© 2017 IJEDR | Volume 5, Issue 2 | ISSN: 2321-9939

Recent Trends in Image 2D to 3D: Monocular Depth Cues

[1]Disha Mohini Pathak,[2]Tanya Mathur
[1]Assistant Professor,[2] Assistant Professor
[1]Computer Science Department,
[1]KIET, Ghaziabad,India

***Abstract*—With the growing interest in 3D hardware counterparts the dependence of 3D contents over its 2D characteristics. Many researchers have worked upon different methods to bridge this gap. This paper addresses various techniques and recent trends in Image 2D to 3D conversion. We have conferred motion parallax, Kinetic Depth Effect, Ambient occlusion, Image Blur, Shading concepts based on several aspects and considerations. The strength and limitations of these algorithms are also discussed to give a broader review on which technique to be used in different cases**
***IndexTerms*—2D to 3D image, Monocular Depth cues**

I. INTRODUCTION

2D to 3D image conversion is the process of transforming an image from 2D form (x,y) to 3D form (x.y.z) by adding another dimension i.e depth(z value) . It can be done basically by two methods: mono and stereo , in mono we take one 2D image as input and in stereo we take two images (like human eye),for mono it is the process of creating imagery for each eye from one 2D image as shown in figure 1.

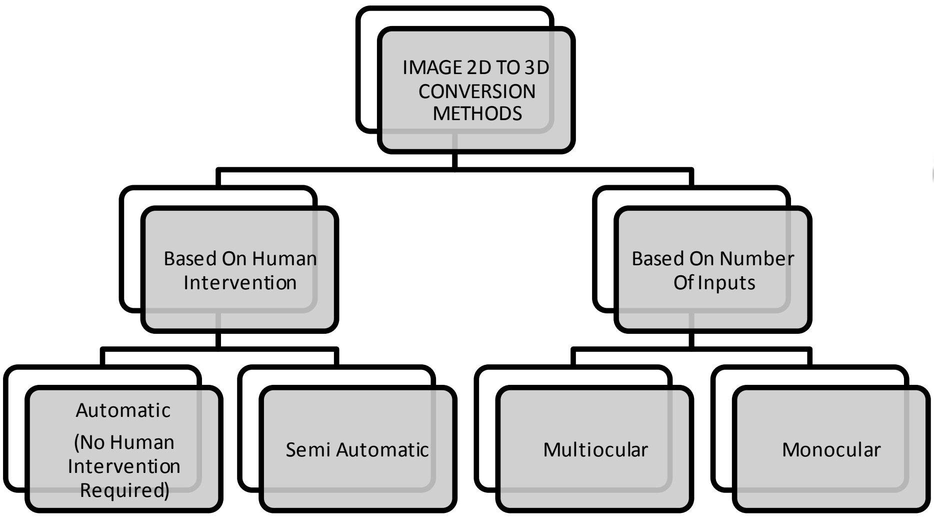

Fig 1: Categorisation of conversion methods

3D equipment has entered in our live, such as 3D display, stereoscopic capture, games and so on. The perception of 3D images is due to the parallax between the viewer's two eyes. Therefore, traditional stereo vision generation requires at least two images with slightly different projections. However, most existing digital photos were captured from monocular 2D format, and lack the corresponding depth maps to generate the perception of 3D image. Thus, generating a depth map from a single monocular image becomes an important issue for 2D-to-3D conversion.[1]
The conversion methods are also categorized into automatic method and semi-automatic method. In automatic method human intervention is not required, but in semi-automatic method human operator is involved. Computational time and design cost are the metrics that should be considered while designing these algorithms.
Depending on the number of input images, we can categorize the existing 2D to 3D conversion algorithms into two groups: algorithms based on two or more images and algorithms based on a single still image. In the first case, the two or more input images could be taken either by multiple fixed cameras located at different viewing angles or by a single camera with moving objects in the scenes. We call the depth cues used by the first group 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. [2]

II. ISSUES

Without respect to particular algorithms, all conversion workflows should solve the following tasks:[3][5]

1. Allocation of "depth budget" – defining the range of permitted disparity or depth, what depth value corresponds to the screen position (so-called "convergence point" position), the permitted distance ranges for out-of-the-screen effects and behind-the-screen background objects. If an object in stereo pair is in exactly the same spot for both eyes, then it will appear on the screen surface and it will be in zero parallax. Objects in front of the screen are said to be in negative parallax, and background imagery behind the screen is in positive parallax. There are the corresponding negative or positive offsets in object positions for left and right eye images.
2. Control of comfortable disparity depending on scene type and motion – too much parallax or conflicting depth cues may cause eye-strain and nausea effects
3. Filling of uncovered areas – left or right view images show a scene from a different angle and parts of objects or entire objects covered by the foreground in the original 2D image should become visible in a stereo pair. Sometimes the background surfaces are known or can be estimated, so they should be used for filling uncovered areas. Otherwise the unknown areas must be filled in by an artist or inpainted, since the exact reconstruction is not possible.

III. CONVERSION TECHNIQUES

Motion parallax

One of the strongest monocular depth cues is motion parallax. It arises when the location of features at different depths results in different retinal velocities. The strength of this cue is relatively high when compared to other monocular cues and also when compared to binocular disparity. This fact has been exploited in several applications, such as wiggle stereoscopy [13] where motion parallax is used as a metaphor for stereoscopic images, or parallax scrolling [12] used in games where, by moving foreground and background at different speeds, a depth sensation is evoked. Striking examples of motion parallax efficiency are species that introduce subtle head movements to enable motion parallax [15]. This mechanism has been incorporated into cameras where apparent depth is enhanced by subtle motion of the sensor [16]. Interestingly, motion parallax is not limited to observer motion, but also provides depth information whenever local motion in the scene follows a predictable transformation [17] (Fig. 2). These facts suggest that motion parallax is a very strong source of depth information for the human visual system (HVS), but it has never been explored in the context of stereoscopic image manipulations.

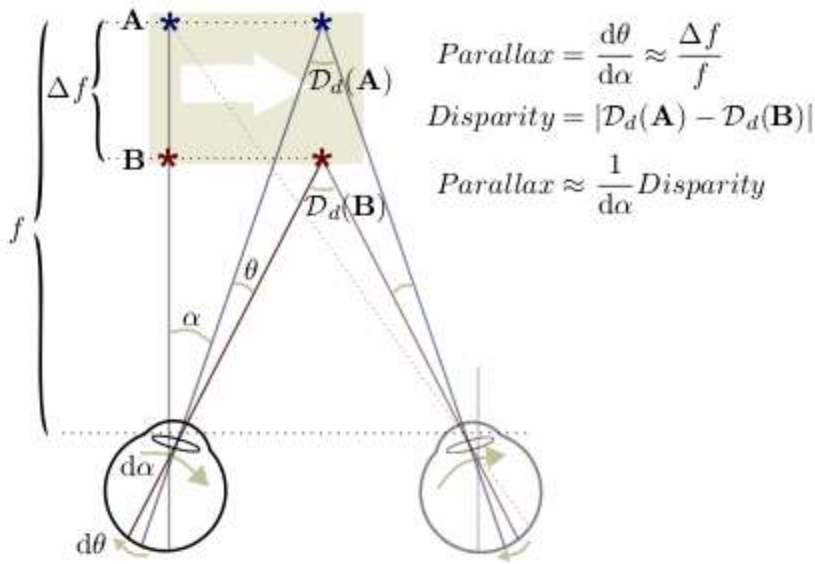

Fig 2: Motion Parallax Mechanics [5]

Perceived depth from motion parallax while processing of depth from disparity and motion parallax seem to engage common neural mechanisms [18][19], there are also notable differences between them. The HVS sensitivity to motion parallax as a depth cue has been measured in the equivalent disparity experiment where static sinusoidal depth corrugations have been used as the reference for perceived depth [18][19][20]. The respective sensitivity function has a shape similar to the disparity sensitivity function (DSF) [21] with a maximum around 0.4 cpd; however, the threshold magnitudes are 2 to 10 times higher than for disparity. [5]

Advantages: It provides strong depth cue and freely reproduce on 2D screen without any limit.

Kinetic Depth Effect

An object structure can be recovered from its rotating motion; this is referred to as the kinetic depth effect (KDE) [22]. Durgin et al. [1995] demon-stated that, as for motion parallax, the recovery of quantitative depth information from object rotation in monoscopic images is weaker than from binocular disparity. Recently, Bista et al. [23] proposed an approach for KDE triggering from a pair of photographs, where first a rough scene geometry (mesh) is reconstructed, then an optimal location for the scene rotation axis that is parallel to the screen is derived, and finally a swinging rotation of the camera around this axis is introduced. In this work we consider motion parallax, which is inherent for an animation, rather than artificially generated rotations as in [23]. Also, we focus on the binocular vision and disparity manipulation, rather than monocular images and mesh deformation, to bring the relative velocity between rotating scene elements into desirable ranges as proposed by Bista et al. [23][5]

In visual perception, the kinetic depth effect refers to the phenomenon whereby the three-dimensional structural form of an object can be perceived when the object is moving. In the absence of other visual depth cues, this might be the only perception mechanism available to infer the object's shape. Being able to identify a structure from a motion stimulus through the human visual system was shown by [22] in the 1950s through their experiments.

There are two propositions as to how three-dimensional images are perceived. The experience of three-dimensional images can be caused by differences in the pattern of stimulation on the retina, in comparison to two-dimensional images. Gestalt psychologists hold the view that rules of organization must exist in accordance to the retinal projections of 3-D forms which happen to form three-dimensional precepts. Most retinal images of two-dimensional forms lead to 2-dimensional forms in experience as well. The other deduction is related to previous experience. Unfortunately, this assumption does not explain how past experience influences perception of images.[22]

In order to model the calculation of depth values from relative movement, many efforts have been made to infer these values using other information like geometry and measurements of objects and their positions.[24] This is related to the extraction of structure from motion in computer vision. In addition, an individual's ability to realize the kinetic depth effect conclusively shows that the visual system can independently figure the structure from motion problem.[25]

As with other depth cues, the kinetic depth effect is almost always produced in combination with other effects, most notably the motion parallax effect. For instance, the rotating circles illusion[12] and the rotating dots visualization[26] (which is similar in principle to the projected wireframe demonstration mentioned above) rely strongly on the previous knowledge that objects (or parts thereof) further from the observer appear to move more slowly than those that are closer.

The kinetic depth effect can manifest independently, however, even when motion parallax is not present. [5][6]

Advantages: The kinetic depth effect can manifest independently, however, even when motion parallax is not present.

Disadvantages: The kinetic depth effect is almost always produced in combination with other effects, most notably the motion parallax effect.[12]

Shading

In natural perception, the human visual system combines several depth cues in order to create a sensation of depth. These include the binocular cues of convergence and dis- parity, as well as a variety of monocular cues, such as motion parallax, interposition, occlusion, size, perspective, accommodation, and others. Stereoscopic 3D (S3D) as enabled by stereo 3D displays, supports binocular cues and from that creates a sensation of depth. Presenting a separate view to each eye does not, however, perfectly emulate real life depth perception so limitations such as the vergence accommodation conflict [SKHB11] and other imperfections remain problematic. Furthermore, all 3D displays are limited to a feasible depth volume.

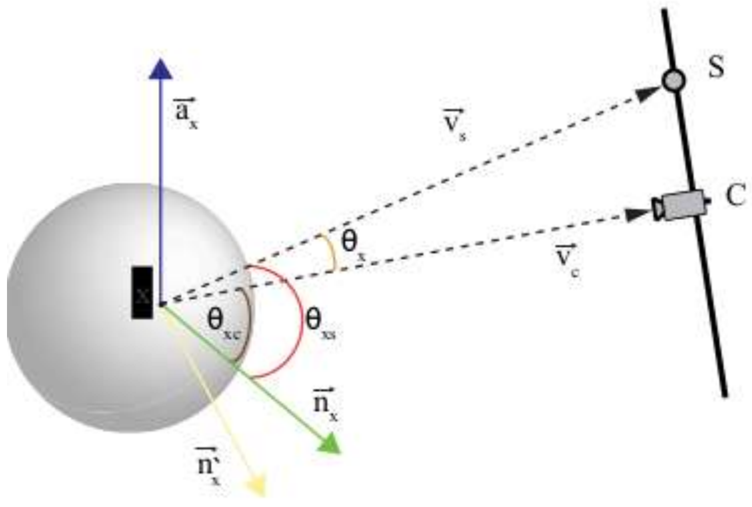

Fig. 3 Shading stereo: the normal of x is adjusted to match the angle of a reference point.

In particular, auto stereoscopic displays are still very restricted **in** the depth range they can reproduce without artifacts due to limited angular resolution. Therefore, enhancement of S3D perception has been a very active field of recent research, which has been focused mainly on disparity enhancement The goal of our new method, shading stereo, is to generate an increase in the S3D sensation by leveraging differences in shading between views, with cast shadows left unchanged between renders.[7]

Advantages: Polygons, more complex than triangles, can also have different colors specified for each vertex. In these instances, the underlying logic for shading can become more intricate. [10]

Disadvantages:

Even the smoothness introduced by Gouraud shading may not prevent the appearance of the shading differences between adjacent polygons.

Gouraud shading is more CPU intensive and can become a problem when rendering real time environments with many polygons.

T-Junctions with adjoining polygons can sometimes result in visual anomalies. In general, T-Junctions should be avoided.[10]

Image Blur

Psychophysical evidence for the use of image blur in depth perception has recently been reported by Marshall et al (1996) and by Mather (1996). Both papers described experiments on ambiguous figure-ground stimuli, containing two regions of texture separated by a wavy boundary. When texture in one region was blurred while texture in the other region was sharp, the two regions appeared to be at different depths. When the boundary itself was sharp, the sharp texture was seen as nearer, and occluding the blurred texture region, but when the boundary was blurred, the blurred texture was seen as nearer. Information available in image blur. The following general expression relates the distance d of a point from a lens to the radius s of its blurred image (Pentland 1987):

$$d = Frv/[rv - F(r + s)] \quad (1)$$

where F is focal length, r is lens aperture radius, and v is the distance of the image plane from the lens. Note that this expression does not distinguish between points that are farther and nearer than the point of focus. In principle, if the values of F, r, and v are known and a measure of image blur s is available, then absolute distance can be calculated (provided that the nearer/farther ambiguity is resolved).

Eq. (1) can be used to predict retinal blur as a function of distance, on assuming typical values for the optical parameters of the human eye (r = 1.5 mm, v = 16 mm).[8]

Ambient occlusion

Ambient occlusion is related to accessibility shading, which determines appearance based on how easy it is for a surface to be touched by various elements (e.g., dirt, light, etc.). It has been popularized in production animation due to its relative simplicity and efficiency. In the industry, ambient occlusion is often referred to as "sky light".[citation needed.The ambient occlusion shading model has the nice property of offering a better perception of the 3D shape of the displayed objects. This was shown in a paper where the authors report the results of perceptual experiments showing that depth discrimination under diffuse uniform sky lighting is superior to that predicted by a direct lighting model.[9]

Advantages:

Independent from scene complexity.

No data pre-processing needed, no loading time and no memory allocations in system memory.

Works with dynamic scenes.

Works in the same consistent way for every pixel on the screen.

No CPU usage – it can be executed completely on the GPU.

May be easily integrated into any modern graphics pipeline.[11]

Disadvantages:

Rather local and in many cases view-dependent, as it is dependent on adjacent Texel depths which may be generated by any geometry whatsoever.

Hard to correctly smooth/blur out the noise without interfering with depth discontinuities, such as object edges (the occlusion should not "bleed" onto objects).[11]

REFERENCES

[1] 2D-TO-3D CONVERSION FOR SINGLE-VIEW IMAGE BASED ON CAMERA PROJECTION MODEL AND DARK CHANNEL MODEL Tien-Ying Kuo, Yi-Chung Lo, and Chia-Chin Lin Department of Electrical Engineering, National Taipei University of Technology, Taiwan, R.O.C. tykuo@ee.ntut.edu.tw, {yclo, cclin}@image.ee.ntut.edu.tw

[2] A Real Time 2D to 3D Image Conversion Techniques Miroslav Nikolov Galabov

[3] Scott Squires. 2D to 3D Conversions

[4] Jon Karafin. State-of-the-Art 2D to 3D Conversion and Stereo VFX International 3D Society University. Presentation from the October 21, 2011 3DU-Japan event in Tokyo.

[5] Motion Parallax in Stereo 3D: Model and Applications: Petr Kellnhofer 1Piotr Didyk 1,2Tobias Ritschel 1,2,3Belen Masia 1,4Karol Myszkowski 1Hans-Peter Seidel 1

[6]WIKIPEDIA,2017.Kinetic depth effect—Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Kinetic_depth_effect

[7] Stereo from Shading:Alexandre Chapiro 1,2 Carol O'Sullivan 3 Wojciech Jarosz 2 Markus Gross 1,2 Aljoscha Smolic 2
[8] The use of image blur as a depth cue: George Mather
[9]WIKIPEDIA,2017. Ambient Occlusion—Wikipedia, the free encyclopedia.https://en.wikipedia.org/wiki/Ambient_occlusion
[10]WIKIPEDIA,2017. Shading -Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Shading
[11] WIKIPEDIA, 2017 . Screen Space Ambient Occlusion- Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Screen_space_ambient_occlusion
[12] StereoKinetic phenomenon from Michael Bach's "Optical Illusions & Visual Phenomena"
[13] WIKIPEDIA , 2017. Parallax scrolling—Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Parallaxscrolling.
[14] WIKIPEDIA , 2017. Wiggle stereoscopy—Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Wigglestereoscopy.
[15] KRAL , K. 2003. Behavioural–analytical studies of the role of head movements in depth perception in insects, birds and mammals.Behavioural Processes 64, 1, 1–12.
[16] P ROFFITT , D., AND B ANTON , T. 1999. Perceived depth is enhanced with parallax scanning. Tech. rep., University of Virginia-Cognitive Science Department.
[17] U LLMAN , S. 1983. Maximizing rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion. Perception 13,255–74.
[18] ROGERS , B., AND G RAHAM , M. 1982. Similarities between motion parallax and stereopsis in human depth perception. Vision Research 22, 2, 261–270.
[19] B RADSHAW , M. F., H IBBARD , P. B., P ARTON , A. D., R OSE , D., AND L ANGLEY , K. 2006. Surface orientation, modulation frequency and the detection and perception of depth defined by binocular disparity and motion parallax. Vision Research 46, 17, 2636–2644.
[20] B RADSHAW , M. F., AND ROGERS , B. J. 1996. The interaction of binocular disparity and motion parallax in the computation of depth. Vision Research 36, 21, 3457–3468.
[21] B RADSHAW , M. F., AND ROGERS , B. J. 1999. Sensitivity to horizontal and vertical corrugations defined by binocular disparity. Vision Research 39, 18, 3049–56.
[22] W ALLACH , H., AND O'C ONNELL , D. N. 1953. The kinetic depth effect. Journal of Experimental Psychology
[23] B ISTA , S., DA C UNHA , ´I. L. L., AND V ARSHNEY , A. 2016. Kinetic depth images: Flexible generation of depth perception. The Visual Computer, 1–13.
[24] Michael S. Landy (1987). "Parallel model of the kinetic depth effect using local computations" (PDF). J. Opt. Soc. Am. A. 4 (5): 864–877. doi:10.1364/josaa.4.000864.
[25] Ullman, S. (1979). The interpretation of visual motion. Cambridge, MA: MIT Press.
[26] George Mather. "Kinetic Depth Effect illustration".