

## **Building Change Detection using Semantic Segmentation on Analogue Aerial Photos**

**Evangelos MALTEZOS, Charalabos IOANNIDIS, Anastasios DOULAMIS and Nikolaos DOULAMIS, Greece**

**Key words:** Change detection; buildings; analogue photos; point cloud; convolutional neural networks

### **SUMMARY**

Spatio-temporal urban monitoring in large-scale is critical in various engineering applications. Periodical building change detection is necessary for applications such as urban and rural planning, updating 3D cadastral objects and databases, identification of informal settlements and constructions, 3D city modeling and valuation purposes. Automatic building change detection is a research topic of high interest including several challenges referring to radiometric, geometric and atmospheric corrections, data registration and multi-modal data fusion as well as difficulty in using images of complex building structures derived from different viewpoints. Deep learning techniques have received increased attention for achieving satisfying results in many classification problems. Semantic segmentation is a pixel-wise classification of images by implementing a deep neural network scheme such as Convolutional Neural Networks (CNNs) under a supervised setting. This paper presents an one-shot building change detection procedure using semantic segmentation on scanned analogue aerial photos. An augmented time period feature band vector is firstly created by fusing 3D geospatial information, that is a 3D point cloud extracted from Dense Image Matching (DIM), with the corresponding orthoimage. A small training set for the classes of “new buildings”, “unchanged buildings” and “other” is created from the same dataset. The training set and the augmented time period feature band vector are fed as input into a CNN, which is responsible for the semantic segmentation. The “new buildings” and the “unchanged buildings” masks are then processed to eliminate noise taking into account the spatial coherency properties. To verify the applicability, usability and functionality of the employed procedure, two complex and real-life urban study areas in Greece (Keratea and Perissa) with various building structures, pixel resolutions and types of data are selected. The CNN results evaluated via pixel-wise success rates of completeness, correctness and quality, and compared with conventional and other “shallow” learning paradigms, such as Support Vector Machines (SVMs). The derived results show the effectiveness of the proposed deep scheme.

# Building Change Detection using Semantic Segmentation on Analogue Aerial Photos

Evangelos MALTEZOS, Charalabos IOANNIDIS, Anastasios DOULAMIS and Nikolaos DOULAMIS, Greece

## 1. INTRODUCTION

### 1.1 Previous Works

Spatio-temporal urban monitoring in large-scale is critical in various engineering and civilian applications such as identification of informal settlements and constructions, disaster management, population estimation, urban and rural planning, transportation, augmented reality, 3D city modeling and housing value (Karantzalos, 2014). All these subjects are actively investigated through 2D or 3D change detection techniques using remote sensing data. The wider scope of the automatic change detection is a research topic in over two decades. In the literature, two types of approaches called *chance enhancement* and *from-to* are proposed for the automatic change detection. The *chance enhancement* approach indicates only the position and the magnitude of the change. In contrary, the *from-to* approach indicates not only the position and the magnitude of the change but also its nature (e.g. building changes, vegetation changes, etc.). However, the scientific methodologies vary with the desired objects of interest (e.g. automatic building change detection) towards the challenge of the increasingly greater demands for accurate and cost-effective applications. Towards this end, a significant amount of research is still, nowadays, focusing on the design, development and validation of novel automated change detection procedures using model-based, graph-based or machine learning techniques (Radke et al., 2005; Karantzalos, 2014; Vakalopoulou et al., 2016). Updating of 3D cadastral objects and databases is also an interesting application that automatic building change detection techniques have considered over the years (Pédrinis et al., 2015; Koeva and Oude Elberink, 2016; Pedrinis and Gesquière, 2017; Maltezos et al., 2017). Detection of building changes from multi-temporal earth observation data, still, remains a challenge due to i) radiometric and atmospheric correction and calibration, ii) geometric correction and data registration, iii) multi-modal data fusion, iv) inherent artifacts of the used data, v) differences in viewpoint and surrounding environment and, vi) complex shape and size of the building structures (Rottensteiner et al., 2014; Gilani et al., 2016).

Dependent on the data source employed, building change detection techniques can be categorized into three groups: i) the ones that use airborne or satellite imagery data, ii) the ones that exploit three-dimensional information and iii) those that combine both of data sources. The first group fully exploits image information as well as available additional information associated with indices (e.g. vegetation indices), band ratios, etc (Bourdis et al., 2011; Barazzetti et al., 2015). The second group implements direct 3D point cloud processing either derived from Dense Image Matching (DIM) techniques or by Light Detection And

Ranging (LIDAR) data (Dini et al., 2012; Xu et al., 2013; Nebiker et al., 2014; Awrangjeb et al., 2015) or a combination of the two (Du et al., 2016). The third group fuses image and 3D information to improve the building change detection accuracy (Rottensteiner, 2007; Ioannidis et al., 2010; Tian and Reinartz, 2013; Sarp et al., 2014; Qin et al., 2015; Hron and Halounova, 2015). The last group is considered appropriate for detecting 2D and 3D building changes associated with informal constructions. According to the existing legal framework, ‘informal construction’ in Greece is characterized by any construction that i) exists without a building permit, ii) has any kind of excess or violation to the building permit and iii) is in violation of any valid urban and spatial regulation, regardless of the existence of a building permit (Ioannidis et al., 2010).

The most common problems of using only image information for building change detection are the presence of shadows and the fact that urban objects usually present similar pixel values (e.g., building rooftops vs. roads, or vegetation vs. vegetation on building rooftops). On the other hand, the use of only 3D data provides often incorrect assignments of building changes when vegetation is significantly grows up between two time periods. To remove low vegetation or insignificant *man-made* objects or cars, height thresholds (e.g. above 2 m) are used. However, a confusion of vegetation with similar height values with buildings still exists. Furthermore, areas of small buildings present low building change detection accuracy due to local under-sampling or mismatches using LIDAR data or DIM point clouds respectively. To overcome these issues, the recent research outcomes focus on fusing *Red, Green, Blue* (RGB) images and additional information, e.g., *Near Infrared band* (NIR) as well as 3D information from LIDAR for each time period (Beumier and Idrissa, 2012; Peng and Zhang, 2016). However, in this case, several research challenges arise such as i) the co-registration of the data coming from different sources, ii) the additional acquisition and processing cost of the multi-modal information, and iii) the fact that LIDAR data or additional information are not always available (e.g. Cadastral agencies commonly have scanned analogue aerial photos or *Red, Green, Blue, Near Infrared* (RGBNIR) aerial images).

## 1.2 Our Contribution

This paper presents a procedure of one-shot building change detection framework through semantic segmentation using scanned analogue aerial photos. It has been proven in machine learning research that deep learning algorithms, such as Convolutional Neural Networks (CNNs), are able to emulate human brain operation (Yu and Deng, 2011). CNNs are an extension of Artificial Neural Networks (ANNs), mainly used in image analysis, which simplifies the feature extraction process (Crommelinck et al., 2016). Also, have been shown to be very efficient for semantic segmentation for many remote sensing applications (Makantasis et al., 2015; Marmanis et al., 2015; Vakalopoulou et al., 2015; Sakurada et al., 2015; Amin et al., 2016; Fujita et al., 2017). Deep CNN classifiers present advantages compared to conventional shallow architectures or model-based approaches (Doulamis and Doulamis, 2012). Semantic segmentation is a pixel-wise classification of images by implementing a deep neural network technique under a supervised setting (Amit et al., 2015).

The main advantage of the machine learning approaches is that they are flexible and data driven methods, requiring only training samples (this also makes machine learning ideal for non-expert users) to generalize well the building change properties and thus to perform an accurate semantic segmentation. On the contrary, model-based schemes consist of many parameters needed to be tuned for each study area. Therefore supervised learning paradigms provide higher generalization capabilities.

On the other hand, scanned analogue aerial photos are still a source of information for surveyor engineers and Urban Planning and Cadastral agencies as historically record the urban land-uses and development of cities. 2D or 3D building change detection through i) direct ground-field measurements and inspections, ii) photointerpretation, and iii) stereoscopic observations in digital photogrammetric workstations, may be expensive, laborious and time consuming. Compared to the aforementioned traditional methods, the proposed methodology provides a cost-effective mapping of building changes. The main contributions are:

- The full exploitation of one type of data (scanned analogue aerial photos) that have i) poor and/or limited spectral capabilities, ii) medium quality and iii) no additional information.
- The fusion of scanned analogue aerial photos with modern derived remote sensing data, e.g., derived from unmanned aerial vehicles (UAVs).
- The fusion of image and 3D information for each time period dataset in order to extract an augmented time period feature band vector. The 3D information is expressed as a 3D point cloud extracted from innovative techniques such DIM, while the image information is expressed as the corresponding orthoimage.
- The use of a sophisticated deep learning model in order to indentify, in one-shot, building changes in terms of the classes of “new buildings”, “unchanged buildings” and “other”.

## **2. BUILDING CHANGE DETECTION USING CNN**

The workflow of the processing procedure is shown in Figure 1. Firstly, an augmented time period feature band vector is created containing the normalized Digital Surface Model (nDSM) and the orthoimage of each time period. Then a training set for the classes of “new buildings”, “unchanged buildings” and “other” is created from the same dataset. The training set and the augmented time period feature band vector are fed as input into a CNN, which in turn is responsible for the semantic segmentation. Finally, a post process is carried out in order to extract the final “new buildings” and “unchanged buildings” masks.

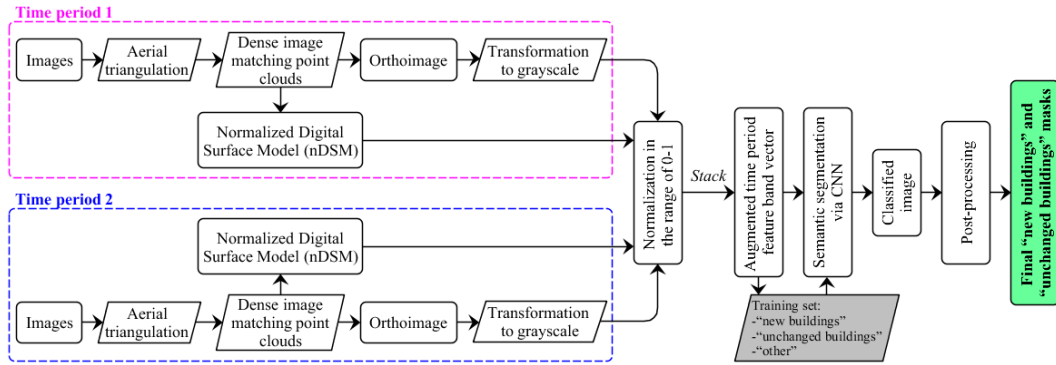


Figure 1. Workflow of the processing procedure

## 2.1 Augmented Time Period Feature Band Vector for Building Change Detection

To increase the building change detection accuracy, fusion of image and 3D information is considered for each time period dataset. To this end, an augmented time period feature band vector, denoted as  $I_t$ , is created which is fed as input into the deep learning framework. The  $I_t$  contains the image and the 3D information for both time periods in terms of bands stack. Let us denote as  $I(x,y) \equiv I(q)$  an image, expressing a transformation  $I:K^1 \rightarrow K^n$  where  $n$  denotes the  $n^{\text{th}}$  image band stacks. For example for an RGB image  $n=3$  while for a grayscale image  $n=1$ . Each image band of each time period can be considered as an element of the augmented time period feature band vector for building change detection. To avoid confusions during the classification performance, the normalized Digital Surface Model (nDSM) is extracted from the DIM/DSM of each time period. Firstly the cloth simulation technique (Zhang et al., 2016) was performed to detect the ground points from the DIM point cloud. Then, a closest point method was implemented between the detected ground points and the DIM point cloud setting the height values of the ground points equal to zero and shifting the remaining respectively. A rasterization to the estimated point cloud is then takes place to create the nDSM. The nDSM of each time period, denoted as  $I_h$ , is used as additional element at the  $I_t$  representing the 3D information. It should be mentioned that each feature band is normalized in the range of 0-1 in order to increase the classification accuracy. In this study, the image  $I$  is expressed by the generated orthoimage of each time period transformed to a grayscale band (see Figure 1 and Table 1). The augmented time period feature band vector of a dataset is given as

$$I_t = \left[ I^{\text{time period 1}} \quad I_h^{\text{time period 1}} \quad I^{\text{time period 2}} \quad I_h^{\text{time period 2}} \right] \quad (1)$$

## 2.2 CNN Structure

A CNN classifier has two main components; the convolutional layer and the classification layer. A convolutional layer is essentially a feature extractor that employs convolution filters (i.e., transformations) to the input data, e.g., in this paper to the augmented time period feature band vector. These extracted features are able to optimize the classification performance.

Spatial coherency is an important element of the transformations involved in the convolutional layer. This is an important property of a deep CNN model since spatial characteristics significantly affect building change detection accuracy. The aim of the classification layer is actually a supervised learning paradigm with a capability of transforming the inputs from the convolutional layer to desired outputs, i.e., the labeled classes. Therefore, a CNN classifier, instead of a shallow machine learning method, first filters the input data in a way to maximize the classification accuracy, and then performs the classification. The output of the CNN is a classified image on a pixel-level including information associated with the label of each class. In this study the three desirable classes are i) new buildings, ii) unchanged buildings and iii) other.

### 2.2.1 Implementation details

We assume that the label of the class of the pixel  $q$  at location  $(x,y)$  must be the same to the label of the patch centered at location  $(x,y)$ . Thus, the corresponding neighborhood  $s$  was set equal to 5. The first layer of the proposed CNN is a convolutional layer with  $C1=3 \times n$  trainable filters of dimension  $3 \times 3$ , where  $n$  is the dimensionality of the augmented time period feature band vector  $I_t$ . The first convolutional layer is followed by a second convolutional layer with  $C2=3 \times C1$  trainable filters of dimension  $3 \times 3$ . The second convolutional layer delivers a vector with  $C2$  elements, which is fed as input to a Multiple Layer Perceptron (MLP) classifier (Makantasis et al., 2015). Concerning the training phase, the standard back-propagation algorithm was used. We split the labelled dataset into three mutually exclusive subsets; the training, the validation and the test dataset. The training set is used to train the network, the validation set to evaluate the CNN model during the training and the test set to benchmark the ability of the CNN model in data outside the training and validation set. In this study the training and the validation dataset were selected as 80% and 20% respectively. The implementation of the deep learning framework was made using the MATLAB (<https://in.mathworks.com/products/matlab.html>) computing environment and the PYTHON programming language (<https://www.python.org/>). A laptop computer (CPU with 2.6 GHz and 16GMemory) with a graphic card NVIDIA GeForce GTX 960M (640 CUDA cores) was used to process the test datasets.

## 2.3 Post-Process

To eliminate possible noise of the semantic segmentation results, a post processing to the output of the classifier is used taking account spatial coherency properties. Initially, only the pixels associated with the classes of “new buildings” and “unchanged buildings” are extracted creating the corresponding “new buildings” and “unchanged buildings” masks. Then, a morphological median operator is implemented at each mask. Further, area criteria are considered depending on the pixel resolution (i.e., the scale) of the used image datasets. The implementation of the area thresholding process is to filter small insignificant buildings or *man-made* structures.

### 3. EXPERIMENTAL RESULTS

#### 3.1 Study Areas and Test Datasets

To verify the applicability, utility and functionality of the proposed deep framework, two complex and real-life urban study areas with various kinds of buildings structures, pixel resolutions and types of data are selected. The selected study areas are located at Keratea area and Santorini island (Perissa area) in Greece (Figure 2). The Keratea is a sub-urban area with small detached houses that consisted of sloped or flat surfaces but also has sporadically some, often high, vegetation. Santorini contains mainly hotel resorts with continuous and complex building structural system with constant depth discontinuities, small extensions or additions of major buildings as well as several objects on the rooftops such solar water heaters, etc. The type of vegetation is characterized as moderate. It should be noted that both study areas have dynamic changes through the recent years associated with new buildings and not with buildings that do not exist anymore (i.e., demolished buildings). Since the proposed methodology is a data driven and from-to approach, the available classes of “new buildings”, “unchanged buildings” and “other” are considered sufficient for the examined study areas. In case that the study areas had demolished buildings, an additional class named “demolished buildings” and its corresponding training samples should be created.

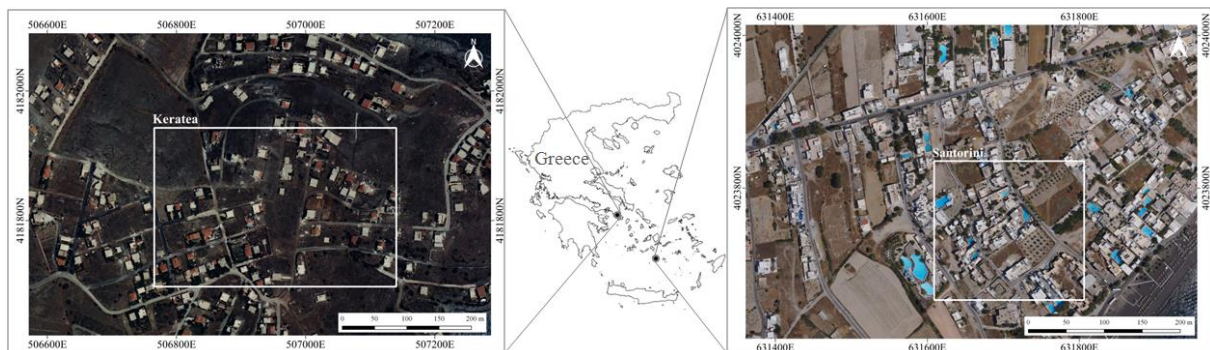


Figure 2. Study areas

		Study areas			
		Keratea		Santorini	
		1995 (Time period 1)	2001 (Time period 2)	1995 (Time period 1)	2012 (Time period 2)
Flying parameters	Type of images	3 Scanned analogue photos Scale 1:12000	3 Scanned analogue photos Scale 1:10000	1 Scanned analogue stereo-pair Scale 1:5000	59 Multiple/Digital images
	Focal length	153.35 mm	153.35 mm	305.13 mm	28.97 mm
	Forward overlap	70%	70%	60%	60%

	Side lap				20%
	Pixel resolution	Scan resolution (21 $\mu$ m) $\rightarrow$ 25 cm	Scan resolution (21 $\mu$ m) $\rightarrow$ 21 cm	Scan resolution (16 $\mu$ m) $\rightarrow$ 8 cm	Pixel size (4.78 $\mu$ m) $\rightarrow$ 6 cm
	Spectral bands	R/G/B	R/G/B	Grayscale	R/G/B
	Ground Control Points	6	6	6	10
	Aerial triangulation accuracy	<1pixel	<1pixel	<1pixel	<1.5pixel
Height (3D) information	Software for DIM	Erdas Imagine (Imagine UAV)	Erdas Imagine (Imagine UAV)	Erdas Imagine (eATE)	Erdas Imagine (Imagine UAV)
	GSD of the DIM/DSM	21 cm	21 cm	7 cm	7 cm
	Software for nDSM	CloudCompare	CloudCompare	CloudCompare	CloudCompare
Image information	Software for orthoimages	Erdas Imagine (Imagine UAV)	Erdas Imagine (Imagine UAV)	Erdas Imagine (Imagine Photogrammetry)	Erdas Imagine (Imagine UAV)
	GSD of the orthoimages	21 cm	21 cm	7 cm	7 cm
Input data at the CNN classifier	Block tile (pixels) of the area of interest	1166 $\times$ 1791		2086 $\times$ 2046	
	Augmented time period feature band vectors ( $I_i$ )	<b>Grayscale<sub>1995</sub><sup>1*</sup></b> <b>+nDSM<sub>1995</sub></b> <b>+Grayscale<sub>2001</sub><sup>2*</sup></b> <b>+nDSM<sub>2001</sub></b> ↓ Composition of the $I_i$ : 4 bands	<sup>1*</sup> The $R_{1995}/G_{1995}/B_{1995}$ transformed to one grayscale band, i.e., $Grayscale_{1995}$  <sup>2*</sup> The $R_{2001}/G_{2001}/B_{2001}$ transformed to one grayscale band, i.e., $Grayscale_{2001}$	<b>Grayscale<sub>1995</sub></b> <b>+nDSM<sub>1995</sub></b> <b>+Grayscale<sub>2012</sub><sup>3*</sup></b> <b>+nDSM<sub>2012</sub></b> ↓ Composition of the $I_i$ : 4 bands	<sup>3*</sup> The $R_{2012}/G_{2012}/B_{2012}$ transformed to one grayscale band, i.e., $Grayscale_{2012}$
	Labeled training set of classes	New buildings:1 Unchanged buildings:2 Other:3		New buildings:1 Unchanged buildings:2 Other:3	

Table 1. Flying parameters and supplementary information about the selected datasets

Table 1 shows the flying parameters, aerial triangulation accuracy and supplementary information about the selected datasets of Keratea and Santorini. For the case of Keratea, overlapped scanned analogue RGB aerial images of two time periods (1995 and 2001) were used in order to extract the DIM/DSMs and the corresponding RGB orthoimages. For the case of Santorini, a scanned analogue greyscale aerial stereo-pair (time period 1995) and overlapped RGB digital aerial images derived from a UAV (time period 2012) were used to extract the DIM/DSMs and the corresponding orthoimages. Figure 3 shows the extracted coloured DIM point clouds and the generated orthoimages of the areas of interest. The Ground Sample Distances (GSDs) of each dataset was selected the same for each time period



for the proper composition of the augmented time period feature band vector. The DIM point clouds were extracted using the Erdas Imagine software (<http://www.hexagongeospatial.com/products>) while the nDSMs were extracted through CloudCompare (<http://www.danielgm.net/cc/>).

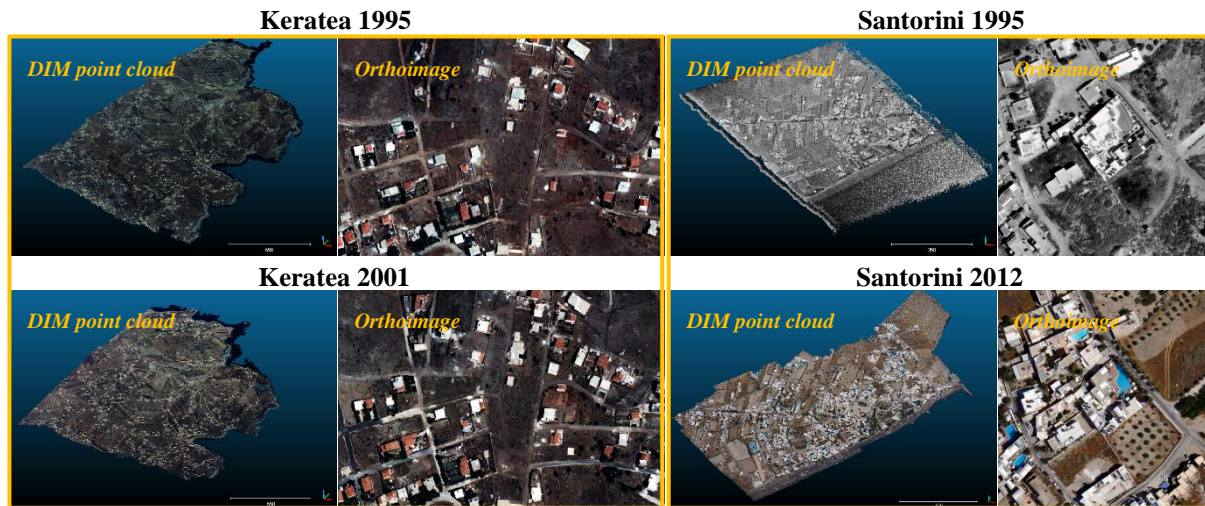


Figure 3. DIM point clouds and orthoimages of the study areas

### 3.2 Training Sets

Three classes are used for the semantic segmentation: “new buildings”, “unchanged buildings” and “other”. The collecting training set varies depending on the complexity and pixel resolution of the image scene. From the available data, we collect a training set of representative samples of the three classes. This set consists of sample polygons for data of each class (Figure 4). The class “new buildings” contains buildings that were constructed at the second time period while the class “unchanged buildings” contains buildings that remain unchanged at both time periods. The class “other” contains all the other urban objects (either they have undergone changes over the time or not) that are not correspond with buildings. Such urban objects are ground, high and low vegetation, cars, insignificant structures, etc). To improve the semantic segmentation process, shadowed areas of each class are also included to the corresponding training sample polygons. In addition, the training sample polygons are spatially created to improve representativity of each class and take into account the spatial coherency of the content. Only a very small percentage of the total data in the dataset, denoted as  $P_{sum}$ , has been included in the learning set.  $P_{sum}$  equals to  $P_{sum} = P_{nb} + P_{ub} + P_{oth}$  where subscripts “nb”, “ub” and “oth” refers to the classes of “new buildings”, “unchanged buildings” and “other” respectively. The exact values selected for the learning set are shown in Table 2.



Figure 4. Collected training sets for the areas of interest

	Class: New buildings	Class: Unchanged buildings	Class: Other	$P_{sum}$
Study area	$P_{nb}$	$P_{ub}$	$P_{oth}$	
Keratea	0.04%	0.07%	0.21%	0.33%
Santorini	0.30%	0.25%	0.79%	1.35%

Table 2. Learning set ( $P_{sum}$ ) expressed in percentage of the total data in the dataset

### 3.3 Semantic Segmentation and Evaluation

#### 3.3.1 Semantic Segmentation and Post-process

As we stated above, the output of the CNN is a classified image on a pixel-level including information associated with the label of each class. Figure 5 shows the semantic segmentation results of the two study areas superimposed to the corresponding orthoimages of the second time period.



Figure 5. Semantic segmentation results applying the CNN

It should be noted that the same CNN was implemented for the two study areas. Concerning the computational time: Keratea was learning in 3 minutes while the test was executing in 30 seconds and Santorini was learning in 15 minutes while the test was about 3 minutes. After the semantic segmentation, the post-process discussed in Section 2.3 is implemented to extract the “new buildings” and “unchanged buildings” masks. The morphological median operator was implemented twice with a window size of  $7 \times 7$  at each mask. The area thresholds that were used were  $20 \text{ m}^2$  and  $4 \text{ m}^2$  for the case of the Keratea and the Santorini datasets respectively.

### 3.3.2 Evaluation and comparisons

To quantitatively evaluate the extracted “new buildings” and “unchanged buildings” masks, the success rates of completeness ( $C_m$ ), correctness ( $C_r$ ) and quality ( $Q$ ) are used using reference data. According to the ISPRS guidelines (Rutzinger et al., 2009)

$$C_m (\%) = \frac{\|TP\|}{\|TP\| + \|FN\|} \cdot 100; \quad C_r (\%) = \frac{\|TP\|}{\|TP\| + \|FP\|} \cdot 100; \quad Q (\%) = \frac{\|TP\|}{\|TP\| + \|FP\| + \|FN\|} \cdot 100 \quad (2)$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. This paper aims to assess the performance of automatic techniques, such as machine learning for the detection of building changes from remote sensing and photogrammetric products. Compared to topographic maps, such products have different characteristics concerning the building structures, i.e., differences in viewpoint, occlusions, 2.5D elevation models, etc. Thus, the reference data did not acquire from ground-field surveying processes but were manually digitized using the orthoimages and the corresponding DIM/DSMs.

The effectiveness of the adopted CNN is also tested in comparison to Support Vector Machine (SVM)-based methods (Mountrakis et al., 2011). More specifically, the Linear kernel SVM and the Radial Base Function (RBF) kernel SVM classifiers were implemented. For fairness reasons, the same  $I_t$  and labeled training set for each study area as well as the same post-process were considered. Figures 6 to 9 show the evaluation results of the extracted “new buildings” and “unchanged buildings” masks of each study area applying the CNN, the Linear SVM and the RBF SVM; the corresponding per-area accuracies in terms of completeness ( $C_m$ ), correctness ( $C_r$ ) and quality (Q) rates are shown in Table 3.

Commonly, the DIM point clouds present surface roughness and local deformations at the boundaries of the objects due to poor performance of the stereo image matching algorithms (mismatches) (Maltezos and Ioannidis, 2015). The main problems during the stereo image matching procedure are associated with the geometry of each stereo-pair as well as with radiometric differences (e.g., shadows and texture-less areas) and complexity of the scene (e.g., intense relief, depth discontinuities/steep slopes and repetitive pattern of objects). The geometry of a stereo-pair (i.e., the height to base ratio) is fully connected with the presence of occlusions as well as with the spatial blend intersection in the 3D space (e.g., a small height to base ratio increases the uncertainty  $\sigma_z$ ). The performance of the DIM process is also highly affected from the quality of the used imagery. Scanned analogue aerial photos usually present radiometric deformations and noise (speckles) as well as suffer from paper extensions due to extensive use. Since scanned analogue aerial photos were used, at least for one time period, the extracted DIM point clouds present significant local mismatches. The situation is aggravated at building change detection since two time period dataset are used. Such problems negatively affect not only the extracted DIM/DSM (e.g., presence of excessive interpolations during the rasterization) but also the generation of the corresponding orthoimages (e.g., deformed objects in the scene).

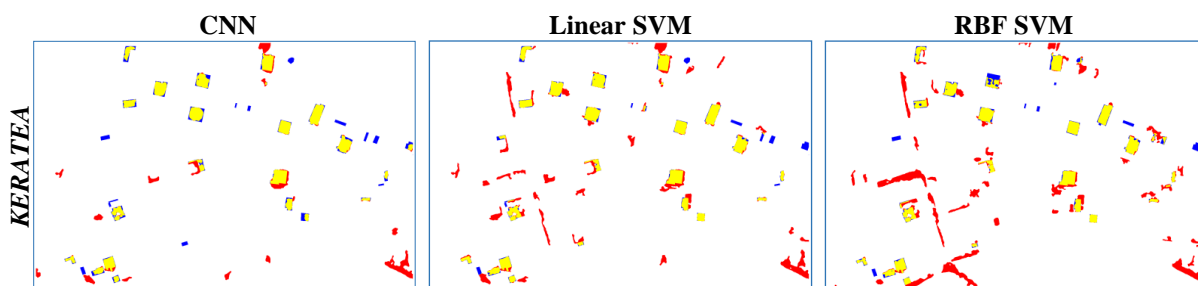


Figure 6. Pixel-wise results of class “new buildings” ■ TP ■ FP ■ FN

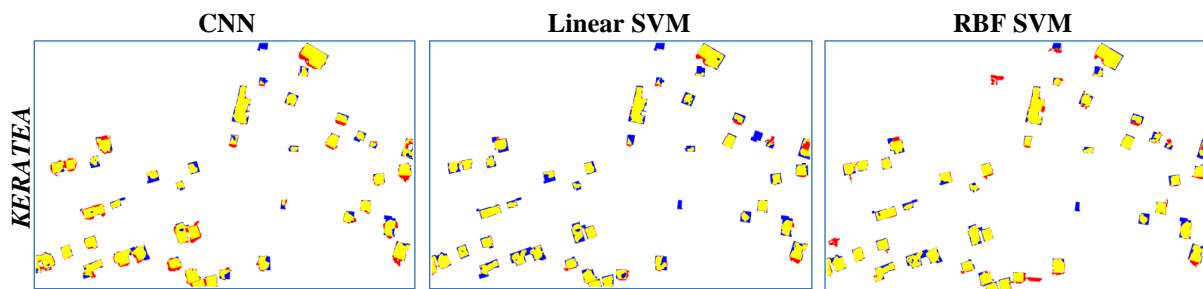


Figure 7. Pixel-wise results of class “unchanged buildings” ■ TP ■ FP ■ FN

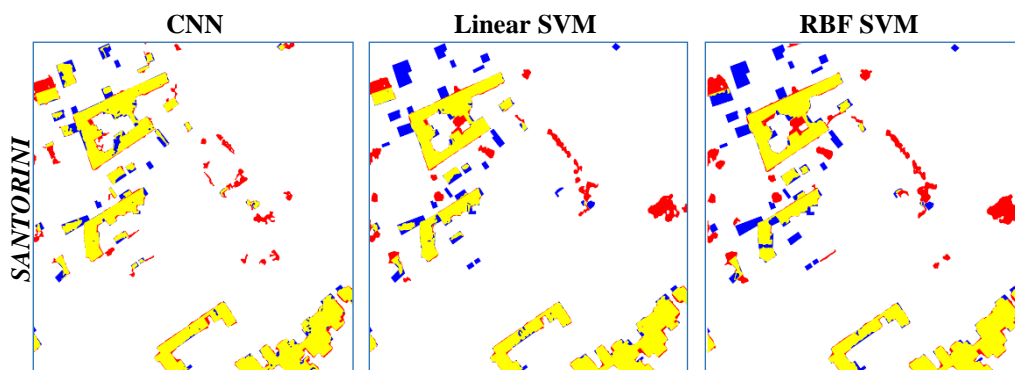


Figure 8. Pixel-wise results of class “new buildings” ■ TP ■ FP ■ FN

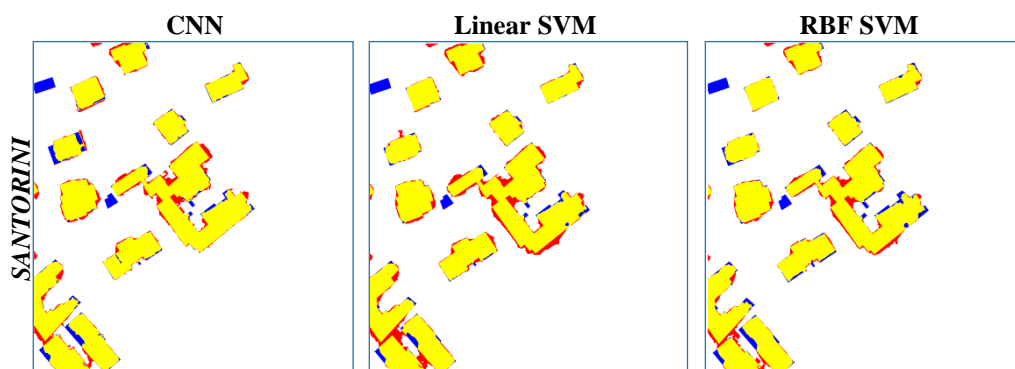


Figure 9. Pixel-wise results of class “unchanged buildings” ■ TP ■ FP ■ FN

Further, inherent artifacts of the used data and differences in viewpoint are also affect the classification performance. Characteristic examples are shadowed areas, relief's diversions of high buildings due to different off-nadir angles of the aquired images, pixel resolution and misregistration effect (e.g., remaining systematic errors). The FN of the “unchanged buildings” are associated with mismatches during the DIM (i.e., insufficient 3D reconstruction of a building at one or both time periods) and incorrect detection of the ground points during the calculation of the nDSM (e.g., incorrect assignment of buildings as ground at one or both time periods). The FP of the “new buildings” contain i) pixels that assigned as

“new buildings” while should be assigned as “unchanged buildings”, ii) artifacts coming from occlusions or mismatches during the DIM at one time period expressed as excessive interpolations during the rasterization of the nDSMs (Du et al., 2016) and, iii) insignificant objects (e.g. umbrella arrays) that present similar characteristics (height and image pixel values) with building changes. A representative case of FP of the “new buildings” is shown in Figure 10. The FN of the “new buildings” mostly come from insufficient 3D reconstruction of a building during the DIM of the second time period or by lack of representative training samples. The FP of the “unchanged buildings” are due to excessive interpolations during DIM and remaining misregistration errors between the datasets of both time periods.

As is observed in Table 3, CNN provides better classification performance compared to the results provided by the SVM classifiers. In particular, CNN achieved balanced average quality (Q) rates of “new buildings” (64.6%) and “unchanged buildings” (77.0%). Instead, the SVM classifiers presented unbalanced quality (Q) rates achieving low classification performance to “new buildings”. This shows the applicability and the functionality of the CNN to detect building changes in terms of various kinds of buildings structures, pixel resolutions and types of data. It should be mentioned that the total computational time for each study area using the SVM classifiers was less than 1 minute. Although the CNN weak in computational complexity compared to the SVM classifiers, the use of a better CPU and more RAM capacity can significantly speed the processing. Furthermore, there are methods that accelerate the learning phase (Doulamis and Voulodimos, 2016). To overcome misclassification problems and therefore to achieve higher success rates, more representative training samples can be utilized to optimize class predictions. Furthermore, additional information coming from GIS or aerial/satellite data (with a proper pixel resolution) for both time periods can be used. Also, additional morphological criteria can be performed to eliminate FP assignments (Du et al., 2016).

Study area	Classes	CNN			Linear SVM			RBF SVM		
		C <sub>m</sub> (%)	C <sub>r</sub> (%)	Q (%)	C <sub>m</sub> (%)	C <sub>r</sub> (%)	Q (%)	C <sub>m</sub> (%)	C <sub>r</sub> (%)	Q (%)
Keratea	New buildings	75.0	67.8	55.3	83.4	53.9	48.7	79.0	43.00	38.6
	Unchanged buildings	81.9	83.2	70.3	68.2	91.8	64.3	79.3	86.6	70.7
Santorini	New buildings	87.0	83.0	73.8	79.4	78.7	65.3	76.1	72.5	59.0
	Unchanged buildings	92.5	89.7	83.7	94.3	87.0	82.6	92.1	90.8	84.2
	Average New buildings	81.0	75.4	64.6	81.4	66.3	57.0	77.6	57.8	48.8
	Average Unchanged buildings	87.2	86.5	77.0	81.3	89.4	73.5	85.7	88.7	77.5

Table 3. Per-area accuracies of classes “new buildings” and “unchanged buildings” obtained by the considered methods

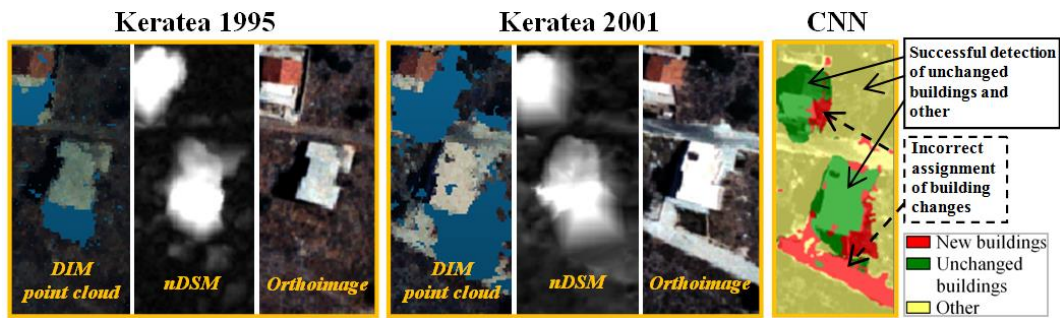


Figure 10. Example of a misclassified case of building change

#### 4. CONCLUSIONS

This paper proposes a novel method for detecting building changes, in one-shot, using scanned analogue aerial photos. A deep scheme based on Convolutional Neural Networks (CNNs) was adopted to classify two complex and real-life urban study areas in Greece to the classes of “new buildings”, “unchanged buildings” and “other”. The CNN was fed with a very small training set (under the 1.5% of the total data of each study area) and with an augmented time period feature band vector consisted of image and 3D information. The proposed method is used only for the detection of changes, in our case the new constructions. It cannot tell whether a new construction is formal or not. A further examination of the existence (or not) of a building permit is needed.

Compared to conventional and shallow learning paradigms such as SVMs, the CNN achieved higher and balanced average quality rates (about 65% and 77% for the classes of “new buildings” and “unchanged buildings” respectively). This shows the flexibility and efficiency of the CNN to identify unchanged buildings as well as new constructions and small extensions or additions of major buildings. However, some buildings were missed and some urban objects were incorrectly identified as buildings. This is mainly due to the inherent artifacts of the used data such as shadows, deformations at the boundaries of the objects during the DIM and quality of the scanned analogue aerial photos. Due to these inherent artifacts, the building outlines are not accurately shaped with sharp edges (see FP and FN near the building boundaries in Figures 6 to 9); building reconstruction techniques can be used to optimize the building boundaries.

The CNN results are considered to be satisfactory as provide reliable building change detection maps that are especially useful for an initial spatial-temporal analysis and predictive assessment. Once such high spatial-temporal probabilities of new and unchanged buildings are localized, more accurate processes may be used to extract the final accurate new and unchanged buildings. Future work is needed to explore and evaluate the performance of several deep learning schemes for building change detection using various test sites, pixel resolutions and types of remote sensing data.

## REFERENCES

- Amin, A., M., E., Liu, Q. and Wang, Y., 2016. Convolutional Neural Network Features Based Change Detection in Satellite Images. International Workshop on Pattern Recognition, Tokyo, Japan, 11-13 May 2016.
- Amit, S., N., K., Saito, S., Sasaki, S., Kiyoki, Y. and Aoki, Y., 2015. Semantic segmentation and difference extraction via time series aerial video camera and its application. International Symposium on Remote Sensing of Environment, Berlin, Germany, 11–15 May 2015.
- Awrangjeb, M., Fraser, C., S. and Lu, G., 2015. Building change detection from lidar point cloud data based on connected component analysis. ISPRS Geospatial Week, La Grande Motte, France, 28 September-03 October 2015.
- Barazzetti, L., Brumana, R., Cuca, B. and Previtali, M., 2015. Change detection from very high resolution satellite time series with variable off-nadir angle. RSCy, Paphos, Cyprus, 16-19 March 2015.
- Beumier, C. and Idrissa, M., 2012. Building Change Detection from Uniform Regions. Lecture Notes in Computer Science 7441, Springer, pp. 648-655.
- Bourdis, N., Marraud, D. and Sahbi, H., 2011. Constrained optical flow for aerial image change detection. IGARSS, Vancouver, BC, Canada, 24-29 July 2011.
- Crommelinck, S., Bennett, R., Gerke, M., Nex, F., Yang, M., Y., Vosselman, G., 2016. Review of automatic feature extraction from high-resolution optical sensor data for UAV-based cadastral mapping. Remote Sensing 8(8), 689, pp. 1-2.
- Dini, G., R., Jacobsen, K., Rottensteiner, F., Al Rajhi, M. and Heipke, C., 2012. 3D Building change detection using high resolution stereo images and a GIS database. XXII ISPRS Congress, Melbourne, Australia, 25 August-01 September 2012.
- Doulamis, N. and Doulamis, T., 2012. Fast and Adaptive Deep Fusion Learning for Detecting Visual Objects. LNCS, 7585, Springer, pp. 345-354.
- Doulamis, N. and Voulodimos, A., 2016. FAST-MDL: Fast Adaptive Supervised Training of multi-layered deep learning models for consistent object tracking and classification. IEEE International Conference on Imaging Systems and Techniques (IST), Chania, Greece, 04-06 October 2016.
- Du., S., Zhang., Y., Qin, R., Yang., Z., Zou, Z., Tang., Y. and Fan, C., 2016. Building Change Detection Using Old Aerial Images and New LiDAR Data. Remote Sensing 8 (1030), pp. 1-22.
- Fujita, A., Sakurada, K., Imaizumi, T., Ito, R., Hikosaka, S. and Nakamura, R., 2017. Damage detection from aerial images via convolutional neural networks. IAPR International Conference on Machine Vision Applications, Nagoya, Japan, 08-12 May 2017.



- Gilani, S., A., N., Awrangjeb, M. and Lu, G., 2016. An automatic building extraction and regularization technique using LIDAR point cloud data and orthoimage. *Remote Sensing* 8 (3), 258.
- Hron, V. and Halounova, L., 2015. Use of aerial images for regular updates of buildings in the fundamental base of geographic data of the Czech Republic. PIA15+HRIGI15- Joint ISPRS conference, Munich, Germany, 25-27 March 2015.
- Ioannidis C., Psaltis, C. and Potsiou, C., 2010. Towards a strategy for control of suburban building permits informal buildings through automatic change detection. *Computers, Environment and Urban Systems* 33, pp. 64-74.
- Karantzalos, K., 2015. Recent Advances on 2D and 3D Change Detection in Urban Environments from Remote Sensing Data. *Computational Approaches for Urban Environments, Geotechnologies and the Environment* 13, pp. 237-272.
- Koeva, M. and Oude Elberink, S. 2016. Challenges for Updating 3D Cadastral Objects using LiDAR and Imagebased Point Clouds. 5<sup>th</sup> International FIG 3D Cadastre Workshop, Athens, Greece, 18-20 October 2016.
- Makantasis, K., Karantzalos, K., Doulamis, A. and N., Doulamis, 2015. Deep supervised learning for hyperspectral data classification through convolutional neural networks. IGARSS, Milan, Italy, 26-31 July, 2015.
- Maltezos, E., Doulamis, N., Doulamis, A. and Ioannidis, C., 2017. Deep convolutional neural networks for building extraction from orthoimages and dense image matching point clouds. *Journal of Applied Remote Sensing* 11(4), pp.- 042620-1-042620-22.
- Maltezos, E. and Ioannidis, C., 2016. LIDAR vs dense image matching point clouds in complex urban scenes. RSCy, Paphos, Cyprus, 04-08 April 2016.
- Marmanis, D., Adam, F., Datcu, M., Esch, T. And U. Stilla., 2015. Deep neural networks for above-ground detection in very high spatial resolution digital elevation models. PIA15+HRIGI15- Joint ISPRS conference, Munich, Germany, 25-27 March 2015.
- Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, pp. 247-259.
- Nebiker, S., Lack, N. and Deuber, M., 2014. Building Change Detection from Historical Aerial Photographs Using Dense Image Matching and Object-Based Image Analysis. *Remote Sensing* 6, pp. 8310-8336.
- Pédrinis F. and Gesquière G., 2017. Reconstructing 3D Building Models with the 2D Cadastre for Semantic Enhancement. LNGC, *Advances in 3D Geoinformation*, pp.119-135.
- Pédrinis F., Morel M. and Gesquière G., 2015. Change Detection of Cities. LNGC, *3D Geoinformation Science*, pp 123-139.
- Peng, D. and Zhang, Y., 2016. Building change detection by combining lidar data and ortho image. XXIII ISPRS Congress, Czech Republic, 12-19 July 2016.

- Qin, R., Tian, J., Reinartz, P., 2015. Spatiotemporal inferences for use in building detection using series of very high-resolution space-borne stereoimages. *International Journal of Remote Sensing* 37, pp. 1-22.
- Radke, R., J., Andra, S. and Al-Kofahi, O., 2005. Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing*, 14(3), pp. 294-307.
- Rottensteiner, F., 2007. Building change detection from digital surface models and multi-spectral images, PIA, Munich, Germany, 19-21 September 2007.
- Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J., D., Breitkopf U. and Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93, pp. 256-271.
- Rutzinger, M., Rottensteiner, F. and Pfeifer, N., 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2(1), pp. 11-20.
- Sakurada, K. and Okatani, T., 2015. Change Detection from a Street Image Pair using CNN Features and Superpixel Segmentation. *British Machine Vision Conference*, Swansea, UK, 07-10 September 2015.
- Sarp, G., Erener, A., Duzgun, S. and Sahin, K., 2017. An approach for detection of buildings and changes in buildings using orthophotos and point clouds: A case study of Van Earthquake. *European Journal of Remote Sensing* 47, pp. 627-642.
- Tian, J. And Reinartz, P., 2013. Comparison of two fusion based building change detection methods using satellite stereo imagery and DSMs, *ISPRS IWIDF*, Antu, Jilin Province, PR China, 20-22 August 2013.
- Vakalopoulou, M., Karantzalos, K., Komodakis, N. and Paragios, N., 2015. Building detection in very high resolution multispectral data with deep learning features. *IGARSS*, Milan, Italy, 26-31 July, 2015.
- Vakalopoulou, M., Karantzalos, K., Komodakis, N. and Paragios, N., 2016. Graph-Based Registration, Change Detection, and Classification in Very High Resolution Multitemporal Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(10), pp. 2940-2951.
- Xu, S., Vosselman, G. and Elberink, S., O., 2013. Detection and classification of changes in buildings from airborne laser scanning data. *ISPRS Workshop Laser Scanning*, Antalya, Turkey, 11-13 November 2013.
- Yu, D. and Deng, L., 2011. Deep learning and its applications to signal and information processing. *IEEE Signal Processing Magazine* 28(1), pp.145-154.
- Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X. and Yan, G., 2016. An Easy-to-Use Airborne LIDAR Data Filtering Method Based on Cloth Simulation. *Remote Sensing* 8(501), pp. 1-22.

## BIOGRAPHICAL NOTES

**Evangelos Maltezos** obtained his BSc degree in civil engineering from the Technological Educational Institute of Patras, Greece, in 2007. He also received a Diploma degree in rural and surveying engineering from the National Technical University of Athens (NTUA) in 2013. He is currently a Ph.D candidate at NTUA's Laboratory of Photogrammetry, School of Rural and Surveying Engineering. His research interest lies in object detection, feature extraction, segmentation, shadow detection and removal, 3D reconstruction, change detection, LIDAR and SAR processing.

**Charalabos Ioannidis** is a professor at the Laboratory of Photogrammetry, School of Rural and Surveying Engineering, National Technical University of Athens, Greece, where he teaches photogrammetry and cadastre. His research interests lie in the fields of computer vision, satellite photogrammetry, 3D modeling, change detection and cultural heritage documentation. He has authored more than 170 papers in the above fields. He is chair of the FIG Commission 3 Working Group 3.2 "Technical Aspects of Spatial Information Management".

**Anastasios Doulamis** received the Diploma degree and Ph.D in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) both with the highest honor. He is currently assistant professor of the NTUA. Prof. A. Doulamis has received several awards in his studies, including the Best Graduate Thesis Award, and Best Student award among all engineers. He is author of more than 280 scientific papers receiving more than 3000 citations.

**Nikolaos Doulamis** received a Diploma and Ph.D in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) both with the highest honor. He is now Assistant Professor at the NTUA. He has received many awards (Best Greek Engineer, NTUA Medal). He has served as Organizer and/or TPC in major IEEE/SPIE conferences. He is author of more than 60 (200) journals (conference) papers and received more than 3000 citations.

## CONTACTS

Evangelos Maltezos  
National Technical University of Athens (NTUA)  
9, Iroon Polytechniou St.  
Athens 15780  
GREECE  
Tel. + 30 2107722686  
Fax + 30 2107722677  
Email: maltezosev@gmail.com  
Web site: <https://www.ntua.gr/en/>

Charalabos Ioannidis  
National Technical University of Athens (NTUA)  
9, Iroon Polytechniou St.

Athens 15780  
GREECE  
Tel. + 30 2107722686  
Fax + 30 2107722677  
Email: [cioannid@survey.ntua.gr](mailto:cioannid@survey.ntua.gr)  
Web site: <https://www.ntua.gr/en/>

Anastasios Doulamis  
National Technical University of Athens (NTUA)  
9, Iroon Polytechniou St.  
Athens 15780  
GREECE  
Tel. + 30 2107722686  
Fax + 30 2107722677  
Email: [adoulam@cs.ntua.gr](mailto:adoulam@cs.ntua.gr)  
Web site: <https://www.ntua.gr/en/>

Nikolaos Doulamis  
National Technical University of Athens (NTUA)  
9, Iroon Polytechniou St.  
Athens 15780  
GREECE  
Tel. + 30 2107722686  
Fax + 30 2107722677  
Email: [ndoulam@cs.ntua.gr](mailto:ndoulam@cs.ntua.gr)  
Web site: <https://www.ntua.gr/en/>