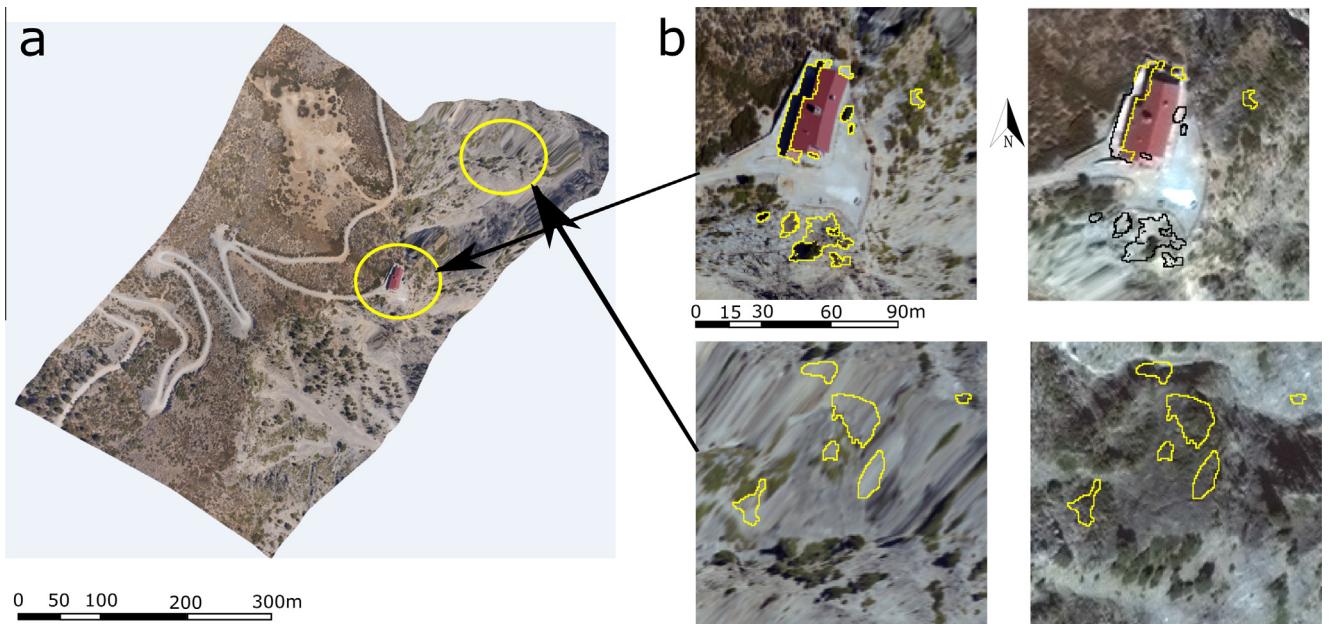


Fig. 15. In (b) some of the changes that were detected in the Samaria dataset are drawn in Im^{UAV} (the image on the left of each pair) while their contours are plotted in the respective places in Im^{sat} . Dark contours are used where necessary to enhance contrast. In (a) the position of these cropped parts are also indicated in the whole Im^{UAV} . In the second pair of images, some false changes have been introduced due to the viewing angle, in relation to the low resolution of the DTM.

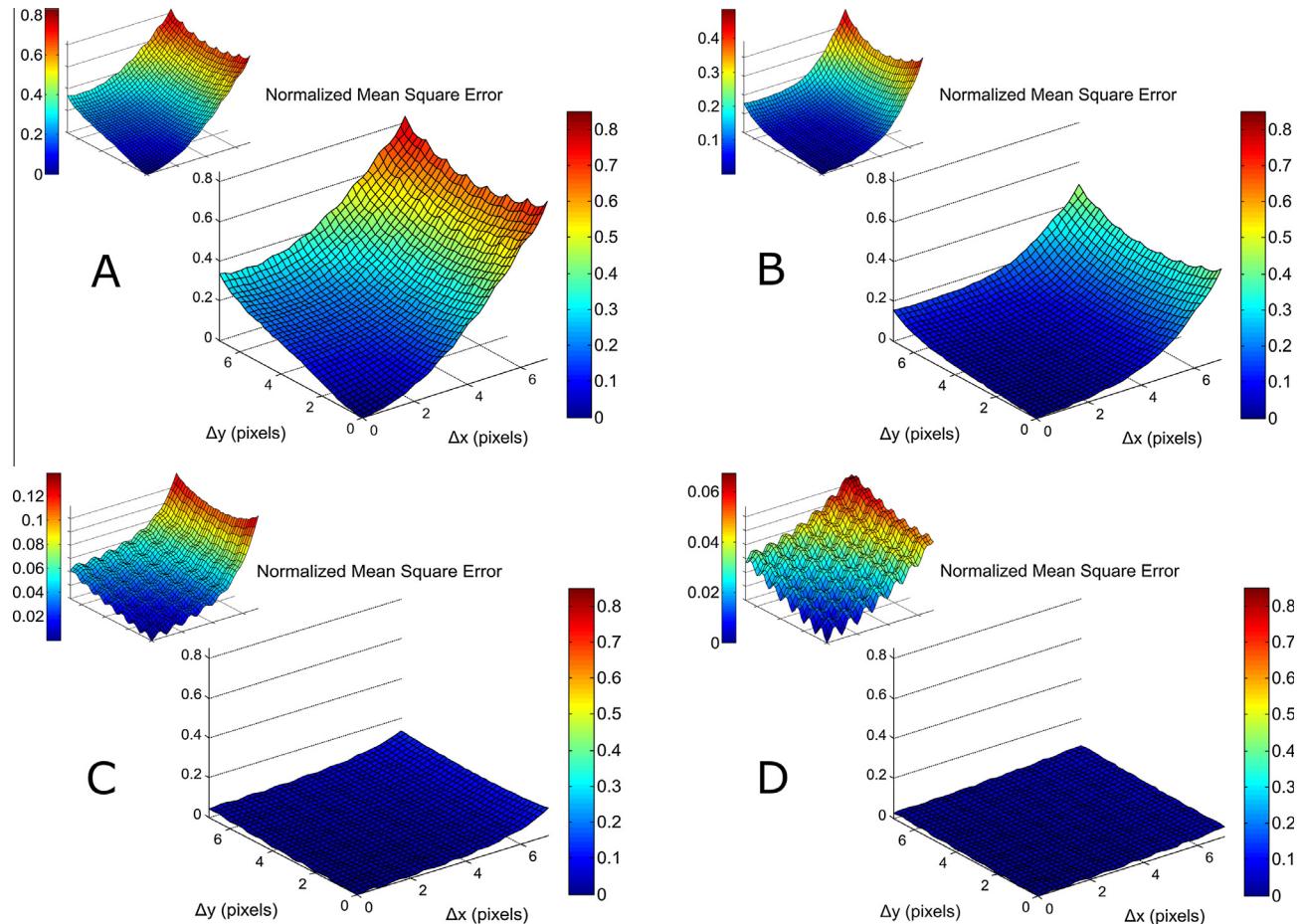


Fig. 16. The NMSE plots for all misregistration offsets, when W_{ij} of size 3×3 (A), 7×7 (B), 11×11 (C) and 15×15 (D) are used. The main figures are scaled to the maximum NMSE value at 3×3 window size, while the top left figures are scaled to their respective maximum values. In large windows, the small holes at integer offsets are an indication of the interpolation.

Last but not least, Fig. 20 shows that the total number of potential change components may change as misregistration errors are introduced. This change may be quite noticeable for large windows ($OIP = \pm 20\%$ of $|D_{th}|$), but it is more significant for the smallest window. This is clearly due to Rosin's thresholding method, as it determines a threshold by finding a characteristic curve point in the pdf of the difference image rather than defining *a priori* the amount of change between the two images.

4.3. Configuration and deployment aspects

This section discusses some aspects concerning the configuration of the proposed methodology so that it can be used effectively in surveillance applications.

An important aspect in the change detection problem under study is the spectral information provided in the related images. This information depends more on the sensor that is loaded on the UAV platform and less on the reference image; as the design of the UAV platform is more important due to restrictions such as its total cost, the reference image can be selected later so that it can be comparable to the images that will be produced by the UAV sensor. In our case, the CD methodology was developed to detect changes between RGB images, which is a cost-effective choice and, in the same time, it is considered to be the simplest case under which changes can be detected and validated automatically by the processing stages described in Section 2. However, the proposed methodology is not bounded to work only with RGB

images, but it can be extended in order to work with multispectral images as well.

This can be accomplished as follows. If more spectral channels are used, the color vector and the descriptor vector (Section 2.1.1) for each pixel become extended (since the information in the additional bands for the current and its neighboring pixels are taken into account). However, the remaining processing stages will remain unaltered. One should be careful though while bringing more channels into use, as some of them may be sensitive to factors that are responsible for false changes, and selecting the most suitable ones (see for example in Bruzzone and Serpico (1997) the use of spectral channels that give prominence to the changes of interest and others that are sensitive to misregistration errors). Another drawback would be that, as the descriptor vector becomes extended, curse of dimensionality problems and additional computational complexity requirements arise, since the generation of the difference image and the region growing and merging stages calculate Euclidean distances of vectors, which is $O(n)$ concerning the number of spectral channels that are used and, thus, the length of descriptor and color vectors of the pixels.

A second way which makes up for these drawbacks would be to reduce the dimensionality of the spectral space by applying a transformation method such as Principal Component Analysis and then select a subset of principal components that contain most of the information contained in the data. Then the whole proposed methodology is applied on the images composed by the selected principal components. However, the SRG should still be performed

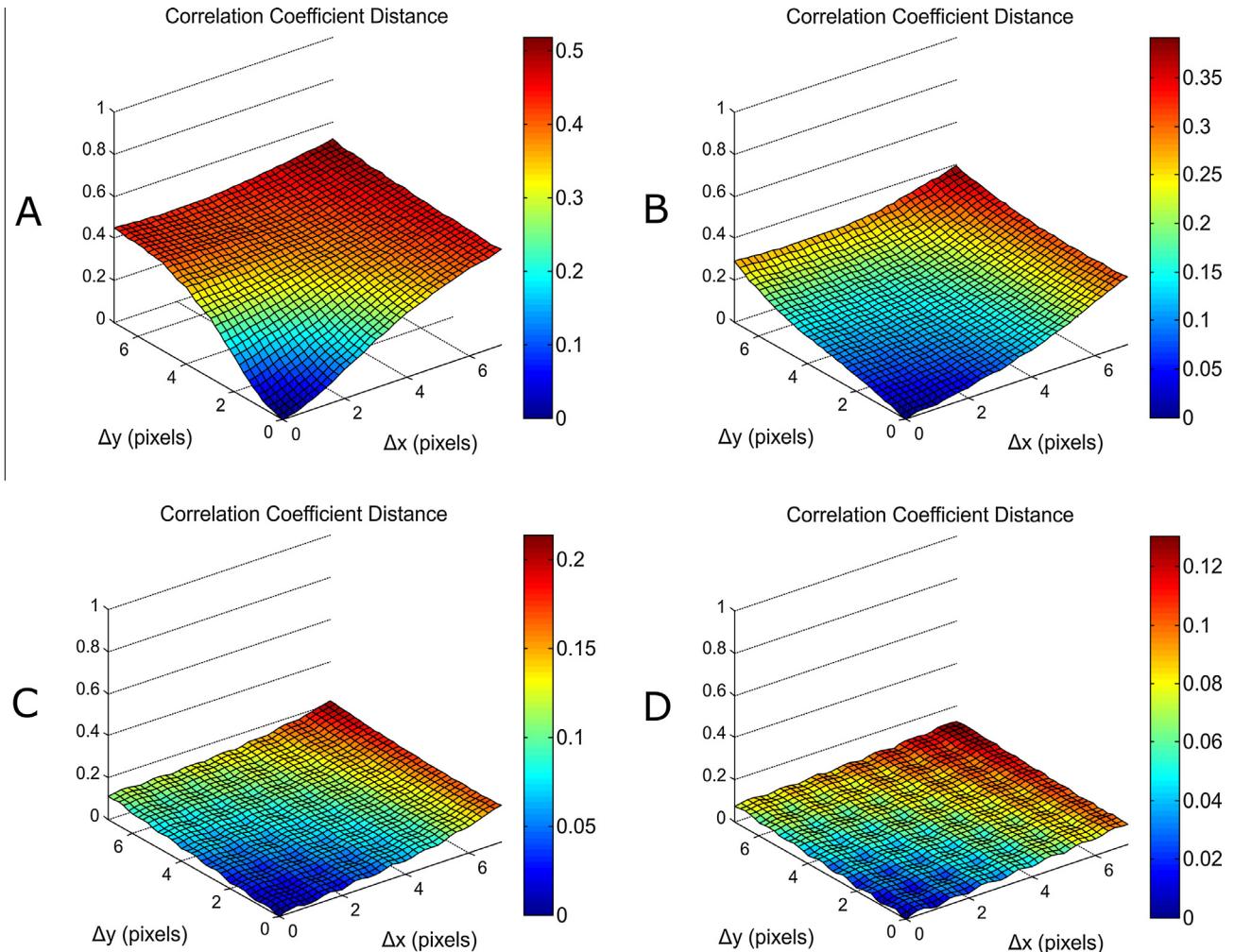


Fig. 17. The CC distance plots for all misregistration offsets, when W_{ij} of size 3×3 (A), 7×7 (B), 11×11 (C) and 15×15 (D) are used. For large windows (C, D), extremely high correlations are calculated for the difference images even for large misregistration offsets.

in the original multispectral images since in the transformed space the texture of the resulting images is expected to change significantly and the SRG algorithm might not be able to define objects properly.

Apart from the above extensions, which mainly transform the input given to the proposed methodology, the use of more spectral channels could bring advantages to individual processing stages of the methodology. For example, we could use more spectral channels for segmenting further the Im^{UAV} in more than two target classes. This knowledge can be useful first, in the region growing and later in the evaluation of the resulting regions. For example, NDVI can be used to extract vegetation and shadow components which could not be separated clearly by using RGB images; then, shadow components could be immediately rejected more confidently.

It is important to mention here that it is required that the multispectral images to be compared to have the same spectral channels so that they are comparable, even if their sensitivity in these channels is not the same (which was the case in our experiments). However, in case one image contains a few channels more than the other, the difference image generation stage (Section 2.2.1) can be replaced by the Multivariate Alteration Detector method (Nielsen, 2007; Nielsen et al., 1998), which also makes up for the difference in sensor gain and sensitivity. However, as this is a pure pixel-based method, it is vulnerable to misregistration errors, which, in

our methodology, is treated by the descriptor vectors. Having this in mind, the combination of the descriptor vector and the search window could be replaced when not many false changes are expected.

In addition, most of the parameters of the proposed methodology control the post-processing (via the SRG algorithm) and the evaluation of the changes. Some of them, for example the minimum size N_2 and the area percentage criteria have a direct physical meaning and their values can be set according to the needs of the application and the scale of the changes of interest. As for the parameters of the SRG algorithm and the regions merging, they need to be adjusted first according to an ortho-rectified UAV image, so that the resulting regions are defined properly. Last but not least, the size of the search window is probably the most crucial parameter in the methodology, as it is related to the misregistration errors. The accuracy of registration relies greatly upon the accuracy and the density of the DTM, while the morphology of the terrain is also important since rough terrains are prone to generating incorrect ortho-images. In our case, we have set empirically the size of the window by inspecting carefully the misregistration errors in both rough and smooth terrain in the set of images available, which is a practical and reliable solution. A more elegant one would be to estimate the search range dynamically, but it requires the modelling of the factors of misregistration errors, which is considered to be a very difficult task.

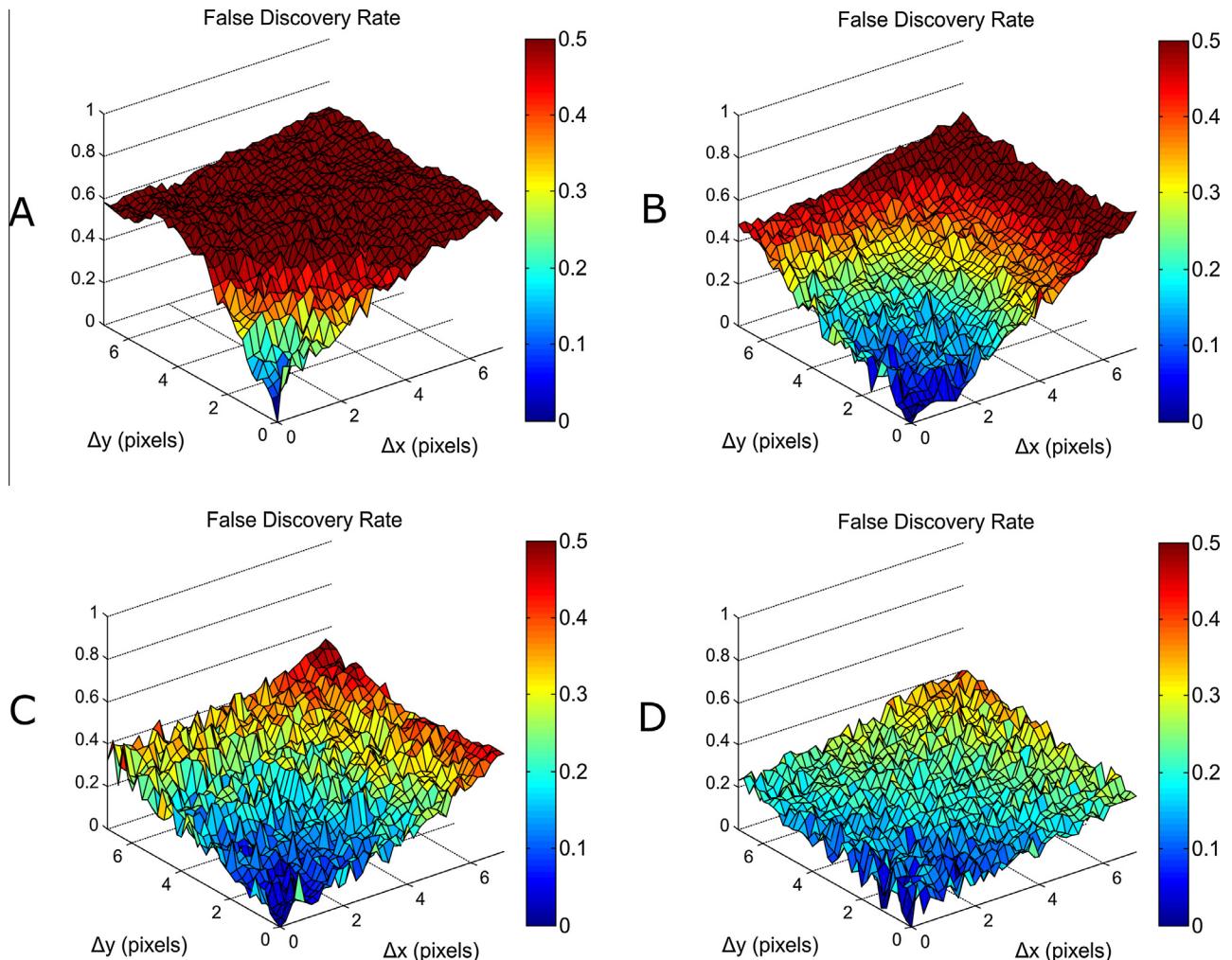


Fig. 18. The FDR values for all misregistration offsets, when W_{ij} of size 3×3 (A), 7×7 (B), 11×11 (C) and 15×15 (D) are used. For small windows, the FDR increases more quickly, which indicates that most of the potential changes detected are due to large misregistration errors (A, B). Larger windows help in rejecting them, despite misregistration errors (C, D).

However, once the parameters are set, they may never need to be adjusted again, especially when the equipment and the acquisition conditions of the UAV images are the same. The regions are also defined properly, with the help of the initial segmentation result that is obtained via a non-parametric thresholding method (Section 2.3.1). The same method is also used for the detection of PCC's, which are the first indications of changes. All things considered, the two datasets have been processed with the use of the same values for the parameters, with the exception of one, which supports further the claim that the proposed methodology is just affected by the choice of their values.

Finally, another important aspect for a system like the proposed one is the processing time required. It is of vital importance for surveillance applications to develop a procedure that completes processing in (at least) near real time, meaning that the delay introduced by the execution of the methodology is not significant, so that the report of a change is produced fast and any intervention is decided short after its detection. The proposed methodology was designed to work in detail with the images, which is accomplished via the generation of the region map, so that the evaluation of the regions takes place. The region growing is a segmentation method for defining objects accurately; in fact, similar SRG (Liu et al., 2015; Wang et al., 2010) and merging (Zhang et al., 2014) algorithms have been recently proposed that work well with VHR images.

However, region growing is a rather time consuming procedure. To this end, since there is no need to segment the entire Im^{UAV} into regions, in the proposed methodology the region growing is performed only on PCC's, so that places of no importance are bypassed. This helps in keeping the required processing time low (each image required up to 20 s for the Markopoulos dataset and up to a couple of minutes for the Samaria dataset), while this time varies according to the number of PCC's, since as more PCC's are extracted, more Reg^{UAV} 's will be grown. This number can be affected by the following parameters; (a) the minimum PCC size, N_1 , and (b) the size of the search window W_{ij} , w . If N_1 is set to a larger value, more PCC's will be excluded from further processing. Concerning w , although a large value would require slightly more time to produce the difference image (its complexity is $O(w^2)$ regarding the number of descriptor distance values to be calculated for each pixel, but w is expected to be small), it will help in rejecting differences that arise due to misregistration errors. As a consequence, less PCC's are extracted (this is also shown in the small increase in OIP, see Section 4.2), thus, the following processing stages (SRG, evaluation of changes) require less time and the total processing time is almost constant. This is illustrated in Fig. 21, where the time required by the four processing stages, when different window sizes of 7, 11 and 15 pixels are used, is displayed (the images of Samaria are used in this example).

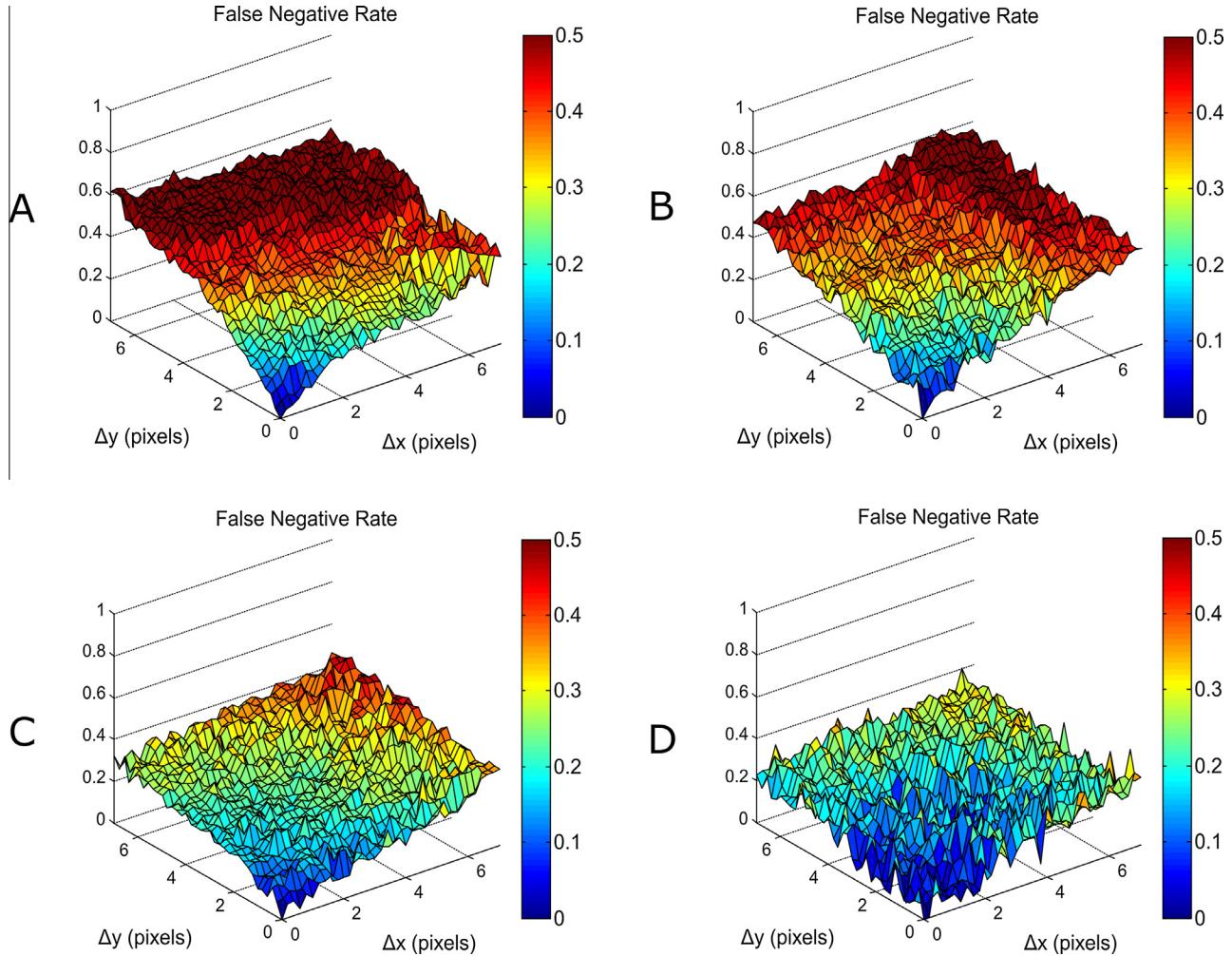


Fig. 19. The FNR values for all misregistration offsets, when W_{ij} of size 3×3 (A), 7×7 (B), 11×11 (C) and 15×15 (D) are used. For small windows, many original potential changes are not detected anymore, because false potential change components are introduced (A, B). However, larger windows are more robust to misregistration errors, as they help in retaining the original potential change components (C, D).

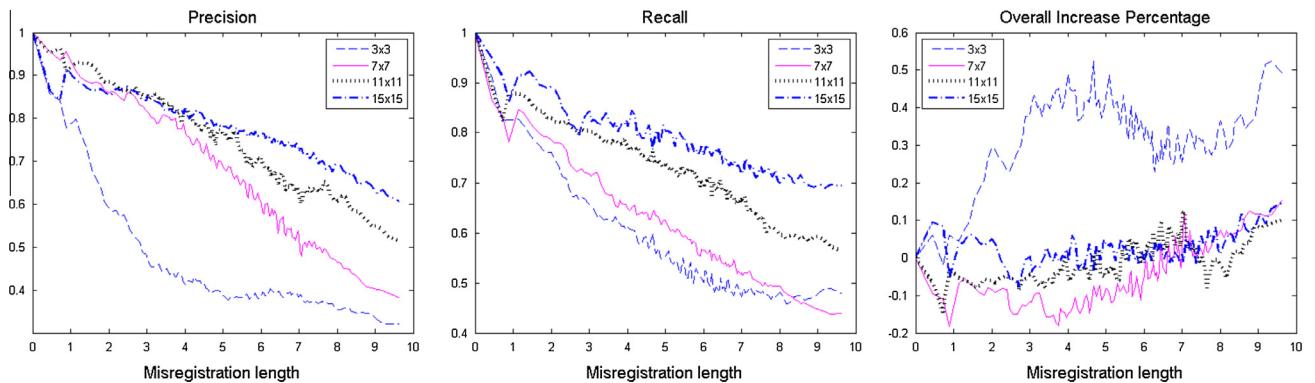


Fig. 20. Precision, Recall and OIP values as a function of the misregistration length. Values were calculated by averaging over the misregistration errors of the same length. For the ease of presentation, a moving average filter was applied to the plots for misregistration errors greater than 1 pixel so that the trend of the curve is observed.

5. Conclusions

The proposed methodology is the first one to detect changes between an image taken from a VHR satellite sensor, which covers a wide area of control and is used as reference of historical data, and an image taken inside this area from a VHR UAV camera. The main challenge in this problem is that misregistration errors are

more evident and affect greatly the accuracy of the change detection. Moreover, the images are taken from different sensors and they are expected to differ in resolution and the level of detail in picturing natural elements.

The processing stages of the proposed methodology (pre-processing, extraction of potential change components, region map generation and evaluation of homogeneous regions) were

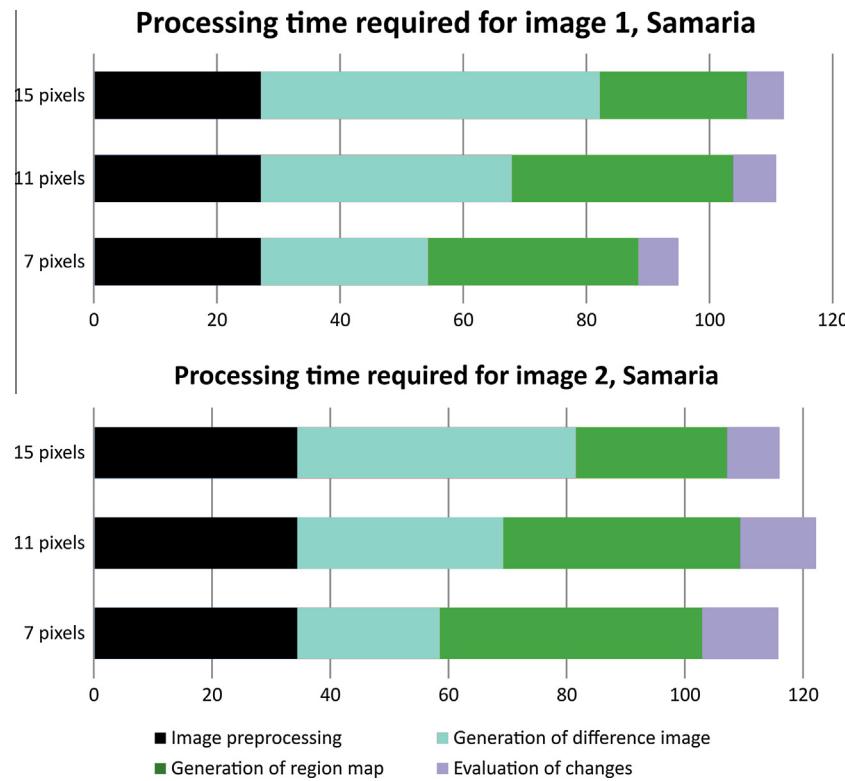


Fig. 21. The individual times (in seconds) required for the four processing stages, while search windows of different sizes are applied, are summed up.

designed so that it addresses many of the challenges encountered in change detection problems. The use of simple pixel descriptors and the search for homologous pixels in a small spatial neighborhood compensates for unavoidable misregistration errors. At the same time, it calculates an initial difference image from which potential change components are extracted. The seeded region growing algorithm defines homogeneous regions on the UAV image, which are a refined version of potential change components. The resulting regions are finally evaluated so that false changes are rejected.

The accuracy in detecting changes and the robustness of the proposed methodology against misregistration errors was assessed via the use of two datasets of UAV images and associated satellite images from different areas, namely "Markopoulos dataset" and "Samaria dataset", and, additionally, via the simulation of misregistration errors. The study areas exhibit significantly different landscape conditions (topography, land use–land cover), while the datasets contain UAV images of different spatial resolution and footprint size. Specifically, the Markopoulos dataset contains images of a small footprint which makes their automatic registration difficult, if not impossible, and also leads to large misregistration errors. Moreover, the Samaria dataset was selected so that to represent challenges such as image misregistration, image distortion and spectral intensity differences to a great extent, in other words, bringing many cases which would produce false changes. The methodology detected successfully the vast majority of changes in the two datasets, which proves that it performs adequately under very different scenarios. Furthermore, the simulation of misregistration errors showed that the pixel descriptors and the search window prevent the difference image from being heavily affected, even if large misregistration errors are introduced. Concerning the potential change components, the precision and

recall measures are notably high. In other words, the vast majority of the true potential change components are still extracted and, in the same time, very few false ones are introduced.

Last but not least, the proposed methodology brings benefits to the specific problem of timely wide area surveillance operations, since (a) it automatically transforms the UAV images and registers them on the reference satellite image, (b) it detects changes in near real time and (c) it rejects as many false changes as possible in a fully unsupervised way, under the assumption that the footprint of the UAV images is not too small. Many of the changes which were detected in our experiment were in fact changes that should be alerted, while a few remained to be subject to re-evaluation by a human expert. The above conclusions make evident that the proposed methodology can be employed for surveillance applications, it can be embedded in a chain of real time acquisition and near real time processing and it can be used effectively to assist operations of institutional surveillance agents.

Acknowledgments

The methodology described in this paper was developed for the purpose of the NSRF project titled "ACRITAS – Advanced Coordination System for Border Surveillance", with the participation of the National Observatory of Athens.

Appendix A

A.1. Region merging

For the sake of presentation convenience, the following definitions are made.

Table 6

The algorithm for merging multiple regions.

```

Set  $Reg_t' = Reg_{UAV}$ 
Set  $T' = \emptyset$ 
Sort  $Reg_t$ 's in ascending order according to  $P_D(Reg_t, Reg^{UAV})$ 
for each  $Reg_t$  examined in this order
  if  $P_C(Reg_t, Reg_t') = \text{true}$ 
     $Reg_t' = Reg_t' \cup Reg_t$ 
     $T' = T' \cup \{t\}$ 
  else
    Perform screening procedure between  $Reg_t'$  and  $Reg_t$ 
  end if
  update the mean value of  $Reg_t'$ 
end for
id = min( $T'$ )
Register  $Reg_t'$  with id
Renumber accordingly the ids in the region map

```

- The similarity distance P_D for two regions, Reg_i and Reg_j , is defined as the Euclidean distance of their mean intensity values, i.e.

$$P_D(Reg_i, Reg_j) = \|\mu_i - \mu_j\|_2 \quad (\text{A.1})$$

- The similarity criterion P_C for two regions, Reg_i and Reg_j , with areas of A_{Reg_i} and A_{Reg_j} respectively, is defined as
 - $P_D(Reg_i, Reg_j) < \delta_2$, or, if this is not true,
 - $P_D(Reg_i, Reg_j) < \delta'_2$, where $\delta'_2 > \delta_2$ and $A_{Reg_i \cap Reg_j} / \min(A_{Reg_i}, A_{Reg_j}) > p_3$,

where δ_2, δ'_2 and p_3 are user defined parameters. The second one indicates that although the regions may not be close enough regarding color intensity values, the overlay percentage may be high and, in this case, they should be merged.

Therefore, we write $P_C(Reg_i, Reg_j) = \text{true}$ (or false) if Reg_i is similar (or dissimilar) to Reg_j .

- ID is the set of unique identification integer numbers that have been attributed to registered regions.
- Reg^{UAV} is a new unregistered region.
- Reg_{id} is an already registered region with identification number $id \in ID$.
- The set Y , that contains the ids of the registered regions that intersect with Reg^{UAV} , each denoted by Reg_y , is defined as

$$Y = \{y : y \in ID, Reg_y \cap Reg^{UAV} \neq \emptyset\} \quad (\text{A.2})$$

- The set F , that contains the ids of the registered regions that intersect with Reg^{UAV} and are dissimilar to it, each denoted by Reg_f , is defined as

$$F = \{f : f \in Y, P_C(Reg_f, Reg^{UAV}) = \text{false}\} \quad (\text{A.3})$$

- The set T , that contains the ids of the registered regions that intersect with Reg^{UAV} and are similar to it, each denoted by Reg_t , is defined as

$$T = \{t : t \in Y, P_C(Reg_t, Reg^{UAV}) = \text{true}\} \quad (\text{A.4})$$

It is obvious that $T \cap F = \emptyset$ and $T \cup F = Y$.

In the simplest case, where there is no intersection between Reg^{UAV} and any of the Reg_{id} 's, Reg^{UAV} acquires $id = \max\{ID\} + 1$, which is registered in region map and we update the ID set as $ID = ID \cup \{id\}$.

In case where Reg^{UAV} intersects with one or more Reg_t 's, the latter are grouped into Reg_f and Reg_t . Firstly, the screening procedure (see Section 2.3.2) is performed on the pixels shared by Reg^{UAV} and each of the Reg_f separately. In this way, Reg^{UAV} discards all of its pixels that should be registered to another region, before proceeding to merging. Then, a merging of the possibly reduced Reg^{UAV} with each of the Reg_t regions is examined. The number of them, $|T|$, defines the complexity of the merging procedure, and the following cases are considered.

- $|T| = 0$ (there is not any Reg_t). In this case, Reg^{UAV} acquires a new id and it is registered on the region map.
- $|T| = 1$ (the exists exactly one Reg_t). In this case, the regions are merged under the existing $id = t$, i.e., $Reg_t = Reg_t \cup Reg^{UAV}$, and Reg_t is augmented.
- $|T| > 1$ (there exist more than one Reg_t 's). This is the most complicated case, as Reg^{UAV} may be merged with more than one Reg_t 's which were not connected before. The new region will replace more than one existing regions and the region map and the set ID must be updated upon their deregistration. Merging of multiple regions is performed by defining a temporary region $Reg_t' = Reg^{UAV}$ and augmenting it sequentially with neighboring similar regions Reg_t 's. The region is accompanied by a temporary set T' , which is used to keep note of regions Reg_t that will be merged in order to form Reg_t' (see also Fig. 6). The merging procedure is described in algorithmic form as shown in Table 6.

We explain that:

- The Reg_{id} 's are initially sorted so that the merging starts from the more to the least similar one.
- The criterion P_C is re-evaluated for each one of the remaining Reg_t 's with the Reg_t' as it has been currently formed. The necessity of this step is justified since every time the Reg_t' merges with one more Reg_t or may dispose of a few pixels, its mean value is updated. Then, as we move to the next one, the criterion may not be satisfied anymore. Note that for computational reasons, the sorting step is not applied after each re-evaluation of P_C .
- The id to be attributed to the region is selected as $id = \min(T')$ from the set T and the rest of the members become temporarily unused. These gaps in ID are filled by shifting the id 's in the region map, and the ones that are then unused are deleted from ID .

References

- Abdel-Aziz, Y.I., Karara, H.M., 1971. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In: Proc. Symposium on Close-Range Photogrammetry. American Society of Photogrammetry, Falls Church, VA, pp. 1–18.
- Al-Khudairy, D.H.A., Caravaggi, I., Glada, S., 2005. Structural damage assessments from Ikonos data using change detection, object-oriented segmentation and classification techniques. Photogramm. Eng. Remote Sens. 71, 825–837.
- Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16.
- Bovolo, F., Bruzzone, L., Marchesi, S., 2009. Analysis and adaptive estimation of the registration noise distribution in multitemporal VHR images. IEEE Trans. Geosci. Remote Sens. 47, 2658–2671.
- Bruzzone, L., Cossu, R., 2003. An adaptive approach to reducing registration noise effects in unsupervised change detection. IEEE Trans. Geosci. Remote Sens. 41, 2455–2465.
- Bruzzone, L., Fernandez, D., 2000. Automatic analysis of the difference image for unsupervised change detection. IEEE Trans. Geosci. Remote Sens. 38, 1171–1182.

- Bruzzone, L., Serpico, S.B., 1997. Detection of changes in remotely-sensed images by the selective use of multi-spectral information. *Int. J. Remote Sens.* 18, 3883–3888.
- Byrne, G.F., Crapper, P.F., Mayo, K.K., 1980. Monitoring land-cover change by principal component analysis of multitemporal landsat data. *Remote Sens. Environ.* 10, 175–184.
- Carlotto, M.J., 1997. Detection and analysis of change in remotely sensed imagery with application to wide area surveillance. *IEEE Trans. Image Process.* 6, 189–202.
- Chen, B., Deng, L., Duan, Y., Huang, S., Zhou, J., 2015. Building change detection based on 3D reconstruction. In: *IEEE International Conference on Image Processing (ICIP)*, Quebec City, 27–30 Sept., 2015, pp. 4126–4130.
- Chen, G., Zhao, K., Powers, R., 2014. Assessment of the image misregistration effects on object-based change detection. *ISPRS J. Photogramm. Remote Sens.* 87, 19–27.
- Collins, J.B., Woodcock, C.E., 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sens. Environ.* 56, 66–77.
- Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* 92, 79–97.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sens.* 25, 1565–1596.
- Coulter, L., Lippitt, C., Stow, D., McCreight, R., 2011. Near real-time change detection for border monitoring. In: *Proc. ASPRS Annual Conference*, Milwaukee, WI, USA, 1–5 May, 2011, pp. 1–5.
- Dai, X., Khorram, S., 1998. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Trans. Geosci. Remote Sens.* 36, 1566–1577.
- Deng, J.S., Wang, K., Deng, H.Y., Qi, G.J., 2008. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* 29, 4823–4838.
- Desclée, B., Bogaert, P., Defourny, P., 2006. Forest change detection by statistical object-based method. *Remote Sens. Environ.* 102, 1–11.
- Fan, B., Du, Y., Zhu, L., Tang, Y., 2010. The registration of UAV down-looking aerial images to satellite images with image entropy and edges. In: *Proc. ICIRA'10 Third International Conference on Intelligent Robotics and Applications*, Shanghai, China, pp. 609–617.
- Fischler, M., Bolles, R., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. Assoc. Comput. Mach.* 24, 381–395.
- Gevaert, C.M., Suomalainen, J., Tang, J., Kooistra, L., 2015. Generation of spectral-temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for agriculture applications. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 8, 3140–3146.
- Gong, P., Ledrew, E.F., Miller, J.R., 1992. Registration-noise reduction in difference images for change detection. *Int. J. Remote Sens.* 13, 773–779.
- Gonzalez, R.C., Woods, R.E., Eddins, S.L., 2002. *Digital Image Processing using Matlab*, second ed. Prentice Hall, Upper Saddle River, NJ.
- Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: from pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* 80, 91–106.
- Im, J., Jensen, J.R., 2005. A change detection model based on neighborhood correlation image analysis and decision tree classification. *Remote Sens. Environ.* 99, 326–340.
- Im, J., Jensen, J.R., Tullis, J.A., 2008. Object-based change detection using correlation image analysis and image segmentation. *Int. J. Remote Sens.* 29, 399–423.
- Kontoes, C.C., 2008. Operational land cover change detection using change vector analysis. *Int. J. Remote Sens.* 29, 4757–4779.
- Liu, J., Li, P., Wang, X., 2015. A new segmentation method for very high resolution imagery using spectral and morphological information. *ISPRS J. Photogramm. Remote Sens.* 101, 145–162.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60, 91–110.
- Lu, D., Mausel, P., Brondizio, E., Moran, E., 2004. Change detection techniques. *Int. J. Remote Sens.* 25, 2365–2401.
- Lunetta, R.S., Knight, J.F., Ediriwickrema, J., Lyon, J.G., Worthy, L.D., 2006. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sens. Environ.* 105, 142–154.
- Marchesi, S., Bovolo, F., Bruzzone, L., 2010. A context-sensitive technique robust to registration noise for change detection in VHR multispectral images. *IEEE Trans. Image Process.* 19, 1877–1889.
- Mas, J.F., 1999. Monitoring land-cover changes: a comparison of change detection techniques. *Int. J. Remote Sens.* 20, 139–152.
- Mendoza-Schrock, O., Patrick, J.A., Blasch, E., 2009. Video image registration evaluation for a layered sensing environment. In: *Proc. IEEE Aerospace & Electronics Conference (NAECON)*, Dayton, OH, 21–23 July, 2009, pp. 223–230.
- Nielsen, A.A., 2007. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Trans. Image Process.* 16, 463–478.
- Nielsen, A.A., Conradsen, K., Simpson, J.J., 1998. Multivariate alteration detector (MAD) and MAF post-processing in multispectral, bi-temporal image data: new approaches to change detection studies. *Remote Sens. Environ.* 64, 1–19.
- Oh, J., Toth, C.K., Grejner-Brzezinska, D.A., 2011. Automatic georeferencing of aerial images using stereo high-resolution satellite images. *Photogramm. Eng. Remote Sens.* 77, 1157–1168.
- Poynton, C.A., 2003. *Digital Video and HDTV: Algorithms and Interfaces*. Morgan Kaufmann Publishers, San Francisco, CA, p. 291.
- Qin, R., 2014. An object-based hierarchical method for change detection using unmanned aerial vehicle images. *Remote Sens.* 6, 7911–7932.
- Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B., 2005. Image change detection algorithms: a systematic survey. *IEEE Trans. Image Process.* 14, 294–307.
- Rango, A., Laliberte, A., Steele, C., Herrick, J.E., Bestelmeyer, B., Schmugge, T., Roanhorse, A., Jenkins, V., 2006. Using unmanned aerial vehicles for rangelands. *Environ. Pract.* 8, 159–168.
- Rosin, P.L., 2001. Unimodal thresholding. *Pattern Recogn.* 34, 2083–2096.
- Rosin, P.L., Ioannidis, E., 2003. Evaluation of global image thresholding for change detection. *Pattern Recogn. Lett.* 24, 2345–2356.
- Shi, J., Wang, J., Xu, Y., 2011. Object-based change detection using georeferenced UAV images. *Int. Arch. Photogramm., Remote Sens. Spatial Inform. Sci.* 38 (1/C22), 177–182.
- Singh, A., 1989. Digital change detection techniques using remotely sensed data. *Int. J. Remote Sens.* 10, 989–1003.
- Singh, K.K., Pal, K., Nigam, M.J., 2012. Shadow detection and removal from remote sensing images using NDI and morphological operators. *Int. J. Comput. Appl.* 42, 37–40.
- Sobel, I., Feldman, G., 1973. A 3×3 isotropic gradient operator for image processing. In: Duda, R., Hart, P. (Eds.), *Pattern Recognition and Scene Analysis*, second ed. John Wiley and Sons, NY, pp. 271–272.
- Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A., 2001. Classification and change detection using landsat TM data: when and how to correct atmospheric effects? *Remote Sens. Environ.* 75, 230–244.
- Stow, D.A., 1999. Reducing the effects of misregistration on pixel-level change detection. *Int. J. Remote Sens.* 20, 2477–2483.
- Tian, J., Cui, S., Reinartz, P., 2014. Building change detection based on satellite stereo imagery and digital surface models. *IEEE Trans. Geosci. Remote Sens.* 52, 406–417.
- Theodoridis, S., Koutroumbas, K., 2009. *Pattern Recognition*, fourth ed. Academic Press, Burlington, MA, pp. 498–501.
- Townshend, J., Justice, C., Gurney, C., McManus, J., 1992. The impact of misregistration on change detection. *IEEE Trans. Geosci. Remote Sens.* 30, 1054–1060.
- Wang, H., Ellis, E.C., 2005. Image misregistration error in change measurements. *Photogramm. Eng. Remote Sens.* 71, 1037–1044.
- Wang, Z., Jensen, J.R., Im, J., 2010. An automatic region-based image segmentation algorithm for remote sensing applications. *Environ. Model. Softw.* 26, 1149–1165.
- Wang, Q., Zhang, X., Wang, Y., Chen, G., Dan, F., 2013. The design and development of object-oriented UAV image change detection system. *Geoinformat. Resour. Manage. Sustain. Ecosyst.* 398, 33–42.
- Wong, A., Clausi, D.A., 2007. ARRSI: automatic registration of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* 45, 1483–1493.
- Zhang, X., Xiao, P., Feng, X., Wang, J., Wang, Z., 2014. Hybrid region merging method for segmentation of high-resolution remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 98, 19–28.