Resolution Enhancement by Vibrating Displays

FLORAINE BERTHOUZOZ
University of California, Berkeley
and
RAANAN FATTAL
Hebrew University of Jerusalem

We present a method that makes use of the retinal integration time in the human visual system for increasing the resolution of displays. Given an input image with a resolution higher than the display resolution, we compute several images that match the display's native resolution. We then render these low-resolution images in a sequence that repeats itself on a high refresh-rate display. The period of the sequence falls below the retinal integration time and therefore the eye integrates the images temporally and perceives them as one image. In order to achieve resolution enhancement we apply small-amplitude vibrations to the display panel and synchronize them with the screen refresh cycles. We derive the perceived image model and use it to compute the low-resolution images that are optimized to enhance the apparent resolution of the perceived image. This approach achieves resolution enhancement without having to move the displayed content across the screen and hence offers a more practical solution than existing approaches. Moreover, we use our model to establish limitations on the amount of resolution enhancement achievable by such display systems. In this analysis we draw a formal connection between our display and super-resolution techniques and find that both methods share the same limitation, yet this limitation stems from different sources. Finally, we describe in detail a simple physical realization of our display system and demonstrate its ability to match most of the spectrum displayable on a screen with twice the resolution.

1. INTRODUCTION

Computational photography is a new research trend that examines how computation, done both in software or through novel hardware, can be used to enhance images or extend the capabilities of digital photography. One challenging and open-ended problem in this area is the design of imaging systems that can display richer information about a scene. While nowadays consumer camera sensor resolution exceeds ten mega-pixels, current displays' resolution falls behind and does not reach half that resolution.

Increasing the resolution of a display is a challenging engineering problem, because it requires aligning more fully functional pixel units into a smaller area. This problem is most acute in projectors, 3D, HDR and stereo displays as well as street TVs. The common practice is to downscale images to the display resolution at the cost of losing fine image details. As an alternative, researchers propose to combine multiple lower-resolution projectors to display a higher resolution image. For example, Damera-Venkata and Chang [2007b] superimpose different images projected from up to ten projectors. These images are optimized to produce the target high-resolution image. This approach must take into account intra- and inter-projector variations and requires careful geometric, photometric and color calibrations of each component projector. Allen and Ulichney [2005] describe a way to jitter light by a precise mirror shifting mechanism, inside a projector, to achieve higher-resolution diamond pixel patterns. Didyk et al. [2010a] bring these resolution enhancement efforts to LCD displays. However, since multiple displays cannot be combined together, as in the projector setting, they exploit properties of the human visual system to increase the apparent resolution of still images by moving them across the screen. When tracking a moving image, the eye integrates light at different offsets along the motion trajectory. By rapidly displaying different images at these sub-pixel offsets, the eye integrates the images into a single perceived image that can be of higher resolution than any of the images displayed. Didyk et al. generate a set of low-resolution images (LRIs) from a single high-resolution input image that are optimized to enhance the apparent integrated-image resolution when rapidly displayed at sub-pixel offsets. This solution is cost-effective, requires a single consumer-level high refresh-rate LCD display and does not require any calibration step.

In this work, we also exploit the retinal integration to increase the apparent resolution of LCD displays by applying small sub-pixel movements to the display. Given a single high-resolution input image we generate several LRIs that we display in synchronization with the display movements and at periods that fall below the retinal integration time. The static eye gaze integrates the images temporally and perceives them as a single higher-resolution image. Thus, we do not rely on eye tracking and avoid the need to move the displayed content across the screen. Our realization moves the display along a small circular trajectory similar to the one described

(a) High-resolution input image   (b) One of the four low-resolution images generated by our method   (c) Perceived image on the vibrating display (showing four low-resolution images)
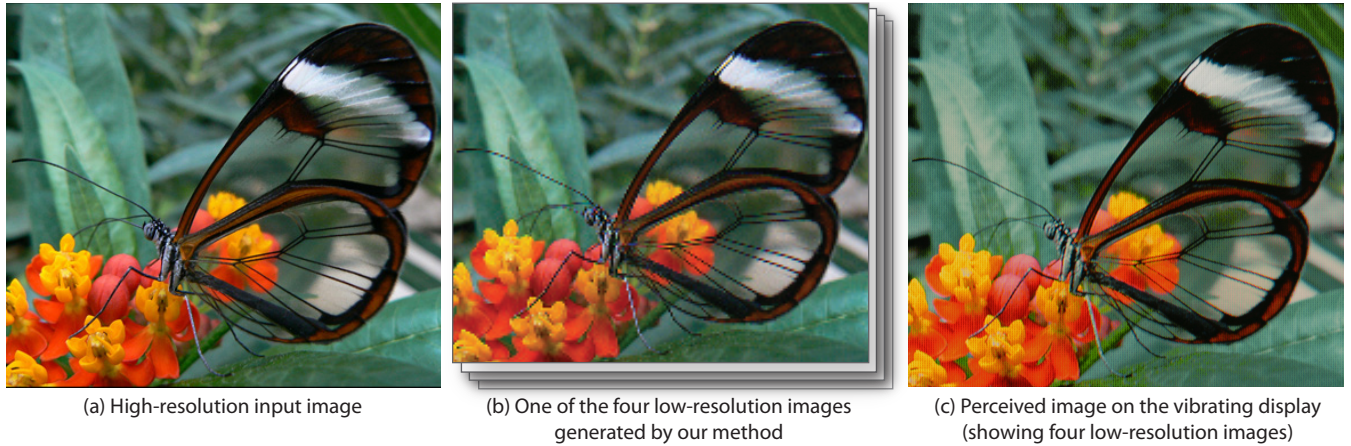
Fig. 1.   Given a high resolution input image (a), our method generates four low-resolution images (b). These four images are integrated by the observer's eye when viewed on our vibrating display and lead to a perceived image of higher resolution (c). The perceived image is an actual photograph taken of the display. Image credits: *Eddy Van Leuven*.

in a patent by Damera-Venkata [2009] which we implement on an LCD. We achieve these very small amplitude vibrations using a simple low-cost mechanism found in cellular phones, see Figure 2. Unlike the linear trajectories, used by Didyk et al., our planar trajectories allow us to achieve resolution enhancement in both image axes.

Besides offering this prototypical solution, in our work we explore various aspects of such display systems, including the derivation of a perceived image model that incorporates accurately measured components, the realization of such systems with a detailed description of necessary calibration steps, and the description of various measurements that quantify the success of such display systems. Moreover, we study various signal processing aspects of our display system. Specifically, we analyze the spectral properties of the matrix that models our system and generates the LRIs. This analysis formally connects our method to multi-frame super-resolution techniques and shows that the resolution enhancement of our system shares a common limitation known to exist in super-resolution [Baker and Kanade 2000; Lin and Shum 2001]. However, we show that the source of the limitation in the two systems is different and discuss options of overcoming it in our case by trading image contrast for spatial resolution.

Figure 1 shows an actual result we obtained, where the perceived (photographed) image shows most of the details available in the input image which is at twice the screen's resolution. Notice how the antennae, the wings and legs of the butterfly appear sharp and well-resolved in the perceived image. To conclude, our system allows us to match images of twice the screen resolution without moving them across the screen. Thus, it offers a practical solution to resolution enhancement. We discuss other potential mechanical and optical alternatives for our implementation.

2.   RELATED WORK

In addition to the works discussed above our display system builds on several other works from different areas.

Perception-Based Techniques.   Our eyes periodically sample information of images projected onto the retina and integrate this information temporally to produce the perceived image. It takes the eye a certain amount of time to process the sampled information
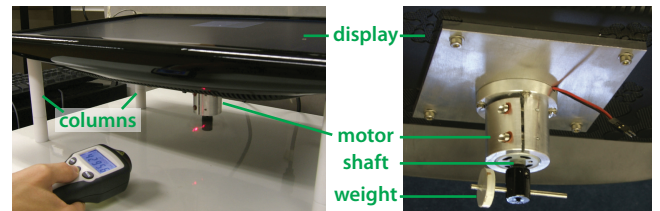

Fig. 2.   To vibrate the display, we attach to the back of it a motor with a weight connected at an offset to the motor shaft (right). We then mount the display on four columns (left). As the motor is rotating centrifugal forces pull the screen towards the weight and cause the screen to vibrate in a circular trajectory.

and it is therefore limited in how fast it can respond to changing stimuli. If the rate of change of a stimulus is faster than what the eye can process, it will be perceived as a single continuous stimulus and the images that are projected onto the retina are integrated temporally [Kalloniatis and Luu 2009]. Several existing display systems exploit this property of the visual system. For example, video interlacing used in CRT television sets is a technique that alternates between displaying the even and odd horizonal video scanlines of the screen. Even though only half of the pixels are shown at every instant, the viewer integrates this information temporally and perceives one continuous image. The same principle is used in digital light processing projectors that display the RGB color components sequentially and rely on the human eye to temporally merge the components. As we discussed in the Introduction, recent works in computer graphics also use the temporal and spatial integration properties of the eye to increase the resolution of displays spatially [Didyk et al. 2010a] as well as the temporal and spatial resolution of computer-generated videos [Didyk et al. 2010b; Templin et al. 2011]. Our display system is inspired by this line of work.

Note that if we fail to display the images within the retinal integration time, the viewer will not fuse the images and experience a sensation of flickering. This display rate at which flickering starts appearing is known as *critical flicker frequency* (CFF). The CFF is a complex phenomenon that varies with many factors, including the

size of the stimulus pattern [Mäkelä et al. 1994; McKee and Taylor 1984] and its intensity [Norwich 1993]. Didyk et al. [2010a] propose an effective post-processing method to prevent flickering in the setting of high refresh-rate displays, but their solution may reduce the amount resolution enhancement gained. While we do not address the problem of flickering in our work, Didyk et al.'s method would apply to our setting without any modification.

Projector Super-Sampling. Jaynes and Ramakrishnan [2003] exploit the ability to overlap images projected from multiple projectors to enhance resolution. They generate one LRI for each projector in a sub-optimal greedy approach. Damera-Venkata and Chang [2007b] push this idea farther and derive an optimal LRI generation algorithm for the same setting of multiple projectors. They also analyze how non-uniform displacements between the LRIs influences the resolution enhancement they obtain [Damera-Venkata and Chang 2007a]. Finally, they show that projector super-resolution may be achieved by designing a bank of alias-canceling rendering filters and derive a practical non-iterative filter bank approach for their previous technique [Damera-Venkata and Chang 2009]. Our work is similar to multi-projector techniques in the sense that we also generate an optimal set of LRIs from a high-resolution input image. But in our case, the target device is a single display. Our method therefore does not have to account for intra- and inter-projection variations that require careful geometric, photometric and color calibrations. Instead, we are able to achieve resolution enhancement by exploiting the temporal integration properties of the human eye. As we noted earlier, the idea of displaying shifted images using a single display was also described by Allen and Ulichney for projectors and in [Damera-Venkata 2009]. However, Allen and Ulichney do not further explore how to optimally generate the displayed LRIs, nor do they discuss the limits of such an approach. Damera-Venkata's patent merely describes the idea of such a display system with no analysis, evaluation or theoretical explanation. Finally, Napoli et al. [2008] suggest an alternative approach where they use a single projector and deviate the light rapidly using an aperture array. This allows them to display up to 300 non-overlapping LRIs that form a new higher resolution grid. In contrast, our approach uses the overlaps between the LRIs to gain resolution.

Multi-frame Super-Resolution. Our work is also closely related to super-resolution techniques [Irani and Peleg 1990; Park et al. 2003], where a high-resolution image is computed from a set of LRIs taken at different translational offsets. We have the opposite goal of super-resolution techniques; given a high-resolution image, we compute a set of LRIs and display them at different offsets. Previous studies [Baker and Kanade 2000; Lin and Shum 2001] show that super-resolution has a practical limit on the maximal magnification factor it can achieve due to the strong amplification of noise at high-frequencies. We formally link our construction of the LRIs to super-resolution and analyze and deduce limits on the resolution enhancement achievable on a given display. Our work shares similarities with the work of Ben-Ezra et al. [2004] on super-resolution techniques for video enhancement. In this work, they also use physical motion and shift the camera sensor using micro-actuators to obtain precise offsets between the LRIs.

3. METHOD

In order to compute the set of LRIs that will optimally display the high-frequency content of the input high-resolution image, we model the perceived image generated by our proposed imaging system. We start this derivation by defining the continuous function that describes how a single image L viewed on a stationary display

appears on the screen,

$$V_{static}(y) = \sum_{x=1}^{N} \int_{0}^{T} P(y-x,t)L(x)dt, \qquad (1)$$

where x is an integer pixel index, N is the number of pixels in L, y is a real-valued point on the screen plane, T is a period in time and P is the point spread function (PSF) of the display. The point spread function describes the shape of a single-pixel turned on as it appears on the display at a given time. We allow P to depend on time in order to account for frame-rate control, overdrive and other techniques typically used by modern displays to control the appearance of a pixel over time. The static image is an aggregate over time that the viewer perceives.

Vibrating the display along some periodic trajectory and rapidly displaying several images within each period, that *falls below* the retinal integration time, yields the following perceived image

$$V_{vibrating}(y) = \sum_{x=1}^{N} \int_{0}^{T} P(y-x-\varphi(t),t)L_t(x)dt, \qquad (2)$$

where T is the time period of a single vibration and $\varphi(t)$ indicates the display position as it moves along the trajectory at time t. However, since we display a small finite number of images, L_t ceases to be a continuous function in time and the above becomes

$$V_{vibrating}(y) = \sum_{x=1}^{N} \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} P(y-x-\varphi(t),t)L_i(x)dt = \sum_{x=1}^{N} \sum_{i=1}^{n} S_i(y-x)L_i(x),$$
$$(3)$$

where n is the number of LRIs and

$$S_i(y-x) = \int_{t_i}^{t_{i+1}} P(y-x-\varphi_i(t),t)dt \qquad (4)$$

where $[t_i, t_{i+1}]$ is the time interval during which the i-th image is displayed. We call the time-averaged display point spread functions, S_i, *smeared point spread functions* (SPSFs). Note that the SPSFs are aggregates over time and thus implicitly take into account variations of P over time. In Figure 3 we show an example of a static PSF, the trajectory along one quarter of a circular trajectory (φ_i shown by the red curve), and the result of smearing the PSF along this trajectory (S_i). In Section 3.1 we explain how we acquire these PSFs for our display.

We compute an optimal approximation to a given high-resolution image $H(y)$, in the L_2 sense, through the following minimization problem

$$\min_{L_1 \dots L_n} \int \Big(\sum_{x=1}^{N} \sum_{i=1}^{n} S_i(y-x)L_i(x) - H(y) \Big)^2 dy. \qquad (5)$$

In practice, we solve this minimization problem on a discrete grid, in which the y coordinate is discretized. The input image H and the SPSFs are specified at this grid. We call it the *computational grid* and set its resolution to be high enough such that the shape of the SPSFs is properly captured. As we explain in Section 3.1, in our implementation we use a computational resolution that is eight times higher than the LRIs' resolution. Note that setting the computational resolution to eight does not mean that we will be able to achieve a resolution enhancement by that factor. We distinguish between the computational resolution used for computational purposes only, and the *target resolution* which indicates the amount of true resolution enhancement that we can obtain with our method. The target resolution is dictated by the number and the arrangement of the LRIs on the display, as well as other factors
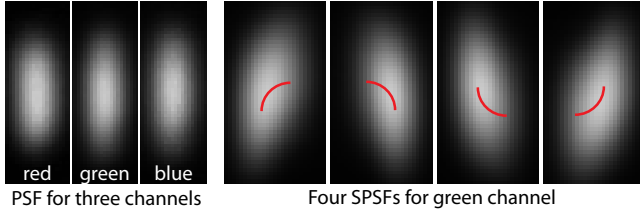
Fig. 3. (left) The time-averaged static PSFs for the red, green and blue channels. (right) The four SPSFs generated for the green channel. We show in red the trajectory along which the static PSFs were smeared. These images were magnified by a factor 32, such that in these photographs the center of two static PSFs of the same color are spaced apart by 32 pixels.

that result from the spectral properties of the SPSFs (see discussion in Section 5). We define this resolution as the maximal number of possible matched Fourier components. In other words, if we use n LRIs then, under ideal conditions, we will at most be able to match n times the number of Fourier components in each of the LRIs. The target resolution is therefore equal to the total number of pixels in the n LRIs, which is typically much lower than the computational resolution. This implies that the high-resolution image $H(y)$, specified at the computational resolution, does not need to contain more content than the amount needed to fill the spectrum of the target resolution. In practice, if we have an input image specified at the target resolution, we interpolate it (e.g., via ideal filtering) to the computational resolution to obtain $H(y)$.

Once discretized, the minimization problem in (5) can be expressed in matrix form. We define a matrix $\mathbf{W}_i$ for each LRI L_i. $\mathbf{W}_i$ contains, as its row vectors, the SPSF for every pixel x in L_i captured at the computational resolution. By combining these matrices into one matrix $\mathbf{W} = [\mathbf{W}_1, ..., \mathbf{W}_n]$, (5) becomes

$$\min_L \left(\mathbf{W}L - H \right)^2, \tag{6}$$

where $L = [L_1, ..., L_n]^T$ is the column vector containing the n LRIs that we are solving for and H is the vector of the input high-resolution image, discretized at the computational grid. The product $\mathbf{W}L$ gives an image, which we call the *approximated image*, that is produced by L and approximates the high-resolution image H at the computational grid. The minimum of (6) is obtained by solving the following linear system

$$\mathbf{W}^T \mathbf{W} L = \mathbf{W}^T H. \tag{7}$$

We define and solve this equation for each color channel (red, green, and blue) independently. The matrix $\mathbf{W}^T \mathbf{W}$ is highly sparse due to the limited overlap between the display SPSF functions. Such positive definite matrices are solved efficiently via iterative linear solvers. We use the Gauss-Seidel iteration with a small modification. Pixel values must be confined to a limited range $[0, 1]$, because an LCD cannot produce 'negative' light intensities and also has a maximal intensity it can emit. Therefore, similarly to [Damera-Venkata and Chang 2007b], we clamp the LRIs pixel values, $L_i(x)$, after every Gauss-Seidel iteration by setting them to zero and one if they run below or above these values respectively. This operation corresponds to a *projections onto convex sets* scheme [Youla 1978] and it is known to converge to the optimal solution. We provide pseudo-code for this procedure in Algorithm 1. The convergence speed and hence the efficiency of iterative linear solvers, including Gauss-Seidel, depend on the condition number of the matrix $\mathbf{W}^T \mathbf{W}$. In Section 5 we analyze this aspect and discuss the implications.

Algorithm 1 Modified Iterative Gauss-Seidel Solver

Define $\mathbf{A} \leftarrow \mathbf{W}^T \mathbf{W}$
$\qquad\quad b \leftarrow \mathbf{W}^T H$
$L^{(0)} \leftarrow b / Diag(\mathbf{A})$ (initialization)

$L_i^* \leftarrow L_i^{(k)} - \beta \mathbf{A}_{ii}^{-1} (\sum_{j<i} \mathbf{A}_{ij} L_j^{k+1} + \sum_{j \geq i} \mathbf{A}_{ij} L_j^k - b_i)$ (update)
$L_i^{(k+1)} = \begin{cases} L_i^* & \text{if } L_i^* \in [0, 1] \quad \text{(clamping)} \\ 0 & \text{if } L_i^* < 0 \\ 1 & \text{if } L_i^* > 1 \end{cases}$

3.1 Implementation

We realize our method using a 120Hz Samsung SyncMaster 2233RZ monitor and display $n = 4$ LRIs at 30Hz. We apply small circular vibrations to the display within the LCD panel plane using the same mechanism used in current cellular phones. Here we describe the various design choices, techniques, and measurements we performed to construct our prototype display system.

Number of LRIs. A 120Hz refresh-rate display offers two natural choices; we can either render $n = 3$ images at 40Hz or $n = 4$ at 30Hz. We need to display the images fast enough so that the eye will temporally integrate all LRIs into a single stable perceived image. As we explain in Section 2, this rate is called critical flicker frequency (CFF) and displaying images below this rate can lead to perceived flickering [Kalloniatis and Luu 2009]. While existing reports quote 40Hz as the safe frequency in terms of apparent flickering [Didyk et al. 2010a], we opted for the choice of $n = 4$ LRIs at 30Hz to achieve higher resolution enhancement. This rate is equal or higher than the frequency at which interlaced television standards operate (NTSC at 30Hz and PAL at 25Hz). Previous research in visual perception indicates that flickering is mostly visible in large uniform regions [Mäkelä et al. 1994; McKee and Taylor 1984]. Because there are no high-frequencies in such regions, the pixel intensities of the four LRIs that we generate are close to one another (and equal to the input image) in these areas. Therefore, our method does not cause any flickering in large uniform regions, which correspond to the majority of the pixels in images. But when displaying the LRIs at 30Hz, some flickering can still take place in high-frequency areas (e.g., along strong edges).

To reduce flickering, we could either use a higher refresh rate display, or use the post-processing step proposed by Didyk et al. [Didyk et al. 2010a]. Since flickering is caused by strong intensity changes in the LRIs, their post-processing approach averages the LRIs with Lanczos-filtered images at pixels where the temporal intensity change exceeds a threshold. Nevertheless, similarly to the option of using $n = 3$ LRIs at 40Hz, averaging with filtered images undermines the amount of resolution enhancement.

Display PSF Acquisition. We acquire the static display PSF (P) for each color channel by photographing the LCD panel as it displays a single red, green and blue pixel each time. These functions were acquired at a very high resolution using a macro magnifying lens. Assuming the PSFs are time-independent we acquired this measurement by long time-averaging exposures. Under this assumption, we generate the SPSFs, S_i, by convolving P with a function that describes the trajectory covered by the image L_i (four quarters of a circle in our case). Figure 3 shows both the acquired static PSFs and the computed SPFSs. Note that this accurate estimation of the PSFs could also be used in the method of Didyk et al. [Didyk et al. 2010a].
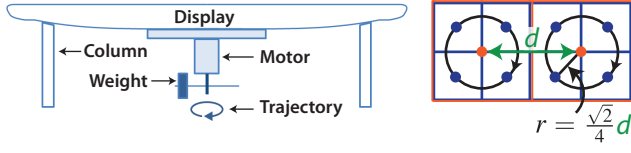
Fig. 4. (left) Vibrating display construction. (right) Computation of the optimal vibration radius: we show two display pixels marked by orange pixel centers and boundaries. The optimal vibration radius of the circular trajectory (black) is such that the pixel centers of all the L_is (shown by blue circles) form a new regular grid of double the resolution shown in blue.

Computational Resolution. We set the computational resolution to be eight times higher than the

resolution of the LRIs. The spectral properties of the SPSFs dictate this choice. By discretizing (5) and the SPSFs in image space, we introduce periodicity to their spectrum in the Fourier domain. In the inset we show the magnitude spectrum[1] of $\hat{S}$ and one of its repetitions, where $\hat{S}$ is the horizontal cross-section of the SPSF in Fourier domain. The proximity of the repetitions in Fourier domain depends on the resolution of the computational grid. The higher the resolution of the grid, the farther away the repetitions will be. Like other super-resolution algorithms [Hardie 2007], our goal is to set the computational resolution high enough, such that $\hat{S}$ is not influenced by any of the other repetitions and we can properly capture the shape of the SPSFs. The inset shows that this is achieved reasonably well by making the signal 16π periodic, since $\hat{S}$ decays enough at 8π. A periodicity of 16π corresponds to a computational resolution of eight times the resolution of the LRI.

Circular Trajectory. For reasons of simplicity and similarly to [Damera-Venkata 2009] we vibrate the LCD in a circular trajectory at the display panel plane. This is done via a simple mechanism found in current cellular phones. We mount a motor with an unbalanced weight at the back of the screen, as shown in Figure 2 and 4. The distance between the weight and the motor shaft allows us to control the vibration radius. We use a mass of 20 grams and offset it by 17 millimeters from the shaft axis. As the motor is rotating, the weight's motion exerts a centrifugal force pulling the screen towards the weight. In our realization we positioned the screen horizontally and mounted it on four columns touching the back of the screen by elastic (rubber) pads (Figure 4 left).

This circular vibration trajectory has several advantages. Besides the simple mechanism needed to produce it, it corresponds to the closed trajectory with a minimal maximal curvature and hence involves minimal forces acting on the LCD panel. Another advantage is that, unlike linear trajectories, this planar motion allows us to achieve resolution enhancement for both image axes. The resolution enhancement is strongest for directions that are perpendicular to the displacement, because each LRI displays new samples along these directions. Linear trajectories, such as the ones used in [Didyk et al. 2010a], affect only a single axis, the direction perpendicular to the trajectory, and therefore edges parallel to the displacement cannot be enhanced. For instance, an image that does not change along the trajectory will not gain resolution.

---

[1]Throughout the paper we use the *magnitude spectrum* to measure and show the spectral behavior of various component. The magnitude spectrum is defined as the absolute value of the Fourier transformed values.

Trajectory Radius. Every LRI L_i is displayed over one quarter of the circular path of the vibration. If we model each SPSF by its center of mass, we can find the vibration radius that produces a new *regular* grid with a resolution that is double the LRIs' resolution. These geometrical considerations, depicted in Figure 4, show that the vibration radius is $r = \sqrt{2}d/4$, where d is the distance between two adjacent pixels (in physical coordinates). We found that this radius is close to the optimal choice by scanning exhaustively every $r \in [0.3d, 0.4d]$ at intervals of 2×10^{-3}. In this process, we generate the SPSF for every radius and use these SPSFs to compute the four LRIs. We then evaluate the L_2 norm between the approximated image at the computational grid (**W**L) and the high-resolution input image (H). We found that the best approximation is found for $r = 0.332d$, which deviates from the theoretical radius by 6%. The difference can be attributed to the fact that the display SPSFs are not points but have elongated and curved shapes.

Display Calibration. In order to be consistent with the model of our display system, which we use to generate the optimal LRIs, we need to calibrate the speed and radius of the LCD vibrations. We measure the rotation frequency using an accurate tachometer and set the rotation speed to 30Hz by varying the voltage applied to the motor. However, small discrepancies in the vibration speed drive the computer (displaying the LRIs) and the screen rotations in and out of sync. While we did not address this problem with our prototypical construction, a full synchronization can be achieved by a simple optical or electronic sensor that will inform the computer once every cycle starts or ends. We set the radius of the display vibrations using a trial and error process. We photograph the vibrating display showing a single pixel turned on, with a camera exposure time set to an integer times $1/30$th of a second in order to capture a full vibration circle. We then convolve the static PSF, P, with circles of different radii and search for the radius that best approximates the photograph in L_2 norm. Based on these readings we modify the weight distance from the shaft axis, until we get a radius of $0.332d$. This calibration step was done once to produce all the results reported in the paper.

3.2 Resolution Enhancement

Having described the method behind our new display system and its prototypical realization, we briefly review the spectral properties that allow it to achieve resolution enhancement. In Fourier space, the spectrum of an image at the display's resolution lies within $[-\pi, \pi]$. Our goal when enhancing the resolution of the diplay is to expand that spectrum to $[-2\pi, 2\pi]$. We explain the spectral properties that make it possible for our display system to produce and pass frequencies within that range. Let us begin by analyzing the static display model equation (1) in Fourier space,

$$\hat{V}_{static}(\omega) = \hat{P}(\omega) \cdot \left(\hat{L}(\omega) * 2\pi \sum_{k=-\infty}^{\infty} \delta(\omega - 2\pi k) \right) \qquad (8)$$

where the spacing between pixels in the LCD is assumed to be one. Naturally, the discretized image values produce a periodic spectrum (bracketed term above) with periods of 2π. The PSF P acts as a reproducing kernel that attenuates the extra high-frequency aliases to produce a continuous image with an appropriate spectrum band. However, due to physical constraints, the PSF must be localized in space and positive and hence does not provide ideal reconstruction filtering. Specifically, $\hat{P}$ does not drop to zero beyond $|\omega| > \pi$ nor does it maintain the same transfer magnitude within $|\omega| \leq \pi$. Figure 5 shows the magnitude spectrum of $\hat{P}$ of the Samsung SyncMaster 2233RZ display we use. This graph shows that
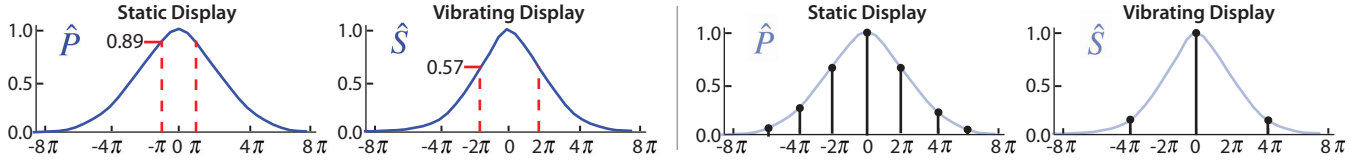
Fig. 5. The magnitude spectrum of the horizontal cross-sections of the display PSF ($\hat{P}$) and SPSF ($\hat{S}$). (left) The dashed red lines show the extents of one period and mark the highest frequency the static or vibrating display can pass. (right) Grid effects when displaying a constant image on the static and vibrating displays. The Fourier transform of a discrete constant image are periodic delta spikes. In the static configuration, the first and strongest repetition of the delta spike appears at $\pm 2\pi$, whereas in the vibrating case it appears in $\pm 4\pi$ where it is considerably lower as it is attenuated by $|\hat{S}(\pm 4\pi)|$.

for the horizontal $\omega = (\pi, 0)$ mode, which is the highest horizontal frequency available in L, the magnitude spectrum of $\hat{P}$ at π reduces by 11% compared to $|\hat{P}(0)|$. Clearly, these design choices attempt to balance between the ability to maintain the same brightness level when displaying content at different frequencies and filtering-out spurious high frequencies.

In our construction we exploit the fact that the PSF $|\hat{P}|$ and the SPSFs $|\hat{S}_i|$ decay slowly and do not completely eliminate frequencies beyond π. Figure 5 shows $|\hat{S}_i|$ which appears slightly narrower due to the smoothing S_i undergoes when being smeared in space. Nevertheless, $|\hat{S}_i|$ still manages to pass 57% of the content at the 2π horizontal mode which corresponds to double the highest frequency in L and is the highest frequency we target with our construction. The ability to display frequencies beyond π and up to 2π requires the system to generate spectrum with periods $\geq 4\pi$. This is what we achieve by vibrating the display and can be easily verified by considering the Fourier transformation of Equation (3) that models our system. For simplicity we do this in one dimension with $n = 2$ and get

$$\hat{V}_{vibrating}(\omega) = \sum_{i=1}^{n} \hat{S}_i(\omega) \cdot \left(\hat{L}_i(\omega) * 2\pi \sum_{k=-\infty}^{\infty} \delta(\omega - 2\pi k) e^{-j\varphi_i \cdot \omega} \right),$$

(9)

where the factor $e^{-j\varphi_i \cdot \omega}$ results from the spatial offset between the SPSFs of different LRIs. Assuming these centers form a uniform grid by setting the vibration radius $r = \sqrt{2}d/4$ then, as shown above, this yields effective pixel spacing which is half the original and hence $\varphi_1 = 0$ and $\varphi_2 = 1/2$. Therefore, the $e^{-j\varphi_i \cdot \omega}$ factor in (9) becomes $(-1)^k$ for $i = 2$ and thus the discretized signal produced by L_1 and L_2 generates a 4π periodic spectrum.

To conclude, we achieve resolution enhancement by mimicking a denser pixel layout by offsetting the display and benefit the spectral properties of the display PSFs that transfer the added resolution.

4. RESULTS

To evaluate our construction, we run our method on a set of 27 high-resolution input images. Figure 6 shows three example results of a text image and two natural images. Please view our supplemental website for additional results. The figure shows: (a) the target image, (b) a single image of the same resolution as one of the low-resolution images displayed on a static display, and (c) the resulting image we get by displaying our four LRIs on the vibrating display. The static image (b) is a reference to the image the display's native resolution can produce. It has the number of pixels of a single LRI and it is computed by filtering and subsampling the target image. This operation is the most common approach used when trying to display an image that has a greater resolution than the display resolution and corresponds to bicubic downsampling in image manipulation software such as Adobe Photoshop. We captured both the

static image (b) and the vibrating image (c) that a user would perceive when viewing our display by photographing the LCD panel at an exposure time equal to an integer times 1/30th of a second. Figure 6 also shows the four LRIs used for displaying the spider's web image. The results on the vibrating LCD show most of the fine detail that exists in the target image. This level of detail is missing in the images displayed on the static LCD. For example, both the text letters and the spider's web threads are much better resolved in the vibrating display.

In addition to resolution enhancement, the appearance of an underlying grid is greatly reduced in the vibrating display compared to the static configuration. This is mostly apparent in flat regions and it can be formally explained when displaying a constant image. This case corresponds to plugging a delta spike as the LRIs, $\hat{L}(\omega) = \hat{L}_i(\omega) = \delta(\omega)$ in equations (8) and (9) that model, in Fourier space, the perceived images in the static and vibrating display respectively. As we explained earlier, the spectrum of the discrete image values will contain the delta spike at periods of 2π in the static setting and at periods of 4π in the vibrating case, as shown in Figure 5 (right). Therefore, in the static case there will be a strong peak at $|\omega| = 2\pi$, because the delta spike is multiplied by a large value $|\hat{P}(\pm 2\pi)| = 0.68$. These strong repetitions of the delta spike cause aliaising that is perceived as an underlying grid in the image. In contrast, in the vibrating display the first repetitions of the delta spike appear at $|\omega| = 4\pi$ by which time it is attenuated by $|\hat{S}(\pm 4\pi)| = 0.15$. Thus, the grid effect is significantly reduced.

Flickering occurs when displaying the LRIs at an insufficient rate. In our experience, when displaying $n = 4$ LRIs at 30Hz we did not perceive any flickering for natural images. But when viewing images with very high-contrast edges such as text images or frequency modes (sinusoidal gratings in Figure 7), we observed flickering comparable to the one in PAL or NTSC displays. As mentioned in Section 3.1 using fewer LRIs, or existing post-processing methods for such display systems [Didyk et al. 2010a] would reduce the flickering, but might also undermine the amount of resolution enhancement.

4.1 Quantitative Evaluation

In order to better quantify the amount of resolution enhancement obtained, we compare the spectral content of the resulting perceived images to that of the input target images. In this evaluation we also compare to the approximated image (**W**L) to evaluate the precision of our construction and calibration steps. This test is performed on two types of images: (a) vertical and horizontal sinusoidal grating images of increasing frequency, and (b) an image composed of four natural images, all shown in Figure 7. The sinusoidal function we use is $\sin(cx^2)$ where $x \in [0, \pi]$. The highest frequency of the sinusoidal gratings is controlled by c and we set it to be the highest frequency available in the target resolution,

(a) Target Image   (b) Static Display Image   (c) Result

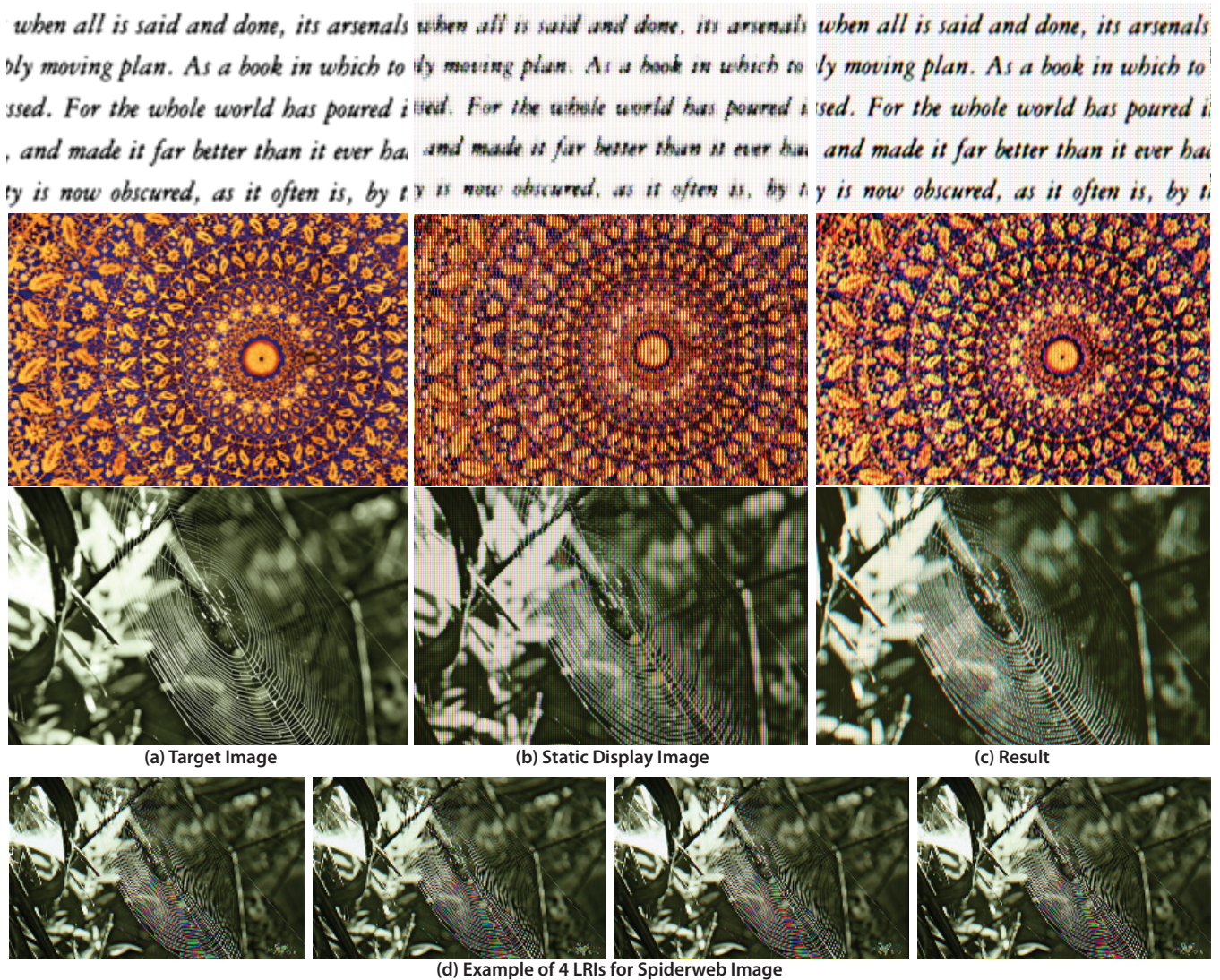(d) Example of 4 LRIs for Spiderweb Image

Fig. 6. A comparison between (left to right): the target input image, an image shown on a static display, and the image seen on our vibrating display using four LRIs. Below we show the four LRIs used to produce the spider's web image. We perform a global linear color correction to the captured (photographed) static and vibrating images to match the colors of the target image. This operation corrects for color shifts that occur in the photographed images due to room lighting and camera specifications when compared to the target image. Image credits for spiderweb image: *Kelsey J. Nelson*.

which is twice the maximal frequency in the LRIs. We also linearly map the range of this sinusoidal function from $[-1, 1]$ to $[0.1, 0.9]$ to reduce effects due to the clamping stage in the LRI construction (see Algorithms 1). We discuss this effect in the next section. In Figure 7 we show the logarithm of the magnitude spectrum of the target image, the approximated image and the result seen on the vibrating display. Note that although each of the LRIs contains frequencies up to π, our results and the approximated images almost double this range and pass frequencies up to 2π. Our results and the approximated images follow closely the target images in the horizontal modes, but show some deviation at the high frequencies in the vertical modes. We attribute this difference to the fact that the display SPSFs are more elongated in the vertical axis (Figure 3), and therefore do not span the vertical high-frequencies as well as the horizontal ones. The figure also shows the percentage

of the magnitude spectrum that our results and the approximated images achieve compared to the target images at every frequency. At the upper 1/16th frequency band ($\omega \in [30/16\pi, 2\pi]$) our result and the approximated image achieve 84% and 90% of the content in the horizontal grating test respectively. These numbers reduce to 61% and 62% in the vertical grating image. For natural images, our results and the approximated images contain 94% and 97% of the spectral content for the upper 1/16th frequency band and 69% and 74% of the vertical spectral content respectively. We conclude that our system allows us to match most of the content available in the input image which is at twice the screen's resolution.
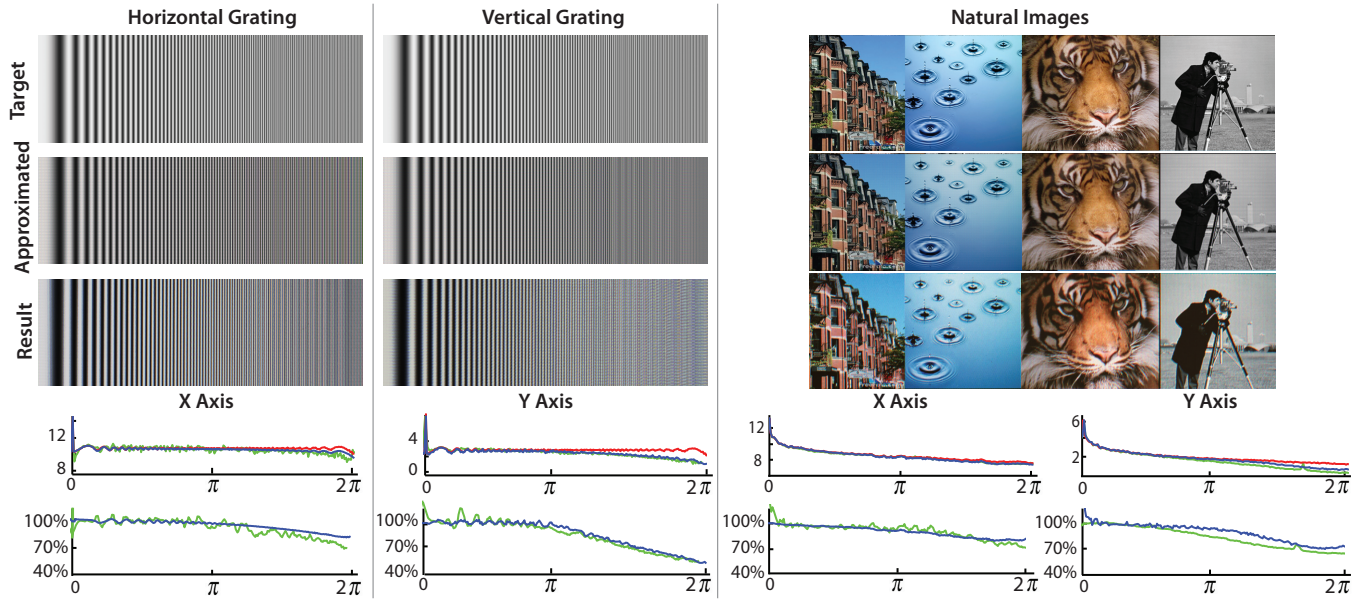
Fig. 7. We compare the spectrum of our result (row 3) to the approximated (row 2) and target (row 1) images. In row 4, we show the logarithm of the magnitude spectrum of the target (red) and approximated (blue) images, and our result (green). We also compute the percentage of Fourier components that our result and the approximated image match in the target image (row 5). Image credits for tiger image: *Moni Sertel*.

4.2 Sub-Pixel Rendering vs. Vibrating Display

Sub-pixel rendering is a technique where the different light emitting components of the LCD, typically red, green, and blue pixels, are considered as independent rather than a single triplet that jointly matches a single color tone. This technique allows to increase the display's luminance spatial resolution, at the cost of introducing some chromatic aliasing (color fringes) [Platt 2002], and it is used in Microsoft's ClearType font technology. Optimizing each color channel independently to match a corresponding channel in the target high-resolution image, as we do in Section 3, falls into the sub-pixel rendering category. We conduct two experiments to quantify the amount of spatial resolution enhancement gained by the display vibrations versus the sub-pixel rendering.

In the first experiment we estimate the resolution gained in the image *luminance* component, which we measure by summing the red, green and blue channels in the physical RGB color space. In this test we compare the spectrum of the target image against three images we display. The first image is a downscaled version of the input, computed by filtering and subsampling. This is the most common option used when the image resolution is greater than the display. Since we use the same filters and subsampling scheme for each of the three color channels this option does not implement sub-pixel rendering. The second image optimizes the three color channels of a single LRF to match the high-resolution input color channels independently. This image is computed using our method applied with $n = 1$ and non-smeared PSFs. Both these images are displayed on a non-vibrating display. The third image is the result of our algorithm with $n = 4$ and smeared PSFs and it is displayed on a vibrating display. Figure 8 shows that the sub-pixel rendering has a limited contribution to the horizontal luminance resolution and no contribution in the vertical axis (which is the expected result for the vast majority of displays whose RGB components are arranged horizontally). The reason for not observing a higher horizontal resolution, despite having three times more active pixels, is

that in natural images only a small fraction of the edges do indeed involved all three color channels. For example, the red and blue pixels do not participate in a sharp transition in the green channel and therefore such an edge will not benefit from sub-pixel rendering. Black-and-white text images are one of the very few cases where sub-pixel rendering is beneficial. This experiment confirms that most of the gain in resolution we obtain for $n = 4$ is indeed due to the vibration mechanism. Furthermore, as we explain in Section 4.1, we attribute the higher increase in resolution that vibrating displays achieve in the horizontal axis compared to the vertical axis to the spectral properties of the display PSF, which are more elongated in the vertical axis.

In Figure 8 we also measure the resolution enhancement in a chromatic channel of the perceived image, namely the green channel of the RGB color space. As expected, there is very little gain in resolution in both image axes since pixels of different color cannot contribute to one another. This test confirms that the vibrating display achieves the same amount of resolution enhancement it achieved in the luminance channel for each of the RGB color channels *independently*. We provide additional comparisons between using sub-pixel rendering on a static and vibrating display showing all three color channels on our supplemental website.

5. LIMITS OF RESOLUTION ENHANCEMENT

The matrix $\mathbf{W}^T \mathbf{W}$ plays a central role in the construction of the LRIs. In our case $\mathbf{W}$ is composed of four different SPSFs that correspond to the display PSF smeared along each of the four quarters of the circle. Therefore, the matrix $\mathbf{W}^T \mathbf{W}$ whose entries are the dot-products between nearby SPSFs, is not a spatially-homogeneous convolution matrix. This fact makes it harder to analyze it using Fourier transform. Nevertheless, in the Appendix we apply Fourier transform to a one-dimensional case where there are two different display SPSFs, one being a mirror of the other. While the transformed matrix does not separate the different frequencies, in
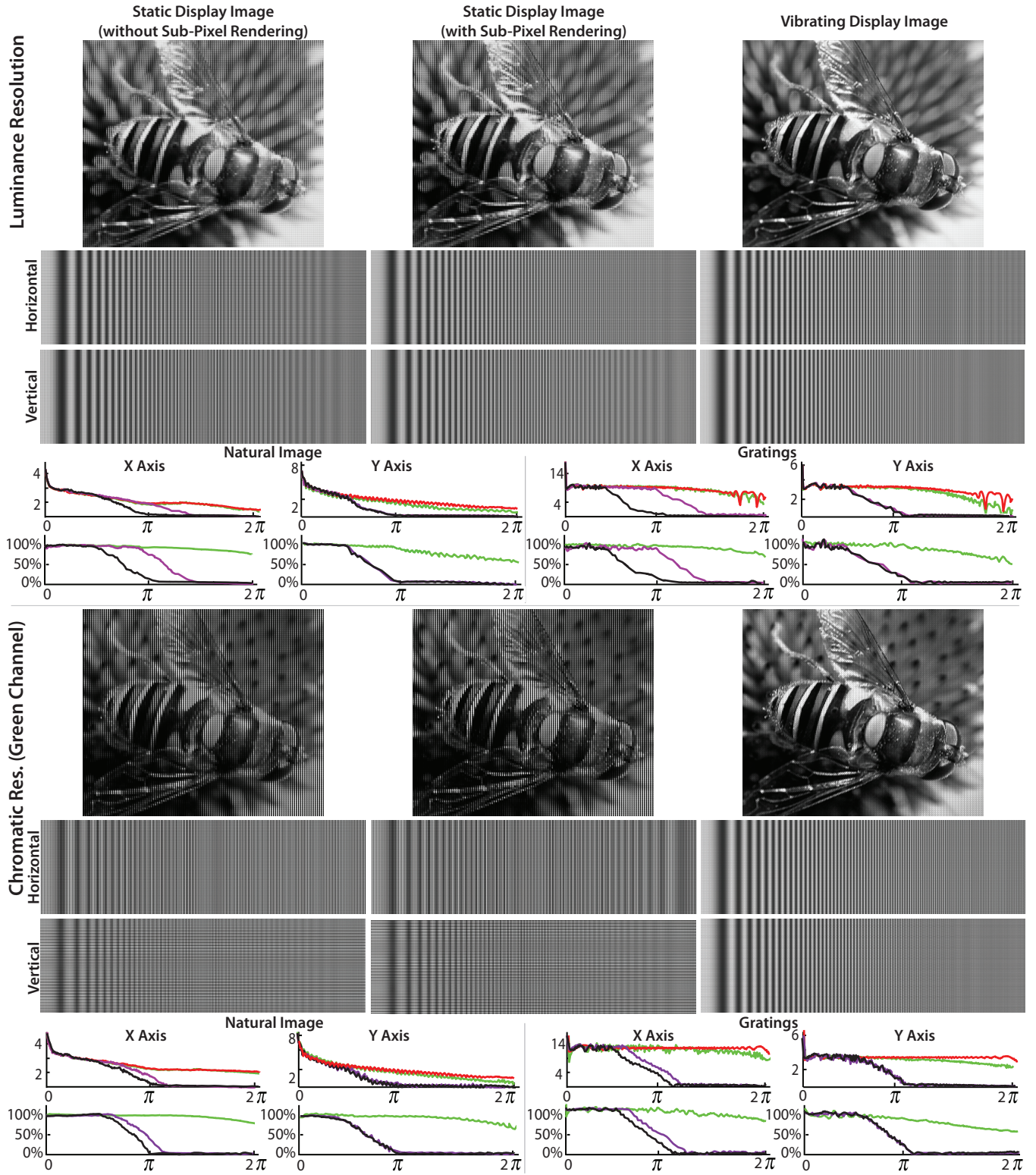
Fig. 8. Comparison between sub-pixel rendering and vibrating display. We conduct the same test for the (top) luminance image component and (bottom) a chromatic channel. The plots show the magnitude spectrum of the different images: (red) the target image, (purple) sub-pixel rendering ($n = 1$), (black) naively-downscaled image, and (green) the vibrating display ($n = 4$). The left and right plots correspond to the natural (Bee) image and the matching grating images respectively. Image credits for bee image: *Dino Abatzidis*.

Fourier space the matrix consists of 2-by-2 block matrices that mix the Fourier modes $\omega/2$ and $\omega/2 + \pi$. We show in the Appendix that, as the variability (the asymmetry in case of mirrored SPSFs) between the smeared SPSFs diminishes, each of the sub-matrices becomes diagonal and the eigenvalue that corresponds to the ω mode is $|\hat{S}(\omega)|^2$. Moreover, we show that if $|\hat{S}|$ obtains its maximum at $\omega = 0$ and minimum at $\omega = \pi$, then the condition number of $\mathbf{W}^T\mathbf{W}$ (ratio between the maximal and minimal eigenvalues) is

$$\kappa\left(\mathbf{W}^T\mathbf{W}\right) = \frac{|\hat{S}(0)|^2}{|\hat{S}(\pi)|^2}, \qquad (10)$$

as long as the difference (the asymmetry) is small enough, as is the case for our display kernels. As we explain in the Appendix, extending this analysis to 2D is very challenging as it requires computing eigenvalues of 4-by-4 block matrices. However, we numerically compute $\kappa\left(\mathbf{W}^T\mathbf{W}\right)$ in the 2D case, and find that it is equal to $|\hat{S}(0,0)|/|\hat{S}(\pi,\pi)|$. We conclude that in the 2D case the condition number also depends on the LCD's ability to span high frequencies compared to low frequencies. This analysis has important consequences on how much resolution enhancement we are able to obtain with our display system as we discuss below.

Limitations and Relation to Multi-Frame Super-Resolution. As we saw in Section 3.2, the display SPSFs play the role of reproduction kernels and act as low-pass filters that attenuate the higher frequencies (e.g., at 2π there is an attenuation of 43%). In order to match the higher frequencies of the target images, the LRIs must compensate for this attenuation by the SPSFs, such that after being multiplied by $\hat{S}_i$, the resulting perceived image will be close to $\hat{H}$ across the entire spectrum. Our analysis of $\mathbf{W}^T\mathbf{W}$ in the Appendix shows that indeed this attenuation takes place in this matrix and hence the inversion process involves an amplification ratio of $|\hat{S}(\pi)|/|\hat{S}(0)|$ between the highest and lowest frequency modes. In image space this amplification translates into getting more extreme pixel values in the LRIs (the oscillations in the LRIs, seen in Figure 6). However, the intensity range in our system is limited and therefore, during the construction of the LRIs in Algorithm 1, we restrict their values to $[0,1]$. As a consequence, we cannot amplify the high-frequency modes indefinitely because more and more LRI pixels will be clamped. In other words, the clamping introduces 'noise' that limits the amount of amplification possible and therefore also limits the amount of resolution enhancement achievable with our display.

A very similar scenario takes place in the context of multi-frame super-resolution. The goal of super-resolution techniques is to reconstruct a high-resolution input image from a set of LRIs taken at different spatial camera offsets [Irani and Peleg 1990; Park et al. 2003]. Our method can therefore be seen as tackling the inverse problem of existing super-resolution techniques. In the super-resolution model, the connection between every input LRI L_i and the output high-resolution image H can be described as

$$L_i(y) = \int P_{cam}(y - x - \varphi_i)H(x)dx + \eta(y) \qquad (11)$$

where P_{cam} is the PSF of the camera offsetted by φ_i and η is a noise term. This equation is very similar to (2) where L and H swap their role. As with the display PSF, the camera PSF, P_{cam}, must be inverted in order to obtain H from the LRIs. This is commonly tackled with a least-squares formulation, analog to (5), which result in linear systems analog to (6) where H is the unknown (see for example [Irani and Peleg 1990]).

These linear systems are known to be poorly conditioned and existing analysis [Baker and Kanade 2000; Lin and Shum 2001]
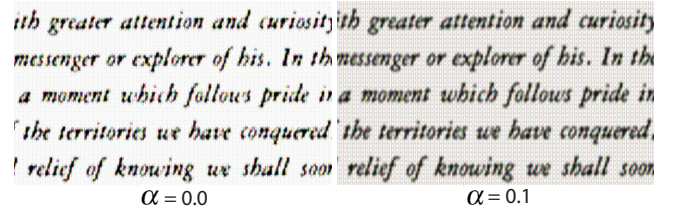

Fig. 9. Reducing the contrast of the high-resolution input image sharpens the resulting perceived image produced with our method.

shows that their condition numbers are proportional to the ratio we obtain in (10). These large condition numbers lead to a fundamental limitation that prevents the possibility to achieve high magnification factors between the input LRIs and the reconstructed high-resolution image. The main cause for this limitation is that, similarly to our display PSFs, the camera sensor PSF also acts as a low-pass filter (designed to avoid aliasing effects in photographs). Setting aside registration misalignments between the LRIs, the noise present in the acquired low-resolution input, η, becomes dominant at the higher frequency band of the image. As a result, it also gets amplified during the reconstruction process and hence the super-resolution inversion process becomes a highly unstable process when aiming for large magnification ratios. In contrast to super-resolution, the noise present in our input, H, is also amplified when generating the LRIs but it is not a limiting factor since it is attenuated back to its original level by the display PSFs. Instead, the source of the limitation in our setting is the 'noise' introduced by the clamping, which does not lead to instabilities, but rather limits the amplification factor achieved by the solver described in Algorithm 1. This limitation also applies to the existing multi-projector super-sampling methods [Damera-Venkata and Chang 2007b; Allen and Ulichney 2005] and to the method of Dydik et al.[Didyk et al. 2010a].

To conclude, both super-resolution and our method can produce a limited amount of resolution enhancement, depending on how their PSFs respond to high frequencies. To avoid this problem in the super-resolution setting, camera sensors with smaller pixel size are needed. However, this also corresponds to a reduction in the amount of light collected by the sensor and hence lower signal-to-noise ratios. In the context of displays, this limitation is less stringent and one could manufacture displays with smaller and more intense pixels without increasing their number.

Contrast Reduction. In image space, the amplified high-frequency modes appear as over- and under-shoots around sharp edges in the image. When the edge values are close to either zero or one, the oscillations get clamped, as shown in Figure 10. Therefore, the resolution enhancement becomes sub-optimal around these edges. This problem can be avoided by mapping the image values away from zero and one, for example using the following linear tone-mapping

$$g\left(H(x)\right) = (1 - 2\alpha)H(x) + \alpha, \qquad (12)$$

with $\alpha > 0$. Clearly, this mapping reduces the image's dynamic range and contrasts.

Specific classes of images, such as text images, contain many edges that reach extremal intensity values and will benefit from such contrast reduction. In Figure 9 we show how a mild contrast reduction, performed using (12) with $\alpha = 0.1$ sharpens the image. Furthermore, we show the effect of α on the ability to match each component of the spectrum. To evaluate this effect, we use the same vertical and horizontal sinusoidal gratings as in the Results Section

Target / Approximated   **LRIs**

Without Clamping
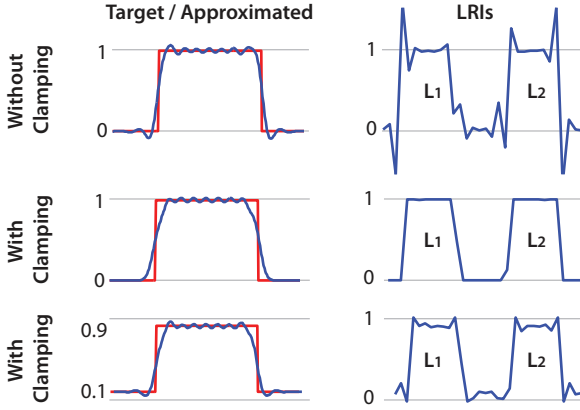
With Clamping

With Clamping

Fig. 10.   We use our method to approximate a 1D step function (red) with two LRIs. Without any clamping, the LRIs and approximated image (blue) over- and under-shoot the high contrast edge. With clamping, the approximated image captures the high contrast edge less well and blurs the edge. When reducing the contrast of the edge ($\alpha = 0.1$), the LRIs can over- and under-shoot the edge without causing any clamping and therefore approximate better the step function.

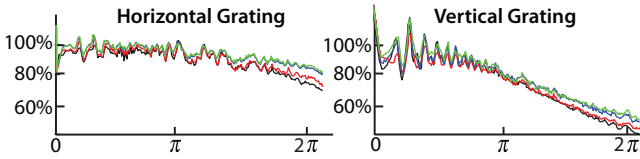
Horizontal Grating   **Vertical Grating**

Fig. 11.   When reducing the contrast of the horizontal and vertical gratings (black: $\alpha = 0$, red: $\alpha = 0.05$, blue: $\alpha = 0.1$, green: $\alpha = 0.2$) our results capture a larger percentage of the highest frequencies in the target image.

but we vary their contrast ($\alpha = 0, 0.05, 0.1, 0.2$). In Figure 11 we compare the magnitude spectrum of the target image H (the sinusoidal function) to the result we obtain using our vibrating display. When $\alpha = 0.1$ our results gain an additional 10% of the highest frequencies in the horizontal axis and 12% in the vertical axis. By further reducing the contrast ($\alpha = 0.2$), we only gain an additional 2% in the highest frequencies compared to $\alpha = 0.1$. Using $\alpha > 0.1$ therefore does not significantly benefit the quality of our results.

To our advantage, regular natural images do not contain many edges that experience clamping and hence contrast reduction may not be necessary. Also, the linear tone-mapping in (12) can be replaced by other contrast reduction techniques that better preserve the image details, such as [Fattal et al. 2002].

Practical Limits.   The convergence speed and hence the efficiency of iterative linear solvers depends on the condition number κ of $\mathbf{W}^T\mathbf{W}$. More specifically, the number of Gauss-Seidel iterations needed to achieve a small fixed error grows linearly with κ [Saad 2003]. This number directly depends on the choice of target resolution and the spectral decay of the display PSF. As the discussion above suggests, the resolution enhancement is undermined by the clamping that limits the achievable amplification factors. Therefore, there is no reason to set the target resolution such that $\hat{S}(\pi, \pi)$ is too small. This limit bounds the condition number κ. For example, in our implementation $\kappa \approx 85$ and we use 100 iterations to achieve accuracy of 0.01. Note that as Equation (10) suggests, the condition number κ does not depend on the number of image pixels. This means that the convergence speed is *independent* of the number of pixels in the input image H.
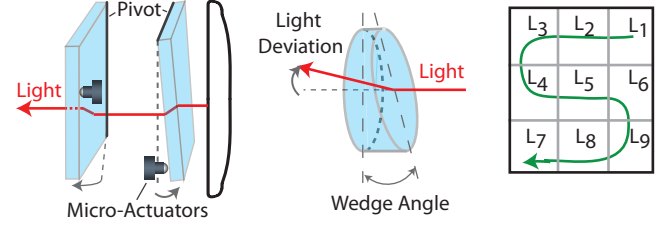

Fig. 12.   (left) When shifting two thin glass plates in front of the display, we can deviate the light rays of the pixels. (middle) Optical wedges shift light rays and the amount of deviation is determined by the wedge angle. (right) One could use a scanline trajectory for the display movements for $n > 4$ LRIs. Here every pixel is divided into a new 3x3 grid using 9 LRIs.

6. DISCUSSION

We presented a method that exploits the retinal integration for increasing the apparent resolution of displays. In this approach we trade temporal resolution for spatial resolution by displaying multiple images at sub-pixel offsets and within the retinal integration time. The images we display are optimized such that their perceived integral best matches a given high-resolution image. We describe a low-cost mechanical realization of this concept where we apply small circular vibrations to an LCD panel. This subtle movement replaces the need to move the displayed content as in previous approaches [Didyk et al. 2010a]. Further contributions of our work are the derivation of the perceived image model that contains accurately measured and calibrated components. We also analyze different spectral aspects of the new display system and formally link it to multi-frame super-resolution. By analyzing the governing matrix we describe the limitations of such systems.

The current implementation of our display system is a prototype that we designed to demonstrate the benefits of deviating the light coming from high refresh rate displays. As such, it has some evident drawbacks that might impede on the practicality of such a display (noise level, display vibrations and horizontal positioning of the screen, energy consumption). Nevertheless, we believe that, as future work, various implementations of this idea can be developed for increasing the resolution of other types of displays that suffer from insufficient resolution (projectors, the newly-emerging 3D LCDs, high-dynamic range panels, stereo displays, street TVs). For example, one straight-forward extension is to embed the panel inside a suspension mechanism which will allow it to vibrate freely while minimizing external vibrations. Another option is to use two thin glass plates in front of the display panel which shift the light by refracting it. Different shifts can be attained using micro-actuators that change the angle of the plates. This configuration is illustrated in Figure 12. In this setting a change by one degree of a 1mm thick glass plate, shifts the image by approximately one pixel. Note that a similar but more complex principle is used in lens-based image stabilization where an actuator moves a lens group based on shake-detecting sensors. Another possible shifting mechanism consists of rotating an optical wedge, depicted in Figure 12, that creates circular light trajectories similar to the ones we used in our construction.

Perhaps the most promising avenue of future research lies in optical systems to deviate the light. Current research in optics has shown that a crystal called Potassium Lithium Tantalate Niobate (KLTN) can act like an optical wedge [Agranat et al. 2010]. The refraction properties of this crystal can be set by applying an electrical field. Researchers successfully constructed small prototypes that allow changing the refractive index on a per pixel basis. Placing such a crystal in front of a display panel will allow a non-

mechanical implementation of our system and provide a more practical solution that avoids the physical movement of the display, the noise associated with it, and reduce the energy consumption. Furthermore, it will open the possibility to generate light movements that are not limited by inertial laws. For example, this construction will allow the use of $n > 4$ LRIs, where the circular trajectory would not be ideal, by following a scanline trajectory (Figure 12).

Another avenue of future research is the design of kernels that better transfer high frequencies. We could construct smaller pixels using masks and, as we noted earlier, compensate for the intensity loss using stronger backlights.

Finally, there are further research directions that relate to the algorithmic aspect of our work. In the setting of stereo displays it is known that a viewer integrates the information seen by each eye individually. This could give an opportunity for achieving resolution enhancement by generating the two views that exploit this property. Our method could also be extended for increasing the apparent resolution of videos. However, in this case one would need to model how the eye tracks objects in the scene as the eye motion affects how the images are integrated into a perceived image.

Acknowledgements

REFERENCES

AGRANAT, A., GUMENNIK, A., AND ILAN, H. 2010. Refractive index engineering by fast ion implantations: a generic method for constructing multi-components electro-optical circuits. In *Proceedings of SPIE*. Vol. 7604.

ALLEN, W. AND ULICHNEY, R. 2005. Wobulation: Doubling the addressed resolution of projection displays. *SID 47*.

BAKER, S. AND KANADE, T. 2000. Limits on super-resolution and how to break them. In *CVPR*. Published by the IEEE Computer Society, 2372.

BEN-EZRA, M., ZOMET, A., AND NAYAR, S. 2004. Jitter camera: High resolution video from a low resolution detector. In *CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Vol. 2. IEEE.

DAMERA-VENKATA, N. 2009. Displaying sub-frames at spatially offset positions on a circle. *United States Patent* 7483044 B2.

DAMERA-VENKATA, N. AND CHANG, N. 2007a. On the resolution limits of superimposed projection. In *ICIP 2007*. Vol. 5. IEEE.

DAMERA-VENKATA, N. AND CHANG, N. 2007b. Realizing super-resolution with superimposed projection. *CVPR 2007*, 1–8.

DAMERA-VENKATA, N. AND CHANG, N. 2009. Display supersampling. *ACM Transactions on Graphics (TOG) 28*, 1, 1–19.

DIDYK, P., EISEMANN, E., RITSCHEL, T., MYSZKOWSKI, K., AND SEIDEL, H. 2010a. Apparent display resolution enhancement for moving images. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2010, Los Angeles) 29*, 3.

DIDYK, P., EISEMANN, E., RITSCHEL, T., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010b. Perceptually-motivated real-time temporal upsampling of 3D content for high-refresh-rate displays. *Computer Graphics Forum (Proceedings Eurographics 2010, Norrköping, Sweden) 29*, 2, 713–722.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. In *SIGGRAPH 2002*. ACM, New York, NY, USA, 249–256.

HARDIE, R. 2007. A fast image super-resolution algorithm using an adaptive Wiener filter. *Image Processing, IEEE Transactions on 16,* 12, 2953–2964.

IRANI, M. AND PELEG, S. 1990. Super resolution from image sequences. In *International conference on Pattern Recognition*. II: 115–120.

JAYNES, C. AND RAMAKRISHNAN, D. 2003. Super-resolution composition in multi-projector displays. In *Proc. of IEEE International Workshop on Projector-Camera Systems (ProCams)*.

KALLONIATIS, M. AND LUU, C. 2009. Temporal resolution. www.webvision.med.utah.edu/temporal.html.

LIN, Z. AND SHUM, H. 2001. On the fundamental limits of reconstruction-based super-resolution algorithms.

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, 1981–1991.

MCKEE, S. AND TAYLOR, D. 1984. Discrimination of time: comparison of foveal and peripheral sensitivity. *Journal of the Optical Society of America A 1,* 6, 620–628.

NAPOLI, J., DEY, S., STUTSMAN, S., COSSAIRT, O., PURTELL II, T., HILL, S., AND FAVALORA, G. 2008. Imaging artifact precompensation for spatially multiplexed 3-D displays. In *Proceedings of SPIE*. Vol. 6803. 680304.

NORWICH, K. 1993. *Information, sensation, and perception*. Academic Press San Diego, CA.

PARK, S. C., PARK, M. K., AND KANG, M. G. 2003. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine 20,* 3 (May), 21–36.

PLATT, J. 2002. Optimal filtering for patterned displays. *Signal Processing Letters, IEEE 7,* 7, 179–181.

SAAD, Y. 2003. *Iterative Methods for Sparse Linear Systems*, 2nd ed. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

TEMPLIN, K., DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. Apparent resolution enhancement for animations. In *Proc. of the 27th Spring Conference on Computer Graphics*. Vinicne, Slovak Republic, 85–92.

YOULA, D. 1978. Generalized image restoration by the method of alternating orthogonal projections. *Circuits and Systems, IEEE Transactions on 25,* 9, 694–702.

7. APPENDIX

The matrix $\mathbf{W}^T\mathbf{W}$ is the Gram matrix whose elements are dot-products between every two SPSFs at the computational grid. If the SPSFs were identical then this matrix would have been a translation invariant convolution matrix which we could easily study using Fourier analysis. However, our system has four different SPSFs which make the analysis of $\mathbf{W}^T\mathbf{W}$ more difficult. Here we try to find a connection between the spectral properties of $\mathbf{W}^T\mathbf{W}$ and the SPSFs S_i that define it. Due to the lack of spatial invariance, this analysis is very challenging and therefore we analyse a 1D model problem that contains two SPSFs.

We denote by L_1 and L_2 the two 1D low-resolution images that are optimized to approximate a higher resolution 1D image. We first upsample L_1 and L_2 to a higher resolution grid by adding zeros between every two values. In Fourier space, this corresponds to

$$\hat{L}_1(r\omega), \quad \text{and} \quad \hat{L}_2(r\omega)e^{-i\omega}, \tag{13}$$

where r is the resolution factor between the LRIs and the finer grid. The factor $e^{-i\omega}$ results from the spatial shift between L_1 and L_2. Since each LRI is reproduced by a different kernel, the Fourier transform of the perceived image $V = \mathbf{W}L$ is given by

$$\hat{V} = \hat{S}_1(\omega)\hat{L}_1(r\omega) + \hat{S}_2(\omega)\hat{L}_2(r\omega)e^{-i\omega}, \qquad (14)$$

where $\hat{S}_1$ and $\hat{S}_2$ are the respective smeared kernels for $\hat{L}_1$ and $\hat{L}_2$. Multiplying the perceived image by $\mathbf{W}^T$ as in (7), translates in Fourier space to convolving (14) with the complex conjugates of the SPSFs and gives

$$|\hat{S}_1|^2(\omega)\hat{L}_1(r\omega) + (\hat{S}_1^*\hat{S}_2)(\omega)\hat{L}_2(r\omega)e^{-i\omega}, \qquad (15)$$

$$(\hat{S}_2^*\hat{S}_1)(\omega)\hat{L}_1(r\omega) + |\hat{S}_2|^2(\omega)\hat{L}_2(r\omega)e^{-i\omega}. \qquad (16)$$

The multiplication also performs a subsampling at rate r (undoing the upsampling in $\mathbf{W}$) and the above becomes

$$\sum_{s=1}^{r}|\hat{S}_1|^2\left(\frac{\omega}{r}+s\frac{2\pi}{r}\right)\hat{L}_1(\omega)+\sum_{s=1}^{r}(\hat{S}_1^*\hat{S}_2)\left(\frac{\omega}{r}+s\frac{2\pi}{r}\right)e^{-i\frac{\omega}{r}+s\frac{2\pi}{r}}\hat{L}_2(\omega), \qquad (17)$$

$$\sum_{s=1}^{r}(\hat{S}_2^*\hat{S}_1)\left(\frac{\omega}{r}+s\frac{2\pi}{r}\right)e^{i\frac{\omega}{r}+s\frac{2\pi}{r}}\hat{L}_1(\omega)+\sum_{s=1}^{r}|\hat{S}_2|^2\left(\frac{\omega}{r}+s\frac{2\pi}{r}\right)\hat{L}_2(\omega). \qquad (18)$$

Note that (18) was multiplied by an additional factor $e^{i\omega}$ to take into account the shift of L_2 w.r.t. L_1 when downsampling. We can now write the 2-by-2 matrix that operates on the vector $\hat{L} = (\hat{L}_1(\omega), \hat{L}_2(\omega))$ for every ω

$$\begin{pmatrix} \sum_{s=1}^{r}|\hat{S}_1|^2(\frac{\omega}{r}+s\frac{2\pi}{r}) & \sum_{s=1}^{r}(\hat{S}_1^*\hat{S}_2)(\frac{\omega}{r}+s\frac{2\pi}{r})e^{-i\frac{\omega}{r}+s\frac{2\pi}{r}} \\ \sum_{s=1}^{r}(\hat{S}_2^*\hat{S}_1)(\frac{\omega}{r}+s\frac{2\pi}{r})e^{i\frac{\omega}{r}+s\frac{2\pi}{r}} & \sum_{s=1}^{r}|\hat{S}_2|^2(\frac{\omega}{r}+s\frac{2\pi}{r}) \end{pmatrix}. \qquad (19)$$

These matrices show that there is a large amount of spectral decoupling in this system; while the Fourier transform did not diagonalize $\mathbf{W}^T\mathbf{W}$, it reduced it to 2-by-2 block matrices.

As in the 2D case, the two smeared display kernels S_1 and S_2 are mirrored w.r.t. one another. Therefore, $\hat{S} := \hat{S}_1 = \hat{S}_2^*$. Let us decompose $\hat{S}_1$ into a symmetric and asymmetric part $\hat{S}_1 = \hat{S}_s + i\hat{S}_a$, where $\hat{S}_s$ and $\hat{S}_a$ are real-valued functions. With this decomposition $\hat{S}_2 = \hat{S}_s - i\hat{S}_a$ and the matrix in (19)

$$\begin{pmatrix} \sum_{s=1}^{r}(\hat{S}_s^2+\hat{S}_a^2)(\frac{\omega}{r}+s\frac{2\pi}{r}) & \sum_{s=1}^{r}(\hat{S}_s-i\hat{S}_a)^2(\frac{\omega}{r}+s\frac{2\pi}{r})e^{-i\frac{\omega}{r}+s\frac{2\pi}{r}} \\ \sum_{s=1}^{r}(\hat{S}_s+i\hat{S}_a)^2(\frac{\omega}{r}+s\frac{2\pi}{r})e^{i\frac{\omega}{r}+s\frac{2\pi}{r}} & \sum_{s=1}^{r}(\hat{S}_s^2+\hat{S}_a^2)(\frac{\omega}{r}+s\frac{2\pi}{r}) \end{pmatrix}. \qquad (20)$$

In this analysis we make another simplification and set $r = 2$ even though the actual resolution factor between the LRIs and the computational grid is 8 in our implementation. This introduces the mild frequency folding we discussed in Section 3.1. As we verify here, it does not introduce a significant inaccuracy for our estimation.

In order to determine the condition number of $\mathbf{W}^T\mathbf{W}$, we have to compute the maximal and minimal eigenvalues of the 2-by-2 matrices in (19) among all ω. The eigenvalues of any 2-by-2 matrix $\mathbf{A}$ are found by solving the quadratic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. If we denote $\hat{S}_s(\frac{\omega}{2})$ by $\hat{S}_s$ and $\hat{S}_s(\frac{\omega}{2} + \pi)$ by $\hat{S}_s'$ and the same for $\hat{S}_a$, the two eigenvalues are given by

$$\lambda_\omega^\pm = |\hat{S}|^2+|\hat{S}'|^2 \pm \sqrt{\left(|\hat{S}|^2-|\hat{S}'|^2\right)^2+2\left[\hat{S}_s^2\hat{S}_a'^2+\hat{S}_s'^2\hat{S}_a^2-\hat{S}_s\hat{S}_s'\hat{S}_a\hat{S}_a'\right]}. \qquad (21)$$

If the kernels were identical, then $\hat{S}_a = 0$ and $\hat{S}_a' = 0$ and the maximum and minimum eigenvalues would be $2|\hat{S}(\frac{\omega}{2})|^2$ and $2|\hat{S}(\frac{\omega}{2} +$

$\pi)|^2$ respectively. In this case, the condition number would be

$$\kappa(\mathbf{W}^T\mathbf{W}) = arg\max_\omega \frac{|\hat{S}(\frac{\omega}{2})|^2}{|\hat{S}(\frac{\omega}{2}+\pi)|^2} = \frac{|\hat{S}(0)|^2}{|\hat{S}(\pi)|^2} \qquad (22)$$

assuming the SPSFs have a monotonically decreasing magnitude spectrum. This implies that the higher frequency response the SPSFs have to high frequencies, the better conditioned $\mathbf{W}^T\mathbf{W}$ becomes. In the general case where the smeared kernels are not symmetric, we can show that the condition number in (22) still holds if $\hat{S}_a$ and $\hat{S}_a'$ are sufficiently small. We do this by bounding $|\hat{S}_a(\omega)| < \varepsilon|g(\omega)|$, where $g(\omega) = -\omega(\pi-\omega)(\pi+\omega)$. The function $g(\omega)$ is zero at 0 and $\pm\pi$ where the Fourier transform of SPSFs are real. The inset shows $\hat{S}_s$, $\hat{S}_a$ and εg for our acquired display kernels. We found that $\varepsilon = 0.008$ is enough to bound $\hat{S}_a$ for our SPSFs. Let us write the resulting upper bound on λ_ω^+

$$|\hat{S}|^2+|\hat{S}'|^2+\sqrt{\left(|\hat{S}|^2-|\hat{S}'|^2\right)^2+2\varepsilon^2g^2\left[\hat{S}_s^2\left(\frac{2\pi+\omega/2}{\pi-\omega/2}\right)^2+\hat{S}_s'^2-\hat{S}_s\hat{S}_s'\left(\frac{2\pi+\omega/2}{\pi-\omega/2}\right)\right]}, \qquad (23)$$

where we used $g(\omega/2+\pi) = -g(\omega/2)\frac{2\pi+\omega/2}{\pi-\omega/2}$.

Finally, if we show that λ_0^+ is greater than the bound (23) of λ_ω^+ (for $\omega > 0$) and with $\varepsilon \geq 0.008$, then we *know* that λ_0^+ is maximal eigenvalue of $\mathbf{W}^T\mathbf{W}$.

We further bound (23) using the inequality $\sqrt{a+b} < \sqrt{a}+\sqrt{b}$

$$\lambda_\omega^+ < 2|\hat{S}|^2 + \sqrt{2}\varepsilon g\sqrt{\hat{S}_s^2\left(\frac{2\pi+\omega/2}{\pi-\omega/2}\right)^2+\hat{S}_s'^2-\hat{S}_s\hat{S}_s'\left(\frac{2\pi+\omega/2}{\pi-\omega/2}\right)} \qquad (24)$$

Now, when requiring $\lambda_\omega^+ < \lambda_0^+$ we get the following condition on ε

$$\varepsilon < \frac{(\lambda_0^+ - 2|\hat{S}|^2)}{\sqrt{2}g\sqrt{\hat{S}_s^2\left(\frac{2\pi+\omega/2}{\pi-\omega/2}\right)^2+\hat{S}_s'^2-\hat{S}_s\hat{S}_s'\left(\frac{2\pi+\omega/2}{\pi-\omega/2}\right)}}. \qquad (25)$$

The right hand side of (25) evaluates to 0.038 which is significantly larger than 0.008. Therefore, our upper bound is tight enough and our analysis succeeded to find the highest eigenvalue of $\mathbf{W}^T\mathbf{W}$. The same procedure applied to λ_ω^- gives $\varepsilon < 0.022$ and shows that λ_0^- is the minimal eigenvalue of $\mathbf{W}^T\mathbf{W}$.

Similarly to the case of a translation-invariant operator resulting from a single PSF, our analysis manages to obtain an estimate for $\kappa(\mathbf{W}^T\mathbf{W})$ based on the Fourier properties of the SPSFs. In the 2D case the Fourier analysis reduces the problem to computing eigenvalues of 4-by-4 matrices. This is a challenging task that we could not fulfill. However, we computed $\kappa(\mathbf{W}^T\mathbf{W})$ for the 2D case using Matlab and obtained the same values as $|\hat{S}(0,0)|^2/|\hat{S}(\pi,\pi)|^2$ where the kernels were discretized at the same resolution as in $\mathbf{W}$, i.e., eight times higher than the LRI resolution. More importantly this analysis explains the condition number we observe and shows that it depends on the LCD pixels' ability to span high-frequencies compared to low-frequencies.