Depth Maps: Faster, Higher and Stronger ?

Ianir. A. Ideses*, Leonid. P. Yaroslavsky and Barak Fishbain
Dept. of Physical Electronics, Tel-Aviv University 69978 Tel-Aviv Israel

ABSTRACT

In this paper we present experimental evidence of the redundancy of depth maps for 3D visualization. To this end, we performed a series of statistical experiments, devised to measure the effect of depth map quantization and the resolving power of 3D perception. The results of these tests show that for good 3D perception and 3D visualization, one does not need to use depth map of the same resolution neither with the same quantization as the original images. These results indicate that depth map based visualization can be based on low resolution, coarsely quantized, depth maps without significant degradation in the perceived 3D image.

Keywords: 3D, Depth Maps, Experiments, Resolution, Quantization

1. INTRODUCTION

With the recent advancement in three dimensional visualization devices over last years, we are seeing a growing market for stereoscopic content. Although 3D visualization devices have improved over recent years, it appears that 3D acquisition and synthetic 3D generation are still lagging behind. Stereo acquisition remains technically problematic and real-time 3D from 2D video synthesis is still largely limited. It appears that the major problem of 2D to 3D conversion, as well as that of multi-view auto-stereoscopic displays, is computation of depth maps of the scenes.

Computing depth maps in a stereo setup is equivalent to computing the disparity of each pixel in the two stereo pair images. Knowledge of the camera parameters enables exact triangulation and computation of depth maps.

There exist many methods for computation of disparity maps, in principle, disparity maps can be computed by direct correlation performed on local neighborhoods of pixels on a running window, these can be computed utilizing classic Optical Flow methods and also by variational methods. Among these methods are the works of Lucas and Kanade [1], Horn and Schunck [2], Senthil Periaswamy and Hany Farid [3], Yu-Te Wu et al 4], L. Alvarez et al [5], Jochen Schmidt [6] and Adee Ran and Nir Sochen [7]. In previous publications we have shown that disparity maps can also be extracted from compression encoders [8].

While most effort and research is aimed at computing depth maps with high precision and full resolution (dense depth maps), much less, to the authors knowledge, has been done to evaluate how accurate or how dense must these depth maps be for different applications. One can indeed consider cases where exact metrics are required, such examples may include micro-surgery or metrology of nano-manufacturing. However, for the purpose of 3D visualization, it was proposed in previous publications that only a fraction of the data is required [9-10]. In this paper, we provide experimental evidence that, in generating stereo-images with acceptable 3D visual quality, one can reduce the volume of depth map data needed for the synthesis of stereo images by the orders of magnitude compared to that of images themselves. We describe results of visual testing of stereoscopic images synthesized using depth maps that were coarsely quantized in their values. Other tests were performed to find the resolution limit of depth maps. The tests were performed with synthetic stereoscopic images using the random dots methodology developed by B. Julesz [11], as well as using other texture and real world images. 3D visualization was performed using anaglyph glasses, LCD shutter glasses and an autostereoscopic display.

*ianir@eng.tau.ac.il; phone 972 3 640-8014; fax 972 3 641-0189

Three-Dimensional Image Capture and Applications 2008, edited by Brian D. Corner, Masaaki Mochimaru, Robert Sitnik
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6805, 68050B, © 2008 SPIE-IS&T · 0277-786X/08/$18

2. METHODOLOGY

2.1 Experiments

The experiments were aimed at finding two parameters, the first set of experiments was intended to find the resolvable accuracy of the depth maps. The second set was aimed at finding the perceived spatial resolution limit of depth maps.

Because depth perception is based not only on stereoscopy, but rather on other depth cues such as texture, shading, focus, long-range motion, reflection, shadows, symmetry, inter-reflection and polarization we chose to eliminate these and include only stereopsis as a depth cue. This was performed by using a random dot field as an image and a controlled depth map. These depth maps were either deterministic shapes (for example a hemisphere) or bandwidth controlled textures. In order to verify these tests in real world conditions, these tests were augmented by using real world images and "ground truth" depth maps taken from public repositories [12].

2.2 Visualization

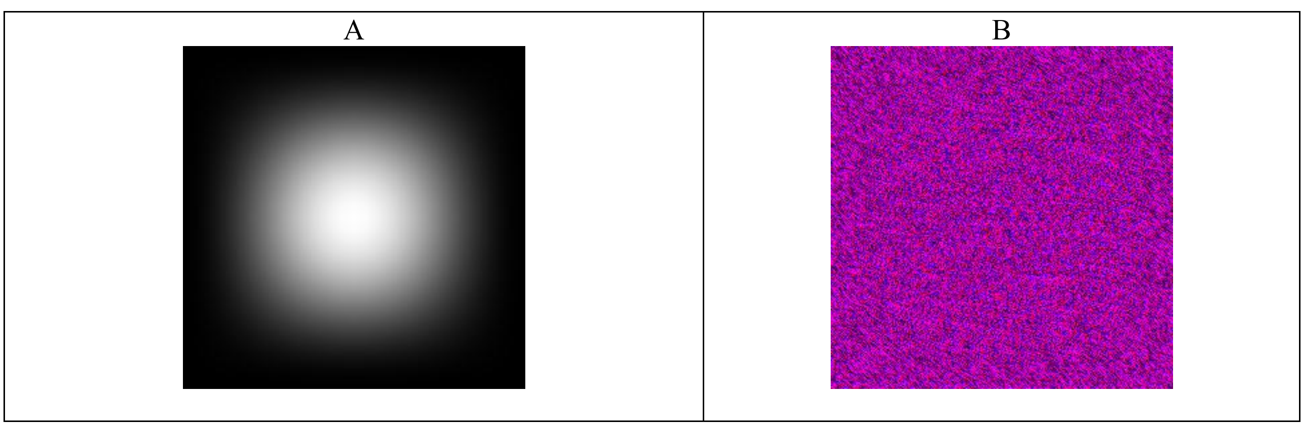
Visualization for these experiments was performed using three types of 3D displays:

a. Anaglyphs. In this technique, a 3D image is constructed using the stereo pair as different color components (typically the red channel for the left image and blue/cyan for the right). The 3D image is viewed with color filtering glasses that are suited to the channels used in the 3D image synthesis. This display technique is very useful due to the fact that it does not reduce spatial or temporal resolution, is very easy to implement and use and does not require special hardware.

b. LCD Shutter glasses. In this technique, special glasses, with shutter in front of each eye are synchronized to a monitor. These glasses toggle from showing only the right image to showing only the left image. Synchronizing the toggling to the monitor (usually in the form of odd and even fields) enables the user to see a different image in each eye. This technique is useful because it does not hamper color perception, but is limited in the sense that it lowers the temporal refresh rate and reduces the vertical resolution by half. Opting to keep spatial resolution equal on both axes resulted in images that appear distorted. An additional drawback of this method is that it can generally only be used with CRT monitors due to the polarization that is used in the shutter mechanism.

c. Autostereoscopic displays. In this technique no glasses or special apparatus, other than the display itself are needed. The display accepts as input a concatenation of the left and right images and displays them in 3D. The major drawback of this display is that the 3D viewing zone is limited, making the experiment difficult to conduct.

Figure 1 shows 3D images constructed for each visualization device.



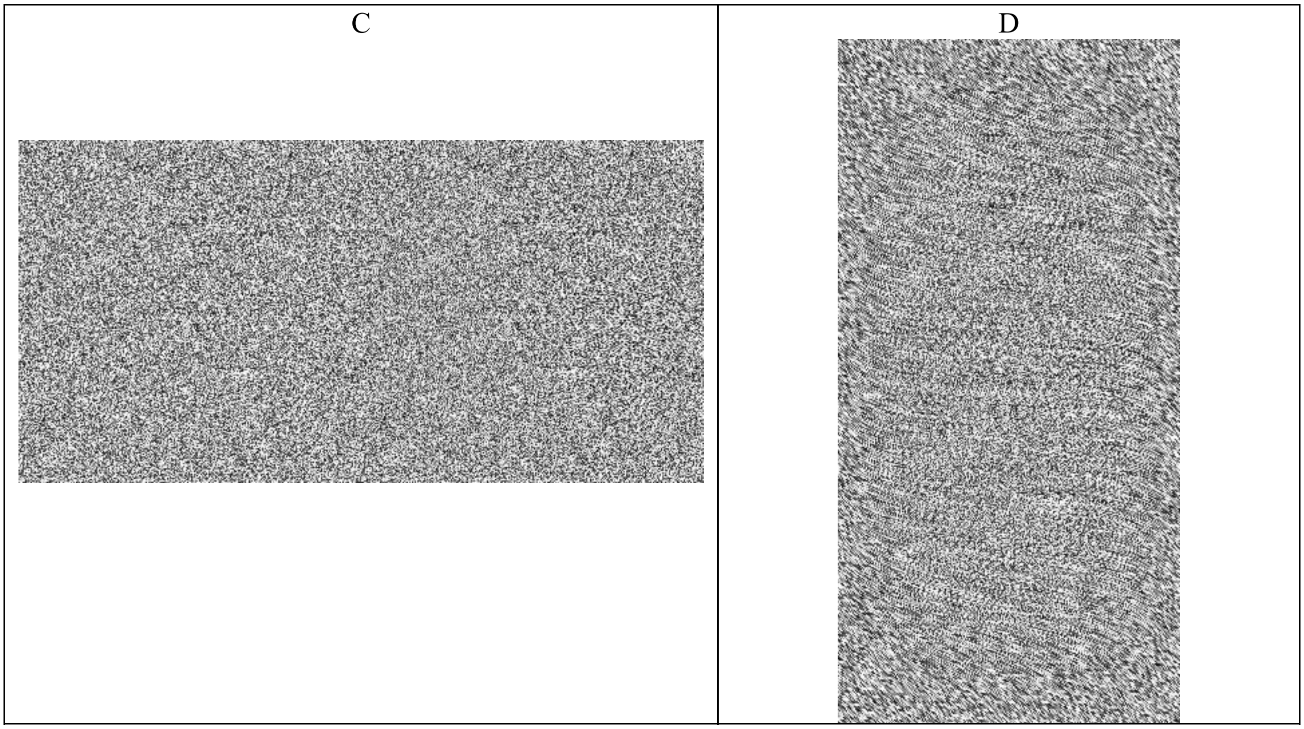


Figure 1 shows the original depth maps (A) in 3 visualization methods: (B) is an anaglyph representation of the stereo pair, viewed with red color filter on the left eye and blue on the right. (C) is a representation fit for autostereoscopic displays and is the concatenation of the stereo pair. (D) is intended for shutter glasses and contains the stereo pair in the odd and even fields of the image.

2.3 Rendering

In order to render the 3D images, one must generate 2 images, corresponding to the stereo pair. This pair is then used differently in each 3D visualization device, but the initial stereo pair is the same in all methods.

Generation of a synthetic stereo pair requires a depth map and one initial image, usually the left image. The process itself is basically a process of image magnification and resampling. In this case resampling is not performed uniformly, but rather on a grid that is governed by the depth map. Magnification is required in order to enable sub pixel resampling. This is required because of the limited perceivable disparity in 3D images. For example, without magnification and sub pixel resampling, a nominal disparity value of 16 pixels would only allow 16 levels of pixel shifting, thus only 16 values of depth. In our experiments, the images were magnified by a factor of 16 to enable a larger range of depth values. In the case of a disparity of 16 pixels this enables 256 levels of depth values.

Magnification of the image was performed using discrete-sinc interpolation, the resampling stage was conducted according to Equation 1.

$$P_R(i,j) = P_L(i, j + MF * D(i,j)), \tag{1}$$

where $P_R(i,j)$ is the pixel value of the rendered right image in location (i,j), $P_L(i,j)$ is the initial left image value at (i,j), MF is the magnification factor, in our case 16 and $D(i,j)$ is the depth(disparity) value at (i,j).

The disparity map is adjusted to the nominal dynamic range by stretching its maximal value to the desired disparity level as shown in Equation 2.

$$D(i,j) = \frac{D_i(i,j)}{\max(D_i(i,j))} * MD, \tag{2}$$

where $D(i,j)$ is the disparity value at (i,j), $D_i(i,j)$ is the original depth map value and MD is the maximal disparity selected by the user.

The depth map is input by the user, either a synthetic image, or a depth map computed by any of the aforementioned methods.

2.4 Quantization Tests

In quantization tests, different types of test objects, both deterministic and bandwidth-controlled textures, were used. The experimenters were shown reference stereoscopic images with non-quantized depth map and images with depth map quantized to a certain number of quantization levels and were asked to determine whether these images appear equivalent. The users could increase or decrease the quantization levels until the point where the two images appeared the same (in case of increasing the quantization levels) or when the quantized image begins to deteriorate (in case of decreasing the quantization levels). In addition to synthetic depth maps, ground truth depth maps taken from public repositories were also quantized. These were used with real-life images rather than pseudo random dot patterns in order to verify the validity of these tests to real-life images.

The depth maps were quantized using linear uniform quantization. Although other quantization methods such as P-law quantization exist [13], for the purpose of these tests we opted to use a simple method that would provide the minimal number of quantzation levels, utilizing more advanced quantization methods with this number can only improve visual results.

Examples of quantized depth maps are shown in Figure 2. The corresponding 3D images that were displayed to the users are shown in Figure 3.

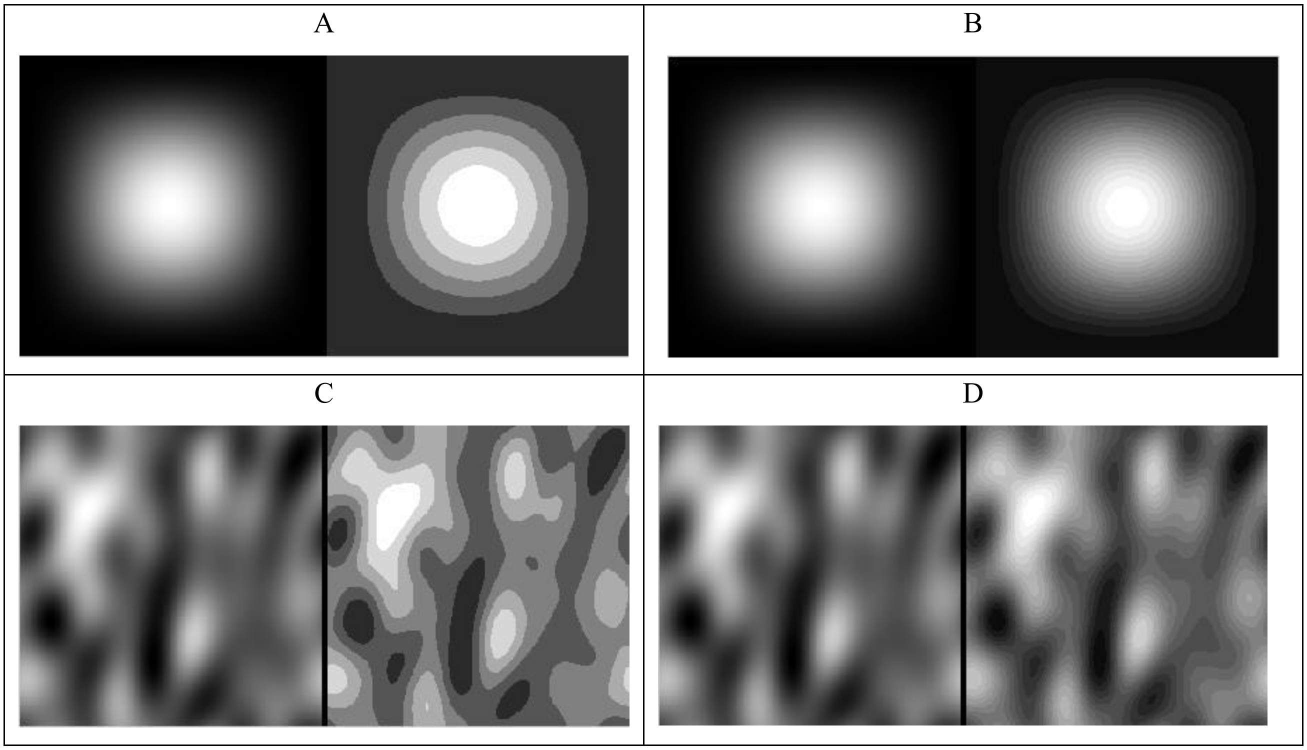
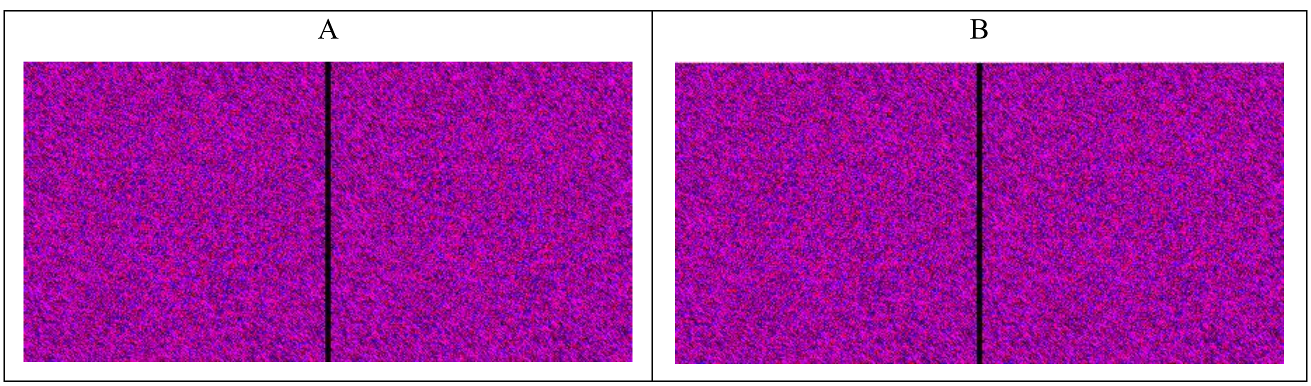


Figure 2. shows the effect of quantization on depth maps. Each image shows the original depth maps side by side its quantized version. (A) shows a deterministic hemisphere depth map quantized to 6 levels. (B) shows the same depth map as (A) but quantized to 20 levels. (C) shows a pseudo random texture generated by filtering white noise to 1% of its base band and quantized to 6 levels. (D) shows the same texture as (C) but quantized to 20 levels.



C D

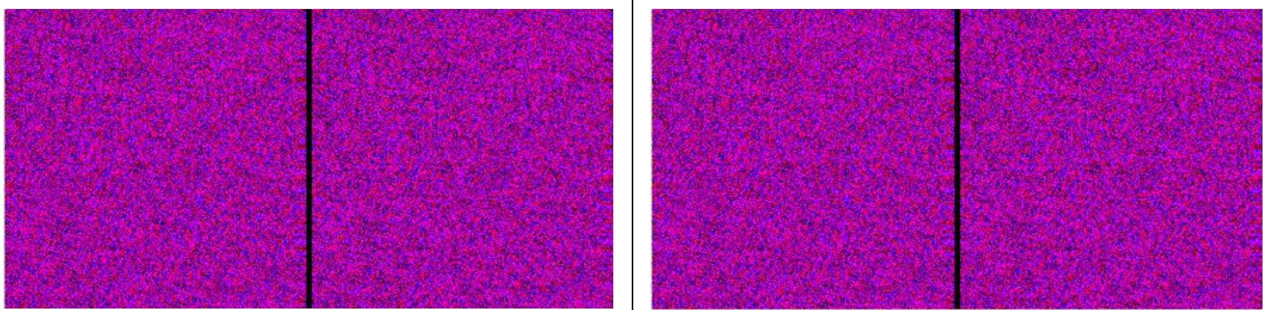


Figure 3. shows the effect of quantization on depth maps using anaglyphs as a visualization device. Each image shows the original anaglyph side by side its quantized version. These images correspond to the depth maps shown in Figue 2. (A) shows a deterministic hemisphere depth map quantized to 6 levels. (B) shows the same depth map as (A) but quantized to 20 levels. (C) shows a pseudo random texture generated by filtering white noise to 1% of its base band and quantized to 6 levels. (D) shows the same texture as (C) but quantized to 20 levels. All images can be viewed using standard anaglyph glasses, use red filter for left eye, blue filter for right eye.

In order to verify that applicability of the results we included a test with real-life images. These images contain additional 3D information in the form of other depth cues as described above and therefore can not serve as a test-bed for stereopsis alone, but can verify the results obtained using random dot patterns.

Figure 4 shows a real world image and a "ground truth" depth map. These images belong to the 3DTV image repository.

A B

Figure 4 – (A) shows a real-life image. (B) shows a depth map regarded as "ground truth". These images are part of the 3DTV image repository and are used in various 3DTV research projects.

Figure 5 and 6 show the depth maps after quantization to 6 and 20 levels and the resulting depth maps respectively.

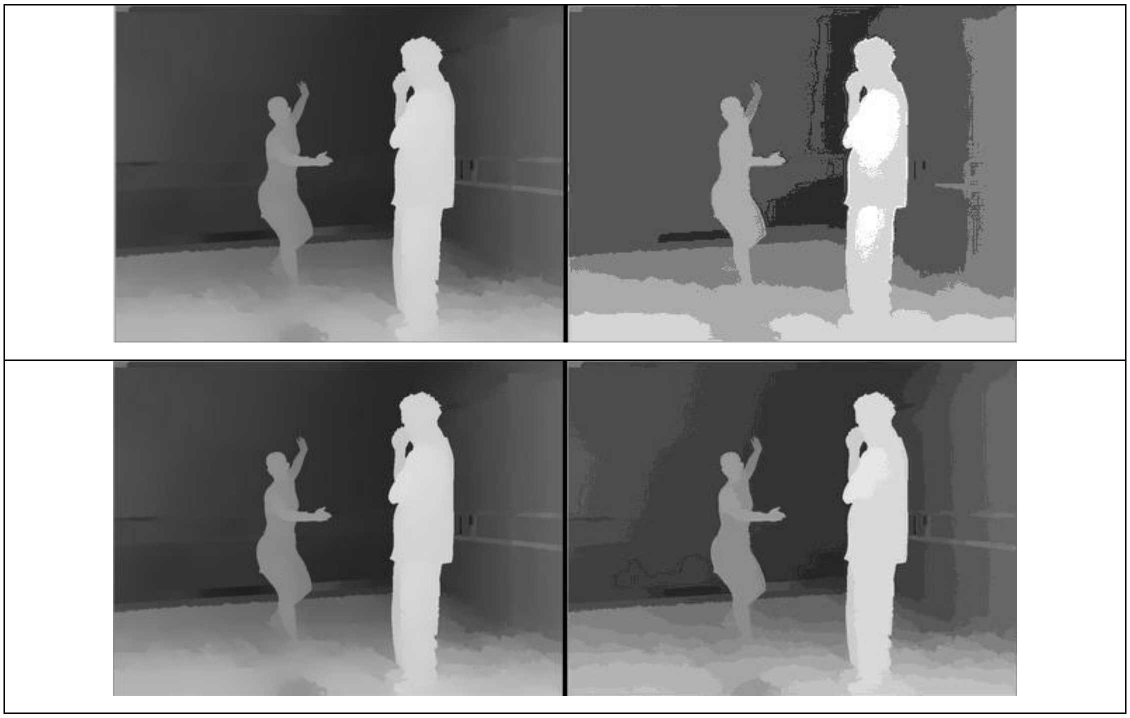


Figure 5 Top images are the “ground truth” depth map, original and quantized to 6 levels. Bottom images are the “ground truth” depth map, original and quantized to 20 levels.

Figure 6 Top images are the anaglyph result of “ground truth” depth map, original and quantized to 6 levels. Bottom images are the the anaglyph result of “ground truth” depth map, original and quantized to 20 levels.

2.5 Resolution Tests

Initially, we attempted to measure the error of localization of a stimulus whose depth map was subjected to blurring. While these results showed that there is very little deterioration in the ability to fuse the stimulus and locate its center, it was very difficult to find a common ground for statistical analysis. Since direct measurement of the effect of depth map blurring on perception was difficult to perform we chose a different test. In this test stereo-images were generated using pseudo-random depth maps with bandwidth of certain fractions of the image base band. As the bandwidth that was selected became higher, the texture elements became smaller, this can be seen in Figure 7.

The users were asked to find the maximal bandwidth in which they were still able to fuse the 3D image. These tests were repeated and statistics regarding maximal bandwidth. This bandwidth is directly related to resolving power of the depth map. An example of such a stimulus in anaglyph format is shown in Figure 8.

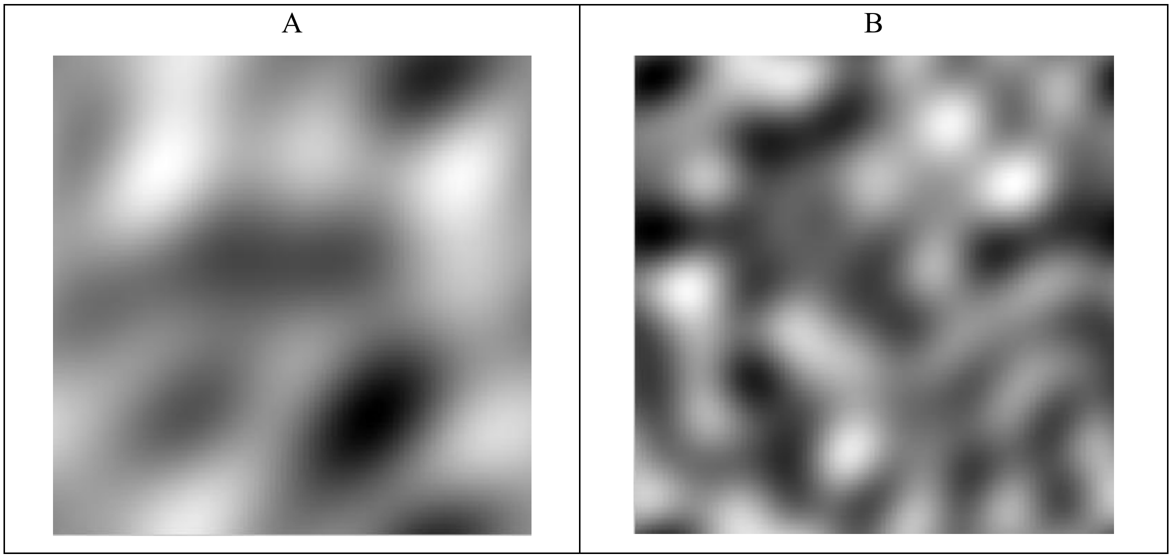


Figure 7 – (A) shows a depth map that corresponds to 0.01 of the baseband. (B) shows a depth map that corresponds to 0.02 of the baseband.

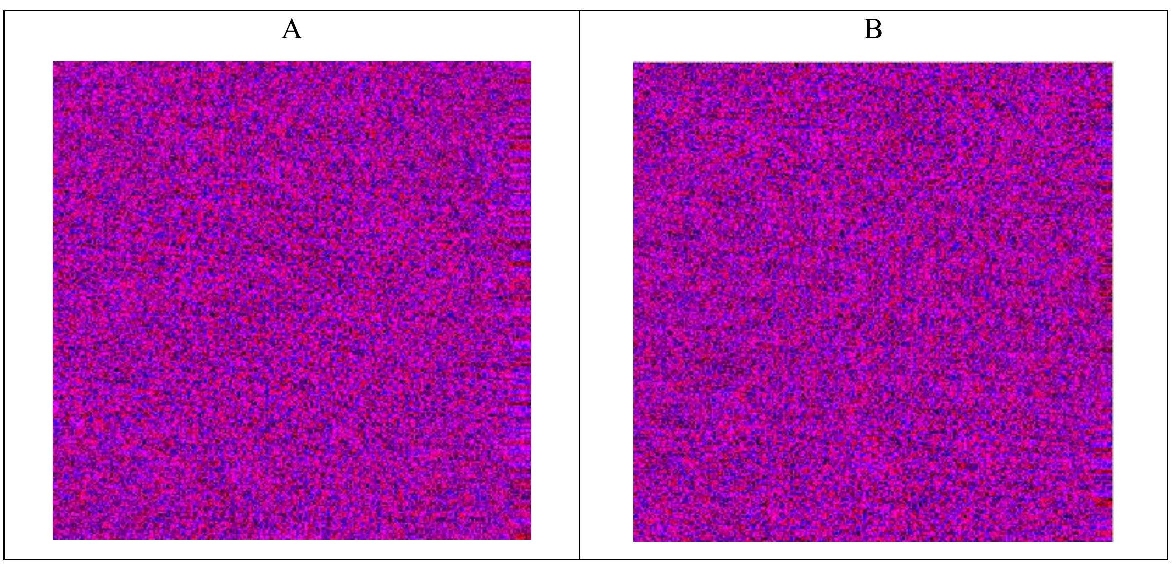


Figure 8 – (A) shows the anaglyph result of a depth map that corresponds to 0.01 of the baseband. (B) shows the anaglyph result of a depth map that corresponds to 0.02 of the baseband.

3. RESULTS

3.1 Quantization tests

Results for quantization tests are shown in Figure 9.

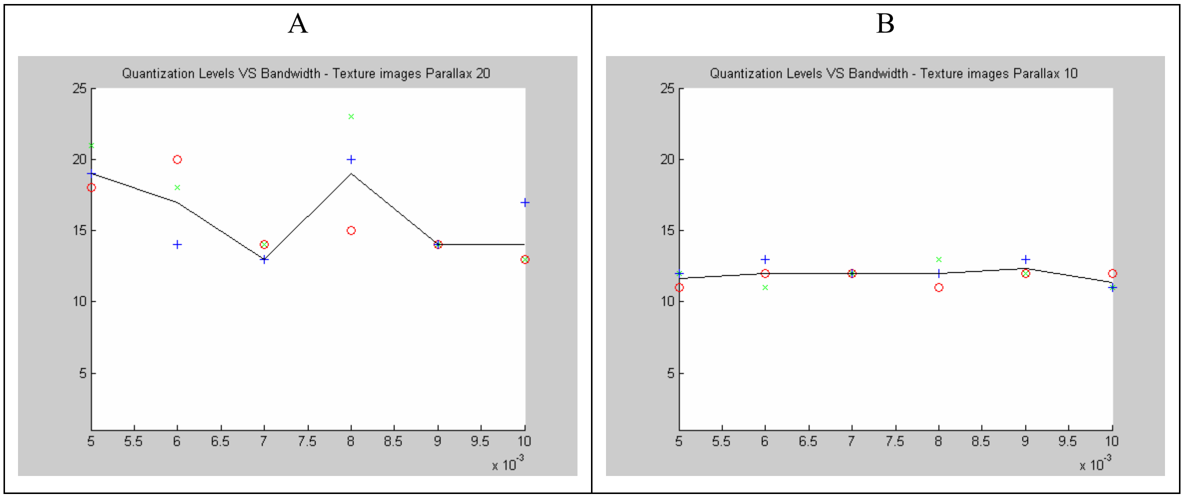


Figure 9 – (A) shows results of the number of quantization levels required, tests were performed with 10 viewers. The graphs show the results for typical 3 viewers (shown with different shapes and colors, as well as average values) for different textures with parallax of 20 pixels. (B) shows results of the number of quantization levels required, for the same viewers for different textures with parallax of 10 pixels.

As can be seen in figure 9, a very low number of quantization levels is required for good 3D perception. This number increases with the maximal parallax (disparity) and remains almost constant regardless of the texture baseband. Experiments with the hemisphere produced very similar results, an average of 13 quantization levels was required for fusing that image with a parallax of 20 pixels. These results remain constant, regardless of the visualization device. Tests performed with anaglyphs, shutter glasses or autostereoscopic devices produced similar results for all viewers.

3.2 Resolution tests

As described before, two types of resolution tests were performed, the first – measurement of the accuracy of finding an object's center vs. depth map blurring. These tests showed that there is very little deterioration in the ability to perceive images with blurred depth maps and that the error actually decreases for moderate blur factors. These results are congruent with previous results for stereo image blurring [9]. However, statistically it was very difficult to find a common ground for these tests. This can be seen in several graphs that were produced by different experimenters, shown in figure 10.

The second set of resolution tests have shown that depth maps whose bandwidth exceeds a fraction of 0.04 of the baseband can not be correctly viewed by the users. From a group of 10 experimenters, none were able to fuse these small details correctly and reported problems fusing the images at all.

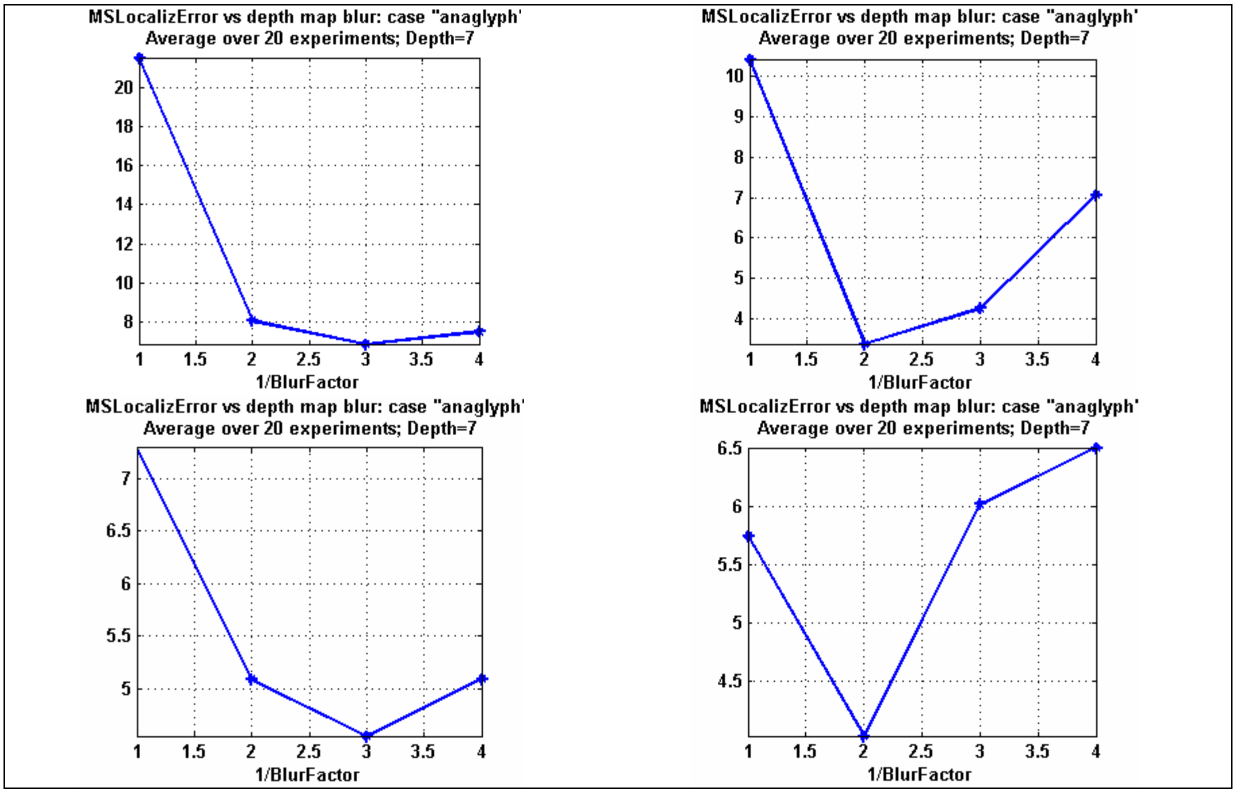


Figure.10. Four typical results for localization errors of different viewers. Note that the spread of error values for different viewers is quite large while the shape and tendency remain the same. The Y axis shows localization error in pixels, the X axis shows the amount of blur, in this case blurring was performed by wavelet decomposition and taking the level that corresponds to the Blurfactor (e.g. Blur factor 2 corresponds to half resolution, 3 to 1/8 etc').

4. CONCLUSIONS AND APPLICATIONS

As shown in the results section, 20 quantization levels of the depth map values are sufficient for viewing comfortable representation of 3D images and video. Using this data may have a significant effect on depth map computation and 3D synthesis.

Many modern depth map computation techniques rely on iterative algorithms. Iterative algorithms are bounded by a stopping criteria and have a step size for each iteration. It is therefore possible, to compute the required step by dividing the maximal parallax value by the number of quantization levels.

In addition to depth map calculation, depth map quantization may be important for artificial stereo pair synthesis. In our process of stereo pair synthesis we used subpixel accuracy for the testing. The results, however, indicate that, for a parallax suitable for visualization, shifting of pixels does not require subpixel accuracy. This has implications on the computational complexity required for 3D video synthesis.

In related work [14], we have shown that by using compressed video formats, we are able to synthesize 3D video from 2D video in real-time. In this algorithm, most of the computation power is spent in the resampling stage, namely in interpolating the image prior to resampling. The results acquired in these new tests shed a new light on this. If interpolation is not necessary for 3D synthesis, we can greatly reduce the computational load in the 3D synthesis stage and achieve above real-time performance. We suggest that similar performance increases are to be expected for other applications as well.

Resolution tests shed light on the denseness of the required depth maps. Previous experimentation [9] has shown that images of the stereo pair need not be of the same resolution and even can be of different color depths. These previous testing hinted that the same would hold for depth maps. In current tests this conjecture was experimentally verified. One can not fuse depth map information beyond a certain fraction of the base band. This means that for the purpose of 3D visualization (as opposed to other applications, such as metrology) one does not need to compute the depth map beyond this point. This has a significant impact on the process of depth map computation. It is therefore, possible, to compute the depth map in lower resolution, namely to subsample the images before depth map computation and upsample the resulting depth map using relatively simple and efficient interpolation methods. One such application is synthesis of depth maps from MPEG streams [8]. In this application the depth maps are acquired at quarter resolution on each axis, in our implementation we used simple nearest neighbor interpolation and low pass filtering to upsample the images. In principle, more advanced interpolation methods could be used, but where rejected in order to achieve real-time performance.

5. SUMMARY

In this paper we have tested the effect of quantization and resolution limits of depth maps on 3D perception. In order to test this, we wrote programs that presented synthetic 3D images to viewers, using different visualization devices. These experiments included testing for the number of quantization levels which is sufficient for viewing without perceptual artifacts. Resolution tests concentrated on finding the depth maps that correspond to the maximal fraction of the baaeband and are still fusable. This fraction is a hint on the denseness of the depth maps that is sufficient for artifact free 3D visualization.

The results of quantization tests show that, for depth map quantization, a relatively low number of about 20 quantization levels of depth map are sufficient for 3D synthesis. This number was acquired for deterministic shapes as well as frequency controlled texture images.

Resolution tests show a fundamental limit to the spatial frequency of depth maps that can be fused in 3D. This limit can serve as a starting point to finding the actual spatial limits of depth maps.

The obtained results can be utilized in different applications, and especially in algorithms for depth map computation and in the process of generating artificial stereo pairs from an image and a depth map.

REFERENCES

1. Lucas, B., and Kanade, T., "An Iterative Image Registration Technique with an Application to Stereo Vision". Proceedings of 7th International Joint Conference on Artificial Intelligence (IJCAI), pp. 674-679 (1981).
2. B. Horn and B. Schunck, "Determining Optical Flow" Artificial Intelligence, 17:185-203 (1981).
3. Senthil Periaswamy, Hany Farid, "Elastic Registration in the Presence of Intensity Variations", IEEE Transactions on Medical Imaging, , Volume 22, Number 7 (2003).
4. Yu-Te Wu, Takeo Kanade, Ching-Chung Li and Jeffrey Cohn, "Image Registration Using Wavelet-Based Motion Model" International Journal of Computer Vision (2000).
5. L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert. "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scalespace Based Approach. Technical Report" RR-3874, INRIA (2000).
6. Jochen Schmidt, Heinrich Niemann, and Sebastian Vogt. "Dense disparity maps in real-time with an application to augmented reality", Orlando, FL USA,. IEEE Computer Society, December 3-4 (2002) . IEEE Workshop on Applications of Computer Vision (WACV 2002).
7. Adee Ran, Nir A. Sochen.: Differential Geometry Techniques in Stereo Vision Proceedings of EWCG, pp 98-103 (2000).
8. Ianir Ideses, Leonid Yaroslavsky, Roni Vistuch and Barak Fishbain, "3D video from compressed 2D video" Proceedings of Stereoscopic Displays and Applications XVIII, SPIE and IS&T, San Jose California (2007).

9. L. P. Yaroslavsky, J. Campos, M. Espínola, and I. Ideses, "Redundancy of stereoscopic images: Experimental evaluation," Opt. Express vol 13, 10895-10907 (2005).
10. L. P. Yaroslavsky, "On Redundancy of Stereoscopic Pictures" Acta Polytechnica Scandinavica, n. 149. Image Science '85. Proceedings. Helsinki, Finland., V. 1, p. 82-85, 11-14 June (1985).
11. B. Julesz, *Foundations of Cyclopean Perception*, The University of Chicago Press, 1971, p. 96.
12. 3DTV web site, http://www.3dtv-research.org/
13. L. P. Yaroslavsky and M. Eden, *Fundamentals of Digital Optics*, Birkhauser, Boston, 1996.
14. I. Ideses, L.P. Yaroslavsky, B. Fishbain, :Real-Time 2D to 3D Video Conversion:, Journal of Real-Time Image Processing, Volume 2, Number 1 / October 2007, (2007).