ISSN: 2321-7782 (Online)
Volume 3, Issue 6, June 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

A Review on Converting 2D Images to 3D Models

Roshnara Nasrin. P. P[1]
PG scholar
MEA Engineering College
Perinthalmanna - India

Sherin Jabbar[2]
Assistant Professor
MEA Engineering College
Perinthalmanna - India

Abstract: 3D reconstruction of authentic world objects is becoming more and more popular among computer graphics and computer vision researchers. One of the more practical approaches of achieving 3D reconstruction is called Multi View Geometry, an approach that are utilizing images taken of the authentic world objects to reconstitute it. Depending on the number of input images, the subsisting conversion algorithms can be categorized into two groups: algorithms predicated on two or more images (binocular depth) and algorithms predicated on a single still image (monocular). In the first case, the two or more input images could be taken either by multiple fine-tuned cameras located at different viewing angles or by a single camera with moving objects in the scenes. The second operates on single image. Cognate works on 3D reconstruction [1], [2], [3], [4] have been reviewed and observed that the 3D reconstruction of multi-ocular images predicated on silhouette is efficient and effective.

Keywords: 3d reconstruction, Depth cues, Shape from silhouette, Shape from motion, Shape from focus, Binocular disparity.

I. INTRODUCTION

The widespread utilization of 3D body scanning is set to become a major technological change in the coming decade. Body Aspect provides 3D body scanning and analysis to enable clinicians, image consultants and habiliments retailers to solve body image, shape and size issues.

Image has become an aspect of life that is increasingly difficult to ignore. The aim in 3D reconstruction is to use a turntable or a robot using which we can obtain photographs of a given object from known angles. Thus this survey does not limit itself to giving the user the freedom of viewing the object at any angle and in any orientation but provides a complete model in space. 3D reconstruction and modeling is used in many fields like Virtual Reality, recognizing and manipulating objects etc.

The world of 3D incorporates the third dimension of depth, which can be perceived by the human vision in the form of binocular disparity. Human ocular perceivers are located at marginally different positions, and these perceive divergent views of the authentic world. The brain is then able to reconstruct the depth information from these divergent views. A 3D exhibit capitalizes on this phenomenon, engendering two remotely different images of every scene and then presenting them to the individual ocular perceivers. With a congruous disparity and calibration of parameters, a correct 3D perception can be realized.

There has been rapid progress in the fields of image capturing, coding and exhibit, which brings the realm of 3D more proximate to authenticity than ever before. Related works on 3D reconstruction [1], [2], [3], [4] have been reviewed and observed that the 3D reconstruction of multi-ocular images predicated on silhouette is efficient and effective.

II. BACKGROUND

Depending on the number of input images can be categorized the subsisting conversion algorithms into two groups: algorithms predicated on two or more images and algorithms predicated on a single still image. In the first case, the two or more input images could be taken either by multiple fine-tuned cameras located at different viewing angles or by a single camera with moving objects in the scenes. The depth cues utilized by the first group can be referred as the multi-ocular depth cues. The second group of depth cues operates on a single still image, and they are referred to as the monocular depth cues. The Table 1 summarizes the depth cues utilized in 2D to 3D conversion algorithms and their representative works. Algorithms using specific cues are reviewed in this work.

A. *Other depth cues*

Apart from the depth cues described in this chapter, which are fairly ascendant in the current computer vision field, a number of other depth cues with different principles subsist and have withal been prosperously translated into algorithms, for example, shadow, dynamic occlusion, static occlusion predicated on T-junction, and so on. Due to reasons like perspicacious property rights and the inhibited scope of the survey, these depth cues are not investigated here.

According to the depth cues on which the algorithms reply, the algorithms are relegated into the following 7 categories: binocular disparity, motion, defocus, focus, silhouette, atmosphere scattering, and shading.

TABLE 1: Depth Cues And Their Representative Algorithms

The Number of Input Images	**Depth Cues**	**Representative Works**
Two or More Images (binocular or multi-ocular)	Binocular disparity[1]	Correlation-based, feature-based correspondence; triangulation
	Focus [2]	A set of images of different focus level and sharpness estimation
	Motion [3	Optical flow ; Factorization; Kalman filter
	Silhouette[4];[5]	Voxel-based and deformable mesh model
One single image (monocular)	Defocus	Second Gaussian derivative
	Atmospheric scattering	Atmosphere Scattering Light scattering model
	Shading	Shading Energy minimization

This survey describes and analyzes algorithms that use multiple cues because the multi-ocular depth cues take both spatial and temporal image information into account, which yield in general a more precise result. The monocular depth cues are less precise but do not require multiple images, which make them more multifarious. The papers referred on multi-ocular depth cues are follows.

Binocular disparity

With two images of the same scene captured from remotely divergent view points, the binocular disparity can be utilized to recuperate the depth of an object. This is the main mechanism for depth perception. First, a set of corresponding points in the image pair are found. Then, by designates of the triangulation method, the depth information can be retrieved with a high degree of precision.

Depth-From-Focus

The depth-from-focus approach is proximately cognate to the family of algorithms utilizing depth from defocus. The main difference is that the depth-from-focus requires a series of images of the scene with different focus levels by varying and registering the distance between the camera and the scene, while depth-from-defocus only needs 2 or more images with fine-tuned object and camera positions and use different camera focal settings.

 ISSN: 2321-7782 (Online)

Shape from Motion

The relative kineticism between the viewing camera and the observed scene provides a consequential cue to depth perception: near objects move more expeditious across the retina than far objects do. The extraction of 3D structures and the camera kineticism from image sequences is termed as structure from motion. The kineticism may be visually perceived as a form of "disparity over time", represented by the concept of kineticism field. The kineticism field is the 2D velocity vectors of the image points, induced by the relative kineticism between the viewing camera and the observed scene. The rudimentary posits for structure-from-motion are that the objects do not deform and their forms of kineticism are linear.

Shape from silhouette

A silhouette of an object in an image refers to the contour dissevering the object from the background. Shape-from-silhouette methods require multiple views of the scene taken by cameras from different viewpoints. Such a process together with correct texturing engenders full 3D model of the objects in the scene, sanctioning viewers to observe a live scene from an arbitrary viewpoint.

Shape-from-silhouette requires precise camera calibration. For each image, the silhouette of the target objects is segmented utilizing background subtraction. The retrieved silhouettes are back projected to a prevalent 3D space with projection centers identically tantamount to the camera locations. Back-projecting a silhouette engenders a cone-like volume. The intersection of all the cones forms the visual hull of the target 3D object, which is often processed in the voxel representation. This 3D reconstruction procedure is referred to as shape-from-silhouette.

III. LITERATURE SURVEY

The field of 3D visualization is a paramount aspect of image processing, because of their immensely colossal applications in many areas of our life. This work reviews some of works related to 3D conversion.

D. Comanducci, A. Maki, C. Colombo, and R. Cipolla[1], presents an efficacious package for 3D reconstruction; with two images of the same scene captured from remotely different viewpoints, the binocular disparity can be utilized to recover the depth of an object. This is the main mechanism for depth perception. First, a set of corresponding points in the image pair are found. Then, by betokens of the triangulation method, the depth information can be retrieved with a high degree of precision when all the parameters of the stereo system are kenned. When only intrinsic camera parameters are available, the depth can be recuperated correctly up to a scale factor.

J. Florczak and M. Petko[2], presents depth-from-focus approach is proximately cognate to the family of algorithms utilizing depth from defocus. The main difference is that the depth-from-focus requires a series of images of the scene with different focus levels by varying and registering the distance between the camera and the scene, while depth-from-defocus only needs 2 or more images with fine-tuned object and camera positions and use different camera focal settings.

L. Torresani, A. Hertzmann, and C. Bregler[3], presents relative kineticism between the viewing camera and the observed scene provides a consequential cue to depth perception: near objects move more expeditious across the retina than far objects do. The extraction of 3D structures and the camera kineticism from image sequences is termed as structure from motion. The kineticism may be optically discerned as a form of "disparity over time", represented by the concept of kineticism field. The kineticism field is the 2D velocity vectors of the image points, induced by the relative kineticism between the viewing camera and the observed scene. The rudimental posits for structure-from-motion are that the objects do not deform and their forms of kineticism are linear.

T. Matsuyama, X. Wu, T. Takai, and T. Wada[4], presents silhouette of an object in an image refers to the contour dissevering the object from the background. Shape-from-silhouette methods require multiple views of the scene taken by cameras from different viewpoints. Such a process together with correct texturing engenders full 3D model of the objects in the scene, sanctioning viewers to observe a live scene from an arbitrary viewpoint. Shape-from-silhouette requires precise camera calibration. For each image, the silhouette of the target objects is segmented utilizing background subtraction. The retrieved silhouettes are back projected to a prevalent 3D space with projection centers equipollent to the camera locations. Back-projecting a silhouette engenders a cone-like volume. The intersection of all the cones forms the visual hull of the target 3D object, which is often processed in the voxel representation. This 3D reconstruction procedure is referred to as shape-from-silhouette.

D. C. Schneider[5], proposes Shape from silhouette (SFS) algorithms compute the (approximate) 3D shape of an object from multiple 2D projections considering only the outline of the object in the projections. The most paramount class of SFS methods are Visual Hull algorithms. The quandary of computing shape from outlines (silhouettes) of projections is generally under constrained. Depending on the object's geometry and the number of available views, approximations can be computed which are sufficient for some applications. Often, the SFS output accommodates as initialization for appearance-predicated shape reconstruction methods such as volumetric reconstruction or multiview stereo. The most general class of SFS approaches, where no posits about the type of objects to reconstruct are made, are the Visual Hull algorithms.

IV. Observation and Analysis

Fair and efficacious performance evaluation of 2D to 3D conversion algorithms requires conscientious design of criteria and data sets. However, there is a lack of uniformity in the framework predicated on which methods are evaluated. Furthermore, an abundance of papers do not provide an explicit quantitative performance analysis, which perplexes the evaluation process. It is thus imprudent to make explicit claims such as which methods indeed have the lowest error rates or which methods are the most expeditious.

The comparison is predicated on 6 qualitative aspects. Some of them are correlated with each other. This chapter is thus dedicated to evaluating and elucidating the results in the sundry aspects in the comparison table.

3.1.1 Image acquisition:

This aspect describes the purposive modification of the image acquisition system’s parameter, that is, whether the method is active or passive. It is observed that virtually all multi-ocular depth cues require a special camera set-up, and most monocular depth cues do not.

3.1.2 Image content:

The image content aspect involves what kind of image characteristics is needed by the algorithms in order for them to work reliably. Some of them have been assigned the term “All” in the table, which betokens that no special requisite of the image content is needed.

3.1.3 Motion presence:

The motion presence aspect concerns the presence of disparity of the same feature point in the input images. It is only applicable for multi-ocular depth cues. Since monocular depth cues operate on a single image, no motion is needed.

3.1.4 Real-time processing:

Some of the investigated papers provide explicit running time and environment parameters, others just claim that the algorithm is opportune for authentic-time application or do not mention the speed at all This is reflected in the comparison table.

3.1.5 Accuracy

As the aspect of authentic time processing, precision comparison additionally lacks experimental data for certain depth cues. And if it subsists, the test data and environment are not predicated on a uniform substructure. We therefore only present the available error quantifications here and do not make any normalization.

After evaluating the comparison table can be concluded that 3d reconstruction based on silhouette is far better than other depth cues. From the table it is clear that the comparison based on accuracy binocular disparity and focus algorithm had errors compared to other algorithms. According to the given parameters for comparison silhouette algorithm is better than other cues.

Table 2: Multi-ocular Depth Cue Comparison

Depth cues	**Image acquisition**	**Image content**	**Motion presence**	**Real-time processing**	**Accuracy**	**Miscellaneous**
Binocular Disparity[1]	Active	All	Yes	Yes	Error >1	Sensitive to occlusions
Motion [3]	Active/ Passive	All	Yes	Yes	High accuracy	Optical flow sensitive to noise
Focus [2]	Active	Object with complex surface characteristics.	No	No	Relative error rate: 0.1%	Sensitive to aberrations of lens system
Silhouette [4], [5]	Active/ Passive	Foreground objects distinguishable	Yes	Yes	High accuracy	

V. CONCLUSION

A prodigious number of 2D to 3D conversion algorithms are dedicated to recover the “structure” or “shape” of objects in the images, which are understood to mean the 3D coordinates of a minuscule set of points in the scene. These algorithms (e.g. shading, silhouette, symmetric patterns) can be possibly put into better use in computing the 3D kineticism of the camera or the objects, robot navigation, surveillance etc. rather than 3D video.

The multi-ocular depth cues take both spatial and temporal image information into account, which yield in general a more precise result. The monocular depth cues are less precise but do not require multiple images, which make them more multifarious. Image sequences where both objects and camera remotely move can best resort to the monocular cues.

A silhouette predicated 3D reconstruction algorithm is a multi-ocular depth cue capable of authentic-time or near authentic-time utilization has been reviewed. It is clear that by using larger resolutions on both the reconstructed volume and the silhouette images the algorithm produces higher quality results. It has also been observed that the increase in accuracy inherit from increasing the number of silhouette images is rapidly decreased between the usage of one to ten silhouette images. When using more than ten silhouette images the increase in accuracy is only visible in smaller details.

ACKNOWLEDGEMENT

The author would like to express her sincere thanks to HOD, project guide and staff in Computer Science department, M E A Engineering College for many fruitful discussions and constructive suggestions during the preparation of this manuscript.

References

1. D. Comanducci, A. Maki, C. Colombo, and R. Cipolla, “2d-to-3d photo rendering for 3d displays," in Proc. Of international Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), 2010.
2. J. Florczak and M. Petko, “Usage of shape from focus method for 3d shape recovery and identication of 3d object position,“ International Journal of Image Processing (IJIP) , vol. 8, no. 3, p. 116, 2014.

3. L. Torresani, A. Hertzmann, and C. Bregler, "Learning non-rigid 3d shape from 2d motion," in Advances in Neural Information Processing Systems, 2003, p. None.
4. T. Matsuyama, X. Wu, T. Takai, and T. Wada, "Real-time dynamic 3D object shape reconstruction and high-fidelity texture mapping for 3D video," Circuits and Systems for Video Technology, IEEE Transactions on , vol. 14, no. 3, pp. 357-369, 2004.
5. D. C. Schneider, "Shape from silhouette," Computer Vision: A Reference Guide, pp. 725-726, 2014.
6. Lorensen, William E., and Harvey E. Cline. "Marching cubes: A high resolution 3D surface construction algorithm." In ACM siggraph computer graphics, vol. 21, no. 4, pp. 163-169. ACM, 1987.
7. Fredriksson, Linus. "Evaluation of 3D Reconstructing Based on Visual Hull Algorithms." (2011).
8. Laurentini, Aldo. "The visual hull concept for silhouette-based image understanding." Pattern Analysis and Machine Intelligence, IEEE Transactions on 16, no. 2 (1994): 150-162.

