SEMI-AUTOMATIC GEO-REFERENCING IMAGES IN MOBILE MAPPING

R. Roncella, G. Forlani

DICATA, Dept. of Civil Engineering, Parma University, Parma, 43100, Italy – (rroncell, gianfranco.forlani)@unipr.it

Commission III, WG 1

KEY WORDS: : Georeferencing, Image Matching, Image Sequences, GIS, Mobile Mapping, GPS/INS, Surveying

ABSTRACT:

This paper presents a purely photogrammetric strategy to orient short image sequences in a MM van equipped with two GPS receivers and a pair of synchronized cameras but currently still lacking an IMU. The motivation for this is twofold: bridge over short GPS outages, so increasing the vehicle productivity; improving the consistency of image georeferencing between consecutive image pairs, which in relative terms is poor due to the limited accuracy of the GPS-supported camera parameters determination. Drawing on techniques developed for structure and motion reconstruction from image sequences, a general method has been tailored to the specific conditions of the MM imaging geometry, trying to ensure reliability of the matches and stability of the solution. Though currently not all constraints between synchronous image pair are yet enforced, the first results suggest that the technique may be working satisfactory, ensuring that the error propagation is within the specifications for GPS outages of about 100 m..

1. INTRODUCTION AND MOTIVATIONS

Since the advent of digital photogrammetry it is apparent that photogrammetry and computer vision increasingly share goals and methods, though retaining their roots, i.e. stressing accuracy and metrology aspects the former, focussing more on real-time applications, image understanding and artificial vision the latter. The progress in digital imaging sensors opens new fields of application but demands for improvement in automation of the processing, to cope with the amount of data and to keep production costs in check. In this context, the automation of image orientation and, to some extent, image restitution, is being addressed in two ways: by developing automatic methods based on image coordinates or by measuring directly the exterior orientation elements.

As a well known example of the former approach, which has been growing steadily in the last years, we may list structure and motion (S&M) reconstruction from multiple views; techniques proposed under this umbrella have been applied to several fields, from visualization, archeological and architectural surveying, computer graphics and so on.

On the other side, integrated GPS/IMU systems are being more and more used in aerial and terrestrial application. The increasing demand on database population and updating is pushing demand towards using so-called high productivity surveying vehicles or mobile mapping vehicles. Through sensors and system integration (GPS, digital cameras, INS, laser scanners…), these vans can collect georeferenced data (mainly images and 3D point clouds) at relatively high speed.

Somehow following a midway path, we are trying to make both approaches to cooperate, to improve the reliability of a mobile mapping system and its productivity. Since about one year, we are developing at our Department a mobile mapping system for the acquisition and updating of road databases and road maintenance. The system specification are of an absolute accuracy of about 1 m in horizontal, 5m in elevation and 10 cm in relative accuracy (on the measurement of distances, mainly the road width). Due to limits in funding, we are upgrading and improving the on-board sensors in steps, trying to get the best out of the available instrumentation, also by developing software for image georeferencing from the navigation data and for road database population from the oriented images. One of the key modules of the software for image orientation, which is the subject of this paper, is being developed to bridge photogrammetrically over small GPS outages and improve the restitution of points in asynchronous image pairs.

1.1 A mobile mapping van without an IMU

Currently our vehicle is just equipped with a pair of Leica SR 530 GPS receivers, mounted on the roof about 3 m apart. The two receivers provide position, pitch and yaw of the vehicle, so the roll angle currently cannot be determined. From the test carried out, the accuracy of the yaw angle is around 0.1 degrees or better, under good GPS conditions (no reliable figure has yet been verified for pitch).

Two B/W Basler AF101 digital cameras with 8 mm focal lenght and resolution of 1300x1030 pixels are mounted on the front, with optical axes parallel and slightly inclined downwards and a base of about 1.70 m. The cameras, with a pixel size of 6.7 micrometers, can acquire up to 12 fps at full resolution and are synchronized to the GPS through the exposure signal sent to the input event port of the receiver. Typical mission parameters are an operating speed between 20-30 km/h and a frame rate of 2 Hz, with the GPS receiver acquiring at 10 Hz. After system calibration, under good GPS conditions and on level road sections, absolute accuracies in the range 20-50 cm have been verified at distances up to 15 m; relative accuracy in a synchronous stereo pair is better, from 2 to 5 cm across track at distances below 10 m.

Obviously, besides the errors arising from the unknown roll angle of the vehicle, the main problem is currently the lack of an inertial measurement unit, capable to make up for the loss of lock of the GPS signal due to occlusions caused e.g. by trees and buildings: this severely limits the productivity of the vehicle at the current stage of development. Though we are not arguing that we can dispense with a IMU to reach a truly operational level (and this is obviously our long term goal), there is still a sizeable set of roads where we can perform a

survey without major GPS outages, for instance along planes or gently hilly areas and small villages with low raise buildings.
We noticed indeed that in many cases the loss of lock is very short in time and space, let say from 20 to 100 m. Often we have a relatively long sequence of small interruptions each followed by small sections where the ambiguity is recovered for some tens of metres or less. In other cases, for instance crossing a tree row, the loss of lock is just a 30-50 m long. In such cases, we may use the information recovered by the image sequence itself.
Although the most important motivation is to bridge over GPS outages, by applying the procedure to successfully georeferenced image sequences, we can improve their orientation parameters, very much like applying integrated sensor orientation in aerial blocks (Heipke et al, 2002). This may allow point restitution also among images of either the left or the right camera or even multiple collimation, to increase accuracy and reliability when needed.
Our goal has therefore been finding a robust algorithm, capable to determine automatically the cameras' motion structure along all the unreferenced image sequence. To this aim we built upon the theories and applications heavily developed in the last few years by the CV community. To our understanding their application to mobile mapping did not received much attention (*Tao et al, 1999; Crosilla, F., Visintini, D., 1998* are two exception) but we believe they may be appropriate to solve this task, provided the loss of lock is not too long. It is well known indeed that, without ground control or auxiliary information, the error propagation on a strip is rather unfavourable and the solution quickly deviates significantly, especially in height.

1.2 The imaging geometry of a mobile mapping

There is a number of issue characteristics of the imaging geometry of an image sequence taken by a van with a pair of synchronized cameras: a large variation in depth (or image scale), a small base, fast moving objects, and so on. They will be discussed later in the detailed description of the method. It is clear nevertheless that while for a robust and efficient image matching and S&M recovery the imperative is to take shots not too different from one another (i.e. the frame rate has to be quite high compared to the vehicle's velocity, especially along curved paths) the position error will rapidly increase with the number of images processed. So, we need a method satisfying both constraints: a robust estimation of the cameras pose and limited systematic errors in the exterior orientation. The basic geometry of our blocks will therefore be a double strip, with longitudinal overlap larger than 60-70 percent of the image format along straight road sections (less on curves), side overlap of about 80%. The relative orientation of each pair is known and constant and the strip ends are constrained to the exterior orientation values provided by the GPS solution (just before and after the loss of lock). Whenever the loss of lock lasts too long, some human interaction may be accepted: in order to constrain the solution of S&M estimation we can bring in (*Crosilla, F., Visintini, D., 1998*) point coordinates from a GIS system. If the number of points in one image is enough, this may allow a spatial resection; in most cases just a partial constraint will be enforced, if just a few points are available. Another (though less reliable) option would be to use the noisy code solution of the GPS, which may be available along the sequence.
In the following section we describe how our general M&S recovery system works; in Section 3 we discuss how we tailored it to the MM application; finally, in Section 4 we show and analyze the results obtained during a test session.

2. STRUCTURE AND MOTION RECOVERY FROM IMAGE SEQUENCES

2.1 Introduction.

The last ten years witnessed the growth several methods and algorithms for recovering structure and motion from an image sequence, exploiting the geometric relationships between the images of a sequence and their similarity. The use and improvement of robust algorithms (MLS, RANSAC,…) capable to eliminate a great percentage of outliers in a data set and of correlation procedures more and more reliable, allowed to develop fully automatic vision systems capable to solve the S&M problem.
As previously pointed out, these algorithms require that the images of the sequence do not differ too much in order to achieve a good match of feature correspondences, which are the basis for a successful camera pose reconstruction. To limit error propagation some constraint are usually called in, such as a closed sequence around an object; besides, a key element is the ability to trace a consistent number of points along a sufficiently long section of the sequence, to allow a good relative geometry among cameras and objects. Developing our general system for M&S recovery, which largely builds up on the techniques presented in (*Hartley, R., Zisserman, A., 2000)*, we therefore tried to specialize it in order to gain advantage of some constraints that apply in the mobile mapping case, in the attempt to overcome some of the restrictions of the general case and optimizing at the same time the error propagation.

2.2 Robust automatic recovery of structure and motion.

2.2.1 Feature extraction and putative correspondences evaluation.
The first step in our workflow is the extraction of interest points from the sequence, possibly ensuring that they can be easily related to the same image points in other images of the sequence. We used the Harris operator (*Harris C., Stephens M. 1987*) but also the Foerstner operator provides good results (*Förstner, W. and E. Gülch, 1987*). The algorithm developed try to achieve a uniform distribution of the extracted point on the image frame, in order to give better results during camera pose estimation and reject points without a sufficient gradient g.v.
In order to compute a first geometry of the camera pose, we need to establish for every extracted point in an image a potential correspondent point (if any) in the next image of the sequence. This correspondence is accepted or rejected on the disparity threshold and on the similarity of the g.v. in the neighbourood. Currently we use, in order to limit computation time, a simple cross-correlation between two windows; a possible improvement might be using LSM to improve accuracy and correctness in the matched points. Even if the algorithm eliminates many wrong correspondences (we use a 0.8 threshold on the cross-correlation coefficient and adopt a bidirectional uniqueness of matching criteria) the data set is still affected by a great amount of outliers.
2.2.2 Outlier detection.
To achieve an error free set of correspondences in the image pair we filter the data set taking into account that all points must satisfy some geometric constraints due to the cameras' relative position (normally unknown). First of all we estimate the epipolar geometry between the first two images of the sequence with a robust estimation algorithm (*Fischler M., Bolles R.,*

1981). Since we do not know how many outliers affect our data set, to reduce computating time we use an adaptive algorithm (as suggested in *Hartley, R., Zisserman, A., 2000)* that starts considering a 99% of outlier presence and than updates the number of iterations required to assure the elimination of all the outliers (at least for the epipolar constraints). When a first camera geometry has been established we then try to find some more correspondences through a guided matching (we use again cross correlation algorithm); the final estimate for the fundamental matrix derive from a least squares solution over all matches.

Since the epipolar constraint cannot filter out all false matches (see next chapter) the data set undergoes another, more restrictive, control: joining three (or more) consecutive images we can estimate the three view geometry (*Hartley, R., Zisserman, A., 2000*) through a robust algorithm, finding a more reliable camera reconstruction and getting rid of the remaining outliers. The tests we carried out and the results published in literature assure that a 99% probability of success in outlier elimination is reached.

2.2.3 Metric reconstruction and bundle adjustment.

Until now we have only determined image points correspondence, filtering the wrong ones; we finally recover the structure and motion of the whole sequence through a self-calibration approach (as in *Pollefeys, M., 1999*). Besides, since in our mobile mapping van we use calibrated cameras, we can estimates the metric frame of the reconstruction directly through the use of the essential matrix. The calibrated approach gives more reliable results (mainly in the errors estimation) even if lead to larger residuals. Once the metric reconstruction of the sequence has been achieved, a bundle adjustment of all the observation leads to an optimal estimation of all the S&M (in terms of minimization of a geometric cost function).

In order to limit error propagation and the probability of finding local minima during bundle adjustment, we adopted a hierarchical approach to compute an initial estimate of the ground point coordinates and the exterior orientation parameters of the cameras. The whole sequence is subdivided in shorter sub-sequences and the set of points is found which was traced in every image of the sequence. The optimal number of sub-sequences may depend on the problem at hand: our goal is to ensure that the relative geometry of the cameras along the sequence changes enough to allow a better intersection of the homologous rays. Of course, if the changes in attitude between consecutive images are not smooth or if the scene changes very quickly (as in curved road sections) or if an object moving fast through the scene (such as a truck on the opposite lane) cuts most points in the background also this strategy may not be enough. Nevertheless, we found that this normally improves the quality of the approximations.

In each sub-sequence the trifocal geometry among the first, last and middle frame is computed, with the rationale that these three images should have the best relative geometry. A metric reconstruction is performed through the essential matrix, yielding by triangulation the coordinates of the common set of points. Based on that, the exterior orientation parameters of the intermediate frames and the approximate coordinates of the remaining points along the sequence will be calculated by alternating resection and intersection using a linear algorithm and the unit quaternion as in (*Horn, B.K.P., 1987*) and (*Quan, L., Lan, Z., 1999*). Optionally, a l.s. bundle block adjustment (Forlani, G., Pinto, L., 1994) with data snooping will be executed to improve the orientation parameters and discard remaining outliers.

Finally, all sub-sequences are joined together by using the points of the last image of the subsequence, which is also the first of the next sub-sequence. This propagates also the scale of the metric reconstruction along the whole sequence. Once the sequence is completed, a final l.s. bundle block adjustment with data snooping is performed using all images and including all available information on the object reference system.

Though the all-purpose algorithm implementation has only recently been completed, test on image sequences around buildings as well as along a rock face showed good results, fairly comparable with those of manual orientation of the same images. As mentioned above, the number of images in the sub-sequences may vary depending on the scene characteristics and on the camera motion: while for movements of a hand-held camera towards a distant subject we found that 15-20 was a good compromise, in the MM case it is likely to be much smaller, as will be discussed in the next section. If cutting the sequence indeed complicates a bit the processing, since an additional step is needed to put them together, we believe this is a price worth paying for increased (at least local) stability of the solution.

3. THE MOBILE MAPPING CAMERA GEOMETRY

As previously pointed out, the geometry of the image acquisition of a mobile mapping system presents some disadvantages for a general structure and motion reconstruction algorithm. On the other hand, since cameras are calibrated and their relative orientation is known with sufficient accuracy, we have in fact two overlapping image strips, a fact we can exploit to eliminate some of the problem’s unknown. We leave open the possibility for the algorithm to manage sequences along the motion direction (i.e. a sequence produced by a single camera) and across the motion direction (i.e. a sequence from a stereoscopic synchronous system); also, the pipeline structure of the program allows to merge single camera and stereoscopic sequences. This flexibility leads to a great improvement in performance, because the system gain in robustness from the combination of both approaches.

If we process a sequence of images from a single camera pointing along the vehicle trajectory, the first difficulty arising is due to the small overlap between consecutive frames. For reliability reasons we consider a tracked point as good only if it has been seen in at least three frames: therefore the accepted points are almost always located in the middle of the scene, quite far from the vehicle. Using the procedure described in Section 2, this would lead to large uncertainties in the estimation of point coordinates and exterior orientation. Moreover, the epipolar constraint used to filter outliers often performs poorly when tracking of well defined points along the motion direction (for instance lane markings): indeed the epipolar line tends to overlap with the vanishing line of the road borders so little discrimination is achieved. This increase the number of wrong matches, later removed by the trifocal tensor, on what if often, in the countryside, the best source of interest points.

Despite this, the forward approach has also advantages: in straight road sections about ¾ of the image frame depicts the same scene in three consecutive pictures: this generally leads to good results of the cross-correlation matching procedure (with more troubles because the increase in scale for points at the frame bottom).

Across the left and right images of the sequence some other useful constraints apply: in the normal stereo configuration, the epipolar lines are almost orthogonal to the vanishing direction of road markings: therefore the fundamental matrix and the

RANSAC estimation of corresponding matches yield a more reliable recovery of the camera structure.
Moreover, if the left and right shots are simultaneous, the epipolar geometry between the images is unchanged through the whole sequence: a first calibration of the relative camera leads to a precise and reliable computation of the fundamental matrix that can be used to "robustify" the matching in a guided procedure. As relative orientation and baselenght are constant along the sequence, the Euclidean reconstruction achieved using this information, allows to join different image subsets and eliminates the scale factor ambiguity.
Merging the two approaches assures a great improvement in the procedure performance: while the *'along sequence'* pair leads to a great number of correspondences (most of which correct), the epipolar constraint arising from the '*across sequence*' pair tends to eliminate some ambiguity (i.e. points on the road markings) or troublesome wrong matches (i.e. building façade texture). This merging is performed through a trifocal tensor estimation using a robust algorithm. The resultant sub-block (using two consecutive trifocal tensor computation we obtain a symmetric four-image block) has a balanced size: the distance between consecutive frames is approximately 4 m, while the baseline of a synchronous pair is about 1.7 m. Since the camera pose estimation is obtained within this symmetric configuration, matched points can be found also in the vicinity of the projection centres (so filling more uniformly the image format); using two trifocal tensors provides a strong filter and leads to a better constrained and redundant bundle block adjustment at the end of the pipeline; finally, though less precise, points far away tends to be tracked along many images of the sequence, providing ties between sub-blocks.

4. FIRST TESTS AND RESULTS

In this chapter we present a summary of the results achieved during the test procedure. Although the testing of the implementation cannot be yet considered completed, we processed a fairly representative set images, in terms of arrangements of the images in the processing sequence, in terms of road traffic, vehicle speed and scene background. Given the variety of situations along roads, no definitive conclusion can be drawn yet, but tests have been useful to understand how and when satisfactory results can be obtained using the proposed approach. In the first part of this section we illustrate results on a block along a small countryside road around Parma, as showed in figure 4; then we evaluate through simulated data set the error propagation along a sequence of about 250 m and how points or camera constraints can improve the motion and structure computation.

4.1 Matching procedure and metric reconstruction

As already said all the tests were performed using two digital cameras (Basler A101f) with a 8 mm focal length and a resolution of 1300x1030 pixel. Since the camera lenses we use produce a strong barrel distortion effect on the images, we first determined their calibration parameters with a build-up test field. The estimated distortion model is correct up to 0.5 pixel. A good camera distortion model is necessary because attempting an automatic self calibration never gave the desired and expected results.
About 2000 feature points have been extracted in every image of the sequence using the Harris operator: the number of feature to accept was determined, considering the camera resolution and quality, by finding a compromise: using too many points the epipolar and trifocal estimation may leads to uncertain results, whenever they are too close to each other so that an ambiguous matching arise; on the contrary with fewer points their ground distribution is poor and the camera pose is affected by a weak geometry.
Through the disparity threshold a first putative matching is computed between the images: though the *along sequence* approach tends to give more correspondences than the *across sequence* one, the difference is negligible (see table 5).
The main difference between the two approaches arise during the first outlier filtering: here, the perspective differences between left and right images leads to a more difficult matching between the putative correspondences; with a least squares matching approach, the differences arising from the disparity in the viewing angle may be taken in account and more matches might be obtained.
In order to limit computation time for the *across sequences* the epipolar geometry is evaluated only on the first pair using a robust algorithm; then, in the other pairs, a guided matching procedure, using the estimated fundamental matrix between left and right images, is performed.

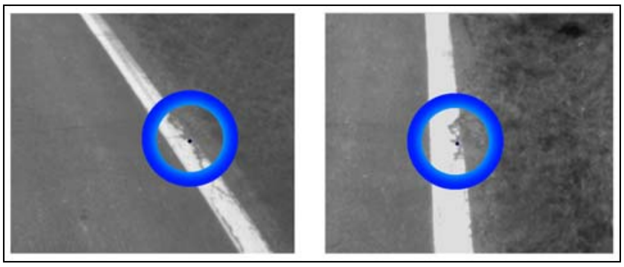

Figure 1. Matched points between left and right image of a synchronous stereo pair (*across sequence*).

Figure 2. Matched points between consecutive left images (*along sequence*).

The algorithm proceeds joining two different epipolar matched data sets in order to perform the trifocal outlier filtering; the common points are therefore fewer than those found separately in the image pairs.
The trifocal tensor manage to eliminate those outlier that satisfied the epipolar geometry constraints: in figure 3 we can see how the parallax effect of the branches of the tree on the background of the white building, arising from the different standpoint of the left and right cameras, satisfies even the epipolar geometry but is spotted by the tensor geometry.

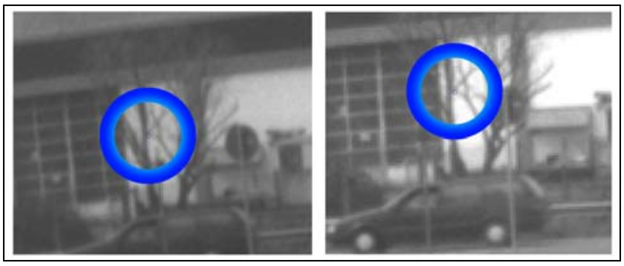

Figure 3. Wrong match due to different viewing angles eliminated by the trifocal tensor test.

The disparity is recognized, and the correspondence rejected. In some other, more malicious, cases, the parallax effect is weaker between two simultaneous frames but became larger and larger along the sequence, leading to a whole set of apparently good matches in adjacent images. This suggest that a final trifocal test may be useful in order to reject all these “moving” wrong matches.

Finally the two trifocal filtered data are joined together leading to a four-image block set. In the following figure the accepted point distribution is presented.

Figure 4. Ground points distribution after the trifocal tensor filtering. Grass presence on the image left area prevents any reliable matching to be found.

Then the two images ahead are used to compute the essential matrix and obtain a first reconstruction of the scene points. The algorithm performs a test on the camera geometry: if the reprojection errors of the ground points is worst than the a priori expected precision (1 pixel seems to be a reasonable figure) space resection and space intersection are alternated until convergence. In straight path sequences, at this stage, at least 60 – 100 points have passed all the outlier detection tests; on these points a half pixel reprojection error on both the images is not unusual. Along curved path, obviously the matching procedure is hindered by the increased disparity between adjacent and consecutive images: less matches may be tracked and the camera geometry suffers of a poorer ground points distribution. Nevertheless, while the solution’s convergence is slower (and in some cases may be unstable), the final reprojection errors are similar to those of the straight road sections.

Once obtained the ground points’ coordinates, the process may be iterated on the back images of the sub-block.

At this time the scene points coordinates and the camera pose are referenced in a local system.

When another block is solved, through all the above mentioned steps, using a conformal transformation estimated by using unit quaternions for the computation of the rotations, the new block is referenced in the first local system. The algorithm proceed in this way until the whole sequence is processed.

The final bundle adjustment, despite possible drift of the solution due to the linking, did not show any convergence problem. Though independent verification of the ground coordinates has not yet been performed, plotting suggest a correct structure and motion reconstruction.

Matching procedure	Sequence 1		Sequence 2	
	across	along	Across	along
Putative correspond.	785	954	710	832
Inliers after epipolar	358	581	370	573
Common pts. in 3 im.	124		137	
Inliers after trifocal	86		107	
S&M reconstruction	Sequence 1		Sequence 2	
Common pt. in 4 im.	62		80	
Reproj. err. after ess.*	0.79 pix		0.65 pix	
Sub-block orient. err.	4.3 cm		3.9 cm	
Reproj. err. after BA.*	0.52 pix		0.51 pix	
			***** mean value	

Table 5. Mean performance of the algorithms in two different sequence of about 100 m length (not including curved paths).

4.2 Simulations on stabilization of the solution

It is well known in aerial photogrammetry that long image strips tend to drift away from the terrain, especially in elevation, unless ground control is provided along. Since at first sight the geometry of the MM van looks similar, we performed a simulation to find out whether it is truly so and, in case, what kind of improvement may be introduced by adding control information. This may come either from maps or GIS data or by the differential GPS code solution, whenever available (e.g. when the ambiguity cannot yet be recovered but 4 satellites are visible). Of course, the former solution would be the most appealing, since no additional work would be necessary, apart in the final block adjustment, while looking for points in the map is always time consuming (unless they are reprojected after an initial solution constrained only at the strip ends) and error prone. Besides, most of the time the information from map points is either horizontal or vertical, rather than 3D.

To this aim, starting from a set of exterior orientation parameters computed by the GPS solutions and a number of suitably distributed points along the road section have been generated and projected over the images, trying to filter out, based on two distance thresholds, those unlikely to be visible. By constraining the first and last image of the sequence, as we would do using the last and first successful GPS solution before and after the loss of lock, the RMS errors on the ground points and on the exterior orientation elements have been computed by variance propagation in a block adjustment. Given a measurement accuracy of 1 pixel to image coordinates, for a sequence of 80 images (250 m long) the RMS of the X,Y coordinates is worse than that for elevation by almost an order of magnitude. With an average number of 10 rays per point and of 17 points per image, theoretical RMS are around 50 cm in horizontal and 4 cm in elevation. Corresponding figures for projection centres are lower, with 15 cm in horizontal and 3 cm in elevation. These results seems somehow too good, so we want to review the likelihood of the image point distribution in the simulation, especially as far as the number of ray per image is concerned. Anyway, if the trend will be roughly confirmed, some conclusion can be drawn: using GPS code solution should not help much, since the accuracy of this solution, under the extreme condition (4 satellites, poor signal reception) we are considering, cannot be expected to be better than 10 cm. On the other hand, 50 cm is well within the tolerance of points taken by 1:2000 maps, but too small for the average accuracy of 1:5.000 maps for the planimetry. Since 1:2.000 maps are restricted to urban areas, also this method may not lead to significant improvements of the length of the recoverable loss of lock.

5. CONCLUSIONS AND FUTURE WORK

Though we cannot claim a verified accuracy level, we are confident that the proposed methodology, as shown in the test images, is capable to achieve reliable result, at least for short sequences. Since to ensure a good motion estimation good interest points well distributed in the scene are necessary, the method may work well especially in sub-urban environments, where buildings and man made objects make the tracking process easier and the loss of lock should be short in time.

There are still a few steps of the process where there significant improvements are possible and necessary. The most important is the evaluation of the putative correspondences: based just on a disparity threshold, is quite rudimental and spends a lot of computation power providing results of different quality depending on the scene characteristics. Since an important problem is still to find correspondences in the bottom part of the images, the use of a LSM algorithm may possibly allow to incorporate more points at the image bottom, where scale changes more dramatically between pair of frames. Introducing a further guided matching, once the camera pose is correctly estimated, in order to gain more correspondent points, it's not unreasonable. Since finding points on the paved road surface is not too easy due to the difficulty to adjust to changing lightning conditions and shutter speed, another possible improvement is to apply some sharpening filter such as the Wallis filter. Finally, the management of the various blocks may be optimized considering different schemes and made more flexible.

As already mentioned, the first step will be to come up with figures on the accuracy of the reconstruction in object space. This may also put some light on the possibility to use the procedure for improvement of the exterior orientation computed by GPS only. To this aim, the GPS solution will provide initial values to speed up the search for correspondences, while the trifocal geometry will serve as filtering of outliers. Finally, an integrated bundle block adjustment including the GPS solution as pseudo-observed values may provide an improved block geometry.

6. REFERENCES

References from Journals:

Crosilla, F., Visintini, D., 1998. External Orientation of a Mobile Sensor System for Cartographic updating by dynamic vision of digital map points. Bollettino di Geodesia e Scienze Affini, 1 (1998), pp. 41-68.

Fischler M., Bolles R., 1981. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, vol. 24:3, pp. 81-95.

Horn, B.K.P.**,** 1987. Closed-form solution of absolute orientation using unit quaternions**,** *J. Opt. Soc. Amer. A* vol. 4, no. 4, pp. 629-642.

Longuet Higgins, H. C., 1981 A computer algorithm for reconstructing a scene from two projections. *Nature*, pp. 133-135, September

Quan, L., Lan, Z., 1999 Linear N-Point Camera Pose Determination in *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 21, No. 8, August

Torr, P.H.S., Zisserman, A. 2000. MLESAC: A new robust estimator with application to estimating image geometry. In *Computer Vision and Image Understanding,* 78(1), pp. 138-156.

References from Books:

Hartley, R., Zisserman, A., 2000. *Multiple View Geometry in computer vision*. Cambridge University Press, Cambridge, pp. 1-496.

Cooper M.A.R., Robson, S., in Atkinson, K.B., 1996. *Close Range Photogrammetry and Machine Vision*. Whittles Publishing, Caithness, pp. 9-50.

References from Other Literature:

Brown, M., Lowe, D.G., 2002. Invariant Features from Interest Point Groups *British Machine Vision Conference, BMVC,* Cardiff, Wales.

Forlani G., Pinto L., 1994. Experiences of combined block adjustment with GPS data. Int. Archives of Photogrammetry and Remote Sensing, Vol. 30 Part 3/1, Muenchen, 1994, 219-226.

Förstner, W. and E. Gülch, 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: *ISPRS Intercommission Workshop*, Interlaken, pp. 149-155.

Harris C., Stephens M. 1987. *A combined corner and edge detector*. In: Proceedings of the Alvey Conference, pp. 189-192.

Hartley, R., Sturm, P., 1994. Triangulation. In *Proc. DARPA Image Understanding Workshop*, pp. 745-753.

Heipke C. et al, 2002. The OEEPE test Integrated Sensor Orientation, OEEPE official publication n. 43.

Hu, M., Yuan, B. 2002 Robust Estimation of Trifocal Tensor using Messy Genetic Algorithm. In *Pattern Recognition, 2002.* In: Proceedings of the 16th International Conference on Pattern Recognition.

Pollefeys, M., 1999. Self-calibration and metric 3D reconstruction from uncalibrated image sequences. PhD thesis. EAST_PSI, K.U. Leuven.

Tao, C., Chapman, M. A., El-Sheimy, N. and Chaplin, B., 1999. Towards automated processing of mobile mapping image sequences. *International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, 2W1*