EXTRACTING 3D EDGEL HYPOTHESES FROM MULTIPLE CALIBRATED IMAGES: A STEP TOWARDS THE RECONSTRUCTION OF CURVED AND STRAIGHT OBJECT BOUNDARY LINES

Franck Jung [a], Vincent Tollu [b], Nicolas Paparoditis [a]
[a] Institut Géographique National/MATIS, 2-4 avenue Pasteur, 94165 Saint-Mandé - franck.jung@ign.fr, nicolas.paparoditis@ign.fr
[b] UTC Compiègne. 60203 - Compigne Cédex, France - vincent.tollu@etu.utc.fr

KEY WORDS: Photogrammetry, 3D Reconstruction, Curved 3D Lines, 3D Edgels, Multiple Images, Feature Matching, Robust LMS bundle adjustment.

ABSTRACT:

This paper describes a novel technique to extract $3D$ elementary linear elements from multiple views, often called edgels, that are 3D points with tangent direction. Techniques based on stereopair feature, edgel or contour, matching have been intensively developed in the 80s. The relatively poor quality of images used at the time and the ill-posed stereo context contributed to focus the research around segment matching, which provided fairly good results on objects with rectilinear boundaries. Here, we take advantage of the multiplication and of the high quality of images provided by a digital frame camera to revisit this feature matching technique and to enhance the accuracy, the detection and the robustness of a 3D elementary edgel feature estimator. 3D edgels introduce the lowest implicit modelisation as possible, thus allowing the characterisation and reconstruction of curved and straight linear structures which are widespread in our man-made world. The 3D points with tangent directions generated by our estimator can be directly injected in a surface reconstruction framework based on tensor voting (Tang and Medioni, 1998). In a more practical and photogrammetric context, e.g. HR aerial images of urban landscapes, these low-level features can be used to build 3D lines that can be injected as constraint lines in addition to other features, such as 3D points and 3D segments, in a triangulation process (Paparoditis et al., 2001) to improve the surface reconstruction of manmade objects (buildings, pavements, roads, etc.). These features can also be used to construct higher level features such as planar facets in aerial images of urban areas for building reconstruction.

1 INTRODUCTION

1.1 Context

Urban areas represent challenging areas for 3D reconstruction. The main goal of many approaches is to reconstruct buildings using various prior knowledge (Model Based Reconstruction, 2D boundaries hypothesis, segments matching (Taillandier and Deriche, 2002), etc.). Here, we take advantage of the multiplication and of the high quality of images (SNR=300) provided by a digital frame camera (Thom and Souchon, 2001) to extract elementary 3D points with a local direction called edgels. These edgels introduce the lowest implicit modelisation as possible, allowing the characterisation and reconstruction of linear structures which are widespread in our man-made world (buildings, roads, complex crossroads,etc.). Hence, edgels can be used as accurate 3D elementary descriptors in aerial scene interpretation, architectural surveys and many other vision applications. We perform a neat evaluation of the reconstruction power of edgels in terms of planimetric accuracy, altimetric accuracy, angular error and false negatives.

1.2 State of the Art

Model-Based Building reconstruction has a long history (Weidner, 1996), (Gülch, 1997), (Fuchs, 2001). A matching process is generally used between a model and extracted features. Within this context, line segment extraction was intensively studied ((Medioni and Nevatia, 1985), (Zhang and Faugeras, 1992), etc.). The use of multiple images supplied reliable results ((Schmid and Zisserman, 1997), (Taillandier and Deriche, 2002)), however many structures cannot be described with large line segments (Roads, Complex Buildings, etc.). In order to get a reconstruction of these complex structures, energetic approaches are often proposed (snakes, level sets, ...). These approaches are generally very sensitive to local minima, hence a good initialization is often required. We try to avoid these difficulties by providing local 3D descriptors of the scene taking advantage of image multiplicity. We start with an image contour extraction.

2 EXTRACTING 2D EDGELS FROM IMAGES

Contours within each image of the set are detected by applying the Canny-Deriche edge detector and by extracting the local maxima in the gradient images. The local maxima are then located with sub-pixel accuracy by determining the peak value of the parabola that fits locally the gradient samples in the gradient direction. The sub-pixel contour points are thereafter chained and the 2D tangent direction for each contour point is calculated by fitting a straight line on the few closest chain samples on each side of it. This estimation can be sensitive to noise, change of direction, etc. thus, in order to get a robust estimation of the associated direction, we use an M-estimator (Rousseeuw and Leroy, 1987). We thus obtain chains of edgels: points with tangent direction. We now want to match each 2D edgel with the $N-1$ other images.

3 IDENTIFYING EDGEL-MATCHING HYPOTHESES IN SLAVE IMAGES THROUGH OBJECT SPACE

First we choose a master image. For each edgel point $\tilde{p}$ in this master image, we determine the corresponding bundle in object space (Figure 1). Having previously corrected the chains of contours of the image distortion, the epipolar curves i.e. the projection of the master bundle in all slave images, are straight lines. Thus the set of homologous 2D edgels hypotheses englobes all the 2D edgel segments intersecting the epipolar lines. A quadtree search is performed to reduce efficiently the combinatory of the hypotheses determination. The 3D object space edgel point P corresponding to $\tilde{p}$ is determined by an adjustment of all the bundles corresponding to all the set of hypotheses as explained in the following paragraph.

4 ROBUST 3D EDGEL ESTIMATION (POINT LOCATION WITH TANGENT DIRECTION)

The first step of our adjustment aims at finding a point matching algorithm related to the master point $\tilde{p}$. During this step, we use

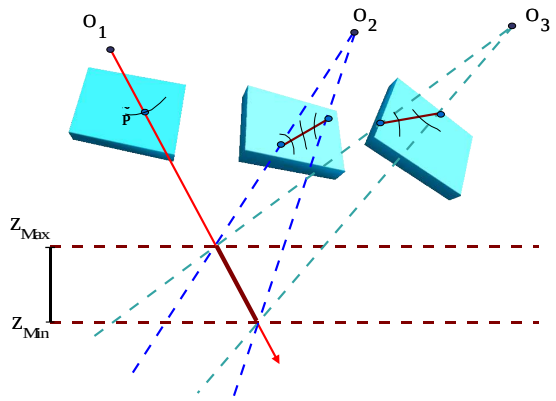

Figure 1: The master point $\tilde{p}$ is reprojected in every slave image using an epipolar constraint and a bound on the altitude $z \in [Z_{min}, Z_{max}]$. Image i contains a set S_i of possible matches with $\tilde{p}$.

a Least Median of Squares (LMS) technique in order to select the best cluster of bundles close to the master bundle. The matching is described by the master point $\tilde{p}$, the set of associated slave points and a 3D location P . Each slave image contains at most one slave point. It is important to notice that we do not systematically find a matching for each slave image. First, we reject a matching candidate as soon as the distance between the bundle and P is larger than a given threshold. This threshold is only a function of the various $\frac{B}{H}$ parameters. Second, the whole matching is rejected as soon as it contains a poor number of points. The lower bound on the number of points depends on the robustness required by the application (in our application, we kept matchings containing at least four points).

Let's suppose that $\tilde{p}$ belongs to image 1. The epipolar constraint gives us a set of possible matches in each slave image. Let S_i be the set of possible matches in image i. We define, for $i \in [2, N]$:

$$S_i^\star \quad = \quad S_i \cup \emptyset \tag{1}$$

We want to find a set $L \in \{\tilde{p}\} \times S_1^\star \times \cdots \times S_N^\star$ containing at least 3 points. We define $P(L)$ as the point in the object space that minimizes the distance between each bundle and $(\epsilon_1(L), \cdots, \epsilon_k(L))$, which is the ordered set of distances of P to each bundle generated by an element belonging to L. We select $\mathcal{L}$

$$\mathcal{L} \quad = \quad \arg\min_L \epsilon_{(p)} \tag{2}$$

Practically, we minimize the third residual ($p = 3$). Bundles having a distance larger than the altimetric resolution (in our case $.5m$) are automatically rejected. Our evaluation procedure (see next section) proved that this threshold was meaningful. We used a greedy strategy to compute L, that is each pair $\tilde{p} \times p_1$ with $p_1 \in S_1^\star \cup \cdots \cup S_N^\star$ is used to get an LMS candidate.

After the LMS location adjustment, we suppose that the remaining measures follow a gaussian error. Hence, it is now possible to refine the 3D location by a least square technique. After this step, the importance of the master bundle is reduced and we get a precise P associated to our matching.

To estimate a 3D tangent direction associated to P, we tailored a direction adjustment. The aim is to find a 3D direction as close as possible to every plane formed by a matching point belonging to $\mathcal{L}$, the associated 2D direction and the projection center of the associated camera. Intersecting all these planes two by two gives us a set of hypothesis for the 3D direction (Figure 2). We first use a LMS technique to get a robust estimation of the direction. We reject all measures that are to far from our LMS solution. The threshold should only depend on the accuracy of the 2D edgels and on the geometry of the different point locations. A threshold of 10° was fixed in our applications (cf section evaluation).

Then a least square technique is used with the directions close to the LMS solution.

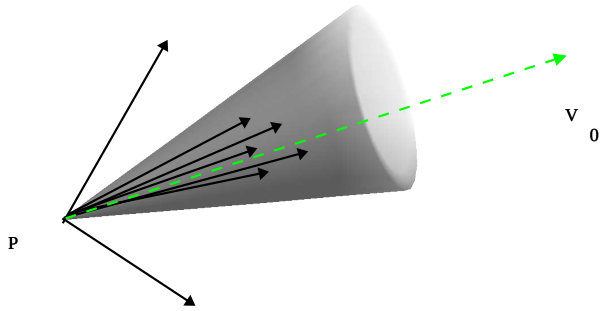

Figure 2: Direction Adjustment. The LMS step aims at eliminating outliers. The second step consists of a L_2 etimation in order to refine the direction.

We solved the following minimisation problem for the vector $\vec{u}$:

$$\begin{cases} \arg\min_{\vec{u}} \sum_{i=1}^{n} \vec{u_i} \wedge \vec{u} \\ ||\vec{u}|| = 1 \end{cases} \tag{3}$$

The aim of this optimization problem is to find a vector $\vec{u}$ that minimizes the sum of the squared sinus between the solution vector and every $\vec{u_i}$. We can notice that for small angles, this optimization problem should roughly minimize the sum of the squared angles. We define $M_{\vec{u_i}}$ the 3×3 matrix verifying $\vec{u_i} \wedge v = M_{\vec{u_i}} \cdot v$ for every 3D vector v. Hence, solving (3) is equivalent to solving the following linear constraint problem :

$$\begin{cases} AX = 0 \\ ||X|| = 1 \end{cases} \tag{4}$$

with $A^t = [M_{\vec{u_1}}^t, \cdots, M_{\vec{u_i}}^t]$. Therefore, we will have to extract the eigenvector corresponding to the smallest eigenvalue of $A^t A$. We can point out that it is even possible to find the best direction by an M-estimator for this linear problem (we reweight the set of equations 3 by 3). However, after several evaluations, the LMS followed by an L_2 optimization seems to give the best results.

We have separated the location adjustment from the directional adjustment. Because the location residuals are metric, while the directional residuals are angular, it is hard to compare these quantities. Besides, considering simultaneously position and direction estimation stays challenging due to the large number of potential matches to compare with a robust technique.

As soon as all master image edge locations are processed, we repeat the very same algorithm choosing a master image from the remaining set of images. This enables the detection of 3D edgels not significantly detected in the previously processed master images. Moreover, seeing that the object sampling changes in each view, repeating the process master images and accumulating the 3D edgels allows a super resolution object characterization.

5 USING A DSM TO REDUCE MATCHING AMBIGUITIES AND COMPUTATIONAL COST

Many applications take advantage of the use of robust external surface data : robust DSM computed by correlation techniques (Paparoditis et al., 2001), laser scanning surface, etc. Using a photogrammetrically computed or a LASER DSM in our process reduces drastically the matching search-space combinatory and consequently the matching ambiguities and erroneous edgel features. A comparison of edgel extraction using a DSM (provided by an a posteriori merging of all elementary DSM calculated on all possible stereopairs) or a LASER DSM (density of 4 to 8 points per square meter, TOPOSYS ®) has been performed. The results are quite identical due to a similar geometric accuracy of both surface models we used.

Having removed most of the erroneous matches, it is possible to associate a geometric uncertainty to each reconstructed feature using Förstner and Heuel's technique of uncertainty propagation (Heuel and Förstner, 2001). This enables a thresholdless grouping and chaining of our low-level 3D features to reconstruct higher-level features and objects.

6 RESULTS

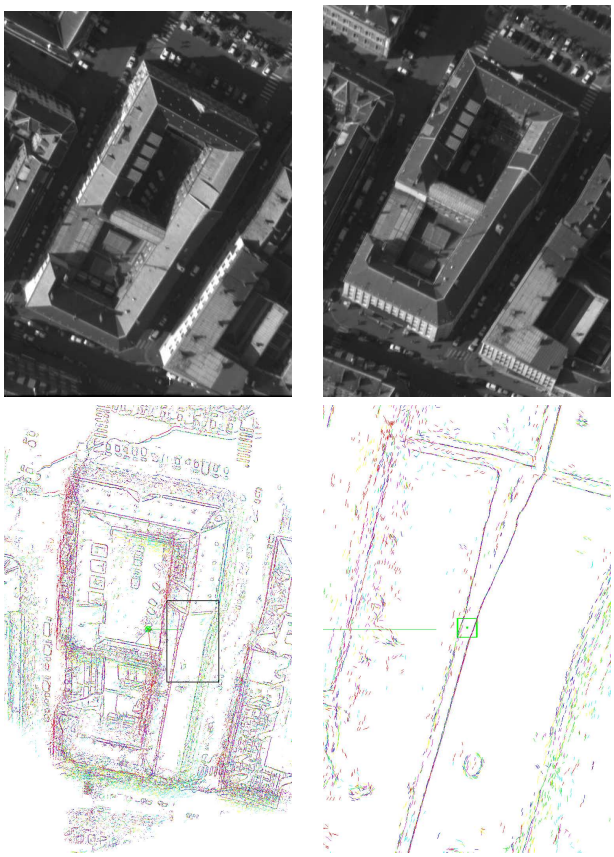

Figure 3: Example of 3D edgel extraction. For this example, we used a set of 9 panchromatic aerial images. Upper left and right image show a subset of the 9 images used for edgels extraction. Bottom left figure shows an edgel extraction for the entire scene. Bottom right figure shows planimetric accuracy of the detailed 3D edgels. The square centered in the bottom right image is used as $1m \times 1m$ metric reference.

In our experiments, we used a set of 9 panchromatic aerial images (20 cm pixel size; 3 strips of 3 images ; 60% along-track overlap; 60% across-track overlap; $\frac{B}{H}$ values varying between 0.35 and 0.85) acquired with IGN's digital frame camera over the french city of Amiens. The results show that this technique is very promising (Figure 3). The edgel features are very precise: planimetric errors are around 2 or 3 cm, altimetric errors are around 10 cm, and angular error are around $5°$. Moreover and above all, the morphology of curved and small objects boundaries is very well preserved and rendered (Figure 4, Figure 5). In order to corroborate the geometric accuracy of edgels we estimated visually on some examples, we compared our extracted features to a model.

Figure 4: Example of non-linear object extraction with edgels. The shape of small objects (cars, zebra crossing, etc.) are well reconstructed using edgels. The square (upper left and bottom left images) is used as a $1m \times 1m$ metric reference.

We compared our feature extraction results on the nine-shaped building with a 3D CAD model manually stereoplotted from analogue aerial photographs at a scale of $1 : 6000$ (see Figure 6) where the altimetric (resp. planimetric) precision is 15 cm (resp. 20 cm). The model is composed of 241 segments. We have removed all segments describing the ground footprints and all vertical lines to retain only 92 segments decribing roof tops. One should remark that the localisation quality of this reference model is not precise enough (the angular precision is precise enough) to evaluate the pure precision performance of our technique.

Our evaluation protocol is very simple. We retain for each segment all the edgels that are within a selective distance of 0.5 meters and which have an angular error beneath 30 degrees. All remaining edgels (with a length of 0.50 meter) are projected onto the segment to estimate the global coverage and the length of the largest coverage gap.

The mean value of planimetric (resp. altimetric) localisation errors is 0.2 (resp. -0.05) meters and the RMS is 0.13 (resp. 0.22) meters, estimated on a set of 10506 edgels. A parametric analysis has shown that localisation and angular errors (histogram shown on Figure 7) do not depend on the length nor on the slope of the building features. Most of the building ridges are perfectly reconstructed and the dispersion of edgels around a main axis does not exceed 0.03 meters and 2 degrees. All 5 dormer windows have been correctly reconstructed and the shapes are rendered in a very satisfactory way. The horizontal gutter lines are not as well

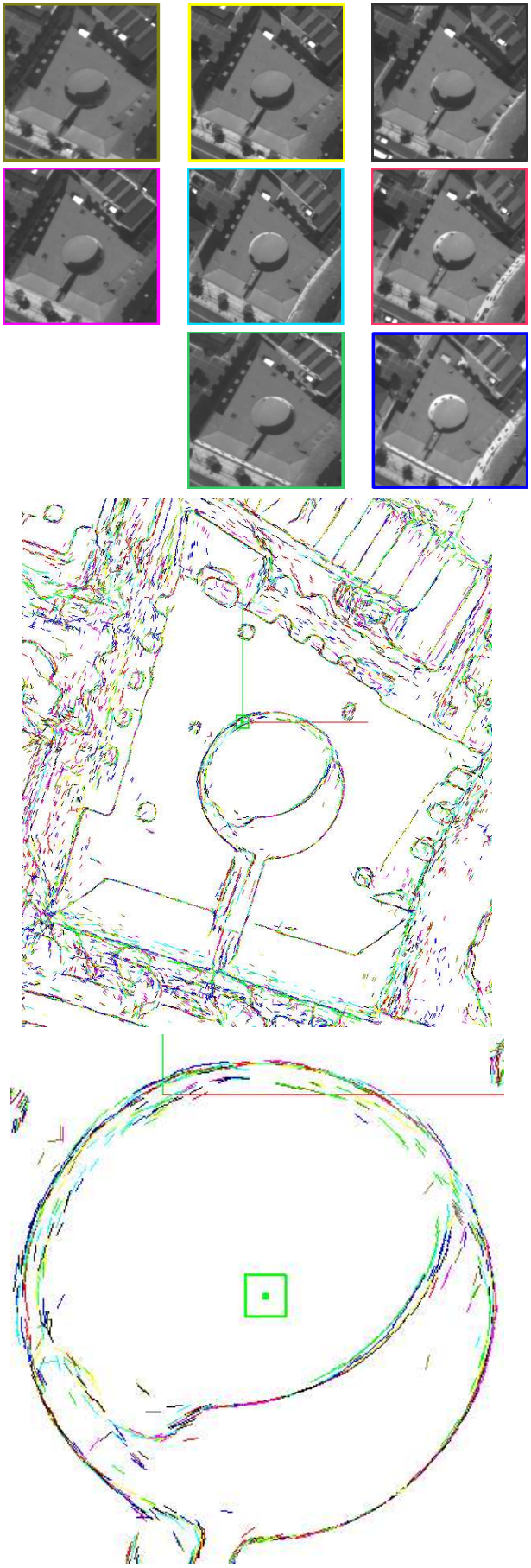

Figure 5: 3D edgel extraction. The shape of the object is not necessary composed by linear structures. Hence, curved structures can be accurately described by edgels. To each color is associated a master image used for edgels computation.

characterised and cleanly reconstructed as the ridge lines due to noisier edges extractions within the images.

Merging all edgels generated from every master image is justified seing that it increases the completion of our reconstruction as shown in Figure 8. Indeed, our elementary master image process will only reconstruct edges that are characterised in the master image. Due to the lighting conditions and viewing angles, some features are not sufficiently contrasted. Merging all elementary master image results will enhance the probability of having all interesting building features reconstructed. Concerning the analysis of each individual master image process, best results are obtained for master images where the building is closest to the image nadirs. Indeed, contours corresponding to building gutters are more regular and cleaner i.e. the texture of the facade is not or nearly not visible.

Concerning the impact of the number of images, our experiments have shown that there are no real improvements over 5 images especially for the roof top features, which are not or very partially hidden within all images.

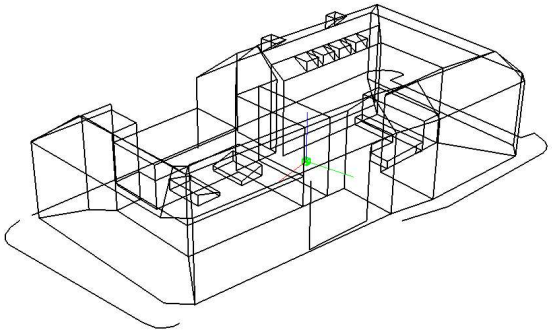

Figure 6: 3D CAD model used for our evaluations. Comparison between edgels and a model is a tricky task due to the generalisation of the CAD model.

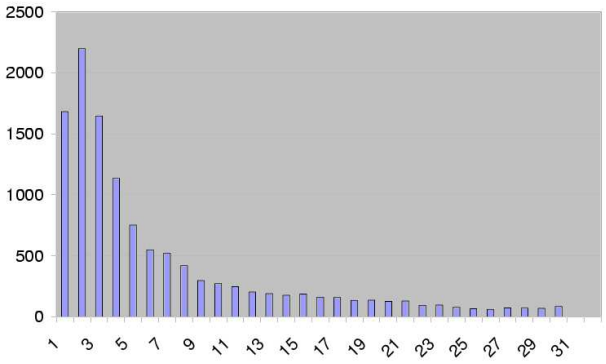

Figure 7: Angular dispersion of edgels. Standard deviation stays between 4-5 degrees.

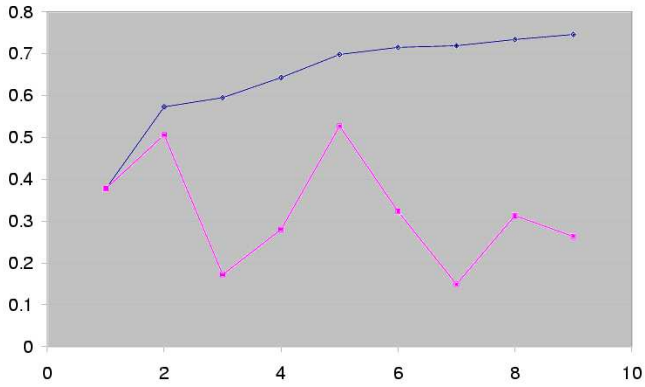

Figure 8: Measurement of the 3D edgel overlap with the CAD model. The upper curve describes the mean completion (percentage of overlap) of edgels as a function of the number of master images. The lower curve describes the mean completion (percentage of overlap) for each master images individually.

7 FUTURE WORK AND CONCLUSION

7.1 CONCLUSION

This paper describes a 3D edgel extraction algorithm. We have shown how to use the multiplicity of images in order to get accurate 3D elementary description (3D location+ 3D direction). The description of a scene by edgels is a promising area of investigation. The geometric accuracy of location and direction of edgels as well as the good completion let us hope to tackle the building reconstruction problem using these features. Even if the current results can already be used in several applications, we can still improve the accuracy and the reliability of 3D edgels using a better estimation step,i.e. keeping for each master point $\tilde{p}$ several 3D locations and then select one of them after the direction adjustment. A neat evaluation in terms of angular accuracy, planimetric accuracy, altimetric accuracy and false negatives has been performed and is used in order to fix our set of parameters. Let us give some possible extensions of our work.

7.2 Application and Future Work

Grouping It is now possible to chain our 3D edgels. We can either use directly a 3D primitive grouping strategy in the object space or take advantage of the 2D topology of the chained contours in each master image. It seems promising to combine 3D edgels with elementary surface estimation (using a DSM).

Object Recognition It is even possible to use these edgels as an input for an object detector. Many pattern recognition algorithms (neural networks, decision trees, etc.) just need a "recoded" input image in order to take a decision. We are currently testing the recognition of planar patches using 3D edgels.

Bundle Adjustment We consider that the edgels can be used in order to estimate a refinement of the calibration parameters of the camera. As soon as we can estimate matches of 2D line elements into multiple images, we will be able to use these matches inside a bundle adjustment algorithm (Triggs et al., 2000) in order to improve the accuracy of the intrinsec and extrinsic orientation parameters.

REFERENCES

Fuchs, F., 2001. Building reconstruction in urban environment : a graph-based approach. In *3rd International Workshop on Automatic Extraction of Man-made Objects from Aerial and Space Images (Ascona).*

Gülch, E., 1997. Application of semi-automatic building acquisition. In *Automatic Extraction of Man Made Objects from Aerial ans Space Images (II)*, pages 129–138, Ascona, Birkhauser, Basel.

Heuel, S. and Förstner, W., 2001. Matching, reconstructing and grouping 3d lines from multiple views using uncertain projective geometry. In *CVPR '01*. IEEE. accepted for publication.

Medioni, G. and Nevatia, R., 1985. Segment-based stereo matching. *CVGIP*.

Paparoditis, N., Maillet, G., Taillandier, F., Jibrini, H., Jung, F., Guigues, L., and Boldo, D., 2001. Multi-image 3D feature and DSM extraction for change detection and building reconstruction. In Verlag, B. B., editor, *Automatic Extraction of Man-Made objects from aerial and space images*, Ascona.

Rousseeuw, P. and Leroy, A., 1987. *Robust Regression and Outlier Detection.* John Wiley & Sons, New-york.

Schmid, C. and Zisserman, A., 1997. Automatic line matching across views. In *CVPR*, pages 666–671.

Taillandier, F. and Deriche, R., 2002. Reconstruction of 3d linear primitives from multiple views for urban areas modelisation. In *Proceedings PCV02, Vol B, to appear.*

Tang, C.-K. and Medioni, G. G., 1998. Integrated surface, curve and junction inference from sparse 3-d data sets. In *ICCV*, pages 818–826.

Thom, C. and Souchon, J. P., 2001. Multi-head digital camera systems. *GIM International*, 15(5):34–37.

Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A., 2000. Bundle adjustment – A modern synthesis. In Triggs, W., Zisserman, A., and Szeliski, R., editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag.

Weidner, U., 1996. An approach to building extraction from digital surface models. In *U. Weidner, An Approach to Building Extraction from Digital Surface Models, in Proceedings of the 18th ISPRS Congress, Comm. III, WG 2, Vienna, Austria, 1996, pp. 924-929. 43 Building Detection from a Single Image.*

Zhang, Z. and Faugeras, O., 1992. *3D scene analysis: a stereo based approach.* Springer.