We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Data Fusion in a Hierarchical Segmentation Context: The Case of Building Roof Description

Frédéric Bretar

Institut Géographique National – MATIS laboratory
France

1. Introduction

Automatic mapping of urban areas from aerial images is a challenging task for scientists and surveyors because of the complexity of urban scenes. The 2D image information can be converted into 3D points provided that aerial images have been acquired in a (multi-) stereoscopic context (Kasser & Egels, 2002). Altitudes are then processed using automatic correlation algorithms to generate Digital Surface Models (DSM) (Pierrot-Deseilligny & Paparoditis, 2005), (Baillard & Dissard, 2000). The DSM helps in the understanding of an urban scene, especially for the 3D building reconstruction problem. There are two main approaches to take to this problem:

1. detecting 3D primitives (segments or planes) before making polyhedric building models (Jibrini & al, 2000),
2. using a parametric model-based approach (Lafarge et al, 2006).

This study aims to present a methodology for detecting building roof facets. These facets are meant to be integrated into an algorithm for building reconstruction. Many researches have been performed on this topic using DSMs as altimetric data. Nevertheless, in the last past years, airborne lidar systems (ALS) have become an alternative source for acquiring altimetric data (Baltsavias, 1999). These systems are based on the recording of the time-of-flight distance between an emitted laser pulse and its response after a reflection on the ground. They provide sets of 3D irregularly distributed points, georeferenced with an integrated GPS/INS device. The accuracy (< 0,15 m in altimetry) and the robustness of such systems are better than photogrammetric derived DSMs. However, ALS do not provide textural information that can be exploited, as they are with optical aerial images.

We therefore propose in this paper to use jointly calibrated aerial images and 3D lidar data to extract 3D roof facets. We built a joint image segmentation paradigm that includes radiometric, geometric and semantic properties of each data set. Very few researches have been performed on the fusion of lidar and aerial images and are mainly focused on image classification (Rottensteiner et al, 2004), (Haala & Walter, 1999).

We will present in the first part the theoretical background of our methodology, especially the hierarchical segmentation framework. We will then show some results of 3D roof facet extraction.

2. Methodology

2.1 Background

A region is defined as a set of pixels sharing the same properties. Segmenting an image I in n regions consists in determining a partition $\Delta_n I$ of I satisfying:

$$\Delta_n I = \bigcup_{i \in [1,n]} R_i, \quad R_i \bigcap_{i \neq j} R_j = \phi, \forall i, R_i \text{ is connected}$$

The segmentation problem may be considered under various points of views seeing that a unique and reliable partition does not exist. Beyond classical region growing algorithms, approaches based on a hierarchical representation of the scene retained our attention. These methodologies open the field of multi-scale descriptions of images (Guigues et al, 2003). Here, we are interested in obtaining an image partition whereon roof facets are clearly delineated and understandable as unique entities.

A hierarchy is defined as a tree structure. It is a graph where nodes are related to image regions and edges (father-child relationships) to region subset inclusions. The root of the tree corresponds to the whole image and the leaves to the initial partition (over-segmentation) of the image. An **eligible partition** onto a hierarchy (or a *cut*) is therefore a set of nodes which related leaf region sets are disjoint. Figure 1 sketches a *cut* in a hierarchy represented as a dendrogram as well as the corresponding partition.
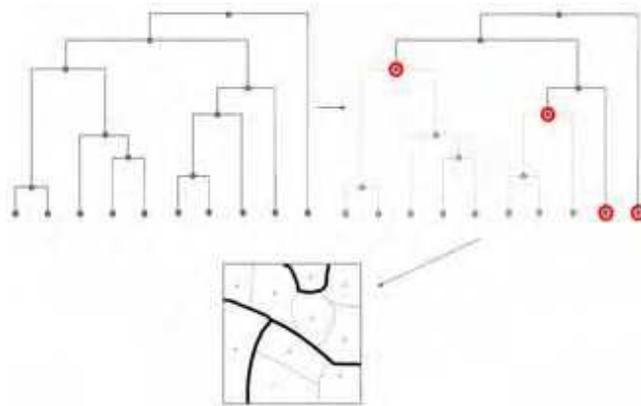

Fig. 1. Sketch of a *cut in* a hierarchy (dendrogram). Red circles are the selected cut nodes and correspond to the presented image partition.

A data structure for representing an image partition is the Region Adjacency Graph (RAG). The RAG is defined as an undirected graph $G(E,V)$ where V is the set of nodes related to an image region and E the set of edges related to adjacency relationships between two neighbouring regions. Each edge E is weighted by a **cost function** (or energy) that scores the dissimilarity between two adjacent regions. The general idea of a hierarchical ascendant segmentation is to merge sequentially the most "similar" pair of regions (or the one that

minimises the cost function) until a single region remains. The fusion of these two regions (or the contraction of the RAG minimal edge) creates a node in the hierarchy and two father-child relationships in case of a binary tree. Figure 2 sketches the generation process of the hierarchy from an initial partition of the image.
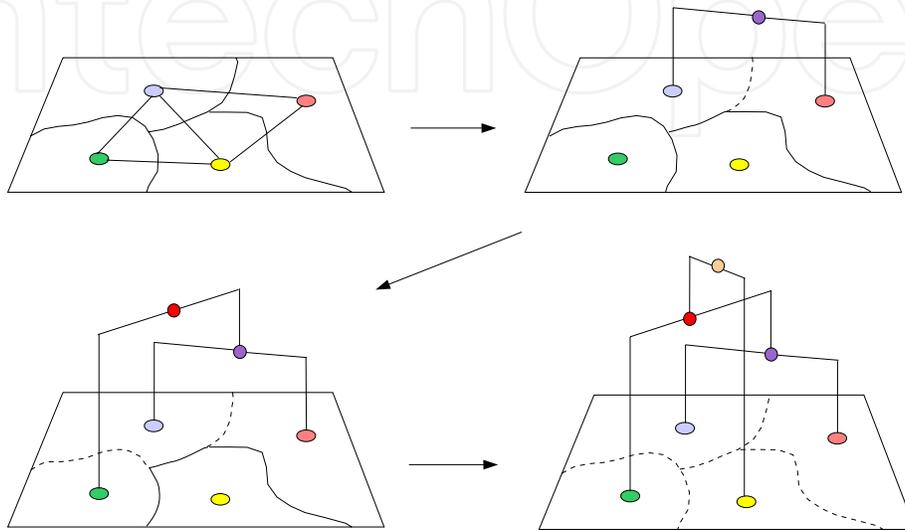

Fig. 2. Construction of a hierarchy based on a RAG (left).

2.2 Theory
The shape of the hierarchy (therefore the region merging order) constraints the existence of an eligible partition. In other words, initial regions that theoretically belong to a roof facet must be mutually merged until a node in the hierarchy appears as a roof facet entity. If it appears that sub-regions of a facet merge with adjacent regions that do no belong to their supporting facet, the embedded geometry is broken.

2.2.1 The cost function
The region merging order depends on the definition of the energy E associated to each edge of the RAG. We can write E as a sum of three terms E_r, E_l and E_s respectively related to the image radiometry, to the lidar geometry and to the semantic extracted from lidar data.

E_r is defined to minimise the loss of information when describing the image from n to $n-1$ regions. We retained the cost function given by *Haris* (Haris et al, 1998) for merging two neighbouring regions R_i and R_j :

$$\mathrm{E}_r(R_i, R_j) = \frac{\|R_i\|_r \|R_j\|_r}{\|R_i\|_r + \|R_j\|_r} (\mu_i - \mu_j)^2$$

Where $\left\| . \right\|_r$ is the number of pixels in each region and $\mu = \dfrac{1}{\left\| R \right\|_r} \sum_k I(k)$ the average value

of the radiometries at image sites k of the region.

In our context, E_l is defined to take advantage of both the **accuracy** and the **regularity** of lidar measurements onto roof surfaces to make appear in the hierarchy nodes corresponding to roof facet entities. It is therefore expected that image regions merge independently over each roof facet of the focused building. Higher levels of the hierarchy are not of interest in this study. The adequation of lidar points to lie on a roof facet is measured by estimating a plane onto those included in $R_i \bigcup R_j$. A non robust least square estimator is applied specifically for neighbouring regions not to merge when the estimated plane is corrupted by non coplanar points. Such is the case when attempting to merge two regions apart from the roof top before other couples of regions belonging to the same roof facet with possible significant radiometric dissimilarities. If $\left\| N_i \right\|_l$ (resp. $\left\| N_j \right\|_l$) is the number of lidar points in region R_i (resp. R_j) and r_p the residuals of a laser point to the estimated plane, ρ_l^2 is the average square distance of laser points to the estimated plane with

$$\rho_l^2 = \frac{1}{\left\| N_i \right\|_l + \left\| N_j \right\|_l} \sum_{p=1}^{\left\| N_i \right\|_l + \left\| N_j \right\|_l} r_p^2$$

If we consider a similar weighting factor as for E_r depending on the number of lidar points $\left\| . \right\|_l$ in image regions, E_l is expressed as:

$$\mathrm{E}_l(R_i, R_j) = \frac{\left\| N_i \right\|_l \left\| N_j \right\|_l}{\left\| N_i \right\|_l + \left\| N_j \right\|_l} \rho_l^2$$

3D lidar points have been previously processed to extract a binary semantics: **ground** and **off-ground** points. Theoretically, the off-ground class includes building and vegetation. However, we will only consider the segmentation algorithm to be focused on buildings. The process is performed with a high level of relevancy over urban areas owing to the sharp slope breaking onto building edges (Bretar et al, 2004). An image region will be classified as **ground** if it contains at least one projected lidar **ground** point. Otherwise, the region is considered to be a built up area. This binary semantics provides a reliable ground mask that can be integrated into the initial segmentation. Two regions of different classes are kept disjoint until the highest levels of the hierarchy. Finally we can write

$$\mathrm{E}_s(R_i, R_j) = \begin{cases} \infty \; \mathit{if}\; R_i \; \mathit{or}\; R_j \; \mathit{is\; a\; ground\; region} \\ \qquad 0\; \mathit{if\; not} \end{cases}$$

2.2.2 The optimal eligible partition

A roof facet is defined as a 3D planar polygon which average square distance to lidar support points (ρ_l^2) is less than a threshold s. The final partition is obtained by recursively exploring the binary tree structure from its root comparing ρ_l^2 of each node to s.

3. Results

The test area is part of the inner city of Amiens, France. Aerial images (resolution 0,2m) are firstly re-projected into ortho-rectified geometry in order to avoid the segmentation of building facades. Using the original geometry of a set of calibrated aerial images is thought as future work. In order to enforce the region borders to lie on real discontinuities, we applied a contour detection algorithm (hysteresis thresholding) on the gradient image. The gradient was computed with a Canny-Deriche operator ($\alpha = 1$) (Deriche, 1987). The watershed algorithm is finally applied on a combination of both images (maximum of gradient and contour images). Figure 3 sketches the flowchart of the entire methodology.
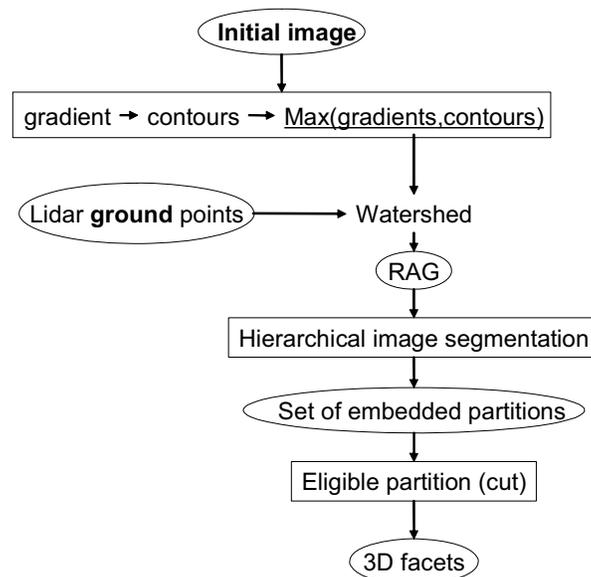
Fig. 3. Flow chart of the algorithm

We present in table 1 a set of embedded image partitions. Region contours are back-projected onto the ortho-rectified image. Parameter s describes a partition set. Following s, one can notice that structures progressively appear as unique entities until adjacent facets merge together. At the time of the study, s is tuned after a visual evaluation of each partition. Indeed, this threshold is highly related to the roof shape and is therefore different from one building to the other. We clearly see on these examples the delineation of the buildings with regard to ground regions as well as to courtyards. Isolated elementary regions remain within the large ground region due to the lack of lidar points inside them.
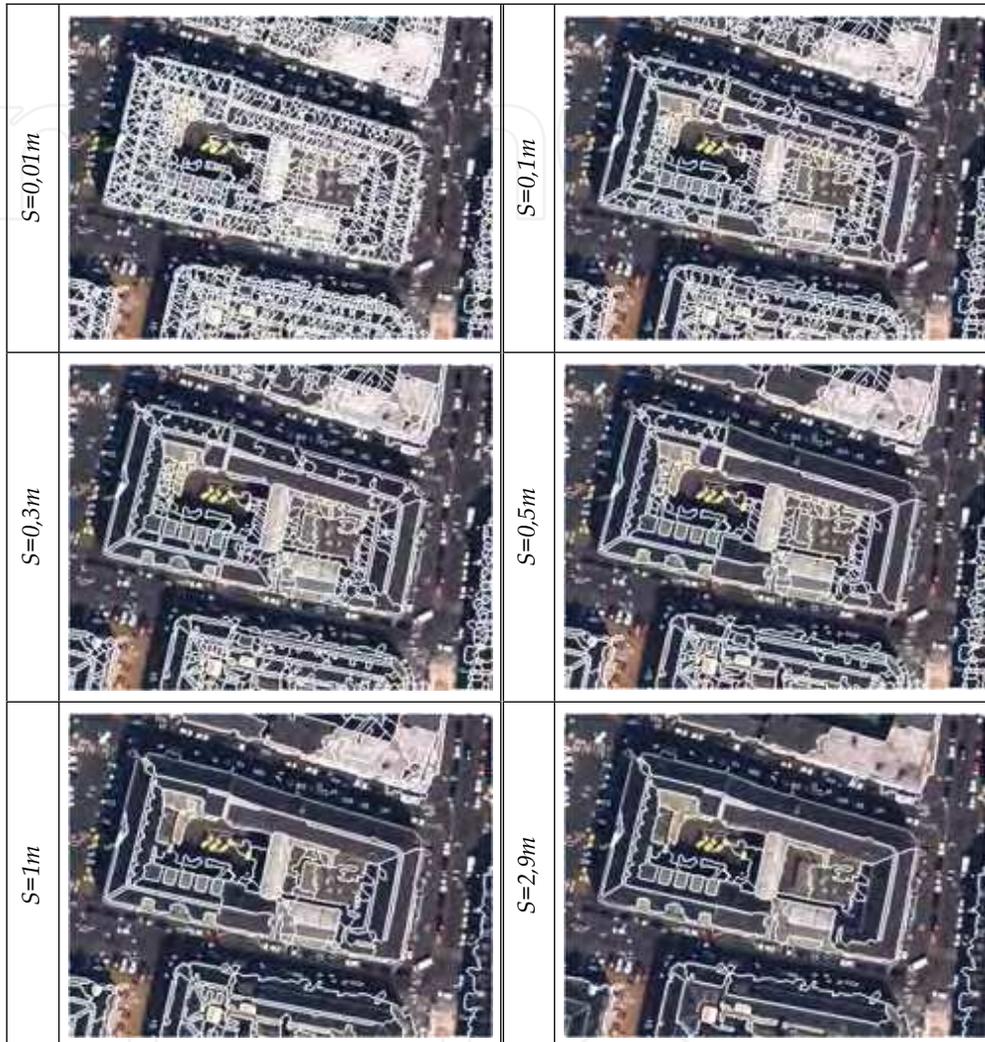
Table 1. Examples of partitions at different scales .White segments are the region borders.

As for the building presented in table 1, we consider that a final eligible partitions is achieved for *s=0,5m*. Figure 4 shows the reconstructed 3D facets of this building. This reconstruction considers lidar points belonging to an image region larger than 30 pixels and which orientation is greater than 30° from vertical. The presented 3D scenes give a realistic representation of the buildings whereon hyper-structures such as dormer windows are particularly visible. There delineation could not have been obtained considering only lidar data due to their low spatial density. The high radiometric contrast of the aerial image over some of these structures is then real complementary information. The accuracy of lidar points gives also the opportunity to detect two neighbouring regions with a low orientation difference as two different facets.
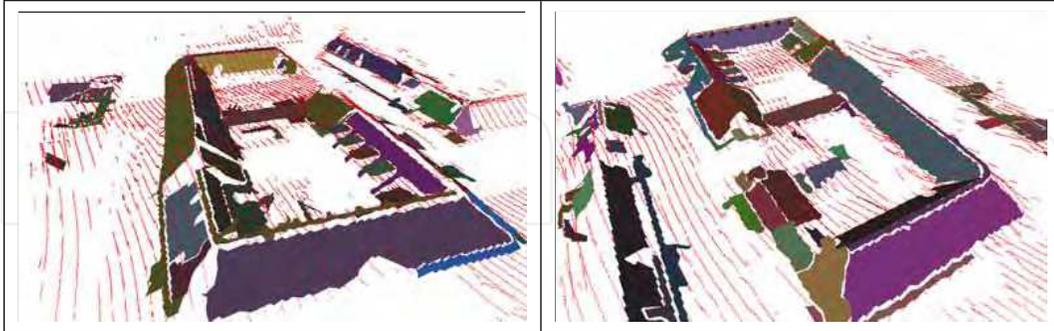
Fig. 4. 3D views of facets estimated from lidar (red) points (same building as in table 1).
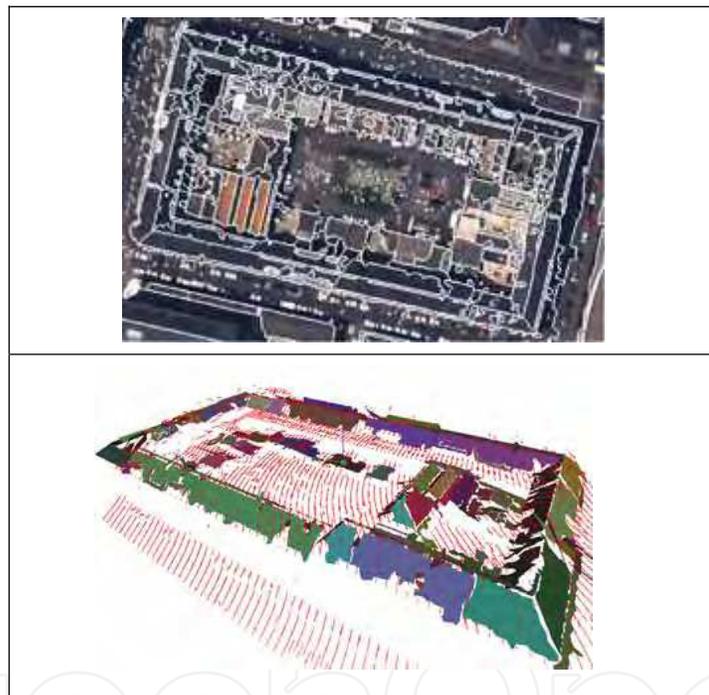
Fig. 5. Result of the algorithm on a complex building.

The 3D region contours are calculated from 2D region contours. In case of overlapping regions, the smallest one is extruded from the largest, which explains the general shape of the presented facets.

4. Conclusion

We have presented a methodology for extracting roof facets over buildings by merging aerial images and 3D lidar data in a hierarchical segmentation framework. Building roof facets are detected using radiometric, semantic and geometric information of images and of

lidar data. We have shown that integrating lidar points in an image segmentation process has enhanced the potentialities of using only 3D lidar points for extracting planar surfaces.
The 3D facet contours are not accurate or realistic even if they are based on the image contours. This is mainly due to remaining small regions located at the region borders. Seeing that images have been resampled to fit the ortho-rectified geometry, facet contours do not take benefit of the original image geometry. The future work consists at first in using aerial images in their original geometry to avoid the resampling artefacts onto building borders. In a second step, we would like to derive global criteria to provide admissible range values of parameter *s*.

5. References

Baillard, C. & Dissard, O. (2000). A stereo matching algorithm for urban digital elevation models. *Photogrammetric Engineering and Remote Sensing*, 66(9), Sept. 2000

Baltsavias, E. P (1999). Airborne laser scanning: Basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:1999-214, 1999

Deriche, R. (1987). Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, volume 1(2), pages 167–187, May 1987

Bretar, F., Chesnier, M., Pierrot-Desseilligny, M. & Roux, M. (2004). Terrain modeling and airborne laser data classification using multiple pass filtering. *Proceedings of the XXth ISPRS Symposium* , pages 314–319, Istanbul, Turkey, 2004.

Guigues, L. Lemen, H. & Cocquerez, J-P. (2003). Scale-sets image analysis. *ICIP'03*, pages 299-306, Barcelona, Spain, Oct. 2003

Jibrini, H., Pierrot-Deseilligny, M. & Maître, H. (2000). Automatic building reconstruction from very high resolution aerial stereopairs using cadastral ground plans. *Proceedings of the XIXth ISPRS Symposium*, Amsterdam , The Netherlands, 2000.

Haala, K. & Walter, V. (1999), Automatic classification of urban environment for database revision using lidar and color aerial imagery. *IASPRS,* Valladolid, Spain, 1999

Haris, K., Efstratiadis, S., Maglaveras, N. & Katsaggelos, A. K. (1998), Hybrid image segmentation using watersheds and fast region merging. *IEEE Transactions on Image Processing*, 7(12):1684–1689, Dec. 1998.

Kasser, M. & Egels, Y. (2002). *Digital photogrammetry*. Hermes – Lavoisier, Paris

Lafarge, F., Descombes, X., Zerubia J. & Deseilligny, M. P. (2006). An automatic 3D city model : a bayesian approach using satellite images. *Proc. IEEE InternationalConference on Acoustics, Speech and Signal Processing (ICASSP),* Toulouse, France, 2006

Pierrot-Deseilligny, M. & Paparoditis, N. (2005). A Multiresolution and Optimization-Based Image Matching Approach : An Application to Surface Reconstruction from Spot5-HRS Stereo Imagery. WG I/5 & I/6 Workshop on Topographic Mapping from Space (with Special Emphasis on Small Satellites), Ankara, 2005

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K. & Lovell, B. C. (2004). Building detection by Dempster-Shafer fusion of Lidar data and Multisepctral aerial imagery. *ICPR*, pages 339-342, 2004

Vision Systems: Segmentation and Pattern Recognition

Edited by Goro Obinata and Ashish Dutta

Research in computer vision has exponentially increased in the last two decades due to the availability of cheap cameras and fast processors. This increase has also been accompanied by a blurring of the boundaries between the different applications of vision, making it truly interdisciplinary. In this book we have attempted to put together state-of-the-art research and developments in segmentation and pattern recognition. The first nine chapters on segmentation deal with advanced algorithms and models, and various applications of segmentation in robot path planning, human face tracking, etc. The later chapters are devoted to pattern recognition and covers diverse topics ranging from biological image analysis, remote sensing, text recognition, advanced filter design for data analysis, etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Frederic Bretar (2007). Data Fusion in a Hierarchical Segmentation Context: The Case of Building Roof Description, Vision Systems: Segmentation and Pattern Recognition, Goro Obinata and Ashish Dutta (Ed.), ISBN: 978-3-902613-05-9, InTech, Available from:
http://www.intechopen.com/books/vision_systems_segmentation_and_pattern_recognition/data_fusion_in_a_hierarchical_segmentation_context__the_case_of_building_roof_description