PSYCHOVISUAL MEASUREMENT AND DISTORTION METRICS FOR IMAGE SEQUENCES

Edmund M. Yeh, Anil C. Kokaram, Nick G. Kingsbury
Department of Electrical Engineering and Computer Science,
MIT, Cambridge, MA 02139, USA
`emyeh@mit.edu`
Department of Engineering,
University of Cambridge, CB2 1PZ, UK
`ack@eng.cam.ac.uk, ngk@eng.cam.ac.uk`

ABSTRACT

A number of perceptual distortion measures have been developed for both still images and image sequences. Many of these measures, however, derive from visual experiments using stimuli quite unlike digital image coding artifacts. We present the results from six visual experiments characterizing human visual sensitivity to edge-like blocking artifacts under a number of spatiotemporal conditions, in an attempt to establish a firm experimental basis for a tuned perceptual distortion measure for blocking artifacts in image sequences.

1 INTRODUCTION

The relatively recent recognition that a perceptually-meaningful distortion measure is essential for the development of efficient and high quality image coding algorithms has led to a number of visual metrics for still images and image sequences [1, 2]. In spite of these encouraging developments, there seem to be two problems with some of the measures developed thus far. First, most of the measures are based purely on general (and incomplete) models of vision, and not specific to the particular kinds of distortions introduced by lossy coding algorithms. In particular, the stimuli used in general psychovisual experiments rarely occur as distortions resulting from image coding. Second, models derived from general psychovisual results usually measure visual sensivities to stimuli at the threshold level, where the HVS behaves essentially linearly. These results, however, often fail to hold at suprathreshold levels, where stimuli are easily detectable and the HVS behaves nonlinearly. In low-bit-rate image coding situations where distortions may be easily detectable, most of the presently available models would fail to apply.

The above problems were addressed in [3], where the authors develop a PDM for still images based on direct experimental measurements of visual sensitivities to blocking distortions resulting from popular transform coding schemes. The main measurement technique used there is the reaction time (RT) technique, whereby visual sensitivities are indicated by the amount of time taken to discriminate an artifact-free reference image (sequence) from a distorted image (sequence) containing an artifact. In the context of still images, the RT technique turns out to be uniquely flexible in its ability to measure sensitivities at both the threshold and suprathreshold levels.

In the present paper, we extend the paradigm of [3] to assess the perceptual significance of coding distortions in time-varying image sequences, with the aim of incorporating the experimental results in a perceptual distortion measure (PDM) for image sequences. Using the RT and other measurement techniques, we carry out a systematic, quantitative characterization of visual sensitivity to blocking artifacts in the context of various spatio-temporal environments commonly found in image sequences. It is found that the RT technique accurately characterizes visual sensitivities to artifacts in moving backgrounds at both the threshold and suprathreshold levels, except in the case where the durations of the artifact are extremely brief. The experimental findings have been fully interpreted, and incorporated into a PDM for image sequences, as described in [4, 5].

2 EXPERIMENTAL SETUP

A dark enclosure was constructed for the experimental setup. All images were displayed on a 256 grey scale monitor (Dell UltraScan Model 15FS-EN) operated at high resolution graphics mode (115) by a Stealth 64 DRAM graphics card (Diamond Computer Systems). The screen size was 800 pixels by 600 pixels while the vertical scan rate was 100 Hz non-interlaced. Subjects viewed images binocularly from a distance of five times the image height. The standard image size used for all psychovisual tests was 160 pixels by 160 pixels. For the display of image sequences, a frame rate of 25 frames per second was achieved by employing direct writing of data screen memory and a system-level timer with a resolution of 1 microsecond. For more details of the experimental setup, see [4].

Blocking artifacts produced by coarse quantization in transform coding schemes are characterized by sharp intensity variations near block boundaries which gradually decay away from the edge [3]. To simulate the blocking artifact, a standard test stimulus was constructed using a two-dimensional waveform with a step-change in the direction of the edge (high frequency) and a raised cosine variation in the direction perpendicular to the edge (low frequency). We refer to the magnitude of the step-change as the amplitude of the artifact.

3 EXPERIMENTAL RESULTS

3.1 Masking by Noise

This is the first of three investigations examining the visibility of blocking artifacts in the presence of noise. Whereas this experiment used still noise as image backgrounds, exper-

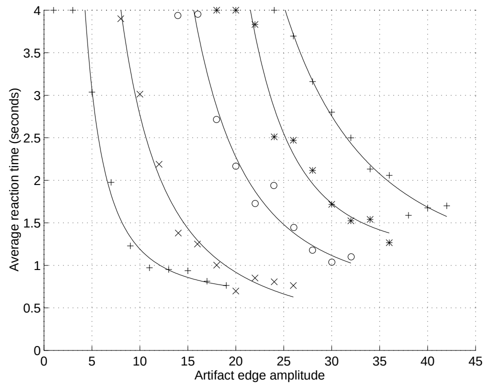

Figure 1: RT curves for Experiment 1, with noise standard deviations of (from left to right) 0(+), 30(×), 60(∘), 90(∗), 120(+).

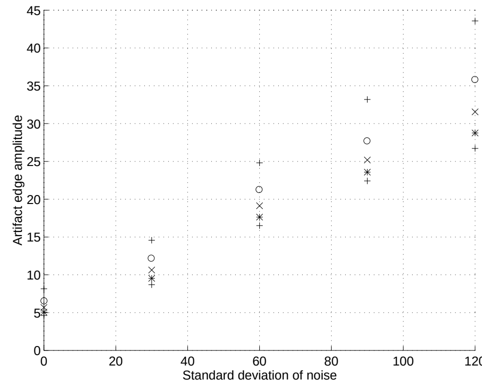

Figure 2: Constant time response plots for Experiment 1, evaluated at (from top to bottom) $t = 1.5(+), 2.0(\circ), 2.5(\times), 3.0(*)$, 3.5(+) seconds.

iments 3 and 4 investigated artifact visibility in the presence of moving noise.

The experiment used the standard blocking artifact described above as the stimulus and uniformly-distributed white noise for the still image backgrounds. The RT technique was used to measure visual sensitivities. During a typical comparison, the reference/distorted pair was presented simultaneously and side by side, with the spatial order of presentation randomly selected. The subject was asked to identify the side containing the artifact as quickly as possible by pressing a key within a maximum allowed time of four seconds. A small RT indicates high sensitivity (or artifact visibility), while a large RT signifies low sensitivity (visibility). To obtain the relationship between artifact amplitude and visibility for a fixed background, RT measurements were made at various artifact amplitudes to yield an RT curve of artifact amplitude versus response time. The influence of background parameters was then assessed by obtaining RT curves under various background conditions and studying the *constant time*, or *constant sensitivity* cross-sections. The contant time response (CTR) plots relate artifact amplitude and values of a background parameter at constant sensitivity.

Artifact visibility was measured by recording reaction times to blocking artifacts superimposed on the noise backgrounds at five standard deviations. Results of the experiment are summarized in in Figures 1 and 2. In Figure 1, each data point represents the reaction time for a fixed artifact and background noise variance, averaged across 28 independent measurements (seven individual responses from four subjects). An RT curve (reaction time as a function of artifact amplitude) was interpolated from the averaged data set for a fixed background noise variance.

The interpolated RT curves demonstrates the ability of the RT technique to measure visual sensitivities at both the threshold and suprathreshold levels. For instance, the upper-left-hand corner of the curve, relating low artifact amplitudes to large (close to maximal) reaction times, corresponds essentially to the *threshold* region of detection, where artifacts are barely visible. As the reaction time decreases with increasing artifact amplitude, the curve enters the *suprathreshold* region, where artifacts become more readily detectable. In this region, the curve is monotonically decreasing and is approximately linear. In the still lower stretches of the curve, the reaction time response quickly approaches an asymptotic level beyond which no further decreases in RT is possible. This part of the curve may be described as the *saturated* region.

Figure 1 shows that RT curves consistently shift rightward and upward as the standard deviation of the noise increases, indicating a general decrease in the visibility of the artifact with increasing noise levels. This trend is confirmed in Figure 2, where constant sensitivity points have been derived for five time cross-sections. The results of this experiment offer no surprises: increasing background noise amplitude translates into decreased visibility of the artifact stimulus. However, these artifact-specific results will prove instrumental in assessing the influence of *additional* background parameters such as movement speed and direction.

3.2 Masking by Background Edges

Sharp background gradients can dramatically decrease visibility of foreground objects. This experiment was designed to quantify the masking effect introduced by background edges on the visibility of blocking artifacts.

The RT technique was again employed here for measurements of artifact visibility. Both the standard blocking artifact stimulus and background edge were vertically-oriented and placed in the center of the image. The background edge was modelled as a raised cosine ramp with a fixed rise of 4.6 arcmin and variable amplitude. RT measurements were made for four background edge amplitudes of 0, 30, 90, and 150 gray levels. For brevity, we summarize the results using the CTR curves in Figure 3. The plot shows that in general, larger backround edge amplitudes lead to greater artifact stimulus values at constant sensitivity. That is, sensitivity generally decreases as the background amplitude levels increase.

3.3 Speed of Background Movement

Years of psychovisual research have pointed to the importance of motion perception as a fundamental sense whose properties cannot simply be derived from spatial and temporal sensitivities. In the perceptual evaluation of image sequences, the motion dimension is all the more paramount. As a first step in the characterization of the motion response to blocking artifacts, this experiment studied the visibility

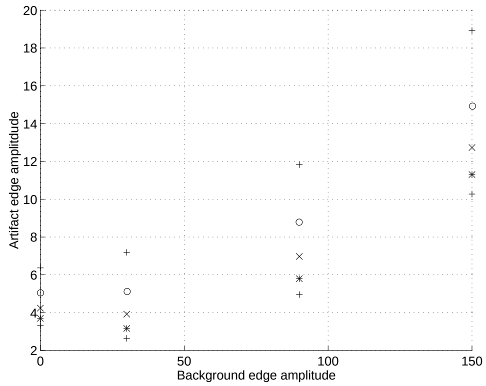

Figure 3: Constant time response plots for Experiment 2, evaluated at (from top to bottom) $t = 1.5(+), 2.0(\circ), 2.5(\times), 3.0(*), 3.5(+)$ seconds.

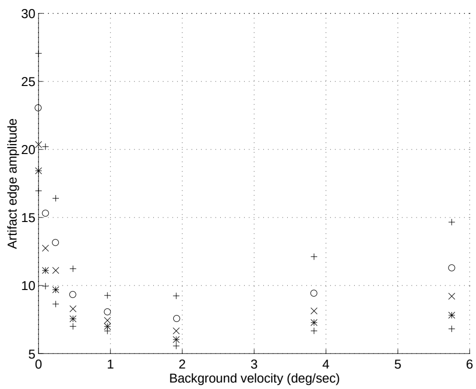

Figure 4: Constant time response plots for Experiment 3. From top to bottom, $t = 1.5(+), 2.0(\circ), 2.5(\times), 3.0(*), 3.5(+)$ seconds.

of artifacts as a function of the *speed* of moving noise backgrounds.

In roughly-quantized image sequences coded using transform schemes, blocking artifacts appear in fixed positions relative to backgrounds moving with different speeds and directions. These conditions were simulated by superimposing a stationary blocking artifact on top of uniformly-distributed white noise (random dot) backgrounds moving at a specified speed. Reaction times were measured for blocking artifacts superimposed on moving noise backgrounds of fixed standard deviation (60 levels) at speeds ranging from 0 to 5.75 deg/sec. The direction of movement ϕ was fixed at 0° from the horizontal (rightward movement). Visual stimuli were displayed at 25 frames/second. The results are summarized in the CTR plot of Figure 4, which shows a bandpass behavior for the visibility of the artifacts superimposed on moving backgrounds. The results indicate that visibility of blocking artifacts generally improve with noise background movement. The visibility increases steadily with speeds of movement at low speeds and approaches a maximum between 1.92 deg/sec and 3.83 deg/sec. For higher speeds, the visibility decreases significantly but stays higher than that for a still background.

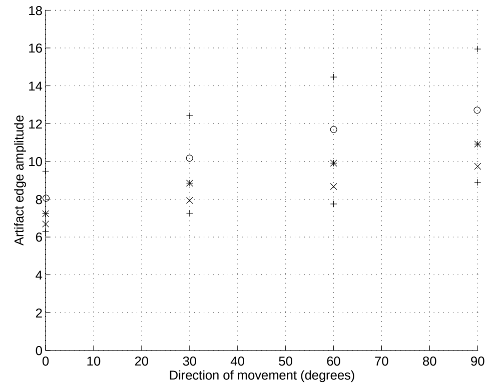

Figure 5: Constant time response curves for Experiment 4. From top to bottom, $t = 1.5(+), 2.0(\circ), 2.5(\times), 3.0(*), 3.5(+)$ seconds

3.4 Direction of Background Movement

Given that the blocking artifact has a specific spatial orientation, we can expect that its visibility in the presence of moving noise backgrounds would vary with not only the speed but also the direction of background movement. In this fourth experiment, the variation of artifact visibility with changing directions of background movement was examined.

RT measurements were made for vertically-oriented artifacts ($\theta = 90^\circ$) superimposed on white noise backgrounds moving at a constant speed $v = 1.92$ deg/sec but with varying directions of movement ϕ, measured from the horizontal (rightward movement). Results are displayed in Figure 5. The CTR plots show that artifact visibility decreases monotonically as the movement direction ϕ diverges from the horizontal and approaches the artifact orientation at $\theta = 90^\circ$. The experimental results have thus demonstrated significance of the movement direction in the evaluation of artifact visibility in moving backgrounds.

3.5 Temporal Summation

The development of a spatiotemporal visual model for blocking artifacts requires that the temporal response of the HVS be adequately simulated. This experiment was designed to quantify the so-called temporal summation effect in human vision by measuring the visual sensitivity to blocking artifacts as a function of the artifact's temporal duration.

The artifact stimuli used in this experiment were presented in brief bursts within a longer image sequence containing otherwise uniform backgrounds. Due to the very brief nature of the stimuli, continuous characterization of sensitivities from the threshold to the suprathreshold regions was no longer possible via with RT method. Subjects' reaction times to brief stimuli tend to be polarized to two extremes, with one region of very short RT indicating immediate detection, and another region of infinite RT indicating no detection at all. In the face of these difficulties, another widely-used psychovisual technique, the *method of constant stimuli* was used to estimate detection thresholds for blocking artifacts of various durations [4]. The estimation of each threshold involved measuring the number of detection responses at a fixed artifact amplitude and comparing the percentage of detections across different artifact amplitudes. After fitting the discrete data set with a continuous form of the *psychometric function*, thresholds were estimated by noting the artifact amplitude

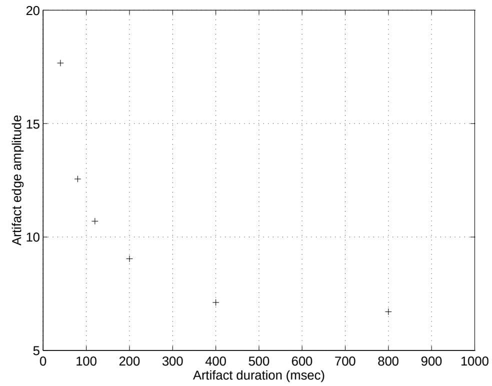

Figure 6: Visual thresholds for Experiment 5

corresponding to a 75% detection level.

Detection data and related psychometric curves were obtained for six temporal durations ranging from 40 to 800 msec. Artifact amplitudes corresponding to the 75% detection level were then determined for each temporal duration. These threshold amplitudes are shown in Figure 6, The results show that visual thresholds decrease monotonically as the temporal duration of the blocking artifact increases, with the rate of decrease gradually diminishing with increasing duration. This effect is a clear illustration of the lowpass temporal filtering behavior of the HVS, an effect that cannot be adequately modelled by purely spatial visual models applied successively in time.

3.6 Temporal Masking

Most image sequences are characterized by an abundance of temporal variations. When considering the visibility of artifacts in coded sequences, it is essential to quantify variations in the visibility of the artifact with global background changes (such as scene changes, panning) over time.

The experiment focused on the measurement of visual sensitivity to blocking artifact stimuli presented in the temporal vicinity of a sharp temporal change in the background luminance. The artifact had a brief duration of 40 msec and occurred a short time before or after the onset of a background luminance change from a uniform level L_1 to L_2. As in Experiment 5, artifact stimulus durations were too short to enable measurements of sensitivity using the RT method. For stability, the experiment used an adaptive *staircase procedure* which estimates the visual thresholds by adaptively adjusting the artifact stimulus amplitude according to the subjects' responses. The threshold determined correspond to a detection probability of 75% [4].

Threshold measurements were obtained for artifacts occurring at six temporal distances from the temporal background luminance transition. These measurements were then repeated for three separate magnitudes of $\Delta L = L_2 - L_1$. The results are presented in Figure 7.

Results for all three values of background luminance discontinuity indicate that artifact detection thresholds rise dramatically in the vicinity of the temporal discontinuity, both before (backward masking) and after (forward masking) the occurrence of the temporal discontinuity, and tapers off quickly as the distance from the temporal edge increases. We can make two observations. First, the threshold elevation

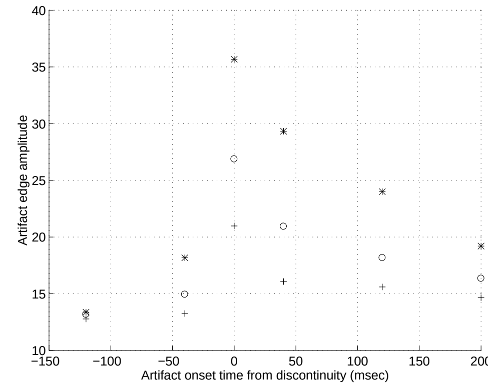

Figure 7: Visual thresholds for Experiment 6, with $\Delta L = 40(+), 80(\circ), 120(*)$. The background luminance change occurs at $t = 0$.

for test stimulus in the vicinity of the temporal discontinuity increases as the magnitude of the luminance change $\Delta L = L_2 - L_1$ increases. Second, the temporal duration in which threshold elevation of the test stimulus occurs increases as ΔL increases.

4 CONCLUSION

In this paper, six new visual experiments regarding HVS sensitivities to blocking artifacts have been presented. A number of psychovisual measurement techniques were used in obtaining the results. Visual sensitivity variations with spatial, temporal, and motion characteristics of both the test artifact and backgrounds were investigated. The reaction time technique was shown to be effectively way of quantifying visual sensitivities to artifacts under various spatiotemporal conditions, except in the case where artifact durations are extremely brief. The experimental results presented have been fully incorporated into a perceptual distortion measure for blocking artifacts in image sequences.

References

[1] S. Daly. "The visible differences predictor: an algorithm for the assessment of image fidelity," in *Human Vision, Visual Processing, and Digital Display III*, Vol. 1666, SPIE, 1992, pp.2-14.

[2] C. J. van den Branden Lambrecht, O. Verscheure. "Perceptual Quality Measure using a Spatio-temporal Model of the Human Visual System," in *Digital Video Compression: Algorithms and Technologies*, Vol. 2668, SPIE, 1996, pp. 450-461.

[3] S. A. Karunasekera, N. G. Kingsbury. "Distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Transactions on Image Processing*, Vol.4, No.6, June 1995, pp. 713 – 24.

[4] E. M. Yeh. *Perceptual Distortion Measure for Image Sequences*. M. Phil. Thesis, Department of Engineering, University of Cambridge, August 1995.

[5] E. M. Yeh, A. C. Kokaram, N. G. Kingsbury. "A Perceptual Distortion Measure for Edge-like Artifacts in Image Sequences," To appear in *Human Vision, Visual Processing, and Digital Display III*, SPIE, 1998.