Automation in Construction 48 (2014) 44–52

Contents lists available at ScienceDirect

Automation in Construction

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

Image-based modeling of built environment from an unmanned aerial system

Pablo Rodriguez-Gonzalvez *, Diego Gonzalez-Aguilera, Gemma Lopez-Jimenez, Inmaculada Picon-Cabrera

Department of Cartographic and Land Engineering, University of Salamanca, Polytechnic School of Avila, Hornos Caleros, 50, 05003 Avila, Spain

A R T I C L E   I N F O

Article history:
Received 30 December 2013
Received in revised form 11 July 2014
Accepted 22 August 2014
Available online 16 September 2014

Keywords:
Automation
As-built constructions
Complex sites
Unmanned aerial system (UAS)
Computer vision
Photogrammetry
Software development
Point clouds
Electrical substations

A B S T R A C T

In recent years, the modeling from images, called image-based modeling, is emerging as a powerful alternative for the as-built three-dimensional (3D) reconstruction of complex scenarios compared with the use of laser scanner technology or classical topographic methods. Image-based modeling is a non-invasive and low-cost technology that allows 3D reconstruction of objects and scenarios by using only images. However, although several tools currently exist for converting from two-dimensional (2D) to 3D, the specific requirements of quality and completeness for these types of scenarios are difficult to determine.

In this paper, a methodology for automatically reconstructing 3D complex scenarios, particularly electrical substations, using images acquired from an unmanned aerial system is analyzed and proposed. This methodology has been incorporated by the Iberdrola Company as an optimal technique for obtaining low-cost 3D as-built metric models of electrical substations. The results obtained show that this camera-based low-cost system is a competitive alternative to laser scanning systems for modeling as-built complex constructions and can become an essential tool for facility planning and management tasks.

1. Introduction

Three-dimensional (3D) models have become an essential tool for experts in various fields that provide quality representations of as-built sites, reducing discrepancies between the design and construction phases. These models are used in disciplines including urban and environmental planning [1–3], cultural heritage documentation [4–6], building and infrastructure inspection [7,8], industrial measurement and reverse engineering applications [9–11] and in the film industry, video game industry and virtual reality applications [12,13]. However, depending on the application, the specific requirements for a 3D model can be different. In particular, as-built models for engineering and architecture demand metric requirements in terms of accuracy, reliability and completeness, whereas 3D models for the entertainment industry require visual quality in terms of texture-mapping, file size, computational cost, level of detail and ease of use. Therefore, trying to enclose all of these requirements under the same 3D model continues to be difficult for the international scientific community.

To describe complex constructions, laser scanner technology [14] and binomial photogrammetry-computer vision [15,16] as active and passive techniques, respectively, provide exhaustive and non-invasive 3D documentation methods and can help generate 3D models.

The use of laser scanning technologies for 3D data capture in industrial sites has grown considerably over the last decade over traditional methods for acquiring as-built information, which consists of manual measurements by metric tape and topography [17–19], due to the improvement in their competitiveness. This competitiveness is due to the rapid increase in the speed and accuracy of the laser scanners in the last decade, while their costs and sizes have been shrinking [20]. However, laser scanner technology requires specialized personnel for the acquisition and processing phases as well as prolonged times for data acquisition and for the processing of the different point clouds.

Conversely, photogrammetric reconstruction, whose use has become popular in recent years due to its hybridization with computer vision [21], has strengthened so-called image-based modeling. The revolution of image-based modeling has been favored because of three main factors: (i) the development of wide-spread computer skills and hence an improvement in the calculation algorithms and possibilities for automation; (ii) flexibility in the camera specifications being allowed to take images, calibrated or not; and (iii) the quality in the results, releasing accurate and reliable as-built 3D models. For example, the computer vision approach for 3D visualization has been applied to document as-planned 3D models for construction sites and to monitor their progress during construction [22,23]. Although currently there are several image-based modeling tools, none of them guarantee these three results, especially when complex constructions such as electrical substations are considered. In particular, several web-based tools (i.e., Photosynth, Autodesk 123 Catch and Photofly) have been

* Corresponding author. Tel.: +34 920 353500; fax: +34 920 353501.
E-mail addresses: pablorgsf@usal.es (P. Rodriguez-Gonzalvez), daguilera@usal.es (D. Gonzalez-Aguilera), glj@usal.es (G. Lopez-Jimenez), ipicon@usal.es (I. Picon-Cabrera).

http://dx.doi.org/10.1016/j.autcon.2014.08.010
0926-5805/

developed for 3D modeling using only images; however, they do not guarantee enough quality in terms of accuracy and reliability [24]. Likewise, other standalone executable tools (i.e., Bundler-PMVS2, VisualSFM) have been developed under the well-known computer vision line, creating structure from motion (SfM), but cannot guarantee enough precision and result in unscaled models [24,25]. In the case of VisualSfM, the software does not include the global coordinates of the ground control points into the orientation process; this requires that the scale for the reconstructed object or its geo-reference to a global frame must be conducted using a Helmert 3D transformation. According to the close-range photogrammetric community, the most popular tool, PhotoModeler™, which guarantees accurate results, requires manual interaction in the orientation and restitution steps. More recently, some open-source (i.e., Apero-Micmac) [26] and commercial (i.e., PhotoScan™) tools have emerged that combine computer vision and photogrammetric capabilities; however, the former requires command-line usage and does not perform well with multiple images that have oblique geometry, and the latter requires a commercial license.

Considering the limitations detailed above, a hybrid methodology supported by photogrammetry and computer vision that guarantees automation (i.e., converts from 2D to 3D automatically), flexibility (i.e., enables the use of calibrated and non-calibrated cameras) and quality (i.e., provides 3D models with higher resolution than 3D laser scanners) for passing from 2D to 3D has been developed. Photogrammetry Workbench (PW) is a multiplatform software with a user-friendly interface that works with terrestrial or aerial images and considers vertical or oblique geometries [27]. The approach employed by PW software improves the current practice of image-based modeling because it integrates computer vision and photogrammetric algorithms into a smart approach that overcomes the issues of complex objects and scenarios. This is accomplished by combining the last generation descriptors in the extraction and matching steps, combining several lens calibration models under the self-calibration process and combining several stereovision and multiple stereovision algorithms.

This paper is presented in the following structure. After this introduction, Section 2 addresses a detailed description of the developed method for automatic aerial image-based modeling; Section 3 shows the experimental results applied to an electrical outdoor substation, where the shape and size of the elements have high complexity and human interaction with the elements to document may involve security risks, economic costs or even both. Finally, the most relevant concluding remarks are outlined in Section 4.

2. Camera-based low-cost system for modeling complex constructions

To automate the image-based modeling of an outdoor electrical substation from oblique aerial images acquired from an unmanned aerial system (UAS) is a challenge of great complexity in both computer vision and photogrammetry disciplines. The reasons for this are diverse: the instability of the UAS platform brings images which violate classical geometric restrictions (i.e., verticality, scale, overlap, etc.) of the aerial photogrammetry [28]; and the need for oblique convergent captures could show the vertical elements on the substation and present a great variability in radiometric (i.e., illumination, shadows) and geometric conditions. To address these problems, an efficient method used by PW software has been developed, allowing robust and geo-referenced automatic image-based modeling from images that do not need to be previously calibrated. In particular, the images' orientation process is addressed in a rigorous way. This process involves two steps: 1) an initial approximation provided by computer vision; and 2) a robust and precise refinement from the photogrammetry bundle block adjustment. Afterwards, the tool allows the final user to select the generation of a 3D dense model according to the previous image orientations; that is, the tool addresses the normal case of photogrammetry through the generation of a dense model from stereo vertical or horizontal images, which makes use of the Semi Global Matching (SGM) strategy [29]. Alternatively, a dense model could be generated from multiple oblique images, which surround the object or scene in a "ring" configuration and employ the Patch-based Multi-View Stereo (PMVS) strategy [30]. This global approach aims to generate high-quality 3D models supported by the rigorous photogrammetric images orientation.

The following figure (Fig. 1) illustrates the process for the automatic image-based modeling method as described.

Note that the georeferencing process could be conducted in two different phases: during the photogrammetric block adjustment; or by applying georeferencing by a Helmert 3D transformation if the dense model was generated in a relative system.

2.1. Extraction and matching

While digital image correlations began in the 1970s [31], local image detectors have not advanced significantly until the last decade by Feature Based Matching (FBM) techniques. One of the greatest advances in the field of singular point detection took place with the emergence of the SIFT (Scale Invariant Feature Transform) algorithm proposed by Lowe [32]. This algorithm exceeds other descriptors [33] because it is invariant to image scaling and rotation and partially invariant to illumination and viewpoint changes [34].

In the workflow developed, the fully affine invariant ASIFT algorithm was used. The Affine Scale Invariant Feature Transform algorithm [34], a variation of the SIFT algorithm, provides image matching with greater robustness to geometric and radiometric variations. ASIFT assigns a descriptor invariant to the scale, rotation, translation and accused related deformations between images (i.e., perspective angles coming from the inclination of the optical axis of the camera). More specifically, ASIFT identifies reliable features that have suffered great distortion due to perspective, allowing an efficient and reliable reconstruction from oblique images.

The most remarkable improvement in this method is the possibility of including two additional affinity parameters to control the perspective of the images, which correspond to the two perspective angles of the optical axis of the camera: the ϖ (tilt) angle and φ (axis) angle (Eq. (1)). The following expression accounts for the resulting descriptor:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = H_\lambda R_1(\kappa) T_1 R_2(\varpi) = \lambda \begin{bmatrix} \cos\kappa & -\sin\kappa \\ \sin\kappa & \cos\kappa \end{bmatrix} \cdot \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\varpi & -\sin\varpi \\ \sin\varpi & \cos\varpi \end{bmatrix} \quad (1)$$

where A is the affine transformation with the scale λ, κ is the rotation around the optical axis (swing) and the perspective parameters for the inclination of the optical axis of the camera: φ (tilt) $= \arccos(1/t)$ is the angle between the optical axis and the normal to the image plane; and ϖ (axis) is the azimuth angle between the optical axis and a fixed vertical plane.

The extraction of keypoints is performed by the same SIFT workflow [32] but adds a simulation of the perspective effect caused by a variation of the camera optical axis direction from a frontal position. The keypoints are computed through the transformation of each image by a perspective effect by a finite number of tilt-φ and ϖ-axis angles. For each simulated image, the keypoints are extracted through gradient magnitude and orientation and are based on histogram analysis for each pixel in a 4×4 neighborhood.

When the keypoints are computed, a matching process is carried out over the simulated images using SIFT descriptors. The matching involves a two-step process, where the descriptors are firstly matched according to the Euclidean distance [35] and secondly filtered by the Moisan–Stival ORSA algorithm (Optimized Random Sampling Algorithm) [36]. This filter is a variant of RANSAC (Random Sample