Geo-Referencing Uncalibrated Photographs using Aerial Images and 3D Urban Models

O. Moslah[1,2], V. Guitteny[1], S. Couvet[1]

[1]THALES Security Solutions and Services, 1 Rue du General de Gaulle, 95523 Cergy-Pontoise, France.
[2]ETIS - UMR CNRS 8051, ENSEA, 6 Avenue du Ponceau, 95014 Cergy-Pontoise, France.
Email: oussama.moslah@thalesgroup.com

Abstract

We present methods and techniques for the geo-referencing of a set of uncalibrated photographs using aerial images and 3D urban models. We use structure and motion techniques to register accurately the set of uncalibrated photographs. The geo-referencing is then achieved either by using a semi-automatic registration with an aerial image or automatically using a 3D urban model.

Keywords: structure from motion, 3d plane fitting, absolute orientation.

1 Introduction

This paper addresses the problem of geo-referencing a set of uncalibrated photographs using either aerial images or existing urban 3d models. The context of this work is the growing interest in 2D/3D GIS-based services (Geographic Information System) in urban environnements. With the constant evolution of mobile hardware technologies and computer vision techniques many services and applications are now made possible.

1.1 Related work

Our work is mainly related to Structure From Motion (SFM) and Image-Based Modeling (IBM). Structure from motion techniques are able to automatically recover the sparse structure of a scene together with the motion of the camera using multiple view geometry techniques [17]. There are three main Structure From Motion methods : (1) Factorization based methods [5, 6] consist of an SVD decomposition of a matrix containing the images of points in all views to recover a projective structure and motion of the scene. The metric reconstruction is then obtained using self-calibration methods [17, 7, 8],(2) Trifocal tensor based methods [17] use image triplets to iteratively recover the structure and motion from images sequences, (3) Sequential methods [8, 14] use the motion computed from the fundamental matrix between a pair of images as initialization and then iteratively update the structure and motion by resection. PhotoTourism is a recent sequential structure and motion research work [13] able to calibrate a large set of digital photographs taken by different cameras. Image-based modeling allows for generation of realistic CAD models from a set of calibrated photographs either by a user interaction [2] or automatically by fitting and recognizing features [4, 3]. The first approach was the source of inspiration for commercial products able to produce high quality CAD models from a set of uncalibrated photographs such as Canoma [21] and ImageModeler [22]. This approach is robust but need a lot of user interactions to manually select and match features in photographs. The second approach allows for an automatic generation of semantic 3d models including some architectural components such as windows, doors and columns but the robustness is strongly dependent on the images.

2 Structure from motion

2.1 Camera model

In this paper we use the well known pinhole camera model that describes how a 3D point M with coordinates (X, Y, Z) in the world coordinate space projects into an image point m with coordinates (u,v) in pixels using the classic perspective transformation.

$$m \cong K\left[R^T| - R^T t\right] M = PM \tag{1}$$

Where R and t respectively represent the camera orientation and position and K the camera matrix or matrix of intrinsic parameters.

$$K = \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

Where s is the skew parameter, (c_x, c_y) are the pixel coordinates of the principal point, and f_x, f_y are focal lengths expressed in pixel-related units. The matrix of intrinsic parameters does not depend on the scene viewed and, once estimated, can be re-used as long as the focal length is fixed.

2.2 Matching keypoints

The first step of any structure and motion recovery technique consists in matching feature points between the set of photos. We held SIFT [12] as the solution to detection and description of keypoints. SIFT allows for the detection and description of points invariant to the change of scale, rotation, illumination and partially to the point of view. We use a SIFT K-d tree based implementation to detect and match keypoints between the pair of photos. To address the problem of estimating robustly the fundamental matrix we use a combined RANSAC [15] and M-Estimator [18] scheme. During the RANSAC iterations we use the bucketing technique described by Zhang [18] to improve the spatial distribution of the keypoints in the image. The matches between each pair of photos are then linked together into tracks.

2.3 Initial reconstruction

The SFM pipeline starts by initializing the structure and motion with a convenient pair of photos. The essential matrix is derived from the fundamental matrix and it represents the calibrated epipolar geometry between two views :

$$E = K_2^T F K_1 = [R_2(T_1 - T_2)]_x \, R_2 R_1^T \quad (3)$$

Where R_1, R_2, T_1 and T_2 are repectiveley the camera orientation and translation of the two cameras and $K1$, $K2$ respectively to their intrinsic matrices. The first camera is chosen so that it is aligned with the world coordinate frame [8] and the second camera is chosen to correspond to the relative camera motion (R,T) computed by an SVD decomposition of essential matrix :

$$E = K^T F K = [T]_x \, R \quad (4)$$

Among the four possible solutions obtained we choose the solution that give a positive depth for the reconstructed 3d points. We use the optimal triangulation method [16] to reconstruct the 3d points and initialize the structure of the scene.

2.4 Adding views

Then, we iteratively select a new camera and compute extrinsic parameters using the direct linear transform method [17] within a RANSAC scheme followed by an optimization of the reprojection error through gradient descent. The structure is updated by removing, adding or refining 3d points [8]. We refine the results through a local bundle adjustment which consists in finding the parameters of cameras and 3d points which minimize the reprojection error. So for m views and n tracks, we try to minimize the following criterion:

$$\min_{P_i, M_j} \sum_{i=1}^{m} \sum_{j=1}^{n} d(P_i M_j, m_j^i)^2 \quad (5)$$

We use the sparse bundle adjustment library of Lourakis and Agyros [19] based on the non linear minimization method of Levenberg-Marquardt to minimize this criterion. Figure 1 and Figure 2 show an example of a photo sequence and a sparse reconstruction generated with our SFM pipeline. Table 1 illustrates the computation times for different photo collections.

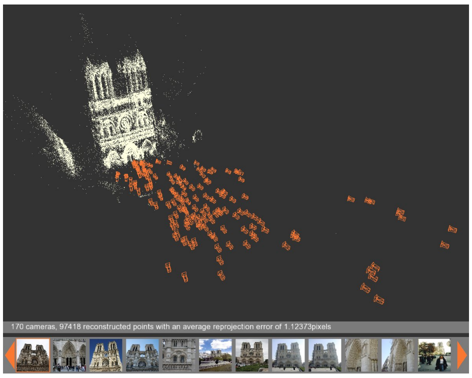


Figure 1: Structure and motion of the Notre Dame de Paris photo collection (170 cameras, 97418 points).

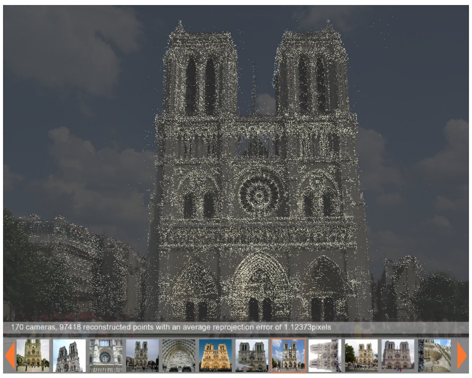


Figure 2: Visualisation of the sparse reconstruction of Notre Dame de Paris from a selected camera viewpoint.

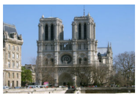
Name	Luxembourg 	Notre-Dame 	Arenberg 	Temple 	Triomphe 
Resolution	$1536 * 1024$	$1896 * 1350$	$768 * 576$	$2050 * 1543$	$1912 * 1440$
N	9	277	22	39	48
N_{cal}	9	170	22	29	25
$t_{detection\ SIFT}$	2 min	3 h	4 min	15 min	25 min
$t_{matching}$	3 min 15 s	4 days	3 min	45 min	2 h
$t_{FM\ estimation}$	15 s	2 h	10 s	1 min	3 min
$t_{linking}$	30 s	8 h	30 s	3 min	12 min
t_{SFM}	1 min	3 days	50 s	12 min	15 min
t_{total} (approx.)	7 min	8 days	9 min	2 h 11 min	2 h 55 min
n_{SIFT} (mean)	9457	11002	6100	8152	10050
$n_{points\ 3D}$	5773	97418	10086	10355	13606
Error	0.38 pixels	1.23 pixels	0.19 pixels	0.84 pixels	1.05 pixels

Table 1: Strcuture from motion computation times for different photo collections

3 Geo-Referencing

Geo-referencing a set of uncalibrated photographs can be formulated and solved as an absolute orientation problem [20]. The unknown parameters to fit are : the scale s (1 parameter), the 3d orientation R (3 parameters) and the 3d translation T (3 parameters). This process can be done either using a semi-automatic registration of the structure from motion reconstruction with a geo-referenced aerial image or automatically by fitting the 3d point cloud to a CAD model. Figures 3 and 4 show respectively the fitting of the Notre Dame de Paris structure and motion to an aeriel image and a CAD model.

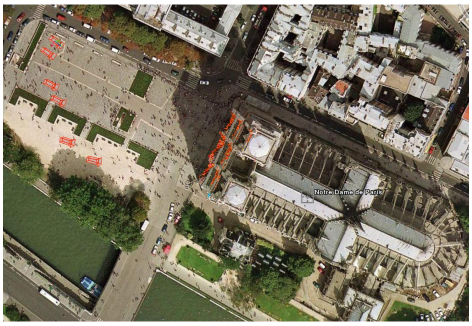


Figure 3: Structure and motion of the Notre Dame de Paris photo collection fitted to an aerial image.

4 Visualisation and rendering

We developed an openGL-based user interface for the visualization and rendering of the CAD model using projective texture mapping technique. The camera selected by the user is used to project texture onto the model. Figure 5 shows two different views for the rendering of a CAD model using our interface.

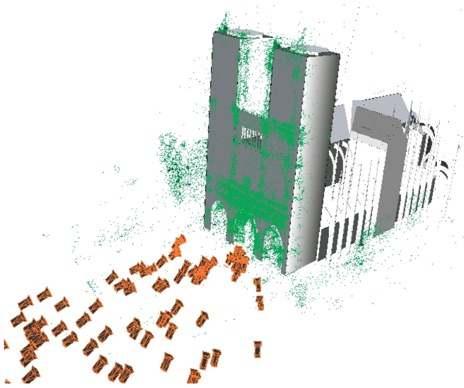


Figure 4: CAD model of the Notre Dame de Paris fitted to the computed structure and motion.

5 Conclusion

This paper introduces methods and techniques for recovering wall planes from a sparse reconstruction and for fitting CAD models to a set of uncalibrated photographs. These two elements form the major contribution of this paper. Our structure and motion pipeline is mainly inspired from state of the art techniques but involved some contribution to improve their robustness.

6 Acknowledgements

We wish to acknowledge the Cap Digital Business Cluster Terra Numerica project for sponsoring the research reported in this paper.

References

[1] Paul E. Debevec, George Borshukov and Yizhou Yu. Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping, In *9th Eurographics Rendering Workshop,* Vienna, Austria, June 1998.

[2] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik, Modeling and Rendering Architecture from Photographs, In *SIGGRAPH,* August 1996.

[3] Dick, A., Torr, R., Ruffle, S., and Cipolla, R., Modeling and interpretation of architecture from several images, In *IJCV,* 2003.

[4] Tomas Werner and Andrew Zisserman, New Techniques fro Automated Architectural Reconstruction from Photographs, In *ECCV,* 2002.

[5] Peter Sturm and Bill Triggs, A Factorization Based Algorithm for Multi-Image Projective Structure and Motion, In *ECCV,* 1996.

[6] Bill Triggs, Factorization Methods for Projective Structure and Motion, In *CVPR,* San Francisco, June, 1996

[7] J. Ponce, T. Papadopoulo, M. Teillaud and B. Triggs. On the Absolute Quadric Complex and its Application to Autocalibration, In *CVPR,* 2005.

[8] Marc Pollefeys, Visual Modeling With A Hand-Held Camera, *IJCV,* 59(3), 207-232, 2004.

[9] Lukas Zebedin, Andreas Klaus, Barbara Gruber and Konrad Karner, Facade Reconstruction from Aerial Images by Multi-View Plane Sweeping, In *ISPRS,* 2006.

[10] Florent Lafarge, Xavier Descombes, Josiane Zerubia, An automatic building reconstruction method: a structural approach using high resolution satellite images, In *ICIP,* 2006.

[11] Susanne Becker and Norbert Haala, Refinement of building fassades by integrated processing of LIDAR and image data, In *PIA,* 2007.

[12] David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *IJCV,* 2003.

[13] Snavely N., Seitz S. M. and Szeliski R. Photo tourism : Exploring photo collections in 3D. In *ACM Transactions on Graphics (SIGGRAPH Proceedings),* 25(3),835-846, 2006.

[14] Beardsley P.A., Zisserman A. Muray D.W. , Sequential Updating of Projective and Affine Structure from Motion, *IJCV*, 23(3), 235-259, 1997.

[15] M.A. Fischler et R.C. Bolles, Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography, In *Communication Association and Computing Machine,* Vol. 24 No. 6, pp. 381-395, 1981.

[16] Richard I. Hartley et Peter Sturm, Triangulation, In *Computer Vision and Image Understanding,* Vol. 62, No. 2, pp. 146-157, Novembre 1997.

[17] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in computer vision.* Cambridge University Press, Second Edition, 2003.

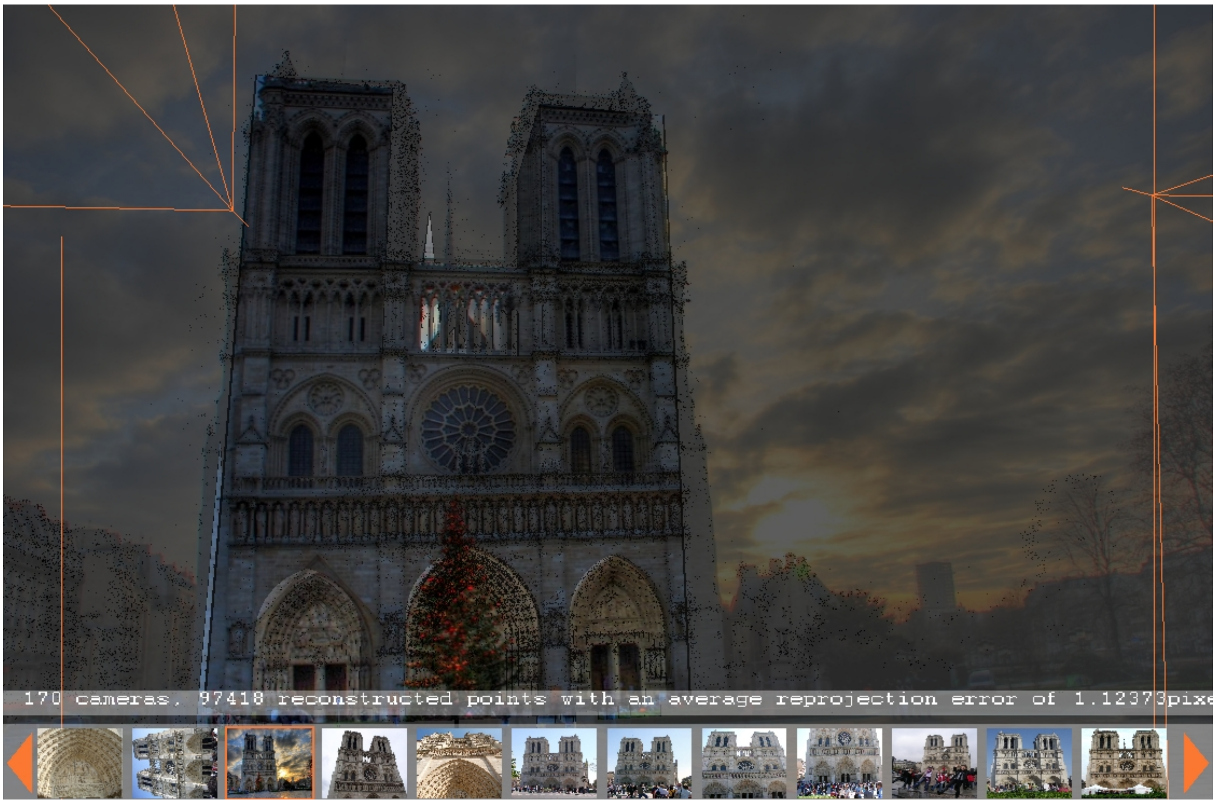
[18] Zhengyou Zhang, Determining the Epipolar Geometry and its Uncertainty: A Review, INRIA Research report n 2927, Juillet 1996.

[19] M.I.A. Lourakis and A.A. Argyros, The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm, ICS/FORTH Technical Report No. 340, 2004.

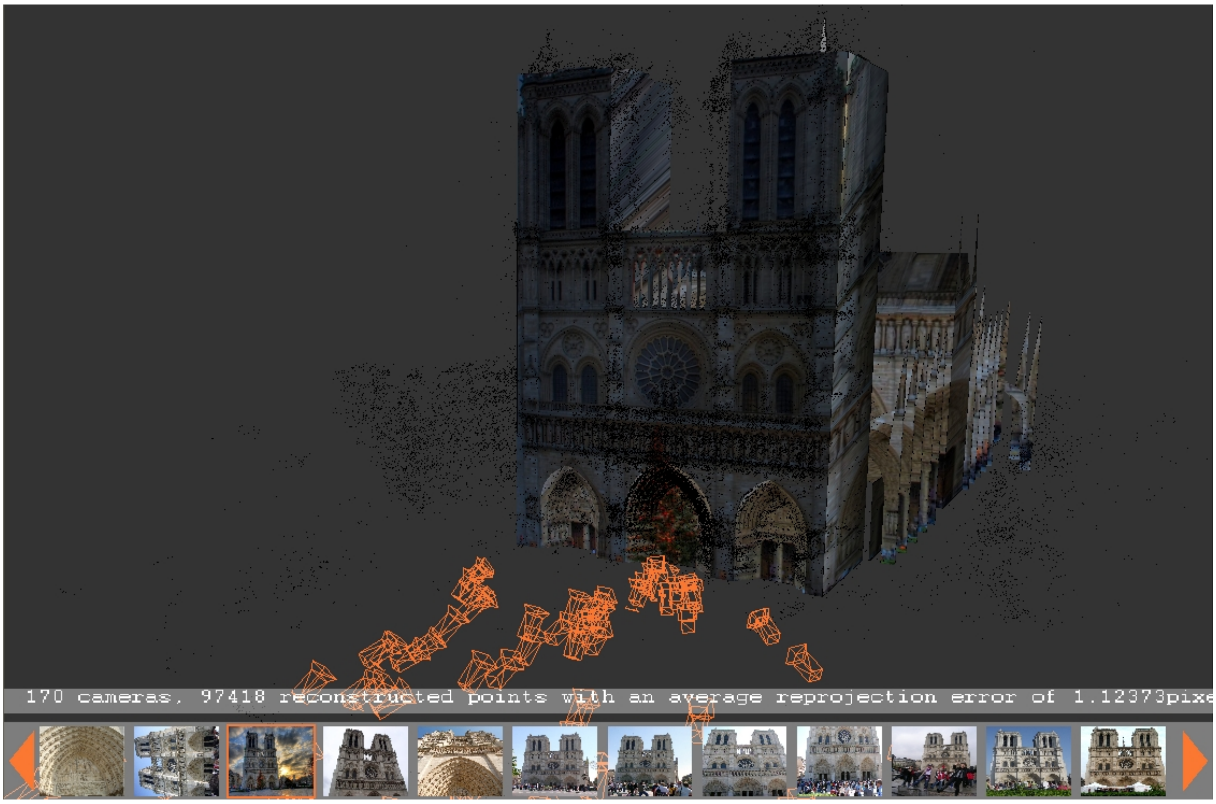
[20] O. Moslah, M. Klee, A. Grolleau, V. Guitteny, S. Couvet, and S. Philipp-Foliguet, Urban Models Texturing from Un-calibrated Photographs, 23rd International Conference Image and Vision Computing New Zealand, IVCNZ, 2008.

[21] Canoma, Software for fast creation of photorealistic 3D models from one or more photographs. http://www.canoma.com, July 2008.

[22] ImageModeler, Photorealistic 3D modeling software. http://imagemodeler.realviz.com/, July 2008.



(a) Rendering from a selected camera viewpoint.



(b) Rendering from an arbitrary viewpoint.

Figure 5: A photorealistic rendering of the CAD model using projective texture mapping.