BUNDLE ADJUSTEMENT AND POSE ESTIMATION OF IMAGES OF A MULTIFRAME PANORAMIC CAMERA

Cannelle Bertrand 1, Craciun Daniela 12, Paparoditis Nicolas 1, Didier Boldo 1

1 : IGN, Laboratoire MATIS,2/4 Avenue Pasteur, 94 165 Saint Mandé cedex, France
Phone :(+33) 1 43 98 80 04, fax : (+33) 1 43 98 85 81
2 : Télécom Paris Tech, CNRS URA 820-TSI Department
46 rue Barrault, 75634 Paris Cedex 13, France
Email : {bertrand.cannelle, daniela.craciun, nicolas.paparoditis, didier.boldo}@ign.fr

Abstract: The STEREOPOLIS mobile mapping system, developed in the MATIS laboratory of IGN for city modelling and multimedia applications, integrates a panoramic head composed of 10 full HD cameras very accurately synchronised. For each pose, a panoramic image is generated from the set of corresponding but poorly overlapping images. In this paper, we evaluate the performance of a three-step method developed in Craciun [1] that computes the relative pose between the different frame camera images composing the panoramic frame. A phototheodolite, i.e. a theodolite coupled with a digital camera, is used to construct a reference data set with a ground truth which is sufficiently accurate (to some extent) to evaluate by comparison the results of our pose estimation process. We present the different steps to compute, then we present the algorithm which is used as well as mathematic concept and finally results are commented the results of our pose estimation process will also be compared with an off-the-shelf high quality software using SIFT-based corresponding points and bundle adjustment, i.e. Autopano[2].

1. Introduction

The STEREOPOLIS mobile mapping system, developed in the MATIS laboratory of IGN, aims at collecting 3D imagery for city modelling and multimedia applications. One if its imaging subsystems is a panoramic head composed of 10 full HD cameras (1924x1080). The different cameras are very accurately synchronised. For each pose, a unified panoramic image can thus be generated from the set of corresponding images. The particularity of our panaoramic system is the very poorly overlapping of the images (less than 1% of image size).

In general, the panoramic generation pipeline is composed of four sequential steps:

- the matching of homologous points between the pair of panoramics,
- the gross estimation of all the stereopair poses using the tie points,
- the fine pose estimation by bundle adjustment,
- the resampling of each elementary image in the panoramic space,
- the blending of all elementary images to obtain a seamless panoramic.

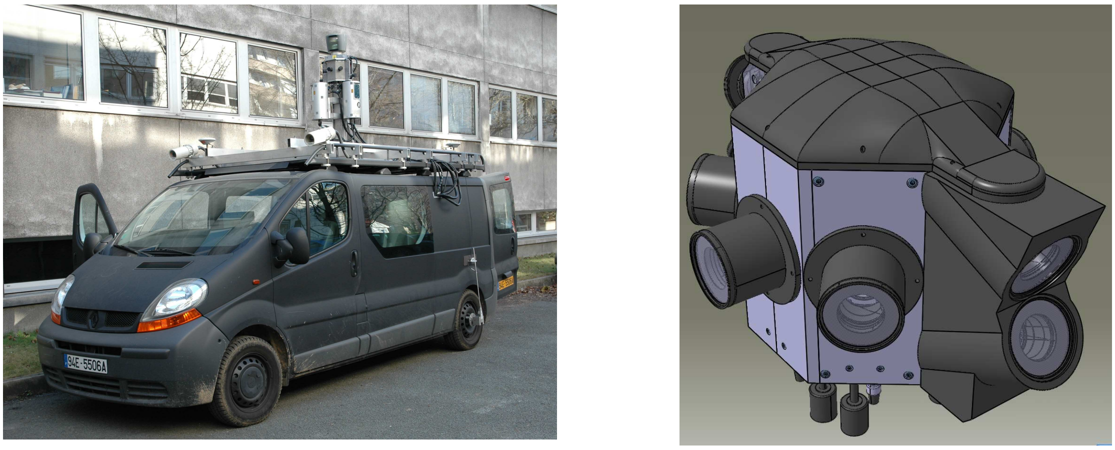

Figure 1: (on the left) the STEREOPOLIS II mobile mapping system. (on the right) the design of the panoramic head with 10 full HD cameras

Several techniques have been developed to find homologous points. One very successful approach in the case of structured environments consists in extracting and matching SIFT interest points and descriptors [3], [4]. In our context, the difficulty lies in the very small image overlap and in the possible very poor number of interest points. Nevertheless, in the context of a calibration procedure where we aim at estimating the relative pose of all the images, we can cumulate the measures coming from different image acquisitions to augment the number of points visible in the image overlaps. This process is valid under the assumption that the geometry of the system is perfectly rigid during the survey. The geometric stability of the mechanics is difficult to guarantee orientation-wise especially when a pixel represents a very narrow field of view (in our case 0.04 degrees). Thus, for applications needing the highest accuracy, it is necessary to estimate these angular variations on the-fly, and thus we face again the problem of an unsufficient number of points in the overlap.

2. Description of our matching process

To avoid these drawbacks, we have developed a matching of homologous points in two neighbouring images based on anonymous features (image patches) [1]. We follow a two step approach. In a similar way to the process describe in S. Coorg and S. Teller [5], the first step of our pose estimation method consists in finding for each pair of overlapping images the best rotation which optimises the Normalised Cross Correlation similarity score computed on the overlap of the first image with the second image rigidly transformed in the geometry of the first image. The search for the best is performed in a greedy way within a multi-resolution framework. In our case, we use an initial solution and bounds for the rotation coming from the head mechanical design plan. Then, we use the refined rotation to initialise the matching of anonymous points, *i.e.* points on a grid covering the overlap, based on the similarity of image patches centered on the considered points. Indeed, these homologous points are necessary to feed a photogrammeric bundle adjustement to estimate accurately the relative pose of all the images within the panorama. The main advantage of this method is to find in all situations corresponding points even if the surface is uniform or regular.

3. Our bundle adjustment process

Many low-cost softwares to generate panoramic can be found on the market. Some of them integrate rigorous photogrammetric models in the geometric adjustment including selfcalibration parameters. Nevertheless, these software are dedicated to non metrological applications. In general they aim at generating seamless transitions between the images by finding the intrinsic and extrinsic parameters minimising the residues on the tie points. The problem is that (for classical panoramics) these tie points are positioned nearly all at the same radius in the image. Thus there are no observations nor constraints within the rest of the image field and the internal geometry of the image can not be determined accurately.

In our case, as we have ten different cameras on the panoramic head with different intrinsic parameters, our general calibration approach to estimate the relative pose of the images is in two steps. We first estimate the calibration of each camera with a 3D network of very accurately surveyed targets. During that inner calibration process, all the cameras are physically tied in order to have the same external orientation for each pose. Then we use a bundle adjustment where we estimate the extrinsic calibration parameters with a bundle adjustment.

Our perspective model separates the principal point of autocollimation and the Principal Point of Symmetry which is the intersection of the focal plane with the optical axis. Our distorsion is considered radial and of amplitude modelled by a polynom of the form $f(r)=k_1.r^3+k_2.r^5+k_3.r^7$ where r is the radius).

3.1. Definitions

The representation of the rotation which is used is a vector with 3 composantes: the norm of this vector is the angle and this vector normalized represents the axe of rotation. Formula (1) and formula (2) explains the correspondence between the axe and angle representation and the rotation matrix representation.

$$\text{Pose} \begin{pmatrix} 0 & -\gamma' & \beta' \\ \gamma' & 0 & -\alpha' \\ -\beta' & \alpha' & 0 \end{pmatrix} = \frac{1}{2}(R - {}^{t}R) \text{ and } \cos(\theta) = \frac{1}{2}(tr(R)-1)$$

$$\text{So } u = \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \end{pmatrix} \text{and} \begin{cases} \theta = \arcsin\left(\sqrt{\alpha'^2+\beta'^2+\gamma'^2}\right) \\ \theta = \pi - \arcsin\left(\sqrt{\alpha'^2+\beta'^2+\gamma'^2}\right) \end{cases} \tag{1}$$

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + A\sin(\theta) + A \times A \times (1-\cos(\theta)) \text{ with } A = \begin{pmatrix} 0 & -\gamma & \beta \\ \gamma & 0 & -\alpha \\ -\beta & \alpha & 0 \end{pmatrix} \tag{2}$$

3.2. Presentation of system

In our system, even though the perspective centers of the cameras are separated by a few centimeters, we will consider at a first stage that this distance is very small relatively to the distance between the cameras and the objets of the scene. Thus in our bundle process we will first suppose that all cameras have the same perspective centre. So for each image we will estimate the rotation unknown.

All measurements are in 3 dimensions and norms of theses are equal to one. For a given tie point of image coordinates (c_a,l_a) in image A and (c_b,l_b) in image B (and corresponding 3D rays of coordinates (c_a,l_a,f_a) and (c_b,l_b,f_b) in each image system), we try as expressed in equation (3), to estimate the unknowns, *i.e.* the rotation between A and a reference image (arbitrary chosen within the set of images of the panoramic) and at the same time rotation B and the same reference image. This will allow to integrate each observation from all pairs within the same reference frame, i.e. the frame of the panoramic.

$$R_{A \to pano} \frac{{}^{t}\left(c_A, l_A, f_A\right)}{\sqrt{c_A{}^2 + l_A{}^2 + f_A{}^2}} = R_{B \to pano} \frac{{}^{t}\left(c_B, l_B, f_B\right)}{\sqrt{c_B{}^2 + l_B{}^2 + f_B{}^2}} \tag{3}$$

We have chosen to solve this system with least square.

Equation (4) explains the system to minimize, where h denotes all corresponding points according to the pair of images (i,j), $i \in [2;N]$and $j \in [1;N]$ and N the number of image in the system.

$$\left(A_{i,j}(h)\right) dR_i = R_j(h) \tag{4}$$

With three images with corresponding points between (1,2) and (2,3), this system will be compensated in the following way

$$\begin{pmatrix} A_{2,1}(h) & 0 \\ A'_{2,1}(h) & A_{3,2}(h) \end{pmatrix} \begin{pmatrix} dR_2 \\ dR_3 \end{pmatrix} = \begin{pmatrix} R_1(h) \\ R_2(h) \end{pmatrix}$$

In practice, to have an initial solution, we use the mechanical design. We cans also calculate the initial solution for each par and propagate in the panoramic referential.

4. Evaluation of our compensation method on images acquired by a photo-theodolite

To evaluate our method of compensation, we have used a photo-theodolite (a Trimble VX) which provides accurate horizontal and vertical angle for each image acquired. We can consider these angles as references for our evaluation process. Indeed the angular pixel size of the camera is 0.008° which is much larger than the angular accuracy of the total stations wich is 0.3 mgon.

For a given pair of images, we have compensated 24 measures covering their overlap. After computation, the mean of angular residuals on the tie points on all measurements is better than 0.005° (*i.e.* 0.625 pixels). There is a mean of 0.00484° and a standard deviation of 0.00221°.

To compare the overall quality of the calculation relatively to the rotation given by the photo-theodolite, we have calculated an intermediary matrix equal to equation (5).

$$R_{res} = R_{compute} \cdot {}^{t}R_{photo-theo} \tag{5}$$

By extracting the angle of this rotation, we obtain the overall angular accuracy of both the image matching process and our compensation process. In this case, it was 0.0246 degrees (*i.e.* 3.5 pixels). A part of this error is due to the parallax between the perspective center of the camera and the rotation axis of the total station.

5. Results

5.1. Comparison with the mechanical design

Figure (2) illustrates the difference in rotation between the theoretical mechanical design and the rotation found with the algorithm for each camera (the camera names are

21,22,23,33,34,43,42,41,31,32). The spatial orientation of the cameras on the diagram grossly corresponds to the spatial orientation of the cameras in the panoramic head. These errors of quite expected amplitudes are due to mechanical drilling and positioning errors and also to optical alignment variations relatively to the frame of the camera.

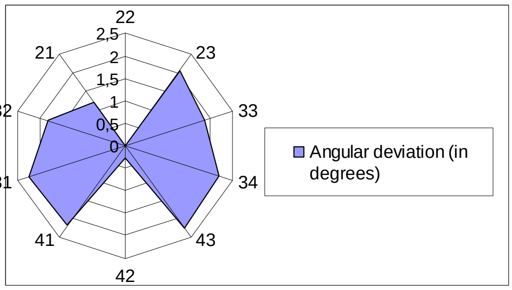

Figure 2: Angular comparison between the theoretical angles of the mechanical design and the angles estimated by our algorithm.

5.2. Comparison with an auto-calibration method of a market software

We have compared our results with an auto-calibration method of the software AutopanoPro [2]. This software can evaluate the intrinsic parameters at the same time as the extrinsic parameters. the distortion on all cameras whereas our solution computes only extrinsic parameters. Figure 3 shows the difference between the focal length and distorsion estimated with our process and with Autopano. Error is calculated on the corner, where there is the most deviation. Mean of errors is 1.02 degrees (i-e 14.5 pixels).

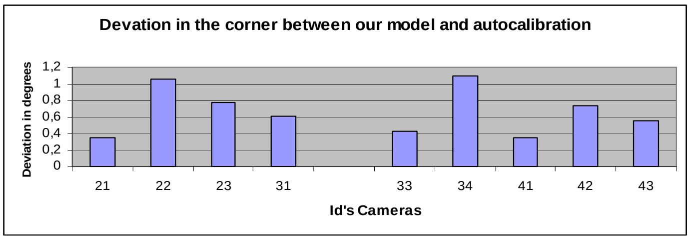

Figure 3: Deviation on corner between our calibration and auto-calibration

5.3. Some qualitative results

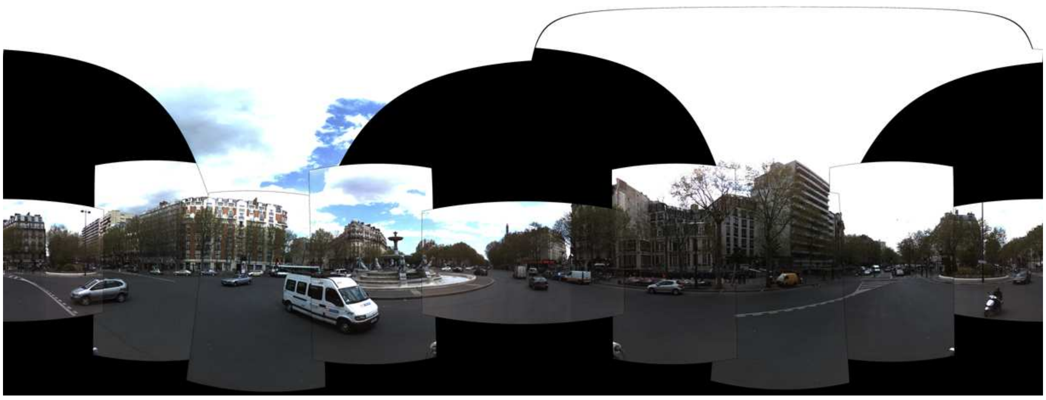

Figure 4: The result of our method

The figure 4 is the result of the compensation with the same perspective centre. At the top you can see the lake of symmetry in the panoramic head. But in all the panoramic scene you can observe the continuity of objects on all 360°.

Figure 5 presents some zooms on the panorama. It illustrates the precision of reconstruction but the image on the right illustrates the effects of parallax. Indeed if the tie points are taken on objects which are far away, the images are under-corrected for objects which are closer. This is particularly visible on the bus and on the car because, due to the mechanical design, the parallax for this couple of cameras is the larger as it can be seen on Fig.1.

Figure 5: Zoom on details

5.3.1. Basis estimation

Each camera within a panoramic head has not exactly the same perspective centre. The bases between the cameras are too small to compute a classical relative orientation between them without having degenerate solutions. Nevertheless, in our process, once the rotations have been determined, the equation of co-planarity can be used to estimate the stereo basis vector. But this computation permitted to have an estimation of the base. In our case, mean of deviation angular between theses estimations with our method and the true bases is 13.4 degrees on ten bases. In any case, the photogrammetric determination of the perspective centers will always be less accurate than the positions of the perspective centers provided by the mechanical design which will be at the millimetre level.

6. Conclusion and future Works

This paper has described a fairly simple bundle procedure to estimate the rotations of images acquired from calibrated cameras within a panoramic head. A fine estimation of these rotations is necessary to assemble the images seamlessly in order to avoid the blending of the images. In order to achieve a panorama with the highest accuracy, future work will take into account the depth of the scene provided by laser scanning devices to correct residual parallaxes to enhance the invisibility of the seamlines.

References:

[1] Craciun, D., Paparoditis, N., Schmitt, F., Automatic Gigapixel Mosaicing in Large Scale Unstructured Underground Environments, in IAPR Machine Vision and Applications, Tokyo, 2009.

[2] Autopano Pro, v. 1.4.2, Kolor, http://www.autopano.net/ last access march, 25 2009.

[3] Lowe, David.G. Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[4] Brown, M., Lowe, D.G., Automatic Panoramic Image Stitching using Invariant Features. International. Journal of Computer. Vision, vol. 74, no. 1, pp. 59-73, 2007.

[5] Coorg, S., Teller S. Spherical mosaics with quaternions and dense correlation. In International Journal of Computer Vision, vol. 37, pp. 259-273,2000.