Automation in Construction 33 (2013) 48–60

Contents lists available at SciVerse ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Image-based 3D scene reconstruction and exploration in augmented reality

Ming-Der Yang [a,*], Chih-Fan Chao [a], Kai-Siang Huang [a], Liang-You Lu [a], Yi-Ping Chen [b]

[a] Department of Civil Engineering, National Chung Hsing University, 250 Kuokuang Rd., Taichung 402, Taiwan
[b] Department of Management, Da-Yeh University, 168 University Rd., Dacun, Changhwa 515, Taiwan

ARTICLE INFO

Article history:
Accepted 29 September 2012
Available online 10 November 2012

Keywords:
Augmented reality
3D reconstruction
Feature extraction
Corresponding feature
Structure from motion

ABSTRACT

Augmented Reality (AR) is an integrated technique of image processing and display system of complex information, which involves real-time computing, motion tracking, pattern recognition, image projection, database linking, feature extraction, and coordinate transformation. In this study, such techniques through Structure From Motion (SFM), Clustering Views for Multi-View Stereo (CMVS), Patch-based Multi-View Stereo (PMVS), and Poisson surface reconstruction were smoothly integrated into a 3D reconstruction system with comparative efficiency in computation. The system can be applied to regular images taken by amateur cameras, smart phones, tablet PCs, and other mobile devices, without need of a priori internal and external camera parameters. To verify the ability of the established reconstruction system, indoor and outdoor objects at various scales, such as a helmet (a small object), a corridor (an indoor medium object), an arbor (outdoor medium object), and a building (outdoor large object) were tested. Through tracking and registration, the reconstructed 3D models were loaded in an AR environment to facilitate displaying, interacting, and rendering that provides AR applications in construction design and management for better and qualitative communication in economical and handy ways.

1. Introduction

Three-dimensional (3D) reconstruction techniques have been developed with substantial effort and can be conducted by either traditional surveying or novel 3D modeling systems. Traditionally, remotely sensed reconstruction techniques include two major methods, such as airborne image photogrammetry and LiDAR (Light Detection And Ranging). Photogrammetry is a sophisticated technique for 3D reconstruction [1], but it is high in time and financial costs. In previous decade, a LiDAR technique has been developed to build a Digital Elevation Model (DEM) with time-efficiency and high accuracy, but the cost of equipment and the difficulty of cloud point processing must still be improved [2].

To overcome the disadvantages of traditional reconstruction methods, computer visual techniques have been introduced into 3D modeling systems in recent years. Such techniques are Simultaneous Localization And Mapping (SLAM) [3], Project Photofly from Autodesk [4], Multi-View Stereo (MVS) [5], Photo Tourism [6], Bundler [7], Clustering Views for Multi-View Stereo (CMVS) [8,9], Patch-based Multi-View Stereo-Version 2 (PMVS2) [10,11], and ARC 3D Webservice [12]. All of the aforementioned techniques adopt computer visual techniques to reconstruct 3D models without requiring photogrammetry information, such as interior and exterior orientations.

Image-Based Modeling (IBM) is considered as a passive system and includes three major methods, such as depth-map-based approaches, volume-based approaches, and surface-based approaches. One of the major surface-based approaches to obtain a 3D structure in the state of motion is called Structure from Motion (SFM), which was first introduced to search for the same feature points in different images, which recover the imaging scene and estimate the location and orientation of the camera. Once the camera orientation parameters are available, the 3D coordinates of the cameras and the image-based point cloud can be found. The main principle of the SFM is to establish the relationship between the different images by analyzing the relative and absolute positions of the field image. Tristrom and Robert used various feature detection techniques, such as Harris, SUSAN, and the Scale Invariant Feature Transform (SIFT) algorithm for corner and feature detection [13,14]. Gabriel et al. introduced a novel method to identify 3D objects using the point cloud of SFM because of its adaptability to construct an acceptable model structure in any scattered level of the cloud point structure [15]. Without requiring any observed measurement, SFM is able to identify location, orientation, and geometry from images [16]. Once the interior and exterior parameters (such as focal length, location and orientation of camera, and object coordinates) are known, Augmented Reality (AR) provides the functions of tracking and registration,

* Corresponding author. Tel.: +886 4 22840440x214; fax: +886 4 22862857.
E-mail addresses: mdyang@dragon.nchu.edu.tw (M.-D. Yang), ven.evanscence@gmail.com (C.-F. Chao), d9762402@mail.nchu.edu.tw (K.-S. Huang), g099062076@mail.nchu.edu.tw (L.-Y. Lu), ypchen@mail.dyu.edu.tw (Y.-P. Chen).

0926-5805/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.autcon.2012.09.017

displaying, rendering, interacting devices and techniques, and presenting and authoring the reconstructed 3D model [17].

The AR system has been developed for more than 40 years since ultimate display was introduced in 1965 [18]. In 1968, the first AR system was created using an optical see-through head-mounted display and was tracked using one of two different 6 Degrees of Freedom (DoF) trackers; however, only extremely simple wireframe drawings could be displayed in real time [19]. Following the inventions of the central processing unit (CPU) and smart-phone in the 1990s, AR was coined to refer to overlaying computer-presented materials on top of the real world [20]. With the development of the webcam and Internet, Milgram and Kishino described a continuum that spans from Augmented Virtuality, which is closer to the virtual environment, to Augmented Reality, which is closer to the real environment [21]. In the previous decade, AR was applied in many fields, such as the GPS (Global Positioning System) receiver, a head-worn electronic compass and GIS (Geographic Information System) database combined as a personal guide [22], and the Chameleon system displaying spatially situated information using a tracked hand-held device [23]. Similar to the Chameleon system, NaviCam used a nearby powerful workstation with a camera mounted on the mobile screen for optical tracking to detect color-coded markers on the live camera image and display context sensitive information on top of the video feed in a see-through manner [24]. 2D matrix markers were introduced to allow camera tracking with 6 DoF [25]. The preliminary spatial information survey in AR provided a widely acknowledged definition [26], and a touring machine was presented by equipping the first mobile augmented reality system (MARS) [27]. Starner et al. created an interconnecting network community for users equipped with wearable computers and explored possible applications of mobile AR, such as an information system for offices, people recognition, and coarse localization using infrared beacons [28]. A backpack-based wearable computer was introduced to construct an AR platform used for several AR projects [29]. The ARToolKit was a pose tracking library with 6 DoF using square fiducials and a template-based approach for recognition as an open source under the GPL license [30,31]. Since the first GSM (Global System for Mobile communications) phone with a built-in GPS receiver and WiFi were released in 1999, GPS and the RTK technique were integrated in a mobile AR system [32]. In 2000s, the synergy of AR created new possibilities in the field of 3D data visualization, navigation, and interaction far beyond traditional static navigation and interaction [33]. Many previous studies have combined the smart-phone, GPS, GIS database, and AR technology to achieve a new 3D world, such as the AR-Quake game [34], a mobile Passive Augmented Reality Device (mPARD) [35], Battlefield Augmented Reality System (BARS) [36], PDA-based BatPortal localizing by measuring the travel time of ultra-sonic pulses between specially built devices worn by the users [37], TOWNWEAR using a fiber optic gyroscope for orientation tracking [38], and a mobile and multi-user AR system [39]. Currently, AR techniques are closer to reality because of the development of mobile AR applications and mobile entertainment games [40–42]. In recent studies, AR has been established in mobile phones as institution guide [43]. Furthermore, additional new techniques and devices have been added to enhance and extend AR applications. The first real-time 6 DoF implementation of natural feature tracking heavily modified the well-known SIFT and Ferns methods to accelerate processing and reduce memory requirements [44]. MapLens is a mobile AR map using a magic lens over a paper map to facilitate place-making by creating a constant need for referencing to the physical world [45]. In surveying applications, 3D virtual rectangular objects with a scale were located on the grid of 3D geographical model, and were displayed in an overlapping manner with the actual landscape from multiple viewpoints using the AR technology [46].

Although image-based point cloud models are less accurate compared to laser scanner point cloud models, it provides the opportunity of automatically visualizing the as-built scene through geo-registered site photographs with much less effort and no cost [16]. Although there are some 3D scene reconstruction products in the market, such as Acute3D and Pix4D, the reconstruction processes of those commercial products are unreleased to the public so that great cost and specific equipment are often required for 3D modeling. To provide economic acquisition equipment and low expertise requirement, this paper employs open sources of multiple image-based modeling approaches to establish an efficient 3D reconstruction process and displays the reconstructed 3D models in an AR environment for construction management. Since AR is considered as a new technique of image processing and displays system of complex information, which involves real-time computing, motion tracking, pattern recognition, image projection, database linking, feature extraction, and coordinate transformation, this study integrates these techniques situated in a reality–virtuality continuum, as shown in Fig. 1, to achieve real-time, time efficient, cost economic, user friendly, and easy-carrying goals.

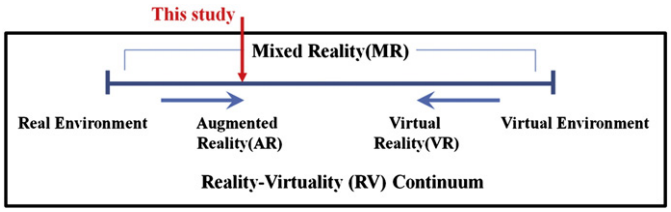

Fig. 1. Role of this research playing in reality–virtuality continuum.

2. Methodology

As shown in Fig. 2, SFM, CVMS, PMVS, and Poisson Surface Reconstruction (PSR) [47] were adopted to reconstruct a real 3D model. The proposed 3D reconstruction method was applied to four indoor and outdoor test sites at various scales, such as a helmet, a corridor, an arbor, and a building, which were displayed in the AR system. The details of the process are addressed as follows.

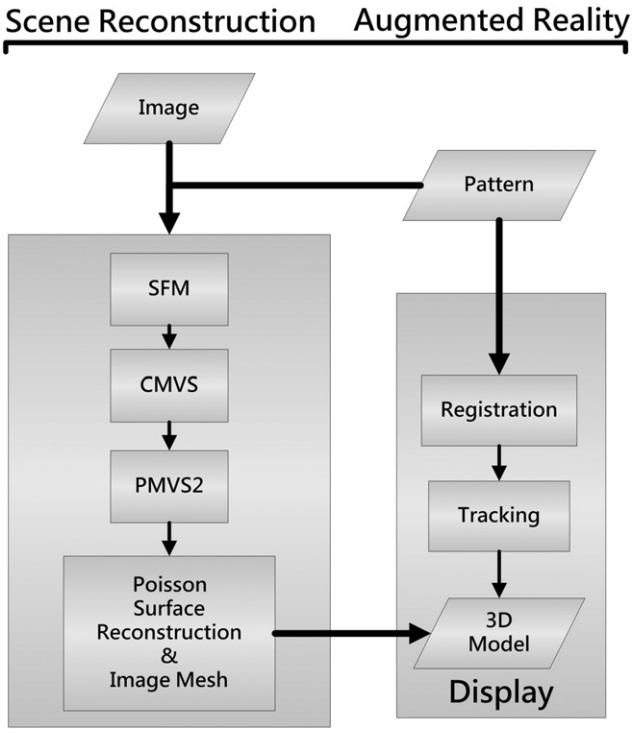

Fig. 2. Schema of the proposed system.

2.1. Structure from motion

Photogrammetry is the main method of the multi-perspective technique to solve the 3D coordination of field points using numerous images. In traditional photogrammetry, if the image orientation is known, the ground coordination can be obtained through aerial triangulation; by contrast, if the ground 3D coordination is unknown, the image orientation should be inversely transformed and calculated. Computer vision and photogrammetry techniques, such as SIFT, RANSAC (RANdom Sample Consensus) [48], and bundle adjustment, were introduced to automatically calculate the point relationships and simultaneously solve the orientation of camera exposure stations and scene structure. Integration of photogrammetry and computer vision is fundamental in Structure From Motion (SFM). Without requiring any observed measurement, SFM is able to identify the 3D camera position and scene geometry by determining the configuration of cameras and 3D points. As shown in Fig. 3, this study searched for the feature points between each pair of images, constructed the relationship of feature points, and computed a series of corresponding fundamental matrix. Based on the fundamental matrix, the best pair images can be considered as the initial images for reconstruction, and then new cameras are iteratively added into sparse bundle adjustment [49] to optimize the parameters and ultimately obtain scattered cloud points.

Structure From Motion has three main tasks: analyzing camera motion and target shooting, recovering the camera moving track, and finally constructing target objects in a 3D scene [50]. The process of SFM in Fig. 4 involves matching the first image to another, repeating the matching process for all images, and deleting mismatched images to obtain the relative camera position.

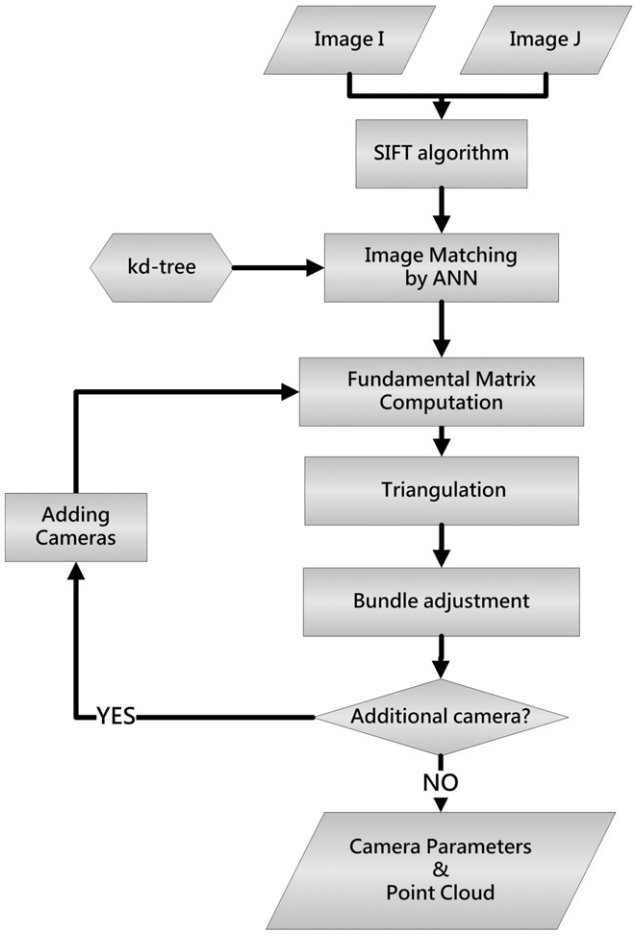

Fig. 3. Flowchart of SFM.

2.1.1. Feature detection by SIFT

The SIFT algorithm extracts feature points for fundamental attributes in experimental images, and records the corresponding images in the database for tracking process and camera continuous motion calculation. Because of highly significant and relatively easy to capture feature points, the SIFT algorithm is commonly used in computer vision. Based on the appearance of local points of interest, the SIFT algorithm can effectively resist the size and rotation difference of images. The tolerance of SIFT in light, noise, view and zoom-in/out is quite high, and can accurately identify object features in the large number of non-parametric database [51].

However, a time-consuming and inefficient problem remains in SIFT implementation, because the SIFT algorithm tries to find a great amount of feature points while a large region is reconstructed. To resolve this problem, this study used GPU (Graphics Processing Unit) to program the SIFT process in a GPU system as SIFTGPU [52].

2.1.2. Feature matching by ANN

In the feature matching process, tracking images according to SIFT-detected feature points is employed to observe the field and the camera motion relationship and build correspondence [51]. For every pair of images I and J, $F(I)$ is the set of features in image I. Let each feature $fi \in F(I)$ and its corresponding feature $fi' \in F(J)$ find its nearest neighbor $f_{nn} \in F(J)$ [53] by

$$f_{nn} = \arg\min_{f' \in F(J)} \|f_i - f_i'\|_2. \tag{1}$$

To raise the efficiency of feature matching, this study adopted ANN (Approximate Nearest Neighbors) library established by Arya and Mount [54] to construct the relationship of feature points between image pairs, so-called track. The feature points in image J are adopted to create less than 200 bins in kd-Tree data structure and search for the nearest points from J to I. Having a candidate pair of matching features (f, f_{nn}), one can calculate the nearest and the second nearest distances between the features, $d1$ and $d2$, respectively. A ratio test ($d1/d2$) proposed by Lowe [51] was adopted as a threshold instead of the nearest neighbor distance to cut improper corresponding features off. In this paper, a pair of corresponding features with a ratio of $d1/d2$ greater than 0.6 are excluded in the correspondence track for the following 3D reconstruction. After matching features I to J, each feature $fi \in F(I)$ was paired with at most one feature in $F(J)$.

2.1.3. Fundamental matrix computation

The EXIF (Exchangeable Image File) format, which is bundled with digital photos by most cameras, provides image information that can preliminarily supply approximations to calculate the possible configurations of the two cameras and their corresponding *epipolar geometry* [55]. The *epipolar geometry* of an image pair can be expressed in a 3×3 rank-2 matrix, so-called fundamental matrix (**F** or F-matrix), which describes the relative positions and orientations of the two cameras as well as internal camera settings such as zoom. Each pair of the corresponding points $(x, y) \rightarrow (x', y')$ in two corresponding images must satisfy the *epipolar constraint* [55]:

$$x' F x^T = 0 \tag{2}$$

$$F \sim K^{-1T} [T]_\times R K'^{-1} \tag{3}$$

where $x = [x\ y\ 1], x' = [x'\ y'\ 1]$. **R** is the rotation matrix, and **K** is the intrinsic camera matrix as

$$K = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ 0 & k_{22} & k_{23} \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

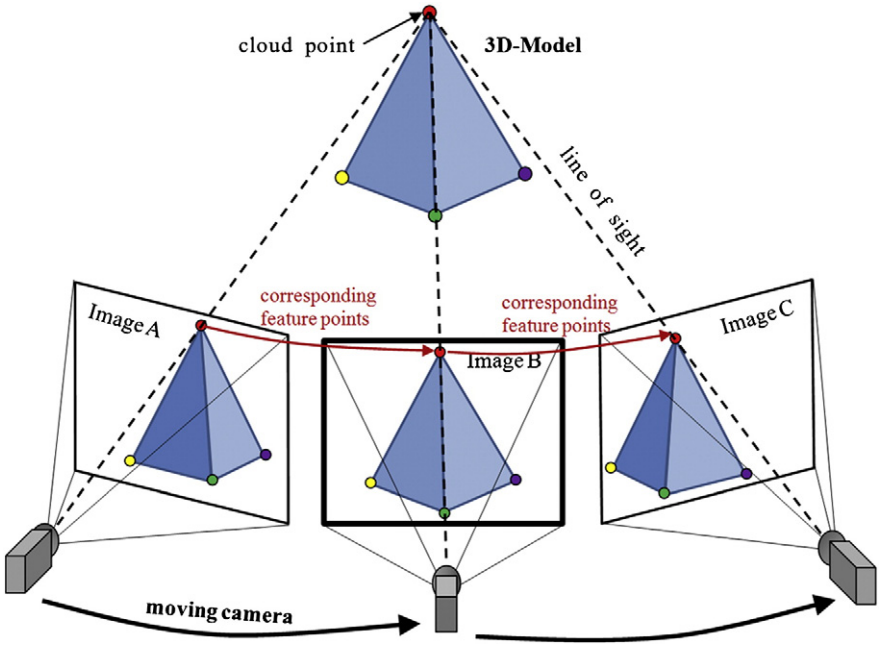

Fig. 4. SFM schematic diagram.

The focal length ratio (f_k) can be estimated using EXIF or k as [6]

$$\begin{cases} f_k = f_{EXIF}, & \text{if } 0.7f_k < f_{EXIF} < 1.4f_k, \\ f_k = \dfrac{k_{11}+k_{22}}{2}, & \text{otherwise.} \end{cases} \tag{5}$$

In addition, $T = [t_x t_y t_z]^T$ is a 3-vector presenting to offset of the camera center as:

$$[T]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \tag{6}$$

Obviously, the F-matrix contains the camera intrinsic matrix and extrinsic matrix. This study adopted RANSAC [48] to determine the F-matrix for all image pairs based on the estimate provided by an eight-point algorithm [55] that was reported to be efficient in regularization problems and noise improvement [56]. The F-matrix computation is shown in Fig. 5. In this paper, the outlier threshold of the RANSAC is set as a mapping error in pixel greater than 0.6% of the maximum of the image width and height to determine if each individual feature match is consistent with the F-matrix [6]. After removing the outlier points, the remaining features were added into RANSAC for iteration until the determination of F-matrix approached 1. The correspondence of the qualified features in all images was stored in the track.

2.1.4. Triangulation

Triangulation involves reconstructing 3D coordinates of the corresponding features by detecting the motion tracking orientation of the features. The more overlap that exists between two images (or the longer the baseline that exists), the more favorable the reconstruction result. To enhance the coordinate accuracy, homography [55] was adopted to screen the corresponding features identified using the F-matrix in 2D pair images by defining a cut-off ratio of a mapping error over the maximum of the image width and height as 0.4% [6]. Direct Linear Transform (DLT) technique described in Hartley's [55] can initialize the exterior parameters of new camera for five-point algorithm [57] to establish the homography of an image pair in RANSAC. At least 100 corresponding features are needed to make an image pair qualified for the following SFM process.

The SFM detects the camera parameters, such as interior and exterior matrices, and 3D coordinates of features, and insists that the F-matrix be the same in the reprojection error. By minimizing the sum of the distances between every track point projection to the correspondence as shown in Fig. 6, the objective function can be [6]:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m} \omega ij \left\| \boldsymbol{q}ij - P\left(C_i, X_j\right) \right\|^2 \tag{7}$$

$$P\left(C_i, X_j\right) = K_i [R_i | T_i] X_j \tag{8}$$

where a set of camera parameters, $C = \{C_1, C_2, \ldots, C_n\}$, consists of the parameter collection of a camera, $C_i = \{c_i, R_i, f_i, k1_i, k2_i\}$ and $X = \{X_1, X_2, \ldots X_m\}$, and 3D point coordinate, $X_j = (X_{jx}, X_{jy}, X_{jz})$, is projected to a point P in 2D image I. ω_{ij} is an indicator, where $\omega_{ij} = 1$ if camera i observes point j otherwise $\omega_{ij} = 0$. By minimizing the distance between the calculated point P and the observed point q_{ij}, the 3D

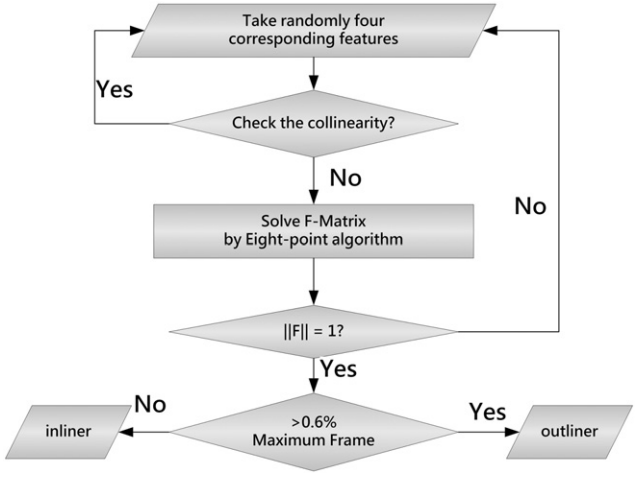

Fig. 5. F-Matrix computation.

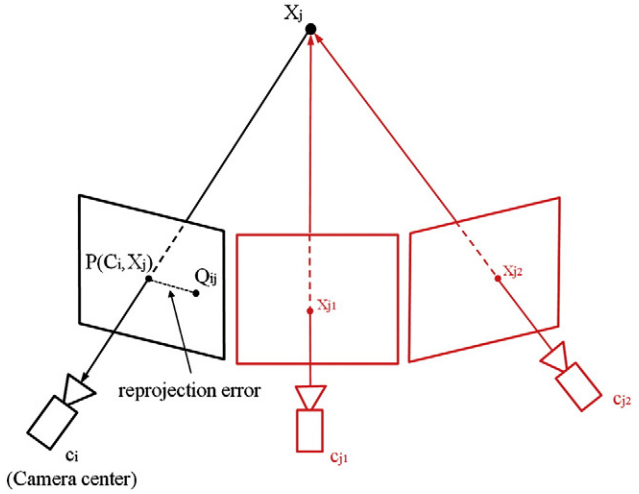

Fig. 6. Reprojection error.

coordinates of the features and camera parameters can be modified. Radial distortion of each camera affects the accuracy of the reconstruction. $k1$ and $k2$ are distortion parameters that can be modeled in a quartic polynomial for many consumer cameras and lenses that produce images with noticeable non-linear distortions. This study estimated the radial distortion parameter for each camera using a 2D mapping function from p (p_x, p_y) to a distortion point $p'(x', y')$ [6]:

$$p' = ap, a = k1\rho^2 + k2\rho^4, \rho^2 = \left(\frac{p_x}{f}\right)^2 + \left(\frac{p_y}{f}\right)^2 \tag{9}$$

where f is the estimated focal length. While initializing a new camera, $k1$ and $k2$ can be set as 0.

2.1.5. Bundle adjustment

Because of the rotation, perspective division, and radial distortion, resolving SFM becomes a non-linear problem. Generally, minimization of the reprojection error through non-linear least squares is a process known as bundle adjustment. Bundle adjustment is used to optimize the point position estimation in SFM using a multiple overlap image mutual restraint and a collinearity formula that is iteratively repeated to obtain point coordination until numerical convergence [58]. Eventually, all interior parameters can be modified through bundle adjustment.

Although the Levenberg–Marquardt algorithm can solve the non-linear problem, it only ensures the local minimum and might possibly miss the global minimum in the large-scale SFM problem. To overcome the localization crisis, this study estimated camera parameters by progressively adding one camera every time instead of estimating all camera parameters at once. This study used the Sparse Bundle Adjustment (SBA) package provided by Lourakis and Argyros [49], which is based on the Levenberg–Marquardt method.

The camera exterior matrix includes a 3×3 rotation matrix ($\mathbf{R}$) and a translation vector ($\mathbf{T}$). To reduce the parameters, an incremental rotation $R(\theta, \hat{n})$ is defined as [6]:

$$R(\theta,\hat{n}) = I + \sin\theta[\hat{n}]_\times + (1-\cos\theta)[\hat{n}]_\times^2, \omega = \theta\hat{n}, [\hat{n}]_\times = \begin{bmatrix} 0 & -\hat{n}_z & \hat{n}_y \\ \hat{n}_z & 0 & -\hat{n}_z \\ -\hat{n}_y & \hat{n}_z & 0 \end{bmatrix} \tag{10}$$

where θ is an angle of rotation with respect to a 3-vector unit axis, $\hat{n}$. $R(\theta, \hat{n})$ is pre-multiplied by the initial rotation matrix to compute the current rotation inside the global optimization. $R(\theta, \hat{n})$ is nearly linear in ω for small incremental rotations.

Through bundle adjustment, non-linear iterative optimization approach, such as the Gauss–Newton iterative method, was adopted to minimize $R(\theta, \hat{n})$. The initial parameters are obtained by Newton's method after iteration. Because Newton's method is suitable for linear problems and may converge to a saddle point rather than to the global minimum, this study used the Levenberg–Marquardt method to find the correct cost function. Let $\theta \to \theta + \delta\theta$ and $f(\theta)$ be smaller. According to the Taylor series:

$$f(\theta+\delta\theta) \approx f(\theta) + g^T\delta\theta + \frac{1}{2}\delta\theta^T H\delta\theta \tag{11}$$

where g is the gradient, and $\boldsymbol{H}$ is the Hessian matrix. Let $\frac{df}{d\theta}(\theta+\delta\theta) \approx g + H\delta\theta \approx 0$

$$\delta\theta = -H^{-1}g \tag{12}$$

Eq. (12) can be rewritten in terms of Jacobian J as Gauss–Newton or normal equations:

$$\left(J^T W J\right)\delta\theta = -J^T W\Delta z. \tag{13}$$

Eq. (7) can be solved using Eq. (13), which can evaluate the minimum for well-parameterized bundle problems under an outlier-free least squares [55]. Finally, the features derived by SFM were loaded in a custom coordinate system according to the relative point positions. In addition, this study recompiled Bundler [7] to structure unsorted image sets of SFM for more efficient implementation.

2.2. Multi-view stereo

Using camera parameters and sparse 3D feature coordinates from SFM as initial information, Clustering Views for Multi-View Stereo (CMVS) [9] provides multi-view stereo vision by grouping images with similar views to improve computation efficiency, as shown in Fig. 7. Based on the grouped images identified by CMVS, PMVS (Patch-based Multi-View Stereo), which is a patch-based algorithm [11], generates additional cloud points into the 3D model through three processes: matching, expansion, and filtering. The calculation of PMVS was established by Furukawa et al. [11] with the denotation as:

$c(p)$	the center of p
$n(p)$	the unit normal vector of p
$R(p)$	reference images of p
$V(p)$	set of estimated images of p
$V^*(p)$	set of visible images of p

where p is a patch that is essentially a local tangent plane approximation of a surface with $\mu\times\mu$ pixels (in this paper, μ is 5 for all experiments). The location of the patch depends on the center $c(p)$. $n(p)$ can be observed from the camera position.

2.2.1. Feature matching

The Harris and DoG (Difference-of-Gaussians) filters are able to capture the features from images, and then calculate the relationship between features for each image. A 2D Gaussian filter with a standard deviation σ can be expressed by $G\sigma$ [51]. To detect features, the DoG filter and Harris filter can be expressed as:

$$\text{DoG filter}: D = \left|\left(G\sigma - G_{\sqrt{2\sigma}}\right) * I\right| \tag{14}$$

where * denotes a 2D convolution operator, and

$$\text{Harris filter } H = \det(M) - \lambda trace^2(M), M = G\sigma^*\left(\nabla I I^T\right), \nabla I = \left[\frac{\partial 1}{\partial x}\frac{\partial 1}{\partial y}\right]^T \tag{15}$$

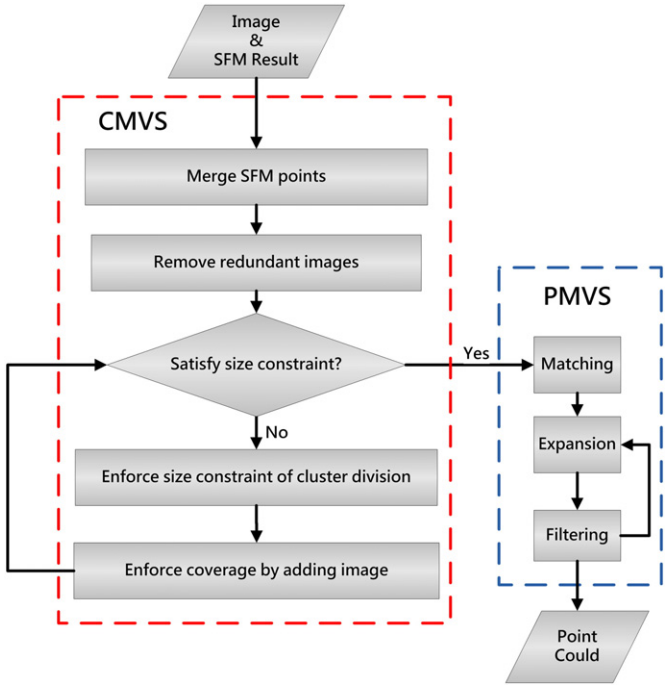

Fig. 7. Flowchart of CMVS and PMVS.

where ∇I is computed by convolving the image I with the partial derivatives of the Gaussian $G\sigma$. M is computed by convolving $G\sigma$ with respect to each element in ∇I [51].

In image I_i using a camera center $0(I_i)$, a feature f has a corresponding feature f' in image J_i. Based on the correspondence of f and f' using the triangulation of 3D points generated by SFM, an epipolar line, which the corresponding feature locates on, can be calculated in image J_i. The feature matching results in the center of the patch, unit normal vector of patch, and reference images of the patch all serve as an initialization:

$$c(p)\leftarrow\left\{\text{Triangulation from } f \text{ and } f'\right\},\quad n(p)\leftarrow\overline{c(p)0(I\iota)}/\left|\overline{c(p)O(I\iota)}\right|, R(p)\leftarrow I_i. \tag{16}$$

In addition, the following constraint must be met:

$$V(p)\leftarrow\left\{I\left|n(p)\cdot\overline{c(p)0(I)}/\overline{c(p)0(I)}\right|>\cos(\iota)\right\} \tag{17}$$

where $\iota=\pi/3$ in the experiments.

2.2.2. Expansion

At this stage, we iteratively added new neighbors to existing patches until they covered the surfaces visible in the scene. More intensive corresponding patches can be expanded by inputting the first partial initial scattered corresponding patch and correcting corresponding points nearby to obtain accurate corresponding points [11]. To confirm the expansion of the unit, this study initialized $C(p)$ through the collection of visual images of adjacent image elements:

$$C(p)=\left\{Ci\left(x',y'\right)\left|p\in Q_i(x,y),\left|x-x'\right|+\left|y-y'\right|=1\right.\right\} \tag{18}$$

$$\left|\left(c(p)-c\left(p'\right)\right)\cdot n(p)\right|+\left|\left(c(p)-c\left(p'\right)\right)\cdot n\left(p'\right)\right|<2\rho_1 \tag{19}$$

where $C_i(x,y)$ is a set of neighboring image patch of p, $C_i(x',y')\in C(p)$. If $C_i(x',y')\in C(p)$, expansion is not necessary, then $C_i(x',y')$ should be removed from $C(p)$. In other words, if the patch p of neighboring image element contains a patch p', p and p' are defined as neighbors.

2.2.3. Filtering

Filtering was applied to the reconstructed patches to further enforce visibility consistency and remove erroneous matches. Relying on visibility consistency, p is filtered as an outlier if the following inequality holds [11]:

$$\left|V^*(p)\right|(1-g^*(p))<\sum_{p_i\in U(p)}1-g^*(p_i). \tag{20}$$

Using updated $V^*(p)$, new photometric discrepancy function $g(p)$ can be in the term of $g^*(p)$:

$$V^*(p)=\{I|I\in V(p),h(p,I,R(p))\le a\} \tag{21}$$

$$g^*(p)=\frac{1}{|R(p)/V^*(p)|}\sum_{I\in R(p)/V^*(p)}h(p,I,R(p)) \tag{22}$$

where $h(p,I,R(p))$ is a pairwise photometric discrepancy function between images I and $R(p)$ [11].

The reliability of the corresponding points can be enhanced by finding multi-view photometric consistency and corresponding points near an epipolar line using these dense corresponding points to obtain a stereoscopic view and object depth for 3D reconstruction through CMVS and PMVS. Deleting incorrect and obstacle corresponding patches according to visibility, PMVS searches dense 3D cloud points to generate a comprehensive mesh in the next step. Both CMVS and PMVS can automatically detect and delete obstacles and outliers, which can be accomplished without any object information. However, this method remains a broken map obstacle because of deficient corresponding points.

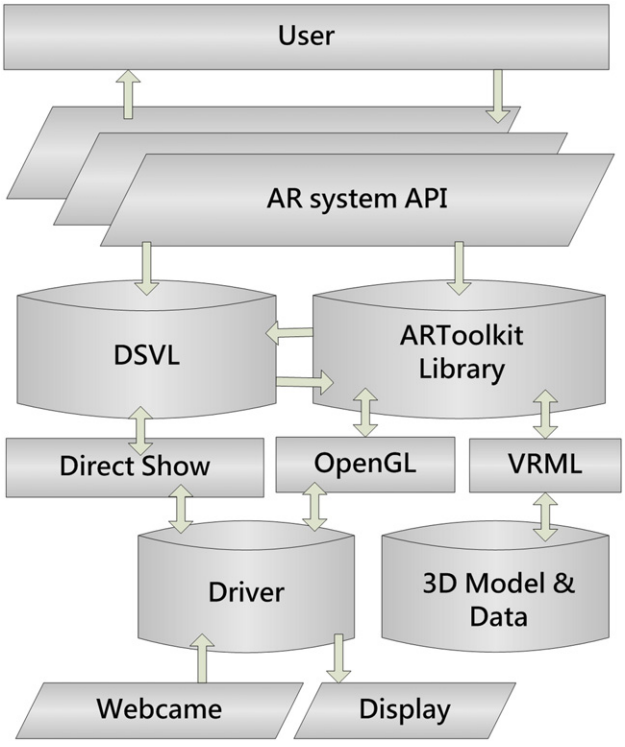

Fig. 8. AR system architecture.

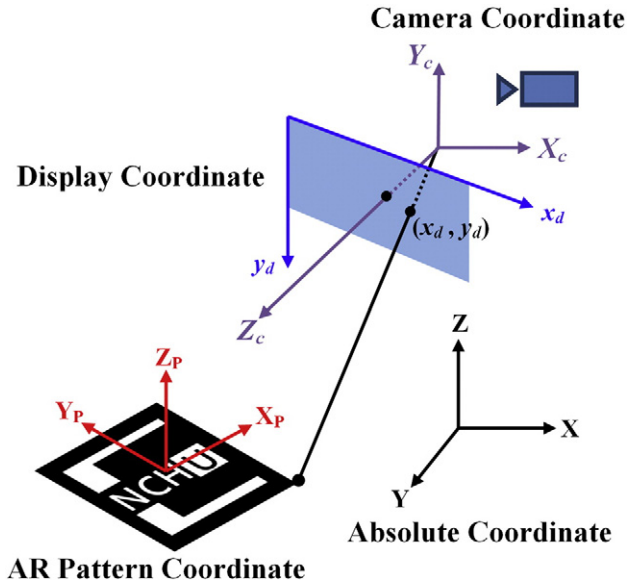

Fig. 9. Geometrical relationship between AR pattern coordinate vs. camera coordinate and real-world coordinate.

2.3. Mesh reconstruction

After the 3D points have been corrected and reconstructed, the Poisson Surface Reconstruction (PSR) algorithm can be adopted to connect 3D points as mesh and form the more complete 3D model [59]. The PSR algorithm can generate a 3D model with a smooth surface and efficiently reduce noise impact by flexibly defining a mesh size. This study used MeshLab, which is open source software using PSR for mesh reconstruction [60].

2.4. Augmented reality

Bimber and Raskar described AR as a technique that is constructed by complex information and image processing, such as tracking, registration, display technology, rendering, synthesizing devices and techniques, presentation and authoring [17]. The most critical component of AR is tracking and registration technique using a camera as the image acquisition device whose real-time interior and exterior parameters, such as focal length, location and orientation of camera, and object coordinates, can be calculated. Imagine that vision tracking is composed of four tasks: 1) loading marked object image from camera; 2) analyzing the angle and position of tracking objects and calculating the location and orientation of the camera; 3) loading 3D stereo virtual objects and conducting stack operation; and 4) presenting the result based on a 3D stereo analysis algorithm. Commonly, a tracking technique in computer vision consists of mark and non-mark types. The mark technique is superior because of the accuracy and efficiency and was thus used in this AR system.

2.4.1. AR system

An AR system consists of numerous imaging processing techniques, such as fundamental digital image processing, computer vision, and computer graphics. There are several developed platforms providing AR developers with tool libraries. The ARToolKit [31] library (version 2.72.1) is commonly used and represents the core of AR application in this paper. Data exchange and interaction in AR are essential components and are structured as an AR system in Fig. 8. All open source codes were transformed into application programming interface (API) coded in C/C++ as an interface to allow the components to communicate with each other, which makes easy to execute on smart handset devices. The generated reconstruction models were transferred into VRML models that can be explored in AR environment to facilitate displaying, interacting, and rendering.

At first, the AR system obtains the image streams and image resolution through DSVL (Direct Show Video Library). Direct Show developed in DSVL controls the camera through the driver. Image acquisition and tracking recognition are simultaneously processed in the AR system. The ARToolKit Library recognizes and tracks patterns defined a priori. The constructed 3D models created in VMRL (Virtual Reality Modeling Language) format are loaded into ARToolKit Library. Once the corresponding pattern is recognized, the 3D model can be called out by the ARToolKit Library. Finally, the 3D model is redrawn and projected on a monitor using OpenGL. The ARToolKit Library calculates the position and perspective of the 3D model based on the geometrical relationship of the camera and patterns. The 3D model fitting in real-time images, which are acquired through the camera controlled by Direct Show, is displayed in VRML format on the monitor using OpenGL. The AR system architecture consists of four steps:

(1) acquiring patterns in images from cameras;
(2) calculating the view point and location of the tracked pattern to estimate the position and perspective of the 3D model;
(3) loading the 3D model and fitting with real images; and
(4) displaying the 3D model within real-world images on the monitor.

In this paper, AR functions, such as recognition, video tracking, and positioning provided in the ARToolKit (version 2.72.1), were programmed in a modular format. Data exchange between modules depends on external variables. OpenVRML (version 0.14.3) is an open source library to process 3D models in VRML format that can be displayed using OpenGL, which is built in Windows 7.

2.4.2. ARToolKit

The ARToolKit is a cross-platform function library which was commonly used in AR computer vision tracking [30]. ARToolKit software is coded in C language to allow AR virtual image stacking in the real world. Through tracking the defined marks by camera, 3D virtual objects can be constructed in OpenGL (Open Graphic Library) and be accurately stacked upon the marks. ARToolKit solves mark tracking and virtual object interaction based on two key techniques, tracking and registration, which are described as follows: (a) capturing the real-world images through the camera, (b) searching AR patterns in the images, (c) calculating the position of the camera relative to the black square in the AR pattern, (d) recognizing the content of the AR pattern, (e) drawing a 3D model of the real-world images, and finally (f) interacting between users and 3D model in the AR system [31].

2.4.3. AR pattern

The AR pattern is one type of Trusted Reference (or Fiducially Marker) for tracking and recognition, which is adopted in ARToolKit using fiducial base tracking developed by Kato et al. [30]. The design of the AR pattern should follow two criteria, (1) a white square within a black square, and (2) asymmetric black-and-white or color signature within the white square [25]. The AR pattern plays two roles: as a feature of SFM in 3D construction and as a tracking and recognition reference in the AR system. By contrast, the black-and-white pattern is a perfect feature to be identified in a pair images for feature detection. The AR pattern is a rectangle with two pairs of parallel lines and four right corners that facilitate F-matrix computation and result in accurate 3D reconstruction. Based on the known dimensions of the AR pattern, the reconstructed model is generated at an absolute scale. By contrast, the position and perspective of the 3D model are called out by the identified pattern and are displayed on the monitor through the geometric transformations among the AR pattern coordinate, camera coordinate, and real world coordinate, as shown in Fig. 9.

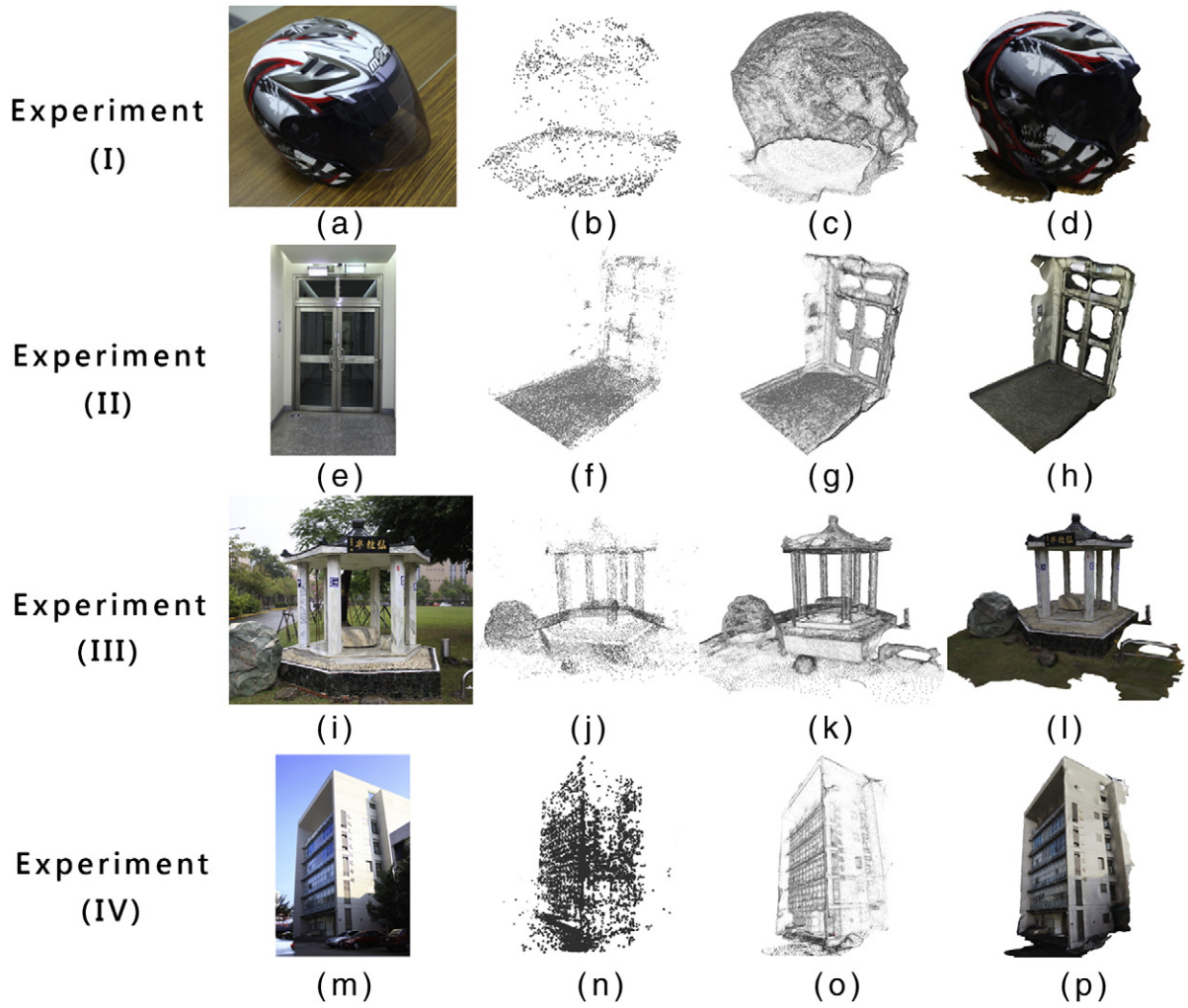

Fig. 10. Experimental objects in scene photos ((a), (e), (i), (m)), sparse point cloud ((b), (f), (j), (n)), dense point cloud ((c), (g), (k), (o)), and 3D reconstruction result ((d), (h), (l), (p)).

2.4.4. OpenGL

OpenGL (Open Graphics Library) developed by Silicon Graphics Inc. is a standard specification defining a cross-language, cross-platform API for applications producing 2D and 3D computer graphics. OpenGL is widely used in CAD, virtual reality, scientific visualization, information visualization, flight simulation, and video games. OpenGL interface consists of over 250 different function calls that can be used to draw complex 3D scenes using simple primitives. In this AR system, OpenGL is employed to redraw and present the 3D model on a monitor according to the simulated scenario from a human's viewpoint. The detailed description refers to the OpenGL Programming Guide [61] and Computer Graphics Using OpenGL [62].

2.4.5. VMRL (Virtual Reality Modeling Language)

Virtual Reality Modeling Language (VRML) is a text file format expressing vertices and edges for a 3D polygon specified along with the surface color, shininess, and transparency. Virtual reality modeling language (VRML) is suitable for numerous data processes such as graphical components on a webpage and in 3D reconstructed models. Animations, sounds, lighting, and other aspects of the virtual world can interact with the user or may be triggered by external signals such as the AR pattern in this study. Currently, many 3D modeling programs save the reconstructed objects and scenes in VRML format, so does this study.

3. Experiments and result analysis

3.1. Study sets

To verify the ability of the established reconstruction system, indoor and outdoor objects at various scales, such as a helmet (a small object), a corridor (an indoor medium object), an arbor (an outdoor medium object), and a building (an outdoor large object) as shown in Fig. 10, were tested as Experiments I, II, III, and IV, respectively. The corridor (Experiment II) is located on the third floor of a 6-story RC building (Experiment IV). The RC building and the arbor (Experiment III) both are on the campus at National Chung Hsing University (NCHU), Taiwan.

The equipment used in the experiments included a PC computer, software, and a camera. The computer was equipped with Intel i7

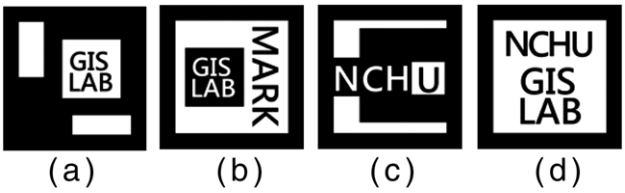

Fig. 11. AR patterns in this paper using Illustrator software.

870@2.93 GHz CPU, 16 GB memory RAM, NVIDIA GeForce GT 420 video card, Windows 7 OS, and VC++ 2010. The software includes Bundler open source established by Noah Snavely and served as the SFM program, CMVS (containing PMVS2) presented by Furukawa and was recompiled as the CMVS program, and Poisson Surface Reconstruction in open source software Meshlab to edit point cloud and construct surfaces. The imaging device was a camera, a Canon EOS 5D Mark II equipped with a 35 mm f2 fixed focus lens. Finally, this study employed ARToolKit library to program AR system. Also, AR patterns shown in Fig. 11(a), (b), (c) and (d) for Experiments I, II, III, and IV, respectively, were introduced into the reconstruction process and can significantly overcome a serious problem of broken map occurred in the 3D reconstructed models.

3.2. Experimental results

Most parts of the helmet were effectively reconstructed, except for the transparent mask at the front of the helmet (Fig. 10(b)) because of a sparse point cloud generated by SFM, which was caused by incomplete light reflectance and transparency. Through CMVS and PMVS, more cloud points (Fig. 10(c)) were generated but did not completely recover the mask shape. In Experiment II, few corresponding features were found on the white wall to cause a serious broken map obstacle in the reconstruction. Even though the AR pattern increased many corresponding features and cloud points, they were limited only around the AR pattern and itself. In Experiment III, an outdoor medium object (the arbor) was reconstructed using surrounding images that provides better multi-view geometry. The outdoor environment easily provides sufficient corresponding features and cloud points to result in a complete reconstruction. Furthermore, the AR patterns pasted on the white columns helped overcome a broken map obstacle for the column reconstruction (see Fig. 12). For a large outdoor object (such as the RC building), reconstruction can be effectively completed with many corresponding features because of the outdoor complex light environment. However, all images of the RC building were taken on the ground. A low imaging position relative to the building height results in a substantial limit to view the object from various perspectives, thus causing distortion at the top of the constructed RC building. If possible, surrounding images from all aspects should be taken for a comprehensive reconstruction. For a large-scale object, UAV (Unmanned Aerial Vehicle) images can be an alternative to overcome the physical limit.

Fig. 13 displays the reconstructed 3D models from various viewpoints. A reconstructed 3D model is more comprehensive if the images could be taken from all view angles toward the object. A short baseline relative to the object size causes deformation and broken spots in the reconstructed 3D model, such as the corridor because of a space limit and the RC building because of a physical limit. In summary, the approach proposed in this paper is proven to effectively reconstruct a single, small object and complex large scale of outdoor objects.

3.3. Estimation of focal length

The 3D reconstruction results were significantly influenced by the focal lengths which were calculated by SFM and are shown in Table 1 for the 3D reconstruct experiments. The relationship between cloud points and faces is positively correlated, and the reconstructed 3D model is obviously more complete and realistic with the increase of the cloud points and faces. The number of cloud points is determined by image resolution and correspondence. In addition, the faces of the 3D model are established based on the cloud points as vertices. A great density of cloud points is helpful to establish dense triangular faces to obtain accurate parameters and reconstruct a more meticulous 3D model with an accuracy of less than 1.5 pixels.

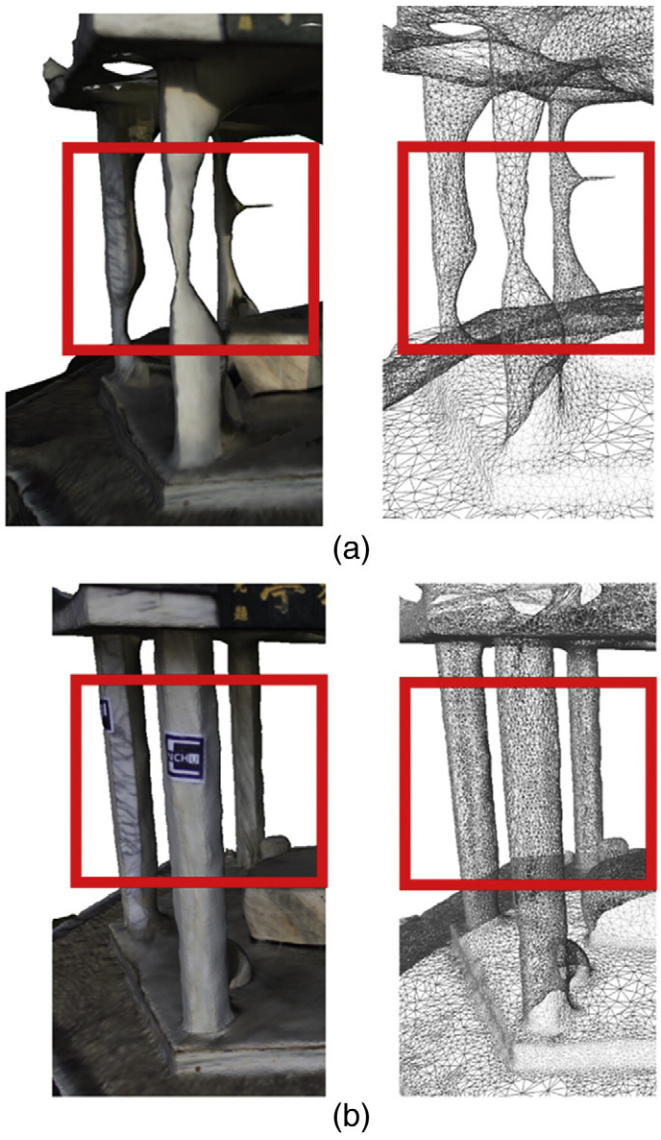

Fig. 12. Point cloud for 3D reconstruction (a) without and (b) with AR pattern.

The estimated focal length can be calculated by multiplying the mean of the focal length ratio derived in the K matrix of SFM by the CCD size. The camera used in this study is a full-frame camera, a 5D Mark II with a CCD size of 35 mm. According to Table 1, the mean of the focal length ratio ranges from 0.999 to 0.986 with a CCD size of 35 mm to obtain the estimated focal lengths ranging from 34.97 to 34.51, which are close to the nominal focal length of 35 mm. With an additional AR pattern, the RMSE (Root-Mean-Square Error) of the focal length ratio in the corridor and arbor experiments being less than the others means that the focal lengths were estimated with superior consistency in the experimental images, thus improving the reconstructed 3D models.

3.4. AR presentation

Fig. 14 shows the reconstructed 3D models viewed at various scales in the AR system. It is demonstrated that in both indoor and outdoor experiments, the accuracy of adopting AR pattern in the image-based point cloud modeling was improved by approximately 15% in all three dimensions. With the AR pattern, the reconstructed 3D models can be registered in an absolute coordinate system so that surveying-level measurement becomes possible. In addition, in

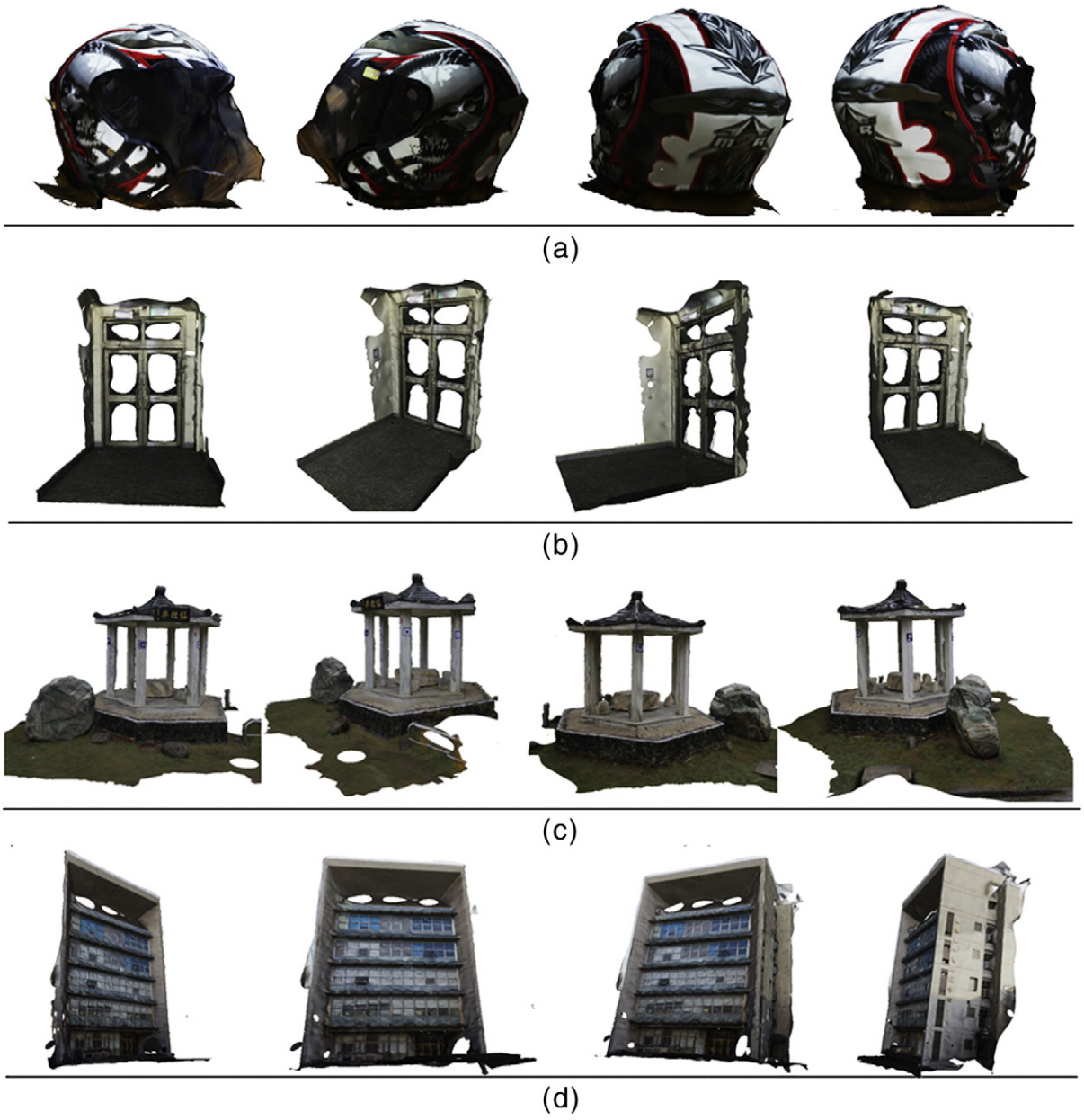

Fig. 13. Reconstructed 3D models from various viewpoints.

the AR environment the 3D models can be displayed with real-world objects at various scales to show the visualizing relationship and interaction between the reconstructed 3D and the real-world objects. This is helpful for construction engineers to display their designs in real environments and explore the interactions.

4. Conclusion

This paper presents an interactive system prototype by integrating multiple image-based 3D reconstruction approaches and AR, which was applied to four sets of 3D spatial models, including two indoor objects and two outdoor construction objects. Image-based point cloud models may not be as accurate and dense as laser scanners' point cloud models, but it has advantages of being economic, easy to use, and having low expertise requirement. Based on the results of this study, several conclusions are drawn which are as follows:

1. Computer version techniques, such as SFM, CMVS, PMVS, and Poisson Surface Reconstruction, were smoothly integrated into a 3D reconstruction system with comparative efficiency in computation, as compared with the traditional 3D reconstruction methods or 3D scanners.

Table 1
Image information and the estimated focal length.

	Helmet	Corridor	Building	Arbor
Number of images	16	73	74	53
Image resolution (pixel)	21 M	21 M	21 M	21 M
Number of cloud points	71,563	126,415	385,463	139,335
Number of faces	134,706	250,809	767,592	276,938
Mean of focal length ratio	0.999	0.991	0.991	0.986
RMSE of focal length ratio	0.0116	0.0088	0.0137	0.0031
Estimated focal length (mm)	34.97	34.69	34.69	34.51

Fig. 14. Reconstructed 3D models viewed in AR system at scales of (a) cup size and (b) human size.

2. Unlike other image-based or conventional photogrammetric techniques acquiring a rigid photo format, the image-based reconstruction model allows to input any kind of photos collected by image-capturing devices, such as amateur cameras, mobile phones, and tablet PCs, without need of a priori internal and external camera parameters.
3. This paper adopts open sources of image-based modeling with much less effort and no cost compared to the commercial 3D reconstruction software. Moreover, all open source codes were transformed into API coded in C/C++, which makes it easy to execute on smart handset devices.
4. The system can significantly reduce intensive labor, professional expertise, and expensive equipment that can be fully automatic and broadly applied to field surveying with visualization.
5. With additional AR patterns, a 3D model can be built with additional details by increasing the detected and matching features and the 3D model accuracy of 15% in all three dimensions. The geo-registered 3D models can be displayed in an absolute coordinate system, instead of a relative coordinate system, that makes measurements available in a surveying level.
6. The 3D reconstruction results show that surrounding images with various perspective views enhance the density of cloud points and result in a more comprehensive 3D model. To overcome the physical limit for outdoor large objects such as buildings, UAV images can be an alternative to capture proper images.
7. The generated reconstruction models were transferred into VRML models that can be explored in an AR environment to facilitate displaying, interacting, and rendering that provides engineers to

perceive the interaction between as-built scene and as-planned construction in construction design and management. In the AR environment, 3D CAD planed models are allowed to overlay to the 3D scene reconstruction and to be visualized from different viewpoints and scales.

In the future, the proposed system integrating image-based reconstruction approaches and AR can be optimized and built in mobile devices, which is convenient for on-field engineers to remotely assess the daily progress of an ongoing construction project using handy equipment and facilitates decision making on the necessary remedial actions.

Acknowledgments

The authors like to sincerely thank four anonymous reviewers whose comments greatly improved the quality of this paper.

References

[1] P.R. Wolf, B.A. Dewitt, Elements of Photogrammetry: with Application in GIS, McGraw-Hill, 2000.
[2] F. Ackermann, Airborne laser scanning — present status and future expectations, ISPRS Journal of Photogrammetry and Remote Sensing 54 (1999) 64–67.
[3] A. Davison, I. Reid, N. Molton, O. Stasse, MonoSLAM: real-time single camera SLAM, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 1052–1067.
[4] Autodesk Labs, http://labs.autodesk.com/2010.
[5] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, A comparison and evaluation of multi-view stereo reconstruction algorithms, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, 2006.
[6] N. Snavely, S.M. Seitz, R. Szeliski, Modeling the world from internet photo collections, in: IJCV08, 2008, pp. 189–210.
[7] N. Snavely, Bundler: structure from motion (SFM) for unordered image collections, http://phototour.cs.washington.edu/bundler/2010.
[8] Y. Furukawa, CMVS, Department of Computer Science and Engineering, University of Washington, Seattle, Washington, 2010. http://grail.cs.washington.edu/software/cmvs/ (7/17/2010).
[9] Y. Furukawa, B. Curless, S.M. Seitz, R. Szeliski, Towards Internet-scale Multi-view Stereo, CVPR, 2010.
[10] Y. Furukawa, PMVS2, Department of Computer Science and Engineering, University of Washington, Seattle, Washington, 2010. http://grail.cs.washington.edu/software/pmvs/ (7/13/2010).
[11] Y. Furukawa, J. Ponce, Accurate, dense, and robust multi-view stereopsis, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (8) (2010) 1362–1376.
[12] M. Vergauwen, L. Van Gool, Web-based 3D reconstruction service, Machine Vision and Applications 17 (6) (2006) 411–426.
[13] N. Govender, Evaluation of feature detection algorithms for structure from motion, in: 3rd Robotics and Mechatronics Symposium, Pretoria, 2009.
[14] T. Cooke, R. Whatmough, Evaluation of corner detectors for structure from motion problems, in: Digital Image Computing: Techniques and Applications, Cairns, 2005, p. 77.
[15] G.J. Brostow, J. Shotton, J. Fauqueur, R. Cipolla, Segmentation and recognition using structure from motion point clouds, European Conference on Computer Vision 1 (2008) 44–57.
[16] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Peña-Mora, Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, Automation in Construction 20 (8) (2011) 1143–1155.
[17] O. Bimber, R. Raskar, Spatial Augmented Reality: Merging Real and Virtual Worlds, in: A.K. Peters, Ltd., Wellesley, MA, USA, 2005, pp. 1–7, (Chap.1).
[18] I.E. Sutherland, The ultimate display, in: IFIP Congress, Munich, 1965, pp. 506–508.
[19] I.E. Sutherland, A head-mounted three dimensional display, in: Fall Joint Computer Conference, San Francisco, 1968, pp. 757–764.
[20] T.P. Caudell, D.W. Mizell, Augmented reality: an application of heads-up display technology to manual manufacturing, in: Twenty-fifth Hawaii International Conference on System Sciences, Kauai, vol. 2, 1992, pp. 2659–2669.
[21] P. Milgram, H. Takemura, A. Utsumi, F. Kishino, Augmented reality: a class of displays on the reality–virtuality continuum, telemanipulator and telepresence technologies, SPIE 2351 (1994) 282–292.
[22] J. Loomis, R. Golledge, R. Klatzky, Personal guidance system for the visually impaired using GPS, GIS, and VR technologies, in: First Annual International Conference on Virtual Reality and Persons with Disabilities, Millbrae, 1993.
[23] G.W. Fitzmaurice, Situated information spaces and spatially aware palmtop computers, Communications of the ACM, special issue on computer augmented environments: back to the real world 36 (7) (1993) 39–49.
[24] J. Rekimoto, K. Nagao, The world through the computer: computer augmented interaction with real world environments, in: Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology, Pittsburgh, 1995, pp. 29–36.
[25] J. Rekimoto, Augmented reality using the 2D matrix code, Workshop on Interactive Systems and Software, 1996.
[26] R. Azuma, A survey of augmented reality, Presence: Teleoperators and Virtual Environments 6 (4) (1997) 355–385.
[27] S. Feiner, B. MacIntyre, T. Höllerer, A. Webster, A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment, IEEE International Symposium on Wearable Computers, Cambridge, Massachusetts, 1997.
[28] T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, R.W. Picard, A. Pentland, Augmented reality through wearable computing, Special Issue on Augmented Reality, 1997.
[29] B. Thomas, V. Demczuk, W. Piekarski, D. Hepworth, B. Gunther, A wearable computer system with augmented reality to support terrestrial navigation, in: IEEE International Symposium on Wearable Computers, Pittsburgh, Pennsylvania, 1998, pp. 168–171.
[30] H. Kato, M. Billinghurst, Marker tracking and HMD calibration for a video-based augmented reality conferencing system, IEEE and ACM International Workshop on Augmented Reality (1999) 85–94.
[31] ARToolKit, http://www.hitl.washington.edu/artoolkit/2007.
[32] T. Höllerer, S. Feiner, J. Pavlik, Situated documentaries: embedding multimedia presentations in the real world, in: IEEE International Symposium on Wearable Computers, San Francisco, California, 1999, pp. 79–86.
[33] C. Portalés, J.L. Lerma, S. Navarro, Augmented reality and photogrammetry: a synergy to visualize physical and virtual city environments, ISPRS Journal of Photogrammetry and Remote Sensing 65 (1) (2010) 134–142.
[34] B. Thomas, B. Close, J. Donoghue, J. Squires, P. DeBondi, M. Morris, W. Piekarski, AR quake: an outdoor/indoor augmented reality first person application, in: IEEE International Symposium on Wearable Computers, Atlanta, Georgia, 2000, pp. 139–146.
[35] H. Regenbrecht, R. Specht, A mobile passive augmented reality device, in: International Symposium on Augmented Reality, Munich, Germany, 2000, pp. 81–84.
[36] S. Julier, Y. Baillot, M. Lanzagorta, D. Brown, L. Rosenblum, BARS: battlefield augmented reality system, in: NATO Information Systems Technology Panel Symposium on New Information Processing Techniques for Military Systems, 2000.
[37] J. Newman, D. Ingram, A. Hopper, Augmented reality in a wide area sentient environment, in: IEEE and ACM International Symposium on Augmented Reality, New York, New York, 2001, pp. 77–86.
[38] K. Satoh, K. Hara, M. Anabuki, H. Yamamoto, H. Tamura, TOWNWEAR: an outdoor wearable MR system with high-precision registration, in: IEEE Virtual Reality, Yokohama, 2001.
[39] G. Reitmayr, D. Schmalstieg, Mobile collaborative augmented reality, in: IEEE and ACM International Symposium on Augmented Reality, New York, New York, 2001, pp. 114–123.
[40] A.D. Cheok, S.W. Fong, K.H. Goh, X. Yang, W. Liu, F. Farzbiz, Human Pacman: a sensing-based mobile entertainment system with ubiquitous computing and tangible interaction, in: Proceedings of the 2nd Workshop on Network and System Support for Games, Redwood City, California, 2003, pp. 71–81.
[41] S. Guven, S. Feiner, Authoring 3D hypermedia for wearable augmented and virtual reality, in: IEEE International Symposium on Wearable Computers, White Plains, New York, 2003, pp. 118–126.
[42] A. Henrysson, M. Billinghurst, M. Ollila, Face to face collaborative AR on mobile phones, in: IEEE and ACM International Symposium on Mixed and Augmented Reality, Vienna, Austria, 2005, pp. 80–89.
[43] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, S. Lieberknecht, An augmented reality museum guide, in: IEEE and ACM International Symposium on Mixed and Augmented Reality, Cambridge, 2008, pp. 103–106.
[44] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, D. Schmalstieg, Pose tracking from natural features on mobile phones, in: IEEE and ACM International Symposium on Mixed and Augmented Reality, Cambridge, 2008, pp. 125–134.
[45] A. Morrison, A. Oulasvirta, P. Peltonen, S. Lemmelä, G. Jacucci, G. Reitmayr, J. Näsänen, A. Juustila, Like bees around the hive: a comparative study of a mobile augmented reality map, in: International Conference on Human Factors in Computing Systems, 2009, pp. 1889–1898.
[46] N. Yabuki, K. Miyashita, T. Fukuda, An invisible height evaluation system for building height regulation to preserve good landscapes using augmented reality, Automation in Construction 20 (3) (2011) 228–235.
[47] M. Kazhdan, M. Bolitho, H. Hoppe, Poisson Surface Reconstruction, ACM Solid and Physical Modeling Symposium, Stony Brook, New York, 2008.
[48] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.
[49] M. Lourakis, A. Argyros, The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg–Marquardt algorithm, in: Technical Report 340, Institute of Computer Science-FORTH, Heraklion, Greece, 2004.
[50] K. Zhou, Structure & Motion, Structure in Pattern Recognition, Vienna University of Technology, Faculty of Informatics, Institute of Computer Graphics and Algorithms, Pattern Recognition and Image Processing Group, 2010.
[51] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[52] C. Wu, M. Pollefeys, SiftGPU Library, Technical Report, University of North Carolina, Chapel Hill, 2005.
[53] N. Snavely, Scene reconstruction and visualization from internet photo collections, Doctoral thesis, University of Washington, Seattle, Washington, 2008.
[54] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, A.Y. Wu, An optimal algorithm for approximate nearest neighbor searching fixed dimensions, Journal of the ACM 45 (6) (1998) 891–923.
[55] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision Second Edition, Cambridge University Press, Cambridge, 2004.

[56] R. Hartley, In defense of the eight-point algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (6) (1997) 580–593.
[57] D. Nister, An efficient solution to the five-point relative pose problem, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, 2003.
[58] M. Lourakis, A. Argyros, SBA: a software package for generic sparse bundle adjustment, ACM Transactions on Mathematical Software 36 (1) (2006) 2:2–2:15.
[59] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3D shape descriptors, in: Symposium on Geometry Processing, Aachen, 2003, pp. 167–175.
[60] MeshLab, http://meshlab.sourceforge.net/2011.
[61] D. Shreiner, The Khronos OpenGL ARB Working Group, OpenGL Programming Guide, Addison-Wesley Professional, USA, 2009.
[62] S.H. Francis, M.K. Stephen, Computer Graphics Using OpenGL, in: 3rd edition, Prentice Hall, USA, 2006, (Chap. 7).