Introducing the Depth Transfer Curve for 3D capture system characterization

Sergio R. Goma, Kalin Atanassov, Vikas Ramachandra
Qualcomm Incorporated.

ABSTRACT

3D technology has recently made a transition from movie theaters to consumer electronic devices such as 3D cameras and camcorders. In addition to what 2D imaging conveys, 3D content also contains information regarding the scene depth. Scene depth is simulated through the strongest brain depth cue, namely retinal disparity. This can be achieved by capturing an image by horizontally separated cameras. Objects at different depths will be projected with different horizontal displacement on the left and right camera images. These images, when fed separately to either eye, leads to retinal disparity. Since the perception of depth is the single most important 3D imaging capability, an evaluation procedure is needed to quantify the depth capture characteristics. Evaluating depth capture characteristics subjectively is a very difficult task since the intended and/or unintended side effects from 3D image fusion (depth interpretation) by the brain are not immediately perceived by the observer, nor do such effects lend themselves easily to objective quantification. Objective evaluation of 3D camera depth characteristics is an important tool that can be used for "black box" characterization of 3D cameras. In this paper we propose a methodology to evaluate the 3D cameras' depth capture capabilities.

Keywords: 3D evaluation, 3D quality, stereo convergence, object disparity, stereopsis, visual fatigue, depth contrast

1. INTRODUCTION

In order to objectively evaluate 3D quality, one needs to :

- Design a special test chart that is practical and contains necessary structure to extract important 3D depth statistics.
- Develop a processing algorithm that extracts important 3D depth statistics.
- Formulate a methodology to visualize and evaluate the 3D depth statistics.

The test chart design should ensure that the camera field of view (FOV) is populated with detectable objects/keypoints at different distances and should be designed to provide an insight into how the depth transfer function depends on the view angle. Once the stereo image of the chart is captured, the comprising left and right images are processed to extract the depth statistics, namely the horizontal and vertical disparities for all the key points in the image. The disparities are extracted by selecting arrays of key points from the captured left images and finding their corresponding counterparts in the right images. The disparity vectors are then separated into vertical and horizontal components. The vertical disparity component is regarded as an undesired artifact and usually results from incorrect camera mounting or incorrect or incomplete image rectification. The horizontal component is indicative of the camera depth capture capability and several useful statistics can be extracted from it. Plotting a depth-to-disparity curve specifies camera depth range capabilities as well as depth rendering granularity. The bias on the curve is indicative of the camera convergence point that can be achieved through toe-in of the cameras or post-processing (shift and crop). Another element of the 3D capability is the depth granularity that will depend on the camera baseline and the spatial resolution of the 3D content. The final goal of the study is to extract important statistics, and to group and display them so that the 3D camera depth capture capabilities can be summarized in a visually intuitive fashion.

Qualcomm Inc., 5775 Morehouse Drive, San Diego CA 92121.

2. EVALUATION METHODOLOGY

3D quality is determined by the stereo camera separation (baseline), camera baseline placement, camera orientation, camera lens distortions, accuracy of camera zoom coordination and 3D post-processing. In order to quantify 3D quality without having access to the exact details of the above six factors, one needs to treat the 3D components together as a black box and extract statistics from the final 3D images. There are several factors that can be used to specify 3D quality.

2.1 Disparity range

Disparity range is determined by the baseline of the cameras. Camera baseline refers to the horizontal distance between the two cameras. To capture the best stereo images for 3D imaging, the baseline must be carefully specified. The camera baseline influences the 3D effect in the following manner:

- A larger baseline produces a larger disparity range. A large disparity range produces a pronounced 3D effect on mid to far depth range scenes. The drawback is that for close scenes the disparity is too large and the scene may not be fusible (Fusibility refers to the left and right eye image being interpreted as a single 3D scene by the brain).
- A smaller baseline produces a smaller disparity range which diminishes the 3D experience for scenes that are in the mid to far depth range. However, the advantage of a smaller baseline is that for scenes with objects close to the camera, the 3D image pair is still fusible.

The baseline effects discussed above are illustrated in the following figure. The figure shows how large of a depth

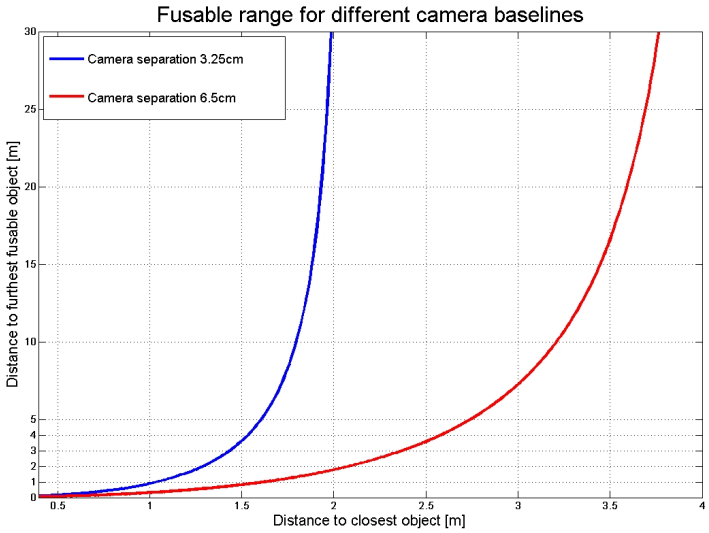


Figure 1. Stereo camera baseline trade offs

range can be captured without eye strain for two different baseline distances. For example, if the closet object is at 2 meters , a camera baseline of 3.25 centimeters can fuse objects up to infinity while a baseline of 6.5 centimeters can fuse objects only up to 7 meters. However, if the closest object is at 3.75 meters and the farthest object is at infinity, a baseline of 6.5 centimeters will have full 3D effect range while a baseline 3.25 centimeters will only use a fraction of available 3D range (less pronounced 3D effect).

2.2 3D Depth Consistency

In an ideal situation, objects that are located at the same distance from the camera should appear with the same horizontal disparity. Failure to do so will cause incorrect depth perception which will manifest itself as curvature of constant depth planes when displayed.[1] Variations in the horizontal disparity for a fixed depth over different positions usually comes from camera misalignment and/or deficiencies in the subsequent processing.

2.3 3D Discomfort Level

3D discomfort usually comes from limits of horizontal and vertical disparity outside of the Panum's fusion range.[2] Of the two, vertical disparity has stronger effects and it is determined by camera position and subsequent 3D processing.

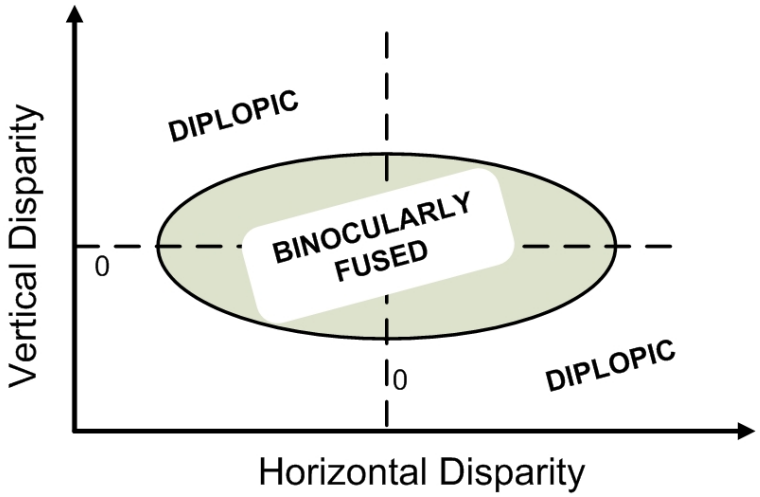


Figure 2. Panum's fusional range

2.4 Evaluation procedure

2.4.1 Test chart

The test chart setup is shown in the following figure. It needs to be three dimensional in order to estimate camera position relative to the chart.[3] This is needed to automatically produce real world coordinates of the detected key points on the images, which is a necessary step in camera calibration. The test chart is made with known dimensions and the green patch is used as the origin of the 3D coordinate system. This helps relate the pixel coordinates of the key points on the camera images to their respective 3D real world coordinates in space.

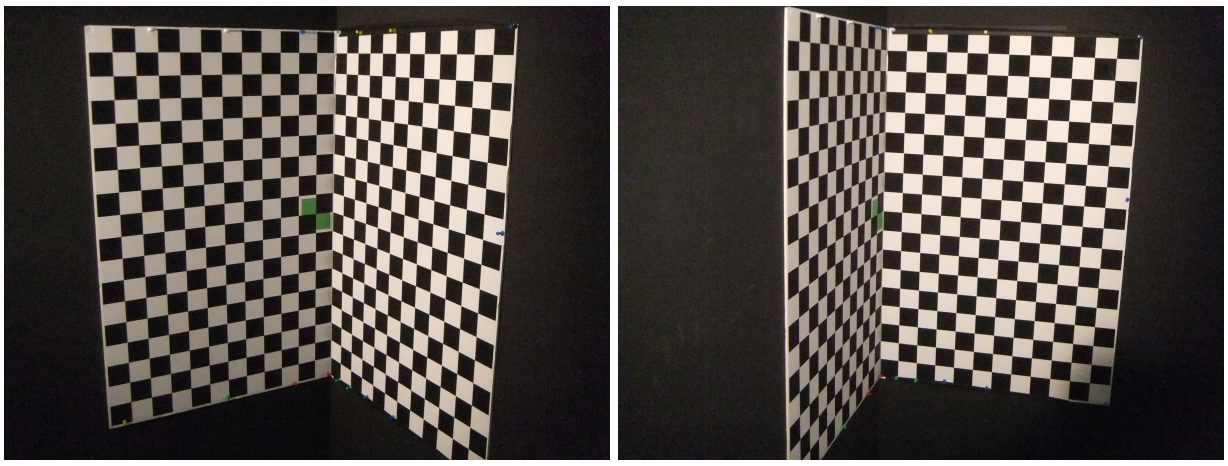


Figure 3. Test chart

2.4.2 Data capture

The purpose of this evaluation is to gather enough key points so that the depth range (in this case, 50cm to 1m) is well represented. Thus, for every depth bin, (a small depth range) key points which lie in that depth

range need to be gathered from camera images taken with cameras looking at the chart from different angles. To achieve the task, several images are captured at different distances and rotations of the camera rig with respect to the chart.

2.4.3 Data processing

Once the stereo image pair is captured, images from each camera are processed independently as follows:

- Detect and label the array of key points using the green patch as an origin.

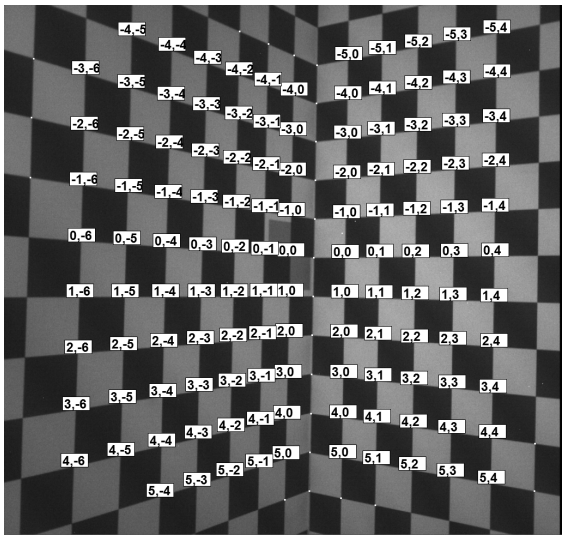


Figure 4. Labeled key points

- Estimate dual camera intrinsic and extrinsic (pose) parameters.
- Place the origin of the coordinate system at the (left) camera center (i.e. remap all real world 3D coordinates of the key point array relative to the camera center).
- Compute disparities for key points that are present in both images.

3. EVALUATION RESULTS

To verify the process, two commercially available 3D cameras have been evaluated: Fuji and Aiptek.

3.1 Baseline estimation

The procedure for baseline estimation is as follows

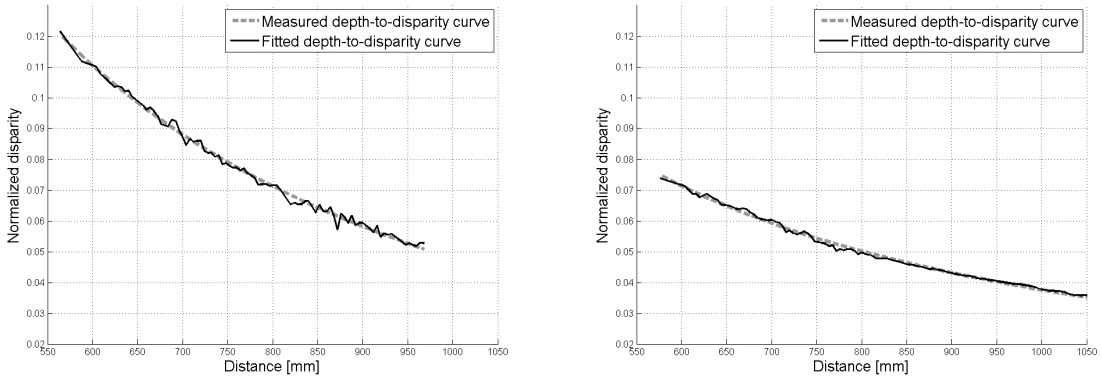
- Compute mean horizonal disparity for every depth bin.
- Measure camera horizontal Field-of-View (hFOV).
- Solve for cameras baseline as

$$CAM_{baseline} = min \sum_{i=0}^{i<N} \left(disp(i) - \frac{c_{base}}{2 \times dist(i) \times \tan(hFOV/2)} - C \right)^2$$

where $disp(i)$ is the normalized (divided by the image width) mean disparity of the depth bin, $dist(i)$ is the distance from the camera, C is a arbitrary constant corresponding to the relative image shift that might be applied, and N is the number of horizontal bins. A normalization process implemented in the above manner has the following properties:

– Is independent of the toe-in between the cameras or the relative horizontal pixel shift (the constant C accommodates for that).
– Is independent of the camera resolution - sub-sampling/up-sampling will not change the final result.
– Is independent of any modification (done through cropping) to the field-of-view of the camera.

The purpose of the normalization is to only observe the factor that affects/determines 3D depth range, namely camera baseline, in the absence of other 2D modifications including change in resolution and/or cropping that can also affect the measured depth of a rendered scene, but are not indicative of the true depth range. Although camera baseline might be available through the camera specs for traditional 3D cameras (two cameras separated by a certain distance) or can be measured, there are several emerging camera technologies that can provide 3D without stereo cameras in the traditional sense. An example for this is the plenoptic camera[4] which does offer horizontal (and vertical) parallax rendering with a baseline that cannot be clearly obtained. Similar technologies split a light beam by mirrors and form left and right camera views through a single lens system.
The synthetic curves for the two cameras are shown in the following plot. The estimated camera baselines were 76mm and 41mm which is consistent with manual baseline measurement.



(a) Fuji (b) Aiptek

Figure 5. Baseline estimation

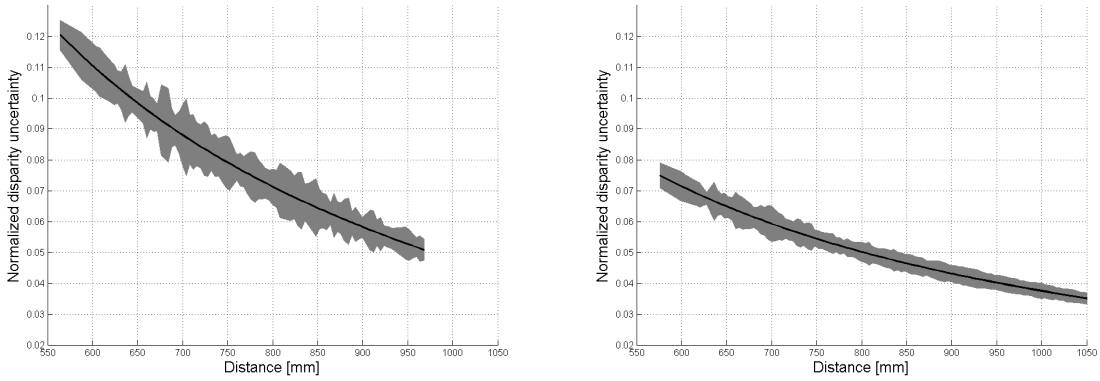
3.2 Depth consistency evaluation

The depth consistency corresponds to the horizontal disparity variation within a depth bin. In the following plot the two $\mathbf{x}$ standard deviation is used to represent the confidence interval.

As indicated before, a depth bin is comprised of several key points that are at the same depth. The size of the bin is around 2mm, i.e., the deviations in the depths of all points in a bin are within 2mm. The deviation range for a depth bin is selected to be as small as possible while ensuring that there are enough points to obtain meaningful statistics. Ideally, all key points in a given bin should produce similar (consistent with the 2mm size of the bin) horizontal disparities. In reality, the measured disparities for the key points in a given bin will not be similar. This will cause an erroneous perception of depth and will be most visible for scenes with a lot of planes. The inconsistency of the depth is usually caused by

- Incorrect camera placement
- Calibration errors
- Images rectification deficiencies
- Inconsistent focus - focus difference between left and right image

- Geometrical lens distortion



(a) Fuji (b) Aiptek

Figure 6. Depth consistency evaluation

3.3 Discomfort level evaluation

So far, the focus has been on horizontal disparities, which are the primary cue for depth perception. Vertical disparities, on the other hand, tend to degrade the perception of depth from horizontal disparities and cause discomfort. The discomfort associated with 3D imaging is highly influenced by the amount of vertical disparities present. With the presence of pronounced vertical disparities, viewers have reported discomfort, headaches, nausea and double vision. Thus, the absence of vertical disparities is necessarily a good indicator of 3D quality.

4. CONCLUSION

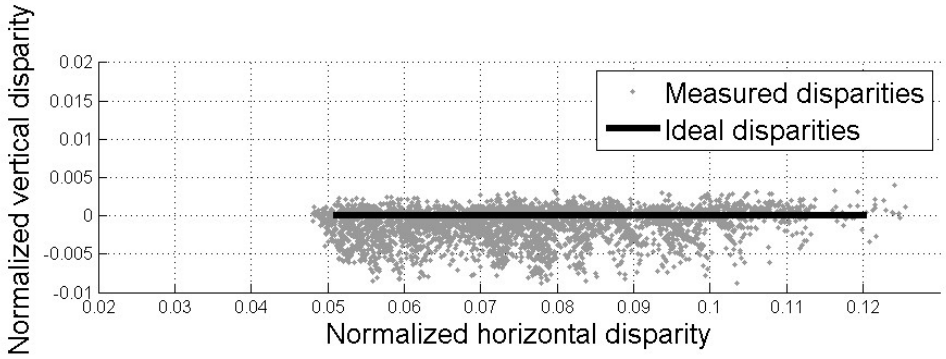
In this paper, a methodology is presented for the evaluation of stereo images based on three features that provide of an objective measure of 3D image quality. These features are: 1) Baseline distance - presenting a trade-off between range of depths and 3D image fusibility. 2) Depth consistency - resulting in different disparities for objects at the same depth from the stereo camera and causing curvature of depth, and 3) Discomfort level - caused by the presence of vertical disparity that results in discomfort and visual fatigue. As eluded before, the baseline estimation is extremely useful for non-traditional 3D cameras that rely on single (and wide) aperture and subsequent mechanical modifications and/or digital processing to generate the left and right image pair. Based on our measurements using two commercial 3D cameras, we were able to extract camera baselines that were consistent with the actual values. It was shown that horizontal disparity consistency and the presence of vertical disparity are also relevant issues that should be evaluated objectively. The latter can be caused by a variety of factors along the image processing pipeline, ranging from incorrect camera placement/calibration to image rectification deficiency and/or inconsistent focus. Note that (the above mentioned) three features are necessary but may not be sufficient in all test scenarios. Further studies are needed in order to evaluate all features (some of which are likely to be scene dependent) automatically to identify a sufficient statistical set for all foreseeable stereo camera calibration issues.

REFERENCES

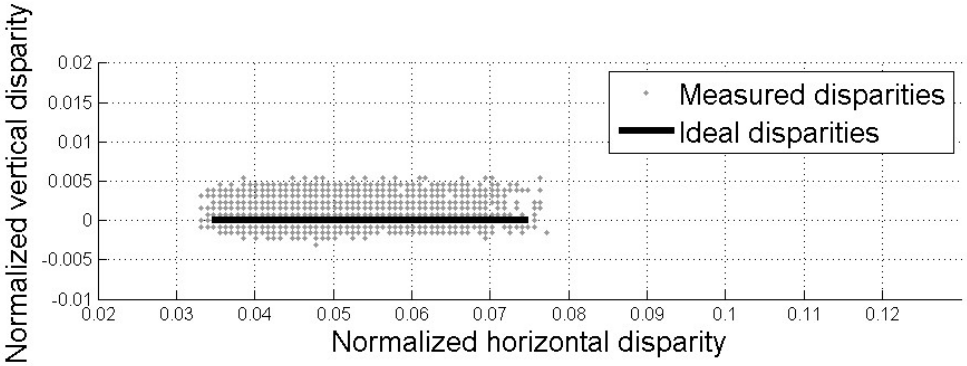
[1] A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," in *Proceedings of the SPIE Volume 1915, Stereoscopic Displays and Applications IV.*, 1993.

[2] C. Tyler, "Sensory processing of binocular disparity," in *Basic and Clinical Aspects of Binocular Vergence Eye Movements.*, 1983.

[3] E. Trucco and A. Verri, "A compact algorithm for rectification of stereo pairs," in *Machine Vision and Applications*, 2000.



(a) Fuji



(b) Aiptek

Figure 7. Disparity scatter plot

[4] N. Ng, M. Levoy, M. Brdif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," in *Stanford University Computer Science Tech Report CSTR 2005-02*, 2005.