Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering

Supplementary material

Gyorgy Denes* Kuba Maruszczyk† George Ash‡ Rafał K. Mantiuk§

University of Cambridge, UK

CONTENTS

The supplementary material contains a theoretical framework for analyzing the motion quality of TRM and the setup and results of Experiment 4, a side-by-side subjective validation experiment on a high-frame-rate LCD display, evaluating TRM against half and full frame rate.

1 ANALYSIS

In this section we provide a theoretical analysis of how the display and rendering parameters, such as refresh rate and reduction factor, affect the motion quality of TRM rendering.

1.1 Spatio-temporal visual difference model

The quality of video generated with TRM depends on (1) by how much the resolution of the odd-numbered frames is reduced, (2) on the display refresh rate, (3) on the velocity of motion, (4) display angular resolution, (5) display luminance range and (6) scene content (contrast). As it is impractical to explore all these dimensions in an experiment, we use a custom visual energy model to analyze the key dimensions and to find the best parameter range for the algorithm. The described model ignores some effects, such as contrast masking and motion sharpening, but it provides a simpler analysis and it well captures the worst-case scenario. The model is partly inspired by the pyramid of visibility [4] and a generalized contrast energy model [3], and is tailored to test all visual factors that are relevant for our application.

We focus on the quality of motion generated by TRM as compared to the motion produced on the same display when every frame is rendered at the full resolution. We analyze the motion in a simple animation, in which a square of 1 visual degree is moving horizontally with a constant speed. The luminance levels of the square and the background were selected to be equidistant from the half-luminance of the display in order to minimize under- and over-shoots. We also model imperfect SPEM and thus reduced sensitivity at high velocities. For better visualization, we consider a single row of pixels from such an animation as all rows containing the square are identical. Figure 1 visualizes the spatial and temporal dimension of the animation. For our analysis, we use the representation that compensates for the SPEM, shown on the right side of the figure. To estimate the visual difference between both renderings, we convert the difference into contrast:

$$C(x,t) = \frac{L_{TRM}(x,t) - L_{full}(x,t)}{L_{TRM}(x,t) + L_{full}(x,t)}, \tag{1}$$

*e-mail: gyorgy.denes@cl.cam.ac.uk
†e-mail: kuba.maruszczyk@cl.cam.ac.uk
‡e-mail: ga354@cl.cam.ac.uk
§e-mail: rafal.mantiuk@cl.cam.ac.uk

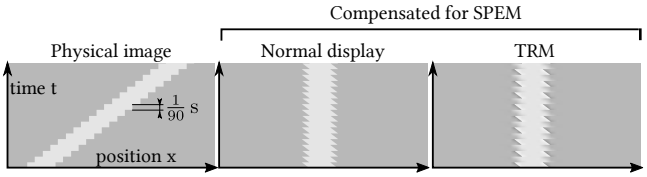

Figure 1: The images represent the spatial and temporal dimensions of an animation consisting of a single bar (box) moving left-to-right. Note that the animation is not smooth (is blocky) because of the finite resolution (30 ppd) and refresh-rate of the monitor (90 Hz). The left image represents the animation as seen with the stationary gaze, the two images on the right assume that the gaze perfectly follows the object. The choppy edge is caused by smooth-pursuit-eye-motion over the pixels that remain static for the duration of the frame, causing hold-type blur when the perceived image is integrated along temporal dimension. TRM can reduce hold-type blur.

where L_{TRM} is the luminance of the sequence rendered with TRM, L_{full} is the luminance for the standard rendering of all frames at the full resolution. To account for the spatio-temporal characteristic of the visual system, we modulate the contrast difference by the stCSF:

$$C_n(x,t) = \mathfrak{F}^{-1}\{\mathfrak{F}\{C\}(\rho,\nu)\cdot S_{st}(\rho,\nu+\rho\, v_r)\}, \tag{2}$$

where $\mathfrak{F}$ denotes the Fourier transform, ρ is spatial frequency in cycles per degree, ν is temporal frequency in Hz, and S_{st} is stCSF from [2]. Note that, since the sensitivity S_{st} is the inverse of the detection contrast, the equation above is equivalent to the normalization (division) by the detection contrast threshold. v_r is the difference between the velocity of the square (v_{sq}) and the velocity of the gaze motion (v_{eye}), accounting for the lag of SPEM (refer to Section 3 in the paper). We use the empirical formula proposed by Daly [1] to estimate the lag of gaze motion:

$$v_r = v_{sq} - v_{eye} = v_{sq} - min(g_{sp}\, v_{sq} + v_{min}, v_{max}), \tag{3}$$

where the gain of SPEM $g_{sp} = 0.82$, the minimum gaze velocity $v_{min} = 0.15$ and the maximum gaze velocity $v_{max} = 80$ deg/s. The term $\rho\, v_r$ expresses a relative object motion as a temporal frequency. A similar transformation is commonly used to transform spatio-velocity into spatio-temporal representation [2]. Finally, the stCSF-normalized contrast difference is pooled to find the visual difference energy:

$$E_{diff} = \left(\sum_{x=1}^{N} \sum_{t=1}^{K} |C_n(x,t)|^{\beta} \right)^{\frac{1}{\beta}}, \tag{4}$$

where N is the number of pixels and K the number of frames. The parameter β controls the efficiency of spatio-temporal integration. We use the value of $\beta = 2.6$, which was shown to provide a good fit

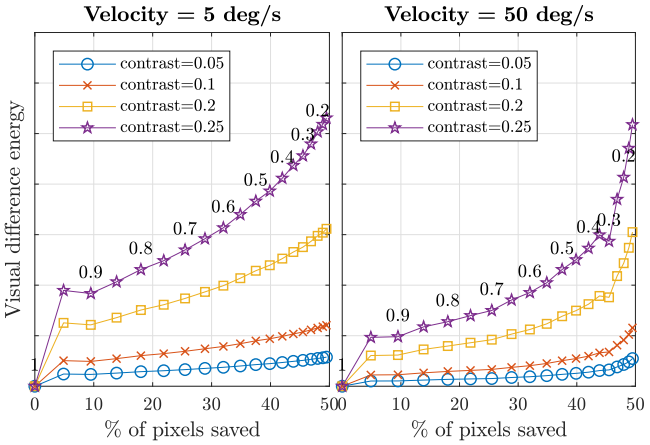

Figure 2: The visual error of TRM for varying contrast is plotted as a function of reduction of the number of rendered pixels. The error is expressed as a visual difference between rendering every frame and TRM. The black numbers above the data points indicate the resolution reduction factor (0.5 corresponds to TRM ½)

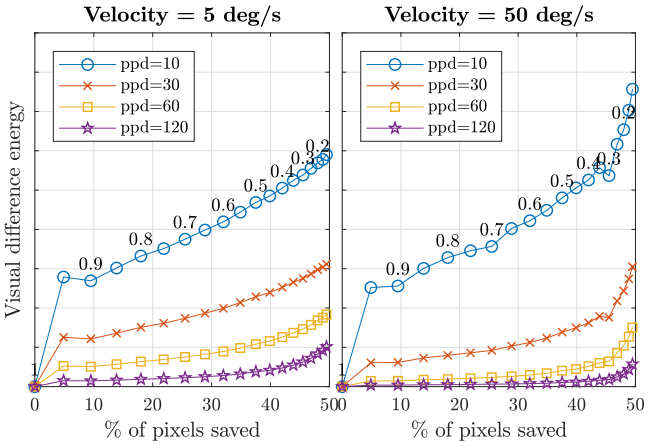

Figure 3: The visual error of TRM for varying angular resolution in pixels per visual degree is plotted as a function of reduction of the number of rendered pixels. Notation identical to Figure 2.

to Modelfest dataset [3]. The resulting energy of visual difference is directly related to the probability of detecting a visual difference. However, the exact mapping from the energy to the probability of detection (psychometric function) is unknown as it depends on the content, observer and many other factors. Our goal here is to compare relative visibility for different parameters of TRM rendering and not to find absolute detection thresholds.

1.2 Analysis of the parameter-space

Figures 2 to 4 illustrate how the visual difference energy (Equation 4) varies with the resolution reduction factor and with one of the selected parameters. The two plots in each figure are computed for different motion velocities. The x-axis denotes the percentage of pixels saved when rendering with TRM at a given resolution reduction factor r: $(0.5-0.5\,r^2)\times 100\%$. If not stated otherwise, the default parameters for the simulation are: the angular resolution of 30 pixels per degree, frame rate of 90 Hz and (Michelson) contrast of 0.2.

The effect of local contrast (between the square and the background) is shown in Figure 2. It shows that potential artifacts are

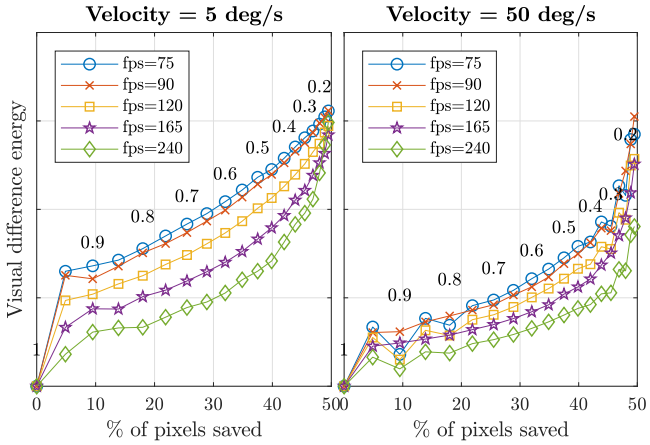

Figure 4: The visual error of TRM for varying frame-rate is plotted as a function of reduction of the number of rendered pixels. Notation identical to Figure 2.

less likely to be seen at smaller contrast and also at higher velocities.

Figure 3 shows that the angular resolution of the display significantly impacts the visibility of distortions, with low pixels-per-degree displays, such as those found in VR headsets, being the most problematic. Since the resolution reduction is performed in the pixel space, larger pixels will naturally result in higher visibility of differences. Finally, Figure 4 illustrates how higher display refresh-rate can reduce the visibility of differences. It is worth noting that the reduction in visibility due to higher frame rate is much lower than the reduction due to lower contrast, higher spatial resolution or higher velocity. Plots indicate that as the visual difference energy function turns steeper when the percentage of pixels saved is more than 40%, therefore reduction factors less than 0.5 are expected to result in more artifacts. However, none of the plots can tell whether the artifacts of TRM are actually visible or not. For that reason, we measure the smallest resolution reduction factor in the experiment, described in the main paper.

One dimension that we could not analyze using our model is the field-of-view (FoV). Although the visibility of flicker can increase for large FoV displays (refer to the CFF in Section 3), we will demonstrate in the experiments in Experiments 2 and 3 that no flicker is visible for the FoV of up to 110 degrees, for two popular VR headsets.

1.3 Experiment 1: frame-rate vs. resolution reduction

1.4 Alternative multiplexing strategies

Higher pixel savings could be possible if we render one full-resolution frame followed by two or more reduced-resolution frames. However, such a multiplexing scheme poses too many challenges to be practical. Firstly, the visibility of flicker increases rapidly with lower temporal frequencies of the multiplexing cycle. For example, when one full-resolution frame is followed by two low-resolution frames on a 120 Hz display, the fundamental frequency of the flicker signal is 40 Hz, making the flicker well visible. Flicker could be reduced on a high-refresh display, but then rendering so many frames would eliminate any savings and provide negligible improvement in the quality of animation. Secondly, each high-resolution frame would need to boost high-frequency signal as many times as the number of reduced-resolution frames. Achieving such a high boost without overshoots is impossible on a display with a limited dynamic range. In summary, rendering interleaved full and reduced resolution frames appears to be the optimal multiplexing technique with regards to computational savings, motion

quality improvement, complexity and flicker.

2 Additional Experiment

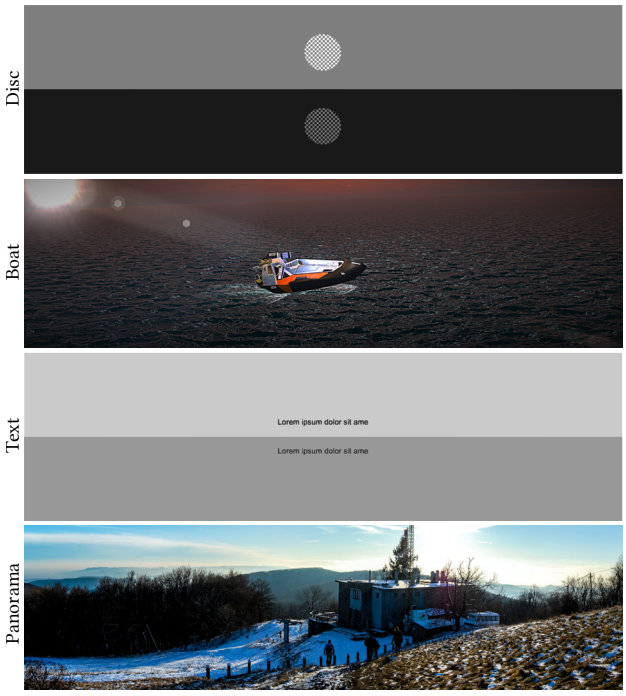

Figure 5: Stimuli used for validation on the HFR monitor.

The primary goal of this experiment was to compare the quality of TRM at three selected resolution reduction factors with standard rendering at 120 Hz and 60 Hz rendering. The setup was identical to the one used in Experiment 1 in the paper (2560×1440 G-Sync capable high-frame-rate Asus ROG Swift P279Q 27" monitor, viewing distance fixed at 75cm using a headrest, Intel i7-7700, NVIDIA GeForce GTX 1080 Ti GPU). However, instead of comparing conditions sequentially, they were shown simultaneously side-by-side.

Stimuli:

In each trial, the participant saw two 10-second looping video clips simultaneously, one in the upper, the other in the lower half of the screen. We considered 5 conditions: (1) 60 Hz, presented at 120 Hz by repeating frames, (2) native 120 Hz video, (3–5) our TRM technique where the odd frames were reduced to 1/2, 1/4 or 1/8 of the original resolution. The clips *Discs*, *Text* and *Panorama* were identical to those used in Experiment 1, while the new clip *Boat* was added to test for more complex animation. The thumbnails of all clips are shown in Figure 5, while Figure 6 visualizes how the eye perceives these videos when SPEM motion is taken into account. The clips were presented using custom software, which played uncompressed frames from the GPU memory.

Task:

The task was identical as in Experiment 1 but the goal of the experiment was different — to measure the quality of each tested condition. We used a pairwise comparison method with the a full design, in which all combinations of pairs were compared. Each observer saw each pair three times, resulting in 120 trials per observer. The order of the stimuli as well as the position of the techniques on screen were randomized.

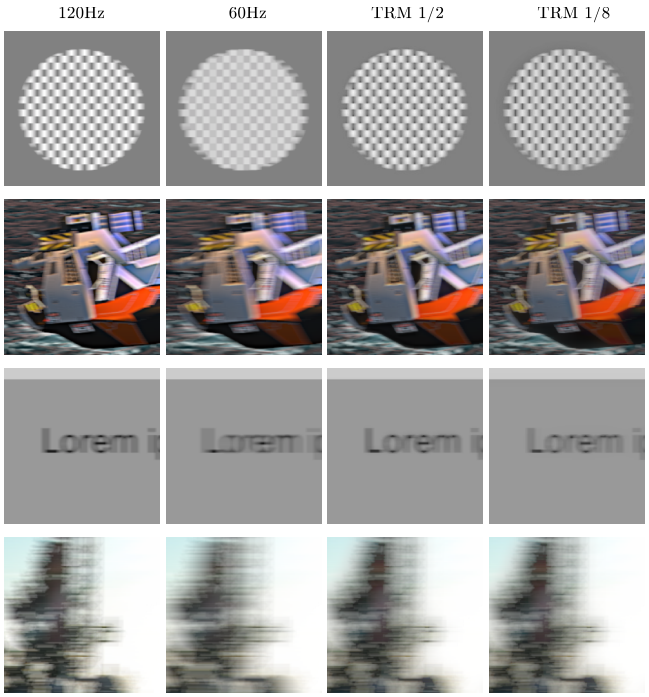

Figure 6: Simulation of perceived video frames. At full frame rate (120 Hz) the stimulus looks sharper than for half frame rate (60 Hz). With TRM low-frequency blur is eliminated. The reduction in contrast for high-frequency signal is usually unnoticeable for moving objects.

Participants:

Eleven paid participants aged 18 – 40 took part in the experiment. All had normal or corrected-to-normal full color vision.

Results:

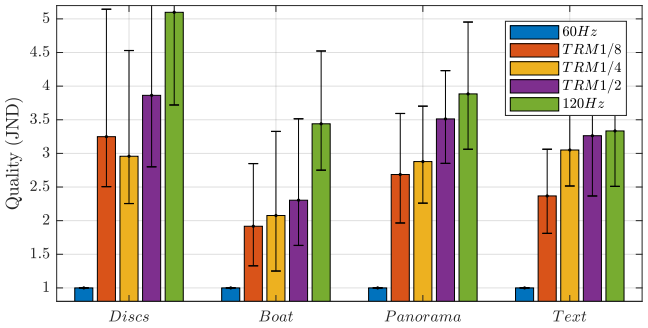

Figure 7: Results of experiment on a HFR monitor. The higher JOD values indicate higher quality. Error bars denote 95% confidence intervals.

The results of the pairwise comparison experiments were scaled using publicly available software as in Experiments 2 and 3. A difference of 1 JOD means that 75% of the population can spot a difference between two conditions. Since JOD values are relative, the 60 Hz condition was fixed at 1 JOD for better presentation.

The results shown in Figure 7 indicate that observers could easily spot the difference between the 60 Hz and 120 Hz videos. TRM was nearly indistinguishable from 120 Hz for *Panorama* and *Text*, but about 75% of the observers could see the difference (1 JOD) for *Boat* and *Discs* clips. This is consistent with our findings in

Experiment 1, only the threshold is shifted due to the side-by-side presentation. Further reduction in the resolution of odd frames did not result in a strong reduction of quality. Unfortunately, the saving in number of rendered or transmitted pixels also becomes negligible as the resolution is reduced.

REFERENCES

[1] S. J. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models. In B. E. Rogowitz and T. N. Pappas, editors, *Human Vision and Electronic Imaging*, volume 3299, pages 180–191, jul 1998.

[2] D. H. Kelly. Motion and vision II Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10):1340, oct 1979.

[3] A. B. Watson. Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express*, 6(1):12, jan 2000.

[4] A. B. Watson and A. J. Ahumada. The pyramid of visibility. In *Human Vision and Electronic Imaging*, volume 2016, pages 1–6, feb 2016.