GHOSTING AND POPPING DETECTION FOR IMAGE-BASED RENDERING

S. Guthe, P. Schardt, M. Goesele

TU Darmstadt, Germany

D. Cunningham

BTU Cottbus - Senftenberg, Germany

ABSTRACT

Film sequences generated using image-based rendering techniques are commonly used in broadcasting, especially for sporting events. In many cases, however, image-based rending sequences contain artifacts, and these must be manually located. Here, we propose an algorithm to automatically detect not only the presence of the two most disturbing classes of artifact (popping and ghosting), but also the strength of each instance of an artifact. A simple perceptual evaluation of the technique shows that it performs well.

Index Terms — image-based rendering, artifact detection

1. INTRODUCTION

Most image-based rendering (IBR) sequences involve a camera moving through a (mostly) static scene and are usually constructed from individual photographs, with the occasional support of depth maps. The fact that such scenes do not require expensive recording equipment has helped to make them increasingly popular (e.g. Google Street View, sports broadcasts). Unfortunately, the rendered sequences often exhibit artifacts like ghosting and popping, which must be manually located. Once located, these artifacts can generally be removed or at least reduced through the use of additional photographs (or through better depth maps). Naturally, the manual search for artifacts is very tedious and time consuming. Thus, the ability to automatically detect the location and strength of artifacts in a sequence would be very helpful, especially if it could also be coupled with additional steps to maximize the final rendering quality while minimizing the amount of additional photographs that are required.

2. RELATED WORK

In order to develop a perceptually-meaningful image-quality measure for artifacts in IBR sequences, we need to examine the relevant aspects of the human visual system, standard IBR algorithms, and artifact detection.

The Human Visual System: In order to quantify the perceived quality of an image sequence, we need to know how images are processed by the human visual system. The fact that over 50% of the cerebral cortex is dedicated solely to visual perception [1] emphasizes the incredible complexity of the human visual system. As a natural consequence of this complexity, most of the research into visual perception over the last 150 years has focused on the early stages visual processing, primarily on how visual features are extracted and represented [2]. Despite the fact that many layers of complex processing follow the early stages—and that these have a strong influence on what information we can see or use—several models of early visual processing have been used to successfully detect changes in an image using only very simple visual features such as contrast and spatial frequency [3].

The very first stage of visual processing is the enhancement of local contrast (using the lateral inhibition between cells in the retina). This, combined with other aspects of the early visual system help to detect and enhance edges [4]. The central role of high-contrast edges in visual processing can be seen in the fact that every model of visual processing focuses—sometimes exclusively—on edges [5].

The perception of changes over time is likewise a central aspect of visual perception [6]. Indeed, there is an increasing body of evidence that the earliest cells in the visual cortex are not static detectors of edges as previously thought, but change the shape of their receptive field over time (i.e., respond to complex, dynamic edges) [7]. Thus, the sensitivity to temporal modulation and edges are inherently linked and combine for impressive sensitivity [8].

Of course, color also plays an important role. The human visual system represents color in a three-dimensional space. Initially, this space is based on the response properties of the three types of cone in the fovea. Rather quickly, though, the system switches over to a color-opponent system, which is approximated well by the CIELAB color space. The three axes of CIELAB—which are statistically independent from each other—are luminance (L), red-to-green (A), and blue-to-yellow (B). We use the CIELAB color space in all our calculations. More detailed information on the relationship between image statistics and the human visual system can be found in [5].

In sum, the early visual system is tuned to temporal discontinuities (i.e., sudden disappearances or appearances) and rapidly-changing, high-contrast edges [9]. It should not be surprising, then, that the most noticeable artifacts in IBR sequences occur when moving edges (or indeed entire objects) suddenly appear/disappear (popping) or when they fade in/out (ghosting).

Image-Based Rendering: One of the first IBR algorithms was based on view interpolation [10]. Even though this approach used depth maps to interpolate between images, it produced strong ghosting artifacts and popping. In contrast, lumigraph-based approaches [11] do not use a depth map, but instead represent a significant part of the plenoptic function (note that depth information may be implicitly encoded in the plenoptic function). Even with a lumigraph, however, interpolation between different views is required, unless an unreasonable amount of data is involved. This will, in turn, cause a lot of ghosting. The unstructured lumigraph [12] tries to compensate for this by providing a better parametrization of the plenoptic function, but ghosting still remains an issue. View-depended texture maps [13], which rely on globally consistent geometry, are prone to producing ghosting artifacts for small details that are not well represented in the geometry. Floating textures [14] try to improve on this by warping the view-dependent textures instead of just interpolating between them. The

Bayesian approach by Cayon et al. [15] is a hybrid approach using depth-based, coherent, superpixel warping together with view interpolation. Although better, this approach still produces some ghosting and popping artifacts, especially if the estimated depth is not entirely accurate.

Artifact Detection: In general, automatic evaluations of IBR quality tend to be reference based [3, 16, 17, 18]. Moreover, most of the related work on artifact detection concentrates on detecting compression [19, 20] or application specific artifacts, such as MRI scanning related artifacts [21]. Vangorp et al. [22] investigated artifacts in the context of IBR and classified them according to type (blending/ghosting, popping, and parallax distortion), visibility, and severity, but did not supply an automatic way to detect these artifacts. Berger et al. [23] presented a method for detecting ghosting artifacts that is based on detecting edges. They explicitly use still images, however, and do not take any temporal effects into account. Schwarz and Stamminger [24] presented a perceptually-motivated popping detection algorithm which can not easily be extended to detecting ghosting artifacts.

3. ARTIFACT DETECTION

The first step in automatically detecting artifacts in IBR should be to detect the most disturbing artifacts: ghosting and popping. Since both of these are explicitly edge- and motion-based, we need to be able to detect and track moving edges. For this, we chose the optical flow algorithm of Farnebäck [25] since it efficiently (i.e., with low computational complexity) produces a robust dense optical flow. Once we know the correspondences between pixels, we can classify each pixel as ghosting, popping or no-artifact. Notice that some pixels will lack a correspondence because the surface they represent has become occluded or left the scene (and are not popping). Thus, we initially focus on static scenes and do not evaluate the border pixels (which generally will leave the scene) which we define as the outermost 1% of the input image width and height.

3.1. Ghosting

Conceptually, we define ghosting as a smooth, almost linear transition of colors for a moving pixel over the duration of several frames. First, we define x_t as the position of a pixel in frame t, u_t as the corresponding flow vector, and $I(x_t)$ as its color in frame t. The corresponding position x_{t+1} in frame $t+1$ is thus $x_{t+1} = x_t + u_t$. We now define a pixel as a candidate for ghosting if the color changes significantly over several $(2n)$ frames, i.e.

$$I(x_{t-n}) \neq I(x_{t+n}) \tag{1}$$

and if the change in color is almost constant over these frames.

$$\begin{aligned} I(x_{t-n}) - I(x_{t-n+1}) &\approx I(x_{t-n+1}) - I(x_{t-n+2}) \\ &\approx \dots \\ &\approx I(x_{t+n-1}) - I(x_{t+n}) \end{aligned} \tag{2}$$

We have to assume that there is a certain amount of noise in the input sequence and that there is some inaccuracy in the optical flow. Thus, we define two thresholds c_{ghost}, $c_{nonlinear}$ and consider a pixel to be a ghosting pixel if

$$\|I(x_{t-n}) - I(x_{t+n})\| > c_{ghost} \tag{3}$$

and

$$\|I(x_{t+i-1}) - 2I(x_{t+i}) + I(x_{t+i+1})\| \leq c_{nonlinear} \tag{4}$$

for all $-n < i < n$. Note that Equations 2 and 4 need to be evaluated in the color space of the IBR algorithm. The strength s_g of the ghosting artifact depends on the color difference found in Equation 3.

3.2. Popping

Similar to ghosting, we look at corresponding pixels. However, for popping artifacts, we are only interested in the difference between the current and the previous frames. Thus, Equation 1 changes to:

$$I(x_t) \neq I(x_{t-1}) \tag{5}$$

To compensate for noise in the input sequence, we again use a detection threshold c_{pop} and only consider a pixel to be popping if

$$\|I(x_t) - I(x_{t-1})\| > c_{pop} \tag{6}$$

Unfortunately, this approach is very sensitive to inaccuracies in the optical flow calculations. Thus, we need to check a small (3×3) neighborhood around the pixel x_{t-1} to see if one of its neighbors might have been a better correspondence. If we find any pixel y such that

$$\|I(x_t) - I(y_{t-1})\| \leq c_{pop} \tag{7}$$

we decide that popping did not occur and discard the pixel x_t. The strength s_p of the popping artifact depends on the color difference found in Equation 6.

4. QUALITY METRIC

After the detection phase, the strength of each artifact is defined for individual pixels in individual frames. To rate the entire sequence, we need to combine these separate numbers into a single quality measure, starting by determining the quality of individual frames. Given a sequence of T frames where each frame t has N pixels, we define D_t to be the collection of all pixels in frame t that were detected as artifacts. Since Vangorp et al. [22] found that ghosting artifacts elicit a stronger response, we scale the artifact strength by a constant factor w_g. If a pixel was detected as both ghosting and popping, we use the stronger detection. Thus, the overall artifact strength S_t for any given frame is

$$S_t = \sum_{x_i \in D_t} \max\left(s_p(x_i), w_g \cdot s_g(x_i)\right) \tag{8}$$

Note that we skip over all frames where more than 25% of the pixels were detected as popping artifacts since we consider these frames to be scene changes.

Since the quality is reciprocal to the number and strength of the artifacts, we define the quality for frames that has artifacts as:

$$Q_t = \frac{N}{S_t} \tag{9}$$

If no artifacts were detected, the quality becomes infinite, similar to the PSNR in the absence of noise. Further, for a whole sequence, we define Q_{avg} as the average quality

$$Q_{avg} = \frac{N \cdot T}{\sum_{t=1}^{T} S_t} \tag{10}$$

and the minimum quality

$$Q_{min} = \min_{t \in [1..T]} Q_t \tag{11}$$

Note that even if a single frame does not contain any artifacts, the quality of a video sequence Q_{avg} will not be infinite as long as there is a single artifact in any of the frames.

(a) outdoor 1 (b) outdoor 2

(c) outdoor 3 (d) outdoor 4

(e) indoor 1 (f) indoor 2

Figure 1: Scenes are sorted in each class according to the results of our quality metric from worst to best. All outdoor sequences (a-d) were taken from Cayon et al. [15]. The second indoor sequence (f) is the ground truth for the first indoor sequence (e).

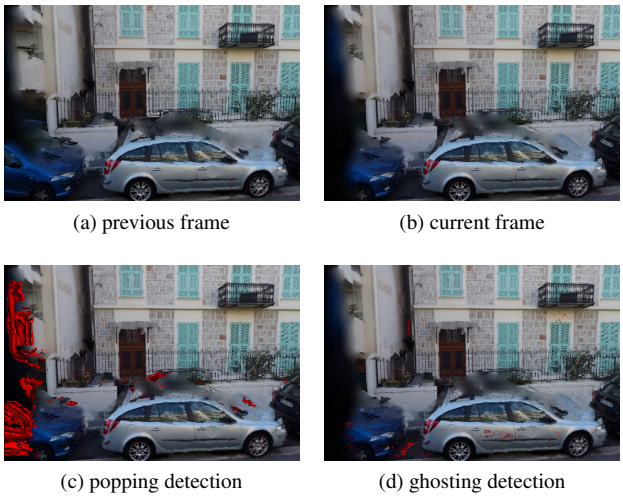
5. RESULTS

We tested our quality metric on 5 IBR sequences, and one real scene. As seen in Figure 1, four of the sequences depict outdoor environments while two show indoor scenes. Note that the second indoor scene is a ground truth video for validation purposes.

We computed the quality metrics (Q_{avg} and Q_{min}) using the following pre-determined thresholds: $c_{ghost} = 7.5$, $c_{nonlinear} = 5$, and $c_{pop} = 10$. Since the color distance calculations are in CIELAB color space, a value of 1 corresponds to roughly 1% of the difference between back and white, and is perceptually just noticeable. Even though most current IBR algorithms use CIELAB space during rendering, the $c_{nonlinear}$ term can compensate for those that use RGB space during interpolation. We analyzed the IBR sequences and searched for the frames with the highest artifact detection. Figure 2 shows the frame that had the highest popping and overall artifact detection for each sequence. Overall, it seems that the ghosting artifact weight w_g needs to be set to approximately 10 to compensate for the difference in detection strengths.

We conducted a small perceptual validation with 18 participants (12 male, 5 female, aged between 25 and 45). In the study, participants were asked to rate the quality of each video based on the 5-point Likert-type scale (with 5 being very good and 1 very bad). Artifacts were never mentioned and participants were explicitly told to use their own intuitive definition of "image quality". The ratings were submitted to a one-way, within-participants ANOVA, which showed that the ratings of sequences differ significantly from each other ($F(5,85)=37.78$, $p<0.0001$)[1].

[1] for more on inferential statistics, see [26].



(a) previous frame (b) current frame

(c) popping detection (d) ghosting detection

Figure 2: Popping (c) and ghosting (d) detection for a single frame of the first outdoor sequence. Most of the detected popping artifacts are easily visible when comparing the current frame (b) with the previous one (a).

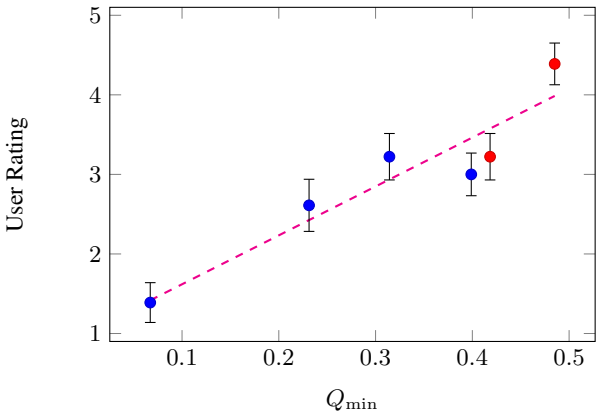


Figure 3: Q_{min} compared against 5-point Likert-type ratings (error bars show the 95% confidence interval). The dashed line shows the linear regression though both the outdoor (blue) and indoor (red) sequences.

As can be seen in Figures 3 and 4, the minimum quality Q_{min} correlates very well with the average user rating (the Pearson correlation coefficient R is 0.9339) but the average quality Q_{avg} does not (with $R = -0.2641$). A possible explanation is that the ghosting detection assumes an almost constant movement and thus linear blending for the 5 frames used in the detection. Since some of the outdoor sequences do not have constant motion for large parts of the video, the detection of ghosting artifacts fails for these frames. Also, the camera comes to a nearly complete stop multiple times during these sequences, so there is no artifact detection at all. Finally, after the experiment, we asked participants how they rated the images and they explained that they rated a sequence based on the worst couple of frames rather than the average.

6. CONCLUSION AND FUTURE WORK

We presented a perceptually-motivated algorithm that is able to detect both popping and ghosting artifacts in an IBR sequence.

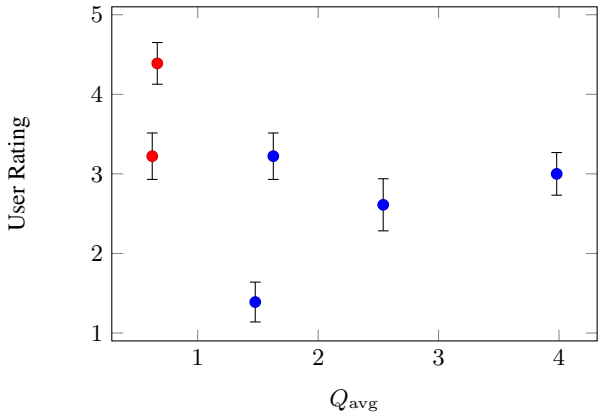


Figure 4: Q_{avg} does not correlate well to the participants' ratings, particularly if the speed of camera movement varies a lot as in the outdoor sequences (blue).

The algorithm includes two ways of generating a single quality score for the whole sequence. One of them, the minimum quality, is able to predict human image-quality ratings. The fact that the minimum quality score can predict human performance, but the average quality score cannot confirmed our assumption that the quality of a IBR sequence is based on the worst artifacts rather than the overall impression.

In the future, we will examined methods for making ghosting detection more robust against changing camera motions. We will also explore how to use our quality estimate to find the next best camera in view planning or view selection problems.

7. ACKNOWLEDGEMENTS

This work was supported in part by the European Commission's Seventh Framework Programme under grant agreements no. ICT-611089 (CR-PLAY).

8. REFERENCES

[1] A. Milner and M. Goodale, *The Visual Brain in Action*, vol. 27, England, 1995.

[2] W. Thompson, R. Fleming, S. Creem-Regehr, and J. Stefanucci, *Visual Perception from a Computer Graphics Perspective*, CRC Press, 2011.

[3] T. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic Range Independent Image Quality Assessment," in *ACM Trans. on Graphics (TOG)*, 2008, vol. 27, p. 69.

[4] H. Hartline, H. Wagner, and F. Ratliff, "Inhibition in the Eye of Limulus," *The Journal of general physiology*, vol. 39, no. 5, 1956.

[5] T. Pouli, D. W. Cunningham, and E. Reinhard, *Image Statistics in Visual Computing*, A. K. Peters, Natick, MA, USA, 2013.

[6] J. J. Gibson, *The Ecological Approach to Visual Perception*, Lawrence Erlbaum, Hillsdale, NJ, 1979.

[7] G. DeAngelis, I. Ohzawa, and R. Freeman, "Receptive-Field Dynamics in the Central Visual Pathways," *Trends in Neurosciences*, vol. 18, no. 10, 1995.

[8] S. Jackman, N. Babai, J. Chambers, W. Thoreson, and R. Kramer, "A Positive Feedback Synapse from Retinal Horizontal Cells to Cone Photoreceptors," *PLoS Biol*, vol. 9, no. 5, 2011.

[9] X. Gao, W. Lu, D. Tao, and X. Li, "Image Quality Assessment and Human Visual System," in *Visual Communications and Image Processing*. International Society for Optics and Photonics, 2010.

[10] S. Chen and L. Williams, "View Interpolation for Image Synthesis," in *Proc. ACM SIGGRAPH*, 1993.

[11] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The Lumigraph," in *Proc. ACM SIGGRAPH*, 1996.

[12] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured Lumigraph Rendering," in *Proc. ACM SIGGRAPH*, 2001.

[13] P. Debevec, Y. Yu, and G. Borshukov, *Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping*, 1998.

[14] M. Eisemann, B. De Decker, M. Magnor, P. Bekaert, E. De Aguiar, N. Ahmed, C. Theobalt, and A. Sellent, "Floating Textures," in *CGF*, 2008, vol. 27.

[15] R. Cayon, A. Djelouah, and G. Drettakis, "A Bayesian Approach for Selective Image-Based Rendering using Superpixels," in *Proc. IEEE 3DV*, 2015.

[16] K. Mueller, X. Zabulis, A. Smolic, and T. Wiegand, "Evaluation of 3D Reconstruction using Multiview Backprojection," 2005.

[17] C. Weigel and F. Fan, "GPU-Based 3D Video Object Synthesis and Its Quality Assessment," in *Proc. 3DTV*. IEEE, 2008.

[18] M. Waechter, M. Beljan, S. Fuhrmann, N. Moehrle, J. Kopf, and M. Goesele, "Virtual Rephotography: Novel View Prediction Error for 3D Reconstruction," *CoRR*, vol. abs/1601.06950, 2016.

[19] T. Vlachos, "Detection of Blocking Artifacts in Compressed Video," *Electronics Letters*, vol. 36, no. 13, 2000.

[20] K. Zhu, C. Li, V. Asari, and D. Saupe, "No-Reference Video Quality Assessment Based on Artifact Measurement and Statistical Analysis," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 25, no. 4, 2015.

[21] Y. Hirokawa, H. Isoda, Y. Maetani, S. Arizono, K. Shimada, and K. Togashi, "MRI Artifact Reduction and Quality Improvement in the Upper Abdomen with PROPELLER and Prospective Acquisition Correction (PACE) Technique," *American Journal of Roentgenology*, vol. 191, no. 4, 2008.

[22] P. Vangorp, G. Chaurasia, P. Laffont, R. Fleming, and G. Drettakis, "Perception of Visual Artifacts in Image-Based Rendering of Façades," in *CGF*, 2011, vol. 30.

[23] K. Berger, C. Lipski, C. Linz, A. Sellent, and M. Magnor, "A Ghosting Artifact Detector for Interpolated Image Quality Assessment," in *Proc. IEEE ISCE*, 2010.

[24] M. Schwarz and M. Stamminger, "On Predicting Visual Popping in Dynamic Scenes," in *Proc. ACM Symposium on Applied Perception in Graphics and Visualization*, 2009.

[25] G. Farnebäck, "Two-Frame Motion Estimation Based on Polynomial Expansion," in *Image Analysis*. 2003.

[26] D. Cunningham and C. Wallraven, *Experimental Design: From User Studies to Psychophysics*, AK Peters, Natick, MA, 2011.