

# THREE-DIMENSIONAL OBJECT RECOGNITION IN LIDAR DATA USING A PLANAR PATCH APPROACH

Paolo GAMBÀ, Fabio DELL'ACQUA, Marcella CESARI

Department of Electronics, University of Pavia, ITALY

{paolo.gamba, fabio.dellacqua}@unipv.it

Working Group III/5

**KEY WORDS:** Urban remote sensing, LIDAR, building extraction, best plane fitting.

## ABSTRACT

In this paper we report on development and testing of some improvements to recently introduced object extraction techniques suitable for 3D remotely sensed data such as LIDAR or InSAR. Typical urban entities such as buildings, trees, roads, etc. are characterized based on their 3D structure without a need for referencing a given model. Regularisation and segmentation algorithms are applied to the original Digital Surface Model to remove noise and other artifacts; then, best-fitting-plane criteria are applied to the cleaned data in order to partition the scene into a set of planar patches which can constitute the base element for a subsequent, model-driven, recognition and refinement step. Results are shown on a LIDAR data set over the city of Parma in Northern Italy.

## 1 INTRODUCTION

A precise characterization of the urban environment from remote sensed data requires efficient and robust algorithms for data interpretation (Maas and Vosselman (1999)-Gamba and Houshmand (2000)). Model-driven characterisation systems need a collection of base elements on which to fit the embedded models and recognise objects. In this paper we propose an algorithm able to extract planar patches aimed at feeding such basic elements into further processing stages for object recognition. This paper is organised as follows: next section describes the approach that allows to incrementally partition the image into planar patches, section 3 describes the LIDAR data sets used for our experiments, section 4 presents the results obtained on the data sets mentioned, section 5 draws some conclusions.

## 2 PLANAR PATCH APPROACH

The methodology developed in this paper relies on the results and the algorithms described in Gamba *et al.* (2000), where each building is detected and isolated from the surroundings by means of a suitably modified machine vision approach, originally developed for range image segmentation (Dell'Acqua and Fisher, 2001). The procedure is based on a local approximation of the 3-D data by means of best-fitting planes. In this way, a building footprint, height and position, as well as its description with a simple 3-D model, are recovered by a self-consistent partitioning of the topographic surface reconstructed from interferometric radar data.

In our case, the procedure is composed of a segmentation phase followed by a reconstruction phase. Finally, elevated objects are detected.

### 2.1 Segmentation phase

The original raster Digital Surface Model (DSM) represented as an image containing one height datum in meters in every pixel, is partitioned into four quadrants. In each of

the quadrants the mean value is computed and the difference is computed between such mean values and each of the pixels. If for at least one pixel the difference is greater than 4 m (a value found heuristically), the involved quadrant is in turn partitioned again and each sub-quadrant is subject to the same procedure. The procedure stops when either the size of the sub-quadrants is  $2 \times 2$  or none of the sub-quadrants satisfies the condition for further partition. Information about each block (quadrant or sub-quadrant) generated in the segmentation phase are stored in a structure array containing the block coordinates, the maximum, the minimum and the average of the height values in the block. The coefficients of the plane best approximating the associated portion of the DSM are computed using a least squares approach and stored together with the blocks.

### 2.2 Reconstruction phase

In turn, each of the blocks generated in the segmentation phase is compared with the adjacent blocks. Two conditions are evaluated:

1. the difference between the maximum of one block and the minimum of the other block must be below  $1.5 \times H_{th}$ , where  $H_{th}$  is the height threshold formerly set, i.e. 4 m;
2. the planes fitting the two blocks are *similarly* tilted, as explained below.

The block pairs complying with both the enumerated conditions are merged together. The second, "co-planarity" condition is based on a representation of the approximating planes like in (1):

$$ax + by - z + d = 0 \quad (1)$$

Two planes are considered *similarly* tilted if the set of conditions (2) is met:

$$\left\{ \begin{array}{l} \cos(\Theta) \geq 0.93 \\ a_1 \cdot a_2 \geq 0 \\ b_1 \cdot b_2 \geq 0 \end{array} \right\} \quad (2)$$

subset name	#columns	#rows	tot#pixels	#building pixels	#terrain pixels
HOUSES	182	121	22,022	5,950	16,072
CATHEDRAL	70	85	5,950	3,610	2,340
BUILDING	11	113	12,543	4,727	7,816
PIAZZA	67	115	7,705	3,304	4,401
UNIVERSITY	102	110	11,220	7,491	3,729

Table 1: Features of the 5 Parma subsets

where  $\Theta$  is the angle between the plane normals, that can be computed with the following formula:

$$\cos(\Theta) = \left| \frac{a_1 \cdot a_2 + b_1 \cdot b_2 + 1}{\sqrt{(a_1^2 + b_1^2 + 1) \cdot (a_2^2 + b_2^2 + 1)}} \right| \quad (3)$$

In (3) and (2), the variables  $a_k, b_k, k = 1, 2$  are defined as in (1) for each of the two considered blocks. The lower bound on the cosine corresponds to two normals which are at most 20.5 deg. apart from each other. This value was set by experiment, and in the case of the Parma data set, it allows to aggregate about 15% of the examined blocks. The two constraints on the sign of  $a_k, b_k$  are necessary to avoid merging planes that are oppositely tilted, as it frequently happens in urban areas with the two opposite sides of pitched roofs. One last note regards a threshold on the slope of the approximating planes: when  $|a| < 0.005$  and  $|b| < 0.005$  the block is assumed to be horizontal and thus both  $a$  and  $b$  are set to zero. Horizontal planes in urban areas are very frequent, while planes with a very light inclination may be simply artifacts resulting from the noise inherent in the sensed data.

### 2.3 Detection of elevated objects

Next phase of processing consists of detecting elevated objects, assumed to be the buildings present in the scene. The first step is the detection of the presumed terrain level; the terrain is assumed to be the area with the maximum extent and the lowest elevation; in the considered data both constraints have always been satisfied by the same areas, so there has been no need to define a priority. In the future, however, with different datasets, it will be probably necessary to do so. Starting from the individuated reference, all the pixels are examined and compared with the presumed terrain level. It is assumed to be belonging to the terrain every pixel satisfying the constraint:  $pixelelevation < terrainelevation + 2 \cdot H_{th}$ .

All the remaining pixels belong to elevated objects. Regions with a majority of non-terrain pixels are assumed to be elements of building profiles.

## 3 THE DATASETS

The above mentioned procedure was applied to two data sets, in different urban areas. These data sets correspond also to different city structures. In particular, Parma is an old European city, while San Francisco is a North American city.

### 3.1 Parma data set

The dataset has been acquired on the town of Parma, Northern Italy, in June 1998 with the Toposys sensor installed on a plane of an Italian company called CGR, Compagnia Generale Ripresearee. The flight height was around 800 meters; the Toposys sensor is able to acquire, flying at that height, approximately five points per square meter, so that the one-meter grid which is usually delivered to the customers, and that we used, can be calculated with a good reliability. Up to now the Toposys instrument is unable to measure the reflected signal intensity, so it gives pure geometric data and it can acquire first pulse or last pulse alternatively: our data has been acquired in the last pulse mode.

From the above data we extracted a series of 5 rectangular sub-areas, whose conventional name and description are: “HOUSES”, a series of residential units with pitched roofs, approximately square in shape, organised in a pattern of nearly parallel lines, “CATHEDRAL”, the town cathedral, latin-cross-shaped, with segmented, pitched roofs, a dome on the cross center and a tower on one corner, “BUILDING”, a single, U-shaped building with pitched roof, surrounded by tall trees, “PIAZZA”, an open market place amidst a series of buildings with different geometrical shapes, with both pitched and flat roofs, plus one dome on top of a building, “UNIVERSITY” the university site, a rectangular-shaped building with two different sized inner yards, pitched roofs, a tiny tower and a dome stemming out of the roof top; a few smaller buildings around. The LIDAR dataset came with a series of aerial pictures of the city which allowed us to build a satisfactory ground truth to verify the results. Care was taken to counteract the effect of aerial perspective, in particular the displacement due to the off nadir observation angle.

### 3.2 San Francisco data set

As second set of LIDAR data, a portion of the Presidio area in San Francisco, California is used for this study. This area contains natural topography such as hills, open areas, tree covered areas in addition to buildings of small to moderate footprint and heights. The LIDAR data is reported on an 80 cm grid. As expected for high resolution LIDAR data, the building geometries are observable for large and small structures. The regions covered by trees are denoted by a texture which is distinct from the neighbouring areas.

## 4 EXPERIMENTAL RESULTS

The building profiles obtained at the building detection stage are compared with the ground truth extracted from the aer-

ial photographs. Table 2 reports the size and the number of non-terrain (building) pixels in the various subsets considered.

Comparison with the ground truth data produced accuracy data for each of the subsets. The following table reports percentages of overlap between pixels assigned to the class *building* using the procedure described in section 2.

subset name	% overlap
HOUSES	77.55%
CATHEDRAL	94.02%
BUILDING	84.94%
PIAZZA	92.98 %
UNIVERSITY	90.01 %

Table 2: Overall accuracy values of the 5 Parma subsets

It can be noted that in terms of percentage of correctly classified pixels, the results are satisfying in most cases. For subset HOUSES, which gave a worse than average result, the main source of errors were the tall trees surrounding the house rows, as visible in figure 1.

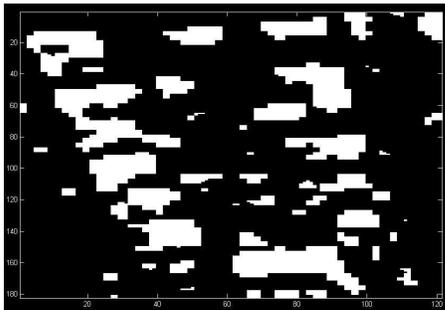


Figure 1: Output on the HOUSES subset. White pixels have been assigned to the class *buildings*

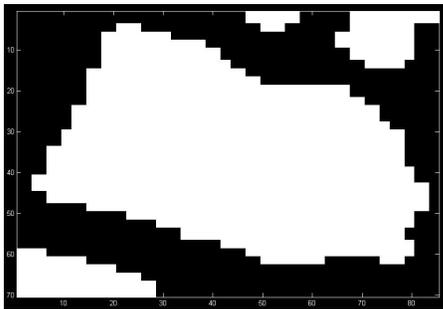


Figure 2: Output on the CATHEDRAL subset. White pixels have been assigned to the class *buildings*

As for the second data set, figure 3 shows an example of the building extraction algorithm, applied to a large structure (an hospital) in the same area. The reconstructed 3D shape, obtained by starting from LIDAR data, is now made by a very small number of planar patches. Moreover, many of these patches are similarly orientated, as we expect in a man-made object. If nearby patches have higher orientation variance, probably it is because they belong to tree canopies or to natural surfaces.

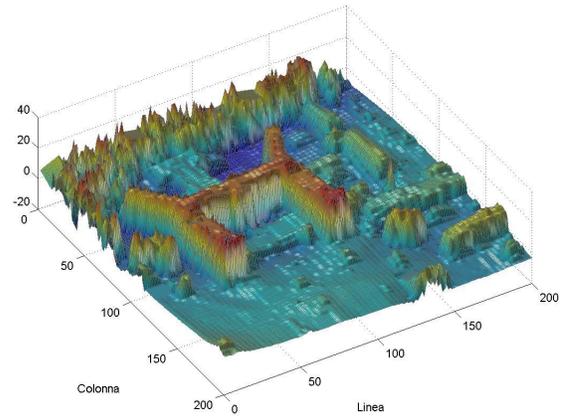


Figure 3: An example of output on the San Francisco dataset.

## 5 CONCLUSIONS

A method to detect building elevated from the terrain in LIDAR images, through partition into planar patches has been described, and the results of its application to an old European city (Parma, Northern Italy) and to a USA city (San Francisco) have been presented. The results on Parma are satisfactory, with accuracy over 90% in 3 cases out of 5. In the two worst cases most wrongly assigned regions were non-building pixels mistaken for buildings due to the presence of canopy trees, which were dense enough to trigger the building detectors despite having used last pulse LIDAR data. The results on San Francisco can be evaluated only qualitatively, as no ground truth were available at the time of the experiments. The appearance of the results is however good, as seen in figure 3.

## References

- H.-G. Maas and G. Vosselman: Two algorithms for extracting models from raw laser altimetry data, ISPRS J. Photogramm. Remote Sensing, Vol. 54, No. 2-3, pp. 153-163, July 1999.
- P. Gamba, V. Casella: Model independent object extraction from digital surface models, International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B3/1, pp. 312-319, 2000.
- P. Gamba, B. Houshmand: Digital surface models and building extraction: a comparison of IFSAR and LIDAR data, IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, n. 4, pp. 1959-1968, July 2000.
- P. Gamba, B. Houshmand, M. Saccani: Detection and extraction of buildings from interferometric SAR data, IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, n. 1, pp. 611-618, Jan. 2000.
- Fabio Dell'Acqua, R. Fisher: Reconstruction of planar surfaces behind occlusions in range images, IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 24, n.4, pp. 569-575, 2001.