International Journal of Research in Advent Technology, Vol.2, No.12, December 2014
E-ISSN: 2321-9637

A Survey on 2D to 3D Image and Video Conversion Techniques

Divya K.P., Sneha Arun.
Computer Science, MTech
Email: divyakrishnadas90@gmail.com , snehanair88@gmail.com

Abstract-Despite of the significant growths in the past few years, the availability of 3D content is still dwarfed to that of its 2D counterpart. In order to fill this gap many 2D to 3D image and video conversion techniques are proposed. The sense of depth in a scene captured by the conventional cameras can now be experienced where virtually no depth was ever provided. This sense of depth is used in many applications ranging from training simulation, gaming, scientific explorations and cinema, to get a sense of realism. Thus the users can really appreciate the content they are viewing. In this paper a detailed survey on the existing 2D to 3D image and video conversion techniques is done.

Index Terms- 3D images; stereoscopic images; depth; disparity.

1. INTRODUCTION

Viewing 3D content has become more popular over the last few years. In the past, much research focused on obtaining a sense of depth of a scene by using only two views, corresponding to the left and right eye viewpoints, known as stereo correspondence. Using a reference viewpoint, whether it is the left or right eye, a sense of depth is determined for each point in the reference, where points in one image are matched with their corresponding points in the other image. The amount of shift that the corresponding points undergo is known as disparity. Disparity and depth have an inverse relationship, where by the greater the horizontal shift, or disparity, the lesser will be the depth, and the closer the point is to the viewer.

By presenting these two views of a scene to a human observer, where one is slightly offset from the other, this produces the illusion of depth in an image, thus perceiving an image as 3D. The work is primarily done by the visual cortex in the brain, which processes the left and right views presented to the respective eyes. In the past, the disparities of each point between both viewpoints are collected into one coherent result, commonly known as a disparity map. The image is monochromatic in nature, and the brightness of each point is directly proportional to the disparity between the corresponding points. When a point is lighter, this corresponds to a greater disparity, and is closer to the viewer. Similarly, when a point is darker, this corresponds to a lesser disparity, and is farther from the viewer. For scenes that have lesser disparities overall, the disparity maps are scaled accordingly, so that white corresponds to the highest disparity.

However, filming directly in 3D can be rather expensive and difficult to set up. So the one we are ultimately interested in, is to take already existing 2D content, and artificially produce the left and right views, or other views for multiview applications.

Although the amount of material is endless, but the methods required to convert monocular video to 3D is a very difficult problem to solve, and still one of the most open-ended problems in computer vision that exist today. When a camera captures a scene, it makes a transformation from 3D to 2D coordinates, which is well understood. However, when considering the reverse situation, which is the one we are interested in, we are trying to produce information from a model that lost the very information we are trying to obtain. The goal for the conversion of monocular video to its stereoscopic or multiview counterpart is to generate a set of disparity maps for each view that we wish to render. Each disparity map will determine the shift in pixels we need from the reference view, or a frame from the video sequence. The caveat here is that we must generate disparity maps for each frame of each view in the video sequence. Regardless, 2D to 3D conversion algorithms have been the subject of many useful applications which are targeted for the industry, as well as the end user. Specifically, 2D to 3D conversion techniques are of primary consideration in the latest 3DTVs. This is due to the fact that 3D technology is now widely available for the home, and they can now experience their existing 2D content in 3D.

2. STATE OF THE ART

There are two styles of 2D-to-3D image conversion strategies: semi-automatic methods, that need human operator intervention, and automatic strategies, that need no such facilitate.

2.1. *Semi Automatic Methods*

Most semiautomatic ways of stereo conversion use depth maps and depth-image-based rendering.

The concept is that a separate auxiliary image referred to as the "depth map" is made for every frame or for a series of homogenized frames to point depths of objects gift within the scene. The depth map may be a separate grayscale image having a similar dimensions because the original second image, with varied reminder grey to point the depth of each a part of the frame. whereas depth mapping will manufacture a reasonably potent illusion of 3D objects within the video, it inherently doesn't support semi-transparent objects or areas, nor will it permit specific use of occlusion, therefore these and different similar problems ought to be treated via a separate technique.

A development on depth mapping, multi-layering works round the limitations of depth mapping by introducing many layers of grayscale depth masks to implement restricted semi-transparency. just like an easy technique, multi-layering involves applying a depth map to over one "slice" of the flat image, leading to a way higher approximation of depth and protrusion. The additional layers square measure processed one by one per frame, the upper the standard of 3D illusion tends to be.

3D reconstruction and re-projection is also used for stereo conversion. It involves scene 3D model creation, extraction of original image surfaces as textures for 3D objects and, finally, rendering the 3D scene from 2 virtual cameras to accumulate stereo video. The approach works tolerably just in case of scenes with static rigid objects like urban shots with buildings, interior shots, however has issues with non-rigid bodies and soft fuzzy edges.

Another technique is to line up each left and right virtual cameras, each offset from the initial camera however rending the offset distinction, then painting out occlusion edges of isolated objects and characters. primarily clean-plating many background, middle ground and foreground components. Binocular inequality can even be derived from straightforward pure mathematics.

2.2. *Automatic Methods*

It is attainable to mechanically estimate depth exploitation differing kinds of motion. just in case of camera motion depth map of the whole scene may be calculated. Also, object motion may be detected and moving areas may be assigned with smaller depth values than the background. Besides, occlusions give info on relative position of moving surfaces.

On "depth from defocus" (DFD) approaches, the depth info is calculable supported the number of blur of the thought of object, whereas "depth from focus" (DFF) approaches tend to match the sharpness of associate degree object over a variety of pictures soft on completely different focus distances so as to search out its distance to the camera. DFD solely desires a pair of to three pictures at completely different focus to properly work, whereas DFF desires ten to15 pictures a minimum of however is additional correct than the previous methodology.

If the sky is detected within the processed image, it also can be taken into consideration that additional distant objects, besides being hazy, ought to be additional desaturated and additional blue attributable to a thick air layer.

The idea of depth from perspective relies on the very fact that parallel lines, like railroad tracks and roadsides, seem to converge with distance, eventually reaching a vanishing purpose at the horizon. Finding this vanishing purpose offers the farthest purpose of the entire image.

The conversion may be rotten into completely different classes, and that we can discuss the relevant and up to date strategies for every of those classes. We'll discuss these strategies intimately; citing works that square measure recent inside the previous few years. The classes that 2D to 3D conversion may be rotten into the following: exploiting color info, edge info, techniques from depth-based image rendering (DIBR), motion, and analyzing scene options. However, the 2 most predominant techniques measure motion estimation and analyzing scene options. to create this literature review additional coherent, we'll mix the discussion of color and DIBR into the analysis of scene options, because it technically falls into this class. the remainder of this literature review are formatted as follows. Section 3.1 discusses the relevant and recent 2D to 3D methods by using motion estimation as the framework. Section 3.2 discusses the 2D to 3D methods using scene analysis, or other features extracted from the scene. Finally, brief summary of the methods in each category, the disadvantages and drawbacks that can be encountered, leading to what can be formulated for future research in this area is discussed.

3. 2D TO 3D CONVERSION TECHNIQUES

3.1. *Using Motion*

The first of the foremost predominant techniques that's used for 2D to 3D conversion is victimization motion estimation to work out the depth or inequality

of the scene. The underlying mechanism is that for objects that area unit nearer to the camera, they ought to move quicker, whereas for objects that area unit way, the motion ought to be slower. Thus, motion estimation will be wont to verify correspondence between 2 consecutive frames, and may therefore be wont to verify what the acceptable shifts of pixels area unit from the reference read (current frame), to the target read (next frame). However, the utilization of the particular motion vectors themselves to get a stereoscopic or multiview video sequence varies between the ways, however the underlying mechanism is that the same. we are going to currently begin our discussion on the utilization of motion estimation for second to 3D conversion.

We begin with the strategy by Ideses et al. [1], wherever they target a period of time conversion from second to 3D. This work is arguably one in all the ways that has established that victimization motion estimation is one in all the key options to use once determinative the depth map. Primarily, they use motion vectors from the MPEG4decoder. They use the magnitude of the motion vector, by taking the euclidian distance of the horizontal and vertical motion estimation elements for every element. For the video sequence in question, every frame is rotten into its RGB elements. In parallel, the magnitude of the motion vector is set. For show, anaglyph pictures area unit created, wherever the red channel of every frame is employed, and therefore the motion vector magnitudes area unit wont to shift pixels to form the proper image. The initial image and changed red channel area unit therefore incorporated to form the anaglyph image for every frame.

In a similar fashion, Huang et al. [2] confirm depth maps by using motion and scene depth. Specifically, the motion vectors from the H.264 decoder are used to generate a motion-based depth map. Additionally, a moving object detection formula is introduced, to diminish the block impact caused by the motion estimation in H.264. Using geometry of the scene, they extract vanishing lines and vanishing points by playacting edge detection and also the Hough remodel. With the Hough transform, and the location of the vanishing lines, the location of points with respect to the vanishing lines—which can be transformed to vanishing planes—can be used to assign a depth value that is dependent on the location of the pixel point with respect to the vanishing lines or planes. With the mixture of those 2 depth maps, a hybrid fusion is performed to merge the 2 depth maps along, and so a depth map is formed that handles each scene geometry, additionally to the pertinent objects shown within the scene. Figure 2 shows the overall diagram of this approach.

3.2. *Using Scene Features*

The second most well liked methodology for changing 2D monocular video sequences into 3D is through analyzing the options of the scene of interest. Features like shape, edges, color, or something involving the direct use of features aside from motion area unit of interest here.

3.2.1. Color

There are attempts to use color data directly for determining depth maps from monocular video sequences. The foremost well-known one recently comes from Tam et al. [3]. Primarily, they decompose the video sequence into the YCbCr color area, and solely use the Cr color channel as a live of depth. The reasoning is that totally objects have different hues, associated so the Cr color channel provides an approximate segmentation of the objects, which are characterised by completely different intensities within the Cr component of the image. Every intensity ought to belong to roughly similar objects, and so a similar depth worth. However, some tweaking is needed, specifically once there's blue and green, as a blue contains a lighter hue than a green tree, which might mean that the sky is nearer to the viewer. There's some postprocessing that's concerned here, however it's awfully a crude approach to determinative a depth map.

3.2.2. Shape and Texture

Interestingly, some approaches to 2D to 3D conversion use form and texture options to come up with the stereoscopic content. One methodology to try to is that the one by Feng et al. [4]. This is often not a real 2D to 3D conversion, however it is of a motivating nature to notice, and that we shall cowl this explicit methodology in additional detail. Specifically, there ought to be some stereoscopic pairs obtainable throughout the video sequence, and therefore the disparities for the remaining monocular sequence are calculated. Three features are extracted using shape. They are major axis, minor axis and therefore the center of mass of the object. This is often primarily performing arts principal part analysis (PCA), moreover as determinative the primary moment. After, dynamic programming is employed to estimate the inequality map between those stereoscopic pairs and therefore the disparities

propagated to the 2D via the Hausdorff distance using shape features.

3.2.3. Edges

In the work by Cheng et al[5], a block based algorithm together with the edges of the image is used to group pixels of regions together. Then the depth for each pixel is generated by a depth from prior hypothesis method. In order to smooth the boundaries bilateral filtering is used.

3.2.4. Miscellaneous Methods

This section details those algorithms that uses the scene features but not come under the categories discussed above.

In the method performed by Konrad et al. [35] the framework entails a machine learning algorithm, in order to determine what the depth values in the video sequence are. They developed two types of methods. The first one is based on learning a point mapping from local image/video attributes, such as color, spatial position, and motion at each pixel, to scene-depth at that pixel using a regression type idea.

The second one globally estimates the entire depth map of a query image directly from a repository of 3D images(image+depth pairs or stereopairs) using a nearest-neighbor searching idea.

4. CONCLUSION

In this literature review, the most recent undertakings in converting monocular video footage to their stereoscopic or multiview counterparts for display on 3D visualization technology are discussed. It is concentrated on the various recent methods to achieve this goal over the last few years, as research into this area has surfaced in higher volumes in this period. As previously mentioned, methods to convert from 2D to 3D can be further subdivided into two major categories: using motion and analyzing scene features.

Nevertheless, with the plethora of different research in this area that has surfaced, it has been demonstrated that it is quite an important topic for realizing a sense of depth in video sequences obtained by only a single camera, and is very useful in many aspects of the industry and to the end user. It is also important, as this is an alternative solution to producing 3D content to alleviate the lack of 3D content that exists, and to also provide a more cost-effective solution.

Acknowledgments

Every success stands as a testimony not only to the hardship but also to hearts behind it. Likewise, the present work has been undertaken and completed with direct and indirect help from many people and I would like to acknowledge all of them for the same.

REFERENCES

[1] I. Ideses, L. P. Yaroslavsky, and B. Fishbain, Real-time 2D to 3D video conversion, *J. Real-Time Image Processing*, 2, 3–9, 2007.

[2] X. Huang, L.Wang, J. Huang, D. Li, and M. Zhang, Adepth extraction method based on motion and geometry for 2D to 3D conversion, *Proc. IEEE Symp. on Intelligent Information Technology Applications*, Nanchang, China, 2009.

[3] W. J. Tam, C. Vazquez, and F. Speranza, Three-dimensional TV: A novel method for generating surrogate depth maps using colour information, *Proc. SPIE Electronic Imaging—Stereoscopic Displays and Applications XX*, Vol. 7237, pp. 72371A-1–72371A-9, San Jose, California, 2009.

[4] Y. Feng, J. Jiang, and S. S. Ipson, A shape-match based algorithm for pseudo-3D conversion of 2D videos, *Proc. IEEE Conf. on Image Processing (ICIP)*, Vol. 3, pp. 808–811, Genoa, Italy, 2005.

[5] C.-C. Cheng, C.-T. Li, and L.-G. Chen, A 2D-to-3D conversion system using edge information, *Proc. IEEE Conf. on Consumer Electronics*, pp. 377–378, Las Vegas, Nevada, 2009.

[6] J.Konrad, M. Wang, P. Ishwar, C. Wu and D. Mukherjee, Learning-Based, Automatic 2D to3D Image and Video Conversion, *IEEE Trans. Image Processing*, Vol.22, No.9. pp. 3485-3496, Sep 2013