High Quality View Synthesis Algorithm and Architecture for 2D to 3D Conversion

Yeong-Kang Lai, Yu-Fan Lai, and Jung-Wei Lin
Department of Electrical Engineering, National Chung Hsing University
No. 250, Kuo Kuang Road, Taichung 402, Taiwan
yklai@dragon.nchu.edu.tw

Abstract—Due to the advances in consumer electronics, 3D display systems have become possible and attractive. In order to execute backward-compatible 2D images/video, 2D-to-3D conversion techniques are developed. The monocular depth cue is to synthesize the virtual images of left-eye and right-eye. During the process of view synthesis, 3D image warping and hole filling are crucial techniques to effect the visual quality of final synthesized views. For 3D-TV applications with high resolution (1920x1080 or higher), this paper proposes a high quality view synthesis algorithm to improve the virtual view frames. The proposed algorithm can provide PSNR gain of 0.2~1.5dB and capture good human visual perception.

I. INTRODUCTION

With the advances in 3D movie, people are familiar with stereoscopic display rendering. 3D-TV is the future trend compared to conventional 2D TV. However, stereoscopic display technology is not the new video standard; it has been the extension of H.264/AVC standard by the Joint Video Team (JVT) [1]. 3D-TV technology is related to Multi-View Coding (MVC) and free-viewpoint rendering [2]. The 3D presentation formats can be divided into two types: frame-compatible representations and depth-based representations [3][5]. Under frame-compatible representations, it can consider two independent video sequences (left and right views) and put them interleaved into a single frame. However, it will lost spatial or temporal resolution and need two real cameras. In order to backward-compatible 2D TV and let it reveal 3D rendering, depth-based representations produce virtual views through depth image-based rendering (DIBR) [6]. The depth information can be extracted from specific camera or binocular parallax. Fig. 1 shows the block diagram of view synthesis for 3D display system. This paper focuses on the algorithm of view synthesis; the depth information is extracted from camera. The proposed algorithms also can be used in other 3D consumer electronics, such as 3D camera, 3D mobile [7] and 3D projector.

This paper is organized as follows. In section II, we propose a high quality view synthesis algorithm. Moreover, the simulation results are shown in Section III. The hardware architecture design and implementation results are shown in Section IV. Finally, a brief conclusion is given in Section V.

II. PROPOSED ALGORITHM

A. 3-D Warping

We adopt DIBR method in 3D warping; the concept is from the intermediate view to generate left and right virtual views. Fig. 2 shows the model of DIBR, the original image points at location \((P_{m,y})\) are transferred to left points \((P_{l,y})\) and right points \((P_{r,y})\). It is shown in Equation (1).

\[
P_l = P_m + \frac{Bf}{2Z} \quad \text{and} \quad P_r = P_m - \frac{Bf}{2Z}
\]  

(1)

Where \(f\) is the focal length of the camera, \(B\) is the distance between the left and right virtual camera. Moreover, \(Z\) represents the depth value of each pixel in the intermediate view. In order to get the depth value of \(Z\) from depth map, the equation is shown as follows,
Where \( d \) is the depth level (0–255) in the depth map, \( Z_{\text{min}} \) and \( Z_{\text{max}} \) are the minimum and maximum depth values in a frame. According to Equations (1) and (2), the virtual views of left and right can be calculated via DIBR techniques. The concept of 3D warping is from the 2D camera image plane to the 3D coordinates plane, the depth cue must be considered in order to render 3D display.

### B. Image Inpainting Algorithm

Although after 3-D warping can produce virtual views. During DIBR, the integer grid points in the reference are warped to irregularly points in the virtual views. It occurs occluded problem and needs hole filling to fix virtual images. Normally, the point of left and right virtual views after 3D image warping should contain decimal place. However, most 3D image warping adopts round off to approach integral point.

Thus we keep decimal place and use weight interpolation to improve the quality of virtual view frames. Fig. 3 shows the dataflow of the proposed algorithm (image inpainting), we modify the precision of 3D image warping and inpaint holes using different pixels. Moreover, we adopt 3 pixels distance to determine the hole is small or not and judge the frame is edge or not. In Fig. 4, we classify holes into three types: small hole, big hole and frame edge hole.

### C. PSNR Evaluation

In order to identify our image inpainting algorithm can improve the quality of virtual view frames. We adopt Book Arrival, Love bird1, and Newspaper [11-12] as our test sequences. The frame size is all 1024x768 pixels. From experimental result in Table I, our algorithm can improve PSNR gain of 0.2–1.5dB than others.
D. Formula Derivation

In Fig. 5, due to the integer grid points in the reference are warped to irregularly points in the virtual views, we use weight interpolation to instead of traditional round off interpolation. As a result, Figs. 6 and 7 shows the image inpainting interpolation among 1 pixels and 2 pixles. The Equations are shown as follows,

\[ A = \frac{(x_a - x_0)}{(x_1 - x_0)} \times P_1 + \frac{(x_1 - x_a)}{(x_1 - x_0)} \times P_0 \]  
\[ B = \frac{(x_b - x_0)}{(x_1 - x_0)} \times P_1 + \frac{(x_1 - x_b)}{(x_1 - x_0)} \times P_0 \]

In order to simply the Equations (3) and (4), we assign \( h_a \) and \( h_b \) as:

\[ h_a = \frac{(x_a - x_0)}{(x_1 - x_0)} \quad \text{and} \quad h_b = \frac{(x_b - x_0)}{(x_1 - x_0)} \]

It can also get the formulas of \( (1-h_a) \) and \( (1-h_b) \) from \( h_a \) and \( h_b \) assignments.

Finally, the Equations (3) and (4) can be rewritten as:

\[ A = h_a \times P_1 + (1-h_a) \times P_0 = h_a \times P_1 + P_0 - h_a \times P_0 \]
\[ B = h_b \times P_1 + (1-h_b) \times P_0 = h_b \times P_1 + P_0 - h_b \times P_0 \]

III. ARCHITECTURE

According to above Equations (5) and (6), we can draw the hardware architecture of image inpainting interpolation. Fig. 9 shows the detail Hardware architecture for 1 pixels or 2 pixels. It only needs 3 adders and 2 multiplier circuits. Because of the values of \( h_a \) and \( h_b \) are all just constant, we consider them as constant factor. Hence, the operation of 2 subtraction and 1 division can be decreased. Table II shows the comparison of operations among various interpolation circuit. For the high definition (HD) video such as 1280×720 (HD720p) or 1920×1080(HD1080p), the computational complexity is very huge. For this reason, our circuit is suitable for real-time applications.

Fig. 8 reveals hardware architecture of 3D view synthesis engine. In order to decrease computational complexity. The intermediate-view texture and their depth map are stored in external memory. We adopt look-up table to speed up calculation time in 3D warping engine. Finally, the hole filling engine adopts image inpainting interpolation to fill the remaining holes, and write two virtual views data (left and right) to output buffer for 3D rendering.

IV. EXPERIMENTAL RESULTS

We implement our proposed architecture and using 3D screen to reveal it. The test sequences are Break-Dancers, Love bird1, and Ballet. Because of we use active shutter technology and need wearing 3D active glasses. The shutters can rapidly block the left and right eye views alternately, hence each eye receives the correct image. Fig. 10 shows the
final results which revealed in 3D screen. The occlusion and hole problems can be resolved via our algorithm and architecture. Moreover, we invited 10 audiences to evaluate the quality of 3D sequences, the average score were 86.

V. CONCLUSION

In this paper, we propose a high quality view synthesis algorithm for 3D display systems. According to pixel distance and edge detection, it can separate hole inpainting into three types. From the experimental result, the proposed algorithm could distinguish between foreground and background information effectively and get better human visual perception for 3D-TV. Moreover, The proposed algorithm can provide PSNR gain of 0.2~1.5dB than other method. Finally, it only needs 3 adders and 2 multiplier circuits in image inpainting interpolation per pixel. Hence it is suitable on hardware platform design for real-time applications.

ACKNOWLEDGMENT

This work was supported by the National Science Council of the Republic of China under Contract NSC 100-2220-E-005-003.

REFERENCES