DEPTH-BASED 3D ANAGLYPH IMAGE MODELING

Makhzani Niloufar ¹, Kok-Why Ng ², Babaei Mahdi ³

¹,²,³ Faculty of Computing and Informatics, Multimedia University, Malaysia.
¹ m.niloufar91@gmail.com, ²kwng@mmu.edu.my, ³mahdi.babaei@gmail.com

ABSTRACT

3D modeling or reconstruction is no longer a new topic in the latest graphics research world. However, they remain to be a challenging and tedious work in this new technology era. Many traditional 2D to 3D conversion techniques are time-consuming and with a lot of human intervention required. Most of them used either stereoscopic camera or two input images to achieve the goal. This paper proposes an algorithm to generate interesting 3D anaglyph image from a single 2D image and depth map. The 2D image is automatically duplicated and a displacement metric is applied to shift the pixels into left and right directions for stereoscopic view purpose. The blending process is then applied to mingle both created images for producing a sense of 3D image. The proposed work has been experimented under different light conditions and is well viewed with anaglyph glasses. This piece of work will be of helpful in future devices for creating 3D content easily.

Keywords: 3D Model, Anaglyph Image, Depth Map, Kinect

1. INTRODUCTION

Recently, many entertainment industries (e.g., movies, games) grow incredibly in the track of three-dimensional (3D) effect. However, due to limited 3D contents, the conversion from 2D to 3D effect is proposed as an alternative solution to meet the viewers’ visual demand. Another concern is that by displaying 3D effect on 2D screen, the 3D effect will again be projected as 2D view. This will be a waste of process as it goes back to the original 2D display. The dispute here is why not the 3D sense be created directly from the 2D images? This creation is called 3D image modeling.

The 3D space refers to a geometric three-parameter model of physical universe, which includes the depth for all the objects. For 3D image modeling, one would need to create view illusion by absorbing the third dimension (depth) into 2D images. This can be done by finding the distance of the objects from the camera. One way to locate in the camera for recognizing and differentiating the distance between the objects is through the image brightness (intensity). The nearer objects will reflect higher intensity (brighter) to the sensor in the camera and vice versa for the further objects (darker).

The 3D anaglyph image is one of the examples of 3D image model. It blends the red and cyan colors or blue-channel filters of a sense captured images with a specific displacement to create the illusion for human eyes. These images allow perception of depth when observed through anaglyph colored glasses.

It would be ideal and expediency if the 3D images are able to be generated from a single camera. However, from our survey, no such a device is available in the digital market for acquiring the 3D images. Though, the breakdown of the 3D image components can simply be
obtained with a normal camera and a depth sensor through using the infrared lights reflection [1]. These characteristics is made available in a device namely Microsoft Kinect. This device able to capture the RGB frame (Red-Green-Blue) and depth frame in a specific range [2].

The goal of this research is to convert and optimize the captured input stream into 3D anaglyph images which can be viewed as a 3D stream. The next section of this paper is our literature review. Section Three will discuss the mapping of the acquired data for producing the anaglyph images. We will show the result and analysis in Section Four. Section Five is our conclusion.

2. LITERATURE REVIEW

The term "3D" was coined in 1850's. In 1853, W. Rollman was the first person who illustrated the principle of the anaglyph using blue and red lines on a black field, and was perceived with red and blue glasses for the effect. However, this was for line drawings only. In 1858 Joseph D'Almeida projected 3D magic lantern slide show using red and green filters, and the audience wore red and green goggles to view the effect [3]. The first printed anaglyph was owed to Louis Ducas du Hauron, produced in 1891. In 1953, the anaglyph was appeared in the newspapers, magazines and comic books. [4]

Scientifically, human eyes are able to see 3D effect is not because of human own a 3D vision. It is human eyes to perceive the same scene but each eye receives slightly different signal (items with slightly different in location) [5][6]. The visual accuracy would be improved with the assistance of the other attributes such as objects’ shadows, curvature and etc. for better depth estimation.

Many existing papers studied the techniques for creating imaging and 3D visualization based on single image camera and IRcamera built-in Kinect to detect depth information. Nagai et al. [7] estimated the depth by performing surface reconstruction from each image for known fixed objects using Hidden Markov model (HMM) framework. Their approach was based on the knowledge of the objects acquired from a number of samples. In our opinion, this approach may not be stable if the knowledge obtained from the samples was not accurate. One way to improve the knowledge accuracy is to have a large number of samples and take the average of the happening in all the samples so that the result does not favor to those excessive samples. But this will burden the process with too many unrelated samples data. Zhang et al. [8] opposed the use of mean and standard deviation of depth error to process synthetic images. They claimed that minimization approach was more robust. Unfortunately, minimization process is well known with heavy computation. Lindeberg and Garding [9] commented on the scale in shape from texture that would lead to depth inaccuracy. This does not affect our work as we do not apply any scale to our images. Hertzmann and Seitz [10] reconstructed high quality 3D models from several depth images with the assistant of the other objects’ shape next to the target object. Same approach had been applied by Torresani and Hertzmann [11] on input data from video sequences. In our work, we do not need to reconstruct the 3D model. Therefore, our process is more efficient than them.

Torralba and Oliva [12] applied Fourier spectrum from the images to compute the mean depth of a scene. This method worked well but it was too complex. Single view metrology [13][14] assumed that vanishing lines and points are known in a scene. Thus, it calculated angles between the parallel lines to infer 3-d structure from Manhattan images. The other method to find the depth map is to use RGB-D cameras, which are available in the market. These cameras work with either one of the two main processes below:

1. Depth from stereo:
   In computer vision, one way to estimate depth for 3D construction is to use pair of stereo images, which can be captured using stereoscopic camera or to capture the images by rotating the camera with a specified degree [15].

2. Depth from defocus of structured light:
   This method works by projecting a known light pattern to a scene and analyzing the reflection of those lights [16].

A quick review: In June 2011, Kinect was designed to track human body and recognize gestures. This device produces RGB color image and 2.5D image (Figure 1). Objects that are farther away have greater range than the near objects. With the propagation of the infrared lights and the reflection of them, Kinect is able to detect the depth between the objects as shown in Figure 2 with the help of its triangular view.
The preliminary stage is to initialize the sensor, color and depth streams in our program. This will capture 30 frames per second as the input data. The color and depth information in each pixel frame will then be processed and categorized. Next, the left and right images will be created based on the depth information. With the images, we further extract the red and cyan channel filters separately. Lastly, we blend the red and cyan frames together with a specific displacement. The detail process of every stage will be discussed in the following subsections.

As RGB camera and IR camera are not built at the same location, the collected depth and color pixel values are not matched accurately in the two frames. A mapping process (which will be discussed in next section) is needed to re-match the depth pixels in the color image for applying depth detail at every valid color pixel.

The anaglyph 3D image is a combination of two images from the perspective of the left and right eyes. One eye will see through a red filter and the other eye will see through a contrasting color filter such as cyan filter. Though both eyes receive signals from the images, they will be recognized differently as the color and distant viewed differently by each eye. With that, the 3D effect would be perceived by wearing the red-cyan glasses. All the traditional methods of forming the 3D image from 2D image involved two input images. For generating anaglyph 3D image using only one image, the depth information of the image needs to be estimated [17]. Our input is one single 2D color image and a depth map. Next section will demonstrate our algorithm for generating anaglyph image.

3. PROPOSED ALGORITHM

Figure 3 below shows our proposed algorithm. The preliminary stage is to initialize the sensor, color and depth streams in our program. This will
same distance from the camera and it is assigned to the largest depth value. Same measure is applied to pixels shorter than two feet to the minimum depth distance. With this assumption, we have depth and color information for all the pixels in our frame.

3.3 Creating left and right images

For creating anaglyph images, we need to blend two RGB images. As our input has only one RGB image, we need to duplicate the color frame and apply pixel shifting to have left and right images that analogous to capturing the images in a stereoscopic view. The pixel shifting (or displacement) can be of any reasonable range depend on the viewer. We may request the user to choose their preferably displacement measure or we may find the displacement base on the distance between the sensor and the object [18]. In our work, we propose to shift the pixels based on depth value. In our loop of pixels, each retrieved depth value will displace the color pixel proportionally.

\[ \text{Displacement, } D_{\text{pixel}} = \text{RGB}_{\text{pixel}} + \text{Depth}_{\text{pixel}} \]

3.4 Channel filtering

Anaglyph image is a result of blending the red and cyan or blue filters of two frames. For this purpose, we make use of Aforge library, which includes the methods for filtering images. It receives 24bpp bitmap as input and does the channel filtering based on the color passed to it. In our work, the red filter is to be applied on the left image and the cyan filter (the mixture of blue and green channel) is to be applied onto the right image. After that we will have both frames ready to be blended in the last process to produce the anaglyph image.

3.5 Generating anaglyph image

The left and right images that we have created, we apply channel filter to them. Blending these two frames will result the anaglyph image, which makes sense of depth when we view by wearing the anaglyph glass. Superimposing the red and cyan frames will result continuous anaglyph models from Kinect RGB captured frames. Below is the metric to blend the RGB images. To synchronize the output, we normalize the result of the maximum RGB value (from the original, left and right RGB images) by dividing the maximum pixel intensity and multiplying to 255 pixel value.

\[ \text{Anaglyph, } A_{\text{image}} = \frac{\text{Max}(\text{RGB}_{\text{original}}, \text{RGB}_{\text{left}}, \text{RGB}_{\text{right}}) \times 255}{\text{Max pixel intensity}} \]

4. RESULTS AND ANALYSIS

Figure 4 shows the result of indoor captured with normal lighting condition in our work. Figure 4(a) and 4(b) are the original RGB frame and depth frame retrieved in first stage respectively; while Figure 4(c) and 4(d) are the created red and cyan channel filters. The blended result of 3D anaglyph image is shown in Figure 4(e). One can perceive that the color balls in the middle region are well demonstrated the 3D effect as compared to nearer or further color balls. Besides, those balls with different colors (or color mixture) from the channel filters present better quality than those with color close to the filters.

Figure 5 shows the result of indoor captured with minimum lighting condition. The targeted objects are aligned within the range of coverage. This produces good 3D effect to our eyes as almost all the objects get blended.

Figure 6 presents the result of outdoor captured with indirect sun light. In this experiment, one can perceive clearly the red and cyan color around the human body. This gives an accurate result. Besides, we find that as sun rays emit infrared light, it would affect our depth sensor. Therefore, it is advised not to capture the frames under the sun light.

In short, we have tested our application in different lighting conditions and it produces comparable results. Though the lighting condition may affect the result (while capturing the images), if one tries to avoid the targeted object to be exposed directly to the extreme light intensity, it can soften the effect on the captured frames. Moreover, the proper processes and high quality of the device will also contribute directly on the accuracy of processing the input images.
Figure 4: Indoor captured with normal lighting condition.

Figure 5: Indoor captured with minimum lighting.

Figure 6: Outdoor captured with indirect sun light.
This paper proposes an algorithm for generating 3D anaglyph images from 2D RGB and depth frames. The experiments are taken under different lighting conditions and our application manages to produce the desired results. The sense of 3D images from our work can well be viewed by anaglyph glasses. This piece of work will be helpful in future devices for creating 3D content easily.

REFERENCES