3D MODEL RECONSTRUCTION AND ACCURACY ASSESSMENT: A CASE STUDY ON PHOTOSYNTH*

Jie-Shi Weng[a], Yi-Ting Tsai[b], Jin-Tsong Hwang[c]

[a]Graduate Student, National Taipei University, No.,151, University Rd., San Shia, New Taipei City, Taiwan; E-mail: jessie8031@hotmail.com
[b]Master, National Taipei University, No.,151, University Rd., San Shia, New Taipei City, Taiwan; E-mail: tsai_ting@hotmail.com
[c] Associate Professor, National Taipei University, No.,151, University Rd., San Shia, New Taipei City, Taiwan, Tel: +886286741111 ext. 67426; E-mail: jthwang@mail.ntpu.edu.tw

KEY WORDS: Close-range Photogrammetry, LiDAR, Point Cloud

Abstract: The main method for three-dimensional model reconstruction of building are close-range photogrammetry and LiDAR scanning, but the former requires more human intervention, and the latter needs expensive equipment to use. Therefore, the approach of using non-metric digital camera to reconstruct the 3D model may easily and affordable. In 2008, Microsoft released free software of Photosynth. It 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. This study uses general consumer digital cameras to shoot the subject of pre-laid target in the experimental area, and uses the difference of a known target coordinates in RMSE as the accuracy evaluation. Experiments show that when the appropriate shooting factors are under control, the check points of three-dimension accuracy are around ± 0.059m, ± 0.068m and ± 0.079m respectively. To construct a suitable and accurate three-dimensional model, a completed space appearance can be established in an economic, convenient and reliable way. This will be helpful for reserving buildings, and also provide another choice for future reconstruction of this three-dimensional model.

INTRODUCTION

With the development of two-dimensional image has matured, and in pursuit of a better visualization and multiple-use, 3D modeling technique is continue developing in spatial surveying and geographic information system. In current approach, we can use the close-range photogrammetry, photogrammetry, LiDAR, etc., to construct three-dimensional model. For example, Chih-An Chang and Liang-Chien Chen (2006), combined LiDAR point cloud with topographic maps to reconstruct housing model. However, there are differences among various approaches on their characteristics and the purpose of use. In terms of the convenience when using, the close-range photogrammetry approach is the fastest and easy way to obtain information. As long as using the general digital camera, and then through the appropriate calibration, it can be used in constructing model efficiently. In 2006, Microsoft introduced Photosynth browser which combined close-range photogrammetry, SIFT, and SfM approaches, etc. This is a new development on photogrammetry which can mosaic and display images. In addition, it can create the point cloud data with three-dimensional coordinates and then construct models based on the data.

In terms of three-dimensional model reconstruction by close-range photogrammetry, Cheng-Chung Wang, (2002), used photogrammetry to construct a coarse three-dimensional building model, and use with close-range photogrammetry to refine the building model. By using the least squares model to solve the exterior orientation parameters and use a semi-automated way to reconstruct the building model. Pomaska, G., (2009), described the download procedure from Photosynth point and how to use point cloud data to construct three-dimensional models. Dowling, et al., (2009), used Photosynth processing and matching point cloud through shooting surface images that can quickly generate small-scale digital elevation model (Digital Elevation Model, DEM) data to identify the raises and falls of surface, and to know the true size of the DEM according to the appropriate reference scale. 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. SfM was developed in the 1980s. Its main purpose

* This work was supported in part by the National Science Council of Taiwan under Grant (NSC 100-2410-H-305-065)

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.

Currently, the common method to construct three-dimensional models is the use of close-range photogrammetry and LiDAR scanning. However, the first method requires more human intervention, while the second method requires more expensive equipments to complete the model. Therefore, shooting building images to obtain the coordinates and color information by using low-cost non-metric digital cameras as an easy equipment to build the three-dimensional models; can be regarded as an economic and fast way. However, the camera in this method has not been used in calibration and the image is uncalibrated. In the detection of building characteristics and the points matching, the tests of its validity and accuracy are necessary to do. In this study, the close-range photogrammetry, Photosynth and terrestrial LiDAR are the different methods to build three-dimensional modeling. Close-range photogrammetry (using software of Photomodeler pro 5) and Photosynth both use digital camera to construct models, and the terrestrial LiDAR uses the instrument scanning point cloud data. This study obtains three-dimensional coordinates of observation targets by these three approaches then compares the observation results with that of measurement by total station as the basis to evaluate the accuracy.

METHODS

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 of SIFT process is to detect scale-space extreme for feature extraction. This is the stage where the interest points, which are called key points 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. Key points 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):

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \quad (1)$$

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

$$G(x,y,\sigma)=(1/2\pi\sigma^2)\exp(-(x^2+y^2)/2\sigma^2) \quad (2)$$

To efficiently detect stable key point locations in scale space, we have proposed (Lowe, 1999), using scale-space extreme 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:

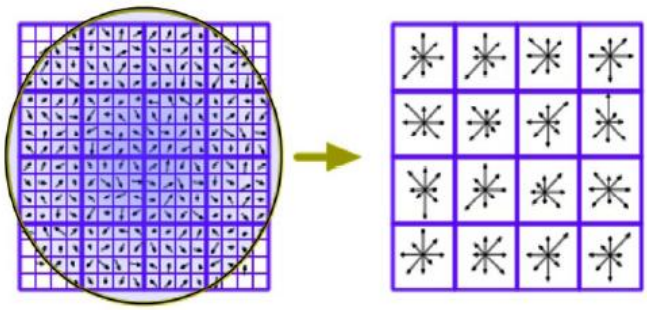
$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)$$

$$=L(x,y,k\sigma)-L(x,y,\sigma) \quad (3)$$

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. It is selected only if it is larger than all of these neighbors or smaller than all of them.

Figure 1 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 achieves 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.



(a) (b)

Figure 1: 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.

2. SfM

SfM is based on analysis the relative or absolute position of several photos, and then achieve the three-dimensional model of shooting object with the coordinate system. The process schema is tracking the feature points by constantly stitching the adjacent photos. It will automatically removed photos which cannot be matched. At the same time, the camera motion parameters can be estimated as well. The process schema is tracking the feature points by constantly stitching the adjacent image, shown as Fig. 4.

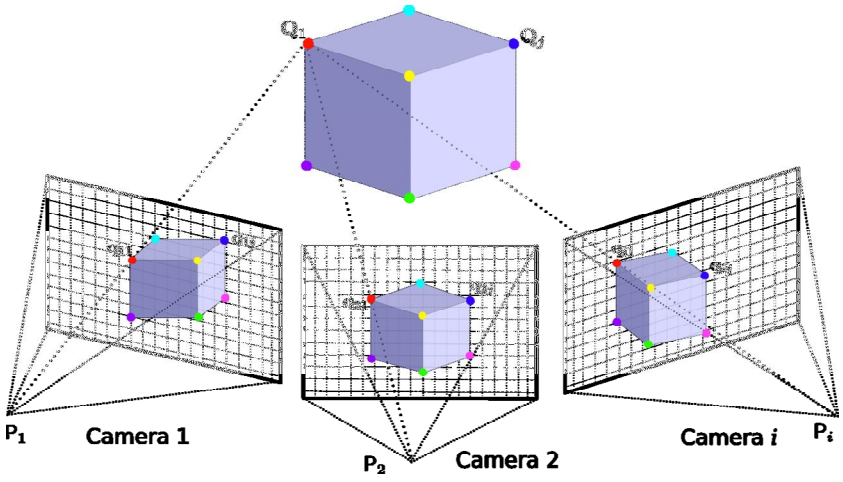


Figure 2: SfM: Tracking the feature points by constantly stitching the adjacent photo.

SfM process is divided into four steps. The first, extracting features point in two-dimensional image based on the SIFT algorithm. The second is estimated the object and camera position by using epipolar plane geometry of feature

points. The third is using bundle adjustment to optimize the coordinate of feature points and camera position. Several overlapped images can be constrained each other by using collinearity equation. The camera orientation parameters and points' coordinate of the shooting object were estimated by using space resection and intersection which required to constantly iterating in order to keep data convergence.

The last is built the geometry of points. SfM creates a coordinates system based on the relative position of camera and shooting object. These estimated feature points are the points cloud structure presented in Photosynth.

EXPERIMENT

1. Study area

The study area is mainly in north-east side of the College of Public Affairs of National Taipei University (NTPU) Sanshia Campus in Taiwan as the experimental area. The test area is about 60m in length and about 35m in height. Figure 3 is the appearance of the experimental area. This study has the artificial observation targets to attach on the building walls for close-range photography, Photosynth and LiDAR measurement in order to create the three-dimensional coordinates and then estimated the positioning accuracy. The wall to the measurement station has different depth of field, so the target should be set equally plane and with depth of field as possible as it can. Figure 3 shows distribution of observation targets, and Figure 4 shows the schema of the target.

Figure 3: Experimental area.

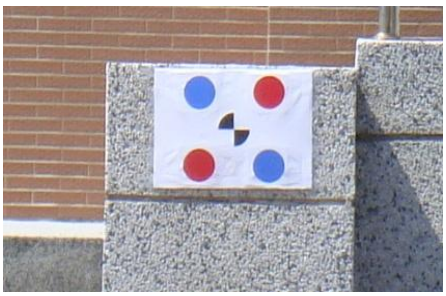


Figure 4: Schema of target

In order to get the real world coordinates of target for accuracy assessment, there is two of ground control points near to experiment area were measured by GPS, and 60th observation targets were measured by theodolite then. Static GPS observation was adopted, and there are keep more than seven satellites to receive. Then do network adjustment with the other two fixed stations (NTPU and HSR1) to obtain accurate TWD97 coordinate system. The coordinate of target was measured by Lecia TPS total station based on ground control points by GPS measurement. After that, the coordinates of observation target would be the referenced for accuracy assessment by that of estimated by Photosynth points cloud.

2. The Research Process

In this study, the close-range photogrammetry, Photosynth and terrestrial LiDAR are the different methods to build three-dimensional modeling. Close-range photogrammetry and Photosynth both use digital camera to construct models, and the terrestrial LiDAR uses the instrument scanning point cloud data.

The positioning accuracy assessment is mainly based on the observation targets. This study sets sixty observation targets on the walls of the building as the measurement of control points and check points. Coordinate system is established by static GPS measurement which uses two ground control points near the experimental area. The coordinates of the two ground control points are solved by measuring adjustment together with two of fixed satellite stations of NTPU and Shulin high speed rail station in Taiwan. Set up the total station in one ground control point and another is for setting orientation. We observe and then solve each observation target's three-dimensional coordinate as a main basis to evaluate the accuracy of three different approaches. In the part of close-range photogrammetry and Photosynth use the digital camera shooting the wall with observation targets, and then continue the following work. The shooting location is away from the wall; the distance between them is 10 to 30 meters. PhotoModeler needs to select the visual feature points of each image point position. After manual specify conjugate point matching on different images, we can obtain object space coordinate system by using direct linear transformation. However, Photosynth needs to upload the photo to the cloud server then download the point cloud to edit and extract the location of observation targets. In LiDAR, this study uses single-station scanning approach in the scanning mode with the distance of 100 meters and the point cloud interval of 2 cm. Because LiDAR can detect the special pattern of targets, it can automatically identify the location of observation targets after scanning. After the location of observation targets extraction, we can use parts of the observation target as control point for coordinate system transformation. This study obtains three-dimensional coordinates of observation targets by these three approaches then compares the observation results with that of measurement by total station as the basis to evaluate the accuracy.

3. The use of equipment

Coordinate system is built by using the GPS observation and total station to measure and calculate the control point coordinates. The GPS receiver uses Topcon GRS-1 satellite receiver with G3-A1 antenna which can effectively eliminate multi-path effects in order to improve the accuracy of the static measurement. Besides, this study uses Leica TPS 700 total station to measure the three-dimensional coordinates of observation targets as a reference for subsequent accuracy analysis. LiDAR data uses Leica Scanstation C10 laser scanner to scan.

4. The set of observation target

The observation targets in this study are mainly to improve the observation targets of terrestrial LiDAR that make them both can be identified by LiDAR and Photosynth. In Figure 3, the black and white part of circle center is the observation target of terrestrial LiDAR by which can be identified and directly extracted points by Leica software. In PhotoModeler, because its black and white image shows the shape of a cross, it is also easy to identify when selecting image points by hand. However in Photosynth, due to the part of circle center is not a good feature point for generating cloud point. According to experiments, we found that the color feature have better identification, so we can use the circle target with a color (blue and red) around to determine the center position of the observation target (Figure 3).

In the distribution of the observation targets, the study area set a total of 60 observation targets, of which there are 20 control points as three-dimensional coordinate transformation basis and the remaining 40 are used for check points. Figure 6 shows the point cloud around the observation target by Photosynth approach.

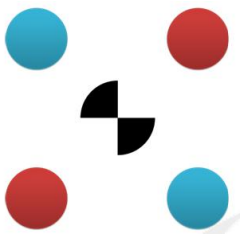


Figure 5: Illustration of the observation target.

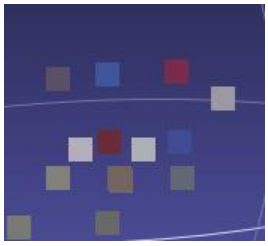


Figure 6: Point cloud near the Photosynth observation target.

THE ANALYSIS OF EXPERIMENTAL RESULTS

The design of the evaluation of positioning accuracy uses 60 observation targets which are based on 20 target points as control points and the remaining 40 points are check points. After extracting these 60 points by LiDAR, close-range photogrammetry, and Photosynth approaches, we compare the coordinate of all observation target with that were measured by total station respectively. The RMSE (Root Mean Square Error) is selected as indicator of positioning accuracy.

Accuracy analysis mainly uses No. 1, 6, 9, 13, 15, 18, 19, 22, 25, 27, 34, 36, 39, 42, 45, 48, 50, 56, 57, 58, the total of 20 points in Figure 3 to be the control points as for coordinate transformation. The selection principle is that these 20 points are shown in the three approaches of point cloud data, and are equally distributed in the study area. However, in the measurement and calculation process of LiDAR and Photosynth, some points are obscured or cannot produce target location from point cloud, etc. In order to avoid the result of different conversions, 20 points are selected and the remaining points are the check points. All of target points generated by three approaches transfer into geodetic coordinate system wit 7-parameters respectively.

Each control points and check points accuracy analysis of LiDAR, PhotoModeler, and Photosynth are listed in Table 1 and Table 2. According to Table 2, the check points accuracy show that three-dimensional positioning accuracy of LiDAR whether in X, Y, or Z axis, RMSE are lower than the other two approaches and achieve the level of millimeter accuracy. The total RMSE was ± 0.005m. In the other two approaches, the accuracy is limited in shooting distance, intersection angle and height of buildings. Only some floors can be completely shot by high-elevation shooting, so the accuracy is only in the centimeter level. In PhotoModeler, the total RMSE is still generally better than that of Photosynth and is around ± 0.065m.

In the case of Photosynth, the accuracy of check points is more stable when the images are taken between 300-500 images in the study area. The total RMSE is about ± 0.214m to ± 0.197m. Too many images are easy to generate too much point cloud and match with error; however, less may not generate enough feature points. As shown in Table 2, when the images are more than 500 or less than 300, the total RMSE is around ± 0.546 m. And when low-angle of shooting is smaller, the accuracy will be improved. This is because low-angle of shooting from the ground is high, the intersection angle is poor. However, the impact will be little in small low-angle shooting. The results show the floors under four can promote the accuracy to ± 0.049m. Therefore, we found that the points on high-floor have a bigger position error.

Table 1: Control point accuracy analysis. (RMSE, units: meter)

Case	No. of points	X	Y	Z	Total
LiDAR	20	0.004	0.002	0.003	0.005
Photo Modeler	20	0.044	0.059	0.025	0.065
Synth564 images	20	0.364	0.276	0.300	0.546
Synth507 images	20	0.121	0.123	0.095	0.197
Synth456 images	20	0.072	0.083	0.067	0.129
Synth401 images	20	0.100	0.109	0.086	0.171
Synth351 images	20	0.067	0.071	0.096	0.136
Synth297 images	20	0.099	0.159	0.105	0.214
Synth251 images	20	0.372	0.283	0.219	0.516
1st to top floors	18	0.044	0.085	0.062	0.114
1st to 8 floors	15	0.030	0.071	0.026	0.080
1st to 4 floors	8	0.027	0.039	0.015	0.049

Table 2: Check point accuracy analysis. (RMSE, units: meter)

Case	No. of points	X	Y	Z	Total
LiDAR	37	0.005	0.003	0.004	0.007
PhotoModeler	40	0.038	0.052	0.024	0.069
Synth564 images	40	0.312	0.425	0.295	0.604
Synth507 images	40	0.087	0.149	0.071	0.187
Synth456 images	39	0.059	0.082	0.089	0.135
Synth401 images	40	0.069	0.118	0.118	0.180
Synth351 images	40	0.059	0.068	0.079	0.119
Synth297 images	38	0.076	0.082	0.062	0.128
Synth251 images	38	0.170	0.330	0.166	0.407
1st to top floors	38	0.058	0.078	0.091	0.133
1st to 8 floors	33	0.034	0.072	0.028	0.084
1st to 4 floors	15	0.028	0.065	0.012	0.072

CONCLUSIONS

Photosynth has developed image processing in recent years. It can be used to shoot many of images to reconstruct three-dimensional object space automatically and also can be viewed directly on Web, and then generates the point cloud data. And for the method SIFT algorithm and SfM, both are improved continuously in order to overcome the image with deformation to provide Photosynth better results.

In this study, the experimental research is mainly to compare the positioning of accuracy among close-range photogrammetry, LiDAR, and Photosynth approaches and to understand these three methods in 3-D reconstruction procedures and the results. Based on the goals above, this study proposes the following conclusions:

1. In the three-dimensional positioning accuracy analysis on the study area, experimental results show that the accuracy of control point by LiDAR is slightly higher than that of the check point. Y direction due to the direction is for the laser ranging, so accuracy of the depth of field is better than X and Z. The check point of RMSE in X and Z direction are around ± 0.005m and ± 0.004m respectively and the depth of field of RMSE in Y direction is around ± 0.003m, while the total RMSE of the control points is around ± 0.005m and the total RMSE of the check points are around ± 0.007m. In close-range photogrammetry, it is limited in shooting angle in this study area. The positioning accuracy of PhotoModeler in direction of the X and Z is ± 0.038m and ± 0.024m respectively, while the accuracy of the depth of field in the Y direction is limited in the shooting intersection angle and distance, so the accuracy is poor, about ± 0.052m. The total accuracy of control points is around ± 0.065m and the accuracy of check points is around ± 0.069m that the results are poor than that of LiDAR.
2. In Photosynth, we can find that the wall about 60mx35m was shot in the experiment and when the shooting location is away from the wall about 10m to 30m, the number of the images shoot in the 300-500 can get a more stable positioning accuracy. In the experiment, the best total positioning accuracy was shooting 351 pieces of images. Their X, Y and Z direction can be the best precision, which are ± 0.059m, ± 0.068m and ± 0.079m respectively. However, if the intersection angle of shooting is too small or too blunt, that will affect the accuracy of the depth of field in Y direction. Therefore, in most cases the accuracy in Y direction is lower than the accuracy of X and Z direction and the total accuracy are about ± 0.119m. And the cases of the images are

less than 300 or above 500, take 564 for example, the total RMSE is up to ± 0.604m. In the analysis of different shooting low-angle shows that the lower the angle, the higher the precision is. Calculating the positioning accuracy only from 1st to 4th floor, the total RMSE of control points can be promoted to ± 0.049m, and the check point can reach ± 0.072m.

REFERENCES:

Cheng-Chung Wang, 2002. Model-based Building Reconstruction Using Close-Range Photogrammetry. National Cheng kung University, Department of Geomatics Master's thesis, 93 pages.

Chih-An Chang, and Liang-Chien Chen, 2006. Building Shaping from LIDAR Data. Journal of Photogrammetry and Remote Sensing, 11(2), June, pp. 175-189.

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

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

Dowling, T., I., Read, A., M. and Gallant, J., C., 2009. Very high resolution DEM acquisition at low cost using a digital camera and free software. The 18th World IMACS Congress and MODSIM09, International Congress on Modeling and Simulation, Cairns, Australia, pp. 2479-2485.

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

Pomaska, G., 2009. Utilization of Photosynth Point Clouds for 3D Object Reconstruction, 22nd CIPA Symposium, Kyoto, Japan.