

## MODEL-BASED BUILDING DETECTION FROM LOW-COST OPTICAL SENSORS ONBOARD UNMANNED AERIAL VEHICLES

K. Karantzalos<sup>a</sup>\*P. Koutsourakis<sup>a</sup>, I. Kalisperakis<sup>b</sup>, L. Grammatikopoulos<sup>c</sup>

<sup>a</sup> Remote Sensing Lab., National Technical University of Athens,  
[karank@central.ntua.gr](mailto:karak@central.ntua.gr), [panagiotis.koutsourakis@gmail.com](mailto:panagiotis.koutsourakis@gmail.com)

<sup>b</sup> up2metric PC, Athens, Greece,

[ilias@up2metric.com](mailto:ilias@up2metric.com), [www.up2metric.com](http://www.up2metric.com)

<sup>c</sup> Laboratory of Photogrammetry, Technological Educational Institute of Athens, Greece,  
[lazaros@teiath.gr](mailto:lazaros@teiath.gr)

Commission I, WG ICWG I/Vb

**KEY WORDS:** Extraction, reconstruction, unsupervised, segmentation, bundle adjustment, dense matching, fusion, DSM, DTM

### ABSTRACT:

The automated and cost-effective building detection in ultra high spatial resolution is of major importance for various engineering and smart city applications. To this end, in this paper, a model-based building detection technique has been developed able to extract and reconstruct buildings from UAV aerial imagery and low-cost imaging sensors. In particular, the developed approach through advanced structure from motion, bundle adjustment and dense image matching computes a DSM and a true orthomosaic from the numerous GoPro images which are characterised by important geometric distortions and fish-eye effect. An unsupervised multi-region, graph-cut segmentation and a rule-based classification is responsible for delivering the initial multi-class classification map. The DTM is then calculated based on inpainting and mathematical morphology process. A data fusion process between the detected building from the DSM/DTM and the classification map feeds a grammar-based building reconstruction and scene building are extracted and reconstructed. Preliminary experimental results appear quite promising with the quantitative evaluation indicating detection rates at object level of 88% regarding the correctness and above 75% regarding the detection completeness.

### 1. INTRODUCTION

Three dimensional building and landscape models are of great interest for various engineering applications such as urban and rural planning, updating geographic information system, smart cities, 3D visualization, virtual tourism, location based services, navigation, wireless telecommunications, disaster management, noise, heat and exhaust spreading simulations. All these applications are actively discussed in the geography, geoscience and computer vision scientific communities both in academia and industry. Organizations like Google and Microsoft are seeking to include extensively up-to-date 2D and 3D urban models in their products (Microsoft Virtual Earth and Google Earth). Such tools, which have gained important public acceptance, have already deployed 3D building reconstruction techniques as a major visualization component in their process.

However, the prohibitively high cost of generating manually such 2D and 3D dynamic models/maps explains the urgent need towards automatic approaches, especially when one considers modeling and monitoring time-varying events within the complex urban areas at ultra high spatial resolution [Haala and Kada, 2010], [Karantzalos, 2015]. Recent quantitative results from the ISPRS (WGIII/4) benchmark on urban object detection and 3D building reconstruction [Rottensteiner et al., 2014] indicated that 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 [Karantzalos and Paragios,

2009, Karantzalos and Paragios, 2010, Ferro et al., 2010, Jabari et al., 2014], while several building detection approaches from optical data are based on morphological filters [Lefevre et al., 2007], texture [Volpi et al., 2013], point descriptors [Wang et al., 2013], gradient orientation [Benedek et al., 2012], deep learning features [Vakalopoulou et al., 2015], etc.

Moreover, elevation data, like airbone or terrestrial LIDAR, can significantly ameliorate the detection procedure [Sun and Salvaggio, 2013]. However, they can not yet be considered as a cost-effective solution especially for large scale mapping and change detection applications which require frequent data acquisition campaigns [Karantzalos, 2015]. As unmanned aerial vehicles (UAVs) are becoming more and more reliable, with more autonomous capabilities while at the same time their cost is also decreasing, the integration of low-cost UAV and imaging systems is becoming gradually efficient [Kim et al., 2013], [Roca et al., 2013]. Moreover, such autonomous acquisition systems, both fixed-wing or drones, can deliver quickly, huge datasets in ultra-high spatial resolution.

To this end, in this paper we have exploited aerial datasets from low-cost UAV and imaging systems towards building detection in 2D and 3D. In particular, the developed approach through hierarchical structure from motion, bundle adjustment and dense image matching generates a DSM and a true orthomosaic from the numerous GoPro images which are characterised by significant geometric distortions due to the sh-eye effect. An unsupervised multi-region segmentation and a rule-based classification is then separating and labelling multiple object classes in the image domain. Scene's DTM is calculated based on an inpainting and mathematical morphology process. A data fusion process between the detected building from the DSM/DTM and the classification map feeds a model-based process where scene buildings

\*Corresponding author: Dr. Konstantinos Karantzalos, Remote Sensing Laboratory, National Technical University of Athens, email: [karank@central.ntua.gr](mailto:karak@central.ntua.gr), www: <http://users.ntua.gr/karak>



(a) The study area along with the UAV flight path and plan



(b) Different views from the low-cost GoPro camera with important geometric distortions due to the fish-eye effect

Figure 1: The study area covers a  $0.8Km^2$  region in the Eastern Perfecture of Attica, Greece. Two autonomous flights were performed and more than 300 RGB images with important geometric distortions were collected.

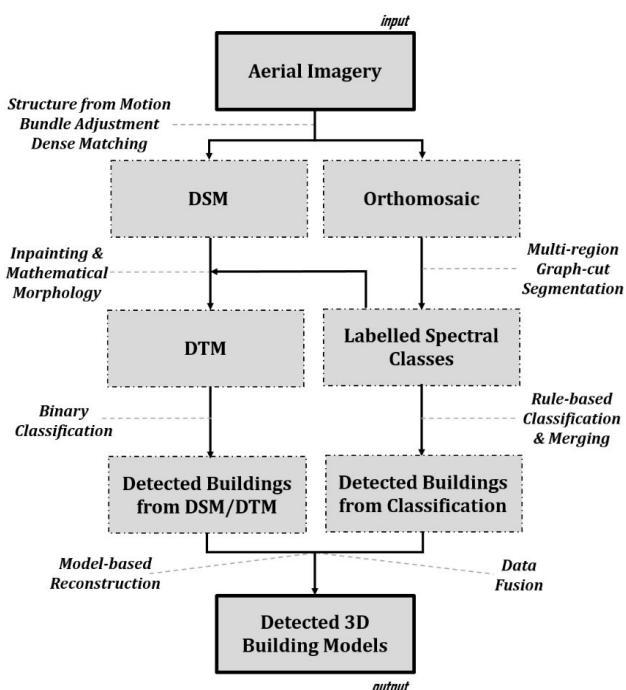


Figure 2: The flowchart of the developed model-based building extraction and reconstruction approach.

are extracted and reconstructed. Preliminary experimental results from an aerial dataset covering a  $0.8Km^2$  region in the Eastern Perfecture of Attica, Greece appear quite promising (Figure 1). The algorithm managed to process more than 300 images and detect with high detection rates more than 215 buildings with a detection completeness above 75% and a detection correctness of 88%.

## 2. METHODOLOGY

The main motivation was to design a building extraction and 3D modelling approach able to exploit aerial datasets from low-cost, lightweight, RGB cameras onboard unmanned aerial vehicles. A

flowchart of the developed approach is shown in Figure 2.

The entire process has been highly automated and once the UAV complete the flight(s), the acquired image data are processed towards the generation of an orthomosaic and a 3D surface model. In particular, all GoPro views are initially rectified from severe fish-eye distortion effect using a rough approximation of their radial distortion model, provided by the manufacturer. All images are then relatively oriented through a hierarchical structure from motion scheme [Pollefeys et al., 2004], [Hartley and Zisserman,



(a) The resulted orthophotomosaic (approximately 13.500x7.000 pixels) with a spatial resolution of 8cm.



(b) The resulted DSM with a spatial resolution of 10cm.

Figure 3: The developed approach through advanced structure from motion, bundle adjustment and dense image matching managed to successfully compute a DSM and a true orthophotomosaic from numerous GoPro images.

2004], [Snavely et al., 2008] at different image scales in order to handle effectively a large number of high resolution images. We start at a low resolution where 2D SIFT [Lowe, 2004] and SURF [Bay et al., 2008] features with descriptors are extracted on all images. Image pairs are identified among the unordered set of images, sparse matching is then performed and outliers are detected and eliminated through RANSAC on the fundamental matrix computation. Combining the inlier matches across different stereopairs, multi-image point correspondences are obtained. A bucketing algorithm is also applied to reduce the number of tie points without affecting their distribution. Then, by means of closed-form algorithms and successive, local bundle adjustment solutions initial image orientations are estimated. These orientations are further refined through a typical self-calibrating bundle adjustment solution.

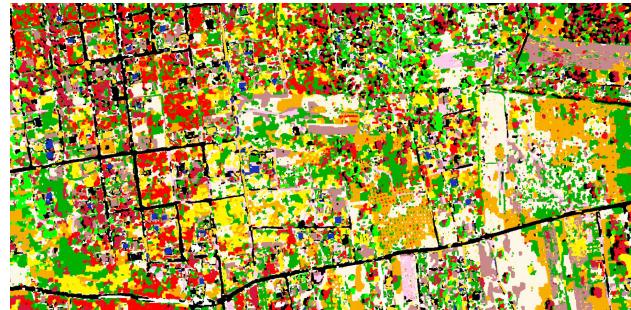
This procedure is repeated at higher image resolutions but this time the matching is restricted by the estimated epipolar geometry and a rough 3D reconstruction from the tie points of the previous resolution. The final interior and exterior orientation of all images is calculated again a bundle adjustment solution. At this point, any remaining lens distortion of the initially rectified GoPro images is correctly modelled and optimally estimated.

Once all GoPro images are oriented a DSM is reconstructed. For this step several state-of-the-art dense stereo and multi-image reconstruction algorithms have been tested [Hirschmüller, 2008], [Furukawa and Ponce, 2010], [Jancosek and Pajdla, 2011], [Stentoumis et al., 2014] which, either combine different disparity maps into a single model, or compute directly an optimal 3D surface from all images. The DSM presented here has been generated using the CMPMVS algorithm9. Using this DSM a true orthomosaic was computed, by employing a multi-image algorithm based on automatic visibility checking and weighted texture blending [Karras et al., 2007].

An unsupervised segmentation approach is taking place dividing the image in approximately 25 spectral classes. We have employed a multiregion graph cut image segmentation in a kernel-induced space [Salah et al., 2011] on the simplified, through anisotropic morphological filtering [Karantzalos et al., 2007], orthomosaic. Based on a supervised rule-based classification image objects are labelled and merged to terrain classes like buildings, roads, vegetation, etc.

Based on the detected terrain classes those belonging to vegetation and buildings are fused with the DSM towards estimating a DTM. In particular, building and vegetation regions were regarded as missing information and an image inpainting technique was responsible to predict -based on the surrounding surfaces- the actual ground height. A morphological reconstruction process further slightly smooth the estimated DTM. The difference between the DSM and the estimated DTM provide a rough but quite accurate approximation of the building heights. This approach has been validated under different terrain, vegetation and building densities.

During the next processing step the detected building footprints from the classification procedure were fused with the detected building footprints from the DSM/DTM. Then a model-based procedure is responsible to fit the appropriate geometry into the detected binary 2D footprints and 3D DSM. The procedure is heavily based on a prior-based approach [Karantzalos and Paragios, 2009, Karantzalos and Paragios, 2010] and integrates a simplified version of a grammar-based modelling [Koutsourakis et al., 2009], [Teboul et al., 2013]. In particular, a vocabulary with a broad set of footprint models can be considered, while the observed 2D building is a projective transformation of the footprint.



(a) The labelled spectral classes after the multi-region, graph-cut segmentation



(b) The detected 2D building footprints after the fusion process



(c) The detected building footprints after the fusion process superimposed to the true orthomosaic

Figure 4: Based on an unsupervised segmentation, rule-based classification and data fusion process the developed algorithm managed to extract building footprints. A model-based approach further recovers their correct 2D and 3D geometry.

Given, the variation of the expressiveness of the grammar, and the degrees of freedom of the transformation, we then focus on the 3D aspect of the model. Similar to [Karantzalos and Paragios, 2010] only building's main height  $h_m$  and building's roof height  $h_r(x, y)$  at every point need to be recovered. However, in this approach, instead of calculating the four angles of the four inclined planes, the gradient along the plane was computed.

### 3. EXPERIMENTAL RESULTS AND EVALUATION

An aerial campaign with an unmanned aerial vehicle was conducted during July 2014 at the Eastern Prefecture of Attica, Greece. In particular, a GoPro HERO3+ Black Edition was mounted and stabilized thought a camera gimbal (AV200, PhotoHigher, New Zealand) onboard a multicopter (OnyxStar BAT-F8, Altigator, Belgium) with electronic controllers and navigation systems (BL-Ctrl V2.0, Navi-Ctrl v2.0, Mikro-kopter, Germnay).

The study area covers a  $0.8 \text{ Km}^2$  region in the Eastern Prefecture of Attica, Greece Figure 1. Two autonomous flights were

performed and more than 300 RGB images with important geometric distortions were collected. The study area along with the UAV flight path and plan are shown in Figure 1a, while different views from the low-cost GoPro camera are shown in Figure 1b with important geometric distortions due to the fish-eye effect.

The developed approach through hierarchical structure from motion, bundle adjustment and dense image matching managed to successfully compute a DSM and a true orthophotomosaic from the numerous GoPro images. The resulted orthophotomosaic (approximately 13.500x7.000 pixels) with a spatial resolution of 8cm and the resulted DSM with a spatial resolution of 10cm are shown in Figure 3.

During the next processing steps, the unsupervised segmentation, the rule-based classification and the data fusion process managed to detect scene buildings and separate them from other terrain objects (like man-made and trees) with similar spectral or height properties. Experimental results with the labelled spectral classes after the multi-region, graph-cut segmentation are shown in Figure 4a. The detected 2D building footprints after the fusion process are shown in Figure 4b and superimposed to the true orthophotomosaic in Figure 4c.

The final step of the developed algorithm managed to recover the geometry of both the 2D footprints and 3D rooftops as shown in Figure 5. In particular, after the application of the model-based procedure the corrected building 2D footprints are shown in Figure 5a in a binary format and superimposed onto the orthophotomosaic in Figure 5b. The recovered geometry of the detected buildings is shown in Figure 5c.

The validation of the developed building detection framework has been performed, also, through the quantitative comparison of the detection results with the reference (ground truth) data. The standard measures of Completeness and Correctness 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.

The performed quantitative evaluation indicated high detection correctness rates above 88%. The detection completeness was relative lower but above 75%. The developed algorithm managed to detect and reconstruct the geometry of more than 215 buildings. The vast majority of the false negatives belonged to buildings with a relative small size and this particular fact was in accordance with the literature and similar efforts [Rottensteiner et al., 2014]. Moreover, it should be noted that the evaluation of the rooftop reconstruction component was performed with a qualitatively manner since 3D reference data at ultra high resolution were not available.

#### 4. CONCLUSIONS AND FUTURE PERSPECTIVES

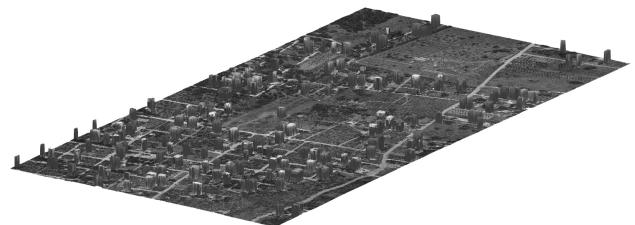
In this paper, building detection in 2D and 3D was addressed through the exploitation of aerial datasets from low-cost UAV and imaging systems. In particular, the developed algorithm integrates advanced structure from motion, bundle adjustment and dense image matching procedures towards the computations of a DSM and a true orthophotomosaic from numerous RGB images which important geometric distortions. Terrain classes are detected through an unsupervised multi-region segmentation and rule-based classification process. Scene's DTM is calculated based



(a) The corrected building 2D footprints after the model-based procedure



(b) The corrected buildings superimposed onto the orthophotomosaic



(c) A 3D view of the detected buildings after the application of the developed approach

Figure 5: The detected buildings in 2D and 3D after the application of the developed approach.

on an inpainting and mathematical morphology procedure. A data fusion process between the detected building from the DSM/DTM and the classification map feeds a model-based process where scene buildings are extracted and reconstructed. Preliminary experimental results from an aerial dataset covering a 0.8 Km<sup>2</sup> region in the Eastern Prefecture of Attica, Greece appear quite promising. The algorithm managed to process more than 300 images and detect with high detection rates more than 215 buildings with a detection completeness above 75% and a detection correctness of 88%. The reliable estimation of building 3D geometry suggested that the proposed method constitutes a quite promising tool for various object extraction and reconstruction application with data from low-cost imaging systems onboard UAVs. A sensitivity analysis and code optimisation is currently under development.

#### 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.

## REFERENCES

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., 2008. Speeded-up robust features (surf). *Comput. Vis. Image Underst.* 110(3), pp. 346–359.

Benedek, C., Descombes, X. and Zerubia, J., 2012. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(1), pp. 33–50.

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

Furukawa, Y. and Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(8), pp. 1362–1376.

Haala, N. and Kada, M., 2010. An update on automatic 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(6), pp. 570–580. *ISPRS Centenary Celebration Issue*.

Hartley, R. I. and Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Second edn, Cambridge University Press, ISBN: 0521540518.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30(2), pp. 328–341.

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

Jancosek, M. and Pajdla, T., 2011. Multi-view reconstruction preserving weakly-supported surfaces. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 3121–3128.

Karantzalos, K., 2015. Recent Advances on 2D and 3D Change Detection in Urban Environments from Remote Sensing Data. In: M. Helbich, J. Jokar Arsanjani and M. Leitner (eds), *Computational Approaches for Urban Environments, Geotechnologies and the Environment*, Vol. 13, Springer International Publishing, pp. 237–272.

Karantzalos, K. and Paragios, N., 2009. Recognition-Driven Two-Dimensional Competing Priors Toward Automatic and Accurate Building Detection. *Geoscience and Remote Sensing, IEEE Transactions on* 47(1), pp. 133–144.

Karantzalos, K. and Paragios, N., 2010. Large-Scale Building Reconstruction Through Information Fusion and 3-D Priors. *Geoscience and Remote Sensing, IEEE Transactions on* 48(5), pp. 2283–2296.

Karantzalos, K., Argialas, D. and Paragios, N., 2007. Comparing morphological leveling constrained by different markers. In: ISMM, G.Banon, et al. (eds), *Mathematical Morphology and its Applications to Signal and Image Processing*, pp. 113–124.

Karras, G. E., Grammatikopoulos, L., Kalisperakis, I. and Petsa, E., 2007. Generation of orthoimages and perspective views with automatic visibility checking and texture blending. *Photogrammetric Engineering and Remote Sensing* 73, pp. 403–411.

Kim, J., Lee, S., Ahn, H., Seo, D., Park, S. and Choi, C., 2013. Feasibility of employing a smartphone as the payload in a photogrammetric UAV system. *ISPRS Journal of Photogrammetry and Remote Sensing* 79(0), pp. 1–18.

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

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

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 20(2), pp. 91–110.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. and Koch, R., 2004. Generation of orthoimages and perspective views with automatic visibility checking and texture blending. *International Journal of Computer Vision* 3(3), pp. 207–232.

Roca, D., Lagela, S., Daz-Vilario, L., Armesto, J. and Arias, P., 2013. Low-cost aerial unit for outdoor inspection of building faades. *Automation in Construction* 36(0), pp. 128–135.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U. and Jung, J., 2014. Results of the ISPRS benchmark on urban object detection and 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 93(0), pp. 256–271.

Salah, M., Mitiche, A. and Ayed, I., 2011. Multiregion image segmentation by parametric kernel graph cuts. *Image Processing, IEEE Transactions on* 20(2), pp. 545–557.

Snavely, N., Seitz, S. and Szeliski, R., 2008. Modeling the world from internet photo collections. *International Journal of Computer Vision* 80(2), pp. 189–210.

Stentoumis, C., Grammatikopoulos, L., Kalisperakis, I. and Karras, G., 2014. On accurate dense stereo-matching using a local adaptive multi-cost approach. *International Journal of Photogrammetry and Remote Sensing* 91, pp. 29–49.

Sun, S. and Salvaggio, C., 2013. Aerial 3D Building Detection and Modeling From Airborne LiDAR Point Clouds. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 6(3), pp. 1440–1449.

Teboul, O., Kokkinos, I., Simon, L., Koutsourakis, P. and Paragios, N., 2013. Parsing facades with shape grammars and reinforcement learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(7), pp. 1744–1756.

Vakalopoulou, M., Karantzalos, K., Komodakis, N. and Paragios, N., 2015. Building Detection in Very high Resolution Multispectral Data with Deep Learning Features. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*.

Volpi, M., Tuia, D., Bovolo, F., Kanevski, M. and Bruzzone, L., 2013. Supervised change detection in {VHR} images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation* 20(0), pp. 77–85. *Earth Observation and Geoinformation for Environmental Monitoring*.

Wang, M., Yuan, S. and Pan, J., 2013. 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*, pp. 508–511.