Advanced Engineering Informatics 27 (2013) 270–282

Contents lists available at SciVerse ScienceDirect

Advanced Engineering Informatics

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

Optimized selection of key frames for monocular videogrammetric surveying of civil infrastructure

Abbas Rashidi [a], Fei Dai [b,*], Ioannis Brilakis [c], Patricio Vela [d]

[a] School of Civil and Environmental Engineering, Georgia Institute of Technology, 130 Hinman Research Building, 723 Cherry Street NW, Atlanta, GA 30332, United States
[b] Department of Civil and Environmental Engineering, West Virginia University, P.O. Box 6103, Morgantown, WV 26506-6103, United States
[c] Department of Engineering, University of Cambridge, BC2-07, Trumptington Street, Cambridge CB2 1PZ, United Kingdom
[d] School of Electrical and Computer Engineering, Georgia Institute of Technology, TSRB 441/Van Leer 368, Mail Code 0250, Atlanta, GA 30332, United States

ARTICLE INFO

Article history:
Received 15 November 2011
Received in revised form 26 September 2012
Accepted 14 January 2013
Available online 16 February 2013

Keywords:
Spatial remote sensing
3D reconstruction
Motion blur
Point cloud
Key frames
Infrastructure

ABSTRACT

Videogrammetry is an inexpensive and easy-to-use technology for spatial 3D scene recovery. When applied to large scale civil infrastructure scenes, only a small percentage of the collected video frames are required to achieve robust results. However, choosing the right frames requires careful consideration. Videotaping a built infrastructure scene results in large video files filled with blurry, noisy, or redundant frames. This is due to frame rate to camera speed ratios that are often higher than necessary; camera and lens imperfections and limitations that result in imaging noise; and occasional jerky motions of the camera that result in motion blur; all of which can significantly affect the performance of the videogrammetric pipeline. To tackle these issues, this paper proposes a novel method for automating the selection of an optimized number of informative, high quality frames. According to this method, as the first step, blurred frames are removed using the thresholds determined based on a minimum level of frame quality required to obtain robust results. Then, an optimum number of key frames are selected from the remaining frames using the selection criteria devised by the authors. Experimental results show that the proposed method outperforms existing methods in terms of improved 3D reconstruction results, while maintaining the optimum number of extracted frames needed to generate high quality 3D point clouds.

1. Introduction

Over the past few years, computer vision based technologies have been considered as cost effective and easy-to-use alternatives to traditional spatial sensing methods, e.g., laser scanning and range cameras [1–4]. In the Architecture, Engineering, and Construction (AEC) domain, the collected spatial data is useful for designers, engineers, and inspectors to control/verify quality issues of construction projects [5,6], identify deviations between as-built and as-designed structures [7–9], analyze safety and productivity issues [10], monitor projects' progress in a more proactive manner [11–14], and assess damages caused by disasters [15–17]. Computer vision based 3D reconstruction methods are generally based on processing a number of captured images (photogrammetry) or video streams (videogrammetry) from the scene. In comparison with arbitrarily taken images, sequential video streams provide more valuable information for processing [18]. In the AEC domain, it is more convenient to videotape a relatively large civil infrastructure scene rather than taking hundreds or thousands of images. However, when it is applied to civil infrastructure, the feasibility of videogrammetry significantly suffers from two issues. During videotaping a scene with off-the-shelf cameras, it is difficult to precisely restrict and control the overall camera speed and motion stability. As a result, motion blur will occur, significantly impacting the performance of the reconstruction. Efficiency in processing of video frames is the second issue. Even a short 1 min video stream consists of 1500–1800 frames. Obviously it is computationally expensive to process all these frames. To achieve a satisfactory result, only a small portion of these frames is necessary for post processing. Thus far, there is a lack of effective and automatic methods for selection of informative, high quality frames for 3D spatial sensing of infrastructure using videogrammetry.

Evaluating the feasibility of applying videogrammetry in spatial sensing of civil infrastructure is an active field of research. However, only very little research has been conducted to develop algorithms for selecting key frames in the 3D reconstruction pipeline. Current key frame selection algorithms consider criteria in regard to sufficient overlap between frames, avoiding degeneracies, and minimizing reprojection errors. Optimizing the number of extracted frames is another important factor that is neglected by current research efforts. For example, increasing the number of extracted key frames will improve the density and completeness

* Corresponding author. Tel.: +1 304 293 9940.
E-mail addresses: rashidi@gatech.edu (A. Rashidi), fei.dai@mail.wvu.edu (F. Dai), ib340@eng.cam.ac.uk (I. Brilakis), pvela@gatech.edu (P. Vela).

1474-0346/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aei.2013.01.002

of the resulting point clouds. However, once the results reach a certain level of density and completeness, processing more frames is redundant and makes the procedure computationally ineffective. In terms of applying videogrammetry to practical settings, none of current research undertakings have dealt with motion blur effects contained in captured video, which leads to significantly decreased accuracy of the 3D reconstruction.

Considering the aforementioned gaps in videogrammetric research for surveying civil infrastructure, this paper proposes a novel method of automating the selection of informative, high-quality frames for processing in the videogrammetric pipeline. According to the method, blurred frames are removed using thresholds determined based on the minimum level of frame quality required to obtain robust results. Then, an optimum number of key frames are selected from the remaining frames using a set of selection criteria devised by the authors. Lastly, the proposed method is incorporated into the videogrammetric pipeline and tested on a concrete highway bridge. The remainder of this manuscript is organized as follows: Sections 2 and 3 outline current practices for handling low image quality and video frame redundancy and recent research efforts in this direction. This is followed by the research statement and objective, and our proposed method for automating the selection of high quality, informative frames for 3D reconstruction of infrastructure in Sections 4 and 5. In Section 6, experiments are conducted to test the validity of the proposed key frame selection algorithms and measure the accuracy and completeness of applying the complete videogrammetric pipeline to reconstructing a concrete highway bridge. Finally, conclusions are drawn in Section 7.

2. State of practice

Over the past two decades, videogrammetry has emerged as a popular reconstruction and measurement tool [18,19]. It is mainly utilized for reverse engineering in the manufacturing industry for the purpose of collecting geometric and spatial data of objects [20]. For most manufacturing reverse engineering cases, objects are small (i.e., in comparison with civil infrastructure), video clips are short (i.e., a couple of seconds) and data are collected in controlled settings (i.e., stabilized camera stations or uniform speeds of camera movements). As a result, quality of frames and necessity to extract a small amount of frames for post processing is not a major issue. For most industrial applications, it is possible to process all frames or just a number of them that are selected at a steady interval based on a specific frame rate.

While videogrammetry works well for the manufacturing industry, applying it to the civil infrastructure domain faces several practical constraints, preventing its adoption in real construction practices [21]. Poor quality of captured frames is a major concern that significantly undermines the performance of 3D reconstruction. Unlike manufacturing applications in indoor environments, it is difficult to control outdoor conditions. Motion blur is inevitable due to occasional jerky movements of the camera that occur when traversing a site. Moreover, considering size and level of complexity, videotaping civil infrastructure usually takes several minutes instead of a few seconds. This means that there are tens of thousands of frames that need processing. Estimating the camera poses and the 3D scene structures is computationally expensive if it is performed on all frames in a video sequence. If the video sequence can be temporally decimated, then this process can be executed more efficiently.

Sufficient baseline between two consecutive frames is another important factor for any robust 3D reconstruction, on the grounds that the short baselines usually induce larger measurement errors than those produced by the long baselines (Fig. 1) [22].

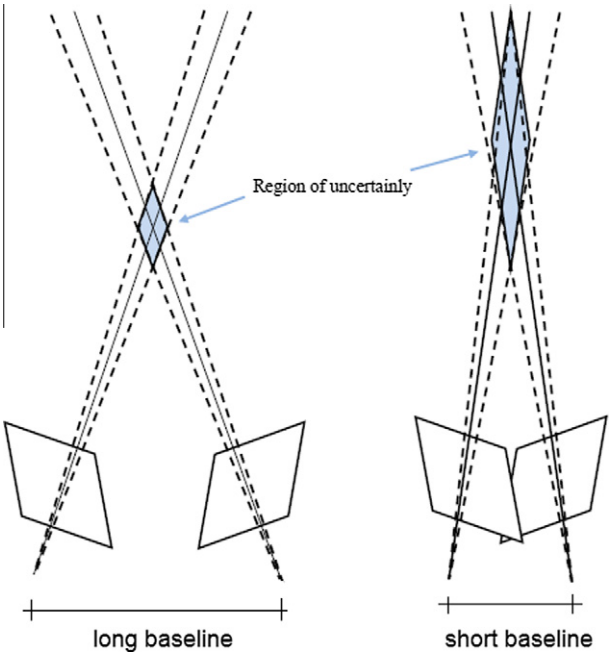


Fig. 1. Impact of length of baseline on accuracy of 3D reconstruction results.

Consequently, consecutive frames must have long enough baselines for accurate 3D reconstruction.

Besides the above mentioned problems, there are failure risks for the 3D reconstruction algorithms known as degeneracy cases. Usually, the 3D reconstruction algorithms perform well if there is a generic camera motion (i.e., a camera movement has both rotations and translations) and generic point positions (i.e., the scene has points with different depths). However, when the generic conditions for camera motion or structure do not hold, the method fails. The degeneracy cases can be divided into two categories:

- Motion degeneracy: If a camera rotates around its center without any translation, epipolar geometry cannot be achieved.
- Structure degeneracy: If all the 3D points in the physical space are coplanar, the fundamental matrix cannot be uniquely determined by the image correspondences themselves [23].

Both cases are common during the data collection of an infrastructure scene. The first case happens since it is very common that the videotaper stops for a while and just rotates the camera. It also might happen if the videotaper changes his direction. The second case happens in many cases that civil infrastructure scenes are coplanar, e.g., a wall or a slab.

3. State of research

Several research undertakings have taken place in creating videogrammetric pipelines to reconstruct 3D buildings and infrastructure [24–26]. The authors in Brilakis et al. [21] created a framework that exploits a binocular stereo camera system to progressively reconstruct an ongoing building structure. Fathi and Brilakis [27] proposed a method to generate the sparse point cloud of the scene from stereo images. These efforts are the basic steps toward applying videogrammetry in 3D spatial sensing of civil infrastructure. Considering the practical constraints (i.e., low frame quality and video frame redundancy) encountered on site, a robust videogrammetric method is needed. Fig. 2 depicts a pragmatic videogrammetric pipeline built upon the existing methods. As shown in Fig. 2, the processes of camera calibration, structure from

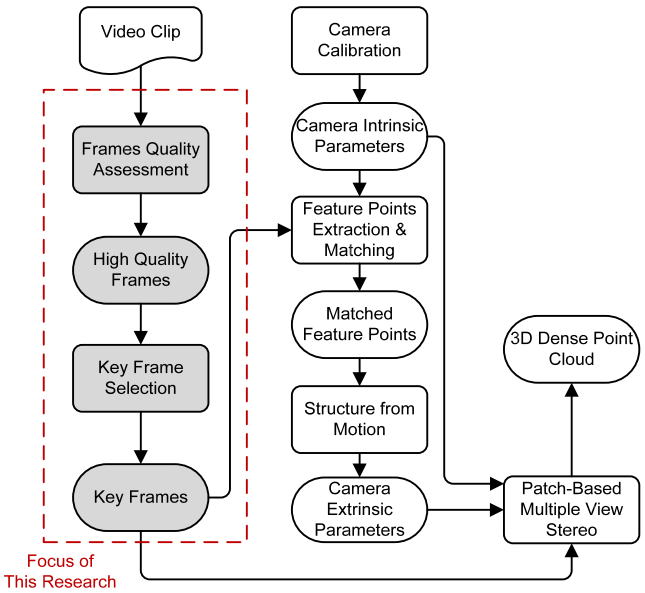


Fig. 2. Videogrammetric pipeline for 3D reconstruction.

motion (SfM), and multi-view stereo (MVS) work with existing techniques (i.e., white color boxes and ellipses). The focus of this paper is methods (i.e., grey color boxes and ellipses) needed to address the practical constraints mentioned above. Research related to this area from computer science is presented in the following sections.

3.1. Strategies for quality control of video frames

One challenge in processing video frames is to deal with the low quality frames, which primarily are associated with motion blur [28]. Motion blur is caused by camera shakes. Even slight shaking may lead to serious blur effects. Normally, there are two strategies for dealing with this problem.

The first strategy is image deblurring, which has recently received much attention among the computer vision community. Deblurring techniques could be divided into two main categories:

3.1.1. Hardware methods

The most common commercial method for reducing image blur is image stabilization (IS). These methods, used in high-end lenses and now appearing in lower-end point and shoot cameras, use mechanical means to dampen camera motion by offsetting lens elements or translating the sensor. Fundamentally, IS tries to dampen motion by assuming that the past motion predicts the future motion [29]. However, it does not counteract the actual camera motion during an exposure nor does it actively remove blur – it only reduces blur.

3.1.2. Software methods

For software methods, different algorithms are utilized to model motion blur. If a motion blur is shift-invariant, it can be modeled as the convolution of a latent image with a motion blur kernel, where the kernel describes the trace of a sensor. Then, removing a motion blur from an image becomes a de-convolution operation. In non-blind de-convolution, the motion blur kernel is given and the problem is to recover the latent image from a blurry version using the kernel. In blind de-convolution, the kernel is unknown and the recovery of the latent image becomes more challenging. Several methods were proposed to handle the motion blur problem [30–32]. While these methods might produce excellent deblurring results, they necessitate intensive computation. It usually takes more than several minutes for the methods to deblur even a single image of moderate size, making this method not applicable for high volume applications.

The aforementioned hardware and software methods reflect the state-of-the-art of current video processing techniques. However, these methods are not feasible for use in infrastructure reconstruction applications considering their computational complexity [32]. Moreover, reconstructing infrastructure scenes only needs to process a few frames rather than process every frame contained in a video stream.

The second strategy involves removing blurred frames. Removing these frames relies on applying blur metrics that are capable of measuring numerical values that represent the extent of the effect. To measure the blur effect, the intensity change along edges was extensively studied [33–35]. Such algorithms are easy to use for predicting sharpness/blurriness of an image. However, evaluation of these algorithms is mainly focused on widths of intensity edges, which does not reflect completely the quality of the measured images. To address this limitation, Chen and Bovik [36] established a multi-resolution decomposition method capable of extracting reliable features regarding image blurs. The advantage of this method is all the pixels of an image are taken into account and the estimate is more reliable. However, each time an image is assessed; a training process of sample images is needed, imposing extra work on the modeler.

Apart from the above methods, the BluM metric method [37] is based on the human blur perception. It uses the blur discrimination properties of the human perception to create a reference image for blur estimation, making this method simple yet practical. Here, blur discrimination means that the human visual system is capable of differentiating a blurred and a sharp image but cannot accurately discern a blurred image with the one that is re-blurred. By this method, the complete scene in the image is considered, and the resulting scores are simply denoted by the numerical range from 0 to 1 representing the worst blur and the best sharpness of the image, respectively. Therefore, the BluM metric is a convenient technique to measure the blur effect of video of an infrastructure, and its feasibility will be evaluated in this research. The end result of this step is extracted high quality frames from a video stream.

3.2. Strategies for key frame selection

Selecting a number of representative (key) frames from a large video sequence is a key step for efficient and robust 3D reconstruction. The criteria that have been applied for key frame selection are primarily focused on four aspects: (1) sufficient overlap, (2) sufficient baseline, (3) degeneracy avoidance, and (4) minimized reprojection errors. Seo et al. [38] considered three criteria when extracting key frames: (a) the ratio of the number of correspondences to the number of features, (b) the homography error, and (c) the distribution of correspondences over the frames. In their method, the criterion (a) is used to ensure the sufficient overlap of two frames, the criterion (b) serves as a good proxy for the baseline distances between two views, and the criterion (c) is used to increase the accuracy of calculating the fundamental matrices, based on which to lead to higher accurate estimations of camera motions and an object structure. Later on, Seo et al. [39] improved their method in [38] to extract informative key frames by incorporating a fourth criterion: that the reprojection error of the reconstruction process is minimized. Nonetheless, these methods did not consider the degeneracy cases. In Pollefeys et al. [25], the degeneracy problem is addressed by employing the Geometric Robust Information Criterion (GRIC) [40]. The next key frame is selected only once the fundamental matrix model explains the

relationship between the pair of images better than the homography matrix model through the scores calculated by the GRIC [40].

In addition, Ahmed et al. [22] proposed another key frame selection method which is based on the weighted score considering the GRIC difference and point to epipolar line cost. In their method, they also use the correspondence ratio to indicate whether long baseline and sufficient overlap between two frames exist. Gibson et al. [41] proposed a method in which the sum of three weighted addends of (1) the fraction of features that were matched in the previous frame pair which cannot be matched in the current pair, (2) the inverse of the square of the homography error, and (3) the squared median epipolar error, is minimized to selects the frames. Sufficient overlap, sufficient baseline, and degeneracy avoidance are ensured in their method. Thormahlen et al. [42] proposed a criterion for the selection of key frames with the lowest expected estimation error of initial camera motion and object structure. At the same time, their method utilized the GRIC to guarantee a sufficient baseline, overlap, and avoid degeneracy cases.

In summary, the above discussed methods are able to guarantee a long enough baseline, sufficient overlap, no degeneracy cases, and minimized reprojection errors of 3D reconstruction. However, these research efforts have not taken into account an optimized number of required frames for processing. Besides, the practical constraints (i.e., speed of camera movements, complexity of the scene) also significantly impact the key frame selection algorithms, and they have not been addressed by the existing research, as there was no need to do so in other application domains. Videotaping large, outdoor, infrastructure scenes poses these additional unique challenges that must be addressed for videogrammetry to be feasible for surveying of civil infrastructure.

4. Problem statement and objectives

While videogrammetry has been conceptually proven to be viable for collecting spatial data of civil infrastructure [21,43], two major issues have not been holistically addressed. The first is the question of how to deal with low quality (i.e., blurry) image frames. Poor quality frames result mainly from motion blur, which is a common issue in videotaping civil infrastructure scenes. Not only do these poor quality image frames significantly increase the reprojection errors in a 3D reconstruction pipeline, in the worst cases, they can also potentially lead to failure if a sufficient number of feature points cannot be extracted from these blurry frames (Fig. 3).

The second major issue concerns the sheer size of the video files collected via video recording. As the number of frames increases, the computation costs of the videogrammetric pipeline exponentially increase. For example, capturing a 5 min video clip using a 25 fps camcorder, which is common to off-the-shelf cameras, will result in 7500 frames. Obviously, processing every frame would be costly in terms of computational time and complexity. In civil infrastructure applications, only a small portion of these frames (e.g., 5–15%) are necessary to generate high quality dense point clouds of scenes. Hence, instead of processing all frames of a video or uniformly selecting a number of frames based on the capturing rate, creating an automated algorithm for selecting a number of informative, high quality frames is vital.

As mentioned previously, several researchers have proposed algorithms for selecting key frames. In their research, criteria which theoretically affect a videogrammetric pipeline (length of base line, sufficiency of overlap, degeneracy avoidance, and minimized reprojection errors) were taken into account. However, none of these methods considered the blur as an obstacle. There is also a lack of an optimization strategy in current key frame selection methods. Though current key frame selection methods reduce the number of frames, there is no guarantee that the number of extracted frames is optimum. If the number of extracted frames is less than the number required, it is not possible to generate a high quality point cloud. On the other hand, processing unnecessary extra frames is redundant and time consuming.

To fill in the gaps mentioned above, the objective of this paper is to propose an innovative key frame selection algorithm tailor made for use in the civil infrastructure domain. The proposed algorithm not only tackles common issues that occur while running a 3D reconstruction pipeline, but also addresses two major practical problems, i.e., low video quality, and optimization of the number of extracted key frames. Also, this paper intends to evaluate the impact of applying the key frame selection algorithm on the videogrammetric pipeline.

5. Automated key video frames extraction: Methodology

This paper presents a novel method of selecting key frames for the purpose of robust 3D reconstruction of infrastructure objects. The proposed method takes into account six significant factors for creating the optimal key frame selection algorithm: high quality frame extraction, determining sufficient overlap between adjacent frames, determining baseline length, data degeneracy avoidance, uniform distributions of features in each frame, and the optimization of the number of the extracted key frames. Considering that the existing key frame selection criteria have been well established in dealing with the first four factors listed, this paper does not intend to re-address these factors. Instead, this paper adopts these factors and expands the key frame selection pipeline by incorporating the factors of: uniform feature distribution, and optimized number of key frames extracted for the video footage. The main contribution of this paper is the creation of these two new elements for the purpose of augmenting the existing key frame selection algorithms.

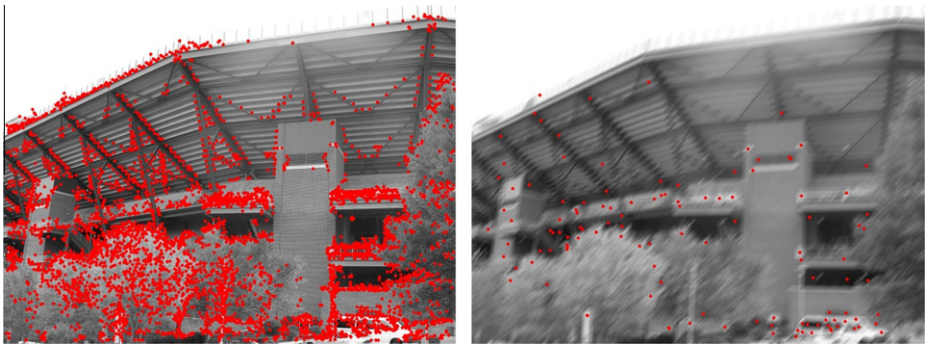


Fig. 3. Impact of blur effects on the numbers of extracted feature points: 1107 feature points for a high quality frame (left) versus 104 feature points for a blurry frame (right).

Fig. 4 shows the main workflow of the proposed key frame selection algorithm. This algorithm first ensures that a sequence of high quality video frames is obtained, and starts with marking the first frame from the sequence as a key frame. From there, a number of subsequent frames are nominated as key frame candidates depending upon sufficient overlaps and baseline lengths. Among these candidates, those that lead to degeneracy cases and large re-projection errors are also removed. Finally, the most suitable candidate is selected as the next key frame based on the uniform distribution of features over the frame. This selection process is then repeated, this time with the newly selected key frame as the starting key frame; this continues through the length of the video footage. Next, the algorithm examines whether the number of the extracted frames is optimized (i.e., the number falls in an optimum range). If yes, the algorithm terminates and exits. Otherwise the algorithm applies a linear programming method to update the parameters of the baseline and overlap criteria, and repeat the selection process. We present the detail of the algorithm in the following.

Input: High quality video frames: In order to address blurry video frames, we follow a low quality frame filtering approach. We utilize the BluM metric [37] to measure the quality of frames and remove the blurry ones.

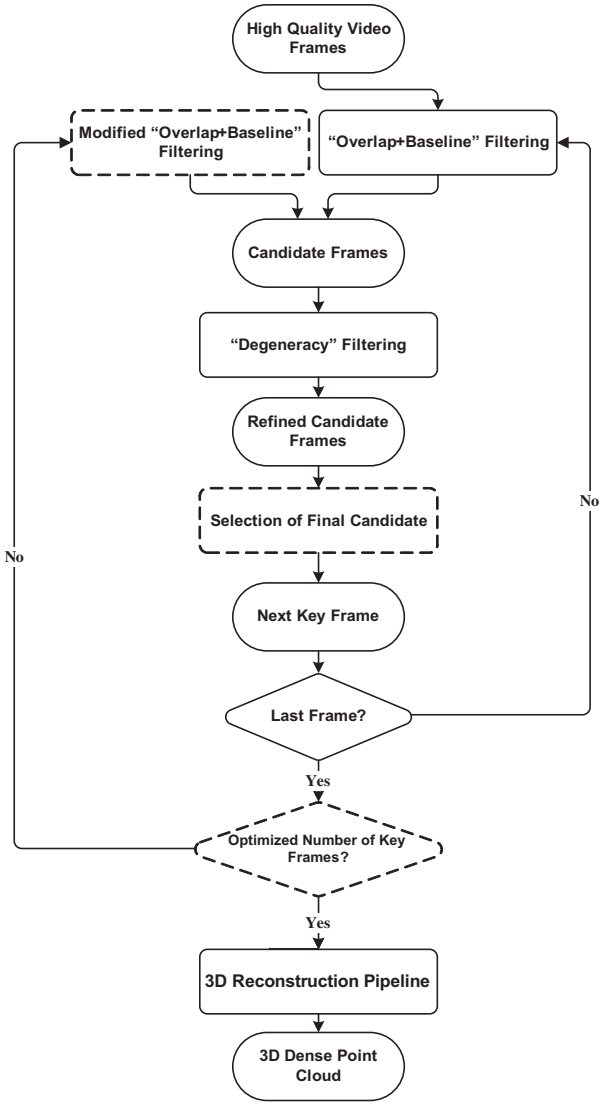


Fig. 4. Workflow of the key frame selection algorithm.

Step1: Overlap and baseline filtering: Once the first frame has been selected as a key frame, it is necessary to select a number of consecutive frames as key frame candidates. The selection criteria must guarantee both enough baseline and sufficient overlap between the candidates and the first key frame. To achieve this goal, we use the correspondence ratio defined by Seo et al. [39] and Ahmed et al. [22].

Step2: Degeneracy filtering: In order to avoid degeneracy cases, we follow a similar strategy suggested by Torr and Pollefeyes [40] by calculating GRIC scores. Using GRIC score also assists the pipeline in retaining frames with minimal re-projection errors.

Step 3: Selecting the next key frame: After filtering a number of frames from the video stream, the next step is selecting the final candidate among the remaining frames. While using both the fundamental and homographic matrices we realized that the ratio of inliers to total number of feature points was an accurate indicator of how compatible two frames were modeled using these two matrices [44]. In addition, this realization correlated with the knowledge that evenly scattered correspondence points between the candidate frame and the key frame were desirable for achieving a more accurate fundamental matrix. This improved matrix could then, in turn, be used to compute a more accurate structure model. Considering this positive relationship, we propose the following procedure for selecting the next key frames from the possible candidates.

After calculating the fundamental and homography matrices between the candidates and the key frame using RANSAC, we calculate the percentage of inliers to the total number of correspondences. Then we calculate the S score to select the final candidate:

$$S = (1-\sigma)\frac{S_F - S_H}{S_F} \tag{1}$$

where S_H is the percentage of inliers for calculating the homography matrix; S_F is the percentage of inliers for calculating the fundamental matrix, and σ is the standard deviation calculated from measuring how uniform the distribution of features is over the frame. To calculate the σ, the frame is divided into sub-regions. The point density for sub-regions and the entire frame are calculated separately. Then, the standard deviation is calculated using the following equation:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(N_i - \frac{N}{n}\right)^2} \tag{2}$$

where N_i and N are number of features in each sub-region and the entire frame respectively; n is the number of sub-regions. The frame with the highest S score will be chosen as the next key frame (Fig. 5).

Step 4: Optimizing the number of key frames: The difference between our method and those of previous researchers lies in the optimization of the number of extracted key frames required for use in the 3D reconstruction pipeline. First, we define the correspondence ratio as follows:

$$R = \frac{R_c}{R_T} \tag{3}$$

$$\tau_1 < R < \tau_2 \tag{4}$$

In Eq. (3), R_c is the number of correspondence points between the key frame and the next candidate while R_T is the total number of feature points in the first key frame. In Eq. (4), τ_1 and τ_2 are the lower and upper thresholds of R. R is inversely proportional to the length of camera motion since, as the camera moves, features tend to leave the scene. Researchers in computer vision have usually set fixed thresholds for τ_1 and τ_2 based on experiments conducted on a few datasets. However, it is not ensured that an optimum quantity

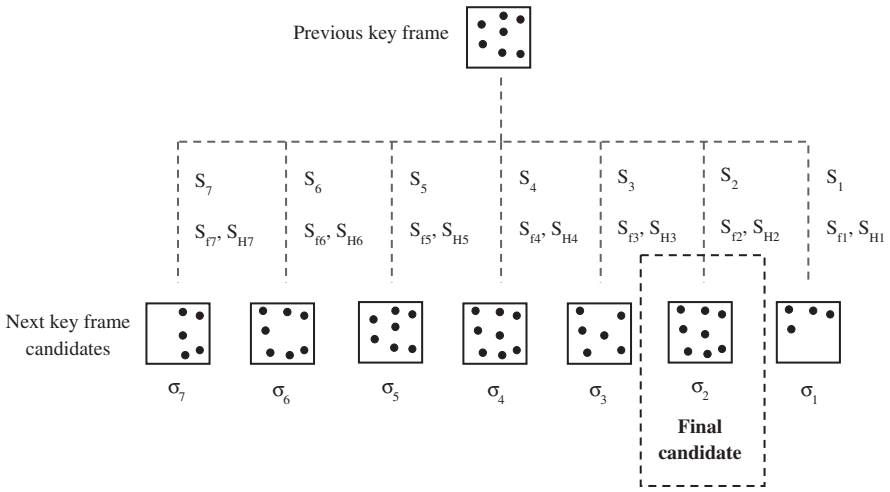


Fig. 5. Selection of the next key frame.

of key frames can be selected. In our case, instead of assigning fixed values as the upper and lower thresholds, we define a specific range for each one based on three important factors: desired number of extracted key frames, approximate speed of camera while traversing, and complexity of the civil infrastructure scene.

Given a defined set of ranges for the upper and lower thresholds, we use a linear programming method (Eqs. (5) and (6)) to optimize the number of the required frames:

$$\text{Goal}: \; P_1 \leqslant P \leqslant P_2 \tag{5}$$

$$\text{Constraints}: \left\{ \begin{array}{l} \tau_{lmin} \leq \tau_1 \leq \tau_{lmax} \\ \tau_{umin} \leq \tau_2 \leq \tau_{umax} \end{array} \right\} \tag{6}$$

In Eq. (5), P is the percentage of key frames over the entire number of frames existing in the sequence and P_1 and P_2 define the optimum range for the percentage of the number of key frames. The optimum range for the number of key frames is pre-defined based on the capturing rate of the camera, and movement speed of the video taper. In Eq. (6), τ_{lmin}, τ_{lmax}, τ_{umin} and τ_{umin} are acceptable ranges for the upper and lower thresholds which can be obtained through experiments in various scenarios. Fig. 6 shows the relationship between the number of frames and correspondence ratios. As discussed earlier, the correspondence ratio is inversely proportional to the number of total frames in the sequence.

Given P_1, P_2, τ_{lmin}, τ_{lmax}, τ_{umin} and τ_{umin} are known, the algorithm uses the average values of τ_{lmin}, τ_{lmax}, τ_{umin} and τ_{umin} as the upper and lower limits (Eq. (7)) for the first round in selecting a set of key frames. The upper and lower limits are used to specify the parameters of the baseline and overlap selection criteria.

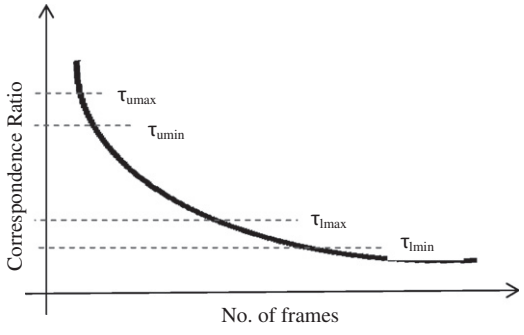


Fig. 6. Acceptable ranges for upper and lower correspondence ratio thresholds.

$$\begin{aligned} \tau_{l(first\ try)} &= \frac{\tau_{lmin} + \tau_{lmax}}{2} \\ \tau_{u(first\ try)} &= \frac{\tau_{umin} + \tau_{umax}}{2} \end{aligned} \tag{7}$$

Then, the optimization begins with checking whether the number of key frames for the first round is within the acceptable range, i.e., between P_1 and P_2. If not, the algorithm will run again. This time, the parameters of the baseline and overlap criteria will be adjusted based on Eqs. (5) and (6) with the current number of extracted key frames. The procedure will be repeated until an acceptable number of key frames is achieved. The result is an optimum number of key frames that can be used for processing within the 3D reconstruction pipeline. It is necessary to mention that since the feature points and their matching information for each frame in the video sequence already exist, repeating the search for a new set of key frames is not computationally expensive.

6. Implementation and results

A C# based prototype was implemented to test the validity of the proposed key frame selection algorithm and the videogrammetric pipeline. It was written in Visual Studio 2010 using Windows Presentation Foundation (WPF) and publicly available libraries such as OpenCV 2.0 (wrapped by EmguCV) for access to computer vision tools and DirectX 10 for the graphic display of results. The Open CV's image structure was the primary data structure. It removed the conversion needs of the image processing tools from that library, which drastically reduced the processing speed.

As the first stage of implementing the videogrammetric pipeline, we used the camera calibration method proposed by Zhang [45] to extract the intrinsic parameters of the camera accurately. As the next step, feature points on different frames must be extracted and matched. SIFT and SURF are two well-known feature detectors and descriptors among the computer vision community [27]. Due to higher dimensionality of descriptors, SIFT outperforms SURF in terms of the number of extracted feature points; however, SURF algorithm is computationally more effective. Selecting the most appropriate feature detector is a trade-off problem between computational costs and number of extracted features. A number of studies have been conducted to compare the performance of the two algorithms. Fathi and Brilakis [27] reported that for a number of civil infrastructure video frames, the average number of extracting feature points using SIFT is 31% greater than SURF. They

also mentioned that the average runtime of SIFT was 4.34 times of that for SURF [27]. In our research, we used SURF as the feature detector since computational efficiency was the more significant issue in processing sheer volumes of video frames.

As the next stage, the camera motion, i.e., camera translation and rotation, as well as the 3D coordinates of the feature points were computed through a procedure known as Structure from Motion (SfM) [44,46]. Considering the fact that there is sufficient baseline and adequate overlap between consecutive key frames, the SfM starts with processing of the first two key frames. Then the 5 point algorithm, proposed by Nister [47,48], is used to calculate the motion of the camera (extrinsic parameters) between these two frames. Next, corresponding feature points in these two views are triangulated, followed by the Sparse Bundle Adjustment [49]. Once this is done, the next key frame is added to the process. The extrinsic parameters of the camera associated with the new key frame are computed using the direct linear transform (DLT) technique inside a RANSAC procedure and points observed by the new frame are added to the process [44]. Finally, a global Sparse Bundle Adjustment is run to refine the results and minimize the propagated errors. This procedure proceeds until the last key frame is processed. Then, the extrinsic parameters derived from SfM along with the intrinsic parameters obtained from the calibration step, are fed to the Patch-based Multi-View Stereo (PMVS) [50] algorithm to generate the dense point cloud.

Considering the variety in civil infrastructure scenes, we captured 25 video streams from eight different scenes, i.e., two highway bridges, three campus buildings, one residential building, one sport facility, and one concrete water reservoir. The lengths of the video streams varied from 4 to 10 min. To validate the degeneracy cases, we captured some video streams that contain planar scenes such as walls, and in some cases during the process of capturing videos, the videotaper intentionally stopped for a while and the camera was simply rotated without any translations.

6.1. Identifying threshold for low quality frame filtering

The first step of applying the frame quality control filter is identifying the BluM metric threshold [37]. The threshold is used for determining the frames that have to be removed. If the measured value of a frame obtained by using the metric is larger than the threshold, the frame is removed. Otherwise, it is retained. In determining the threshold, the criterion is that a minimum level of frame quality required to obtain robust results be ensured. To achieve this, a sample set of satisfactorily high quality images (20 from each video stream) were selected through human observations. According to Sheikh et al. [51], an obvious way of measuring the quality of an image or video is to solicit opinion from human observers. Applied with the blur metric, the sample set of frames resulted in the sample mean 0.283, and the sample standard deviation 0.01 of the metric evaluation scores. The left-sided 95% confidence limit for a normal distribution was then used to determine the statistical estimate of the threshold as $0.283 + 1.64 \times 0.01 = 0.299$. This means that with 95% likelihood, any "satisfied" image would have a score measured by the proposed metric of no more than 0.299. It is noted that the sample size is statistically significant to the threshold analysis in consideration of the expected accuracy level being in the order of 0.001 and the relatively small variation on the sample standard deviation of the resulting scores [52].

6.2. Identifying thresholds for key frame selection

In order to implement the key frame selection algorithm, the values of the thresholds depicted in Eqs. (3) and (4) have to be determined. Fig. 7 shows the relationship between the correspondence ratio and the number of frames for a number of captured video streams from various civil infrastructure scenes. As shown in Fig. 7, rapid changes in R might occur at different points. From the experiments on civil infrastructure scenes, we inferred that the changes in the correspondence ratio mainly depended on two factors:

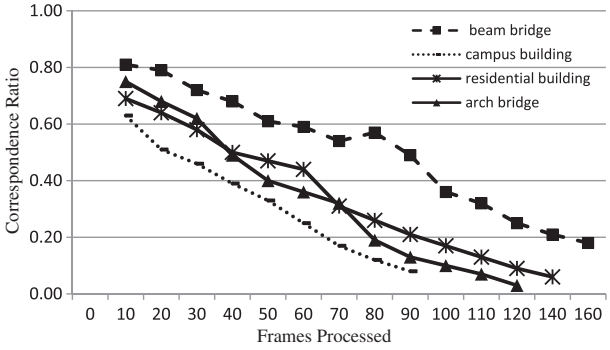


Fig. 7. Samples of correspondence ratios for different video streams.

- *Speed of camera movements:* It is difficult to measure the real speed of a camera when it was traversing around the jobsite. However, considering 25 fps as a common rate for a regular video camera, we defined two categories for camera movement speed:
  - Normal movement: The videotaper traversed with the handheld camera at a speed of more than around 0.7 m/s (one step per second).
  - Slow movement: The videotaper traversed with the handheld camera at a speed of less than around 0.7 m/s.

As an example, the correspondence ratios for a number of frames of a sample infrastructure video clip, in both cases of normal and slow movements, are presented in Fig. 8.

- Uniformity of the scene: Some of civil infrastructure scenes contain uniform texture, such as curtain walls or exposed concrete surfaces. On the other hand, some other civil infrastructure scenes contain complex variable texture. The correspondence ratio declined rapidly in complex scenes where the texture of the scene changed rapidly. However, in relatively uniform scenes, e.g., curtain walls or exposed concrete surfaces, the changes were relatively slow. Given these observations, scenes were defined as either complex or uniform.

Other than uniformity in the texture of different surfaces, factors such as reflection or transparency might also affect the results in terms of numbers of extracted feature points, values in correspond-

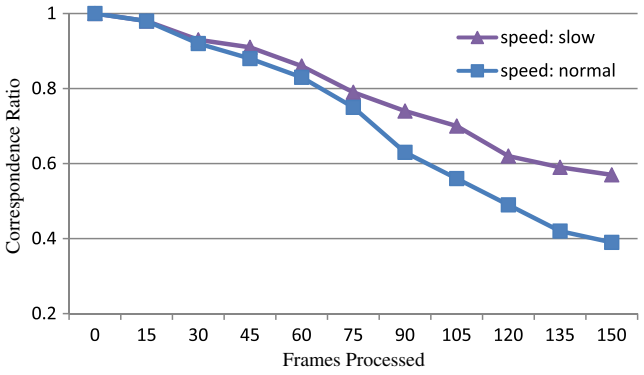


Fig. 8. Impact of camera speed on corresponding ratio: Arch bridge scene.

Table 1
Upper and lower thresholds for correspondence ratios.

Complexity of the scene	Camera motion speed	Lower threshold (τ_{lmin}–τ_{lmax})	Upper threshold (τ_{umin}–τ_{umax})
Complex	Normal	0.85–0.7	0.55–0.4
Uniform	Normal	0.8–0.65	0.5–0.35
Complex	Slow	0.75–0.6	0.4–0.3
Uniform	Slow	0.8–0.65	0.45–0.3

ing ratios, and accuracy of the feature point matching algorithm. It is well known that reflective or transparent materials, such as glass, have a negative impact on the overall performance of the 3D reconstruction algorithm. Assessing the impact of these categories of surfaces on the performance of 3D reconstruction algorithms is a potential topic for future research.

Based on the above mentioned different scenarios, the ranges proposed in Table 1 are used as acceptable ranges for upper and lower thresholds.

As the next step, the optimum range of the number of frames should be estimated. Infrastructure scenes could be reconstructed using various numbers of frames. Obviously, this is a trade-off problem. Processing more frames results in higher quality point clouds with higher computational costs. In this research, in order to find the optimum number of frames required for processing, different numbers of key frames were extracted and processed for different video streams captured in one specific scene, i.e., a concrete highway bridge. The completeness of the generated dense point clouds from different numbers of key frames were calculated using a method explained later. The results of measuring completeness rates for different numbers of key frames from one video sequence with capturing rate of 25 fps are illustrated in Fig. 9.

Based on Fig. 9, for a capturing rate equal to 25 fps, the maximum density of a point cloud is achievable by processing a small selection of frames (7–11%), indicating that processing more

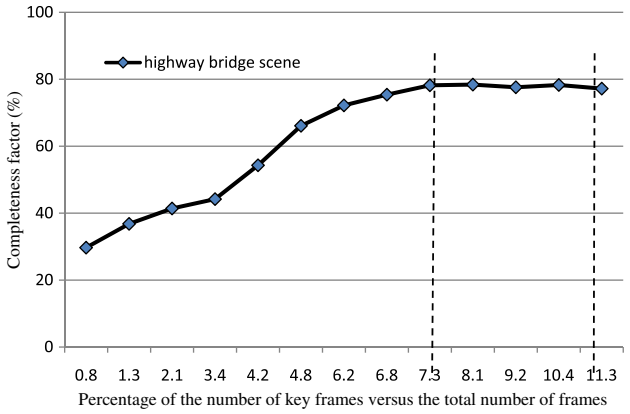


Fig. 9. Completeness of generated dense point clouds versus the percentage of the extracted key frames to total number of frames.

Table 2
Optimum ranges for the percentage of number of key frames to total number of frames.

Capturing rate (fps)	Optimum ranges for ratio of number of key frames to total number of frames (%)	
	P_1	P_2
Up to 5	25	75
6–10	15	50
11–15	10	25
16–20	8	15
21–25	7	12

frames is redundant. By conducting the same experiments on eight different civil infrastructure scenes with different video capturing rates, optimum ranges for number of key frames for civil engineering scenes are suggested in Table 2.

It is necessary to mention that the proposed optimum ranges are based on experiments on a limited number of civil infrastructure scenes; and in other specific scenes these values might be slightly different.

6.3. Validation of the proposed method

The validation procedure focused on two aspects:

1. Comparing the performance of the key frame selection algorithm with other existing methods, and
2. Evaluating the quality and accuracy of generated point clouds obtained from the videogrammetric pipeline, with and without using the proposed key frame selection algorithm.

Step 1: Validating the key frame selection algorithm: We firstly qualitatively evaluated the impact of blur effects on the results of 3D reconstruction. To this end, the frames from the video of a highway bridge were artificially blurred. Then, the blurred frames were passed into the reconstruction pipeline. For each case, the BluM value, the number of extracted and matched feature points, and reprojection errors were computed. As an example, six sample blurred frames are illustrated in Table 3. As observed, in all cases, blur has a significant effect on the number of extracted features and reprojection errors; in most cases, the number of features extracted from the blurred frames was not even sufficient to reconstruct the scene.

In the next step, we measured the performance of the proposed key frame selection algorithm. A number of key frames were extracted from each video sequence using our key frame selection algorithm and the methods developed by [25,40,42]. In addition, a number of frames equal to the number of key frames by our method were selected at equally distributed intervals without using any key frame selection method. These extracted key frames were processed in the videogrammetric pipeline and the percentages of failure cases as well as reprojection errors for successfully reconstructed cases of each method were computed. The obtained results are summarized in Table 4.

As evident from Table 4, our method outperforms all other methods in terms of failure cases and average re-projection errors. The only exceptional case is that the Thormahlen method generates a lower average re-projection error than our method. However, in this specific case, the number of extracted key frames by our method is within the desired range. Despite the fact that our method is slightly more time consuming in the phase of key frame selection, the optimized number of extracted key frames drastically reduces the computational time of post-processing, and thereby optimizes the entire efficiency of the videogrammetric pipeline.

Step 2: Validating the videogrammetric pipeline: The second phase is to test the performance of the entire videogrammetric pipeline. To this end, a reinforced concrete bridge located on the intersection of Boat Rock Road SW and Camp Creek Parkway, Atlanta, GA was selected as the test-bed for the experiment. It is a four-span bridge with three rows, and each row contains three rectangular columns (Fig. 10).

In order to evaluate the impact of the proposed key frame selection algorithm on the performance of the videogrammetric pipeline, we planned on two scenarios as described in the following:

1. Evaluating the pipeline with the proposed key frame selection algorithm implemented, and

Table 3
Impact of blur on number of extracted feature points and reprojection errors.

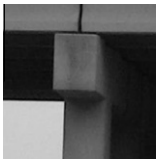
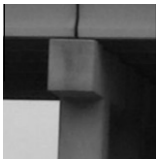
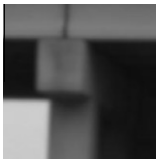
	BluM value	Number of extracted feature points	Number of matches	Reprojection error		BluM value	Number of extracted feature points	Number of matches	Reprojection error
	0.259	2984	657	0.0225		0.622	1017	316	0.0548
	0.329	2562	579	0.0394		0.655	299	107	0.0751
	0.465	2077	474	0.0423		0.704	147	24	0.132

Table 4
Failure percentages and re-projection errors for different key frame extraction methods.

	Method	Average failure percentage (%)	Average reprojection error	Average number of extracted key frames	Running time (h)	
					Key frame selection	3D reconstruction
1	Uniformly extracted frames	54.54	0.0754	432	–	12
2	Pollefeys et al. [25]	36.36	0.0358	619	0.02	21
3	Thormahlen et al. [42]	27.27	0.0208	509	0.05	15
4	Seo et al. [39]	45.45	0.0439	573	0.07	18
5	Our method	22.72	0.0275	432	0.14	12

2. Evaluating the pipeline by selecting the same number of frames as obtained from the key frame selection algorithm but based on equally distributed intervals.

In collecting the on-site videos, we used an off-the-shelf Canon Vixia HF S100 with its resolution set at 2 MP. The camera was calibrated using Bouguet's camera calibration toolbox [45]. Then by applying the videogrammetric pipeline, dense point clouds of the bridge were generated. Fig. 11 shows a snapshot of a generated dense point cloud using the pipeline with the key frame selection algorithm incorporated.

There are some clarifications regarding the use of PMVS in generating the point clouds. In order to effectively run the algorithm, a number of parameters should be adjusted. Except for the following two parameters, the authors used the pre-defined values suggested in [50] for setting of the PMVS parameters:

- MaxAngle: This parameter defines the maximum acceptable angle between two visible cameras. We changed the predefined value, i.e., 10°, to 8° to improve the reconstruction process for the scenes located far from the camera (which is the case for several civil infrastructure scenes).
- Sequence: The PMVS algorithm is primarily designed for use with unordered images. In the case of using sequential key video frames it is possible to narrow down the reconstruction process to only a few consecutive frames. We set this parameter to 7 so only 7 frames before and 7 frames after the target frame are considered for.

Fig. 10. Snapshot of the test-bed bridge taken from the distance 55 m.

As shown in Fig. 11, there might be some gaps or outliers associated with generated the point clouds. This might be the result of several factors including:

- Specific views of the scene are not fully covered during videotaping or extracting the key frames.
- Surfaces of concrete members are generally texture-less, making the reconstruction procedure more challenging.
- Assigned values to the PMVS parameters, e.g., MaxAngle, patch size or cell size are not optimum. More information could be found at [50].
- Environmental and hardware parameters such as lens distortions or occlusions might have negative impact on the performance of the 3D reconstruction algorithm.

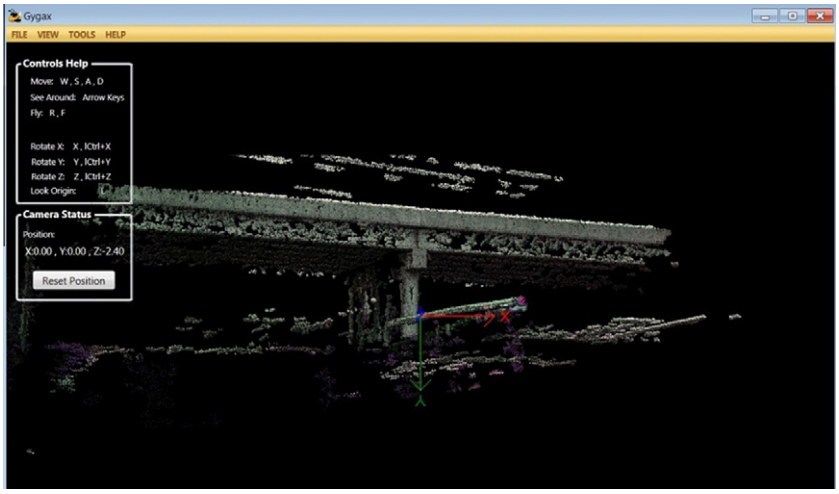


Fig. 11. Snapshot of produced dense point cloud.

Along with the video capturing, a SOKKIA Series 30R reflectorless total station was used for collecting data as the ground truth (Fig. 12a). As a result, around 2000 points on the surfaces of the bridge were captured and their spatial coordinates were derived. These points were then rendered into a surface model for this bridge by exploiting the surface reconstruction Poisson algorithm [53], which generates a water-tight surface given the spatial points and each of their normal vectors, as shown in Fig. 12b.

Automated generation of surfaces using 3D points might be a challenging task. In this research, since the points are collected by total station as ground truth for validation purpose, there were no outliers involved in the surface model reconstruction. In order to rendering the 3D ground truth model of the bridge, each surface element is automatically generated followed by a manual process of stitching all surfaces into a model based on the priori knowledge of geometric relationships (i.e., parallel, perpendicular, coincident) of those elements.

Automated generation of surfaces using 3D points might be a challenging task. In this research, since the points are collected by total station as ground truth for validation purpose, there were no outliers involved in the surface model reconstruction. In order to rendering the 3D ground truth model of the bridge, each surface element is automatically generated followed by a manual process of stitching all surfaces into a model based on the priori knowledge of geometric relationships (i.e., parallel, perpendicular, coincident) of those elements. Fig. 13 shows an example of the registration results.

In order to measure the accuracy of the point clouds, the Euclidian distance (error) from the point to the surface of the ground truth model where the point was supposed to be located was considered the metric to measure the accuracy. We denoted the ith point's coordinate as (X_i^j, Y_i^j, Z_i^j); it was supposed to lie on the jth surface of the ground truth bridge model, as $a_jX + b_jY + c_jZ + d_j = 0$. The average error of the point cloud could accordingly be calculated as follows:

$$err = \frac{1}{\sum_{j=1}^{n} m_j} \sum_{j=1}^{n} \sum_{i=1}^{m_j} \frac{\left| a_j X_i^j + b_j Y_i^j + c_j Z_i^j + d_j \right|}{\sqrt{a_j^2 + b_j^2 + c_j^2}} \tag{8}$$

In this equation, m_j is the number of points supposed to belong to the jth surface, and n is the number of surfaces.

For each surface, 4 boundary offset planes are defined and points fall within this space are supposed to belong to this surface. If a point's distance to the surface is far beyond the average value, it will be deemed as an outlier and removed from the testing data set. More information about the procedure of measuring accuracy can be found at [3].

In order to measure the completeness of the generated point clouds, different surfaces of the bridge model were divided into 2.5 cm × 2.5 cm squares. For each square, if a corresponding 3D point existed, that area was considered a successfully reconstructed region. By calculating the percentage of successfully reconstructed areas, the completeness of the generated point cloud

Fig. 12. (a) Using a total station to collect ground truth data and (b) the actual surface model of the bridge built on ground truth data.

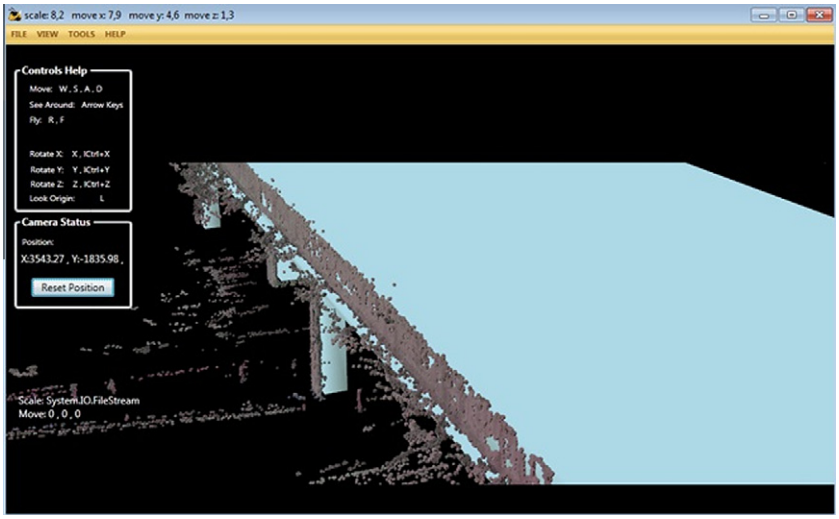


Fig. 13. Registration of the generated dense point cloud into the ground truth model.

Table 5
Average errors and completeness rates for the generated point clouds.

Method	Proposed method	Uniform distributed intervals	Pollefeys et al. [25]	Thormahlen et al. [42]	Seo et al. [40]
Number of video clips	12	12	12	12	12
Number of failure in 3D reconstruction	1	5	1	2	4
Average number of frames per sequence	4500	4500	4500	4500	4500
Average number of extracted key frames per sequence	481	481	682	591	786
Average error (cm)	6.28	7.79	6.34	6.12	7.42
Variance of Euclidian distance error (cm)	2.87	2.96	2.76	2.49	3.68
Average completeness rate (%)	77.43	72.08	76.59	77.76	78.02

was accordingly measured. The obtained results of accuracy and completeness for different point clouds, as well as the number of failure cases are summarized in Table 5.

Based on the results in Table 5, we observe that for the videogrammetric pipeline, using the proposed key frame selection algorithm significantly reduces failure risks in generating point clouds. For this specific scene, the number of failure cases was reduced from 5 to 1. We also observe that by applying the proposed key frame selection algorithm, the average accuracy and completeness of the generated point clouds were increased by 24% and 7.5% respectively. These results demonstrate the efficiently of the proposed method.

7. Conclusion

Video clips captured from civil infrastructure sites contain a large volume of blurry, noisy, or redundant frames, which can significantly affect the performance of the videogrammetric pipeline. This problem is associated with several factors, e.g., lens distortions, motion blur, and high speed rates of frame capturing. As a result, filtering redundant, low quality frames and selecting an optimized number of informative high quality frames is a challenging task. This paper presented a novel method for extracting high-quality, informative frames from a video stream. The resulting key frames could be fed into the videogrammetric pipeline to effectively generate dense point clouds of civil infrastructure. The proposed algorithm automated the processes of removing blurry frames and selecting a number of frames in a way by which computational efficiency was achieved and common degeneracy cases were minimized. The experimental results revealed that the proposed key frame selection algorithm eclipsed the existing methods at a relatively high successful rate of the 3D reconstruction, while maintaining the best reprojection accuracy. In addition, applying the proposed key frame selection algorithm significantly reduced the risk of failure in the 3D reconstruction pipeline. In addition, this study validated the performance of the proposed key frame selection algorithm within the entire videogrammetric pipeline in terms of the accuracy and completeness of the resulting 3D point clouds.

Selection of high-quality and informative frames from a long video sequence is mandatory in order to achieve the satisfactory performance of the 3D videogrammetric reconstruction. By applying the proposed key frame selection algorithm within the videogrammetric pipeline, we achieved up to 6.28 cm accuracy as well as 77.4% completeness. This level of accuracy and density have the potential for a number of practical applications in the AEC domain, including rapid/comprehensive emergency building assessment, remote visual inspection, as-built documentation, safety and productivity analysis and progress monitoring.

In order to improve the accuracy and completeness of generated point clouds, two approaches should be taken into account:

- Hardware improvement, e.g., using cameras with higher resolutions and high quality lenses.
- Software improvement, e.g., utilizing improved algorithms for extracting feature points, effective matching, camera motion estimation and optimizations.

As the next step of this ongoing research project, we plan to investigate the impact of different practical settings, i.e., various

types of cameras and lens, resolution configurations and data collection distances on the performance of the videogrammetric pipeline. In addition, a natural extension of this research is the comparison of the proposed videogrammetric method with the other existing vision-based reconstruction algorithms. The factors that influence the accuracy of applying videogrammetry will also be meticulously investigated in order to lend this technology effectively to broader applications in civil infrastructure scenes. Another extension that will take place is the utilization of the recognition techniques and contextual information for the detection of the detailed information of the scene. Recognition of important structural object has utility in terms of removing cluttering scene structure and irrelevant points from the reconstructed 3D point cloud, thereby more effectively modeling the infrastructure object of interest. Furthermore, from the observation of the obtained point clouds (e.g., Fig. 10), the completeness of the 3D surface of an object needs further improvement (e.g., fixing the gaps on the surfaces of the point cloud of the bridge).

Acknowledgments

The presented research was funded by the U.S. National Science Foundation (NSF) under Grant CMMI-1031329. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF. The authors would also like to thank Craig Burgess, undergraduate student at the School of Civil and Environmental Engineering, Georgia Tech, for his assistance in collecting ground truth data of the bridge.

References

[1] Z. Zhu, I. Brilakis, Comparison of optical-sensor-based spatial data collection techniques for civil infrastructure modeling, Journal of Computing in Civil Engineering 23 (3) (2009) 170–177.
[2] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Pena-Mora, Evaluation of image-based and laser scanning accuracy for emerging automated performance monitoring techniques, Automation in Construction 20 (8) (2011) 1143–1155.
[3] F. Dai, A. Rashidi, I. Brilakis, P. Vela, Comparison of image- and time-of-flight-based technologies for three dimensional reconstruction of infrastructure, ASCE Journal of Construction Engineering and Management 139 (1) (2013) 69–79.
[4] V.R. Kamat, J.C. Martinez, M. Fischer, M. Golparvar-Fard, F. Peña-Mora, S. Savarese, Research in visualization techniques for field construction, Journal of Construction Engineering and Management 137 (10) (2011) 853–862.
[5] F. Dai, M. Lu, Assessing the accuracy of applying photogrammetry to take geometric measurements on building products, Journal of Construction Engineering and Management, American Society of Civil Engineers 136 (2) (2010) 242–250.
[6] C.A. Quiñones-Rozo, Y.M.A. Hashash, L.Y. Liu, Digital image reasoning for tracking excavation activities, Automation in Construction 17 (5) (2008) 608–622.
[7] L. Klein, N. Li, B. Becerik-Gerber, Imaged-based verification of as-built documentation of operational buildings, Automation in Construction 21 (1) (2012) 161–171.
[8] M. Golparvar-Fard, F. Peña-Mora, S. Savarese, Integrated sequential as-built and as-planned representation with D4AR tools in support of decision-making tasks in the AEC/FM industry, Journal of Construction Engineering and Management 137 (12) (2011) 1099–1116.
[9] A. Bhatla, S.Y. Choe, O. Fierro, F. Leite, Evaluation of accuracy of as-built 3D modeling from photos taken by handheld digital cameras, Automation in Construction 28 (2012) 116–127.
[10] S. Han, F. Peña-Mora, M. Golparvar-Fard, S. Roh, Application of a visualization technique for safety management, in: Proc. The ASCE International Workshop on Computing in Civil Engineering, Austin, Texas, USA, 2009.
[11] M. Golparvar-Fard, F. Peña-Mora, S. Savarese, Application of D4AR – A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication, Journal of Information Technology in Construction 14 (Special Issue) (2009) 129–153.
[12] H. Kim, N. Kano, Comparison of construction photograph and VR image in construction progress, Automation in Construction 17 (2) (2008) 137–143.
[13] Z.A. Memon, M.Z.A. Majid, M. Mustaffar, An automatic project progress monitoring model by integrating AutoCAD and digital photos, in: L. Soibelman, F. Peña-Mora (Eds.), Proc. The International Conference on Computing in Civil Engineering, American Society of Civil Engineers (ASCE), Cancun, Mexico, 2005.
[14] M. Golparvar-Fard, F. Peña-Mora, S. Savarese, Automated progress monitoring using unordered daily construction photographs and IFC-based building information models, Journal of Computing in Civil Engineering, in press. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000205
[15] M.R. Jahanshahi, J.S. Kelly, S.F. Masri, G.S. Sukhatme, A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures, Structure and Infrastructure Engineering 5 (6) (2009) 455–486.
[16] M.R. Jahanshahi, S.F. Masri, G.S. Sukhatme, Multi-image stitching and scene reconstruction for evaluating defect evolution in structures, Structural Health Monitoring 10 (6) (2011) 643–657.
[17] M.R. Jahanshahi, S.F. Masri, C.W. Padgett, G.S. Sukhatme, An innovative methodology for detection and quantification of cracks through incorporation of depth perception, Machine Vision and Applications (2011) 1–15.
[18] D.T. Kien, A Review of 3D Reconstruction from Video Sequences, Intelligent Sensory Information Systems, Department of Computer Science, University of Amsterdam, Netherlands, 2005.
[19] J. Greenwood, Large component deformation studies using videogrammetry, in: Proc. The 6th International Workshop on Accelerator Alignment (IWAA 99), Grenoble, France, 1999, pp. 1–33.
[20] R.S. Pappa, J.T. Black, J.R. Blandino, T.W. Jones, P.M. Danehy, A.A. Dorrington, Dot-Projection Photogrammetry and Videogrammetry of Gossamer Space Structures, in: Proc. The 21st International Modal Analysis Conference (IMAC), Kissimmee, FL, USA, 2003.
[21] I. Brilakis, H. Fathi, A. Rashidi, Progressive 3D reconstruction of infrastructure with videogrammetry, Automation in Construction 20 (7) (2011) 884–895.
[22] M.T. Ahmed, M.N. Dailey, J.L. Landabaso, N. Herrero, Robust Key Frame Extraction for 3D Reconstruction from Video Streams, in: Proc. The VISAPP, 2010, pp. 231–236.
[23] R. Hartley, A. Zisserman, Multiple View Geometry, Cambridge University Press, Cambridge, UK, 2004.
[24] D. Nistér, Automatic passive recovery of 3D from images and video, invited paper, in: Proc. The Second International Symposium on 3D Data Processing, Visualization & Transmission (3DPVT04), Thessaloniki, Greece, 2004.
[25] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, Visual modeling with a hand-held camera, International Journal of Computer Vision 59 (3) (2004) 207–232.
[26] M. Pollefeys et al., Detailed real-time urban 3D reconstruction from video, International Journal of Computer Vision 78 (2–3) (2008) 143–167.
[27] H. Fathi, I. Brilakis, Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features, Journal of Advanced Engineering Informatics 25 (4) (2011) 760–770.
[28] Jian-F. Cai, H. Ji, C. Liu, Z. Shen, Blind motion deblurring using multiple images, Journal of Computational Physics 228 (14) (2009) 5057–5071.
[29] S. Cho, S. Lee, Fast motion deblurring, ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH Asia 28 (2009) 5.
[30] Q. Shan, J. Jia, A. Agarwala, High-quality motion deblurring from a single image, ACM Transactions on Graphics 27 (3) (2008). 73:1–73:10.
[31] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, Removing camera shake from a single photograph, ACM Transactions on Graphics 25 (3) (2006) 787–794.
[32] N. Joshi, S.B. Kang, C. Lawrence Zitnick, R. Szeliski, Image deblurring using inertial measurement sensors, ACM Transactions on Graphics 29 (2010) 4.
[33] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, A no-reference perceptual blur metric, in: Proc. The International Conference on Image Processing, Rochester, NY, 2002, pp. 57–60.
[34] Y. Chung, J. Wang, R. Bailey, S. Chen, S. Chang, A non-parametric blur measure based on edge analysis for image processing applications, in: Proc. IEEE Conference on Cybernetics and Intelligent Systems, Singapore, 2004, pp. 356–360.
[35] S. Varadarajan, L.J. Karam, An improved perception-based no-reference objective image sharpness metric using iterative edge refinement, in: Proc. The 15th IEEE International Conference on Image Processing, San Diego, CA, 2008, pp. 401–404.
[36] M.J. Chen, A.C. Bovik, No reference image blur assessment using multiscale gradient, in: Proc. The 1st International Workshop on Quality of Multimedia Experience (QoMEX), 2009, pp. 70–74.
[37] F. Crete, T. Dolmiere, P. Ladret, M. Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric, in: B.E. Rogowitz, T.N. Pappas, S.J. Daly (Eds.), Proc. The SPIE, vol. 6492, pp. 64920I.
[38] J. Seo, S. Kim, C. Jho, H. Hong, 3D estimation and keyframe selection for match move, in: Proc. The ITCCSCC, 2003.
[39] Y.H. Seo, S.H. Kim, K.S. Doo, J.S. Choi, Optimal keyframe selection algorithm for three-dimensional reconstruction in uncalibrated multiple images, Journal of the Society of Photo-Optical Instrumentation Engineers 47 (5) (2008) 53201–53400.
[40] P.H.S. Torr, A.W. Fitzgibbon, A. Zisserman, Maintaining multiple motion model hypotheses over many views to recover matching and structure, in: Proc. The 6th International Conference on Computer Vision, Bombay, India, 1998, pp. 485–491.
[41] S.Gibson, J. Cook, T. Howard, R. Hubbold, D. Oram. Accurate camera calibration for off-line, video-based augmented reality, in: IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002), Darmstadt, Germany, 2002.

[42] T. Thormahlen, H. Broszio, A. Weissenfeld, Keyframe selection for camera motion and structure estimation from multiple views Thorsten, in: Proc. The 8th European Conference on Computer Vision, 2004, pp. 523–535.
[43] A. Rashidi, H. Fathi, I. Brilakis, Innovative stereo vision-based approach to generate dense depth map of transportation infrastructure. Transportation research record, Journal of the Transportation Research Board 2215 (2011) 93–99.
[44] N. Snavely, S.M. Seitz, R. Szeliski, Modeling the world from Internet photo collections, International Journal of Computer Vision 80 (2) (2008) 189–210.
[45] Z. Zhang, Flexible camera calibration by viewing a plane from unknown orientations, in: Proc. The 7th IEEE International Conference on Computer Vision, Kerkrya, Greece, 1999, pp. 1–8.
[46] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[47] A. Rashidi, F. Dai, I. Brilakis, P. Vela, Comparison of camera motion estimation methods for 3D reconstruction of infrastructure, in: Proc. The ASCE International Workshop on Computing in Civil Engineering, Miami, FL, USA, 2011.
[48] D. Nistér, An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 26 (6) (2004) 756–770.
[49] M.A. Lourakis, A.A. Argyros, SBA: a software package for generic sparse bundle adjustment, ACM Transactions on Mathematical Software (TOMS) 36 (1) (2009).
[50] Y. Furukawa, J. Ponce, Accurate, dense, and robust multi-view stereopsis, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (8) (2010) 1362–1376.
[51] R.H. Sheikh, C.A. Bovik, G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Transactions on Image Processing 14 (12) (2005) 2117–2128.
[52] M.H. Hansen, W.N. Hurwitz, W.G. Madow, Simple Random Sampling. Sample Survey Methods and Theory, Methods and Applications, vol. I, John Wiley & Sons, New York, 1953. pp. 126–129 (Chapter 4).
[53] M. Kazhdan, M. Bolitho, H. Hoppe, Poisson, surface reconstruction. in: Proc. The Forth Symposium on Geometry Processing, Sardinia, Italy, 2006, pp. 61–70.