Title:
Automatic and Improved Camera Information Processing for a 3D Photo-Modelling System

Authors:
Tsung-Chien Wu, rabbit94577@gmail.com, National Central University, Taoyuan, Taiwan
Jiing-Yih Lai, jylai@ncu.edu.tw, National Central University, Taoyuan, Taiwan
Douglas W. Wang, dwmwang@gmail.com, Ortery Technologies, Inc., Taiwan
Chao-Yaug Liao, cyliao@ncu.edu.tw, National Central University, Taoyuan, Taiwan
Ju-Yi Lee, juyilee@ncu.edu.tw, National Central University, Taoyuan, Taiwan

Keywords:
Camera Calibration, Camera Alignment, 3D Textured Model

DOI: 10.14733/cadconfP.2019.158-163

Introduction:
Two-dimensional (2D) photos are commonly used for product presentation in e-commerce because they can reveal an object's high-quality texture and are easy to process. However, this requires an efficient (automatically) and high-quality collection of the object photos. The photo-taking device and corresponding software play an important role in the e-commerce industry [7]. In addition, as 2D photos only represent discrete views of an object, it is often necessary to capture hundreds of 2D photos with the photo-taking system and orient a photo at a given viewing angle via a viewing interface on the website [7]. However, storing and displaying so many photos while maintaining high photo quality on a website requires significant memory and CPU time for data downloading. In addition, the actual three-dimensional (3D) shape and dimensions of an object cannot be obtained in this representation. Orienting the object photos on the website is not continuous as only a limited number of views are recorded. A technique known as 3D photo-modelling [4], [7-9] was developed to reconstruct the 3D model of an object using multiple 2D photos, while maintaining its texture on the model (called 3D textured model hereafter) [9]. This technology is an alternative option for product presentation because a 3D textured model requires less memory and can be freely oriented in 3D space.

High-quality object photos can be obtained using either a single-camera device that applies a digital single-lens reflex (DSLR) camera with a camera rail, or a multi-camera device that applies several DSLR cameras mounted on an arm to capture an object placed on a turntable from different latitudes. These devices are integrated with photo capturing software to form a photo capturing system. This system can position the camera precisely and orient the object on the turntable to different angles to capture the object photos in different views. It also provides a controlled environment, e.g., a single background color and adjustable lighting, for easily separating the object from the background. As this system is already used in the field of product presentation, we use it here as the photo capturing source of the 3D photo-modelling technology.

The following kinds of data need to be obtained for the 3D photo-modelling technology: object and camera calibration mat photos, camera parameters, a 3D mesh model, and texture data. The object photos are employed to generate the 3D textured model. The mat photos are employed in camera calibration to evaluate camera parameters. The camera parameters are mainly used to transform 2D contour points extracted from each object photo into a 3D space on which the 3D mesh model is generated. The 3D mesh model describes the object's surface geometry in triangular meshes. The texture data covered on the 3D mesh model denotes the 3D textured model. The photo capturing is performed by a capturing system [7] to obtain high-quality object and mat photos. The camera calibration can be performed by matching the feature points between mat photos and the actual mat dimensions to figure out camera parameters [3-5], [10-11].

Main Idea:

This study develops algorithms for processing camera information to yield satisfactory object photos and camera parameters for generating 3D textured models. The camera information is evaluated using a system's workflow; this workflow requires a photo capturing device and accompanying software, and 3D photo-modelling software to generate the 3D textured model. The issues addressed in this study are a part of the system's workflow, which primarily comprises the following techniques: capturing object photos from multiple views, pre-processing object photos, calibrating the camera, and aligning the camera coordinates in multiple views. To simplify manual operation, these techniques are integrated into two software workflows. Several realistic examples are presented to demonstrate the feasibility of the proposed 3D photo-modelling system.

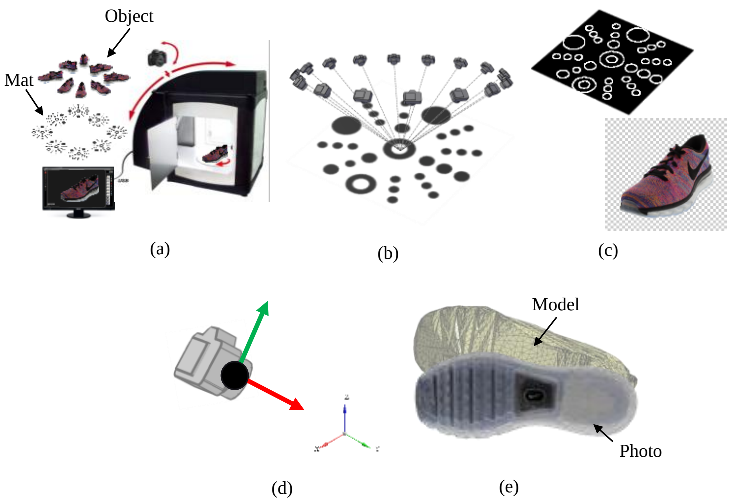

Fig. 1: Overall processes of camera information processing: (a) photo capturing, (b) camera calibration for each camera, (c) mat and object photo processing, (d) camera information, and (e) camera alignment.

The main contribution of this study is that we describe the technologies involved for camera information—photo capturing, photo pre-processing, camera calibration, and camera alignment—for generating and presenting 3D textured models. The entire procedure from photo capturing to 3D model generation is almost fully automatic, except for the camera alignment, which requires manual operation to change the pose of an object. The proposed system is more flexible and suitable for dealing with various kinds of realistic objects.

The proposed camera information processing is integrated with a photo capturing device and software for generating a 3D textured model of an object captured in multiple viewing angles. The main techniques involved are photo capturing, photo pre-processing, camera calibration, and camera alignment. Fig. 1 shows the overall flowchart of the proposed camera information processing. The photo capturing device can capture a series of photos for the calibration mat and the object under a controlled environment (Fig. 1(a)). The mat photos are employed in the camera calibration procedure to evaluate a set of camera parameters for each viewing angle (Fig. 1(b)). Pre-processing of object photos and calibration mats is required to remove the background color and detect the mat features, respectively (Fig. 1(c)). The first set of camera information is calculated by applying the mat information (Fig. 1(d)). An initial mesh model can be generated by using the first set of object photos and camera parameters with an SFS method [8]. The camera information of the flipped object is then calculated by applying the initial mesh model and the silhouette of the flipped object (Fig. 1(e)).

Photo Capturing:

The basic idea of the proposed conformal mesh parameterization is to unwrap 3D meshes onto the UV domain while preserving the angles of each triangle on the UV domain. A critical issue of the preservation is that all angles of 2D meshes on the UV domain cannot be exactly the same as those of 3D meshes. Therefore, an optimization problem is formulated to minimize the deviation of angles on 2D meshes and determine the optimized positions of all vertices [2-3], [6].

The proposed photo capturing system is composed of a capturing device, a control, and processing software [7]. The capturing device contains a set of lighting equipment and a turntable. The lighting equipment can build up an environment to make the object more appealing for the consumers. The turntable is equipped to rotate the object to capture the photos in 360°. However, an interface is needed to control the features of this integrated device, which is achieved by the integrated software. This software can also remove the background of the object to create a transparent or pure color background for different applications.

The components of the photo capturing device include the front, back, left, and right diffusion panels. These panels are the light source for the capturing device. They can adjust the lighting environment of the device, and they can be controlled separately to set different directions of lighting strength. The turntable driving mechanism can drive the turntable to rotate the desired angles. All of the components are controlled by two control boards, one for the turntable, and one for the lighting panels. The control signal is sent from the integrated software, which can control the device components using a separated module in a user-friendly interface.

This controlling software is also bundled with a background-removing feature to generate a transparent background photo. The purpose of this feature is to retain the perfect color of the object and remove the background color. The background-removing interface, including several parameters for setting, e.g., single or multiple loops, and removing strength. It can also preview the result after removing the background. Once the parameters are set, the object photos are captured, with the background color removed automatically. Further, the software also provides a tool to recover the background in case the result after removing the background is unsatisfactory.

Camera Calibration:

The basic idea of camera calibration is to find the corresponding feature points between each mat photo and the actual mat dimensions, and to solve the extrinsic and intrinsic matrices to obtain the camera parameters on each viewing angle [3]. As each mat photo and object photo are captured at the same viewing angle, the camera parameters at each viewing angle can be used to generate the position and orientation of the corresponding object photo in 3D space. All object photos in 3D space can thus be related, allowing the 3D points on the model to be calculated.

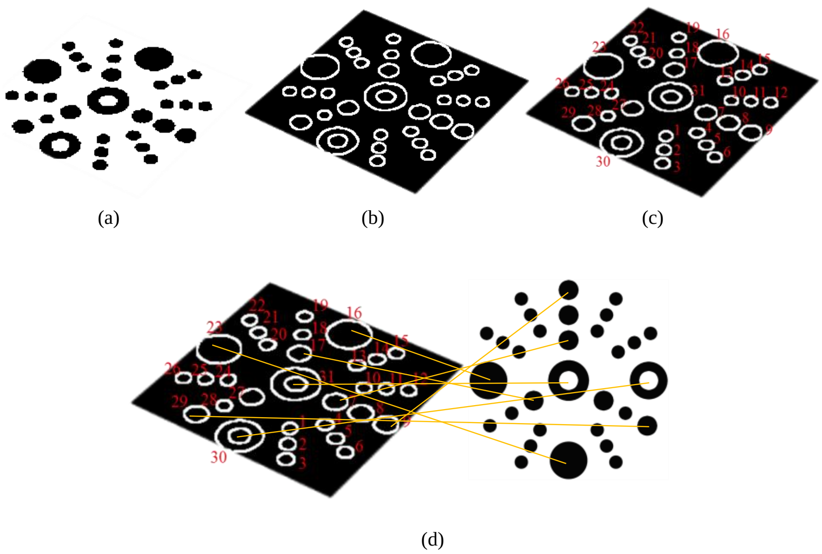

Fig. 2: Image processing for camera calibration: (a) image thresholding, (b) feature extraction, (c) feature encoding, and (d) feature matching.

The first step of camera calibration is thresholding the mat photos to distinguish the feature points from the background. As the purpose of this step is to filter out the background noise, the threshold is very important. For the automatic workflow, an automatic thresholding process has been developed. The original mat photos are first converted into grayscale; the automatic thresholding can determine the filtering value on each photo by detecting its highest grayscale value. The threshold of a mat photo is set to 80% of the highest grayscale value. Once the threshold value is determined, each mat photo can be converted individually into a binary photo, as shown in Fig. 2(a).

The next step is to recognize feature points; each feature point is recognized as an ellipse. The original method calculates the distance between the center of the bounding box and the corner vertex

[3]. Previously, if the distance is less than a threshold, the corresponding feature is detected as an ellipse. However, for the automatic procedure, the past criterion of this detection is not robust enough. In this study, there are two new criteria for detecting the ellipse. The first criterion compares the area difference between the bounding box and the feature. The second criterion compares the perimeter difference between the bounding box and the feature. If both values are less than two respective thresholds, the feature is detected as an ellipse, as shown in Fig. 2(b).

The third step involves decoding the features by calculating the distance of each center of the ellipse to the central ellipse on the 2D domain. Each feature point is assigned a number, which matches the corresponding feature point on the ideal mat. In the original decoding procedure [3], the number of total features recognized was not checked, thus resulting in an insufficient number of features for calibration. Therefore, in the proposed method, a checking process has been added to make sure the feature number is correct. Because the features on the mat have an arrangement, the system can continuously check the number of detected ellipses on adjacent groups to ensure the detected features are correct. Once all features are detected, each of the features is assigned a corresponding number, as shown in Fig. 2(c). The feature points are matched one by one according to their assigned number, as shown in Fig. 2(d).

Camera Alignment:

The purpose of the camera alignment process is to obtain the information about the bottom of the object. However, the object must be flipped to capture photos of its underneath. Once the object is moved, the entire coordinate system must be reset. The camera alignment process can merge the new coordinate system into the original coordinate system so that both sets of object photos use the same coordinate system. A procedure for camera alignment was previously developed [4], but it was not accurate enough. We have therefore developed an improved method, as described below.

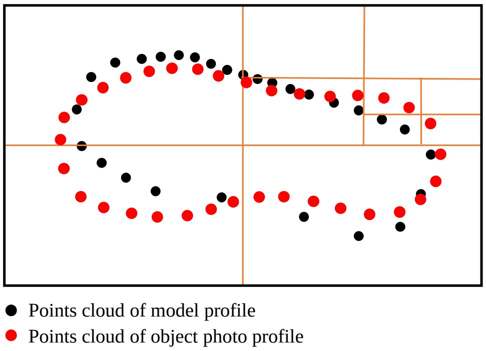

Fig. 3: Quad-tree structure to improve the search for the closest point.

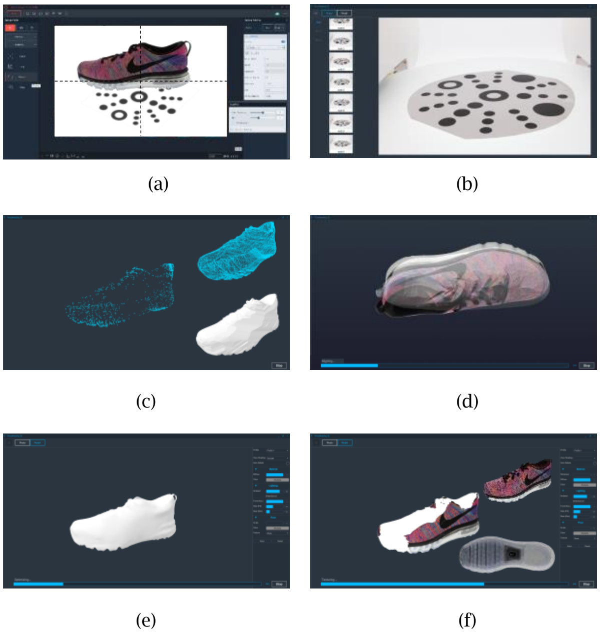

Fig. 4: Operation workflow for 3D photo-modeling system: (a) object photo capturing software, (b) input mat photos and camera calibration, (c) model reconstruction, (d) camera alignment, (e) model optimization, and (f) texture mapping.

The basic idea of camera alignment is to apply the original calibrated camera information to evaluate the flipped angle. However, the original camera position cannot be changed when taking the second layer of object photos. The input data is the original calibrated camera information and the initial mesh model generated using the first set of object photos. First, the model is projected onto the 2D domain. The purpose of this projection is to compare the object profile on the flipped photo with that of the projected model. In the original procedure [4], the entire mesh model is projected onto the 2D domain, which is time-consuming. In the modified method, only meshes near the boundary are projected onto the 2D domain, which can save significant computational time.

The next step is to adjust the model orientation to fit the photos. Before adjusting the orientation, artifacts of the initial mesh model are ignored to prevent possible errors. As artifacts usually occur on the bottom of the mesh model, the meshes with a y coordinate value less than 0 are ignored. Next, the position and orientation of the projected model are adjusted using the iterative closest points (ICP) method [2]. The boundary profile of the projected model and the silhouette of the object photo are respectively regarded as two groups of point cloud. The original centroid of the point cloud was calculated based on the shape. The ICP accuracy is affected by the error of the center position. Thus, the improved center determination is based on the centroid of the point cloud. When the density of the point cloud is high near a particular region, the centroid of the points cloud can approach that region.

Through the ICP process, the initial alignment can be performed. However, the alignment is still not accurate enough to merge the two sets of data. The next process is fine-tuning the new set of camera data. The fine tuning process is based on the Golden Section method [1]. To increase the searching accuracy and efficiency, we ignore errors that are larger than a threshold, and to search for the closest points, we consider the computational efficiency. We therefore apply a quad-tree structure to increase the efficiency of the searching process, as shown in Fig. 3.

Examples and Discussion:

The algorithms of the object photo capturing, camera calibration, and camera alignment are integrated into a capturing software and a modelling software. The modelling software is also combined with (1) an SFS modelling procedure to create an initial mesh model [8], (2) a mesh optimization process to optimize the shape and surface smoothness of the initial mesh model [8], and (3) a texture mapping process to generate the texture map of the 3D textured model [9].

Fig. 4 depicts the entire workflow of this integrated system for a shoe example, which shows the screenshot of the process from the object photo capturing to the viewing of the final model. The capturing device can be controlled by this software (Fig. 4(a)). Fig. 4 (b) depicts the camera calibration process. Fig. 4 (d) depicts the camera aliment process. The model reconstruction, model optimization, and texture mapping processes are shown as Fig. 4 (c), (e), and (f), respectively. Figs. 5(a) to (f) depict the collected results from the object to the 3D textured model. By applying the proposed improvement method, sufficient information about the model shape, texture, and bottom can be obtained.

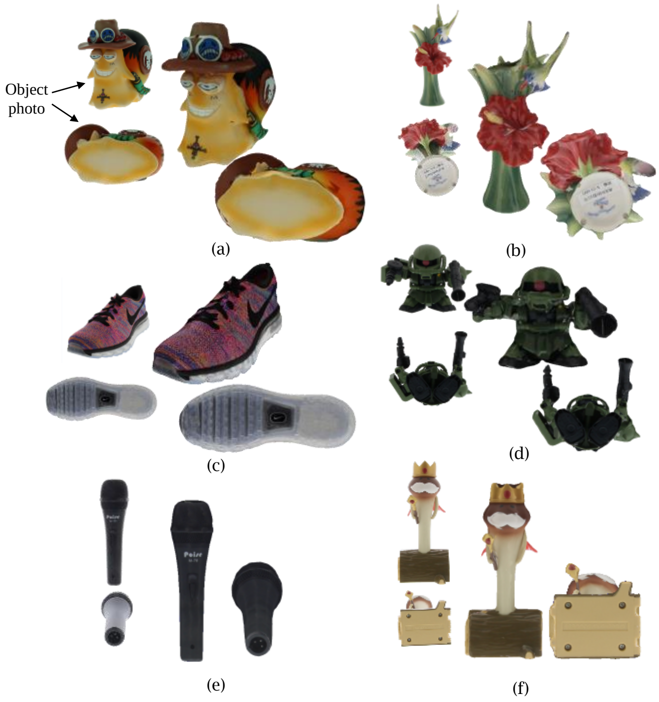

Fig. 5: Results of 3D textured model that includes bottom information: (a) snail doll, (b) vase, (c) shoe, (d) Gundam, (e) microphone, and (f) doll.

Conclusion:

In this study, we proposed a process of photo capturing, improved camera calibration, and improved camera alignment to generate 3D textured models for presentation on a website. The proposed photo capturing device can capture object and mat photos in a controlled environment and easily remove the background color on object photos. The camera calibration can automatically and accurately reconstruct the capturing coordinate system for further applications. The camera alignment can combine two different coordinate systems accurately and efficiently to merge the object photos from two different poses. These processes are combined with mesh model reconstruction, mesh model optimization, and texture mapping to generate the 3D texture model of the object. A high-quality 3D textured model can be output accurately and efficiently by applying this series of processes for presentation in e-commerce.

References:

[1] Kiefer, J.: Sequential minimax search for a maximum, American Mathematical Society, 4(3), 1953, 502-506. http://dx.doi.org/10.2307/2032161

[2] Low, K.-L.: Linear Least-Squares Optimization for Point-to-Plane ICP Surface Registration, Technical Report TR04-004, University of North Carolina at Chapel Hill, NC, 2004.

[3] Liao, C.-Y.; Xiong, Y.-S.; Wang, D. W.; Lai, J.-Y.; Lee, J.-Y.: A camera calibration process for 3D digital model reconstruction of huge objects, Machining, Materials and Mechanical Technologies (IC3MT 2016), Matsue, Japan, 2016.

[4] Liao, C.-Y.; Ren, Y.-L.; Wang, D. W.; Lai, J.-Y., Lee, J.-Y.: Model for 3D printing created from multiview 2D silhouette images, International Conference on Leading Edge Manufacturing in 21st Century: LEM21, 2017. http://dx.doi.org/10.1299/jsmelem.2017.9.130

[5] Marquardt, D. W.: An algorithm for least-squares estimation of nonlinear parameters, SIAM Journal on Applied Mathematics, 11, 1963, 431-441. http://dx.doi.org/10.1137/0111030

[6] Mulayim, A. Y.; Yilmaz, U.; Atalay, V.: Silhouette-based 3D model reconstruction from multiple images, IEEE Transactions on Systems Man and Cybernetics, 33(4), 2003, 582-591. http://dx.doi.org/10.1109/TSMCB.2003.814303

[7] Ortery. https://www.ortery.com/.

[8] Phothong, W.; Wu, T.-C.; Lai, J.-Y.; Yu, C.-Y.; Wang, D. W.; Liao, C.-Y.: Quality improvement of 3D models reconstructed from silhouettes of multiple images, Computer-Aided Design and Applications, 15(3), 2018, 288-299. http://dx.doi.org/10.1080/10.1080/16864360.2017.1397881

[9] Wu, T.-C.; Lai, J.-Y.; Phothong, W.; Wang, D. W.; Liao, C.-Y.; Lee, J.-Y.: Editable texture map generation and optimization technique for 3D visualization presentation, Computer-Aided Design and Applications, 15(3), 2018, 378-389. http://dx.doi.org/10.1080/16864360.2017.1397888

[10] Zhang, Z.: A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 2000, 1330-1334. http://dx.doi.org/10.1109/34.888718

[11] Zhang, Z.: Camera calibration with one-dimensional object, IEEE Transactions on Pattern Analysis and Machine Intellgence, 26(7), 2004, 892-899. http://dx.doi.org/10.1109/TPAMI.2004.21