Visual Equivalence in Dynamic Scenes

Peter Vangorp
Timothy S. Condon
James A. Ferwerda
Kavita Bala
Roeland Schoukens
Philip Dutré

Report CW 557, July 2009

Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

Visual Equivalence in Dynamic Scenes

Peter Vangorp *Timothy S. Condon*
James A. Ferwerda *Kavita Bala*
Roeland Schoukens *Philip Dutré*

Report CW 557, July 2009

Department of Computer Science, K.U.Leuven

Abstract

Characterizing the fidelity of an image is a fundamental problem in computer graphics. Visible difference predictors are used to detect when images are perceptually indistinguishable. Visual equivalence introduces a new appearance-based standard for image fidelity, where two visibly different images of an object can still be visually equivalent as representations of the object's appearance if they convey the same shape, material and lighting properties. The original visual equivalence study focused on static scenes. In this paper, our goal is to understand and evaluate visual equivalence in the context of dynamic scenes.

We conducted a series of psychophysical experiments that explore visual equivalence for objects undergoing rotational and linear motion under a variety of illumination transformations. We demonstrate that equivalence continues to exist with motion for transformations that maintain some important properties. Specifically, they should not affect the sense of structured motion in the patterns of surface reflections. Based on these insights, we propose a new interpolation transformation that preserves some of these dynamic visual features.

Keywords : perception, human visual system, appearance, motion.
CR Subject Classification : I.3.7, J.4

Visual Equivalence in Dynamic Scenes

Peter Vangorp* Katholieke Universiteit Leuven
Timothy S. Condon† Cornell University
James A. Ferwerda‡ Rochester Institute of Technology
Kavita Bala† Cornell University
Roeland Schoukens* Katholieke Universiteit Leuven
Philip Dutré* Katholieke Universiteit Leuven

(a) Rotating object (b) Rotating object and camera (c) Translating object and camera

Figure 1: A few frames from the animation stimuli we used to study visual equivalence in dynamic scenes. (a) Experiment 1(a) studies rotating objects with a static camera. (b) Experiment 1(b) studies rotating objects with synchronous camera rotation. (c) Experiment 2 studies translating objects with a tracking camera. See the accompanying video for the complete animations.

Abstract

Characterizing the fidelity of an image is a fundamental problem in computer graphics. Visible difference predictors are used to detect when images are perceptually indistinguishable. Visual equivalence introduces a new appearance-based standard for image fidelity, where two visibly different images of an object can still be visually equivalent as representations of the object's appearance if they convey the same shape, material and lighting properties. The original visual equivalence study focused on static scenes. In this paper, our goal is to understand and evaluate visual equivalence in the context of dynamic scenes.

We conducted a series of psychophysical experiments that explore visual equivalence for objects undergoing rotational and linear motion under a variety of illumination transformations. We demonstrate that equivalence continues to exist with motion for transformations that maintain some important properties. Specifically, they should not affect the sense of structured motion in the patterns of surface reflections. Based on these insights, we propose a new interpolation transformation that preserves some of these dynamic visual features.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture; Animation; J.4 [Social and Behavioral Sciences]: Psychology

Keywords: perception, human visual system, appearance, motion

1 Introduction

Characterizing the fidelity of an image is a fundamental problem in computer graphics. Visible differences predictors (VDPs) [Daly 1993] model early vision to characterize image fidelity by predicting when two images will be indistinguishable. Recently, Ramanarayanan et al. [2007] introduced a new appearance-based standard for image fidelity: *visual equivalence*. Two visibly different images of an object can still be visually equivalent as representations of the object's appearance if they convey the same shape, material and lighting properties. Ramanarayanan et al.'s study focused on the degree to which representations of lighting can be approximated or manipulated without affecting object appearance. In this paper, we seek to understand and evaluate visual equivalence in the context of dynamic scenes: does the motion of an object or of the surrounding scene affect equivalence positively or negatively? Previous work indicates that motion can either obscure differences [Yee et al. 2001] or instead bring them out [Hartung and Kersten 2002].

To investigate this issue, we conducted a series of psychophysical experiments that explore visual equivalence for objects undergoing rotational and translational motion. We explored visual equivalence over a variety of illumination transformations, including warps and interpolation of the environment illumination field. Interpolating such environment maps is common practice in interactive applications like games. We demonstrate that equivalence continues to exist with motion for some illumination transformations, though the transformations must maintain some important properties. Specifically, transformations like warp which do not fundamentally alter the spatial and temporal statistics of the illumination field provide equivalence. On the other hand, simple interpolation often fails to convey realistic structured motion in the patterns of surface reflections of objects, resulting in visual inequivalence. On the basis of these insights, we propose a new interpolation transformation

---

*e-mail: {peter.vangorp,roeland.schoukens,philip.dutre}@cs.kuleuven.be
†e-mail: {tcondon,kb}@cs.cornell.edu
‡e-mail: jaf@cis.rit.edu

that preserves some of these dynamic visual features and thereby efficiently produces animations rendered with interpolated illumination maps that are visually equivalent to reference animations.

In Section 2 we discuss previous work. Section 3 describes the general design of our psychophysical experiments. In Sections 4 and 5 we present our experiments and results for two specific types of motion. Finally, we discuss these results and conclude in Section 6.

2 Previous Work

2.1 Visual equivalence

Ramanarayanan et al. [2007] introduced the concept of visual equivalence and demonstrated that it can be used as the basis of advanced perceptual metrics for realistic image synthesis. Visual equivalence is the phenomenon that two visibly different images can be equivalent as representations of object properties such as shape and material, and scene properties such as illumination. Through a series of psychophysical experiments, the authors found that images rendered with blurred or warped illumination maps are perceived as visually equivalent even though the images themselves are visibly different. They also found significant differences between the blur and warp transformations, and saw a significantly greater range of equivalence for the warps than for the blurs. Ferwerda et al. [2008] extended the analysis of this work and related the results to changes in natural image statistics.

Ramanarayanan et al. [2007] and Ferwerda et al. [2008] provide recent reviews of the shape, material, and illumination perception literature relevant to the concept of visual equivalence, and the reader is referred there for more information. However, since our goal is to understand visual equivalence in dynamic scenes, key findings from the motion perception literature must be considered.

When an object moves, it produces a changing pattern of stimulation on the retina. Gibson [1950] demonstrated that the information in these transformations provided reliable visual information for object motion. Dynamic occlusion, motion parallax, and optic flow are the classic examples of dynamic cues for depth and the spatial layout of scenes, and an observer's movement through it. But more generally, motion perspective and the transformation of visual cues such as texture gradients have also been shown to provide reliable information for object shape and dynamics [Koenderink 1986; Cutting and Millard 1984; Palmer 1999].

Transformations of illumination features such as surface shading, shadows, highlights, and reflections have also been shown to provide reliable information for object and scene properties. Todd and Mingolla [1983], Mingolla and Todd [1986], and Norman et al. [2004] have demonstrated that dynamic transformations of shading on diffuse objects and specular highlights on glossy objects provide information for object shape. More recently Hartung and Kersten [2003], Savarese et al. [2004], and Fleming et al. [2004] have shown that observers can also use the complicated patterns of environmental reflections seen in mirror-like objects to perceive the shapes of geometrically complex objects.

2.2 Rendering with environment maps

Environment maps were originally introduced in computer graphics by Blinn and Newell [1976] to add convincing specular reflections of the distant environment to rendered images. The technique was later extended to include glossy reflections as well [Miller and Hoffman 1984; Greene 1986; Cabral et al. 1999; Heidrich and Seidel 1999; Kautz et al. 2000; Kautz and McCool 2000]. Environment maps capture complex illumination from either synthetic or natural scenes by representing all light sources as purely directional and infinitely distant. They may then be used as light sources for artificial scenes, producing compellingly realistic results.

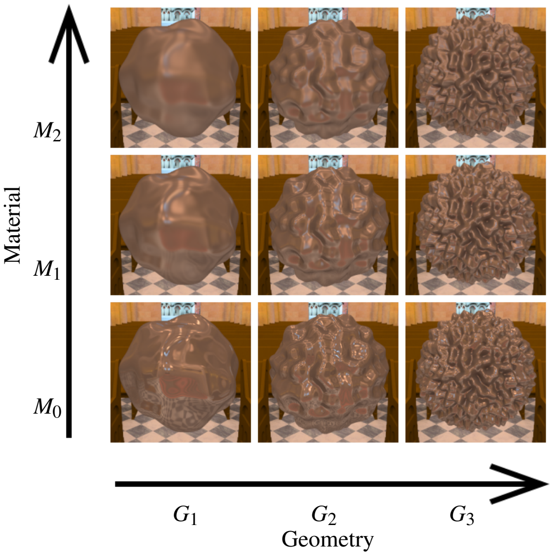


Figure 2: Geometry and materials of the stimuli.

Precomputed radiance transfer techniques model the so-called transfer function, which describes global light transport in the scene. Both this transfer function and the environment map are projected onto lower dimensional bases such as spherical harmonics [Sloan et al. 2002] or wavelets [Ng et al. 2003] to allow interactive relighting of static scenes.

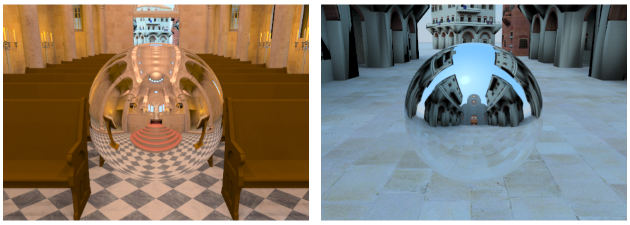
These techniques have been widely adopted in interactive media such as video games, where environment lighting is often calculated at sample points and interpolated to approximate dynamic lighting. These environment samples, usually low-frequency spherical harmonics, are either scattered manually by an artist [Chen and Liu 2008; Valient 2007] or spaced uniformly across the scene [Shopf et al. 2008].

3 Experimental Design and Procedure

To understand visual equivalence in dynamic scenes, we must study objects with a range of shapes and materials moving through a complex environment under varied lighting conditions. To do this, we created a set of animations that would allow us to systematically explore interactions between geometry, motion, and illumination and their impact on the visual equivalence of objects in the animations.

3.1 Stimuli

First, we created a set of animations that spanned a range of geometry, materials, motion, and illumination. Figure 1 shows a few frames from our animations. The scene model consists of a bumpy sphere-like test object moving (either rotating or translating) through an environment. The environment transitions from a high-frequency indoor lighting colored in warm tones to a lower-frequency outdoor lighting in cooler tones. Below, we describe the object, scene, and rendering parameters used to generate the various animations used in our study, and the motivations behind our decisions.



(a) Inside Sibenik Cathedral (b) Outside courtyard

Figure 3: Mirror spheres in the 2 distinct parts of the scene.

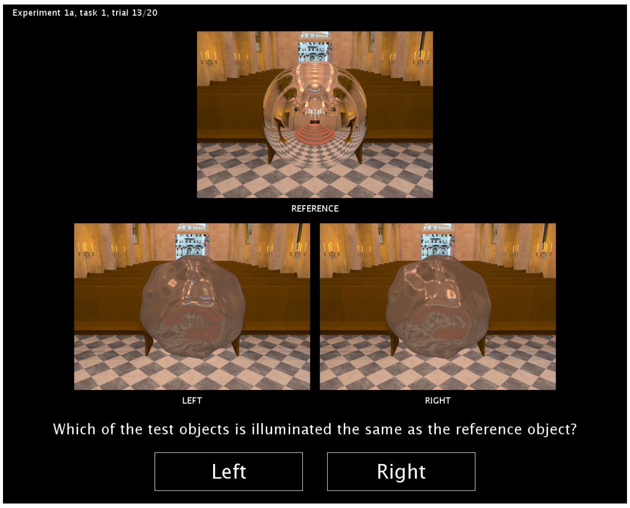
Geometry and Materials: Our goal was to pick shapes and materials from a perceptually uniform space. For our experiment, we selected 9 of the 16 objects used in [Ramanarayanan et al. 2007]. Those authors used a sphere and 3 blobby objects (G_1, G_2, G_3), following the precedent of perception literature [Todd and Mingolla 1983; Todd 2004]. $G_1 - G_3$ are modified spheres with bumps of uniform amplitude and increasing spatial frequency with the aim of producing objects of roughly uniformly increasing "bumpiness". We eliminated the sphere from our experiment since it is an extreme case and unrepresentative of typical scene geometry. We also eliminated the most diffuse of the original four materials, reducing the material dimension to only three isotropic Ward materials [Ward 1992], with diffuse reflectance $\rho_d = 0.15$, specular reflectance $\rho_s = 0.19$, and surface roughness (specular lobe width) values $\alpha = \{0.01, 0.06, 0.11\}$ for $M_0 - M_2$ respectively. These parameters were selected to represent perceptually equal changes in gloss appearance [Pellacini et al. 2000].

Illumination: We sought to render our animations under illumination environments that mimic real-world conditions, in particular maintaining properties such as high resolution and high dynamic range (Figure 3). In order to generate meaningful motion, we also needed our environment map sequence to be spatially coherent throughout the entire animation. Given that an arbitrary sequence of environment maps will not exhibit this property, we chose to design our own scene and render the requisite maps along the path of motion. Our scene includes a large indoor space (Sibenik Cathedral) with a doorway into an outdoor courtyard area surrounded with buildings. The interior was inspired by the well-known natural illumination environment map of Grace Cathedral [Debevec 1998], featuring a large number of small interior lights, as well as global illumination, and other objects were added to increase the visual complexity of the setting. The exterior was designed to represent a courtyard flanked by buildings and illuminated by an environment map based on the daylight model of Preetham et al. [1999].

Rendering and display: The high quality background of the scene and the synthetic environment maps used as illumination sources were both rendered with the multidimensional lightcuts global illumination rendering algorithm [Walter et al. 2005; Walter et al. 2006]. The animations were rendered at a resolution of 484×342 and a frame rate of 30 frames per second.

For display, the images were tone mapped using the global photographic operator of Reinhard et al. [2002] with constant parameters tuned to the characteristics of the LCD display. The images were viewed under office lighting conditions. At a nominal 60 cm viewing distance each image subtended 11.6° of visual angle and each test object subtended 7°.

Each experiment used a specific type of **motion** of the object and camera (rotation or translation) and different **illumination transformations** (warp or interpolation).



Which of the test objects is illuminated the same as the reference object?

Figure 4: Interface used in the experiments: The objects in the left and right images have the same geometry and material properties and they always move in the same way. In this condition, one is rendered with the reference illumination map and the other is rendered with one of the transformed maps. The observer's task is to identify the object that is illuminated the same as the reference object.

3.2 Procedure

Participants were presented with a pair of animated stimuli using the interface shown in Figure 4. Both sequences were renderings of the same object – identical geometry and material – with the same motion, but one was rendered with reference illumination, and the other with illumination that had been manipulated in some way, such as a warp or interpolation. Some trials included a third reference animation of either the same test object again or of a chrome sphere, performing the same motion under the reference illumination. For any given trial, the participant was asked to perform one of four tasks that compared animations based on different criteria:

1. *Indistinguishability Task*: In this task, participants were shown a reference animation and a pair of test animations, all showing the same object and motion. One of the test animations was identical to the reference animation, while the other test animation was rendered with a transformed illumination map. The question posed to participants was: "Which of the test movies is the same as the reference movie?", with "left" or "right" being the possible responses. The purpose of this task is to determine when movies rendered with the transformed maps were visibly different (in the VDP sense) from movies rendered with the reference map.

2. *Shape Difference Task*: In this task, participants were shown two animations of the same object and motion, one rendered with the reference map, the other with a transformed map. Participants had to answer the question: "Do the objects have the same shape?". Possible responses were "same" or "different". The intent of this task was to determine if the transformed maps produced illusory differences in the apparent shape of the objects.

3. *Motion Difference Task*: In this task, participants were again shown two animations of the same object and motion, one rendered with the reference map, the other with a transformed map. Participants were asked: "Are the objects moving in the same way?". Possible responses were "same" or "different".

The intent of this task was to determine if the altered maps produced illusory differences in the apparent motion of the objects.

4. *Illumination Difference Task*: In this task, participants were shown a reference animation of a chrome sphere under reference illumination, and a pair of test animations with identical test objects, all performing the same motion. One test animation was rendered with the reference map, the other with a transformed map. Participants were asked: "Which of the test objects is illuminated the same as the reference object?". Possible responses were "left" or "right". The intent of this task was to determine whether observers can use surface reflection patterns to detect differences in scene illumination.

The Indistinguishability and Illumination difference questions are 2-alternative forced choice (2AFC) designs: the objective is to select out of two stimuli the one that satisfies a certain criterion. The Shape and Motion difference questions are same/different designs: the objective is to compare two stimuli for the equality of a certain property [Gescheider 1997].

Note that the four tasks can be divided into two conceptual categories. In the *Indistinguishability* task, participants are being asked to simply report on any differences between *movies*. In the *Shape*, *Motion*, and *Illumination difference* tasks participants are being asked to report on apparent differences between *objects*. We chose these tasks because they allow us to dissociate the effects of visible differences on image and object appearance, and quantify when different configurations of object geometry, motion, and illumination produce animations that are visually equivalent.

Visual equivalence was previously defined as the preservation of all visual appearance properties, including shape, material, and illumination, despite potential visible differences. Motion in dynamic scenes becomes an equally important visual appearance property. We decided to omit the material task because it was geared towards detecting the material ambiguity arising from blurred environment maps.

Within each experiment, the Shape, Motion, and Illumination tasks were delivered in random order. Contrary to [Ramanarayanan et al. 2007] we decided always to ask the Indistinguishability task last to avoid a learning effect we observed informally in early trials. Most stimulus movies have an easily distinguishable feature at some point in the sequence. Participants would learn to recognize the reference stimulus during the indistinguishability task and they would notice that shape and motion differ rarely if ever. If they had some experience in computer graphics, they could deduce that illumination was the only variable left. In the illumination task, they would assume correctly that the reference stimulus always matches the illumination of the mirrored sphere, giving them an unfair advantage.

Within each task, both the overall order of presentation and left/right positions of the images were randomized across trials. On each trial, participants entered their responses with a keyboard or mouse. The trials were open-ended, and participants could take as much time as they needed. On average, participants took about 7 seconds to complete each trial.

Overall, 17 volunteers participated in each of the experiments. The participants were university students, staff, and faculty (ages 20 to 65). Most had technical backgrounds, with about half in graphics and imaging. All were naive to the design and purpose of the experiments. All had normal or corrected-to-normal vision.

We calculated the response categories using the following methods (after [Ramanarayanan et al. 2007]):

- The performance of the *average observer* was calculated by averaging the binary scores of the participants into continuous probabilities. If a participant failed the Indistinguishability task for a certain stimulus pair, their answers to the corresponding Shape, Motion, and Illumination tasks were excluded from the averages for those tasks.
- If the average probability of correctly selecting the identical stimulus in the Indistinguishability task was below the conventional 75% (2AFC) threshold, then stimulus pair was classified as visually *indistinguishable*.
- If the average probability of correctly selecting the matching stimulus in the Illumination task was above the 75% threshold, that stimulus was classified as visibly *different*.
- If the probability of mistaking an illumination difference for a difference in shape or motion was above the conventional 50% (same/different) threshold, that stimulus pair was also classified as visibly *different*.
- Otherwise, the stimulus pair was classified as visually *equivalent*.

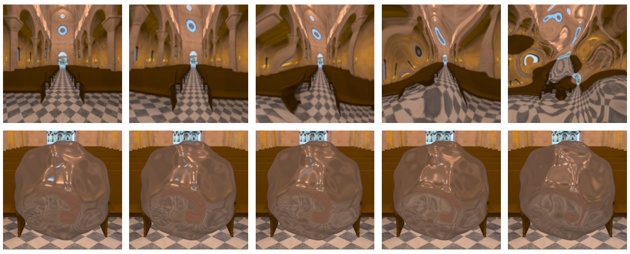


Figure 5: The illumination warps used in Experiment 1 (rotating object inside Sibenik Cathedral), and the resulting renderings on object G1/M0.

4 Experiment 1: Rotation

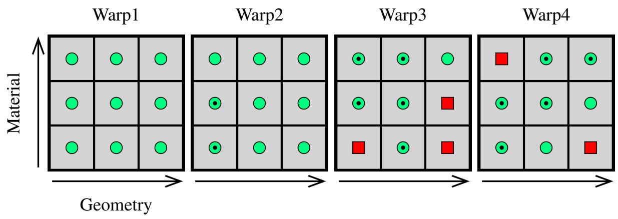
In order to study the effects of motion incrementally, Experiment 1 introduces only simple rotational motion.

In Experiment 1(a) the object oscillated back and forth around a vertical axis through an angle of $\pm 15^\circ$ with respect to the view direction. The object's velocity was modulated sinusoidally to assure smooth transitions at the reversal points. Peak velocity was $30^\circ/s$. The fixed camera was positioned at a distance of 1 m from the rotating object. The object was lit by an environment map that represented the remainder of the scene geometry.

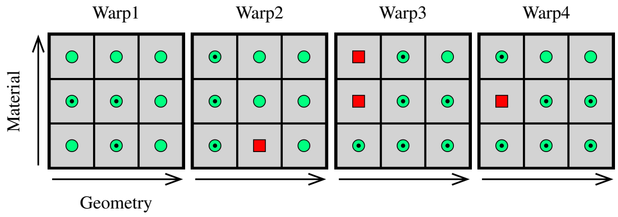
In Experiment 1(b) the object's motion was the same as described above, but the scene camera moved in synchrony with the object. Thus, the camera always saw the same side of the object, but both the scene behind the object and the reflections on the object's surface changed according to the motion.

In both studies, the transformed illuminations were spatially warped versions of the reference illumination. We selected four of the 5 warps from [Ramanarayanan et al. 2007] (Figure 5). We eliminated the most extreme warp (warp5) from our studies since we expected it would always be visually inequivalent. Observers performed the Indistinguishability, Shape, Motion, and Illumination difference tasks for each object.

The final animation frames combined the rotating object with the pre-rendered environment illumination, using a state-of-the-art prefiltered environment map technique [Green et al. 2007]. In total 92



(a) Experiment 1(a) – static camera



(b) Experiment 1(b) – tracking camera

● Indistinguishable ⊙ Equivalent 🟥 Different

Figure 6: Results of Experiment 1 – rotation.

animations were rendered: (3 geometries × 3 materials × (4 warps + 1 reference illumination) + 1 mirror ball reference) × 2 motions.

4.1 Results and Analysis

In Experiment 1, we were interested in evaluating if visual equivalence holds for objects undergoing rotation. The results are summarized in Figure 6. Three trends stand out:

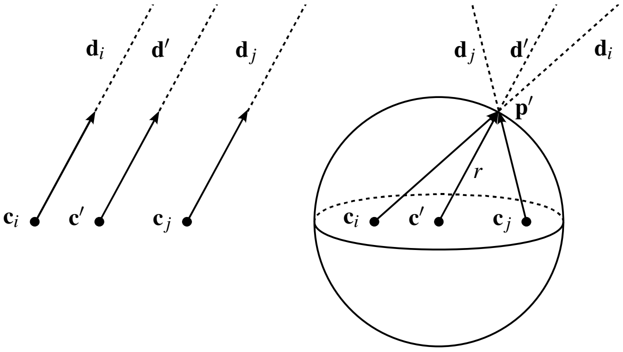
- Lower levels of environment map warping produced animations that were indistinguishable or visually equivalent to reference animations. Higher levels of environment map warp exhibit some cases where the object's appearance was classified as visibly different from the reference object. However, closer inspection revealed that 5 of the 9 cases were actually very close to the visual equivalence threshold (Table 1). This pattern of results underscores the fact that psychophysical data is noisy and binary thresholds are somewhat arbitrary. More accurate would be a continuous scale (perhaps red/green) that indicates the distance of the responses for the test stimulus from the threshold.
- Participants rarely misinterpreted illumination differences as shape or motion differences. In only 4 out of 72 cases (6%) was the equivalent/different classification decided by the Shape or Motion difference tasks. This will allow us to omit these tasks in Experiment 2.
- These results are consistent with the previous work of Ramanarayanan et al. [2007], although the motion introduces slightly more visual equivalence than they found for similar warps. We speculate that motion helps to reduce confusion in the Shape perception task. However, motion apparently does not help enough to make the illumination easily distinguishable.

5 Experiment 2: Translation

In Experiment 2 we studied linear translation of the object through space. The object moves at a speed of 21.6 km/h for 6.7 s, covering a total distance of 40 m. The camera "chases" the object at a constant distance of 1 m. The start and end positions, as well as trajectory, were selected to ensure that the object spends an equal portion of the time indoors and outdoors (3.3 s).

Criterion: Equivalent if:	Illum. $< 75\%$	Shape $> 50\%$	Motion $> 50\%$	Indist. $< 75\%$
1(a) Warp3 G1/M0	62%	38%	77%	76%
1(a) Warp3 G3/M0	46%	92%	46%	76%
1(a) Warp3 G3/M1	86%	86%	71%	82%
1(a) Warp4 G1/M2	79%	64%	57%	82%
1(a) Warp4 G3/M0	53%	93%	47%	80%
1(b) Warp2 G2/M0	77%	85%	62%	76%
1(b) Warp3 G1/M1	82%	41%	53%	100%
1(b) Warp3 G1/M2	59%	41%	59%	100%
1(b) Warp4 G1/M1	82%	35%	53%	100%

Table 1: Details of the inequivalent cases of Experiment 1. These cases would become equivalent if all equivalence criteria or the distinguishability criterion were met. Highlighted values are not significantly different from the threshold and would change the classification if they crossed it.



(a) Pixelwise interpolation (b) Reprojection

Figure 8: Illumination transformations.

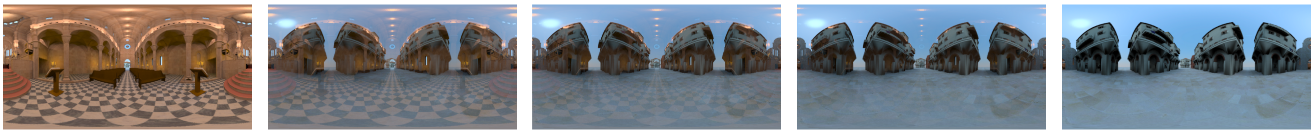
In this experiment we considered three diverse illumination transformations for the animated environment maps:

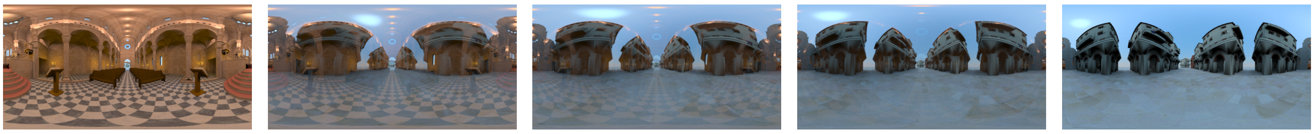
Warp: We only considered one warp (warp2), because such transformations are rarely applicable to real time rendering engines. It would require an engine that computes a new approximate environment map for each frame, with temporally coherent warp-like distortion artifacts.

Interpolation: We considered various interpolation schemes for the animated environment maps. One common technique used in interactive applications such as games is to precompute only "key frame" environment maps at a small number of scene locations, and interpolate between them for intermediate positions. We study two interpolation schemes of environment maps, illustrated in Figure 8.

1. **Pixelwise interpolation (P):** Each intermediate environment map is a pixel-by-pixel linear interpolation of the adjacent keyframe environment maps. This is identical to linear interpolation in other orthogonal function bases such as spherical harmonics. Pixelwise interpolation will induce a sense of motion if corresponding features overlap sufficiently in adjacent key frame environment maps.
2. **Reprojection followed by pixelwise interpolation (RP):** Key frame environment maps $\mathbf{M}_i$ are centered at the points


(a) Reference


(b) Pixelwise interpolation (P)


(c) Reprojection followed by pixelwise interpolation (RP)

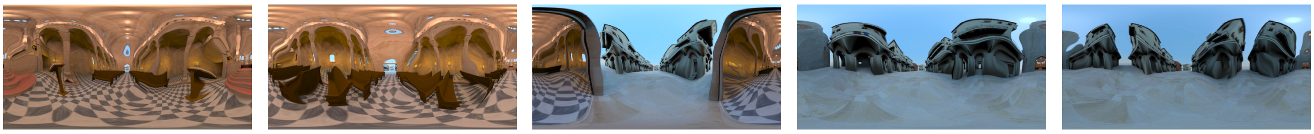

(d) Warp (W)

Figure 7: Example illumination transformations. (a) Reference environment maps. (b) Pixelwise interpolation between the start and end frames ($k = 200$). This extreme case demonstrates the behavior of this operation: stationary features fade in and out. (c) Reprojection followed by pixelwise interpolation between the start and end frames ($k = 200$). Features move in a plausible way while fading in and out. (d) Temporally coherent warp of each individual reference frame.

> $\mathbf{c}_i$ in the scene. An intermediate environment map $\mathbf{M}'$ is centered at the point $\mathbf{c}'$. To compute the value of the intermediate environment map in a certain direction $\mathbf{d}'$, project that direction $\mathbf{d}'$ onto a sphere centered at $\mathbf{c}'$ with a finite radius r that should enclose the adjacent key frame centers $\mathbf{c}_i$. Call this projected point $\mathbf{p}' = \mathbf{c}' + r\mathbf{d}'$. For each adjacent key frame $\mathbf{M}_i$, that point $\mathbf{p}'$ lies in the direction $\mathbf{d}_i = \mathbf{p}' - \mathbf{c}_i$. The value of $\mathbf{M}'(\mathbf{d}')$ is the linear interpolation of the values $\mathbf{M}_i(\mathbf{d}_i)$. This simple reprojection produces a plausible optical flow and an improved sense of motion.

In practice, both interpolation schemes use linear interpolation between 2 key frame environment maps, but they trivially generalize to arbitrarily weighted barycentric, bi-, or trilinear interpolation.

For both interpolation schemes we must pick key frames at some distance k apart. For a 200 frame sequence, a keyframe interval $k = 200$ would mean only the two end frames are interpolated for all intermediate frames, and $k = 1$ would mean there is no interpolation. To keep the study size manageable we selected the keyframe intervals $k = \{50, 25, 10\}$ that span the threshold between equivalence and inequivalence.

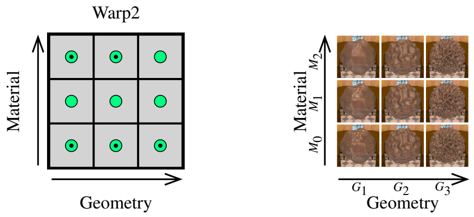
The test object was illuminated by either the reference map animation, a warped map animation, pixelwise interpolation (P), or reprojected pixelwise interpolation (RP). The final animations, combining the translating object with the transforming illumination, were rendered using multidimensional lightcuts [Walter et al. 2006]. In total 73 animations were rendered: 3 geometries × 3 materials × (2 transformations × 3 key frame distances + 1 warp + 1 reference illumination) + 1 mirror ball reference.

Not all tasks were performed in this experiment; based on the results of Experiment 1, which demonstrated that the shape and motion properties were almost always unambiguous, we removed those tasks. Observers performed the Indistinguishability task and the Illumination difference task for each object.

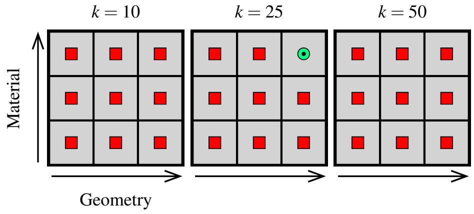
5.1 Results and Analysis

In Experiment 2 we were interested in evaluating whether visual equivalence holds for objects undergoing translation, also for different kinds of illumination transformations besides warping. The results are shown in Figures 9(a), 9(c), and 9(d). We found two major trends:

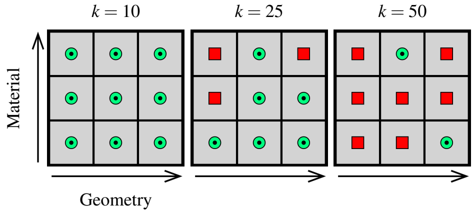
- Reprojection exhibits the most interesting behavior. Figure 9(d) clearly shows that visual equivalence increases when the keyframe interval k decreases. As compared to pixelwise interpolation, reprojection consistently exhibits more equivalence. Holding material constant and comparing changes in geometry reveals a clear trend that increasing geometric complexity increases visual equivalence in most cases. This matches our expectations based on previous work. With respect to material changes, however, the behavior is somewhat unexpected. Ramanarayanan et al. [2007] observed that more diffuse materials display more visual equivalence, but our study indicates the opposite effect: shiny materials are more likely to be equivalent. The reason for these results is not entirely understood, and is certainly worthy of further study; it may point to a relationship between speed of motion and visual equivalence. One possibility is that the high frequency features of highly reflective materials add so much information that it confuses the observer and makes it difficult for them to track features across the surface of the object.
- Reprojection exhibits only visual equivalence, and never vi-



(a) Experiment 2 – warp



(c) Experiment 2 – pixelwise interpolation



(d) Experiment 2 – reprojection

Figure 9: Results of Experiment 2 – translation.

> sual indistinguishability. Thus, despite the fact that this transformation creates animation sequences that are clearly different, these renders are still plausible to human observers. This indicates that visual equivalence in the context of motion is a useful standard, as it opens the door to more aggressive techniques that manage to create a compelling sense of motion, even if they do not perfectly match ground truth.

The parameter space of dynamic scenes is huge and even studying the portion of the space we explored was a massive undertaking. Without a more extensive study of scene/stimulus parameters such as speed of motion, environment map statistics, and interpolation rates, it is inadvisable to construct an analytic prediction metric. While we did develop a support vector machine (SVM) capable of classifying our data using a single separating plane with 85% accuracy, the breadth of illumination variation makes it unlikely that our SVM would generalize to an arbitrary environment map sequence. Likewise, determining the appropriate keyframe interval threshold of visual equivalence remains an open question. Nonetheless, the high-level trends we have shown through our studies are clear and indicate that visual equivalence can serve as valuable heuristic when designing and illuminating dynamic scenes.

We have validated our results using sequences that were not part of our stimuli. Space restrictions do not enable us to include these results.

6 Discussion and Conclusions

In this paper, we have conducted a series of psychophysical experiments to investigate whether the phenomenon of visual equivalence exists for animations of dynamic scenes. We studied two kinds of motion: rotation and translation, and three classes of transformations on the scene illumination maps: warp, pixelwise interpolation, and reprojected pixelwise interpolation. Our results show that different degrees of the transformations produce the same three classes of visual response observed by [Ramanarayanan et al. 2007]: animations *indistinguishable* from reference animations, animations visibly different than the reference but *visually equivalent* in appearance, and animations visibly different and *visually inequivalent* in appearance.

In particular, in the rotation experiments which examined warp transformations, we found that the animations were visually indistinguishable from the references for low warp levels, visually equivalent at moderate warp levels, and inequivalent at higher levels. We posit that this arises because a moderate warp fundamentally retains the important properties of lighting, such as intensity and directionality, while primarily changing the shape of the lights, something we are insensitive to particularly when the lights are only visible in reflections from complex geometry or more diffuse materials. The translation experiments for moderate warp also validate the use of warp transformations in more general settings than studied in [Ramanarayanan et al. 2007]. We conclude that the proposed use of warp for compression or lighting acceleration should be valid in general dynamic scenes.

The translation experiments with illumination interpolation methods showed the same trends, but also revealed significant differences between both classes of interpolations. In particular, pixelwise interpolation which is ubiquitous in interactive applications such as games, was visually inequivalent for all objects at all keyframe intervals tested. This suggests that pixelwise interpolation is a poor method to use for generating illumination mapped animations. On the other hand, our new reprojection method shows significant levels of equivalence for short and moderate keyframe intervals because it better captures a sense of motion. Our proposed reprojection algorithm could itself use refinement: for example, it does not conserve the numerosity or total energy of the light sources. Also, while it can potentially be integrated with wavelet representations, it is unclear if it can be made to work efficiently with the more common spherical harmonics representations.

Overall, these studies confirm that the phenomenon of visual equivalence does hold for animations of dynamic scenes, and that the concept should provide a solid foundation for developing advanced perceptually-based animation rendering algorithms. That said, these studies represent some tentative first steps into a very large domain, and much more work is necessary before quantitative metrics can be developed. In particular, several limitations of the current studies suggest directions for future work. First, although we tried to create a rich and varied environment with interior and exterior characteristics, further tests should be conducted with a range of different illumination maps. Second, we only tested objects moving at one velocity. In Experiment 2, we observed some unexpected interactions between geometry, material, and object velocity that suggest this velocity is likely to affect the salience of illumination transformation; in the same vein, we have no analytic method for selecting the keyframe interval k – these related concerns need further exploration. Third, we need to develop more robust methods for mapping from the geometric properties of our test objects to the properties of real-world objects. Finally, on the basis of these and the abovementioned studies, as future work, we to want to use the insights gained about visual equivalence in dynamic scenes to develop efficient new animation rendering algorithms.

Acknowledgements

The authors thank all volunteers who participated in our experiments. We thank Jurgen Laurijssen for sharing his prefiltered environment map viewer. We would also like to acknowledge Marko Dabrovic (RNA Studios) for providing the Sibenik Cathedral model, Pascal Müller (Procedural.com) for the CityEngine Roman building models, and the Stanford 3D Scanning Repository, Baumgarten Enterprises (`http://www.baument.com`), and various TurboSquid.com contributors for various models used in the cathedral. Peter Vangorp is supported by grant CREA/08/017 from the Research Fund K.U.Leuven (Onderzoeksfonds K.U.Leuven).

References

BLINN, J. F., AND NEWELL, M. E. 1976. Texture and reflection in computer graphics. *Communications of the ACM 19*, 10, 542–547.

CABRAL, B., OLANO, M., AND NEMEC, P. 1999. Reflection space image based rendering. In *Proceedings of SIGGRAPH 99*, ACM Press, A. Rockwood, Ed., Computer Graphics Proceedings, Annual Conference Series, 165–170.

CHEN, H., AND LIU, X. 2008. Lighting and material of *Halo 3*. In *Advances in Real-Time Rendering in 3D Graphics and Games, ACM SIGGRAPH 2008 Course Notes*, N. Tatarchuk, Ed.

CUTTING, J. E., AND MILLARD, R. T. 1984. Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General 113*, 2, 198–216.

DALY, S. 1993. The visible differences predictor: an algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, A. B. Watson, Ed. MIT Press, 179–206.

DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of SIGGRAPH 98*, ACM Press, M. Cohen, Ed., Computer Graphics Proceedings, Annual Conference Series, 189–198.

FERWERDA, J., RAMANARAYANAN, G., BALA, K., AND WALTER, B. 2008. Visual equivalence: an object-based approach to image quality. In *Proceedings IS&T 16th Color Imaging Conference*, 347–354.

FLEMING, R. W., TORRALBA, A., AND ADELSON, E. H. 2004. Specular reflections and the perception of shape. *Journal of Vision 4*, 9, 798–820.

GESCHEIDER, G. A. 1997. *Psychophysics: The fundamentals*. Lawrence Erlbaum Associates, Mahwah, NJ.

GIBSON, J. J. 1950. *The perception of the visual world*. Houghton Mifflin, Boston, MA.

GREEN, P., KAUTZ, J., AND DURAND, F. 2007. Efficient reflectance and visibility approximations for environment map rendering. *Computer Graphics Forum 26*, 3, 495–502.

GREENE, N. 1986. Environment mapping and other applications of world projections. *IEEE Computer Graphics and Applications 6*, 11, 21–29.

HARTUNG, B., AND KERSTEN, D. 2002. Distinguishing shiny from matte. *Journal of Vision 2*, 7, 551.

HARTUNG, B., AND KERSTEN, D. 2003. How does the perception of shape interact with the perception of shiny material? *Journal of Vision 3*, 9, 59.

HEIDRICH, W., AND SEIDEL, H.-P. 1999. Realistic, hardware-accelerated shading and lighting. In *Proceedings of SIGGRAPH 99*, ACM Press, A. Rockwood, Ed., Computer Graphics Proceedings, Annual Conference Series, 171–178.

KAUTZ, J., AND MCCOOL, M. 2000. Approximation of glossy reflection with prefiltered environment maps. In *Graphics Interface*, 119–126.

KAUTZ, J., VÁZQUEZ, P.-P., HEIDRICH, W., AND SEIDEL, H.-P. 2000. Unified approach to prefiltered environment maps. In *Proceedings of the Eurographics Workshop on Rendering Techniques 2000*, Springer-Verlag, London, UK, 185–196.

KOENDERINK, J. 1986. Optic flow. *Vision Research 26*, 161–180.

MILLER, G., AND HOFFMAN, C. 1984. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH 84 Advanced Computer Graphics Animation seminar notes*.

MINGOLLA, E., AND TODD, J. T. 1986. Perception of solid shape from shading. *Biological Cybernetics 53*, 3, 137–152.

NG, R., RAMAMOORTHI, R., AND HANRAHAN, P. 2003. All-frequency shadows using non-linear wavelet lighting approximation. *ACM Transactions on Graphics 22*, 3 (July), 376–381.

NORMAN, J. F., TODD, J. T., AND ORBAN, G. A. 2004. Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science 15*, 8, 565–570.

PALMER, S. E. 1999. *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA.

PELLACINI, F., FERWERDA, J. A., AND GREENBERG, D. P. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of SIGGRAPH 2000*, ACM Press, K. Akeley, Ed., Computer Graphics Proceedings, Annual Conference Series, 55–64.

PREETHAM, A. J., SHIRLEY, P. S., AND SMITS, B. E. 1999. A practical analytic model for daylight. In *Proceedings of SIGGRAPH 99*, ACM Press, A. Rockwood, Ed., Computer Graphics Proceedings, Annual Conference Series, 91–100.

RAMANARAYANAN, G., FERWERDA, J., WALTER, B., AND BALA, K. 2007. Visual equivalence: Towards a new standard for image fidelity. *ACM Transactions on Graphics 26*, 3 (July), 76:1–76:11.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. In *Proceedings of SIGGRAPH 2002*, ACM Press, J. F. Hughes, Ed., Computer Graphics Proceedings, Annual Conference Series, 267–276.

SAVARESE, S., FEI-FEI, L., AND PERONA, P. 2004. What do reflections tell us about the shape of a mirror? In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization (APGV)*, ACM, New York, NY, USA, 115–118.

SHOPF, H., BARCZAK, J., OAT, C., AND TATARCHUK, N. 2008. March of the Froblins: Simulation and rendering massive crowds of intelligent and detailed creatures on the GPU. In *Advances in Real-Time Rendering in 3D Graphics and Games, ACM SIGGRAPH 2008 Course Notes*, N. Tatarchuk, Ed.

SLOAN, P.-P., KAUTZ, J., AND SNYDER, J. 2002. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics 21*, 3, 527–536.

TODD, J. T., AND MINGOLLA, E. 1983. Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance 9*, 4, 583–595.

TODD, J. T. 2004. The visual perception of 3D shape. *TRENDS in Cognitive Science 8*, 3, 115–121.

VALIENT, M. 2007. Deferred rendering in *Killzone 2*. In *The Develop Conference and Expo*.

WALTER, B., FERNANDEZ, S., ARBREE, A., BALA, K., DONIKIAN, M., AND GREENBERG, D. P. 2005. Lightcuts: a scalable approach to illumination. *ACM Transactions on Graphics 24*, 3 (Aug.), 1098–1107.

WALTER, B., ARBREE, A., BALA, K., AND GREENBERG, D. P. 2006. Multidimensional lightcuts. *ACM Transactions on Graphics 25*, 3 (Aug.), 1081–1088.

WARD, G. J. 1992. Measuring and modeling anisotropic reflection. In *Computer Graphics (Proceedings of SIGGRAPH 92)*, ACM, vol. 26, 265–272.

YEE, H., PATTANAIK, S., AND GREENBERG, D. P. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics 20*, 1, 39–65.