Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

Article

Fusion of Infrared Thermal Image and Visible Image for 3D Thermal Model Reconstruction Using Smartphone Sensors

Ming-Der Yang [1,2], Tung-Ching Su [3,*], Hung-Yu Lin [1]

[1] Department of Civil Engineering, National Chung Hsing University, 145 Xingda Rd., South Dist., Taichung 402, Taiwan

[2] Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University, 145 Xingda Rd., South Dist., Taichung 402, Taiwan

[3] Department of Civil Engineering and Engineering Management, National Quemoy University, 1 Da Xue Rd., Kinmen 892, Taiwan

* Correspondence: spcyj@nqu.edu.tw; Tel.: +886-82-313-527

Abstract: Thermal infrared imagery provides temperature information on target objects, and has been widely applied in non-destructive testing. However, thermal infrared imagery is not always able to display detailed textures of inspected objects, which hampers the understanding of geometric entities consisting of temperature information. Although some commercial software has been developed for 3D thermal model displays, the software requires the use of expensive specific thermal infrared sensors. This study proposes a cost-effective method for 3D thermal model reconstruction based on image-based modeling. Two smart phones and a low-cost thermal infrared camera are employed to acquire visible images and thermal images, respectively, that are fused for 3D thermal model reconstruction. The experiment results demonstrate that the proposed method is able to effectively reconstruct a 3D thermal model which extremely approximates its corresponding entity. The total computation time for the 3D thermal model reconstruction is intensive while generating dense points required for the creation of a geometric entity. Future work will improve the efficiency of the proposed method in order to expand its potential applications to in-time monitoring.

Keywords: 3D thermal model; image fusion; smart phone; thermal IR

1. Introduction

Currently 3-dimensional (3D) mapping techniques fall into two categories: active or passive mapping techniques. Active mapping techniques, such as laser scanning, directly and quickly acquire an enormous point cloud dataset with high spatial precision. Coupled with visible images, a 3D model with true color information can then be reconstructed. Passive mapping techniques, such as image-based modeling (IBM) [1,2], use image datasets with multiple fields of view (FOV) to reconstruct 3D models, so these techniques usually require certain imaging conditions, including high spatial resolution, a high percentage of imaging lap, accurate camera calibration, and precise camera positioning [3].

Laser scanning can directly provide an enormous point cloud dataset to reconstruct precise 3D models, but requires a high degree of training and is very costly. A light detection and ranging (LiDAR) instrument coupled with a thermal camera has also been applied to 3D thermal model reconstruction [4]. Despite the spatial data acquired by LiDAR instruments having high spatial precision, a LiDAR-based 3D thermal model reconstruction method is still impractical due to its prohibitively high instrument cost. Unlike laser scanning and LiDAR, IBM adopts an indirect method to derive a dataset of transfer parameters, as well as equations for 3D model reconstruction.

© 2018 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

Furthermore, IBM needs no knowledge about the interior orientation elements of cameras before 3D model reconstruction, and is a low-cost spatial information acquisition technique. With IBM, an effective extraction of conjugate features from adjacent images is necessary. Lowe (2004) [5] indicated that scale-invariant feature transform (SIFT) has invariance to image rotation and scale, and is robust across a substantial range of affine distortion, the addition of noise and change in illumination; thus, it would be useful in key point feature selection. Based on the conjugate features, structure from motion (SFM) can derive the interior orientation elements of a camera in order to reconstruct the 3D model without preprocessing camera calibration [6]. However, the procedure from SIFT to SFM usually requires a central processing unit (CPU) with powerful computation efficiency. In order to address this issue, Wu (2013) [7] ran bundler reconstructions on a PC with an Intel Xenon 5680 3.33Ghz CPU (24 cores), 12GB RAM, and an nVidia GTX 480 Graphics Processing Unit (GPU). Jancosek and Pajdla (2014) [8] noted that SFM and multiview-stereo (MVS) have become the most popular methods for obtaining input points with visibility information in 3D model reconstruction.

A thermal camera can detect the spectral radiances of objects under a certain temperature, based on the blackbody radiation theory. The detected spectral radiances are usually displayed as digital numbers (DN) obtained by a conversion of radiation level to electric signal. Thus, a thermal image usually consists of numerous DN values, and records temperature information in the 2-dimensions of rows and columns. Temperature information can be directly extracted from the thermal image, but geometric information, such as edges, feature points and others, cannot; this hampers the understanding of geometric entities consisting of temperature information. Therefore, fusion of visible images with thermal images is helpful for 3D thermal model reconstruction.

Recently, integrations of multisource remote sensing data have been widely applied to environmental monitoring [9,10,11], building inspection [12,13,14,15], heritage preservation [16,17] and non-destructive testing [17,18,19,20,21,22,23]. A 3D thermal model cannot be successfully reconstructed without considering the geometric information in thermal images, thus limiting the above applications. IBM is a computer vision technique for effectively integrating multisource image data into a unique 3D model. Lagüela et al. (2012) [24] used image fusion and image matching techniques to combine thermographic and metric information to obtain 3D thermal models. Several literatures also presented the fusion systems for multiple sensors, including 3D laser scanner, RGB camera, and thermal camera, to generate 3D thermal models [17,25,26,27,28]. Therefore, image fusion is an important pre-processing for multiple sensors-based 3D thermal model reconstruction.

With the popularization of smart mobile devices, carried nonmetric sensors have been widely applied in the collection of spatial data due to their convenience, low-cost, and availability [29,30,31]. This study aims to reconstruct 3D thermal models by fusing visible images with thermal images, in which geometric and corresponding temperature information are involved, for the scenes of different scales. The reconstructed 3D thermal models will better display stereo-temperature information compared to conventional thermal images. Building envelop inspection of green buildings, for instance, used 3D thermal models to test the insulation of energy [32,33,34] and to assess the failure of external wall tiles [35]. In this paper, a cost-effective sensor on smart phones for 3D thermal model reconstruction based on IBM is proposed, and the performances, including model quality and computation time, of the 3D thermal model reconstruction are discussed.

2. Experimental Equipment

This study used two smart phones (iPhone SE) and a low-cost thermal camera (FLIR ONE for iOS) to acquire visible images and thermal IR images, respectively. The iPhone SE specifications are 12 million pixels (3,000 pixels in the row direction and 4,000 pixels in the column direction) and a focal length of 4.15 mm. The FLIR ONE specifications are 76,800 pixels (240 pixels in the row direction and 320 pixels in the column direction), a focal length of 3 mm, and a range of detected temperature between -20 and 120℃ (±3℃ or ± 5%). The FLIR ONE also carries a visible light lens with the same number of pixels and focal length as the thermal one.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

Figure 1 shows the arrangement of the experimental sensors. Two smart phones were used to acquire stereopairs of visible images with high spatial resolution. The FLIR ONE was arranged between the two smart phones to acquire the thermal infrared (IR) images. Thus, an alignment arrangement of the three sensors on a theodolite ensures that the exterior orientation elements of the sensors are known. Moreover, the shifts among the lenses are fixed, which facilitates the geometric calibration for the acquired images.

3. Methodology

Figure 2 shows the proposed 3D thermal model reconstruction method. The first stage is to acquire the visible and thermal IR images. Next, the acquired images are calibrated by geometric translation and image registration using normalized cross correlation (NCC) [36]. The geometric translation addresses the shift between the visible and thermal IR lenses on the FLIR ONE. The image registration matches the visible images of the FLIR ONE with those of the iPhone SE. In the final stage, the visible images of the iPhone SE are used to reconstruct a 3D model by the conventional IBM technique, and the geometric translated thermal IR images are textured onto the reconstructed 3D model to produce a 3D thermal model.

3.1. Image Calibration

Generally, a baseline between two photographs with stereoscopic viewing must be fixed so that the shift between the two camera stations is known a priori in order to produce a good geometric calibration. In addition to the known baseline, the correct interior orientation elements of the experimental equipment are also required. However, acquiring the correct interior orientation elements from a manufacturer calibration is extremely difficult due to the limited budget. In this research, several algorithms, such as NCC and three-step search (TSS) [37], encoded in the C++ and OpenCV programming languages were tested for image calibration.

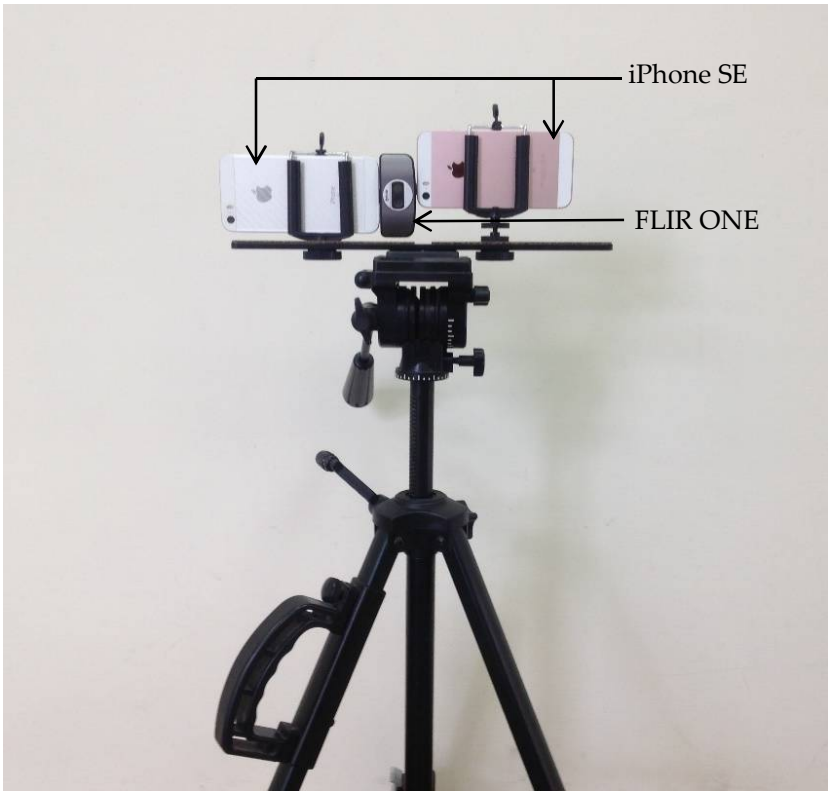


Figure 1. Arrangement of experimental equipment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

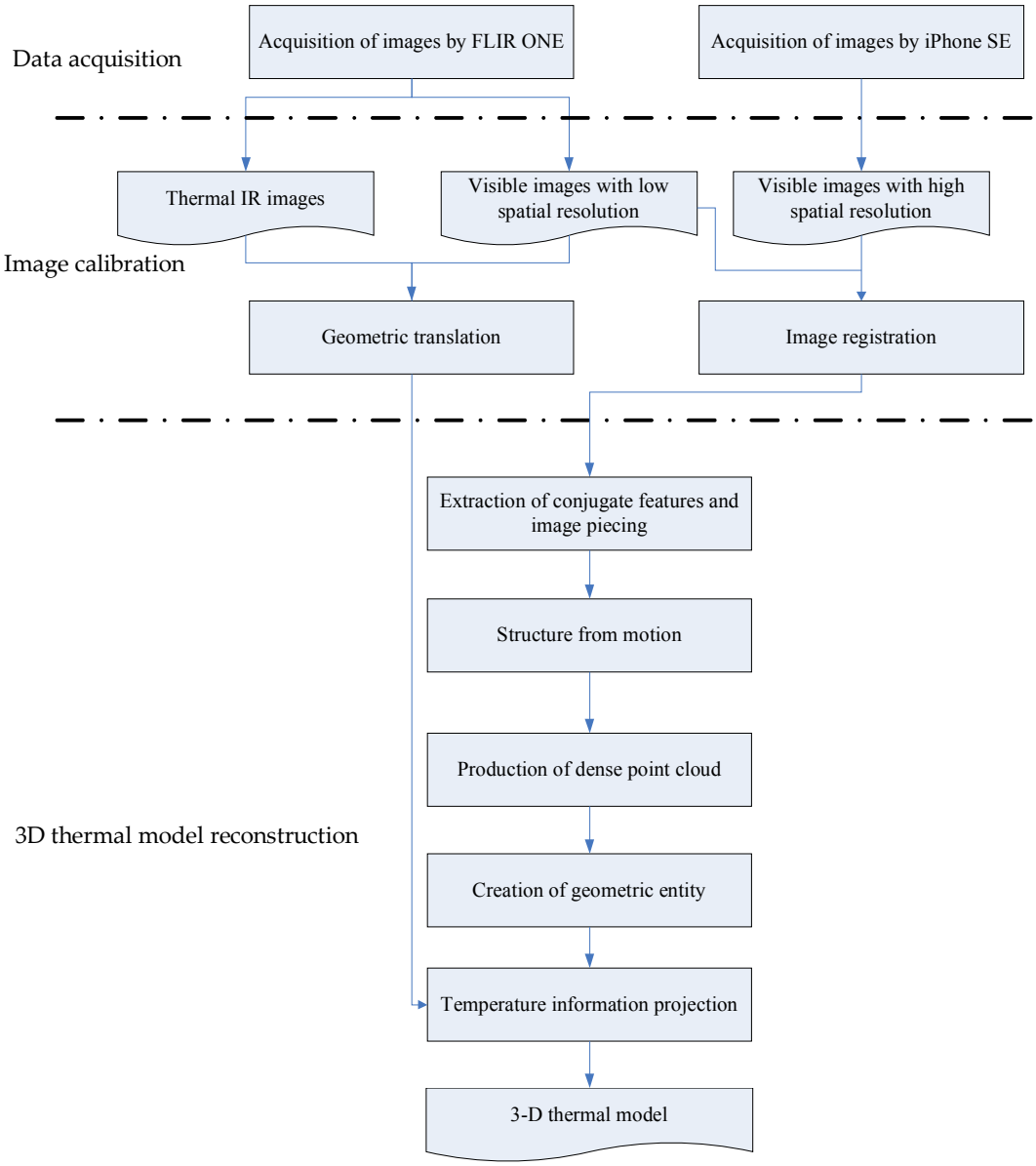


Figure 2. Scheme of proposed method for 3D thermal model reconstruction.

3.1.1. Normalization of Sensed Temperature

The FLIR ONE offers relative sensed scenes rather than absolute temperature information, so that a certain DN value among different thermal IR images can stand for different sensed temperatures. In this paper, a normalization technique is presented to appropriately adjust the obtained DN values in order to achieve consistency in sensed temperature. The normalization equations are as follows:

$$\mathrm{DN}_{\min} = [(255-0)/(\mathrm{T}_{\max}-\mathrm{T}_{\min})] \times (\mathrm{T}_{\mathrm{L}} - 10.0), \quad (1)$$

$$\mathrm{DN}_{\max} = [(255-0)/(\mathrm{T}_{\max}-\mathrm{T}_{\min})] \times (\mathrm{T}_{\mathrm{H}} - 10.0), \quad (2)$$

$$\frac{\mathbf{DN}_a - \mathbf{DN}_{\min}}{\mathbf{DN}_b - 0} = \frac{\mathbf{DN}_a - \mathbf{DN}_{\max}}{\mathbf{DN}_b - 255}, \quad (3)$$

$$\mathbf{DN}_a = \frac{\mathbf{DN}_b(\mathbf{DN}_{\max} - \mathbf{DN}_{\min}) + 255\mathbf{DN}_{\min}}{255}, \quad (4)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

where DN_{max} and DN_{min} are the normalized DN values of the highest and lowest sensed temperatures, respectively. DN_b and DN_a express the DN values before and after the normalization, respectively. T_H and T_L are the highest and lowest sensed temperatures before the normalization, respectively, and T_{max} and T_{min} are the highest and lowest sensed temperatures after the normalization, respectively. As the thermal IR images are 8-bit images, the sensed temperature information can be expressed as the DN values between 0 (corresponding to T_{min}) and 255 (corresponding to T_{max}).

3.1.2. Geometric Translation

Due to the shift between the thermal IR and visible lenses on the FLIR ONE, an image mapping result could have the phenomenon of relief displacement. To ensure that the percentage of end lap is fixed, a unity object distance among the acquired images is also needed. Consequently, the amount of relief displacement can be calculated prior to the 3D model reconstruction. A calibration template, shown in Figure 3, is employed to acquire ten pairs of thermal IR and visible images for the geometric translation, where firstly the centers of the circles on the calibration template are detected by OpenCV and secondly a root mean square error (RMSE) of the coordinates of the detected circular centers between the thermal IR and visible images is calculated. According to RMSE, the circular centers between the thermal IR and visible images can be translated to the consistent positions.

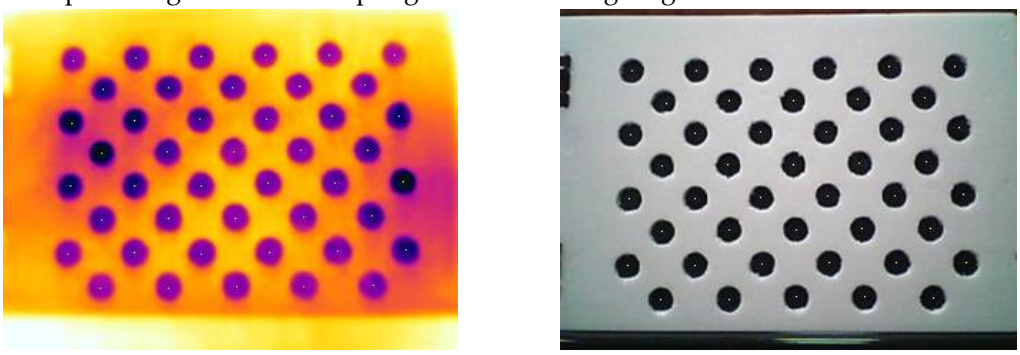
3.1.3. Image Registration

Image registration matches the visible images of the FLIR ONE with those of the iPhone SEs. Because the visible images of the FLIR ONE and iPhone SE differ in size, an image resampling process is necessary before image registration. Thus, the iPhone SE images need to be down-sampled so to have the same spatial resolution as those captured by the FLIR ONE. The percentage of down-sampling affects the image registration performance. The actual focal lengths of the FLIR ONE and iPhone SE are 3 mm (f_1) and 4.15 mm (f_2), respectively, and have a relationship with object distance (d) and size of object (l) as follows:

$$\frac{f_1}{l_1} = \frac{f_2}{l_2} = \frac{d}{l}, \tag{5}$$

$$\begin{cases} l_1 = P_1 \times S_1 \\ l_2 = P_2 \times S_2 \end{cases}, \tag{6}$$

where P_1 and P_2, and S_1 and S_2 represent the pixel numbers, and the sizes of charge-coupled devices (CCD) of the FLIR ONE and iPhone SE, respectively. Thus, l_1 and l_2 express the sizes of CCD arrays of the FLIR ONE and iPhone SE, respectively. According to Equation (5), if S_1 and S_2 are known, a certain percentage of down-sampling exists between P_1 and P_2. In order to obtain an optimal percentage of down-sampling, the percentages of down-sampling from 8 to 50% were tested. The optimal percentage of down-sampling leads to an image registration result with the least RMSE.



(**a**) (**b**)

Figure 3. Calibration template for geometric translation: (**a**) False color thermal IR image; (**b**) Gray scale visible image.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

NCC is a common image registration technique as:

$$\text{NCC Index} = \frac{1}{n}\sum_{x,y}\frac{(f(x,y)-\bar{f})(t(x,y)-\bar{t})}{\sigma_f\sigma_t}, \quad (7)$$

where f(x,y) and t(x,y) represent the DN values of the registered image and the template image, respectively. $\bar{f}$ and $\bar{t}$ and σ_f and σ_t are the means and the standard deviations of DN values of the registered image and the template image, respectively. x and y are the numbers of pixels in the row and column directions, respectively. The values of x and y are given as 320 and 240, respectively, because of the 320 × 240 thermal IR image (registered image). Figure 4 is an illustration of image registration using NCC. The image region within the red frame is regarded as a template image and slips on the original image of iPhone SE from the upper left to the lower right to calculate the NCC indices. The k % is a percentage of down-sampling. Consequently, the NCC index is between -1 and 1. The NCC index approaching 1 means a high similarity between the registered image and template image.

However, the computation complexity of NCC depends on the sizes of the processed images, and the process is usually time-consuming. Li et al. (1994) [36] noted that TSS is an efficient algorithm in processing an enormous imagery dataset, and thus could be frequently applied to film detection. In this research, TSS was adopted to assist NCC in the image registration. Figure 5 shows an illustration of the TSS operation. TSS starts the search from the central pixel with the search radius S = 4. The search parameter value P is set to 7; thus, TSS searches eight pixels around the central one with search radius ±S. Of these eight searched pixels, the pixel with the least difference of DN values is selected to be a new search origin [38]. Then, the search radius is reduced by half, i.e. S = 2, and repeats a similar search for two more iterations until the search radius is equal to 1. Finally, TSS finds a template image with the best match with the registered image. The computation time and number of search pixels are reduced. For the search range P = 7, the full search method requires x × y pixels whereas TSS only needs 25 (9 + 8 +8) pixels [39]. In Figure 5, the green, blue and orange pixels are the critical pixels in the first, second and third iterations, respectively.

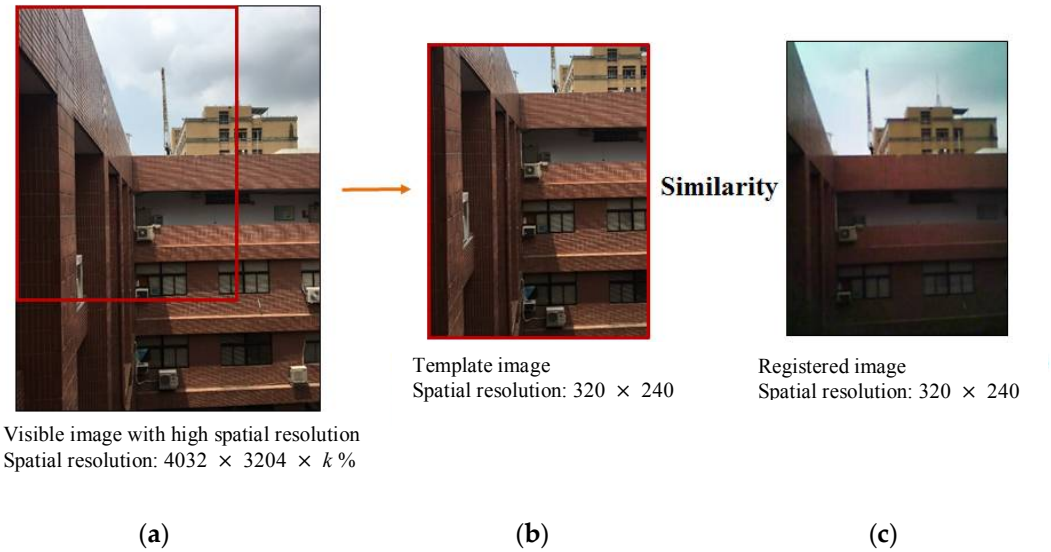
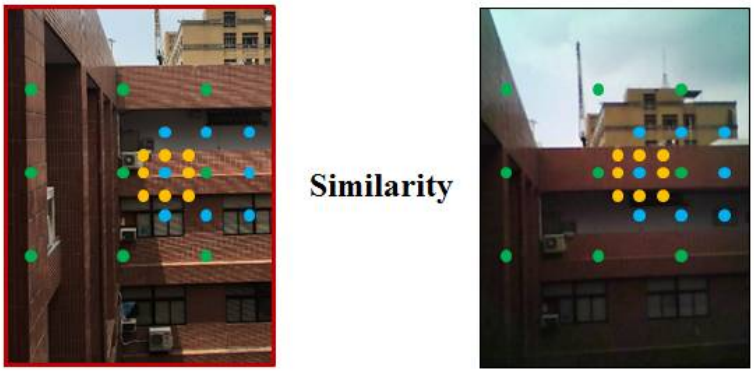


Figure 4. Illustration of image registration using NCC: (**a**) Original image of iPhone SE; (**b**) Template image; (**c**) Registered image.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003



(**a**) (**b**)

Figure 5. Illustration of TSS operation: (**a**) Template image; (**b**) Registered image.

3.2. 3D Thermal Model Reconstruction

A 3D thermal model reconstruction procedure includes extraction and matching of conjugate features, structure from motion, production of dense point cloud, creation of geometric entity, and temperature information texturing. The above steps are described below.

3.2.1. Extraction of Conjugate Features and Image Piecing

The conjugate feature extraction performance affects the creation of a geometric entity. Scale-invariant feature transform (SIFT), a popular computer vision technique, offers image matching with invariance of scale, rotation and brightness [5]; therefore, SIFT can extract appropriate and sufficient conjugate features from an enormous imagery dataset, and is useful in image piecing under any imaging condition.

3.2.2. Structure from Motion (SFM)

Based on the conjugate features extracted by SIFT, the camera stations can be determined according to the geometries of the epipolar planes; thus, a geometric relationship can be established to reconstruct the 3D model for an imaged scene. In order to optimize the geometric relationship, a bundle adjustment, based on the derived triangular points and camera stations, is employed in this study to repeatedly implement the collinearly forward or backward intersections until the residual errors are converged.

3.2.3. Production of Dense Point Cloud

After the bundle adjustment, the optimized triangular points are obtained and assembled as a point cloud. The preliminary point cloud is fairly sparse, so this study introduces a nearest-neighbor field (NNF) algorithm [40] for the production of a dense point cloud. In NNF, the first step is to build a few initial kernels based on several triangular points. The following step is to randomly move the kernels and simultaneously execute an imagery consistency test along with the moving kernels. The final step is to produce new points within the kernels, which can pass the imagery consistency test, and to repeat the second through final steps to improve the density of a point cloud.

3.2.4. Creation of Geometric Entity

Visual-Hull presented by Laurentini (1994) [41] is used to create the geometric 3D model. Visual-Hull assumes the foreground object in an image can be separated from the background, thus the original image can be segmented into a foreground/ background binary image, called a silhouette image. The foreground mask, known as a silhouette, is the 2D projection of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

corresponding 3D foreground object. Along with the camera viewing parameters, the silhouette defines a back-projected generalized cone that contains the actual object, so-called a silhouette cone.

3.2.5. Temperature Information Texturing

Using the image calibration described in Section 3.1, the relationship between the images acquired by the FLIR ONE and iPhone SEs is established. Consequently, the thermal IR images can be directly textured onto the surface of the created 3D model by the collinearity theory. The collinearity condition equations for temperature information texturing are expressed as follows:

$$\begin{cases} x = x_0 - f_1 \left[\frac{m_{11}(X-X_L)+m_{12}(Y-Y_L)+m_{13}(Z-Z_L)}{m_{31}(X-X_L)+m_{32}(Y-Y_L)+m_{33}(Z-Z_L)}\right] \\ y = y_0 - f_1 \left[\frac{m_{21}(X-X_L)+m_{22}(Y-Y_L)+m_{23}(Z-Z_L)}{m_{31}(X-X_L)+m_{32}(Y-Y_L)+m_{33}(Z-Z_L)}\right] \end{cases}, \quad (8)$$

where (x, y) is the space coordinate of an object point in a thermal IR image plane; (x_0, y_0) is the principal point of a thermal IR image; f_1 is the focal length of the FLIR ONE; (X_L, Y_L, Z_L) is the coordinate, with respect to the entity coordinate system XYZ, of exposure station of a thermal IR image; the m's are functions of the rotation angles along the XYZ axes of the FLIR ONE.

4. Results and Discussion

4.1. Geometric Translation of Double Lenses of FLIR ONE

Based on the ten pairs of thermal IR and visible images of the calibration template and an object distance of 2 m, the geometric translations in the column and row directions were calculated in Table 1. The average geometric translations in the column and row directions are 4.771 and -2.321 pixels, respectively. According to the average geometric translations, the phenomenon of relief displacement is effectively removed from the FLIR ONE thermal IR and visible images.

Table 1. Calculation results of the geometric translation.

ID of pair of images	Geometric translation in the column direction (pixels)	Geometric translation in the row direction (pixels)
IMG_L_1	3.579	-1.289
IMG_L_2	4.777	-1.095
IMG_L_3	6.805	-3.532
IMG_L_4	5.789	-3.193
IMG_L_5	3.153	-1.238
IMG_L_6	6.210	-2.764
IMG_L_7	4.446	-2.374
IMG_L_8	4.274	-2.737
IMG_L_9	4.498	-2.684
IMG_L_10	4.177	-2.305
Average	4.771	-2.321

4.2. Image Registration of Visible Images of FLIR ONE and iPhone SE Images

Down-sampling percentages from 8% to 50% were tested for the registration of the visible FLIR ONE and iPhone SE images. It is worth noting that the spatial resolution of iPhone SE images is lower than that of FLIR ONE images, if the down-sampling percentage is lower than 8%. Table 2 lists the RMSE values of the image registration for different down-sampling percentages, and shows that 9% down-sampling results in the best image registration. Consequently, this paper further tested the 9% down-sampling with the centesimal percentages to derive the RMSE values of the image

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

248 registration (see Table 3). The experiment results indicate the best down-sampling percentage of
249 9.24%.

250 **Table 2.** RMSE of image registration varying with down sampling percentages.

Down sampling percentage	RMSE (pixels)
8%	15.712
9%	3.403
10%	9.560
20%	211.380
30%	427.958
40%	856.391
50%	1064.691

251 **Table 3.** RMSE of image registration under 9% down sampling with centesimal percentages.

Down sampling percentage	RMSE (pixels)
9.20%	2.818
9.22%	2.738
9.24%	1.846
9.26%	2.041
9.28%	2.564
9.30%	2.629
9.40%	2.831
9.60%	4.230
9.80%	6.712

252

253 Three scenes of the exterior wall of the departmental building of Civil Engineering and
254 Environmental Engineering at National Chung Hsing University (NCHU), Taichung, Taiwan were
255 imaged and provided for an image registration test, as shown in Figure 6. Table 4 compares the
256 performances of the image registration by NCC with and without assistance from TSS. In addition,
257 the computation times required by the Intel(R) Core(TM) i7-3610QM CPU are listed in Table 4. NCC
258 without the assistance of TSS needs about 321 s to derive the NCC indices of approximately between
259 0.95 and 0.96. Nevertheless, it is demonstrated that TSS can effectively assist NCC in improving the
260 NCC indices from 0.95 to 0.99, and simultaneously reduce computation time significantly.

261 *4.3. 3D Thermal Model Reconstruction*

262 This section discusses the 3D thermal model reconstruction performance of the proposed
263 method for small, medium, and large-scale scenes and the computation time for each. A concrete
264 sample for strength testing was selected as the small-scale 3D thermal model reconstruction scene
265 (see Figure 7). The volume of the concrete sample was found to be 3662.5 and 3404.7 cm^3 by the
266 drainage method and the proposed method, respectively. Thus, the volume error of the 3D thermal
267 model reconstruction is 7.04%. Additionally, conventional IBM was also used to determine the
268 volume of the concrete sample. The determined volume and volume error are 3517.7 cm^3 and 3.95%,
269 respectively. In terms of 3D model reconstruction accuracy, the proposed method is slightly inferior
270 to the conventional IBM, but the 3D model reconstructed by the proposed method contains
271 temperature information. Consequently, the proposed method is demonstrated as a cost-effective
272 3D thermal model reconstruction method.

273

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

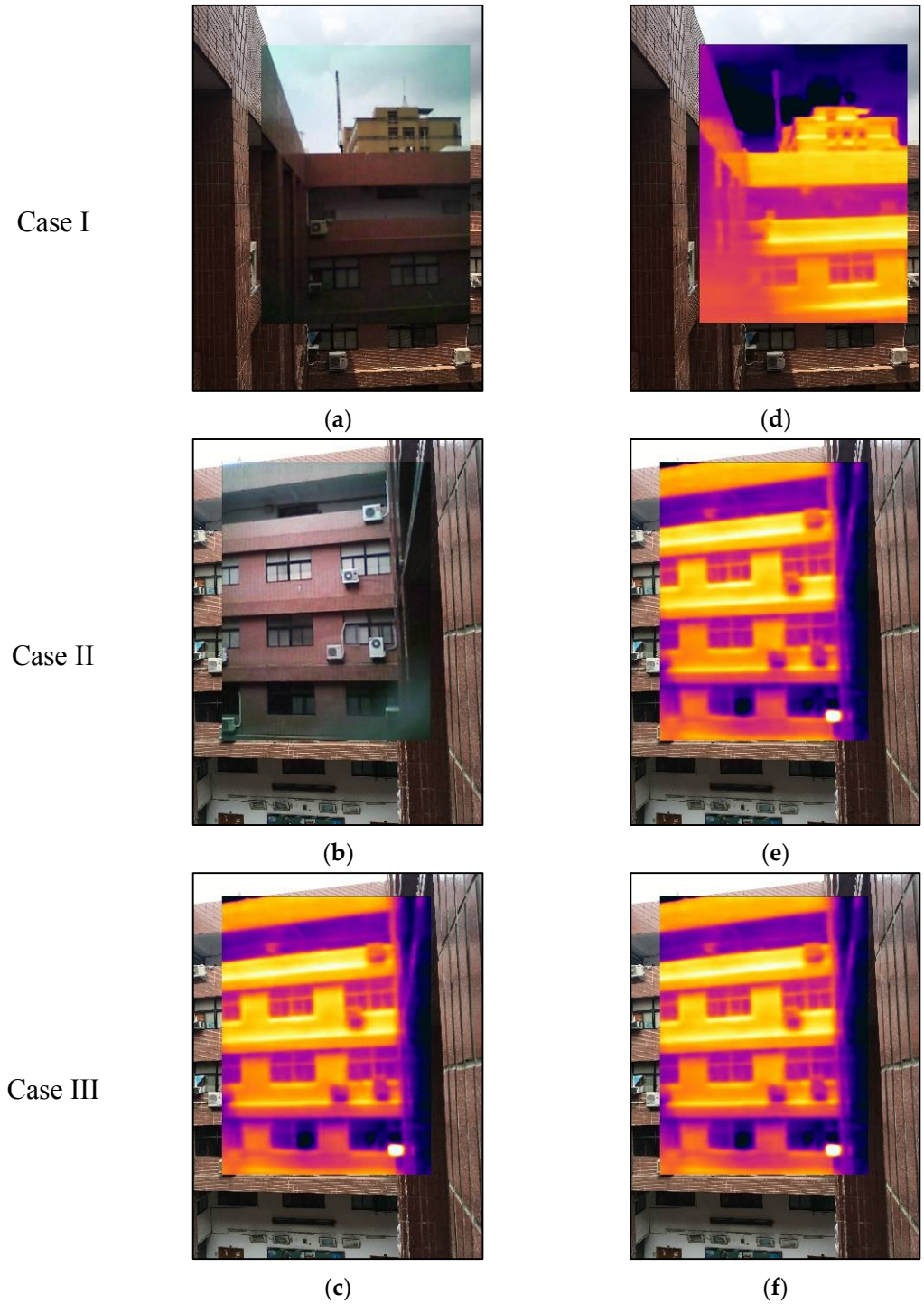
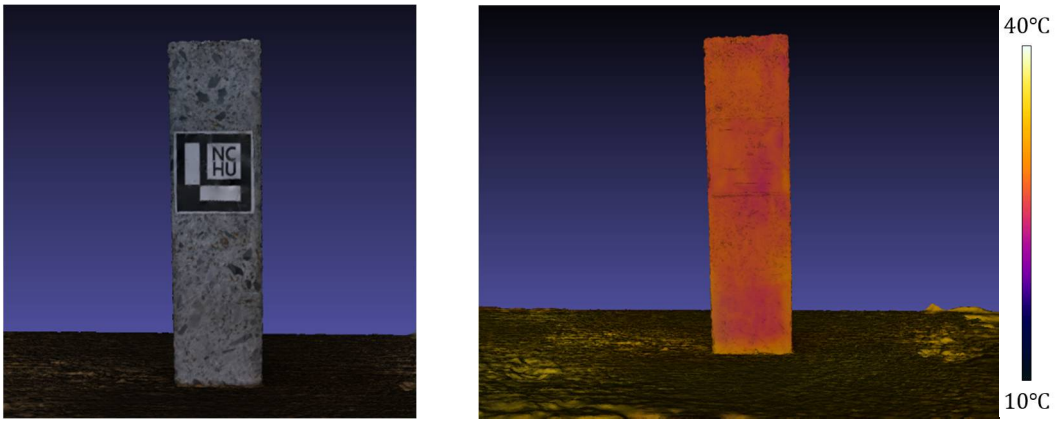


Figure 6. Image registration of exterior wall of departmental building: (**a**)-(**c**) Visible image of FLIR ONE registered for that of iPhone SE; (**d**)-(**f**) Thermal IR image of FLIR ONE registered for visible image of iPhone SE.

A classroom in the departmental building was selected as the medium-scale 3D thermal model reconstruction scene (see Figure 8). The 3D thermal model shows that the detected temperature of the classroom is between 35°and 45°, and the precision of the temperature information is ± 2° according to a temperature measurement *in situ*. The temperature of the glass curtain wall is over 40°, which is significantly higher than that of the interior of the classroom. If the classroom is intended to be part of a green building policy, the performances of proposed energy efficiency strategies can be presented and compared using this 3D thermal model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003



(**a**) (**b**)

Figure 7. Model reconstruction of small scale scene: (**a**) 3D model of concrete sample; (**b**) 3D thermal model of concrete sample.

Table 4. NCC indices and computation time of image registration.

Case	NCC		NCC + TSS	
	NCC index	Computation time (s)	NCC index	Computation time (s)
I	0.955	321.188	0.977	1.676
II	0.957	323.016	0.995	1.584
III	0.963	321.089	0.991	1.721

An exterior wall of the departmental building was selected as the large-scale 3D thermal model reconstruction scene (see Figure 9). The temperature of the tile wall is higher than that of the window glass. Moreover, the upper tile wall compared to the lower one seems to have a higher temperature. Due to the shading of the floors, the corridors have a lower temperature than the tile wall. In the winter of 2016, the departmental building experienced an extreme cold-front, causing several tiles to detach from the exterior wall. The proposed method could, in future, also provide multi-temporal 3D thermal models for exterior wall tests to pre-empt such deterioration.

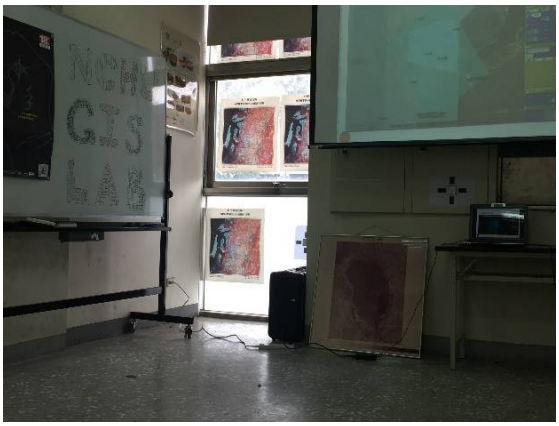
4.4. Computation Time

Table 5 lists the number of processed images and the required computation time for each step of the proposed scheme in Figure 2. The larger the scene, the longer the computation time requires; however, the total computation time does not seem to increase with an increase of the number of processed images. Table 5 indicates that the bottleneck step should be "creation of geometric entity" due to its needs of computation time significantly more than other steps.

Table 6 shows the qualities of the 3D models derived by structure from motion, production of dense point cloud, and creation of geometric entity. The production of a dense point cloud could produce a dozen times the number of points than by deriving structure from motion. Despite this, there is no significant relationship between the numbers of the dense points and the silhouette cones. In other words, only a portion of the dense points are useful in the creation of geometric entity, and most of the dense points may be superfluous. If the appropriate number of dense points required for the creation of a geometric entity can be determined in advance, the computation time of the creation of a geometric entity should be significantly reduced, thus reducing total computation time.

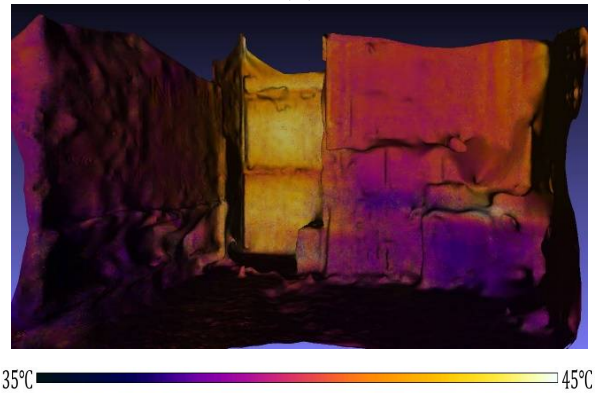
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003



(**a**)

(**b**)

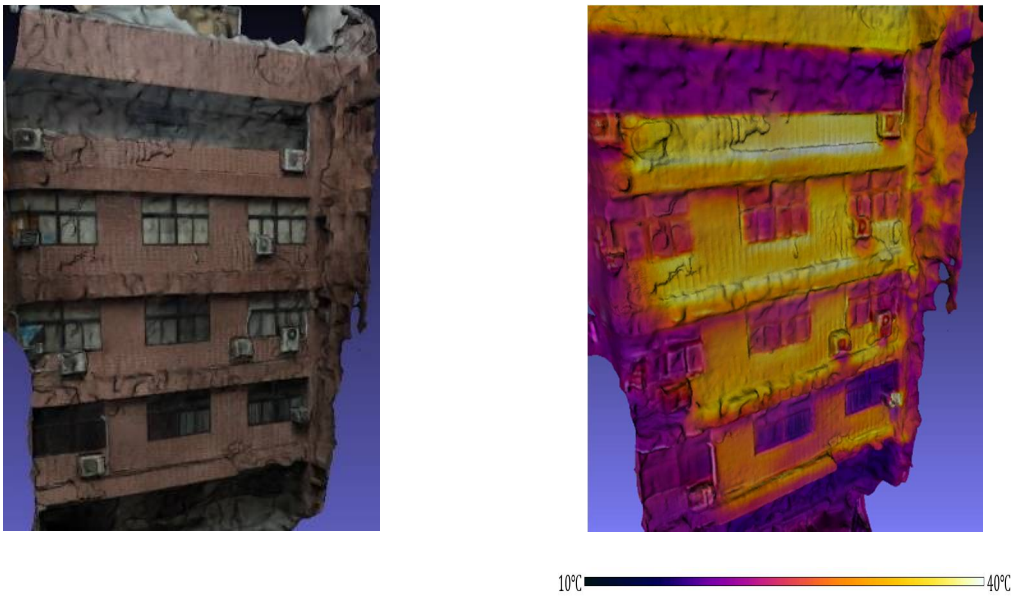


(**c**)

Figure 8. Model reconstruction of medial scale scene: (**a**) Visible image of classroom scene; (**b**) 3D model of classroom scene; (**c**) 3D thermal model of classroom scene.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003



(**a**) (**b**)

Figure 9. Model reconstruction of large scale scene: (a) 3D model of exterior wall of departmental building; (b) 3D thermal model of exterior wall of departmental building.

5. Conclusions

This paper presents an IBM-based method for 3D thermal model reconstruction; two smart phones and a low-cost thermal camera were employed to acquire visible images and thermal images, respectively, that was fused for constructing 3D thermal models. In the IBM-based method, image calibration, which includes geometric translation and image registration, is an important pre-processing for 3D thermal model reconstruction. It was demonstrated that TSS can effectively assist NCC in reducing the computation time required for image registration, and that the optimal percentage for down-sampling the smartphone image size to that of the thermal camera is 9.24%.

A small scene, i.e. concrete sample, was tested to reconstruct its 3D thermal model, and the obtained volume error of the 3D thermal model is 7.04%. A classroom within the departmental building and an exterior wall of the departmental building were tested to reconstruct their 3D thermal models. The experiment result shows that the obtained precision of the temperature information in the 3D thermal models is ± 2 ℃.

According to the required computation time of 3D thermal model reconstruction, the creation of a geometric entity, one of the steps of 3D thermal model reconstruction, is the critical step. Moreover, the production of superfluous points in the point cloud increases the computation time, reducing the efficiency of 3D thermal model reconstruction. At present, it has been demonstrated that the proposed method is cost-effective in 3D thermal model reconstruction. If the appropriate number of dense points required for the creation of a geometric entity can be determined in advance, the computation time of 3D thermal model reconstruction can be significantly reduced. Future work will continue to improve the method in order to further reduce computation time and to offer better model quality and in-time monitoring.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

339 **Table 5.** Computation time of 3D thermal model reconstruction.

	Step	Small scale scene		Medial scale scene		Large scale scene	
		No.(I) [1]	CT(s) [2]	No.(I) [1]	CT(s) [2]	No.(I) [1]	CT(s) [2]
1.	Image calibration		219		508		69
2.	Extraction of conjugate features and image piecing		165		231		192
3.	Structure from motion	81	486	202	1,686	51	1,091
4.	Production of dense point cloud		912		167		2,365
5.	Creation of geometric entity		2,650		31,453		36,323
6.	Temperature information projection		131		311		373
	Total		4,563		34,356		40,413

340 [1] Number of images; [2] Computation time.

341 **Table 6.** Qualities of 3D models derived by structure from motion, production of dense point cloud,
342 and creation of geometric entity.

Scene	Step	No. of points	No. of silhouette cones
Small scale	Structure from motion	46,911	-
	Production of dense point cloud	377,785	-
	Creation of geometric entity	-	753,627
Medial scale	Structure from motion	10,515	-
	Production of dense point cloud	270,368	-
	Creation of geometric entity	-	540,022
Large scale	Structure from motion	735,440	-
	Production of dense point cloud	13,587,715	-
	Creation of geometric entity	-	509,281

343

344 **Acknowledgments:** This work was financially supported in part by the "Innovation and Development Center
345 of Sustainable Agriculture" from The Featured Areas Research Center Program within the framework of the
346 Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

347 **Author Contributions:** All authors contributed to the main idea of this paper; Ming-Der Yang conducted the
348 experiment and analyzed the data; Tung-Ching Su and Hung-Yu Lin were in charge of encoding the algorithms;
349 Tung-Ching Su wrote the paper, which was revised by Ming-Der Yang.

350 **Conflicts of Interest:** The authors declare no conflict of interest.

351 **References**

352 1. Oliveira, M.M. Image-based modeling and rendering techniques: A survey. *RITA* **2002**, *9*, 37-66.
353 2. Yang, M.D.; Chao, C.F.; Lu, L.Y.; Huang, K.S.; Chen, Y.P. Image-based 3D scene reconstruction and
354 exploration in augmented reality. *Automation in Construction* **2013**, *33*, 48-60,
355 doi:10.1016/j.autcon.2012.09.017.
356 3. Nguyen, T.T.; Slaughter, D.C.; Max, N.; Maloof, J.N.; Sinha, N. Structured light-based 3D reconstruction
357 system for plants. *Sensors* **2015**, *15(8)*, 18587-1861, doi:10.3390/s150818587.
358 4. Yandún Narváez, F.J.; Salvo del Pedregal, J.; Prieto, P.A.; Torres-Torriti, M.; Auat Cheein, F.A. LiDAR and
359 thermal images fusion for ground-based 3D characterisation of fruit trees. *Biosystems Engineering* **2016**, *151*,
360 479-494, doi:10.1016/j.biosystemseng.2016.10.012.
361 5. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer*
362 *Vision* **2004**, 1-28.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

6. Yang, Z. Fast Template Matching Based on Normalized Cross Correlation with Centroid Bounding. Proceedings of 2010 International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha City, China, 13-14 March 2010.
7. Wu, C. Towards Linear-time Incremental Structure From Motion. Proceedings of 2013 International Conference on 3D Vision - 3DV 2013, Seattle, WA, USA, June 29 2013-July 1 2013.
8. Jancosek, M.; Pajdla, T. Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. *International Scholarly Research Notices* **2014**, *2014*, 1-20, doi:10.1155/2014/798595.
9. Giuliani, G.; Dao, H.; De Bono, A.; Chatenoux, B.; Allenbach, K.; De Laborie, P.; Rodila, D.; Alexandris, N.; Peduzzi, P. Live monitoring of earth surface (LiMES): A framework for monitoring environmental changes from earth observations. *Remote Sensing of Environment* **2017**, *202*, 222-233, doi: 10.1016/j.rse.2017.05.040.
10. Manzo, C.; Mei, A.; Zampetti, E.; Bassani, C.; Paciucci, L.; Manetti, P. Top-down approach from satellite to terrestrial rover application for environmental monitoring of landfills. *Science of The Total Environment* **2017**, *584-585*, 1333-1348, doi:10.1016/j.scitotenv.2017.01.033 .
11. Gulbe, L.; Caune, V.; Korats, G. Urban area thermal monitoring: Liepaja case study using satellite and aerial thermal data. *International Journal of Applied Earth Observation and Geoinformation* **2017**, *63*, 45-54, doi:10.1016/j.jag.2017.07.005.
12. Roca, D.; Lagüela, S.; Díaz-Vilariño, L.; Armesto, J.; Arias, P. Low-cost aerial unit for outdoor inspection of building façades. *Automation in Construction* **2013**, *36*, 128-135, doi: 10.1016/j.autcon.2013.08.020.
13. Natephra, W.; Motamedi, A.; Fukuda, T.; Yabuki, N. Integrating building information modeling and virtual reality development engines for building indoor lighting design. *Visualization in Engineering* **2017**, *5*, 1-21, doi:10.1186/s40327-017-0058-x.
14. Iwaszczuk, D.; Stilla, U. Camera pose refinement by matching uncertain 3D building models with thermal infrared image sequences for high quality texture extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *132*, 33-47, doi:10.1016/j.isprsjprs.2017.08.006.
15. Rea, P.; Ottaviano, E. Design and development of an Inspection Robotic System for indoor applications. *Robotics and Computer-Integrated Manufacturing* **2018**, *49*, 143-151, doi:10.1016/j.rcim.2017.06.005.
16. Merchán, P.; Adán, A.; Salamanca. S.; Domínguez, V.; Chacón. R. Geometric and colour data fusion for outdoor 3D models. *Sensors* **2012**, *12(6)*, 6893-6919, doi:10.3390/s120606893.
17. Costanzo, A.; Minasi, M.; Casula, G.; Musacchio, M.; Buongiorno, M.F. Combined use of terrestrial laser scanning and IR thermography applied to a historical building. *Sensors* **2015**, *15(1)*, 194-213, doi:10.3390/s150100194.
18. Wang, X.; Witz, J.F.; Bartali, A.E.; Jiang, C. Infrared thermography coupled with digital image correlation in studying plastic deformation on the mesoscale level. *Optics and Lasers in Engineering* **2016**, *86*, 264-274, doi:10.1016/j.optlaseng.2016.06.001.
19. Capozzoli, L.; Rizzo, E. Combined NDT techniques in civil engineering applications: Laboratory and real test. *Construction and Building Materials* **2017**, *154*, 1139-1150, doi:10.1016/j.conbuildmat.2017.07.147.
20. Yang, M.D.; Su, T.C. Segmenting ideal morphologies of sewer pipe defects on CCTV images for automated diagnosis. *Expert Systems with Applications* **2009**, *36(2)*, 3562-3573, doi:10.1016/j.eswa.2008.02.006.
21. Yang, M.D.; Su, T.C.; Pan, N.F.; Yang, Y.F. Systematic image quality assessment for sewer inspection. *Expert Systems with Applications* **2011**, *38*, 1766-1776, doi:10.1016/j.eswa.2010.07.103.
22. Yang, M.D.; Su, T.C.; Pan, N.F.; Liu, P. Feature extraction of sewer pipe defects using wavelet transform and co-occurrence matrix. *International Journal of Wavelets Multiresolution and Information Processing* **2011**, *9(2)*, 211-225, doi:10.1142/S0219691311004055.
23. Su, T.C.; Yang, M.D.; Wu, T.C.; Lin, J.Y. Morphological segmentation based on edge detection for sewer pipe defects on CCTV images. *Expert Systems with Applications* **2011**, *38(10)*, 13094-13114, doi:10.1016/j.eswa.2011.04.116.
24. Lagüela, S.; Armesto, J.; Arias, P.; Herráez, J. Automation of thermographic 3D modelling through image fusion and image matching techniques. *Automation in Construction* **2012**, *27*, 24-31, doi:10.1016/j.autcon.2012.05.011.
25. Ham, Y.; Golparvar-Fard, M. An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery. *Advanced Engineering Informatics* **2013**, *27(3)*, 395-409, doi: 10.1016/j.aei.2013.03.005.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 May 2018 doi:10.20944/preprints201805.0225.v1

Peer-reviewed version available at Sensors 2018, 18, 2003; doi:10.3390/s18072003

26. Rangel, J.; Soldan, S.; Kroll, A. 3D Thermal Imaging: Fusion of Thermography and Depth Cameras. Proceedings of the 12th International Conference on Quantitative InfraRed Thermography. QIRT 2014, Bordeaux, France, 7-11 July 2014.
27. Adán, A.; Prado, T.; Prieto, S.A.; Quintana, B. Fusion of Thermal Imagery and LiDAR Data for Generating TBIM Models. Proceedings of 2017 IEEE Sensors, Glasgow, UK, 29 Oct.-1 Nov. 2017, doi: 10.1109/ICSENS.2017.8234261.
28. Schramm, S.; Rangel, J.; Kroll, A. Data Fusion for 3D Thermal Imaging Using Depth and Stereo Camera for Robust Self-localization. Proceedings of 2018 IEEE Sensors Applications Symposium (SAS), Seoul, Korea (South), 12-14 March 2018, doi: 10.1109/SAS.2018.8336740.
29. Kazi, A.M.; Ali, M.; Ayub, K.; Kalimuddin, H.; Zubair, K.; Kazi, A.N.; Artani, A.; Ali, S.A. Geo-spatial reporting for monitoring of household immunization coverage through mobile phones: Findings from a feasibility study. *International Journal of Medical Informatics* **2017**, *107*, 48-55, doi:10.1016/j.ijmedinf.2017.09.004.
30. Matarazzo, T.; Vazifeh, M.; Pakzad, S.; Santi, P.; Ratti, C. Smartphone data streams for bridge health monitoring. *Procedia Engineering* **2017**, *199*, 966-971, doi:10.1016/j.proeng.2017.09.203.
31. Zhang, H.; Wei, Q.; Jiang, Z. 3D reconstruction of space objects from multi-views by a visible sensor. *Sensors* **2017**, *17(7)*, 1689, doi:10.3390/s17071689
32. Lu, Y.; Wu, Z.; Chang, R.; Li, Y. Building Information Modeling (BIM) for green buildings: A critical review and future directions. *Automation in Construction* **2017**, *83*, 134-148, doi: 10.1016/j.autcon.2017.08.024.
33. Lin, Y.H.; Tsai, K.T.; Lin, M.D.; Yang, M.D. Design optimization of office building envelope configurations for energy conservation. *Applied Energy* **2016**, *171*, 336-346, doi:10.1016/j.apenergy.2016.03.018.
34. Yang, M.D.; Lin, M.D.; Lin, Y.H.; Tsai, K.T. Multiobjective optimization design of green building envelope material using a non-dominated sorting genetic algorithm. *Applied Thermal Engineering* **2017**, *111*, 1255-1264, doi: 10.1016/j.applthermaleng.2016.01.015.
35. Rajeev, P.; Sanjayan, J.G.; Seenuth, S.S. Assessment of thermal cracking in concrete roof tiles. *Materials & Design* **2016**, *107*, 470-477, doi:10.1016/j.matdes.2016.06.072.
36. Li, R.; Zeng, B.; Liou, M.L. A new three-step search algorithm for block motion estimation, *IEEE Transactions on Circuits and Systems for Video Technology* **1994**, *4(4)*, 438-442, doi:10.1109/76.313138.
37. Bracewell, R. *The Fourier Transform and Its Applications*, 1st ed.; McGraw-Hill: New York, USA, 1965, pp. 46-243.
38. Yu, H.; Lin, Z.; Pan, F. Applications and improve of H.264 in medical video compression. *IEEE Transactions on Circuits and Systems-I* **2005**, *52(12)*, 2707-2716, doi:10.1109/TCSI.2005.857869.
39. Lakamsani, P.; Zeng, B.; Liou, M. An Enhanced Three Step Search Motion Estimation Method and Its VLSI Architecture. Proceedings of 1996 IEEE International Symposium on Circuits and Systems. Circuits and Systems Connecting the World. ISCAS 96, Atlanta, GA, USA, 15 May 1996.
40. Barnes, C.; Shechtman, E.; Finkelstein, A.; Goldman, D.B. A Randomized Correspondence Algorithm for Structural Image Editing. Proceedings of ACM SIGGRAPH 2009, New York, NY, USA, 3-7 August 2009.
41. Laurentini, A. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1994**, *16(2)*, 150-162, doi:10.1109/34.273735.