Algorithms and Applications of Structure from Motion (SFM): A Survey

Article · November 2017

CITATIONS
4

READS
2,941

3 authors, including:

Abdou Shalaby
Faculty of Computers & Information
3 PUBLICATIONS **8** CITATIONS

SEE PROFILE

Mohammed Elmogy
Mansoura University
258 PUBLICATIONS **2,775** CITATIONS

SEE PROFILE

International Journal of Computer and Information Technology (ISSN: 2279 – 0764)
Volume 06 – Issue 06, November 2017

Algorithms and Applications of Structure from Motion (SFM): A Survey

Abdou Shalaby
Information Systems Department,
Faculty of Computers and Information,
Mansoura University, Egypt

Mohammed Elmogy
Information Technology Department,
Faculty of Computers and Information,
Mansoura University, Egypt
Email: melmogy [AT] mans.edu.eg

Ahmed Abo El-Fetouh
Information Systems Department,
Faculty of Computers and Information,
Mansoura University, Egypt

Abstract - Structure from motion (SFM) is one of the most popular problems that researchers interested in the field of computer vision and computer graphics. Structure from motion is the problem of reconstructing 3D from 2D images. According to the data they use are a set of features matches or an optical flow field, 3D reconstruction Approaches can be classified as feature-based or flow based Techniques. Feature-based reconstruction is carried out using corresponding features in pairs of images of the same scene taken from different viewpoints. In flow-based Matching is replaced by optical flow the features velocity field generated by the camera motion. In this paper, we discuss and analyze the efforts of scientists in this field to cover all aspects of structure from motion. We discuss the techniques; the most critical applications depend on SFM, challenges, and the current research topics.

Keywords - *Structure from motion (SFM), 3D reconstruction, Bundle Adjustment (BA), Factorization, Omnidirectional, triangulation, rectification, correspondence search, incremental bundle adjustment (IBA).*

I. INTRODUCTION

In the field computer vision the problem of Structure from motion one of the most significant problems, and has taken care of scientists and researchers frequently over the past decade. It deals with reconstructing 3D from 2D images. It can be seen as an automation and extension of photogrammetry [1]. The computation of 3D reconstruction from 2D images generally consists of 3 steps: (1) rectification, (2) correspondence search, and (3) reconstruction, as shown in Fig. 1.

In the step of rectification, determines a transformation of each image (pairs of conjugate epipolar lines become collinear and parallel to the horizontal image axis) to reduce the correspondence problem from 2D search to just 1D search. In the step of correspondence search, the correspondence between pixels is determined in the left and right image. To find the correspondence of a pixel in the left image, we need to search for the same row in the right image. In the step of reconstruction, by using triangulation algorithm input each pixel and its correspondence we can compute the 3D at that pixel.

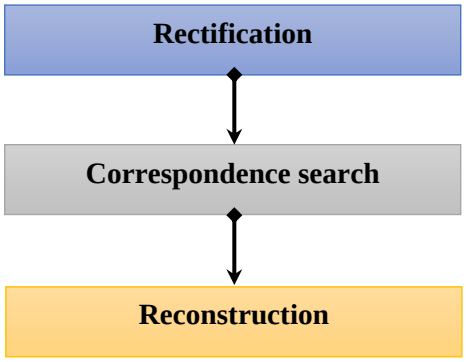

Figure 1: The computation of 3D reconstruction.

Structure from motion solves two problems, the surveying an unknown structure from known camera positions, and determining camera motion from known fix-points, as shown in Fig. 2. The main model to build an SFM system is illustrated in "Fig. 3". The interaction between estimation of the multiple view geometry and the feature tracking consists of the multiple view relationships being used to regularize the feature tracking. The 3D structure of the extracted features estimation based on the features and estimate of the camera parameters. If the extracted features are image points, then the estimated 3D structure is a 3D point cloud.

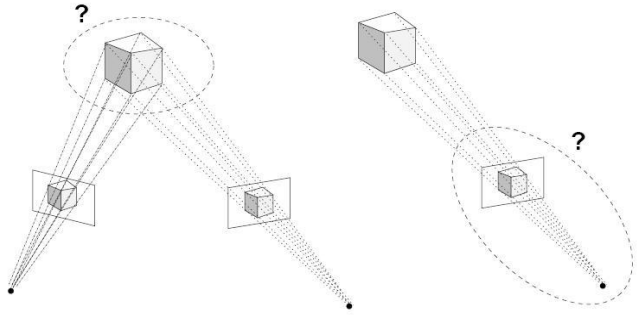

Figure 2: The structure from motion solves two problems [1].

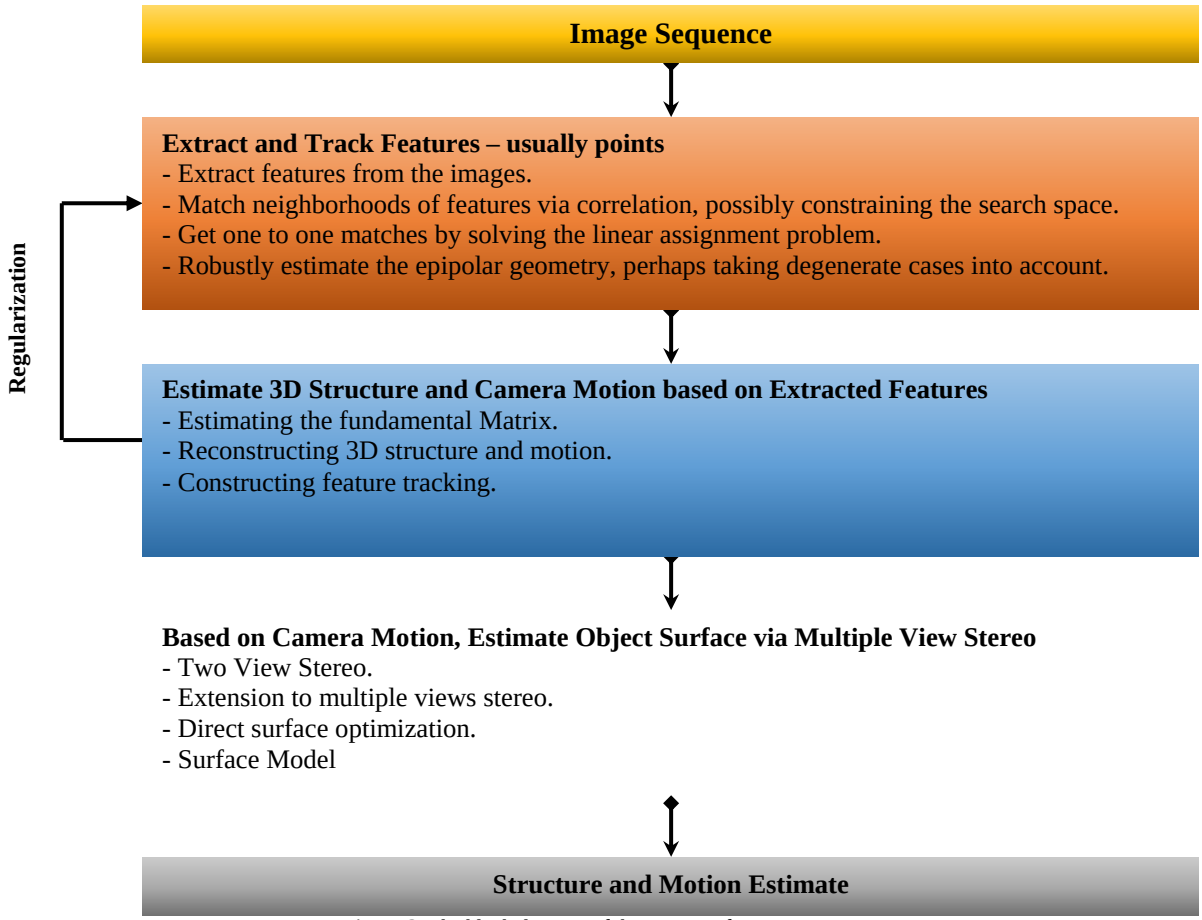

Figure 3. The block diagram of the structure from motion system.

II. STRUCTURE FROM MOTION TECHNIQUES

As previously mentioned, for over 50 years Computer vision community focuses on low-level 3D reconstruction. Now there is a large group of researchers is focusing on the higher-level of 3D reconstruction. In this section, we summarize the work and the research and analysis in the efforts of these researchers [2].

A. Prehistory: Single view 3D Reconstruction

The dominant approach to two-view or multiple-view 3D reconstruction is on the low-level reconstruction using local patch correspondences. The performance of such approach usually significantly outperforms other alternatives, such as reasoning about the lines, because parallax is the strongest cue in this situation. Therefore, there are very few works on higher-level reconstruction in this domain. However, for single view image as input, because there is no parallax between images to utilize, many approaches are forced to try to be smarter to reason about the higher-level reconstruction task. Therefore, in this subsection, we will only focus on the 3D reconstruction of single-view images.

Pixel-wise 3D Reconstruction: There is a line of works on the reconstruction of pixel-wise 3D property. Using Manhattan world assumption, Coughlan and Yuille [3] proposed a Bayesian inference to predict the compass direction from a single image. Hoiem et al. [4] used local image feature to train a classifier to predict the surface orientation for each patch. Moreover, Saxena et al. [5] also used local image feature to train the classifier, but to infer the depth value directly, under a conditional random field framework.

Photo Pop-up: Beyond prediction of 3D property for local image regions, a slightly higher-level representation is to pop-up the photos. Hoiem et al. [6] built a system on top of work proposed in [4] to use local geometric surface orientation to fit ground-line that separates the floor and objects to pop-up the vertical surface. Photo pop-up is not only useful for computer graphics application but also introduce the notion of higher-level reconstruction, by grouping the lower level surface estimation output with regularization (i.e., line fitting).

Line-based Single View Reconstruction: Work in [7, 8, 9, and 10] focused on using lines to reconstruct 3D shapes for indoor images or outdoor buildings, based on exploring the

Manhattan world property of artificial environments. In particular, Han and Zhu [7] designed several common rules for a grammar to parse an image combining both bottom-up and top-down information.

B. Beginning: 3D Beyond Low Level

The volumetric 3D reasoning of the indoor layout marked the beginning of 3D reconstruction beyond a low level. Yu et al. [11] built 3D from a single 2D image by grouping; edges, quadrilaterals, and finally depth. Because it aimed to infer the layout of a room, it is forced to reason about the 3D structure beyond a low level. Since then, several groups independently started working on 3D geometric reasoning. Lee et al. [12] built the 3D structure of the interior of a building by using automatically extracted a collection of line segments from one indoor image. Then, by using geometric reasoning, many structure hypotheses are proposed and verified to find the best suitable model for line segments that converted to a full 3D model.

Beyond lines, Hedau et al. [13] made use of geometric surface prediction [4] modeling the global room space with a parametric 3D box. Going one step further, not only the room layout can be estimated, but we also desire to estimate the objects in the cluster. Hedau et al. [14] developed 3D based object detector and used a probabilistic model to refine the 3D object estimates. Lee at al. [15] proposed a parametric representation of objects in 3D, which allows us to incorporate volumetric constraints of the physical world. On the other hand, going beyond indoor scenes, we can also reason about 3D structure for outdoor scenes. Gupta et al. [16] built a model for 3D of an outdoor scene where objects have volume and mass, and relationships describe the 3D structure and mechanical configurations. Work in [17, 18, 19, 20] reconstructed the 3D at a higher level, especially at extraction of the 3D spatial layout of indoor scenes.

C. Unifying Recognition and Reconstruction

All the approaches mentioned above only focus on 3D reconstruction without any semantic meaning. Xiao et al. [21, 22] designed a system to predict: the scene category, the 3D boundary of space and camera parameters. All objects in the scene represented by their 3D bounding boxes and categories. Xiao et al. [23] proposed a unified framework for parsing an image to infer geometry and semantic structure jointly. Using a structural SVM to encode many novel image features and context rules into the structural SVM feature function and automatically weigh the relative importance of all these rules based on training data. It demonstrates some initial results to infer semantics and structures jointly. For real applications, the unification of recognition and reconstruction may also be very useful, such as [24, 25, 26, and 27].

D. Shape Recognition

Although higher-level 3D understanding starts with indoor room layout estimation, it is also very critical for individual 3D object shape recognition. Very recently, there is a line of works that the emphasis on this problem. In particular, for many objects, their 3D shape can be entirely explained by a simple geometric primitive, such as cuboids. It is the case for most human-made structures [28, 25, and 29]. Therefore, for such an image with cuboids, it would be very useful to parse the image to detect all the cuboids. Our desired output is not simply an indication of the presence of a geometric primitive and its 2D bounding box in the image as in traditional object detection. Instead, Xiao et al. [30] proposed a 3D object detector to detect rectangular cuboids and localize their corners in uncalibrated single-view images depicting everyday scenes. Along the same line, work in [31, 32, 26] proposed 3D detectors for some object categories, such as cars and motorbikes. In particular, Hejrati and Ramanan [31] proposed a two-stage model to propose candidate detection and 2D estimates of shape

E. Human Activity for 3D Understanding

Human activity is a very strong cue for a 3D understanding of scenes. Very recently, there is a line of works [33, 34, 35] pursuing this idea. [33] presented a model for scene understanding by estimates 3D scene geometry and predicted the workspace of a human. [34] presented an approach that exploits the coupling between human actions and scene geometry. They investigated the use of human pose as a cue for single-view 3D scene understanding. Their method used still-image pose estimation to extract functional and geometric constraints about the scene. These limitations were then used to improve single-view 3D scene understanding approaches.

III. 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. In the following, we survey ten categories of applications in detail mainly by illustrating the specific role of SFM in the context of each category of applications [36].

A. 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. There are two methods of 3D reconstruction contact and non-contact, in the non-contact 3D acquisition techniques, there is an image-based 3D modeling technique aims to reconstruct a 3D model from images.

B. Hand-eye Calibration

Hand-eye Calibration problem is the need to determine the measurements obtained by a digital camera attached to a robotic gripper to its coordinate system that corresponds to a homogeneous transformation between the gripper and the camera's coordinate frame.

The early methods of 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 that may not always be 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.

C. 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.

D. Autonomous Navigation/Guidance

Another major category of applications is 'autonomous navigation/guidance,' for which location recognition and object positioning are two important tasks. One popular localization device for robotic vehicles is the Global Positioning System (GPS), which can achieve an accuracy of 1 cm in localization. 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.

E. Motion Capture

The most straightforward approach to capturing human motion is the use of wearable motion sensors. Although this kind of approach can measure the motion directly, it is intrusive and not comfortable to wear. In most existing vision-based motion capture techniques, recording in a closed stage with controlled imaging conditions is required, which restricts their use in the large outdoor area.

F. Image/Video Processing

There are also some interesting and useful applications to image and video processing, such as video enhancement, curved document rectification, and video stabilization.

G. Remote Sensing

In the area of remote sensing and aerial photogrammetry, SFM has also been shown to be able to provide effective solutions. Relevant applications including aerial image mosaicking, 3D mapping, earthwork planning, and landslide investigation.

H. Photo Organization/Browsing

Image organization or browsing is another vast area for applications of SFM. 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.

I. 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 the scene, the performances of both segmentation and recognition have been further improved with the use of SFM. Recent studies have also integrated SFM into segmentation and recognition for more sophisticated algorithms. While modeling the spatial layout and context, the authors combined features in the image projected from 3D cues. Using motion and sparse 3D structure showed that more accurate segmentation and recognition could be achieved.

J. Military applications

In addition, to the aforementioned civilian applications, SFM has also been exploited in military applications. Gather information quickly from a broad range of sensors and process it into useful and dependable data. The use of SFM in situational awareness (SA) appeared in recent research. Using the SFM technique with vision sensors mounted on a moving robotic vehicle.

IV. RELATED WORK

A collection of papers on SFM was surveyed and classified according to the techniques are applied. In this section, we discuss techniques used to reconstruct 3D based on Structure from Motion (SFM).

Dellaert et al. [37] presented a tool to solve the SFM problem without the need for a priori correspondence information. It can deal with images given in any order and taken from widely separate viewpoints. The goal is to Find the maximum likelihood structure and motion given only the 2D measurements, integrating all possible assignments of 3D features to 2D measurements. The goal achieved using an algorithm which iterative refines a probability distribution over the set of all correspondence assignments. The final algorithm is simple, easy to implement and fast.

Jin et al. [38] presented an algorithm to estimate structure and motion using a sequence of images collected in a causal fashion. The algorithm integrates visual information by using a finitely parameterizable class of geometric and photometric models for the scene. The image region is tracking, and 3D motion estimation is combined into a closed loop. They cast the problem of SFM in the framework of nonlinear filtering. The unknown structure and motion are estimated by reconstructing the state of a nonlinear dynamical system via an extended Kalman filter. Furthermore, they have shown that the dynamical system is observable under the assumption that the

scene contains, at least, two planar patches with different normal directions and sufficiently exciting texture, and the translational velocity is non-zero. The recursive nature of the algorithm makes it suitable for real-time implementation.

Koch et al. [39] designed a system for 3D surface reconstruction from image streams of an unknown but static scene. The system operates fully automatic, estimates camera pose and 3D scene geometry. System divided into offline data acquisition (estimate calibration and depth maps for each view) and online rendering (the given data set is used to render novel views at interactive rates). For large and complex scenes the system addresses these issues; Selection of best real camera views, a fusion of multi-view geometry from the views, viewpoint-adaptive mesh generation and viewpoint-adaptive texture blending.

Yang et al. [40] discussed Symmetry to build 3D reconstruction from perspective images. They presented a framework for extracting poses and structures of 2D symmetric patterns from calibrated images. The framework includes all three fundamental types of symmetry reflective, rotational, and translational-based on a systematic study of the homographic groups in image induced by the symmetry groups in space. The system can automatically extract and segment multiple 2-D symmetric patterns present in a single perspective image. The result of segmentation called symmetry cells and complexes, whose 3-D structures and poses are fully recovered.

Scaramuzza et al. [41] discussed a flexible technique for accurate omnidirectional Camera calibration and structure from motion. The method requires only a camera to observe a planar pattern shown in a few different orientations, and the only assumption is that a Taylor series expansion can describe the image projection function. So that no need for a priori knowledge of the motion or a particular model of the omnidirectional sensor. This work on omnidirectional camera calibration is motivated by the use of panoramic vision sensors for SFM and 3D reconstruction. The reconstruction results are very good, and the proposed procedure is easy to use and flexible.

Szeliski and Torr [42] developed techniques used to recover the structure and motion of points seen with 2 or more cameras. They estimated the position of each point in two or more images, assumed that some of the points are coplanar and given one or more image regions where the inter-frame homography are known. These techniques enable us to exploit homography between different regions of the image directly.

Schwartz and Klein [43] proposed a method of merging different connected components resulting from a lack of good input images, aimed to overcome the fact that for the first second no initial global estimation could be found. Into the method, one might manage to speed up the SFM process by automatically dividing the images into subsets by using global image descriptors like the histogram and merging them afterward.

Agarwal [44] Build a system to reconstruct 3D geometry from large, unorganized collections of photographs. The system used algorithms for image matching and 3D reconstruction, which maximize parallelism at each stage of the pipeline. The system is divided into (1) Pre-processing: images are available at a central store from which they are distributed to the cluster nodes on demand in chunks of fixed size. Each node down-samples its images to a fixed size and extracts SIFT features. (2) Verification: propose and verify (via feature matching) candidate image pairs and (3) Track generation: group these features together so that the geometry estimation algorithm can estimate a single 3D point from all the features. The system runs on a cluster of computers (nodes) with one node designated as the master node, responsible for job scheduling decisions.

Sturm and Triggs [45] proposed a method used only fundamental matrices and epipoles estimated from the image data to recover the projective shape and motion from multiple images of a scene by the factorization of a matrix containing the images of all points in all views. Factorization is only possible when the image points are correctly scaled. The algorithm runs quickly and provides accurate reconstructions and results be presented for simulated and real images. Quantitative evaluation of numerical simulations shows the robustness of the factorization and the good performance on noise. The results also show that it is essential to work with normalized image coordinates.

Crandall [46] presented a method for unstructured image collections which considers all the photos at once rather than incrementally building up a solution. Using all available photos, the approach computes an initial estimate of the camera position and then using bundle adjustment to refines that estimate for scene structure. The technique uses a two steps process. The first step, discrete belief propagation (BP) technique is used to estimate camera parameters, the second step, Levenberg-Marquardt non-linear optimization, related to bundling adjustment, but involving additional constraints. The method gives better reconstructions and faster than current incremental bundle adjustment (IBA) approaches.

V. CURRENT RESEARCH TOPICS

- The method of Dellaert et al. [37] has some problems that need to resolve in the future. These problems are features must be picked automatically, results must be shown on sequences with occlusions and the problem of how many features need to be instantiated.
- Koch et al. [39] designed a system has many issues; Selection of best real camera views, a fusion of multi-view geometry from the views, viewpoint-adaptive mesh generation, and viewpoint-adaptive texture blending.
- The system of Agarwal [44] is designed to reconstruct 3D geometry from large, unorganized collections of

photographs has many Issues: all algorithms are operating on the level of connected components; this means that the largest few components completely dominate these stages. The system produces a set of disconnected reconstructions, if the images come with geotags/GPS information, the system can try and geo-locate the reconstructions. However, this information is frequently incorrect, noisy, or missing. The runtime performance of the matching system depends on how well the verification jobs are distributed across the network and Making the system incremental.

- Sturm and Triggs [45] proposed an algorithm to recover the projective shape and motion from multiple images of a scene should be able to treat points that are not visible in all images. Work on the development of the algorithm to use trilinear and quadrilinear matching tensors.
- The Crandall's method [46] had many future work such as; characterize the performance and trade-offs of the algorithm (including studying its scalability to even larger collections of hundreds of thousands of images) and improve the approach (including solving for rotations and translations in a single optimization step).

VI. CONCLUSION

For more than 50 years researchers in the field of 3D reconstruction focus on very low-level reconstruction, such as recovering an accurate depth map or 3D point cloud. In this paper, we argue that just like recognition, reconstruction is a task that contains all low-level, mid-level and high-level representation. We analyzed and studied the efforts of researchers at a higher level in 3D reconstruction. We focused our efforts on the study of new applications and methods for building 3D. We hope that this paper will help the 3D reconstruction research community to focus more on the understanding of the mid-level and high-level 3D, to build an intelligent vision machine eventually. We will concentrate our efforts in the future to improve and develop some of these methods and solve the problem that experienced former researchers to gain access to better 3D reconstruction based on SFM.

REFERENCES

[1] H. Aanaes, "Methods for Structure from Motion," Informatics and Mathematical Modelling Technical University of Denmark Ph.D. Thesis No. 122, Kgs. Lyngby 2003.

[2] J. Xiao, "3D reconstruction is not just a low-level task: retrospect and survey", MIT9.S912: What is Intelligence? Technical Report, 2003.

[3] J. Coughlan and A. Yuille, "Manhattan world: Compass direction from a single image by Bayesian inference," In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2, pp 941–947. IEEE, 1999.

[4] D. Hoiem, A. Efros, and M. Geometri, "context from a single image," In Computer Vision 2005, ICCV 2005, Tenth IEEE International Conference on, volume 1, pp 654–661, IEEE, 2005.

[5] A. Saxena, S. Chung, and A. Ng, "Learning depth from single monocular images," Advances in Neural Information Processing Systems, 18: pp1161, 2006.

[6] D. Hoiem, A. Efros, and M. Hebert, "Automatic photo popup," In ACM Transactions on Graphics (TOG), Volume 24, pp 577–584. ACM, 2005.

[7] F. Han and S. Zhu, "Bottom-up/top-down image parsing by attribute graph grammar," In Computer Vision 2005, ICCV 2005. Tenth IEEE International Conference on, volume 2, pp 1778–1785, IEEE, 2005.

[8] O. Barinova, V. Konushin, A. Yakubenko, K. Lee, H. Lim, and A. Konushin, "Fast automatic single-view 3-d reconstruction of urban scenes", Computer Vision–ECCV 2008, pp 100–113, 2008.

[9] O. Barinova, V. Lempitsky, E. Tretiak, and P. Kohli, "Geometric image parsing in man-made environments," Computer Vision–ECCV 2010, pp 57–70, 2010.

[10] P. Zhao, T. Fang, J. Xiao, H. Zhang, Q. Zhao, and L. Quan, "Rectilinear parsing of architecture in an urban environment," In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp 342 – 349, 2010.

[11] S. Yu, H. Zhang, and J. Malik, "Inferring spatial layout from a single image via depth-ordered grouping," In Computer Vision and Pattern Recognition Workshops, CVPRW'08, 2008.

[12] D. Lee, M. Hebert, and T. Kanade "Geometric reasoning for single image structure recovery," In Computer Vision and Pattern Recognition, IEEE Conference, pp 2136–2143, 2009.

[13] V. Hedau, D. Hoiem, and D. Forsyth, "Recovering the spatial layout of cluttered rooms," In Computer vision, 2009 IEEE 12th international conference on, pp 1849–1856, 2009.

[14] V. Hedau, D. Hoiem, and D. Forsyth, "Thinking inside the box: Using appearance models and context based on room geometry," Computer Vision–ECCV 2010, pp 224–237, 2010.

[15] D. Lee, A. Gupta, M. Hebert, and T. Kanade "Estimating spatial layout of rooms using volumetric reasoning about objects and surfaces," Advances in Neural Information Processing Systems (NIPS), 24, pp1288–1296, 2010.

[16] A. Gupta, A. Efros, and M. Hebert, "Blocks world revisited: Image understanding using qualitative geometry and mechanics," Computer Vision–ECCV 2010, pp 482–496,2010.

[17] Y. Zhao and S. Zhu, "Image Parsing via stochastic scene grammar," In Advances in Neural Information Processing Systems, 2011.

[18] L. Del Pero, J. Guan, E. Brau, J. Schlecht, and K. Barnard, "Sampling bedrooms," In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp 2009–2016 IEEE, 2011.

[19] B. Kermgard, "Bayesian geometric modeling of indoor scenes," In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CVPR '12, pp 2719–2726, Washington, DC, USA, IEEE Computer Society, 2012.

[20] V. Hedau, D. Hoiem, and D. Forsyth, "Recovering free space of indoor scenes from a single image," In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp 2807–2814, IEEE, 2012.

[21] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene cognition from abbey to zoo", In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp 3485 – 3492, 2010.

[22] J. Xiao, K. Ehinger, A. Oliva, and A. Torralba, "Recognizing scene viewpoint using panoramic lace representation," In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp 2695 – 2702, 2012.

[23] J. Xiao, B. C. Russell, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "Basic level scene understanding: from labels to structure and beyond", In SIGGRAPH Asia 2012 Technical Briefs, SA '12, pp 36:1–36:4, New York, NY, USA, 2012.

[24] J. Xiao and L. Quan, "Multiple view semantic segmentation for street view images", In Computer Vision, 2009 IEEE 12th International Conference on, pp 686 –693, 2009.

[25] J. Xiao, T. Fang, P. Zhao, M. Lhuillier, and L. Quan, "Image-based street-side city Modeling", In M SIGGRAPH Asia 2009 papers, SIGGRAPH Asia '09, pp 114:1–114:12, New York, NY, SA, 2009.

[26] J. Xiao, J. Chen, D.-Y. Yeung, and L. Quan, "Structuring visual words in 3d for arbitrary-view object localization", In Proceedings of the 10th European Conference on Computer Vision: Part II, ECCV '08, pp 725–737, Berlin, Heidelberg, 2008.

[27] H. Zhang, J. Xiao, and L. Quan, "Supervised label transfer for semantic segmentation of street scenes", In Proceedings of the 11th European Conference on Computer Vision: Part V, ECCV'10, pp 561–574, Berlin, Heidelberg, 2010.

[28] J. Xiao and Y. Furukawa, "Reconstructing the world's museums", In Proceedings of the 12th European Conference on Computer Vision - Volume Part I, ECCV'12, pp 668–681, Berlin, Heidelberg, 2012.

[29] J. Xiao, T. Fang, P. Tan, P. Zhao, E. Ofek, and L. Quan, "Image-based facade modeling" ACM rans Graph., 27(5): pp 161:1–161:10, 2008.

[30] J. Xiao, B. Russell, and A. Torralba, "Localizing 3d cuboids in single-view images", In Information Processing Systems 25, pp 755–763. 2012.

[31] M. Hejrati and D. Ramanan, "Analyzing 3d objects in cluttered images", In Advances in Neural Information Processing Systems 25, pp 602–610, 2012.

[32] S. Fidler, S. Dickinson, and R. Urtasun, "3d object detection and viewpoint estimation with a deformable 3d cuboid model", In Advances in Neural Information Processing Systems 25, pp 620–628, 2012.

[33] A. Gupta, S. Satkin, A. Efros, and M. Hebert, "From 3d scene geometry to human workspace", In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp 1961– 1968, IEEE, 2011.

[34] D. F. Fouhey, V. Delaitre, A. Gupta, A. A. Efros, I. Laptev, and J. Sivic, "People watching: Human actions as a cue for single-view geometry", 12th European Conference on Computer Vision, 2012.

[35] V. Delaitre, D. Fouhey, I. Laptev, J. Sivic, A. Gupta, and A. Efros, "Scene semantics from long-term observation of people", 2012.

[36] Y. WEI, L. KANG, B.YANG and L. daWU, "Applications of structure from motion: a survey", Journal of Zhejiang University-SCIENCE 2013, Volume 14, pp 486-494, 2013.

[37] F. Dellaert, M. Seitz, E. Thorpe and S. Thrun, "Structure from Motion without Correspondence", Computer Vision and Pattern Recognition, 2000, Proceedings IEEE Conference, vol.2, pp 557 - 564, 2000.

[38] H. Jin, P. Favaro and S. Soatto, "A Semi-direct Approach to Structure from Motion", the visual computer October 2003, Volume 19, pp 377-394, 2003.

[39] R. Koch, J. Evers, J. Frahm and K. Koeser, "3D Reconstruction and Rendering from Image Sequences", Proc. of WIAMIS, 2005.

[40] A. Yang, K. Huang, S. Rao, and Y. Ma, "Symmetry-based 3D reconstruction from perspective images", CVIU, 99: pp 210-240, 2005.

[41] D. Scaramuzza, A. Martinelli and R. Siegwart, "A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion", ICVS, page 45, IEEE Computer Society, 2006.

[42] R. Szeliski and P. Torr,"Geometrically Constrained Structure from Motion: Points on Planes", Workshop on 3D Structure from Multiple Images of Large-scale Environments (SMILE), Freiburg, Germany, June 1998.

[43] C. Schwartz and R. Klein, "Improving Initial Estimations for Structure from Motion Methods", 13th Central European Seminar on Computer Graphics (CESCG 2009), 2009.

[44] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S.Seitz, and R. Szeliski, "Building Rome in a Day", ICCV, pp 72-79, IEEE, 2010.

[45] P. Sturm and W. Triggs, "A Factorization Based Algorithm for multi-Image Projective Structure and motion", CCV (2), volume 1065 of Lecture Notes in Computer Science, pp 709-720, 2011.

[46] D. Crandall, A. Owens, N. Snavely and D. Huttenlocher, "Discrete-Continuous Optimization for Large-Scale Structure from Motion", CVPR, pp3001-3008, IEEE, 2011.