THE STUDY ON 3D MODEL RECONSTRUCTION SOFTWARE ON THE INTERNET— A CASE STUDY OF PHOTOSYNTH AND PHOTOFLY

Jie-Shi Weng[1], Yi-Ting Tsai[2], and Jin-Tsong Hwang[3]
[1]Graduate Student, Dept. of Real Estate and built Environment, National Taipei University,
No.,151, University Rd., San Shia, New Taipei City, Taiwan; Tel: +886-2-26748189#67426;
E-mail: jessie8031@hotmail.com

[2]Graduate Student, Dept. of Real Estate and built Environment, National Taipei University,
No.,151, University Rd., San Shia, New Taipei City, Taiwan; Tel: +886-2-26748189#67426;
E-mail: tsai_ting@hotmail.com

[3]Associate Professor, Dept. of Real Estate and built Environment, National Taipei University,
No.,151, University Rd., San Shia, New Taipei City, Taiwan; Tel: +886-2-26748189#67426;
E-mail: jthwang@mail.ntpu.edu.tw

KEY WORDS: Point Clouds, 3D Model, Reconstruction, LiDar

ABSTRACT: In recent years, many commercial 3D modeling software combined with cloud computing have been released. People can use these free software for getting point clouds data and building 3D models quickly with digital camera. The most commonly used software available on the Internet for 3D model reconstruction were Photosynth (Microsoft) and Project Photofly (Autodesk). The advantages of these two software include convenience to use and ease to operate. Users just need to take digital photos and upload them onto the website on the Internet. Point clouds data could then be generated for users to download point clouds from the website. After point clouds data editing, 3D models would be easily reconstructed. However, there may be insufficient point clouds generated for 3D model reconstruction caused by shooting angle, number of photos taken, and other factors. This study compares the positioning accuracy of point clouds by adjusting factors such as shooting rotation, photo resolution, and brightness of image.

1. INTRODUCTION

One of the goals in the field of geoinfomatics is technological development in 3D model construction. There are many approaches to 3D model construction, which include close-range photogrammetry, photogrammetry, and LiDar. Cheng (2011) adopted Plane map of Building with height attribute to construct 3D model of buildings. From the perspective of operational convenience, the approach of 3D model construction by close-range photogrammetry is the fastest. All that is required is for the digital camera to be first calibrated to take pictures of the building and 3D reconstruction can then begin.

Photofly technology was acquired on May 2008 from Realviz. After a few years of research and development conducted by AutodeskLab, Photofly was released on July 22, 2010. which was released on August 20, 2008, is a software application developed by Microsoft Live Labs and the University of Washington. These two software analyze digital photographs and generate a 3D model of the photos and a point clouds of a photographed object. They are using scale-invariant feature transform (SIFT) and Structure from Motion (SfM) techniques integrated with close-range photogrammetry. In this paper, the position accuracy assessment of point clouds generated by Photosynth and Photofly will be compared with that estimated using close-range photogrammetry and Lidar.

2. Literature review

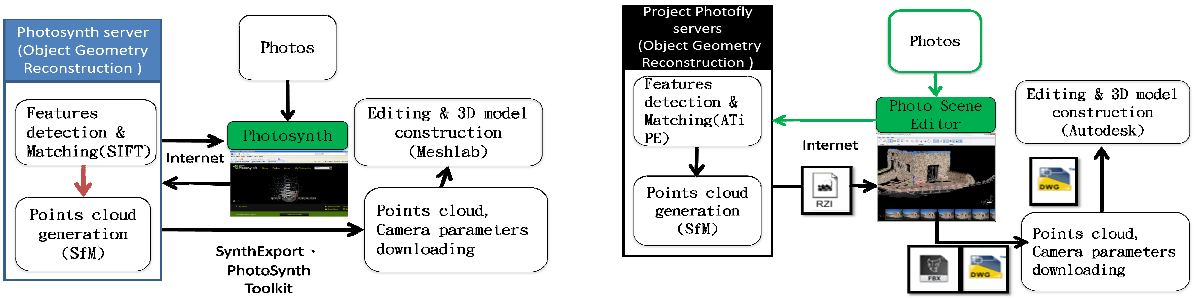
In 2008, Microsoft Photosynth allowed the public to upload their own photos to create a 3D panoramic image. To date, studies on positioning accuracy assessment with these two software are few, so the literature review will focus on the procedures of point clouds generation. The theory of Photosynth and Photofly comprised both SIFT and SfM. The SIFT algorithm is used for feature extraction and image matching, and SfM is employed to restore camera motion parameters; and above all, SIFT and SfM serve to find coordinates of the shooting object, and then construct 3D models.

David Lowe first proposed the scale-invariant feature transform algorithm in 1999, a computer vision algorithm about detecting local image features. SIFT algorithm has been employed to perform image stitching in 2003, as well as to find feature points and automatic panoramic image stitching in 2006. SIFT algorithm can capture the local feature of an image, and is robust and invariant for spatial scale, rotation angle and brightness of image. It is widely used in image recognition, image matching, and 3D model construction. Huang (2009) used SIFT algorithm to recognize human faces. Huang (2009) applied SIFT algorithm in the study of stitching and matching images. Chang (2008) and Wu et al. (2009) used the SIFT algorithm for feature matching and obtained good results. Chen (2008) used the SIFT algorithm in automatic aerial triangulation to obtain more feature points, and estimate accuracy and reliability by space intersection. The result shows that it can reduce the rate of failure caused by the adjustment of scale, rotation, and brightness of image. SfM was developed in the 1980s. Its main purpose is to calculate the correlation between feature points in the image, estimate the camera position and shooting angle, restore camera motion parameters and build the coordinates of the object in 3D modeling by a continuous image. Snavely et al. (2008) proposed how to make use of photos on the Internet for 3D scene reconstruction or visualization.

Project Photofly is the free software which combines close-range photogrammetry with cloud computing by Autodesk Labs, providing users a way to address rapidly 3D images. Users need to download first the “Photo Scene Editor”, and then upload pictures to the cloud database by the scene editor, and down loading from cloud database after computing the point clouds. Finally, the follow-up editing process can be completed using the scene editor (Abate et al., 2010). Abate et al. (2010) compared 3D models constructed by Photofly using photos of open squares, independent buildings as well as building elements and sculptures taken by different kinds of camera. The results showed that all the objects can be constrained rapidly by tie points among images, but failure occurs easily when facing large or complex data. Current studies focus mainly on how to build 3D models and exploring the strengths and weaknesses of Photofly. This article aims to analyze positioning accuracy of 3D models. The comparison between Photosynth and Photofly is summarized in Table 1. The flow chart of Photosynth and Photofly are shown in Figs. 1(a) and (b), respectively.

Table 1. Comparison between Photosynth and Photofly.

Items	Photosyhth	Photofly
Version(2011.08.14)	2.110.317.1042	2.1
Service	Cloud computing	Cloud computing
Photo format	.jpg	.jpg, .tif
Algorithm	Feature Extraction: SIFT, and SfM	Feature Extraction: (Automatic Tie Point Extraction, ATiPE) , and SfM
Point clouds downloading	No (have to use SynthExport or PhotoSynthToolkit)	yes
Export format	obj, ply, vrml, x3d	DWG（only exports the manual points and lines you create in Phtofly）, las, ipm, rzi, obj
dense clouds	No(have to use PMVS)	Yes (provided Mobile, standard, and Maximum output quality)
Speed	fast	Slow (depend on number of photos)
Manual Photo stitch	No	Yes
Setting coord. system	No (Local coord. system)	Yes (WCS and Reference distance)
Camera prarmeters	No (have to down loading by SynthExport or PhotoSynthToolkit)	Yes (includes in .rzi file)
3D Model reconstruction	No(have to use the other editing software, for example, meshlab)	No (Export to DWG combined with .las format can be editing on Autocad 2010 or we can use the AutoCAD import FBX to get the mesh into)
Animation	No	Yes (Generating the .avi format animation file by user defined the path of observation)


Photosynth server (Object Geometry Reconstruction)
Features detection & Matching(SIFT)
Points cloud generation (SfM)
Photos
Photosynth
Internet
Editing & 3D model construction (Meshlab)
Points cloud, Camera paramters downloading
SynthExport、PhotoSynth Toolkit
Project Photofly servers (Object Geometry Reconstruction)
Features detection & Matching(ATiPE)
Points cloud generation (SfM)
Photos
Photo Scene Editor
Internet
RZI
Editing & 3D model construction (Autodesk)
DWG
FBX
DWG
Points cloud, Camera parameters downloading

(a) (b)

Figure 1 Flow chart of (a)Photosynth and (b)Photofly.

3. METHODLOGY

Photosynth and Photofly proceed with Image-based Modeling through the reconstruction of spatial geometry by searching for feature points to match and stitch images, restore the camera position, inquire the coordinates of shooting subject, and navigation systems to provide users browse their own image location. The spatial geometry reconstruction process is divided into three steps: feature points extraction, feature points matching, and restoring the camera position. The first two steps are based on SIFT algorithm, the last step is based on SfM.

3.1 SIFT

SIFT is a computer vision algorithms used to describe and search the local features of image. Local features means the locations have larger and more significantly different of gray value in the neighbor pixel, such as edge and corner. The SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint.

The first step is to detect scale-space extrema for feature extration. This is the stage where the interest points, which are called keypoints in the SIFT framework, are detected. For this, the image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images are taken. Keypoints are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales. The scale space of an image is defined as a function L (x, y, σ) that is produced from the convolution of a variable-scale Gaussian G (x, y, σ) with an input image I(x, y), shown as equation(1):

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where * is the convolution operation in x and y, and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp(-(x^2 + y^2)/2\sigma^2) \tag{2}$$

To efficiently detect stable keypoint locations in scale space, we have proposed (Lowe, 1999), using scale-space extrema in the Difference-of-Gaussian function convolved with the image, D (x, y,σ), which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k:

$$\begin{aligned} D(x, y, \sigma) &= \big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{3}$$

An efficient approach to construction of D (x, y, σ) is shown in Figure 1. The initial image is incrementally convolved with Gaussians to produce images separated by a constant factor k in scale space, shown stacked in the left column. We choose to divide each octave of scale space into an integer number, s, of intervals, so $k = 2^{1/s}$. We must produce s + 3 images in the stack of blurred images for each octave, so that final extrema detection covers a complete octave. Adjacent image scales are subtracted to produce the Difference-of-Gaussian images shown on the right. In order to detect the local maxima and minima of D(x, y, σ), each sample point is compared to its eight neighbors in the current image and nine neighbors in the scale above and below (see Figure 3). It is selected only if it is larger than all of these neighbors or smaller than all of them.

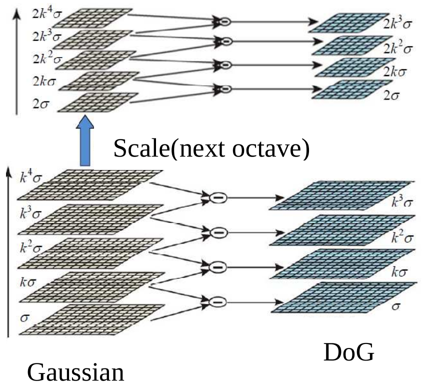


Figure 2 Difference of Gaussian and image pyramid .

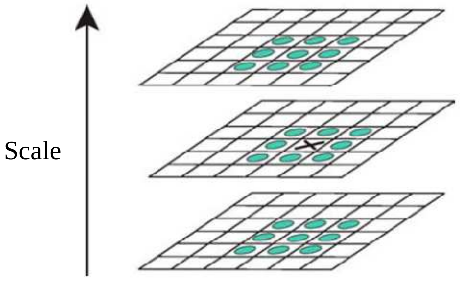


Figure 3 Extreme value detection.

Figure 4 illustrates the computation of the keypoint descriptor. Each keypoint is assigned one or more orientations based on local image gradient directions. This is the key step in achieving invariance to rotation as the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation. The previous stage found keypoint locations at particular scales and assigned orientations to them. This ensured invariance to image location, scale and rotation. The final stage computes descriptor vectors for these keypoints such that the descriptors are highly distinctive and partially invariant to the remaining variations, like illumination, 3D viewpoint, etc. The feature descriptor is computed as a set of orientation histograms on (4 x 4) pixel neighborhoods. The orientation histograms are relative to the keypoint orientation and the orientation data comes from the Gaussian image closest in scale to the keypoint's scale. Just like before, the contribution of each pixel is weighted by the gradient magnitude, and by a Gaussian with σ 1.5 times the scale of the keypoint. Histograms contain 8 bins each, and each descriptor contains a 4x4 array of 16 histograms around the keypoint. This leads to a SIFT feature vector with (4 x 4 x 8 = 128 elements). This vector is normalized to enhance invariance to changes in illumination.

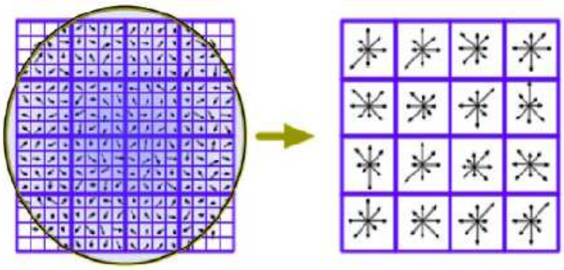


Figure 4 This figure shows (a)image gradients, and (b) keypoint descriptor.

Feature points matching between adjacent image is based on Lowes (2004). The research on matching shows that by finding both the closest matching descriptor as well as the second closest, and then discarding matches where the distance ratio between the closest and second closest descriptor is greater than 0.8 eliminates 90% of the false matches and only 5% of the correct matches. In other words: Matches where the closest and second closest descriptors are too close will be discarded, resulting in the elimination of most false matches. To estimate the transformation between the two images a RANSAC approach is used. The RANSAC algorithm estimates the fundamental matrix containing the scaling, rotation and translation of the image features. RANSAC is short for RANdom SAmple Consensus, it is a robust estimator.

3.2 SfM

In computer vision SfM refers to the process of finding the three-dimensional structure of an object by analyzing local motion signals over time. In vision science, SfM refers to the general phenomenon by which humans can recover 3-D structure from the projected 2D motion field of a moving object. The application of projective geometry techniques in computer vision is most notable in the stereo vision problem which is very closely related to SfM. Unlike general motion, stereo vision assumes that there are only two shots of the scene. In principle, one could apply stereo vision algorithms to a SfM task (Robertson, D.P. and R. Cipolla, 2009).

SfM is work by incorporating successive views one at a time. As each view is registered, a partial reconstruction is extended by computing the positions of all 3D points that are visible in two or more views using triangulation. There exist several strategies for registering successive views included epipolar constraints resection and merging partial reconstructions. The process schema is tracking the feature points by constantly stitching the adjacent image, shown as Fig. 5. From image features, SfM gives an initial estimate of projection matrices and 3D points. Usually it will be necessary to refine this estimate using iterative non-linear optimisation to minimize an appropriate cost function. This is bundle adjustment. Bundle adjustment works by minimising a cost function that is related to a weighted sum of squared reprojection errors. SfM creates a coordinates system based on the relative position of camera and shooting object. These estimated feature points are the point clouds structure presented in Photosynth and Photofly.

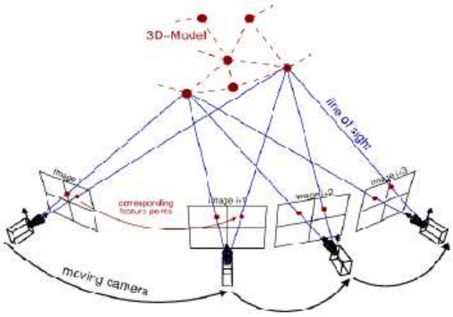
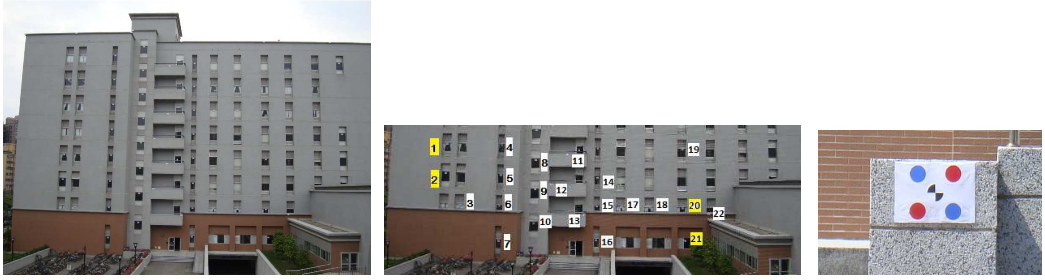


Figure 5 SfM: Tracking the feature points by constantly stitching the adjacent image.

4. EXPERIMENT

4.1 Study Area

The experimental area is based on the National Taipei University collage of Public Affairs building's northeast wall of 1 to 4 floor which range long about 60m, height about 16m. There are 22 observation targets pasted on the wall with uniform distribution. Photos are taken in different angle, distance, and location of shooting. A full view and part view of facades are considerate. Figure 6(a) is the appearance of the experimental area, Figure 6(b) shows distribution of observation targets, and Figure 6(c) shows the schema of the target.



(a) (b) (c)

Figure 6(a) Experimental area; (b) Distribution of observation targets, and (c) Schema of target

In order to get the real world coordinates of target for accuracy assessment, two of ground control points near experimental area were measured by GPS with static observation approach. The network adjustment with the other two fixed stations (National Taipei University (NTPU), and High Speed Railway (HSR1)) was processed to obtain accurate TWD97 coordinate system. The coordinate of 22th targets were measured by Lecia TPS total station based on ground control points by GPS measurement. After that, the coordinates of observation target would be as the referenced for accuracy assessment by that of estimated by Photosynth and photofly point clouds.

5. RESULTS AND ANALYSES

The number of generated point clouds and positioning accuracy of target under different conditions were the main items for comparison. In this article, just selected the photos which cover 1st to 4th floor of building facade because higher floor may generate too few point clouds caused too large shooting angle (Chen, 2011). We selected 150 photos based on covering facade of the building floor of 1 to 4. The 150 photos are available owing to the consideration of time consumer and of number of point clouds which was good enough to identify the position of the target among point clouds. In this study, processing time of using 150 images to generate point clouds by Photosynth for about 5 minutes, and the number of generated point clouds at about 60,000 points. In all 150 photos, there are each of 50 photos covering almost full view of the building facade on the side of north-east wall in different shooting angles; another 100 photos which camera is closer to the building and each of photos covered at least 1 to 4 targets. Fig. 7 shows each of selected 50 photos almost covered full view of building facade, and Fig. 8 shows the selected 100 photos just covered with parts of it.

Figure 7 A full view of facades.

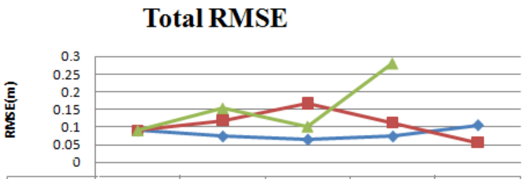
Figure 8 A part view of facades.

In this paper, there are three of adjustment factors on image which includes changing of brightness, rotation, and resolution. To estimate the positioning accuracy of observation marks by Phosynth and Photfly as adopted changing of three factors, the photos are divided into two categories. There are about 50 photos which each of photos covered full view of façade of building, the rest are just covered parts of façade. In order to make two parts of photos can be evenly distributed on shooting position, the first 50 "full view" photos were arranged in shooting position order then selected two take one. And then, adjusting the brightness, rotation, and resolution of the half of photos, respectively, and keeping the other (25 photos) unchanged. Considering about another category, 100 "part view" photos, the 50 photos changed and keep the other (50 photos) unchanged. In this study, there are five of cases be considered on changing the brightness of photos included the original, original-40, original-20, original+20, and original+40 of brightness. The photo was rotated into original, original+15, original+30, original+45, and original+60 degrees, respectively. The original photo resolution is 3264 * 2448 pixels. In the case of resolution change on this study is to adjust the original picture resolution by reduced 50%, 25%, and 12.5% from original resolution, respectively, and then resulted in 1632 * 1224 pixels, 816 * 612 pixels, and 408 * 306 pixels resolution, respectively.

According to the well distribution of control points for coordinate transformation, the observation mark of 1, 2, 19, and 21 should be a good selection. But we found the number 19 is not clear to identify among the point clouds, so we select the point of 1, 2, 20, 21 instead. All of the coordinate of observation marks on Photosynth and Photofly have their local coordinate system respectively. They have to be transformed into real world coordinate system, based on these 4 control points. The accuracy assessment of positioning is compared the transformed coordinate of observation marks with that of measurement by theodolite.

Figure 9 shows the total RMSE of the trend of the three adjustment factors. According to Fig. 9, the results of the resolution changes have larger RMSE relatively, while brightness adjustment has the smallest one. From the number of generated point clouds of view, in the case of brightness adjustment can produce about 67,000 of point clouds, the angle adjustment have about 40,000~5000 of points, and adjusted resolution down to 816 * 612 only generated 28,000 of points. It could found adjusted the brightness shows the impact is small.

Figure 9 shows when adjusted for the 60-degree angle result in the smallest positioning error with total RMSE of ±0.056m. It is interested results and need to further research.



Brightness	0.09 (0,67185)	0.074 (-40,61753)	0.065 (-20,67293)	0.074 (20,49077)	0.105 (40,48700)
Angle	0.09 (0,67185)	0.119 (15,51236)	0.167 (30,42909)	0.112 (45,45993)	0.056 (60,44689)
Resolution	0.09 (3264*2248 ,67185)	0.153 (1632*1224, 58469)	0.101 (816*612, 31710)	0.281 (408*306,. 28563)	

Figure 9 Total RMSE of three factors of changing on photos. The value of adjustment(left) and number of generated point clouds(right) are shown in parenthesis.

6. CONCLUSION

The experiment makes the appropriate adjustment on the part of photographs by brightness, rotated angle, and resolution. Using photosynth and photofly generate point clouds. Positioning accuracy of RMSE is estimated by comparing the coordinate of observation marks generated by Photosynth and Photfly respectively with measurement by theodolite. In photosynth, the original photos gives total RMSE of ±0.090m with the number of point clouds of 67185. From the view of the three adjustment factors, the rotation angle of 60 degrees and adjust the brightness to -20 have better results with total RMSE of ±0.056m and ±0.065m, respectively. Number of point clouds can be generated more than sixty thousand points by original photos, after adjustment, the average number of point clouds produced about 40,000. According to the experiment in this paper, the resolution factor have the greatest influence on positioning accuracy among these three. Photofly cannot work while adjustment photo brightness, and rotation so it did not have this part of the experimental results. In this paper, there still have some factor did not taken into account, such as factor of shooting distance.

References

Abate, D., Furini G., Migliori S., and Pierattini S., 2010. Project Photosynth:New 3D Modeling Online Web Service(Case Study and Assessments).

Chang, Ting-Rong, 2008. The Application of Stereo Image Matching and Image Retrieval Based on SIFT Algorithm, National Kaohsiung University of Applied Sciences of Department of Civil Engineering Institute of Civil Engineering and Mitigating Technology of Disasters Master's thesis, 139 pages.

Chen, Yi-Ting, 2008. The Application and Analysis of Automatic Aero-triangulation Based on SIFT Feature Matching, National Kaohsiung University of Applied Sciences of Department of Civil Engineering Institute of Civil Engineering and Mitigating Technology of Disasters Master's thesis, 85 pages.

Chen, Szu-Han, 2011. 3D Construction and Accuracy Assessment for Uncalibrated Images, National Taipei University of Department of Real Estate and Build Environment Master's thesis, 110 pages.

Huang, Han-che, 2009. The Study of Aerial Imageries Stitching Based on SIFT Algorithm, National Sun Yat-sen University Department of Marine Environment and Engineering Master's thesis, 89 pages.

Lowe, D.G. 1999. Object recognition from local scale-invariant features. In International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2): 91-110.

Robertson, D.P. and R. Cipolla, 2009. Structure from Motion. In Varga, M., editors, Practical Image Processing and Computer Vision, John Wiley.

Snavely, N., Seitz, S., M., and Szeliski, R., 2008. Skeletal Sets for Efficient Structure from Motion, Proc. Computer Vision and Pattern Recognition(CVPR), pp. 1-8.

Wu, Joz, Chi Chang, and Ming-Che Liub, 2009. Precise Multisource Image Registration Based on Scale Invariant Feature Transform Points, Journal of Photogrammetry and Remote Sensing 14(2) , pp. 141-155.