RELATIVE LASER SCANNER AND IMAGE POSE ESTIMATION FROM POINTS AND SEGMENTS

Matthieu Deveau[a][b], Marc Pierrot-Deseilligny[a], Nicolas Paparoditis[a], Xin Chen[b]

[a] Institut Géographique National/MATIS Laboratory, 2-4 avenue Pasteur, 94165 Saint-Mandé - firstname.lastname@ign.fr
[b] MENSI S.A. 30 rue de la Fontaine du Vaisseau, 94120 Fontenay-sous-Bois - firstname.lastname@mensi.fr

KEY WORDS: scanner laser, image data, terrestrial surveying, pose estimation.

ABSTRACT:

This paper presents an approach involving linear features for pose estimation. Here we are interesting in surveys mixing image and laser scanning, for metrological applications. Since data need to be registered with the best accuracy, we are faced to a 2D-3D pose estimation problem. In most cases, scenes contain numerous segments, which are good orientation clues. We use these segments to find pose. Therefore, targets are less prevalent for location and orientation estimation purpose. This means less field operations during data acquisition. Since some scenes with very few straight lines can leave insufficient spatial constraints, we reintroduce points. We can deal with feature points to reinforce the system. Then, the algorithm simultaneously minimizes an energy function managing distances between 3D points projection in images and image points, and distances on segments ends. Precise determination of primitives in 2D and 3D data leads to fine orientation. Using subpixelar regression after an edge detection gives high-quality estimates for 2D segments. In point clouds, 3D segments come from plane intersection. We discuss relative influence of features through uncertainty assessment.

1 INTRODUCTION

1.1 Context

Our research study deals with 3D reconstruction of terrestrial scenes such as cultural heritage main buildings or industrial environments. Combining laser data with image is supposed to ease and automate surface reconstruction. We are betting on geometrical complementarity of such heterogeneous data. Hence, our goal is to reach a very precise orientation of data sets so as to perform parallel segmentation of both data. We are presenting here our work towards image pose estimation relative to a point cloud involving metric distances between segments and between points.

For many laser scanner systems nowadays available, a digital video camera is interdependent with the scanner body. Thus, calibration is carried out only once (maybe regularly, if needed). Yet, some of these cameras have got low resolution that gives poor geometrical information and poor texture for final model. Furthermore, ratio between image resolution and scan resolution depends only on scan resolution. With a free camera, one can choose higher image resolution than scan resolution, without altering scan resolution. Moreover, it allows to take pictures from different points of view, so to handle occlusions, by means of convergent photogrammetry. We are also putting ourselves in this way in a general context. Finally, this frame enables exact model overlay, with high resolution image.

1.2 Related work

Many solutions have been studied in the pose estimation framework from 2D-3D correspondences. Most of the methods coming from the photogrammetry community use point matches. There are direct solutions using three, four or six points (Wang and Jepson, 1994). More accurate results, when data present noise, come from least-square resolution with a larger points set. Even in computer vision, points correspondence remains the most common pose estimation technique (Haralick et al., 1994). In some cases, fundamental matrix estimate drives to both intern and extern parameters compute. Thus, resolution needs more points, minimum seven, eight for Hartley's algorithm (Hartley, 1997). One main problem with fundamental matrix estimate comes from interdependence between intern and extern parameters. If camera calibration is available, essential matrix estimate provides better results. Here, we are considering rigid body digital cameras which **intern parameters are known** and computed independently. Focal length, principal points and radial distortion are determined precisely by a calibration procedure on a target field.

As it has been pointed formerly in (Habib, 1999) where one can find an overview of previous works about pose estimation, man-made environments are rich in linear features. These features can be found often on planar intersections. Earlier works (Dhome et al., 1989) used three lines in the image corresponding to three ridge lines on the object. (Hanek et al., 1999) have shown that better results come from exploiting segments ends points rather than lines. Other approaches (Van den Heuvel, 1999) use geometric constraints between features. Kumar and Hanson (Kumar and Hanson, 1994) have studied two main models : one estimates distances between 3D segment ends projected on image plane and a line extracted from image ; another minimizes distances between 2D segment ends and the line built on 3D segment ends projection into image plane. Second model performs better, regarding to final solution. We have chosen this approach in this field application case, in a (photogram)metric context.

1.3 Our approach

In this framework, segments matching avoids using targets. Till now, we have been managing surveys with spheres and targets for points correspondences. To get targets' center fine position, high resolution scan is needed, operation which is time-consuming : much of field work is spent in scanning particular points which may not be useful in reconstruction...

Moreover, targets can not always be spread correctly all over the imaged overlaps and thus leads to imprecise geometrical determination. Thus identifying and matching scene invariants in the data acquired is a real trend for automating and increasing the quality of surveys through the quality of pose.

Although using segments reduces time waste, we are sometimes faced with weak configurations where few segments are present in the scene. Such scenes can leave indetermination because of several straight lines in the same plane or in the same direction. When scanning a facade, many segments lie on the same plane, and most of them are parallel (vertical or horizontal). For a gothical frontage(our example), they are mostly vertical. In other cases, there are too few segments to highlight faults. For instance, with industrial environments made of pipes, perspective avoids matching on cylinders edges. To overcome these difficulties, we have chosen to reintroduce points, but without extra field operation work (sphere or target high resolution scan). We can then choose feature points where constraints coming from segments are too weak.

2 DATA SETS

On the one side, we are dealing with a point cloud coming from one station scan. Scan is performed by constant angle ray tilting in a vertical plane, followed by rotation around vertical axis. Range measures present noise that can be reduced by multiple-shots. Points coordinates are defined in the scanner reference system.

On the other side, we have got a digital image and its optical model, in an image coordinate system centered on camera point of view. It is a conic projection where distortion is corrected. Camera devices record 5 Mpixels color image, coded on 12 bits.
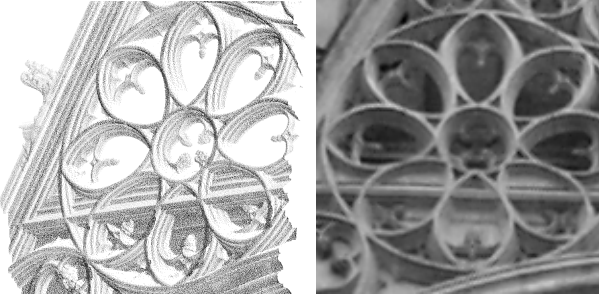

Figure 1: Point cloud, digital image (details).

3 SYSTEM FORMALIZATION

In this section, we develop the explicit system where orientation and location unknowns are considered. It is a bundle adjustment frame, where we look for translation and rotation between a terrain reference system (the scanner one) and an image system.

3.1 Distance between segments.

Let us project the ends of a 3D segment into the image plane. $\vec{p_1}$ and $\vec{p_2}$ are the image projections of the segments ends. Expressed in polar coordinates, in the image coordinate system, the straight line passing through these points is defined by :

$$x \cdot \cos \theta + y \cdot \sin \theta = \rho \qquad (1)$$

with :

$$\cos \theta = \frac{(p_2 - p_1)_y}{\|\overrightarrow{p_1 p_2}\|} \qquad (2)$$

$$\sin \theta = \frac{(p_1 - p_2)_x}{\|\overrightarrow{p_1 p_2}\|} \qquad (3)$$

$$\rho = \frac{p_{1x} p_{2y} - p_{2x} p_{1y}}{\|\overrightarrow{p_1 p_2}\|} \qquad (4)$$

The projection of a 3D point $\vec{P}$ into image is given by :

$$p_x = f \frac{\mathbf{R}(\vec{P} - \vec{T})_x}{\mathbf{R}(\vec{P} - \vec{T})_z} \quad p_y = f \frac{\mathbf{R}(\vec{P} - \vec{T})_y}{\mathbf{R}(\vec{P} - \vec{T})_z} \qquad (5)$$

where :

- $\mathbf{R}$ rotation between world system and image system
- $\vec{T}$ translation between world system and image system
- f the focal length.

Note $\vec{P_1}$, $\vec{P_2}$ the 3D segment ends in world coordinates. Considering a vector $\vec{n}$ lying into the plane which contains the 3D
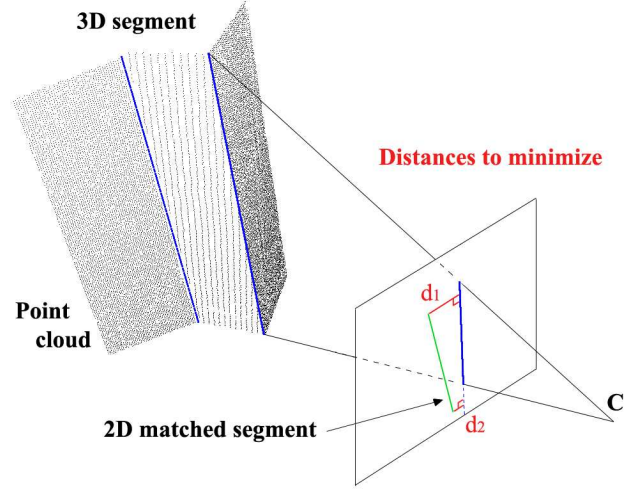

Figure 2: Distance between segments.

segment and the image projection center, and orthogonal to the image plane, such as :

$$\vec{n} = \mathbf{R}(\vec{P_2} - \vec{T}) \wedge \mathbf{R}(\vec{P_1} - \vec{T}) \qquad (6)$$

Replacing equation (5) into relations (2), (3) and (4) gives :

$$\cos \theta = \frac{\vec{n_x}}{\|\vec{n}\|} \quad \sin \theta = \frac{\vec{n_y}}{\|\vec{n}\|} \quad \rho = -f \frac{\vec{n_z}}{\|\vec{n}\|} \qquad (7)$$

A 2D segments end named $\vec{p} = (I_x, I_y, f)$ lies on the projected line ; this yields :

$$\frac{\vec{n} \cdot \vec{p}}{\|\vec{n}\|} = 0 \qquad (8)$$

Left term in equation (8) is the distance between the ends of the segment detected in images and the line supported by the 3D segment projection. This distance is associated as residual for each segment.

3.2 Distance between points

On the same scheme, one can define distance between two matched points : reaching least-square distance between the image point and the 3D point projection amounts to minimize the sum of squared distances on the X-axis and on the Y-axis.

Expected distance annulation along X-axis gives :

$$I_x - f \frac{\mathbf{R}(\vec{P} - \vec{T})_x}{\mathbf{R}(\vec{P} - \vec{T})_z} = 0 \qquad (9)$$

Along Y-axis :

$$I_y - f \frac{\mathbf{R}(\vec{P} - \vec{T})_y}{\mathbf{R}(\vec{P} - \vec{T})_z} = 0 \qquad (10)$$

3.3 Global energy function

For segments, energy function is derived from (4) :

$$E_1 = \sum_{l=1}^{N_1} \left(\frac{\vec{n_l} \cdot \vec{p_l}}{\|\vec{n_l}\|} \right)^2 \qquad (11)$$

$$E_1 = \sum_{l=1}^{N_1} \left(\frac{1}{\|\vec{n_l}\|} {}^T \mathbf{R} \vec{n_l} (\vec{P_2} - \vec{T}) \wedge (\vec{P_1} - \vec{T}) \right)^2 \qquad (12)$$

For points, corresponding energy is :

$$E_2 = \sum_{p=1}^{N_2} \left(I_x - f \frac{\mathbf{R}(\vec{P}_p - \vec{T})_x}{\mathbf{R}(\vec{P}_p - \vec{T})_z} \right)^2 + \left(I_y - f \frac{\mathbf{R}(\vec{P}_p - \vec{T})_y}{\mathbf{R}(\vec{P}_p - \vec{T})_z} \right)^2 \tag{13}$$

Finally, expression to minimize will be :

$$E = \frac{1}{\sigma_1^2} E_1 + \frac{1}{\sigma_2^2} E_2 \tag{14}$$

σ_1 and σ_2 express expected standard deviation on estimation of segments and points. They balance relative importance between points and segments during compensation.

4 SYSTEM RESOLUTION

It is quite clear that equations (8), (9) and (10) are non-linear with respect to R and T. This entails linearization , which is done with a formal calculator, and iterative estimation. At each iteration stage, global energy E in equation 14 is minimized through Gauss-Newton algorithm.

Equation (5) is applied to each 2D segment end. So, one matching gives two constraints. To solve the system for R and T, e.g. seven unknowns(we use quaternion representation for rotation), without points, four segments minimum are necessary. Of course, in most cases, four segments are not enough : they could be coplanar or/and parallel.

4.1 Approximate solution

Classically, resolution by linearization requires to find an initial estimate close to function global minimum. Approximate solution is achieved by space resection on three or more well-distributed matched points.

Visualizing clouds under an image topology The perception of objects structures and limits from point cloud is difficult and not very appropriate in a 3D viewer, even if the operator is well trained. A way of representing the 3D points acquired from one laser scanner station is an image topology. Indeed, scans are angular resolution constant. This representation has major advantages, e.g. ability of visualization of huge clouds, but the one among all is the easiness of interpretation.

Image topology is recovered from scan angular resolution : from points in Cartesian coordinates, we need to go to spherical coordinates to produce a range image. We estimate scan angular resolution looking for couples of consecutive points.

Then, we can plot points into this image, where each pixel corresponds to a 3D point. As for each scanned point, we have got on top of coordinates, retro-diffusion information (coming from the laser beam signal) and radiometry (coming from the low resolution camera), we can create clearer images (see Figure 3) and select points into these images. From geometry, we can also compute images more understandable, such as normal image, shaded range image or distance to principal plane. Small 3D details are perfectly highlighted in the range image by shading the surface. The retro-diffusion image also provides complementary very detailed information. We can then switch over different layers to choose point position.

Accuracy of points plotted into these images obviously depend on scan resolution. As system converges well from initial estimate rather far from final solution (Kumar and Hanson, 1994), the problem of finding approximate solution is not so crucial. Nevertheless, initial solution remains important for matching reasons.

Figure 3: Laser data in image topology : retro-diffusion image.

4.2 Points matching

We carry out points matches by correlation between retro-diffusion image and intensity image. Since we have got approximate solution, we can project laser points into image and compute a new image holding radiometry RGB mean value. We call it intensity image. A pixel from intensity image points at 3D coordinates. The same pixel in retro-diffusion image points at the same 3D coordinates.

Feature points extraction in intensity image is then achieved using Harris detector (Harris and Stephens, 1988). Generally, far too much corners are extracted. Since for some scenes most of the "strongest" corners are located in the same area, this scheme is refined further to ensure that in every part of the image a sufficient number of features is found. To achieve this, image is divided into a regular grid. For each area, the corner with the maximum value is selected. The number of areas can be tuned to yield the desired number of features.

Assuming that homologous point should lie near from its counterpart in intensity image, because of the fine approximate solution calculated, we look at it in retro-diffusion image in a window centered on the Harris' point. Correlation value is calculated for each window's pixel. The maximum correlation score position is considered as the matched point. To get a better estimate of this position, interpolation in the correlation window would bring matched point's subpixelar position.

Figure 4: Gradient on digital and retro-diffusion image (right).

Since radiometry variations in two images cannot be compared, because of the difference between captors responses, correlation is done in gradient images (see Figure 4). Correlation scores on 15x15 windows present various aspects (see Figure 5). Some edges and multiple peaks appear. They should be filtered to keep only unique sharp peaks.
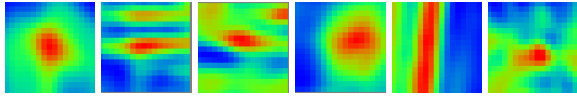

Figure 5: Correlation windows on gradient images on Harris edges.

This approach is fully automatic and leads to comparable results to interactive points matching.

4.3 Segments extraction

3D segments : Segments extraction in point cloud is fulfilled by planes intersection. Planes are extracted using region growing in the range image. Seeds are chosen by click into the retro-diffusion image overlayed by range image. They could also be chosen at random, or on a regular grid, on the same scheme as for points. Single value decomposition on the 3D points set leads to plane's parameters. For the growing process, an aggregation criterion is put on distance to the plane, which removes outliers. This threshold t is set according to the noise of measure($t = 3\sigma$). Noise on data is integrated so on by robust estimate on numerous points.

2D segments : We are using Canny-Deriche edge detector (Deriche, 1987). So, we are handling the alpha parameter and a compromise between localization and sensitivity to noise. Then, comes a hysteresis threshold, chaining and pixel chains polygonisation by split and merge algorithm. At last, to increase edge localization quality, a subpixelar estimate on the pixel chains is performed by least-square fitting.

Figure 6: Repartition of the segments.

At the moment, segments matching is interactive. Automation should be easy, since we can predict segments' position in image with a simple projection from approximate pose estimate.

5 STATISTICS

The approach presented in this paper uses rigorous least-squares adjustment at three stages : edge fitting, plane fitting and global system minimization. For each case, residuals are normalized. This representation allows error propagation and thus, assessment of the quality of pose results.

The covariance matrix on pose parameters results from propagation of the covariance matrix of the observations, which can be determined by propagation of the variance of the measures. Formalization of error propagation for linear and non-linear systems is described by Hartley and Zisserman (Hartley and Zisserman, 2000). For more details about uncertainty propagation, see (Förstner, 2004) for single view geometry and (Jung and Boldo, 2004), where the same mathematical model is studied for bundle adjustment. Therefore, error propagation has been used at three stages of our study.

5.1 Primitives consistency.

Error propagation needs prior estimation of error for observations on each primitive.

Points

- Two cases have been tested to find 2D points.

  - Points are plotted into digital images. Here, we consider that accuracy is ranging about 1 pixel.

  - Points are extracted by Harris detector. Since detection is followed by correlation, this method provides subpixelar accuracy on localization.

  If targets have been placed into the scene, their center would be recovered with a precision from 0.1 to 0.01 pixel (depending on the image quality).

- 3D points have direct relation with laser scanner measure. We are using scanner with range measure standard deviation of 6 mm. As we perform multiple-shots measures, standard deviation is reduced : $\sigma_n = \sigma_1/\sqrt{(n)}$ For instance, we use four measures for each points. This leads to $\sigma_4 = 3$ mm. In first approximation, we consider homogeneous standard deviation around point.

Segments As linear features come from least-square estimate, we can predict their respective variance from the variance of data measures.

- 2D segments are fitted on edge points detected by the Canny-Deriche operator. The line parameters are retrieved by regression (Taillandier and Deriche, 2002). The variance-covariance matrix of these parameters are estimated from variance on pixels, which depends on signal to noise ratio in image.

- 3D Segments Assuming that range measures follow a Gaussian law, the forward error propagation frame enables to calculate the variance-covariance matrix of the parameters of the normal to the plane. We may then spread variance to cross product, considering the Jacobian matrix of the application. This part of our work is still under development.

5.2 Pose quality evaluation

In this paragraph, uncertainty on location and orientation has been investigated for pose analysis.

Experiments have been performed on sets of 15 segments and 15 points. 20 random sampling/trials of n segments amongst 15 have been proceeded. We have also randomly sampled 3 and 12 points amongst 15 points. Configuration with 3 points corresponds to the minimum number of points needed for space resection.

Mean values have been computed on three outputs :

- residuals on points (see Figure 7)
- residuals on segments (see Figure 8)
- main axis of orientation error ellipsoid (see Figure 9)

We do not present location error ellipsoid parameters, since bad spatial configurations, especially occurring with few segments and points, returned wrong values. With few observations, our approximation which enables error propagation is not valid anymore.

Mean residuals on points goes down to less than 0.4 pixel. Residuals on segments are also satisfactory. There is still a weighting between primitives issue, which is not yet overcome.
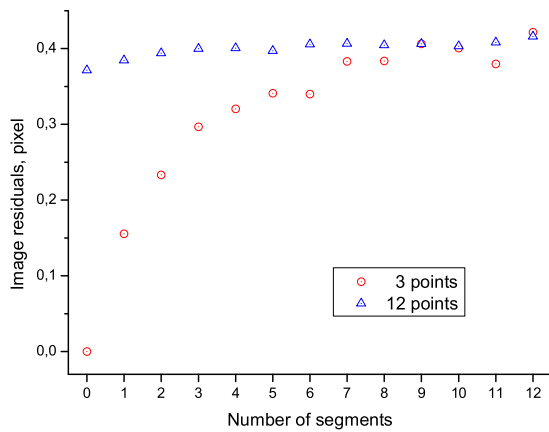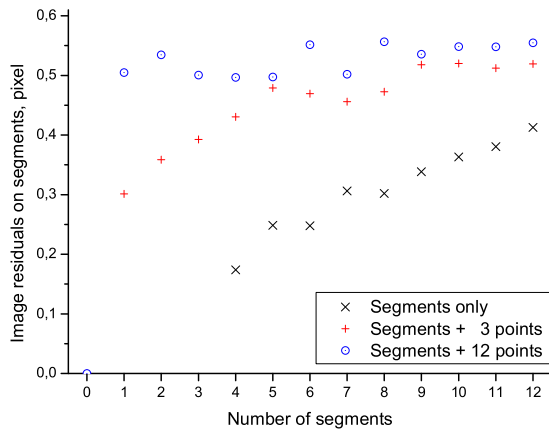

Figure 7: Image residuals on points.

Figure 8: Image residuals on segments.

First results show that using few segments (less than 6) is quite dangerous (it is a little bit obvious). Much interesting is the fact that combining 3 points with segments seems to be quite equivalent to taking 12 points with segments (even if the scale representation in y axis is good).

Some results are presented through figure (10) and (11). Here, each 3D point is drawn with its corresponding image radiometry. These 3D views highlight pose accuracy, especially on radiometric discontinuities.
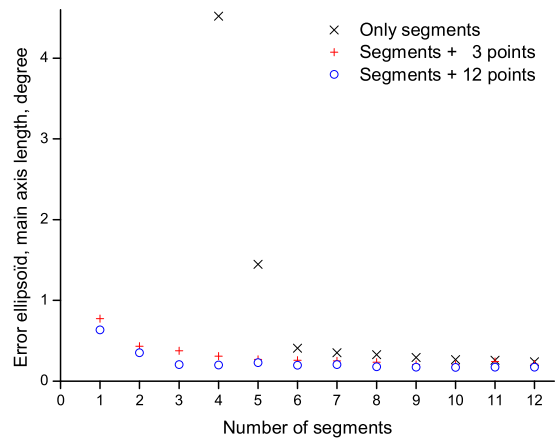

Figure 9: Relative effects of features on orientation estimate.

Figure 10: Results on the Vincennes castle chapel. Here, we show the point cloud, where each point has been assigned its image corresponding RGB pixel.
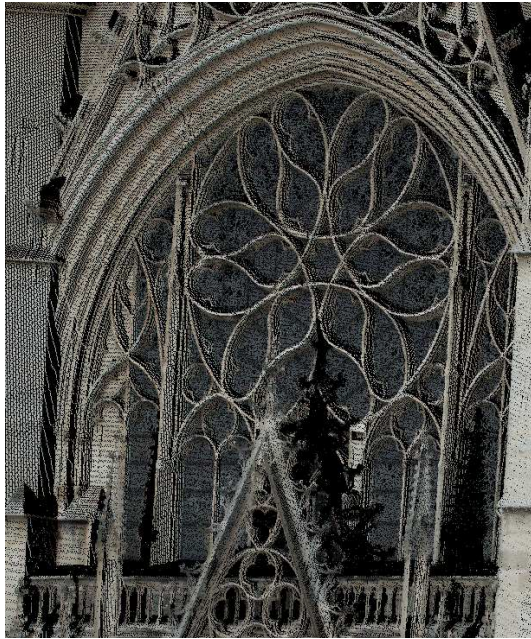
Figure 11: Results (detail).

6 CONCLUSION

We have proposed an approach for pose estimation using points and segments features for a special case : range and digital image recording. Points allow to constrain geometry, filling gaps in region of space where there are too few segments. Thus, even though segments are geometrically very constraining, to ensure a high quality of the geometrical determination, points and segments have both to be considered. It has been shown that very few observations can bring reasonable results.

Segments should be used with cautiousness. Indeed, rectilinear features are straight to a certain extent. The longer the segments the higher the risk of reality being far from straightness hypotheses. Anyway a very easy solution to this problem is to split linear features in smaller parts.

Error propagation will enable to predict if an a priori expected accuracy on pose estimation is reached. Error will be also estimated on primitives to qualify 3D reconstruction.

This approach can be easily extended to solve 3D to 3D registration, simulating image projection from station point of view. It has already been extended to aerial and terrestrial bundle adjustment (Jung and Boldo, 2004).

Further evaluations should be carried out. First of all, we should look to ground control primitives which do not play any role during compensation. Besides, we will compare more precisely results provided by correlation points. We would like also to examine distortions in scanner data.

Further improvements should be done towards automation. Approximate solution could be recovered automatically by vanishing points detection in the image (Van den Heuvel, 1998) and by finding horizontal and vertical directions in points cloud.

To improve reliability of adjustment, residuals should be compared to a threshold at each iteration. This threshold should be tuned thanks to expected precision of pose estimation, in relation with primitives variance. Through process, segments and points would be disabled or activated, depending on their residuals.

Finally, further extensions should pay attention for spatial repartition of matched primitives. Spatial repartition should take into account the fact that points and segments have not the same geometric influence on the global system.

REFERENCES

Deriche, R., 1987. Using canny's criteria to derive a recursively implemented optimal edge detector. IJCV 1(2), pp. 167–187.

Dhome, M., Richetin, M., La Preste, J. and Rives, G., 1989. Determination of the attitude of 3-d objects from a single perspective view. PAMI 11(12), pp. 1265–1278.

Förstner, W., 2004. Uncertainty and projective geometry. In: E. Bayro-Corrochano (ed.), Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics, Springer.

Habib, A., 1999. Aerial triangulation using point and linear features. In: ISPRSGIS99, pp. 137–142.

Hanek, R., Navab, N. and Appel, M., 1999. Yet another method for pose estimation: A probabilistic approach using points, lines and cylinders. In: CVPR99, pp. II: 544–550.

Haralick, R., Lee, C., Ottenberg, K. and Nolle, M., 1994. Review and analysis of solutions of the 3-point perspective pose estimation problem. IJCV 13(3), pp. 331–356.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. In: Alvey88, pp. 147–152.

Hartley, R., 1997. In defense of the eight-point algorithm. PAMI 19(6), pp. 580–593.

Hartley, R. and Zisserman, A., 2000. Multiple view geometry in computer vision. In: Cambridge.

Jung, F. and Boldo, D., 2004. Bundle adjustment and incidence of linear features on the accuracy of external calibration parameters. In: International Archives of Photogrammetry and Remote Sensing.

Kumar, R. and Hanson, A., 1994. Robust methods for estimating pose and a sensitivity analysis. CVGIP 60(3), pp. 313–342.

Taillandier, F. and Deriche, R., 2002. Reconstruction of 3D linear primitives from multiple views for urban areas modelisation. In: PCV'02, Vol. 34:3B, pp. 267–272.

Van den Heuvel, F. A., 1998. Vanishing point detection for architectural photogrammetry. In: H. Chikatsu and E. S. Editors (eds), International Archives of Photogrammetry and Remote Sensing, Vol. 32, pp. 652–659.

Van den Heuvel, F. A., 1999. A line-photogrammetric mathematical model for the reconstruction of polyhedral objects. In: Videometrics VI, Vol. 3641, pp. 60–71.

Wang, Z. and Jepson, A., 1994. A new closed-form solution for absolute orientation. In: CVPR94, pp. 129–134.