Perceptual considerations for motion blur rendering

FERNANDO NAVARRO, Microsoft Games Studios
SUSANA CASTILLO, FRANCISCO J. SERÓN, DIEGO GUTIERREZ, Universidad de Zaragoza

Motion blur is a frequent requirement for the rendering of high quality animated images. However, the computational resources involved are usually higher than those for images that have not been temporally antialiased. In this paper we study the influence of high-level properties such as object material and speed, shutter time and antialiasing level. Based on scenes containing variations of these parameters, we design different psychophysical experiments to determine how influential they are in the perception of image quality.

This work gives insights on the effects these parameters have and exposes certain situations where motion blurred stimuli may be indistinguishable from a gold standard. As an immediate practical application, images of similar quality can be produced while the computing requirements are reduced.

Algorithmic efforts have traditionally been focused on finding new improved methods to alleviate sampling artifacts by steering computation to the most important dimensions of the rendering equation. Concurrently, rendering algorithms can take advantage of certain perceptual limits to simplify and optimize computations. To our knowledge, none of them have identified nor used these limits in the rendering of motion blur. This work can be considered a first step in that direction.

Categories and Subject Descriptors: I.3.7 [**Computer graphics**] Three-Dimensional Graphics and Realism

General Terms: Algorithms, Performance, Experimentation, Human Factors

Additional Key Words and Phrases: Motion blur, temporal antialiasing, perceptual rendering

ACM Reference Format:

Navarro, F. and Castillo, S. and Serón, F. and Gutierrez, D. 2011. Perceptual considerations for motion blur rendering. ACM Transactions on Applied Perception ACM Trans. Appl. Percept. 2, 3, Article 1 (May 2010), 15 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

Motion blur is an important cue for the perception of objects in motion. The phenomenon results from the light integration at the imaging device's sensor combined with object or camera motion. This generates a visible streak that follows the trajectory of the movement. It is naturally produced by both film and digital cameras and it has been leveraged as an artistic resort. It is also a frequent requirement for computer generated images where it needs to be explicitly simulated.

While the methods to produce synthetic motion blur have been studied and classified [Navarro et al. 2011], the perceptual aspects associated with the observation of motion blurred images have not received the same

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

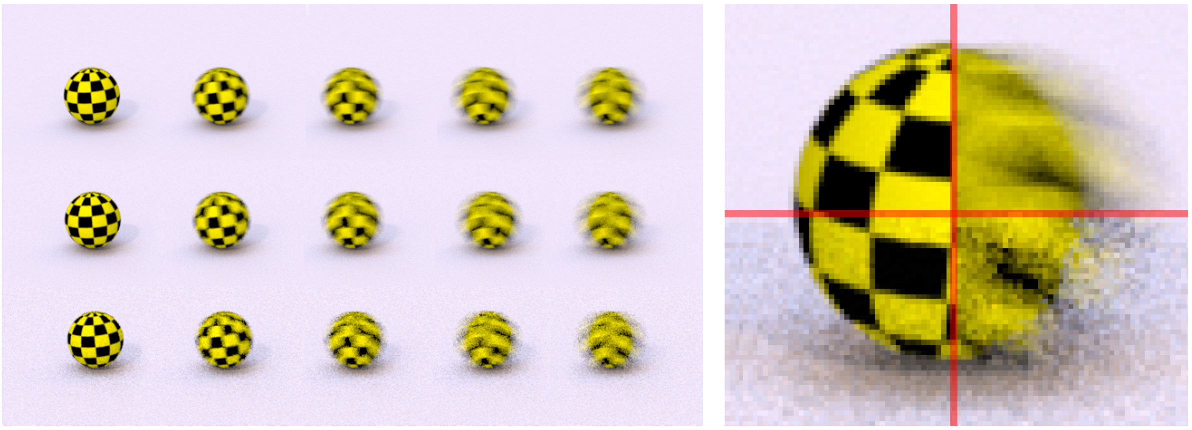

Fig. 1. In this work, we explore the implications of alternative scene properties and rendering parameters on the perception of motion blurred images. In the figure, a matrix showing a subset of the stimuli used in a series of psychophysical experiments: Columns, from left to right, use shutter times T1 (0% of the frame time, no motion blur), T2 (25%), T3 (50%), T4 (75%) and T5 (100%). Rows, from bottom to top, use antialiasing levels A1 (4 samples per pixel), A3 (256 samples) and A5 (512 samples). To the right, croppings of the images at the four corners show detailed views of the artifacts generated by the rendering process.

attention. Research has mostly been focused on determining the lower level capabilities of the HVS (human visual system). These works have relied on controlled studies where the stimuli are composed of gratings, gradients and moving dots. There is still a need to determine how these mechanisms work with imagery that is representative of the complexity of real world scenes. We believe this work is one of the first that exposes some insight in this area.

In the next sections we design a series of psychophysical experiments to determine how the HVS reacts to complex stimuli. This study also links the perceptual and practical algorithmic aspects of the problem. We have used path traced images of objects under the influence of complex lighting with the intention of associating standard rendering parameters such as shutter time and antialiasing level to their corresponding perceptual effects.

This problem is inherently multidimensional and the interplay of the different aspects involved is complex. However, this study is targeted to give answers to simple questions such as: When the scene is composed of simple materials, can a render be produced at a reduced cost without perceived quality loss? Do users detect the difference between alternative shutter times? How important are high sampling levels in the computation of motion blur?

The conclusions and recommendations contained in this work may prove useful for the production of photorealistic images. We hope that CG practitioners, render technical directors, shader writers and hobbyists will be able to apply them to minimize render times and reduce scene complexity without visual quality loss. These results may help settle some ground for the implementation of perceptually inspired motion blur rendering methods. This is especially interesting for the production of feature films and VFXs, where the cost of generating temporally antialiased images may be several times higher with respect to their non-motion-blurred equivalents. It is also important given that render times are usually optimized by following labor-intensive processes.

2. PREVIOUS WORK

The literature covering the perception of computer generated stimuli has identified different psychophysical trends. Recent examples include models for the psychophysics of photorealistic materials [Vangorp et al. 2007; Rushmeier 2008; Jimenez et al. 2009], natural illumination [Fleming et al. 2003; Čadík 2004; Křivánek et al. 2010; Lopez-Moreno et al. 2010], occlusions [Kozlowski and Kautz 2007; Yu et al. 2009], the ability to differentiate close variations of the stimuli [Ramasubramanian et al. 1999; McNamara 2006] and has also described the influence of object variety [McDonnell et al. 2008; Ramanarayanan et al. 2008]. However, none of these focus on the specifics of motion blurred images nor time varying images.

Motion blur rendering has received an important degree of attention. Existing works have mostly focused on the algorithmic aspects of the problem and describe methods capable of performing temporal antialiasing on different frameworks. We refer the reader to the work of Navarro et al. which contains detailed classifications and extensive descriptions of the state-of-the art of the existing techniques [Navarro et al. 2011].

Even if not specific to computer rendered images, the perceptual aspects of temporally varying images have also been described [Adelson and Bergen 1985; Watson et al. 1986] and their corresponding sensitivity functions [Lange 1958; Robson 1966; Virsu et al. 1982] have been clearly determined. The work of Burr and colleagues is fundamental in this area, being the first ones to describe motion smear, identify the window of temporal integration and quantify the perception of motion blurred stimuli [Burr 1980; Burr et al. 1986]. Even if many of the psychophysical factors have been clearly determined, the literature covering how the HVS reacts to dynamic photorealistic stimuli is still scarce. Our study focuses on this aspect while exploring the implications of using Monte Carlo rendering algorithms.

Other works have focused on determining how visible certain features are when they are observed in a picture or a rendered image. As such, the corresponding saliency functions have been proposed [Anson et al. 2006; Longhurst et al. 2006] which, in some cases, are temporally aware [Zhai and Shah 2006; Wang and Li 2007; Vangorp et al. 2009]. Other studies have proposed alternative perceptual metrics that can determine how different two images are [Mantiuk et al. 2004; Wang et al. 2004; Ramanarayanan et al. 2007]. The temporal component has been considered notably in the context of video processing [wei Tang 2007; Xia et al. 2009] and quality assurance [Watson et al. 1999].

3. PERCEPTION OF MOTION BLUR

The algorithms associated with motion blur rendering will be briefly described in Section 4. A byproduct of these approaches is the introduction of certain amount of blur in the images. A different set of factors are also capable of changing the nature of the image by adding extra blur, giving higher importance to certain temporal frequencies or diverting attentional resources. Some of them are intrinsic to the HVS while others are produced by the rendering pipeline and display device.

The existence of a temporal window of integration produces an effect that is known as motion smear. During a certain time span, the light hitting the eye is accumulated. As a result, each spatial location contains an image that represents the evolution of the scene being observed and not a discrete snapshot of it. Depending on the stimuli and environmental conditions, the window of integration can span from a few milliseconds to several seconds [Watson et al. 1986].

Peripheral vision [Burr 1981] is specialized in the detection of moving objects. Even if it is limited to low spatial frequencies, it is tuned to detect high temporal frequencies. This increases the chances of detecting image flickering, an artifact that is frequently present in low quality rendered images. In contrast, wide fields of view use an alternative visual channel that is specialized in identifying objects. While it is adapted to prioritize high spatial frequencies, high temporal frequencies are processed with reduced efficiency. This, combined with faster angular speeds in the retinal image, may lead to increased perceived blurriness even if this extra blur is not present in the original image.

The opposite is also true: Images may look sharper than they are. Motion deblurring [Ramachandran et al. 1974; Burr and Morgan 1997] is triggered for moving objects and in some cases, perceived images may be less blurry than their static incarnations. This is thought to be a mechanism to increase the resolved amount of detail in the image of a moving target.

The detection of motion itself has been explained using several models that are based on space-time filtering. They focus on the idea of finding moving brightness patterns in the retinal image of an object [Adelson and Bergen 1985; Watson and Ahumada 1985]. Alternative explanations exist for second-order motion generated by variations in contrast and flicker which may be helpful in the presence of occlusions and transparency [Fleet and Langley 1994; Hegde et al. 2004]. Finally, the surrounding environment has also been identified as a cue to detect motion [Loomis and Nakayama 1973]. Our study has been designed to analyze perceived quality as the absence of rendering artifacts. As such, the stimuli exceed the minimum threshold to trigger motion detection, no object occlusions need to be resolved and a single object can be comfortably tracked by the eyes. These images do not make use of any depth of field, so this source of static image blur is also excluded from the study.

Display devices contribute additional blur to the image. With common technologies, such as LCD monitors, the color of each pixel is held for certain amount of time. In this case, blur is generated by the display's inertia to keep its current state and its relative inefficiency to instantaneous change between different values [Pan et al. 2005]. Existing models can determine whether a given edge will appear sharp or not [Watson 2006]. The discrete nature of the media produces hold-type blur. The imaging pipeline discretizes the original stimuli so when tracking an object, the HVS is forced to reconstruct the original image from a reduced set of independent snapshots. The lower the number, the bigger the chances the temporal window of integration produces a blurred image that differs from the one that would otherwise be produced from the initial stimuli. This issue has been dealt with using several methods [Feng 2006; Didyk et al. 2010]. Both types of blur are intrinsic of the technology and are always present to some degree.

According to Block's Law, for stimulus with similar spatial characteristics, detectability is determined by their energy [Gorea and Tyler 1986]. Taking into account temporal integration, longer stimulus of reduced intensity may seem equivalent to shorter ones with higher luminance. In the specific case of display devices, several frames are fused together to produce a single unique perception that ultimately is responsible of the perception of continuous motion.

In summary, the perception of moving objects and motion blur is a complex process. Several perceptual mechanisms can interfere with the original images. Apart from the human factors, there is a group of technologically related issues that further complicate any predictions. In the following sections we will concentrate on determining the final appearance of rendered stimuli based on a series of scene and rendering parameters. In some cases, we will see that the explanation for certain results comes hand in hand with the phenomena we just described.

4. ALGORITHMIC MODEL

The formulations of Sung et al. [Sung et al. 2002] and Meredith [Meredith-Jones 2000] explain the generation of motion blurred CG images with the following expression:

$$I_{xy} = \sum_{l} \int_{\Omega} \int_{\Delta T} r(\omega, t) g_l(\omega, t) L_l(\omega, t) dt d\omega \tag{1}$$

In this equation, I_{xy} represents the contents of the image pixel with coordinates (x, y) and Ω is its corresponding subtended solid angle. Independently of their geometrical representation, the overall contribution of all primitives in the scene is considered by iterating over each individual object l. $g_l(\omega, t)$ is a geometrical term that accounts for occlusions between objects. Its value is 1 if object l is directly visible in the direction ω, 0 otherwise. Shutter shape and efficiency, lens aberrations and film type can influence the final image. The reconstruction filter $r(\omega, t)$ accounts for their overall effect. $L_l(\omega, t)$ represents the radiance of object l without explicitly establishing the method by which is calculated.

In order to account for the complex spatio-temporal relationships taking place in an animated scene, all terms are evaluated at an instantaneous time t over the aperture time ΔT, and over the solid angle Ω. In

Table I. Dimensions and values considered in the study. Between parenthesis, the short identifier that is used to refer to it.

Dimension	Values
Material	Flat (M1), checker (M2), noise (M3)
Object speed	Slow (S1), medium (S2), fast (S3)
Shutter time (%)	0 (T1), 25 (T2), 50 (T3), 75 (T4), 100 (T5)
Antialiasing level (samples)	4 (A1), 128 (A2), 256 (A3), 384 (A4), 512 (A5)

some cases and depending on the desired filter footprint, Ω can represent narrower or wider solid angles than the one defined by the pixel.

$$I_{xy} \approx \frac{1}{N_j} \sum_{j}^{N_j} i_d(\omega_j, t_j) \quad (2)$$

$$i_d(\omega_j, t_j) = \sum_{l} r(\omega_j, t_j) g_l(\omega_j, t_j) L_l(\omega_j, t_j) \quad (3)$$

Equation 1 represents the general case for any motion blur algorithm. As far as this study is concerned, rendered images are generated using distribution ray tracing. Equation 2 approximates Equation 1 taking into account the sampled nature of this method. Each image pixel is calculated as a sum of N_j discrete point samples. For simplicity, we will consider that motion blur is based on sampling locations in the spatio-temporal domain only. A sample $i_d(\omega_j, t_j)$ accounts for the contribution of each object l as seen in the direction ω_j at an instantaneous time t_j. This value can be calculated by different means, but in general it will be the result of evaluating the visibility g_l and radiance L_l functions for each of them. These contributions are weighted with the value determined by the filter $r(\omega_j, t_j)$. The resulting samples can be simply added together and averaged, or alternatively more sophisticated weighting methods can be applied.

5. DESCRIPTION OF THE STIMULI

The tasks associated with the perception and recognition of motion-blurred objects can potentially be affected by many elements. The geometry, material and animation of the objects, scene illumination, number of objects in the scene, complexity of the visibility changes, the render settings and even viewing conditions can have an impact on both the resulting images and how they are perceived. Among these factors, we have selected a subset that, while making the scope of the study tractable, still leaves a range of cases that is rich enough. Our intention is focusing on the issues associated with scene and rendering technology and, as such, we will analyze object material, object motion, shutter aperture time and antialiasing level. The first two are fully determined by the geometry and animation of the scene. The rest, far from being specific of a certain framework, are parameters are representative of a wide range of rendering algorithms.

We must note that the workflows for VFX and feature film pipelines have heavily inspired the selection of the dimensions and the design of the tests. In these environments, rendering technical directors receive scenes whose materials and object animations have been polished by the look development and animation departments and accepted by the director of art. In some cases, shutter settings are also determined by artistic decisions or by the need to match a real camera. As such, even if we will determine the influence of all the dimensions, we will insist on the algorithm related ones as they are the subset that can be freely modified.

We have considered a number of values for each of the four dimensions of the study which are also listed in Table I:

Object material: Three physically plausible material variations with different frequency contents. The intention is evaluating the responses of the participants using shading functions with alternative complexities.

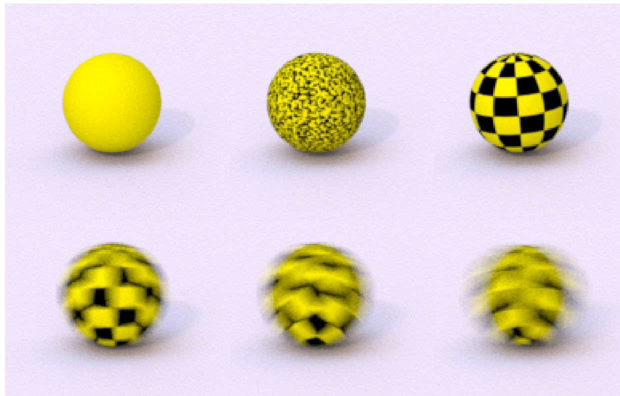

Fig. 2. Materials and object speeds. Top row: Non motion blurred image rendered with the materials used in the study, from left to right: flat (M1), noise (M3), checker (M2); bottom row, from left to right: slow (S1), medium (S2) and fast animation speeds (S3) rendered using the longest shutter setting (T5).

Figure 2, top row, shows the cases considered: A checker pattern containing abrupt changes in color, a noise pattern containing a moderate amount of high frequencies and small details and a simple flat color material. All of them use an approximation of a lambertian surface. Apart from the perceptual implications, the higher the frequencies contained in the pattern, the more challenging it is to antialias, and the higher the chances for the algorithm to produce noisy and strobing artifacts.

Object speed: Given the camera settings and viewing conditions, we have considered three object speeds that are equivalent to 5, 10 and 15 degrees of visual angle per second. These values, shown in Figure 2, bottom row, provide enough variation to produce images where the object can be easily recognized to other where the sphere is almost reduced to the motion trail.

Shutter times: Five shutter times simulating the fraction of the frame when the camera accepts incoming light. In Equation 1, this parameter has been represented as ΔT. We have considered five values: instantaneous exposure with no motion blur as well as percentages of 25, 50, 75 and 100% of the frame time. Apart from the aesthetic and perceptual considerations associated to the amount of blur, longer values imply an expanded strata and sparser sampling for the same number of samples. Although more accurate models can be considered [Stephenson 2007], temporal samples are combined using a box filter. Example images using different settings are displayed in Figure 1, left to right.

Antialiasing level: Five different antialiasing levels, N_j in Equation 2, representing the amount of samples that are averaged per pixel. The number of rays required has been determined empirically so they generate images ranging from very noisy to noiseless. Intermediate values have been chosen to cover this range from 4 to 512 samples using roughly equidistant steps. The average render times range between 15 seconds to 2 minutes per frame and roughly follow a linear progression. The highest level, that we consider the gold standard, medium and lowest levels are displayed from top to bottom, in Figure 1.

Given that temporal integration tends to soften the details in the images, we have decided to exclude the influence of the objects' geometry. Complex object occlusions have also been avoided. As such, this study uses a generic scene showing a sphere of approximately 20cm of diameter rolling over a flat surface. The trajectory is linear at constant speed and parallel to the camera. The camera is static and is placed 3 meters away from the object and 50 cm over the floor and uses the equivalent of a 50 mm lens.

In order to improve material discrimination, an HDR map provides natural illumination for all the objects in the scene [Fleming et al. 2003]. Arnold engine [Křivánek et al. 2010] has been used to produce the rendered images. Its unbiased Monte Carlo radiosity algorithm produces effects such as photorealistic soft shadowing

Table II. Psychophysical tests completed by the participants in the study.

Test	Type	Dimensions
#1 - Broad comparison	2AFC	All (2 each)
#2 - Shutter time / Antialiasing level	Matching/2AFC	Shutter time (5) / Sampling level (5)
#3 - Shutter time and antialiasing level	2AFC	Shutter time (3), sampling level (3)

Note: The "Dimensions" column lists the dimensions considered by each test, with the number of values used for each axis between parentheses. Any dimensions not listed are sampled at a single value.

and color bleeding. A template scene has been modified to use each possible combination of the previous values, resulting in a total of 225 rendered sequences each 4 seconds long. The images were rendered at a resolution of 640x360 pixels and played at 30 frames per second. The accompanying video shows a subset of the sequences generated for the study.

6. PERCEPTUAL TESTS

Our study comprises a series of psychophysical experiments whose targets are: first, finding how influential each of the dimensions described in the previous section are; second, determining the perceptual trends within each dimension; third: identifying the interplay between different dimensions and how they affect what a human subject sees.

We have designed several tests, based on 2AFC (two alternate forced choice) and matching tasks. Initial tests determine the most generic phenomena and the values that are more representative of each dimension. Further tests build upon them and provide refined results. They are summarized in table II.

Twenty-one participants, sixteen males and five females, aged between 25 and 35, completed the tests. Nine of them had experience in computer graphics and although they were aware of the overall purpose of the study they ignored the details of it. The remaining twelve were naive observers without formal knowledge of computer graphics but exposure to mainstream graphics technology. All of them self-reported normal or corrected to normal vision. The sequences where displayed in similar dim lighting conditions using a LCD monitor 17.3" diagonal, 1600x900x60Hz, sRGB color space, contrast 1000:1, luminosity 300 cd/m^2, gamma 2.2. Participants were placed at roughly 60 cm of the screen and each of the sequences subtended 14.4 degrees of visual angle.

6.1 Test #1: Broad comparison

Given the lack of previous results using photorealistic motion blurred stimuli, the first task is targeted to identifying and gaining an initial knowledge of the most salient trends. As such, we simultaneously consider all the dimensions described in section 5.

In this test, participants are asked to select one out of two videos according to the question *Which sequence has higher quality?* Before starting the task, every subject was briefed about the most salient image features associated with high quality media. Different example sequences containing noticeable levels of noise, flickering patterns and excessive blurriness were used as a method to identify each feature. In order to ensure these concepts were clearly understood, each participant was requested to perform the same task using a pair comparing the most extreme cases for each algorithmic axis as well a sample of different values for the other two dimensions. This process was performed before each test and the answers used for the training were not taken into account in the final results.

Being a preliminary test, each axis is represented by two values that produce visually different renders without making the decisions trivially obvious. These values are: checker (M2) and noise materials (M3), medium (S2) and fast (S3) speeds, medium (T3) and long (T5) shutter times and medium (A3) and high (A5) antialiasing levels. For each trial, the two videos are displayed in sequence one after the other and subjects are allowed to repeat them as many times as required. The pairs are configured so all combinations from a pool of 16 videos are used.

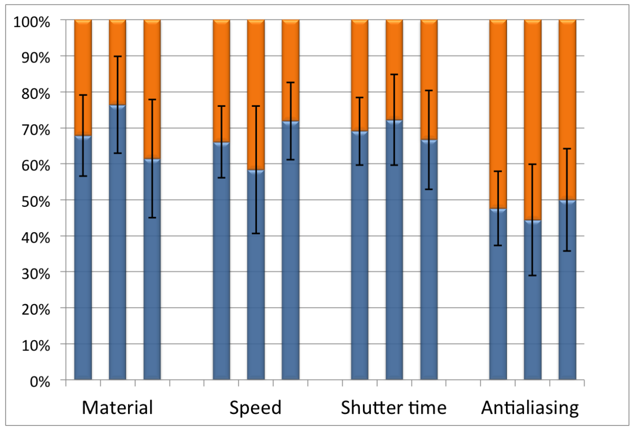

Fig. 3. Percentage of trials where a particular animation is perceived with higher quality. Blue/orange bars represent the corresponding values for noise/checker material (M3/M2), medium/high speed (S2/S3), medium/long shutter time (T3/T5) and medium/high antialiasing level (A3/A5). Each group of bars shows from left to right, the results for all, experienced and naive users, respectively. Error bars represent the standard error.

The results show that users perceive higher quality in the renders containing the noise material, medium speed and medium shutter time. A certainty of 95% has been confirmed using a Chi-square test. The answers, represented in Figure 3, have different qualitative interpretations:

—The preference for the noise material (67%) seems to be related to the fact that objects whose shading looks too clean or very regular are perceived as unrealistic. Also, the checker material, while not affected by sampling noise, produces more stroboscopic artifacts at every antialiasing level. These results are consistent both for medium and high sampling levels.

—Users manifest a preference for the medium speed in 66% of the situations. The reasons reported were excessive blur, lack of detail and a noticeable level of noise in the images using the fastest speed.

—Medium shutter times are chosen in 70% of the cases as longer values produce slightly unnatural images due to excessive softness.

—More significative is the absence of a clear preference for any of the antialiasing levels (the highest quality setting has received just 5% more votes). However using a visual differences predictor operator [Mantiuk et al. 2005], the images of individual frames are reported as perceptually different. This suggests that while an accepted model can detect the differences on static images, the HVS has difficulties identifying the same when the temporal dimension is involved.

Segmenting the results according to the users' knowledge of computer graphics has not shown any statistically significative differences. That said, experienced users took notably longer to complete the test with an average of 11:23 minutes versus 8:28 minutes. We attribute this to this prior knowledge preconditioned them to ensure the answer was the correct one which, given their subtle differences, was in some cases hard to find.

In practical terms the previous results support the idea that certain types of renders can be optimized without visible perceptual degradation, that is, scenes rendered with reasonable levels of quality may be indistinguishable from a gold standard even if the later may have been calculated using more computational resources. This insight will be refined in Sections 6.2 and 6.3 and gives us a good foundation for the remaining tasks of the study.

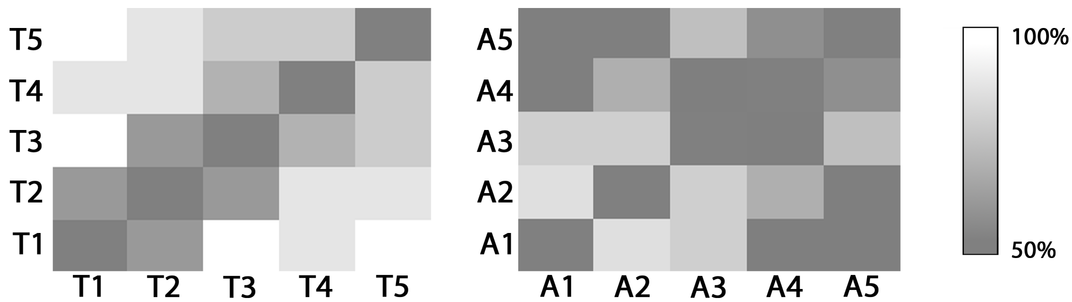

Fig. 4. Answers to the shutter speed (left) and antialiasing level tasks (right). The color in cell (i, j) represents the percentage of answers that were correct when the matching task involved stimuli i and j with darker colors meaning less accurate responses. All users responses are averaged and normalized to a percentage so, they are equivalent to an imaginary user providing all the answers.

6.2 Test #2: Tolerance to variations in shutter time and antialiasing level

The next study focuses on the parameters that can be considered specific of the rendering algorithm: shutter time and antialiasing level. These are the only dimensions that someone involved in the most practical aspects of the rendering pipeline will be allowed to have a direct impact on. In this section we will determine how efficient observers are when two different stimuli need to be matched, or alternatively, in which cases two sequences cannot be differentiated even if they have been rendered with distinct settings.

The test is composed of two matching tasks. Each task fixes the values for three of the dimensions and use the settings that a majority of users preferred in the previous test. The free dimension, shutter time for the first task and antialiasing level for the second, uses the five values available. All possible combinations of pairs of videos are used. One of them is randomly used as a reference and shown at the top of the screen. Both videos are then played at the bottom of the screen and the user is requested to select which one matches the reference. Videos are played in sequence and can be repeated as many times as required. Sixteen participants completed each task.

Shutter times can be accurately matched in 86.25% of the cases. An overall view of the answers is represented in Figure 4, left, with brighter areas for those values with the most accurate answers. For all the pairs comparing the most differing stimuli the matching is done with high success rate. Lower accuracy levels are obtained in those pairs that look more similar so darker regions surround the diagonal. The high number of incorrect matches that is found when comparing pairs using T1-T2-T3 suggests that the strobing artifacts present in these sequences are so objectionable than in practice they are visually equivalent. Values T4 and T5 seems to be distinguishable. We attribute it to the excessive blur that was already detected in Section 6.1 is being used as an extra cue.

Matching sampling levels seems to be a more challenging task in which the percentage of success drops to 63.75%. We have also observed an increment in the effort required to complete it (an average of 7:49 minutes compared to 6:12 for the previous task). In Figure 4, right, the most accurate answers are concentrated in the lowest sampling levels A1, A2 and A3 where the artifacts are more visible. Comparisons between higher sampling levels, A4 and A5, produce percentages near chance. For these values, the videos can be considered indistinguishable from a gold standard. This suggests the HVS is forced to work near its perceptual limits and supports the intuition that given a certain level of quality, it becomes less interesting to make further improvements as they become increasingly less noticeable. A lack of accuracy can be observed in those pairs comparing really low and high sampling levels. We have attributed this singularity to the complexity associated with the task itself. Users informally confirmed that in the shutter time task, the amount of blur and artifacts associated with each level where visible enough to allow an accurate matching. However, we believe that for this task, the lack of strong cues and the importance of visual memory and retention may be factors that need to be considered as a source of bias [Desimone et al. 1995; Treisman and Sato 1990]. This may be the reason behind the reduced accuracy exhibited throughout the test.

In order to discount the previous effects, we performed an alternative 2AFC subtest with the five sampling levels. Users were requested to select the stimulus they preferred from a pair of alternative sequences. The selection criteria was the perceived quality under similar conditions explained in Section 5. Ten users completed a task composed of twenty pairs of sequences, twice for each possible pairing.

Based on their answers, we perform a significance test following the approach described in previous works [Gutierrez et al. 2008; Rubinstein et al. 2010]. With their method, answers can be clustered in statistically indistinguishable groups based on an analysis of score differences. Two groups can be considered similar when the difference between their scores R is below $\lceil R_c \rceil$. Using a significance level α of 0.01, we compute R_c so that $P(R \geq \lceil R_c \rceil) \leq \alpha$ [Setyawan and Lagendijk 2004]. It can be shown [David 1963] that R_c can be obtained from $P(W_{t,\alpha} \geq (2R_c - 0.5)/\sqrt{nt})$, with n the number of participants and t the number of samples we compare (20 and 5 in our case). Values of $W_{t,\alpha} \approx 4.604166667$ and $\lceil R_c \rceil = 24$ can be retrieved from Pearson and Hartley's tables [Pearson and Hartley 1966].

Results show that users consistently prefer high over low sampling levels: A5 was selected over any other level in 59 cases; A4 in 51; A3 in 48; A2 in 34 and A1 in 8. Three groups can be built based on these values: the most voted containing A5, A4 and A3; second most voted grouping A4, A3 and A2; and a final group containing A1. These groups contain stimulus that are perceptually indistinguishable even if users are capable of finding differences between each group. Lower sampling is unequivocally associated with lower quality levels, whereas other stimuli are not so clearly differentiated. In general, these results confirm the capability of the HVS to detect alternative quality levels, and also confirm the suspicion that other factors related to the complexity of the task may affect the performance in the previous matching test.

In general, CG-aware participants show a slightly higher accuracy in their answers: 11% when shutter time is considered and 4% for the sampling level task. Since the videos are not played simultaneously, visual memory and other acquired abilities may be influencing factors. The ultimate intention is determining how effectively spectators perceive motion blurred stimuli and in which cases the lack of quality is noticeable. In the most frequent situations, a single sequence will usually displayed. The previous tasks can be considered a compromise between this visual experience and the case where several videos are played simultaneously.

With independence with respect to the HVS' capabilities to detect such subtle changes in the individual images of a sequence, there are other factors that we have consciously left outside this study. There are many studies that support the influence of phenomena such as inattentional blindness [Mack and Rock 1998], the attentional spread associated with multiple object tracking [Pylyshyn and Storm 1988], change blindness [Rensink et al. 1997] and attentional blink [Shapiro et al. 1997]. On the other hand, there is evidence that, in situations where a single element receives all the attentional resources, perception of contrast and spatial frequency is enhanced [Liu et al. 2009; Abrams et al. 2010]. As such, the participants in this study have completed the test under conditions where any issues associated to the quality of the images should be easily detected. We believe that under less constrained viewing conditions, the detection mechanisms involved in the detection of those artifacts may probably fall shorter and more aggressive simplifications could be possible.

6.3 Test #3: Interactions between shutter time and sampling level

The objective of this test is uncovering the existence of any interactions between the shutter time and antialiasing dimensions that may affect the performance of the observers to differentiate between different stimuli. Monte Carlo algorithms, namely those inspired on distribution ray tracing, consider the strata defined for all dimensions of the rendering equation simultaneously. Thus the more complex the geometry and shading functions are, the higher the chances that sampling methods will require further computation to

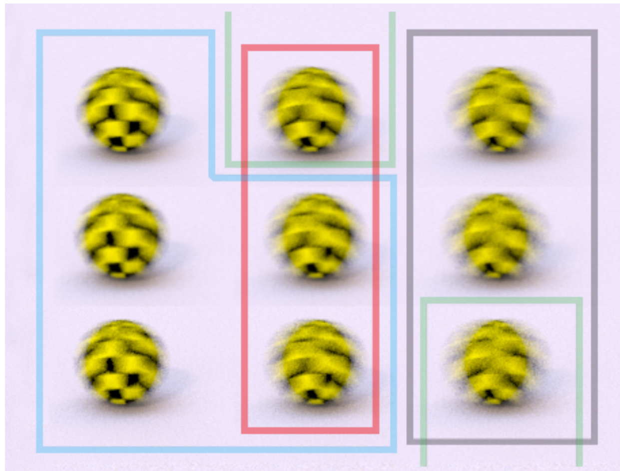

Fig. 5. Grouping of the scenes rendered with different antialiasing and shutter time settings. From left to right: Shutter time T3, T4, T5; from bottom to top: Sampling rates A3, A4, A5. Group1 to Group4, most to least voted, have been highlighted using blue, red, green and gray respectively.

produce an accurate approximation and avoid aliasing. Using any rendering engine will immediately reveal these underlying correlations. This section will help support this interpretation based on the results of a new psychophysical test.

In this test, participants have been requested to complete a 2AFC routine. In this case, the checker material and medium speeds have been fixed in accordance with the preferences manifested in previous tests. Shutter time take the values T3, T4, T5 and antialiasing level use A3, A4 and A5. Observers are requested to select the sequence with the highest quality from each pair taken from a set of 36 tuples containing all possible combinations of the two free parameters. The rest of the conditions and associated explanations are similar as with the previous tests.

Circular triades have been searched as a method of detecting inconsistencies (i.e. users manifesting preferences that result in contradictions: sequence X_1 is preferred over X_2, X_2 over X_3 but also X_3 over X_1). The coefficient of consistency ξ [Kendall and Babington-Smith 1940] is found to be 0.6492, which suggest an relatively low number of inconsistencies. The coefficient of agreement u, in a theoretical range of $[-1/21, 1]$, is 0.2794, that is, a moderate agreement among users.

The results have been analyzed using a significance test similar to the one detailed in the previous section. In this case, we set the significance level α to 0.01 with n the number of participants and t the number of samples equal to 21 and 9. We compute $W_{t,\alpha} \approx 5.0769231$ and $\lceil R_c \rceil = 36$.

Figure 5 shows the resulting groups. Group1 to Group4 have been ordered in decreasing number of votes. Group1, in blue, packages all the images calculated with the shortest shutter time T3 and those computed using T4 and the highest sampling levels. Group2 and Group4, in red and gray, contain all the sequences that have been rendered using shutter settings T4 and T5 respectively. From this partitioning, we can conclude shutter time drives users' answers with the antialiasing level taking a secondary role. The implications are important and twofold: first, all the renders computed with any of the settings belonging to a given group are perceptually equivalent to any other render in the same group; second, the most expensive renders can be optimized by just selecting the parameters that belong to the same group that require the least computational resources. In terms of ordering, each sequence has been preferred over any other according to the following number of votes: A4T3, 125; A3T3, 124; A5T3, 119; A5T4, 96; A4T4, 90; A3T4, 78; A5T5, 46, A4T5, 40, A3T5, 38.

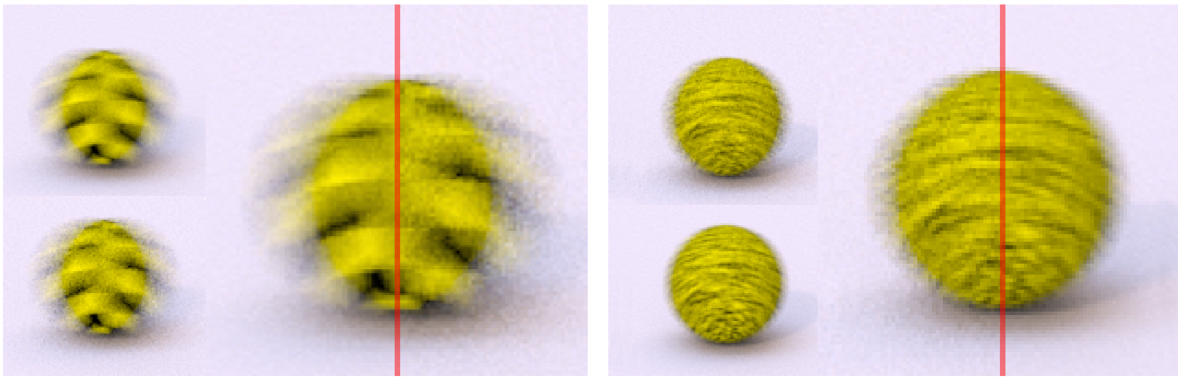

Fig. 6. Two pairs of perceptually equivalent renders. Left images are computed with alternative sampling settings (bottom: T5 and A3; upper: T5 and A5), while right images where rendered using alternative sampling and shutter time settings (bottom: T3 and A5; upper: T4 and A3). In both cases the image with lower sampling rate took half the time to render.

All the previous confirms our first findings, suggest the interplay between different dimensions can be useful in the presence of perceptual limits and opens interesting possibilities to reduce computing requirements.

7. CONCLUSIONS AND FUTURE DIRECTIONS

In this work, we have expanded the existing knowledge on the perception of motion blurred images to those situations where the stimuli is based on simple photorealistic images. We have found evidence that psychophysical limits in line with the ones found with simpler stimuli are applicable and take an important role on perceptual tasks. We believe there is also a spread of the attentional resources to complete more demanding tasks, as such; the tolerance to noise and high frequency artifacts is notable.

This insight comes from different psychophysical experiments that are based on high level properties of the 3D scenes and parameters of the render algorithm. These are object material, object speed, the shutter time simulated by the virtual camera and the antialiasing levels applied by the algorithm. The results suggest that, in some cases, images can be produced using aggressive simplifications without degradation of the perceived quality. In a practical situation, this can result in a drastic reduction of the render times and the computational resources required. As an example, Figure 6 shows two pairs of images, each one using the same object rendered using alternative settings. In each pair, the image that was rendered with the highest sampling level took double the time to render. Even if these images are notably different when compared side by side, they are perceptually indistinguishable when played at the viewing conditions considered in the study.

Even if this represents just a starting point, we have provided clear directions on the most influential elements and the interplay between the different aspects involved in the rendering of temporally varying scenes. These observations can be directly applied on demanding fields such as those related to VFX and film production.

There is still work that needs to be completed. In the future, we would like to provide more detail for the material and object speed dimensions. The results can also be extended to a wider set of scenarios covered by the rendering equation. Elements such as geometrical occlusion, evolving shading, level of detail or morphing geometry need to be explored as they are know to produce a shift in the appearance of the objects. This may potentially lead to formal models explaining the existing perceptual limitations as well as an expansion of concepts such as visual equivalence [Ramanarayanan et al. 2007].

Finally, we believe we have delineated the first steps to reducing the computing requirements of motion blur rendering processes without altering the perceived quality of the results. Equivalent approaches have already been successfully implemented as part of several rendering algorithms where the time dimension is ignored. Using complex stimuli may also increment the chances for a variety of render settings to be perceived with similar degrees of quality. This is especially interesting for those values that may be less expensive to evaluate.

These are exciting directions that need to be explored as the benefits for the rendering of motion blurred images can be significative.

ACKNOWLEDGMENTS

The authors would like to thank the participants that took part in the study. We would like to express our gratitude to Adonis Stevenson and the anonymous reviewers for their suggestions and useful comments. All sequences used in this study have been rendered using Arnold licenses loaned by Solid Angle. This research has been funded by a Marie Curie grant from the Seventh Framework Programme (grant agreement no.: 251415), the Spanish Ministry of Science and Technology (TIN2010-21543) and the Gobierno de Aragón (projects OTRI 2009/0411 and CTPP05/09).

REFERENCES

Abrams, J., Barbot, A., and Carrasco, M. 2010. Voluntary attention increases perceived spatial frequency. *Atten Percept Psychophys 72,* 6, 1510–21.

Adelson, E. H. and Bergen, J. R. 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2,* 2 (Feb), 284–299.

Anson, O., Sundstedt, V., Gutierrez, D., and Chalmers, A. 2006. Efficient selective rendering of participating media. In *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*. APGV '06. ACM, New York, NY, USA, 135–142.

Burr, D. 1980. Motion smear. *Nature 284*, 164–165.

Burr, D. C. 1981. Temporal Summation of Moving Images by the Human Visual System. *Royal Society of London Proceedings Series B 211*, 321–339.

Burr, D. C. and Morgan, M. J. 1997. Motion Deblurring in Human Vision. *Royal Society of London Proceedings Series B 264*, 431–436.

Burr, D. C., Ross, J., and Morrone, M. C. 1986. Seeing objects in motion. *Proc. R. Soc. Lond., B, Biol. Sci. 227*, 249–265.

Čadík, M. 2004. Human perception and computer graphics. Tech. Rep. DC-PSR-2004-04, Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague.

David, H. 1963. *The method of paired comparisons*. Charles Griffin and Company.

Desimone, R., Chelazzi, L., Miller, E. K., and Duncan, J. 1995. *Neuronal mechanisms of visual attention*. Massachusetts Institute of Technology, Cambridge, MA, USA, 219–226.

Didyk, P., Eisemann, E., Ritschel, T., Myszkowski, K., and Seidel, H.-P. 2010. Perceptually motivated realtime temporal upsampling of 3D content for high refresh rate displays. *Computer Graphics Forum 29,* 2, 713–722.

Feng, X. 2006. Lcd motion blur analysis, perception, and reduction using synchronized backlight flashing. In *Society of PhotoOptical Instrumentation Engineers Conference Series*. Vol. 6057. 213–226.

Fleet, D. J. and Langley, K. 1994. Computational analysis of non-fourier motion. *Vision Research 34,* 22, 3057–3079.

Fleming, R. W., Dror, R. O., and Adelson, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision 3*, 347–368.

Gorea, A. and Tyler, C. W. 1986. New look at bloch's law for contrast. *J. Opt. Soc. Am. A 3,* 1 (Jan), 52–61.

Gutierrez, D., Seron, F. J., Lopez-Moreno, J., Sanchez, M. P., Fandos, J., and Reinhard, E. 2008. Depicting procedural caustics in single images. In *ACM SIGGRAPH Asia 2008 papers*. SIGGRAPH Asia '08. ACM, New York, NY, USA, 120:1–120:9.

Hegde, J., Albright, T., and Stoner, G. 2004. Second order motion conveys depth order information. *Journal of Vision 4,* 10, 1–1.

Jimenez, J., Sundstedt, V., and Gutierrez, D. 2009. Screen-space perceptual rendering of human skin. *ACM Transactions on Applied Perception 6,* 4, 23:1–23:15.

Kendall, M. G. and Babington-Smith, B. 1940. On the method of paired comparisons. *Biometrica 31,* 3/4, 324–345.

Kozlowski, O. and Kautz, J. 2007. Is accurate occlusion of glossy reflections necessary. In *Proceedings of the 4th symposium on Applied perception in graphics and visualization*. APGV '07. ACM, New York, NY, USA, 91–98.

Křivánek, J., Fajardo, M., Christensen, P. H., Tabellion, E., Bunnell, M., Larsson, D., Kaplanyan, A., Levy, B., and Zhang, R. H. 2010. Global illumination across industries. In *ACM SIGGRAPH Course Notes*.

Lange, H. D. 1958. Research into the dynamic nature of the human fovea cortex systems with intermittent and modulated light. i. attenuation characteristics with white and colored light. *J. Opt. Soc. Am. 48,* 11 (Nov), 777–783.

Liu, T., Abrams, J., and Carrasco, M. 2009. Voluntary attention enhances contrast appearance. *Psychol Sci*.

Longhurst, P., Debattista, K., and Chalmers, A. 2006. A gpu based saliency map for high-fidelity selective rendering. In *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*. AFRIGRAPH '06. ACM, New York, NY, USA, 21–29.

Loomis, J. M. and Nakayama, K. 1973. A velocity analogue of brightness contrast. *Perception 2,* 4, 425–428.

Lopez-Moreno, J., Sundstedt, V., Sangorrin, F., and Gutierrez, D. 2010. Measuring the perception of light inconsistencies. In *Symposium on Applied Perception in Graphics and Visualization (APGV)*. ACM Press.

Mack, A. and Rock, I. 1998. *Inattentional blindness.* MIT Press, Cambridge, MA.

Mantiuk, R., Daly, S., Myszkowski, K., and Seidel, H.-P. 2005. Predicting visible differences in high dynamic range images - model and its calibration. In *Human Vision and Electronic Imaging X, IS&TSPIE's 17th Annual Symposium on Electronic Imaging (2005)*, B. E. Rogowitz, T. N. Pappas, and S. J. Daly, Eds. Vol. 5666. 204–214.

Mantiuk, R., Myszkowski, K., and Seidel, H.-P. 2004. Visible difference predicator for high dynamic range images. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*. 2763–2769.

McDonnell, R., Larkin, M., Dobbyn, S., Collins, S., and O'Sullivan, C. 2008. Clone attack! perception of crowd variety. *ACM Trans. Graph. 27*, 26:1–26:8.

McNamara, A. 2006. Exploring visual and automatic measures of perceptual fidelity in real and simulated imagery. *ACM Trans. Appl. Percept. 3*, 217–238.

Meredith-Jones, R. 2000. Point sampling algorithms for simulating motion blur. M.S. thesis, University of Toronto.

Navarro, F., Serón, F. J., and Gutierrez, D. 2011. Motion blur rendering: State of the art. *Computer Graphics Forum 30,* 1, 3–26.

Pan, H., fan Feng, X., and Daly, S. 2005. Lcd motion blur modeling and analysis. In *International Conference on Image Processing*. 21–24.

Pearson, E. and Hartley, H. O. 1966. *Biometrika Tables for Statisticians, 3rd ed.* Cambridge University Press.

Pylyshyn, Z. W. and Storm, R. W. 1988. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision 3,* 3, 179–197.

Ramachandran, V., Rao, V., and Vidyasagar, T. 1974. Sharpness constancy during movement perception (short note). *Perception 3,* 1, 97–98.

Ramanarayanan, G., Bala, K., and Ferwerda, J. A. 2008. Perception of complex aggregates. *ACM Trans. Graph. 27*, 60:1–60:10.

Ramanarayanan, G., Ferwerda, J., Walter, B., and Bala, K. 2007. Visual equivalence: towards a new standard for image fidelity. In *ACM SIGGRAPH 2007 papers*. SIGGRAPH '07. ACM, New York, NY, USA.

Ramasubramanian, M., Pattanaik, S. N., and Greenberg, D. P. 1999. A perceptually based physical error metric for realistic image synthesis. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. SIGGRAPH '99. 73–82.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science 8*, 368–373.

Robson, J. G. 1966. Spatial and temporal contrast-sensitivity functions of the visual system. *J. Opt. Soc. Am. 56,* 8 (Aug), 1141–1142.

Rubinstein, M., Gutierrez, D., Sorkine, O., and Shamir, A. 2010. A comparative study of image retargeting. In *ACM SIGGRAPH Asia 2010 papers*. SIGGRAPH ASIA '10. ACM, New York, NY, USA, 160:1–160:10.

Rushmeier, H. 2008. The perception of simulated materials. In *ACM SIGGRAPH 2008 classes*. SIGGRAPH '08. ACM, New York, NY, USA, 7:1–7:12.

Setyawan, I. and Lagendijk, R. L. 2004. L.: Human perception of geometric distortions in images. In *SPIE, Security, Steganography and Watermarking of Multimedia Contents*. Vol. IV. 272–280.

Shapiro, K., Raymond, J., and Arnell, K. 1997. The attentional blink. *Trends in Cognitive Sciences 1,* 8, 291 – 296.

Stephenson, I. 2007. Improving motion blur: Shutter efficiency and temporal sampling. *journal of graphics, gpu, and game tools 12,* 1, 9–15.

SUNG, K., PEARCE, A., AND WANG, C. 2002. Spatial-temporal antialiasing. *IEEE Transactions on Visualization and Computer Graphics 8,* 2, 144–153.

TREISMAN, A. AND SATO, S. 1990. Conjunction search revisited. In *Journal of Experimental Psychology: Human Perception and Performance*. 459–478.

VANGORP, P., CONDON, T., FERWERDA, J., BALA, K., SCHOUKENS, R., AND DUTRÉ, P. 2009. Visual equivalence in dynamic scenes. Tech. rep., Katholieke Universiteit Leuven.

VANGORP, P., LAURIJSSEN, J., AND DUTRÉ, P. 2007. The influence of shape on the perception of material reflectance. *ACM Trans. Graph. 26.*

VIRSU, V., ROVAMO, J., LAURINEN, P., AND NSNEN, R. 1982. Temporal contrast sensitivity and cortical magnification. *Vision Research 22,* 9, 1211 – 1217.

WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. 2004. Image quality assessment: From error measurement to structural similarity. *IEEE Transactions on Image Processing 13,* 1, 1–14.

WANG, Z. AND LI, Q. 2007. Video quality assessment using a statistical model of human visual speed perception. *J. Opt. Soc. Am. A 24,* 12 (Dec), B61–B69.

WATSON, A. B. 2006. The spatial standard observer: A human vision model for display inspection. *SID Symposium Digest of Technical Papers 37,* 1, 1312–1315.

WATSON, A. B. AND AHUMADA, A. J. 1985. Model of human visual-motion sensing. *J. Opt. Soc. Am. A 2,* 2 (Feb), 322–341.

WATSON, A. B., ALBERT J. AHUMADA, J., AND FARRELL, J. E. 1986. Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *J. Opt. Soc. Am. A 3,* 3 (Mar), 300–307.

WATSON, A. B., HU, Q. J., MCGOWAN, J. F., AND MULLIGAN, J. B. 1999. Design and performance of a digital video quality metric. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, B. E. Rogowitz & T. N. Pappas, Ed. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Series, vol. 3644. 168–174.

WEI TANG, C. 2007. Spatiotemporal visual considerations for video coding. *IEEE Transactions on Multimedia 9*, 231–238.

XIA, J., SHI, Y., TEUNISSEN, K., AND HEYNDERICKX, I. 2009. Perceivable artifacts in compressed video and their relation to video quality. *Image Commun. 24*, 548–556.

YU, I., COX, A., KIM, M. H., RITSCHEL, T., GROSCH, T., DACHSBACHER, C., AND KAUTZ, J. 2009. Perceptual influence of approximate visibility in indirect illumination. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*. APGV '09. ACM, New York, NY, USA.

ZHAI, Y. AND SHAH, M. 2006. Visual attention detection in video sequences using spatiotemporal cues. In *Proceedings of the 14th annual ACM international conference on Multimedia*. MULTIMEDIA '06. ACM, New York, NY, USA, 815–824.