An overview of free viewpoint Depth-Image-Based Rendering (DIBR)

Wenxiu SUN, Lingfeng XU, Oscar C. AU, Sung Him CHUI, Chun Wing KWOK
The Hong Kong University of Science and Technology
E-mail: {eeshine/lingfengxu/eeau/shchui/edwinkcw}@ust.hk

***Abstract*—3D Video has caught enormous attention in consumer market due to the progress in technical maturity of 3D displays, digital video broadcasts and computer vision algorithms. Video-plus-depth representation makes the concept of free viewpoint video feasible allowing the user to freely navigate within real world visual scenes. In this paper, we review the existing methodologies and 3D video processing framework. Previous image-based rendering techniques can be classified into three categories according to the amount of geometric information being used, of which DIBR combines the advantages of the two extremes (geometry based and image based representations). We then discuss the techniques for depth map creation, the method for rendering given the depth information and post-processing for corresponding artifacts in the multi-view adaptation. We also issue the quality measurement of the generated depth and synthesized view. In respect to the clear demand from industry and user side for such new types of visual media, there is still a large improving room in the 3D data representation and rendering quality of 3D perception.**

I. Introduction

3D video as a new type of visual media nowadays has highly expanded the user's sensation over the traditional 2D video. With the development of different display technologies and the requirement of user's visual enjoyment to the real 3D world, how to represent and render the realistic 3D impressions is a research area with excellent prospects and great challenges. A viewer will create a 3D depth impression if each eye receives its corresponding view. These views must correspond to images taken from different viewpoints with human eye distance. 3D impression is attractive in applications such as medical imaging [1][2], multimedia services [3] and 3D reconstruction [4].

Different display technologies also encourage the applications into different aspects. For instance, in a 3D cinema theater, the viewer is supposed to wear polaroid glasses without much possibility to move around. All viewers sitting at different seats have the same 3D impression. 3D cinema with display technology based on glasses is therefore expected to grow further and this will also increase the acceptance and create demand for 3DV applications at home. In a living home environment, however, the user's expectation is quite different. When the user moves through a scene, he/she should have different 3D impression like looking from different viewpoints in a real 3D scene, which is referred to as motion parallax. Nowadays, in an autostereoscopic display system, several images are emitted at the same time and the technology ensures that users only see a stereo pair from a specific viewpoint without the necessity to wear glasses. When moving around, a natural motion parallax impression can be supported if consecutive views are arranged properly as stereo pairs.

However, transmitting 9 or more views of the same 3D scene is extremely inefficient. Earlier in 1990s, MPEG-2 proposed multi-view profile which was included in ITU-T Rec. H.262/ISO/IEC 13818-2. Besides, MPEG organized a group named 3DAV (3D Audio-Visual) which was devoted to researching the technologies of 3D audio and video in 2001. One of the most practical approaches appeared in the following years was to use multiple views to represent a 3D scene. Due to the large amount of inter-view statistical dependencies among these multiple views, it needs to be implemented by exploiting combined temporal/inter-view prediction, referring to Multi-view Video Coding (MVC) [5].

Apart from MVC, conventional color video plus an associated depth map is another popular format. Further more, multiview-plus-depth (MVD) gives more flexibility on high-quality 3D video rendering and viewpoints covering which do not lie on the camera baseline [6]. At the receiver side, virtual views are interpolated using Depth-Image-Based Rendering (DIBR) from the transmitted MVD data.

3DV is an advanced system based on MVD and DIBR which is now in process [7]. This coding format allows rendering 9 or more views out of two or three views. Although DIBR provides backward compatibility and scalability to the traditional 2D video, there is still the possibility that part of the synthesized views cannot be correctly rendered due to occlusions and depth map accuracy.

The 3D video representation is of central importance for the design of 3DV and FVV systems [9]. More details will be presented in Section II. The success of 3DV concept highly relies on the high-quality view synthesis algorithms. And the accuracy of depth map is crucially important. The depth map creation techniques are presented in section III. Section IV shows the general formulation of DIBR and 3D warping and associated quality improvements of synthesized view. Quality Assessment method of the 3D system is discussed in Section V. Finally, Section VI concludes the paper.

II. 3D video representation

3D representation sets the requirements of scene acquisition and signal processing methods, eg. camera setting, data arrangement, sender-side and receiver-side data processing [10]. On the other hand, the 3D representation determines

Proceedings of the Second APSIPA Annual Summit and Conference, pages 1023–1030,
Biopolis, Singapore, 14-17 December 2010.

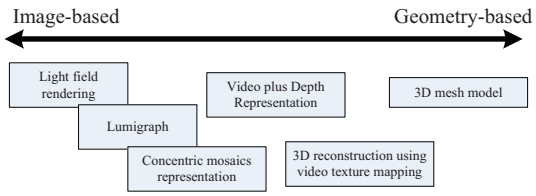


Fig. 1. 3D representation Categories used in this paper

the rendering algorithms as well as coding and transmission. The previous work [11][12] of 3D representation and the associated rendering techniques can be classified into three categories, namely representing with explicit geometry, with implicit geometry and without geometry, as depicted in Fig. 1 .

A. Geometry based representation

At one end of rendering is spectrum with explicit geometry which is represented by classical 3D computer graphics. It has direct 3D information encoded in it, mostly on the basis of 3D meshes. Real world objects are reproduced using geometric 3D surfaces with an associated texture mapped onto them. Unfortunately, computer vision techniques are generally not robust enough to work for textureless regions without prior structural knowledge. In addition, it is very difficult to capture complex visual effects which can make the 3D models more realistic such as highlights, reflections and transparency using a texture mapped 3D model.

B. Image based representation

Another extreme of rendering spectrum is image based representation and rendering which does not require any geometric information or correspondence. In this case, virtual intermediate views are generated from an over-sampling image sets to counter undesirable aliasing effects in output display, such as light field rendering [21], lumigraph [22] and concentric mosaics representation [23]. The main advantage is a potentially high quality of virtual view synthesis avoiding any 3D scene reconstruction. However, the advantage has to be paid by dense sampling of the real world (over-sampling) with a sufficiently large number of camera view images. And a tremendous amount of images has to be processed and transmitted in order to achieve high-quality rendering. In other words, if the number of captured views is too small, interpolation artifacts such as disocculusions will be obvious, possibly affect the synthesized quality.

C. Depth-plus-video concept

In between of the two extremes, some image-based rendering systems rely on implicit geometry [24]. Implicit expresses the fact that geometry is not directly available but on geometric constraints. The geometric constraints can be of the form of known depth value at each pixel, epipolar constraints between pairs of images, or trifocal/trilinear tensors that link correspondences between triplets of images. See Fig. 2 of depth map.

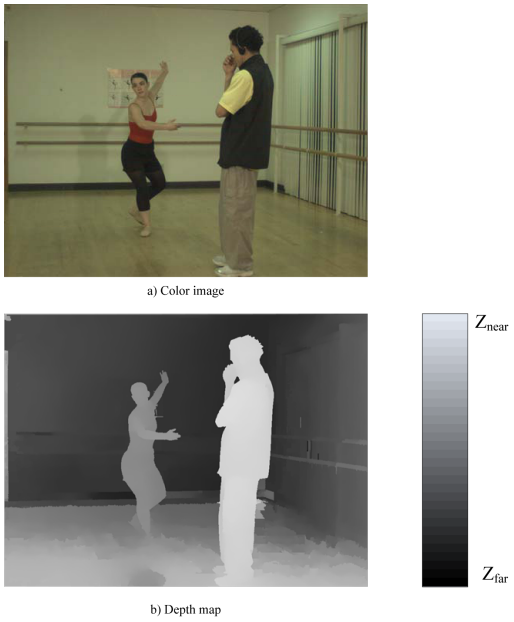


Fig. 2. 3D data representation using video-plus-depth format

These methods make use of both approaches and combine the advantages in some ways. Previous research on warping using DIBR from one reference image has two inherent limitations, which are viewpoint dependency of textures and disocclusions [25]. To overcome these limitations, most recent methods [26][27][28][29][30]employ warping from two surrounding reference images to a virtual viewpoint. Disocclusions from one reference view are compensated by the other view to minimize errors. Zitnick et al. [27] pointed out that the three main challenges in rendering a high quality virtual viewpoint. First of all, empty pixels and holes due to insufficient sampling of the reference images need to be interpolated. Secondly, pixels at borders of high discontinuities tend to cause contour artifacts which need to be fixed. Thirdly, the remaining disocclusions after blending the projected images (this area can not be viewed from any of the two reference views) need to be generated. The details will be explained more explicitly in Section IV.

III. Depth Map Creation

One of the most challenging tasks for DIBR is to estimate accurate depth maps from stereo images. Most algorithms for depth generation make assumptions of epipolar geometry and stereo camera calibration. In this section, the stereo camera model together with the disparity equation will be presented. Then a framework of depth generation including rectification, stereo matching, disparity calculation ,etc will be covered.

A. Stereo camera model

Fig. 3 is the stereo camera model with C_0 and C_1 as the camera centers and I_1 and I_2 as their corresponding image planes. 2D points m_1 and m_2 are the projections of an arbitrary 3D space point M on the image planes. Based on

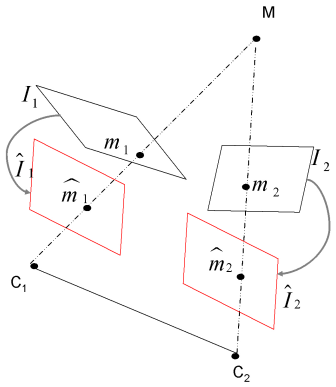


Fig. 3. Stereo camera model: black - original views; red - rectified parallel views

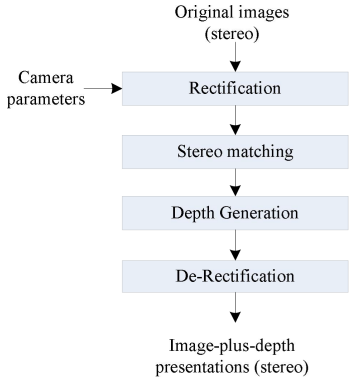


Fig. 4. Block diagram of stereo matching processing

the assumption of pinhole camera model, the two projective equations [13] are :

$$m_1 = \frac{1}{Z} K_1 \cdot R_1[I|-C_1]M \quad (1)$$

$$m_2 = \frac{1}{Z} K_2 \cdot R_2[I|-C_2]M \quad (2)$$

Where m_1, m_2 and M are symbolized by homogeneous notations and K_1, R_1, C_1 are the intrinsic parameter matrix, rotation matrix and shift matrix for the first camera, so are K_2, R_2, C_2 for the second camera. Z is the depth of the 3D point M. Take the coordinate system of camera 1 as the world coordinate system, the above two projective equations can be simplified into:

$$m_1 = \frac{1}{Z} K_1 \cdot [I|0]M \quad (3)$$

$$m_2 = \frac{1}{Z} K_2 \cdot R[I|-C]M \quad (4)$$

R and C in Eq. (3)(4) are the rotation and shift matrices of the second camera referring to the first.

Substitution of Eq. (3) to Eq. (4) then leads to the disparity equation, which is:

$$Zm_2 = ZK_2RK_1^{-1}m_1 + K_2C \quad (5)$$

Disparity equation shows the relationship of m_1 and m_2, which are the coordinates in I_1 and I_2 respectively.

B. Depth matching

Fig. 4 describes the framework for depth creation from stereo images. The input stereo images are rectified initially so that the corresponding points can be searched along the scan line in the following stereo matching. After that, a suitable stereo matching algorithm is performed to get the disparity map. Using the disparity equation derived above, the depth information is calculated for the rectified image. And finally, the depth map will be de-rectified in order to achieve the depth corresponding to the original images.

The advantage of rectification is that 2D searching problem in stereo matching is converted into 1D searching, which reduces the computation significantly [15]. As illustrated in Fig. 3, the view planes are only rotated and the camera centers are not shifted. After adjusting the intrinsic matrix, the disparity equation can be simplified into

$$Zm_2 = Zm_1 + KC \quad (6)$$

Then, we can derive the following equation directly

$$Z = \frac{KC}{m_1 - m_2} \quad (7)$$

Eq. (7) means that, the depth can be obtained as long as we know the disparity m_1 - m_2. In order to search for the point correspondences along the scan lines, stereo matching (or called disparity matching) is applied to the stereo pair. A tremendous amount of algorithms have been developed for stereo correspondence such as window based HRM [14], max-flow [16], graph cut [17], belief propagation [19], dynamic programming [18] , etc.

In general, those stereo matching algorithms can be classified into two groups [20]: local algorithms and global algorithms. The local algorithms calculate the matching cost in a supporting region such as the square window, and then simply choose the disparity associated with the minimum cost value. Some commonly used matching costs include squared intensity difference, absolute intensity difference, and normalized cross-correlation. Local algorithms usually make the assumption of smoothness implicitly while global algorithms make explicit smoothness assumptions. The objective of global algorithm is to find the disparity function d which minimizes the global energy [20],

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \quad (8)$$

E_{data} measures how well the disparity function d agrees with the input image pair and E_{smooth} measures the extent to which d is not piecewise smooth. Usually the squared difference is chosen for E_{data} [17], and the choice of E_{smooth} is a critical issue and diversity of functions are proposed such as [17][18].

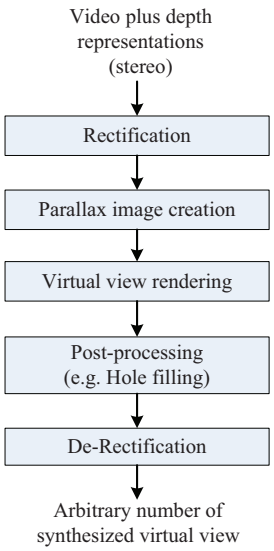


Fig. 5. Block diagram of view synthesis processing.

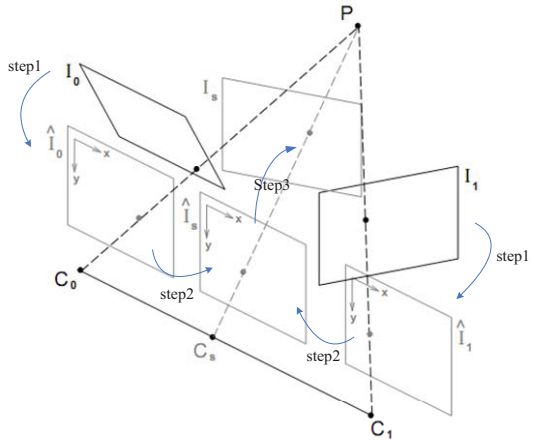


Fig. 6. Projecting Models in view synthesis.

After the step of stereo matching, a rough disparity map containing ambiguities and occlusions is derived. Ambiguities are caused by the ill-posed matching in the textureless or periodic areas, which can be reconstructed by segmentation based interpolation [34]. In contrast, occlusion happens when points correspondences do not exist at all. One possible solution to the problem is to extrapolate the background layer to the occluded area based on the assumption that the disparity are spatially consistent in the color segments and stationary background [34]. Finally, depth map Z can be calculated by Eq. (7).

IV. Depth Map Based Rendering

A. Rendering model

Depth-Image-Based Rendering (DIBR) is the process of synthesizing virtual views of the scene from captured color images or videos with associated depth information [31][33]. Generally, the synthesize approach can be understood by first warping the points on the original image plane to the 3D world coordinates and then back-projecting the real 3D points onto the virtual image plane which is located at the required viewing position defined by user at the receiver side.

For DIBR purposes, it is assumed that all camera calibration data, the quantization parameters and the 3D position of the convergence point M_c are transmitted as metadata and are known at the receiver side [34]. One requirement is that both the multiview capturing system and rendering system are on the same 3D coordinate such that the relative positions between the real cameras of the capturing system and the virtual cameras of the 3D display system are well defined for further processing.

Based on the geometric relations, the rendering process chain follows the steps as depicted in Fig. 5.

Kauff et. al. [34] use a real view and a virtual view to do the image rectification. In order to render the virtual view V_{ij}, They first select the closest view C_i out of the N available camera views in the first step. This camera view C_i and the associated depth map D_i are rectified together with the virtual view V_j, resulting in one rectified image pair (C_{ij}, V_{ji}). Then a parallax map is derived from the rectified depth map and associated information (eg. the convergence point M_c, baseline, and sensor shift) and the virtual view is calculated on the basis of rectified color image and parallax map. The virtual view is then de-rectified to fit into the display 3D system again. In case of disocclusions, the virtual view generation process is repeated for the next but one nearest view C_k to reconstruct the missing data in the disoccluded area.

Seitz and Dyer's view morphing technique [24] directly uses two camera views as reference as shown in Fig. 6. They first warp the views to the same plane so that the two views are parallel to each other, then reconstruct any viewpoint on the line linking two optical centers of the original cameras (known as baseline). Intermediate views are exactly linear combinations of two warped views only if the camera motion associated with the intermediate views is parallel to the reference views. If not, a pre-warp stage can be used to rectify the intermediate virtual images.

Usually, the displaced points will not lie on the pixel raster of the output view after 3D warping and the novel view must be carefully resampled [35]. Fortunately, the resampling does not need to be operated on the two-dimensional space but rather one-dimensional in the horizontal line based on the rectified images. The main discriminating feature of the different resampling approaches lies in the interpolation kernel (eg. nearest neighbor, linear, cubic convolution, cubic spline or sinc function).

B. Post-processing

During the generation of virtual view, it can happen that two different original points are warped to the same integral point location in the virtual image. This situation occurs when one of the two points are occluded by the other in the 3D world

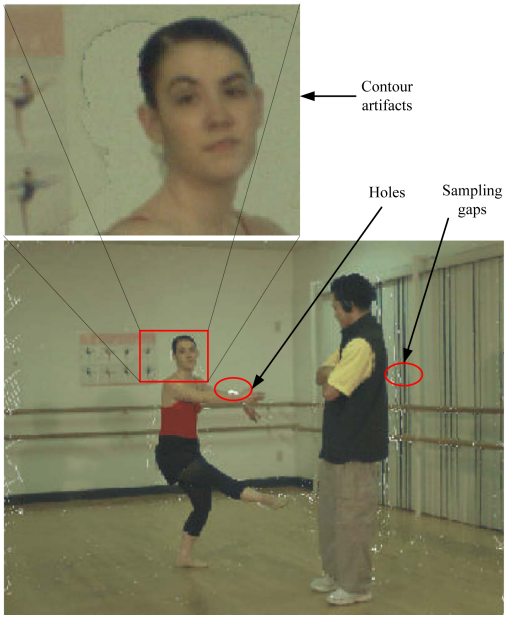


Fig. 7. Visual artifacts in the synthesized view before post-processing.

coordinate. One simple way to resolve the problem is to always choose the front point to warp in the virtual view which is the so called occlusion-compatible warp order [31]. Visibility can still be resolved in this way for more general camera configurations, as long as the image planes stay parallel. The correct order in this case depends on the position of the epipole in the new image [32].

Several visible artifacts after 3D warping may appear as shown in Fig. 7. These artifacts need to be removed in the post-processing step.

Holes in the new view occur if the new viewpoint uncovers previously invisible scene points. Gaps due to the forward-mapping process and the real holes caused by occlusion boundaries in the disparity map have to be distinguished carefully.

1) Filling sampling gaps: Sampling gaps occurs when the small disparity difference between adjacent pixels is reprojected to the virtual view. Holes are caused by large difference between adjacent pixels (depth discontinuity) in the disparity map. Since depth map is discrete, distinguish between the two cases can present a problem [37]. One possibility is to impose a disparity gradient limit that acts as a threshold. For example, a gradient limit of 1 would mean that if two neighboring disparity values differ by an amount $d \leq 1$, then they are considered to belong to the same object and forward mapping can create a sampling gap which needs to be filled. If they differ by $d > 1$, on the other hand, they would be considered to be separated by an occlusion boundary and thus forward mapping can create a hole.

Typical techniques developed to deal with sampling gaps and holes are: (a) replacing the missing area during the view synthesis with "useful" color information, or (b) preprocessing of the depth information in a way that no disocclusions appear in the "virtual" view [35]. In the IST (European Information Society Technologies) project ATTEST (Advanced Three-Dimensional Television System Technologies) 3D-T, depth information is preprocessing (smoothing) with a suitable Gaussian filter in a way that no disocclusions occur to achieve visually less perceptable impairments. Scharstein [37] noticed that decreasing sampling gaps can be achieved by increasing the sampling rate proportionally to the distance of the new camera to the reference camera. Luat Do et. al [38] proposed to process the projected depth map with a median filter. Depth maps consist of smooth regions with sharp edges, so that median filtering will not degrade the quality. Afterwards, they compare the input and output of the median filter and perform an inverse warping when pixels have changed.

2) Filling holes: Holes can be largely reduced by combining the information from both reference images. Although displaying the identical view from two reference views, these two images can differ in the following ways: (a) The global intensities can be different due to different camera characteristics of the original two cameras. (b) The quality can be different due to the different distortions created by the two warps. (c) The holes (i.e., locations of previously invisible scene points) are at different positions.

To compensate for the first two effects, it is useful to blend the intensities of the two images, possibly weighting the less-distorted image more or the one closer to the new viewpoint. For example, the weights could be proportional to the distance between the virtual viewpoint and the reference viewpoint. Partially occluded holes could be filled using the information from the other reference views. However, after blending the two projected images into the virtual images, disocclusions may still occur. These are areas that cannot be viewed from any of the the reference cameras. Some inpainting techniques for filling those missing areas are developed to reconstruct small regions of an image. These regions can be inpainted by texture extrapolation while maintaining the structure of the textures. An easy way is to spread the intensities of the neighboring pixels. But this often yields "blurry" regions. A different approach is to mirror the intensities in the scanline adjacent to the hole which gives noticeable better result than simple intensity spreading. It is very important to prevent intensities from being spread across occlusion boundaries that is to avoid smearing of foreground and background [38]. More sophisticated texture synthesis methods based on neighboring intensity distributions are possible. For example, those developed in the context of image restoration [39].

3) Contour artifacts removal: Contour artifacts are caused by the mixing of foreground and background color at the discontinuity of depth map. Muller et al [7] separate input images in reliable and unreliable areas based on edge detection in high-quality depth images, since that edges correspond to depth discontinuities. Reliable and unreliable image areas are treated separately and the results are merged depending on reliability criteria.

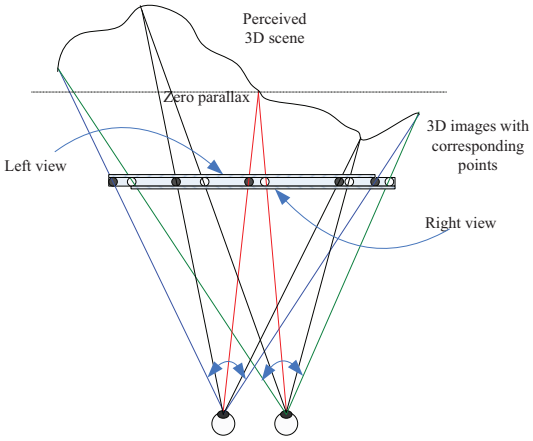


Fig. 8. 3D perception in the rendering system.

TABLE I
CONTROL PARAMETERS FOR DEPTH PERCEPTION

Parameter	±	Parallax	Perceived depth	Object size
Interaxial distance t_c	+	Increase	Increase	Constant
	−	Decrease	Decrease	Constant
Focal length f	+	Increase	Increase	Increase
	−	Decrease	Decrease	Decrease
Convergence distance Z_c	+	Decrease	Shift(fore)	Constant
	−	Increase	Shift (aft.)	Constant

C. User control of depth parameter

Table I shows how the virtual 3D reproduction can be influenced by the choice of these main system parameters [35].

Two different depth parameters can be controlled by the user in the framework of the rendering process to adapt the 3D reproduction to the viewing conditions and individual preferences [34].

The first one is the interaxial distance t_c between two views to be rendered in 3D system. By changing the parameter, the position of the virtual views are shifted along the baseline. An increase of t_c will increase the depth impression and vice versa.

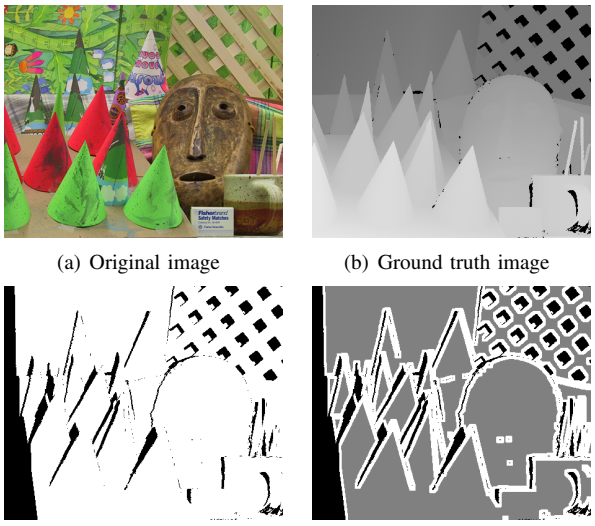
The Convergence distance Z_c represents a second, very important depth parameter. It can be used to shift the position of the 3D scene relatively to the display surface. Note that the changing of Z_c only affects the rectification processes of the whole chain but not the depth structure. Thus, Z_c can be changed afterwards during rendering.

V. Quality Assessment

In this section, we describe the quality assessment for evaluating the performance of stereo matching and image rendering. Some commonly used datasets are also introduced here.

A. Assessment for stereo matching

The most commonly used approach to the assessment of stereo matching is to compute the error statistics with respect to the ground truth depth. Usually the following two measurements [20] with the reference of ground truth depth are calculated.



(a) Original image (b) Ground truth image

(c) Non-occluded region (white) (d) Depth discontinuities (white)

Fig. 9. Segmented region maps

1. Percentage of bad matching pixels:

$$B = \frac{1}{N}\sum_{(i,j)} f_{cnt}(|d_C(i,j) - d_T(i,j)|), \tag{9}$$

$$where\ f_{cnt}(x) = \begin{cases} 1, if\ x > T_B \\ 0, else \end{cases}$$

where $d_C(i,j)$ is the computed depth map, d_T is the ground truth depth, T_b is a error tolerance and N is the total number of pixels.

2. Root mean square error (RMS):

$$R = \left(\frac{1}{N}\sum_{(i,j)} (d_C(i,j) - d_T(i,j))^2\right)^{\frac{1}{2}} \tag{10}$$

In order to describe these statistics more accurately, four different kinds of regions [20] are focused here, as shown in Fig. 9:

- Non-occluded regions: regions that are not occluded in the matching images;
- Depth discontinuity regions: pixels whose neighboring disparities differ by more than disparity jump threshold;
- Textureless regions: regions where the average intensity gradient is below a given threshold;
- All regions: all the pixels in the image.

The Middlebury computer vision page [36] provides the evaluation for different kinds of stereo matching. 21 datasets obtained by the technique of structured light [40] are also provided in that site.

B. Assessment for image rendering

1) Subjective Quality Assessment: The objective measures of rendered left and right views may not account for the perceptual factors such as depth reproduction, stereoscopic impairments, and visual comfort. Therefore, subjective evaluations which incorporate human observers need to be carried out to get a true representation of human 3D perception. ITU-Recommendation BT.1438 [41] describes the methodology of carrying out appreciation oriented subjective evaluation tests for stereoscopic television applications. In the explorative study described in this paper, subjective tests are performed in order to obtain human feedback on the effect of different combinations of symmetric/asymmetric coding levels and different packet loss rates. The feedback provides information on the perceived quality of color and depth map based stereoscopic video. The perceived quality of the reconstructed binocular video is measured using two image attributes, namely perceived overall image quality and depth.

2) Objective Quality Assessment: The accuracy of a synthesized image I is quantified as the registration error with respect to a reference image I' [42]. An image I is represented by a set of pixels $p \in I$ and the registration error at each pixel is computed as the minimum distance to a similar pixel in I'. The error in view synthesis is now characterized by the distribution of pixel-wise error distances $d(p, I')$.

$$d(p, I') = \|p - p'\|_2, max \prod_{p' \in I'} S(I(p), I'(p')) \qquad (11)$$

S(.) defines the similarity. A single error metric can be defined using the root mean square error (RMS) across the entire image. However, a simple mean can mask visually distinct errors in highly structured regions which is the case that mean errors are small while visual effects are bad. The Hausdorff distance is adopted to measure the maximum distance from image I to the reference I'.

$$d(I, I') = max \prod_{p \in I} (p, I') \qquad (12)$$

In practise the Hausdorff metric is sensitive to outliers and the generalized Hausdorff distance is taken as the k-th ranked distance in the distribution.

$$d^k(I, I') = Q^k_{p \in I} d(p, I') \qquad (13)$$

Intuitively the distance measure is related to the geometric error in the underlying geometry of the scene. The metric is however specifically tailored to measure the registration of distinct image regions where the effect of geometric error is most apparent to an observer.

VI. Conclusions

3D video service is attracting more and more attentions in the recent years, especially after the extremely hot movie 'Avatar'. In this paper, a DIBR based free viewpoint system which allows the user to interactively control the viewpoint and generate new virtual views of a dynamic scene from any 3D position is presented. We first briefly introduce three common used 3D video representations and then give a detailed description for depth-plus-video structure. Then we outline the complete analysis for depth map generation and depth-map-based image rendering and give a brief introduction of the most important algorithms for them. Depth map quality assessment as well as the synthesized view quality assessment is depicted finally.

There are still many open questions we would like to address here. During stereo matching, what kind of shiftable/adaptive windows works better? How to generate more accurate depth in the regions of occlusion and textureless areas? Is the existing quality assessment useful enough for gauging the quality of stereo matching as well as image rendering? Also, high quality view synthesis makes the concepts of auto and multiview stereoscopic 3D displays feasible. But it is known that annoying artifacts (eg. holes) along depth discontinuities may occur during rendering. Errors in depth map generation and camera calibration parameters will cause mis-registration errors. Transparent and reflectance surfaces will also introduce problems.

By reviewing the existing methodology and the framework depicted in this paper, we hope to give the readers a deeper understanding of IBDR.

Acknowledgment

This work has been supported in part by the the Research Grants Council (GRF Project no. 610210) and the Hong Kong Applied Science and Technology Research Institute Project (ART/093CP).

References

[1] S. Zinger, D. Ruijters, and P. H. N. de With, "iGLANCE project: free-viewpoint 3D video," *17th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG),*2009.

[2] S. Zinger, D. Ruijters, "iGLANCE: transmission to medical high definition autostereoscopic displays," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video,*2009.

[3] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Processing Magazine,* 24(6),pp.10-21,2007.

[4] C. Leung and B. C. Lovell, "3D reconstruction though segmentation of multi-view image sequences," *Workshop on Digital Image Computing*,1,pp. 87-92,2003.

[5] P. Merkle, K. Mller, A. Smolic, and T. Wiegand,"Efficient compression of multi-view video exploiting inter-view dependencies based on H.264/MPEG4-AVC," *IEEE International Conference on Multimedia and Exposition,*2006.

[6] S. Shinya, K. Hideaki and Y. Ohtani,"Real-time free-viewpoint viewer from Multi-video plus depth representation coded by H.264/AVC MVC extension," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video,*1-6,2009.

[7] K. Mller, A. Smolic, K. Dix, P. Merkle, P. Kauff, T. Wiegand, et al. "View Synthesis for Advanced 3D Video Systems," *EURASIP Journal on Image and Video Processing,* 1-12,2008.

[8] "Introduction to 3D video," *ISO/IEC JTC1/SC29/WG11 coding of moving pictures and audio/N9784, Archamps, France,*2008.

[9] A. Smolic, P. Kauff,"Interactive 3D Video Representation and Coding Technologies," *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery 93(1),* 2005.

[10] A. Smolic, "An Overview of 3D Video and Free Viewpoint Video," *Signal Processing,* 1-8,2009.

[11] R. Szeliski, and P. Anandan, "The geometry-image representation trade-off for rendering," *Proceedings 2000 International Conference on Image Processing,* 13-16 vol.2,2000.

[12] H. Shum, "Review of image-based rendering techniques," *Proceedings of SPIE,* 2-13, 2000.

[13] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press,*,2000,pp. 239-259.

[14] Atzpadin, N. Kauff, and P. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Transactions of Circuits and Systems for Video Technology ,*, March 2004,pp. 321- 334.

[15] A. Fusiello, E. Trucco, A. Verri, A. Fusiello, E. Trucco, and A. Verri, "Rectification with Unconstrained Stereo Geometry," *British Machine Vision Conference*,1997

[16] Roy, S. Cox, and I.J., "A maximum-flow formulation of the N-camera stereo correspondence problem," *IEEE Transactions of Circuits and Systems for Video Technology ,*, March 2004,pp. 321- 334.

[17] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence*, Nov 2001,pp.1222-1239.

[18] A. Bobick, and S. Intille , "Large occlusion stereo,"*International Journal of Computer Vision,*,1999

[19] J. Sun, NN. Zheng, and HY. Shum, "Stereo matching using belief propagation," *Pattern Analysis and Machine Intelligence*,July 2003,pp. 787- 800.

[20] D. Scharstein, and R. Szeliski,"A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*,2002.

[21] M. Levoy, and P. Hanrahan, "Light field rendering," *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH,* 31-42,1996.

[22] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH'96,*43-54,1996.

[23] H. Shum, and L. He, "Rendering with concentric mosaics," *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99,* 299-306,1999.

[24] S. M. Seitz, and C. R. Dyer, "View morphing," *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96,* 21-30,1996.

[25] M. M. Oliveira, "Relief texture mapping," *PhD thesis,University of North Carolina,* 2000.

[26] Y. Morvan, "Acquisition, Compression and Rendering of Depth and Texture for Multi-View Video," *PhD thesis, Eindhoven University of Technology,*2009

[27] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics,* 23(3), 600,2004.

[28] K. Oh, S. Yea, and Y. Ho, "Hole-Filling Method Using Depth Based In-Painting For View Synthesis in Free Viewpoint Television (FTV) and 3D Video,"*Picture Coding Symposium,* 2009.

[29] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," *Signal Processing: Image Communication,* 24(1-2), 65-72,2009.

[30] A. Smolic, K. Mller, K. Dix, P. Merkle, P. Kauff, T. Wiegand, et al. "Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems," *Image Processing,* 2008.

[31] L. McMillan, "An Image-Based Approach to Three-Dimensional Computer Graphics," *PhD thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA,* 1997.

[32] L. McMillan, G. Bishop, "Plenoptic modeling: An image-based rendering system," *Computer Graphics (SIGGRAPH'95),* 39-46, 1995.

[33] W. R. Mark, "Post-Rendering 3D Image Warping: Visibility, Reconstruction, and Performance for Depth Image Warping," *PhD thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA,* 1999.

[34] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Processing: Image Communication,* vol. 22, pp. 217-234, 2007.

[35] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," *Proc. Stereoscopic Displays Appl,* pp. 93-104, 2002.

[36] Middlebury Computer vision: *http://www.middlebury.edu/stereo*

[37] D. Scharstein "View Synthesis using stereo vision," *PhD thesis,Cornell University,* 1997.

[38] L. Do, S. Zinger, and P.H. de With, "Quality improving techniques for free-viewpoint DIBR," *Solutions,* 2010.

[39] A. C. Kokaram and S. J. Godsill. "A system for reconstruction of missing data in image sequences using sample 3D AR models and MRF motion priors," *Fourth European Conference on Computer Vision (ECCV'96),* 613-624, Cambridge,UK, 1996.

[40] D. Scharstein and R. Szeliski. "High-accuracy stereo depth maps using structured light,"*IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* June 2003, volume 1, pages 195-202

[41] Int. Telecommunication Union/ITU Radio Communication Sector, "Subjective assessment of stereoscopic television pictures," ITU-R BT.1438, Jan.2000.

[42] J. Starck, J. Kilner, and A. Hilton, "Objective Quality Assessment in Free-Viewpoint Video Production," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video,* 2008, pp. 225-228.