

MODEL INDEPENDENT OBJECT EXTRACTION FROM DIGITAL SURFACE MODELS

Paolo GAMBÀ*, Vittorio CASELLA**

*University of Pavia, Italy

Department of Electronics

gamba@ele.unipv.it

**Department of Building and Territory Engineering

casella@unipv.it

KEY WORDS: DSM, building extraction, urban characterization.

ABSTRACT

This paper is devoted to analyze and improve a recently introduced technique that may be applied to Digital Surface Models (DSMs) produced by different sensors (Interferometric SAR, for instance, or LIDAR, or even photogrammetry) for object detection and extraction. The proposed approach, based on modified image analysis techniques, does not need to specify any surface or object model of the object to be recognized. To this aim, we apply a jointly regularization and segmentation algorithm to the original data, so that it is easy to partition them into significant structures, the background, and uninteresting areas from a 3D point of view). The criteria applied to segment the data are geometric ones, involving the principle of segment and plane-fitting.

By means of this approach we will show that it is possible to better characterize urban scenes in terms of buildings, trees, roads and other natural and geometrical structures. In particular, looking at their 3D structure we may be able to discard natural objects without referring to given object models. The original data resolution is exploited as much as possible, and the method provides a high robustness to noise.

1 INTRODUCTION

Urban three dimensional (3D) geometry and land cover are among the widely required for an efficient urban analysis. In turn, this analysis may be useful for a number of applications, including urban monitoring, change detection, damage assessment, as well as traffic and heat modeling, and finally urban scene visualization for architectural purposes.

Currently, only a few sensors are able to produce data sufficiently accurate for such an analysis. In particular, Digital Surface Models (DSMs) for this complex environments may be computed starting from aerial stereo high resolution photographs, from Laser Ranging (LIDAR) systems and from Interferometric Radar (IFSAR) measurements. These different sources provides differently characterized DSMs, so that their analysis requires the capability to cope with a number of different problems and noise sources.

This paper is devoted to develop and test some recently introduced techniques that may be indifferently applied to these DSMs for a suitable object detection and extraction. These techniques are based on machine vision and pattern recognition techniques, and do not need to specify a surface or object model to which the data is compared. Indeed, to this kind of *parametric* approach may be reduced most of the systems recently proposed in literature (Gruen and Wang, 1998, Gruen, 1998), and only a few *non-parametric* algorithms may be also found (see the second part of Haas and Vosselman, 1999, for instance). The goal of this paper is instead to characterize urban scenes in terms of buildings, trees, roads and other natural and geometrical structures looking at their 3D structure, but without the reference to a given model.

More in detail, for extracting geometrical structures we apply regularization and segmentation algorithms to the original Digital Surface Models. The goal is to exploit their resolution, maintain a high robustness to noise, and partition the original data into background and significant objects. In particular, the criteria applied to segment the data are geometric ones, involving the principle of segment and plane-fitting. With this approach, objects and structures may be characterized as a set of planar regions whose relationships may be used, in a second time, for a model-driven recognition and refinement step. This approach corresponds to looking for the building roofs and walls, discarding noisy measurements by a clustering algorithm that aggregates points belonging to the same surface by means of a region growing technique.

The algorithms presented in this paper were originally applied to the IFSAR DSM of a part of the Los Angeles urban area composed of isolated financial buildings (Gamba and Houshmand, 1999), and discarding the problems due to shadowing/layover effects. In this paper we focus on a more densely built area, with trees, roads and a more complex topography. Moreover, we use LIDAR measurements, that provide a more accurate representation of the urban three dimensional geometry and a urban Digital Surface Model with cm level accuracy.

2 THE DSM OBJECT EXTRACTION ALGORITHM

No question that one of the open problems of remote sensing data interpretation, especially when dealing with 3D (also called *range*) images is how to segment the data into significant parts. This is usually accomplished by looking inside the same data for some kind of common features (classification) or by comparing data subsets with known structure models for the area (object recognition). The dilemma is between a rapid recognition process (that usually discards minor details and retrieve simple geometrical characteristics) and precise classification (when every pixel is assigned to a cluster).

However, different kinds of data often correspond to different problems (for instance, noise sources) and different retrieval algorithm. In urban areas this is especially true, because of the large interaction among the geometric, dielectric and topographic characteristics of the objects involved. In this situation it is clear that best results can be obtained only by using additional information, like those already introduced in a GIS of the same area or in an other kind of data (Haala, 1999). Anyway, even in large populated areas it is difficult to have these data readily achieved, or even available. That's why there is a number of approaches dealing with fusion of data coming from different source, possibly mounted on the same platform.

In this work we want to introduce an algorithm based on machine vision techniques for surface retrieval and detection, based only on the (*a priori* known) geometrical properties of the structures to be identified. With such an approach, different sources of imprecision and noise may be treated in the same way, as errors in the range data acquisition. Even if we do not exploit the noise characteristics to discard it, we delete its effects on the geometrical structures by recovering from the distortion that it introduces. For instance, this algorithm may be very interesting to enhance the 3D geometry retrievable from interferometric SAR measurements. In this case, the so called *layover* and *shadowing* effects, due to multiple bounces of the electromagnetic backscattered pulse, usually produce relevant changes in the 3D structure of the buildings in a urban, crowded environment. Still, the algorithm introduced in Gamba and Houshmand, 1999 and Gamba et al., 2000 succeeds in retrieving to a large extent the main objects in such an environment. In the following we give a quick outline of the procedure, trying to highlight the advantages of our proposal for LIDAR data analysis as well as the drawbacks still to be overcome. The algorithm is based on a fast segmentation procedure outlined in Jiang and Burke, 1994, even if it provides a different way to aggregate range points and has been improved to cope with remote sensed data.

As a final comment before presenting the algorithm, we should note that the algorithm requires a regular data grid, and works on lines (or columns of this grid). While this allows a simple and intuitive way to aggregate pixels into segments (the first step of the procedure), it has all the drawbacks of a discretization process, especially when looking for building footprints. Moreover, LIDAR data, that are naturally sparse and not regularly equally spaced, must be pre-processed before entering the algorithm.

The data flow may be subdivided into five steps.

1. As noted above, the regular data grid is first subdivided into segments, working by rows or columns. To this aim, the algorithm introduces two parameter (α_1 and α_2 , see Table 1), which rule the process. The first one defines the maximum height difference that is allowed between successive points to be in the same segments. The second one, instead, provides the inferior limit to the segment lengths. Both of them must be carefully chosen, in order to avoid excessive segmentation (and more difficult aggregation in the following steps). Moreover, from the opposite point of view, we should not low-pass too much the data (as possible with thresholds allowing long segments), because we may loose important information hidden in the data. As usual, this leads to a compromise, that must even take into account the data resolution and the mean dimension of the objects to be extracted.
2. Once the segments have been individuated, they are compared two at a time, to detect areas where collinear couples are present. They provide the *seeds* of the planar patches that we want to define. Here the parameter ruling the process (α_3) is the minimum similarity value that we have to find to allow to segments to be a seed. The choice suggested by the original algorithm in Jiang and Burke, 1994 is to take the similarity between the *i*-th and *j*-th segment as measured by

$$s_{ij} = \frac{1}{2} \left(\frac{m_i m_j + 1}{\sqrt{m_i^2 + 1}} + \frac{n_i n_j + 1}{\sqrt{n_i^2 + 1}} \right) \tag{1}$$

where $z = m_i x + n_i$ is the generic segment equation. We will see in the following subsection that this choice could be improved.

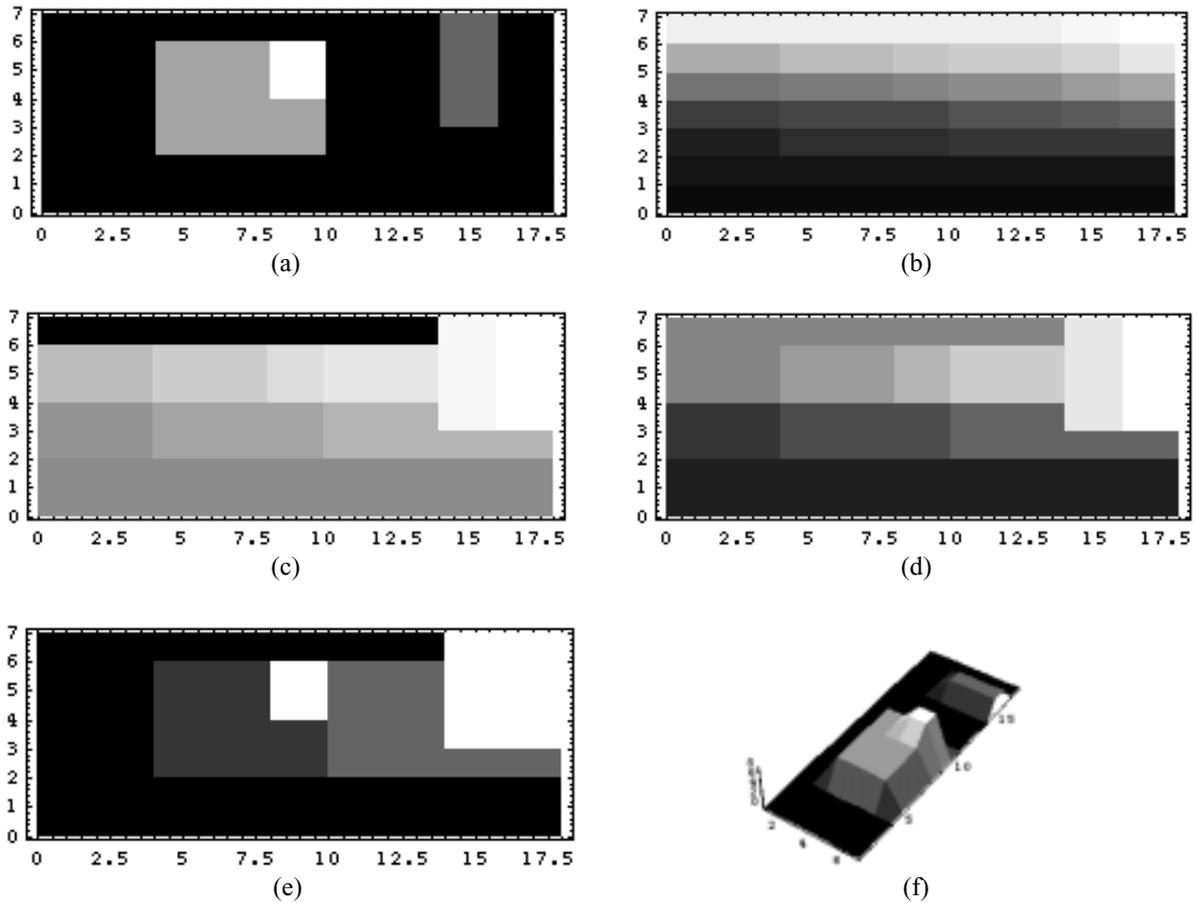


Figure 1. An example of data analysis by the proposed grouping algorithm: (a) the original data, (b) the segment extracted, (c) initial plane seed detected, (d) planes after segment and point aggregation, (e) final planar reconstructed surfaces with (f) their 3D profile.

- The third step is a region growing algorithm, looking for other segments to be aggregated to the original seeds. After each aggregation the parameters of the surface are updated, and a new segment search is performed. Planar patches and linear elements are considered as belonging to the same plane if their orientations in space are sufficiently similar, i.e. if the projection of the patch on the segment direction is sufficiently large. In this case no additional threshold is needed, since the segment similarity required for seed detection is used also for this step. Because the value of α_3 is usually very high, not many segments are aggregated, at the end of this step, especially in areas where many objects crowds or the data is affected by noise. From the other end, a lower values of this parameters may provide a larger regularization of the detected surfaces, possibly losing smaller details.

Parameter	Meaning	Suggested value depending on
α_1	max. step between points in the same segment	data vertical precision (as large as possible)
α_2	min. segment length	data horizontal precision (as small as possible)
α_3	min. similarity value	data imprecision (between 0.5 and 0.8)
α_4	max. distance point-plane	data vertical precision (as large as possible)

Table 1. The parameters of the proposed algorithm.

4. Finally, a further merging step takes into account separately all pixels left. In other words, we now neglect the segment analysis of the first step for all those pixels that have not been aggregated into a plane, and we try to merge them one at a time with the nearest plane. The largest value of the distance that a point is allowed to have from a plane and be anyway aggregated correspond to the fourth parameter in Table 1, α_4 . Even in this processing step the computations is repeated after each computation of the plane parameters; by this way, we take into account the newly added points in the successive iteration. However, since at this point the planar patches are usually sufficiently large, the adjustments are very limited.
5. The last step involves all the planar surfaces already recognized, that may be further aggregated into larger planes to improve the process output and a better recognition of the objects in the scene. The mean for realizing this operation is again a similarity measure:

$$s = \frac{1}{3} \left(\frac{aa' + 1}{\sqrt{a^2 + 1}} + \frac{bb' + 1}{\sqrt{b^2 + 1}} + \frac{cc' + 1}{\sqrt{c^2 + 1}} \right) \quad (2)$$

where a, b, c are the three parameters that identify each planar patch.

The approach does not guarantee that all the original range image is subdivided into planar regions, and isolated points or even small regions may not considered in the final output. This could be considered a drawback, in some cases, since these areas represent the “noise floor” of the algorithm, positions where the procedure is not able to recover the true surface from the distorted measurements. However, it could be also an advantage for some kind of data. For instance, we should take in mind that LIDAR data is affected by random noise at the edge of the buildings, because of false responses. So, it could be useful to discard these points both for a successive edge analysis and also to prevent them from being cause of large errors in the object reconstructed shape.

Figure 1 provides a way to understand how the plane growing procedure acts through steps 1, 3, 4, and 5 of the algorithm. Indeed, we may observe the original data (in this case, a very simple test image), the boundaries of the planar regions initially detected, the final regions obtained after segment, point and plane aggregation, and the final result. Black pixels in fig. #(c) represent areas that we were not able to characterize at the corresponding processing point.

2.1 Some considerations on similarity evaluation

Similarity is the key of the preceding algorithm, both for the initial data mining and the successive object reconstruction process. In this subsection we shall discuss the formulas used for this evaluation, and explain a better formulation. Moreover, we shall introduce with an example why this similarity evaluation could be modulated for a refinement of data analysis and to discriminate as much as possible between natural and artificial objects.

The similarity index used for both the segment and the plane aggregations are based on the work in Jiang and Burke, 1994, but reflects a similarity evaluation with some drawbacks, that could be considered looking at the following figure, where we plot s_{ij} with respect to m_j , taking $m_i=10, n_i=n_j=0$ and following (1) or a different formulation (to be discussed next and presented in (3)).

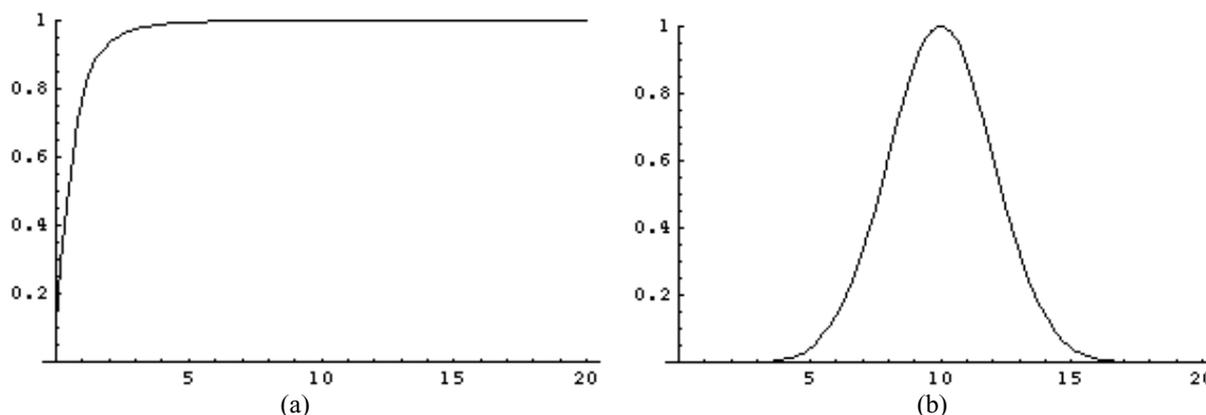


Figure 2. Similarity index for a monodimensional segment with length equal to 10 with respect to a second segment: in (a) the index is computed according to (1), while in (b) according to (3).

It is clear that fig. 2 (b), where similarity is computed according to

$$s_{ij} = \frac{1}{2} \left(e^{-\frac{25(a-a')^2}{2a'^2}} + e^{-\frac{25(b-b')^2}{2b'^2}} \right) \quad (3)$$

provides a better way to characterize the “differences” between two segments, allowing to discard samples that are either too long or too short with respect to the reference.

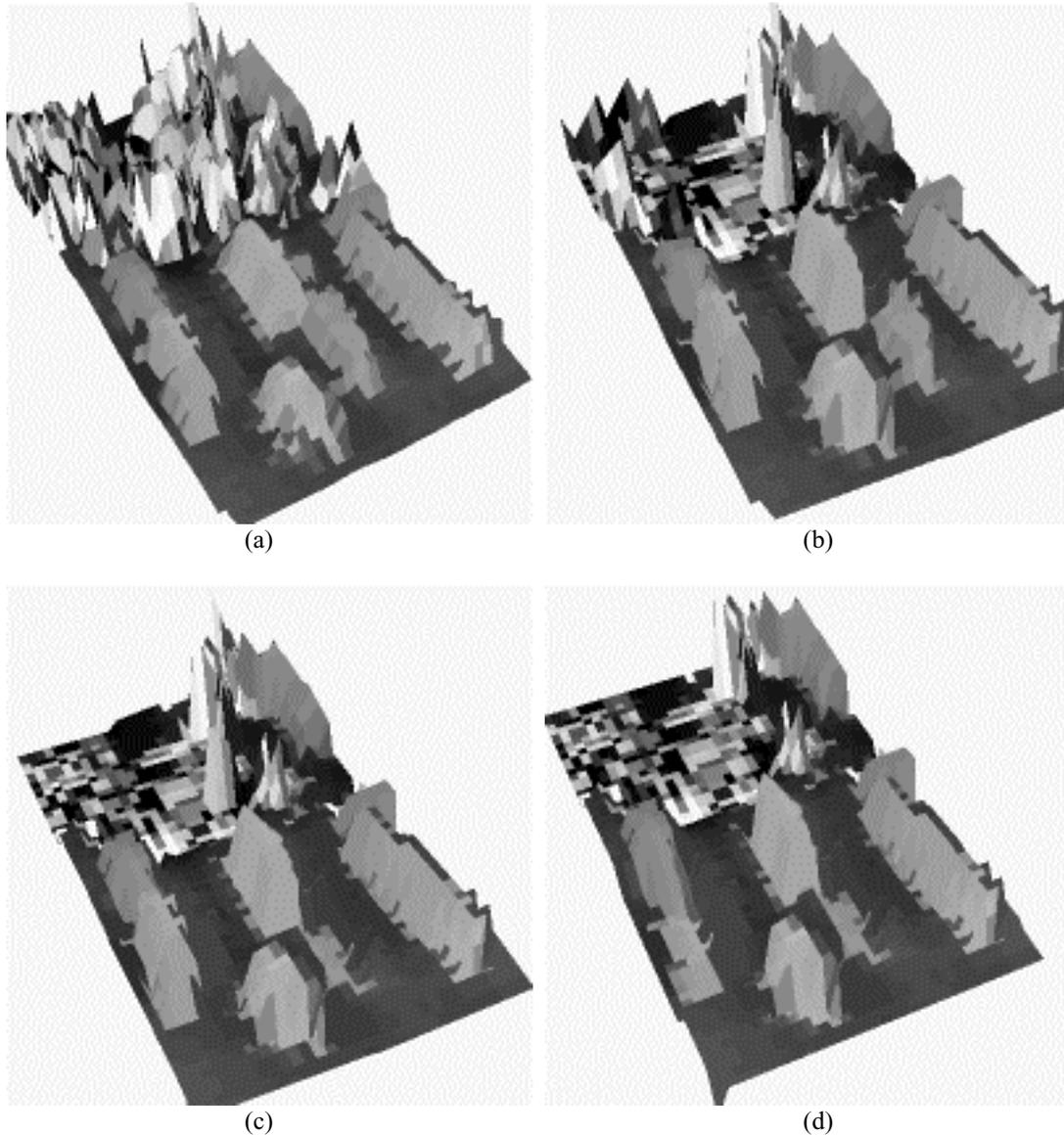


Figure 3. An area with both vegetation and buildings: the original LIDAR data is in (a), while the objects recognized with a segment similarity threshold of 0.5, 0.7 and 0.8 are respectively in (b), (c) and (d).

The second point to be discussed, i. e. the importance of index s_{ij} for object discrimination, may be more evident by looking at some results on a DTM where both trees and buildings are present, as depicted in fig.3 (a). vegetation corresponds to an area where the DTM is extremely irregular, the initial data segmentation provides only very short segments, and similarity between neighboring segments, in the sense that we defined above, is, when applicable, a very small value. In the figure the output of the algorithm changing only the minimum similarity value required to aggregate a segment to a plane (either a seed, or a grown plane) is changed, from 0.5 to 0.8. We observe that of course the buildings, if decomposed into segments, define segment sets with strong similarity, while the vegetation provides a surface where segments may be considered as equally oriented in all directions. When applying larger and larger similarity thresholds, only a few of these segments survive the second and third step of the above delineated procedure, providing a way to recognize where interesting or useless areas are located. If the threshold is too large, however, also

parts of the building start to be (erroneously) discarded, due to the small acquisition errors or noise in range measurements; this is clear, in fig. 3 (b) and (c), looking at the central and leftmost built structures.

3 EXPERIMENTS

The preceding technique has been applied to a number of different sites corresponding to urban structures in downtown areas. In particular, we refer here to some examples exploiting data over the town of Parma, Northern Italy. The dataset has been acquired on the town of Parma 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 isn't able 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. Nevertheless, thanks to the well equipped and powerful plane, we could also acquire, during the laser flight, aerial photogrammetric images that will be exploited in a further development step of our algorithm.

The first example is depicted in fig. 4, and represents Piazza Garibaldi in Parma. We observe that the 3D shapes of the buildings in the area are much more delineated after the building extraction procedures than in the original data: all the significant features are considered, but the vegetated area in front of the tower building has disappeared, thanks to a sufficiently high similarity threshold (see previous section). Even the 2D shape (or better, footprint) of the buildings has been improved by the applied algorithm.

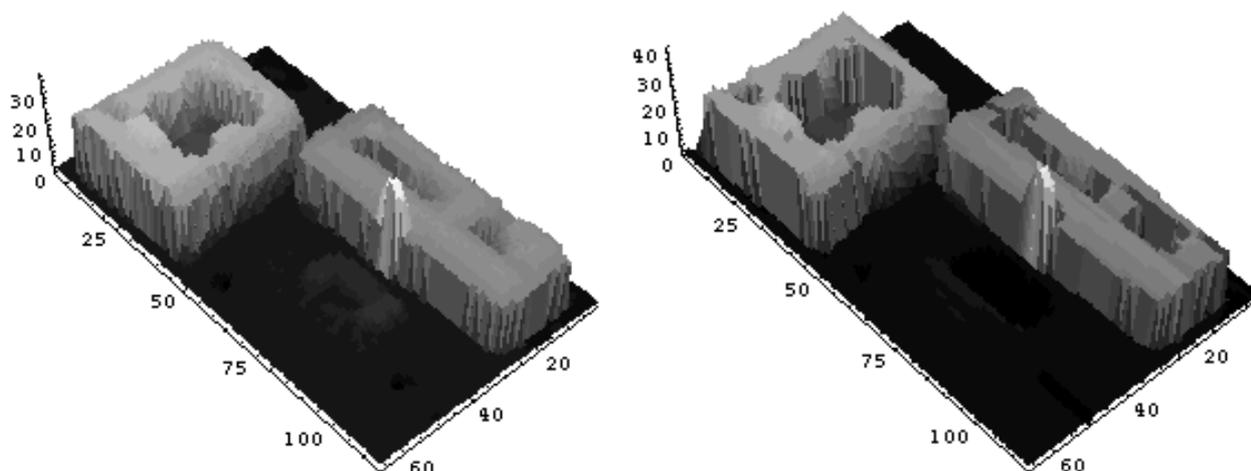


Figure 4. Data over Piazza Garibaldi, Parma, before and after object extraction.

The second example portrays a different area in Parma, a residential area with a group of large houses and some trees. Even in this case the regularization of the 3D and 2D shapes of the artificial structures is evident. This is true especially looking at the small errors in the LIDAR DTM near the walls; after the building extraction procedure this kind of imprecision has almost completely been discarded. Moreover, the trees and bushes in the area have been either discarded or extremely regularized, so that their presence could be easily recognized with further simple processing steps.

However, fig. 5 evidences also some problems in the procedure, already noted in Gamba and Houshmand, 2000, and related to the need to have a regular grid as the input of the analysis algorithm. When the searched objects are not exactly along the rows or columns of this grid, as in this case, we have possible reconstruction error, like those in the internal (with respect to the viewer) wall of the largest building. A second problem source is the use of a unique value for the threshold values in Table 1 through all the analyzed image. This is the cause, for instance, of the fact that the elevator small towers on the roof of the same large building are only partially considered in the output image (more precisely, the foremost one has been discarded, while the one in the background has been considered). A possible solution to this drawback is the introduction of an adaptive threshold, taking into account the local regularity (to be somehow defined) of the area around each considered segment.

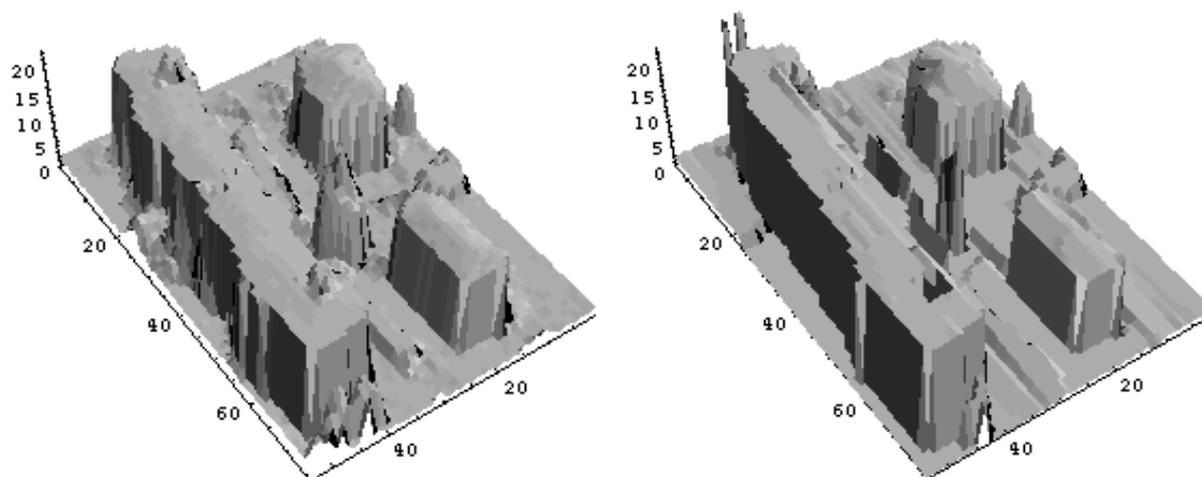


Figure 5. Data over a residential area in Parma, before and after object extraction.

4 CONCLUSIONS

The above presented results show that natural and artificial structures are differently treated by the proposed algorithm. In fact, even if not based on particular assumptions about object shapes, the *similarity* concept provides a powerful mean to discriminate between buildings and vegetation. This possibility, clearly susceptible of further refinements, may be useful for many different applications, like some kind of support to classification algorithms and to the procedures for Digital Terrain Models (DTMs) generation. Indeed, DTM characterization starting from range data is one of the more interesting open research topics, and this algorithm may provide some hints on how to define adaptive filtering techniques suitable for discarding equally well buildings and trees.

More in general, the aim of this work was to prove that even when no additional information on a site is available, and only range data could be considered, model-independent algorithms may provide useful (and, sometimes, enough) information to understand the environment.

We do not pretend that this and similar algorithms are able to analyze completely this kind of data sets, and believe that *data fusion*, i. e. the use of many different sources of data, is the right way to characterize complex environments like towns and cities. However, the effort to extract as much as possible from any single sensor's data may be helpful to this wider process.

ACKNOWLEDGMENTS

We are grateful to the Italian National Research Project entitled *Digital surface modelling by laser scanning and GPS for 3D city models and digital ortophoto*, financed by the Italian Ministry of the University for the 1998 year, and chaired by the prof. Galetto of the University of Pavia for providing the data over Parma and Pavia. The research presented in this paper was funded by the University of Pavia under the "Progetti d'Ateneo" project.

REFERENCES

- Gamba P., Houshmand B., 1999. Three-dimensional urban characterization by means of IFSAR measurements, Proc. of the 1999 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'99), Vol. V, pp. 2401-2403, Hamburg, Germany.
- Gamba P., Houshmand B., Saccani M., 2000. Detection and extraction of buildings from interferometric SAR data, IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, n. 1, pp. 611-618.

-
- Gamba P., Houshmand B., 2000. Digital surface models and building extraction: a comparison of IFSAR and LIDAR data, accepted on IEEE Transactions on Geoscience and Remote Sensing.
- Gruen A, Wang X., 1998. CC-modeler: a topology generator dor 3-D city models, ISPRS J. Photogramm. Remote Sensing, Vol. 53, No. 5, pp. 286-295.
- Gruen A., 1998. TOBAGO -- a semi-automated approach for the generation of 3-D building models, ISPRS J. Photogramm. Remote Sensing, Vol. 53, No. 2, pp. 108-118.
- Haala, N., 1999. Combining multiple data sources for urban data acquisition. Proc. of the Photogrammetric Week 1999, Stuttgart, Germany.
- Jiang X. and Bunke H., 1994. Fast segmentation of range images into planar regions by scan line grouping, Machine Vision and Applications, No. 7, pp. 115-122.
- Maas H.-G., Vosselman G., 1999. Two algorithms for extracting models from raw laser altimetry data, ISPRS J. Photogramm. Remote Sensing, Vol. 54, No. 2-3, pp. 153-163.