Slimane LARABI

Computer Vision

From Bidimensional Images to Three Dimensional Scene

January 02, 2025

Springer Nature

Chapter 1

Computer Vision in the Beginning of Artificial Intelligence

1.1 Introduction

The transition from primitive to the sophisticated computer vision models we know today has been made possible by the many contributions of scientists over the past six decades. The results of these works have allowed computer vision to evolve to be able to interpret and understand visual information in a manner similar to the human brain.

During the 1960s, universities seriously considered the computer vision system project as a stepping stone to artificial intelligence. Researchers of this era were extremely optimistic about the future of these related fields and promoted artificial intelligence as a technology with the potential to transform the world.

The objective of the various research projects was to provide robots with a universal artificial vision system, which functionally is similar to the human visual system. Very quickly, the researchers realized the difficulty of the posed problem. This forced them to direct their research works towards the development of specific vision systems. Constraints have been issued on the nature of the objects processed to facilitate the task of the system of vision. The class of polyhedral objects was chosen in all the first approaches proposed. Despite this simplification of the nature of the images to be processed, the resolution of the different problems encountered was not made until around twenty years later.

This chapter explores the scientific events made in order to reproduce human vision using computational methods that actually advanced computer vision.

One of the earliest contributions to the field is the work of L. G. Roberts in 1965 [26], which laid the groundwork for future developments in image processing and understanding. His approach to contour labeling was a significant leap forward in recognizing the importance of image structure in visual interpretation.

The Summer Vision Project in 1966 [1] further expanded this research, introducing learning from visual data to better understand image perception.

A crucial breakthrough came in 1971 with the development of random dot stereograms by Bela Julesz, [31] which provided new insights into depth perception and the human visual system. We present at the end of this chapter the influential model of David Marr, published in 1982 [32], which offered a comprehensive framework for understanding the stages involved in human vision and inspired subsequent computer vision algorithms.

1.2 Approach by L. G. ROBERTS (1965) [26]

Lawrence GILMAN Roberts, was the first one got a Ph.D. in computer vision. His Ph.D. thesis entitled ”Machine Perception of Three-Dimensional Solids“ is considered one of the foundational works of the field of computer vision [26].

L. G. Roberts proposed a computer system for perception and recognition of 3D scenes of polyhedral objects (see figure 1.1). From an image taken by a camera, this system performs the following tasks:

- Detection of contrast points (contours).
- Connection of contour points for the formation of contour lines.
- Drawing of lines (straight lines in the sense of least squares)
- recognition of observed objects

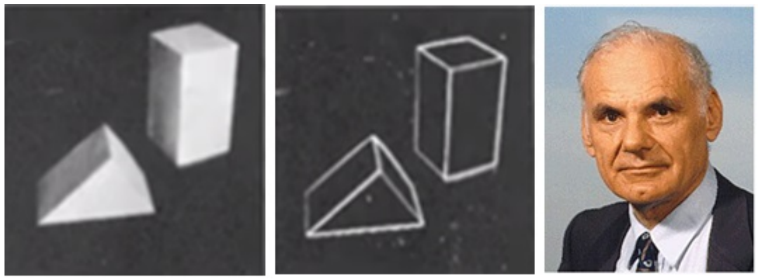

Fig. 1.1 (Left) Example of processed scene, (Middle) the located contours, (Right) L. G. ROBERTS.

Image is processed by local differential operator in order to locate lines. The procedure found at that time makes mistakes in complex pictures. Figure 1.2 shows the image after computer sampling of the initial image, the features selected and the final lines drawing.

Fig. 1.2 (Left) Initial image, (Middle) Extracted features and (Right) The drawing of lines.

The next step is the construction of the dimensional object from line drawing using three models (cube, wedge, and hexagonal prism). A matching of the 2D representation obtained with the 3D models is performed. In the example of Figure 1.3, the set of corners A, B, C, D, E is first identified with a model cube. The defined contours from these corners are then subtracted from the global drawing, and the same process is applied to the rest of the drawing.

After tests performed on different combinations of objects, the main difficulty encountered in the development of this system was the identification of the lines of contours corresponding to the same surface of the object.

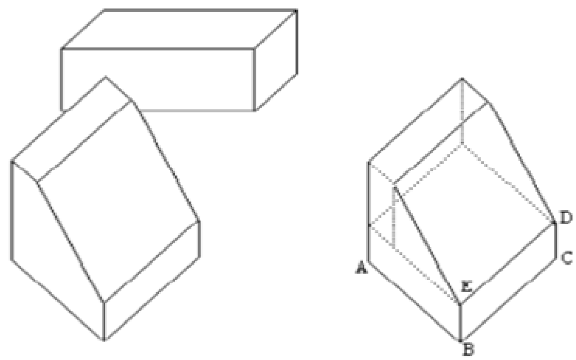

Fig. 1.3 Applying the recognition process The set of surfaces defined by the corners A, B, C, D, E are first recognized as a part of a parallelepiped object.

1.3 Labeling contours of images for understanding

1.3.1 Heuristic Approach of GUZMAN [27]

A. GUZMAN [27] proposed an heuristic approach applied to the world of polyhedral objects, and unlike L. G. ROBERTS, A. GUZMAN performed a grouping of lines based on heuristics. In his work, he tried to find how a computer can identify and recognize objects in visual scene. An example of image of a scene is shown in Figure 1.4. The objects O_i, the aimed grouping of surfaces is:
$O_1 = \{1, 8, 9\}$, $O_2 = \{7, 2\}$, $O_3 = \{3, 5, 6\}$, $O_4 = \{4, 13, 14\}$, $O_5 = \{10, 15, 16\}$ (see Figure 1.4).

The proposed system does not identify objects, only performs grouping. Moreover, a failure for certain designs was observed. The fundamental problem in the analysis of images of polyhedra is the partition of an image into separate objects. The proposed approach is guided by intuition and observation, but is not based on theoretical analysis.

The vertex is defined as a point of intersection of two or more boundaries of regions. These regions might or might not be faces of a single body. The analysis

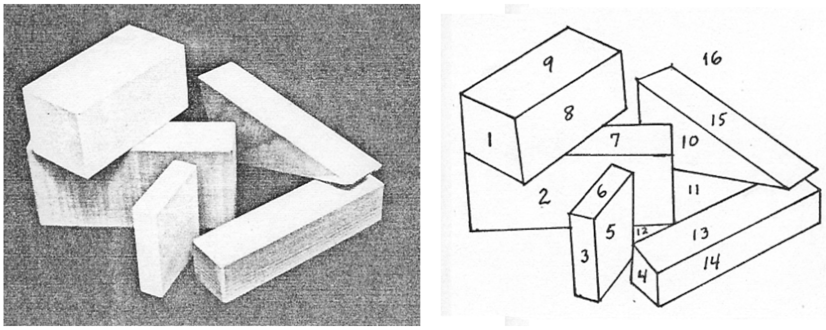

Fig. 1.4 Example of scene and the aimed grouping of lines forming the objects.

of images of polyhedral scenes revealed five (05) more significant types of vertexes described in table 1.1, (see Figure 1.5):

Type of Node	Number of corners	Nature of grouping	Example
V	3	01 region	(BAF)
W	5	Grouping of 02 regions	(AFG, GFE)
Y	4	Grouping of 03 regions	(BGD, BGF, FGD)
X	5	Grouping of 04 regions	(CAQ, QAH, CAK, KAH)
T	4	Grouping of 02 regions	
		the more distant part of the scene is hidden	

Table 1.1 Different types of nodes.

1.3.2 Labeling contour lines

Still for the world of blocks M. B. Clowes [28] proposed the notion of labeling contour lines. Each line in a drawing delimiting one or more surfaces is:

- labeled by $(+)$ if the angle formed by the two surfaces is convex,

- labeled by $(-)$ if the angle formed by the two surfaces is convex,

- labeled by $(->)$ if the line is part of an obscured surface. The direction of the

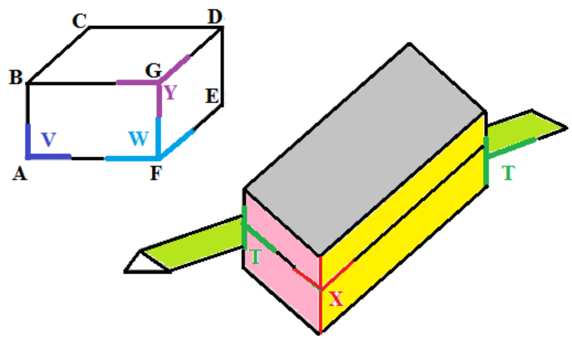

Fig. 1.5 Different kinds of nodes proposed by A. Guzman [27];

arrow is chosen so that the hidden surface is to the left of the labeled line. The figure 1.6 shows an example of labeling.

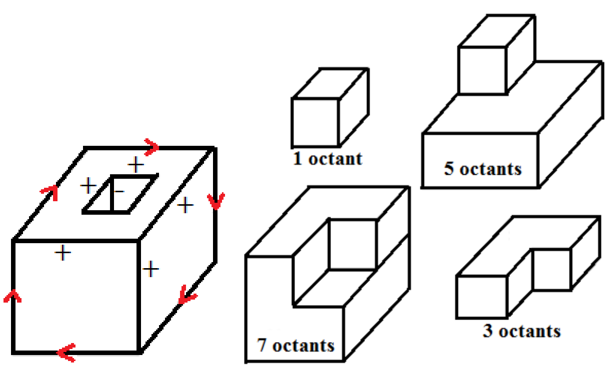

Fig. 1.6 Contour labeling

Therefore, only a few labeling configurations are possible. Using this notion of labeling, it is possible to recognize entities from 2D images. The problem posed by this approach is the large number of combinations obtained for a number of nodes greater than 10.

By focusing on the labeling of the nodes listed by A. GUZMAN, there are actually only 3 possibilities for node Y, 3 possibilities for W, 4 for T and 6 for V.

The possible labelings of each node are obtained by the perception of the node from the empty octants by looking at four types of nodes (1, 3, 5, 7) where each number indicates the number of full octants (see figure 1.7).

In his work, D. L. WALTZ [29] proposed the extension of the method labeling for a larger set of features including shadow edges. So from new labelings and new nodes are generated. Although the number of possible labelings of the different nodes becomes very high, D. L. WALTZ discovered that only a very small percentage of labels are possible (3 for some nodes and much less for others).

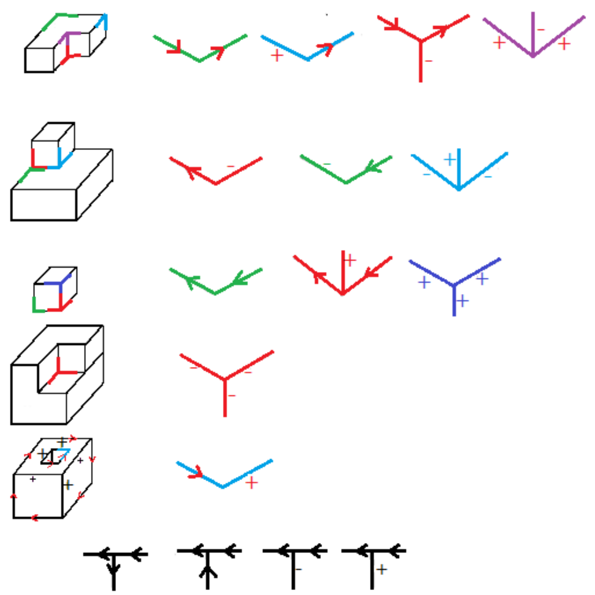

Fig. 1.7 The possibilities of the different junctions;

In addition, D. L. WALTZ [29] proposed a prediction, verification, and propagation algorithm which allows us to properly assign labels to all the contours of the drawing. This technique applied for the recognition of the surface $ABCDE$ is illustrated by the graph of Figure 1.8.

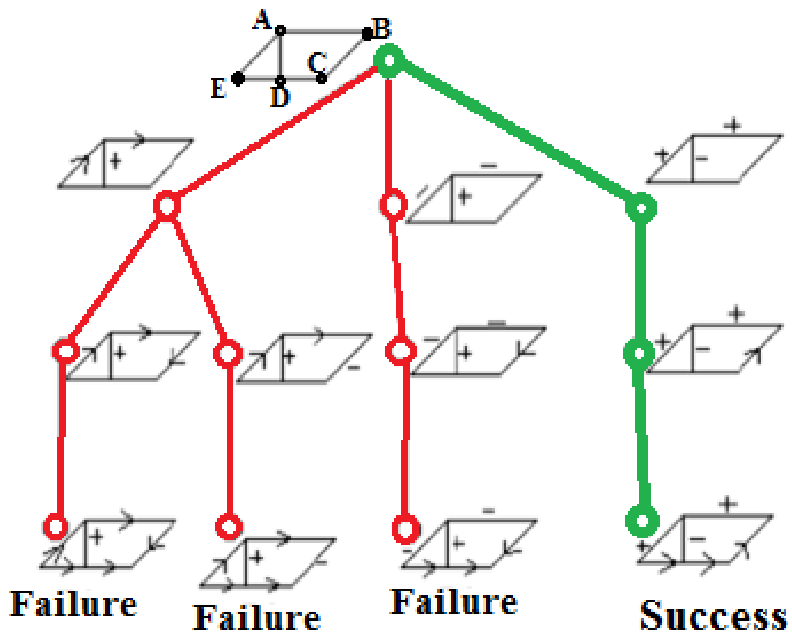

Fig. 1.8 Applying the prediction, propagation, verification Algorithm

1.4 Summer Vision Project, 1966

In 1966, Seymour Papert proposed to MIT students a Summer Vision Project [1] in which the graduate student would connect a camera to a computer and ask it to describe what it sees, then come up with an important part of a visual system.

They made the first attempt to mimic the human brain, triggering further research into computers' ability to process information to make intelligent decisions.

The main goal of this project is to build a system of programs that divide an image into objects, background. The second objective was the description of the region through the analysis of shape and surface properties. The final goal was the identification of objects by matching them to a vocabulary of known objects.

This main goal has been considered as computer vision research and conducted during the 1960s at the MIT Laboratory for Artificial Intelligence, under the direction of Marvin Minsky, with contributions from Patrick Winston and Berthold Horn

[5]. The implemented programs extract line drawings from images, utilize knowledge about the three-dimensional world, and incorporate new ideas about artificial intelligence into these processes.

1.5 Random Dot Stereograms of Bela Julesz, 1971

Béla Julesz (1928–2003) [31] invented the random dot stereogram for understanding the Human Visual System and especially what are the origins of Binocular Fusion?

A random dot stereogram consists of two random arrays of dots that, when viewed stereoscopically, with one array seen by each eye, appear to contain a shape such as a triangle or square lying in a plane either in front of or behind the rest of the dots, bounded by an illusory contour (see Figure 1.9).

It is constructed by generating two arrays of randomly placed dots, identical except for a clearly defined region that is slightly shifted sideways in one of the arrays, and it is usually presented for viewing by printing one of the arrays in red and the other in green or cyan (blue-green), with a slight horizontal displacement so that the unshifted dots do not fall exactly on top of one another, and it is viewed with spectacles having one red and one green or cyan lens.

1.6 The model of David Marr, 1980

David Marr (1945-1980): is a British neuroscientist and physiologist. He integrated results from psychology, artificial intelligence, and neurophysiology into a new model of visual processing.

Inspired by the results of Neurophysiology, the proposed model for Computer Vision is based on three levels of representation (see Figure 1.10):

-**The primary sketch**: it is the collection of characteristics of the image

-**Sketch 2 ½ D**: Acquisition of data concerning the orientation of the surfaces, surface depth, discontinuity contours. This is a representation that is centered on the

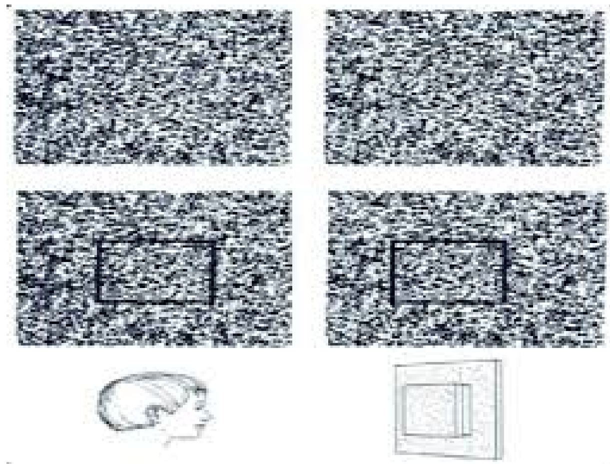

Fig. 1.9 Classic Julesz random dot stereogram of a square defined by a disparity from its background.

observer.

-**The 3D sketch**: Correspondence of the $2\frac{1}{2}D$ representation with the 3D model (Knowledge). The volumetric and surface primitives are used to obtain a description of the scene in terms of objects and a relationship between objects. This representation is focused on stage.

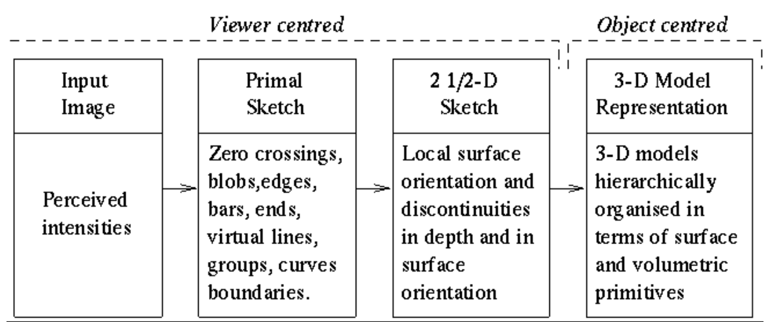

Fig. 1.10 The three levels of information processing.

1.7 Conclusion

The evolution of computer vision has been shaped by a series of fundamental advances, each of which has contributed to the vision systems we recognize today. From the early work of L. G. Roberts on edge detection, through the edge labeling proposed by Guzman and Waltz, to the pioneering stereograms of Bela Julesz. These early studies provided fundamental advances to allow machines to begin to interpret visual data. The Summer Vision Project of 1966 exemplifies the spirit of collaboration that led to further exploration, while David Marr's 1972 model established a theoretical framework that would guide much of the field's subsequent progress.

These early efforts not only advanced the understanding of our visual system but also laid the groundwork for the development of algorithms that could eventually replicate human-like perception. The progression from basic image processing techniques to the sophisticated, machine learning-driven computer vision systems of today underscores the enduring impact of these early milestones. As the field continues to evolve, the foundations discussed in this chapter remain integral to the ongoing quest to enhance how computers understand and interact with the visual world.

In the next chapter, we assume that the set of features are extracted from image such as edges, corners and SIFT descriptors. We will study how can we compute transformations of image using Homography. The homography is a useful mathematical tool, it allows to perform many applications such as image stitching, camera calibration.

References

1. Seymourt Papert. The Summer Vision Project. Artificial Intelligence Group, MIT, 1966. https://dspace.mit.edu/handle/1721.1/11589
2. Bob Sumner. Augmented Creativity, TEDxZurich, 19th of January, 2015. https://www.youtube.com/watch?v=AJJOWemfOYI
3. Sachiko Iwase and Hideo Saito. Tracking soccer players based on homography among multiple views", Proc. SPIE 5150, Visual Communications and Image Processing 2003, (23 June 2003); https://doi.org/10.1117/12.502967
4. Martin A. Fischler, Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, Volume 24, Issue 6, pp 381–395, https://doi.org/10.1145/358669.358692
5. "The Eye of a Robot: Studies in Machine Vision at MIT" and "TX-O Computer". Courtesy of MIT Museum.
6. Richard Hartley and Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
7. D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), volume 1, pages 195-202, Madison, WI, June 2003.
8. Zhengyou Zhang. A Flexible New Technique for Camera Calibration. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 11, NOVEMBER 2000
9. A computer algorithm for reconstructing a scene from two projections. Nature volume 293, pages 133–135 (1981)
10. OD Faugeras, QT Luong, SJ Maybank. Camera self-calibration: Theory and experiments. Second European Conference on Computer Vision ECCV'92, 1992
11. Olsson, Carl; Enqvist, Olof, Stable Structure from Motion for Unordered Image Collections Scandinavian Conference on Image Analysis, 2011.
12. Marc Pollefeys and Luc Van Gool. 3-D Modeling from Images. COMMUNICATIONS OF THE ACM July 2002/Vol. 45, No. 7.
13. Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Xin ; O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis , July 2017. ACM Transactions on Graphics, 36(4):1-11
14. M. Michalkiewicz, J. K. Pontes, D. Jack, M. Baktashmotlagh, A. Eriksson. Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. 2019. arXiv:1901.06802 [cs.CV]. https://arxiv.org/pdf/1901.06802.pdf

15. A. Podlozhnyuk, S. Pirker, C. Kloss. Efficient implementation of superquadric particles in Discrete Element Method within an open-source framework. Computational Particle Mechanics 4(1), 2016. DOI: 10.1007/s40571-016-0131-6
16. M.A. Turk et al. Face recognition using eigenfaces,CVPR 1991.
17. S.K.Nayar et al., Real-Time 100 Object Recognition System, ICRA, 1996
18. C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese. 3dr2n2: A unified approach for single and multi-view 3d object reconstruction. In European conference on computer vision, pages 628–644. Springer, 2016
19. R. Girdhar, D. F. Fouhey, M. Rodriguez, and A. Gupta. Learning a predictable and generative vector representation for objects. In European Conference on Computer Vision, pages 484–499. Springer, 2016
20. D. J. Rezende, S. A. Eslami, S. Mohamed, P. Battaglia, M. Jaderberg, and N. Heess. Unsupervised learning of 3d structure from images. In Advances in Neural Information Processing Systems, pages 4996–5004, 2016
21. S. R. Richter and S. Roth. Matryoshka networks: Predicting 3d geometry via nested shape layers. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1936–1944, 2018
22. C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., 1(2):4, 2017.
23. Y. Liao, S. Donne, and A. Geiger. Deep marching cubes: ´ Learning explicit surface representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2916–2925, 2018
24. H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3d object reconstruction from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., volume 2, page 6, 2017
25. Mateusz Michalkiewicz et al. Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. arXiv:1901.06802v1 [cs.CV], 21 Jan 2019.
26. Roberts, Lawrence G. 1963. Machine perception of three-dimensional solids. Outstanding PhD dissertations in the computer sciences. Garland Publishing, New York. 1963.
27. Adolfo Guzman-Arénas, Computer Recognition of Three-Dimensional Objects In a Visual Scene. PhD Dissertations in the computer sciences, MIT 1968.
28. M.B. Clowes. On seeing things. Artificial Intelligence, Volume 2, Issue 1, Spring 1971, Pages 79-116.
29. David L. Waltz. GENERATING SEMANTIC DESCRIPTIONS FROM DRAWINGS OF SCENES WITH SHADOWS. MIT Artificial Intelligence Laboratory. Technical Report 271, November 1972.

30. Kemp, M. Julesz's joyfulness. Nature 396, 419 (1998). https://doi.org/10.1038/24753
31. Julesz B. Foundations of Cyclopean Perception. Chicago University Press; Chicago, IL, USA: 1971.
32. David Marr. Vision, A Computational Investigation into the Human Representation and Processing of Visual Information. Originally published: San Francisco : W. H. Freeman, c1982.
33. Arthur Coste. Affine Transformation, Landmarks registration, Non linear Warping. https://www.sci.utah.edu/ãcoste/uou/Image/project3/ArthurCOSTE_Project3.pdf October 2012.
34. Sid Yingze Bao, Mohit Bagra, Yu-Wei Chao, Silvio Savarese. Semantic Structure From Motion with Points, Regions, and Objects. CVPR 2012.
35. Marco Crocco, Cosimo Rubino, Alessio Del Bue. Structure from Motion with Objects, CVPR 2016.
36. D.P. Robertson and R. Cipolla. Structure from Motion. In Varga, M., editors, Practical Image Processing and Computer Vision, John Wiley, 2009.
37. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. International Journal of Computer Vision, 9(2):137–154, 1992.
38. P. F. Sturm andW. Triggs. A factorization based algorithm for multi-image projective structure and motion. In European Conference on Computer Vision (ECCV'96), pages 709–720, 1996.
39. F. Schaffalitzky, A. Zisserman, R. I. Hartley, and P. H. S. Torr. A six point solution for structure and motion. In European Conference on Computer Vision (ECCV'00), pages 632–648, 2000.
40. W. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: A modern synthesis. In W. Triggs, A. Zisserman, and R Szeliski, editors, Vision Algorithms: Theory and Practice, LNCS, pages 298–375. Springer Verlag, 2000.
41. D. C. Brown. The bundle adjustment - progress and prospects. International Archives of Photogrammetry, 21(3), 1976.