Simultaneous registration, segmentation and change detection from multisensor, multitemporal satellite image pairs.

Maria Vakalopoulou, Christos Platias, Maria Papadomanolaki, Nikos Paragios, Konstantinos Karantzalos

▶ **To cite this version:**

Maria Vakalopoulou, Christos Platias, Maria Papadomanolaki, Nikos Paragios, Konstantinos Karantzalos. Simultaneous registration, segmentation and change detection from multisensor, multitemporal satellite image pairs.. International Geoscience and Remote Sensing Symposium (IGARSS), Jul 2016, Beijing, China. 10.1109/IGARSS.2016.7729469 . hal-01413373

HAL Id: hal-01413373

https://inria.hal.science/hal-01413373v1

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.

SIMULTANEOUS REGISTRATION, SEGMENTATION AND CHANGE DETECTION FROM MULTISENSOR, MULTITEMPORAL SATELLITE IMAGE PAIRS

M. Vakalopoulou[*1], *C. Platias*[1], *M. Papadomanolaki*[1], *N. Paragios*[2], *K. Karantzalos*[1]

[1]Remote Sensing Lab., National Technical University of Athens, Athens, Greece
[2]Center for Visual Computing, Ecole Centrale de Paris, Paris, France

ABSTRACT

In this paper, a novel generic framework has been designed, developed and validated for addressing simultaneously the tasks of image registration, segmentation and change detection from multisensor, multiresolution, multitemporal satellite image pairs. Our approach models the inter-dependencies of variables through a higher order graph. The proposed formulation is modular with respect to the nature of images (various similarity metrics can be considered), the nature of deformations (arbitrary interpolation strategies), and the nature of segmentation likelihoods (various classification approaches can be employed). Inference of the proposed formulation is achieved through its mapping to an over-parametrized pairwise graph which is then optimized using linear programming. Experimental results and the performed quantitative evaluation indicate the high potentials of the developed method.

Index Terms— Multimodal, Deimos, Iris, Satellite, Video, Classification, Deep learning, CNN

1. INTRODUCTION

Currently, there are more than forty satellites in operation offering remote sensing data solutions from multiple sensors with various spatial, spectral and temporal resolutions. By 2024, this number is expected to climb to 160, including earth observation satellites from both private enterprises and governments, excluding the numerous nanosatellites and cubesats. It is, therefore, critical to build the capacities for operational exploitation of these multisensor, multiresolution, multitemporal datasets towards adequate return on initial investments as well as efficient spatio-temporal environmental monitoring.

Along with data handling, serving and pre-processing issues, multimodal data fusion techniques hold a primary role in information extraction and exploitation. Despite the important research effort during the last decades ([1] and the references therein) there are still important challenges to be addressed. Towards this direction, the Image Analysis and Data Fusion Committee of the IEEE Geoscience and Remote Sensing Society organises annual data fusion contests [2] in order to highlight efficient approaches [3–5] and way forward. In this paper, a novel methodology has been developed able to ingest information from multisensor datasets of different spatial, spectral and temporal resolutions as those of the 2016 contest.

In particular, we extended the formulation of [6, 7] by adding another graph which is related to the segmentation problem. The proposed formulation jointly considers data-driven costs regarding the segmentation likelihoods (various classification approaches can be employed *e.g.,* [8]), registration metrics (*e.g.,* similarity metrics) and change detection scores. These energies are efficiently coupled with local geometric constraints in the context of a higher order graph. Reduction methods are used to map this graph into a pairwise one which is then optimized using efficient linear programming. Promising experimental results indicating less than 2 pixels in mean displacement errors regarding the registration, above 77% in most cases regarding the segmentation completeness and correctness rates and around or above 70% in change detection demonstrate the high potentials of the developed method.

2. METHODOLOGY

2.1. Energy Formulation

Following the notations of [6, 7], here, we add another graph G_{seg} associated with the image segmentation problem. The first graph G_{reg} involves nodes where the labels correspond to deformations vectors from the registration process (mapping between the source and the target images), the second G_{ch} refers to nodes with binary labels expressing changes in the temporal domain, while the last G_{seg} refers to the labels representing the segmentation structures being present in the image. The proposed energy function (1) consists of three terms and couples the three different graphs to one:

$$\boldsymbol{E_{reg,ch,seg}(l_p^{reg}, l_p^{ch}, l_p^{seg})} = E_{ch} + E_{seg} + E_{reg} \quad (1)$$

The labels for each node of the coupling graph will be $l_p = [l_p^{ch}, l_p^{seg}, l_p^{reg}]$; $l_p^{ch} \in \{0, 1\}$ represent the labels for change detection, $l_p^{seg} \in \{0, 1\}$ are the labels for binary segmentation and l_p^{reg} are the registration labels (with $l_p^{reg} \in \Delta$

*Corresponding author's email: mariavak@central.ntua.gr

where $\Delta = [d^1, \ldots, d^n]$ corresponds to all possible displacements). Concluding the label space can be summarized as $L \in \{0,1\} \times \{0,1\} \times \Delta$.

Both the registration and change detection terms are following the same formulations as in [6, 7], while the goal of segmentation is to assign the correct segmentation label to each node of the target image. The segmentation graph, therefore, contains a potential term with the classification score for each class (2) and a pairwise term (3) which penalises different segmentation labels. Let us denote by $v_{s,I_t}(x)$ the feature vector for the target image at every point x and by $\Psi_{l_p^{seg}}(\cdot)$ the classification score for each label l_p^{seg}. $\hat{\eta}$ is the projection function.

$$V_{seg}(l_p^{seg}) = \int_{\Omega} \hat{\eta}(||x-p||)\Psi_{l_p^{seg}}(v_{s,I_t}(x))dx \quad (2)$$

$$V_{pq,seg}(l_p^{seg}, l_q^{seg}) = ||l_p^{seg} - l_q^{seg}|| \quad (3)$$

It should be noted that for simplicity, the segmentation graph is defined only at the target image, and co-segmentation will be achieved through the coupling of the segmentation labels with the change detection ones.

2.2. Coupling the Energy Terms

The coupling between the three terms is performed by one potential term (4) which penalises different segmentation labels in the absence of change and reverse, between the source and the target images for all possible displacements.

$$V_{reg,ch,seg}(l_p^{reg}, l_p^{ch}, l_p^{seg}) = (1 - l_p^{ch}) \int_{\Omega} \hat{\eta}(||x-p||)\Psi_{l_p^{seg}}(v_{s,I_s}(x + d^{l_p^{reg}}))dx + l_p^{ch} \int_{\Omega} \hat{\eta}(||x-p||) (\Psi_{(1-l_p^{seg})}(v_{s,I_s}(x + d^{l_p^{reg}})))dx \quad (4)$$

Let us abuse the notation and consider a node with an index $p \in G$ (we recall that the three graphs are identical) corresponding to the same node throughout the three graphs (G_{reg}, G_{ch}, G_{seg}). Then the global energy (5) can be expressed as follows:

$$\begin{aligned} \boldsymbol{E_{reg,ch,seg}(l_p^{reg}, l_p^{ch}, l_p^{seg})} = w_1 \sum_{p \in G} V_{reg,ch}(l_p^{reg}, l_p^{ch}) + \\ w_2 \sum_{p \in G} V_{reg,ch,seg}(l_p^{reg}, l_p^{ch}, l_p^{seg}) + w_3 \sum_{p \in G_{seg}} V_{seg}(l_p^{seg}) + \\ w_4 \sum_{p \in G_{reg}} \sum_{q \in N(p)} V_{pq,reg}(l_p^{reg}, l_q^{reg}) + \\ w_5 \sum_{p \in G_{ch}} \sum_{q \in N(p)} V_{pq,ch}(l_p^{ch}, l_q^{ch}) + \\ w_6 \sum_{p \in G_{seg}} \sum_{q \in N(p)} V_{pq,seg}(l_p^{seg}, l_q^{seg}) \end{aligned} \quad (5)$$

3. EXPERIMENTAL RESULTS AND EVALUATION

The evaluation of the developed framework was performed on the '2016 IEEE GRSS Data Fusion Contest' dataset which includes one Deimos-2 multispectral image acquired in March'15 [***D1***], one Deimos-2 multispectral image acquired in May'15 [***D2***] and one video sequence from the Iris RGB sensor acquired in July'15 [***V***]. The two Deimos images were radiometrically corrected and then pansharpened based on the standard High Pass Filter method, resulting into an overlapping image pair of approximately 12760 by 11000 pixels. The overlapping image pairs with frames from the Iris video sequence were approximately 4720 by 2680 pixels. In order to employ an additional image/map which could serve as a reference/target map while contributing on the automation of the subsequent training procedure, an image mosaic [***G***] as well as the corresponding map were downloaded from Google Map APIs. All raw unregistered data as well as several experimental results can be viewed here:
`http://users.ntua.gr/karank/Demos/DemoContest16.html`

Classification scores: In order to estimate the required in (2) and (4) classification scores $\Psi_{l_p^{seg}}(\cdot)$ for each label l_p^{seg} and each modality, a deep neural network classification approach was employed. Such deep learning frameworks have reported high classification accuracy rates for a number of cases [8–10]. Training, testing and validation polygons were created under a semi-automated procedure and after an initial registration of all datasets to Google's image mosaic in order to relate every pixel to a Google's map color (*i.e.,* terrain class). Spectral analysis on the derived numerous polygons and probabilities were employed (as in [9]) in order to define 8 terrain classes *i.e.,*: *Roads*, *Buildings*, *Building Shadows*, *Soil*, *Sea*, *Ship/vessels*, *Vegetation* and *Vegetation shadows*. Clouds on the Deimos March image were manually annotated, while in the Iris video sequence/frame the class *Vegetation shadows* were merged with *Vegetation*.

The training for the two Deimos images (***D1***, ***D2***) was performed on the large (12.760 by 11.000 pixels) overlapping region with 8 classes (*e.g.,* Figure 1), while for the Iris (***V***) 7

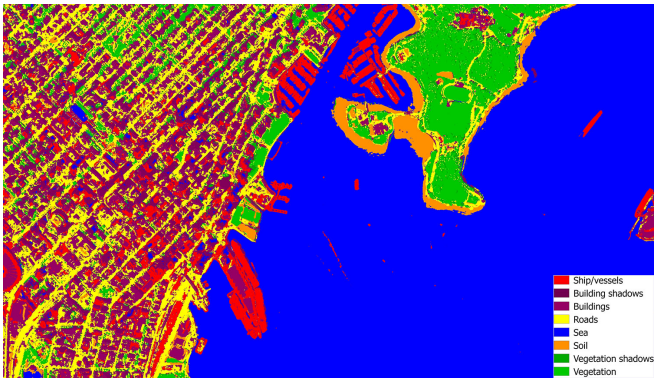

Fig. 1: Classification map with the dominating scores for the Deimos-2 May'15 (***D2***) image.

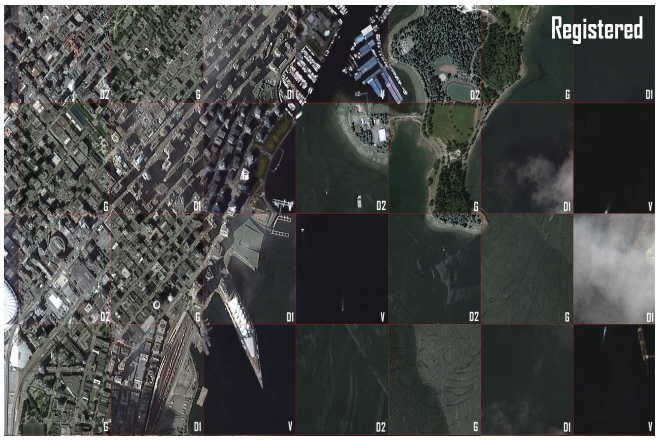

Fig. 2: Chessboard visualization with the resulted registered multisensor data.

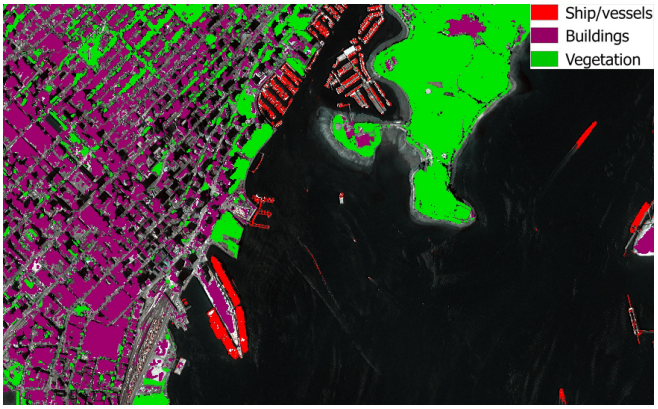

Fig. 3: Segmentation results after the application of the developed framework on the ***D1-D2*** pair. Results on ***D2***.

classes were employed. Based on the derived polygons numerous patches of size 21x21 were created by centring each patch to annotated pixels [10]. Approximately 200.000 randomly selected patches for each class were used for the satellite Deimos data, while 50.000 for the Iris video frames. The training procedure was accomplished with the use of a Convolutional Neural Network (CNN) which consisted of 10 layers. The network was made of two convolutional layers each followed by tangent and max pooling layers, and ending to two fully connected, a tangent and a linear layers. The model was trained with a learning rate of 1 for 40 epochs, while every 2 epochs the learning rate was reduced at half. Regarding the CNN architecture for the Iris video RGB frames, the overall implementation was the same, whereas the size of each patch was 3x21x21.

Implementation details: Regarding the employed parameters, in a similar way as in [6] we used 3 image and 4 grid levels. The parameter C was set to 1800. The Sum Absolute of Differences plus Gradient Inner Products (SADG) was used as a similarity metric. All the parameters and the weights had been set using grid search. For the optimisation procedure, the FastPD was employed[1].

Registration results: For the validation of the registration results (Figure 2), several Ground Control Points (GCPs) were manually collected in all resulting image pairs. In Table 1 the mean displacement errors for both axis and the distance in pixels are presented. It should be mentioned, that the registration process didn't manage to address the largest relief displacements of the tallest buildings/skyscrapers of this part of Vancouver and these errors hindered both the segmentation and change detection results. All other building rooftops, roads, terrain classes were registered with a sub-pixel accuracy. A quantitative comparison is also provided in Table 1 (first four columns) based on [11].

[1] ***D2-D1***: $w_3 = 340, w_4 = 30, w_5 = 4.5, w_6 = 7.5$, ***V-D2***: $w_3 = 340, w_4 = 40, w_5 = 10, w_6 = 14$

Mean Displacement Errors

in pixels	D1 to G	D2 to G	V to G	V *frame to frame*	D1 to D2	V to D2
Dx	1.09	1.22	0.93	0.84	1.12	1.04
Dy	1.62	1.49	1.73	0.92	1.59	1.61
DS	1.95	1.93	1.97	1.24	1.94	1.92

Table 1: Quantitative evaluation for the registration results.

Segmentation results: For the segmentation and change detection results, the quantitative evaluation of the framework was performed using the completeness, correctness and overall quality criteria at object level. The resulting true positives, false negatives and false positives were calculated on the validation dataset after the application of the developed framework. The framework was tested for three different classes *i.e., Buildings*, *Ship/vessels* and *Vegetation*. After the optimisation and based on the polygon of the class *Sea* derived automatically from Google map all segmented as *Buildings* objects in the *Sea* were ignored and respectively all *Ship/vessels* in the land, as well.

As one can observes in Figure 3, although the classification scores (*e.g.,* Figure 1) constrain significantly the solution, the developed framework integrates both scores and similarities ameliorating the segmentation results in several image regions. The quantitative evaluation (Table 2) indicated that the detection completeness rates were above 78% (apart from the class *Buildings* in the ***D1-D2*** pair) while the detection correctness rates were above 72% in all cases. The higher rates were for the class *Vegetation* indicating that the NIR Deimos-

	Completeness	Correctness	Overall Quality
	Deimos March - Deimos May		
Ship/vessel	81.4%	78.0%	66.2%
Vegetation	83.9%	88.3%	75.6%
Buildings	68.9%	77.4%	57.4%
	Iris July - Deimos May		
Ship/vessel	79.0%	77.9%	65.6%
Vegetation	82.5%	86.2%	72.8%
Buildings	78.8%	72.2%	60.5%

Table 2: Quantitative evaluation for the segmentation results

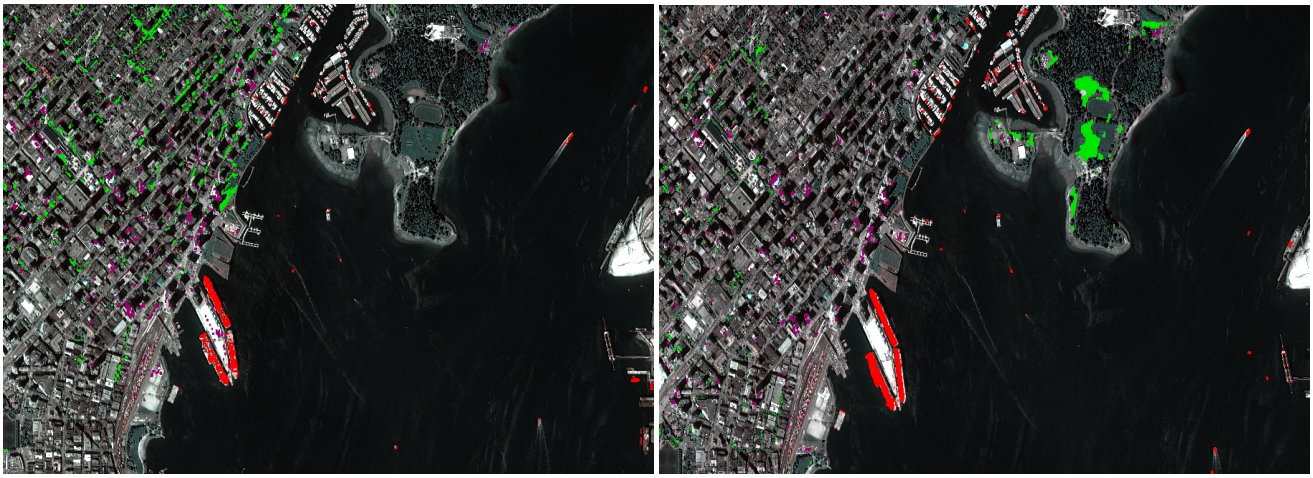

Fig. 4: Change Detection from multitemporal, multi-sensor between: (a) a Deimos March'15 and a Deimos May'15 (***D1-D2***, left), (b) an Iris video sequence (first frame) and a Deimos May'15 (***V-D2***, right).

	Completeness	Correctness	Overall Quality
	Deimos March - Deimos May		
Ship/vessel	68.6%	66.7%	51.1%
Vegetation	80.2%	82.3%	68.4%
Buildings	69.2%	67.4%	51.8%
	Iris July - Deimos May		
Ship/vessel	70.6%	69.5%	53.9%
Vegetation	81.1%	79.6%	67.2%
Buildings	71.3%	65.6%	51.9%

Table 3: Quantitative evaluation for the change detection.

2 band significantly contributed in class separation. Most segmentation errors were due to false alarms near the port, pier and ship wake on the sea, while *Buildings* and *Roads* were in certain cases confused.

Change Detection results: Qualitatively, the same errors were observed on the change detection results (Figure 4) for both image pairs. Quantitative evaluation results (Table 3) indicated lower completeness and correctness rates than the segmentation task, as expected. This was mainly due to a number of false positives in the dense urban regions were the relief displacements were significant due to the tallest buildings and skyscrapers.

4. CONCLUSION

A novel generic graph-based framework was developed addressing simultaneous registration, segmentation and change detection in multisensor data of different spectral, spatial and temporal resolutions. The quite promising experimental results indicated: *(i)* for the registration task a mean displacement error of less than 2 pixels, *(ii)* for the segmentation task in most cases completeness and correctness rates above 77% and *(iii)* for the change detection around or above 70%. The main errors derived from the important relief displacements around the tallest buildings and from the quite similar reflectance spectra of the different man-made objects including ship/vessels.

5. ACKNOWLEDGMENTS

The authors would like to thank Deimos Imaging for acquiring and providing the data used in this study, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee.

6. REFERENCES

[1] L. Gmez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal Classification of Remote Sensing Images: A Review and Future Directions," *Proceedings of the IEEE*, vol. 103, pp. 1560–1584, 2015.

[2] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1585–1601, Sept 2015.

[3] 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," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 331–342, Feb 2012.

[4] C. Berger, M. Voltersen, and R. Eckardt et al., "Multi-Modal and Multi-Temporal Data Fusion: Outcome of the 2012 GRSS Data Fusion Contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1324–1340, June 2013.

[5] W. Liao, X. Huang, F. Van Coillie, and S. Gautama et al., "Processing of Multiresolution Thermal Hyperspectral and Digital Color Data: Outcome of the 2014 IEEE GRSS Data Fusion Contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2984–2996, 2015.

[6] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Simultaneous Registration and Change Detection in Multitemporal, Very High Resolution Remote Sensing Data," in *The IEEE Conference on Computer Vision and Pattern Recognition Workshops*, June 2015.

[7] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. 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, (in press)*, 2016.

[8] A. Lagrange, B. Le Saux, A. Beaupre, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Randrianarivo, and M. Ferecatu, "Benchmarking classification of earth-observation data: From learning explicit features to convolutional networks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 4173–4176.

[9] Volodymyr Mnih, *Machine Learning for Aerial Image Labeling*, Ph.D. thesis, University of Toronto, 2013.

[10] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building Detection in Very High Resolution Multispectral Data with Deep Learning Features," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 1873–1876.

[11] K. Karantzalos, A. Sotiras, and N. Paragios, "Efficient and Automated Multi-Modal Satellite Data Registration through MRFs and Linear Programming," *IEEE CVPR Workshops*, 2014.