Advanced Engineering Informatics 29 (2015) 149–161

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges

Habib Fathi [a,1], Fei Dai [b,*], Manolis Lourakis [c,2]

[a] CTO, Pointivo Inc., Atlanta, USA
[b] Department of Civil and Environmental Engineering, West Virginia University, Morgantown, USA
[c] Institute of Computer Science, Foundation for Research and Technology, Greece

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 13 December 2014
Accepted 22 January 2015
Available online 24 February 2015

Keywords:
Structure from motion
Civil infrastructure
Image-based 3D reconstruction
Spatial data collection
Point cloud

ABSTRACT

Image-based 3D reconstruction of civil infrastructure is an emerging topic that is gaining significant interest both in the scientific and commercial sectors of the construction industry. Reliable computer vision-based algorithms have become available over the last decade and they can now be applied to solve real-life problems in uncontrolled environments. While a large number of such algorithms have been developed by the computer vision and photogrammetry communities, relatively little work has been done to study their performance in the context of infrastructure. This paper aims to analyze the state-of-the-art in image-based 3D reconstruction and categorize existing algorithms according to different metrics that are important for the given purpose. An ideal solution is portrayed to show what the ultimate goal is. This will be followed by identifying gaps in knowledge and highlighting future research topics that could contribute to the widespread adoption of this technology in the construction industry. Finally, a list of practical constraints that make the 3D reconstruction of infrastructure a challenging task is presented.

1. Introduction

Collecting as-built spatial data (i.e., shape, size, position, and orientation) of civil infrastructure is an almost everyday task in the Architecture, Engineering, Construction (AEC) and Facilities Management (FM) industry. Currently, the undertaking of this task in field is primarily with four technologies: tape measuring, total station surveying, laser scanning, and image-based 3D reconstruction. Manual measurement with a tape measure is the simplest method that needs minimal expertise. However, it is time consuming, expensive in terms of labor costs, and not always accurate. A total station measures angles and distances to target points (typically key points on the surface of an infrastructure element). The method suits projects whose profile is relatively simple. Otherwise, collecting thousands of points of a project using a total station would be tedious and time-consuming. Laser scanners operate based on the Time-of-Flight (ToF) or Phase-Shift principle and generate a point cloud of a target structure. Comparatively, Phase-Shift-based scanners are generally faster, lighter in weight, and cheaper. The time spent to capture 3D coordinates of a point is very small in this technology (compared to tape measuring and total station surveying) while the accuracy of the obtained points is high. Its primary downside is high equipment costs. The equipment itself can be bulky and it may not be convenient to move it from one location to another. Image-based 3D reconstruction relies on image processing and multiple-view geometry to determine the spatial location of a point covered by at least two photos. It converts a set of corresponding image points (in pixel coordinates) into a cloud of 3D points. It is inexpensive, easy to use, and time-efficient. However, the generated point cloud could be noisy depending on the visual characteristics of the scene.

In the AEC & FM industry, image-based 3D reconstruction has been used to document construction activities after the assembly of a construction project. For example, engineers use this technology to document a rebar installation for later inspection. It is also applied to document and measure existing buildings for conservation and preservation. In parallel to the industrial use, researchers have led an effort to expand this technology to broader civil infrastructure applications. Several research studies have evaluated the accuracy and efficiency of this technology [1–9]. There also exist another group of studies that customize its analytics for applications in civil

* Corresponding author. Tel.: +1 304 293 9940.
E-mail addresses: habib@pointivo.com (H. Fathi), fei.dai@mail.wvu.edu (F. Dai), lourakis@ics.forth.gr (M. Lourakis).
[1] Tel.: +1 404 713 3667.
[2] Tel: +30 2810 391716.

http://dx.doi.org/10.1016/j.aei.2015.01.012
1474-0346/

infrastructure [10–21]. In addition to these, image-based technology has been applied in construction engineering and management applications with respect to site visualization [22], progress monitoring and assessment [10,23], quantity measurement and takeoff [24,25], geometric quality control and verification [26], assessment of damage caused by catastrophes [27], mobile contextual awareness [28], and preservation of evidence for dispute resolution [29].

While promising, implementing image-based 3D reconstruction for civil infrastructure purposes remains a challenging task. The following highlights a number of these issues from a practical perspective. The construction site entails complicated geometries and it is ever changing. This sometimes leads to difficulty in precisely recovering a site. The level of accuracy required for certain applications such as structural deflection measurement is high. Meeting these accuracy requirements via the image-based technology is not always successful. Repeatability hinders the application of this technology in real construction projects and reduces its reliability. There is also a lack of proper guidelines for data collection in the field. Finally, field crew and project managers want site actionable information instead of raw 3D point clouds.

Technical challenges in image-based 3D reconstruction are best understood by studying the image formation process. When photographing a scene, the 3D geometry is projected into the 2D image sensor and hence the depth information is lost. Recovering the lost information is a challenging problem because the image formation process is not invertible [30]. Given a single image, this means that a 3D point in the scene can only be recovered up to a one-parameter ambiguity which is the distance from the camera. Corresponding image points in multiple views could be used to resolve the ambiguity and extract the 3D information. The geometrical theory behind this process (monocular, binocular, and/or multi-camera 3D reconstruction) is so called Structure from Motion (SfM). SfM has been successfully used in a wide range of applications from modeling the world from Internet photo collections [31] to photogrammetric surveying and topographic mapping using unmanned aerial vehicles [32]. Besides the previously highlighted practical issues, there are a number of technical problems that arise when SfM is used in the context of reality capture of civil infrastructure including: presence of poorly-textured surfaces that are covered with uniform material, necessity of large scale and far-range 3D reconstruction, unlikelihood to be able to cover all angles of the scene while collecting the visual data, dynamic and crowded nature of a site under construction, uncontrolled environment and lots of unwanted background, existence of repetitive patterns, and occlusion.

In comparison to the SfM process, the automation level of converting a 3D point cloud to information (e.g., meshed surfaces, metric measurements) that is amenable to civil infrastructure applications is still low. This might influence the widespread adoption of this technology. To consider the reasons, most of image-based reconstruction algorithms generate the output in the form of a 3D point cloud that maintains no information describing relationships between individual points; and it may suffer from imperfections such as noises, clutters, and holes. As a result, it is challenging if one wants to efficiently and effectively extract usable information from the point cloud.

This study aims to analyze the state-of-the-art. Qualitative performance metrics are identified to evaluate the aforementioned issues and highlight major research challenges ahead. The main contribution of this comprehensive analysis is identifying the gaps in knowledge and providing a research roadmap that could pave the road for a reliable and practical 3D reconstruction technique to collect civil infrastructure spatial data.

The rest of this paper is organized as follows. Section 2 comprehensively synthesizes the current state-of-the-art in image-based 3D reconstruction with a special focus on civil infrastructure applications. In Section 3, consideration for practical implementation is discussed based on engineering requirements in the construction industry. Section 4 explains the gaps in knowledge and the next section lists all practical constraints. Finally, Section 6 presents concluding remarks.

2. State-of-the-art

The existing computer vision-based technologies for capturing and modeling the reality using imagery can be broadly categorized into active and passive sensing. Active sensors, such as a RGB-D camera that produces depth maps by analyzing a speckle pattern of infrared laser light (i.e., structured light projection), have attracted considerable attention since the introduction of Microsoft Kinect as a consumer-level 3D sensor. The basic working principle is to project a known pattern into the scene and infer depth from the deformation of that pattern by using depth from focus and depth from stereo as two classic computer vision techniques. Projecting a known infrared pattern enables data capturing in low illumination conditions, eliminates the dependency on having textured objects, and prevents any confusion by repeated scene textures. However, it restricts the working range of the sensor to few meters, results in more energy consumption, limits the resolution of the data to the resolution of the projector, and becomes problematic in dark surfaces that do not reflect enough light. Additional problems arise if more than one device scans the scene because each structured light source disrupts the scanning of surrounding sensors. A theoretical and experimental analysis on the Euclidean accuracy of data captured by a calibrated RGB-D camera shows that the acceptable working range should be limited to 1–3 m as the random error of depth measurements increases quadratically in relation to the distance from the sensor [33].

Passive sensors (e.g., RGB cameras), on the other hand, do not inject any sort of optical energy into the scene and only rely on natural light in the environment. This allows working anywhere with sufficient illumination (e.g., indoor or outdoor), using less energy, and eliminating the limit imposed on data capturing resolution by the projector. Considering the flexibility and lower cost of passive sensors (no need for special hardware) as well as the fact that data capturing in most civil infrastructure applications necessitates a working range of more than few meters, this section only focuses on the passive sensing category (henceforth is referred to as image-based 3D reconstruction).

2.1. General framework for image-based 3D reconstruction

Research efforts have undertaken extensive study regarding establishment of an image-based 3D reconstruction technique that could be used to capture as-built geometric information of civil infrastructure [10,12,14,25,31,34–38]. Such a technique generally comprises (1) collecting optical sensor data, (2) processing the raw sensor data into 3D points, and (3) modeling and extracting measurements from generated 3D points. Fig. 1 provides an overview of the framework in which the established technique is carried out.

Having an appropriate data capture device, a user moves slowly around a target structure. Depending on camera type, the result will be a set of images or a video footage. For the latter case, an extra procedure is required to pre-process the raw video into a set of key frames [16]. Usually, videotaping an infrastructure scene results in a large video file filled with blurry, noisy, or redundant frames. Key frame selection techniques have been developed particularly to address this issue: remove blurred frames [39–43] and then extract an optimum number of key frames based on