International Journal of
Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.04,Issue.27,
July-2015,
Pages:5228-5233

www.ijsetr.com

3D Model Distortion Estimation Based on Rendering Process

K. NAVEEN[1], P. KASTHURAIAH[2]
[1]PG Scholar, Dept of ECE(DSCE), KMM Institute of Technology and Science, Tirupati, AP, India,
Email: theppadevendertheppa@gmail.com.
[2]Assistant Professor, Dept of ECE, KMM Institute of Technology and Science, Tirupati, AP, India,
Email: ambica.mangalagiri@gmail.com.

Abstract: We propose a logical model to gauge the combined perspective quality in 3d feature. The model relates mistakes in the profundity pictures to the union quality, considering surface picture qualities, composition picture quality, and the rendering process. Particularly, we disintegrate the blend mutilation into composition mistake impelled twisting and profundity slip affected twisting. We dissect the profundity slip instigated mutilation utilizing a methodology consolidating recurrence and spatial space procedures. Test results with feature arrangements and coding/rendering apparatuses utilized as a part of MPEG 3dv exercises demonstrate that our logical model can precisely evaluate the union commotion power. Subsequently, the model can be utilized.

Keywords: Model Distribution.

I. INTRODUCTION

3-D video services are becoming reality in the consumer market thanks to recent improvements on high quality 3-D display technologies and 3-D content production skills, which is a significant step towards a more realistic multimedia experience. For the advanced 3-D video services, two application scenarios have been discussed; first, providing ability to vary the baseline distance for stereo video to adjust the depth perception to help avoiding fatigue and other viewing discomforts, and secondly, supporting auto-stereoscopic displays, which do not require for the viewer to wear special glasses to feel the depth. In both cases, multiple number of views are required at the display side. As 3-D video requires multiple video sequences captured from different camera positions, it becomes challenging to transmit and store such a large amount of data. This has lead to significant interest in investigating efficient view rendering methods. For this purpose it has been studied how to sample the plenoptic function or the light field to reconstruct missing views using image based rendering techniques. To improve the rendered view quality, multiview video plus depth formats are being developed, where only selected views are coded along with their corresponding depth maps, and other views are interpolated at the decoder using depth image based rendering (DIBR) technique Since depth maps have to be sent to the decoder, it is necessary to develop efficient compression methods for depth maps.

Even though a depth map can be treated as a gray scale image, and compressed using standard image or video coding techniques, better coding efficiency can be achieved if depth-map-specific characteristics are exploited, since some of these properties are quite different from those of standard image or video. Many researchers have aimed to improve the efficiency of depth map coding by exploiting these characteristics. For example, typical depth maps tend to lack texture so that they can generally be well approximated as piecewise smooth signals, with relatively constant depth areas separated by sharp edges where each smooth depth region may correspond to an object at a different depth. Many approaches have been proposed along these lines. Morvan et al and Merkle et al used platelet coding, and Maitre and Do , Daribo et al , and Sanchez et al used edge-adaptive wavelet transforms, which seek to design transform that avoid filtering across edges or object boundaries in a depth map. In our work, we have shown that better coding efficiency can be achieved by edge adaptive coding tools such as the edge adaptive intra prediction and the edge adaptive transform Other such approaches include compressed sensing based methods and the reduced resolution with edge-preserving up sampling . Besides, other depth map specific characteristics are utilized to achieve better coding efficiency in the researches, such as dynamic range reshaping , 3-D motion estimation , warping based inter-view prediction , reuse of video motion information to reduce encoding complexity , and sparsity-based in-loop filtering,In these previous works depth-map-specific characteristics are utilized to improve coding efficiency. However, it is also very important to understand how the depth map is used for the view synthesis.

In standard video compression, quantization errors directly affect the rendered view quality by adding noise to the luminance or chrominance level of each pixel. In contrast, the distortion in the depth map will affect indirectly the rendered view quality. Considering that the depth map itself is not displayed but used to provide additional geometry information to help the view synthesis process, it would be inefficient to maintain same level of quality for the whole depth map area using the aforementioned methods. For example, a small amount of geometry error can cause large distortion in the rendered view in regions of complex texture or at object boundaries with high contrast to the background intensity. Instead, large amounts of geometry error can lead to negligible artifacts in the rendered view in the case of flat regions. Therefore, it is crucial to analyze the relationship between the depth map error and the distortion in the rendered view in order to achieve optimal performance of depth map coding. There has been research to evaluate the impact of depth error to the rendered view distortion. M¨uller et al studied experimentally the impact of bitrate distribution between the video and depth map on the rendered view quality, but no theoretical analysis was given to find the relationship between the depth map error and the rendered view distortion. Merkle et al measured geometry error caused by depth map error by calculating the distance between the 3-D surfaces constructed from the original and the coded depth map, respectively, using the Hausdorff distance; however, no method was given to find how the geometry error causes the rendered view distortion, and the depth map distortion itself is used to optimize the depth map coding.

Nguyen and Do derived an upper bound on the mean squared error in the rendered view due to geometry error using Taylor series expansion. Ramanathan and Girod used the power spectral density with Gaussian modeling of image signal in order to estimate the distortion in rendered view due to geometry error, where global relationship can be found between the geometry error and the rendered view distortion. These approaches can be used to analyze the effect of geometry error on the rendered view quality; however, both the global Gaussian model and the upper bound derivation do not provide a precise estimation of how local depth coding distortion (e.g., within a block) lead to distortion in the rendered view. Because of this, distortion estimates obtained using the aforementioned methods may not be sufficiently accurate for rate-distortion (RD) optimization in depth map coding. In our previous work , the relationship between depth map error, geometry error, and distortion in the rendered view was analyzed, with a global linear model used to characterize the distortion in the rendered view as a function of depth map error. A problem with such a global distortion is that there may exist significant local mismatches, since the rendered view distortion varies according to local characteristics of the reference video. For example, the amount of the distortion caused by the geometry error will be small for a smooth region of the video as compared to a region with complex textures.

In this thesis, we start by focusing on the different types of distortion affecting the DIBR process. Clearly, the quality of decoded video frames that are used for view interpolation is an important factor, but since they are likely to be coded using standard tools (e.g., H.264/AVC) it is relatively simple to control their distortion. Instead, we focus on how the quality of depth maps transmitted to the decoder affects overall quality, which has non-linear characteristics (i.e., increases in intensity errors in the interpolated view are not proportional to increases in depth map coding errors) and is highly localized (i.e., more significant errors occur only in a small subset of pixels in a video frame). The main contribution of our research is to improve the rendered view quality using the new coding tools and new data format specifically designed to reduce the distortion occurred in the rendered view. First, we propose a simple and precise local estimation method to estimate the distortion generated in the rendered view due to depth map coding. Of course, the rendered view distortion can be exactly measured if the intermediate view is synthesized and compared to the ground truth. However, this is not practical, since the ground truth for an arbitrary view position may not exist, and the view synthesis process would be too complex to be used during the depth map coding. Instead, we propose a simple and precise estimation method, which reflects local video characteristics, so that it can be used during depth map coding to achieve optimal performance.

II. LITARATURE SURVEY

Several algorithms have been proposed to estimate the rendering quality. Nguyen and Do analyzed the rendering quality of image-based rendering (IBR) algorithms and used Taylor series expansion to derive the upper bound of the mean absolute error (MAE) in the synthesis output. They also quantified the effect of sample jitters caused by depth errors. On the contrary, our work analyzes the rendering quality with a combination of frequency and spatial domain techniques that are quite different from their work, and our model estimates the value (instead of upper bound) of the mean squared error (MSE). We also test the model with video sequences and coding/rendering tools used in MPEG 3DV activities. Liu et al. proposed a distortion model to evaluate the synthesized view. Their work approximated errors due to depth map artifacts using a linear model of average magnitude of mean-squared disparity errors over an entire frame and a motion sensitivity factor computed from the energy density. This was motivated by earlier work of using linear distortion model to characterize the effect of motion warping. Our work is different in that we characterize the disparity errors using their distribution rather than their average, and use a different analysis technique to derive the distortion caused by the disparity error distribution. We also notice that spatial variant signals would cause non-negligible discrepancy this necessitates compensating with a video sequence specific constant. Therefore, we propose to augment frequency-domain analysis with spatial-domain analysis of spatial-variant signals.

Our analysis framework is also different and leads to a different formulation for synthesis error decoupling. Yuan et al. proposed a frequency approach to estimate synthesis distortion. Similar to Liu et al, their work was 1This is based on our conference paper work in. motivated by the linear distortion model characterizing the effect of motion warping error . In Yuan et al. , they derived a linear model that relates synthesized view distortion with the quantization steps of the texture and depth videos. Model parameters are estimated by synthesis of three virtual views using compressed texture/depth videos with different texture/depth quantization steps. Note that their model parameters are specific to particular virtual view positions, scene characteristics, coding algorithms and encoding options (since quantization step is used as input in their model). That is,new model parameters need to be estimated using synthesis when these variables change. The author has also extended the work to wiener filter design in .The approach proposed by Wang et al. is similar to Yuan et al., with focus on rate-distortion analysis of free viewpoint coding. An autoregressive model was proposed by Kim et al.to estimate the synthesis distortion at the block level and was shown to be effective for rate-distortion optimized mode selection.

In their work, rendering distortion of a block is approximated by the local video signal variance and a first order autoregressive model for the correlation coefficients. A distortion model as a function of the view location was also proposed by Velisavljevic et al. for bit allocation. Takahashi proposed an optimized view interpolation scheme based on frequency domain analysis of depth map error. Some of his frequency analysis is similar to our previous work, but there is no accuracy comparison provided in his work. Our work is different as we take into account both distortions in the texture images and depth maps, and estimate the distortion due to depth map artifacts using probability distribution of position errors, PSD of texture, and linear approximation of local spatial-variant signals. The present work significantly extends our previous work by augmenting the frequency domain analysis with a spatial domain analysis. The frequency domain analysis is also modified in order to accommodate the new approach. Elaborated experiment results and analysis are provided in this work.

III. PROPOSED METHOD

A. Depth-Image Based Rendering (DIBR)

Depth-image-based rendering (DIBR) is the process of synthesizing “virtual” views of a scene from still or moving images and associated per-pixel depth information. Conceptually, this novel view generation can be understood as the following two-step process: At first, the original image points are reprojected into the 3D world, utilizing the respective depth data. Thereafter, these 3D space points are projected into the image plane of a “virtual” camera, which is located at the required viewing position. The concatenation of reprojection (2D-to-3D) and subsequent projection (3D-to-2D) is usually called 3D image warping in the Computer Graphics (CG) literature and will be derived very briefly in the following.

B. D Image Warping

Consider a system of two cameras and an arbitrary 3D space point □ with the projection in the first-, resp. the second view. Under the assumption that the world coordinate system equals the camera coordinate system of the first camera1 , the two perspective projection equations result

$$\begin{aligned} \tilde{\mathbf{m}} &\cong \mathbf{A}\mathbf{P}_n\tilde{\mathbf{M}} \\ \tilde{\mathbf{m}}' &\cong \mathbf{A}'\mathbf{P}_n\mathbf{D}\tilde{\mathbf{M}} \, , \end{aligned} \tag{1}$$

Where two 2D image points, resp. the 3D space point in homogeneous notion and the symbol denotes ‘equality up to a non-zero scalefactor’ . The matrix contains the rotation and the translation that transform the 3D point from the world coordinate system into the camera coordinate system of the second view and the matrices and specify the intrinsic parameters of the first-, resp. the second camera. Finally, the identity matrix designates the so called normalized perspective projection matrix. Rearranging gives an affine representation of the 3D space point □ that is, however, still dependent on its depth value

$$\mathbf{M} = Z\mathbf{A}^{-1}\tilde{\mathbf{m}} \tag{2}$$

Substituting into then leads to the classical affine disparity equation, which defines the depth dependent relation between corresponding points in two perspective images of the same 3D scene:

$$Z'\tilde{\mathbf{m}}' = Z\mathbf{A}'\mathbf{R}\mathbf{A}^{-1}\tilde{\mathbf{m}} + \mathbf{A}'\mathbf{t} \, . \tag{3}$$

This disparity equation can also be considered as a 3D image warping formalism, which can be used to generate an arbitrary novel view from a known reference image. This only requires the definition of the position and orientation of a “virtual” camera relative to the reference camera as well as the declaration of the “virtual” camera’s intrinsic parameters. Then, if the depth values of the corresponding 3D space points are known for every pixel of the original image, the “virtual” view can be synthesized by applying to all original image points. (The “real-world” problems of image re sampling, resolving the visibility problem and handling of disocclusions in the novel view will be considered later in Section.

C. Distortion Analysis in Rendered View

In this chapter, the view synthesis process using depth map is reviewed, and the distortion in the rendered view is analyzed. Through the analysis various factors affecting the rendered view quality are determined. Among them we focus our effort on investigating the relationship between the depth map error and rendered view distortion, since the analysis reveals that the depth map error has great impact on rendered view quality. We qualitatively analyze the depth map quantization error, which reveals non-linear and localized features of the rendered view distortion from depth map error, and propose a quality measurement method to reflect these features. In addition, flickering artifacts are observed due to depth map estimation error, and we propose a quantitative way to measure this.

D. View synthesis process using depth maps

In a DIBR system, a few views are transmitted to the decoder, while the other views are synthesized using the decoded views . A depth map sequence is included for each transmitted view along with video sequence in order to improve the quality of the synthesized video quality. To analyze the distortion in the rendered view it will be helpful to understand how an intermediate view is generated using the neighboring views that are available at the decoder side. In this section the view synthesis process using depth map is reviewed.To synthesize a view (target view) using the decoded view (reference view), each pixel in the decoded video frame can be mapped to the target view using the camera parameters, such as baseline distance, focal length, and rotation and translation matrix, and per-pixel depth value. This is done by first mapping from the reference view to the world coordinate, and then mapping from the world coordinate to the target view, as illustrated in Fig.. A more detailed mapping procedure is given in Section.

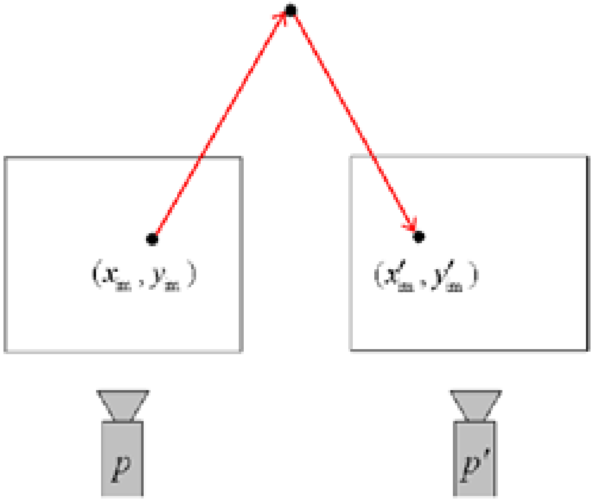


Fig1. Mapping of a pixel from reference view (p) to the world coordinate and to the target view (p′), where (xim, yim) and (x′ im, y′ im) are the location of pixel in the reference view and the target view, respectively.

To improve rendered view quality, it is important to obtain depth information with enough precision. There are various ways to acquire depth information. For example per-pixel depth values can be estimated from neighboring video using stereo triangulation, or it can be directly captured using various range cameras such as laser scanner and time-of-flight camera .Per-pixel depth values can be converted into depth map value for efficient storing. The relationship between the actual depth value, z, and the 8-bit depth map value, Lp(xim, yim), is given as

$$z = \left(\frac{L_p(x_{im}, y_{im})}{255} \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}} \right)^{-1} \quad (4)$$

Where, Znear and Zfar are the nearest and the farthest clipping planes, which correspond to value 255 and 0 in the depth map, respectively, with the assumption that z, Znear and Zfar have all positive or all negative values. Figs. and show an example of the view synthesis procedure , where Fig. and are the warped views from left and right views to the target view, and Fig. is generated by combining these two warped views. When all the pixels in the reference view are mapped to the target view, it is possible that multiple of them are mapped to the same position in the target view. On the other hand it is also possible that none of them can be mapped to certain positions in the target view, which is called as the hole area. The hole area in the left warped view is filled from the right warped view, and vice versa. Methods to generate the target view by combining more than one warped view are called as a blending processes, in which a weighted averaging can be used with the baseline distance or depth value of each reference pixel used as the weight. It is also possible to select only one reference pixel to avoid blurring artifact. When there remains a hole area after blending, it can be filled using a hole filling process, e.g., based on the in painting.

IV. SIMULATION RESULTS

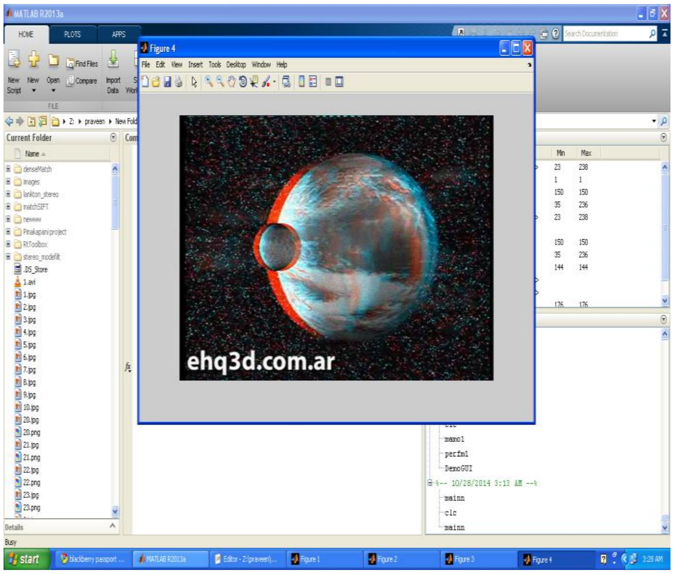


Fig2. Sample for 3D.

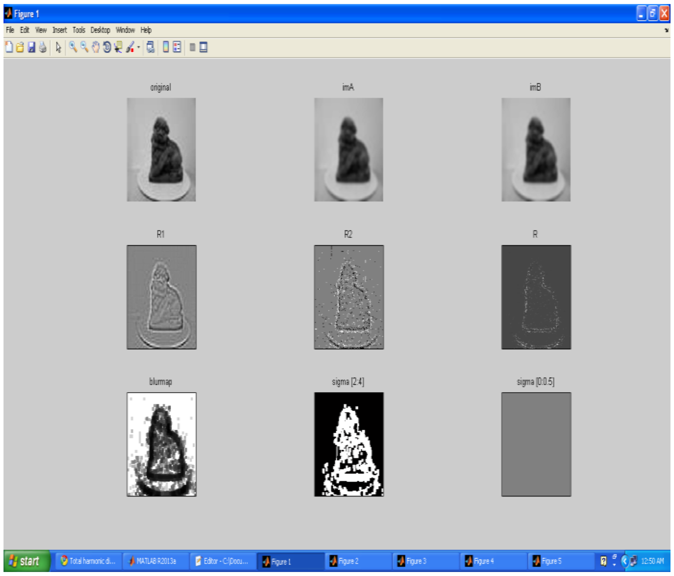


Fig3. Different type of output images.

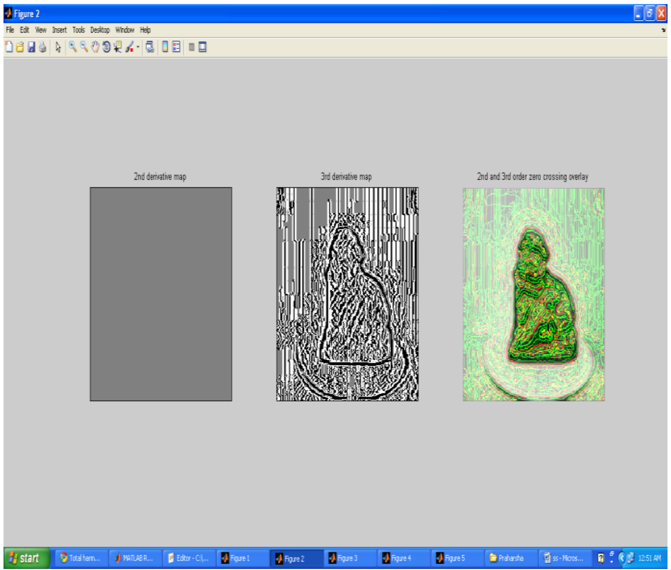


Fig4. Image using 2nd and 3rd derivative Map.

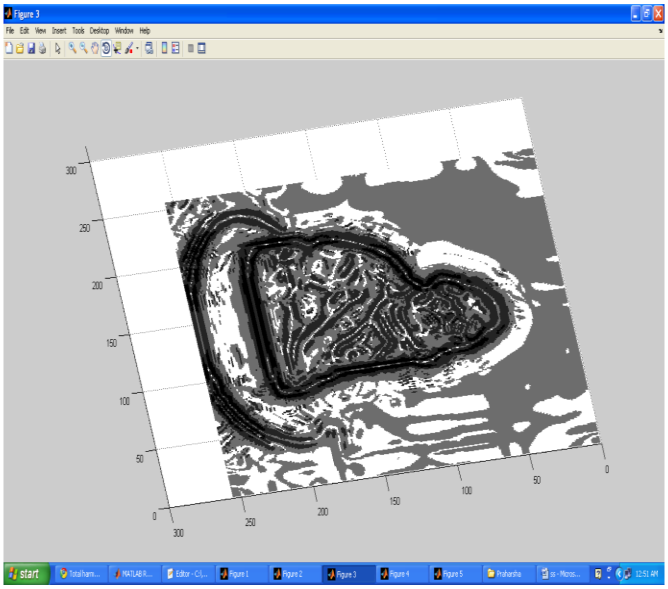
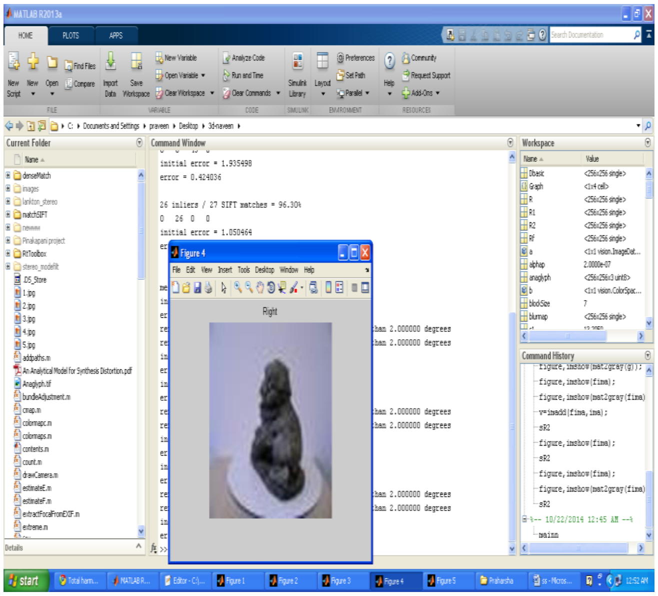


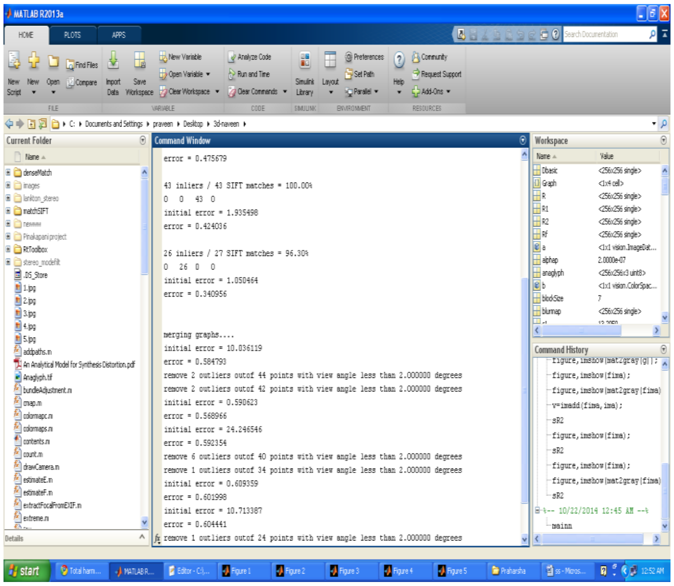
Fig5. 3D Map.



```
initial error = 1.935498
error = 0.424036

26 inliers / 27 SIFT matches = 96.30%
0   26  0   0
initial error = 1.050464
```

Fig6. Right Image of Original.



```
error = 0.475679

43 inliers / 43 SIFT matches = 100.00%
0   0   43  0
initial error = 1.935498
error = 0.424036

26 inliers / 27 SIFT matches = 96.30%
0   26  0   0
initial error = 1.050464
error = 0.340956


merging graphs....
initial error = 10.036119
error = 0.584793
remove 2 outliers outof 44 points with view angle less than 2.000000 degrees
remove 2 outliers outof 42 points with view angle less than 2.000000 degrees
initial error = 0.590623
error = 0.568966
initial error = 24.246546
error = 0.592354
remove 6 outliers outof 40 points with view angle less than 2.000000 degrees
remove 1 outliers outof 34 points with view angle less than 2.000000 degrees
initial error = 0.609359
error = 0.601998
initial error = 10.713387
error = 0.604441
remove 1 outliers outof 24 points with view angle less than 2.000000 degrees
```

Fig7. Command window.

V.CONCLUSION

We have proposed a systematic model to gauge the orchestrated perspective quality in 3d feature. The model relates lapses in the profundity pictures to the rendering quality, taking into account surface picture qualities, composition picture quality furthermore the rendering procedure. We decoupled the estimation of the force of the union mutilation into two steps, one centering on the surface slip impelled contortion, and the other centering on the profundity lapse incited bending. We demonstrated that the PSD of the rendering bending because of profundity coding is the item of the PSD of surface information and a recurrence envelope depending on the likelihood dissemination of the position slips. We likewise determined mathematical statements to gauge the rendering contortions in spatial variation locales along solid edges. Trial results demonstrated that the model can precisely appraise the combination clamor power. The model can be utilized to anticipate the rendering quality for distinctive framework outlines. The examination can likewise help illuminate outlines of coding, transmission and rendering frameworks for 3d feature.

VI. REFERENCES

[1] A. Vetro, W. Matusik, H. Pfister, and J. Xin, “Coding approaches for end-to-end 3-D TV systems,” in Proc. PCS, 2004, pp. 1–6.

[2] T. Maugey, P. Frossard, and G. Cheung, “Consistent view synthesis in interactive multiview imaging,” in Proc. 19th IEEE ICIP, Oct. 2012, pp. 2717–2720.

[3] Z. Zhang, R. Wang, C. Zhou, Y. Wang, and W. Gao, “A compact stereoscopic video representation for 3-D video generation and coding,” in Proc. IEEE DCC, Apr. 2012, pp. 189– 98.

[4] D. Min, J. Lu, and M. N. Do, “Depth video enhancement based on weighted mode filtering,” IEEE Trans. Image Process., vol. 21, no. 3, pp. 1176–1190, Mar. 2012.

[5] D. Min, D. Kim, S. Yun, and K. Sohn, “2-D/3-D freeview video generation for 3DTV system,” Image Commun., vol. 24, nos. 1–2, pp. 31–48, 2009.

[6] C. Fehn, “Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV,” Proc. SPIE, vol. 5291, pp. 93–104, May 2004.

[7] Q. Wang, X. Ji, Q. Dai, and N. Zhang, “Free viewpoint video coding with rate-distortion analysis,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 6, pp. 875–889, Jun. 2012.

[8] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Mueller, P. H. N. de, et al., “The effects of multiview depth video compression on multiview rendering,” Signal Process., Image Commun., vol. 24, nos. 1–2, pp. 73–88, 2009.

[9] D. Tian, A. Vetro, and M. Brand, “A trellis-based approach for robust view synthesis,” in Proc. 18th IEEE ICIP, Sep. 2011, pp. 605–608.

[10] I. Daribo, C. Tillier, and B. Pesquet-Popescu, “Adaptive wavelet coding of the depth map for stereoscopic view synthesis,” in Proc. IEEE 10th Workshop MSP, Oct. 2008, pp. 413–417.

[11] A. Sanchez, G. Shen, and A. Ortega, “Edge-preserving depth-map coding using tree-based wavelets,” in Proc. 43rd Asilomar Conf. Signals, Syst., Comput., Nov. 2009, pp. 578–582.

[12] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, “Depth map distortion analysis for view rendering and depth coding,” in Proc. 16th IEEE ICIP, Nov. 2009, pp. 721–724.

[13] N.-M. Cheung, D. Tian, A. Vetro, and H. Sun, “On modeling the rendering error in 3-D video,” in Proc. 19th IEEE ICIP, Sep. 2012,

[14] B. Girod, “The efficiency of motion-compensating prediction for hybrid coding of video sequences,” IEEE J. Sel. Areas Commun., vol. 5, no. 7, pp. 1140–1154, Aug. 1987.

[15] P. Ramanathan and B. Girod, “Rate-distortion analysis for light field coding and streaming,” Signal Process., Image Commun., vol. 21, pp. 462–475, Mar. 2006.

[16] Moving Picture Experts Group, Geneva, “Call for proposals on 3D video coding technology,” Tech. Rep. MPEG N12036, Mar. 2011.

[17] Moving Picture Experts Group, Geneva, “View synthesis software manual release 3.5 (VSRS 3.5),” Tech. Rep., ISO/IEC JTC1/SC29/WG11 MPEG, Sep. 2009.

[18] H. T. Nguyen and M. N. Do, “Error analysis for image-based rendering with depth information,” IEEE Trans. Image Process., vol. 18, no. 4, pp. 703–716, Apr. 2009.

[19] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, “Joint video/depth rate allocation for 3-D video coding based on view synthesis distortion model,” Signal Process., Image Commun., vol. 24, no. 8, pp. 666–681, Sep. 2009.

[20] A. Secker and D. S. Taubman, “Highly scalable video compression with scalable motion coding,” IEEE Trans. Image Process., vol. 13, no. 8, pp. 1029–1041, Aug. 2004.