Perception of Visual Artifacts in Image-Based Rendering of Façades

Peter Vangorp, Gaurav Chaurasia, Pierre-Yves Laffont, Roland Fleming, George Drettakis

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

Volume 30 (2011), Number 4

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

Image-based rendering (IBR) techniques allow users to create interactive 3D visualizations of scenes by taking a few snapshots. However, despite substantial progress in the field, the main barrier to better quality and more efficient IBR visualizations are several types of common, visually objectionable artifacts. These occur when scene geometry is approximate or viewpoints differ from the original shots, leading to parallax distortions, blurring, ghosting and popping errors that detract from the appearance of the scene. We argue that a better understanding of the causes and perceptual impact of these artifacts is the key to improving IBR methods. In this study we present a series of psychophysical experiments in which we systematically map out the perception of artifacts in IBR visualizations of façades as a function of the most common causes. We separate artifacts into different classes and measure how they impact visual appearance as a function of the number of images available, the geometry of the scene and the viewpoint. The results reveal a number of counter-intuitive effects in the perception of artifacts. We summarize our results in terms of practical guidelines for improving existing and future IBR techniques.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

1. Introduction

The advent of digital cameras, automated camera calibration [SSS06] and partial geometric reconstruction [FP10] makes image-based rendering (IBR) a very attractive solution to capture and render high-quality images of everyday scenes. A popular example of such an approach is Google Street View™, which uses blending between panoramas.

However, visually objectionable artifacts that occur when exploring sparsely sampled or poorly modeled scenes often limit the application of IBR methods, as do potentially huge acquisition and storage costs. These two problems are intimately connected because fewer images and simpler geometry lead to worse artifacts. A deeper understanding of the causes and relative severity of different kinds of artifacts can potentially overcome these two key barriers to allow wider and more compelling deployment of IBR.

The most common artifacts in IBR techniques include ghosting/blurring due to blending of images from different viewpoints (Fig. 1(b)), or alternatively popping artifacts if image switching is used instead. Another important issue is parallax error, i.e., error occurring because an image taken at a given viewing angle is projected onto a plane viewed from a different angle (Fig. 1(c)). This error is caused by the differences in capture and display viewing angles, and the lack of geometric depth reconstruction. Despite some initial studies [SLW*08, MO09b], little is known about how such artifacts are perceived by humans, and no systematic classification and consequent perceptual study have been performed.

In this paper we perform three psychophysical experiments to systematically map out the causes and perception of the most common IBR artifacts so that they can be avoided or minimized in typical usage scenarios. To do this, we restrict the conditions so that we can isolate each class of artifact and measure how it is affected by different parameters. Our target use-case involves simple geometry representing architectural façades, typically a few boxes or planes such as those constructed rapidly using a simple modeling tool (e.g., Google Sketchup™), and a small number of captured photographs; typically between 8–10 for a given building.

The first experiment studies blurring/ghosting, as a func-

(a) (b) (c)

Figure 1: *(a) One of the environments used in our perceptual tests, with the input cameras shown. Examples of two of the artifacts we studied, namely (b) blending and (c) parallax distortion.*

tion of number of images blended. The second experiment studies parallax artifacts as a function of viewing angle and depth range in the scene. These first two experiments require specific conditions for stimuli and setup, to correctly isolate the perceptual effects of each artifact (see Sect. 3). We thus perform a third experiment that examines the link between these conditions and the corresponding effects.

To our knowledge, our experiments are the first to perform a systematic perceptual study comparing real ground truth (i.e., video) to image-based rendering algorithms. In recent years, such formal studies yielding new perceptual insights into existing rendering methods have had a significant impact on the field [MLD*08, LCTS05].

The main contribution of our work is thus in the design and execution of the perceptual study for IBR algorithms and the results of this study. We first provide principled perceptual confirmation of "intuitive" assumptions, which are to be expected based on analysis of geometry or projection, e.g., that blending more images improves rendering quality or that oblique viewing angles degrade the result. More interestingly, our study reveals surprising results on the perception of IBR artifacts, e.g.:

- when only a small number of images are captured, it may be preferable not to use blending;
- variations in scene depth have little influence on quality when using a wide-angle *single* image rendering;
- when cross-fading between panoramic images, shorter transition durations are preferred.

It is interesting to note that the IBR method studied to obtain the second and third results is very similar to that used in transitions of Google Street View™.

In the discussion of the results of our study, we provide a number of such intuitions, or guidelines. These conclusions can be used to help decisions on the capture process, and various algorithmic choices used in image-based rendering systems. Such systems (e.g., Google Street View™, Microsoft Photosynth™ etc.) are gaining widespread popularity; better perceptual insights, such as those offered by our study, can be central in improving quality and efficacy.

2. Previous Work

Image Based Rendering (IBR) is a wide field which can be broadly defined as including any method that visualizes a real scene based on input photographs. We limit this overview to methods that produce novel viewpoints, rather than novel materials or illumination conditions. In general these techniques (implicitly) reconstruct a lower dimensional subset of the 5D plenoptic function [MB95, LH96, GGSC96].

Static Panoramic Images. Single viewpoint panoramas [Che95] can give an overview of large scenes, but by themselves do not allow novel viewpoints. Multi-viewpoint panoramas [AAC*06, RL06] can produce novel viewpoints along the input path by panning over the static image, but they result in perspective distortions especially with curved paths and lack motion parallax.

View Interpolation Methods. View interpolation methods compute a transformation between input photographs based on corresponding image features [Low04]. The transformation should represent a plausible optic flow field [MHM*09]. Novel viewpoints are constructed by warping the adjacent input images and applying a smart blending operation that avoids visible artifacts [SLW*08]. These methods generally require relatively small *baselines*, i.e., small distances between the input cameras and the novel viewpoint.

Geometry-Aware Methods. View-dependent texture mapping [DYB98] uses projective texturing to project photographs of real scenes onto a simplified geometry *proxy*. Overlapping photographs are blended based on the angle between the view directions of the novel viewpoint and the input photographs. Unstructured Lumigraph Rendering (ULR) [BBM*01] generalizes the blending framework by

(a) (b) (c)

Figure 2: *(a) Overview of the Town Hall scene. The input camera positions are represented in white, and the part of the path that was used in the stimuli is highlighted in green. A selection of input images of (b) the Town Hall scene and (c) the Corner scene (overview in Fig. 1(a)).*

introducing specific weights to take into account multiple criteria including view direction and resolution.

A major difficulty in geometry-aware IBR methods is the task of aligning the input photographs to the geometry, because even small misalignments can result in troublesome artifacts. The Façade system [DTM96] calibrates cameras and allows simple geometry creation from a set of photographs with user input. Recent advances in structure-from-motion [SSS06], multi-view stereo [FP10], and surface reconstruction [KBH06] have made the process of camera registration and geometry reconstruction almost completely automatic.

These geometry-aware methods support wide baselines and allow novel viewpoints far from the input cameras. This also means that the available image data is typically much sparser than for view interpolation methods, leading to various artifacts.

Perception of Visual Artifacts in IBR. There has been recent interest in studying perception for image-based techniques. The majority of these approaches use perceptually-inspired algorithmic measures to develop their algorithms, sometimes accompanied by a perceptual study to confirm algorithmic choices. Examples include the work in [MO09b] in which the storage space and processing time required for large amounts of overlapping image data inspired perceptual compression techniques for Unstructured Lumigraphs. Another example is the detection of ghosting artifacts in images [BLL*09]. Although not originally intended for IBR, the work of [SS09] provides a way to detect popping in image sequences using a model of spatio-velocity contrast sensitivity.

To our knowledge, the most closely related perceptual study on image-based techniques is the work on the overall visual quality of panoramic transitions [MO09a]. They concluded that the magnitude of the depth discontinuity at occlusion boundaries is a key factor in visual quality. This work was an important first step in the goal of understanding the perception of IBR artifacts. In our experiments however, we perform a systematic study of artifacts in the more general case of lumigraph-style rendering, and perform direct comparisons with ground truth (video).

Applications. Google Street View™ [Vin07] uses a very sparse set of panoramic images. Transitions between captured viewpoints employ cross-fading and geometry-aware warping to approximate the expected optic flow. Street Slide [KCSC10] uses a denser set of photographs to create a detailed representation of viewpoints perpendicular to the façade, at the expense of other kinds of visual artifacts such as distortion. Microsoft Photosynth™ [SSS06] displays an unstructured collection of photographs in the reconstructed spatial layout and applies image-space transformations and blending transitions.

The recent increased interest in these kinds of applications indicates that it is important to better understand the perception of visual artifacts.

3. IBR Artifacts and Experiments

Our goal is to systematically evaluate the perception of the most common artifacts in IBR, namely blending, popping and parallax distortions. In this section, we start by describing these artifacts, and then present an overview of the experiments performed.

Artifacts. Parallax can be described as the difference in perspective seen from different viewpoints. When a captured photograph is projected back onto an inaccurate geometric proxy, some features will be projected at the incorrect depth and cause the perspective from a different output viewpoint to appear distorted. The artifacts are accentuated by increased distance or angle between capture and output cameras.

Blending and popping artifacts appear during transition

between frames rendered through IBR. If multiple images containing misaligned features are blended at each pixel, these features will show up as clearly separate repetitions (ghosting) or as merging repetitions (blurring) in the output image. On the other hand, using a single source image at any given pixel results in popping artifacts, where image features appear to "jump" between frames.

It is worth noting that these two sets of artifacts are closely related: the main cause of transitional artifacts is the difference in parallax distortions in the images involved in the transition. One of the goals of this paper is to develop the appropriate experimental methodology to study these artifacts *separately*.

Experiments. The first experiment studies popping and blending artifacts, using Unstructured Lumigraph Rendering (ULR) [BBM*01] with a simple planar geometric proxy. This can be achieved using real video data, since we only vary the number of images used.

The second experiment focuses on parallax artifacts in isolation by examining the distortions in a single wide-angle image (equivalent to a panorama) projected onto a planar geometric proxy viewed from different angles. We use this *single* image without transitions so that no blending is required, isolating the two types of artifacts. Parallax distortions depend on the amount of depth range in the façade: if the façade is almost completely flat, the new view will be (relatively) accurate. To map out how this affects artifacts, we need to systematically vary the depth range in a controlled manner. Therefore, we cannot use real images, and instead created realistic synthetic stimuli.

The parallax experiment setup allows systematic control of the angle and depth parameters. However, only a single wide-angle image is used, in contrast to the blending/popping experiment which involves many images. We thus perform a third experiment to investigate the connection between blending/popping and wide-angle image IBR solutions, using both an artificial and a real scene. We investigate a new condition, that of cross-fading, i.e., using linear instead of ULR weights [BBM*01] for blending between wide-angle images. Cross-fading is used in popular panorama-based IBR techniques, giving effects similar to that in Google Street View™.

4. Experiment 1: Popping and Blending

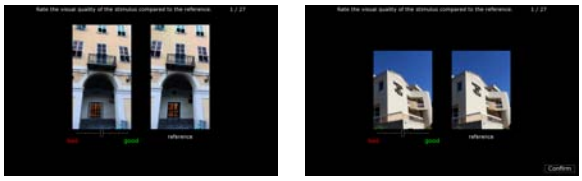
The purpose of this experiment was to measure how popping and blending artifacts affect the perceived quality of image based renderings of real façades. We ask the following questions: Under which conditions do the artifacts become objectionable? Which type of artifact is worse? What is the optimal display strategy when there are restrictions on the number of images that can be captured or stored?

The two parameters that control these artifacts for a given level of geometric reconstruction, are the *coverage* between input images, and the *number of images blended*, or *mixed*, at any given pixel. Coverage is a way to measure input image density, and thus defines the total number of images used to generate the output result. We define coverage in a canonical fronto-parallel viewing condition, as the number of images covering a given point on the planar proxy on average.

Popping causes high-salience motion transients, that draw attention to objects changing perspective, or to differences in brightness if the input images have illumination differences [YJ84]. Popping can be characterized by the frequency of the transitions and the distance that features appear to jump. A sparse set of input images causes *slow popping* with infrequent but long jumps; a dense set of input images causes *fast popping* with frequent but short jumps (see video). Because of their complementary disadvantages, it is not a priori obvious which of these should be preferable.

4.1. Stimulus Generation

We captured steady video sequences of a Corner of a large city square and of a Town Hall (Fig. 2), which allows us to make direct comparisons between image-based renderings and real video. We then extract a regular subsampling of frames from the video and use Bundler [SSS06] to calibrate the cameras and provide a sparse 3D point set. We use this 3D point set to guide the creation of a simplified version of the geometry, similar to the piecewise planar geometry obtained from simple geometric modeling tools such as Google Sketchup™. The stimuli are generated by ULR [BBM*01] with per-pixel weights. More details are available in the supplemental material.



(a) (b)

Figure 3: *Experimental interface for the visual quality rating experiment with real stimuli. (a) Corner scene. (b) Town Hall scene.*

4.2. Procedure

The parameters we vary for the approximate renderings are (1) the coverage, and (2) the number of images blended for any given pixel. For coverage, we use low (lo), medium (me) and high (hi) values corresponding to approximately 3, 6 and 12 images covering any point on the proxy. We need 18, 36 and 65 (Town Hall) or 69 (Corner) input images to achieve these values of coverage. For the number of images mixed

per pixel, we use values of 1, 2 and 3, as commonly used for this class of IBR techniques [EDDM*08, SSS09].

The participant is presented with a pair of videos: an IBR approximation and the corresponding video reference. The videos play in a loop of approximately 16 s with the camera moving forward then backward along the path. The participant is asked to "rate the visual quality of the approximation with respect to the reference" using a continuous slider (Fig. 3 and video). This provides a direct measure of quality. Each of the 3×3 stimuli is repeated 3 times in random order, in separate blocks for both scenes. All stimuli, rating results and procedure details are available as supplemental materials.

4.3. Results

In what follows, visual quality levels will be reported as percentages. The extent of the slider controls 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. We report only differences between *groups* of conditions rather than differences between or even within individual conditions to ensure the necessary statistical power.

Intuitively, we would expect a monotonic progression of quality as we increase the number of images used overall. The key question is how this is affected by popping and blending artifacts.

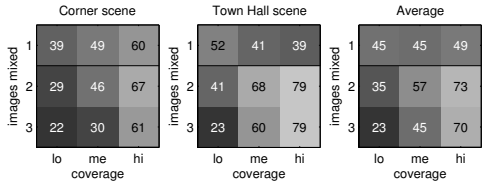


Figure 4: *Average visual quality ratings for Experiment 1, ranging from the worst quality (0%, black) to the best (100%, white). Note that higher values means the sequence looked better, i.e., fewer artifacts.*

Fig. 4 summarizes the overall visual quality ratings, averaged over all 8 participants. Details of the participants are available in supplemental material. It is standard practice in the visual psychophysics literature to use a similar number of participants (e.g., [VGB05, AW05]), as once an effect is statistically significant, adding more participants has an ever decreasing probability of changing the conclusion. The significance levels reported below also imply that the number of participants was sufficient.

Popping. The top rows of Fig. 4 refer to popping, since only a single image is being used at any given pixel. For this case, the overall visual quality appears to depend on the severity of the artifacts which varies from scene to scene. This dependence on the scene is revealed by linear regression of the quality as a function of coverage. There is a significant preference for faster popping in the Corner scene (significantly positive slope of 10.38 pp per approximate doubling of coverage, $p < 0.0005$). A more surprising outcome is the preference for *slower* popping in the Town Hall scene (significantly negative slope of -6.53 pp per approximate doubling of coverage, $p < 0.005$). This result is of interest since it means that it is not necessarily advantageous to have larger coverage, i.e., a larger number of images in total.

Blending. In contrast, for blending (Fig. 4, bottom rows), linear regression confirms our expectation that the overall visual quality improves as the coverage grows (significantly positive slope of 21.45 pp per approximate doubling of coverage, $p < 0.0001$). With a sparser set of input images, the images blended were captured further from the output camera position on average and therefore have larger distortions when projected onto the planar geometric proxy which results in feature misalignment.

We might expect that mixing more images together at every pixel improves appearance by smoothing out transitions. Interestingly, however, we find that visual quality tends to improve when *fewer* images are mixed per pixel. The average quality increase from 3 to 2 images mixed per pixel is 9.14 pp, $p < 0.005$. When geometry is not sufficiently accurate, mixing fewer images at any given pixel reduces blurring or the number and spatial extent of ghost images.

Popping vs. Blending. It is interesting to study whether there is a clear difference in quality between popping (using 1 image per pixel) or mixing 2 images per pixel. We find that the relative unpleasantness of popping and blending artifacts depends on the preference for fast or slow popping in the scene. However, in both scenes there is a crossover point between popping, which is preferred for low coverage, and mixing 2 images, which is preferred for high coverage. Figure 5 confirms this observation with the equivalence groups for all combinations of the number of images mixed and the coverage. This set of equivalence classes can be seen as a basic ranking of quality vs. number of images used (per pixel and total).

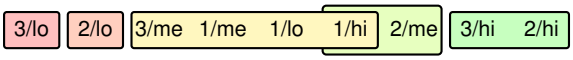


Figure 5: *Equivalence groups for the combinations of the number of images mixed at any pixel and coverage.*

5. Experiment 2: Parallax

In the second experiment, we study how parallax distortions affect appearance. An important design choice we make is to use a single wide-angle image so we can study parallax

artifacts separately from blending. When using very approximate geometry (e.g., planar proxies), and when the viewpoint is far from the input camera positions, parallax distortions can lead to substantial misperception of the depicted scene. It is known that when pictures are viewed from incorrect viewpoints, they do not appear as distorted as one might expect [Kub86, VGB05]. Thus, it is interesting to ask to what extent such distortions interfere with IBR. To study this, we parametrically map out the effects of parallax errors on perceived quality as a function of the geometrical properties of the scene and view position.

5.1. Stimulus Generation

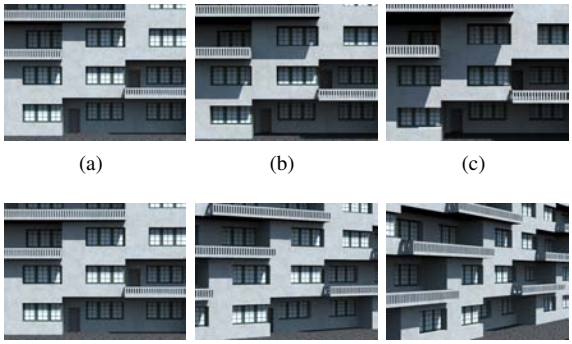


Figure 6: *Synthetic stimuli used for parallax error experiment. Linear depth variations: (a) low, (b) medium, (c) large, and angle variations: (d) 0°, (e) 30°, (f) 60°.*

We modeled an artificial façade in which we can scale the depth range, much as real façades vary from almost perfectly planar (e.g., a skyscraper), to containing large variations in depth (e.g., balconies or alcoves). The output camera is oriented at an angle and travels back and forth parallel to the façade. The camera path is chosen so that each viewing angle condition shows the same part of the façade, namely the part that is seen frontally by the wide-angle input image. Specifically, we created three different depth ranges of relative scales 1, 2, and 3, and 3 different viewing angles of 0°, 30°, and 60° from the normal of the façade. Examples are shown in Fig. 6 and the video.

To create IBR approximations for each of these scenes, they were first raytraced onto a single wide-angle image which was then mapped onto a planar proxy and visualized from the output camera. We use fully raytraced movies as the ground truth, which is equivalent to using the video for the real stimuli.

5.2. Procedure

We presented the stimuli to participants and asked them to "rate how much the artifacts bothered them" by adjusting a continuous slider. Specifically, on each trial, participants were presented with three pairs of movies that varied in depth or view angle. Each pair consisted of the IBR approximation and the corresponding reference. Each trial consisted of three steps (see Fig. 7). First the participants are shown each individual movie separately (a) and adjust the slider. Then each individual pair is shown (b), allowing slider adjustment to ensure comparison with the reference. Finally all six movies were shown simultaneously (c) to allow final cross-checking and minor adjustments (see also video). This procedure was designed to ensure maximum consistency in the use of the rating scale across stimuli.

Figure 7: *Interface for the parallax error experiment. The participant is presented first with (a) each video separately, (b) each IBR/reference pair and finally (c) an entire set of videos.*

5.3. Results

Parallax distortions get progressively larger when the planar proxy is viewed from steeper glancing angles. Thus, we expect to see a monotonic decrease in visual quality as a function of viewing angle. Likewise, perspective distortions also increase as the range of depths in the scene increases, again predicting visual quality should go down as depth range increases.

Fig. 8 summarizes the visual quality ratings, averaged over 14 participants, for the IBR approximation stimuli only.

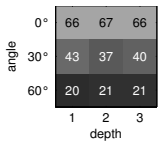


Figure 8: *Average visual quality ratings for Experiment 2, ranging from the worst quality (0%, black) to the best (100%, white).*

Angle. As can be seen in the Fig. 8, view angle has a substantial effect on visual quality. Linear regression shows a significant decrease of visual quality when the façade is

viewed from increasingly oblique angles (significantly negative slope of -23.10 *pp* per 30° increment in viewing angle, $p < 0.0001$), because the novel camera orientation deviates more from the frontal view. Thus, when the façade is viewed at a shallow angle the artifacts become highly noticeable.

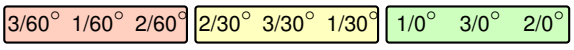


Figure 9: *Equivalence groups for depth range and angle.*

Depth. Surprisingly, the visual quality is not significantly affected by the depth range of the façade, as evidenced by the relatively homogeneous rows in Fig. 8 and the equivalence groups in Fig. 9. Overall, large variations in depth had relatively little effect on visual quality, and interacted only very weakly with the effects of glancing view angles. Thus parallax errors depend much more on the view orientation than the underlying geometry of the scene in the case of a single wide-angle image.

6. Experiment 3: Cross-fading vs. Blending Many Images

Experiment 2 maps out the conditions under which parallax artifacts become problematic when a *single* wide-angle or panoramic image is used as input, while Experiment 1 studies the case of *multiple* images with blending/popping. In this experiment, we compare transitions between two wide-angle images to the multi-image blending/popping condition of the first experiment. The parameter we study for wide-angle images is the duration of transitions. We chose to transition wide-angle images using linear cross-fading weights instead of ULR weights because the simpler approach allows direct control of the duration of the transition.

The goal of Experiment 3 is thus to address the following questions: How does cross-fading panoramas compare to ULR in terms of artifacts? Should transitions be fast (potentially too abrupt), or slow (potentially causing misalignment artifacts to be visible for longer durations)?

We performed this experiment first with artificial stimuli (Experiment 3a), which allowed precise control of the experimental conditions (depth, angle, reference etc.), remaining close to the conditions of Experiment 2. We also investigate how the results generalize to a *real* scene (Experiment 3b), even though the control of the experimental conditions is necessarily less precise.

6.1. Experiment 3a: Artificial Stimuli

6.1.1. Stimulus Generation

In the same spirit as Experiment 2, we first perform this experiment with artificial stimuli because it allows us to control the conditions. We created a variant of the artificial façade

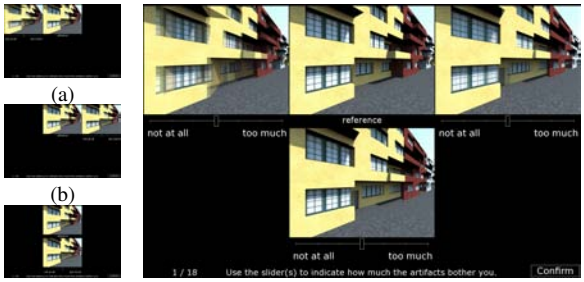


Figure 10: *Interface for the cross-fading experiment. The participant is presented first with (a–c) one of three IBR approximations and the reference, and then with (d) all three IBR approximations and the reference.*

used in Experiment 2 with depth range 1 and viewing angle 45°.

To create ULR approximations this façade was first rendered from frontal cameras, evenly spaced along the output camera path at a density equivalent to the densest set of the Corner scene of Experiment 1. The stimulus videos were rendered by mixing 1 or 2 out of all, half, or a quarter of the input cameras, thus varying coverage. We did not mix 3 images because it did not improve the visual quality in Experiment 1 (Fig. 4).

To create panoramic cross-fading approximations we rendered partial panoramas at opposite ends of the output camera path. The stimulus videos were rendered by projecting these panoramas onto the planar proxy as in ULR. The blending was done using linear interpolation weights over the full output camera path or over the middle 40% or 10% only. Before and after this blending transition only a single reprojected panorama was displayed. As before we created a raytraced reference video.

6.1.2. Procedure

We presented the stimuli to participants and asked them to "rate how much the artifacts bothered them" by adjusting a continuous slider, as in Experiment 2. Each trial consisted of two steps. Participants were first presented with the identified reference stimulus in the center of the screen, with one additional stimulus corresponding to one of blending, popping or cross-fading. These were presented in randomized order, to the left, right and below the reference (see Fig. 10 and video). Blending and popping in a given trial use the same total number of images. The participant rates each stimulus w.r.t. to the reference. After the three stimuli have been rated, the participant is presented with all three stimuli and sliders, with the reference present, and may adjust the relative ratings. As for Experiment 2, the adjustment step ensures maximum consistency.

6.1.3. Results

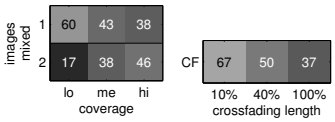


Figure 11: *Average visual quality ratings for Experiment 3a (artificial scene), ranging from the worst quality (0%, black) to the best (100%, white).*

Cross-fading. Figure 11 summarizes the visual quality for the cross-fading experiment with artificial stimuli, averaged over 10 participants. As we can see, short cross-fading is given the highest quality rating overall, while longer cross-fading received very low ratings, demonstrating a preference for shorter cross-fading (significantly negative slope of -3.20 *pp* per 10% increase in cross-fading length). Short cross-fading results in stronger parallax artifacts towards the middle of the path, but less prolonged blending artifacts during the transition. This suggests that the parallax distortions are less objectionable than the blending artifacts in these stimuli.

Popping vs. Blending. The design of the experiment allows us to revisit the question of whether popping or blending artifacts are preferable. In contrast to the Corner scene, slow popping is preferred (significantly negative slope of -11.23 *pp* per doubling of coverage, $p < 0.0001$). The trend that blending improves with higher coverage (Experiment 1) is also confirmed.

6.2. Experiment 3b: Real Stimuli

6.2.1. Stimulus Generation and Procedure

We also conduct essentially the same experiment with real stimuli to confirm that the conclusions generalize to real scenes. Frontal input photographs, evenly spaced along a city street, were used for the ULR approximations. Partial photographic panoramas were captured at both ends of the path. Due to obstacles in the rather narrow street, it was impossible to create a smooth reference video, so none was presented in the experiment interface. All other details remained the same as in Experiment 3a.

6.2.2. Results

Figure 12 summarizes the visual quality for the cross-fading experiment with the real scene, averaged over 8 participants. This confirms the trends within each technique. Most importantly there is again a clear preference for shorter cross-fading (significantly negative slope of -4.13 *pp* per 10% increase in cross-fading length). There is a slight preference for slow popping (significantly negative slope of -8.36 *pp* per approximate doubling of the coverage) and for blending with denser coverage.

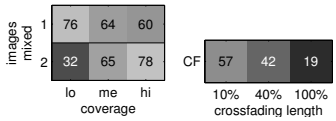


Figure 12: *Average visual quality ratings for Experiment 3b (real scene), ranging from the worst quality (0%, black) to the best (100%, white).*

Cross-fading vs. Popping/Blending. From Fig. 13 we can see that the cross-fading stimuli are comparable to the worst examples of popping and blending. The best cross-fading stimulus, short cross-fading, has a much lower quality rating than the best ULR stimuli (19.45 *pp* lower than 2/hi and 1/lo, $p < 0.0001$) and even a marginally lower quality rating than the medium coverage ULR stimuli (6.66 *pp* lower than xx/me, $p < 0.05$).

Experiment 3a showed a higher relative quality of cross-fading compared to ULR (the equivalence groups for Experiment 3a are provided in supplemental material). We believe this difference is most likely caused by the lack of detail and complexity in the artificial façade and by the high accuracy of its geometric proxy and camera positions. In real scenes the misalignments between the two panoramic images are typically larger and more noticeable than between consecutive ULR input images. Therefore we hypothesize that Experiment 3b is more suitable as a basis for guidelines which will generalize to other real world scenes.

Figure 13: *Equivalence groups for the stimuli in Experiment 3b. The cross-fading stimuli are indicated in bold.*

7. Discussion and Practical Guidelines

We now discuss our results, organized by artifact studied. We include both a perception-oriented discussion followed by practical guidelines for IBR which result from our study.

7.1. Blending and Popping

Our results (Fig. 4 and 5) show a systematic ranking of popping and blending. Clearly, the best overall result is achieved when coverage is high. While this is to be expected, we consider it important to provide a systematic evaluation of this hypothesis. When coverage is low (i.e., the xx/lo case) popping is clearly superior to blending. This result was unclear before performing the experiment. We suggest that it occurs because popping reduces the temporal extent of transitional artifacts, presenting a plausible image for longer.

One interesting observation from Fig. 4 is that the preferred popping speed appeared to be scene-dependent. This

depends on geometry, coverage and camera velocity. Scene features may also be important; e.g., in the Town Hall stimuli, popping was mainly visible on the corner balcony which only covered a small number of pixels. Fast popping was thus akin to a "flashing" stimulus, known to attract attention [YJ84].

In informal debriefing sessions following the experiments it became clear that any artifacts that caused fragmentation or doubling of features in the scene were considered worse than sudden transitions that kept the scene structure intact.

Just as the severity of popping is content dependent, so too blurring and ghosting also vary according to the features that they affect. When blurring and ghosting makes salient text illegible, or disturbs key features like edges of archways, doors or windows, it is considered highly undesirable. By contrast, blurring in the middle of a wall is often barely noticeable. This suggests that future methods could benefit considerably from content-aware transition strategies.

Guidelines. Clearly, when storage and acquisition are not an issue, two images out of a dense set should be mixed. However, storage is often limited, and thus popping is probably the best option when only a small number of images are available. Mixing more than two images at a given pixel reduces quality; rendering algorithms should thus either pop (one image) or mix two images.

7.2. Parallax

Experiment 2 indicates that as the angle of view becomes more oblique, parallax errors are more perceptible. Again, this is an intuitive result, but our study provides a systematic demonstration.

In contrast, we were surprised by the fact that, for the case of single wide-angle IBR, depth differences do not appear to be important. However, when multiple wide-angle images are used as in Experiment 3, depth becomes an important factor because the parallax distortions cause transitional artifacts.

From empirical observation of the participants and our various pilot studies, it seems that parallax artifacts were harder to spot for participants. Some indirect experimental evidence of this is discussed in Sect. 7.3.

It is worth noting that because participants were presented with the corresponding ground truth and IBR stimuli simultaneously, they could directly compare the errors in the approximation to the appearance of the ground truth, allowing them to detect subtle errors, which they may otherwise not have noticed. Because of this, our method tends to set an upper limit on the detectability of parallax artifacts—in other words if subjects tend not to notice errors in this experiment, they are unlikely to notice them in other conditions. Parallax errors may cause subjects to misperceive the shape of features in the scene. However, when there is no ground truth to compare against, subjects may be *unaware* that they are misperceiving the scene, and thus do not find the errors disturbing.

Guidelines. The dependency of quality on angle should be taken into account when capturing input photographs. Clearly, the angle depends on the expected *output* (viewing) camera. It is thus best to avoid novel camera positions which result in oblique viewing angles with respect to the captured images. The result on depth is useful, since it means that depth differences do not affect the quality of the results, and can thus be ignored in capture and display for the single, wide-angle image case.

7.3. Cross-fading vs. Blending Many Images

Both Experiments 3a and 3b showed that for cross-fading, a short transition was preferred. In the short cross-fade condition, parallax artifacts become quite acute towards the middle of the path; despite this, the condition is ranked as highest quality among cross-fading stimuli. This indirectly indicates that parallax artifacts are quite tolerable, as suggested by Experiment 2.

Experiment 3b used real stimuli and is therefore appropriate as a basis for guidelines that generalize to other real scenes. In particular, Experiment 3b showed that slow popping or dense coverage blending performed better than cross-fading (Fig. 13). Clearly long transitions or blending with sparse coverage should both be avoided.

Guidelines. Our experiment indicates an interesting way to improve image-based navigation applications based on panoramas, such as Google Street View™ which currently appears to use a technique akin to long cross-fading. By switching to shorter cross-fading perceived quality would be enhanced, despite parallax artifacts. The slightly more complex rendering technique of ULR is able to produce better results, and in addition taking a simple picture every few steps is simpler for the casual user compared to creating accurate, ghost-free panoramas which require a tripod and post-processing.

8. Conclusions and Future Work

We presented an extensive and principled study of perceptual artifacts for the domain of image-based rendering. This is a vast topic, with a very large number of interdependent parameters. Our goal was to present an initial methodology of systematically investigating these artifacts, and provide first results and guidelines to deal with them.

To enable such a systematic study, we had to restrict the set of conditions that we examined; we believe that our work opens up a number of interesting avenues for future research. One restriction we imposed was piecewise planar reconstruction of proxies. Studying the effect of progressively improved geometry is an entire topic on its own. Our two first

experiments separated out blending and parallax artifacts, while the third experiment starts investigating the combined case of the relative importance of blending vs. parallax.

There are also many additional questions that merit further investigation: our study permitted to identify these as relevant. In particular, the scene-dependency of popping speed is worthy of further investigation. Similarly, the question of the influence of depth variation in the presence of blending merits an in-depth study.

Acknowledgments

The authors would like to thank the reviewers for their helpful comments and suggestions, and all the volunteers for participating in the experiments. The authors acknowledge the support of the INRIA ARC NIEVE project, NVIDIA (Professor Partnership Program), Adobe Systems (research gift) and Autodesk (Maya donation).

References

[AAC*06] AGARWALA A., AGRAWALA M., COHEN M., SALESIN D., SZELISKI R.: Photographing long scenes with multi-viewpoint panoramas. *ACM Trans. Graph. 25* (2006), 853–861. 2

[AW05] ANDERSON B. L., WINAWER J.: Image segmentation and lightness perception. *Nature 434* (2005), 79–83. 5

[BBM*01] BUEHLER C., BOSSE M., MCMILLAN L., GORTLER S., COHEN M.: Unstructured lumigraph rendering. In *Proc. ACM SIGGRAPH* (2001), pp. 425–432. 2, 4

[BLL*09] BERGER K., LIPSKI C., LINZ C., SELLENT A., MAGNOR M.: A ghosting artifact detector for interpolated image quality assessment. In *Proc. APGV* (2009), pp. 128–128. 3

[Che95] CHEN S. E.: Quicktime VR: an image-based approach to virtual environment navigation. In *Proc. ACM SIGGRAPH* (1995), pp. 29–38. 2

[DTM96] DEBEVEC P. E., TAYLOR C. J., MALIK J.: Modeling and rendering architecture from photographs: a hybrid geometry-and image-based approach. In *Proc. ACM SIGGRAPH* (1996), pp. 11–20. 3

[DYB98] DEBEVEC P., YU Y., BOSHOKOV G.: *Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping*. Tech. rep., Berkeley, CA, USA, 1998. 2

[EDDM*08] EISEMANN M., DE DECKER B., MAGNOR M., BEKAERT P., DE AGUIAR E., AHMED N., THEOBALT C., SELLENT A.: Floating textures. *Comput. Graph. Forum 27* (2008), 409–418. 5

[FP10] FURUKAWA Y., PONCE J.: Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell. 32* (2010), 1362–1376. 1, 3

[GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The lumigraph. In *Proc. ACM SIGGRAPH* (1996), pp. 43–54. 2

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. In *Proc. SGP* (2006), Eurographics Association, pp. 61–70. 3

[KCSC10] KOPF J., CHEN B., SZELISKI R., COHEN M.: Street slide: browsing street level imagery. *ACM Trans. Graph. 29* (2010), 96:1–96:8. 3

[Kub86] KUBOVY M.: *The psychology of perspective and renaissance art*. Cambridge University Press, 1986. 6

[LCTS05] LEDDA P., CHALMERS A., TROSCIANKO T., SEETZEN H.: Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. Graph. 24* (2005), 640–648. 2

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proc. ACM SIGGRAPH* (1996), pp. 31–42. 2

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision 60* (2004), 91–110. 2

[MB95] MCMILLAN L., BISHOP G.: Plenoptic modeling: an image-based rendering system. In *Proc. ACM SIGGRAPH* (1995), pp. 39–46. 2

[MHM*09] MAHAJAN D., HUANG F.-C., MATUSIK W., RAMAMOORTHI R., BELHUMEUR P.: Moving gradients: a path-based method for plausible image interpolation. *ACM Trans. Graph. 28* (2009), 42:1–42:11. 2

[MLD*08] MCDONNELL R., LARKIN M., DOBBYN S., COLLINS S., O'SULLIVAN C.: Clone attack! Perception of crowd variety. *ACM Trans. Graph. 27* (2008), 26:1–26:8. 2

[MO09a] MORVAN Y., O'SULLIVAN C.: Handling occluders in transitions from panoramic images: A perceptual study. *ACM Trans. Appl. Percept. 6* (2009), 25:1–25:15. 3

[MO09b] MORVAN Y., O'SULLIVAN C.: A perceptual approach to trimming and tuning unstructured lumigraphs. *ACM Trans. Appl. Percept. 5* (2009), 19:1–19:24. 1, 3

[RL06] ROMÁN A., LENSCH H. P. A.: Automatic multiperspective images. In *Proc. EGSR* (2006), Eurographics Association, pp. 83–92. 2

[SLW*08] STICH T., LINZ C., WALLRAVEN C., CUNNINGHAM D., MAGNOR M.: Perception-motivated interpolation of image sequences. In *Proc. APGV* (2008), ACM, pp. 97–106. 1, 2

[SS09] SCHWARZ M., STAMMINGER M.: On predicting visual popping in dynamic scenes. In *Proc. APGV* (2009), ACM, pp. 93–100. 3

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph. 25* (2006), 835–846. 1, 3, 4

[SSS09] SINHA S. N., STEEDLY D., SZELISKI R.: Piecewise planar stereo for image-based rendering. In *Proc. ICCV* (2009), IEEE, pp. 1881–1888. 5

[VGB05] VISHWANATH D., GIRSHICK A. R., BANKS M. S.: Why pictures look right when viewed from the wrong place. *Nature Neuroscience 8* (2005), 1401–1410. 5, 6

[Vin07] VINCENT L.: Taking online maps down to street level. *Computer 40* (2007), 118–120. 3

[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. 4, 9