SYSTEM FOR AUTOMATIC DETECTION OF DISTORTED SCENES IN STEREO VIDEO

Alexander Voronov, Alexey Borisov, Dmitriy Vatolin

Graphics & Media Lab, Moscow State University
{avoronov, aborisov, dmitriy}@graphics.cs.msu.ru

ABSTRACT

In this paper we describe our methods for estimating stereo video quality and its suitability for certain displays and projectors. The first issue we investigate is that not every 3D video can be viewed comfortably on a given 3D display owing to its particular features, especially parallax. The second issue is the correspondence between left and right views. Because of the complexity of calibrating and synchronizing video capture devices, some mismatch problems and artifacts may appear. These artifacts can affect the 3D viewing experience. Our goal is to detect and, in some cases, suggest how they can be suppressed. We demonstrate our current results and describe directions for further development and improvement of the implemented techniques.

1. INTRODUCTION

Nowadays 3D video is growing in popularity. The amount of available 3D video content and the number of different 3D displays and projectors are growing. The first problem we encounter is that not every 3D video can be viewed comfortably using a given device. Different devices have different comfortable parallax ranges, and (for example) 3D effects in videos prepared for a large screen will be unnoticeable on home displays. Thus, we need a tool to estimate stereo video parallax, and we must also determine what parallax is suitable for a particular device.

A variety of techniques are available for 3D video creation. Video can be captured by a stereo camera, converted from traditional 2D format or rendered from a 3D model. Sometimes combinations of these approaches are used. When creating 3D content, new requirements arise for achieving high-quality results. Capturing video using a stereo camera requires precise camera settings with fine-tuned angles, and we therefore assume that the color and geometry of the pictures taken by two cameras in a stereo rig may differ slightly. Furthermore, zooming shots may be unsynchronized. We must control sharpness in each view; the same object must be in focus in both the left and right views.

When converting from 2D format to stereoscopic 3D, processing of edges and occlusion areas is the most important consideration. With improper processing of semitransparent edges and thin objects, some parts of the foreground may end up stuck to the background in the resulting video. For complex boundaries or backgrounds, insufficiently accurate processing can produce specific jitter, which is noticeable in moving objects. Such problems in stereo production can spoil the experience of viewing 3D video.

2. RELATED WORK

The problems described above are under active investigation. We have conducted an overview of techniques for detecting and correcting artifacts and problems in stereo video – mostly focusing on detection.

Different methods enable control and enhancement of color correspondence between views, using both a global transform for the entire frame [1] and a local transform for analyzing areas in a given frame separately [2]. These methods allow us to estimate the color difference between views. Most of the works on correction of geometric distortions refer to the stereo rectification problem [3, 4] and describe calculations for parameters such as vertical parallax, tilt and vertical perspective distortion. A special method for zoom mistiming detection was suggested in [5]. To detect different sharpness levels in a stereomate, it is possible to match feature points between views, estimate blur measure [6] in the images and compare results in corresponding areas. Another important technique is time synchronization between views. A wide variety of methods exist for solving this problem. Some of these methods match a videos 3D structure in $width \times height \times time$ axes on the pixel or superpixel level [7], and some reduce the videos complexity by extracting a special trace from the sequence and then matching it. Examples of such traces are pixels in epipolar lines [8] and the number of points of interest in a frame [9]. Estimation of time desynchronization is also important because sometimes a stereo motion picture is produced from mono by time shifting one view. But if motion in the video is uneven, the parallax of the resulting stereo is unstable.

3. PROPOSED METHOD

Our goal is to create a stereo video analysis tool that will control the maximum parallax and automatically detect problems in stereo video shots. We have thus far implemented the following features:

- maximum parallax estimation
- color distortion estimation
- geometric distortion estimation

To perform matching between views we use a block-based motion estimation algorithm, which is described in [10]. This algorithm estimates motion with quarter-pixel accuracy for block sizes of 4×4 to 16×16 pixels. One of the reasons for using this block-based approach is processing speed: block-based matching is much faster than optical flow or point-based matching.

3.1. Maximum parallax estimation

This procedure comprises the following steps:

1. Estimation of the motion vector field between views
2. Motion vector field filtering
3. Estimation of the rough disparity map and distribution of depths in the frame
4. Analysis of depth distribution throughout the entire video and detection of problem shots

Filtering of the motion vector field (MVF) eliminates outliers, which mostly appear in uniform areas. Also, a noteworthy point is that depth artifacts are usually almost imperceptible in such uniform untextured areas, so these areas don't play an important role in the final result. Filtering of the MVF is based on the confidence measure from [11]. In that work the authors estimate confidence according to the variance in an image block, the SAD (sum of absolute differences) error in compensation, and the correspondence between the vector direction and its neighbors. This step yields the confidence map for the MVF; this map is used later when calculating the depth distribution in a frame.

We estimate disparity as the horizontal component of a motion vector for a given block and then compute the disparity histogram for the frame; only blocks with high confidence are taken into account. Then we clip 5 percent of all blocks from the boundaries of the histogram and assume the final left- and rightmost values are the minimum and maximum parallax values.

To analyze depth we plot a distribution graph that shows the number of pixels according to the disparity in the timeline. Beforehand, we asked user to input the display system characteristics, which can clearly indicate the maximum artifact-free parallax level; an example parallax monitoring graph is shown in Fig. 1. We then mark frames that have a parallax exceeding a certain value.

(a) Left view (b) Right view

Fig. 3: Example of color distortion from "Avatar Pandora Discovered" at frame 1,251. A large difference in luminance between the views was detected.

3.2. Color distortion estimation

This procedure detects shots with color mismatch between the left and right views. The reason of such distortions is typically in cameras. First, even for identical cameras there may be some difference in sensor readout that yields different color tones in different views for the same object. Second, not all cameras allow manual exposure, white balance setting or high-quality automatic correction, thus potentially causing differences in images' gamma.

For each view we form three histograms (one for each color plane in RGB), then we smooth them using a Gaussian kernel and calculate the SAD among the histograms. When this value is greater than a predefined threshold, we assume that a noticeable color distortion is present. Next we place the estimated values in the timeline (Fig. 2); the shots for which our measure exceeds the acceptable threshold are visible. Fig. 3 shows an example of a problem stereomate.

3.3. Geometric distortion estimation

This procedure enables detection of shots for which the cameras were inaccurately set and calibrated.

Currently we estimate vertical parallax and tilt without using epipolar geometry computation or stereo rectification. Vertical parallax is estimated as the median value of the y-component among all the vectors with high confidence. To estimate tilt we subtract the vertical parallax value from each vector; the estimate of the possible rotation angle is the median tilt angle for each confident vector. In these calculations we take into account the distance of a point from the image center as well as its new y-coordinate.

Analysis of the estimated values in the timeline enables detection of problem shots (Fig. 4). A sample of visualization is shown in Fig. 5.

4. IMPLEMENTATION

The algorithm is implemented as combination of a VirtualDub filter and a MATLAB script. The VirtualDub filter performs video processing and estimates different measures for each

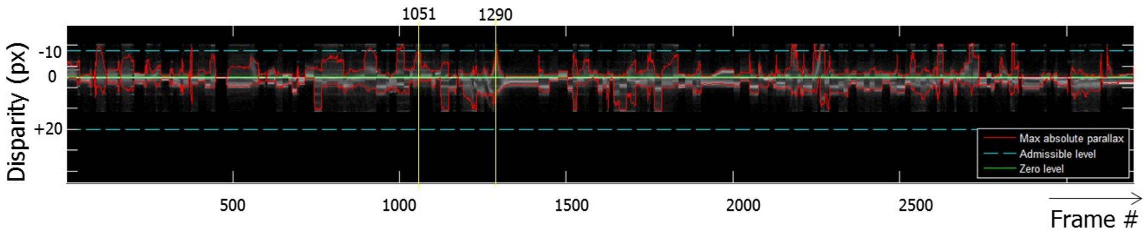


Fig. 1: Example of parallax control graph for the movie "Pirates of the Caribbean: on Stranger Tides." The parallax clearly differs from scene to scene, and a large parallax is achieved only in a small portion of scenes.

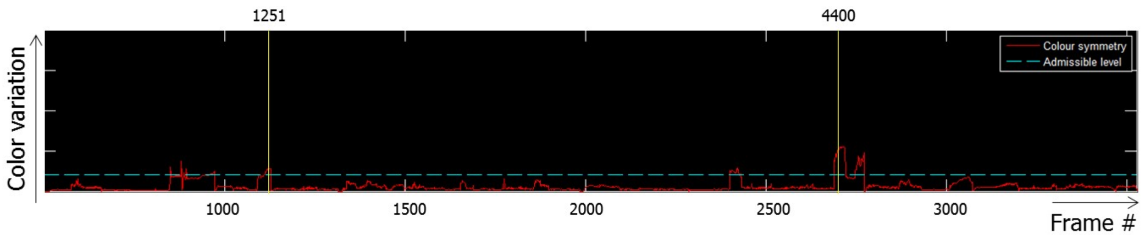


Fig. 2: Color distortion graph for "Avatar Pandora Discovered." The locations of detected problems are marked with yellow lines.

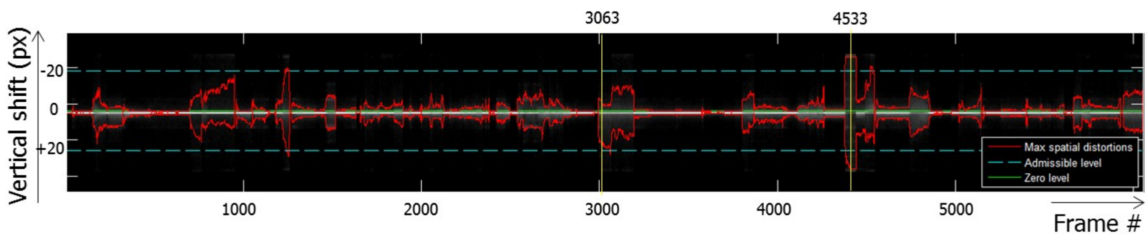


Fig. 4: Geometric distortion graph for "Avatar Pandora Discovered." The locations of detected problems are marked with yellow lines.

frame. The filter writes the data to a text file, and the MATLAB script analyzes the data and plots graphs.

5. RESULTS

To demonstrate our tool, we tested 10 trailers for new and well-known 3D movies:

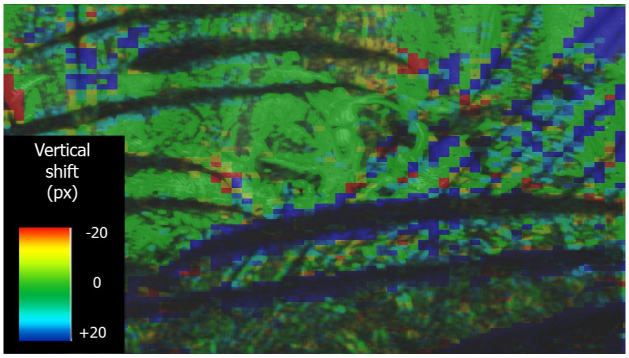
- Avatar Pandora Discovered
- Pirates of the Caribbean: On Stranger Tides
- Transformers 3: Dark of the Moon
- Alice in Wonderland
- The Priest
- Resident Evil 4: Afterlife
- The Smurfs
- Bitwa Warszawska
- Tron: The Legacy
- The Chronicles of Narnia: The Voyage of the Dawn Treader

Our analysis revealed some problems in these trailers. For the "Avatar" trailer, differences in luminance were detected, as Fig. 3 shows. The trailer also contains some geometric distortions; for example, Fig. 6 demonstrates a problem frame that is a combination of captured video and computer rendering. The rendered parts are placed correctly, but the captured video appears mistimed between views. For some shots in "Pirates of the Caribbean," rather noticeable differences in color tones are apparent (Fig. 7). Other trailers also contain some shots with color distortions (Fig. 8).

Another interesting point is analysis of parallax control in different movies. In "Pirates of the Caribbean" we see appreciable 3D effect only in some scenes (specifically, in close-up scenes), but most shots have little noticeable 3D character. For other trailers – for example, "Transformers 3" or "Bitwa Warszawska" – we can see that the depth grading procedure was accurately performed: parallax is adjusted for the whole video. Fig. 9 presents graphs demonstrating different strategies for depth grading. Also we can select

(a) Left view

(b) Right view



(c) Geometric distortion visualization

Fig. 5: Example of geometric distortion visualization for “Avatar Pandora Discovered” at the frame 3,063. Color hue is the vertical shift value given in the legend, and color brightness is the confidence measure for this point: points with a higher confidence are marked with a brighter color, and points with a lower confidence are marked with grey.

scenes with maximum parallax for proper testing. For example, in “The Smurfs” the maximum parallax appears in the scene with fast motion, and it doesn’t cause discomfort in 3D perception (Fig. 10).

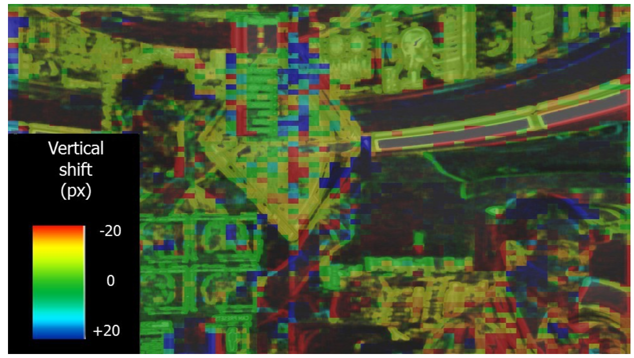
6. FURTHER WORK

We are pursuing two main directions of further work. The first is improvement in the stability and reliability of results. Currently some problems arise in low-contrast areas and for fading and cross-fading (dissolving) between scenes. Processing of low-contrast areas can be improved using color correction and a better motion vector confidence measure. When processing fading between scenes, we are attempting to detect this effect and also estimate the resulting confidence; for a low-confidence result, values will be interpolated using neighbors that have a higher confidence. Using a scene-change detector can also improve stability, as it can split the video into distinct scenes and analyze the results in each scene more accurately.

The second direction of further work is increasing the number of estimated features and types of detectable artifacts. We intend to implement a detection scheme for time and zoom desynchronization, as well as for focus variations

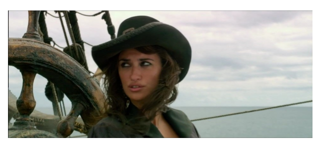
(a) Left view

(b) Right view

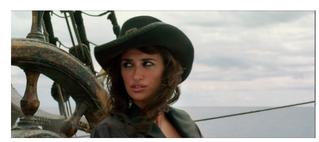


(c) Geometric distortion visualization

Fig. 6: Example of geometric distortion for “Avatar Pandora Discovered” at frame 4,533. This frame is a combination of captured video and computer graphics; that computer graphics portion is placed correctly, but the captured video has geometric distortion.



(a) Left view



(b) Right view

Fig. 7: Example of color distortion from “Pirates of the Caribbean” at frame 1,140.

between views. Then we plan to improve detection of object boundaries and to create a database of videos with different stereo-capture artifacts to enable proper testing of our tool. Using analysis data for some samples captured by a stereo camera, it may be possible to provide instructions for calibrating video capture devices (e.g., to reduce geometric distortions or variations in sharpness).

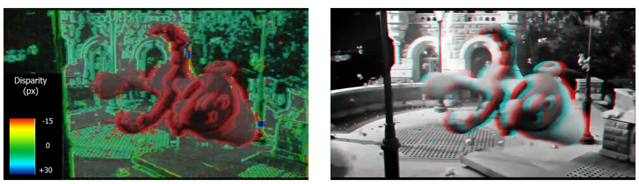
7. CONCLUSIONS

In this paper we have demonstrated our results thus far in developing an automatic stereo video quality assessment tool. Our tool detects color and geometric distortions between views and performs parallax control. This tool is useful for automatic detection of problem shots in video. Example results were demonstrated for trailers of well-known

(a) Left view (b) Right view

Fig. 8: Example of color distortion from "Resident Evil 4: Afterlife" at frame 2,790.

(a) Left view (b) Right view



(c) Disparity map visualization (d) Anaglyph stereo

Fig. 10: Frame 1,962 from "The Smurfs" trailer with example of maximum parallax visualization. This scene has fast motion, so a large parallax doesn't cause discomfort in 3D perception.

movies, and we showed that these trailers contained some artifacts. Our tool is in an early development stage, but in the future, we plan to increase the number of solutions to recognizable problems and improve the stability of results.

8. ACKNOWLEDGEMENTS

This work is partially supported by the Intel/Cisco Video-Aware Wireless Network (VAWN) Program and by grant 10-01-00697a from the Russian Foundation of Basic Research.

9. REFERENCES

[1] Kenji Yamamoto and Ryutaro Oi, "Color correction for multi-view video using energy minimization of view networks," *International Journal of Automation and Computing*, vol. 5, no. 3, pp. 234–245, 2008.

[2] J.J. Yu, H.D. Kim, H.W. Jang, and S.W. Nam, "A hybrid color matching between stereo image sequences," *Proceedings of 3DTV-Conference 2011*, pp. 1–4, 2011.

[3] Guo-Yu Lin, Xu Chen, and Wei-Gong Zhang, "A robust epipolar rectification method of stereo pairs," *Measuring Technology and Mechatronics Automation, International Conference on*, vol. 1, pp. 322–326, 2010.

[4] J. Zhou and B.X. Li, "Image rectification for stereoscopic visualization without 3d glasses," in *Image and Video Retrieval*. 2006, pp. 495–498, Springer Berlin / Heidelberg.

[5] Colin Doutre, Mahsa T. Pourazad, Alexis M. Tourapis, Panos Nasiopoulos, and Rabab K. Ward, "Correcting unsynchronized zoom in 3d video," *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 3244–3247, 2010.

[6] Luhong Liang, Jianhua Chen, Siwei Ma, Debin Zhao, and Wen Gao, "A no-reference perceptual blur metric using histogram of gradient profile sharpness," *Proceedings of the 16th IEEE International Conference on Image Orocessing*, pp. 4309–4312, 2009.

[7] Yaron Caspi and Michal Irani, "Spatio-temporal alignment of sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1409–1424, 2002.

[8] V. Guitteny, R. Benosman, and C. Charbuillet, "Synchronizing video sequences from temporal epipolar lines analysis," in *Advanced Concepts for Intelligent Vision Systems*, October 2008.

[9] Jingyu Yan and Marc Pollefeys, "Video synchronization via space-time interest point distribution," *Proceedings of Advanced Concepts for Intelligent Vision Systems*, 2004.

[10] Karen Simonyan, Sergey Grishin, Dmitriy Vatolin, and Dmitriy Popov, "Fast video super-resolution via classification," pp. 349–352, 2008.

[11] Karen Simonyan, Sergey Grishin, Dmitriy Vatolin, and Dmitriy Popov, "Confidence measure for block-based motion vector field," *Graphicon*, pp. 110–113, 2008.

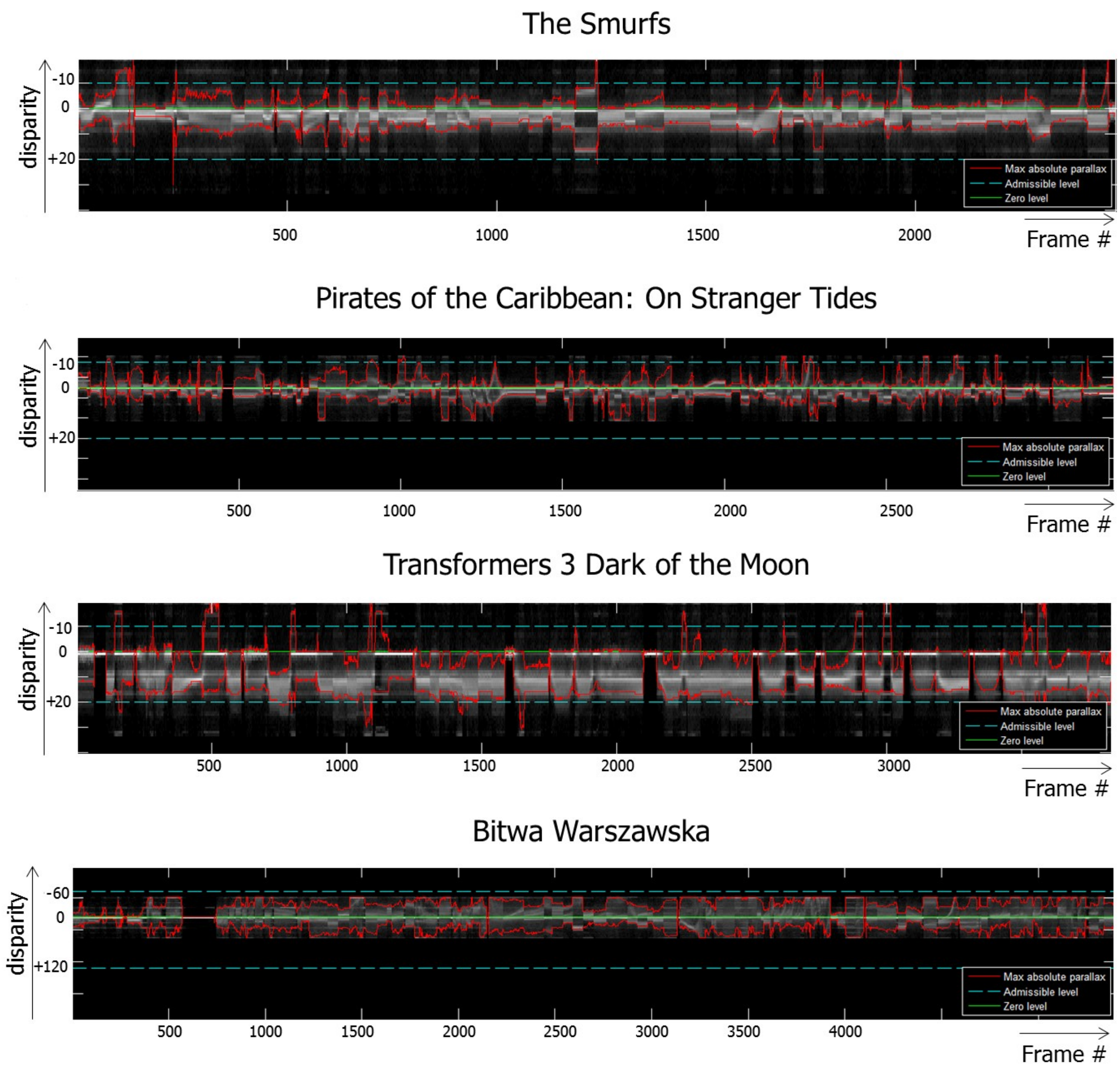


Fig. 9: Graph demonstrating different parallax control strategies for different movie trailers. For the movies “Transformers 3” and “Bitwa Warszawska,” the depth was consistent throughout the entire video, but in “Pirates of the Caribbean” and “The Smurfs,” scenes have different 3D effects.