

# STRATEGY FOR THE EXTRACTION OF 3D ARCHITECTURAL OBJECTS FROM LASER AND IMAGE DATA ACQUIRED FROM THE SAME VIEWPOINT

Matthieu Deveau<sup>a, b</sup>, Nicolas Papanoditis<sup>a</sup>, Marc Pierrot-Deseilligny<sup>a</sup>, Xin Chen<sup>b</sup>, Guillaume Thibault<sup>c</sup>

<sup>a</sup> IGN/MATIS Laboratory, 2-4 avenue Pasteur, 94165 Saint-Mandé - France - [firstname.lastname@ign.fr](mailto:firstname.lastname@ign.fr)

<sup>b</sup> MENSI S.A. 30 rue de la Fontaine du Vaisseau, 94120 Fontenay-sous-Bois - France - [firstname.lastname@trimble.com](mailto:firstname.lastname@trimble.com)

<sup>c</sup> EDF-R&D, CAD and VR Group, 1 Avenue du Général de Gaulle, 92141 Clamart - France - [guillaume.thibault@edf.fr](mailto:guillaume.thibault@edf.fr)

**KEY WORDS:** Laser scanner, image data, terrestrial surveying, 3D reconstruction, semi-automation, data fusion

## ABSTRACT:

This paper presents a general strategy for 3D reconstruction of complex architectural scenes. This approach makes use of both laser data (range and retro-diffusion) and image data, and of their great complementarity. The proposed approach starts with pose estimation between the two data sets. This part is largely automated by means of correlation of fine feature points and linear primitives. The second part deals with segmentation of the scene. Here, each element of the scene must be detected. This can be fulfilled in two ways. The first one is semi-automatic extraction of the objects where user interaction is used to initialize the object localisation, the second one is fully automatic segmentation, driven by image and 3D data. Then, each object needs to be reconstructed. We consider a large range of objects since we can describe planar surfaces, general geometric primitives and generalised cylinders. When these models do not fit with the surface, that is to say on complex parts, triangulation is performed. This leads to two potential reconstruction tools: one similar to laser data editing software, with more efficient object selection and object boundary estimation, the other would perform automatic reconstruction, which would be followed by a verification procedure. The advantage of this approach, compared to existing techniques, is to reduce dramatically reconstruction time, without having any hard limitation on the set of shapes that can be obtained.

## 1 INTRODUCTION

The 3D reconstruction of large scale scenes' geometry is studied for numerous reasons. Researches within this field are motivated by both the practical applications and the automation difficulties involved. One great goal is to complete 3D city models generated from aerial imagery, on historically or architecturally significant buildings. The second main goal lies in cultural heritage digital archiving. A further aim is to answer to geometric reconstruction needs, either for restoration purposes or for archaeological analysis.

Terrestrial environments yield scenes that are complex with respect to the number, the shape and the difference in size of geometric primitives involved. In this context, the description we are looking for is a compact structured surface description, with a high geometric accuracy.

With the laser scanner soaring since the beginning of the 90's, laser scanning has been competing with classical photogrammetry on many surveys. As a consequence meaningful questioning rises about sensing lay-out and sensors complementarity (El-Hakim and al., 2003; Dekeyser and al., 2004). One may emphasize the interest of laser scanning for complex surfaces, while images provide relevant surface boundaries and of course texture. Moreover, scanners have the best of both worlds in bringing a quick description of the scene. But when a structured modelling of the scene is needed, either photogrammetry or laser data editing is appealing for interactive drawing. These operations are extremely time consuming.

Promising work has addressed the automation of terrestrial scene reconstruction either from images (van den Heuvel, 1999; Werner and Zisserman, 2002; Pollefeys and al., 2003; Dick and al., 2004), or from range data (Stamos and Allen, 2002; Han and al., 2003; Yu and al., 2001). Most of these deal with piecewise planar description, or mesh generation. Planar description remains too simple for complex geometry while triangulation has also major drawbacks: non-optimal description length, dependence to point density on discontinuity description (if object boundaries are not available), lack of measurement redundancy on same features, and no object oriented description of the scene. Most of current known work dealing with combining laser data and image address the problem through an additive scheme: laser data and image

add a contribution independently from each other. Here, we want to combine data in the heart of the reconstruction process. (Dias and al., 2001) have used images and triangulation to complete locally the point cloud, on complex structures or on edges. This technique is useful for occluded parts filling, but does not deal with reconstruction into geometric features.

### 1.1 Sensing lay-out

As laser scanning provides accurate and dense 3D information, it can be preferred to stereo based techniques, which encounter difficulties when faced with deep scenes or poorly textured surfaces. Moreover, scanner constructors offer imaging devices, at the moment mainly for texturing purposes. These imaging devices can have a higher resolution than that of the scanner, and so give good geometric clues, especially on object boundaries. We are thus considering a sensing lay-out close to that of the laser scanner: one scan station and one imaging sensor from the same point of view or from two close viewpoints. The scanning resolution is chosen slightly lower than that of the image, in order to reduce the scanning time and the post-processing time.

This lay-out has got ideal properties for reconstruction: precise geometry is supplied by range data while higher image resolution improves that of the scanner, especially on object boundaries. This lay-out also yields communal occluded regions for both the image and 3D data, which means no need for Z-buffering, and which already provides the right image for texturing. Then, it allows to deal with 3D data through a range image geometry because laser data are still organised (after the registration step, they would not be anymore organised). Finally, it will ease pose estimation automation since details have the same aspect in the range image geometry and in the image geometry.

### 1.2 Data description

Laser data has angular and distance noise, which can be reduced by multiple shots. Precision on distance measurements is lowered to 2 mm in our case studies. Scan resolution for our first tests is around one point every 2 cm. In addition to multiple shots on each point, the large number of measurements on each primitive (plane, cylinder or other...) allows to get a very good estimation of the primitive, through least square fitting.

Image resolution is around one pixel every 5 mm on the object. The radiometric dynamic range is spread out over 12 bits, and the signal to noise ratio is over 300. The CCD array used is 4096 by 4096 pixels. Intern parameters are known, computed independently. Optical distortions are corrected at a tenth of a pixel.

## 2 GENERAL PROCESSING STRATEGY

The general strategy described in figure 1 is split into three stages. The first one solves the orientation between the point cloud and the image. Once pose is settled, the reconstruction question is divided into two very-linked parts: segmentation into consistent primitives and estimation of each primitive. Since segmentation remains a complex problem (a problem that we are currently addressing, through a hierarchical image based segmentation), we propose a semi-automatic restitution scheme, which belongs to the same strategy: detection of each primitive and reconstruction. Thus, the reconstruction of a global 3D scene model will be split into as many elementary reconstruction problems as occurring models: planar surface, surface of revolution, surface of translation, parametric model, surface without any acceptable geometric description. This follows the same framework as 3D data editing software, or CAD software, or even photogrammetric workstations, but it brings powerful automation on the extraction of surface boundaries and accurate primitives' estimation.

As the whole scene cannot be reconstructed from a single viewpoint, we assume that the same reconstruction strategy can be reproduced on the next stations, with minimum overlapping parts, and that reconstructions can be merged afterwards. Overlapping reconstructed parts may improve pose estimation, and allow automatic estimation, again without any target placed in the field.

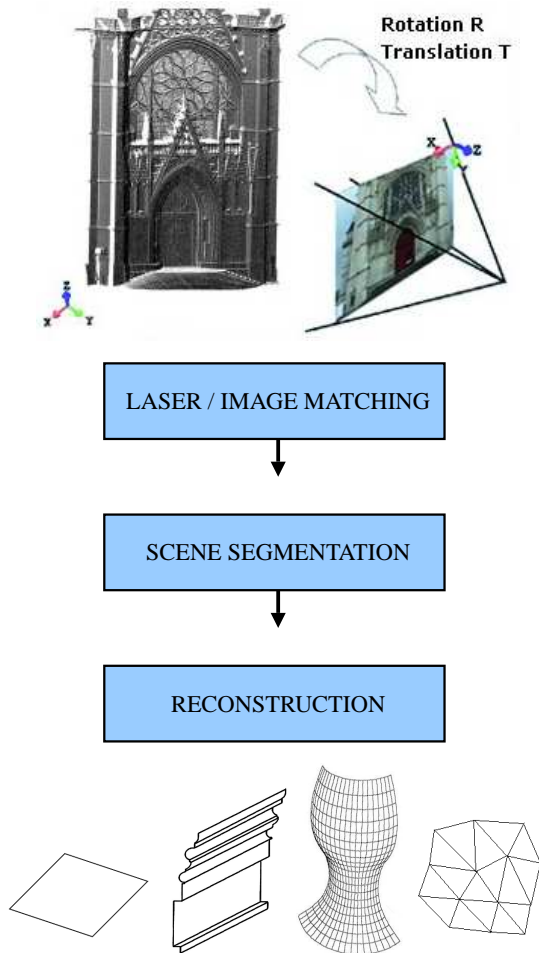


Figure 1: General strategy

## 3 LASER/IMAGE MATCHING, POSE ESTIMATION

The first step consists of estimating the relative pose between laser data and image. We are only considering the external parameters estimation, making the hypothesis, as stated before, that camera intern parameters are known and computed independently. Thus, we want to find a rotation and a translation between two frames, i.e. six independent unknowns. Most of the methods coming from the photogrammetry or the vision community use points matches to find pose between two overlapping data sets. Matching is commonly performed manually on feature points. This needs target positioning in the field and dense scanning in the laser survey case, which takes up a lot of time in the field. Here, we propose a method where we do not need any target: we rely on punctual and linear primitives found in the laser and image data. When considering laser data sets, one may notice that laser points constitute a sampling of the surface. So choosing a few points within the point cloud and matching them with image pixels will not bring a good pose estimation: we will not match the same features. But if we choose many feature matches, this effect will be absorbed by a global compensation. Furthermore, laser data has measurement noise, which will also be compensated by dense matching.

Since laser scanners perform a scan under an image topology (sweeping around two axes), an image can be recovered, where each pixel stores a 3D point or the retro-diffusion intensity of the laser beam (see Figures 2 and 3). Because the points of view between the camera and the scanner are supposed to be close, and because retro-diffusion of the laser beam is sensitive, as for any light beam, to the surface material, such images are very similar to the digital image. Thus, correlation is possible, if the two data sets are resampled at the same resolution. Here, we suppose that we can get a first approximate pose from a few points. We can then resample the laser image into the digital image grid: for each pixel of the digital image, the ray intersects the point cloud triangulation, where interpolation is made. Correlation is then carried out on a square window around the feature point. Let us remark that the laser beam wavelength lies in a narrow bandwidth (usually red or green); so the correlation may be done in the corresponding image channel (red or green).

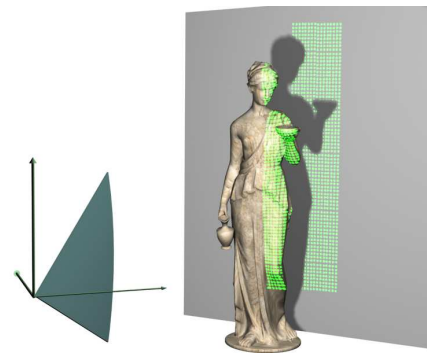


Figure 2: Scan topology

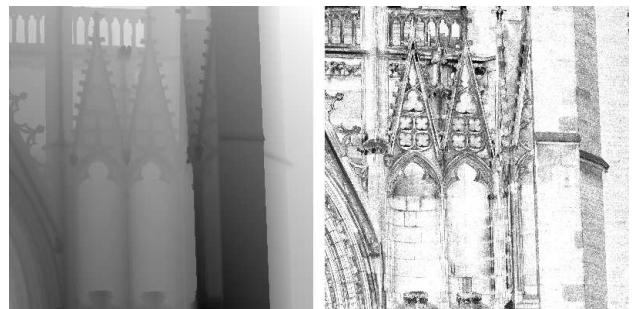


Figure 3: Range image and laser retro-diffusion image - detail

This scheme has been proposed in (Deveau and al., 2004), where more details can be found. Here, we will discuss some improvements. In our first attempts, we used to manage correlation around best feature points homogeneously spread out on the digital image. This was done by selecting the  $n$  best feature points belonging to squares of a regular grid. This approach yields a good spread of points at our disposal. But the main fact is that feature points are often located on geometric discontinuities (Gregor and Whitaker, 2001). Even if we can perform sub-pixel estimation of the point match, it will lead to an interpolation in the 3D data, and a bad approximation of the 3D point. To prevent this interpolation issue, a focalization (or focusing) of feature point detection is carried out on planar surfaces.

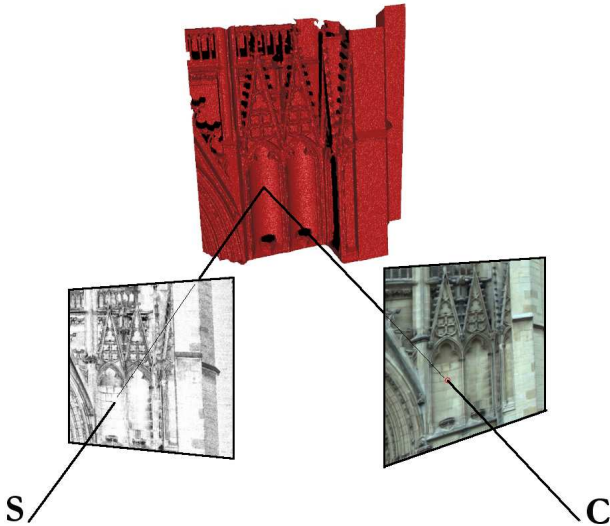


Figure 4: Resampling and correlation geometry

Planar surfaces are extracted from the point cloud and then re-projected into the image thanks to the first approximated pose estimation. Once a feature point is matched in these regions (correlation is no longer done on a square window, but on a mask corresponding to the plane), the 3D point is determined by intersection between the corresponding ray and the plane. Since the plane has been estimated on numerous points (generally more than one thousand) the localisation of the 3D point is very good. One drawback of this focusing step is the loss of the regular spread of points. This is compensated by adding measurements to the system: linear features enhance constraints on the general system.

Root mean square errors on ground control points are around 0.8 pixels. This result seems good, but could be better: this may be caused partly by residual distortions on the laser data, and an ill-conditioned system caused by many measurements in approximately the same plane - that of the facade. The good quality of pose estimation is illustrated in Figure 5, where a high resolution image is projected onto the laser mesh.

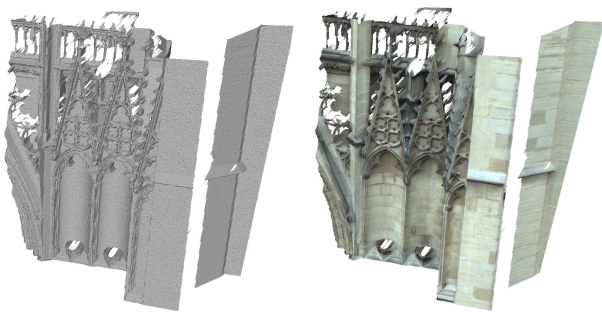


Figure 5: 3D point shaded mesh and textured mesh

## 4 SEGMENTATION

The segmentation issue has been largely studied over the last few decades (see (Haralick and Shapiro, 1985) and (Hoover and al., 1996) for a review of existing techniques). Here, we want to use both image data and 3D data to separate meaningful objects. This is done through a hierarchical segmentation scheme (Guigues and al., 2003). This approach has many advantages: it can lead to a multi-scale representation of the scene, i.e. to several levels of description for the same scene (coarse to fine); it allows fusion of geometric and radiometric energy to take advantage of both; it can supply a set of partitions from which an interactive selection of good regions can speed up efficiently the modelling.

The hierarchical segmentation is formulated through the minimization of a multi-scale separable energy which involves two terms: a goodness-of-fit term and a regularization term, the two terms being balanced by the scale parameter. A hierarchy is built from an initial partition, merging regions according to the global energy minimization criterion.

In our case, initial segmentation is derived from a watershed algorithm applied to the image. So the region boundaries lie on the image edges, and are very well localized. To reduce computation complexity, the watershed basins are merged according to a previous geometric segmentation of the range image into continuous regions, thanks to region growing. The function used to encode the energy is for the moment radiometric, but should take advantage of a geometric factor.

Even if we try this way to get an automatic segmentation without any interaction needed, this multi-scale representation can ease interactive focalization on significant objects. The operator can be proposed to move into the hierarchy, choose the right level, and then catch the region(s) corresponding to the object.

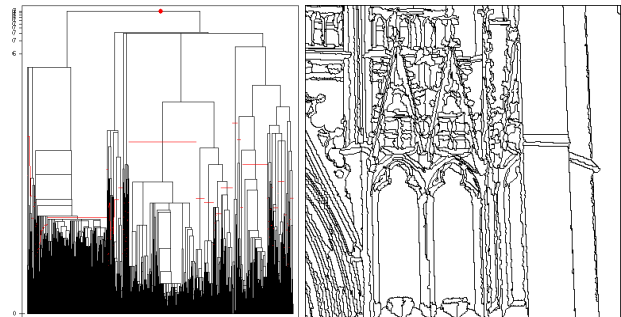


Figure 6: Hierarchy and one possible segmentation

## 5 RECONSTRUCTION

The reconstruction step is always split into two stages: focalisation or focusing, and primitive estimation. Focusing returns the region in which the object lies. In a segmentation process, this region would be given automatically. In a semi-automatic process, the region can come from a polygonal capture, or from region growing, the seed being retrieved manually. In this context, one picking in the image returns instantaneously the primitive. In the following paragraphs, we present reconstruction algorithms from the semi-automatic point of view, in the knowledge that the focusing stage could be replaced in a segmentation process.

### 5.1 Plane reconstruction

Let us consider a point set belonging to a plane. Its spatial resolution is lower than that of the image. The high resolution of the image is recovered on the region boundaries by the point set projection into the image watershed. Each watershed basin where a point is projected may belong to the good region. This implies a good correspondence between watershed partitioning and geometric discontinuities. Once the region boundaries are known in



the image, they are reprojected onto the plane estimated on the point set.

### 5.2 Surface of revolution reconstruction

Here again, let us work with a point set belonging to a surface of revolution. This kind of object is characterized by a profile which turns around its axis of revolution. First, we need to work out the axis of revolution pose. The method makes use of the property of coplanarity between the axis of revolution and each normal to the surface. Each line normal to the surface adds a contribution to a voxel space surrounding the axis location. This focus is achieved by detecting the symmetry axis in image space. This first estimation is followed by a finer estimation, where the rms error between point cloud and the reconstructed surface of revolution is used as the quality criterion. This algorithm keeps dependant axis estimation and profile reconstruction. It could be replaced by an axis estimation quality criterion regarding the morphology of the point cloud projection on a plane perpendicular to the estimated axis. The ratio between the radii of the larger included circle and the smaller including circle in relation to the projected point cloud should be minimum on the right axis.

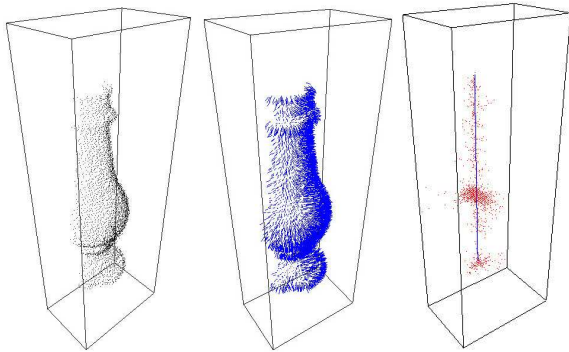


Figure 7: Point cloud (left) - normal estimate - axis of revolution estimate (right)

Given the axis of revolution, one may try to recover the object profile. The point cloud is projected onto the plane rotating around the axis of revolution. The recovery of the profile is carried out by noise reduction with the Moving Least Squares (MLS) method (Lee, 2000). The profile is then reconstructed by sampling the profile along the axis direction. This sampling may be improved by including discontinuities, which would provide a better description of the surface.

The rms errors of distances between the 3D points and the reconstructed surfaces are less than 2 millimetres, which has to be compared with the scanner measuring distance noise (2 mm).

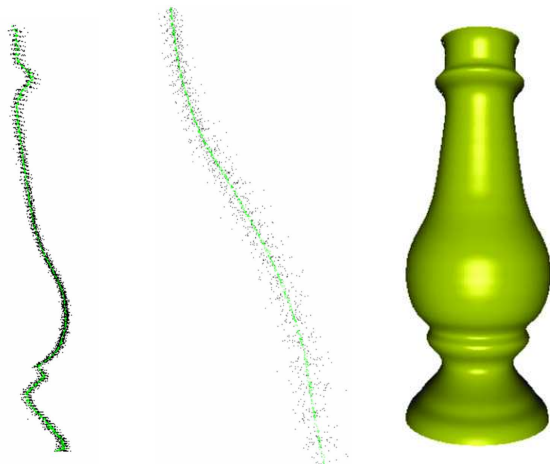


Figure 8: MLS point cloud thinning - reconstructed surface

### 5.3 Cylinder reconstruction

Cylindrical primitives are very common features in architectural scenes. They are characterized by a profile extruded along a generating line. We are considering here only cylinders with straight generating line. The cylinders' reconstruction follows mainly the surface of revolution reconstruction scheme. It differs in the way the axis is determined: we consider the Gaussian image. The Gaussian image of a cloud is the result of the Gauss map applied to the whole set of points. The Gauss map corresponds to the mapping from a point on a surface to the unit normal of the surface at this point. For a cylinder, unit normals focus in a plane perpendicular to the generating line. They belong to a great circle of the Gaussian sphere. So, the method consists in estimating robustly the plane on the unit normals, with the constraint of joining the origin. To get the profile of the cylinder, points are projected on the same plane. Then, the profile is estimated the same way as for a surface of revolution.

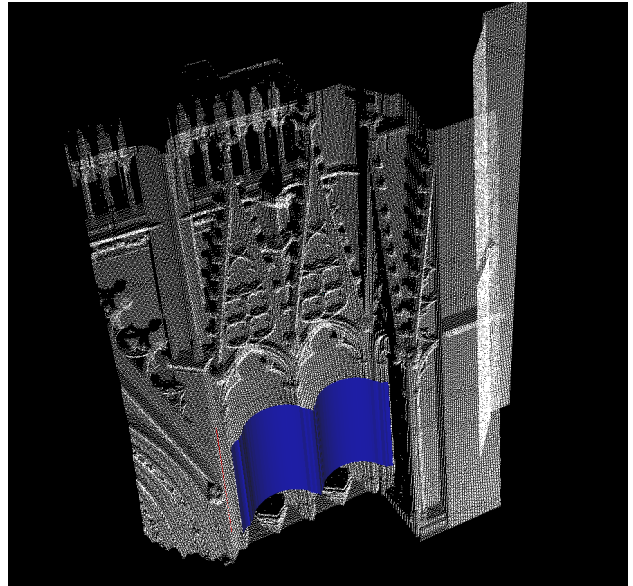
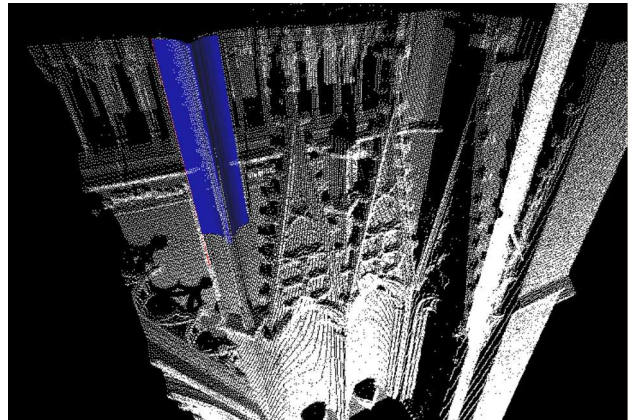


Figure 9: Cylinder reconstruction examples

## 6 CONCLUSION

We have proposed a general strategy for 3D reconstruction of large scale complex terrestrial scenes using both image and laser data. The three main steps are registration, segmentation and reconstruction. The registration step uses dense matching between punctual features and between linear features. These features can be fully automatically extracted with a good accuracy from both image and laser data. The main contribution is the proposition of a global semi-automatic modelling architecture with very few interactions needed. Those interactions concern the segmentation step, which could be replaced by an automatic segmentation. The global principle is based on the idea of building as many interactive tools as primitives describing the scene. The chosen sensing lay-out allows quick field work, without losing geometric accuracy on reconstruction. This approach may speed up the reconstruction tasks, part which should be evaluated in future work. Further work will concern adding geometric primitives to extend the current library. Improvements must also be made towards automatic segmentation.

## REFERENCES

- Dekeyser, F., Gaspard, F., De Luca, L., Florenzano, M., Chen, X., Leray, P., 2004. Cultural Heritage Recording with Laser Scanning, Computer Vision and Exploitation of Architectural Rules, ISPRS.
- Deveau, M., Pierrot-Deseilligny, M., Paparoditis, N., Chen, X. Relative Laser Scanner And Image Pose Estimation From Points And Segments, ISPRS, 2004.
- Dias, P., Sequeira, V., Vaz, F., Goncalves, J., 2001. Combining Intensity and Range Images for 3D modelling In proc. of IEEE Int. Conf. on Image Processing.
- Dick, A., Torr, P., Cipolla, R., 2004. Modelling and Interpretation of Architecture from Several Images. International Journal of Computer Vision, Vol. 60, No. 2.
- El-Hakim, S.F., Brenner, C., Roth, G., 1998. A multi-sensor approach to creating accurate virtual environments, ISPRS Journal for Photogrammetry and Remote Sensing, Vol. 53(6). pp. 379-391, 1998.
- El-Hakim, S.F., Beraldin, J.A., Picard, M., and Vettore, A., 2003. Effective 3D modelling of heritage sites, 3DIM03, pp.302-309.
- Gregor, J. and Whitaker, R.T., Indoor Scene Reconstruction from Sets of Noisy Range Images, in Graphical Models, Vol. 63, No.5, pp. 305-332, september 2001.
- Guigues, L., Le Men, H., Cocquerez, J.P., 2003. Scale-Sets Image Analysis In proc. of IEEE Int. Conf. on Image Processing.
- Han, F., Tu, Z.W. and Zhu, S. C., 2003. Range Image Segmentation by an Effective Jump-Diffusion Method, IEEE Trans. on Pattern Analysis and Machine Intelligence. Image segmentation techniques.
- Haralick, R.M. and Shapiro, L.G., 1985. Image segmentation techniques, APPLICATIONS OF ARTIFICIAL INTELLIGENCE II., vol. 548, pp. 2-9.
- Hoover, A., Jean-Baptiste, G., Jiang, X.Y., Flynn, P.J., Bunke, H., Goldgof, D.B., Bowyer, K.W., Eggert, D.W., Fitzgibbon, A.W. and Fisher, R.B., 1996. An Experimental Comparison of Range Image Segmentation Algorithms, IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. 18, No.7, pp.673-689.
- Kumar, R., and Hanson, A.R., 1994. Robust Methods for Estimating Pose and a Sensitivity Analysis, CVGIP, Vol 60, pp. 313-342.
- Lee, I-K., 2000. Curve reconstruction from unorganized points. Computer Aided Geometric Design 17(2), pp. 161-177.
- Pollefeys, M., Van Gool, L., Vergauwen, M., Cornelis, K., Verbiest, F., Tops, J., 3D Capture of Archaeology and Architecture with a Hand-Held Camera, Proc. of the ISPRS workshop on Vision Techniques for Digital Architectural and Archaeological Archives, July 1-3 2003, Ancona, Italy, The International Archive of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 5/W12, pp. 262-267, 2003.
- Pottmann, H., Leopoldseider, S., Wallner, J., Peternell, M., 2002. Recognition and reconstruction of special surfaces from point clouds Archives of The Photogrammetry, Remote sensing and Spatial Information Sciences.
- Stamos, I. and Allen, P. K., 2002. Geometry and Texture Recovery of Scenes of Large Scale Computer Vision and Image Understanding, Vol. 88 , pp. 94-118.
- van den Heuvel, F. A., 1999. A Line-photogrammetric mathematical model for the reconstruction of polyhedral objects, Videometrics VI, Vol. 3641, pp. 60-71.
- Werner, T. and Zisserman, A., 2002. Model Selection for Automated architectural Reconstruction from Multiple Views, British Machine Vision Conference.
- Yu, Y., Ferencz, A., Malik, J., 2001. Extracting Objects from Range and Radiance Images, IEEE Transactions On Visualisation And Computer Graphics, Vol. 7, No. 4.