

AUTOMATIC TEXTURE MAPPING WITH AN OMNIDIRECTIONAL CAMERA MOUNTED ON A VEHICLE TOWARDS LARGE SCALE 3D CITY MODELS

Fei Deng, Duanshun Li, Li Yan, Hao Fan

School of Geodesy and Geomatics, Wuhan University, 430079, Wuhan, China –
(fdeng, d.shun.lee, liyan, fhao)@sgg.whu.edu.cn

KEY WORDS: Extended Kalman filter, Structure from Motion, Texture mapping, Panoramic images

ABSTRACT:

Today high resolution panoramic images with competitive quality have been widely used for rendering in some commercial systems. However the potential applications such as mapping, augmented reality and modelling which need accurate orientation information are still poorly studied. Urban models can be quickly obtained from aerial images or LIDAR, however with limited quality or efficiency due to low resolution textures and manual texture mapping work flow. We combine an Extended Kalman Filter (EKF) with the traditional Structure from Motion (SfM) method without any prior information based on a general camera model which can handle various kinds of omnidirectional and other kind of single perspective image sequences even with unconnected or weakly connected frames. The orientation results is then applied to mapping the textures from panoramas to the existing building models obtained from aerial photogrammetry. It turns out to largely improve the quality of the models and the efficiency of the modelling procedure.

1. INTRODUCTION

The emergence of panoramas can date back to the panoramic paintings by Robert Barker even before the invention of camera. A short history of panoramas can be found in Thomas's work [1]. During the past hundreds of years, various panorama cameras have been built with the invention of smaller and more flexible cameras based on line scanning or stitching techniques. Large amounts of panoramas distributed all over the world have been collected through panorama sensors mounted on vehicles. Although panoramic images have been widely used for rendering in image-based immersive systems such as QuickTime VR[2], Google street view (Vincent, 2007), Bing streetside and streetslide[3], and etc., the applications of the panoramas have not been limited due to the complexity of the imaging model and the difficulty of accurate orientating, compared to ordinary digital images which have played a significant role in scene reconstruction. Panorama-based reconstruction and modelling also have been studied in several researches. On the one hand, panorama-based method is inferior to conventional methods such as aerial photogrammetry and LiDAR concerning the construction of large-scale 3D city; on the other hand, high resolution panoramas enable the extraction of high quality texture for rendering which is still unattainable through conventional methods, therefore panoramas can be combined with aerial images and LiDAR to make better results.

The geometric model of various panorama cameras has been widely studied. Single perspective camera can be simply modelled as the pinhole model with projective geometry (Hartley,2004), while due to the large distortion, the geometry of fisheye lens cameras has to be treated separately (Brauer, 2001; Schwalbe, 2005, Ying, 2006). Baker (1998) formulated the imaging model of catadioptric cameras with different types of mirror, and many researches (Kang, 2000; Micusik, 2003; Mei, 2004; Scaramuzza, 2009;) have been done for calibration

and the epipolar geometry of those omnidirectional images. The geometry of polycentric cameras (Tang, 2001; Huang, 2001; Shum,2004) and stitching-based multi-camera rigs(Bakstein, 2004; Szeliski,2006) have also been studied. At the same time, numerous general camera models (Seitz, 2001; Yu, 2004; Sturm, 2005; Ponce,2009) have been introduced during the last decades. Geyer (2001) proposed a unified theory for centric panoramic system with a projective mapping from the sphere to a plane with a projection centre on the perpendicular to the plane.

The SfM techniques have been widely used in both photogrammetry and computer vision society, including early fraction-based methods, the commonly used bundle adjustment (BA) based methods and filters based methods. These techniques emphasise on accuracy and speed respectively. Before Snavely (2008) introduced the Skeletal graphs based methods, most of the BA based methods can only handle image sets with limited amounts. Key frame based methods (Klein, 2007; Klein, 2008) only consider a subset of the image sets, while local optimization techniques propagating information forwards and backward with a window of frames through filters (Andrew, 2003; Hauke, 2010) or local bundle adjustment (Nister, 2005; Pollefeys, 2008) make full use of the continuity of the sequential images; however the local techniques usually suffer from shifts of the camera trajectory and both of them introduce loss in accuracy especially when double-backs or loops exist. For panorama sequences, revisiting of the existing features is considerably common, so a full bundle adjustment is necessary for accurate pose estimation. Nevertheless, due to the large distortion of panoramic image and occlusions caused by pedestrian bridge and etc., unconnected frames and numerous miss matching of some frames are unavoidable and will defeat the pure bundle adjust method.

Accurately oriented panoramas can be used for reconstruction (Luhmann, 2004; Micusik, 2004; Hwang, 2006; Micusik, 2009). Debevec (1996) introduced the image based modelling (IBM) technique which can model building and texturing simultaneously through a view-dependent texture mapping method. Debevec's pioneering work Façade inspired the commercial development of IBM softwares such as Canoma, Google Sketchup, ImageModeler, PhotoModeler and Videotrace. Foreseeing the potential of panoramas, Autodesk Inc even extended the panorama-based IBM modules. However, usually we can only get the streetside vertical facades of the buildings due to the reachability of the vehicles. Fortunately the whole model of the buildings can be obtained from aerial images or LIDAR automatically or semi-automatically. Fusing panoramic images with other data sources tends to be a better choice. Panoramas can be fused with point clouds for rendering and species classification (Scheibe, 2004; Salemi, 2005; Haala, 2004;Schwalbe, 2005). Haala (2005) proposed a method for texture extracting and mapping method with panoramas however needs interactions.

Considering that the most commonly used cameras are stitching based cameras such as Ladybug and Eysis which can be treated as a single perspective panoramic camera, we introduced the bundle-based projection model for calibrated cameras due to its simplicity and its linearity which makes it adaptive to the projective geometry of ordinary single perspective camera algebraically. We also adopted the Extended Kalman Filter (EKF) which is widely used in simultaneous localization and mapping (SLAM) in the SfM procedure to "jump" out of the snare of the unconnected frame through smoothing.

2. GENERAL GEOMETRIC MODEL

2.1 Imaging model

Generally, all of the images with a large field of view in one direction or more can be regarded as panoramic images. During the last century, various kinds of panoramic sensors have been invented. According to the imaging systems, most panoramic sensors can be catalogued into fisheye lens cameras, catadioptric cameras, polycentric cameras, line-scanning cameras and stitching-based cameras. Fisheye lens cameras are the simplest device though with limited field of view and low quality. Catadioptric cameras suffer from the same problem. Polycentric cameras are cheap and can be easily constructed but the projection geometry is quite complicated due to multiple perspective centres. Line-scanning cameras get rid of the problems stated above, but its usage is limited to still scene because of low imaging rate. Due to the smart size, flexibility, high quality and imaging rate, stitching-based cameras became the most commonly used panoramic sensor. Such sensors have been mounted on unmanned vehicles and various mobile mapping systems.

For standard single perspective cameras, all scene points X on the same line of sight project to single image point \mathbf{x} which can be formulated with projective geometry as

$$\lambda \mathbf{x} = \mathbf{P}X, (\lambda \neq 0) \quad (1)$$

where λ is a nonzero scalar value, \mathbf{x}_w and \mathbf{x}_i are homogeneous coordinates of 3D object points and 2D image points. The 3×4

matrix \mathbf{P} is the projection matrix formulated with the camera matrix and camera pose. Such a representation allows the whole line passing through the perspective centre project to a single point on the image plane which will not cause any problem for ordinary digital images. However, for panoramic image with an omnidirectional field of view, any line passing through the perspective centre should have two images corresponding to the two object points the line reaches. Thus the line of sight should be regarded as two rays instead of one line

$$\lambda \mathbf{x}_c = \mathbf{P}X, (\lambda > 0) \quad (2)$$

where \mathbf{x}_c is the Cartesian ray under the camera coordinates system. For panoramic images, the relationship between the image points and the ray cannot be simply represented with a linear function just like the camera matrix for standard perspective camera. The imaging model can be regarded as a pose recover process and an imaging process under Cartesian coordinates as

$$\lambda \mathbf{x}_c = \mathbf{R}X + \mathbf{t}, (\lambda > 0, \mathbf{t} = -\mathbf{R}\mathbf{c}) \quad (3)$$

$$\mathbf{x}_i = h(\mathbf{x}_c) \quad (4)$$

where g is the imaging function, \mathbf{R} and \mathbf{c} are rotation matrix and the perspective centre of the camera. All of the coordinate X , \mathbf{x}_c and \mathbf{x} are in Cartesian coordinates, unless otherwise specified, the rest of the paper will follow such a notation.

With a calibrated panoramic sensor, the ray corresponding to any image points can be recovered with g^{-1} . Then, the general bundle based imaging model can be simply represented by a linear formula (3). Considering that any ray in the 3D space has only 2 degree of freedom, we chose the longitude and latitude parameterization of such rays, and the imaging function should be

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = h(\mathbf{x}_c) \begin{pmatrix} \arctan(y_c, x_c) \\ \arctan(z_c, \sqrt{x_c^2 + y_c^2}) \end{pmatrix} \quad (5)$$

The imaging model can be represented as

$$\mathbf{x}_i^j = f(C_i, X_j) \quad (6)$$

Where C_i is the camera parameters of camera i , X_j is the objective Cartesian coordinates of feature j , and \mathbf{x}_i^j is the image of feature j on the parameterized panoramic image i .

The camera parameters including the perspective centre of the camera and the direction of the camera, are of 3 degree of freedom. Considering the stability, we parameterize the direction of the camera with the vector part $\mathbf{p}=(q_1, q_2, q_3)^T$ of a unit quaternion $\mathbf{q}=(q_0, q_1, q_2, q_3)^T$ during bundle adjustment instead of the traditional Euler angles. Thus we can get

$$C_i = (\mathbf{c}_i^T, \mathbf{p}_i^T)^T = (c_x, c_y, c_z, q_1, q_2, q_3)^T \quad (7)$$

2.2 Epipolar geometry

According to epipolar geometry, giving two correspondences image points and their homogeneous coordinates \mathbf{x}_i and \mathbf{x}_i' captured by two calibrated perspective camera, we have

$$\mathbf{x}_i'^T \mathbf{F} \mathbf{x}_i = 0 \quad (8)$$

Where \mathbf{F} is the rank $2 \ 3 \times 3$ fundamental matrix with 7 degree of freedom. Once the camera is calibrated, the rays \mathbf{x}_c and \mathbf{x}_c' can be recovered from the calibrated camera matrix, 2 freedom

related to the camera matrix can be eliminated from the fundamental matrix.

$$\mathbf{x}_c'^T \mathbf{E} \mathbf{x}_c = 0 \quad (9)$$

Where \mathbf{E} is the rank $2 \ 3 \times 3$ essential matrix with 5 degree of freedom. For corresponding rays in two panoramas, the epipolar geometry can directly adopt the essential matrix formulation and can be calculated through direct linear 8-point algorithm or more stable 5 points algorithms.

3. SFM WITH EKF

Automatic feature detection and matching are the most important part of the automation of many of the computer applications including SfM. Though lots of feature detectors and descriptors (Tuytelaars, 2007) have been invented, the problem of feature detection seems always a matter. Because of the large illumination difference in the open space, descriptors use the photometric information directly can hardly get an acceptable result. What's more, except from the vary of illumination and perspective distortion due to view point change and occlusion and etc., the anisotropic distortion caused by parameterizing of sphere panorama is the most troublesome problem for matching and tracking. For equirectangular projection the distortion varies sharply from the equator to the pole. Scale Invariant Feature Transform (SIFT) detector (Lowe, 2004) which is invariant to scale and rotation and partially invariant to illumination and affine transformation has been believed overmatching other detectors (Mikolajczyk, 2005). KTL (Tomasi, 1991) is one of the most commonly used for feature tracking of image sequences. During our research, we compared the performance of the KTL tracker and SIFT matching on several panoramic image sequences, the KTL tracker cannot provide sufficient matching results due to street trees and moving crowds and vehicles. By contrast SIFT matching always obtains better performances. To remove bad correspondences, we adopt the hypothesize-and-test architecture based on the epipolar geometry through RANSAC () and Nister's 5 points algorithm (Nister, 2005) on each pair of images with sufficient connecting coarse correspondences.

The SfM pipeline begins with the choice of initial image pairs. For SfM with unordered image collections the initial pair needs to be carefully chosen because the reconstruction of the initial pair may sticks in the local minimum or with limited numbers of triangulated points with sufficient accuracy. Because most of the objects in panoramic sequences are trees, road, vehicles and buildings, we want the initial pair full of textures coming from buildings which are much easier for feature detection and matching. Once the initial pair is chosen, Nister's 5 points algorithm followed with a non-linear optimization is applied for the estimation of the relative pose of the image pair. The initial scene is then triangulated from the correspondences with the initial pose. After that bundle adjustment which aims to minimize the following function is applied to the existing cameras and the triangulated parameters.

$$g(C, X) = \sum_{i=1}^n \sum_{j=1}^m (\mathbf{x}_i^j - h(C_i, X_j)) \quad (10)$$

Where

$$\begin{aligned} C &= (C_1^T, C_2^T, \dots, C_n^T)^T \\ X &= (X_1^T, X_2^T, \dots, X_n^T)^T \end{aligned} \quad (11)$$

For the task of automatic texture mapping with panorama sequences, there is no claim for real time reaction, we prefer to estimate the pose of every frame incrementally and run full bundle adjustment after every coarse initialization with the direct linear transformation (DLT) method. However, losing tracking due to the trees covering the street or even a single pedestrian bridge make the situation more complicated. What if a frame without sufficient correspondences or even unconnected exists? Filtering based SfM which needs less correspondences (Civera, 2009) and has been used in SLAM for long will be the solution.

We combined an EKF filter during our SfM process. Because we can do bundle adjustment after every initialization, we introduced an easy constructing EKF filter. To predict the camera pose we employ the instantaneous constant velocity assumption proposed in (Davison, 2007). The dynamic of the camera motion is described by a linear velocity $\mathbf{v}=(v_x, v_y, v_z)^T$ and an angular velocity $\mathbf{w}=(w_x, w_y, w_z)^T$ parameterized as the vector part of a unit quaternion. The state vector \mathbf{X} of the filter is the made up of the pose and dynamic parameters of the camera, thus

$$\mathbf{X}_i = (C_i^T, \mathbf{v}_i^T, \mathbf{w}_i^T) \quad (12)$$

By simply choosing the camera parameters after bundle adjustment as the measurements, the observation \mathbf{Z} of the filtering system is just the state vector itself. The transition model and the observation model will be

$$\begin{aligned} \mathbf{X}_{i,i-1} &= f(\mathbf{X}_{i-1}) + \mathbf{d}_x = \begin{pmatrix} \mathbf{c}_{i-1} + \mathbf{v}_{i-1} \\ q(\mathbf{p}_{i-1}) \odot q(\mathbf{w}_{i-1}) \\ \mathbf{v}_{i-1} \\ \mathbf{w}_{i-1} \end{pmatrix} + \mathbf{d}_x \\ \mathbf{P}_{i,i-1} &= \mathbf{F}_{i-1} \mathbf{P}_{i-1} \mathbf{F}_{i-1}^T + \mathbf{Q}_{i-1} \end{aligned} \quad (13)$$

And

$$\begin{aligned} \mathbf{Z}_i &= h(\mathbf{X}_{i,i-1}) + \mathbf{d}_z = \mathbf{X}_i + \mathbf{d}_z \\ \mathbf{S}_i &= \mathbf{H}_i \mathbf{P}_{i,i-1} \mathbf{H}_i^T + \mathbf{R}_i = \mathbf{P}_{i,i-1} + \mathbf{R}_i \end{aligned} \quad (14)$$

Where q is a function mapping the vector part to a unit quaternion, \odot is the multiplication of two quaternions, \mathbf{d}_x and \mathbf{d}_z are the process and observation noises which are both assumed to be zero mean multivariate Gaussian noises with covariance \mathbf{Q}_k and \mathbf{R}_k obtained from the bundle adjustment result accordingly. \mathbf{F}_i and \mathbf{H}_i are the Jacobians of function f and h respect to the updated state vector \mathbf{X}_{i-1} and the predicted state vector $\mathbf{X}_{i,i-1}$.

The update is performed with the standard EKF equations:

$$\begin{aligned} \mathbf{X}_i &= \mathbf{X}_{i,i-1} + \mathbf{K}_i (\mathbf{X}_i - \mathbf{X}_{i,i-1}) \\ \mathbf{P}_k &= (\mathbf{I} - \mathbf{K}_i) \mathbf{P}_{k,k-1} \end{aligned} \quad (15)$$

where

$$\mathbf{K}_i = \mathbf{P}_{i,i-1} \mathbf{S}_i^{-1}$$

The SfM module and the filter are relatively parallel with only data exchange. the SfM module provides the bundle

adjustment results as the measurements of the filter (see figure 1). The filter will not intervene in the SfM process except the unconnected or weakly connected frame occurs. Once it happens, the filter predicts an estimate of the initial pose of the camera with the information accumulated during the long run. The accuracy of the initial estimate though coarse but is sufficient for the initialization of bundle adjustment.

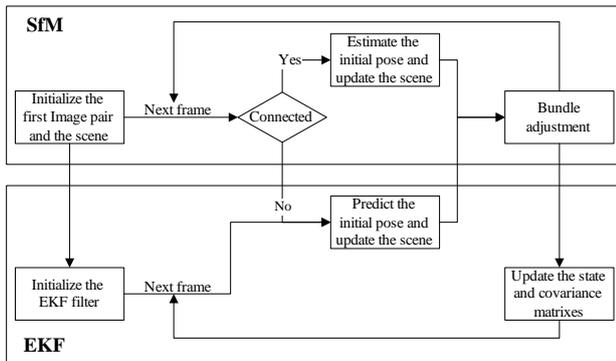


Figure 1. Pipeline of the SfM algorithm with an EKF filter

4. AUTOMATIC TEXTURE MAPPING

Previous 3D building models were constructed manually, while nowadays it can be automated or semi-automated with airborne imagery, ground imagery airborne LiDAR and ground LiDAR (Musialski, 2012). For large 3D city modelling, the state-of-the-art aerial image based automatic urban reconstruction method (Zebedin, 2008) is believed to be the best solution. Though the building can be reconstructed and textured properly, the detail of the facades of the buildings will lose due to the limitation of the view point of airborne images.

Most of the texture mapping methods adopted the view-dependent methods introduced by Debevec (1996). We follow such a method, and because of the omnidirectional coverage of the panoramic images, we only have to concern the viewpoint without considering the orientation of the standard digital images. For simplicity, we chose the nearest panorama as the sources for the facades of the buildings. To get the image patches for texture mapping we project the points of the facade onto the equirectangular panoramic image according to the imaging model, and resample the largely distorted image patch (see figure 2) onto a plane parallel to the facade.



Figure 2. Largely distorted equirectangular panoramic image

We tested our method with a Pointgrey Ladybug 2 camera mounted on a vehicle. The panoramic image sequences are captured along the Weijing Street located in Tianjing where the buildings along the street have been reconstructed accurately through aerial photogrammetry (see figure 3). The

pose of the panoramic image sequences are accurately estimated from the combined method proposed above following with a registration between the triangulated points and the 3D digital models. Finally, we apply the automatic texture mapping algorithm on the data sets (see figure 4 and figure 5).



Figure 3. 3D model reconstructed through aerial photogrammetry



Figure 4. Buildings with texture from panoramic images



Figure 5. Detail of the texture extracted from panoramas

5. CONCLUSION

3D building models reconstructed from aerial images or point clouds can be automated or semi-automated with sufficient accuracy. While for immersive virtual reality applications, the models should have rich textures. Nowadays, high resolution panoramic image sequences with rich texture information can be easily obtained through state-of-the-art multi-camera rigs. Within the paper, we aims to automatic texture mapping with panoramic image sequences for existing 3D models and address the key problem as the accurate pose estimation of the image sequences.

Due to moving objects and occlusions in the image sequence, even the continuative pair may be unconnected or weakly connected without sufficient correspondences. We combined an

EKF filter with the traditional SfM procedure. Through the intervention of the filter by providing camera pose predictions the unconnected frames are “skipped” from the initialization process of SfM. The drawback of the method is that it cannot handle such sequences with long continuous unconnected frame segments. Finally, we apply the accurately oriented panoramas for automatic texture mapping with the view-dependent method.

6. REFERENCES

- Luhmann, T., 2004. A historical review on panorama imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Dresden Vol. XXXIV, Part 5/W16.
- E. Schwalbe, 2005. Geometric modeling and calibration of fisheye lens camera systems, Proc. ISPRS.
- Xianghua Ying and Hongbin Zha, 2006. Using Sphere Images for Calibrating Fisheye Cameras under the Unified imaging Model of the Central Catadioptric and Fisheye Cameras. ICCV 2006, pp.539 – 542.
- Micusik, B and Pajdla, T, 2003. Estimation of omnidirectional camera model from epipolar geometry. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.I-485-I-490.
- S. B. Kang, 2000. Catadioptric self-calibration. In Proc. CVPR, pp.I:201–I:207.
- C. Mei and P. Rives, 2007. Single view point omnidirectional camera calibration from planar grids. In IEEE International Conference on Robotics and Automation. pp. 3945–3950.
- Tang, WK and Hung, YS, 2001. Concentric Panorama Geometry. 8th IEEE Conference on Mechatronics and Machine Vision in Practice, pp.507-512.
- F. Huang, S.K. Wei, and R. Klette, 2001. Geometrical Fundamentals of Polycentric Panoramas, ICCV 2001, pp.560-565.
- Backstein, H.; Pajdla, T. (2005): Calibration procedure for a 360×360 mosaic camera. Proceedings of 2nd Panoramic Photogrammetry Workshop. ISPRS, Vol. XXXVI, Part 5/W8.
- D. Scaramuzza, F. Fraundorfer, M. Pollefeys, and R. Siegwart, 2009. Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints. ICCV 2009.
- Y. Li et al., 2004. Stereo reconstruction from multi perspective panoramas, PAMI 2004, 26(1), pp.44–62.
- Richard Szeliski, 2006. Image alignment and stitching: a tutorial. Foundations and Trends® in Computer Graphics and Vision. 2006,2(1).
- S.M. Seitz, 2001. The Space of All Stereo Images. Proc. Int'l Conf. Computer Vision, pp.26–33.
- Yu and L. McMillan, 2004. General linear cameras. In Proc. ECCV.
- P. Sturm, 2005. Multi-view geometry for general camera models. CVPR 2005, pp.206–212.
- Ponce, 2009. what is a camera. J. CVPR 2009, pp.1526-1533.
- CHEN, S. E. 1995. QuickTime VR – an image-based approach to virtual environment navigation. Computer Graphics (SIGGRAPH'95)(August), pp. 29–38.
- Vincent, L, 2007. Taking online maps down to street level. Computer, 40(12), pp.118-120.
- Johannes Kopf, Billy Chen, Richard Szeliski, Michael Cohen, 2010. Street Slide: Browsing Street Level Imagery. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2010), 29(4), article no. 96.
- Snively, N.; Seitz, S.M.; Szeliski, R, 2008. Skeletal graphs for efficient structure from motion. ICCV 2008, pp.1–8.
- Klein, G.; Murray, D, 2007. Parallel Tracking and Mapping for Small AR Workspaces. 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp.225 - 234
- G. Klein and D. W. Murray, 2008. Improving the agility of key frame-based SLAM. ECCV, 2008.
- Andrew J. Davison, 2003. Real-Time Simultaneous Localisation and Mapping with a Single Camera. ICCV 2003, pp.1403-1410.
- Hauke Strasdat, J. M. M. Montiel, Andrew J. Davison, 2010. Real-time monocular SLAM: Why filter? ICRA 2010, pp.2657-2664
- David Nister, 2005. Preemptive RANSAC for live structure and motion estimation. Mach. Vis. Appl, 16(5), pp.321–329.
- Micusik B., Martinec D. and Pajdla T, 2004. 3D Metric Reconstruction from Uncalibrated Omnidirectional Images, ACCV 2004.
- Yong Ho Hwang, Jaeman Lee, Hyun-Ki Hong, 2006. Omnidirectional Camera Calibration and 3D Reconstruction by Contour Matching. ISVC 2006.
- Micusik, B.; Kosecka, J, 2009. Piecewise Planar City 3D Modeling from Street View Panoramic Sequences. CVPR 2009, pp.2906 – 2912.
- Luhmann, T. & Tecklenburg, W., 2004. 3-D object reconstruction from multiple-station panorama imagery. ISPRS Archives. Vol. XXXIV, 5/W16.
- Karsten Scheibe, Martin Scheele, and Reinhard Klette, 2004. Data fusion and visualization of panoramic images and laser scans. ISPRS Archives. Vol. XXXIV, 5/W16.
- G.Salemi, V.Achilli, C.Cervato, 2005. Data acquisition for cultural heritage navigation integration of panoramic imaging

terrestrial laser scanning and anaglyphs. ISPRS Archives. Vol. 36, Part 5/W1.

N. Haala, R. Reulke, M. Thies, and T. Aschoff, 2004. Combination of terrestrial laser scanning with high resolution panoramic images for investigations in forest applications and tree species, ISPRS Archives. Vol. XXXIV, 5/W16.

N. Haala, and M. Kada, 2005. Panoramic scenes for texture mapping of 3-D city models. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Berlin, Germany, Vol. 36, Part 5/W8.

Tinne Tuytelaars, Krystian Mikolajczyk, 2007. Local Invariant Feature Detectors: A Survey, Computer Graphics and Vision, 3(3).

J. Davison, I. D. Reid, N. Molton, and O. Stasse, 2007. MonoSLAM:Real-time single camera SLAM. PAMI 2007, pp.1052–1067.

D. Lowe, 2004. Distinctive image features from scale-invariant keypoints. IJCV, 60(2), pp.91–110.

Carlo Tomasi and Takeo Kanade, 1991. Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132.

Nister, D., 2005. A Minimal Solution for Relative Pose with Unknown Focal Length, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.789–794.

Paul E. Debevec, Camillo J. Taylor, Jitendra Malik, 1996. Modeling and rendering architecture from photographs: a hybrid geometry and image-based approach. SIGGRAPH '96.

K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, 2005. A Comparison of Affine Region Detectors. International Journal of Computer Vision, 65(1):43–72.

Przemyslaw Musialski and Peter Wonka and Daniel G. Aliaga and Michael Wimmer and Luc van Gool and Werner Purgathofer, 2012. A Survey of Urban Reconstruction. EUROGRAPHICS 2012 State of the Art Reports.

Zebedin L., Bauer J., Karner K., BISCHOF H., 2008. Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery. ECCV 2008, pp. 873–886.