

Apparent Display Resolution Enhancement for Moving Images

Piotr Didyk¹ Elmar Eisemann^{1,2} Tobias Ritschel¹ Karol Myszkowski¹ Hans-Peter Seidel¹

¹ MPI Informatik

² Telecom ParisTech / CNRS-LTCI / Saarland University

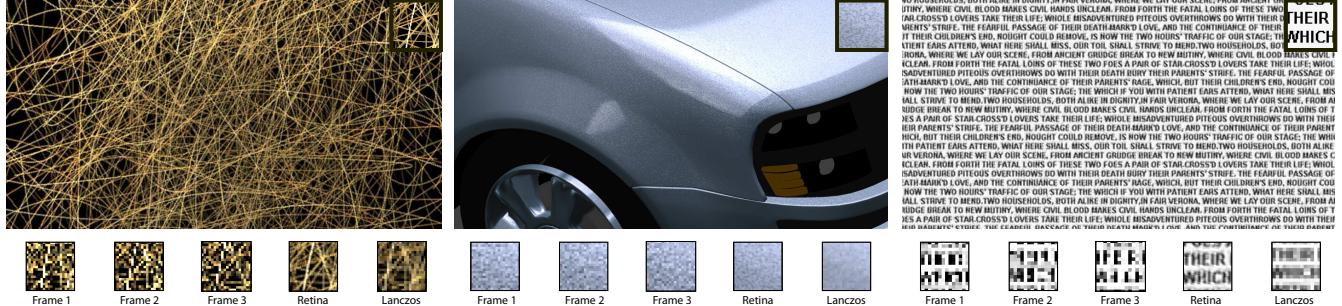


Figure 1: Depicting fine details such as hair (left), sparkling car paint (middle) or small text (right) on a typical display is challenging and often fails if the display resolution is insufficient. In this work, we show that smooth and continuous subpixel image motion can be used to increase the perceived resolution. By sequentially displaying varying intermediate images at the display resolution (as depicted in the bottom insets), subpixel details can be resolved at the retina in the region of interest due to fixational eye tracking of this region.

Abstract

Limited spatial resolution of current displays makes the depiction of very fine spatial details difficult. This work proposes a novel method applied to moving images that takes into account the human visual system and leads to an improved perception of such details. To this end, we display images rapidly varying over time along a given trajectory on a high refresh rate display. Due to the retinal integration time the information is fused and yields apparent super-resolution pixels on a conventional-resolution display. We discuss how to find optimal temporal pixel variations based on linear eye-movement and image content and extend our solution to arbitrary trajectories. This step involves an efficient method to predict and successfully treat potentially visible flickering. Finally, we evaluate the resolution enhancement in a perceptual study that shows that significant improvements can be achieved both for computer generated images and photographs.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—display algorithms, viewing algorithms;

Keywords: image reconstruction, perception, eye tracking

1 Introduction

Due to physical limitations of existing display devices, real-world luminance, colors, contrast, as well as spatial details cannot be directly reproduced. Even though hardware is constantly evolving

and images are skillfully tone and gamut mapped to adapt them to the display’s capabilities, these limitations persist. In order to surmount the physical limitations of display devices, modern algorithms started to exploit characteristics of the human visual system (HVS) such as apparent image contrast [Purves et al. 1999] based on the Cornsweet Illusion or apparent brightness [Zavagno and Caputo 2001] due to glare. In this work we address the problem of image resolution and extend apparent details beyond the physical pixel resolution of display devices by exploiting the time dimension, as well as high refresh rates.

One context in which resolution plays a crucial role is scale-preserving rendering. In the real world, we can clearly distinguish individual hair strands, while such details are usually rendered much thicker, hence affecting realism. Metallic paint, as often applied to cars, can have sub-pixel size sparkling effects where a higher resolution increases faithfulness. Fidelity sensitive applications (e.g., product design, virtual hair styling, makeup design, even surgical simulations) suffer from such shortcomings.

Further, there is a clear mismatch between available sensors exceeding 10 mega-pixel resolution and current display capabilities. Panorama stitching or Gigapixel images [Kopf et al. 2007] can make such a mismatch even more profound. While zooming allows to explore details, seeing the whole image or larger parts in full detail is often more appealing. Downsampling to the resolution of an output device is common practice to display such mega-pixel images, but it filters out high-frequency spatial image details.

In computer graphics, details are easily modeled, but the image-display stage may ruin the visual effect. This is particularly striking for the current trend of smaller devices where resolution is often very limited. In many cases, scrolling is used to present textual information or larger images to the user. Our work addresses the display of such scrolled information and increases readability.

An important aspect, not only in the context of portable devices, is energy consumption that is linked to the increase of physical resolution, whereas producing high-refresh rates is energy-efficient. In particular, the upcoming OLED technology will make another step in this direction enabling refresh rates of thousands of Hz.

In this work, we are faced with several input image pixels that map to the same display pixel. We want to present them to the observer without applying detail-destructive resolution adjustments. Our main idea is to transform a high-resolution input image into N images of the target display resolution (Fig. 2), that we call *subimages*. We then render the subimages sequentially on a high refresh rate display (120 Hz) and apply a one-pixel shift at the end of each rendering cycle. The process is then restarted from the new position. The result gives the impression of a smooth motion. When an observer focuses on an interesting image detail, the eye will track the feature and a *smooth pursuit eye motion* (SPEM) of matching velocity is established. This is critical in our approach because then the subimage details are consistently projected to predictable locations of the fovea region which features the highest density of cone photoreceptors in the retina. By exploiting the integration in the human eye (both temporal, via quickly displayed subimages, and spatial, via rigid alignment of moving and retinal images), the effect of apparent resolution enhancement is achieved.

Motion is key to our approach because it ensures that the pixel grid projects to different locations on the retina, which we exploit in our approach. Consequently, there is a link between the motion direction and the apparent resolution increase, e.g., horizontal / vertical motion only enables horizontal / vertical improvements.

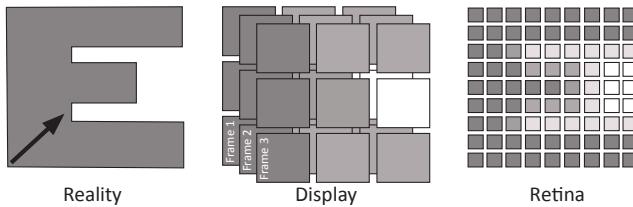


Figure 2: Fixational eye tracking over an ROI in combination with a low-resolution image sequence leads to an apparent high-resolution image via integration in the retina.

In this paper, we explore apparent resolution enhancement in various motion scenarios. Our contributions are:

- An apparent resolution model for moving subimages
- A method to derive optimal subimages according to our model
- A perceptual experiment confirming our findings.

This paper is structured as follows: We present previous work and related perceptual findings in Sec. 2. Sec. 3 introduces our model. We describe our algorithm in Sec. 4 which addresses potential flickering as detailed in Sec. 5. Our perceptual experiment is analyzed in Sec. 6, before we conclude in Sec. 7 and 8.

2 Previous Work

We will first discuss the problem of spatial resolution enhancement beyond the physical pixel resolution in image displays or projectors. Particularly, we are interested in solutions that rely on active steering of spatial and temporal signal integration in the retina which we employ in our work as well.

Color Matrix Displays Subpixel rendering is used to increase the image resolution by breaking the assumption that R , G , or B channels are unified in a single pixel. If their arrangement is known, one can use channels from neighboring pixels to increase spatial resolution. Platt [2000] showed an optimal filtering for liquid crystal displays (LCD), which was used in the ClearType font technology.

The resulting resolution enhancement is limited to the horizontal direction only and works best for black-and-white text. Subpixel rendering is advantageous for complex images as well, but saturated colors or naïve compensations of spatial color-plane misalignments may lead to color fringes at sharp edges, as well as color moiré patterns for high frequency textures [Messing and Kerofsky 2006]. This is underlined by Klompenhouwer and de Haan [2003] who found that subpixel rendering shifts luminance aliasing caused by frequencies over the display's Nyquist limit into the chrominance signal. For this reason, optimization frameworks are often used that involve the precise physical layout of subpixels including inactive or defective subpixels [Messing and Kerofsky 2006]. Hara and Shiramatsu [2000] observe that a special pixel-color mosaic extends the pass band of the image spectrum when moving an image with a specific velocity across the display. However, they conclude that no improvement can be achieved for the standard $|RGB|RGB|\dots$ arrangement, which is predominant in LCD displays, including the ones considered in this work.

Color Sequential Displays Temporal integration in the HVS is exploited for color fusion in digital light processing (DLP) video projectors. While many devices rely on a spatial RGB -subpixel integration, DLPs display the RGB color components sequentially with a temporal frequency over the perceivable flickering limit of the HVS. On the other hand, spatial color disintegration (“rainbow effect”) can be observed, particularly for quickly moving objects of high contrast. Also, eye saccades and SPEM may lead to this effect because images representing different channels are displayed with some delay and, thus, can appear misaligned on the retina.

Wobulation and Display Supersampling In wobulated projectors, multiple unique slightly-shifted subimages are projected on the screen using an opto-mechanical image shifter [Allen and Ulichney 2005], which is synchronized with the rapid subimage projection to avoid flickering. The perceived image resolution and active pixel areas (otherwise limited by the door grid between physical pixels) are enhanced. A similar effect is achieved by *display supersampling* using multiple carefully-aligned standard projectors [Damera-Venkata and Chang 2009], where also an optimization for arbitrary (not raster aligned) subpixel configurations is performed. Our work is similar to [Allen and Ulichney 2005] and [Damera-Venkata and Chang 2009] in the sense that a high-resolution image is transformed in a set of low-resolution images, but we aim at a single desktop display or projector with a limited resolution and fixed pixel layout. We overcome these limitations by making use of SPEM that aligns with a slow linear image motion.

Subpixel information Subpixel information acquired via subtle camera motion has proven useful in many applications, such as super-resolution reconstruction [Park et al. 2003] or video restoration [Tekalp 1995]. In these schemes, subpixel samples from subsequent frames are merged into explicitly reconstructed images, which, finally, are downsampled to match the display resolution. Note that our problem is, in some sense, an inverse problem. We do not need to reconstruct high frequency information because it is available in the original content. Instead, our task is to decompose high resolution images into low resolution subimages which are finally perceived as a high resolution image when displayed sequentially. Our approach avoids any explicit reconstruction and relies on a perceptual processes to ensure detail enhancement.

Krapels et al. [2005] reported better object discrimination for subpixel camera panning than for corresponding static frames (independently confirmed in [Bijl et al. 2006]). Object discrimination improved regardless of the subpixel sensor motion rate, except for

critical velocities [Tekalp 1995, C. 13] such as a one-pixel shift. A similar observation applies to rendering with supersampling where several images, rendered with slightly differing camera positions, are integrated in order to gain information. In Sec. 4 we explicitly discuss this case and confront it with our approach, which leads to subpixel detail visibility.

Image Resampling Displaying continuous or high resolution input images on a finite, lower resolution display is a sampling problem [Mitchell and Netravali 1988]. The display image is reconstructed by convolving the input image with a *reconstruction filter* for every output pixel. Popular reconstruction filters are Lanczos and Mitchell's filter [Mitchell and Netravali 1988].

In our case of moving high resolution images, resampling should first move the image and then filter, before sampling. This conforms with previous findings from computer vision [Krapels et al. 2005].

Background on Perception

This section presents limitations of visual acuity due to the photoreceptor density in the retina, moving-image perception, and temporal-flicker visibility, which are pivotal for our approach.

Spatial Visual Acuity When discussing image-details it is important to investigate the HVS capabilities in resolving spatial detail for existing devices [Deering 2005]. The highest anatomically determined density of cones in the fovea is estimated as 28'' (arc seconds) [Curcio et al. 1990] which, according to the Nyquist's theorem, enables an observer to distinguish 1D sine gratings of roughly 60 cycles/deg resolution. At the same time, the pixel size of a typical full-HD desktop display, such as a 120 Hz Samsung SyncMaster 2233, when observed from 50 cm distance amounts to roughly 1.5' (arc minutes), which means that 1 pixel covers roughly 9 cones. Further, in many situations an observer might actually be closer to the screen, as is the case for hand-held devices. Note that this estimate is valid for the central fovea region. The cone density drops quickly with the eccentricity [Curcio et al. 1990]. We also do not consider low-pass filtering, induced by the optical properties of the eye, which perfectly matches the foveal photoreceptor density and removes frequencies that would lead to aliasing. Interestingly, the HVS is still able to interpolate a feature position with an accuracy higher than 20 % of the distance between cones in the fovea, although this visual hyperacuity is more a *localization* than a *resolution* task (the position of one image element must be located relative to another, e. g., slightly shifted lines in the vernier acuity task [Wandell 1995, p. 239]).

Perception of Linearly Moving Images The perception of motion, where information on objects moving in a 3D world is inferred from 2D retinal images is a complex process [Wandell 1995, C. 10]. In this work, we are concerned with a simpler case of moving 2D images that are stabilized on the retina through SPEM. As confirmed in an eye tracking experiment [Laird et al. 2006] such a stabilization is almost perfect for steady linear motion with velocities in the range of 0.625–2.5 deg/s, which we consider in this work. The performance stays very good up to 7 deg/s. SPEM initialization is a fast process and good tracking is achieved in less than 100 ms. This is faster than typical saccades (200 ms [Krauzlis and Lisberger 1994]), which makes switching the tracking between moving objects with different velocities and in different directions an easy and effortless process. While the eye undergoes additional fixational eye movements, such as tremors, drifts, and microsaccades, these are similar to static fixation, and it is believed that the HVS suppresses their influence on perception [Martinez-Conde et al. 2004]. Schtz et al. [2008] reported

a 16 % increase of visual sensitivity during SPEM for foveally presented luminance stimuli of medium and high spatial frequencies compared to the static case. This HVS mechanism serves towards a better recognition of tracked object, which contributes to human survival skills.

Also visual hyperacuity is maintained for moving targets at uniform velocity in the range 0–4 deg/s [Fahle and Poggio 1981]. Moreover, an illusory displacement can be observed when displaying two parts of a line with a few milliseconds delay [Burr 1979] because for both targets the HVS assumes a smooth motion and their different perceived locations are correlated with the delay between their exposure. Fahle and Poggio [1981] stress the role of the constant velocity assumption as an important constraint in the target position interpolation by the HVS.

Temporal Flickering Our approach relies on a sequential display of subpixel values, which potentially causes temporal flickering. Due to the eye pursuit the resulting signal affects photoreceptors with a frequency imposed by the display device refresh rate. Over the *critical flicker frequency* (CFF) [Kalloniatis and Luu 2009], flickering is not perceivable, the subpixel intensities are fused, and a steady appearance is reached. Flickering perception is complex and the CFF depends on many factors such as the adaptation luminance, spatial extent of flickering pattern, and retinal region (the fovea or periphery) at which this pattern is projected. The CFF rises roughly linearly with the logarithm of time-averaged background intensity (the Ferry-Porter law). The specific CFF values for different adaptation luminance have been measured as the *temporal contrast sensitivity function* [de Lange 1958] for stimuli of the spatial extent of 2° (angular degrees). One important observation is that the CFF is significantly reduced for smaller stimuli [McKee and Taylor 1984; Mäkelä et al. 1994]. The CFF is the highest in the fovea, except for bright adaptation conditions and large stimuli, when flickering is better perceived at the periphery. In Sec. 5, we analyze the flickering perceptibility specifically for viewing conditions that are relevant for our technique.

3 Model

First, we will concentrate on a simple case of a single receptor and a constant velocity. We then use this model to predict the perceived image for arbitrary subpixel images. We then show how an optimization process can be used to transform a high-resolution input into an optimal sequence of subimages.

Photoreceptors The light response of the human photoreceptor is a well-studied issue by neurobiophysicists [Van Hateren 2005]. In this work, we need to rely mostly on psychophysical findings, which take into account the interaction between photoreceptors as well as higher-level vision processing.

One particular element is the CFF, introduced in Sec. 2. The eye has a certain latency and rapidly changing information is integrated over a small period of time which depends on the CFF. In most cases, we used a 120 Hz screen and displayed three subimages, before advancing by one pixel and displaying the same three subimages again. Hence, each subimage sequence takes $1/40^{th}$ of a second. Although this frequency is generally below the CFF and a special processing is needed (Sec. 5), 40 Hz is usually close to the CFF in our context. Higher refresh rates would allow us to add even further subimages. We will detail these points in Sec. 5 and assume for the moment that the subimage sequence is integrated by the eye.

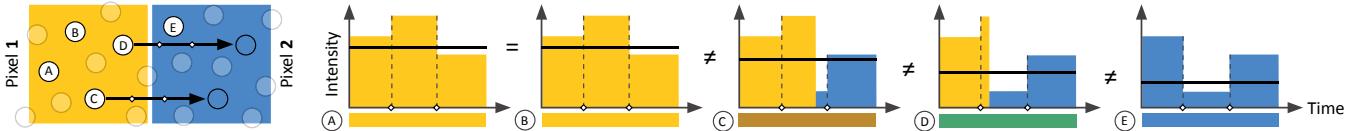


Figure 3: Spatio-temporal signal integration. Left: Two pixels (yellow and blue square) covered by receptors (circles). Right: Intensity response of receptors A–E over time for changing pixel intensity in three frames (dotted lines). For static receptors (A, B and E) the resolution cannot be improved because the same signal is integrated over time. Due to motion (arrow), receptors C and D, although beginning their integration in the same pixel, observe different signals which we exploit for resolution enhancement.

Receptor vs. Changing Image First, we investigate the retinal response for a standard moving picture observed under smooth eye pursuit. In contrast to the real world where, during tracking, the signal arriving on a photoreceptor is basically constant, the situation is different on today’s displays. A single frame is usually displayed over an extended period of time (hold-type displays) or multiple times (general high-refresh rate screens) instead of flashing the information only once (CRT displays). Thus, for an insufficient frame rate, the eye movement over the screen mixes neighboring pixel information. As a consequence, tracking of screen elements leads to the undesirable *hold-type blur*. In our case, we will make use of this observation to increase the perceived resolution.

We first consider a static photoreceptor with an integration time of T that observes a pixel position p of an image I . If I changes over time and is thus a function of time and space, the response is given by $\int_0^T I(p, t) dt$. If the receptor moves over the image during this duration T on a path $p(t)$, the integrated result is:

$$\int_0^T I(p(t), t) dt. \quad (1)$$

Retina In order to predict a perceived image, we need to make simplifying assumptions about the layout of photoreceptors on the retina. While the real arrangement is complex and non-uniform [Curcio et al. 1990, Fig. 2], we assume a uniform grid-aligned positioning with a higher density than the image resolution. The latter assumption reflects that in the dense region of the fovea several receptors observe a pixel (Sec. 2).

4 Resolution Enhancement

Our goal is to use the temporal domain to increase spatial information and, hence, to enhance the apparent resolution. Unfortunately, as indicated by Eq. 1, it is not possible to increase the resolution of a *static* image without eye movement. In such a case, neighboring receptors that observe the same display pixel also share the same integrated information (Fig. 3, cases A,B).

Precisely, this observation implies that for a given time t_0 , $I(p(t), t_0)$ is constant for all $p(t)$ in the same pixel and $I(p(t_0), t)$ is constant during the time that we display the same pixel intensities. Therefore Eq. 1 becomes a weighted finite sum of pixel values:

$$\sum_{t=0}^T w_t I(p(t), t). \quad (2)$$

This equation reveals two crucial elements. First, the simulation can be achieved via a simple summation which will allow us to apply a discrete optimization strategy. Second, for differing paths $p(t)$ (even if only the starting points differ) the outcome of the integration generally differs. This will be key in increasing the apparent

resolution. Due to the changing correspondence between pixels and receptors during SPEM, as well as the temporally varying pixel information, differing receptors usually receive differing information (Fig. 3, cases C,D). Consequently, we can control smaller regions in the retina than the projection of a single pixel.

Simple Case Before generalizing our approach, we will first illustrate the simple case of a static high-resolution 1D image I_H . For each high-resolution pixel we assume a single receptor r_i , while our 1D display can only render a low-resolution image I_L . Let’s assume for now that the resolution of I_H is twice as high as the resolution of I_L and that the image is moved with a velocity of half a display pixel per frame. In theory, we could change the value of each display pixel on a per-frame basis. Nevertheless, we assumed that all receptors track the high-resolution image perfectly. Hence, after two frames, all receptors have moved exactly one screen pixel. We find ourselves again in the initial configuration and the same two-frame subimage sequence can be repeated.

For this particular case, each receptor will, while tracking the image, either see the color of exactly one pixel during the duration of two frames or of two consecutive pixels. More precisely, following Eq. 2, receptor i captures:

$$r_i = \begin{cases} (I_L(i, 0) + I_L(i, 1))/2 & : i \% 2 == 0 \\ (I_L(i, 0) + I_L(i + 1, 1))/2 & : i \% 2 == 1 \end{cases} \quad (3)$$

In order to make the retinal response best match I_H , r_i should be close to $I_H(i)$. This can be formulated as a linear system:

$$W \begin{pmatrix} I_L^1 \\ I_L^2 \end{pmatrix} = I_H, \quad (4)$$

where I_L^t is the subimage displayed at time t and W a matrix that encodes the transfer on the receptors according to Eq. 2. In the scenarios we considered, the matrix W is overdetermined, meaning that there are more independent equations in our system than variables (unknown pixels in subimages). We usually assume that there are fewer pixels displayed over time than the total resolution of the original image and the resolution of the retina is considered to be at least as high as the resolution of the original image. We find the final solution using a constrained quadratic solver [Coleman and Li 1996]. While a standard solver would also provide us with a solution that is coherent with respect to our model, a constrained solver respects the physical display limitations with respect to brightness. Therefore, this approach guarantees that the final subimages can be displayed within the range of zero (black) to one (white). Our problem is convex and so convergence can be guaranteed.

It is natural that subimages contain aliasing. The receptors will integrate the image along the motion path and therefore filter the values. On the other hand, our optimization minimizes the residual of the perceived final image with respect

to the original high resolution version. Therefore, as long as the original frame does not exhibit aliasing problems, the optimization should avoid aliasing in the perceived image as well. Although it is difficult to formally prove this cancellation, no aliasing problems were observed during our experiments.

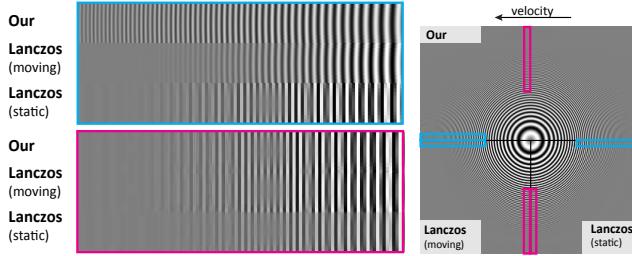


Figure 4: Our method vs. Lanczos. Our method uses image motion to improve the perceived resolution along this movement direction by showing 3 subimages on a rapid display. While we rely on a frame optimization, moving Lanczos resampling derives subimages by filtering the translated original image. The eye integration is computed by blending the subimages, assuming perfect tracking.

Fig. 4 shows an example of a horizontal movement. The resulting spatial apparent resolution is much higher horizontally (blue) than for a standard bandwidth-filtered image, while vertical resolution (red) is similar to the case of a moving Lanczos resampling.

General Case An important property is that an integer movement allows us to reuse the subimage sequence after a few iterations. This is interesting for static images where one can choose a displacement direction and enhance resolution using only a small amount of texture memory.

It is possible to treat more general movements by adapting the integration weights w_t from Eq. 2. Basically, the weights should be proportional to the time, that a pixel's color is seen by a photoreceptor. To formally compute these weights, we introduce one weight $w_{x,y,t}$ for each pixel value $I_{x,y}^t$ where x, y is a discrete pixel position and t the discrete time interval during which the pixel's color is constant, such that: $\int_0^T I(p(t), t) dt = \sum w_{x,y,t} I_{x,y}^t$. It follows:

$$w_{x,y,t} := \int \chi_{(i,j)}(p(t)) \chi_k(t) dt, \quad (5)$$

where χ describes a characteristic function. Precisely, $\chi_{(i,j)}(p(t))$ equals one if $p(t)$ lies in pixel (i, j) , else it is zero, $\chi_k(t)$ is a similar function to test the time intervals. One underlying assumption is that the receptor reaction is immediate with respect to a changing signal. Consequently, temporal integration corresponds to a box filter in the temporal domain.

5 Flicker Reduction

The previous result respects the limits of the display device, but it does not necessarily respect the limits of the HVS. We made the crucial assumption that the HVS integrates a fixed number of subimages and our method only works if their pixel information is fused without producing objectional flickering. To arrive at a flicker-free solution, we proceed as follows: First, a perceptual flicker model computes the amount of flicker for every pixel in the optimized sequence (the *reduction map*). Second, we use this map to define a pixel-wise blending between our potentially flickering, but optimized sequence and a never-flickering, non-optimized standard filtering sequence (Fig. 5).

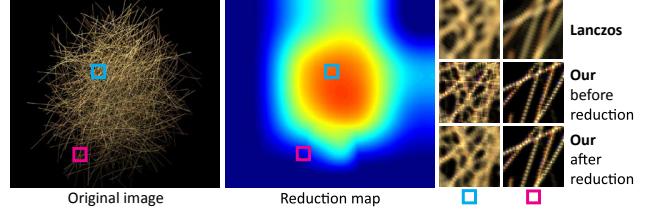


Figure 5: Flickering reduction. Left: Original high resolution image. Center: Reduction map. Right: Outcome of Lanczos filtering, as well as our approach before and after flickering reduction for the marked regions. Note that in the regions of strong temporal contrast reduction an improvement over Lanczos filtering is visible. Similar to Fig. 4, images for our approach are simulations of perceived images assuming motion and perfect eye tracking.

Flicker Detection Model The flicker detection model, used in our solution, is multi-scale, conforming to the scale-dependence of the CFF. It derives per-scale reductions that are pushed to the pixel level where the final contrast reduction happens. In detail, we first compute the maximal intensity fluctuation in each pixel of our subimages. Because flickering is strongly scale-dependent [Mäkelä et al. 1994], we cannot just rely on these values. We use a gaussian pyramid to add a scale component. For each level, this results in a fluctuation measure of the corresponding area in the original image. We can then rely on the perceptual findings in [Mäkelä et al. 1994, Fig. 1], to predict the maximally-allowed temporal variation that will not lead to perceived flickering for such an area (measured as an angular extent). If we find that these thresholds are exceeded, we compute by how much the temporal fluctuation needs to be reduced. We then propagate these values to the lowest-pixel level by taking the maximum reduction that was attributed to it on any of the higher levels (refer to the flickering map in Fig. 5). The maximum ensures that the final flickering will be imperceptible on all scales.

Flicker Sensitivity vs. Pattern Spatial Extent Related experiments with flickering visibility of thin line stimuli (with the angular length up to 2°) indicate a low flickering sensitivity, both in the fovea and periphery [McKee and Taylor 1984, Fig. 5]. Further evidence exists that the sensitivity generally drops rapidly for small patterns [Mäkelä et al. 1994]. This is of advantage to our method as it hints at flickering being mostly visible in large uniform regions. As these uniform regions are those lacking detail and, consequently, our subimages will strongly resemble the original input, any value fluctuation is eliminated.

Hecht and Smith [Kalloniatis and Luu 2009, Fig. 10] found that for a stimuli of 0.3° angular extent and adaptation luminance below 1000 cd/m^2 , the CFF does not exceed 40 Hz. Similar observations can be made in the Ferry-Porter law that indicates a roughly linear CFF increase with respect to the logarithm of time-averaged background intensity up to 40 Hz where the CFF starts to stagnate and the law no longer holds. This seems to indicate that the choice of three intermediate images for a 120 Hz display is very appropriate. In practice, we encountered very few flickering artifacts when displaying a three-subimage solution unprocessed. Consequently, our postprocess leaves most of the original solution unaltered. Nevertheless, when longer integration times are needed, either because more subimages are added or the display's refresh rate is reduced, the processing improves the result significantly. On a 120 Hz display, four subimages became possible without visible flickering. Such a case is illustrated in Fig. 5. Four subimages lead to more details than the three subframe solution and we can work with lower velocities.

Discussion Our approach keeps the highest amount of detail possible while ensuring that the outcome does not result in a perceivable flickering as detected by our flickering model. The blur in Fig. 5 (bottom-right) is a natural consequence of this trade-off between detail/flickering and low-resolution/no-flickering. Since our optimization guarantees that the resulting image fits to the display range, which is also the case for energy-preserving Lanczos filter, any interpolation between such a pair of images cannot cause intensity clipping. Artifacts, e.g., ringing cannot occur, because the reduction map, used for blending, needs only to be a conservative bound in order to detect perceived flickering. Hence, it is possible to find a conservative band-limited image (in practice, a dilation followed by smoothing).

One alternative flicker suppression would be to incorporate the constraints on the maximal temporal fluctuations of signal into the optimization, but this has disadvantages. The process would no longer be quadratic, endangering convergence. It would increase computation times and put pressure on the hard constraints needed to match the display’s dynamic range.

A second alternative would be to suppress flickering via temporal smoothing, but such attempts prove inadequate. Temporal smoothing combines information that should be kept separate to achieve the resolution enhancement according to our model. To illustrate this, consider the receptor C in Fig. 3 moving from one pixel to the next at time t . Filtering over time, would introduce information in the first pixel that occurs after time t , this information was not supposed to be seen by C which at time t is already in the second pixel. We exploit this combination of time and space in our model.

Our flicker reduction, is general and is executed in milliseconds on the GPU. It could be used in other contexts, e. g., to detect and then remedy temporal aliasing for real-time rendering.

6 Experiments

To illustrate the versatility of our approach, we present several application scenarios and tested them in a user study in order to illustrate their effectiveness. We used a 22 inch (diagonal) 120 Hz Samsung SyncMaster 2233 display at its native resolution 1680×1050 . We investigated also lower resolutions to address the fact that displays constantly grow, often already exceeding 100 inches, but keeping their resolution on the level of full HD. On a 100-inch screen, pixels would approximately be four times bigger than in our experiments. The monitor was viewed by the subjects orthogonally at a distance of 50–70 cm. Because some experiments required that two images are simultaneously shown next to each other in a horizontal arrangement, the video sequences and images of resolution 600×600 have been used in all studies. We considered a 120 Hz refresh rate, decomposing the original images into three subimages to illustrate that the details are also visible for the faster-moving variant (compared to four subimages).

14 participants with normal or corrected-to-normal vision took part in the main part of experiments. In an additional 3D rendering part five participants were considered. Subjects were naïve regarding the goal of the experiment and inexperienced in the field of computer graphics. The participants were seated in front of a monitor running the experimental software in a room with controlled artificial lighting. They received standardized written instructions regarding the procedure of the experiment. In all experiments the time for each trial has been unlimited.

High-resolution Images In our study we considered five stimuli, including detailed rendering and text (Fig. 1) as well as natural images (photographs of a cat and a car included in the accompanying

video). The hair and car images have been rendered with a high level of detail and include subpixel information from elongated hair strands and tiny sparkles in the metallic paint. Text was used to evaluate readability as an indicator of detail visibility. Finally, we used photographs to check the performance of our method for real images, which often exhibit slightly blurry edges with respect to synthetic images. Our aim was to show that our method outperforms standard image-downsampling techniques. We tested various velocities and compared our method to Lanczos resampling as well as Mitchell and Netravali [1988], asking people to compare the detail visibility.

In the first part of the study, subjects compared the static reference image of high-resolution that was placed on the right to a moving image on the left. The left image was per-frame Lanczos-filtered or our solution, initialized randomly and not labeled. We did not consider more naïve solutions like nearest-neighbor filtering, as their lower quality and objectionable flickering are readily visible. Subjects could toggle between two methods via the keyboard without any time limit. Subjects were asked to choose the method for which the reproduction of details is closest to the reference version. The results of this part of experiment are shown in Fig. 6(left). The second part of the study was similar to the first (Lanczos), but moving full-HD resolution images have been compared without any reference (Fig. 6, middle). For this first experiment, the pixel size in the moving image was three times enlarged to match the scale of the reference image. All other experiments used the native resolution.

Next, we tested our method against Mitchell-Netravali [1988] filtering. The filter can be balanced between sharpening and smoothing using two parameters which makes it adequate for a large variety of images. We asked subjects to adjust the parameters to match their preferences with respect to the high-resolution image. Later, they were asked to compare their result with our technique, again by toggling between the methods (Fig. 6, right).

Our technique performed better in terms of detail reconstruction, even when allowing filter parameter adjustments. During all experiments no flickering or temporal artifacts have been observed. A series of t-tests (Tab. 1) showed statistical difference in all cases with a significance level of .05.

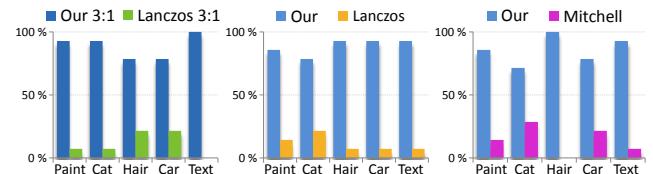


Figure 6: Our method against Lanczos in 3:1 scale (left) and original scale (middle) as well as Mitchell-Netravali (right).

Minimal Text We also investigated horizontally moving text often used for TV news channels, as well as hand-held devices. To push our technique to the limits, we attempted to produce a 2×3 pixel sized font containing English capital letters. We created it by hand at a 6×9 resolution, but did not invest much time in optimizing the characters. We showed all the letters in random order to subjects asking for identification and compared our method to Lanczos filtering. The characters have been placed in chunks of five characters rather than isolated fonts to mimic a text document. We did not compare to static text because any attempts to produce such a small font were futile.

Although not perfect (Fig. 7), the results indicate the quality-increase due to our apparent resolution enhancement. Performed series of t-tests showed significant difference between our font and standard downsampling for 13 out of 26 (Fig. 7) for a significance level of .05.

	Paint	Cat	Hair	Car	Text
Our vs. Lanczos (3:1 scale)					
$t(26)$	8.485	8.485	3.551	3.551	∞
p	< .001	< .001	.002	.002	< .001
Cohen's d	3.207	3.207	1.342	1.342	∞
Our vs. Lanczos					
$t(26)$	5.204	3.551	8.485	8.485	8.485
p	< .001	.002	< .001	< .001	< .001
Cohen's d	1.967	1.342	3.207	3.207	3.207
Our vs. Mitchell					
$t(26)$	5.204	2.419	∞	3.551	8.485
p	< .001	.023	< .001	.001	< .001
Cohen's d	1.967	0.914	∞	1.342	3.207

Table 1: High-resolution images experiment: The table contains t - and p -values as well as effect size (Cohen's d) for pairwise comparison of our method with respect to Lanczos and Mitchel filtering.

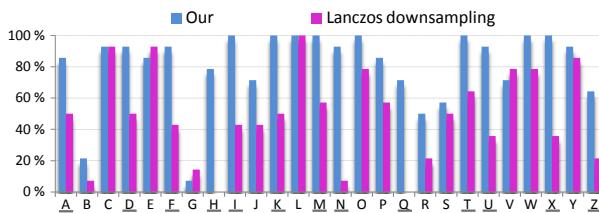


Figure 7: Character recognition: Standard filtering vs. Ours. Dark horizontal lines indicate significant statistical difference.

The biggest improvement was, as expected, in the horizontal direction that coincides with the movement. *H, K, M, N, R, X* contain much horizontally oriented information, making them easy to read. On the other hand, the lack of improvement in vertical direction, affects letters such as: *B, G*.

3D Rendering We also conducted a smaller study for 3D rendering applications. We estimated the eye tracking based on a derived motion flow. We assumed that the motion is piecewise linear for different image regions, and, thus, we can apply our technique locally. We used a scene showing highly detailed hair and a 3D terrain in a fly-over view similar to Google Earth. The stimuli for this experiment are included in the accompanying video. Similar to the high-resolution image experiment, subjects could toggle between moving images for our method and respectively Lanczos and Mitchel filtering. All five subjects chose our solution over Lanczos and Mitchel for both scenes.

7 Discussion

Slightly moving images have become common practice of web-page designers that present scrolling photos, or scrolling text (e.g., news), and small animations. Besides guiding attention and looking more natural and lively (the Ken Burns effect), improved detail perception, as shown in our experiments, might explain this trend. Our experiments suggest that the strongest enhancement can be obtained using our technique, but even frame-wise downsampling (taking into account the current mapping to physical pixels) is a better strategy than naïve resampling of a downsampled image.

Proper filtering becomes even more important for large displays, as illustrated by our study, but big velocities imply the need for higher refresh rates to counteract the hold-type effect.

The optimization scheme delivers a high-quality result, but is computationally costly (e.g., double full-HD image 3840×2160 using three subimages is processed in approx. 18 minutes, standard full HD needed 5 min). Although our CPU-based optimization could be improved, especially using a GPU implementations. Our first experiments with a gradient-descent GPU solver (enforcing constraints in each iteration via clamping), showed that the computation time can be reduced to below 1 s.

Our model does not rely on any profound hardware-specific assumptions (e.g., not on the RGB subpixel layout) which makes our technique relatively immune to technological and perceptual differences. In our experiments with an 120 Hz CRT display as well as 60 Hz DLP and LCD projectors we have obtained a clear resolution enhancement. Therefore, we also expect that our technique works for OLED displays where very high frame rates should lead to an even stronger resolution enhancement. Essentially, our model conforms with the major goals of display manufacturers to reduce the visibility of RGB subpixel layout and screen door effect, which otherwise could ruin the impression of image integrity and continuity.

8 Conclusion

Due to the limited spatial resolution of current displays, the depiction of very fine spatial details is difficult. This work proposed a novel reconstruction for moving images taking human perception into account to improve the detail display. Our work is general in the sense that it extends to arbitrarily high frame rates. This was underlined by a perceptual study. We presented various applications including improved photo details, panorama pop-up views, online rendering, and scroll texts (where we pushed the limits by showing that a 2×3 pixel font can still be legible).

In the future, we want to combine our technique with schemes to exploit the display mosaics similarly to the ClearType fonts. Initial attempts have not led to a clear quality improvement and this issue requires further investigation. We demonstrated the applicability of our approach to offline rendering. In the future, we would like to opt for an online context, eventually combining our solution with an eye trackers to only locally perform the optimization computation. In a first attempt, we also tried to construct directional filters from the results of the optimization process, but (because our optimization is not a filtering) the values often exceed the dynamic range of the display. Via blending we can reduce the contrast, but then the results are clearly inferior to the full optimization and show color aberrations. This remains an exciting avenue for future work.

Acknowledgments

We would like to thank Reinhard Klein, Rafał Mantiuk, Robert Strzodka for helpful discussions as well as Gernot Ziegler and David Luebke of NVIDIA corporation for providing a Samsung SyncMaster 2233 RZ display. This work was partially supported by the Cluster of Excellence MMCI (www.m2ci.org).

References

- ALLEN, W., AND Ulichney, R. 2005. Wobulation: Doubling the addressed resolution of projection displays. In *Proceedings of the Symposium Digest of Technical Papers (SID)*, vol. 47.4 of *The Society for Information Display*, 1514–1517.
- BIJL, P., SCHUTTE, K., AND HOGERVORST, M. A. 2006. Applicability of TOD, MTDP, MRT and DMRT for dynamic image enhancement techniques. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 6207.

- BURR, D. C. 1979. Acuity for apparent vernier offset. *Vision Research* 19, 7, 835–837.
- COLEMAN, T. F., AND LI, Y. 1996. A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM J. on Optimization* 6, 4, 1040–1058.
- CURCIO, C. A., SLOAN, K. R., KALINA, R. E., AND HENDRICKSON, A. E. 1990. Human photoreceptor topography. *The Journal of Comparative Neurology* 292, 4, 497–523.
- DAMERA-VENKATA, N., AND CHANG, N. L. 2009. Display supersampling. *ACM Trans. Graph.* 28, 1, 9:1–9:19.
- DE LANGE, H. 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, 777–783.
- DEERING, M. F. 2005. A photon accurate model of the human eye. *ACM Trans. Graph. (Proc. SIGGRAPH 2005)* 24, 3, 649–658.
- FAHLE, M., AND POGGIO, T. 1981. Visual hyperacuity: Spatiotemporal interpolation in human vision. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 213, 1193, 451–477.
- HARA, Z., AND SHIRAMATSU, N. 2000. Improvement in the picture quality of moving pictures for matrix displays. *J. SID* 8, 2, 129–137.
- KALLONIATIS, M., AND LUU, C. 2009. Temporal resolution. <http://webvision.med.utah.edu/temporal.html>.
- KLOMPENHOUWER, M. A., AND DE HAAN, G. 2003. Subpixel image scaling for color-matrix displays. *J. SID* 11, 1, 99–108.
- KOPF, J., UYTTENDAELE, M., DEUSSEN, O., AND COHEN, M. 2007. Capturing and viewing gigapixel images. *ACM Trans. Graph. (Proc. SIGGRAPH 2007)* 26, 3.
- KRAPELS, K., DRIGGERS, R. G., AND TEANEY, B. 2005. Target-acquisition performance in undersampled infrared imagers: static imagery to motion video. *Applied Optics* 44, 33, 7055–7061.
- KRAUZLIS, R., AND LISBERGER, S. 1994. Temporal properties of visual motion signals for the initiation of smooth pursuit eye movements in monkeys. *J Neurophysiol.* 72, 1, 150–162.
- LAIRD, J., ROSEN, M., PELZ, J., MONTAG, E., AND DALY, S. 2006. Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli. In *Human Vision and Electronic Imaging XI*, vol. 6057 of *SPIE Proceedings Series*, 32–43.
- MÄKELÄ, P., ROVAMO, J., AND WHITAKER, D. 1994. Effects of luminance and external temporal noise on flicker sensitivity as a function of stimulus size at various eccentricities. *Vision Research* 34, 15, 1981–91.
- MARTINEZ-CONDE, S., MACKNIK, S. L., AND HUBEL, D. H. 2004. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience* 5, 3, 229–239.
- MCKEE, S. P., AND TAYLOR, D. G. 1984. Discrimination of time: comparison of foveal and peripheral sensitivity. *J. Opt. Soc. Am. A* 1, 6, 620–628.
- MESSING, D. S., AND KEROFSKY, L. J. 2006. Using optimal rendering to visually mask defective subpixels. In *Human Vision and Electr. Imaging XI*, vol. 6057 of *SPIE Proc. Series*, 236–247.
- MITCHELL, D. P., AND NETRAVALI, A. N. 1988. Reconstruction filters in computer-graphics. *Proc. SIGGRAPH* 22, 4, 221–228.
- PARK, S., PARK, M., AND KANG, M. 2003. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine* 20, 3, 21–36.
- PLATT, J. 2000. Optimal filtering for patterned displays. *Signal Processing Letters, IEEE* 7, 7, 179–181.
- PURVES, D., SHIMPI, A., AND LOTTO, B. R. 1999. An empirical explanation of the Cornsweet effect. *J. Neuroscience* 19, 19, 8542–8551.
- SCHÜTZ, A. C., BRAUN, D. I., KERZEL, D., AND GEGENFURTNER, K. R. 2008. Improved visual sensitivity during smooth pursuit eye movements. *Nat. Neuroscience* 11, 10, 1211–1216.
- TEKALP, A. 1995. *Digital Video Processing*. Prentice Hall.
- VAN HATEREN, J. H. 2005. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *J. Vision* 5, 4, 331–347.
- WANDELL, B. 1995. *Foundations of Vision*. Sinauer Associates.
- ZAVAGNO, D., AND CAPUTO, G. 2001. The glare effect and the perception of luminosity. *Perception* 30, 2, 209–222.