I J C T A, 9(19) 2016, pp. 9231-9238
© International Science Press

A Review on 2d to 3d Image Conversion Techniques

Jagriti Mungali[*], Anand Prakash Shukla[*] and Swati Chauhan[*]

ABSTRACT

In this paper, a technical survey on various techniques of 2D to 3D image conversion has been performed. The survey includes automatic and semiautomatic method of conversion The automatic method of conversion include conversion by learning depth, 2D to 3D image and video conversion using learning technology, conversion algorithm using multi depth cues, depth extraction from video. The semiautomatic method of conversion include semiautomatic stereo extraction from video footage, depth map generation using unconstrained images and video sequence, 2D to 3D conversion combining motion analysis with user interaction, conversion based on RGB-D images. The techniques which are compared in this paper are categorized into monocular and multicolor depth cues. Monocular depth cues include algorithms based on a single image. Multicolor depth cue include algorithms based on two or more images.The performances of various techniques are compared on the basis of some qualitative sets.

Keywords: automatic method, semiautomatic method, monocular, multicolor, depth cue

I. INTRODUCTION

Only two dimensions are present in a 2D image they are width and height. 2D images does not have depth, but 3D images have the additional depth field. 3D display requires depth information which is not available in conventional 2D content. Various 3D applications have evolved therefore it is very important to convert 2D images to 3D images.2D images are monochrome images. In 2D to 3D conversion the monochrome images are converted to digitized images. 2D is a flat image using horizontal and vertical (x and y) dimensions. A digitized image is one where: Descretization of spatial and grayscale values is supposed to be done. x and y directions are used to measure intensity across a regularly spaced grid. Intensities sampled to 8 bits (256 values).

We can assume a digital image as a 2D array where x and y index an image point. This is done for estimation purposes. Single element in the array is called a pixel. Stereoscopy formulates the illusion of 3D depth from given 2D images. The third dimension contains depth. Binocular disparity helps in sensing third dimension in an image. Binocular disparity calculates the change in the location of an object. According to human vision an image can be seen by left and right eye. The distinction between the image location helps the brain to extract depth. Two slightly different images of every scene are created at both the eyes during the 3D view. After this by correct crafting of parameters and according to binocular disparity a correct 3D perception is observed. Most significant step in 3D images is its output. Many cameras have been designed to generate 3D model [1].

II. CONVERSION OF IMAGES FROM 2D TO 3D

A 2D to 3D conversion process consist of 2 steps: Depth estimation for a given image and then depth rendering of a query.

[*] Department of Computer Science, K.I.E.T, Ghaziabad, Uttarpradesh, India, *E-mail: jagriti.mungali@gmail.com; ap.shukla@kiet.edu; swati.chauhan@kiet.edu*

(A) Automatic Classification for 2D to 3D Conversion of Image

1) Conversion by learning depth:
   In this method a simplified algorithm learns the depth of the scene from a large repository of image and depth pairs. Their proposed method is based on observation that among millions of images and depth pairs avail- able there are many pairs available whose 3D content matches that of a 2D input. An assumption has been made that: two images have similar structure if they are photometrically same and are likely to have similar 3D structure i.e., depth.

2) 2D to 3D image and video conversion using learning technology:[8]
   Here two methods are mentioned: One 2D to 3D conversion by learning local point transformation.
   2D to 3D conversion based on global nearest neighbor depth learning.
   Images or videos having attributes at a pixel level that is learned by a point transformation. When the point transformation is learned it is applied to a monocular image. Depth is assigned to a pixel based on its attribute. In this algorithm a key element is a point transformation which is used for computation of depth from image attributes.

3) Conversion Algorithm using multi depth cues:
   Here 3 depth generation procedures are used. They are perspective geometry defocus and visual saliency with adaptive depth models. In this first color image is converted into grayscale image. Then the vanishing point is detected using Canny edge detection. Then lines in the images and intersections are calculated between the detected lines by using Hough transformation.

4) Depth extraction from video:
   The method presented is to automatically convert a monoscopic video into stereoscopic for 3D visualization. They have shown a 729 framework using temporary information for improved and time coherent depth when multiple frames are available [3]

(B) Semiautomatic Classification for 2D To 3D Conversion of Image

Guttmann proposed [12] diffusion scheme based on semiautomatic method of conversion to convert a conventional video into stereoscopic video. Phan[13] proposed more efficient and better method. This method used scale spaced random walks and graph based depth prior that solve using graph cut. Lio[14] simplified involvement of operator by calculating optical flow and after applying structure from motion estimation. In this technique finally moving object boundaries are extracted.

There are four different techniques for semiautomatic 2D to 3D conversion[3]:

1) Extraction From Video Footage Using Semiautomatic Conversion: Here in this system they have elected to estimate disparities directly without estimating depth. Depth is linked to disparity through limited amount of parameters. Depth can range between zero and infinity and is inversely proportional to disparity.

2) Depth Map Generation Using Unconstrained Images and Video Sequence : [9] In this system a depth map is produced that can be used to create stereoscopic 3D image pair. The method constitutes of 2 stages process using the smoothing properties of random walks and the hard segmentation returned by graph cut. The solution to a linear system is random walks. It has problems preserving strong edges but graph cuts are better.

3) 2D to 3D Conversion Combining Motion Analysis With User Interaction: It is a semiautomatic system which converts a conventional video into stereoscopic videos by combining motion analysis with user interaction which transfers possible labeling work from the user to the computer. The user interface design benefits from the already defined 3D cues by the pre processing of movement [4].

4) Conversion Based on RGB-D Images: This contracted a depth image optimization method to correct the quality and minimize errors and noise. The input is given as RGB-D image which serve color information and original depth information. They examined a surface normal which represents object's surface shape information. The surface normal is considered useful for construction of depth information. The first surface normal is retrieved by computing a normal vector for every pixel after the original depth map is pre processed. Consistency of normal vector with RGB value is combined and surface normal is segmented using mean shift algorithm.

III. 2D TO 3D CONVERSION ALGORITHM

The existing conversion algorithms are based on 2 groups:

- Algorithms based on two or more images
- Algorithms based on a single image.

In the first 2 cases the two or more input images could be taken by multiple fixed camera located at different viewing angles or using moving objects in the scene by a single camera. Multioculor depth cue is used by the first group. The second group of depth cues operates on a single still image. This is called monocular depth cues. Table 1 summarizes the categories of depth cue.

Table 1
Categorization of Depth Cues[2]

Two Or More Images(binocular Or Multicolor)	*One Single Image(monocular)*
Binocular disparity	Defocus
Motion	Linear Perspective
Defocus	Atmospheric scattering
Focus	Patterned Texture
	Symmetric Patterns
Silhouette	Occlusions
	Statistical Pattern

1) BINOCULAR DISPARITY: The difference in image location of an object checked by left and right eye is binocular disparity. Human brain uses this method to extract depth information from 2D retinal images. This is often used to perceive depth. The main implementation for depth perception is to capture two images of same scene from slightly disparate viewpoint. Binocular disparity is used to observe the depth of the object. In the below figure 4 Pl , Pr are the projections of 'P' on left image and right image. Ol , Or are the origin of camera coordinate system of the left and right cameras.(P, Pl , Pr) and (P, Ol , Or) are the similar triangles. The depth value 'Z' can be obtained as: Z = f " T /d. 'f' is the focal length and 'T' is the real size of the bar which is perpendicular to the optical axis.d = xr " xl shows the difference in the retinal position between image points[2].

2) MOTION: The standard illustration of 3D motion when it is projected onto a camera image is motion field. An important cue to depth perception is noticed between the viewing camera and observed scene. Near objects move faster across the retina than far objects [11]. Structure from motion is the term given to the removal of 3D structures and the motion of camera from different image is defined as structure from motion. Motion field is well understood as the projection on the image plane using 3D velocity field[5].

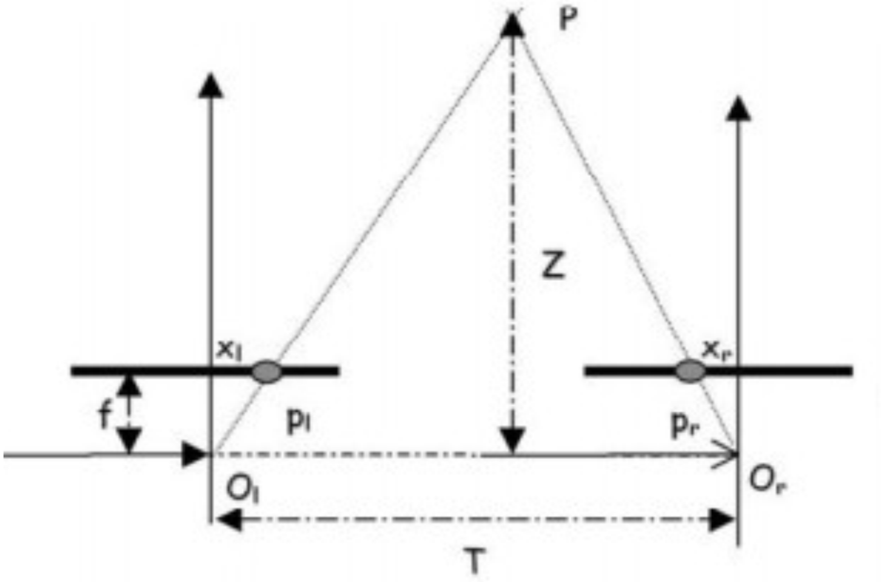


Figure 1: Triangularisation method: It is used to calculate depth information

3) DEFOCUS USING MORE THAN 2 IMAGES: The assessment of 3D surface depth from a set of two or more images is done by the techniques termed asdepth from focus and defocus. Images are taken by changing the camera parameter and taken from same viewpoint. In a thin lens system, objects that are under focus are clearly pictured and on the other hand the other objects are defocused. The defocus is caused by the convolution of the ideal projected image and the camera PSF(Point Spread Function). To simplify the system Gaussian function is used to simulate PSF.

   With the help of blur radius and cameo parameters a depth map can be generated which is based on:

$$U= f_s/(s-f-kf\sigma) \qquad \text{if } u> v$$
$$U= f_s/(s-f+kf\sigma) \qquad \text{if } u<v$$

where

- u is the depth.
- v is the distance between lens and position of perfect focus.
- S is the distance between lens and image plane.
- F is the focal length.
- K is constant

4) FOCUS: It is very similar to depth from defocus approach. The difference between the two is that depth from focus requires an array of images of the scene with different focus levels by changing the distance between the camera and the scene. On the other hand depth from defocus only needs two or more images with fixed objects and camera positions and use different camera focal settings[7].

5) SILHOUTTE: Silhouette is the darken image contour which separates the image from the background. A silhouette is a solid shape of the image of a single color usually black in color. Shape from silhouette: The 3D reconstruction procedure of any image is termed as shape from silhouette.

6) DEFOCUS USING A SINGLE IMAGE: When many images are used ambiguity is reduced in blur radius estimation where the focal setting is not known. In defocus using a single image the camera position object is fixed but different focal lengths are used. When filtering an edge of blur radius with

second derivative of Gaussian filter using variance s the response has positive and negative peak. ‘d’ denotes the distance between peaks which can be measured directly by filtered image. Blur radius is computed by the formula: [6]

$$\sigma 2 = (d/2)2 - s2$$

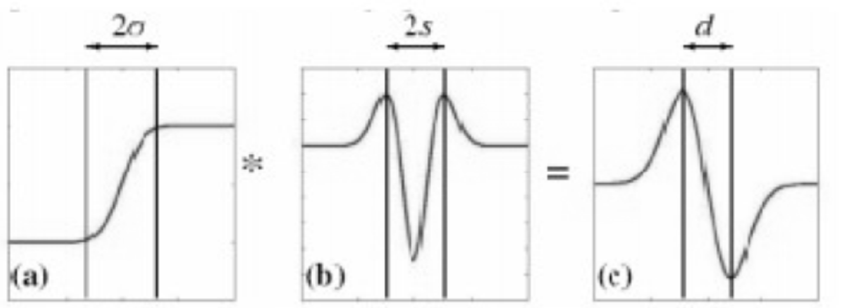
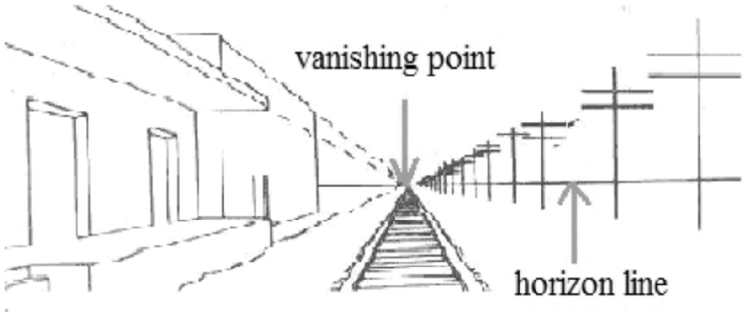


Figure 2: Blur radius estimation

In figure 2 the following terms are shown.

(a) Blurred edge (b) Second derivative of Gaussian filter with standard derivation s (c) response of the filter, the distance d between peaks can be measured from the filtered image[2].

7) LINEAR PERSPECTIVE: It refers to the fact that parallel lines like rail board tracks appear to converge with distance ultimately reaches a vanishing point at the horizon. The more the lines converge the farther away they appear to be. A recent work in this field is the gradient plane assignment approach. The method performs well for single images which contain objects of rigid and geometric appearance. Edge detection is used to check the lines in the image. Then the intersection points of these lines are observed. Where most of the intersection points converge, it is considered as the vanishing point. The lines closer to the vanishing point are the vanishing lines. The pixel closer to the vanishing point is assigned a larger depth value. The given figure shows the horizon line and the vanishing point.



8) ATMOSPHERIC SCATTERING: Atmosphere envelopes the earth. When light is propagated through the atmosphere in such a way that its direction and power is altered through diffusion of radiation by small particles. This results in atmospheric scattering and haze. The relationship between scattering and distance opens the possibility of recovering depth cues from images. A relationship is derived between radiance of an image and distance between object and viewer.

$$\tilde{C} = C_0 e^{-\beta z} + S(1 - e^{-\beta z}) \qquad (3)$$

The above equation has 2 parts. The first part describes that the light power is attenuated when a beam of light is projected onto an object through a scattering medium. The second part is the opposite method. It shows that a gain occurs in intensity due to scattering.[16]

9) SHADING: The slow variation of surface shading in the images encodes the shape information of the objects. Shape from shading technique is used to reconstruct 3D shapes from intensity images by using relationship between surface geometry and image brightness.

10) PATTERNED TEXTURE: This technique offers a good 3D impression. There are 2 key ingredients: The distortion of texels and the rate of change of texel across the texture region.

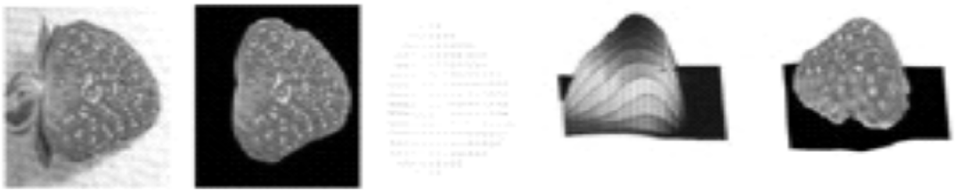


Figure 4: Shape from texture (From left to right: Original image, Segmented texture region; Surface normal; Depth map; Reconstructed 3D shape) (Fig reference 15)

In figure 4 it is shown how a simple image is reconstructed into a 3D shape. The distortion is exploited by shape reconstruction which makes the texels appear smaller. Texel is the fundamental unit of texture space. Textures are represented by arrays of texels.

11) BILATERAL SYMMETRY: Natural and manmade objects possess symmetry. If an object separates into two chiral parts then 3D symmetry exists. The plane is called symmetry plane. The line joining two symmetry points is called symmetry lines. Symmetry lines and planes are perpendicular to each other. Faces, animals, birds, leaves etc are all examples of symmetry[10].

12) OCCLUSION: It is well understood as overlapping objects. Occlusion is considered as the strongest cue for depth. It overrides all other cues when a conflict occurs[17]. The near objects occlude objects that are farther away. Transparency: in this a peak behind the occluded object is created Parallax: It is a technique which shows moving objects occluding each other. In this the information is organized in such a manner that more important information partially occludes less important information.[18]

IV. COMPARISION

The comparison of the algorithm is based on qualitative sets. Some of them are correlated to each other. The comparison is shown in Table 1 and Table 2. The following parameters are taken:

1) **Image Acquisition:** It describes that whether the method used is active or passive.
2) **Image Content:** The image content refers to the image characteristics for them to work reliably.
3) **Motion Presence:** It is concerned with the presence of disparity of the same feature point input image.
4) **Real-time Processing:** A simple conversion rule is applied when qualitative performance data is mentioned. Let us consider the speed of 25 frame in a second (fps) is the speed of real-time processing, running on one regular computer of current standards with a frame size of 640x480 pixels. If an algorithm runs on a normal computer with a speed of 25 fps and a frame size 256x256, this speed is then converted to 5.3 fps ((256* 256)*25)/(640*480).
5) **Absolute/Relative Depth:** The real distance can be calculated between the viewing camera and the objects. By this the actual size of the object can be estimated. Algorithms that rely on camera parameters can recover the real depth. Monocular depth cues cannot be used to predict real depth.
6) **Dense/Sparse Depth Map:** In this the intensity of the depth map is focused. It checks if each pixel of the image is occupied a depth level. The global image feature helps to construct a dense depth map. A sparse depth map checks values for feature points.

Table I
Monocular Depth Cue Comparision

	Atmospheric scattering	*Defocus*	*Shading*	*Linear Perspective*	*Patterned Texture*	*Symmetric Patterns*	*Occlussion*	*Statistical patterns*
Image Acquisition	Passive	Passive	Passive	Passive	Passive	Passive	Passive	Passive
Image Content	Hazy Scene	Two regions exist one infocus and one out of focus	Image must not be too dark.	Geometric appearance exists in images	algorithm requires segmented texture region, other not	Non-frontal image of bilateral symmetric objects	All	All
Motion Presence	No	No	No	No	No	No	No	No
Real-time Processing	N/A	Yes	No	Yes	No	N/A	Yes	Yes
Absolute/ relative depth	Relative	Relative	Relative	Relative	Relative	Relative	Relative	Relative/ Absolute
Dense/ Sparse scenes	900 - 8000 meter; suitable for distant objects	All ranges	All ranges	All ranges	All ranges	All ranges	All ranges	All ranges

Table II
Multioculor Depth Map

	Binocular	*Motion*	*Defocus*	*Focus*	*Silhouette*
Image Acquisition	Active: 2 images of the scene taken from different viewpoints so that corresponding points can be observed	Active/passive: Image sequences of moving objects or static scene taken by moving cameras	Active:2 or more images taken by one camera using different camera parameters	Active: a series of images taken by one camera by varying the distance between the camera and objects	Active: Images taken by multiple cameras surrounding the scene
Image Content	All	All	Objects with complex surface characteristics (e.g. textured images either due roughness of the surface or reflectance variations).	Objects with complex surface characteristics	Foreground objects must be distinguishable from background
Motion Presence	No	No	No	No	No
Real-time Processing	Yes	Yes	Yes	No	Yes
Absolute/ relative depth	Absolute	Absolute	Absolute	Absolute	Absolute
Dense/ Sparsescenes	Dense/sparse	Dense/sparse	Dense	Dense	Sparse

V. CONCLUSION

Different algorithms are presented for converting a 2D image to a 3D image. We can combine various depth cues , both multiocular and monocular and check experimentally using different images. Many 2D to 3D conversion algorithms are found to recover the structure or shape of the objects in the images. These algorithms can be used with the 3D motion of the camera, robot navigation, surveillance etc. In this paper a comparison is also stated between the different depth cues. The comparison is done on the basis of the parameters like Image Acquisition, Image Content, Motion Presence and many more. The comparison is based on the basis of some qualitative sets.

The multiocular depth cue takes both spatial and temporal images into account which in general yield a more accurate result. The monocular depth cues do not require multiple images but are not as much accurate as the multiocular depth map . From the above analysis we can say that not a single cue is enough to perceive depth accurately. It is necessary to combine suitable depth cues in order to get desirable results. Each depth cue has its own advantage and disadvantage.

REFERENCE

[1] Rafael C. Gonzalez, Digital image processing, Pearson Education, 3rd edition, 2009.

[2] Wei, Qingqing. “Converting 2d to 3d: A survey.” International Conference, Page (s). Vol. 7. 2005.

[3] Konrad, Janusz, Meng Wang, and Prakash Ishwar. “2d-to-3d image conversion by learning depth from examples.” Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, 2012.

[4] Konrad, Janusz, et al. “Learning-based, automatic 2d-to-3d image and video conversion.” Image Processing, IEEE Transactions on 22.9 (2013):3485-3496.

[5] Trucco, E; Verri, A (1998) Introductory Techniques for 3-D Computer Vision,Chapter 7, Prentice Hall.

[6] Franke, U.; Rabe, C. (2005), Kalman filter based depth from motion with fast convergence, Intelligent Vehicles Symposium, Proceedings. IEEE, Page(s): 181 186.

[7] Nayar, S.K.; Nakagawa, Y. (1994) Shape from Focus, Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 16, Issue 8, Page(s): 824 831.

[8] Matsuyama, T. (2004) Exploitation of 3D video technologies , Informatics Research for Development of Knowledge Society Infrastructure, ICKS 2004, International Conference, Page(s) 7-14.

[9] Battiato,S.,Curti,S.,La Cascia,M.,Tortora,M.,Scordato,E. Depth map generation by image classification, SPIE Proc. Vol 5302, EI2004 conference Three dimensional image capture and applications VI, 2004.

[10] Francois, A.R.J.; Medioni, G.G.; Waupotitsch, R. (2002) Reconstructing mirror symmetric scenes from a single view using 2-view stereo geometry, Proceedings, 16th International Conference on Pattern Recognition, Vol. 4, Page(s):12 16 vol. III, G.T. Rado and H. Suhl, Eds. New

[11] Po, Lai-Man, et al. “Automatic 2D-to-3D video conversion technique based on depth-from-motion and color segmentation.” Signal Processing (ICSP), 2010 IEEE 10th International Conference on. IEEE, 2010.

[12] M. Guttmann, L. Wolf, and D. Cohen-Or, Semi-automatic stereo ex- traction from video footage, in Proc. IEEE Int. Conf. Comput. Vis.,Oct.2009, pp. 136142.

[13] M. Liao, J. Gao, R. Yang, and M. Gong, Video stereolization: Com- bining motion analysis with user interaction, IEEE Trans. Visualizat. Comput. Graph., vol. 18, no. 7, pp. 10791088 Jul.2012.

[14] R. Phan, R. Rzeszutek, and D. Androutsos, Semi-automatic 2D to 3D image conversion using scale-space random walks and a graph cuts based depth prior, in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 865868.

[15] Loh, A.M.; Hartley, R. (2005) “Shape from Non-Homogeneous, Non-Stationary, Anisotropic, Perspective texture”, Proceedings, the British Machine Vision Conference 2005 (fig 4)

[16] Cozman, F.; Krotkov, E. (1997) “Depth from scattering”, IEEE Computer society conference on Computer Vision and Pattern Recognition, Proceedings, Pages: 801–806

[17] Redert, A. (2005) Patent ID: WO2005091221 A1, “Creating a Depth Map”, Royal Philips Electronics, the Netherlands.

[18] Redert, A. (2005) Patent ID: WO2005083630 A2, WO2005083630 A2, WO2005083631 A2, “Creating a Depth Map”, Royal Philips Electronics, the Netherlands.