Handling Occluders in Transitions from Panoramic Images: A Perceptual Study

YANN MORVAN and CAROL O'SULLIVAN
Trinity College Dublin

Panoramic images are very effective at conveying a visual sense of presence at very low cost and great ease of authoring. They are, however, limited in the navigation options they offer, unlike 3D representations. It is therefore desirable to provide pleasing transitions from one panorama to another, or from a panorama to a 3D model. We focus on motions where the viewers move toward an area of interest, and on the problem of dealing with occluders in their path. We discuss existing transition approaches, with emphasis on the additional information they require and on the constraints they place on the authoring process. We propose a compromise approach based on faking the parallax effect with occluder mattes. We conduct a user study to determine whether additional information does in fact increase the visual appeal of transitions. We observe that the creation of occluder mattes alone is only justified if the fake parallax effect can be synchronized with the camera motion (but not necessarily consistent with it), and if viewpoint discrepancies at occlusion boundaries are small. The faster the transition, the less perceptual value there is in creating mattes. Information on view alignment is always useful, as a dissolve effect is always preferred to fading to black and back.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Picture/Image Generation; I.3.6 [**Computer Graphics**]: Methodology and Techniques; I.3.m [**Computer Graphics**]: Miscellaneous

General Terms: Experimentation, Human Factors, Algorithms

Additional Key Words and Phrases: User study, panorama, transitioning, content mixing, occlusion

ACM Reference Format:

Morvan, Y. and O'Sullivan, C. 2009. Handling occluders in transitions from panoramic images: A perceptual study. ACM Trans. Appl. Percept. 6, 4, Article 25 (September 2009), 15 pages.
DOI = 10.1145/1609967.1609972 http://doi.acm.org/10.1145/1609967.1609972

25

1. INTRODUCTION

As evidenced by the popularity of services that deliver interactive street view panoramas (e.g., Vincent [2007]), there is a worldwide demand for content that lets users experience remote locations visually on their computer screen.

Recent combined advances in remote sensing and photogrammetry [Snavely et al. 2007; Pollefeys et al. 2008; Frueh et al. 2004] and procedural reconstruction [Müller et al. 2007] have made it possible to recover highly detailed and realistic models of urban environments. However, because of the sheer

Authors' addresses: Y. Morvan (contact author), C. O'Sullivan, School of Computer Science and Statistics, Trinity College Dublin, College Green, Dublin 2, Ireland; email: ymorvan@cs.tcd.ie.

DOI 10.1145/1609967.1609972 http://doi.acm.org/10.1145/1609967.1609972

size of the datasets and the effort needed to realistically capture the appearance of real-world objects, compromises in the representations used have to be made.

In this respect, panoramic images probably possess the best visual value per byte of storage and ease of capture. However, they are tied to a fixed vantage point and this introduces limitations in the navigability of the captured environment. 3D models do not suffer these limitations, but their cost may limit their use to only portions of the environment, such as individual buildings.

Thus, two closely related problems arise: providing appealing transitions between panoramic images taken from distant vantage points; and providing appealing transitions between a panorama and a 3D model.

We focus on a central issue in transitioning, namely how to handle occluders such as trees, cars, and pedestrians when the viewer is moving toward an area of interest. Lateral translations are not considered.

We explore which kind of transitions are feasible without losing the advantages of the panorama (size and simplicity), and identify the requirements that they entail (Section 3). A compromise approach based on faking the parallax effect is proposed (Section 4). We then describe a user study we conducted to determine if these added requirements are justifiable in terms of visual added value (Section 5).

2. BACKGROUND

Research in the field of panoramic imaging has predominantly focused on enriching this mode of representation. McMillan and Bishop [1995] led the way in computing convincing transitions between panoramas taken from different viewpoints. All image-based approaches of this kind rely on the estimation of a dense set of correspondences or depth measurements. Moreover, to avoid warping artifacts, distances between vantage points have to be kept low.

The *photosynth* system [Snavely et al. 2008] conveys aesthetically pleasing, if not always realistic, transitions using a smaller number of correspondences. A large variety of panoramic navigation services exist on the Internet. The most ambitious ones, such as *everyscape*, use prerecorded flash videos to convey transitions. They are rendered using common optical-flow/image warping techniques.

A review of the field of panoramic photography can be found in Gledhill et al. [2003]. State-of-the-art techniques are implemented in the *panotools* free software suite. Techniques have been proposed to integrate 3D models and panoramas [Chiang et al. 1997], where the panorama provides the background for the 3D object model. Research topics that have potential applications in transitioning from panoramas to 3D models include image in-painting [Bertalmio et al. 2000] and automatic matte extraction (such as depth from defocus [Watanabe and Nayar 1998]).

There is much literature on the perception of motion from optical flow [Vaina et al. 2004], but in a sense this topic is the reverse of our problem, which is how to generate a convincing and visually appealing stimulus from sparse data to convey a known motion. Recent work by Stich et al. [2008] takes a perceptual approach to image interpolation. However, our problem is different in that occluders are absent by nature in the target view.

3. TRANSITION METHODS AND THEIR REQUIREMENTS

There are a number of options available to transition smoothly from a panorama to another panorama, or from a panorama to a 3D model of a particular area of interest. These options require varying amounts of information about the scene, and deal more or less well with occluders. The motivation behind this work is to determine whether a viewer's experience improves with more advanced options that require more information. If it does not, time and effort could be saved during the authoring process and upon

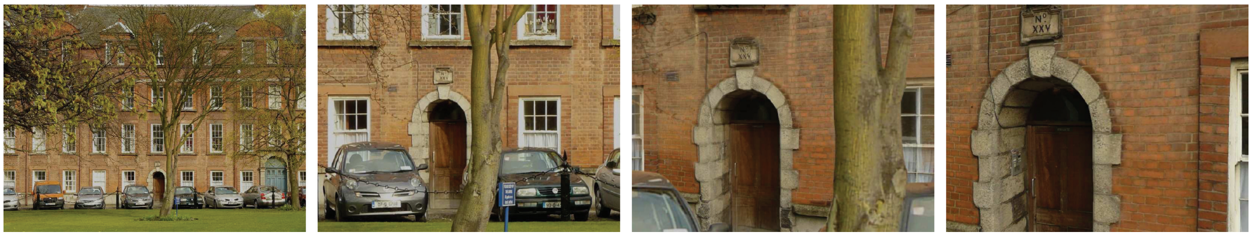

Fig. 1. From left to right: zoomed-out view of the panorama; zoomed-in view of the panorama before transition; frame of a fake parallax transition (see Section 4); view of the 3D model after transition.

Table I. Information Requirements per Transition Type

	Transition Type					
				Parallax		
Required Information	Cut	Fade	Blend	Fake	TIP	Full
Tags	Yes	Yes	Yes	Yes	Yes	Yes
Registration	No	No	Yes	Yes	Yes	Yes
Occluder Matte(s)	No	No	No	Yes	Yes	More
Coarse geometry	No	No	No	No	Yes	More
Extra Background Colour Layer	No	No	No	No	Yes	More
Layered Depth and Colour	No	No	No	No	No	Yes

rendering. Figure 1 shows a sequence of frames displayed to a user interacting with a panoramic picture and moving toward a 3D model.

The various options are discussed shortly, and their information requirements are summarized in Table I. All options make the assumption that the transition occurs from time $t = 0$ to time $t = 1$ during a closing-up on the scene of interest: At time $t = 0$, the panorama is displayed; at time $t = 1$ the target panorama or 3D model is displayed.

Fade and cut. These transitions modes necessitate the least information about the scene. The only requirement is to identify (i.e., tag) which subarea of the panorama corresponds to the area of interest one wishes to move toward. This makes it possible to adjust the panorama view so that it encompasses that subarea, then fade through black, or cut, from the initial panorama and into a view of the target panorama or 3D model. In this case, occluders present in the panorama are dealt with implicitly: One assumes that the new view has left the occluders behind.

Blend. This type of transition consists of progressively blending the pixel values of the initial panorama view and of the target panorama or model view. When the two views are not properly aligned at each frame, the effect achieved is that of an awkward fade. In order to align the views, a number of feature correspondences (i.e., registration) are necessary. This is sufficient because the projection parameters of both panoramas (2D to 2D case), or of the panorama and the virtual camera used to render the model (2D to 3D case), are known. However, the alignment will rarely be perfect because of perspective and disocclusion effects, the magnitude of which is a function of the area of interest's depth range.

Such effects are expected in the 2D to 2D case because of the vantage point discrepancy between the two panoramas. They also occur in the 2D to 3D case because when the transition starts, the panorama view corresponding to the subarea will typically be taken from a far distance with a long focal length, while the virtual camera will be placed closer to the model. This is because the transition cannot start

until the 3D model, whose extent is limited, covers enough of the view to avoid seams where it stops and the panorama begins.

The choice of feature correspondences used will determine which elements of the area of interest stay within focus during the transition. The effect of *blending* on occluders is that they will appear to dissolve into the area of interest.

All subsequent transition modes discussed assume that blending is performed outside of the regions affected by the occluders.

Parallax.

—*Full Parallax*. The correct treatment of occluders as viewers approach the area of interest is to update the angle from which they are seen until they leave the viewer's field of view. In order to reproduce this parallax effect faithfully, it is necessary to know the depth of each pixel belonging to an occluder. Moreover, since the parallax effect will result in disocclusion between objects located at different depths, color and depth information for elements that can be revealed is needed. This type of representation was proposed by Zhu and Hanson [2001] in a more general context.

—*Tour into the Picture* (TIP). The main function of modeling the parallax effect is to gracefully slide the occluders out of view during a transition. Horry et al.'s [1997] technique provides just such a feature. It uses a coarse geometric approximation of the scene, mattes for the occluders, and additional background color information in the areas that may be revealed. The geometry can be recovered with user input through single-view metrology [Criminisi et al. 2000]. Triangulation of registered features can also be used if multiple views were captured when the panorama was authored. Multiple views would also ease the task of recovering a background; in their absence, labor-intensive photo editing to fill the holes left by occluders is necessary.

—*Fake Parallax*. We propose to cut the requirement of a background color layer, as well as the need for a geometric approximation, by faking the parallax effect with simple scrolling and scaling. Since we now control when the movement of the occluders will occur, we can delay starting until a time when the visual artifacts it would cause are out of view. We describe the procedure in the following section.

4. FAKE PARALLAX

The main idea behind our approach is to use the content of the target panorama or 3D model to fill in the void left by occluders as they move out of view.

There are two kinds of artifacts to address when mimicking the parallax effect with mattes, but without knowledge of depths and of occluded colors in the initial panorama. First, occluding objects are attached most of the time to other parts of the scene, typically the ground. Moving their matte in a way that is not consistent with the location of the attachment results in a sliding effect that is very distracting. To avoid this, we only start moving the occluder mattes when none of them has a visible attachment anymore. We achieve this by tagging the edges that contain an attachment for each occluder matte. Movement only starts when all those edges have moved out of view.

Second, the content of the initial panorama at an occlusion boundary may not match the content of the target panorama or 3D model under the occlusion. This may be because the 3D model does not extend far enough (whereas panoramas always do) that is, a boundary of the 3D model is located under an occluder, or because of the viewpoint discrepancy, or both. In the first case, there is no target content available for us to fill voids. We therefore delay occluder movement until this is not the case. This case also explains why more information has to be gathered for tour into the picture or full parallax even if a 3D model of the destination is available: Because the 3D model does not cover the whole initial view, it does not provide necessary information to convey the parallax effect outside of its extent. The second

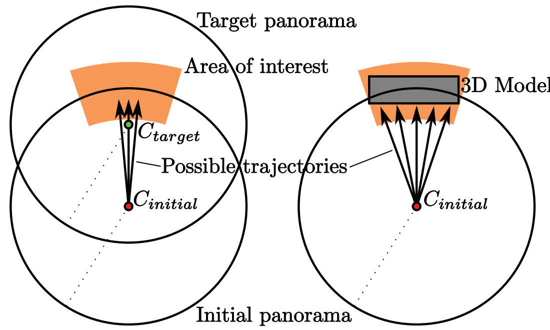

Fig. 2. Transition trajectories most likely to avoid large perspective and disocclusion discrepancies. Left: 2D to 2D case, right: 2D to 3D case.

case is unavoidable but can be alleviated. When transitioning toward another panorama, viewpoint discrepancies increase as one looks away from the viewing direction generated by the line joining the two panorama centers. Figure 2 illustrates the transition trajectories that avoid large discrepancies.

In order to attenuate the hard seam generated by the mismatch at the occlusion boundary, we progressively blend the initial view into the target view in a feathered region around the occluder matte before we start moving the occluder. In our implementation, this preliminary blending lasts 10% of the transition duration, and the feathering radius is of 20 pixels.

We determine the translation and scaling for each occluder as follows: Translation occurs along the vector formed by the center of the viewpoint and the centroid of the occluder matte. If those points coincide, either the right or left direction is picked at random. Translation and scaling increase linearly with the camera position along the transition segment. Their coefficients are chosen so that no pixel of the matte that is nontransparent remains visible at the end of the transition.

5. USER STUDY

Both the full parallax and the tour into the picture approaches require information that is much harder to acquire than taking a panoramic picture, as they involve measuring/estimating depth and sampling colors from several viewpoints. They also add significant rendering complexity. Thus, the transition types that we investigate are *fading*, *blending*, and *fake parallax*. We left *cutting* out because it is not well suited to continuous navigation.

5.1 Overview

We first carried out a feasibility study to compare the general efficacy and user appeal of the transition types under investigation. Users were invited to browse a panorama, transition to a 3D model of a building within it, and then navigate through the model. Our hypotheses that the fake parallax transition would be well received and that it would compare well to fading and blending were confirmed. We also observed that participants judged blending to be superior to fading through black.

We then ran an experiment to evaluate how different transition types compare in terms of appeal to viewers as a function of two factors: the transition speed; and whether the target is another panorama or a 3D model. Four different scenes were presented. We expected this experiment to confirm the preference for fake parallax over blending observed in the feasibility study but this was not the case. The inferiority of fading through black was confirmed, implying that registration is worthwhile. There was no conclusive evidence of an effect of transition speed or target type.

A follow-up experiment was then carried out to test how little the difference in preference between fake parallax and blending is. The number of participants was increased to 50 following a power analysis. We found that transition speed had an effect, with marginal preference for fake parallax at slow speed, and no preference at fast speed.

To investigate the cause of the difference between this result and that of the feasibility study, we carried out a final experiment. While a pause had been added between the panning and zooming and the transition proper in the feasibility study, this had not been the case in the first and second experiments. We replicated this pause in this final experiment and found a preference for fake parallax as hypothesized. We conclude that adding a pause before sliding occluders helps suspend disbelief in their fake motion.

Throughout the experiments we observed that fake parallax compares less favorably to blending the higher the depth complexity of the target region.

5.2 Feasibility Study

This preliminary investigation involved the implementation of the three transition types within a panorama navigation application, and their evaluation by a number of volunteers. We decided to focus on the case of transitions toward a 3D model, as they exhibit view discrepancy effects similar to transitions toward a different panorama, but also contain additional view-dependent parallax and disocclusion effects due to the geometry of the model.

5.2.1 *Content Creation.*

Panorama. Mature panorama authoring tools have been available for some time and are now popular. We used the free software *panotools* suite. For the purpose of implementing the fake parallax transition, we manually created a matte of the cars and tree which occlude a region of the building facade in our scene, shown in Figure 1.

Model. Some of the transitioning approaches that we investigate rely on a proper alignment of views of the 3D model with corresponding panorama views. It is therefore crucial that the model faithfully reproduces the real-world geometry that appears in the photographs used to create the panorama. To achieve this, the building was digitized using a laser scanner (resolution 5mm, accuracy $<$ 2mm). Poisson surface reconstruction [Kazhdan et al. 2006] was used to obtain a mesh which was manually edited to sharpen right-angle features. The model was textured by projecting photographs taken with a calibrated camera. Figure 3 shows a view of the model geometry and a texture photograph. The end result can be seen in the rightmost picture of Figure 1.

Registration. For convenience, we manually performed the registration of the model with the panorama by rendering the model as an overlay and adjusting its pose until its features were aligned to their counterpart in the panorama. In a production environment, registration would be automated, using either surveying equipment at the time of capture or a feature matching algorithm [Lowe 1999] followed by the solving of a linear system. Registration would also be needed for transitions using tour into the picture or full parallax, in addition to the requirements listed in Section 3.

5.2.2 *Procedure.* The start and end points of the transitions were manually set and identical between transition types. Alignment of the virtual camera and panorama views at each frame of the transition was achieved as follows: The optical axis of the virtual camera was chosen to be collinear with that of the panorama view; the vertical was shared; the field of view of the virtual camera was set to 50^0; and its dolly distance from the center of the panorama was manually set at both ends of the transition and interpolated linearly. Each transition lasted 800ms.

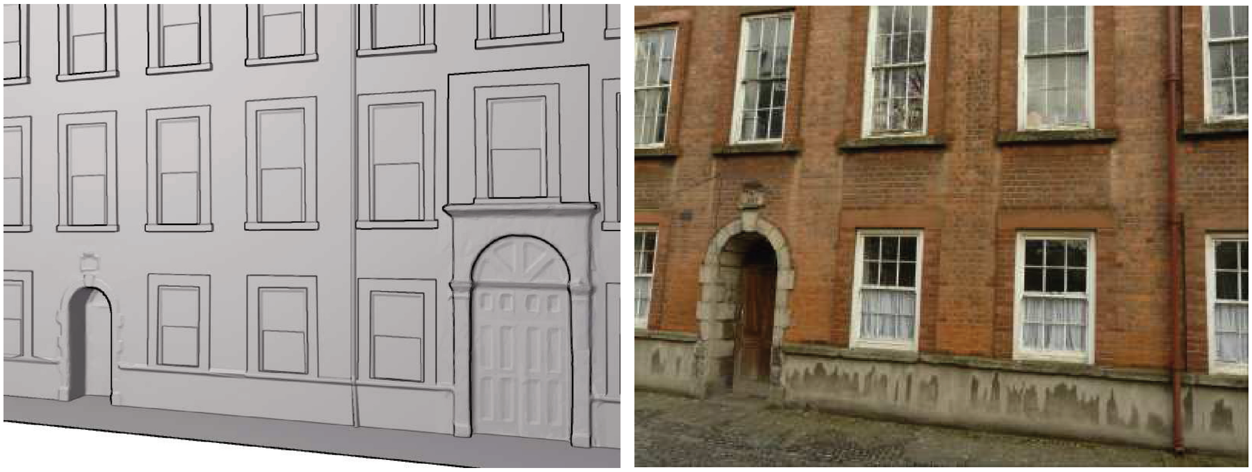

Fig. 3. 3D model geometry and a texture photograph.

Participants were told that they would be repeatedly requested to locate a red star on the brick building; that transitions would be different from one presentation to the next; and that they would have to state a transition preference between pairs of presentations. The choice was worded as follows: "Which transition did you find most visually appealing?" At the end of the session, they were asked to give an informal appraisal of their experience. The mode of interaction was the following for each presentation.

(1) The participant is shown the panorama and can interact with it in the usual way (pan, tilt, and zoom).
(2) When the space bar is pressed, the panorama view smoothly pans and zooms to the transition starting position and pauses.
(3) The transition is shown.
(4) The participant navigates around the model to locate the red star and presses the space bar when it is found, at which point the panorama is shown again for a new presentation.

We used the sigmoid function in Eq. (1) to ease both the initial pan and zoom motion and the transition itself. The pause between the two phases is a consequence of the quasi-null slope of this easing function at $t = 0$ and $t = 1$.

$$f(t) = \frac{1}{1 + e^{-9(t-0.5)}} \text{ linearly scaled so that } [0, 1] \text{ maps to } [0, 1] \quad (1)$$

There were three transition pairs: fading versus blending, blending versus fake parallax, and fake parallax versus fading. Each pair of transitions was presented 4 times. There were therefore 24 presentations in total. The total duration of each individual session was between 8 and 12 minutes, depending on participant behavior.

5.2.3 *Results.* A total of 18 naive volunteers (7 female, 11 male) took part in the feasibility study. All had normal or corrected-to-normal vision. For each transition pair, we collapse the four choices of each participant into a categorical variable with three values corresponding to whether the first or second transition was preferred the majority of the time, or whether preferences were evenly split. Such a setup gives balanced expected frequencies of $\frac{5}{16} \times 18$, $\frac{5}{16} \times 18$, and $\frac{6}{16} \times 18$, respectively.

Fig. 4. Scenes presented during the experiment, from top left to bottom right: simple, tree, columns, and glass. The red frames indicate the extent of the view of the target photograph/video at the end of the transition.

Results are summarized in Figure 5. They indicate very clearly that participants were not indifferent to the transition type shown: $\chi^2(2, N = 18)$ is respectively 37.97 for fading versus blending, 44.8 for fading versus fake parallax, and 21.17 for blending versus fake parallax, all with $p < 0.001$. Binomial tests show a very clear-cut preference for the fake parallax transition over both fading and blending:

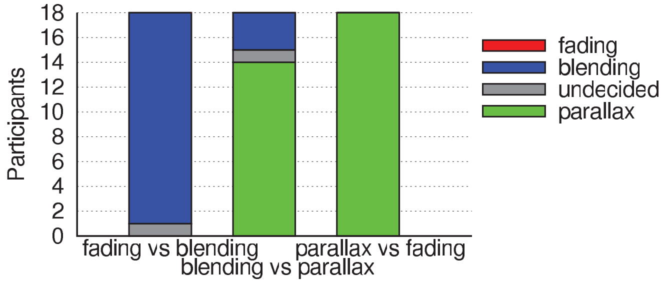

Fig. 5. Number of participants who preferred each transition type over each other type.

18 out of 18, $p \simeq 0$ and 14 out of 17, $p = 0.006$, respectively. Blending is in turn strongly preferred over fading: 17 out of 17, $p \simeq 0$.

Some participants commented that the time between two transitions due to interaction with the panorama and the model made it awkward to recall the first transition for comparison. One participant reported that the pause during the closing-up motion felt contrived.

5.3 Main Experiment

The aim of the main experiment was to consolidate the results obtained in the preliminary feasibility study. Of interest are the effect of the nature of the *scene*, and of the transition *speed*, and whether we are transitioning toward another panorama or switching to a model.

5.3.1 *Content.* Four scenes exhibiting a representative set of properties were chosen, and are shown in Figure 4. The *simple* scene is the simplest and contains the least view-dependent features. The *tree* scene contains vegetation and a more complex façade: It has high depth variation, but in a low range. The *columns* scene's depth range is higher, with columns protruding in front of the building. A glass surface appears in the *glass* scene, where a reflection of the environment interacts with a view of the interior.

Because we were not interested in user interaction before or after each transition, each panorama, both initial and target in the 2D to 2D case, was represented by a single photograph. Furthermore, each stimulus presented to the participants could be fully controlled and replicated. This meant that for transitions toward a 3D model (2D to 3D case), a dynamic rendering of the model could be replaced by a prerecorded video sequence due to the deterministic camera path. Because photo-realistic 3D models are labor intensive to create, we decided to skip the modeling step altogether and use video sequences of the real world. This approach had the added benefit of bypassing potential issues of model quality, which could have distracted participants by generating rendering artifacts. Since this is visually equivalent to displaying renderings of a perfect 3D model, we expect results to be as valid as if an actual model had been used. In the following discussion, graphs, and tables, the 2D to 2D case will be referred to as *photo* and the 2D to 3D case as *video*. The video sequences were taken with a HDV camera mounted on a dolly. Mattes for the occluders were again created manually.

Following feedback from the feasibility study, the closing-in motion was created to be continuous, starting with a zoom into the initial photograph, followed by the transition smoothly starting. This was achieved by applying the easing function described in Section 5.2.2 to the forward motion as a whole, rather than to two consecutive segments (pan/zoom then transition).

5.3.2 *Design.* As a compromise between the number of participants needed and the length of each session, we chose a design with one group for each level of the *toward* factor (whether we are transitioning toward a photograph (2D to 2D) or a video (2D to 3D)), with the *scene* and *speed* factors crossed

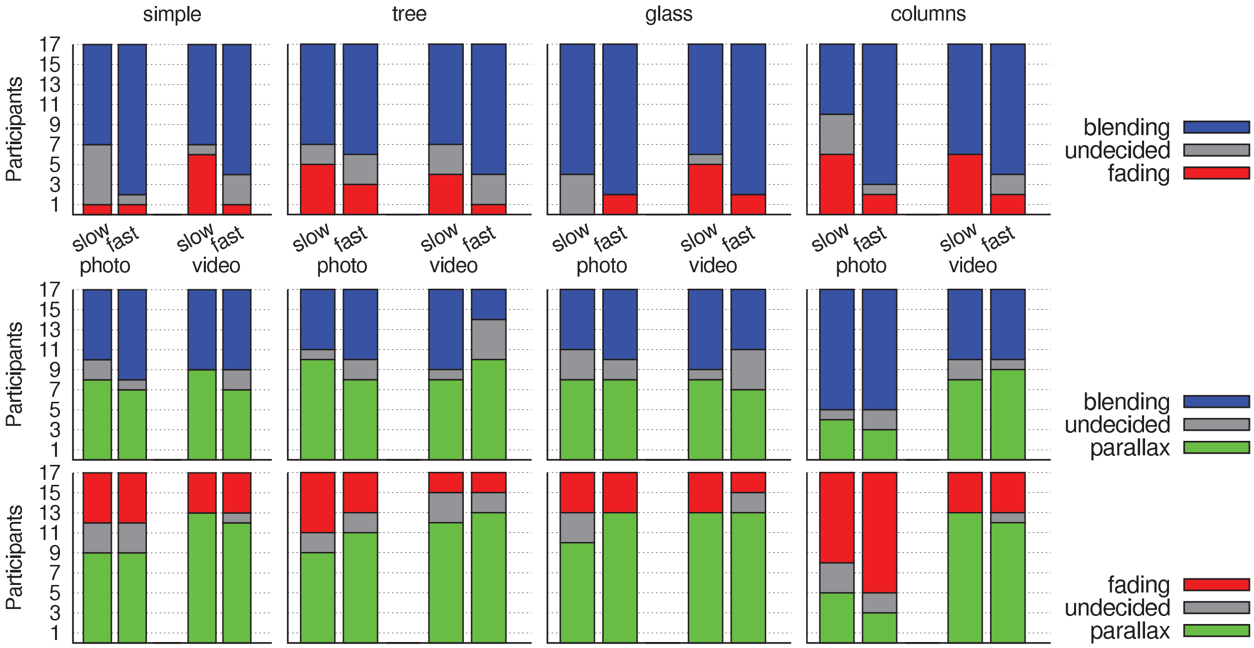

Fig. 6. Breakdown of participants in terms of preference or indecision for each case of the experiment. Top: blending vs. fading, middle: blending vs. fake parallax, bottom: fading vs. fake parallax.

within subjects. Two levels of the speed factor were used, corresponding to transition durations of 1 and 4 seconds. The initial pure zoom duration was chosen to be equal to the transition length, resulting in total presentation times of 2 and 8 seconds, respectively. Each case consisted of showing two transitions one after the other. The participant's task was worded as follows: "Pick the transition that you found most visually appealing (first/second)."

As in the feasibility study, cases were repeated four times. This design was applied to the three possible transition pairs, with results to be analyzed independently. Cases and transition pairs were interleaved randomly, and the order in which transitions within a pair were presented was randomized as well. Both groups saw $2(\textit{speed}) \times 4(\textit{scene}) \times 4(\textit{repetitions}) \times 3(\textit{pairs}) = 96$ transition pairs, for a total session duration of slightly over 20 minutes. The transitions were presented in a controlled environment on an LCD monitor subtending a horizontal field of view of roughly 21°. As results were expected to confirm those obtained in the feasibility study, we aimed for approximately the same number of participants (18) in each group since this had previously proved enough to achieve statistical significance.

5.3.3 *Results.* A total of 34 participants took part in the experiment (17 in each group). Repetitions were collapsed in the same way as explained in the feasibility study. Figure 6 summarizes the participants' answers. At first glance, results are remarkably similar across scenes.

We first tested the probability of obtaining these answers if participants had been indifferent to the type of transition shown: $\chi^2(2, N = 17)$ with expected frequency of a preference occurring of $\frac{5}{16} \times 17$, and expected frequency of a tie (indecision) of $\frac{6}{16} \times 17$ (the *null* hypothesis here is that participants are equally likely to pick the first or second transition). Results are shown in Table II, even rows. They indicate that participants almost always had an opinion. We then tested for the probability of each transition being preferred to its counterpart by even numbers of participants (refer to Table II, odd rows). Results indicate a sharp and consistent preference for the blending transition over fading, while

Table II. Test of Indifference (χ^2) and Equal Preference (p) for Each Transition Pair in Each Experiment Case

		Simple				Tree			
		Photo		Video		Photo		Video	
		Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
fading vs blending	χ^2	**.02**	**0**	**.01**	**0**	**.03**	**.01**	**.04**	**0**
	p	**0**	**0**	0.22	**0**	.15	**.02**	.09	**0**
blending vs parallax	χ^2	.08	**.02**	**0**	.08	**.01**	.08	**.02**	.2
	p	.5	.4	.5	.5	.22	.5	.59	**.04**
parallax vs fading	χ^2	.11	.11	**0**	**0**	.05	**0**	**0**	**0**
	p	.21	.21	**.02**	**.03**	.3	.05	**0**	**0**

		Glass				Columns			
		Photo		Video		Photo		Video	
		Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
fading vs blending	χ^2	**0**	**0**	**0**	**0**	.47	**0**	**0**	**0**
	p	**0**	**0**	.1	**0**	.5	**0**	0.16	**0**
blending vs parallax	χ^2	**.04**	.19	.08	**.02**	**0**	**0**	.08	**.02**
	p	.39	.5	.59	.5	**.03**	**.01**	.5	.4
parallax vs fading	χ^2	**.04**	**0**	**0**	**0**	.11	**0**	**0**	**0**
	p	.09	**.02**	**.02**	**0**	.21	**.01**	**.02**	**.03**

Results at the .05 significance level are shown in bold.

fake parallax is also generally preferred to fading, except when transitioning to a photo in the columns scene at slow speed. At the .05 significance level, the probability of making at least one Type I error over all the experiment cases for each transition pair is quite high: $0.05^{16} = 0.56$. However, we have obtained enough significant results to be confident about our conclusions.

The blending versus fake parallax comparison is inconclusive, with only two significant results that are in contradiction (*tree* × *video* × *slow* and *columns* × *photo* × *fast*). Because of the low number of participants, the power of our tests is low. That is to say, the fact that we do not observe effects does not mean that they are not there. We therefore conducted a first follow-up experiment focusing on the blending versus fake parallax comparison.

5.3.4 *Follow-Up Experiment 1.* The aim of the first follow-up experiment was to reach a conclusion as to whether viewers were split evenly in their pick between the blending and fake parallax transitions. Following Cohen [1988], we determined that obtaining 50 volunteers in each group would let us achieve a power of 0.8 for the binomial test that viewers were less than twice as likely to prefer one transition over the other (.05 significance level). We restricted the experiment to the group transitioning toward videos. The design was identical in other respects. Results are shown in Figure 7 and Table III. At first glance, they indicate an effect of *speed* and *scene*. A hierarchical log-linear analysis confirms the main effect of *speed*: $\chi^2(2, N = 50) = 11.16$, $p \simeq 0$, but not of *scene*: $\chi^2(6, N = 50) = 10.92$, $p = .09$ or of an interaction between the two $\chi^2(6, N = 50) = 4.27$, $p = .6$. There is thus evidence to support the statement that the faster the transition, the better the advantage of blending. There are only two cases were we can reject the even split hypothesis: *tree* × *slow*, in favor of fake parallax, and *columns* × *fast*, in favor of blending. The proportion of participants who preferred the blending transition over the fake parallax transition is shown in Figure 8, along with the confidence intervals at the 95% confidence level.

These results confirm that there is no advantage of using the fake parallax approach for fast transitions toward videos (and consequently 3D models). For slow transitions, the advantage is marginal (observed in one scene with a preference ratio less than 2 : 1, rejected in others).

Table III. Test of Indifference (χ^2) and Equal Preference (p) of Blending vs Fake Parallax, when Transitioning Toward Videos

	Simple		Tree		Glass		Columns	
	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
χ^2	**0**	.23	**0**	.07	**.01**	**.03**	.07	**0**
p	.05	.21	**.04**	.2	.1	.31	.5	**0**

Results at the .05 significance level are shown in bold.

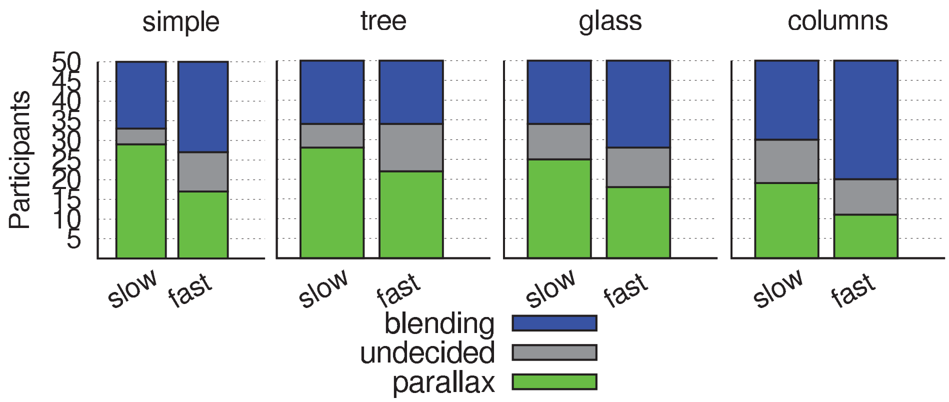

Fig. 7. Breakdown of participants in terms of preference or indecision for each case of the first follow-up experiment (blending vs. fake parallax).

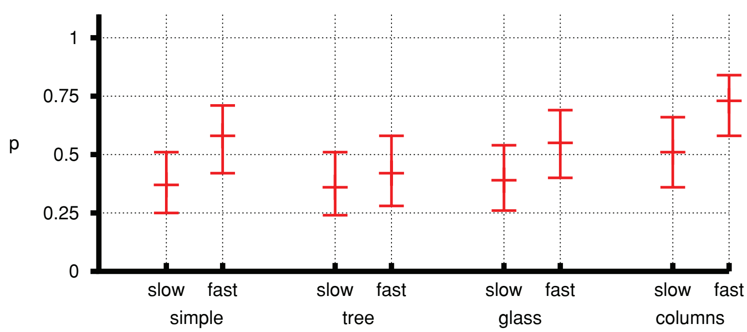

Fig. 8. Proportion and confidence interval of participants who preferred blending over fake parallax.

This is a very different result from what was obtained in the feasibility study. We look for a possible explanation in the main presentation difference: In the feasibility study, a pause was added between the panning/zooming and the transition proper. This caused the occluders to start sliding at the same time as the forward motion resumed. In this experiment, the absence of a pause means that occluders suddenly started sliding while the forward motion was in progress. We pose the hypothesis that a pause helps to suspend disbelief that the occluders would begin to slide after having remained still during the zoom. We conducted a second follow-up experiment to test this hypothesis.

5.3.5 *Follow-Up Experiment 2.* The design of this second follow-up experiment was identical to that of the main experiment (Section 5.3.2), but restricted to the comparison of the fake parallax transition with the blending transition. The number of participants was the same: 17 in the group

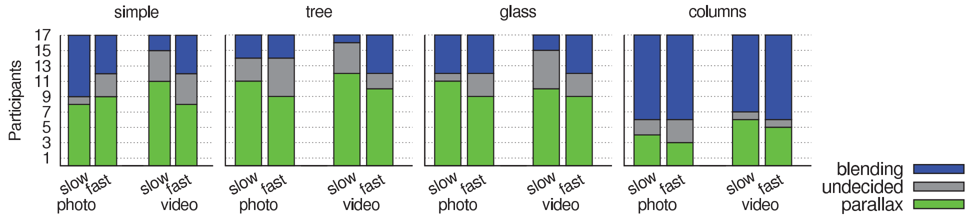

Fig. 9. Breakdown of participants in terms of preference or indecision for each case of the second follow-up experiment (blending vs fake parallax, with pause).

Table IV. Test of Indifference (χ^2) and Equal Preference (p) of Fake Parallax against Blending when a Paused in Marked, for each Experiment Case Results at the .05 Significance Level are Shown in Bold

	Simple				Tree			
	Photo		Video		Photo		Video	
	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
χ^2	**.03**	.11	**.01**	.33	**.01**	.14	**0**	**.03**
p	.6	.21	**.01**	.29	**.03**	.07	**0**	.15

	Glass				Columns			
	Photo		Video		Photo		Video	
	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
χ^2	**0**	.11	**.04**	**.01**	**.01**	**.01**	**.01**	**0**
p	.1	.21	**.02**	.21	.59	**.03**	.22	.1

shown transitions toward another panorama, and 17 in the group shown transitions toward a video. The same scenes were presented, but this time a pause was added between the initial zoom and the transition proper, in the same way as in the feasibility study (Section 5.2.2).

Figure 9 summarizes the outcome of this experiment. It is to be compared with the middle row of Figure 6. Results for the indifference and preference tests (refer to Section 5.3.3) are shown in Table IV. For the first three scenes (simple, tree, and glass), the preference for the fake parallax transition over blending is clear. This confirms our hypothesis (stated at the end of the previous section).

The preference for fake parallax is slightly less clear cut than in the feasibility study, considering that the transition speed used was fast. It is possible that, in the feasibility study, the presentation order of each transition in a pair introduced a bias in favor of fake parallax by chance, given that the time interval between presentations was lengthened by user interaction (panorama browsing and model navigation) and that we did not enforce balance when randomizing the order. However, an analysis of the software log shows that in the worst-case scenarios (first or second systematically picked in case of imbalance), the generated presentation order within each pair could not have produced a bias that would invalidate the statistical significance of the result.

In the case of the columns scene, the preference for blending observed in previous experiments is confirmed. We are confident that this outcome is caused by the greater depth range of this scene, which causes a large mismatch at the occlusion boundary between the van and the columns. Since the target photograph used corresponds to the last frame of the forward dollying video, the mismatch between the initial photograph and the target one is greater than that with the first video frame because the viewpoint discrepancy is greater. In light of this, the fact that the fake parallax effect performs better when transitioning toward a video in the columns scene, both with and without pause, is consistent

with our explanation. It is interesting to note how the video case behaves: While the preference was not clear in the absence of a pause (Figure 9, middle row, last two bars), it becomes so when a pause is marked. This is because the feathered blending of the occlusion boundary mismatch now occurs against a static background (paused video), as in the photo case.

6. DISCUSSION AND FUTURE WORK

We have investigated different transition types, with varying authoring requirements, between panoramas and 3D models, including a variant that we introduce: fake parallax.

We can conclude that view alignment is useful, since transition types that make use of it (blending, fake parallax) are preferred to fading through black. The fake parallax effect is only preferred to blending when the fake occluder motion does not start in the middle of camera motion. In other words, it is only better when the former is *synchronized* with the latter. However, in scenes that exhibit large depth range behind the occluders, the mismatch at occlusion boundaries cannot be concealed well by our feathering approach and causes the blending transition to be preferred. We plan to investigate ways to better conceal this mismatch, in particular using image in-painting techniques. It will then be interesting to study how occluder movement realism (i.e., depth information) alone affects viewer preferences. As could be expected, the faster the transition, the less clear it is that the creation of occluder mattes to fake the parallax effect is justified.

Our conclusions remain limited to forward motions toward an object of interest, rather than empty space. Other cases generate much larger perspective and disocclusion effects, caused by the incompatibility in viewing distance. An exploration of such cases as a function of the scene's depth properties would be of interest.

REFERENCES

Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C. 2000. Image inpainting. In *Proceedings of the Siggraph Computer Graphics Conference*, K. Akeley, Ed. ACM Press/Addison Wesley Longman, 417–424.

Chiang, C.-C., Huang, A., Wang, T.-S., Huang, M., Chen, Y.-Y., Hsieh, J.-W., Chen, J.-W., and Cheng, T. 1997. Panovr sdk—A software development kit for integrating photo-realistic panoramic images and 3-d graphical objects into virtual worlds. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST'97)*. ACM, New York, 147–154.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences (2nd Ed.)*. Lawrence Erlbaum.

Criminisi, A., Reid, I., and Zisserman, A. 2000. Single view metrology. *Int. J. Comput. Vis. V40,* 2, 123–148.

Frueh, C., Sammon, R., and Zakhor, A. 2004. Automated texture mapping of 3d city models with oblique aerial imagery. In *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04)*. IEEE Computer Society, Los Alamitos, CA, 396–403.

Gledhill, D., Tian, G., Taylor, D., and Clarke, D. 2003. Panoramic imaging—A review. *Comput. Graph. 27*, 435–445.

Horry, Y., Anjyo, K.-I., and Arai, K. 1997. Tour into the picture: Using a spidery mesh interface to make animation from a single image. In *Proceedings of the 24th Annual SIGGRAPH Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing, New York, 225–232.

Kazhdan, M., Bolitho, M., and Hoppe, H. 2006. Poisson surface reconstruction. In *Proceedings of the 4th Eurographics Symposium on Geometry Processing (SGP'06)*. Eurographics Association, 61–70.

Lowe, D. G. 1999. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision (ICCV)*. 1150–1157.

McMillan, L. and Bishop, G. 1995. Plenoptic modeling: An image-based rendering system. In *Proceedings of the SIGGRAPH'95 Computer Graphics Conference*. 39–46.

Müller, P., Zeng, G., Wonka, P., and Van Gool, L. 2007. Image-Based procedural modeling of facades. *ACM Trans. Graph. 26,* 3.

Pollefeys, M., Nistér, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S. J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G., and Towles, H. 2008. Detailed real-time urban 3d reconstruction from video. *Int. J. Comput. Vis. 78,* 2-3, 143–167.

Snavely, N., Garg, R., Seitz, S. M., and Szeliski, R. 2008. Finding paths through the world's photos. In *Proceedings of ACM SIGGRAPH Papers*. ACM, New York, 1–11.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2007. Modeling the world from Internet photo collections. In *Int. J. Comput. Vis.* to appear.

STICH, T., LINZ, C., WALLRAVEN, C., CUNNINGHAM, D., AND MAGNOR, M. 2008. Perception-Motivated interpolation of image sequences. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV'08)*. ACM, New York, 97–106.

VAINA, L. M., BEARDSLEY, S. A., AND RUSHTON, S. K., Eds. 2004. *Optic Flow and Beyond*. Kluwer Academic Publishers, Norwell, MA.

VINCENT, L. 2007. Taking online maps down to street level. *Comput. 40,* 12, 118–120.

WATANABE, M. AND NAYAR, S. K. 1998. Rational filters for passive depth from defocus. *Int. J. Comput. Vis. 27,* 3, 203–225.

ZHU, Z. AND HANSON, A. R. 2001. 3d lamp: A new layered panoramic representation. In *Proceedings of the International Conference on Computer Vision (ICCV'02)*, 723.

Received July 2009; accepted August 2009