Image Capture Through TFT Arrays

Neil Emerton*, David Ren**, Tim Large*

*Microsoft Applied Sciences Group, Redmond WA
** Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA

Abstract
Images captured through displays are dim and blurred owing to attenuation and diffraction. Display structures can result in angular frequencies in the scene being completely blocked. We describe an optical method to recover these frequencies resulting in a more complete capture of the scene.

Author Keywords
Under-display camera; display panel; diffraction; modulation transfer function (MTF); image restoration; structured light

1. Introduction
There are several motivators driving location of a front-facing camera behind the main display screen:

- A cleaner industrial design owing to the absence of camera windows or notches – especially in mobile products
- The potential to make the display the same size as the physical device envelope: 'bezel-free'
- The potential to site the camera(s) on-screen where the main image is displayed so that, in a videoconference, the camera is located near the on-screen image of the remote participant so enabling gaze-correct imaging.

However, display screens are not usually designed with the needs of through-screen imaging in mind. Some emerging transmissive OLED displays [1] offer improved through-screen performance but even these result in image artifacts owing to diffraction, scattering and attenuation of the signal. In mobile device applications even the state-of-the-art unimpeded camera performance is limited by photon noise for daylight imaging and additive noise for low-light imaging. Further attenuation of the signal by the display therefore decreases the already modest signal-to-noise ratio in the image.

For some display types deconvolution and denoising methods along with machine learning methods have been shown to result in significant improvements in image quality [2]. In the next section we explain why there are some classes of displays and some types of object features that will prevent a good image reconstruction even with a well-trained machine learning method. In any case, for some biometric applications it is essential that e.g. facial features are derived from a true image and are not composed of synthetic data.

Having identified these potential failure cases, we proposed and implemented an optical hardware-based method to recover these types of objects imaged through the most challenging screen designs.

2. Diffraction characteristics of displays
Most displays consist of a square grid of pixels. In the case of LCDs the individual sub-pixels are generally rectangles or slanted rectangles and each transmits approximately 1/3 of the visible spectrum. For transmissive OLED displays the manufacturers provide a clear 'window' aperture adjacent to the emitting area to facilitate imaging. The requirements on the layout of the emitters,

active electronic elements and electrical conductors often result in a roughly rectangular shape to this window also.

Although the detailed shape of the aperture varies with display design, the collection apertures through which the camera views the world can be approximated as a periodic array of near-rectangular slits, an example of which shown in Figure 1. The area fraction of this slit can be quite high (~40%) for large (e.g. 55" diagonal) HD displays but becomes smaller (~20% or less) for 4K displays and may be much smaller for mobile displays. The relatively large aspect ratio means that the diffraction is much worse in the (typically) horizontal direction than the vertical direction. Therefore, for the purposes of this paper, we restrict our attention to the horizontal diffraction and the consequent impact on the imaging in this direction.

As the relative slit width diminishes, the proportion of diffracted light relative to straight-through light increases; subjectively the image becomes more blurred. However, this subjective assessment masks the fact that the blur can be such as to eliminate entirely certain patterns in the scene from the through-screen image. Conventionally, the modulation transfer function (MTF) is used to characterize the spatial or angular frequency response of imaging systems [3]. The MTF is the autocorrelation of the pupil function [3] and, in the case of an under-display camera, the screen itself forms the pupil, bounded by the aperture of the imaging lens. Accordingly, in Figure 2, we plot the simulated MTF of an under-display camera system for a range of relative slit widths. The angular frequency, f, is in units of λ/p, where p is the pixel pitch. The MTF of a nominal imaging lens is indicated by the dotted line

Figure 1. Image of 3x3 pixels of proposed 4K tOLED structure. White regions are clear on the design. Pixel pitch is 315 µm.

in Figure 2, and the solid line is the combination of the lens with the screen. When the relative slit width is greater than 50%, 'null bands' appear in the MTF, meaning that these angular frequencies are not passed by the screen. Figure 2 shows that the smaller the slit, the wider the band of frequencies excluded. An object in the
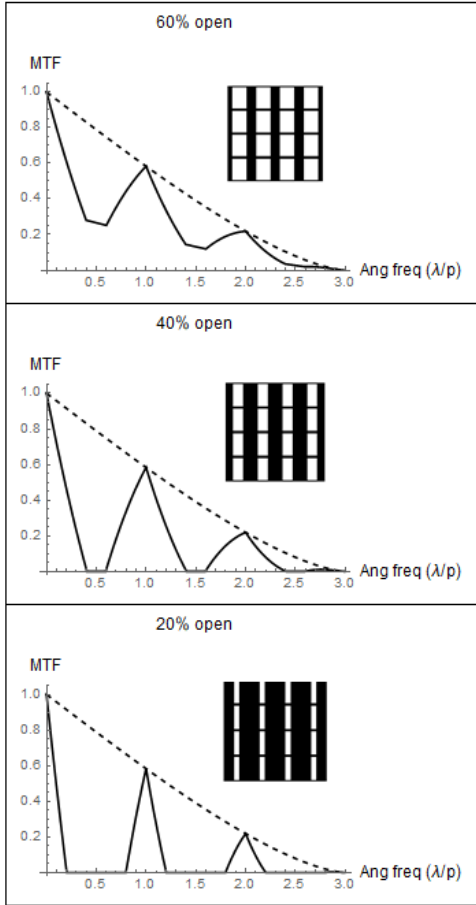
Figure 3. Screen designs and simulated MTF curves for three different open area fractions. White indicates a clear region. The solid line is the MTF through the screen, the dotted line is the MTF of the imaging lens.

scene with angular frequency information principally clustered in this band (e.g. a striped shirt, a fence, a zebra) will not be visible through the screen at all. In such cases, classical deconvolution methods and even well-trained machine learning systems cannot faithfully reconstruct the object.

3. Proposed solution

Consider the sample MTF curve and object angular spectrum shown in Figure 3. If the object frequencies blocked by the null band were shifted down to frequencies that are visible through the display, then they could be captured. This can be achieved by illuminating the scene with a sinusoidal pattern at the cut-off frequency f_c. This could be achieved by a suitable projection system [4]. However, to do that, the projection system must be located outside of the screen as, if located behind the screen, then f_c would be blocked. Locating the projection system outside the screen defeats two of the main motivations for this work.
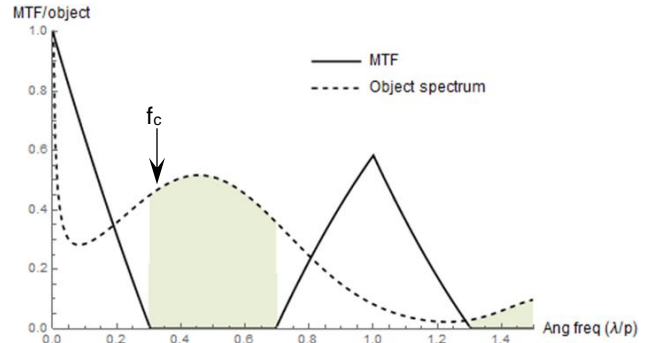

Figure 2. Screen and MTF with sample object spectrum. Shaded region denotes lost frequency content starting at f_c.

The proposed solution is to create the frequency f_c using coherent light and spatial filtering. If we consider a simplified model of the screen as consisting of a square point grid with a substantially rectangular slit of width d at each point on the grid and consider only a narrow range of wavelengths centred on λ then, in incoherent light f_c is λ/d. It is this angular frequency that we wish to project. If we create a point source at each edge of the slit separated by d then, in the far-field, we will produce a set of interference fringes whose angular frequency is indeed λ/d [5]. This is because, by artificially constructing the pair of point sources, we have removed the zeroth order term that would normally be present in the pupil of a projection system and we have effectively doubled the angular frequency of the fringes. We need both spatially and temporally coherent light (or at least partially so) because: 1) the sources need to be small so that they can be placed at the extremities of the slit, 2) we will need to illuminate a large scene with sufficient light to be detectable without wasting optical power and 3) the sources need to generate a large number of sinusoidal interference fringes. 1) - 3) imply spatial coherence and 3) additionally implies temporal coherence or equivalently narrow spectral bandwidth. The spectral bandwidth required can be estimated by demanding that the fringes fill the field of view of the camera with half-angle $\theta_{1/2}$; thus, the optical path difference at the edge of the field must be much less than the coherence length of the source [6]:

$$d \sin(\theta_{1/2}) \ll \frac{\lambda^2}{\delta\lambda}$$

For a 55" 4K screen and for typical pixel layouts, d may be of order 80μm and the half field of view of the through-screen camera may be around 35°. For the wavelength used in the experimental work, 780 nm, then we have $\delta\lambda \ll 13$nm. This, together with the requirements for small effective source size and power efficiency, dictates either a laser source or a superluminescent source.

4. Experimental configuration

With reference to Figure 4, our experiment used near infra-red light at 780nm from a fibre-coupled single-mode laser diode collimated by lens L1 to illuminate a Thorlabs 4K Exulus spatial light modulator (SLM). The SLM was used to create a sinusoidal or square-wave phase pattern to diffract the reflected light into a series of diffraction orders. The SLM enabled flexibility and tuning of the effective point source positions in the screen slit along with the ability to shift the phase of the projected pattern. This is achieved by varying the spatial frequency of the pattern on the SLM and its spatial phase. The phase excursion of the SLM pattern was adjusted to maximise the amount of light diffracted in to the +1st and -1st diffraction orders. Ideally, 100% of the light would be distributed
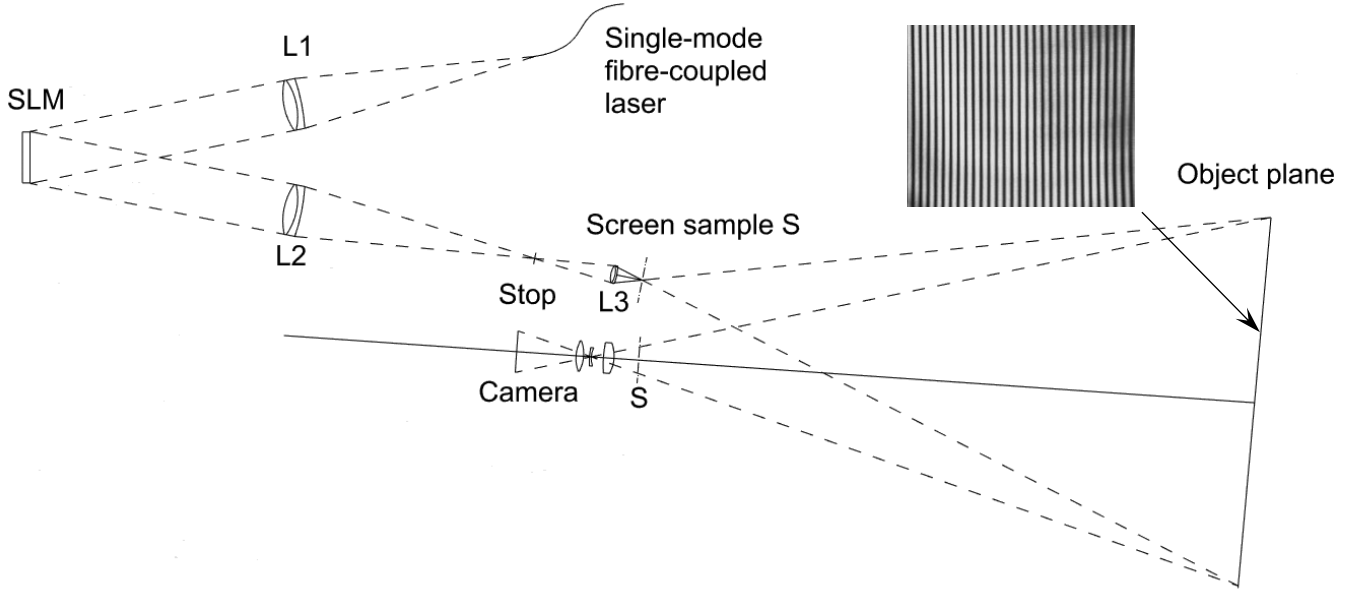
Figure 4. Schematic of optical system layout for capture of structured illumination images.

equally in the +1st and -1st orders but, with a phase-only SLM, this is not achievable. Instead we rely on spatial filtering of the diffracted light by a stop and by the screen itself to limit the light propagating to the scene to the required orders. Light reflected from and modulated by the SLM was focused by a lens L2 and, in the focal plane of L2, a stop was used to block the zeroth diffraction order. The +1st and -1st orders and higher orders were then allowed to propagate to the lens L3 (a 10X microscope objective) and the screen sample S was placed in the conjugate imaging plane so that the new spacing of the +1st and -1st orders was 1/10th of that previously. The lateral position of S and the spatial frequency on the SLM were adjusted so that high-contrast sinusoidal fringes were obtained in the far-field relative to the screen apertures. The spatial frequency of the fringes was maximized in this procedure by varying the SLM spatial frequency and adjusting the screen lateral position so that the spatial frequency was close to cut-off. Higher diffraction orders were then blocked by the screen itself. The optical power launched through the screen sample was approximately 10mW.

A Point Grey Grasshopper camera GS3-U3-23S6C-C with a Sony IMX 174 image sensor with 5.86 μm pixel size was used to acquire the images. The infra-red cut-off filter was removed. A 16mm focal length Navitar lens was used to image the object through the same screen design as the illumination. The camera aperture was set to f/1.8, the frame rate was 30 fps with 30 millisecond integration time and the analog gain was set to 28 dB.

Three images were acquired with the spatial phase of the projected pattern set to 0, $2\pi/3$ and $4\pi/3$ respectively. In the spatial domain each captured image is the original object distribution multiplied by the projected fringe pattern. In the Fourier domain this multiplication is equivalent to a convolution and we know that the convolution kernel for each image consists of three δ-functions at $-f_c$, 0 and f_c, and that the relative phases of the side peaks are determined by the spatial phase of the projected pattern. As a result, the spectrum of the scene, $\tilde{o}(f)$, is duplicated and shifted to the locations of the δ-functions and summed. In the spatial domain, we observe moiré patterns as evidence of the high frequency signal being modulated to the low frequency band. Under the structured illumination, the spectrum of each acquired image, denoted by $\tilde{I}(f)$, can be described as the following linear relationship:

$$\begin{bmatrix} \widetilde{I_0}(f) \\ \widetilde{I_{\frac{2\pi}{3}}}(f) \\ \widetilde{I_{\frac{4\pi}{3}}}(f) \end{bmatrix} = M \begin{bmatrix} \tilde{o}(f) \\ \tilde{o}(f - f_c) \\ \tilde{o}(f + f_c) \end{bmatrix},$$

where M is the linear multiplexing matrix:

$$\mathbf{M} = \begin{bmatrix} 1 & \alpha & -\alpha \\ 1 & \alpha e^{i\left(\frac{2\pi}{3}\right)} & \alpha e^{-i\left(\frac{2\pi}{3}\right)} \\ 1 & \alpha e^{i\left(\frac{4\pi}{3}\right)} & \alpha e^{-i\left(\frac{4\pi}{3}\right)} \end{bmatrix},$$

and α is a constant that defines the modulation depth of the fringes; for fully modulated fringes $\alpha = \frac{1}{2}$, but in a real system it is typically somewhat less than this. We note that this assumes perfect knowledge of the illumination pattern. However, this assumption is not generally valid, and more estimation and calibration are required to refine the image recovery [7].

We recover the unshifted and shifted spectra of the scene then by simply inverting the process above. Notice that M is invertible and only needs to be computed once for a given system set-up. Let this inverse matrix be M^{-1}.

We first form an image stack consisting of the Fourier transforms of the acquired images for the three different phase patterns. We refer to one image in terms of its height and width dimensions and depth to refer to the images in the stack. For each pixel in the output we form the dot product of each row of M' with the vector composed of the three pixels in the depth of the image stack. At level 0 in the stack we have the object spectrum and at levels 1 and 2 we have a frequency shifted form of the object spectrum. We then shift the Fourier transform of the images at levels 1 and 2 along the width dimension by an amount corresponding to the spatial frequency of the illumination pattern. Finally we average the stack in the depth dimension to arrive at a single Fourier transform of the spectrum and transform back to the spatial domain to obtain the image with the recovered resolution.
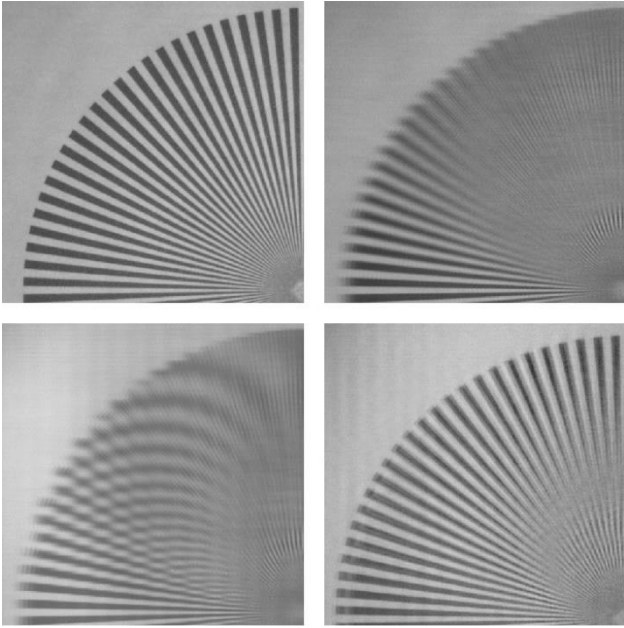
Figures 5.1-5.4. Clockwise from top left: Original object viewed without screen, object captured in unstructured near-IR illumination, single capture of object under structured illumination (note moiré fringes), processed image from 0, 2π/3 and 4π/3 captures.
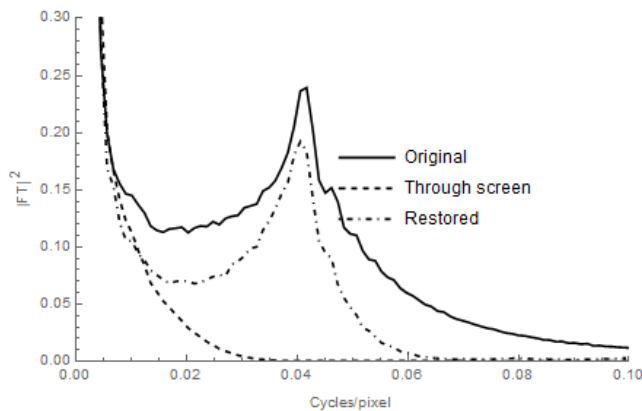

Figure 6. Image spectral power distributions for original, through-screen and restored images.

5. Results

We used a Siemens sector star chart as our test image as it conveniently contains a wide range of spatial frequencies at all angular orientations. Figure 5.1 shows one quadrant of the chart as imaged with the camera without a screen in place. Figure 5.2 shows the chart imaged through the screen sample. Note that, to the right and left of the chart the spokes are imaged well, but at the top and the bottom the base frequencies of the spoke structures are totally lost. These frequencies lie in the first null band region of the MTF. Close inspection of Figure 5.2 shows that some of the higher spatial frequencies closer to the center of the chart are visible. These frequencies lie within the second peak of the MTF curve.

Figure 5.3 is one of the three images acquired with the structured illumination; the aliased frequencies are visible at the top of the image. Figure 5.4 is the processed recovered image. The region of the sector chart at the top is substantially recovered.

Figure 6 is a plot of the mean power in the 1-D Fourier transform of the image. Specifically, each image was Fourier transformed and the square modulus calculated, this was then averaged down the image. This shows that much of the spectral power in the image has been restored in the reconstruction.

6. Conclusion and discussion

In this work, we propose an active structured illumination method to recover lost frequency content due to diffraction when imaging through transparent displays. First, we demonstrated experimentally the possibility of modulating the scene with a IR sinusoidal pattern so that the frequency is not blocked by the display. Then, we showed a simple inverse model to recover lost frequency content in the scene using modulated images acquired. Although monochromatic results are presented, in the future this method can be combined with color imaging and further image processing to enhance the quality of the image captured by the under-display camera system.

7. Acknowledgements

8. References

1. Planar Systems Corp. https://www.planar.com/products/transparent-oled-displays/

2. Lim S, Zhou Y, Emerton N, Large T, Bathiche S. Image restoration for display-integrated camera. Submitted to SID DisplayWeek 2020

3. Welford W T. Aberrations of Optical Systems. 1st ed. Bristol and Boston: Adam Hilger; 1986. 249-260 p.

4. Gustafsson M G L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J. Microsc. 2000 May; 198, Pt 2: 82-87p

5. Born M, Wolf E. Principles of Optics. 7th ed. New York: Cambridge University Press; 1999. 290 p.

6. Born M, Wolf E. Principles of Optics. 7th ed. New York: Cambridge University Press; 1999. 356 p.

7. Yeh LH, Tian L, Waller L. Structured illumination microscopy with unknown patterns and a statistical prior. Biomed. Opt. Express. 2017 Feb 1;8(2):695-711.