Hybrid Modulation for Near Zero Display Latency

Trey Greer, Josef Spjut, David Luebke, Turner Whitted
NVIDIA, Durham NC

Abstract

Binary displays for virtual reality can achieve low latency by integrating view tracking with modulation. We present a novel modulation scheme that combines tracking, pulse density modulation, and pulse width modulation to minimize grayscale artifacts. The hybrid modulator is applied to an AMOLED display at an update rate of 1.7 kHz on which we observe nearly zero latency in the perceived image.

Keywords

virtual reality; low latency; AMOLED; delta-sigma modulation

1. Introduction

The resurgence of virtual reality (VR) and new emphasis on augmented reality (AR) has placed a premium on low-latency in displays. Processing delays in conventional rendering pipelines plus double buffered raster updates do not satisfy the need to maintain minimal delay between motion of a head-mounted display (HMD) and the perceived image presented to viewers. Researchers have bypassed the traditional rendering path and applied motion tracking updates directly to the refresh signals driving displays, a step commonly called post-rendering warp [4]. However this approach is still limited by the raster update rate of the display.

While perceptual latency thresholds for just noticeable differences in images vary widely depending on individual viewers and environments, recommended thresholds for VR are as low as 3 ms and for AR less than 1 ms [3]. The 60 to 120 Hz refresh rate of common liquid crystal displays (LCDs) is simply too slow for VR and AR applications. As an alternative, various binary displays can control light modulation or emission far faster than typical gray scale displays. Digital micro-mirror device frame rates, for example, extend to 20 KHz. Similarly we have measured electrical-to-optical response time of an individual active matrix organic light-emitting diode (AMOLED) pixel, when driven with a binary signal, to be in the microsecond range.

In a typical binary display application a grayscale frame is rendered at 60 Hz and presented as input to a temporal modulation process such as pulse width modulation (PWM) which generates a binary output to drive the display at rates in the kHz range. These binary signals are integrated by the human visual system so that over the input frame period the output image perceived by a viewer converges to the input. Latency in this case is still limited 1/60 sec.

If, however, the input imagery is updated more frequently than 60 Hz, or at the same frame rate as the binary display the door is opened to lower latency [7]. Another approach, called adaptive frameless rendering, does not update pixels in raster refresh order; instead prioritizing regions in which objects in a scene are moving or to regions of a scene which are more likely to attract viewer attention [1]. Post-rendering warp, or alternatively "timewarp," for HMDs can rapidly update pixels in response to viewer head motion. All such techniques generate pixel updates with far less latency than conventional rendering. Presenting such pixels directly to the display driver process is, in effect, the merger of tracking and modulation for displays, but it places unique demands on the modulation algorithm.

This paper describes a display system incorporating an AMOLED display driven by binary updates from a circuit that combines tracking with a novel hybrid modulation technique. It achieves both high perceptual grayscale accuracy and low latency.

2. Hybrid Delta-Sigma

As a means of representing grayscale images on a binary display, delta-sigma ($\Delta\Sigma$) modulation produces the equivalent of 1.5 bits of increased intensity resolution per octave of increased display frame rate compared to gains of .5 bits per octave for dithered pulse code modulation [5]. The route from an incoming pixel to a displayed pixel is through the first order $\Delta\Sigma$ modulation path illustrated in Figure 1. The $\Delta\Sigma$ state is simply a running sum that integrates the difference between the input and the instantaneous output of the modulator clocked by the display update clock.

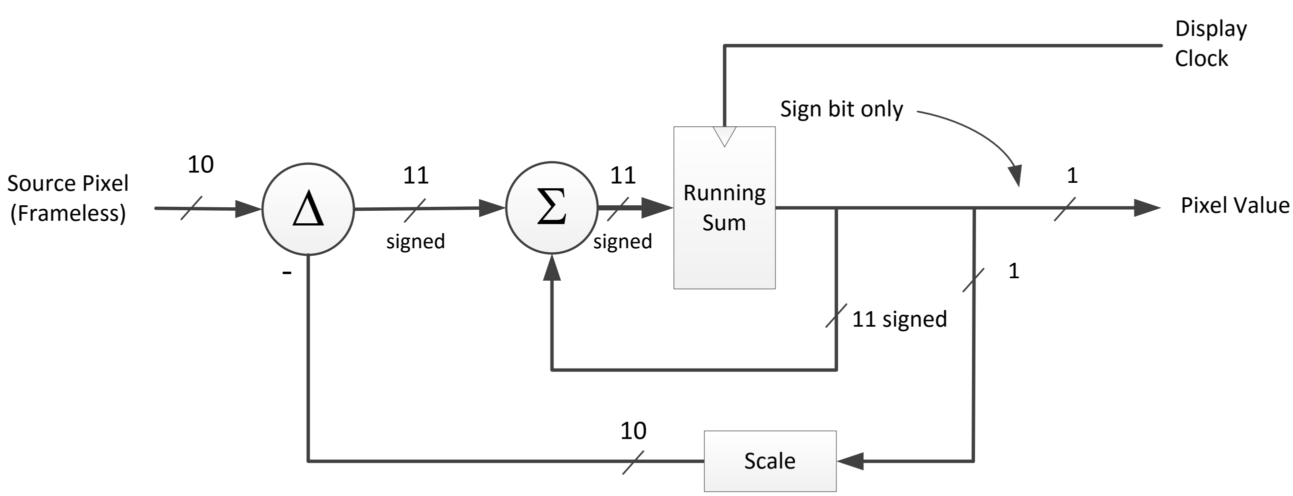

Figure 1. Conventional Delta-Sigma modulation.

 ISSN 0097-966X/16/4701-0076-$1.00

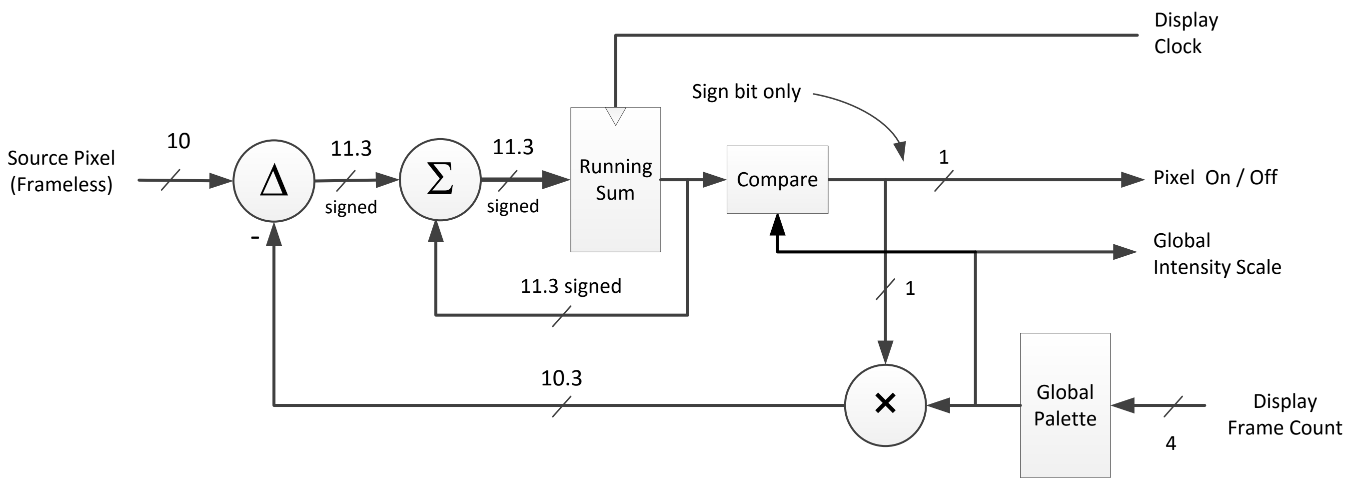

Figure 2. Delta-Sigma with global intensity scaling palette.

Depending on the sign bit from the running sum, binary output to the display is either "on" or "off," effectively controlling the display pixel via pulse density modulation (PDM) [2]. Note that the initial ΔΣ state should be uncorrelated for independent sub-pixels to avoid correlation artifacts.

In the next section we describe an experimental circuit that drives an AMOLED with binary signals at a 1.7 kHz frame rate with 10 bits of intensity resolution. Even this oversampling rate isn't high enough for our requirements. To illustrate, note that a pixel target value of 1 in a signal with 10-bit resolution results in one output pulse every 1024 clock periods. If the display is clocked at 1.7 kHz, the pixel will flash at 1.72 Hz. What viewers perceive is a twinkle artifact due to low intensity pixels 'flashing' at a rate less than the flicker fusion threshold [6]. This artifact may be overcome by applying an intensity reduction to some of the pixel pulses. The low intensity pixels are then represented by lower intensity pulses occurring at a higher rate. This pulse scaling may be applied per pixel, per row, or per frame. In our case we apply scaling globally, i.e. per frame, using scaled pulse widths. Judicious choice of the sequence of global frame intensities can give the needed intensity resolution while suppressing flicker.

We have chosen a global frame intensity sequence of 16 pulse width values ranging from 3/64 of the display clock period up to the entire clock period. The pattern, (1, 1, 1, 3/8, 1, 1, 1, 3/16, 1, 1, 1, 3/32, 1, 1, 1, 3/64), is applied to all pixels in any given frame and repeats every 16 frames.

Using this global intensity sequence the value of 1 in the example above excites the display pixel for 3/64 of a clock cycle every 48 cycles resulting in a pulse rate of 35 Hz, well above the flicker fusion threshold for low intensities. The modified ΔΣ modulator is illustrated in Figure 2.

This removal of flicker comes at a cost in latency. The simple delta sigma modulator of Figure 1 can be modeled as a 1 frame delay plus uncorrelated quantization noise [5]. But the new modulator of Figure 2 adds up to 16 frames of latency. This latency appears as 'motion tails' trailing behind moving objects in the display. But what do we mean by 'moving objects'? We find it useful to define a new space in the vein of the traditional graphics pipeline spaces 'object space,' 'world space,' and 'eye space'. This new space is defined as eye space mapped onto the viewer's retina, and we refer to it as 'retina space'.

Closely observing the motion tail artifacts due to the 16-frame sequence of global intensity modulation, we noticed that the motion tails only appear for objects moving in retina space. This prompted us to transport the delta-sigma state from frame to frame such that it was fixed in retina space. The results are dramatic. The motion tails for objects fixed on the retina (objects the user is tracking) disappear, while the motion tail for objects moving across the retina (the background) are not apparent to the viewer.

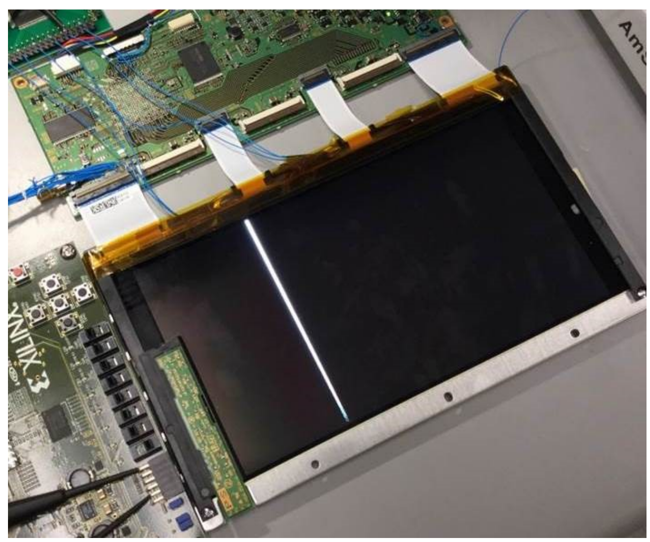

Figure 3. Custom driver applied to AMOLED panel. Bright tapered region across the panel is the 6x540 region driven by the initial circuit prototype.

3. Experimental Apparatus

The AMOLED panel used for this experiment is an older model extracted from a Sony PVM-741 monitor with a resolution of 1026x540 pixels. This display has implemented on glass a pair of shift registers which control row addressing, pixel drive transistor threshold offset compensation, and pixel duty cycle. We found that by bypassing the threshold offset compensation features we could overclock these shift registers at a row rate of 918 KHz. We also found that we could use the duty cycle control feature of the panel's row address scheme to implement the intensity control needed by our hybrid modulation scheme.

In the original product the display columns were driven with 10-bit resolution analog voltages from digital to analog converters (DACs) mounted on three flexible circuit boards across the top edge of the display. These DACs provided 10-bit pixel voltages at a row rate of roughly 30 KHz for a frame rate of 60 Hz. Using test points on the flex circuits and a repurposed plasma display driver chip (STMicroelectronics STV7620), we found we could drive binary pixel values into six adjacent columns of the display at our 918 KHz row rate. With 540 rows this gives us a binary display rate of 1700 frames per second.

Output from the global palette to the display, shown as global intensity scale in Figure 2, is used to control the duty cycle of the row select shift register clock in the AMOLED panel. Applying the hybrid modulation approach described in the previous section yields a full color display with 10 bits per pixel at a resolution of 6x540. We are in the process of adding driver and processing circuitry to extend the resolution across the entire 1026x540 panel.

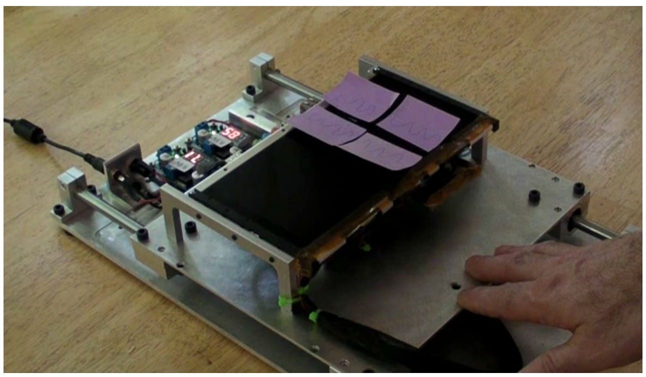

Figure 4. Linear motion platform for tracking and display.

To inject motion tracking into the experiment the display panel is mounted onto a precision machined linear table with 1-D mechanical tracking shown in Figure 4. The linear motion of the table acts as a proxy for head rotation in an HMD. The US Digital EM2 Transmissive Optical Encoder module has a resolution of 2000 counts per inch with 4 quadrature edges per count yielding a resolution of 3.2 microns. The FPGA tracks the quadrature pulses, using the current linear table position to generate each row of binary pixel updates. With this simple setup we have driven the display tracking latency to zero to allow us to consider only display latency as the limiting factor in our experiments.

4. Evaluation

The subjective response of viewers is that the image presented by the display is not physically attached to the panel, but exists in its own space, i.e. no matter how rapidly the display platform is moved there is no noticeable movement of the displayed image. The ideal use of this independent display space would be to map the space of the viewer's retina in order to display an arbitrary image directly in retina space. Thus the perceived illumination can be directly controlled and the image synthesis can appropriately exploit the human visual system.

The motion trail artifact described in an earlier section is evident in the upper image of Figure 5. In essence a pixel's modulation state is attempting to display an image value that has moved to a new location. With state tracking enabled the lower image in Figure 5 is the result. This translation of modulator state is implemented in 1-D by simply offsetting state memory addresses using a motion vector generated by the tracker. On the surface this seems like a limited approach, but it is perfectly compatible with rapid post rendering warp implemented as image translation [7].

Achieving near zero latency with minimal artifacts comes at a cost. The need to maintain and update a spatio-temporal error model results in a significant memory bandwidth load. At the 1.7 kHz frame rate 13 bits of modulator state are read and written for each color sub-pixel, i.e. 4,420 bits per second. Multiplied to full resolution the load is 73.5 gigabits per second. However, the use of one modulation data path per sub-pixel column is clearly overkill given the relatively low clock speeds. An implementation at moderate clock speeds with blocks of shared modulation datapaths is clearly feasible.

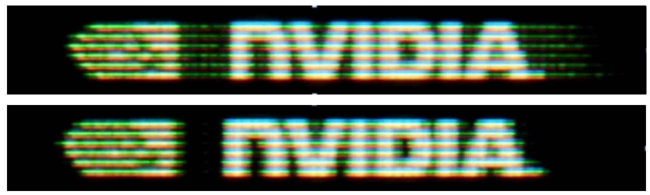

Figure 5. Rapidly moving image without modulation state tracking (above) and with state tracking (below). (Note that display is rotated 90 degrees and subpixel columns are horizontal.)

5. Conclusion

Viewer feedback for the initial demo at a resolution of 6x540 has been enthusiastically positive. The unanimous comment is that this display is the most stable any of our viewers have seen. Whether this reaction will hold true in the transition to a full 1026x540 pixel image remains a question for future evaluation. However the initial prototype does adequately illustrate the benefit of combining just-in-time tracking with the display modulation circuitry. The quality improvement gained by augmenting PDM display modulation with a PWM global palette is abundantly clear.

6. References

[1] Dayal, C. Woolley, B. Watson, D. Luebke, "Adaptive Frameless Rendering," Proceedings of Eurographics Symposium on Rendering (2005).

[2] J-H Jang, M. Kwon, E. Tjandranegara, K. Lee, B. Jung, "A digital driving technique for an 8b QVGA AMOLED display using ΔΣ modulation," in IEEE International Solid-State Circuits Conference - Digest of Technical Papers (2009).

[3] J. Jerald, "Scene-Motion Thresholds and Latency Thresholds for Head-Mounted Displays," PhD dissertation, UNC Computer Science Department (2009).

[4] W. R. Mark, L. McMillan, and G. Bishop. "Post-Rendering 3D Warping," in Proc. ACM I3D, pages 7–16 (1997).

[5] S. Norsworthy, R. Schreier, G. Temes, eds., Delta-Sigma Data Converters, Theory, Design, and Simulation, IEEE Press (1996).

[6] http://webvision.med.utah.edu/book/part-viii-gabac-receptors/temporal-resolution/

[7] F. Zheng, T. Whitted, A. Lastra, P. Lincoln, A. State, A. Maimone and H. Fuchs. "Minimizing Latency for Augmented Reality Displays: Frames Considered Harmful," IEEE ISMAR 2014 (2014).