2D to 3D Video Conversion

Arpita M. Limbachiya
Department of Computer Engineering, Gujarat Technological University Post Graduates School
Gujarat, India

Abstract: 3D video application is becoming popular in our daily life, especially at home entertainment. Although more and more 3D movies are being made, 3D video content is still not 100% satisfactory. This future 3D video market. There is a rising demand on new techniques for automatically converting 2D video content to stereoscopic 3D video displays. However, in recent years, there has been rapid progress in the field’s image capture, coding and display which brings the realm of 3D closer to reality than ever before. The survey investigates the existing 2D to 3D conversion algorithms developed in the past years by various computer vision research communities across the world. Each algorithm has its own strengths and weaknesses. Most conversion algorithms make use of certain depth cues to generate depth maps. Among 2D-to-3D image conversion methods, those involving human operators have been most successful but also time-consuming and costly. Fully-automatic methods typically make strong assumptions about the 3D scene. Although such methods may work well in some cases, in general it is very difficult to construct a deterministic scene model that covers all possible background and foreground combinations. In this, we study about 2D to 3D video conversation. A stereo matching algorithm based on image segments is presented. We propose the hybrid segmentation algorithm that is based on a combination of the Belief Propagation and Mean Shift algorithms with aim to refine the disparity and depth map by using a stereo pair of images. This algorithm utilizes image filtering and modified SAD (Sum of Absolute Differences) stereo matching method. Depth map can be obtained by application of segment disparities to the original images. Hybrid algorithm HSAD gives a good performance. And then we using rendering method that is DIBR there is many advantages of using DIBR method for depth generation DIBR. Which is gives and better visual quality of 3D stereoscopic display.

Keywords: Depth map, HSAD, Disparity map, DIBR.

I. INTRODUCTION

Rapid development of 3D displays technologies and digital video processing has brought 3DTV into our life. As more facilities and devices are 3D capable, the demand for 3D video contents is increasing sharply. However, the tremendous amount of current and past media data is in 2D format and 3D stereo contents are still not rich enough now. Compared to the direct capture of 3D video contents, video conversion from 2D to 3D is a low-cost and backward compatible solution. The term “3D” in the context denotes “stereoscopic,” meaning at a viewing system is used for visualization. Stereoscopic images that are displayed on 3D displays can increase the visual impact and heighten the sense of presence for viewers. The successful adoption of 3D TV by the general public will depend not only on technological advances in 3D display and 3D TV broadcasting systems, but also on the availability of wide variety of program content in stereoscopic 3D (S3D) format for 3D TV services [1]. The visual cortex in our brain is disparity-totagainsenseofdepth inscence. Essentially, the bigger the shift between the views, the closer the object to the viewer (i.e., small depth), and vice-versa. For any image, or frame in a video, we want to create depth maps which will allow us to convert things into 3D. Most of the time, the original image itself is used as the left view, while the use of a depth map would generate the right view. Once the left and right views are created, these can be presented on any 3D compatible technology, as long as you format both the views properly for the technology you want to view them on. Depth maps are black and white images that are of the same size as the images you want to convert, where each pixel in these images or videos give the sense of depth in the scene. Darker pixels denote a pixel being very far away, while lighter pixels denote a pixel being very close. Stereo Vision - The slightly different perspectives from which 2 or more cameras perceive the world lead to different images with relative displacements of objects—disparities—in the different monocular views of the scene [5].

Figure: Camera for a single object
Most of the stereoscopic techniques are based on 2D offset images each simulating the different views: the right eye sees the right portion and the left eye sees the left portion [4]. The conventional 3D imaging depicts the concept that the left eye sees the right portion of the scene and the right eye sees the left portion of the image. This principle varies among different cameras that have been developed. These cameras use 2 lenses. Thus, it takes 2 images, one right side image and one left side image.

II. STATE OF THE ART

Two approaches to 2D to 3D conversion can be loosely defined: quality semiautomatic conversion for cinema and high quality 3D TV, and low-quality automatic conversion for cheap 3D TV, VOD and similar applications. [13] In semiautomatic conversion a skilled operator assigns depth to various parts of an image or video. Based on this sparse depth assignment, a computer algorithm estimates dense depth over the entire image or video sequence. In the case of automatic methods, no operator intervention is needed and a computer algorithm automatically estimates the depth for a single image or video. Automatic methods estimate shape from shading, structure from motion or depth from defocus.

Electronics manufacturers use stronger assumptions to develop real-time 2D-to-3D converters. Such methods may work well in specific scenarios. But generally it is very difficult to construct heuristic assumptions that cover all possible background and foreground combinations. An important step in any 3D system is the 3D content generation. Several companies have been designed to generate 3D model directly. For example, a stereoscopic dual-camera makes use of a co-planar configuration of two separate, monoscopic cameras, each capturing one eye’s view, and depth information is computed using binocular disparity. A depth-range camera is another example. It is a conventional videocamera enhanced with an add-on laser element, which captures a normal two-dimensional RGB image and a corresponding depth map. A depth map is a 2D function that gives the depth (with respect to the viewpoint) of an object point as a function of the image coordinates. Usually, it is represented as a gray level image with the intensity of each pixel registering its depth. The laser element emits a light wall towards the real world scene, which hits the objects in the scene and reflected back. This is subsequently registered and used for the construction of a depth map. All the techniques described above are used to directly generate 3D content, which certainly contribute to the prevalence of 3D TV. However, the tremendous amount of current and past media data is in 2D format and should be possible to be viewed with a stereoscopic effect. This is where the 2D to 3D conversion method comes to rescue. This method recovers the depth information by analyzing and processing the 2D image structure. Figure 1 shows the typical product of 2D to 3D conversion algorithm – the corresponding depth map of a conventional 2D image. A diversity of 2D to 3D conversion algorithms has been developed by the computer vision community. Each algorithm has its own strengths and weaknesses. Most conversion algorithms make use of certain depth cues to generate depth maps. So here describe of our proposed system then experiments results and then at last our conclusion.

III. PROPOSED SYSTEM

Here we proposed a robust automatic 2D to 3D conversion using multiple cue for depth map estimation. For 2D to 3D Conversion the main two steps are:

1) Depth map Generation
2) Rendering
3) Depth Map Generation

For depth map generation the following steps are followed:

1) **Image segmentation**
   
   For color image segmentation we use two segmentation algorithm Belief Propagation and the Mean Shift segmentation. Belief Propagation is the algorithm that was developed mainly with aim to find marginal probabilities in Bayes networks. In addition, the algorithm can also handle other graphical models such as Markov Random Field (MRF) models, which are of certain interest in the optimization of global energy functions found in computer vision. The MRF model is the undirected graph model, in which nodes represent random variables. While The Mean Shift algorithm was proposed by Fukunaga and Hostetler.

Figure 2: Flow For Depth Map generation

© 2015, IJARCSSE All Rights Reserved
The algorithm is based on the kernel density estimation. This algorithm is non–parametric iterative algorithm. The Mean Shift algorithm is based on iterative computing of the mean shift vector and consistently actualizing kernel’s position. Here we examine the hybrid algorithm that is created by a combination of the two techniques: Belief Propagation [14] and the Mean Shift segmentation algorithms [20]. This approach combines the advantages of both segmentation methods. The Mean Shift algorithm is fast and Belief Propagation is very accurate segmentation. Examples of hybrid algorithms are in [16]. Shown in Figure 3, first, we apply image filtering by Mean Shift algorithm. This step is very useful for noise removing, smoothing, and image segmentation. For each pixel of an image, the set of neighboring pixels is determined. Secondly, the image is split into segments using Mean Shift algorithm. In the third step, means of segments are retrieved by applying mean shift theory. Fourth, the small segments are merged together to the most similar adjacent segments by the Belief propagation method. Finally, we have integrated our proposed hybrid segmentation algorithm with the Sum-of-Absolute-Differences (SAD) stereo matching algorithm. This hybrid SAD method (HSAD) is able to produce highly accurate depth map.

2) **Disparity Calculation**

In the process of disparity computation where input images are segmented first and then the same matching points in the left and right images are found. This procedure plays a very important role in our proposed stereo system. This idea is illustrated for an arbitrarily located 3D point P in Fig. 4. Let a distant object is viewed by two cameras positioned in the same orientation but separated by a distance known as the baseline. Then, the object will appear in a similar position in both stereo images. The distance between the objects in left and right images is known as disparity d defined by (8), where xL and xR are x coordinates of the projected 3D coordinate onto the left and right image planes IL and IR [17].

\[
d = X_L - X_R = f \left( \frac{X_p + l}{Z_p} - \frac{X_p - l}{Z_p} \right)
\]

\[
Z_p = \frac{2fl}{d}
\]

Since the left and the right camera image planes are located in the same plane, y-coordinates of these two images are the same (yL=yR), and the disparity equals to the difference between the horizontal coordinates (xL-xR).

![Figure 4: A simple stereo system](image)

This means that once the disparity is computed, depth may in turn be found for the camera parameters: focal length f and baseline distance B=2l.

3) **Stereo Matching Algorithm**

A reconstruction of the disparity map from the left and right stereo pair is known as the stereo matching algorithm. The detection feature points must be matched. There exist several matching techniques based on various algorithms, e.g. Correlation (C), Normalized Cross Correlation (NCC), Sum of Squared Differences (SSD) and Sum of Absolute Differences (SAD) algorithms. The SAD algorithm is one of the simplest of dissimilarity measures of the left and right stereo images corresponding with square windows. Hence, the algorithm described in the next section, was chosen for the proposed algorithm.

1) **Sum of Absolute Differences**: It computes the intensity differences for each center pixel (i, j) in a window W(x, y) as follows:
\[ SAD(x, y, d) = \sum_{(i,j) \in W(x, y)} |I_l(i, j) - I_R(i - d, j)| \]

where IL and IR are pixel intensity functions of the left and right image, respectively. W(x, y) is square window that surrounds the position (x, y) of the pixel. The disparity SAD (x, y, d) calculation is repeated within the x-coordinate frame in the image row, defined by zero and maximum possible disparity dmax of the searched 3D scene. The minimum difference value over the frame indicates the best matching pixel, and position of the minimum defines the disparity of the actual pixel [17]. Quality of 3D disparity map depends on square window size, because a bigger window size corresponds to a greater probability of correct pixel disparity calculated from matched points, although the calculation gets slower [18].

2) Sum of Squared Differences (SSD)

The area-based SSD algorithm is similar to the previously described SAD algorithm. Instead of computing the absolute value, the SSD computes squares of the intensity differences as follows:

\[ SSD(x, y, d) = \sum_{(i,j) \in W(x, y)} |I_l(i, j) - I_R(i - d, j)|^2 \]

here IL and IR are pixel intensity functions of the left and right image, respectively. W(x, y) is square window that surrounds the position (x, y) of the pixel [19].

3) Feature Matching

The feature matching algorithm improves precision of the disparity calculation. This kind of algorithms extracts object’s suitable features in 3D scene, e.g., segments of edges or contours in the left and right stereo images. In the following stage, the disparity map is calculated from the corresponding points of the features. The matching algorithm based on the proposed hybrid segmentation is much faster, since a small portion of all the left and right images pixels are used for matching, while only precision is taken into account. Although note the calculated disparity map is sparse. A final depth map is then calculated from disparity map using the relation with disparity.

4) Depth estimation

Depth maps can also be estimated from motion parallax and relative-height. Motion parallax is one type of binocular depth cues. Because a video provides commonly motion parallax between the adjacent frames, motion parallax is one of the most commonly used depth cues in 2D to 3D video conversion. Nearby things pass quickly, while far off objects appear stationary. Thus, it is reasonable that near objects move faster across the retina than far objects. This is usually called the principle of depth from motion parallax [1]. According to the principle, dense motion vectors are estimated for calculating the depth map. Generally, the motion estimation can be classification into two categories: feature matching and block matching. In our work, the motion between two consecutive frames is estimated by feature matching as explained earlier. The relative-height information is an assistant depth cue for depth generation in our work. Depth from relative-height reconstructs the depth of the non-moving regions. Depth from relative-height overcomes the drawback that depth from motion parallax does not work for non-moving regions when the video is captured by the fixed camera. Relative-height in images denotes that the object locates at the bottom of the image plane is closer than the one at the top of the image.

5) Rendering

To synthesize left and right virtual views from depth, we use depth-image-based rendering (DIBR). DIBR has a lot of advantages over stereoscopic video generation compared with the other 3D warping methods [13-14]. The principle of DIBR compatible with existing 2D display systems. The object is projected the position \( Xc \) in original center image plane, and it is projected the position \( Xl \) in left image plane and \( XR \) in the right image plane [15]. The relationship between the positions in different image plane is described by the following equations:

\[
X_l = X_c + \frac{b}{2} \frac{f}{z} \\
X_R = X_c - \frac{b}{2} \frac{f}{z}
\]

where \( Z \) is the depth between the object and the image plane; \( f \) is the focal length; and \( b \) is the baseline between two virtual point; \( Xc, XR \) and \( Xl \) are the position where the point is projected in the center image plane, right image plane, and left image plane, respectively. DIBR renders a virtual view of any nearby viewpoint from the pixels of the original image. However, holes, i.e., disoccluded regions, appear in the virtual views. We simply fill the hole by using its neighboring pixels.

IV. IMPLEMENTATION

Using the stereo matching algorithm i.e. Sum of Absolute Differences and Sum of Squared Differences and the hybrid algorithm which are explained in the above in the proposed method we try to generate the disparity and depth map. The red color indicates the closest object to the camera while blue indicate the farthest object.

Implementation is in to following steps:

1) Left and right images
2) Depth map and color segmented image
3) Hole filling and 3D warping images

Step 1: Using stereo matching algorithm and SAD and hybrid algorithms we are trying to generate Depth map and Disparity map. Shown in below figure: Here is different images and their depth map.

Fig. i) Shows Left and right images
Fig. ii) Shows depth map
Fig. iii) Shows color segmentation

Take different images and check all the results shows the images below:
Step 1: Reference and target images:

Fig. a) Images of Tsukuba
Fig. b) Images of Cones
Fig. c) Images of Papers

Sequence Series of Original Images Reference and target images shown in Fig. a), b) and c).
Step 2: Depth map and Color segmentation:

Fig. d) Tsukuba
Fig. e) Cones
Fig. f) Paper

Sequence Series of Depth map and disparity map shown in Fig d), e) and f)

Fig. g) Tsukuba
Fig. h) Cones
Fig. i) Papers

Sequence Series of Color segmentation shown in Fig g), h), and i)
Step 3: Hole Filling and 3d warping.

Fig. j) Left Image
Fig. k) Right image
Fig. l) Left Image
Fig. m) Right Image
Sequence Series of Hole filling & warping shown in Fig j), k),l) and m)

Fig. p) Tsukuba
Fig. q) Cones
Fig. r) Teddy

Sequence Series of 3D images shown in Fig p), q), r), x), y) and z).

Fig. x) Jennifer
Fig. y) Hat Girl
Fig. z) Harry Potter

V. CHALLENGING ISSUES

Even though much research has been done to enable automatic 2D-to-3D conversion, the techniques are still far from mature. Most available products and methods are only successful in certain circumstances. In addition to the limitations imposed on each approach, the following are some key challenging issues to be solved and some of not solved.

1) One issue that directly affects the image quality is the occlusion/disocclusion problem during the generation of the stereoscopic images. In that camera is horizontally located so any new scene information is generated in view current viewpoint the texture information is missing in the current frame to another frame [2]

2) The depth ambiguity from monocular depth cues is one issue that impacts the depth quality. [2] The solution to solve the depth ambiguity of moving objects is to use additional depth cues, such as depth from geometrical information, to generate the depth information for each video frame.

3) The integration of various depth cues is another issue affecting the success of automatic 2D-to-3D video conversion. To retrieve the depth from such video sequences, different strategies of depth generation are required. The challenge is how to integrate all the extracted depths from different cues to form, not only spatially, but also temporally stable and reasonable depths. From the literature, only some research on the integration of various depth cues in the spatial domain has been proposed. However, more investigations are still required to provide spatially and temporally consistent depths.

4) The real-time implementation of 2D-to-3D conversion is also a critical issue for the adoption of the proposed techniques by the general public. The availability of real-time 2D-to-3D conversion will allow broadcasters and consumer electronics (CE) manufacturers to remove the natural fear from users of not having enough content for their new 3D-enabled TV set. Real-time, however, adds a new hard constraint that is difficult to meet while maintaining a high quality standard for the converted material. There are several real-time implementations incorporated into TV sets (e.g., Samsung’s 3D-TVs), sold as stand-alone equipment (e.g., JVC’s IF-2D3D1 Stereoscopic Image Processor), or incorporated into software packages (e.g., DDD’s TriDef-Player). However, the quality of the resulting stereoscopic images, with respect to the depth sensation, is still an outstanding issue that requires more research. [2]

5) The major challenge of 2D-to-3D video conversion lies in the disparity estimation of monocular videos, which is closely related to the scene depth estimation.

VI. APPLICATIONS

[1] The three-dimensional (3D) displays provide a dramatic improvement of visual quality over the 2D displays. The conversion of existing 2D videos to 3D videos is necessary for multimedia application.

[2] 3D Display have many Applications which are broadcasting, movies, gaming, photography, camcorders, and education.

[3] That 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

[4] 2D videos to 3D videos is necessary for multimedia application.

[5] Some commercial software, such as DDD’s TriDef 3D player and Samsung’s 3DTV, can generate stereoscopic views from monocular videos in real-time.

[6] The three-dimensional (3D) displays provide a dramatic improvement of visual quality over the 2D displays. The conversion of existing 2D videos to 3D videos is necessary for multimedia applications.
VII. CONCLUSION AND FUTURE WORK

In this thesis describes a procedure for recovering depth map based on the proposed hybrid segmentation algorithm using process filtering. From the Dissertation Phase results one can see a disparity map, which presents displacements between two images and finally is used to estimate the depth value. The algorithms based on the Sum of Squared Differences (SSD) and Sum of Absolute Differences (SAD) can find depth map directly. Depending on the matching techniques based on algorithms SAD and SSD a compromise between runtime and quality needs to be chosen. Finally, the SAD stereo matching algorithm is selected for the implementation of further experiments, because the SSD algorithm requires almost double time. The proposed system gives combination of both segmentation algorithm Mean Shift Algorithm and Belief Propagation are fact and accurate segmentation. And then we using rendering method that is DIBR there is many advantages of using DIBR method for depth generation. Which we will see in our future work. DIBR has a lot of advantages over stereoscopic video generation compared with the other 3D warping methods. Our thesis is partially completed we are working on that to conversion of 2d to 3d video conversion. With the development of more advanced techniques for 2D-to-3D video conversion, the vast collection of 2D material currently available will be converted into stereoscopic 3D to boost the general public interest in purchasing 3D displays and 3D TV services.

ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude to my guide Prof. S.A Vohra for the invaluable guidance and moral support rendered by him in the fulfillment of my dissertation work. I wish to express my deep sense of gratitude to my help and guidance this Thesis would not have materialized. I am also thankful to all my friends who supported me during the preparation of my work. And my sincere acknowledgement is to Internet for inestimable help in providing me the necessary information and help for the literature survey work. Finally, yet importantly, I would like to express my heartfelt thanks to my beloved parents for their blessings and wishes for the successful completion of this Colloquium.

REFERENCES

[4] Zhenyao Li,Xun Cao,Qionghai Dai “ A Novel Method for 2D to 3D video conversion using Bi- Directional motion estimation” – ICASSP 2012.
[6] Chao –Chung Cheng,Chung- Te Li, Po-Sen Huang,Tsung-Kai Lin,Yi-Min Tsai,and Liang-Gee Chen “A Block-based 2D to 3D Conversion System with Bilateral Filter”.
[15] BENCO, M., HUDEC, R. The advances image segmentation techniques for broadly useful retrieval in large image database. In NSSS IX. TatranskeZruby (Slovakia), 2006,
