Extracting 3D Polyhedral Building Models from Aerial Images using a Featureless and Direct Approach

F. Dornaika [1]
[1] French Geographical Institute (IGN)
94165 Saint-Mandé, Cedex, France
fadi.dornaika@ign.fr

K. Hammoudi [1,2]
[2] Université Paris-Est
77454 Marne-la-Vallée, Cedex 2, France
karim.hammoudi@ign.fr

Abstract

We describe a model driven approach for extracting simple 3D polyhedral building models from aerial images. The novelty of the approach lies in the use of featureless and direct optimization based on image rawbrightness. The 3D polyhedral model is estimated using a stochastic and genetic optimizer that minimizes a global dissimilarity measure. The proposed approach gives more accurate 3D reconstruction than feature-based approaches since it does not involve intermediate noisy data (e.g., the 3D points of a noisy Digital Elevation Maps). We provide experiments and evaluations of performance. Experimental results show the feasibility and robustness of the proposed approach.

Keywords: Image based 3D building modeling, featureless approach, Differential Evolution algorithm

1 Introduction

The extraction of 3D models of buildings is currently a very active research area since it is a key issue in urban planning, virtual reality, and updating databases for geo-information systems, to name a few. For roof building reconstruction the main source of data is aerial images [9]. The proposed methods for building reconstruction differ by the assumption made as well as by the type of input data. However, one can easily classify these approaches into two main categories: bottom-up and top-down approaches. In theory, bottom-up approaches can handle the case where there is no prior knowledge about the sought building model. On the other hand, top-down approaches rely on some prior knowledge (e.g., using parametric models).Both categories have been used with features that are extracted and matched in at least two images. For roofs, the most used image features are 2D segments and junctions lines that are converted into 3D features. The final polyhedral model is then estimated from these 3D features. Model-based reconstruction techniques were first applied in digital photogrammetry for the (semi-)automatic reconstruction of buildings in aerial images with the help of generic building models [1, 3, 6]. Several researchers have tried to fully automate the process by using automatically detected lines. A typical feature-based approach is described in [4].

In this paper, we propose a featureless approach that extracts simple polyhedral building models from the rawbrightness of calibrated aerial images where the footprint of the building in one image is obtained either manually or automatically [5]. We were inspired by the featureless image registration techniques where the goal is to compute the global motion of the brightness pattern between them (e.g., affine or homography transforms) without using matched features [7]. Unlike existing approaches, our approach derives the polyhedral building model by minimizing a global dissimilarity measure based on the image rawbrightness. It is carried out using a stochastic and genetic optimizer. To the best of our knowledge the use of featureless and direct approaches has not been used for extracting polyhedral models of buildings. In any feature-based approach, the inaccuracies associated with the extracted features, in either 2D or 3D, will inevitably affect the accuracy of the final 3D model.

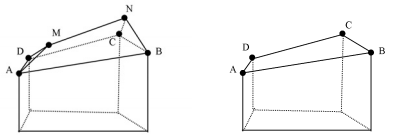
Recently, many researchers proposed methods for extracting polyhedral models from Digital Elevation Maps (DEMs) (e.g., [3]). Compared to these approaches, our method has the obvious advantage that the coplanarity constraints are implicitly enforced in the model parametrization. On the other hand, the approaches based on DEMs impose the coplanarity constraint on the 3D points of the obtained surface in the process of plane fitting. DEMs are usually computed using local correlation scores together with a smoothing term that penalizes large local height variation. Thus, correlation-based DEMs can be noisy. Moreover, height discontinuities may not be located accurately. In brief, our proposed approach can give more accurate 3D reconstruction than feature-based approaches since the process is more direct and does not involve intermediate noisy data (e.g., the 3D points of a noisy DEM). The remainder of the paper is organized as follows. Section 2 states the problem we are focusing on and describes the parametrization of the adopted polyhedral model. Section 3 presents the proposed approach. Section 4 gives some experimental results.

2 Problem statement and model parametrization

Since aerial images are used only roof models can be estimated. In this work, we restrict our study to simple polyhedral models that are illustrated in Figure 1. The model illustrated in Figure 1.**(a)** can describe a building roof having two, three, or four facets by varying the 3D location of the inner vertices.

These models can describe all typical situations: non symmetric shape, sloping ground or roofs (i.e., every vertex can have a different height). Because a complex building can be described as an aggregation of simple building models, our approach can also deal with complex buildings once a partitioning of the building into

simple building-parts is done.



(a) **(b)**

Figure 1: The adopted simple polyhedral models. **(a)** The multi-facet model. **(b)** The one facet model.

The problem we are focusing on can be stated as follows. Given the footprint of a building in one aerial image we try to find the polyhedral model (an instance of the models depicted in Figure 1) using the raw-brightness of the aerial images that views this building. Without loss of generality, we restrict our study to the use of two images. Extending the approach to the multi-view case is straightforward.

The flowchart of the proposed approach is depicted in Figure 2. Since the images are calibrated and since the outer vertices are known in one image, our polyhedral model can be parameterized by eight parameters: four parameters for the image location of the inner vertices M and N and four parameters for the height of the vertices A, M, N, and C. The remaining vertices are determined by intersecting the corresponding line of sight with the estimated support planes.

The eight parameters are encapsulated into one single vector $\mathbf{w}$:

$$\mathbf{w} = (U_M, V_M, U_N, V_N, Z_A, Z_M, Z_N, Z_C)^T \qquad (1)$$

where (U_M, V_M) and (U_N, V_N) are the image coordinates of the vertices $M$and N, respectively. Recall that the 3D coordinates are expressed in a local coordinate system whose Z axis coincides with the ground normal (the aerial images are geo-referenced). In practice, although the location of inner vertices is not known, the 2D line (the projection of a ridge segment) going through them can be easily extracted from the image by using a simple edge detector followed by a Hough transform. Once the equation of this line is known, the parametrization of the building model can be simplified to:

$$\mathbf{w} = (\lambda_M, \lambda_N, Z_A, Z_M, Z_N, Z_C)^T \qquad (2)$$

where λ_M and λ_N parameterize the location of the inner vertices along the 2D segment obtained by intersecting the 2D line with the building footprint. Thus, finding the model boils down to finding the vector $\mathbf{w}$.

3 Proposed approach

The goal is to compute the parameters of the polyhedral model given two images one of which contains the external boundary of the building. This boundary is provided either manually or automatically. The basic idea relies on the following fact: if the shape and the geometric parameters of the building (encoded by the vector $\mathbf{w}$) correspond to the real building shape and geometry, then the pixel-to-pixel mapping between the image I_1 and the image I_2 will be correct for the entire building footprint. Recall that $\mathbf{w}$ is defining all

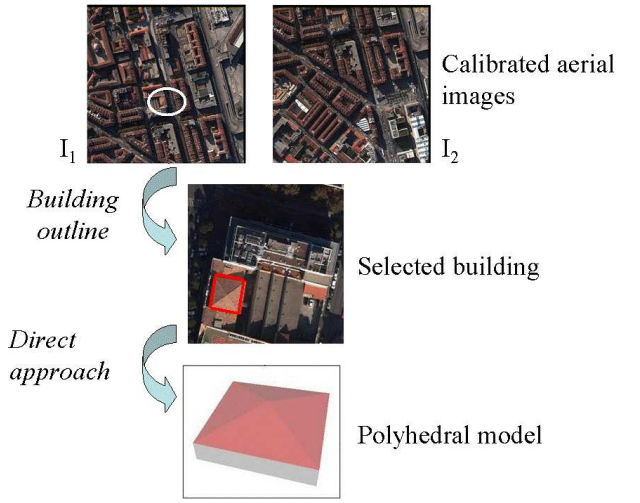


Figure 2: The flowchart of the proposed approach for extracting 3D polyhedral model from image rawbrightness.

support planes of all the building's facets and thus the corresponding pixel $\mathbf{p}'$ of any pixel $\mathbf{p}$ is estimated by a simple image transfer through homographies (3×3 matrices) based on these planes. Therefore, the associated global similarity over these pixels reaches a maximum. In other words, the global dissimilarity measure reaches a minimum. The global dissimilarity is given by the following score:

$$e = \sum_{\mathbf{p} \in S} \rho(|I_1(\mathbf{p}) - I_2(\mathbf{p}')|) \qquad (3)$$

where S is the footprint of the building in the first image I_1, $\rho(x)$ is a robust error function, and $\mathbf{p}'$ is the pixel in the second image I_2 that corresponds to the pixel $\mathbf{p}$. The choice of the error function $\rho(x)$ will determine the nature of the global error (3) which can be the Sum of Squared Differences (SSD) ($\rho(x) = \frac{1}{2}\, x^2$), the Sum of Absolute Differences (SAD) ($\rho(x) = x$), or the saturated Sum of Absolute Differences. In general, the function $\rho(x)$ could be any M-estimator [2]. In our experiments, we used the SAD score since it is somewhat robust and its computation is fast.

We seek the polyhedral model $\mathbf{w}^\star = (\lambda_M^\star, \lambda_N^\star, Z_A^\star, Z_M^\star, Z_N^\star, Z_C^\star)^T$ that minimizes the above dissimilarity measure over the building footprint. This is given by:

$$\mathbf{w}^\star = arg\min_{\mathbf{w}} e \qquad (4)$$

$$= arg\min_{\mathbf{w}} \sum_{\mathbf{p} \in S} \rho(|I_1(\mathbf{p}) - I_2(\mathbf{p}')|) \qquad (5)$$

Recall that during the whole process there is no feature extraction nor matching with the second image. In order to minimize (5) over $\mathbf{w}$, we use the Differential Evolution algorithm [8]. This is carried out using generations of solutions—populations. The population of the first generation is randomly chosen around a rough solution. The rough solution is simply given by a zero-order approximation model (flat roof model) which is also obtained by minimizing the dissimilarity score over one unknown (the average height of the roof). We use

the Differential Evolution optimizer since it has two interesting properties: (i) it does not need an accurate initialization, and (ii) it does not need the computation of partial derivatives of the cost function.

In brief, the proposed approach proceeds as follows. First, the algorithm decides if the building contains one or more facets, that is, it selects either the model of Figure 1.**(a)** or the model of Figure 1.**(b)**. This decision is carried out by analyzing the 3D normals associated with four virtual triangles forming a partition of the whole building footprint. Second, once the model is selected, its parameters are then estimated by minimizing the corresponding dissimilarity score.

4 Experimental results

The proposed approach has been tested with a set of calibrated aerial images depicting a part of the city of Marseille. The resolution of these aerial images is 4158×4160 pixels. The ratio between the baseline to the camera height is about 0.18. One pixel corresponds to a 10 cm square at ground level.

Figure 3 shows the results obtained with two images depicting a four facet building. **(a)** shows the footprint of the building (manually selected) in the first image. **(b)** shows the estimated 3D polyhedral model. **(c)** shows the projection of the this polyhedral model onto the same image. **(d)** displays the variation of the best SAD as a function of the iteration number. One can see easily that the Differential Evolution algorithm succeeded to estimate the correct 3D polyhedral model, i.e., the six unknown parameters (see the estimated location of the two inner vertices). Moreover, one can notice that the convergence was obtained in about five iterations/generations. The population size was 40.

Figure 4 illustrates the best model obtained at different iterations of the Differential Evolution algorithm. The projection of the model onto the first and second images is shown in the first and second columns, respectively. The third column illustrates the associated 3D model.

Figure 5 shows the results obtained with a three facet building. Figure 6 illustrates the estimated model in cases where buildings are affected by shadows. Despite the presence of significant shadows the estimated polyhedral models are correct.

Method comparison. So far the performance evaluation was qualitative, i.e., the goodness of the model was assessed by the projection of the estimated model segments and vertices onto the two images. In order to get quantitative evaluation we compared our method with a 3D reconstruction obtained from DEMs.

Table 1 depicts the 3D reconstruction results associated with one facet having three vertices (only the heights are shown). The first column corresponds to the reconstruction obtained with a DEM (robust plane fitting), the second column to our approach adopting the SSD function, and the third column to our approach adopting the SAD function. The last row shows the average deviation between the estimated model and the model obtained with the DEM. We point out that the DEM based model is not providing the ground-truth data.

	DEM	SSD	SAD
Vertex1 height	41.96m	42.75m	42.22m
Vertex2 height	41.36m	40.87m	40.98m
Vertex3 height	39.78m	40.10m	40.22m
Average deviation	0.0m	0.53m	0.36m

Table 1: Method comparison associated with one facet having three vertices. The first column depicts the estimated height of the model vertices obtained with a DEM. The second (third) column displays the estimated heights using our approach with SSD function (SAD function).

Model extraction in the presence of superstructures. In some cases, roofs include superstructures (detailed volumes present on roofs such as chimneys, dormer windows...). Since the image regions of these superstructures do not obey the image transfer law associated with their support roof plane, they should be excluded in the global cost function (5) in order not to bias the reconstruction results.

To this end, two stages are invoked. In the first stage, the Differential Evolution optimizer is run in a straightforward manner using the total footprint of the building. Using this solution, the individual residuals (the absolute difference for each pixel) are computed. Outlier pixels are then identified by using a threshold depending on the median value of all residuals. Once the outlier pixels are detected, the Differential Evolution optimizer is used with all inlier pixels.

Figure 7 shows the application of this strategy on a selected facet. The upper part of this figure shows the selected facet. The middle part shows the detected outlier pixels. One can notice that detected outliers have a significant overlap with the superstructures. The lower part shows the average deviation (in Z) of the estimated facet model and the model obtained with the DEM with and without outlier filtering.

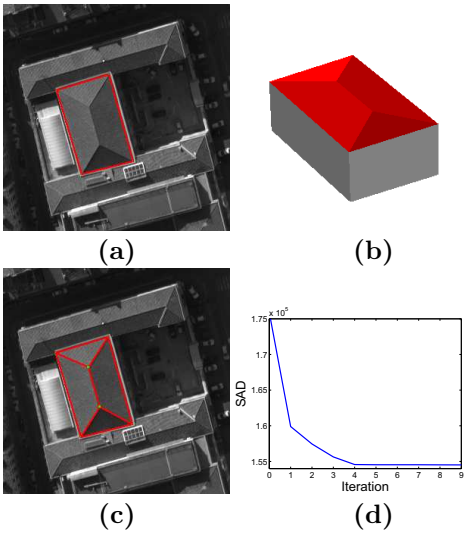


Figure 3: **(a)** the selected building in the first image. **(b)** the estimated 3D polyhedral model. **(c)** the projection of the model onto the image. **(d)** the evolution of the best SAD as a function of the iteration number.

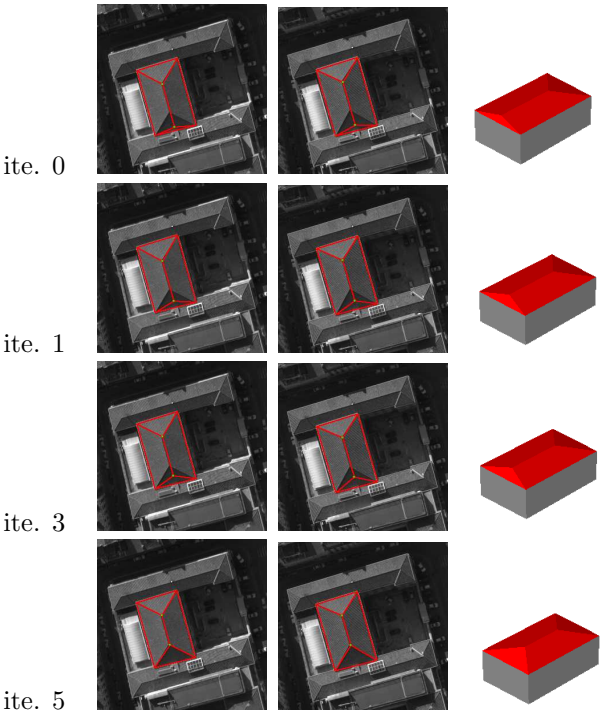


Figure 4: The best solution at several iterations of the Differential Evolution algorithm.

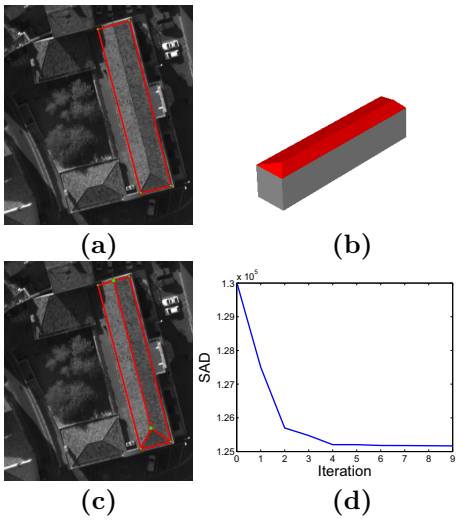


Figure 5: **(a)** the selected building in the first image. **(b)** the estimated 3D polyhedral model. **(c)** the projected model onto the image. **(d)** the evolution of the best SAD as a function of the iteration number.

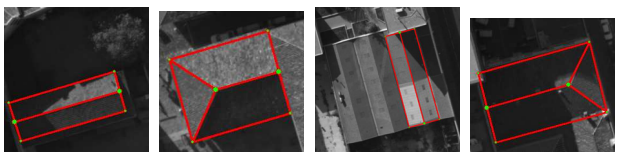
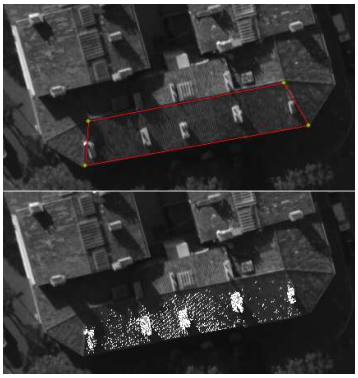


Figure 6: Estimated 3D polyhedral models from aerial images. One can notice that despite the presence of significant shadows the approach has provided the correct models.



	DEM	without filtering	with filtering
Ave. deviation	0.0m	0.46m	0.36m

Figure 7: **Top:** the selected facet. **Middle:** the detected outlier pixels. **Bottom:** the deviation of the 3D facet without and with outliers filtering. The outliers have a significant overlap with the superstructures.

5 Conclusion

We presented a direct model driven approach for extracting 3D polyhedral building models from calibrated aerial images. The approach does not require feature extraction and matching in the images. Moreover, it does not rely on Digital Elevation Maps. Experimental results show the feasibility and robustness of the proposed approach. Future work may investigate the extension of the approach to generic building models.

Acknowledgment The authors would like to thank Mathieu Brédif from IGN for his help with issues related to the data used in this work.

References

[1] A. Brunn, E. Gulch, F. Lang, and W. Forstner. A multi layer strategy for 3D building acquisition. In *IAPR TC-7 Workshop: Mapping Buildings, Roads and other Man-Made Structures from Images*, 1996.

[2] J. Chen, C. Chen, and Y. Chen. Fast algorithm for robust template matching with M-estimators. *IEEE Trans. on Signal Processing*, 51(1):230–243, 2003.

[3] H. Jibrini, N. Paparoditis, M. Pierrot-Deseilligny, and H. Maitre. Automatic building reconstruction from very high resolution aerial stereopairs using cadastral ground plans. In *XIXth ISPRS Congress*, 2000.

[4] Z. Kim and R. Nevatia. Automatic description of complex buildings from multiple images. *Computer Vision and Image Understanding*, 94(1):60–95, 2004.

[5] S. Krishnamachari and R. Chellappa. Delineating buildings by grouping lines with MRFs. *IEEE Trans. on Image Processing*, 5(1):164–168, 1996.

[6] C. Lin and R. Nevatia. Building detection and description from a single intensity image. *Computer Vision and Image Understanding*, 72(2):101–121, 1998.

[7] T. Romero and F. Calderón. *Scene Reconstruction, Pose Estimation and Tracking*, chapter A Tutorial on Parametric Image Registration. I-Tech, 2007.

[8] R. Storn and K. Price. Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[9] L. Zebedin, J. Bauer, K. Karner, and H. Bischof. Fusion of feature- and area-based information for urban buildings modeling from aerial imagery. In *European Conference on Computer Vision*, 2008.