

Patch scanning displays: spatiotemporal enhancement for displays

KAAN AKŞIT* 

NVIDIA, 2788 San Tomas Expressway, Santa Clara, CA 95051, USA

*kaksit@nvidia.com

<https://kaanaksit.com>

Abstract: Emerging fields of mixed reality and electronic sports necessitate greater spatial and temporal resolutions in displays. We introduce a novel scanning display method that enhances spatiotemporal qualities of displays. Specifically, we demonstrate that scanning multiple image patches that are representing basis functions of each block in a target image can help to synthesize spatiotemporally enhanced visuals. To discover the right image patches, we introduce an optimization framework tailored to our hardware. In our method, spatiotemporally enhanced visuals are synthesized using an optical scanner scanning image patches from an image generator illuminated by a locally addressable backlight. As a validation of our method, we demonstrate a prototype using commodity equipment. Our method improves pixel fill factor to hundred percent and enhances spatial resolution of a display up to four times. An inherent constrain regarding to spatiotemporal qualities of displays could be solved using our method.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

The advent of emerging fields such as Virtual Reality (VR), Augmented Reality (AR), and electronic sports (eSports) necessitate greater pixel density and higher presentation rates in next generation displays [1]. This need for greater spatiotemporal quality stems from the demanding needs of the Human Visual System (HVS). A healthy HVS is usually described as a 20/20 vision, which is often considered as 30 cycles per degree (cpd) resolution at the fovea [2]. Although commodity desktop displays often meet this criteria, state-of-the-art VR/AR near-eye displays at the market are still far behind in pixel density, typically offering 5 – 12 cpd. Therefore, having greater pixel density in displays still remains to be an open problem. Today's displays are believed to offer presentation rates (typically 60 – 120 Hz) above the critical flicker threshold (CFF) [3] for a healthy HVS. Nevertheless, various studies [4–7] suggest HVS can perceive flicker artifacts beyond the flicker fusion threshold (>500 Hz). Various studies also suggest human brain responses [8] and perceptual judgments [9] improve greatly when viewing visuals at a display with higher presentation rates. Therefore, higher presentation rate is a key component for improving realism in next generation displays.

Manufacturing displays with greater pixel count and pixel density is a never-ending active branch of investigation [10]. Image generators in displays also known as Spatial Light Modulators (SLMs) with greater number of physical pixels possess a major challenge for a manufacturer with large production volumes as those pixels must be defect free, and also matched in output luminance over usage.

To this end, displays with multiple SLMs are investigated for achieving greater pixel count and pixel density. Sajadi et al. [11] shows that combining two SLMs and a lenslet array can help to generate enhanced spatial resolutions, in which a high resolution edge image, and a complementary low resolution image are combined using two SLMs. Furthermore a single display with multiple SLMs, researchers explore combining multiple displays. Jaynes and Ramakrishnan [12] overlap multiple images from multiple projectors to achieve higher spatial resolutions and contrasts. By cascading two Liquid Crystal Displays (LCDs) on top of each

other in a slightly shifted way in the order of a subpixel, Heide et al. [13] report enhanced spatiotemporal resolutions. Extending many of these techniques to next generation displays causes a major shortcoming regarding to poor pixel fill-factor. This leads to a phenomena called screen-door effect, where fine lines in between pixels become visible when a display is magnified the black matrix between the subpixels becomes visible. A most recent example, work by Sitter et al. [14] demonstrate a technique to reduce screen-door effect using a diffractive film, however such techniques don't improve spatiotemporal qualities while providing a solution to screen-door effect.

Alternatively, researchers are seeking temporal means to generate more pixels from a given number of physical pixels in an SLM. Allen and Ulichney [15] introduce a method known as "wobulation" and "e-shift", where sub-frames from a Digital Micromirror Device (DMD) are shifted optically by a fraction of a pixel to form a complete frame with higher spatial resolutions. Following the steps of the work of Allen and Ulichney [15], Berthouzoz et al. [16] physically displace a Liquid Crystal Display (LCD) using a mechanically vibrating rotating mass to shift sub-frames by a fraction of a pixel. Sajadi et al. [17] shift the whole image with sub-pixel precision and superimposes the shifted image on top of the original image. Although sounding similar to "wobulation", this method of Sajadi et al. [17] doesn't time-multiplex the sub-frames, rather optically superimposes an image on a shifted version of itself. While Zhan et al. [18] shows that "wobulation" as an idea can be helpful for resolution enhancement in conventional near-eye displays using Pancharatnam-Berry deflectors, Wu et al. [19] shows that "wobulation" as an idea can also be helpful for resolution enhanced lightfield near-eye displays using a birefringent plate and an electronically controllable twisted nematic cell. Note that all these techniques demand faster SLMs to resolve the issue of generating greater pixel count and pixel density, which can be considered as a trade-off between representation rate and spatial resolution.

Many researchers have explored exploiting properties of the HVS for greater spatial resolutions. Early on Didyk et al. [20] introduced spatial resolution enhancement for images rapidly varying in time by exploiting retinal integration time. Lee et al. [21] extend methods of Didyk et al. [20] to near-eye displays. To trick one's perception that looking at a high resolution visual, researchers explore varying pixel density across an image according to gaze fixations, this technique is also known as foveated imaging, whereas previously introduced resolution enhancement strategies aims to provide highest resolution possible at all locations across an entire image. Recently, foveated imaging started to appear in the form of hardware [22–24] in near-eye displays. However, current efforts for a foveated display hardware are far from practical implementations. We believe techniques exploiting properties of the HVS are complementary to physical hardware implementations and can be combined for greater spatiotemporal qualities.

We propose a new type of scanning display method that enhances spatiotemporal qualities of Spatial Light Modulators (SLMs) as depicted in Fig. 1, in which multiple patches that represent the bases of blocks in a target image are scanned across a predefined trajectory to reconstruct the image. Our method combines a computational approach that discovers the right patches to be scanned with a new hardware design. The hardware part of our method consists of a locally addressable high refresh rate backlight with an off-the-shelf low refresh rate SLM, which are used to build a display module. The image generated in a combined display module is projected and scanned with the help of optical components. Using the hardware in our approach, we tile bases of a target image side by side over an SLM, and use the backlight as coefficients to select the intensities of each presented basis. As the backlight varies in intensity over time synchronous to the scanning trajectory, a complete image can be reconstructed as bases are scanned and projected optically. A system layout of our display method is depicted in Fig. 2 and our method is summarized as in [Visualization 1](#).

To our knowledge we are first to combine optical scanning with learning based optimization approaches in displays, while providing a hundred percent pixel fill-factor and enhanced

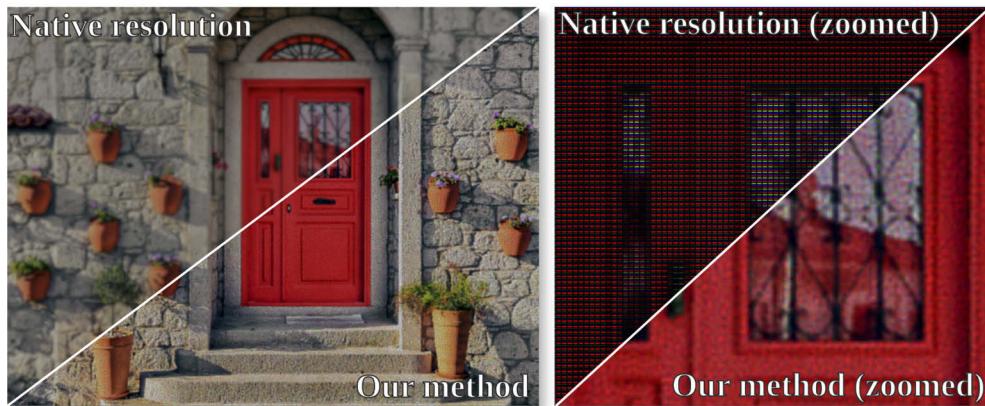


Fig. 1. Spatiotemporal enhancement and pixel fill-factor enhancement by patch scanning displays. (Left) Image generation of a spatial light modulator (SLM) at its native resolution is compared with a patch scanning display based on the same SLM. (Right) Zoomed comparison on the same data provided in left photograph is shown. Both images are simulated. Source images courtesy Erhan Meço. See [Visualization 1](#) for more data.

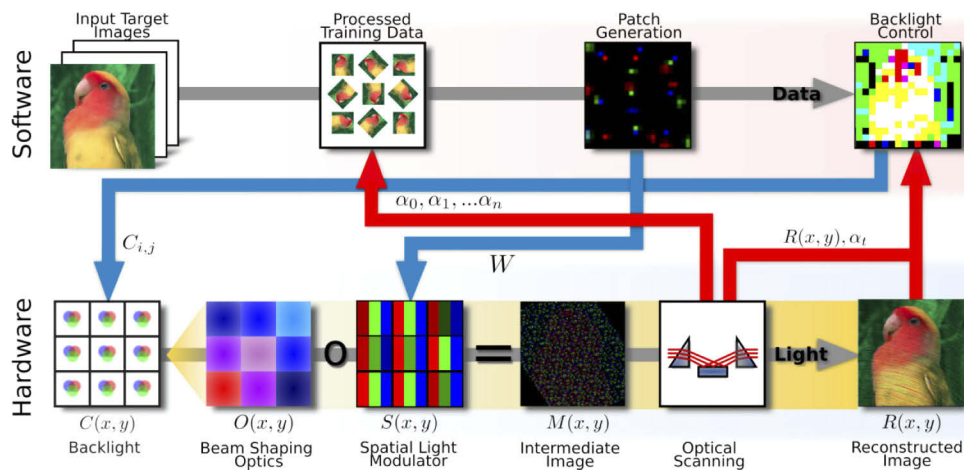


Fig. 2. System layout of patch scanning display method. An intermediate image is formed by combining multicolor locally addressable incoherent backlight with a multicolor Spatial Light Modulator (SLM), in which the backlight is updated at a fast pace in a binary fashion and the SLM is updated at a much lower rate. An optical scanner scans an intermediate image to reconstruct a target image at a final image plane. The resultant reconstructed image has enhanced spatiotemporal qualities. The SLM shows tiled patches, which are learned from a training dataset for a given set of input target images. The backlight array acts as coefficients to reconstruct different portions of a target image, and updated for each step during a scan.

spatiotemporal qualities. Our work distinguishes itself from the rest of the literature by introducing scanning patches rather than one or two pixels at a time [25]. Our work offers image patches to the image reconstruction problem in display hardware rather than per pixel optimization [15]. Our technique provides an opportunity to run a SLM at presentation rates beyond what a SLM was designed for. We believe brute-force methods that targets larger pixel

counts [10] will continue to advance, and our technique promises to be helpful for brute-force way to enhance spatiotemporal qualities.

2. Method

2.1. Image formation model

Patch scanning display method combines a fast backlight, a SLM, and a scanner. Our method partitions an SLM into blocks, where each blocks represent a set of patches for reconstructing a final image. A scanner in our method scan patches with the help of a fast binary backlight. For patch scanning displays, we formulate an image reconstruction model relying on Non-Negative Matrix Factorization (NMF). Figure 2 shows a block diagram of our imaging model in relation with our hardware.

Various kinds of SLMs can modulate either amplitude or phase of light. We choose to focus only on amplitude modulating SLMs and incoherent light sources (i.e. LEDs) as they are the most common types in today's display products on the market. We combine a single amplitude-only SLM with color filters on each subpixel with a multicolor locally addressable LED backlight. Usage of incoherent light sources and scanning in our proposed hardware leads to a multiplicative nature in spatial means, and an additive nature in time,

$$R(x, y) = \sum_{t=t_0}^{t_n} T(M_t, t) = \sum_{t=t_0}^{t_n} T((O_t \odot S_t), t), \quad (1)$$

where R represents reconstructed image over the time, O represents a binary three-color backlight behind a SLM, S represents a three-color SLM, M represents element-wise multiplication of S and O , and T represent the transformation of intermediate image as a result of optical scanning at a given time. In this model, S can be updated relatively slowly with respect to the backlight as off-the-shelf SLMs have typically low representation rates ($\sim 60 - 120$ Hz), where M can be generated at a much higher representation rate ($>0.1 - \dots$ kHz) as it is controlled by a multicolor backlight array.

Our proposed system's additive nature in time brings us to an important problem: How will a human observer perceive an image reconstructed by our imaging model described in Eq. (1) as each subimage $T(M_t, t)$ is presented for a brief moment at a fast rate? The literature on light and dark adaptation of HVS [26] suggest that there are three causes for a visual stimuli to persists and slowly decay until ceases to exist at a HVS, and these causes are photoreceptors bleaching and regeneration, fast neural adaptation and slow neural adaptation. Light and dark adaptation due to photoreceptors bleaching and regeneration is a slow process, which has a regeneration time constant of 110 seconds for rods and 400 seconds for cones according to the work of Hood et al. [27]. According to Pattanaik and colleagues [28], out of fast and slow neural adaptation, fast one is the one most relevant to describe dynamic response of a display. Rinner and Gegenfurtner [29] suggest that fast neural adaptation of rods and cones can be modeled using an exponential decay function,

$$F(t) = I_i(t)(1 - e^{-\frac{\Delta t}{\tau}}), \quad (2)$$

where $I_i(t)$ represents a time varying light source input, Δt represents a discrete time step, and τ is the time constant for a photoreceptor. Work by Pattanaik et al. [28] claims that time constant for rods is $\tau_{rod} = 150$ ms, and time constant for cones is $\tau_{cone} = 80$ ms. Our imaging model found in Eq. (1) is updated with fast neural adaptation property of HVS in Eq. (2) as

$$R(x, y) = \sum_{t=t_0}^{t_n} T((O_t \odot S_t), t)(1 - e^{-\frac{t-t_0}{\tau}}). \quad (3)$$

For sake of simplicity, we choose to use $\tau = 80$ ms for all regions of a reconstructed image.

2.2. Finding the right image patches

The additive nature of incoherent light sources in our backlight have lead us to a second important problem: To reconstruct a target image by optical scanning, how can we determine multiple non-negative N image patches $f^i(x, y)$ with m columns and n rows that can be tiled over a SLM side by side? A variant of SVD entitled as Truncated Singular Value Decomposition (t-SVD) and Non-negative Matrix Factorization (NMF) [30] are known to provide non-negative basis functions. t-SVD are known to be less accurate with respect to NMF [31]. The decomposition model that we choose to use in this study is a special case of Projective Non-negative Matrix Factorization (P-NMF), where learned basis functions are considered by taking a scan trajectory into account.

Basis functions to be identified are in shape of $m \times n$. An input target image $I(x, y, i)$ is provided with k columns, l rows and $i = 3$ color channels. An input target image is first transformed using T for each step across a scan trajectory. For each color channel, a transformed input target image $T(I(x, y))$ can be vectorized to N images each with $M = m \times n$ pixels to form a column of our input data matrix V_r, V_g, V_b with the shape of $N \times M$.

Considering an input data matrix V from a single color channel, a P-NMF searches for a solution to the following optimality problem

$$\arg \min_{W \geq 0} \|V - WW^T V\|, \quad (4)$$

where the Euclidean distance is used as a matrix norm $\|\cdot\|$, W represents an orthogonal matrix with $M \times r$ shape that contains vectorized basis and r represents matrix rank, or in another words, number of basis requested from P-NMF's optimality problem. The Euclidean distance between two sample matrices of A and B can be calculated as follows:

$$\|A - B\|^2 = \sum_{x,y} (A_{xy} - B_{xy})^2. \quad (5)$$

Using the Euclidean distance matrix norm in Eq. (5), we rely on a multiplicative update rule to update W matrix,

$$W_{xy} \leftarrow W_{xy} \frac{(VV^T W)_{xy}}{(WW^T VV^T W)_{xy} + (VV^T WW^T W)_{xy}}. \quad (6)$$

As P-NMF's optimality is met in Eq. (6), we are left with a basis matrix W , where each column represents a basis function. A naive approach for tiling patches from a basis matrix W over a SLM is depicted. Depending on a scan trajectory, various different kinds of tiling can be considered such as along the direction of columns or rows or circular tiling. Patches can also be computed locally as each patch can have an unique trajectory depending on the scan hardware. Patches are generated per target frame, they can be dramatically different across multiple frames. Therefore training per frame or per sequence of frames of a video is needed accordingly.

2.3. Controlling light sources

Similar to DMDs, we choose to derive a binary control (on-off control) for each light source in a backlight. For each time step t , we calculate what is left of a target image J_t after each step towards reconstructing a final image with an element wise subtraction from previous steps of scanning

$$J_t(x, y) = I(x, y) - R_t(x, y), \quad (7)$$

where I represents a target image and R_t represents reconstructed image at a given time step of t , which can be calculated using Eq. (3) for a given set of O_t and S_t . Using J_t from Eq. (7), next

step of a light source in a backlight can be calculated with a greedy approach

$$O_t(x, y) = \begin{cases} \text{if } J_t(x, y) \geq 0 & 1 \\ \text{if } J_t(x, y) < 0 & 0 \end{cases}, \quad (8)$$

where the contribution of each light source in a backlight is checked and helps towards full image reconstruction. Following Eq. (3) and Eq. (8), we choose to calculate each O_t by choosing t randomly using an uniform probability distribution. As each O_t is calculated, we move on to next randomly chosen t from set of steps in an optical scan trajectory to randomize any noise patterns in an image reconstruction for each frame to be reconstructed.

2.4. Scan trajectory considerations

All of the problems explored in previous sections are highly dependent on a chosen scan trajectory of a patch scanning display as presented in Eq. (3). A scan trajectory is also highly dependent on a chosen scan hardware. Common approaches for scanning displays typically relies on scanning a single pixel with a raster scan [32], a sinusoidal scan trajectory over a plane. Result of our survey points us towards solutions in the direction of traditional raster scanning an entire image as such scanners are available in consumer level products or in precision equipment for experimentation. However, we suspect that having a scanner per image patch and scanning image patches in freeform would be the ideal for our proposal.

For all our experimentation, we set the trajectory of the optical scanner to a predefined scan trajectory defined as

$$\begin{aligned} \Theta_x &= \arctan\left(\frac{p_x c_x \cos(2\pi \frac{t}{\Delta t_f} k_x)}{z}\right), \\ \Theta_y &= \arctan\left(\frac{p_y c_y \cos(2\pi \frac{t}{\Delta t_f} k_y)}{z}\right) + 45, \end{aligned} \quad (9)$$

where c_x and c_y represent constants to define the amplitude of a scan pattern in number of pixels, p_x and p_y represent size of a pixel in a SLM along X and Y axes, Δt_f represents time duration to scan entire trajectory once, k_x and k_y represent constants to define frequency of a sinusoidal scan pattern along X and Y axes, and z represents distance in between an optical scanner and a SLM. We choose $p_x = 0.187$ mm, $p_y = 0.167$ mm, $c_x = 10$, $c_y = 10$, $k_x = 1$, $k_y = 64$, $z = 100$ mm in our experimentation.

3. Implementation

In order to demonstrate that we can leverage our computational image reconstruction model in practice, we have built a functional prototype. Our functional prototype is a means to emulate various different situations in physical means and provide guidance in practice towards trade-offs in design choices for a future reference design.

3.1. Hardware

Our functional prototype is based on off-the-shelf optical, optomechanical, mechanical and electronic components augmented with 3D printed pieces. Following Fig. 3, we describe the entire assembly of the hardware used in a functional prototype. Figure 4 shows an actual photograph of our functional prototype.

Optical Assembly. The optical path of a patch scanning display starts from a light source array that forms a backlight. We use a Luckylight KWM-20882XWB-Y 8 by 8 dot matrix white LED array with 1.9 mm pitch size as a light source array. Light rays from the backlight is modulated by going through a transmissive SLM at 1 mm away. We use 320×240 px 60 Hz Adafruit PiTFT LCD 2.8" display as a transmissive SLM, in which we remove the built-in backlight and the

touch panel and glued an isotropic diffuser on one side. In given configuration, a light source in the backlight illuminates a space of 15×15 pixels on the transmissive SLM. Modulated light from the transmissive SLM travels 100 mm in free space and bounces off an optical scanner (Optics in Motion OIMLA02B). We connect two MCP4725 Digital-to-Analog converters (DAC) to our computation modules to control an optical scanner in XY axes, where a raster scanning is imitated following Eq. (9). We choose $c_x = 0.187$ mm, $c_y = 0.167$ mm, $p_x = 10$, $p_y = 10$, $k_x = 1$, $k_y = 64$, $z = 100$ mm in all of our . An example transformation T as in Eq. (3) caused by our raster scanning is simulated as in Visualization 2 for the case of a static image loaded on a SLM. An observer at the end of red highlighted optical path in Fig. 3 can view modulated light bouncing off the optical scanner.

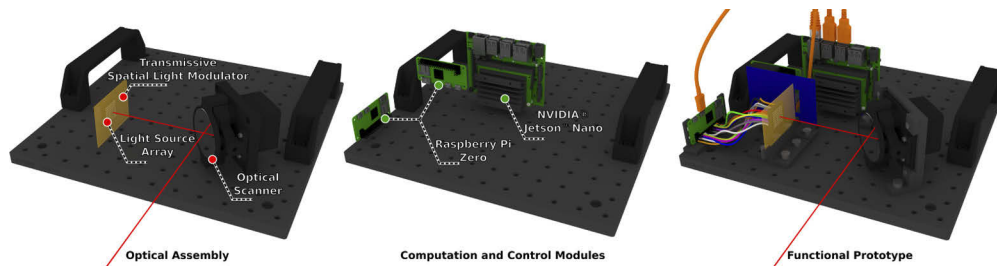


Fig. 3. Renderings of a virtual prototype for a patch scanning display. (Left) This image shows an optical assembly for the functional prototype, where light from a light source array follows the optical path highlighted with a red line representing a chief ray, and gets modulated using a transmissive Spatial Light Modulator (SLM), and modulated light bounces off a mirror of an optical scanner. (Middle) The image shows computation and control modules used in the functional prototype. (Right) The image shows an entire assembly of a functional prototype. Only omitted part in the right image is an optical scanner control circuitry. See Visualization 2 for a mirror scan simulation with a static input image.

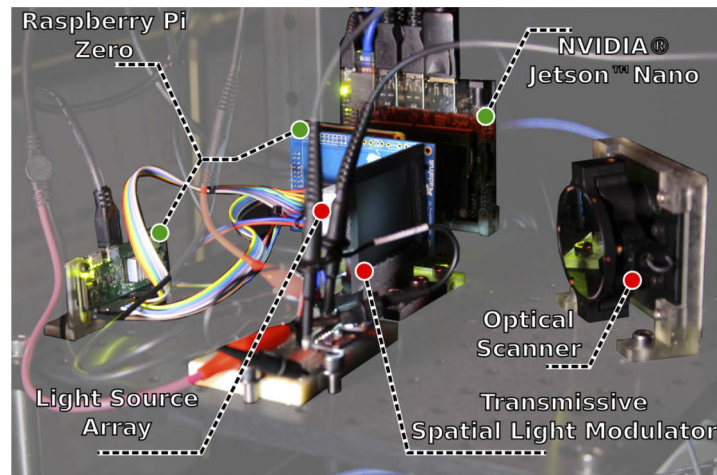


Fig. 4. A photograph of our functional prototype. The photograph shows two raspberry pis, a NVIDIA Jetson Nano, a light source array, a transmissive spatial light modulator and an optical scanner.

Computation and Control Modules. At the heart of our computation hardware lies an NVIDIA Jetson Nano, where we control the backlight, the transmissive SLM, and the optical scanner.

The transmissive SLM of our prototype is connected to one of two Raspberry Pi Zeros using GPIO pins. The light source array of our prototype is directly connected to the second Raspberry Pi Zero using GPIO pins. Both of the Raspberry Pi Zeros are connected to USB ports of the NVIDIA Jetson Nano, where each USB connection is emulating an ethernet device and powering both Raspberry Pi Zeros from a NVIDIA Jetson Nano.

3.2. Software

A body of software components are required to simulate and to operate our functional prototype. We introduce these software components, accordingly.

Virtual prototype. A virtual prototype can assist in verifying our image reconstruction model in Fig. 2 and generating a dataset for determining image patches using Eq. (4). Using Blender, we build a virtual prototype of our functional prototype as rendered in Fig. 3. We capture photographs of our SLM in our functional prototype using a microscope, and we use the photographs as a reference to generate texture for the SLM in a virtual prototype. Using our virtual prototype, we are able to identify geometric transforms caused by optical scanning at each angle. Using the virtual prototype, we render images at 4k resolution from an user's perspective.

Training. Using CuPy, a GPU accelerated implementation of NumPy, we represent our imaging model described in Eq. (1) and our training model described in Eq. (4). Having a virtual prototype helps our training routine by providing the geometric transformation of an intermediate image T caused by a given trajectory of optical scanning and enables us to prepare a dataset of images for identifying image patches for a given scanning trajectory using our training model. A training session is an offline process. We can run our imaging model and our training on a NVIDIA GeForce RTX 2070 or a NVIDIA Jetson Nano.

Functional prototype. The light source array of our functional prototype is driven using bcm2835 library, a C/C++ library for Broadcom BCM 2835 chip found in Raspberry Pi, in which we are able to provide binary update rates up to 70 kHz. To control an optical scanner, DACs are driven using I2C communication protocol using C++. The SLM in our functional prototype is driven by a code base written in Python using PyGame to control the content on the SLM. All the software components of a functional prototype are controlled using sockets to form a client/server relation over TCP/IP written in C++.

4. Results

While patch scanning display method can be used in various types of SLMs (i.e. DMD, LCoS), we choose to rely on commodity hardware accessible at a relatively low cost. Details of our hardware and our optimization framework is provided in Materials and Methods section. We refer to points from that section while explaining our results. We design a virtual prototype using a physically accurate model as rendered in Fig. 3. We use our virtual prototype to assess theoretical limits in our image reconstruction model for our proposed display method. In order to demonstrate that we can realize our image reconstruction model in practice, we have built a functional prototype basing on our virtual prototype. A photograph of our functional prototype can be seen as in Fig. 4. Our functional prototype enables to simulate various situations in physical means and provide guidance in practice towards trade-offs in design choices for a future reference design. All our results are based on a predefined scan trajectory as described in Eq. (9).

4.1. Image quality

To quantitatively analyze the outcome of our technique, we rely on the most widely used perceptual metrics today, Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index. Note that PSNR and SSIM can be insufficient for image quality assessments in many cases as it can quickly fail to account for human perception. In order to address this issue, we also evaluate image quality with a learned perceptual metric closely resembling choice of a human subject [33].

Image quality of a patch scanning display is highly dependent on several different items, those are including the number of binary updates during one pass of a scan trajectory n in our image formation model presented in Eq. (3), basis size $m \times n$ px and optical scan trajectory Θ_x, Θ_y .

Binary updates. We compile a visual image quality comparison in Fig. 5, demonstrating a clear improvement in terms of image quality as more binary updates for a locally addressable backlight is used in one pass of a scan trajectory. The presentation rate of a patch scanning display is a function of number of binary updates required to generate a single frame, $f = \frac{n_{source}}{n_t}$, where n_{source} indicates how many binary updates a light source array has. As an example, a light source array that can modulate at 70 kHz and $n_t = 600$ can in principle provide a presentation rate of $f = \frac{70k}{600} \approx 117$ Hz, providing a gateway to double the presentation rate of the base SLM used in this study.

Qualities of bases and scan trajectory. The number of bases is equal to the rank r in our optimality problem presented in Eq. (4). According to Yuan and Oja [31], the number of bases are chosen so that $(N + M)r < NM$, in other words, $(\frac{kl}{mn} + mn)r < kl$. Given number of time steps in a trajectory $n_t = 600 \dots 4800$, specific size of our SLM 320×240 px, a training dataset V can be as large as $\frac{n_t \times 320 \times 240}{m \times n}$. In a training dataset V , provided images are likely to be very similar to each other as they are from the same scan trajectory, therefore we further simplify the suggestion of Yuan and Oja [31] as $r = mn$ specifically for our case. At this point, we start to observe a good visual quality in our simulations. For example, in a training session for a patch size of $m \times n = 10 \times 10$, $r = 100$ bases are needed, where all $r = 100$ bases can again be tiled on a SLM in 10×10 fashion following the naive approach for presenting image patches. Having a large number of bases stresses larger Θ_x and Θ_y , as scanned virtual images of a SLM has to cover a larger space to let each basis visit different portions of a target image.

4.2. Functional prototype

Considering the physical capabilities of our optical scanner and our SLM, we identify a sweet spot $m \times n = 6 \times 6$ px as basis size, leading to $r = 36$ bases tiled in a 6×6 fashion with Θ_x and Θ_y ranging from -1 degrees to 1 degrees. Note that the identified sweet spot may vary with different optical designs and hardware components. In our functional prototype, a light source in our backlight can illuminate a space of 15×15 pixels on our SLM, which poses a challenge to building a backlight array with denser amount of light sources. We choose to avoid the challenge of manufacturing denser light sources as we believe a display manufacturer can overcome the limitation in the future. We follow a routine where we set the optical scanner to a specific angle in a trajectory, flash the light source array all at once for a time of 500 usec, modulate the SLM as if illuminated by a light source array and repeat the process until all sample points in a single optical scan trajectory are visited. Due to this limitation in hardware, our routine allow us to generate images in the presentation rate slower than we suggest, but allows us to verify our imaging model. Therefore, we use a camera as a proxy to a human eye and capture photographs of our image reconstruction routine with long exposures. Specifically, we use a Canon Rebel T6 body with a 50 mm objective lens, F5.6 and ISO1600 with exposure times of 50 seconds to capture a single pass of an optical scan trajectory. Using $n_t = 600$ discrete evenly sampled time steps in a scan trajectory, $m \times n = 6 \times 6$, $r = 36$, we verify outcome as in Fig. 6, where the results from our virtual prototype closely matches our results from our functional prototype. Speckle pattern across all color channels in our results is due to long exposure times. A human observer is expected to see visuals free from such artifacts in a full product. We believe a display manufacturer could easily overcome slight differences in between functional and virtual prototype as the right engineering resources are available.

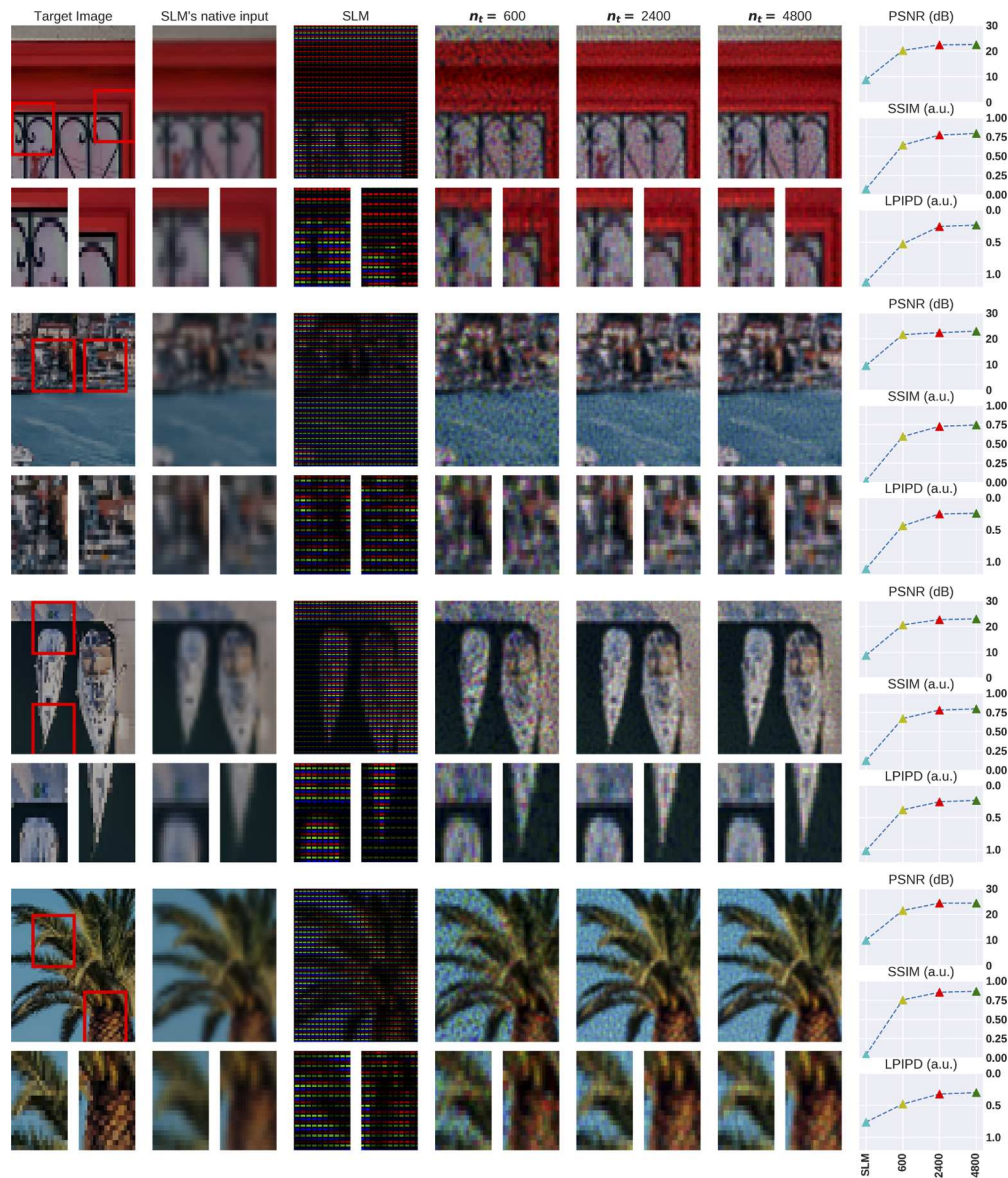


Fig. 5. Image quality assesment of patch scanning displays. Target image has four times more spatial resolution with respect to a base Spatial Light Modulator (SLM). As more binary updates n_t for a light source array is used in one pass of a scan trajectory, Peak signal-to-noise ratio (PSNR), structural similarity (SSIM) index and a learned perceptual image patch distance (LPIPD) [33] show improvements over a base SLM with a certain pixel fill factor. Displayed images are generated using our virtual prototype with $r = 36$ basis with $m \times n = 6 \times 6$ pxs patches. Source images courtesy Erhan Meço. See [Visualization 1](#) and [Visualization 3](#) for sample image reconstructions.

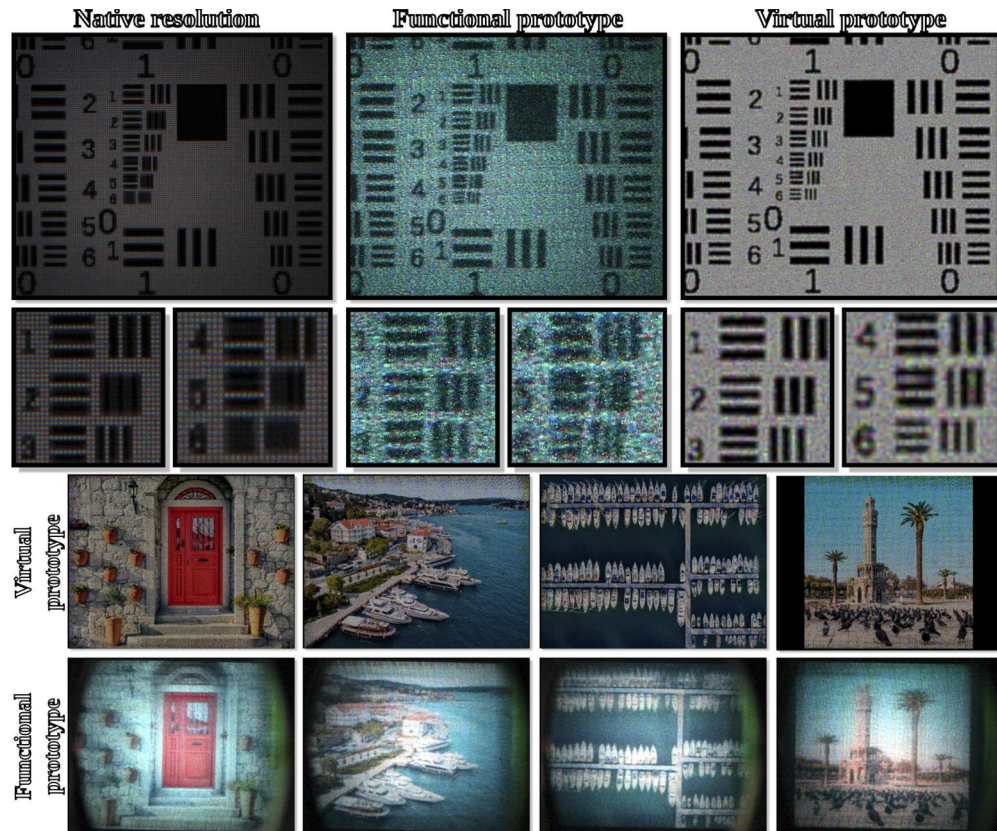


Fig. 6. Photographs from a functional prototype of a patch scanning display in comparison to a simulated virtual prototype and a base Spatial Light Modulator (SLM). Displayed results are generated using $r = 36$ bases with $m \times n = 6 \times 6$ pxs for both virtual and functional prototypes. Source images courtesy Erhan Meço.

5. Discussion

As SLMs and other constituent components of patch scanning displays improve, we believe patch scanning displays stand to reap the benefits in enhancing spatiotemporal qualities of a base SLM. Therefore, patch scanning displays can be best thought as a design method to improve future and today's displays.

Pixel fill factor. Inactive regions separating pixels in a SLM can cause a visual artifact entitled as screen-door effect, which can be visually observed in display systems under optical magnification (i.e. AR/VR near-eye displays and conventional projection displays). Eliminating screen-door effect in display systems is an active branch of research [14], where some of the solutions known to degrade the observed image quality. As can be observed in our demonstrated results, our proposal improves image quality and provides a solution against screen-door effect.

Optical scanner utilization. Designing a conventional optical scanner for high resolution displays (i.e. 4k, 8k) is a major research challenge [34]. A typical conventional scanning display choose k_y in Eq. (9) same as the rows to be addressed and typically suffers from visual separation in between rows, whereas in our approach with the given configuration, we can address 10 times more rows with the same scan trajectory without causing any visual separation in between rows. Having a smaller k_y is also known to increase overall scan speed of an optical scanner. Therefore, our proposal provides a potential to relax the optical scanner hardware requirements greatly.

Given our proposal also provides benefits of a conventional scanning display, designing an optical system with diffraction-limited performance over a large field of view remains a difficult task in near-eye display design, in which optical scanning is known to alleviate the design problem [35]. Beyond conventional scanners or conventional scan trajectories, our proposal support unconventional scanning methods such as rotating scan, consult the [Visualization 1](#) for examples on rotational scan based patch scanning displays. If non-mechanical scanners are required for a final implementation implementation, Liquid Crystal based switchable polarization gratings cascades [36] or Liquid Crystal based Phased Arrays [37] may provide fast enough alternatives in the future.

Programmable. Various aspects of a patch scanning display can be programmed on-the-fly such as qualities of basis, update order and scan trajectory. Therefore, an image can be generated using different combinations. We believe programmable aspect of our proposal may potentially serve as a hardware layer of protection for content distributors, allowing images to be viewed as long as correct combination is set for programmable display configuration.

5.1. Challenges

Fast and dense light sources. For a patch scanning display, faster light sources is important for increasing presentation rates. A light source has to have very short response times so that while scanning a single pixel footprint of a pixel doesn't extend to beyond intended in design. Current generation light sources such as lasers, LEDs or micro LEDs satisfy this need, lasers have already proven themselves as being part of a scanning display (i.e. Sony MPL-CL1A) with modulation rates in the order of GHz. Hoang et al. [38] demonstrate light source that can modulate at the rates of ~ 90 GHz, promising ultrafast light sources with short response times beyond available at a near future. Current generation of light sources such as OLEDs and micro LEDs can be densely populated on a substrate (i.e. pixel pitch of 10-100 μm).

Faster response time in SLMs. Response time of a SLM indicates how quickly a SLM settle at a desired value. This is especially relevant for our proposal as longer settling times in a SLM can cause poor performance during a scan. Going forward, for a possible product, usage of DMDs can lead to a robust design as DMDs offer settling times in the range of few microseconds.

Fast digital interface. The ability to update a display in a binary fashion at a larger representation rate can lead to a larger bandwidth requirement and isn't considered in digital display interfaces of today (i.e. DisplayPort). Currently available displays that can update in a binary fashion chooses to use a static drive logic, therefore not exposing their ability of binary updates to a regular user (i.e. DMDs or OLEDs). For rendering to a display that can update in binary in synchronization, a custom drive logic and a new digital interface is needed in the future. Lincoln et al. [39] demonstrates an interface that can update a binary display with 80 microsecond latency, promising a new digital fast interface that can support larger bandwidths at a near future.

Product design. Our hardware design for a functional prototype resembles a proof-of-concept and demonstrates one possibility out of many different possible designs. A possible future product design can consider usage of a DMD as a SLM, microLEDs or lasers as a light source array in a backlight and a MEMS scanner as an optical scanner among with a tailored projection optics, so that it can lead to a compact display design suitable for mobile miniaturized versions.

5.2. Future work

Due to wide spread usage in today's displays, patch scanning displays are designed with conventional incoherent light illumination in mind. Therefore, non-negativity as a constrain is required in our optimization. Coherent light illumination can enable us to take advantage from interference nature of light. Therefore, a display taking advantage from coherent illumination and interference can also relax the non-negativity constrain in optimization (i.e. using Principle Component Analysis). Extending findings from patch scanning displays to three dimensional

scanning image generation routines using coherent illumination, specifically for computer generated holography (CGH) based on optical scanning, can be a promising avenue for future work as CGH is widely promoted as the next frontier in display and computer graphics industries.

In terms of computation similar to patch scanning displays, there is a growing interest in decomposing a target visual into binary representation for next generation near-eye and volumetric 3D displays [40,41]. This necessitate a common methodology for binarization of visuals suitable to multiple display scenarios for distribution and display purposes. Perhaps, a new variant of patch scanning display in the future can also rely on methodologies from those other displays [40,41].

6. Conclusion

We present a new display method that synthesizes images with enhanced spatiotemporal resolutions by scanning learned image patches. Furthermore, we discuss the design constraints of patch scanning displays affecting image quality both qualitative and quantitative means. We build a functional prototype as a proof-of-concept, which can serve as a guide for constructing future variants of patch scanning displays. With further developments, we believe that patch scanning displays will be a cornerstone for research on the next generation of computational displays. Our approach demonstrates that patch scanning displays can address inherent hardware constraints with a highly programmable computational approach that ultimately leads to faster, better, and smarter displays.

Acknowledgments

The author thanks the anonymous reviewers for their useful feed-back. The author also thanks Duygu Ceylan, Michael Bauer and David Pająk for the fruitful and inspiring discussions improving the outcome of this research.

Disclosures

The authors declare no conflicts of interest.

Data and materials availability

All data needed to evaluate the conclusions in the manuscript are provided in the manuscript. Additional data related to this paper may be kindly requested from the author. See [Visualization 1](#), [Visualization 2](#), and [Visualization 3](#) for supporting content.

References

1. G. A. Koulieris, K. Akşit, M. Stengel, R. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," in *Computer Graphics Forum*, vol. 38 (Wiley Online Library, 2019), pp. 493–519.
2. D. B. Elliott, K. Yang, and D. Whitaker, "Visual acuity changes throughout adulthood in normal, healthy eyes: seeing beyond 6/6," *Optom Vis Sci.* **72**(3), 186–191 (1995).
3. D. Kelly, "Motion and vision. ii. stabilized spatio-temporal threshold surface," *J. Opt. Soc. Am.* **69**(10), 1340–1349 (1979).
4. J. Roberts and A. Wilkins, "Flicker can be perceived during saccades at frequencies in excess of 1 khz," *Light. Res. Technol.* **45**(1), 124–132 (2013).
5. J. Liu, S.-M. Morgens, R. C. Sumner, L. Buschmann, Y. Zhang, and J. Davis, "When does the hidden butterfly not flicker?" in *SIGGRAPH ASIA Technical Briefs*, (2014), pp. 1–3.
6. J. Davis, Y.-H. Hsieh, and H.-C. Lee, "Humans perceive flicker artifacts at 500 hz," *Sci. Rep.* **5**(1), 7861 (2015).
7. K. Noland, "The application of sampling theory to television frame rate requirements," BBC Res. Dev. White Pap. **282**, 1–22 (2014).
8. M. A. Khoei, F. Galluppi, Q. Sabatier, P. Pouget, B. R. Cottureau, and R. Benosman, "Faster is better: Visual responses to motion are stronger for higher refresh rates," *bioRxiv* p. 505354 (2018).

9. S. Kime, F. Galluppi, X. Lagorce, R. B. Benosman, and J. Lorenceau, "Psychophysical assessment of perceptual performance with varying display frame rates," *J. Disp. Technol.* **12**(11), 1372–1382 (2016).
10. C. Vieri, G. Lee, N. Balram, S. H. Jung, J. Y. Yang, S. Y. Yoon, and I. B. Kang, "An 18 megapixel 4.3 1443 ppi 120 hz oled display for wide field of view high acuity head mounted displays," *J. Soc. Inf. Disp.* **26**(5), 314–324 (2018).
11. B. Sajadi, M. Gopi, and A. Majumder, "Edge-guided resolution enhancement in projectors via optical pixel sharing," *ACM Trans. Graph.* **31**(4), 1–122 (2012).
12. C. Jaynes and D. Ramakrishnan, "Super-resolution composition in multi-projector displays," in *IEEE Int'l Workshop on Projector-Camera Systems* vol. 8 (2003).
13. F. Heide, D. Lanman, D. Reddy, J. Kautz, K. Pulli, and D. Luebke, "Cascaded displays: spatiotemporal superresolution using offset pixel layers," *ACM Trans. Graph.* **33**(4), 1–11 (2014).
14. B. Sitter, J. Yang, J. Thielen, N. Naismith, and J. Loneragan, "78-3: Screen door effect reduction with diffractive film for virtual reality and augmented reality displays," in *SID Symposium Digest of Technical Papers*, vol. 48 (Wiley Online Library, 2017), pp. 1150–1153.
15. W. Allen and R. Ulichney, "47.4: Invited paper: Wobulation: Doubling the addressed resolution of projection displays," in *SID Symposium Digest of Technical Papers*, vol. 36 (Wiley Online Library, 2005), pp. 1514–1517.
16. F. Berthouzoz and R. Fattal, "Resolution enhancement by vibrating displays," *ACM Trans. Graph.* **31**(2), 1–14 (2012).
17. B. Sajadi, D. Qoc-Lai, A. H. Ihler, M. Gopi, and A. Majumder, "Image enhancement in projectors via optical pixel shift and overlay," in *IEEE International Conference on Computational Photography (ICCP)*, (IEEE, 2013), pp. 1–10.
18. T. Zhan, J. Xiong, G. Tan, Y.-H. Lee, J. Yang, S. Liu, and S.-T. Wu, "Improving near-eye display resolution by polarization multiplexing," *Opt. Express* **27**(11), 15327–15334 (2019).
19. J.-Y. Wu, P.-Y. Chou, K.-E. Peng, Y.-P. Huang, H.-H. Lo, C.-C. Chang, and F.-M. Chuang, "Resolution enhanced light field near eye display using e-shifting method with birefringent plate," *J. Soc. Inf. Disp.* **26**(5), 269–279 (2018).
20. P. Didyk, E. Eisemann, T. Ritschel, K. Myszkowski, and H.-P. Seidel, "Apparent display resolution enhancement for moving images," *ACM Trans. Graph.* **29**(4), 1 (2010).
21. H. Lee and P. Didyk, "Real-time apparent resolution enhancement for head-mounted displays," *Proc. ACM on Comput. Graph. Interact. Tech.* **1**(1), 1–15 (2018).
22. G. Tan, Y.-H. Lee, T. Zhan, J. Yang, S. Liu, D. Zhao, and S.-T. Wu, "Foveated imaging for near-eye displays," *Opt. Express* **26**(19), 25076–25085 (2018).
23. J. Kim, Y. Jeong, M. Stengel, K. Akşit, R. Albert, B. Boudaoud, T. Greer, J. Kim, W. Lopes, and Z. Majercik et al., "Foveated ar: dynamically-foveated augmented reality display," *ACM Trans. Graph.* **38**(4), 1–15 (2019).
24. K. Akşit, P. Chakravarthula, K. Rathinavel, Y. Jeong, R. Albert, H. Fuchs, and D. Luebke, "Manufacturing application-driven foveated near-eye displays," *IEEE Trans. Visual. Comput. Graphics* **25**(5), 1928–1939 (2019).
25. M. K. Brown, "Scanning projector with vertical interpolation onto horizontal trajectory," (2013). US Patent 8,371,698.
26. E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting* (Morgan Kaufmann, 2010).
27. D. C. Hood and M. A. Finkelstein, "Sensitivity to light," *Handbook of Perception and Human Performance. (Vol. 1: Sensory Processes and Perception)* John Wiley and Sons 10–20. (1986).
28. S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg, "Time-dependent visual adaptation for fast realistic image display," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, (ACM Press/Addison-Wesley Publishing Co., 2000), pp. 47–54.
29. O. Rinner and K. R. Gegenfurtner, "Time course of chromatic adaptation for color appearance and discrimination," *Vision Res.* **40**(14), 1813–1826 (2000).
30. D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, (2001), pp. 556–562.
31. Z. Yuan and E. Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," in *Scandinavian Conference on Image Analysis*, (Springer, 2005), pp. 333–342.
32. M. Leblanc, "Etude sur la transmission électrique des impressions lumineuses," *La Lumière électrique* **2**, 477–481 (1880).
33. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (xxx), 2018586–595.
34. H. Urey and D. L. Dickensheets, "Display and imaging systems," *MOEMS and Applications* (2005).
35. H. Urey, "Optical advantages in retinal scanning displays," in *Helmet-and Head-Mounted Displays V*, vol. 4021 (International Society for Optics and Photonics, 2000) pp. 20–26.
36. J. Buck, S. Serati, R. Serati, H. Masterson, M. Escuti, J. Kim, and M. Miskiewicz, "Polarization gratings for non-mechanical beam steering applications," in *Acquisition, Tracking, Pointing, and Laser Systems Technologies XXVI*, vol. 8395 (International Society for Optics and Photonics, 2012), p. 83950F.
37. P. F. McManamon, P. J. Bos, M. J. Escuti, J. Heikenfeld, S. Serati, H. Xie, and E. A. Watson, "A review of phased array steering for narrow-band electrooptical systems," *Proc. IEEE* **97**(6), 1078–1096 (2009).
38. T. B. Hoang, G. M. Akselrod, C. Argyropoulos, J. Huang, D. R. Smith, and M. H. Mikkelsen, "Ultrafast spontaneous emission source using plasmonic nanoantennas," *Nat. Commun.* **6**(1), 7788 (2015).

39. P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs, "From motion to photons in 80 microseconds: Towards minimal latency for virtual and augmented reality," *IEEE Trans. Visual. Comput. Graphics* **22**(4), 1367–1376 (2016).
40. K. Rathinavel, H. Wang, A. Blate, and H. Fuchs, "An extended depth-at-field volumetric near-eye augmented reality display," *IEEE Trans. Visual. Comput. Graphics* **24**(11), 2857–2866 (2018).
41. Y. Jo, S. Lee, D. Yoo, S. Choi, D. Kim, and B. Lee, "Tomographic projector: large scale volumetric display with uniform viewing experiences," *ACM Trans. Graph.* **38**(6), 1–13 (2019).