Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Review:

Applications of structure from motion: a survey*

Ying-mei WEI[1], Lai KANG[1], Bing YANG[1], Ling-da WU[1,2]

([1]*College of Information System and Management, National University of Defense Technology, Changsha 410073, China*)
([2]*National Laboratory of Electronic Information Equipment System, Academy of Equipment Command and Technology, Beijing 101416, China*)

E-mail: weiyingmei126@126.com; lkang.vr@gmail.com; mmlab@126.com; wld@nudt.edu.cn

Received Dec. 29, 2012; Revision accepted May 10, 2013; Crosschecked June 6, 2013

Abstract: Structure from motion (SfM) has been an active research area in computer vision for decades and numerous practical applications are benefiting from this research. While no previous work has tried to summarize the applications appearing in the literature, this paper deals with a comprehensive overview of recent applications of SfM by classifying them into 10 categories, namely augmented reality, autonomous navigation/guidance, motion capture, hand-eye calibration, image/video processing, image-based 3D modeling, remote sensing, image organization/browsing, segmentation and recognition, and military applications. The goal is to provide insights for researchers to position their work more appropriately in the context of existing techniques, and to perceive both new applications and relevant research problems.

Key words: Structure from motion (SfM), Image-based 3D modeling, Application
doi:10.1631/jzus.CIDE1302 **Document code:** A **CLC number:** TP391

1 Introduction

The creation of a 3D model of a real scene from 2D images is a fundamental task in computer vision, which is commonly referred to as image-based 3D modeling. Among the techniques developed for image-based modeling, structure from motion (SfM) (Hartley and Zisserman, 2004; Szeliski, 2010) is quintessential due to its simplicity in concept and wide applicability.

While the SfM technique itself has been an active area of research for decades, it has recently experienced a renaissance due to a few significant breakthroughs and more and more practical applications that have arisen. There is a large literature reporting novel algorithms and interesting applications of SfM in academic journals and conferences every year (Szeliski, 2010). In this paper we give a comprehensive overview of recent applications of SfM by surveying related applications and classifying them into appropriate categories. While there is still ongoing interest in this topic, the recent applications and trends have not been reviewed in previous work. Hence, the authors feel that a survey would be useful for researchers to position their work more appropriately in the context of existing techniques, and to perceive both new applications and relevant research problems.

2 Background of structure from motion

Image-based 3D modeling, the reverse process of image formation, poses great challenges in computer vision. Given a projected position in the 2D image plane, the corresponding 3D scene point cannot be determined uniquely due to the existence of ambiguity in its depth, which is lost in image formation. Thus, additional information is required to obtain the estimate of 3D geometry from images. In computer vision, SfM refers to the process of simultaneously estimating the 3D geometry of a scene

* Project (No. 61070140) supported by the National Natural Science Foundation of China
©Zhejiang University and Springer-Verlag Berlin Heidelberg 2013

(structure) and the poses of cameras (motion). SfM exploits corresponding image points in two or more views to reduce the number of degrees of freedom.

In the literature, effective SfM, including two- and multi-frame frameworks, appeared as early as the 1980s. In particular, a relative orientation estimation technique was introduced by Longuet-Higgins (1981). The development of SfM techniques for multiple frames occurred later on. These multi-frame techniques include both factorization methods (Tomasi, 1992) and global optimization methods (Spetsakis and Aloimonos, 1991; Szeliski and Kang, 1994; Oliensis, 1999). Recently, bundle adjustment (Triggs *et al.*, 2000) from photogrammetry has made its way into computer vision for estimating optimal 3D geometry and camera parameters (Hartley and Zisserman, 2004).

The computation of SfM is based solely on image correspondences (Lowe, 2004). Thus, SfM is conceptually simple. Also, since it is a bottom-up approach that makes few assumptions about the input data, it is quite general. Thus, SfM has been one of the most popular image-based 3D modeling algorithms. A typical pipeline of SfM performs incrementally in several passes, including detecting image features (Lowe, 2004; Tuytelaars and Mikolajczyk, 2007), establishing image correspondences (Mikolajczyk and Schmid, 2005; Muja and Lowe, 2009), estimating camera poses and locations of 3D points (Hartley and Zisserman, 2004), and optional bundle adjustment (Triggs *et al.*, 2000; Lourakis and Argyros, 2009). It is noteworthy that modern SfM frameworks can handle large scale data sets (Agarwal *et al.*, 2009; 2010; Snavely *et al.*, 2010; Wu *et al.*, 2011). The interested reader is referred to Hartley and Zisserman (2004) and Szeliski (2010) for a more complete review of relevant algorithms.

Specifically, bundle adjustment is an effective non-linear optimization procedure often used to simultaneously refine the camera parameters and scene structure in the SfM pipeline as a final stage. Denote by $\boldsymbol{P} = \boldsymbol{K}[\boldsymbol{R}|\boldsymbol{T}]$ the camera projection matrix, where $\boldsymbol{K}$ is the intrinsic camera calibration matrix, $\boldsymbol{R}$ the rotation matrix, and $\boldsymbol{T}$ the translation vector of a camera (Hartley and Zisserman, 2004). Let N be the number of cameras and M the number of 3D points in the SfM problem. Bundle adjustment optimization is aimed at minimizing the following cost function (Lourakis and Argyros, 2009):

$$\sum_{i=1}^{N}\sum_{j=1}^{M} v_{ij}\left\|\boldsymbol{x}_{ij} - \boldsymbol{P}_i\boldsymbol{X}_j\right\|_2^2, \tag{1}$$

where v_{ij} is a binary variable indicating the visibility of the jth 3D point $\boldsymbol{X}_j$ in the ith image, and $\boldsymbol{x}_{ij}$ the projection of 3D point $\boldsymbol{X}_j$ onto the ith image.

3 Applications of structure from motion

A collection of papers on the applications of SfM was surveyed and classified according to the scenarios to which SfM techniques are applied. Existing applications of SfM are categorized into 10 categories, namely augmented reality, autonomous navigation/guidance, motion capture, hand-eye calibration, image/video processing, image-based 3D modeling, remote sensing, image organization/browsing, segmentation and recognition, and military applications. In the following, we survey these 10 categories of applications in detail mainly by illustrating the specific role of SfM in the context of each category of applications.

3.1 Image-based 3D modeling

Three-dimensional digital models are widely used in applications such as navigation, visualization, and animation. While creating a simple 3D model seems to be easy using CAD tools (e.g., 3DMax and Maya), obtaining an accurate and realistic 3D model of a complex real scene or object remains difficult. Existing 3D reconstruction methods can be roughly divided into contact methods and non-contact methods (Remondino and El-Hakim, 2006). Among the non-contact 3D acquisition techniques there is an image-based 3D modeling technique, which aims to reconstruct a 3D model from images. Image-based 3D modeling based on SfM is of particular interest to researchers in the computer vision area due to the wide applicability and low cost of SfM.

Since SfM is able to recover the 3D geometry of real scenes, this leads to an ideal application in image-based 3D modeling. Pollefeys *et al.* (2001a; 2001b) applied SfM to acquire 3D archaeological heritage based on images. The proposed approach is able to obtain both the 3D shape and the appearance of real scenes such as objects, monuments, and landmarks from images or videos. The whole framework can be divided into the following steps. In the first step, sparse feature correspondences among consecutive images are obtained by feature extraction and matching. In the second step, the SfM algorithm is applied to obtain a sparse 3D scene reconstruction and the motion of the camera. In the third step, both structure and motion are refined by bundle adjustment. In the final step, a standard stereo algorithm is applied on rectified images to estimate the disparity maps, from which a dense surface reconstruction can be extracted. This method can obtain photo-realistic models by making use of texture mapping. To promote cultural heritage sites, Manferdini (2012) developed a methodology to recover 3D information from a sequence of 2D images acquired using uncalibrated cameras. Manferdini (2012) also defined different levels of details of information for different communication purposes, such as efficient visualization on mobile devices and PC, and made a detailed investigation of digital models.

Another popular scenario in which SfM is commonly used is reconstruction in urban environments (Schindler *et al.*, 2006; Sinha *et al.*, 2008; Xiao *et al.*, 2008; Musialski *et al.*, 2012; Pylvanainen *et al.*, 2012). By employing the urban scene structure in a line-based SfM technique, Schindler *et al.* (2006) demonstrated that this knowledge can be used to improve the performance of feature extraction, feature matching, and optimization in the entire process. Sinha *et al.* (2008) presented a system for creating realistic 3D models of architectural and urban scenes from a collection of unordered images. By combining user interaction with geometric information obtained from SfM analysis, the 3D structure is computed from the input images in an automatic fashion. Similarly, Xiao *et al.* (2008) presented a semi-automatic image-based method to facade modeling relying on SfM to retrieve camera locations and point clouds as initialization. In the case where LIDAR data is available, SfM has also been used successfully to address the limited range of the LIDAR system (Pylvanainen *et al.*, 2012). As reviewed in Musialski *et al.* (2012), SfM has been widely employed in image-based urban reconstruction methods.

For large scale data sets, the applicability of SfM in practical use is limited by its efficiency. Recent works have addressed this issue. Frahm *et al.* (2010) developed a powerful approach for dense 3D reconstruction from a huge number of unregistered urban/architecture photos within 24 hours on a single PC. Efficient incremental SfM is combined with bundle adjustment to register images, based on which dense geometry estimation can be performed. Wang and Olano (2011) proposed to accelerate the computational speed of SfM in large scale 3D reconstruction problems by leveraging high performance of modern graphics processing units (GPU).

3.2 Hand-eye calibration

In many robotic applications, there is a need to determine the measurements obtained by a digital camera attached to a robotic gripper to its coordinate system, which corresponds to a homogeneous transformation between the gripper and the camera's coordinate frame. This problem is usually called hand-eye calibration (Andreff *et al.*, 2001; Schmidt *et al.*, 2005; Heller *et al.*, 2011). An imaging system needs hand-eye calibration (Fig. 1).

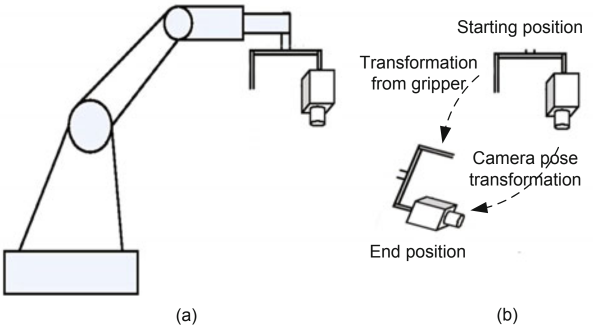

Fig. 1 Hand-eye calibration system (a) and a relative movement of camera and gripper (b)

As reviewed by Wang (1992), most early methods for hand-eye calibration estimate the translation and rotation separately. Conventional hand-eye calibration methods usually estimate the camera poses by viewing a calibration object with known dimensions, which may not be always practical in many situations. The restriction of the use of a calibration object has been recently removed owing to the development of hand-eye calibration methods based on SfM.

Andreff *et al.* (2001) proposed several ways to simplify the procedure of hand-eye calibration. The resulting approach does not need any calibration object and can be applied to small calibration motions.

Unlike traditional methods using calibration objects, in Andreff *et al.* (2001) camera motions were computed, up to an unknown scale factor, through SfM algorithms rather than commonly used pose estimation algorithms. The unknown scale factor which characterizes rotations with orthogonal matrices allowing both large and small motions is then included in a linear system. The limitation of Andreff *et al.* (2001) is that, as pointed out by Schmidt *et al.* (2005), the orthogonality of the rotation matrix $\boldsymbol{R}_{\mathrm{HE}}$ needs to be imposed using singular value decomposition (SVD) (Hartley and Zisserman, 2004), since the extended equations possess no such guarantee. The further limitation of Andreff *et al.* (2001) is that, as pointed out by Schmidt *et al.* (2005), the orthogonality of $\boldsymbol{R}_{\mathrm{HE}}$ is not guaranteed by the extended equations, but has to be enforced afterwards using SVD (Hartley and Zisserman, 2004). This limitation was addressed in Schmidt *et al.* (2005), who presented different SfM extensions to handle hand-eye calibration involving scale factor estimation in addition to rotation and translation.

A more recent application of SfM to hand-eye calibration is based on SfM under the L_∞ norm. One such method was proposed by Heller *et al.* (2011), where the correct scale can be recovered by second-order cone programming (SOCP). Given image correspondences and robotic measurements, the globally optimal estimate of hand-eye transformation with respect to the L_∞ norm can be obtained.

3.3 Augmented reality

The augmented reality (AR) system aims to enhance the real-world environment by integrating virtual objects. To place the virtual objects with correct poses at the proper locations of a real scene. Two critical issues, including camera tracking and depth perception, have to be addressed. In early works, a pre-calibrated camera was often required to observe calibration objects (or markers) placed in real-world environments, which becomes incontinent for many situations (e.g., outdoor environment). These limitations can be overcome by integrating SfM into AR systems.

Much research on AR applications described in the literature is also very closely related to SfM. Cornelis *et al.* (2001) demonstrated an elaborated SfM framework that can recover accurate camera motions from a video sequence based on image feature tracking. The above knowledge allows one to create AR video products without noticeable jitter or drift or virtual objects integrated into the video sequence. Later on, researchers developed many real-time SfM algorithms to augment reality systems, involving such setups as head-mounted camera and display (Pupilli and Calway, 2002; Yao and Calway, 2002; Streckel *et al.*, 2005; Schweighofer *et al.*, 2008).

Mooser *et al.* (2009) demonstrated a complete system for markerless AR using robust SfM, which consists mainly of two components. The first component aims to learn the structure of complex real scenes and augment them with synthetic annotations (an example is shown in Fig. 2). The output of the above component is a database of known landmarks with the 3D descriptions of synthetic objects. In the second component, the database is used to recognize known landmarks, to recover the path of the camera, and to visualize the associated synthetic objects. Recent optimization algorithms for SfM have been used in both components. More recently, Yang *et al.* (2013) achieved an AR system involving many recently developed techniques, such as SfM, view clustering, multi-view stereo, and surface reconstruction from point clouds. The system is verified on both indoor and outdoor objects at various scales, such as a helmet (a small object), a corridor (an indoor medium object), an arbor (outdoor medium object), and a building (outdoor large object). By tracking and registration using SfM, the reconstructed 3D models can be loaded in an AR environment to facilitate displaying, interacting, and rendering.

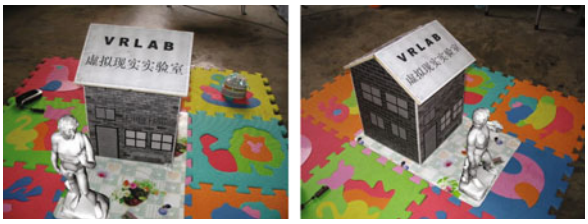

Fig. 2 Creating virtual annotations: the virtual label has been placed properly given a sparse representation of the scene

3.4 Autonomous navigation/guidance

Another major category of applications is 'autonomous navigation/guidance', for which location recognition and object positioning are two fundamental tasks. One popular localization device for robotic vehicles is the Global Positioning System (GPS), which can achieve an accuracy of 1 cm in localization (Royer *et al.*, 2007). However, the GPS signal may not be strong enough or not be available at all due to occlusions (e.g., indoors), causing a dramatic decrease in localization accuracy. Therefore, vision-based autonomous navigation techniques are attractive to researchers.

Given sparse point cloud output by the SfM method, recent studies have developed a number of methods for fast location recognition (Irschara *et al.*, 2009; Moslah *et al.*, 2009; Li *et al.*, 2010; 2012). It is noteworthy that the most recent work by Li *et al.* (2012) in this direction can scale to datasets with hundreds of thousands of images and tens of millions of 3D points through the use of two new techniques: a co-occurrence prior for RANSAC and bidirectional matching of image features with 3D points.

Recently, SfM techniques have also been exploited for object positioning (Zelek *et al.*, 2010; Manweiler *et al.*, 2012). Although object positioning can be fulfilled using GPS, the GPS is known to be sensitive to orientation and satellite positions and it may even drop out in urban canyons and indoor environments (Zelek *et al.*, 2010). Thus, the solutions proposed in Zelek *et al.* (2010) and Manweiler *et al.* (2012) are convenient and reliable alternatives for object positioning.

3.5 Motion capture

The most straightforward approach to capturing human motion is the use of wearable motion sensors. Although this kind of approach is able to measure the motion directly, it is intrusive and not comfortable to wear. On the other hand, for most existing vision-based motion capture techniques, recording in a closed stage with controlled imaging conditions is generally required, which restricts their use in motion capture for outdoor setting as well as traversal of large areas (Shiratori *et al.*, 2011).

The applications of SfM to motion capture have appeared in the last few years (Hasler *et al.*, 2009; Shiratori *et al.*, 2011). Hasler *et al.* (2009) presented an approach for markerless motion capture (MoCap) by recording the articulated objects using several unsynchronized moving cameras. In this system, the reconstruction of static background and camera registration are performed using the SfM method, based on which both the positions and the joint configurations of subjects can be recovered.

To achieve motion capture using body-mounted cameras based on SfM, Shiratori *et al.* (2011) presented both theoretical analysis and a practical motion capture system. In this system, SfM is used to obtain an initial guess of the pose of joint angles and root from video captured by cameras mounted on the subject. The estimate was further refined using a non-linear optimization technique. Shiratori *et al.* (2011) demonstrated that the proposed system performs well even in challenging settings where motion capture is impossible or difficult with traditional motion capture systems.

3.6 Image/video processing

There are also some interesting and useful applications to image and video processing, such as video enhancement (Bhat *et al.*, 2007), curved document rectification (Sato *et al.*, 2006), and video stabilization (Srinivasan and Chellappa, 1999; Kurz *et al.*, 2009; Liu *et al.*, 2009; Zhang *et al.*, 2009).

Bhat *et al.* (2007) presented an automatic framework for enhancing a video sequence using a few images of the same static scene. By transferring image qualities to video, the proposed system allows one to improve the quality of the input video in various ways, e.g., enhancing the resolution of video and enhancing lighting conditions of video. Moreover, unwanted objects and camera shake can be quickly removed from the input video as required by editing a few 2D images. As a key component of this system, SfM provides camera projection matrices for each photograph and video frame, a sparse cloud of 3D scene points, and a list of the viewpoints from which each scene point is visible, which can be used to improve the accuracy of depth-map estimation.

Sato *et al.* (2006) presented a novel method based on 3D reconstruction using SfM for video mosaicking for curved documents. The proposed method is able to restore the target document with a curved surface in the mosaic image. For applications to video stabilization, SfM can be used to reconstruct camera poses and a sparse 3D model of the observed scene, based on which a desired camera path can be computed either automatically or interactively and the resulting warping errors can be determined and minimized (Liu *et al.*, 2009; Zhang *et al.*, 2009).

3.7 Remote sensing

In the area of remote sensing and aerial photogrametry, SfM has also been shown to be able to provide effective solutions. Relevant applications including aerial image mosaicking (Strucl and Quartisch, 2001; Helala *et al.*, 2012; Turner *et al.*, 2012), 3D mapping (Nicosevici and Garcia, 2008; Nilosek and Walli, 2009), earthwork planning (Nassar *et al.*, 2011), landslide investigation (Niethammer *et al.*, 2011), etc.

The unmanned aerial vehicle (UAV) aerial image mosaicking is usually affected by a weak homography due to the use of unleveled ground control points (GCPs) in image registration. Mosaic methods using SfM techniques can address the above issue. Strucl and Quartisch (2001) enhanced initial image transformations derived from inaccurate position and orientation data of UAVs by the camera pose obtained using SfM. Corresponding points are then selected on a common ground plane to find accurate image transformations, which preserves distances and minimizes distortions. To construct a mosaic of nearby objects from a video taken with a moving camera, Helala *et al.* (2012) applied an SfM algorithm to extract the camera parameters, together with the depths of some feature points. Instead of stitching them together, the authors interpolated the video frames to estimate a mosaic using a plane sweep algorithm from a virtual camera, which is conveniently placed to maximize the visual information that each real view contributes to the mosaic. More recently, Turner *et al.* (2012) also presented an SfM-based approach for geometric correction and mosaicking of images captured with UAV.

Nicosevici and Garcia (2008) presented a method to create accurate 3D terrain with texture from video sequences using sequential SfM, in which a 3D structure of scene is maintained and updated incrementally when new visual information becomes available. Nilosek and Walli (2009) developed a system automatically producing synthetic terrain and architecture with calibrated camera remote sensing.

3.8 Photo organization/browsing

Image organization or browsing is another vast area for applications of SfM. Perhaps the first attempt to organize photo collection based on SfM was made by Schaffalitzky and Zisserman (2002). The task of image organization is essentially the problem of estimating the viewpoints of cameras from a large number of unordered images, which can obviously be resolved by SfM. Generally speaking, feature matching in SfM for an image collection with a small number of images can be fulfilled by pairwise image matching. However, the matching problem quickly becomes prohibitively expensive as the number of views increases if the above naive matching strategy is employed. Thus, Schaffalitzky and Zisserman (2002) developed a matching algorithm which is linear in the number of views, and which can significantly reduce the computational complexity of SfM.

The other well-known photo organization system has been proposed by Snavely *et al.* (2006), whose backbone is a robust SfM method for reconstructing the 3D structure of a scene. The proposed method first extracts image features and establishes feature correspondences, and then runs a non-linear optimization to reconstruct the camera motion and 3D locations corresponding to those features within the framework of SfM. The output correspondences and 3D estimates lead to some interesting and useful features of the proposed photo explorer, including 3D navigation and exploration of unordered image collection, automatic image annotation, construction of photo tours, etc.

3.9 Segmentation and recognition

Numerous solutions have been proposed for object segmentation and recognition from 2D still images without the use of SfM, and exciting results have been achieved. For instance, image segmentation algorithms are based on thresholding, pixel clustering, region-growing, graph partitioning, etc. While most existing methods do not exploit the depth information of scene, the performances of both segmentation and recognition have been further improved with the use of SfM (Brostow *et al.*, 2008).

Recent studies have also integrated SfM into segmentation and recognition for more sophisticated algorithms (Brostow *et al.*, 2008; Sturgess *et al.*, 2009; Bao *et al.*, 2012). Brostow *et al.* (2008) proposed a semantic segmentation algorithm based on 3D point clouds obtained using an SfM method. While modeling the spatial layout and context, the authors combined features in the image projected from 3D cues. Using motion and sparse 3D structure, Brostow *et al.* (2008) showed that more accurate segmentation and recognition can be achieved.

Unlike the above method based solely on the output of SfM, Sturgess *et al.* (2009) investigated a method combining both the appearance and SfM output for pixel-wise object segmentation of road scenes. The proposed method was evaluated on the challenging Cambridge-driving labeled video data set (CamVid) (Brostow *et al.*, 2009) both quantitatively and qualitatively. The experimental results showed an overall recognition accuracy of 84% compared to the state-of-the-art accuracy of 69% (Brostow *et al.*, 2008).

Recently, the concept of traditional SfM was generalized by Bao *et al.* (2012), such that high level semantic components (e.g., 3D objects and regions) are also recognized in addition to the recovering of a set of sparse 3D points. The authors demonstrated that the generalized SfM framework can achieve more robust estimates of camera motion compared with traditional SfM algorithms using only points. Moreover, compared with state-of-the-art recognition algorithms based on a single image, the new method is able to obtain more accurate object and region recognition results.

3.10 Military applications

In addition to the aforementioned civilian applications, SfM has also been exploited in military applications. Modern military operations require the military to gather information quickly from a wide variety of sensors and process it into useful and dependable data either automatically or in the man-in-the-loop fashion, which increases the survivability of combatants and fulfills missions successfully in the battlefields. The process of forming useful data from the raw information is called situational awareness (SA) (Shim *et al.*, 2008). The use of SfM in SA appeared in recent research.

Shim *et al.* (2008) exploited the SfM technique with vision sensors mounted on a moving robotic vehicle that computes 3D geometry from observed 2D features over several frames or views and provides visual cues to an SA system for further tracking and recognition of moving objects. The proposed SfM framework is capable of providing robust perception functions, such as passive ranging for autonomous mobility, mid-range sensing for tactical behavior, and moving target indication (MTI) and appearance based automatic target recognition (ATR) for SA.

As argued in Jackson (2008), there exists a need to create a methodology to achieve near autonomous spatial and temporal tracking of entities, providing position and orientation (pose) data for use in tactical displays. Aiming at providing an increased level of SA for soldiers on and off the battlefield, and using a combination of images and spatial information, Jackson (2008) concentrated on the development of a precision force tracking framework by employing recursive SfM based algorithms to perform image reconstruction.

4 Conclusions

This paper deals with a comprehensive overview of recent applications of structure from motion (SfM) by surveying related applications and classifying them into 10 categories. This survey may be useful for researchers to position their work in the context of existing techniques, and to perceive both new applications and relevant research problems. Despite numerous applications of SfM, a large number of potential applications remain unexplored as the SfM techniques develop; important open issues such as robustness, accuracy, and efficiency remain for future work.

References

Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R., 2009. Building Rome in a Day. Proc. IEEE Int. Conf. on Computer Vision, p.72-79. [doi:10.1109/ICCV.2009.5459148]

Agarwal, S., Snavely, N., Seitz, S.M., Szeliski, R., 2010. Bundle Adjustment in the Large. Proc. European Conf. on Computer Vision: Part II, p.29-42. [doi:10.1007/978-3-642-15552-9_3]

Andreff, N., Horaud, R., Espiau, B., 2001. Robot hand-eye calibration using structure-from-motion. *Int. J. Robot. Res.*, **20**(3):228-248. [doi:10.1177/02783640122067372]

Bao, S.Y., Bagra, M., Chao, Y.W., Savarese, S., 2012. Semantic Structure from Motion with Points, Regions, and Objects. CVPR, p.2703-2710. [doi:10.1109/CVPR.2012.6247992]

Bhat, P., Zitnick, C.L., Snavely, N., Agarwala, A., Agrawala, M., Cohen, M., Curless, B., Kang, S.B., 2007. Using Photographs to Enhance Videos of a Static Scene. Eurographics Symp. on Rendering, p.327-338. [doi:10.2312/EGWR/EGSR07/327-338]

Brostow, G.J., Shotton, J., Fauqueur, J., Cipolla, R., 2008. Segmentation and Recognition Using Structure from Motion Point Clouds. Proc. 10th European Conf. on Computer Vision: Part I, p.44-57. [doi:10.1007/978-3-540-88682-2_5]

Brostow, G.J., Fauqueur, J., Cipolla, R., 2009. Semantic object classes in video: a high definition ground truth database. *Pattern Recogn. Lett.*, **30**(2):88-97. [doi:10.1016/j.patrec.2008.04.005]

Cornelis, K., Pollefeys, M., Gool, L.V., 2001. Tracking Based Structure and Motion Recovery for Augmented Video Productions. Proc. ACM Symp. on Virtual Reality Software and Technology, p.17-24. [doi:10.1145/505008.505012]

Frahm, J.M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.H., Dunn, E., Clipp, B., Lazebnik, S., *et al.*, 2010. Building Rome on a Cloudless Day. Proc. 11th European Conf. on Computer Vision: Part IV, p.368-381. [doi:10.1007/978-3-642-15561-1_27]

Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK. [doi:10.1017/CBO9780511811685]

Hasler, N., Rosenhahn, B., Thormahlen, T., Wand, M., Gall, J., Seidel, H.P., 2009. Markerless Motion Capture with Unsynchronized Moving Cameras. CVPR, p.224-231. [doi:10.1109/CVPR.2009.5206859]

Helala, M.A., Zarrabeitia, L.A., Qureshi, F.Z., 2012. Mosaic of Near Ground UAV Videos under Parallax Effects. Proc. 6th ACM/IEEE Int. Conf. on Distributed Smart Cameras, p.1-6.

Heller, J., Havlena, M., Sugimoto, A., Pajdla, T., 2011. Structure-from-Motion Based Hand-Eye Calibration Using l_∞ Minimization. CVPR, p.3497-3503. [doi:10.1109/CVPR.2011.5995629]

Irschara, A., Zach, C., Frahm, J.M., Bischof, H., 2009. From Structure-from-Motion Point Clouds to Fast Location Recognition. CVPR, p.2599-2606. [doi:10.1109/CVPR.2009.5206587]

Jackson, N.L., 2008. Precision Reconstruction Based Tracking for Autonomous Synthetic Battlefield Displays Acquired from Unmanned Aerial Vehicle Video Streams. Dissertation, Morgan State University, Baltimore, Maryland, United States.

Kurz, C., Thormahlen, T., Seidel, H.P., 2009. Scene-Aware Video Stabilization by Visual Fixation. Proc. Conf. for Visual Media Production, p.1-6. [doi:10.1109/CVMP.2009.9]

Li, Y., Snavely, N., Huttenlocher, D., 2010. Location Recognition Using Prioritized Feature Matching. Proc. 11th European Conf. on Computer Vision: Part II, p.791-804. [doi:10.1007/978-3-642-15552-9_57]

Li, Y., Snavely, N., Huttenlocher, D., Fua, P., 2012. Worldwide Pose Estimation Using 3D Point Clouds. Proc. 12th European Conf. on Computer Vision: Part I, p.15-29. [doi:10.1007/978-3-642-33718-5_2]

Liu, F., Gleicher, M., Jin, H., Agarwala, A., 2009. Content-preserving warps for 3D video stabilization. *ACM Trans. Graph.*, **28**(3), Article 44, p.1-9. [doi:10.1145/1531326.1531350]

Longuet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, **293**(5828):133-135. [doi:10.1038/293133a0]

Lourakis, M.I.A., Argyros, A.A., 2009. SBA: a software package for generic sparse bundle adjustment. *ACM Trans. Math. Software*, **36**(1):1-30. [doi:10.1145/1486525.1486527]

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, **60**(2):91-110. [doi:10.1023/B:VISI.0000029664.99615.94]

Manferdini, A.M., 2012. A Methodology for the Promotion of Cultural Heritage Sites Through the Use of Low-Cost Technologies and Procedures. Proc. 17th Int. Conf. on 3D Web Technology, p.180. [doi:10.1145/2338714.2338747]

Manweiler, J., Jain, P., Choudhury, R.R., 2012. Satellites in Our Pockets: an Object Positioning System Using Smartphones. Proc. 10th Int. Conf. on Mobile Systems, Applications, and Services, p.211-224. [doi:10.1145/2307636.2307656]

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**(10):1615-1630. [doi:10.1109/TPAMI.2005.188]

Mooser, J., You, S., Neumann, U., Wang, Q., 2009. Applying Robust Structure from Motion to Markerless Augmented Reality. Workshop on Applications of Computer Vision, p.1-8. [doi:10.1109/WACV.2009.5403038]

Moslah, O., Guitteny, V., Couvet, S., 2009. Geo-referencing Uncalibrated Photographs Using Aerial Images and 3D Urban Models. CORESA, p.1-5.

Muja, M., Lowe, D.G., 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. Proc. 4th Int. Conf. on Computer Vision Theory and Applications, p.331-340.

Musialski, P., Wonka, P., Aliaga, D.G., Wimmer, M., van Gool, L., Purgathofer, W., 2012. A Survey of Urban Reconstruction. Eurographics State of the Art Reports, p.1-28. [doi:10.1111/cgf.12077]

Nassar, K., Aly, E.A., Jung, Y., 2011. Structure-from-Motion for Earthwork Planning. Proc. 28th ISARC, p.310-316.

Nicosevici, T., Garcia, R., 2008. Online Robust 3D Mapping Using Structure from Motion Cues. OCEANS, p.1-7. [doi:10.1109/OCEANSKOBE.2008.4531022]

Niethammer, U., Rothmund, S., Schwaderer, U., Zeman, J., Joswig, M., 2011. Open Source Image-Processing Tools for Low-Cost UAV-Based Landslide Investigations. Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, p.1-6.

Nilosek, D., Walli, K., 2009. Aerial Scene Synthesis from Images. SIGGRAPH Posters, Article 65. [doi:10.1145/1599301.1599366]

Oliensis, J., 1999. A multi-frame structure-from-motion algorithm under perspective projection. *Int. J. Comput. Vis.*, **34**(2-3):163-192. [doi:10.1023/A:1008139920864]

Pollefeys, M., Gool, L.V., Vergauwen, M., Cornelis, K., Verbiest, F., Tops, J., 2001a. Image-Based 3D Acquisition of Archaeological Heritage and Applications. Proc. Conf. on Virtual Reality, Archeology, and Cultural Heritage, p.255-262. [doi:10.1145/584993.585033]

Pollefeys, M., Vergauwen, M., Cornelis, K., Verbiest, F., Schouteden, J., Tops, J., Gool, L.V., 2001b. 3D Acquisition of Archaeological Heritage from Images. CIPA Conf., Int. Archive of Photogrammetry and Remote Sensing, p.1-8.

Pupilli, M., Calway, A., 2002. Real-Time Structure from Motion for Augmented Reality. University of Bristol, Bristol, UK.

Pylvanainen, T., Berclaz, J., Korah, T., Hedau, V., Aanjaneya, M., Grzeszczuk, R., 2012. 3D City Modeling from Street-Level Data for Augmented Reality Applications. 2nd Int. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission, p.238-245. [doi:10.1109/3DIMPVT.2012.19]

Remondino, F., El-Hakim, S., 2006. Image-based 3D modelling: a review. *Photogrammetr. Rec.*, **21**(115):269-291. [doi:10.1111/j.1477-9730.2006.00383.x]

Royer, E., Lhuillier, M., Dhome, M., Lavest, J.M., 2007. Monocular vision for mobile robot localization and autonomous navigation. *Int. J. Comput. Vis.*, **74**(3):237-260. [doi:10.1007/s11263-006-0023-y]

Sato, T., Iketani, A., Ikeda, S., Kanbara, M., Nakajima, N., Yokoya, N., 2006. Video Mosaicing for Curved Documents by Structure from Motion. ACM SIGGRAPH Sketches. [doi:10.1145/1179849.1180007]

Schaffalitzky, F., Zisserman, A., 2002. Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?". Proc. 7th European Conf. on Computer Vision: Part I, p.414-431. [doi:10.1007/3-540-47969-4_28]

Schindler, G., Krishnamurthy, P., Dellaert, F., 2006. Line-Based Structure from Motion for Urban Environments. Proc. 3rd Int. Symp. on 3D Data Processing, Visualization, and Transmission, p.846-853. [doi:10.1109/3DPVT.2006.90]

Schmidt, J., Vogt, F., Niemann, H., 2005. Calibration Free HandEye Calibration: a Structure from Motion Approach. Proc. 27th DAGM Conf. on Pattern Recognition, p.67-74. [doi:10.1007/11550518_9]

Schweighofer, G., Segvic, S., Pinz, A., 2008. Online/Realtime Structure and Motion for General Camera Models. IEEE Workshop on Applications of Computer Vision, p.1-6. [doi:10.1109/WACV.2008.4544016]

Shim, M., Yilma, S., Bonner, K., 2008. A robust real-time structure from motion for situational awareness and RSTA. *SPIE*, **6962**:1-11. [doi:10.1117/12.778074]

Shiratori, T., Park, H.S., Sigal, L., Sheikh, Y., Hodgins, J.K., 2011. Motion capture from body-mounted cameras. *ACM Trans. Graph.*, **30**(4), Article 31, p.1-10. [doi:10.1145/2010324.1964926]

Sinha, S.N., Steedly, D., Szeliski, R., Agrawala, M., Pollefeys, M., 2008. Interactive 3D architectural modeling from unordered photo collections. *ACM Trans. Graph.*, **27**(5), Article 159, p.1-10. [doi:10.1145/1409060.1409112]

Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph.*, **25**(3):835-846. [doi:10.1145/1141911.1141964]

Snavely, N., Simon, I., Goesele, M., Szeliski, R., Seitz, S.M., 2010. Scene reconstruction and visualization from community photo collections. *Proc. IEEE*, **98**(8):1370-1390. [doi:10.1109/JPROC.2010.2049330]

Spetsakis, M., Aloimonos, J., 1991. A multi-frame approach to visual motion perception. *Int. J. Comput. Vis.*, **6**(3):245-255. [doi:10.1007/BF00115698]

Srinivasan, S., Chellappa, R., 1999. Fast Structure from Motion Recovery Applied to 3D Image Stabilization. Proc. IEEE Int. Conf. on the Acoustics, Speech, and Signal, p.3357-3360. [doi:10.1109/ICASSP.1999.757561]

Streckel, B., Evers-Senne, J.F., Koch, R., 2005. Lens Model Selection for a Markerless AR Tracking System. Proc. 4th IEEE and ACM Int. Symp. on Mixed and Augmented Reality, p.130-133. [doi:10.1109/ISMAR.2005.38]

Strucl, D.W., Quartisch, M., 2001. A Structure Based Mosaicking Approach for Aerial Images from Low Altitude of Non-planar Scenes. Proc. 16th Computer Vision Winter Workshop, p.51-58.

Sturgess, P., Alahari, K., Ladicky, L., Torr, P.H.S., 2009. Combining Appearance and Structure from Motion Features for Road Scene Understanding. Proc. British Machine Vision Conf., p.1-11. [doi:10.5244/C.23.62]

Szeliski, R., 2010. Computer Vision: Algorithms and Applications. Springer, New York.

Szeliski, R., Kang, S.B., 1994. Recovering 3D shape and motion from image streams using nonlinear least squares. *J. Vis. Commun. Image Represent.*, **5**(1):10-28. [doi:10.1006/jvci.1994.1002]

Tomasi, C., 1992. Shape and motion from image streams under orthography: a factorization method. *Int. J. Comput. Vis.*, **9**(2):137-154. [doi:10.1007/BF00129684]

Triggs, B., Mclauchlan, P., Hartley, R., Fitzgibbon, A., 2000. Bundle adjustment: a modern synthesis. *LNCS*, **1883**:298-375. [doi:10.1007/3-540-44480-7_21]

Turner, D., Lucieer, A., Watson, C., 2012. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, structure from motion (SfM) point clouds. *Remote Sens.*, **4**(12):1392-1410. [doi:10.3390/rs4051392]

Tuytelaars, T., Mikolajczyk, K., 2007. Local invariant feature detectors: a survey. *Found. Trends Comput. Graph. Vis.*, **3**(3):177-280. [doi:10.1561/0600000017]

Wang, C., 1992. Extrinsic calibration of a vision sensor mounted on a robot. *IEEE Trans. Robot. Autom.*, **8**(2):161-175. [doi:10.1109/70.134271]

Wang, Y., Olano, M., 2011. A Framework for GPU 3D Model Reconstruction Using Structure-from-Motion. ACM SIGGRAPH Posters, p.27. [doi:10.1145/2037715.2037748]

Wu, C., Agarwal, S., Curless, B., Seitz, S.M., 2011. Multicore Bundle Adjustment. CVPR, p.3057-3064. [doi:10.1109/CVPR.2011.5995552]

Xiao, J., Fang, T., Tan, P., Zhao, P., Ofek, E., Quan, L., 2008. Image-based facade modeling. *ACM Trans. Graph.*, **27**(5), Article 161, p.1-10. [doi:10.1145/1409060.1409114]

Yang, M.D., Chao, C.F., Huang, K.S., Lu, L.Y., Chen, Y.P., 2013. Image-based 3D scene reconstruction and exploration in augmented reality. *Autom. Construct.*, **33**:48-60. [doi:10.1016/j.autcon.2012.09.017]

Yao, A., Calway, A., 2002. Robust Estimation of 3-D Camera Motion for Uncalibrated Augmented Reality. University of Bristol, Bristol, UK.

Zelek, J.S., Fazl-Ersi, E., Asmar, D.C., Fakih, A.H., 2010. Computer Vision Geo-location, Awareness & Detail. Proc. 1st Int. Conf. and Exhibition on Computing for Geospatial Research & Application. [doi:10.1145/1823854.1823906]

Zhang, G., Hua, W., Qin, X., Shao, Y., Bao, H., 2009. Video stabilization based on a 3D perspective camera model. *Vis. Comput.*, **25**(11):997-1008. [doi:10.1007/s00371-009-0310-z]