BUILDING DETECTION IN VERY HIGH RESOLUTION MULTISPECTRAL DATA WITH DEEP LEARNING FEATURES

M. Vakalopoulou[1,2], *K. Karantzalos*[1], *N.Komodakis*[2], *N. Paragios*[3]

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

mariavak@mail.ntua.gr; karank@central.ntua.gr; nikos.komodakis@enpc.fr; nikos.paragios@ecp.fr

ABSTRACT

The automated man-made object detection and building extraction from single satellite images is, still, one of the most challenging tasks for various urban planning and monitoring engineering applications. To this end, in this paper we propose an automated building detection framework from very high resolution remote sensing data based on deep convolutional neural networks. The core of the developed method is based on a supervised classification procedure employing a very large training dataset. An MRF model is then responsible for obtaining the optimal labels regarding the detection of scene buildings. The experimental results and the performed quantitative validation indicate the quite promising potentials of the developed approach.

Index Terms— Machine learning, deep convolutional networks, ImageNet, man made objects, extraction

1. INTRODUCTION

Building detection from satellite remote sensing data has been one of the most challenging tasks with important research and development efforts during the last decades. Recent quantitative results from the ISPRS (WGIII/4) benchmark on urban object detection and 3D building reconstruction [1] indicated that, in 2D, buildings can be recognized and separated from the other terrain objects, however, there is room for improvement towards the detection of small building structures and the precise delineation of building boundaries. Depending on the type and resolution of the remote sensing data a lot of different approaches have been proposed in the literature both pixel- and object-based ones [2–5].

Very high resolution (VHR) satellite data are absolutely required for tackling the specific problem, while any spectral information more than the standard RGB enhance the discrimination capabilities between the different man-made objects, soil, *etc*. Moreover, elevation data like Lidar, DSMs, *etc* can significantly ameliorate the detection procedure, however, they are not yet regarded as a cost-effective solution for large scale mapping and change detection applications [2, 6].

Fig. 1. Automated building detection from very high resolution multispectral data. The developed algorithm managed to detect efficiently scene's buildings. The ground truth data are shown with a green color, while the detected buildings with red.

Regarding the classification and the detection procedure, the way one constructs the feature vector which contains the different characteristics of every class is of major importance and in particular, determines the accuracy of the final result. Generally speaking, in most approaches and research efforts the feature vector can consist of a combination of spectral bands, morphological filters [7], texture [8] and point descriptors [9], gradient orientation [10], *etc*. Recently, Convolution Neural Networks (CNN) have been largely employed in object detection [11] and classification [12] setting the state-of-the-art in many computer vision and machine learning applications.

In particular, a large deep convolutional neural network, trained in the ImageNet dataset, has been created and applied in various classification problems [13] with quite promising results. The network contains eight learned layers from which five convolutional and three fully connected, 60 million parameters and 650000 neurons. In order to reduce overfitting a regularization method, namely dropout, has been applied in the fully connected layers.

Inspired from [13], in this paper we propose a supervised building detection procedure based on the ImageNet framework, while integrating certain spectral information by

employing multispectral band combinations into the training procedure. The building detection was addressed through a binary classification procedure based on SVM classifier. During the last processing step, the classification result was refined by solving an MRF problem using powerful linear programming. The experimental results and the performed quantitative validation indicate the quite promising potentials of the developed approach Figure 1.

2. METHODOLOGY

The developed methodology is highly based on a comprehensive training procedure. The training model has been contracted in order to solve a binary problem *i.e.,* building or not-building. Based on ground truth data, the training dataset contains patches centred on buildings and randomly created patches for the not-building classes. Each of these patches is inserted into the ImageNet framework and the features of the FC7 layer are extracted and used as the feature vector (feature dimensionality 4096) for training an SVM classifier. Spectral information has been integrated into ImageNet since it has been trained using standard RGB images. To this end, we have created for every patch (both building and not-building classes) two different band combinations using the red, green, blue and near infrared spectral bands. The dimensionality of the feature vector was therefore n x (2*4096) where n is the number of patches and 4096 is the feature dimensionality of the FC7 layer.

Regarding the learning procedure, a patch of 28x28 pixels (which corresponds approximately to the average size of building patches) is created every 3 pixels in the image. Then the feature vector, using the same procedure as in the training, is generated. For the classification, the same SVM model has been employed for all the test images. The resulting classification map delivers for each pixel a calculated score for each class. During the last processing step, building are extracted through the application of an MRF-based model. The goal is to minimize an energy function which is capable to segment the image into building and not-building classes based on the classification scoring.

Let us firstly consider an undirected graph $G = (N, E)$ with nodes $N = (1, 2, \ldots, i, \ldots, n)$ and edges $E\,(i, j)$. Each node of the graph corresponds to a pixel in the image and each pixel is connected to the four neighbours pixels. The label space for the specific problem will be: $l \in \{0, 1\}$ where 0 is the label for the not-building class and 1 the label for the building. The energy that we will try to minimize using the G can be formulated as:

$$E_{seg} = w_1 \cdot \sum_i V_i(l_i) + w_2 \cdot \sum_i \sum_j V_{i,j}(l_i, l_j) \quad (1)$$

The first term (V_i) corresponds to the unary term and contains the scores of the classifier for each node. The second term ($V_{i,j}$) corresponds to the pairwise term that penalizes neighbouring nodes labelled differently.

The unary term for each node and for all the labels is formulated as bellow:

$$V_i(l) = l_i \cdot e^{-P_b(i)} - (1 - l_i) \cdot e^{-P_{nb}(i)} \quad (2)$$

$P_b(i)$ is the score of the classifier for the building class and $P_{nb}(i) = 1 - P_b(i)$ is the score for the not-building class.

The pairwise term which penalizes neighbouring nodes with different segmentation labels is formulated as follow:

$$V_{i,j}(l) = \alpha \cdot ||l_i - l_j|| \quad (3)$$

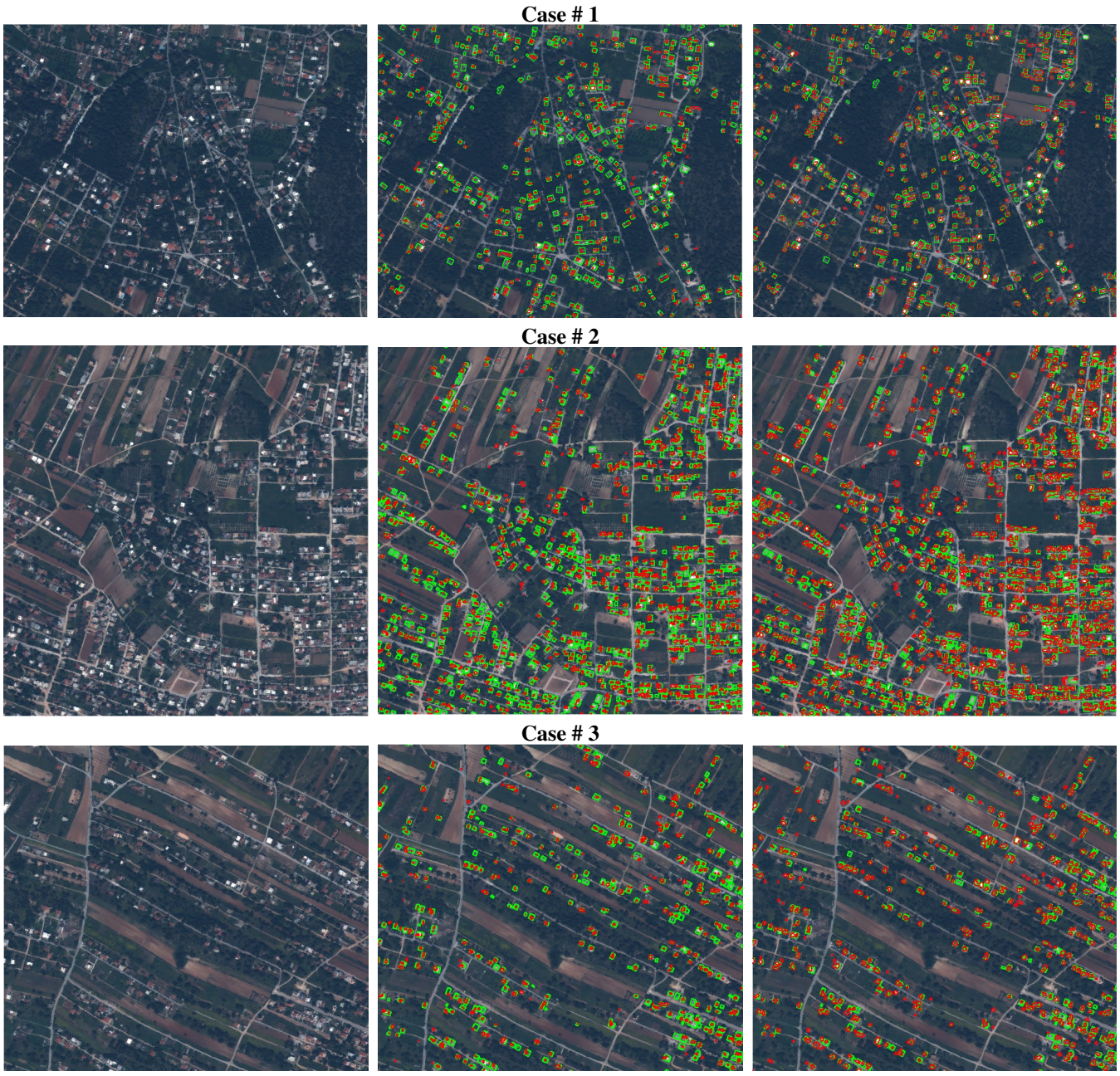
The penalty (α) is a constant value and it is responsible for the smoothness in the final result. For the optimization we use FastPD framework [14] which is based on the dual theorem of linear programming.

3. EXPERIMENTAL RESULTS AND EVALUATION

The proposed building detection framework was applied to various VHR multispectral satellite data. In particular, different VHR images like QuickBird (with 4 spectral bands) and WorldView-2 (with 8 spectral bands) have been used for the validation procedure. All the satellite imagery was acquired between the years of 2006 and 2011 and cover approximately a 9 km^2 region in the East Prefecture of Attica in Greece. All data have been pre-processed with standard radiometric correction algorithms, while pansharpening algorithms have been applied to increase the spacial resolution (*i.e.,* 0.6m for the QuickBird and 0.5m for the WorldView-2 images).

Moreover, in order to create the training data, each image has been divided into twelve subimages and half of them have been randomly chosen for training and the other half for validation. Overall 19000 patches containing buildings have been used as training data and 3 times more randomly selected patches for the not-building category. The ground truth which contained the accurate location of the buildings has been manually created and annotated after an intensive, attentive and laborious photo-interpretation done by an expert.

The validation of the developed building detection framework has been performed through the qualitative (Figure 2) and quantitative (Table 1) comparison of the detection results with the ground truth. The standard measures Completeness, Correctness and Quality have been employed and calculated at object level. To this end, based on the ground truth data, the True Positives (TP), False Positives (FP) and False Negatives (FN) were calculated for every case. In particular, TP represent the buildings that have been identified correctly, the FN represent the buildings that have not been detected and the FP correspond to false alarms *i.e.,* objects that were detected but are not actually buildings.



(a) A natural RGB composite (b) The SVM classification output (c) Final result after the MRF optimization

Fig. 2. Building detection results after the application of the developed algorithm. The ground truth data are shown with a green color, while the detected buildings with a red one.

$$Compl = \frac{TP}{TP + FN}, \quad Corr = \frac{TP}{TP + FP}, \quad Qual = \frac{TP}{TP + FP + FN} \tag{4}$$

In Figure 2, results after the application of the developed automated building detection framework from very high resolution multispectral data are shown. The developed algorithm managed to detect efficiently scene's buildings. In all the cases Table 1, the detected Correctness (and Completeness) rates are above 86% (78% respectively) while the average score reaches the 90% (80% respectively). Moreover, the optimization MRF-based procedure ameliorates the final detection results as it can significantly diminish the number of false positives objects. Finally, after a close look in Figure 2 one can observes that the number of false alarms (FP) in all cases is lower than the buildings that the algorithms did

Images	method	TP	FN	FP	Compl.	Corr.	Qual.
Case #1	class	381	102	61	79%	77%	70%
	mrf	388	83	34	**82%**	86%	**76%**
Case #2	class	682	220	108	76%	86%	68%
	mrf	633	179	59	78%	91%	73%
Case #3	class	278	77	60	78%	82%	67%
	mrf	297	80	27	79%	**92%**	74%
All Cases (mrf)		**1318**	**342**	**120**	**80%**	**90%**	**74%**

Table 1. Quantitative Evaluation Results. The calculated detection Completeness, Correctness and Quality rates from three different cases with and without the MRF refinement step.

not detect, which has been verified also by the quantitative validation Table 1.

4. CONCLUSIONS

In this paper, building detection has been addressed through a binary classification task based on deep learning features. By employing the powerful CNNs, the huge pretrainted ImageNet network and by integrating additionally spectral information during the training procedure, the calculated deep features can account for the building to not-building object discrimination. The quantitative validation indicated quite promising results with significant high detection completeness and correctness rates. The integrated MRF optimization significantly ameliorated the final building detection map. Future work involves multi-class learning towards the classification of various classes in both single and multi-temporal VHR datasets.

5. ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF)-Research Funding Program THALES: Reinforcement of the interdisciplinary and/or inter-institutional research and innovation.

6. REFERENCES

[1] Franz Rottensteiner, Gunho Sohn, Markus Gerke, Jan Dirk Wegner, Uwe Breitkopf, and Jaewook Jung, "Results of the ISPRS benchmark on urban object detection and 3d building reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 93, no. 0, pp. 256 – 271, 2014.

[2] K. Karantzalos and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 1, pp. 133–144, Jan 2009.

[3] K. Karantzalos and N. Paragios, "Large-scale building reconstruction through information fusion and 3-d priors," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 5, pp. 2283–2296, May 2010.

[4] S. Jabari, Yun Zhang, and A. Suliman, "Stereo-based building detection in very high resolution satellite imagery using ihs color system," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, July 2014, pp. 2301–2304.

[5] A. Ferro, D. Brunner, and L. Bruzzone, "Building detection and radar footprint reconstruction from single vhr sar images," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, July 2010, pp. 292–295.

[6] Konstantinos Karantzalos, "Recent advances on 2d and 3d change detection in urban environments from remote sensing data," in *Computational Approaches for Urban Environments*, Marco Helbich, Jamal Jokar Arsanjani, and Michael Leitner, Eds., vol. 13 of *Geotechnologies and the Environment*, pp. 237–272. Springer International Publishing, 2015.

[7] S. Lefevre, J. Weber, and D. Sheeren, "Automatic building extraction in vhr images using advanced morphological operators," in *Urban Remote Sensing Joint Event, 2007*, April 2007, pp. 1–5.

[8] Michele Volpi, Devis Tuia, Francesca Bovolo, Mikhail Kanevski, and Lorenzo Bruzzone, "Supervised change detection in {VHR} images using contextual information and support vector machines," *International Journal of Applied Earth Observation and Geoinformation*, vol. 20, no. 0, pp. 77 – 85, 2013, Earth Observation and Geoinformation for Environmental Monitoring.

[9] Mi Wang, Shenggu Yuan, and Jun Pan, "Building detection in high resolution satellite urban image using segmentation, corner detection combined with adaptive windowed hough transform," in *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, July 2013, pp. 508–511.

[10] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 1, pp. 33–50, Jan 2012.

[11] Y. LeCun, Fu Jie Huang, and L. Bottou, "Learning methods for generic object recognition with invariance to pose and lighting," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, June 2004, vol. 2, pp. II–97–104 Vol.2.

[12] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *CoRR*, vol. abs/1312.6229, 2013.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[14] 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, Aug 2007.