ISSN No. 2455-5800
Journal of Scientific Research in Allied Sciences
Review Article

A REVIEW PAPER ON DEPTH IMAGE BASED RENDERING PROCESS FOR APPLICATION OF 2D TO 3D CONVERSION

Ankita Goswami*, Nitin Jain+

**M.Tech. Scholar, Dept. of Electronics Engg., CEC Bilaspur*
+Assistant Professor, Dept.of Electronics Engg.,CEC Bilaspur

Article history:
Submitted on: August 2016
Accepted on: September 2016
Email: info@jusres.com

ABSTRACT:

Three- dimension (3D) technology increases the visual quality as compared to two-dimensional (2D) technology. In present era every multimedia device needs 3D technology. So for generation of 3D content there is need of Depth image based rendering (DIBR) process which will generate left and right image through depth image and original image. Basically DIBR is following the concept of actual 3D recording camera setup. Through original camera setup there is virtual camera formula is generated which will create left and right image. Using both images, 3D content is created. In this paper we are representing a comparative study of previous existing DIBR process.

KEYWORDS: DIBR, PSNR, SSIM, GMSD, FSIM, Gaussian.

INTRODUCTION

A digital image is a representation of a two dimensional image as finite set of digital values, called pixels. Pixel values typically represent gray levels, colors, heights; opacities etc. Digitization implies that a digital image is an approximation of a real scene. Common image formats include:

- 1 sample per point (B&W or Gray scale)
- 3 samples per point (Red, Green, and Blue)
- 4 samples per point (Red, Green, Blue, and "Alpha")

Digital image processing focuses on two major tasks:

For human interpretation, improvement of pictorial information is needed. For processing of image data for storage, transmission and representation for autonomous machine perception is needed. Some argument about where image processing ends and fields such as image analysis and computer vision.

The history of stereoscopy, i.e. stereoscopic imaging or three-dimensional (3D) imaging can be traced back to 1833 when Sir Charles Wheatstone had created a mirror device that provides a view to the illusion of depth., in his description of the"Phenomena of Binocular Vision". The process consists of merging two slightly different views of the same painting anddrawing both of them into one

stereoscopic image, resulting in a compelling 3D perception of the original picture.

With the realization of still photography in 1839, it was only years before the paintings and drawings were replaced by photographs in thestereoscopic image. In 1844, Sir David Brewster further developed the stereoscopic effect by utilizing prismatic lenses to enlarge and fuse the stereo images. Three-dimensional television (3DTV) has a very long history as over the years a consensus has been found that a successful introduction of 3D television broadcast services can only reach success, if the perceived image quality and the viewcomfort are comparable to conventional two-dimensional television (2D Television). In addition to this, 3DTV technology should be compatible with conventional 2DTV technology to ensure a gradual transition from one system to another. This can be referred to as backward-compatibility. On the other hand, 3D televisions has the ability to generate a compelling sense of physical image, and allows images to emerge out from the screen and enter further into the spectator's space, not possible with the conventional 2D or "flat" televisions. This effect has been often exaggerated by throwing or poking objects from the screen at the viewer. Although many good stereoscopic movies were produced in late 1950s, stereoscopic cinema got a bad reputation within the public because of the discomfort experienced when viewing gets misaligned due to inefficiency of conversion of the images.

Today, stereoscopic cinema iscommercially successful related to 2D cinema, with 3D-IMAX theaters being perhaps the most well-known technology. The improvement of 3D technologies raised more interest in 3DTV and in free viewpoint television (FTV).

LITERATURE REVIEW

Over the years many authors have contributed to the field of Depth Image Based Rendering process:

Less compression of image:

First the image should be converted from RGB into a different color space called YCbCr. It has mainly three components Y, Cb and Cr. The Y component represents the brightness of the image pixel, the Cb and Cr components represent the chrominance (splitting into blue and red components of the image). This is the same color space which is used mainly by the digital color television as well as digital video technologies including DVD images, and is very much similar to the way color is represented in any analog PAL video and MAC video but not by analog NTSC, which always uses the YIQ color space. The YCbCr color space conversion allows great compression ratio without effecting on perceptual image quality (or greater perceptual image quality for the same type of compression). The compression was more efficient with brightness information of the image, which is most important for the eventual perceptual quality of the image, which is confined to a single channel image conversion technique, more closely representing a human vision system. This conversion to YCbCr is specified in the JFIF standard, and should be performed in that technique for the resulting JPEG files which should have maximum compatibility. However, some JPEG images in "highest quality" mode need not to be appliedand instead keeping the color information in the RGB color format, where the image is stored in separate channels in red, green and blue luminance. This results in less efficient compression of the image, and would not likely to be used if the file size was an issue.

Occurrence of slight difference in the image after conversion:

ITU-R BT.601 conversion [1]: The form of Y′CbCr that was defined for standard-definition television is used in the ITU-R BT.601 (formerly CCIR 601) standard for using with digital component video which is derived from the corresponding RGB space as follows:

$$K_B = 0.114 \quad (1)$$

$$K_R = 0.299 \quad (2)$$

From the above equations, constants and formulas, the following can be derived for ITU-R BT.601. Analog YCbCr from analog R'G'B' will be derived as follows:

$$Y' = 0.299.R' + 0.587.G' + 0.114.B' \quad (3)$$

$$P_B = -0.168736.R' - .331264.G' + 0.5.B' \quad (4)$$

$$P_R = 0.5.\ R' - 0.418688.G' - 0.081312.B' \quad (5)$$

Digital Y′CbCr (8 bits per sample) is derived from analog R'G'B' as follows:

$$Y' = \quad 16- \quad (65.481.R' – 128.553.G' - 24.966.B') \quad (6)$$

$$C_B = \quad 128- \quad (-37.797.R' - 74.203.G' - 112.0.B') \quad (7)$$

C_R= 128- (112.0.R'-93.786.G- 18.214.B') (8)

Or simply component wise

$(y'.C_B.C_R) = (16.128.128) - (219.y.224.P_B.224.P_R)$(9)

The resultant signals after the analysis will range from 16 to 235 for Y' (Cb and Cr range from 16 to 240); the values from 0 to 15 are called asfoot room, while the values from 236 to 255 are called asheadroom.

Alternatively, digital Y′CbCr can be derived from digital R'dG'dB'd (8 bits per sample, each using the full range with zero representing black and 255 representing white) with the help of the following equations:

$$y' = 16 - \frac{65.738.R'_D}{256} - \frac{129.057.G'_D}{256} - \frac{25.064.B'_D}{256} \quad (10)$$

$$C_B = 128 - \frac{37.945.R'_D}{256} - \frac{74.494.G'_D}{256} - \frac{112.439.B'_D}{256} \quad (11)$$

$$C_R = 128 - \frac{112.439.R'_D}{256} - \frac{94.154.G'_D}{256} - \frac{18.256.B'_D}{256} \quad (12)$$

Here, the scaling factors are multiplied by$\frac{256}{255}$. This will be allowed for the value of 256 in the denominator, which could be calculated with a single bit shift. If the R'dG'dB'ddigital source including foot room and headroom, the foot room offset 16 needed to be subtracted from each signal, and a scale factor of $\frac{255}{219}$needed to be included in the equations. This form of Y′CbCr is used primarily for2d televisions i.e. old standard-definition television systems, as it uses an RGB model that fits the phosphor emission characteristics of older CRTs.

$$y = \frac{77}{256}E'_{R_D} + \frac{150}{256}E'_{G_D} + \frac{29}{256}E'_{B_D} \quad (13)$$

$$C_R = \frac{131}{256}E'_{R_D} - \frac{110}{256}E'_{G_D} - \frac{21}{256}E'_{B_D} + 128 \quad (14)$$

$$C_B = -\frac{44}{256}E'_{R_D} - \frac{87}{256}E'_{B_D} + \frac{131}{256}E'_{B_D} + 128 \quad (15)$$

Nowwe will take the nearest integer coefficients, with base 256. To obtain the 4:2:2 ratio in components Y, Cr, Cb, low-pass filtering method and sub samplingmethod must be performed on the 4:4:4ratio of Y, Cr, Cb signals described above. There were a very slight difference which exists between Cr and Cb components and those which are derived by analogue filtering prior to sampling.

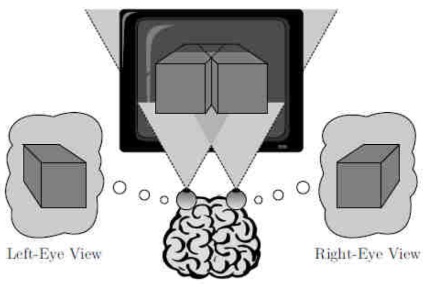

Fig 1: Binocular depth reproduction on a stereoscopic 3D-TV display

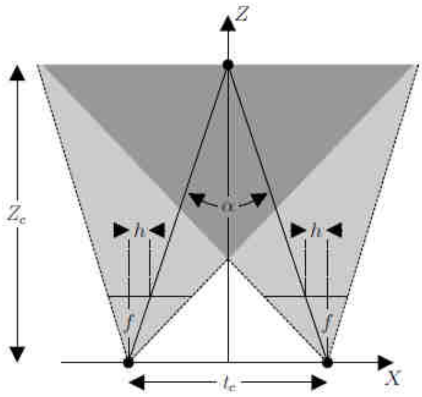

Fig 2: Shift-sensor stereo camera setup for 3D image

Depth-image-based rendering (DIBR) is the process of synthesizing "virtual" views of a particular scene from still or moving image of any sequence of images and associated per-pixel depth information [2]. Conceptually, this view generation can be understood by the following two-step process: Firstly, the original image pixel points are reprojected into the 3D panel, utilizing the respective depth data. Thereafter, these 3D space points are reprojected into the image plane by a "virtual" camera, which is located at the required view position[3]. The concatenation of the reprojection of 2D image (2D-to-3D) and thesubsequent projection of 3D image(3D-to-2D) is usually called 3D image warping in the Computer Graphics and Optimization (CGO) literature[4].

Parallelism in image:

Stereoscopic Image Creation: On a stereoscopic- or auto stereoscopic 3D-TV display, the author [5] has taken two different perspective views of a 3D imagewhich is reproduced (quasi)simultaneously on a joint image plane. The horizontal difference between the left- and right-eye views, means the screen parallax values, are then interpreted by the human brain and the two images were fused into a single image, the three-dimensional percept used by human brain.

In the process of DIBR, the author has generated virtual views from a monocarpic video and the associated depth information of the image [5].Considering an arbitrary 3D space point M, the two perspective projection equations in two views are as follows:

$m = AP_nM$ (16)

$m^* = A^*P_nDM$ (17)

where m and m* represent two 2D image points in the left and right view images, respectively.

The matrix D containsthe rotation matrix and the translation matrix that willtransform the 3D point of the image from the world coordinate system into the camera coordinate system. The matrices A and A* specifies the intrinsic parameters of the camera.The identity matrix Pndesignates the normalized perspective projection matrix. The parameters f and tc represent focal length and the baseline distance between two virtual camera Cland Cr, respectively. Zc means the depth value of the ZPS (Zero Parallax Setting). The parallel camera setup rather than the convergent camera setup is used not in this technique to generate vertical disparities and it is much easier to implement with DIBR technically, since the position of the 3D space point of the image M depends onthe depth value of the image.

$$\tilde{m}^{*} = \tilde{m} + \frac{A^{*}t}{depth} + \begin{bmatrix} h \\ 0 \\ 0 \end{bmatrix}, \text{ with } t = \begin{bmatrix} t_x \\ 0 \\ 0 \end{bmatrix} \quad (18)$$

So, the pixel position (*x, y*) of each warped image point can simply be calculated as:

$$x^{*} = x + \frac{\alpha_u t_x}{depth} + h, \text{ with } y^{*} = y \quad (19)$$

where auis a parameter related to left-right-eye distance and eye-screen distance. The pixel position (*xc, y*), (*xl, y*) and (*xr, y*) of the reference view and two virtual views corresponding to the point *P* with depth will have the following relationship:

$$x_l = x_c + \frac{t_c f}{2depth(x_c, y)} + h \quad (21)$$

$$x_r = x_c - \frac{t_c f}{2depth(x_c, y)} - h \quad (22)$$

From the following equation, the offset "h" between reference view and target view can be computed:

$$h = -\frac{t_c f}{2Z_c} \quad (23)$$

Division in image

After hole-filling, the whole left-and-right view images should be translated horizontally by h pixels so as to replace the addition and subtraction of h [6][7].And as the division operation in this technique is time and area which is consumed by hardware implementation, the DIBR algorithm will be optimized to eliminate the division operation in [8].

1/Depth ranges between 0 and 1. Since the value of 1/depth only represents the relative distance of the pixels but not real distance, researchersare sensitive about the objects which have smaller depth value. In that place, (256−depth)/256 could be used to replace 1/depth.

$$x_l = x_c + k\frac{t_c f}{2}\frac{256 - depth(x_c, y)}{256} \quad (24)$$

$$x_r = x_c - k\frac{t_c f}{2}\frac{256 - depth(x_c, y)}{256} \quad (25)$$

In the implementation, for each depth value between 0 and 255, the average deviation between the practical value and the theoretical value is about 3.6%. Then, the division operation can be implemented simply by shifting operation in the hardware. In the implementation of the operation, the white color in the depth map (i.e. 255) represents the nearest plane in the image while the black color (i.e. 0) represents the farthest. It is opposite to the real depth given in [2]. We use *D* to represent the value that is obtained from the depth map:

D=256-depth (26)

l_offsetD. pos/256 (27)

r_offset= pos-l_offset (28)

l_offsetand r_offset are used to represent the value of offset in the left-and-right view images, respectively. So while implementing the DIBR algorithm in hardware, we are using computation of left-view offset 'l'and right-view offset 'r' offset.

pos is a parameter related to the eye-screen distance and the screen size. *pos*is set to 1/32 of the width of the screen [4]. This gives the researcher a good visual experience and can be simply calculated by shifting operation. After the *l offset* and *r offset* are calculated, the left-view image and the right-view image can be obtained.

$l_{pic}(x, y) = pic(x - l\_offset, y)$ left – view (29)

$r_{pic}(x,y) =$ pic(x –r_offset,y) right –view (30)

where *lpic* and *rpic* represent left-view and right-view images, respectively. *pic*is the original 2D image, and (*x*, *y*) representsthe pixel position in the image. Pre-processing of depth image is usually a smoothing filter operation because depth image with horizontal sharp transition can result in formation of big holes after warping. Hence, smoothing filter is applied to smooth sharp transition to reduce the number of big holes in the image. However, if whole depth image is blurred, it will reduce big holesand also degrade the warped view. This result in smoothing of the depth image and will get the warped view.

3D Image Warping

This will give a good visual experience and can be simply calculated by shifting operation 3D Image Warping.3D image warping is used for mapping the intermediate view pixel by pixel to left or right view of the image according to pixel depth value of the image. In other words, 3D imagewarping is used for transforming pixel location according to the depth value. The formula given below shows that the 3D warping maps pixel of the intermediate view to left and right view of the image in the horizontal direction.

$$x_l = x_c + (\frac{t_x}{2}\frac{f}{z}) \quad (31)$$

$$x_r = x_c - (\frac{t_x}{2}\frac{f}{z}) \quad (32)$$

x*l*is the horizontal coordinate of the left view of the image,

*xr*is the horizontal coordinate of the right view of the image,

*xc*is the horizontal coordinate of the intermediate view,

Z is depth value of current pixel,

*f*is camera focal length and

*tx*is eye distance.

Time complexity

In the Latest DIBR Algorithm [9], the author proposed a newapproach for implementation of DIBR approach. According to this approach, the author has tried to reduce the hole filling problem. In this algorithm, author proposed a new formula for generation of Left and Right Image. According to the formula given by the researcher, he has generated left and right image to fill hole with the help of corresponding depth map. Problem with this approach is that it requires heavy amount of time, means this algorithm has increased time complexity in the implementation of DIBR approach.

RESEARCH GAP

Previous DIBR model [6] [7] arehaving problem same problem of time complexity and image quality.The algorithm reduces some amount of time complexity but still it has some issues of image quality. Some algorithm [9] reduces the image quality problem, but on the other side increases the time complexity issue.In this way the previous approaches does not justice completely with the problem of timing issue and image quality.

FUTURE OBJECTIVE

Form he above literature review done on the image processing methods and there limitations, here we are developing a fast algorithm thatwill reduce time complexity and increase image quality for generation of left and right part of the main image. It is an algorithm which will improve the quality of generated image in context of the pixels of the image and also the brightness of the image. Here our algorithm helps us to generate a good quality of image conversion from 2D to 3D.

APPLICATIONS

The three-dimensional (3D) displays provide an improvement of visual quality over the 2D displays. The conversion of existing 2D videos to 3D videos is necessary for multimedia application like artistic effect, image enhancement restorationand communication. 3D Display has many applications which are broadcasting, movies, gaming, photography, camcorders, education and Wi-Fi.

The system achieves the goals by first sensing the real-time drawing, calculating the time delay and hence the distances from it and then displaying its 3D image simultaneously from all angles.In this way it provides a real image of any object. 2D video to 3D video conversion is necessary for multimedia application. Medical visualization,Industrial inspection.

CONCLUSION

According to the literature survey of conversion of the 2D to 3D image and video all the authors used the Depth image Based rendering which is used in the methodology.There are several works has done under this field but they are included various issues like time consuming, image quality,memory storage problem.So the broader aim of our research is to identify these problem according to the condition and try to removethem.

REFERENCES

[1] Rec. ITU-R BT.601-5, Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-screen 16:9 Aspect Ratios, (1982-1986-1990-1992-1994-1995), Section 3.5.

[2] Kwanghee Jung; Young Kyung Park;JoongKyu Kim; Hyun Lee; Kugjin Yun; Hur, N.; Jinwoong Kim, "Depth Image Based Rendering for 3D Data Service Over T-DMB," 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008 , vol., no., pp.237,240, 28-30 May 2008.

[3] Hang Shao; Xun Cao; GuihuaEr, "Objective quality assessment of depth image based rendering in 3DTV system," 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009 , vol., no., pp.1,4, 4-6 May 2009.

[4] Nguyen, Q.H.; Do, M.N.; Patel, S.J., "Depth image-based rendering with low resolution depth," Image Processing (ICIP), 2009 16th IEEE International Conference on , vol., no., pp.553,556, 7-10 Nov. 2009.

[5] Fehn, Christoph. "A 3D-TV approach using depth-image-based rendering (DIBR)." Proc. of VIIP. Vol.3. 2003.

[6] Wan-Yu Chen; Chang, Yu-Lin; Shyh-Feng Lin; Li-Fu Ding; Liang-Gee Chen, "Efficient Depth Image Based

Rendering with Edge Dependent Depth Filter and Interpolation," Multimedia and Expo, 2005. ICME 2005, IEEE International Conference on, vol., no., pp. 1314-1317, 6-6 July 2005.

[7] Chao-Chung Cheng; Chung-Te Li; Liang-Gee Chen, "A novel 2Dd-to-3D conversion system using edge information," Consumer Electronics, IEEE Transactions on , vol.56, no.3, pp.1739,1745, Aug. 2010.

[8] Hao Liu; Wei Guo; Chao Lu; Jizeng Wei, "An Efficient Stereoscopic Game Conversion System for Embedded Platform," Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on , vol., no., pp.1235-1240, 16-18 Nov. 2011.

[9] Yu-Cheng Fan; Yi-Chun Chen; Shih-Ying Chou, "Vivid-DIBR Based 2D–3D Image Conversion System for 3D Display," Display Technology, Journal of , vol.10, no.10, pp.887,898,Oct. 2014doi: 10.1109/JDT.2014.2331064.

[10] .Xue, W.; Zhang, L.; Mou, X.; Bovik, A., "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," Image Processing, IEEE Transactions on ,vol. pp. no.99,pp.1,1

[11] Lin Zhang; Zhang, D.; XuanqinMou, "RFSIM: A feature based image quality assessment metric using Riesz transforms," Image Processing (ICIP), 2010 17th IEEE International Conference on , vol., no., pp.321-324, 26-29 Sept. 2010.

[12] .Lin Zhang; Zhang, D.; XuanqinMou; Zhang, D., "FSIM: A Feature Similarity Index for Image Quality Assessment," Image Processing, IEEE Transactions on , vol.20, no.8, pp.2378,2386, Aug. 2011.

[13]. Zhou Wang; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P., "Image quality assessment: from error visibilityto structural similarity," Image Processing, IEEE Transactions on, vol.13, no.4, pp.600-612, April 2004.

[14]http://www02.lps.ufrj.br/~eduardo/MMP/3D_results.html

[15] Tsai, Sung-Fang, et al. "A real- time 1080p 2D-to-3D video conversion system"Consumer Electronics, IEEE Transactions on 57.2 (2011): 915-922.

[16] Kaushik, Sharda, and N.V.S.V. Kumar. "Energy-efficient approximate 2D Gaussian smoothing filter for error tolerant applications." Advanced Computing Conference (IACC), 2015 IEEE International.