Eurographics Symposium on Rendering 2011
Ravi Ramamoorthi and Erik Reinhard
(Guest Editors)

Volume 30 (2011), Number 4

Supplemental Material: Perception of Visual Artifacts in Image-Based Rendering of Façades

P. Vangorp[1], G. Chaurasia[1], P.-Y. Laffont[1], R. W. Fleming[2], and G. Drettakis[1]

[1]REVES / INRIA Sophia Antipolis, France
[2]Justus-Liebig-Universität Gießen, Germany

Abstract

This document contains supplemental materials for the paper "Perception of Visual Artifacts in Image-Based Rendering of Façades" *[VCL*11]. It contains additional details describing the stimuli and procedure of the experiments, and a few additional results.*

1. Experimental Procedure

One author (not a computer graphics expert) participated in all experiments except Experiment 3b. All other participants were volunteers, naive to the specific purpose of the experiments but experienced in participating in visual psychophysics experiments. No other volunteers participated in both balcony conditions of Experiment 2 or in both the artificial and real parts of Experiment 3. Two volunteers participated in all experiments (but only one balcony condition of Experiment 2 and only Experiment 3a), and three participated in Experiment 1 (real stimuli) and either Experiment 2 or Experiment 3a (both with artificial stimuli). The others participated in only one experiment. The participants were ages 25-40, predominantly male, and had normal or corrected-to-normal visual acuity. A few participants had experience in computer graphics but not in IBR. All experiments were conducted on 21″ widescreen LCD monitors at the native resolution of 1680 × 1050 pixels, in typical office lighting conditions and viewing distance.

We chose to use a continuous rating task to enable participants to give a wide range of potential responses to the necessarily subjective question of visual quality. This allows participants to directly express the extent to which a given artifact bothers them. We considered, but opted against, a forced choice paradigm, because its restrictive binary nature is known to increase the "apparent consistency" of the data. Forced choice would also tend to encourage participants to base their judgments on any detectable differences, thereby exaggerating the effect of minor artifacts.

We chose to show simultaneous videos to allow direct comparison. We opted against sequential presentation because it relies on memory and would make artifacts more difficult to detect. Especially the parallax distortions of Experiment 2 might go unnoticed without a direct comparison.

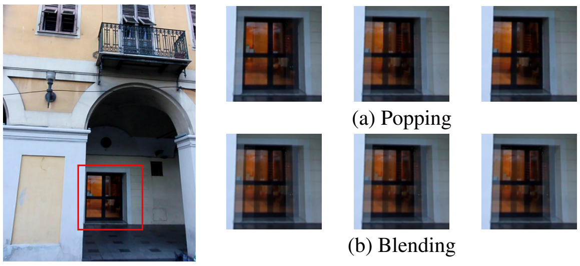
2. Statistical Tests

Visual quality levels are reported as percentages. The extent of the slider controls of the experiment interfaces will be interpreted as 0% to 100%. Differences in visual quality levels will be reported as percentage points (pp). Statistical significance will be reported with p-values for the appropriate hypothesis tests (two-tailed two-sample t-test with unequal variances for differences between groups, and F-test for non-zero linear regression slopes). Equivalence groups are based on the Tukey-Kramer multiple comparison method at significance level $\alpha = 0.05$.

3. Experiment 1: Popping and Blending

Unstructured Lumigraph Rendering. Using a Canon EOS 550D digital video camera, we captured steady video sequences (i.e., with a dolly) of a Corner of a large city square and of a Town Hall. We avoided having large depth discontinuities (e.g., from foreground objects close to the camera), since they cause very strong artifacts.

The stimuli are generated by Unstructured Lumigraph Rendering (ULR) [BBM*01] with per-pixel weights based on input and output camera parameters. For the densest set of input images we extract a regular subsampling of approximately 70 frames from the video, corresponding to 1 frame per second or a camera translation of approximately 50 cm. The sparser sets of images were obtained by using only half



(a) Popping

(b) Blending

Figure 1: *Transitional artifacts visible in 3 consecutive frames (25 fps) on the door under the arch. (a) Popping, with some visible blending in the middle frame. (b) Mixing 3 input images.*

or one quarter of the number of images in the densest set. The input images are reprojected onto the simplified geometry. For each pixel of each output frame, a predefined number of corresponding input pixels with the highest weights are mixed to produce the output pixel. We mix 1, 2 or 3 images at any given pixel. When more than the predefined number of input images have identical highest weights, the choice of "best" images becomes somewhat arbitrary. A commonly applied improvement is to allow up to 3 "best" images to be mixed to avoid the worst popping and boundary artifacts. In the stimuli where we specified to use only one input image per output pixel, this will still cause some blending to occur localized in space and time (Figure 1(a), middle frame). This contributes up to 10.5% of the final pixel values when the coverage of input cameras is high and there are consequently many input image boundaries.

Procedure. The stimulus resolution was 360 × 640 pixels for the Corner scene and 360 × 480 pixels for the Town Hall scene. Each stimulus and reference pair played in a continuous loop of approximately 16 s. Each participant did 3 repetitions of each condition (3 variations of the number of images mixed per pixel × 3 variations of the coverage) in separate blocks for the 2 scenes. Eight participants completed such a session lasting approximately 20–25 min.

Personal Preference for Fast or Slow Popping. Along with the strong scene-dependent preference for either fast or slow popping reported in the paper, there also appears to be a personal preference. This was informally observed and could be based on a trade-off between the frequency of the distracting artifact [YJ84] and the distance of the jump [dBYB*02]. This preference also affects the trade-off between popping and blending.

Equivalence Groups. The equivalence groups for the Corner and Town Hall scenes separately (Fig. 3) exhibit a very similar ranking of stimuli, except for the reversal of the popping stimuli (1/xx).

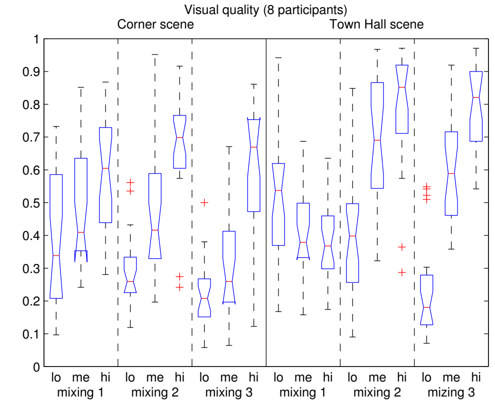
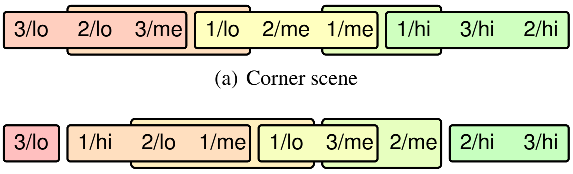


Figure 2: *Box plot of Experiment 1. The horizontal red marks are the medians, the blue box encompasses the interquartile range, the whiskers extend to the most extreme non-outliers, and the red crosses are the outliers. The notches represent comparison intervals: two medians are significantly different at the 5% level if their notches do not overlap.*



(a) Corner scene

(b) Town Hall scene

Figure 3: *Equivalence groups for the combinations of the number of images mixed at any pixel and coverage, separately for the Corner and Town Hall scenes.*

4. Experiment 2: Parallax

Procedure. Realistic artificial stimuli were created using physically based lighting [PH10] and procedural textures. All synthetic images are generated at a sufficiently high resolution to avoid aliasing or blurring artifacts. The final stimulus resolution was 556 × 417 pixels. Each stimulus played in a continuous loop of approximately 16 s. The stimuli were first presented one by one, then paired with the corresponding reference, and finally in a broader comparison between 3 pairs. Each participant did 2 repetitions of each condition (3 depth ranges × 3 viewing angles) for only 1 balcony condition. Such a session lasted approximately 25–30 min. Nine participants completed the condition without a balcony, and 5 participants completed the condition with a balcony.

Balconies. The surprising absence of any effect of the depth range could be caused by the design of the façade. We hypothesized that there could be a perceptually important dis-

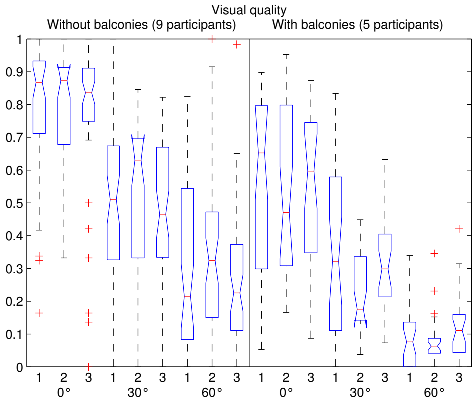
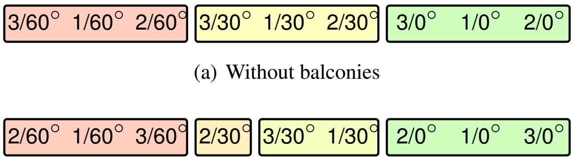


Figure 4: *Box plot of Experiment 2.*



(a) Without balconies

(b) With balconies

Figure 5: *Equivalence groups for the combinations of depth range and angle, separately for the stimuli with only alcoves and the stimuli with balconies as well.*

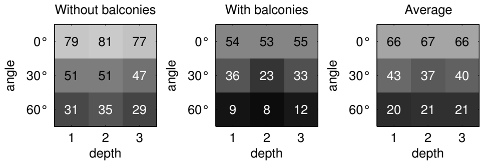


Figure 6: *Average visual quality ratings, ranging from the worst quality (0%, black) to the best (100%, white). We show the stimuli without balconies, those with balconies and the average.*

tinction between balconies and alcoves. Since we put the planar proxy at the depth of the front wall of the building, features which are further away than the proxy will pop in the direction of the camera path, while features which are closer than the proxy will pop in the opposite direction. Adding the balconies also increased the depth range by 50%. Adding the balcony decreased the average quality rating by -21.77 *pp* ($p < 0.0001$). However, it is unclear if this is due solely to the change in the stimuli, as the balcony condition also used different participants, who may have had a different average criterion when using the sliders.

There was still no significant effect of the depth parameter in either condition. However, the visual quality at depth range 2 and angle 30° in Fig. 6(b) deviates from the trend and forms a separate equivalence group in Fig. 5(b). This effect reaches statistical significance ($p < 0.01$) but it is not a monotonic trend.

Direct comparison. The stimuli were presented one at a time without direct comparison, then in pairs (a given IBR approximation vs. its corresponding reference), and then in a broader comparison of 3 pairs at once. Each time the participants were given the opportunity to adjust their previous ratings, and they did so 26.38% of the time for the pairwise comparison, and 12.90% of the time for the broader comparison. When they did adjust their rating, the average magnitude of the adjustments was 13.57 *pp*, up or down.

5. Experiment 3: Cross-fading

Procedure. For popping and blending, we use the same coverage parameters as those used for Experiment 1: each point on the proxy is covered by approximately 3, 6, or 12 images. We only retain two conditions: 1 image used at every pixel, and 2 images mixed at every pixel, since mixing 3 images did not improve the visual quality in Experiment 1.

The stimulus resolution was 556×417 pixels. Each stimulus and reference played in a repeating loop consisting of approximately 600 ms static camera at the start, 8 s translation along the path, 600 ms static camera at the end, and 600 ms blank to transition back to the start of the path.

The artificial stimuli (Experiment 3a) were first presented paired with the reference, and then in a broader comparison of one cross-fading stimulus, one popping stimulus, and one blending stimulus, with the same reference. For the real scene (Experiment 3b) no reference video was available, so the stimuli were first presented individually, and then in a comparison of one cross-fading stimulus, one popping stimulus, and one blending stimulus. The façade used for Experiment 3b is also available in Google Street View$^{\text{TM}}$. Click here to see it and compare the quality of the transitions: http://goo.gl/kpsVg

Each participant did 2 repetitions of each condition (3 cross-fading lengths × 3 variations of the coverage for the popping and blending stimuli). Ten participants (Experiment 3a) and 8 participants (Experiment 3b) completed such a session lasting approximately 20–25 min.

Equivalence Groups. The equivalence groups for the artificial stimuli (Fig. 8(a)) and the real stimuli (Fig. 8(b)) exhibit a similar ranking, except for the position of the cross-fading stimuli among the ULR stimuli. The ranking within each stimulus type (cross-fading, popping, blending) is identical.

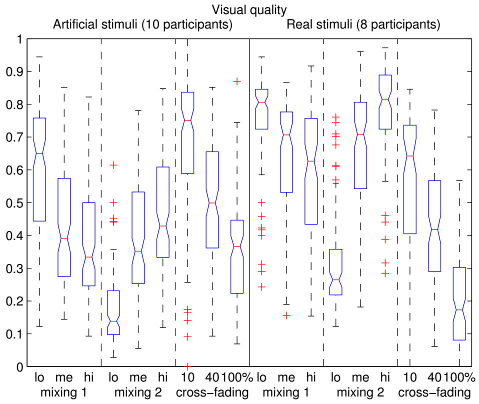
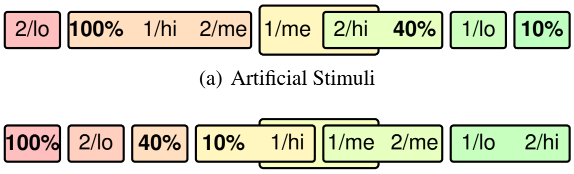


Figure 7: *Box plot of Experiment 3.*



(a) Artificial Stimuli

(b) Real Stimuli

Figure 8: *Equivalence groups for (a) the artificial stimuli of Experiment 3a, and (b) the real stimuli of Experiment 3b. The cross-fading stimuli are indicated in bold.*

References

[BBM*01] Buehler C., Bosse M., McMillan L., Gortler S., Cohen M.: Unstructured lumigraph rendering. In *Proc. ACM SIGGRAPH* (2001), pp. 425–432. 1

[dBYB*02] de Brouwer S., Yuksel D., Blohm G., Missal M., Lefèvre P.: What triggers catch-up saccades during visual tracking? *J. Neurophysiol. 87*, 3 (2002), 1646–1650. 2

[PH10] Pharr M., Humphreys G.: *Physically Based Rendering: From Theory To Implementation*, 2nd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2010. 2

[VCL*11] Vangorp P., Chaurasia G., Laffont P.-Y., Fleming R. W., Drettakis G.: Perception of visual artifacts in image-based rendering of façades. *Computer Graphics Forum 30* (2011). Proc. Eurographics Symposium on Rendering (EGSR 2011). 1

[YJ84] Yantis S., Jonides J.: Abrupt visual onsets and selective attention: Evidence from visual search. *J. Exp. Psychol.: Human Perception and Performance 10*, 5 (1984), 601–621. 2