Lensless Imaging with a Controllable Aperture

Assaf Zomet[*]   Shree K. Nayar

Computer Science Department
Columbia University
New York, NY, 10027
E-mail: zomet@humanEyes.com, nayar@cs.columbia.edu

Abstract

In this paper we propose a novel, highly flexible camera. The camera consists of an image detector and a special aperture, but no lens. The aperture is a set of parallel light attenuating layers whose transmittances are controllable in space and time. By applying different transmittance patterns to this aperture, it is possible to modulate the incoming light in useful ways and capture images that are impossible to capture with conventional lens-based cameras. For example, the camera can pan and tilt its field of view without the use of any moving parts. It can also capture disjoint regions of interest in the scene without having to capture the regions in between them. In addition, the camera can be used as a computational sensor, where the detector measures the end result of computations performed by the attenuating layers on the scene radiance values. These and other imaging functionalities can be implemented with the same physical camera and the functionalities can be switched from one video frame to the next via software. We have built a prototype camera based on this approach using a bare image detector and a liquid crystal modulator for the aperture. We discuss in detail the merits and limitations of lensless imaging using controllable apertures.

1. Lensless Imaging with Apertures

Virtually all cameras today have lenses. Lenses are useful as they focus the light from the scene on the image plane to form bright and sharp images. We are so accustomed to cameras with lenses that we often overlook their fundamental limitation; lenses severely constrain the geometric and radiometric mapping from the scene to the image. The goal of this paper is to develop a new approach to imaging that facilitates a new class of mappings from the scene to the image, thereby enabling the camera to perform a wide set of imaging functionalities.

We propose in this paper a novel, flexible video camera that does not have a lens. The camera design is simple. It consists of two components, an image detector and a special aperture, that are placed at a small distance apart. Figure 1(a) shows the aperture in its simplest form - a flat light attenuator whose transmittances are controllable in space and in time.
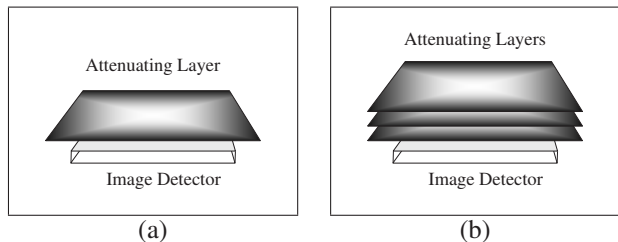

Figure 1. **The proposed camera has two components: a detector and an aperture. (a) In its simplest form, the aperture is a light attenuating layer whose transmittances are controllable in space and time. A practical way to implement a controllable attenuating aperture is by using liquid crystal sheets. (b) In its general form, the aperture is a stack of parallel attenuating layers. This approach leads to a flexible imaging system that can achieve a wide range of mappings of scene points to image pixels.**

Conventional apertures can be realized as a special case of this aperture, by using a constant binary transmittance pattern. Figure 1(b) show the aperture in its complete form. A stack of several controllable light attenuating layers at different distances from the detector are used. One way to implement controllable attenuating layers is by using liquid crystal sheets [1].

Figure 2 highlights the difference between a conventional lens camera and our lensless camera. An ideal lens camera, shown in Figure 2(a), focuses the scene on the image plane. Each point on the image detector integrates light emanating from a single point in the scene. Therefore, the aperture influences only the total brightness of the image and the local blur in defocused areas. In contrast, our lensless imaging system, shown in Figure 2(b), has no focusing. Each point on the image detector integrates light emanating from the entire field of view. Prior to the integration, the 2D light field associated with each image point is modulated by the attenuating aperture. Therefore, it is the aperture that determines the geometry and photometry of the imaging process.

We show that a careful selection of the transmittance pattern of the aperture makes it possible to modulate the light in useful ways that cannot be achieved with conventional cameras. Moreover, since the aperture is controllable, the imag-

---

[1]Other spatial light modulators can be used as well such as a Digital Micromirror Device (DMD) or Liquid Crystal On Silicon (LCOS).

ing properties of the camera can be changed from one video frame to the next. The following are some of the distinctive capabilities of the camera.

- **Instantaneous Field of View Changes**: The camera can change its viewing direction instantaneously to arbitrary directions by merely modifying the transmittance pattern of the aperture. In contrast, conventional cameras rely on pan-tilt motors, which are limited by mechanical constraints and produce motion blur.

- **Split Field of View**: The camera can capture disjoint parts of the scene in a single frame without capturing the regions in between them. A system that uses the camera can select which parts of the scene are captured at each time instance. This way, the camera can capture far apart *moving* objects with higher resolution. In contrast, conventional cameras are forced to distribute the limited resolution of the detector uniformly over a wide field of view.

- **Camera as a Computational Sensor**: The camera can modulate the light such that the captured images are the results of computations applied optically to scene radiances. This way, the camera can be used to perform expensive computations during image formation. In contrast, conventional cameras cannot perform such computations due to the rigid scene-to-image mapping performed by lenses.

Therefore, by removing the lens, we obtain a highly flexible imaging system. One might consider that this flexibility is obtained by sacrificing the overall resolution and brightness of the image. After all, this is the main reason for using a lens in an imaging system. However, we will show that these limitations can be overcome by using a larger video detector (Appendix A). Moreover, we show that image brightness can be further intensified using special aperture designs called "coded apertures".

The ideal design of our camera involves the fabrication of the detector and the attenuating layers as one physical device. In our prototype implementation, we used an off-the-shelf digital still camera without the lens as the detector and an off-the-shelf LCD in front of it as the controllable aperture. In cases where multiple attenuating layers were needed, we have used physical apertures with constant transmittance functions. Using our prototype, we demonstrate the use of our imaging system in different applications.

2. Related Work

This work was inspired by the recent work by Nayar et al. [8] that coined the term "programmable imaging." This previous work proposed a camera with lenses and an array of micro-mirrors. By controlling the orientations of the micro-mirrors, the authors showed that pixel-wise multiplications and instantaneous changes of viewing directions can be done in the optics. In other work [7], the authors proposed a camera with a lens and a light attenuator that can also perform pixel-wise multiplications. While both our camera and the
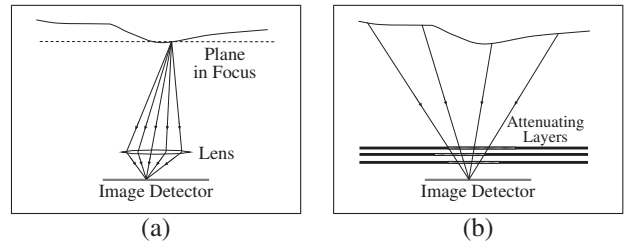

Figure 2. **Comparison between a lens-based camera and the proposed lensless camera. (a) With a lens, each point on the image detector ideally collects light emanating from a single point in the scene. (b) With the lensless camera, each point on the detector collects light emanating from the entire scene and attenuated by the aperture. Manipulations can be done to a 4D set of light rays before the final 2D image is captured. This allows the camera to perform new imaging functionalities.**

cameras in [7, 8] are controllable, there are fundamental differences in the way images are formed. Specifically, the cameras in [7, 8] use a lens to focus the scene on the detector. Therefore, at each image point they modulate the light emanating from a single point in the scene. In contrast, our camera modulates at each image point the light coming from the entire field of view. In other words, the cameras in [7, 8] modulate the 2D image whereas our camera modulates the 4D light field associated with the image detector prior to the capture of the 2D image (see Figure 2). As a result, our camera can perform several new imaging functionalities that have not been possible in the past. Finally, from a practical viewpoint, our camera can be very inexpensive and compact (essentially a thicker detector).

A modulation of the 4D light field with a light attenuator was proposed by Farid et. al. [2]. A defocused lens in conjunction with an attenuating layer were used for scene depth estimation. In contrast, our camera has no lens, therefore providing different modulations to the incident light field. Moreover, our camera includes multiple attenuating layers, which, as we show, provide more general modulations.

3. The Camera Prototype

Our prototype implementation of the camera is shown in Figure 3. It includes an off-the-shelf LCD (MTR-EVUE-4BW from EarthLCD) for the aperture and an off-the-shelf digital camera (EOS-20D from Canon) without the lens for the image detector. The major considerations in selecting the LCD were the pixel size and the contrast ratio. In order to capture high quality images, ideally, the pixels on the LCD should be as close as possible to the optimal pinhole size. The optimal pinhole size depends on the distance of the pinhole from the detector (see Appendix A for details). Since in our case the LCD had to be attached to the lens mount of the camera at a distance of $55mm$, the optimal pinhole size was $0.33mm \times 0.33mm$. The LCD should also have a high contrast ratio to be able to approximate zero transmittance. Unfortunately, at this point in time, most commercially-available high-contrast LCDs have 3 sub-pixels (R;G;B) per pixel, so that the physical pixels have an aspect ratio close to 1:3. The LCD we selected has
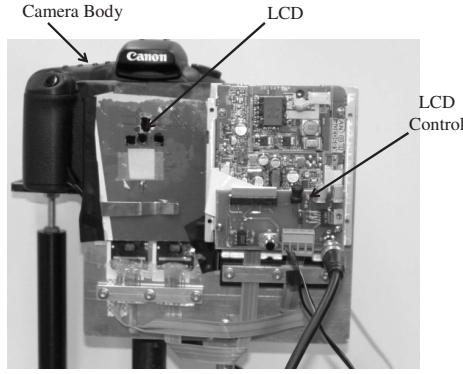
Figure 3. **Our camera prototype consists of the body of a Canon EOS-20D digital still camera with an LCD in front of it. To overcome the low contrast ratio of the LCD, most of its unused area was covered with a cardboard. In experiments that required the use of multiple attenuating layers, the additional layers were physical apertures.**

close-to-square pixels ($0.21mm \times 0.26mm$) and a published contrast ratio 1:50. In practice, we found that the contrast ratio was 1:14 or less. We therefore blocked most of the unused area of the LCD with cardboard, as can be seen in Figure 3. Due to the low contrast ratio of our LCD, LCD pixels that were supposed to block the light, in practice, transmitted considerable amounts of light. To overcome this limitation, the images used in our experiments were captured as follows: We first applied the desired transmittance pattern to the LCD and captured image I_1. Then, we applied a uniform zero transmittance pattern and captured image I_0. The image used as the output of the camera was the difference between these images: $I = I_1 - I_0$.

Most LCDs are coated with diffuse layers that improve the display quality, but are harmful for our purposes. Our LCD had a mild diffuse coating that introduced additional image blur. In addition, the attenuation of LCDs depends on the viewing angle [1]. In the case of the LCD we used, we observed a lower contrast ratio in large angles. To account for this effect, in our experiments we used a pre-calibrated photometric correction function to our captured images.

It should be emphasized that while we used an off-the-shelf grayscale LCD, color LCDs are becoming available with a contrast ratio of 1:1000 and wider viewing angle responses. In other words, in the near very future it will be possible to develop a prototype of our camera that produced images of much higher quality. Due to the limitation of our current prototype, all the experiments reported in this paper were done indoors under strong lighting and all presented videos were obtained by capturing sequences of still images.

Finally, for our second attenuating layer (needed for the optical correlation and for the split field of view experiments) we used a physical aperture with the appropriate attenuation function.

4. Imaging Without Lenses

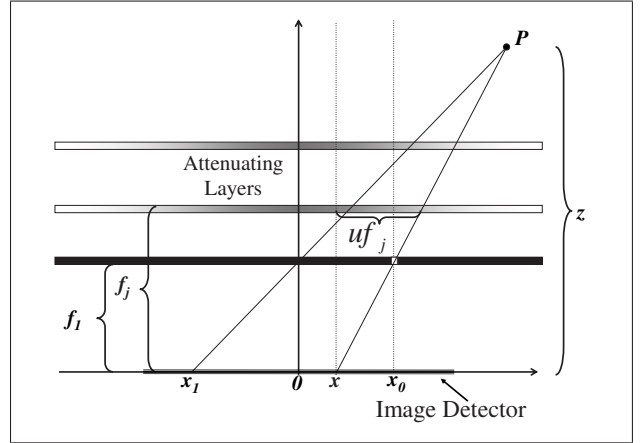We first explore the set of scene-to-image mappings that can be implemented with the proposed camera. We derive

Figure 4. **The scene-to-image mappings that can be implemented with the camera. This illustration is used for the proof of Proposition 1.**

a simple relation for the mapping when the scene is far relatively to the camera size and show the difference between imaging with the usual aperture and imaging with a multi-layered aperture. To keep notations simple, the derivation is given for transmittance patterns in which diffraction effects are negligible and for a one-dimensional camera. The generalization to a 2D camera is straightforward. It is further assumed the scene is a plane [2] parallel to the image plane at distance z.

Proposition 1 *Define a camera composed of an image plane and an attenuating aperture. The aperture is a set of K parallel flat layers at distances $f_1..f_K$ from the image plane. Let $0 \leq T_j(x) \leq 1, j = 1..K$ be the transmittance functions of the layers. The image is captured by the image detector, a finite rectangular area centered at the origin of the image plane. Let $S_f(u)$ be an image of an ideal (diffraction-free) pinhole camera with the pinhole at distance f_1 from the center of the image detector.*

Then the image brightness at point x is given by:

$$I(x) = \int \prod_{j=1}^{K} T_j \left(x - u\frac{f_j}{f_1} \right) S_f \left(u + (u-x)\frac{f_1}{z-f_1} \right) du. \quad (1)$$

Define w as an upper bound on the width of the camera (aperture and detector). Then, in the limit, when $z >> w$ and $z >> f_1$, we get:

$$I(x) = \int \prod_{j=1}^{K} T_j \left(x - u\frac{f_j}{f_1} \right) S_f(u) du. \quad (2)$$

Proof: *Figure 4 shows the camera and a scene point P. We first consider a particular case, in which the first layer is blocked except for a pinhole located at an offset x_0. Scene point P is projected through the pinhole to image point x. Were the pinhole located at offset 0, the point P would be projected to the point x_1. Therefore, the image brightness at point x is given by*[3]:

---

[2]This assumption is made only to simplify the notation. Otherwise, one can associate a z value with each image point and derive a similar result.

[3]Note that here we assume that the radiance emanating from P towards x_1 equals the radiance emanating from P towards x. This approximation depends on the distance of the scene point and its reflectance properties. For a scene that is distant relatively to the size of the camera, the solid angle at point x subtended by a pinhole at location x_0 and the solid angle at point x_1 subtended by a pinhole at location 0 can be approximated to be the same.

$$I_{x_0}(x) = \prod_{j=2}^{K} T\left(x + (x_0 - x)\frac{f_j}{f_1}\right) S_f(x_1). \qquad (3)$$

Note that S_f was defined as the pinhole image for a pinhole location at offset 0. From similarity of triangles, $\frac{x - x_1}{z} = \frac{x_0}{z - f_1}$. Reorganizing terms, $x_1 = x - x_0 \frac{z}{z - f_1} = x - x_0 - x_0 \frac{f_1}{z - f_1}$. Substituting $u = x - x_0$ in the above equation and then plugging into equation 3 gives:

$$I_{x_0}(x) = \prod_{j=2}^{K} T\left(x - u\frac{f_j}{f_1}\right) S_f\left(u - (x - u)\frac{f_1}{z - f_1}\right). \qquad (4)$$

So far we have considered the case in which the first layer has only a pinhole open. For a general transmittance pattern in the first layer, we integrate equation 4 over u to get:

$$I(x) = \int \prod_{j=1}^{K} T\left(x - u\frac{f_j}{f_1}\right) S_f\left(u - (x - u)\frac{f_1}{z - f_1}\right) du. \qquad (5)$$

In the limit, when the camera dimensions are negligible with respect to z, we get equation 2:

$$I(x) = \int \prod_{j=1}^{K} T_j\left(x - u\frac{f_j}{f_1}\right) S_f(u) du. \qquad (6)$$

∎

When the aperture is a plane at distance f_1 with transmittance function T, the brightness equation 2 becomes:

$$I(x) = \int T(x - u) S_f(u) du. \qquad (7)$$

Specifically, a shifted ideal pinhole corresponds to a shift of S_f:

$$I(x) = \int \delta(x - u - d) S_f(u) du, \qquad (8)$$

where δ denotes Dirac's delta function.

Therefore, when the scene is far and the aperture is flat, then all image mappings can be formulated as a convolution of the scene S_f with the transmittance pattern of the aperture, as seen from equation 7. When the aperture has multiple layers, the set of mappings of the camera is richer, as can be seen from equation 2, and includes spatially-varying mappings. We shall now show a few examples of useful mappings, both convolutions and spatially-varying mappings. An approach that further extends the set of achievable mappings is described in Appendix B

5. Examples of New Imaging Functionalities

5.1. Controllable Single-Layer Aperture

The ability to control the aperture in space and time is arguably the most compelling feature of our camera. It allows us to change the imaging characteristics of the camera dramatically from one frame to the next. Consider the case of a *controllable pinhole camera*. In this case, the transmittance pattern corresponds to a pinhole disk. At each time instance, the system that uses the camera can instantaneously shift the pinhole to any arbitrary location on the aperture. In order to understand the effect of this on the captured image, consider Figure 5. Figure 5(a) and (b) show two different pinhole locations and the corresponding fields of view. Figure 5(c) and (d) show the corresponding images captured by our prototype. As can be seen in the figures, a shift of the pinhole
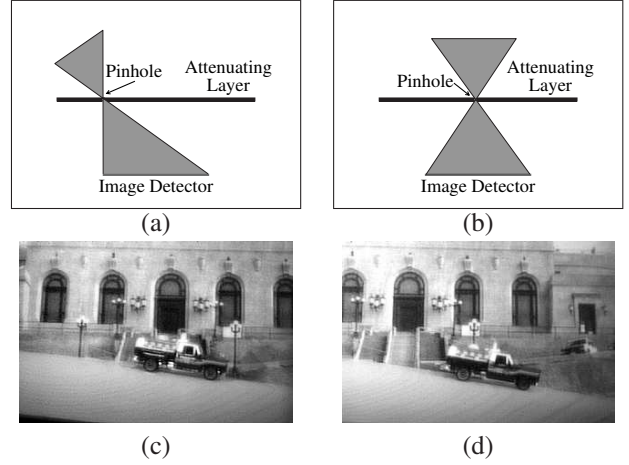

(a) (b)

(c) (d)

Figure 5. **Controllable pinhole camera. By controlling the attenuating aperture, it is possible to shift a pinhole to arbitrary locations from one frame to the next. (a) and (b) show two different pinhole locations and the corresponding fields of view. (c) and (d) show two images captured by our prototype, without physically moving the camera. This allows us to track the moving object without the use of any moving parts, unlike conventional pan-tilt cameras.**

induces a change of the viewing direction. More precisely, equation 8 shows that for a distant scene, a shift of the pinhole induces a shift of the image.

In other words, by electronically shifting the pinhole location on the aperture, the camera can shift the projected image and effectively change its viewing direction arbitrarily. This is in contrast to pan/tilt lens-based cameras that change their viewing direction continuously and are limited by motion blur and mechanical constraints.

The second property unique to our design is that each point on the detector integrates attenuated light from the entire field of view, as was shown in Figure 2. This property can be exploited to utilize the camera as a computational sensor. In other words, by selecting the appropriate transmittance pattern for the aperture, the camera can be programmed to perform a desired computation *by the optics*, so that the image detector captures the computed results. In particular, with a flat aperture the camera can be used to perform convolutions (or correlations) of the scene with pre-defined patterns (see equation 7) [4]. This can be useful in object detection tasks in which the image is typically convolved with a set of patterns (e.g. [11], [9]). A better solution would be to capture both a conventional image of the scene and one or more convolutions of the scene in parallel. This is shown in the next section.

5.2. Controllable Multi-Layered Aperture

As we have shown in section 4, a multi-layered aperture can produce spatially-varying scene-to-image mappings.

---

[4] Note that the camera performs computations on non-coherent light and that the computations are embedded within the imaging process. There is rich literature on the use of light attenuators for optical computations, but these are mostly performed with coherent light. Some examples of non-coherent optical computations are given in [10]. These works use a lens-based camera in conjunction with repetitive Fourier-like patterns for the apertures.
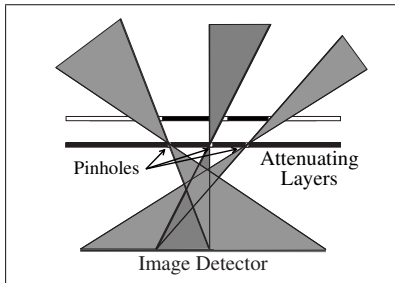
(a)

(b)

(c)

(d)

(e)

Figure 6. **Split field of view imaging. (a) Conventional cameras capture a continuous field of view. (b) In contrast, the proposed camera can split the field of view to disjoint parts and capture only these parts. This way, the camera captures objects of interest with higher resolution and avoids capturing less interesting scene parts. (c) Split field of view imaging is implemented with two or more attenuating layers. (d),(e) The aperture is dynamically adjusted to account for moving objects in the scene. In (d) and (e) the car is maintained within the field of view while the background changes.**

Here, we show two applications that exploit this feature.

Consider the scene shown in Figure 6(a). In order to capture all three subjects in the scene, a conventional video camera needs to maintain a wide field of view, that forces it to capture the three subjects with relatively low resolution. In contrast, the proposed camera allows us to split the image into sub-images and assign disjoint parts of the scene to each sub-image. Figure 6(b) shows an image captured by our prototype camera, with the aperture shown in Figure 6(c). Note that only the three subjects and their surrounding regions were captured by the camera and therefore all three subjects are captured with higher resolution. Since the camera is programmable, the system that uses the camera can determine which parts of the field of view are captured. Therefore, using an appropriate object tracking algorithm, the system can dynamically change the transmittance pattern of the aperture according to the motion in the scene, as shown in Figure 6(d) and (e). In our experiments, the aperture was adjusted manually.

As an alternative way of splitting the image, multiple optical operations can be applied to the *same* scene region, so that each sub-image captures the result of a different optical operation. Specifically, we propose to capture a part of the scene with a pinhole in one sub-image and in parallel apply convolution optically to the same scene part and capture the result in another sub-image. The application of this idea to face detection is shown in Figure 7, using normalized correlation with a face template. Figure 7(a) shows the scene captured with a lens and Figure 7(b) shows the same scene captured with our prototype camera. In this particular scene, the bottom part of the scene is less interesting. Therefore, in our implementation, we capture only the top part of the scene with a pinhole. The bottom sub-image is the result of the correlation of the top part of the scene with a face template (or a convolution with the flipped template). The transmittance pattern that was used in this example is shown in Figure 7(d) (see the face template).

Since the most computationally-intensive part was already done optically during image formation, in order to compute the normalized correlation, we only need to compute the norms of the image blocks. These can be computed efficiently using integral images [12]. Note that normalized correlation with a single template may not be sufficient for accurate and robust face detection, as evidenced by the few false positives (boxes around non-faces) in Figure 7(c), but it can be used to significantly reduce the required computations, as was done in [5]. Furthermore, a given template can only be used to detect faces at a certain distance range from the camera. Detecting faces over a larger range requires using multiple templates. Multiple templates can be used sequentially in subsequent video frames. Alternatively, one video frame can be used for convolving a scene with multiple templates. Figure 7(e) shows how multiple templates can be convolved simultaneously.

Note that Branzoi et. al. [8] also proposed to compute correlation by the camera optics for object detection. However, in their work the optics only performed pointwise multiplication of the image and the template. This permits computing a single correlation of the template rather than computing cor-

relations of the template at all offsets simultaneously. Moreover, the multiplication with different templates in parallel required imaging *a display* with copies of the face rather than imaging the scene. Branzoi et. al. [8] further proposed convolution by the optics, but restricted to convolution kernels that are smaller than the detector's pixel.

6. Summary

In this paper, we proposed a novel lensless camera that is considerably more flexible than conventional cameras, but requires a larger video detector. We have shown that our camera can make better use of the limited resolution of video detectors both in time and in space. In time, the camera can dramatically change its imaging properties from one video frame to the next, thus making it possible to collect significant amounts of visual information within a few video frames. In contrast, conventional cameras capture videos with considerable temporal redundancy. With respect to space, we have shown that the camera can better utilize the resolution of a detector. For example, it can capture objects of interest with higher resolution, while irrelevant scene parts are not captured at all. Alternatively, parts of the detector can be used for computational tasks. In contrast, conventional cameras are limited to a strict set of scene-to-image mappings. We believe that with the enormous advances currently underway in LCD technology and image detector technology, our camera can be a practical alternative to conventional cameras.

Acknowledgment

References

[1] A. Badano, M. Flynn, S. Martin, and J. Kanicki. Angular dependence of the luminance and contract in medical monochrome liquid crystal displays. *Med. Phys.*, 30(5):2602–2613, 2003.

[2] H. Farid and E. Simoncelli. Range estimation by optical differentiation. *Journal of the Optical Society of America*, 15(7):1777–1786, July 1998.

[3] E. Hecht. *Optics*. Addison Wesley, 1998.

[4] J. In 't Zand. A coded-mask imager as monitor of galactic x-ray sources, ph.d. thesis, university of utrecht, 1992.

[5] D. Keren, M. Osadchy, and C. Gotsman. Anti-faces for detection. In *European Conf. on Computer Vision*, pages I: 134–148, 2000.

[6] K. Mielenz. On the diffraction limit for lensless imaging. *Journal of Research of the National Institute of Standards and Technology*, 104(5):479–485, 1990.

[7] S. Nayar and V. Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Int. Conf. on Computer Vision*, pages 1168–1175, 2003.

[8] S. Nayar, V. Branzoi, and T. Boult. Programmable imaging using a digital micromirror array. In *Conf. on Computer Vision and Pattern Recognition*, pages I: 436–443, 2004.

[9] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *Conf. on Computer Vision and Pattern Recognition*, pages 193–199, 1997.

[10] G. Rogers. *Noncoherent Optical Processing*. Wiley, John and Sons, Incorporated, Cambridge (UK) and New York, 1977.

[11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–96, 1991.

[12] P. Viola and M. Jones. Robust real-time face detection. *Int. J. of Computer Vision*, 57(2):137–154, May 2004.

A. On the Limitations of Lensless Imaging

Existing lensless cameras, such as pinhole cameras, are inferior to lens cameras in terms of image brightness and sharpness. Our lensless camera is more general than pinhole cameras, but it suffers from similar limitations. In the following we show that these limitations are minimum in wide field of view setting, and in general can be eliminated by using a large detector.

Lensless imaging is limited in resolution. This was extensively studied in the context of pinhole imaging [6], where it was shown that there exists an optimal pinhole diameter [5] a (measured in mm):

$$a = \sqrt{3.6 \times 5.5 \times 10^{-4} \times f}, \qquad (9)$$

where f is the distance between the pinhole and the image plane. A larger pinhole produces larger blur due to the overlap of the solid angles at adjacent image points subtended by the pinhole. A smaller pinhole produces a larger blur due to diffraction [3]. A key observation with respect to the optimal pinhole equation 9 is that the optimal pinhole diameter grows with f since diffraction blur grows with f, whereas the blur due to solid angle overlap is largely constant as a function of f. In other words, *the sharpness of a pinhole camera is improved when the pinhole is placed closer to the image detector.*

The second limitation of lensless imaging systems is image brightness. The light-gathering power of the optics is expressed by the F-number, which is the ratio between the distance of the aperture and its diameter. *The F-number of a pinhole camera becomes lower (better) when the pinhole is closer to the image detector.*

We have shown that both image sharpness and the F-number (and hence image brightness) improve when the aperture is placed close to the detector. Therefore, given a detector size, a wider field of view allows us to capture sharper and brighter images. In order to obtain the same sharpness and brightness with a narrow field of view, a larger detector is required with the focal distance f scaled accordingly(note that the optimal pinhole equation 9 is non-linear in f). Therefore, the minimal detector size to ensure a desired resolution and brightness depends on the desired field of view. Table 8 shows minimal detector sizes for different field of view angles. These sizes were computed based

---

[5]Here we present the Rayleigh formula for a distant scene and light wavelength $5.5 x 10^{-4} mm$. The full formula is available in [6], equation (12).

Conventional Camera   Proposed Camera   Faces Detected
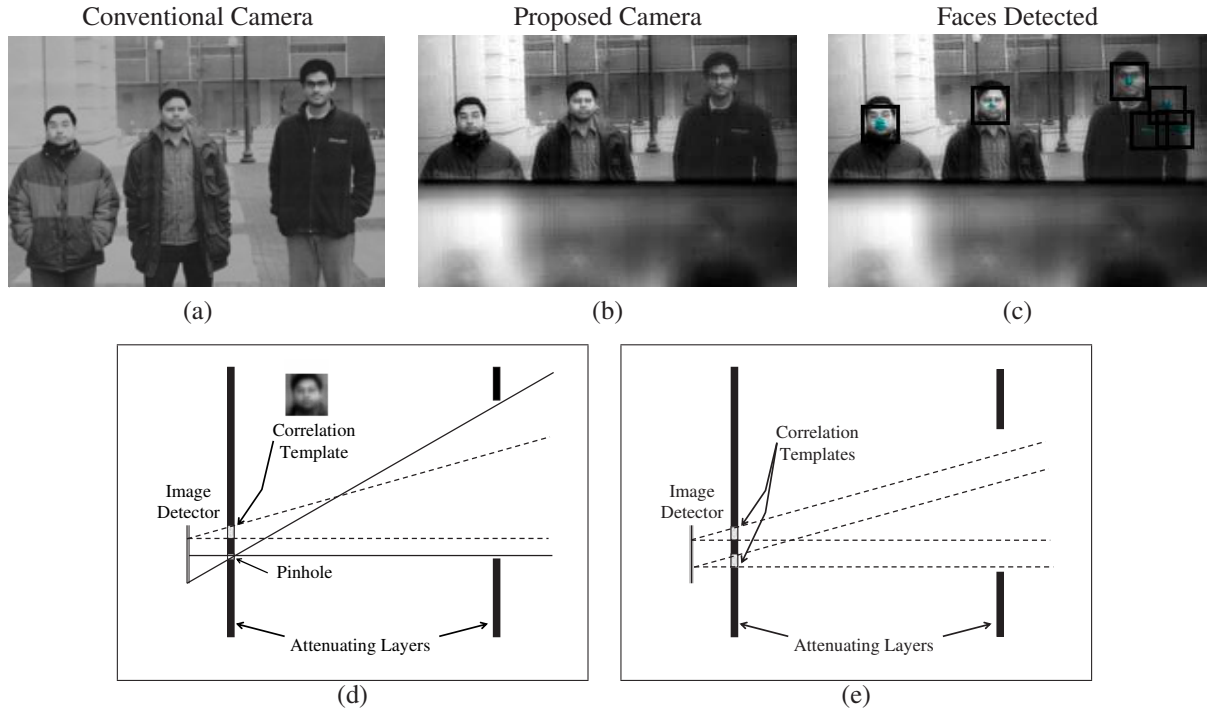
(a)   (b)   (c)

(d)   (e)

Figure 7. **Optical computations during image formation. (a) A scene captured by a conventional lens-based camera, and (b) the same scene as captured by the proposed camera. The bottom part of the captured image is the correlation of the top part of the image with the face template shown in (d). This correlation was performed by the optics. In order to detect faces, we compute the normalized correlation with the template. The optically computed correlation greatly reduces the computations. (c) The detected faces (boxes with blue centers) and a few false detections. This approach can serve as an attention mechanism for more sophisticated detectors. (d) The attenuating layers with their transmittance patterns. (e) The camera can compute convolutions of the scene with multiple templates. In this example, two convolutions are applied simultaneously to half the field of view. In the case of a two-dimensional detector, four convolutions can be applied simultaneously to a quarter of the field of view.**

Half FOV	Detector(mm)	f(mm)
15.000	148.200	276.544
20.000	109.102	149.878
25.000	85.158	91.311
30.000	68.780	59.565
35.000	56.712	40.496

Figure 8. **Detector size as a function of the field of view (FOV). Lensless imaging is limited in resolution. In order to achieve a desired resolution (here 200 pixels), a large detector should be used. The minimal detector size (and f, the distance of the pinhole) vary as a function of the field of view.**

on geometric optics[6]. In order to determine detector size, we constrained the pixel size such that adjacent pixel centers view non-overlapping angular domains, with the pinhole size determined using equation 9. So far we have addressed image sharpness. As for the light gathering power, the F-numbers of the cameras in Table 8 are large. However, taking into account the large pixel size, the cameras in Table 8 are comparable with a standard 1/3" video detector with F-number 12. One way to improve the light gathering is to use a larger detector. There is an alternative ways to increase the amount of light, as we elaborate in the following.

So far we have addressed the resolution and brightness of the image when the aperture is a pinhole. However, there exists a powerful method for substantially increasing image

---

[6]The exact point spread function is affected by diffraction. It is not well defined since it depends on the extent of light coherence

brightness in lensless imaging. This method has been widely studied in high energy astronomy [4] and is called "coded aperture imaging". The key idea is to use multiple pinholes and therefore capture brighter images. The pinholes are placed at a special arrangement that enables optimal reconstruction of a sharp pinhole image from each captured image. For this approach to be effective [4], the solid angle viewed by all image points should be bounded. This can be done in our camera by using a multi-layered aperture - one layer for the coded aperture, and another layer to limit the solid angles, as shown in Figure 9. Our current prototype did not allow us to implement this approach as the LCD we used could not be controlled to the required accuracy. We intend to supplement our approach with coded aperture imaging in the next version of the camera.

B. Extending the Set of Scene-Image Mappings

From a practical perspective, it is desirable to use a small number of attenuating layers. This, however, limits the set of scene-to-image mappings that can be implemented with our lensless camera to those defined in Proposition 1.

In the following we propose an alternative approach that extends the set of mappings of the camera. The key idea is that if a certain desired mapping cannot be achieved with the camera, then an alternative mapping is implemented and the desired mapping is obtained computationally. This approach is demonstrated using the example of imaging with a
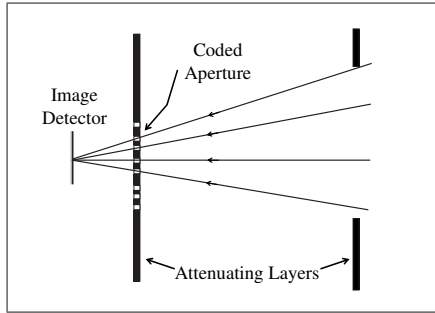
Figure 9. **Coded aperture imaging. Image brightness in lensless imaging can be significantly improved by using multiple open pinholes on the aperture [4]. The captured image can be computationally deblurred to recover a high quality image. In order to allow for a well-posed deblurring, the pinholes are arranged in a special configuration and a second layer limits the solid angles viewed by each image point.**

spatially-varying zoom. We plan to implement this functionality in the next version of our camera, which will include more attenuating layers. Here, to illustrate the feasibility of the approach, we show computer simulations.

Consider a surveillance system for detecting and analyzing moving objects of interest over a large area, as shown in Figure 10(a). In order to detect objects with a minimal delay, the system must maintain an updated wide view of the scene. On the other hand, in order to be able to analyze the objects, the system must capture them at a high resolution. Due to the limited resolution of video cameras, a wide field of view video will not have the required spatial resolution. The proposed solution is a camera with a controllable spatially-varying optical zoom factor. This way, the camera maintains a wide coarse view of the scene and at the same time can capture moving objects of interest with higher zoom, all in the same image.

Figure 10(b) shows an image with varying zoom in the x direction, and Figure 10(c) shows varying zoom in both the x and the y direction. The desired scene-to-image mapping associated with these images cannot be implemented with three attenuating layers (proof is omitted due to space limitations). In order to obtain the desired mapping with only three layers, we propose to implement an alternative mapping with the transmittance patterns shown in Figure 11, and then reconstruct the desired image computationally. The front layer in Figure 11 contains a pinhole[7] that corresponds to a long focal length and therefore large zoom factor. Note that it is assumed that the detector is large enough to allow full optical resolution with this long focal length (as explained in Appendix A). The back layer contains two pinholes, that correspond to a short focal length and therefore small zoom factor. In order to allow light through a pinhole in one layer to pass through the other layer, portions of these layers have a small non-zero transmittance ϵ (we used $\epsilon = 0.1$). Since some light passes in not through any of the pinholes, the resulting captured image is blurry, as shown in Figure 10(d).

---

[7]We show the transmittance patterns for a 1D camera. A 2D camera is also implemented with three layers. The 2D pattern of each layer is then the outer product of the 1D transmittance pattern vectors of the corresponding 1D layers.
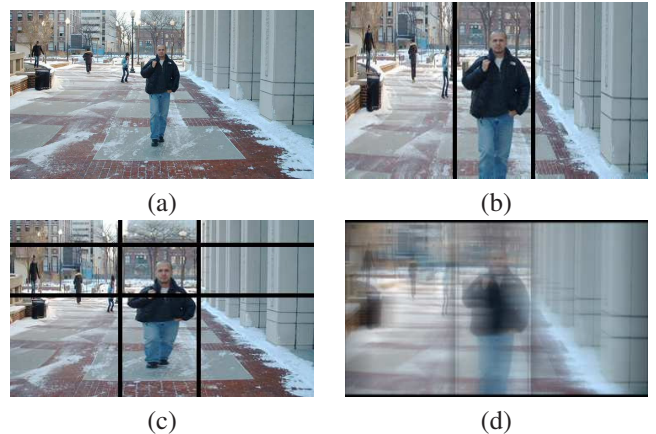

| (a) | (b) |
| (c) | (d) |

Figure 10. **Results for simulations of spatially-varying zoom. Spatially-varying zoom allows us to capture objects of interest with a high zoom while maintaining a wide view of the scene. (a) The scene as captured by a conventional camera. (b) An image subdivided horizontally into three parts, each with a different horizontal zoom factor and all with a high vertical zoom(the black lines were overlaid for visualization). Such an image cannot be captured by our camera with a small number of attenuating layers. Instead the camera can capture the image in (d) (here, (d) was created by simulation). Then, (b) is reconstructed computationally from (d). Non-uniform zoom can also be applied in the vertical direction(c).**
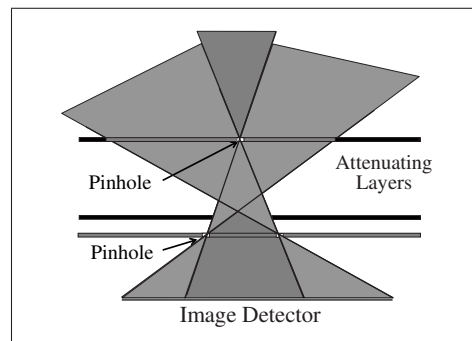

Figure 11. **Spatially-varying zoom with 3 attenuating layers.**

Then, in order to reconstruct the desired image shown in Figure 10(b) from the captured image shown in Figure 10(d) we apply a deblurring algorithm to the captured image as follows. We represent the scene by a high resolution image. The desired mapping of the scene to the image (associated with Figure 10(b)) and the camera mapping (associated with Figure 10(d)) are linear and can be represented by matrices W_x and C_x, respectively. In the case of varying zoom in both x and y, similar matrices are used for the y direction, namely, W_y and C_y. The reconstruction can be applied separately to the rows and columns of the captured image $I_{captured}$ as:

$$I_{desired} = W_y C_y^+ I_{captured} (C_x^T)^+ W_x^T \qquad (10)$$

where C_y^+ denotes the pseudo-inverse of C_y. Here, the matrix $(C_x^T)^+ W_x^T$ multiplies the image rows and the matrix $W_y C_y^+$ multiplies the image columns.

The image presented as the desired result in Figure 10(d) was actually reconstructed from the image in Figure 10(b), quantized to 8 bits.