HAL
open science

Graph-Based Registration, Change Detection, and Classification in Very High Resolution Multitemporal Remote Sensing Data

Maria Vakalopoulou, Konstantinos Karantzalos, Nikos Komodakis, Nikos Paragios

To cite this version:

Maria Vakalopoulou, Konstantinos Karantzalos, Nikos Komodakis, Nikos Paragios. 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, 2016, 9, pp.2940 - 2951. 10.1109/JSTARS.2016.2557081 . hal-01413419

HAL Id: hal-01413419

https://inria.hal.science/hal-01413419v1

Submitted on 9 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Graph-based Registration, Change Detection and Classification in Very High Resolution Multitemporal Remote Sensing Data

Maria Vakalopoulou, Konstantinos Karantzalos, *Senior Member, IEEE,* Nikos Komodakis, Nikos Paragios, *Fellow Member, IEEE*

***Abstract*— In this paper we propose a modular, scalable, metric-free, single-shot change detection/registration method. The developed framework exploits the relation between the registration and change detection problems, while under a fruitful synergy the coupling energy term constrains adequately both tasks. In particular, through a decomposed interconnected graphical model the registration similarity constraints are relaxed in the presence of change detection. Moreover, the deformation space is discretized, while efficient linear programming and duality principles are used to optimize a joint solution space where local consistency is imposed on the deformation and the detection space as well. The proposed formulation is able to operate in a fully unsupervised manner addressing binary change detection problems *i.e., change* or *no-change* with respect to different similarity metrics. Furthermore, the framework has been formulated to address automatically the detection of *from-to* change trajectories under a supervised setting. Promising results on large scale experiments demonstrate the extreme potentials of our method.**

***Index Terms*—Markov random fields, deformable, registration, change trajectories, multisensor, land cover, buildings**

I. Introduction

Detecting and modeling spatio-temporal changes over the structured environment is critical in various engineering, civilian and military applications. One can cite for example, urban and rural planning, mapping and updating geographic information systems, surveillance, transportation, virtual tourism and location based services. Despite important research efforts [6], [17], [19], [23], [25], [26], [30], [36] and recent advances [1], [3], [12] accurate detection and modeling of geometric/man-made changes is, still, challenging. This is particularly the case when one considers addressing the problem in the context of (i) large-scale (e.g., updating geospatial databases, Microsoft Virtual Earth, Google Earth), (ii) detailed and spatially accurate mapping (based on very high resolution data) and (iii) sparse multi-temporal and/or massive streams of current earth observation data.

M. Vakalopoulou and K. Karantzalos are with the Remote Sensing Laboratory, National Technical University of Athens, Zographou campus, 15780, Athens, Greece (e-mail: *mariavak@central.ntua.gr*, *karank@central.ntua.gr*).

N. Komodakis is with the LIGM Laboratory, École des Ponts ParisTech, Marne-la-Vallée 77455, France (e-mail: *nikos.komodakis@enpc.fr*)

N. Paragios is with the Laboratoire de Mathématiques Appliquées aux Systèmes, École Centrale Paris, 92290 Châtenay-Malabry Cedex, France and also with the GALEN Team, INRIA, 91893 Orsay Cedex, France (e-mail: *nikos.paragios@ecp.fr*).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier

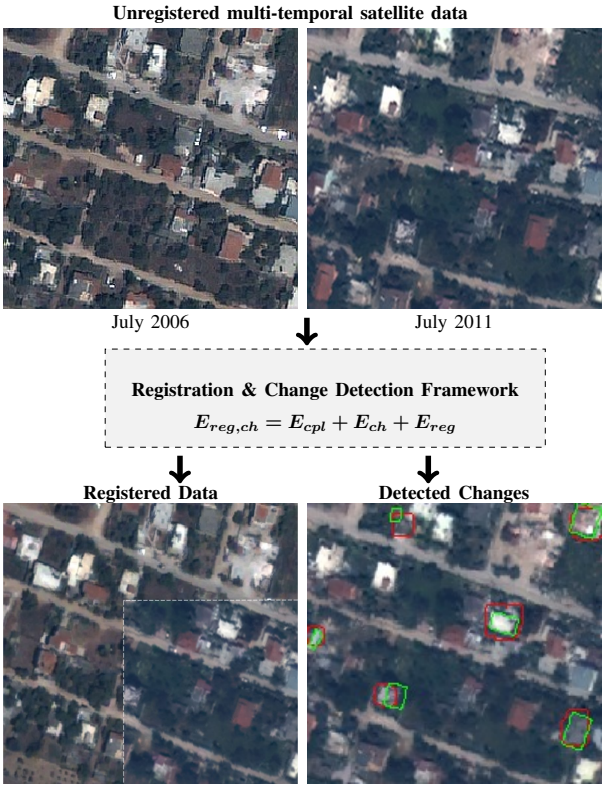

Fig. 1: The developed framework addresses simultaneous the registration and change detection tasks.

Addressing such challenges is becoming currently a necessity since earth observation missions now provide cost-effective data for large-scale urban and peri-urban monitoring. Such data consists of very-high resolution (e.g., less than 50cm) multispectral satellite images and can cover several square kilometers (e.g., 50Km^2) with a single image and daily revisit capabilities. However, the development of accurate and automated registration methods for very large, multispectral, high resolution satellite data is not a trivial task [18], [25], [31].

In the past decade, a number of image registration methods have been proposed and summarized in various surveys on the topic [5], [38] or in related areas [29]. The objective of these methods is to estimate an appropriate transformation model from one image to the other. The existing registration methods can be classified into two main categories: feature-based and

area-based. Feature-based methods employ region descriptors [32] or more recently deep neural networks [37] that are robust to illumination and viewpoint changes. Descriptors similarity criteria are used to provide potential sparse correspondences that are used as basis for the estimation of the transformation parameters. Deformable methods [11], [29], used in computer vision, medical imaging and remote sensing can be considered as area based methods. However, the majority of the automatic image registration algorithms assume that the two images depict objects that are visible in both spaces. Such hypothesis is violated in the presence of changes between two successive acquisitions and is often addressed through change detection.

However, change detection from multitemporal earth observation data, still, remains a challenge. Kernels [2], [34], Markov Random Fields [1], [8], [28] and neural networks [22], [24] have gained attention in the recent years. In the context of man-made object change detection [4], [21] for urban and peri-urban regions, several approaches have been proposed based on very high resolution optical and radar data [7], [20], [21], [24]. However, these change detection techniques require accurately co-registered data in order to perform pixel-by-pixel or region-by-region multi-temporal data fusion, correlation or change analysis as largely spurious results of change detection will be produced if images are misaligned.

To this end, in this paper, we propose an one-shot registration/ change detection framework where the registration of very high resolution data is optimally addressed through deformation grids and powerful discrete optimization [11], while the desired changes correspond to regions for which correspondences between the unregistered multi-temporal data cannot be established (Fig. 1). Moreover, we extend the recently proposed change detection framework [33] by providing information about the type of detected *from-to* change trajectories.

In particular, our contribution refers to a scalable, modular, metric-free, single-shot change detection/registration method. The framework exploits a decomposed interconnected graphical model formulation where in the presence of changes the iconic similarity constraints are relaxed. We employ a discretized, grid-based deformation space. State-of-the-art linear programming and duality principles have been used to optimize the joint solution space where local consistency is imposed on the deformation and the detection space. The unsupervised framework has been designed to handle and process large, very high resolution multispectral remote sensing data, while was optimized towards man-made object change detection in urban and peri-urban regions. Furthermore, we have extended the formulation in order to detect automatically *from-to* change trajectories based on a supervised manner. The developed method has been validated through large scale experiments on several multi-temporal very high resolution optical satellite datasets.

The main contributions of the developed method are *(i)* the novel, single and modular joint registration and change detection framework, *(ii)* the metric-free formulation which allows numerous and change-specific implementations, *(iii)* the classification of the different types of changes, *(iv)* the low computational complexity which would allow near real-time performance once modern parallel programming architectures are considered.

II. Methodology

A. MRF formulation

We have designed and built an MRF model over two different graphs of the same topology, number of nodes and connectivity system (Fig. 2). The first graph corresponds to the registration term (G_{reg}) and the second one to the change detection term (G_{ch}). The assumption of local consistency on the retained solution for each space is imposed by a smoothness term which each graph contains. Moreover, the interaction between the two graphs is performed by the similarity cost which connects the registration with the change detection terms.

Each graph is superimposed on the image [9] and therefore every node of the graph acts and depends on a subset of pixels in its vicinity (depending on the interpolation strategy). With such a manner every pixel can participate through a certain weight, related to its distance from the nodes, to the graph. The dimensions of the graph are related to the image dimensions forming a trade off between accuracy and computational complexity. In particular, the computational complexity is lower as graph's dimensions are smaller than the unregistered raw images. In such a setting the deformation of a pixel is defined through an interpolation of the displacement of the proximal graph nodes:

$$T(x) = x + \sum_{p \in G} \eta(||x - p||) d_p \tag{1}$$

where d_p is the displacement vector of the control point p, x is an image pixel and $\eta(.)$ is the projection function which connects with a weight propositional to the distance the pixels with the nodes of the grid and reverse. That way every pixel participates to each node depending to its distance from the node. A typical example of a projection function would be cubic B-splines which is the one employed here.

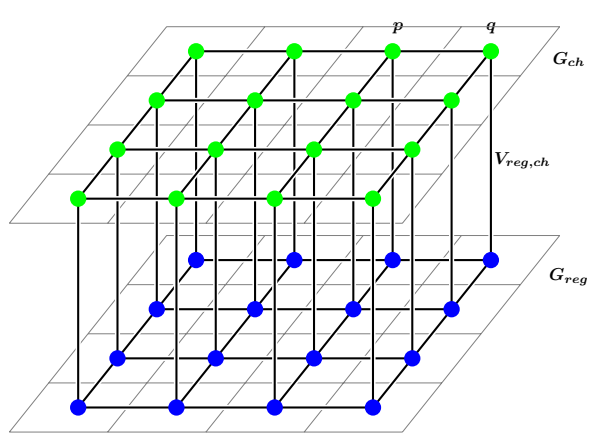

Fig. 2: Each graph contains a smoothness term which imposes the necessary homogeneity within the graph. The interaction between the two graphs is performed by the similarity cost which connects the registration with the change detection terms.

After the optimization, the same projection function $\hat{\eta}(.)$ will be used to project the optimal displacements (d_p) to the image pixels. Once the similarity criterion has been defined, the next step consists of imposing certain continuity on the deformation space which is discussed in the next subsection.

The energy function $E_{reg,ch} = E_{cpl} + E_{ch} + E_{reg}$ introduces connectivity between the two different graphs. The label for each node p belonging to the graph G is $l_p = [l^c, l^{reg}]$. The labels for the registration are $l^{reg} \in \Delta$, where $\Delta = [d^1, \ldots, d^n]$ corresponds to all possible displacements. The labels for the change detection l^c are different for the unsupervised and the supervised approach. In particular, during the unsupervised approach the change detection labels are $l^c \in \{0, 1\}$ where 1 indicates the presence and 0 the absence of change. On the other hand, during the supervised approach the changes are classified in different classes (*i.e., from-to* change trajectories) and therefore the labels are $l^c \in \{0, \ldots, k\}$ where 0 indicates the absence of change and all the others indicate a specific type of change (*e.g.,* vegetation-to-man-made, vegetation-to-soil, soil-to-man-made, *e.t.c.*). Finally, the label space can be summarized as $L = l^c \times \Delta$.

B. The Registration Energy Term

For tackling the registration problem we employ a non-rigid framework based on recently proposed and validated algorithms [9], [11], [29]. It is not based on any single geometric model but can be regarded as an interpolation-based approach employing a free-form deformation strategy [29] coupled with detection labels. In particular, displacements which can be considered known in a restricted set of locations are interpolated for the rest of the image domain. The model is rich enough to describe the transformations that exist in the images in order to reach an optimal solution by employing *e.g.* a free-form deformation. This strategy employs a grid which is superimposed on the image and the transformations are calculated using these control points (nodes). The implemented multiscale approach can address large initial displacements (note that the solution is calculated over the superimposed grid and not over image pixels).

More specifically, let us consider a pair of images where A is the source image and V is the target image defined on a domain Ω. The goal of image registration is to define a transformation map T which will project the source image to the target image as presented below:

$$V(x) = A(x) \circ T(x) \tag{2}$$

Let us denote a discrete set of labels $L_{reg} = [1, \ldots, n]$, and a set of discrete displacements $\Delta = [d^1, \ldots, d^n]$. We seek to assign a label l_p^{reg} to each grid node p, where each label corresponds to a discrete displacement $d^{l_p^{reg}} \in \Delta$. The energy formulation for the registration comprises of a similarity cost (that seeks to satisfy (2)) and a smoothness penalty on the deformation domain.

$$V_{pq,reg}(l_p^{reg}, l_q^{reg}) = ||d^{l_p^{reg}} - d^{l_q^{reg}}|| \tag{3}$$

where p and q are neighboring nodes and $d^{l^{reg}}$ their calculated displacements for each registration label l^{reg}.

The smoothness term penalizes neighboring nodes that have different displacement labels, depending on the distance of the labeled displacements as in (3). The similarity cost depends on the presence of changes and will be subsequently defined at the following Section II.D.

It is worth mentioning that we have designed the framework to work with any type of input image pairs (*e.g.,* raw, rectified based on RPCs, rectified based on a reference image, *etc.*). It can address cases with large initial displacements (*e.g.,* raw data), address relief displacements or already rectified data based on rigid (*e.g.,* affine) transformations. This is mainly a matter of the number of multiresolution grid levels, image scales, initial grid size, *etc*. It should be also noted that the smoothness term highly constrains the displacements of the grid nodes and does not let any grid cross-overs during the optimization at finer scales or at regions with important relief displacements.

C. The Change Detection Energy Term

The goal of the change detection term is to determine the changed and unchanged image regions and at the same time the *from-to* change trajectories depending on the labels of change. The energy formulation for the change detection corresponds to a smoothness term which gives a penalty to neighboring nodes with different detection labels. Depending on the approach the labels of the change detection are different.

*1) **Unsupervised Change Detection: 'Change' or 'No-change'**:* We employ two labels in order to address the change detection problem $l_p^c \in \{0, 1\}$. The energy term in this case can be formulated as follow:

$$V_{pq,ch}(l_p^c, l_q^c) = ||l_p^c - l_q^c|| \tag{4}$$

where p and q are neighboring nodes.

*2) **Supervised Change Detection: 'from-to' change trajectories**:* In this case the total number of the change detection labels is k and depends on the number of different change classes that are defined $l_p^c \in \{0, \ldots, k\}$. Respectively the energy term can be formulated as follows:

$$V_{pq,ch}(l_p^c, l_q^c) = \begin{cases} 0 & l_p^c = l_q^c \\ c_1 & l_p^c \neq l_q^c \quad \& \quad (l_p^c | l_q^c) = 0 \\ c_2 & otherwise \end{cases} \tag{5}$$

where p and q are neighboring nodes, c_1, c_2 are constant values which penalize different change values where $c_1 > c_2$ and (.|.) is the or operator.

D. Coupling the Energy Terms

The coupling between change detection and registration is achieved through the interconnection between the two graphs. Assuming a pair of corresponding nodes belonging to each graph, one would expect that in the absence of change the similarity cost should be satisfied. By coupling the two terms, we achieve a more relaxed deformation field in the changed

areas. In particular, we formulate with the following manner the two cases:

*1) **Unsupervised Change Detection: 'Change' or 'No-change'**:*

$$V_{reg,ch}(l_p^{reg}, l_p^c) = l_p^c \cdot C + (1 - l_p^c) \cdot \int_\Omega \hat{\eta}(||x-p||)\rho(V(x), A(x + d^{l_p^{reg}}))dx \quad (6)$$

where we simply take all pixels in the vicinity of the graph node and project their similarity values $\rho(\ .\ )$ back to the grid node with a weight that is proportional to the distance using the projection function $\hat{\eta}$. Different similarity functions $\rho(.)$ can be used (as demonstrated in Section IV) in order to compare the two images, while a constant value C is used in order to define the changed areas. These two terms are integrated as in (6) which simply uses a fixed cost in the presence of changes and the similarity value in their absence.

*2) **Supervised Change Detection: 'from-to' change trajectories**:*

$$V_{reg,ch}(l_p^{reg}, l_p^c) = \mathbb{1}_{l_p^c \# 0}(C + \int_\Omega \hat{\eta}(||x-p||)e^{-\Psi_{l_p^c}(x)}dx) + \mathbb{1}_{l_p^c = 0}(\int_\Omega \hat{\eta}(||x-p||)\rho(V(x), A(x + d^{l_p^{reg}}))dx) \quad (7)$$

where apart from the fixed cost C, we use the classification scores, $\Psi_{l_p^c}(x)$, for all the different labels of change. The classification scores are independent from the model and had been calculated on pixel level. Using with the same way the projection function, the classification scores are projected to each node. The computed classification scores are used to define the different type of changes (*from-to* change trajectories from (7)) when required. With $\mathbb{1}$ we denote the indicator function.

E. Energy Formulation

With a slight abuse of notation we consider a node with an index $p \in G$ (we recall that the two graphs are identical) corresponding to the same node throughout the two graphs (G_{reg}, G_{ch}). We can now integrate all terms to a single energy which detects changes, establishes correspondences and imposes smoothness in the change detection and the deformation map as follows:

$$\begin{aligned} &E_{reg,ch}(l^c, l^{reg}) = \\ &E_{cpl}(l^c, l^{reg}) + E_{reg}(l_p^{reg}, l_p^{reg}) + E_{ch}(l_p^c, l_q^c) = \\ &w_1 \sum_{p \in G} V_{reg,ch}(l_p^{reg}, l_p^c) + w_2 \sum_{p \in G_{reg}} \sum_{q \in N(p)} V_{pq,reg}(l_p^{reg}, l_q^{reg}) \\ &+ w_3 \sum_{p \in G_{ch}} \sum_{q \in N(p)} V_{pq,ch}(l_p^c, l_q^c) \end{aligned} \quad (8)$$

where $V_{reg,ch}(l_p^{reg}, l_p^c)$ represents the coupling term for each node at each label, $V_{pq,reg}(l_p^{reg}, l_q^{reg})$ the pairwise or binary term for the registration and $V_{pq,ch}(l_p^c, l_q^c)$ the pairwise or binary term for the change detection. Moreover w_1, w_2, w_3 are the weights of each term and $N(p)$ the neighbourhood of each node p.

In such a setting, optimizing an objective function seeking similarity correspondences is not meaningful and deformation vectors should be the outcome of the smoothness constraints on the displacement space. However, the areas of changes, and their type if required, are unknown and is one of the objectives of the optimization process. Without loss of generality we can assume that the matching cost addressing to change can correspond to a value that can be determined from the distribution of these costs on the entire domain (it is metric dependent). Let us consider that this value is known and that is independent from the image displacements and thus we can distinguish the regions that have been changed.

Therefore, the advantage of the developed single-shot framework is mainly that by solving the two problems simultaneously first we have less false positives due to unregistered data and secondly the registration is robust to multitemporal datasets since the corresponding energy terms are relaxed in regions with possible changes. Moreover, the framework can detect the optimal labels which indicate a specific *from-to* change trajectories.

It should be noted that the developed registration and change detection framework is generic and modular and one can integrate any training procedure, classifier, computed features, number of classes making it ideal for various applications [27]. Finally, the pairwise costs for both terms have been described in (3) and (4) or (5).

F. Optimization

There are several techniques for the minimization of an MRF model which can be generally summarized into those based on the message passing and those on graph cuts [35]. The first category is related to the linear programming relaxation [14]. The optimization of the implementation is performed by FastPD which is based on the dual theorem of linear programming [15], [16]. In particular FastPD is an generalisation of α-expansion and it exploits information coming not only from the original MRF problem, but also from a dual problem. That way FastPD computes exactly the same solution as α-expansion but with substantial speedup. Finally, another advantage of FastPD is that it guarantees an almost optimal solution for a wide class of NP-hard MRF problems.

III. Implementation

A multi-scale framework has been designed for the minimization of the MRF energy. Concerning the image, iteratively different levels of Gaussian image pyramids are used. Concerning the grid, we again consider different levels of it, beginning with a sparse one. The objective of the multi-scale approach is twofold: (i) allows an efficient sampling of the search space that is critical given the product label space (this allows to progressively move closer and closer to the solution and do not get stuck to a local minimum), (ii) accelerates convergence of the method and decreases computational load. A single shot framework without a multiscale approach, would

require tremendous sampling of the search space to handle large deformations, and will make the approach suboptimal both in terms of convergence as well as in terms of solution quality. The different levels of the images and the grid with the consistency of nodes in the grid can be defined by the user. For optimizing very high resolution multispectral satellite data we used 2 image and 3 grid levels.

Continuing with the label space, we search for possible displacements along 8 directions (x, y and diagonal axis), while the change labels are (*e.g.,* for the first unsupervised case) always two and correspond to change or no-change description. As far as the registration labels are concerned, their values are not the same at each level, but depending on the parameter label factor, they change in order to be closer to optimal. Consider that in each grid level the source image is deformed according to the optimal labels and it is updated for the next level. In our case, we used the value 0.8 for updating the label values for registration. Last but not least, the maximum displacement has to be smaller than 0.4 times the distance of two consecutive nodes in order to preserve the right displacement of every node. The maximum displacement depends on the initial distance between nodes. For important displacements, the grid should be sparse during the initial grid levels, while gradually the grid is becoming more and more dense allowing the recovery of relative small displacements.

Furthermore, a number of methods for block matching can be considered. Semantic changes in multitemporal imagery affect the local intensities and also change the structure of the region. One of the problems in traditional unsupervised change detection techniques, is that change in intensities does not directly mean semantic change. In our case, this was crucial since we focused and optimizing the unsupervised framework for urban and peri-urban regions and man-made objects changes. Therefore, in our implementation we compare not only the intensity differences but also the differences in the edges of the image, calculated by the gradient inner product (SADG function). On the other hand any other similarity measure such as mutual information, normalized cross correlation, ratio correlation can be used. In Section IV we have tested different similarity functions. Depending on the similarity function and the required system sensitivity the value of the constant parameter can be modified. In all our experiments with the unsupervised change detection framework and the use of the SADG metric, which was optimized for detecting man-made changes, the corresponding constant parameter was set to 100.

Regarding the integrated supervised classification procedure, we have defined five different types of change (*from-to* change trajectories) and we have trained respectively a SVM classifier. In particular, a 50% splitting ratio was employed for the training procedure *i.e.,* half of the images in the datasets have been used for training and the other half for testing. The features used for the classification were based on different spectral bands, the NDVI and MSAVI-2 indexes, as well as the similarity $\rho(\cdot)$ between the two multitemporal data. In particular NDVI and MSAVI-2 indexes are following the standard equations as presented below:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

$$MSAVI_2 = \frac{2NIR + 1 - \sqrt{(2NIR+1)^2 - 8(NIR - RED)}}{2}$$

For the supervised approach, the value of the constant parameter regarding the change detection energy term and the SADG metric was set to 190.

As far as the other parameters of the developed framework are concerned (regarding both the unsupervised and supervised approaches), the number of iterations per level was set to 10, the regularization parameter for the registration to 35 and for change detection to 3.5. The function used for the projection from the pixel to nodes and reverse was the Cubic B-splines. Concluding, the parameter that controls the balance between the absolute difference and the gradient inner product was set to 0.2.

IV. Experimental Results and Validation

The developed framework was applied to several pairs of very high resolution, multispectral images from different satellite sensors (*i.e.,* Quickbird and WorldView-2). All datasets were acquired between the years of 2006 and 2011. The multi-temporal dataset covers approximately a 14 km^2 region in the East Prefecture of Attica in Greece (Fig. 3 and Table II). All raw images were atmospherically corrected, while after the pansharpening their size were approximately 8000 by 7000 pixels with a spatial resolution of approximately 50 centimeters. The dataset is quite challenging both due to its size and the pictured complexity derived from the different acquisition angles. For the quantitative evaluation the ground truth was manually collected and annotated after an attentive and laborious photo-interpretation done by two different experts.

Extensive experiments were performed over several images pairs and based on several similarity metrics namely the

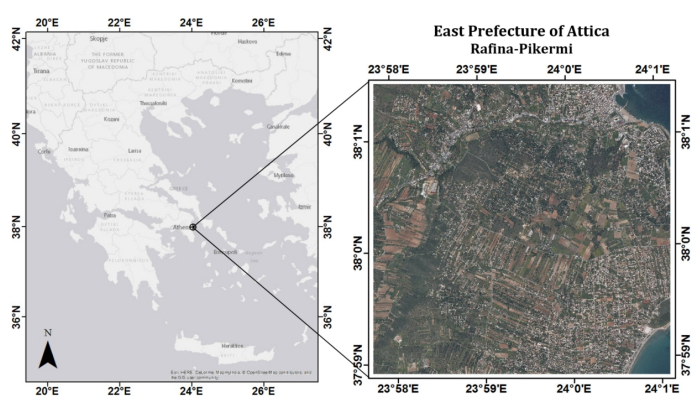

Fig. 3: The multitemporal dataset is covering approximately a 14km^2 region in the East Prefecture of Attica, Greece. The dataset contains very high resolution, multispectral satellite optical data over a complex terrain with urban, peri-urban, agricultural, coastal and forest regions.

Satellite Data	**Date of Acquisition**	**Spectral Bands** (number)	**Size** (GBs)
QuickBird PanSharpened	May 2006	4	0.32
QuickBird PanSharpened	April 2009	4	0.32
Worldview-2 PanSharpened	April 2010	8	0.90
Worldview-2 PanSharpened	April 2011	8	0.90

TABLE II: The multimodal dataset includes satellite optical multispectral images all of them pansharpened on 0.5m and with different acquisition dates.

Sum of Absolute Differences (SAD),the Sum Absolute of Differences plus Gradient Inner Products (SADG), the Sum of Square Differences (SSD), the Normalized Cross Correlation (NCC), the Normalized Mutual Information (NMI), the Correlation Ratio (CR), the Sum of Gradient Inner Products (GRAD), the Normalized Correlation Coefficient plus Sum of Gradient Inner Products (CCGIP), the Hellinger Distance (HD), the Jensen-Renyi Divergence (JRD) and the Mutual Information (MI). The experimental results were evaluated both qualitative and quantitative for the registration and the change detection tasks.

Moreover, in order to evaluate quantitatively the developed algorithm for the *change* or *no-change* detection task, the standard quality metrics of Completeness, Correctness and Quality were calculated at object level.

$$Completeness = \frac{TP}{TP + FN} \tag{9}$$

$$Correctness = \frac{TP}{TP + FP} \tag{10}$$

$$Quality = \frac{TP}{TP + FP + FN} \tag{11}$$

The True Positives (TP), False Negatives (FN) and False Positives (FP) were calculated in all cases. In particular, TP is the number of correctly detected changes, FN is the number of changes that have not been detected by the algorithm and FP is the number of false alarms.

A. *Evaluating the performance of the registration procedure*

For the evaluation of the registration procedure we have calculated the mean displacement errors before and after the application of the developed framework. Several ground control points (GCPs) were manually collected in both unregistered and registered image pairs. It should be noted that most of the GCPs were collected on building roof tops and corners since the goal was to evaluate the performance mainly against relief displacements where the largest errors occur. In general, in all other regions the framework resulted in sub-pixel accuracy. Moreover, we validated the registration energy term alone based on the similar framework of [11]. In particular, for exactly the same experimental setup, results from the proposed energy formulation (8) and results with the registration term alone were compared.

In Fig. 4 and Table I the mean displacement errors (in pixels) in both directions (x, y) and the mean distance before and after the registration are presented. In order to validate the framework, results from experiments with several similarity metrics are shown. It can be observed that the developed framework acts quite robustly regarding the registration procedure since in all cases regardless of the employed similarity metric the mean displacement errors were lower than 3.2 pixels in both axis. The initial mean distance (displacement DS) of the unregistered image pairs was more than 11 pixels and after the application of the developed framework based

	Registration term E_{reg}			**Developed Framework** $E_{cpl} + E_{reg} + E_{ch}$			**Comparison**
	Dx (pixels)	Dy (pixels)	DS (pixels)	Dx (pixels)	Dy (pixels)	DS (pixels)	DS Difference
Unregistered data	7.61	7.31	11.04	7.61	7.31	11.04	-
SADG	2.60	1.93	3.24	2.45	2.03	3.18	0.06
SAD	2.63	1.27	2.92	2.57	1.32	2.89	0.03
SSD	3.04	2.09	3.69	3.12	2.04	3.73	**-0.04**
NCC	**2.23**	**1.33**	**2.60**	**2.13**	**1.23**	**2.46**	0.14
NMI	2.60	1.86	3.20	2.53	1.92	3.18	0.02
CR	2.62	1.51	3.02	2.67	1.04	2.87	0.15
GRAD	3.18	1.66	3.59	3.23	1.74	3.67	**-0.08**
CCGIP	2.69	2.41	3.61	2.84	2.5	3.78	**-0.17**
JRD	2.38	1.29	2.71	2.34	1.34	2.70	0.01
HD	2.38	1.12	2.63	2.42	1.08	2.65	**-0.02**
MI	2.68	1.25	2.96	2.76	1.02	2.94	0.02

TABLE I: Quantitative evaluation regarding the performance of the registration procedure. The mean displacement errors for both axis (Dx & Dy) and the mean distance (DS) before and after the convergence of the developed algorithm are presented. Different similarity metrics are considered, while a comparison with the registration term E_{reg} alone is provided (left).

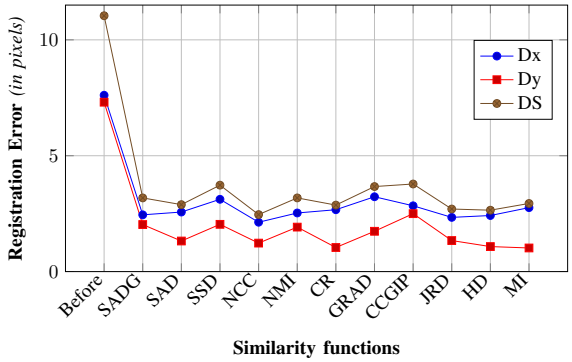

Fig. 4: The performance of different similarity functions during registration in terms of mean displacement error for both axis (Dx, Dy) and the mean distance (DS).

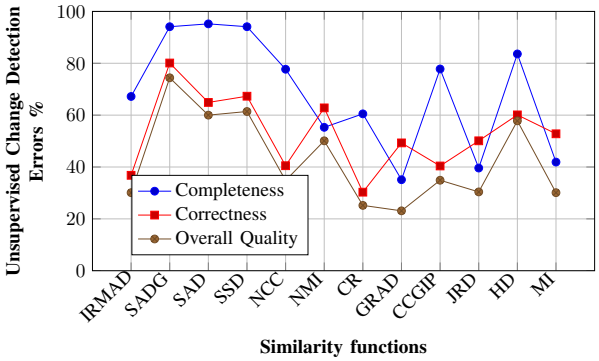

Fig. 5: Evaluating the performance of different similarity functions for the unsupervised change detection formulation based on the standard measures of detection completeness and correctness at object level.

on the NCC similarity measure the mean displacement error was less than 2.5 pixels (2.13 along the x axis and 1.23 along the y, respectively). Regarding the registration the NCC metric outperformed in all our experiments the other similarity measures and this was in accordance with the literature [11].

Regarding the comparison of the developed framework and the performance of the registration term alone, quantitative results (Table I, left) indicate that the average (DS) difference including all examined similarity measures was less than 0.07 pixels (Table I, right). Thus, the coupling of the energy terms didn't affect the registration performance, indicating that the registration similarity constraints (l^{reg}) were efficiently relaxed in the presence of change.

B. Evaluating the performance of the Unsupervised Change Detection: 'Change' or 'No-change'

Regarding the evaluation for the unsupervised change detection task, experimental results after the application of the developed method are shown in Fig. 6. In particular, the detected changes are shown with a red color while the ground truth polygons are shown with green. It can be observed that in most cases the detected changes are in accordance with the ground truth. There are sporadically a few FN and FP due to, in most cases, high spectral variations between soil, gravel and other construction materials between the two acquisition periods. In all cases, the experimental results were derived based on the SADG similarity measure and parameters optimized for man-made object changes as mentioned in Section III. It can be observed that the framework can detect changes related to man-made objects based both on the spectral (*i.e.,* absolute difference) and geometric features (*i.e.,* inner gradient products).

Quantitative evaluation results are presented in Table III and Fig. 5 towards the detection of changes regarding man-made objects. Several experiments have been performed with different similarity measures based on the unsupervised formulation. As it can be observed the SAD, SSD and SADG similarity measures resulted into the highest detection completeness and correctness rates. However, SADG outperformed the other ones with more than 14% regarding the Overall Quality measure. SAD and SSD were significantly more sensitive

	Change Detection Term			**Developed Unsupervised Framework**		
	E^*_{ch}			$E_{cpl} + E_{reg} + E_{ch}$		
Similarity	Completeness %	Correctness %	Ov. Quality %	Completeness %	Correctness %	Ov. Quality %
SADG	83.2	52.6	51.3	92.2	**80.1**	**74.4**
SAD	85.3	54.8	50.07	**95.2**	64.9	60.01
SSD	71.3	60.4	48.6	94.1	67.3	61.4
NCC	56.7	42.9	32.3	77.7	40.5	34.8
NMI	49.5	51.4	33.7	55.3	62.8	41.5
CR	57.1	24.4	20.6	60.5	30.3	25.2
GRAD	33.5	31.6	19.4	35.1	40.3	23.1
CCGIP	59.5	28.5	23.9	77.8	38.8	34.9
JRD	30.2	45.2	22.1	39.6	56.7	30.4
HD	53.2	46.7	33.1	83.6	65.1	57.8
MI	40.1	49.3	28.4	41.9	51.7	30.1

$*$ *unsupervised formulation*

TABLE III: Quantitative evaluation regarding the performance of the unsupervised change detection procedure. The standard measures of detection completeness and correctness at object level have been calculated for different similarity measures for the unsupervised approach.

Fig. 6: Experimental results after the application of the developed change detection framework based on the unsupervised formulation (E_{ch} from equation (6)) for two sub-regions (case #1 and case #2) of the study area. The detected changes are shown with red color and the ground truth data with green. Both are superimposed onto the very high resolution satellite images acquired in 2006 (Quickbird, left) and 2011 (Worldview-2, right).

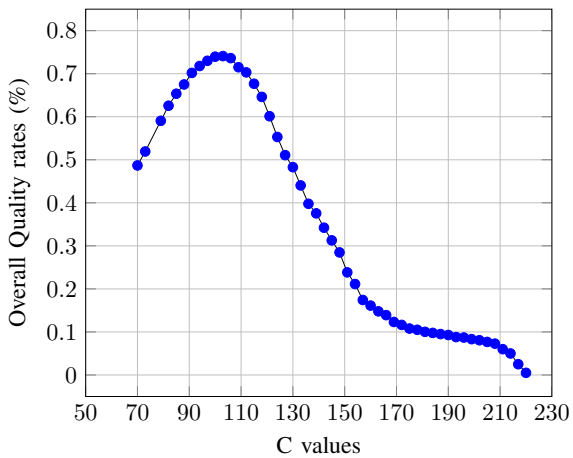

Fig. 7: The resulting Overall Quality (%) measure as a function of the value C (equation (6)) which corresponds to the fixed $Cost$ for the unsupervised formulation. The highest performance rates were acquired for values around 100 with the Overall Quality reaching a 74.4%

scoring higher in the detection completeness but produced relative more false alarms. SADG metric which is based both on spectral and edge differences delivered the highest correctness rates while managed to detect more than 92% of the existed changes.

Moreover, in order to evaluate the performance of the change detection term alone exactly the same experiments were performed with the same similarity and parameter settings with the developed framework. Note that for these experiments with only the E^*_{ch} (Table III, left), image pairs were registered beforehand based on a non-rigid procedure in order to account optimally for relief displacements which were in all images, region and dates important due to the complex terrain and acquisition angles. Results are presented in Table III (left) and indicate the significant lower performance regarding the Overall Quality *i.e.,* lower than 25% than the one achieved through the developed framework (74.4%). In particular, with the same similarity measures, E_{ch} produced several false alarms affecting significantly the detection correctness rates, indicating that the coupling and registration labels (under the proposed joint $E_{cpl} + E_{reg} + E_{ch}$ formulation) constrained more the detection process, reducing the false alarms while increasing the true positives.

In Fig. 7 the measured Overall Quality (%) is presented as a function of the Cost parameter. It can be observed that for the SADG metric the values between 90 and 120 delivered the highest rates *e.g.,* more than 70%. The performed experiments and the acquired quite promising results demonstrate the extreme potentials of the developed unsupervised framework in detecting changes related to man-made objects.

In addition, the computational efficiency of the proposed framework was also evaluated. In particular, in Table IV the computational time required for different image sizes with the SADG metric is presented. With a standard laptop equipped with an Intel Core i7-4700HQ CPU at 2.40GHz and 8GB RAM, it took less than 30 minutes for the convergence of the unsupervised registration/change detection algorithm. The computational time is mainly depending on the selected set of parameters which specify *e.g.,* the number of nodes in the grid, the number of labels or the number of iterations per level, as well as the selected similarity metric. In accordance with the literature, in all our experiments the SAD, SSD, NCC and SADG were the quickest ones, while the JRD and MI were the most time consuming.

Computational Performance

Region size *in* km^2	Image size in *pixels*	Convergence time in *min.*
0.25	1000x1000	6.3
0.5	1500x1300	20.5
1	2000x2000	27.7
1.5	2800x2140	53.3

TABLE IV: The computational performance of the developed unsupervised framework with the SADG metric optimised for man-made change detection.

C. Evaluating the performance of the Supervised Change Detection: 'from-to' change trajectories

For the evaluation of the supervised change detection framework and the detection of *from-to* change trajectories, reference/ground truth data were collected based on an attentive, laborious manual annotation. The ground truth data contained the main *from-to* change trajectories and in particular, five different classes (*i.e,* change trajectories) were annotated and are briefly described below:

- No change.
- Class # 1: Soil to vegetation (yellow color)
- Class # 2: Soil to man-made object (magenta)
- Class # 3: Vegetation to soil (cyan)
- Class # 4: Vegetation to man-made object (blue)
- Class # 5: Man-made to man-made object (red)

In Fig. 8, results after the application of the developed supervised framework and the detection of specific change trajectories are presented. In particular, on the left hand side of Fig. 8, the detected changes are superimposed with different colors on the Quickbird image of 2006. On the right hand side, a zoom into two subregions is presented for both images (dates) *i.e.* Quickbird (2006) and Worldview-2 (2011). The different types of change trajectories are indicated with different colors and are described in the corresponding legend. The majority of the detected changes belonged to Class #5 indicating that the dominant change trajectories were from man-made objects to other man-made object types. Moreover, results from the quantitative evaluation are presented in Table V.

In particular, the calculated confusion matrix after the application of the supervised approach (for the image pairs of Fig. 8) is presented based on the SADG metric. The resulting Overall Accuracy was 73.4%, while the total number of changes were more than 430. The larger number of false alarms came from the misclassification errors between Class #5 and Class #2, where 'soil to man-made' and 'man-made

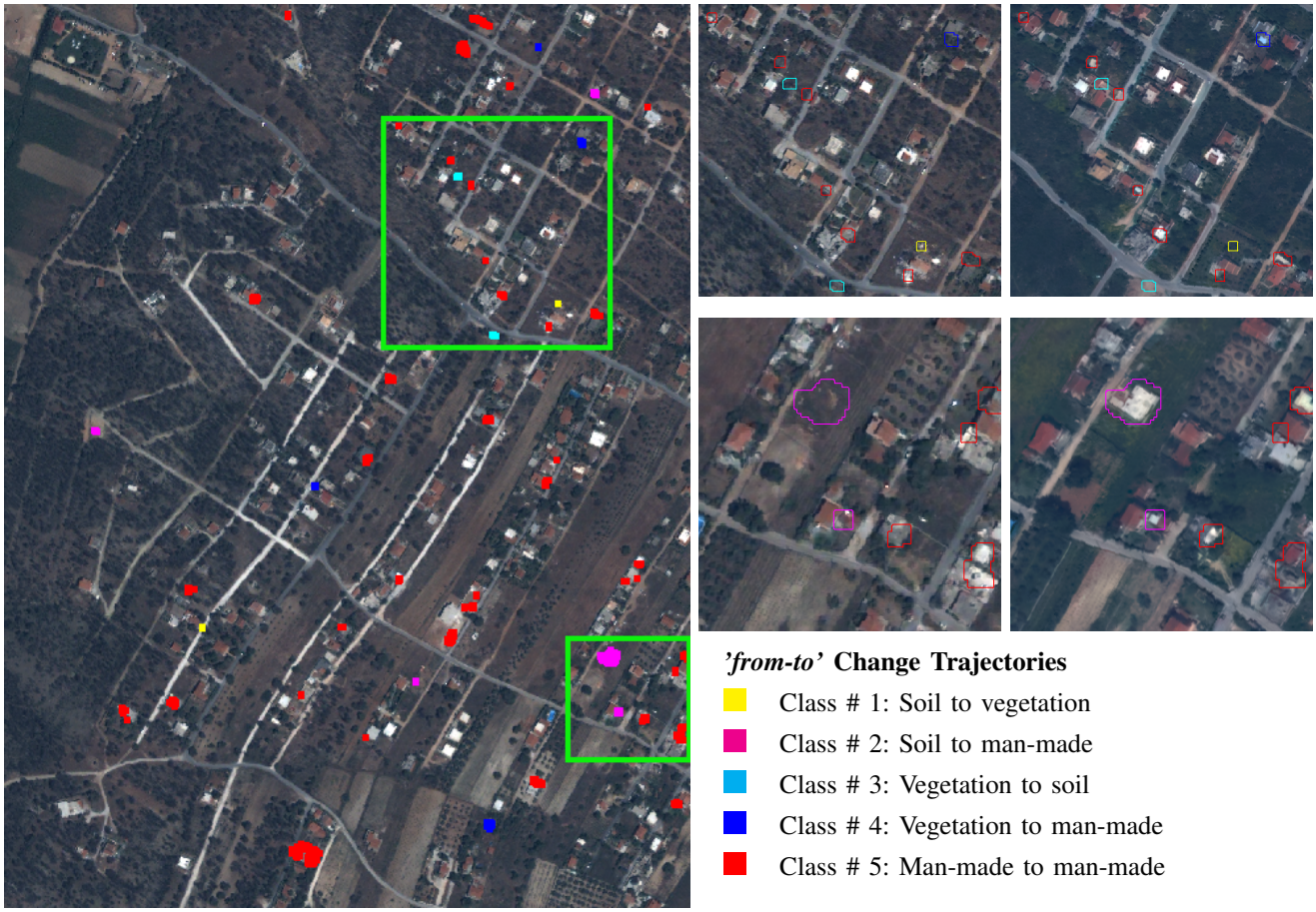

Fig. 8: Experimental results after the application of the supervised change detection framework towards the detection of *from-to* change trajectories. The detected changes (belonging to different change classes) are projected onto the Quickbird 2006 image (left). A zoom into two subregions is shown on the right hand side where the detected change trajectories are shown in both images (dates) *i.e.,* left: Quickbird (2006) and right: Worldview-2 (2011).

Evaluating the performance of the supervised change detection

	Reference Data						
# of objects	Class # 1	Class # 2	Class # 3	Class # 4	Class # 5	Total	UA (%)
Classification							
Class # 1	**38**	2	8	1	0	49	77.6
Class # 2	2	**54**	0	3	17	76	71.1
Class # 3	2	5	**58**	11	3	79	73.4
Class # 4	5	0	10	**59**	6	80	73.8
Class # 5	5	21	2	8	**117**	153	76.5
Total changes	52	82	78	82	143	**437**	
PA (%)	73.1	65.9	74.4	72.0	81.8		

Overall accuracy = 73.4% Kappa coefficient = 67.5%

TABLE V: The resulting confusion matrix after the application of the developed supervised change detection framework (for the image pairs of Fig. 8). The resulting Overall Accuracy was 73.4%, while the total number of changes were more than 430.

to man-made' changes are confused. As it can be observed, changes belonging to Class #5 resulted into the higher PA rates (82%). On the other hand, Class #2 resulted into the lower PA and UA rates. This is mainly due to the fact that the employed classification features contained mainly spectral information based on certain spectral bands and indexes which could not address the similar spectral behaviour between soil and man-made objects *e.g.,* similarities between red roof tiles and red clay soil.

V. Conclusion

In this paper we designed, developed and validated a novel framework which addresses concurrently the registration and change detection tasks in very high resolution multispectral multitemporal optical satellite data. Furthermore, an extension for classifying different types of changes is proposed. The developed method is modular, scalable and metric free. The formulation exploits a decomposed interconnected graphical model formulation where registration similarity constraints are relaxed in the presence of change detection. The unsupervised framework was optimized for the detection of changes related to man-made objects in urban and peri-urban environments. Moreover, the supervised formulation was able to detect several *from-to change* trajectories. The performed large scale experiments and the acquired quite promising results demonstrate the extreme potentials of the developed method. The integration of prior knowledge regarding texture and geometric features under a higher order formulation [10], [13] is currently under consideration and a *GPU* implementation is among the future perspectives as well.

Acknowledgment

We would like to thank the anonymous reviewers for their constructive comments and suggestions. Part of this research was funded by the 'ELKE' PhD Scholarship of the National Technical University of Athens.

References

[1] C. Benedek, M. Shadaydeh, Z. Kato, T. Szirányi, and J. Zerubia, "Multilayer markov random field models for change detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, Special Issue on Multitemporal Remote Sensing Change Detection, p. 18, 2015.

[2] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. Wiley, 2009.

[3] M. J. Canty, *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL and Python*. CRC Press, 2014, 576p.

[4] N. Champion, D. Boldo, M. Pierrot-Deseilligny, and G. Stamon, "2D building change detection from high resolution satelliteimagery: A two-step hierarchical method based on 3d invariant primitives," *Pattern Recognition Letters*, 2010.

[5] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey," English, in *Image and Signal Processing*, ser. Lecture Notes in Computer Science, A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass, and J. Meunier, Eds., vol. 6134, Springer Berlin Heidelberg, 2010, pp. 103–112.

[6] G. Doxani, K. Karantzalos, and M. Tsakiri-Strati, "Monitoring urban changes based on scale-space filtering and object-oriented classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 15, no. 0, pp. 38 –48, 2012.

[7] N. Falco, M. Mura, F. Bovolo, J. Benediktsson, and L. Bruzzone, "Change detection in vhr images based on morphological attribute profiles," *Geoscience and Remote Sensing Letters, IEEE*, 2013.

[8] A. Ghosh, B. Subudhi, and L. Bruzzone, "Integration of gibbs markov random field and hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images," *Image Processing, IEEE Transactions on*, 2013.

[9] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, "Deformable medical image registration: setting the state of the art with discrete methods," *Annual Review of Biomedical Engineering*, vol. 13, pp. 219–244, 2011.

[10] J. H. Kappes, M. Speth, G. Reinelt, and C. Schnrr, "Higher-order segmentation via multicuts," *Computer Vision and Image Understanding*, vol. 143, pp. 104 –119, 2016.

[11] K. Karantzalos, A. Sotiras, and N. Paragios, "Efficient and automated multi-modal satellite data registration through mrfs and linear programming," *IEEE Computer Vision and Pattern Recognition Workshops*, 2014.

[12] K. Karantzalos, "Recent advances on 2d and 3d change detection in urban environments from remote sensing data," English, in *Computational Approaches for Urban Environments*, ser. Geotechnologies and the Environment, M. Helbich, J. Jokar Arsanjani, and M. Leitner, Eds., 2015, pp. 237–272.

[13] D. Khandelwal, K. Bhatia, C. Arora, and P. Singla, "Lazy generic cuts," *Comput. Vis. Image Underst.*, vol. 143, no. C, pp. 80–91, Feb. 2016.

[14] N. Komodakis, N. Paragios, and G. Tziritas, "Mrf energy minimization and beyond via dual decomposition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2011.

[15] N. Komodakis and G. Tziritas, "Approximate labeling via graph cuts based on linear programming," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 8, pp. 1436–1453, 2007.

[16] N. Komodakis, G. Tziritas, and N. Paragios, "Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies," *Comput. Vis. Image Underst.*, vol. 112, no. 1, pp. 14–29, Oct. 2008.

[17] P. Koutsourakis, L. Simon, O. Teboul, G. Tziritas, and N. Paragios, "Single view reconstruction using shape grammars for urban environments," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 1795–1802.

[18] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, *Image Registration for Remote Sensing*. Cambridge University Press, 2011, 497p.

[19] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Q. Du, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009-2010 data fusion contest," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 2012.

[20] S. Marchesi, F. Bovolo, and L. Bruzzone, "A context-sensitive technique robust to registration noise for change detection in vhr multispectral images," *Image Processing, IEEE Transactions on*, 2010.

[21] C. Marin, F. Bovolo, and L. Bruzzone, "Building change detection in multitemporal very high resolution sar images," *Geoscience and Remote Sensing, IEEE Transactions on*, 2015.

[22] F. Pacifici and F. Del Frate, "Automatic change detection in very high resolution images with pulse-coupled neural

networks," *Geoscience and Remote Sensing Letters, IEEE*, 2010.

[23] T. Pollard and J. Mundy, "Change detection in a 3-d world," in *Computer Vision and Pattern Recognition CVPR*, 2007.

[24] C. Pratola, F. Del Frate, G. Schiavon, and D. Solimini, "Toward fully automatic detection of changes in suburban areas from vhr sar images by combining multiple neural-network models," *Geoscience and Remote Sensing, IEEE Transactions on*, 2013.

[25] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *Image Processing, IEEE Transactions on*, vol. 14, no. 3, pp. 294–307, 2005.

[26] K. Sakurada, T. Okatani, and K. Deguchi, "Detecting changes in 3d structure of a scene from multi-view images captured by a vehicle-mounted camera," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013.

[27] A. Shekhovtsov, "Higher order maximum persistency and comparison theorems," *Computer Vision and Image Understanding*, vol. 143, pp. 54 –79, 2016.

[28] P. Singh, Z. Kato, and J. Zerubia, "A multilayer markovian model for change detection in aerial image pairs with large time differences," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, 2014.

[29] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 7, pp. 1153–1190, 2013.

[30] A. Taneja, L. Ballan, and M. Pollefeys, "City-scale change detection in cadastral 3d models using images," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013.

[31] F. Tupin, J. Inglada, and J.-M. Nicolas, *Remote Sensing Imagery*. Wiley, 2014, 368p.

[32] M. Vakalopoulou and K. Karantzalos, "Automatic descriptor-based co-registration of frame hyperspectral data," *Remote Sensing*, vol. 6, no. 4, p. 3409, 2014.

[33] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Simultaneous registration and change detection in multitemporal, very high resolution remote sensing data," 2015.

[34] M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, "Unsupervised change detection with kernels," *Geoscience and Remote Sensing Letters, IEEE*, 2012.

[35] C. Wang, N. Komodakis, and N. Paragios, "Markov random field modeling, inference & learning in computer vision & amp; image understanding: A survey," *Computer Vision and Image Understanding*, vol. 117, no. 11, pp. 1610 –1627, 2013.

[36] C. Wang, M. de La Gorce, and N. Paragios, "Segmentation, ordering and multi-object tracking using graphical models," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 747–754.

[37] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[38] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977 –1000, 2003.

Maria Vakalopoulou received an Engineering Diploma (2011) in survey engineering from the National Technical University of Athens (NTUA), Greece. She is currently pursuing a PhD degree from the same university at the Remote Sensing Laboratory. The subject of her PhD includes image registration, classification and change detection from multitemporal very high resolution data. During 2014 and 2015 she was a visiting student at Universite Paris-Est, Ecole des Ponts ParisTech under the supervision of Prof. Nikos Komodakis and Prof. Nikos Paragios. Her research interests includes remote sensing, computer vision and machine learning.

Konstantinos Karantzalos (Senior Member, IEEE) received his engineering diploma from the National Technical University of Athens (NTUA, Greece) and his PhD (2007, NTUA) in collaboration with Ecole Nationale de Ponts et Chaussees (CERTIS, ENPC, France). In 2007, he joined the Center for Visual Computing of the Department of Applied Mathematics at Ecole Centrale de Paris (France). He is currently an Assistant Professor of Remote Sensing at the National Technical University of Athens, joining the Remote Sensing Laboratory. His teaching and research interests include geoscience and remote sensing, big data, hyperspectral image analysis, computer vision and pattern recognition, environmental monitoring and precision agriculture. He has several publications in top-rank international journals and conferences and a number of awards and honors for his research contributions. He has more than 15 years of research experience, involved with more than 14 EU and national excellence/competitive research projects. *http://users.ntua.gr/karank*

Nikos Komodakis is an associate professor at Universite Paris-Est, Ecole des Ponts ParisTech and a research scientist at the UMR Laboratoire d'informatique Gaspard-Monge, CNRS. He is also an affiliated adjunct professor at Ecole Normale Suprieure de Cachan. He currently serves as an editorial board member for the International Journal of Computer Vision and the Computer Vision and Image Understanding journal. His research interests are in the areas of computer vision, image processing, machine learning and medical imaging, and has published numerous papers in the most prestigious journals and conferences from the above domains.

Nikos Paragios (B.Sc'94, M.Sc'96, PhD'00, HDR'05) is professor of applied mathematics at CentraleSupelec/Universit Paris-Saclay & affiliated research scientific leader at Inria. He is an IEEE fellow, senior fellow of the Institut Universitaire de France and member of the scientific council of Safran conglomerate. Professor Paragios is the editor in chief of the Computer Vision and Image Understanding journal and serves on the editorial board of the Medical Image Analysis and SIAM journal on Imaging Sciences. He has published more than two hundred papers in the areas of computer vision, biomedical imaging and machine learning.