TOWARDS A SUB-DECIMETRIC GEOREFERENCING OF GROUND-BASED MOBILE MAPPING SYSTEMS IN URBAN AREAS: MATCHING GROUND-BASED AND AERIAL-BASED IMAGERY USING ROADMARKS

O. Tournaire[1,2], B. Soheilian[1,2], N. Paparoditis[1]

1. Institut Géographique National, Laboratoire MATIS – 4, avenue Pasteur, 94165 Saint-Mandé Cedex France
2. Université de Marne-La-Vallée, Laboratoire Géomatériaux et Géologie de l'Ingénieur – 5, Bd Descartes, Champs sur Marne – 77454 Marne la Vallée France
(olivier.tournaire, bahman.soheilian, nicolas.paparoditis)@ign.fr

Commission I, ICWG V/I - Integrated Systems for Mobile Mapping

KEY WORDS: Mobile Mapping, Autonomous navigation, Georeferencing, Pose estimation, Automation

ABSTRACT:

We propose a novel image-based strategy for a sub-decimetre quality georeferencing of Mobile Mapping Systems in urban areas. The presented approach uses very high resolution multiple-view calibrated aerial images to extract a database of 3D invariant objects, namely road marks. These objects are also extracted in 3D in the ground-based imagery and are matched with the aerial data source. These road features can then be used as Ground Control Objects for a photogrammetric resection to finely register the images of a terrestrial mobile mapping system. All the results in this paper will be shown on a "zebra crossing" road mark. Nevertheless, the process we describe could be applied to any other road mark.

1. INTRODUCTION

Most of the autonomous navigation systems are based on a real-time extraction of visual landmarks and on a real time matching of these features with the ones collected from a first acquisition with an image-based mobile mapping system. The accuracy and the reliability of the navigation will directly depend on the quality of the data acquired in the first place.

The pose estimation of mobile mapping systems is most of the time based on direct georeferencing devices. Even though absolute decimetre accuracies are achievable in open areas, in dense urban areas, GPS masks and multiple paths do corrupt the quality of the measurements. Even though Inertial Navigation Systems (INS) can help filtering errors and interpolating between GPS interruptions, intrinsic drifts of INS will soon cumulate and lead to metric accuracies which is insufficient for many applications such as generation of databases for autonomous navigation

The images acquired by MMS systems have already been used to improve the relative georeferencing of MMS for example by determining a better relative orientation between intra-date and inter-date images thanks to tie points (Chaplin and Chapman 1998) (Bayoud *et al*, 2004) or tie segments (Bauer *et al*, 2002). Nevertheless if the GPS interruptions are long, even with the help of tie points, systematic errors of camera calibration or tie point extraction pile up and can become important. Very few attempts have been made to use external sources of data to anchor (by matching) the vehicle trajectory. As an example, an aerial-based Digital Surface Model has been used to register a ground-based Lidar-acquired Digital Surface Model with a Monte-Carlo localization technique (Frueh and Zakhor, 2003).

2. OUR MMS AND OUR GEOREFERENCING STRATEGY

Many authors have investigated mobile mapping devices for the automatic 3D modelling of large scale city models In these papers, the 3D geometry is most of the time obtained using a laser scanner mounted on the vehicle (Zhao and Shibasaki 2001).

Figure 1. The Stereopolis Mobile Mapping System

Our system, Stereopolis, is mainly image-based. It is a mobile mapping system developed within the IGN's MATIS and LOEMI laboratories for automated acquisition of georeferenced terrestrial images in urban cities. As shown in figure 1, the platform is equipped with two couples of 4000 by 4000 CCD cameras and georeferencing devices. The system provides goods imaging geometry and coverage of object space.

The road-looking cameras form a horizontal stereoscopic baseline (1.5 m) allowing the stereo-plotting of urban features (e.g. lampposts, traffic lights, road marks and signs, trees, etc.)

and a short stereo vertical baseline on each side of the vehicle (1 m) to survey the façades (only one vertical baseline was mounted in the experiment of figure 2). The short vertical stereoscopic baselines are divergent to enhance the vertical field of view covered by the two cameras. With 28 mm focal length lenses, the field covers a six storey building at a distance of 6 m.

Figure 2. Two stereopairs simultaneously acquired by Stereopolis. (Up): the façade left-looking stereopair, (Bottom): the road-looking stereopair.

Our aim with our MMS is to provide very high quality georeferencing providing 3D measurements on objects with an absolute positioning accuracy of less than one decimetre thus compliant for autonomous navigation applications but also for a "clean" merging of aerial-based and ground based reconstructions for the generation of complete 3D city models as in (Frueh *et al*. 2005).

Our idea is to inject in terrestrial mobile mapping georeferencing an external source of data, i.e. fine automated GCOs (Ground Control Objects) extracted from high resolution aerial images (<20 cm GSD) to anchor the absolute localisation of the ground-based acquisition system. The general strategy we will develop in the paper is to extract and reconstruct 3D road marks from multiple digital aerial images (zebra crossings, discontinuous white lines, etc). These features will be matched with the same road marks extracted from the ground-based stereopairs in order to achieve an off-line decimetre resolution registration of the mobile mapping platform (see figure 3). Indeed, road marks are extremely good features because of their invariant shapes and very constrained specifications, making their extraction a quite easy pattern recognition problem.

The detection and the coarse 3D reconstruction of the zebra crossing will be performed from the calibrated aerial images (Section 3). This step will provide focusing areas to limit the space of research in the ground-based imagery. The detailed 3D shape and surface of the road marks will be reconstructed from the ground-based images of our road-looking MMS stereo-rig (Section 4). We then present our strategy for road mark matching between aerial and terrestrial images and present some results on real data (Section 5) and on simulations (Section 6). This matching process provides a highly georeferenced road mark that is used to estimate the pose of the ground-based images (Section 7).
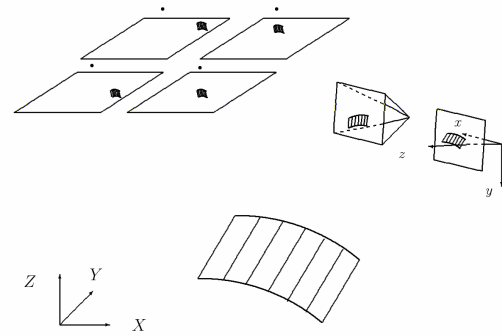

Figure 3. Aerial images provide road mark ROIs; road marks are then reconstructed in 3D from the ground-based stereopairs and finely repositioned with respect to the aerial images; finally these road marks are used as Ground Control Objects for estimating the pose of the ground-based images.

In the following of the paper, all the results will be shown on the same running example and on a "zebra crossing" road mark. Nevertheless, the process described in this paper could be applied to any other road mark.

3. ZEBRA CROSSING DETECTION AND PARTIAL RECONSTRUCTION FROM MULTIPLE CALIBRATED AERIAL IMAGES

The algorithm developed for automatic reconstruction of zebra crossings from aerial imagery takes benefit of external geometric knowledge. Indeed, road-marks are governed by careful specifications (a fixed stripe width of 0.5 m and a minimal length of 2.5 m; a regular inter stripe distance varying from 0.5 m to 0.8 m).

As described in (Tournaire *et al*. 2006) our strategy relies on the matching of 2D segment lines features. Indeed in our aerial images the GSD is such that these objects can be characterised and reconstructed using these features. We use the Canny edge detector (Canny, 1986) to extract contour points. Several classical steps are performed to obtain 2D line segments: hysteresis threshold, subpixel contour points localization, chaining and straight line fitting.

In order to detect 2D segments belonging to zebra crossings, we use specifications and analyse their relative organisation. First, segments are filtered on their lengths, and we then search for groups of parallel segments respecting stripes width and distance between stripes. A length homogeneity criterion is also computed. This process is repeated independently in each image.

To reconstruct the stripes' 3D segments, we use a true multi-image matching algorithm as defined in (Collins, 1996) of sweep-planes giving us a set of possible matches of 2D segments from each image (Taillandier and Deriche, 2002). For a given match, we intersect two by two planes – a plane is defined by the projection centre and the extremities of a segment in focal plane –, obtaining thus a set of 3D segments for each stripe border. The final 3D segment border is then computed by minimising the angular deviation within the set.

Each zebra crossing stripe solution is considered as an initial solution and is refined independently to obtain a better absolute position. We translate, rotate and stretch the initial shape (in a hierarchical frame to speed up the process) and try to find the

best one by maximising the image similarity with a multi image correlation score (Paparoditis *et al*, 2000). A result obtained with four images is shown on figure 4. Our algorithm was evaluated with both simulated and real data (25 cm GSD) and shows a great relative accuracy (a few centimetres). In terms of robustness, the algorithm has shown its ability. The multi image framework is very important because if one object is occluded in one or more images, it is nevertheless possible to reconstruct it. More than 95% stripes are detected and well reconstructed (our data set contains 114 stripes).
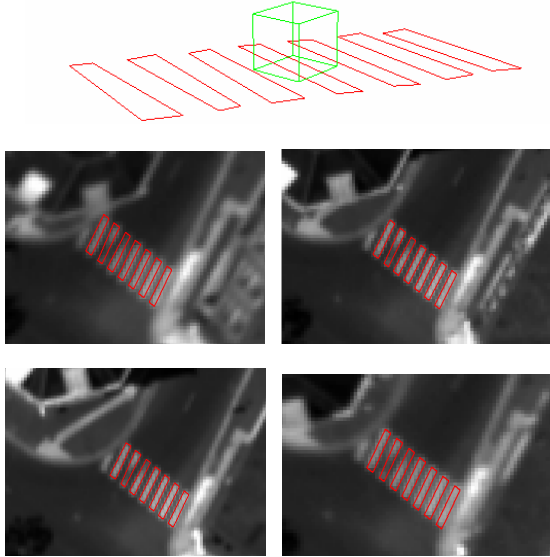

Figure 4. 3D reconstruction and images projections of the zebra crossing from the multi-view aerial images of IGN's digital frame camera. In green, a one meter cube gives the scale.

4. 3D ZEBRA CROSSING RECONSTRUCTION FROM STEREOPAIRS OF THE GROUND BASED MMS

The zebra crossings reconstructed from aerial images will provide ROIs (Region Of Interest). Thus using the initial pose of the ground-based imagery provided by the GPS/INS, one can severely cut down the number of ground-based images to investigate and the search spaces within the images. Within these search spaces, we will reconstruct the zebra crossing borders in 3D. A contour detection is performed on our MMS road-looking calibrated stereopair by applying here again the Canny filter. The contour points are then chained.

These contour chains are matched by a dynamic programming optimisation approach on conjugate epipolar lines (Baillard and Dissard, 2000). The output of this step is a group of 3D chains of contours. Within these 3D chains of contours we look for 3D segment lines that are potential long sides of zebra crossings. It consists in a 3D line segment regression of 3D chained contour points followed by a grouping step for coping with fragmentation of contours. Some *a priori* known specifications of zebra crossing like width and minimum length of their bands are used in order to generate the zebra crossing bands candidates. The last step consists in the fine reconstruction of the zebra crossing shape assuming a quasi parallelogram form for the bands. The final result is a 3D reconstruction of zebra crossing in the relative 3D coordinate system of left camera.

The algorithm has been successfully applied to 15 images containing 83 zebra crossing bands. The rate of detection is about 92% and 70 bands are correctly reconstructed. The 3D internal accuracy of reconstruction is about 1cm. The method is robust to different size, shape, orientation and position of zebra crossing relatively to the images. figure 5. shows an example of our reconstruction results with a stereo pair and the same reconstructed zebra crossing in 3D.
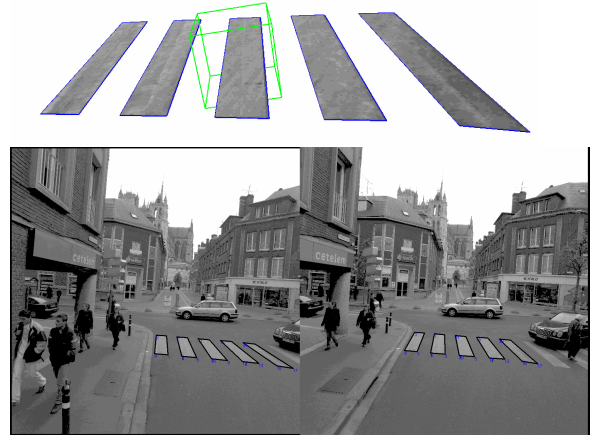

Figure 5. 3D Reconstruction of the zebra crossing from the ground-based stereopair. The zebra is partially reconstructed because not completely seen and occluded in the right image.

5. ROAD MARK MATCHING OF AERIAL AND TERRESTRIAL IMAGERY

As we have shown, we have built robust algorithms both for ground based and aerial imagery in order to obtain 3D reconstructions of zebra crossings.

On the one hand, we can have a good absolute localisation from the aerial reconstruction based on a classical aerial triangulation process. Nevertheless the extracted content is geometrically poor due to the object size / ground pixel size ratio. On the other hand, we can have a fine shape description of road mark patterns from the ground based imagery. The main idea is thus to combine both descriptions by matching objects acquired from the two points of view to achieve the finest accuracy in terms of geometry and absolute positioning (figure 6).
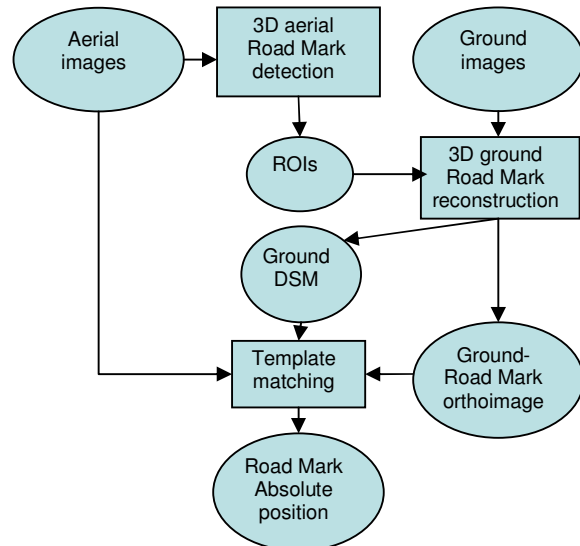

Figure 6. Synthetic diagram of our matching framework.

Our matching is based on the "analysis-synthesis" principle in computer graphics and in computer vision (Gagalowicz, 1994). The synthesis – physics of image formation – links the object to its image following equation 1.

$$I_k(i,j) = R_I(i,j) * f_{I_k}(O(x,y,z))\qquad(1)$$

where R_I is the point spread function, O is the road mark object, I_k is an image within the set, f_{I_k} is the image function model, (i,j) are the coordinates in image space and (x,y,z) the coordinates in object space.

From our ground-based images we can extract a very accurate 3D surface and orthoimages of the road mark. Nevertheless, we do not know the exact absolute position of the road mark object but only an approximate position (x,y,z) provided by the GPS/INS mix. Thus we are only looking for an optimised position and orientation of the 3D road mark object in the neighbourhood of the initial solution with respect to the aerial images. Thus for one possible instance of the road mark, we simulate what the aerial images looking this road mark should be and we compare those images with the real images.

The transformation we are looking for is a composition of a translation and a rotation. To cut down the search space, we estimate the rotation from the set of 3D long-side segments coming from both ground-based and aerial-based reconstructions (providing mean planes and mean directions). This rotation is a priori corrected to simplify the problem to a (dx,dy,dz) translation offset estimation. To conclude, we would like to find (dx,dy,dz) following equation 2.

$$\underset{(dx,dy,dz)}{\arg\min} \|I_k(i,j) - R_I(i,j)* $$
$$f_{I_k}(O(x+dx,y+dy,z+dz))\|\quad(2)$$

How do we do this matching in practice? Thanks to the ground-based road mark reconstruction, we can generate a Digital Surface Model by Delaunay's triangulation on the zebra crossing corners (see figure 7) and as a consequence build an extremely high resolution road mark orthoimage (5 mm pixel size).
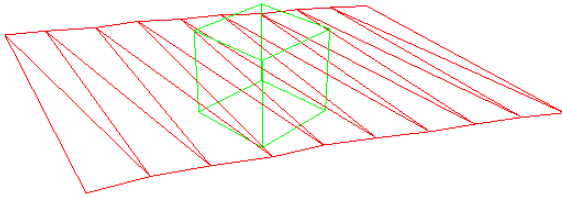

Figure 7. 3D surface obtained by triangulation of the zebra crossing stripes corners.

From the road mark DSM and orthoimage we generate a subsampled image at the resolution of the aerial images as shown in figure 8. We call that image the "ground-based reference pattern". All possible offsets along each 3D axis (for example with a sampling step of one centimetre, within a search space of one meter) have to be tested.
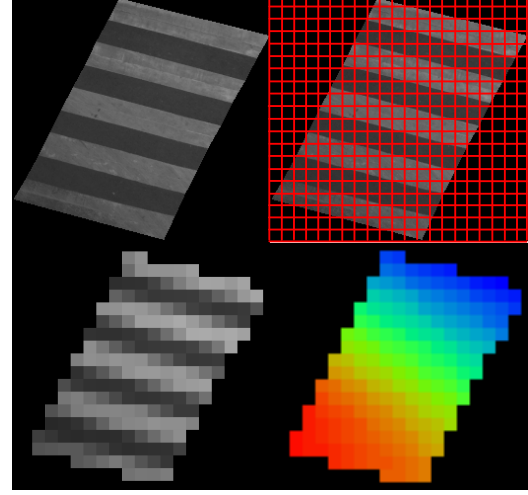

Figure 8. (Up left) Orthoimage computed from a ground-based image with the road surface DSM. (Up right) Same orthoimage but with a superimposition of the grid of pixels at the resolution of the aerial images for a given (dx,dy,dz). (Bottom left) Result of the integration of the object with the image grid at aerial resolution. (Bottom right) Corresponding DSM.

For each offset of our DSM, a new set of aerial ortho-rectified-images is generated (see figure 9.), and a similarity/correlation score between the "ground-based reference pattern" and these rectified images is calculated with equation 3.

$$Score(dx,dy,dz) = \sum_{k=1}^{n} Cov\Big(R_I * P_{ortho}(dx,dy),$$
$$g_{I_k}\Big(f_{I_k}\big(P_{DSM}(x+dx,y+dy,z+dz)\big)\Big)\Big)\quad(3)$$

Where Cov is the normalised covariance, P_{ortho} is the ortho-template of the zebra crossing, and P_{DSM} is the set of samples of the 3D surface of the zebra crossing centred on (x,y,z), f_{I_k} is the photogrammetric model of image I_k, g_{I_k} provides the grey-levels of image I_k.
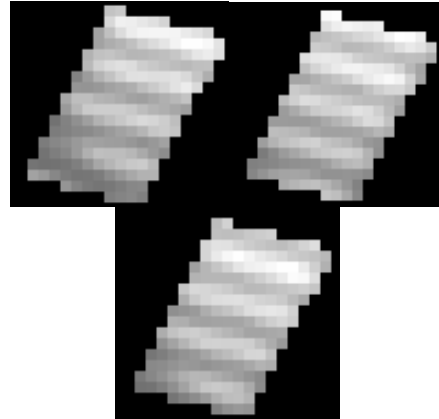

Figure 9. One aerial image resampled in ortho-geometry for different offsets (dx,dy,dz).

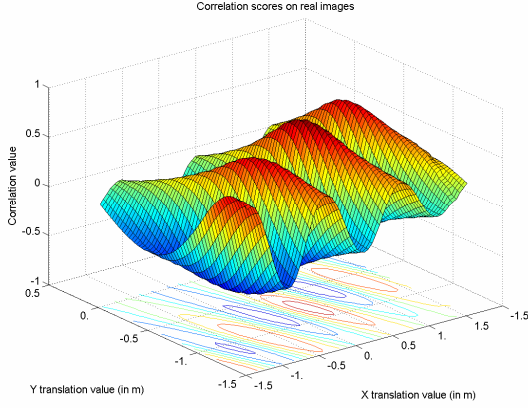We finally choose the offset providing the best score (see figure 10).

Figure 10. Correlation surface at Z=0 on real images.

As we match the complete pattern (or at least as complete as possible) we have no problem of matching ambiguities. If we had matched only one stripe of the zebra-crossing, we would of course encounter matching problems. Nevertheless, some objects (e.g. cars) can sometimes occlude some of the marking, thus leading to possible ambiguous matches. In any case, this lack of robustness can be taken into account in further processing as described in section 8.

In practice, we do not take into account the point spread function for the subsampling of the ground-based orthoimage at aerial resolution. We would like to stress that this is correct if and only if the point spread function is symmetrical. If not a solution would be to estimate the PSF by another mean for example in a laboratory calibration and to convolute our object with the PSF model.

6. VALIDATING MATCHING RESULTS ON SIMULATIONS

The described matching protocol was tested also on simulations to evaluate its potential accuracy. The simulations are both created for aerial and terrestrial images from a known 3D zebra crossing reconstruction. Stripes corners are projected in all the images within the set thus providing us with 2D polygons. Using the AGG® C++ library, we draw anti-aliased polygon for each stripe in all the images. Thus, the polygons represent the exact image position of the 3D surface with correctly integrated grey-values (figure 11).

Figure 11. Simulation of the ground-based orthoimage which is very similar to the real image itself.

Two tests were performed. The first one consists in setting a huge search space (4 meters on the X and Y axes and 1 meter on the Z one with a 2 cm sampling distance). All the translations within the search spaces are performed and a correlation score is computed. The algorithm must find the best correlation score for the translation (0, 0, 0). It is effectively the case as it can be seen on the figure below (figure 12). Note that two negative peaks symmetrically around the positive peak are found showing the shift of the 3D surface of exactly one stripe width. The represented surface is also symmetric and successive maxima (a given stripe of the surface is correlated with its neighbour in the image) are decreasing.
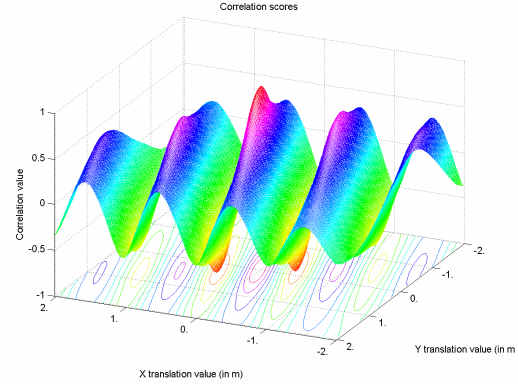

Figure 12. Correlation surface at Z=0 on simulated images.

The second test being performed consists in simulating images in the same way as in the first described test. Then, we added a known translation to the 3D surface. The algorithm found again the highest correlation score for the simulated translation.

These two tests validate the strategy employed for the positioning of the 3D surface obtained from ground based images using the aerial images.

7. ABSOLUTE POSE ESTIMATION OF GROUND-BASED IMAGES

At this stage, we have achieved adjusted 3D ground coordinates for the zebra crossing. These 3D coordinates are interesting as such to enrich a road mark database. But a more important result is that together with the corresponding image coordinates in the ground-based images (see section 4.), these 3D measures and the corresponding image observations can be injected in a photogrammetric bundle adjustment to improve the absolute localisation of all the ground-based images of the block.

We have not yet performed the "complete" bundle integrating these Ground Control Road Marks (GCRMs) together with more classical tie points in between intra-date images and inter-date images. This is a work in progress. Nevertheless, we have estimated the pose of the stereopairs in which the road marks were extracted and compared them with a centimetre accurate ground truth acquired with classical field surveys (GPS and total station). The pose is sub-decimetric and the precision is currently limited by the "poor" standard quality of aerial-triangulations. These results are preliminary and have been performed on too reduced a set to provide significantly enough quantitative quality figures.

8. CONCLUSION

We have presented a novel concept and approach mixing aerial and ground-based imagery, to achieve what we will call a "GCO-based georeferencing" (for Ground Control Objects), i.e. a very high quality absolute georeferencing of ground-based Mobile Mapping devices in dense urban areas using an external source of data. This is, we believe, an interesting contribution to the state of the art. This approach has mixed computer vision and photogrammetric techniques to achieve both robust and accurate results: extraction and matching of invariants, 3D object reconstruction, and adaptive shape image matching.

All the methodology was shown on the same running example: a zebra crossing. Of course, this approach can be completely generalised to all other road marks to densify the distribution of GCOs to improve this "image-map matching" with the highest spatial sampling as possible. This higher density should also limit the impact of possible false 3D road mark measurements (as described in section 5) with the help of robust estimators and robust filtering within the bundle adjustment.

REFERENCES

Baillard, C. and Dissard, O., 2000. A stereo matching algorithm for urban digital elevation models. *ASPRS Journal of Photogrammetric Engineering and Remote Sensing,* Vol. 66(9), 1119-1128.

Bauer, J., Klaus, A., Karner, K., Zach, C., and Schindler, K., 2002. METROPOGIS: A Semi-Automatic City Documentation System. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, proceedings of ISPRS Commission 3 Symposium PCV02,* Vol. 34, Part 3B :22-27, Graz, Austria.

Bayoud, F. A., Skaloud, J. and Merminod, B., 2004. Photogrammetry-derived Navigation Parameters for INS Kalman Filter Updates, *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences, Proceedings of the XXth ISPRS Congress,* Vol. 35, Part B5, Istanbul, Turkey

Bentrah, O., Paparoditis, N. and Pierrot Deseilligny, M. 2004. Stereopolis : an image based urban environments modeling system, *In Mobile Mapping Technology 2004, Proc. 4th International symposium on Mobile Mapping Technology, Kumming,* China, 29-31March 2004, (on CD-ROM).

Canny, J., 1986. A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 8(6): 679-698.

Collins, R.T., 1996. A sweep-space approach to true multi-image matching, *Proceedings of the 15th Conference on Computer Vision and Pattern Recognition*, San Francisco.

Chaplin, B.A. and Chapman, M. A., 1998. A procedure for 3D motion estimation from stereo image sequences for a mobile mapping system. *International Archives of Photogrammetry and Remote Sensing,* Vol. 32(3W1).

Gagalowicz, A., 1994, Modeling Complex indoor scenes using an analysis/synthesis framework. *Chapter In A. Rosenblum, Data Visualization,* Academic Press.

Frueh, C., and Zakhor, A. 2003. Constructing 3D City Models by merging Ground-based and Airborne views. *Proceedings of Computer Vision and Pattern Regonition Conference.*

Frueh, C., Jain, S. and Zakhor, A. 2005. Data processing algorithms for generating textured 3D building façade meshes from laser scans and camera image. *International Journal of Computer Vision,* Vol. 61(2): 159-184.

Paparoditis, N., Bentrah, O., Pénard, L., Tournaire, O., Soheilian, B. and Deveau, M., Automatic 3D Recording and Modeling of Large Scale Cities: the ARCHIPOLIS Project, *In: E.P. Baltsavias et al. (eds), Recording, Modeling and Visualization of Cultural Heritage*, Taylor & Francis Group, London.

Paparoditis, N., Thom, C. and Jibrini, H. 2000. Surface reconstruction in urban areas from multiple views of aerial digital frames. *International Archives of Photogrammetry and Remote Sensing,* Vol. 33, Part B3 (supplement): 43-50.

Roncella, R. and Forlani, G., 2004. Semi-automatic geo-referencing images in mobile mapping, *International Archives of Photogrammetry and Remote Sensing, proceedings of the XXth ISPRS Congress, Vol. 35, Part B5,* Istanbul, Turkey.

Taillandier F. and Deriche R., 2002. 3D reconstruction of linear primitives from multiple images for urban area modelisation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, proceedings of ISPRS Commission 3 Symposium PCV02*, Vol. 34, Part B: 267-272, Graz, Austria.

Tournaire, O., Paparoditis, N., Jung, F. and Cervelle, B., 2006. 3D reconstruction of road marks from multiple calibrated aerial images. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, proceedings of ISPRS Commission 3 Symposium PCV06*, Bonn, Germany, to appear.

Zhao, H. and Shibasaki, R. 2001. Reconstructing textured CAD model of urban environment using vehicle-borne laser range cameras and line cameras. *Proc. 2nd International Workshop on Computer Vision Systems (ICVS'01),* 284-297, Vancouver, Canada, 7-8 July.