Synthetic Aperture Confocal Imaging

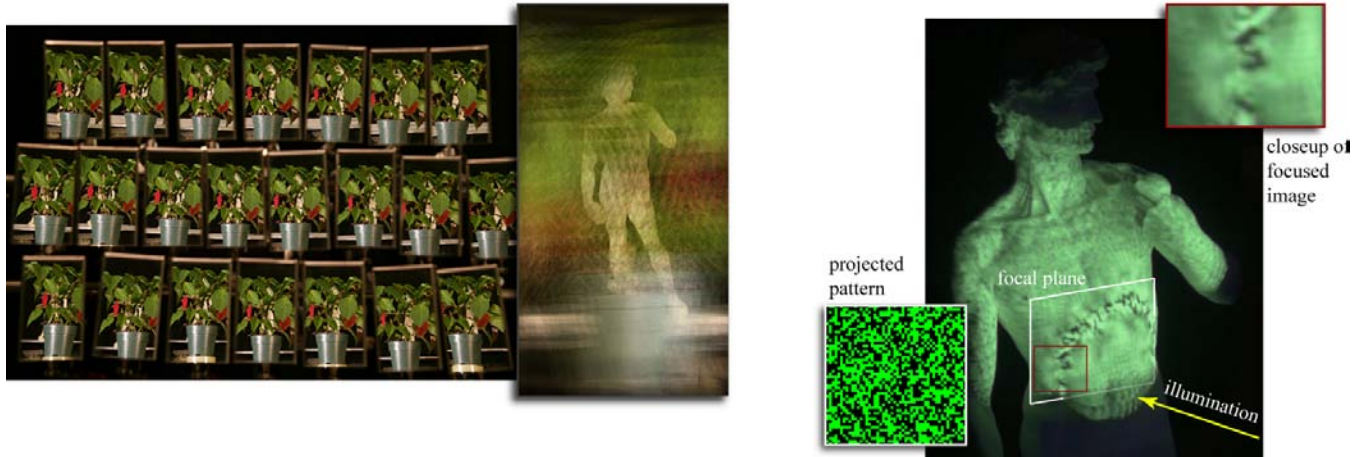Marc Levoy [1]   Billy Chen [1]   Vaibhav Vaish [1]   Mark Horowitz [2]   Ian McDowall [3]   Mark Bolas [3]

[1]Computer Science Department
Stanford University

[2]Electrical Engineering Department
Stanford University

[3]Fakespace Labs

Figure 1: The techniques in this paper employ two computer-assisted optical effects: synthetic aperture photography and synthetic aperture illumination. On the left, we aim a camera at an array of planar mirrors, yielding 22 different views of a statuette partially obscured by a plant. By rectifying, shifting, and adding these views together, we simulate a camera with a wide aperture and a shallow depth of field. Using appropriate shifts, we can position the focal plane of this synthetic camera astride the statuette, blurring out the plant. On the right we replace the camera with a video projector. By shifting, keystoning, and projecting multiple copies of a binary pattern, we produce a real image with a similarly shallow depth of field. Using appropriate shifts, we can position this image astride the statuette. On this plane the image is well focused; elsewhere, it is blurry.

Abstract

Confocal microscopy is a family of imaging techniques that employ focused patterned illumination and synchronized imaging to create cross-sectional views of 3D biological specimens. In this paper, we adapt confocal imaging to large-scale scenes by replacing the optical apertures used in microscopy with arrays of real or virtual video projectors and cameras. Our prototype implementation uses a video projector, a camera, and an array of mirrors. Using this implementation, we explore confocal imaging of partially occluded environments, such as foliage, and weakly scattering environments, such as murky water. We demonstrate the ability to selectively image any plane in a partially occluded environment, and to see further through murky water than is otherwise possible. By thresholding the confocal images, we extract mattes that can be used to selectively illuminate any plane in the scene.

CR Categories: I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture — imaging geometry, sampling

Keywords: Light fields, camera arrays, projector arrays, synthetic aperture, shaped illumination, confocal microscopy, coded aperture.

[1]{levoy,billyc,vaibhav}@cs.stanford.edu   [2]horowitz@ee.stanford.edu
[3]{bolas,ian}@well.com

1. Introduction

The use of image arrays to create a single synthetic image with a wide aperture and shallow depth of field is well known. In remote sensing, it constitutes the basis for synthetic aperture radar (SAR). In medical imaging, it underlies X-ray tomosynthesis, in which the source and detector move laterally and in opposite directions on either side of a common focal plane. For incoherent visible light, the idea of averaging multiple views in a light field to simulate a synthetic aperture was proposed in [Levoy 1996]. The application of this idea to seeing through foliage was demonstrated in [Isaksen 2000] using CG imagery, in [Coorg 1999] using real imagery captured with a moving camera, and in [Vaish 2004] using a dense camera array designed by Wilburn [2002]. We call this technique *synthetic aperture photography (SAP)*.

This idea can also be applied to illumination. Unfortunately, physical systems for generating light fields have been limited by available technology to small numbers of image-producing sources, as in autostereoscopic displays [Okoshi 1976], or to large numbers of point sources [Malzbender 2001, Debevec 2002] or area sources [Han 2003, Masselus 2003, Schechner 2003] for measuring object reflectance. However, the size and cost of projectors is dropping. A dense array of projectors allows us to simulate a projector with a wide aperture. Such a system produces a real image with a depth of field so shallow that it ceases to exist a short distance from the focal plane. Figure 1 demonstrates this technique, which we call *synthetic aperture illumination (SAI)*.

(a) a confocal image   (b) confocal scanning microscope   (c) aperture correlation microscope
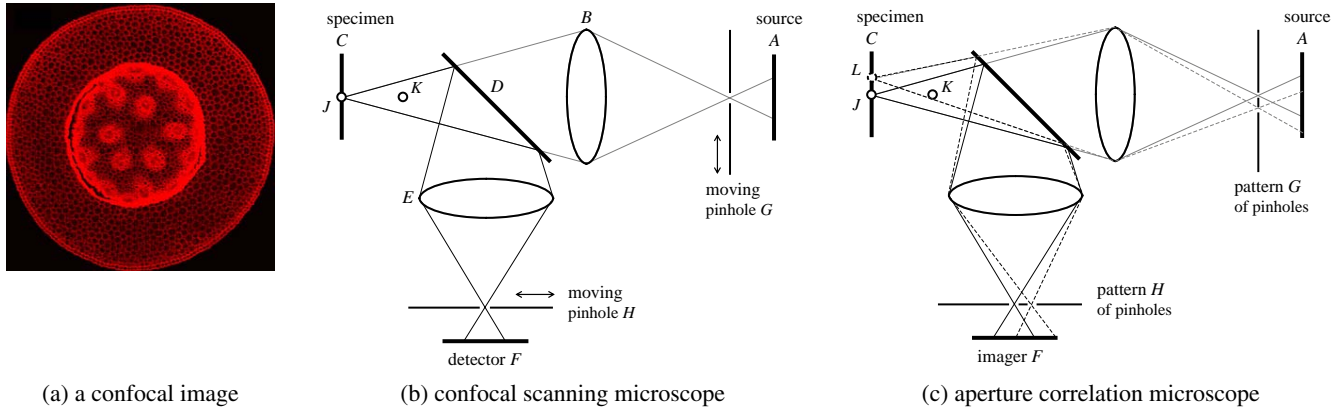
Figure 2: The principle of confocal microscopy. (a) Confocal laser scanning micrograph of fluorescently stained Convallaria rhizom (UMIC SUNY/Stonybrook). (b) A reflection mode confocal scanning microscope. An illumination source at A is imaged by an optical system B onto a 3D specimen that sits astride focal plane C. The specimen is imaged through a beamsplitter D and a second optical system E onto a detector F. A pinhole at G focuses the source on point J, which therefore receives light through the full aperture of the illumination system (the lens). However, the illumination received by point K off the focal plane falls off as the square of the distance from this plane, making it dimmer. A second pinhole at H masks out everything but that portion of the image that is focused on J - hence the term confocal. Assuming the specimen scatters light diffusely, and single scattering dominates over multiple scattering, then the amount of light gathered from K will be lower than from J, making it even dimmer. By moving the pinholes in tandem, the specimen can be scanned, (c) A reflection mode aperture correlation microscope. The single pinholes have been replaced by matched patterns of pinholes at G and H, and the detector has been replaced by an imager at F. This system requires no scanning. Instead, a sequence of trials is performed. On each trial, a randomly chosen 1/2 of the points on the focal plane are illuminated. The light falling on K that is attributable to the light focused on J will be lower than the light falling on J, as before. K is also illuminated by the light focused on nearby point L, but only 1/2 of such points on the focal plane are illuminated at once, so K is still dimmer than J.

In this paper, we employ synthetic aperture photography and illumination to implement discrete adaptations of two techniques from confocal microscopy. In section 2, we briefly review confocal microscopy. In section 3, we describe our adaptation of it, and in section 4 we describe a macroscopic implementation using a projector, a camera, and an array of mirrors. This implementation permits us to selectively image any plane in a partially occluded or weakly scattering volume measuring 10cm on a side. In section 5, we image a toy soldier hiding behind a plant, and we read an AT&T calling card through murky water. By thresholding these images and loading them back into the projectors, we can selectively illuminate any plane in the volume. This lets us spotlight the soldier without lighting up the plant, or vice versa.

2. Confocal imaging with optical apertures

In a conventional microscope, portions of the specimen not lying on the focal plane are blurry, but they still contribute to the image, reducing its contrast and impeding its interpretation. Confocal microscopy, invented by Marvin Minsky in 1955, employs the optical principle described in figure 2(b) (adapted from [Corle 1996]) to reduce the amount of light falling on, and recorded from, points off the focal plane. As a result these points become both blurry and dark, effectively disappearing. This yields a cross-sectional image of the specimen where it intersects the focal plane. By moving the specimen through the focal plane and stacking the resulting images, 3D image arrays can be created. These can be displayed using volume rendering techniques.

The major disadvantage of confocal scanning microscopy is that acquisition is slow, since the specimen must be illuminated and imaged one point at a time. To address this limitation, researchers have proposed a variant called aperture correlation microscopy 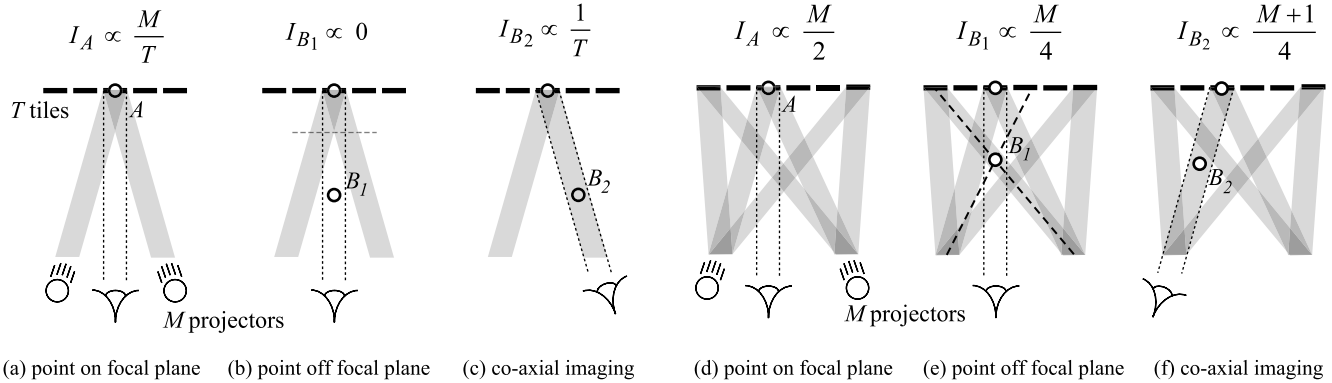[Wilson 1996], in which the specimen is illuminated over a sequence of trials as explained in figure 2(c). By performing and summing a sequence of such trials, one produces an image that is the sum of a confocal and a fully illuminated image. Subtracting a separately captured, fully illuminated image yields a confocal image. If too few trials are acquired, this image is noisy, but it converges in the limit to a correct result.

Since the number of trials Wilson acquires is typically less than the number of points used in scanning, his technique is faster. It has a second advantage; since more than one point on the focal plane is illuminated on every trial, the light efficiency of his approach is higher. Expressed as a fraction of the possible light that could be delivered to the specimen on each trial, scanning microscopy has an efficiency of P/S, where S is the area of the light source and P is the area of pinhole G in figure 2(b). By comparison Wilson's technique has an efficiency approaching 1/2. In later sections, we call this fraction the *fill factor*.

3. Confocal imaging with synthetic apertures

In this paper, we propose replacing the optical apertures in figure 2 with synthetic apertures formed by arrays of (real or virtual) projectors and cameras. In so doing, we obtain discrete approximations of the two confocal imaging techniques described in the previous section. Our approximations differ in three ways from these techniques:

(1) **Discretely sampled aperture.** By replacing one large aperture with a number of smaller apertures, we reduce the light-gathering ability of our system. However, it enables us to operate at large scales, where it would be prohibitively expensive to build a lens. The sampling issues associated with this replacement have been well studied [Chai 2000].

$$I_A \propto \frac{M}{T} \qquad I_{B_1} \propto 0 \qquad I_{B_2} \propto \frac{1}{T} \qquad I_A \propto \frac{M}{2} \qquad I_{B_1} \propto \frac{M}{4} \qquad I_{B_2} \propto \frac{M+1}{4}$$

(a) point on focal plane   (b) point off focal plane   (c) co-axial imaging   (d) point on focal plane   (e) point off focal plane   (f) co-axial imaging

Figure 3: Illumination counting in different cases of confocal imaging using a synthetic aperture. At the bottom of each diagram are M projectors and one or more cameras, aligned so that they independently address each of a set of T finite-size tiles (at top) spanning a common focal plane. Illumination beams are denoted with gray polygons, and the camera's view of a tile is denoted with a pair of parallel dotted lines. These diagrams are not drawn to scale.

(2) **Finite-size tiles.** Due to practical considerations, the smallest "pinhole" we can illuminate and mask is a tile several projector pixels across. In confocal microscopy, larger pinholes create a volumetric zone inside which all points are illuminated and imaged brightly, lead to a larger depth of field and lower axial resolution. In our context, large tiles limit our ability to discriminate between objects lying near versus on the focal plane.

(3) **Intra-tile imaging.** Since our tiles are typically larger than one camera pixel, we can capture each tile as an image, rather than merely recording its average intensity. Why is this useful? In microscopy, specimens are assumed to be of uniform opacity, so lateral spatial resolution is proportional to pinhole size. In our applications, opaque objects are typically embedded in a transparent (or less opaque) medium. In this regime, we can preserve more information if we record an image within each tile.

The first two differences are incidental and are covered by existing theory. The last difference is fundamental and requires us to develop a new explanation for the behavior of our algorithms.

Algorithm #1: scanned aperture confocal imaging

Let us first treat the task of adapting confocal scanning microscopy to the discrete setting. Referring to figures 3(a-c), we perform a scanning sequence of N trials over T tiles. On each trial, we illuminate one tile using all the projectors at once, so $N = T$. We then capture an image, extract the pixels corresponding to that tile, and insert these into the output image. By discarding pixels outside the tile, we effectively focus our image where the light is focused, making the system confocal.

Let us compare the illumination falling on a point A on the focal plane in figure 3(a) with that falling on a point off the focal plane but along a line connecting A and the camera. Since A receives illumination from M projectors, but only on 1 of T trials, its intensity averaged over the duration of the scanning sequence is proportional to M/T. For points off the focal plane, let us first consider the case when the camera is not coincident with any projector, as shown in figure 3(b). Assuming that B_1 lies out of the hot spot generated by the beams converging on A, i.e. below the gray dashed line in the figure, then it will receive no illumination.

If B_1 occludes A, then the corresponding pixel remains dark. Let us now consider when the camera is coincident with one of the projectors, i.e. coaxially imaged, as shown in figure 3(c). This is an important special case, whose utility we discuss later. Here B_2 will receive illumination from one projector on one trial, leading to a time-averaged intensity of $1/T$.

Summarizing, using this algorithm points on the focal plane will have intensity M/T, and points off the focal plane will have intensity 0 or $1/T$. The ratio between them will be at least $1/M$. This constitutes the contrast for this imaging algorithm. As the number of projectors increases, this contrast tends toward infinity. Of course, it is limited in practice by the black level of the projectors, the dynamic range of the camera, and other factors.

Algorithm #2: coded aperture confocal imaging [1]

Referring to figures 3(d-f), we perform a sequence of N trials, where N is typically much smaller than T in algorithm #1. On each trial, we pseudo-randomly illuminate 1/2 of the tiles, a different set on each trial. We discuss suitable illumination patterns in section 4.1. If a tile is chosen to be illuminated on a given trial, then it is illuminated using all the projectors at once. On each trial, we capture an image, extract the pixels known to be illuminated on that trial, and add these pixels to the output image.

Let us again compare the illumination falling on points on and off the focal plane. Since A receives illumination from M projectors on 1/2 of the trials, its time-averaged intensity is proportional to $M/2$. For points off the focal plane, if the camera is not coincident with any projector (figure 3(e)), then by tracing rays from the projectors through B_1 to the focal plane (diagonal dashed lines on the figure), we can identify those M tiles that affect it. By construction, approximately 1/2 of these will be illuminated on any given trial. Thus, if B_1 occludes A, then the time-averaged intensity of point B_1 will be proportional to $M/2$. However, since B_1 falls in the tile containing A, and we only extract this tile on 1/2 of

---

[1] We depart from Wilson's terminology to reflect the greater flexibility projectors gives us over our patterns. Our approach is similar to coded-mask imaging, one form of coded aperture imaging used in astronomy [Zand 1996]; however, unlike those methods no reconstruction step is required.
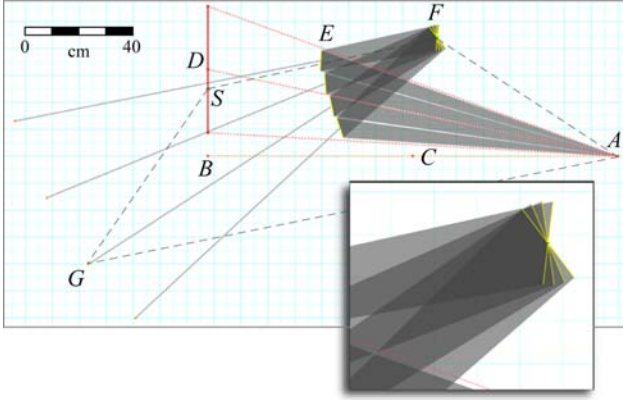
Figure 4: A visualization of our optical layout. A projector at *A* is focused at distance *B* onto a plane perpendicular to line *C*. It has an off-axis perspective, placing its central pixel at *D*. A set of mirrors at *E* partition the projector's field of view into subimages, which reconverge at *F*. The placement of each mirror is such that these subimages are individually well focused when they reach *F*. The reflection of the real projector in each mirror forms a set of virtual projectors *G*. The locus of these points is called the ortho-tomic; it is our synthetic aperture. It can be constructed by plotting the locus of 4th vertices of a family of isosceles trapezoids, the three other vertices of which are points *A*, *F*, and a variable point *S* on the projector's focal plane. One such trapezoid is shown in dashed gray lines. In the closeup at lower-right, note that the subimages (intersecting yellow line segments) vary in orientation; the resulting system does not have a single plane of best focus.

Figure 5: Our optical bench, set up to record a scene similar to figure 9. When performing scattering experiments, this scene is replaced by a water tank. An image loaded into the projector at *A* is reflected by an adjustable 4 x 4 array of planar mirrors at *B*, reconverging on the scene at *C*. The returning image is diverted by a pellicle-type beamsplitter at *D* to a camera at *E*. Stray light lands in a light trap at *F*. Our projector was a Compaq MP1800 (1024 x 768 pixels) with an 18-degree field of view. The camera was a Canon 10D (3072 x 2048 pixels) with about the same field of view. For scattering experiments, the camera was operated in 16-bit RAW mode to preserve low-order bits, and exposures were kept below 1 second to minimize noise.

the trials, then the intensity we record for the pixel that sees B_1 will be proportional to $M/4$. When the camera is coincident with one of the projectors (figure 3(f)), then point B_2 will always receive illumination destined for one tile, plus 1/2 of the remaining tiles. This makes the time-averaged intensity for the pixel that sees B_2 proportional to $(M+1)/4$. As the number of projectors increases, this ratio tends toward $M/4$.
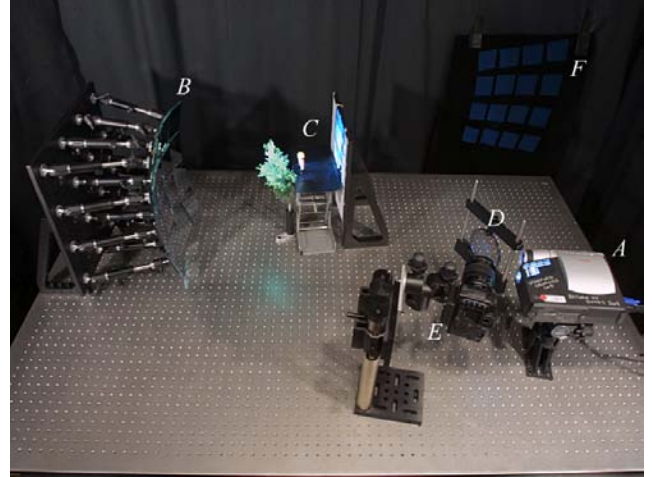
Comparing these results, we see that objects at the focal plane will be brighter by a factor of about 2 than objects off the focal plane. As in Wilson's method, this is not a confocal image; it is a confocal image plus a fully illuminated (floodlit) image. To remove the floodlit contribution, we capture one additional trial in which all tiles are illuminated. On this trial, our distinguished pixel will have an intensity proportional to M regardless of whether it sees an object on or off the focal plane. We now remove the floodlit contribution by computing

$$I_{confocal} \;=\; I_{trials} - \frac{1}{4} \, I_{floodlit} \qquad (1)$$

In our discrete setting, this equation applies only in the non-coaxial case. For the coaxial case, a similar equation can be derived:

$$I_{confocal} \;=\; \frac{M+1}{M-1} \left(\frac{M}{M+1} \, I_{trials} - \frac{1}{4} \, I_{floodlit} \right) \qquad (2)$$

Equations (1) and (2) become the same as the number of projectors M tends to infinity. It can be easily checked that, given the intensities indicated in figure 3, these equations produce images in which points on the focal plane have intensity $M/2$, and points off the focal plane have intensity 0. Compared to algorithm #1, points on the focal plane in this algorithm are $T/2$ times brighter. This represents the fundamental advantage of coded aperture over scanned aperture confocal imaging.
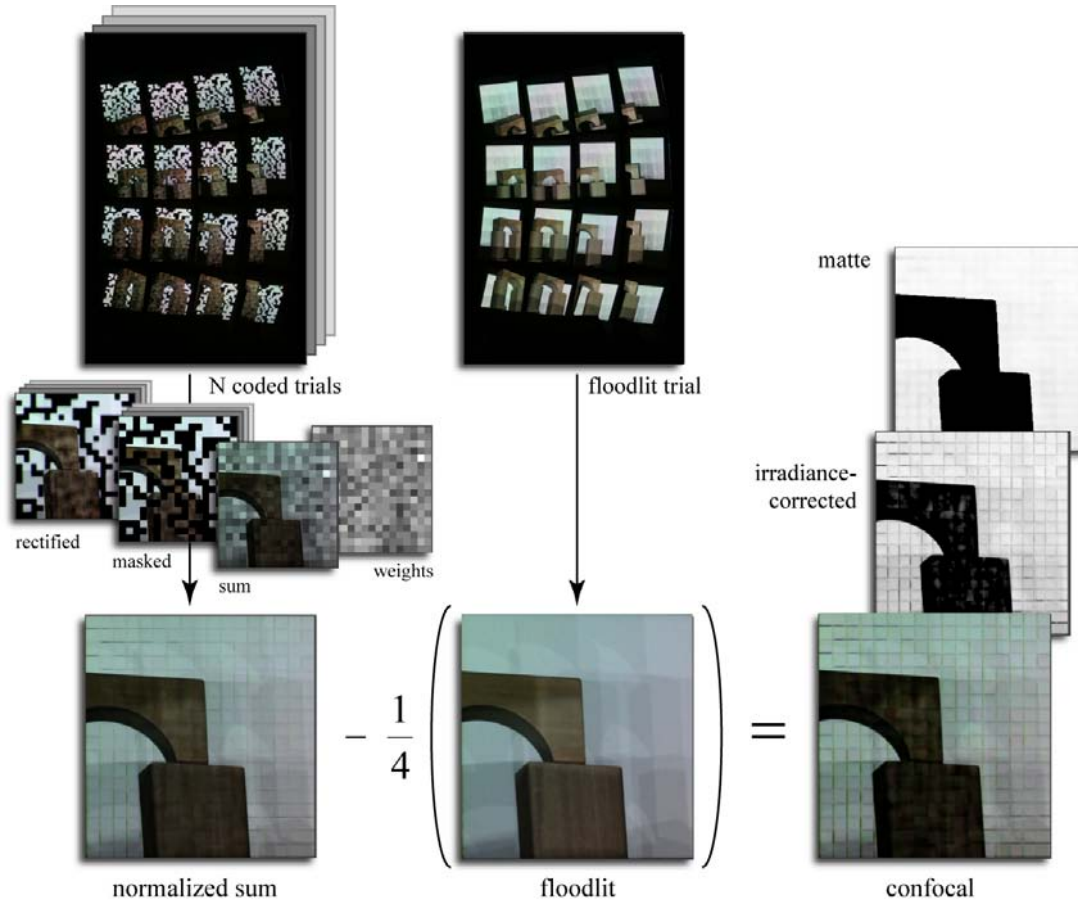
So far we have ignored vignetting, which occurs if the object leaves the field of view of one or more projectors. Fortunately, the confocal effects described in these algorithms require only that $M \geq 2$. In a 2D array of projectors, the number of projectors will fall below 2 only in the corners of the working volume. We have also ignored irradiance - which falls off as the square of the distance to the projector as well as depending on surface orientation, vignetting (described above), and shadows or interreflections. Finally, we have ignored albedo, which changes the intensity returned for a given irradiance. However, these effects apply equally to the N coded trials and the single floodlit trial. It can be shown that although the confocal image will exhibit these effects, in so far as objects on the focal plane may be darker than expected, the ratio of intensities returned for points on and off the focal plane will remain as derived above.

4. Implementation

To experimentally verify the algorithms we have proposed, we built an implementation using a single projector, a single camera, and an array of planar mirrors. Figure 4 abstractly depicts our optical layout, and figure 5 shows the components positioned on an optical bench. Preparing our system for an experiment consists of aiming and focusing the optical components, adjusting their locations to establish coaxial imaging of the projector and camera, and performing geometric and radiometric calibration.

The goal of geometric calibration is to align the virtual projectors and cameras to a common reference coordinate system. To accomplish this, we place a diffuse screen at the vergence point, display a target of squares on each virtual projector, capture its

matte

irradiance-
corrected

$$- \frac{1}{4} \left(\quad \right) =$$

normalized sum

floodlit

confocal

N coded trials

floodlit trial

rectified masked sum weights

Figure 6: *Our implementation of algorithm #2 - coded aperture confocal imaging. The scene was a stack of wooden blocks in front of a diffuse white screen, which sits at the synthetic focal plane. The illumination pattern was a pseudo-random tiling (see section 4.1). If we replace this pattern with a lexicographic enumeration of tiles, and we omit the floodlit trial, then the diagram also describes our implementation of algorithm #1.*

image using a camera, and use standard vision techniques to detect features and compute homographies between the virtual projectors and an arbitrarily chosen reference coordinate system. Illumination patterns are generated in this coordinate system, then warped to each virtual projector using these homographies. To geometrically calibrate the camera, we display the same target on one virtual projector, image it using all virtual cameras, and again use standard techniques to compute homographies between the virtual cameras and the reference coordinate system. The goal of radiometric calibration is to ensure that all images are in a linear luminance space; otherwise, equations (1) and (2) will not work. We do this by imaging any scene using a sequence of exposures one f/stop apart, then fitting a curve to the resulting sequence of values for each pixel. To ensure stability of this calibration, we disable auto-exposure and auto-white balance.

Starting from this calibrated arrangement, figure 6 depicts our implementation of the two algorithms described in section 3. With a scene placed at the vergence point, a sequence of N trials is performed. On each trial, we generate a pattern as described in section 4.1. Using the homographies computed during calibration, we coalesce 16 copies of this pattern to form a single 1024 x 768 pixel image, which we display on the projector. We then record the scene using our camera. One such camera image is shown in the top-left corner of the figure. We capture N such

images, one per trial, then crop out and warp the subimage representing each mirror to place it in the reference coordinate system. This produces $16N$ rectified images, one of which is shown below the camera image. These images are small, typically a few hundred pixels on a side. Examining the rectified image, we see that the pattern produced by the 16 virtual projectors is focused and clearly visible behind the blocks. The illumination falling on points off the focal plane, i.e. on the blocks themselves, contain contributions from many parts of the illumination pattern.

We know from our pattern which points on the focal plane were illuminated on each trial, so we extract only those pixels from the rectified image, masking out the others. This produces the second image in this row in the figure. Note that some tiles within the wooden blocks are black in this image. Summing these masked images over N trials produces the third image in that row. Since the probability that any point on the focal plane is illuminated is 1/2, then over a sequence of N trials, all portions of the image that see the focal plane should converge to a homogeneous color. However, if the number of trials is small ($N = 16$ in this example), variation may remain. For tiled patterns, this variation exhibits itself as color differences between adjacent tiles. This problem is discussed in section 4.1. Lacking an ideal solution, we can improve the quality of our results by normalizing the pixels in each tile by the number of trials on which the corresponding focal

plane point was illuminated. This produces the normalized sum image at lower-left. As required by equation (1), we now acquire one additional trial under floodlit illumination, crop out and warp the subimage for each mirror, and subtract 1/4 of this subimage from the normalized sum, producing the confocal image at lower-right. (For coaxial imaging, we would instead apply equation (2).)

Three aspects of this image are worth noting. First, for a finite number of trials, the confocal image may contain excursions below zero. We clamp these to zero. Second, weak lines can be seen along the boundaries of each tile. These are due to imperfect masking, which is in turn due to imperfect alignment of our projectors. For the results reported in section 5, we surround each tile with a margin of fixed width to ensure that the pixels we extract are fully illuminated. This margin raises our fill factor above 50%. Equations (1) and (2) can be adjusted to compensate for this, and doing so actually increases our light efficiency. However, as the fill factor rises, we find ourselves subtracting two images of similar magnitude, leading to noisy results. A better solution is to reduce the number of tiles placed so that after their margins have been added, the fill factor remains 50%. Since we remove these margins before extracting the pixels for each trial, the number of extractable pixels drops. We call this new fraction the *duty factor*. Its value depends on the quality of our alignment; 40% is common. As the duty factor drops, we need more trials to control variability.

Third, although the wooden blocks have become dark relative to the floodlit image, they have not become completely black. Although we normalize to remove variability in the number of trials with which objects on the focal plane are illuminated, we cannot normalize pixels that see objects off the focal plane, since their illumination depends on their (unknown) depth. These unnormalized variations lead to the mottled appearance of the wooden blocks. This mottling can be reduced two ways: by increasing the number of trials or by increasing the number of projectors. The former is easy, so in later experiments we use 32 or more trials.
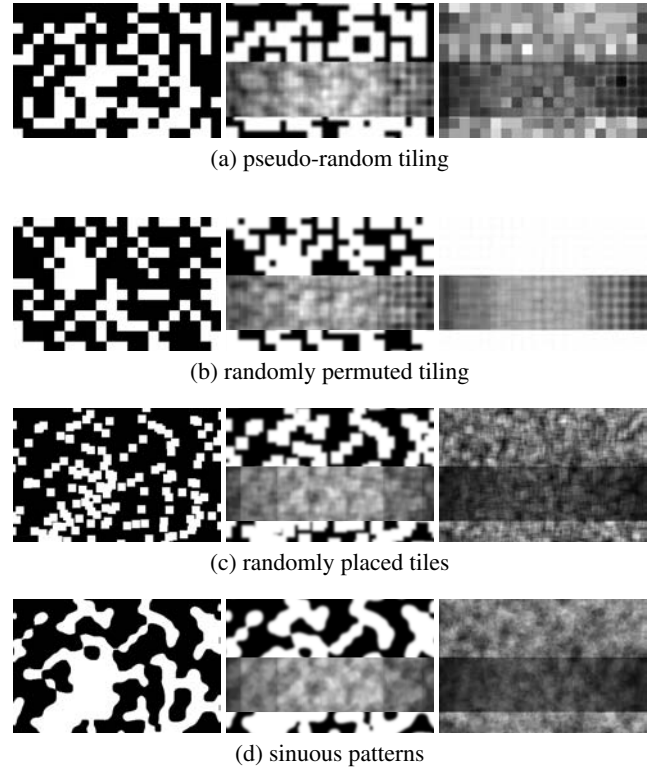
Once we have a confocal image, we can convert it to a matte that isolates objects on the focal plane from other objects in the scene. To create this matte, we divide the confocal image by the floodlit image, thereby eliminating variations in irradiance and albedo (if not too dark), then apply thresholding or contrast-stretching. The latter produces mattes with grayscale silhouettes. Compared to other matte extraction techniques, our technique is active rather than passive as in [Chuang 2001], so its performance is largely independent of scene content, and it uses frontal illumination instead of changing the backing color as in [Smith 1996].

4.1. What are good patterns to use?

For the algorithm just described, we seek a sequence of illumination patterns satisfying the following properties:

(1) Each tile is illuminated in the same number of trials. This avoids variability between points lying on the focal plane.

(2) Any two tiles should be illuminated independently. This avoids variability between points off the focal plane.

(3) Each trial should have the same number of tiles on. This ensures an adequate fill factor (as defined in section 2).

Formally, for N trials and T tiles, an illumination pattern can be represented by an $N \times T$ matrix of 0's and 1's, with 1's corresponding to illuminated tiles. To satisfy properties (1) and (3), we

(a) pseudo-random tiling

(b) randomly permuted tiling

(c) randomly placed tiles

(d) sinuous patterns

Figure 7: Properties of different coding patterns, visualized for a synthetic scene composed of a horizontal strip in front of the focal plane. In each row of the figure, the left column shows the generated pattern for one trial before margins are added, the middle column shows a simulated camera view of the scene, including projector and camera blur and misalignment, and the right column shows the sum of 32 such trials before normalization. In (a)'s sum, notice that tiles in the focal plane exhibit significant variations in brightness; this variability is gone in (b), which now appears white. However, in (b)'s sum the foreground strip contains aliases of the pattern; this aliasing is broken up in (c) and more so in (d).

seek a matrix in which all rows and columns have the same fraction of 1's. To satisfy (2), we seek a matrix whose autocorrelation function is zero except at the origin. Patterns based on Hadamard matrices, often used in spectroscopy [Harwit 1979], satisfy all the above properties. However, for these patterns we need $N = T$, which for us would imply too many trials [Schechner 2003]. Here are some patterns we have successfully tried:

Pseudo-random tiling. By flipping a coin for each tile and trial, we obtain patterns satisfying the above properties as the number of trials approaches infinity. However, for a practical number of trials, this strategy yields poor patterns. In an experiment with 16 trials and 1000 tiles, the binomial distribution tells us we can expect 10 tiles to be illuminated in fewer than 4 trials.

Randomly permuted tiling. For any fill factor expressible as m/n for integers m and n where n is a common factor of N and T, we can partition the matrix into an $n \times n$ grid of blocks of size $N/n \times T/n$, and set all entries in m blocks of each row and column of the grid to 1. This satisfies properties (1) and (3). To approximately satisfy (2), we repeatedly permute random sets of four matrix entries by searching for 1's at indices (i_1, j_1) and (i_2, j_2) such that (i_1, j_2) and (i_2, j_1) are 0 and inverting these four entries.

Randomly placed tiles. Another strategy is to dispense with a regular tiling of the plane. Tiles may be placed anywhere and may even overlap. We add tiles in this way until we reach the desired fill factor. This strategy is no more likely to satisfy properties (1) and (2) than pseudo-random tiling. However, by randomizing the location of tile edges, we break up visually objectionable aliasing.

Sinuous patterns. An extension of the previous strategy is to randomize the orientation and shape of tiles. A further extension is to randomly place tiles, then blur the image with a large filter and threshold the result. This generates patterns with sinuous edges. The threshold is chosen to ensure a fill factor of 50%.

Figure 7 shows these patterns and discusses some of their properties. Figure 6 was generated using pseudo-random tiling, 8 and 9 using sinuous patterns, and 11(d) using randomly permuted tiling. Sinuous patterns usually look best.

5. Results

In this section we demonstrate the use of synthetic confocal imaging on two kinds of scenes: partially occluded environments and weakly scattering environments.

5.1. Partially occluded environments

As noted earlier, our algorithms require only a wide illumination aperture, not a wide imaging aperture. Therefore, let us begin by demonstrating confocal imaging using multiple virtual projectors and a single virtual camera. Figure 8 demonstrates this case for a scene consisting of a plant positioned in front of a diffuse white screen. The focal plane coincides with the screen. This is a relatively easy case, and the resulting confocal image (b) and derived mattes (c-d) look good.

If we replace the single virtual camera with an array of virtual cameras, we can combine confocal imaging with synthetic aperture photography. This allows us to make a partially occluding foreground object disappear, revealing the object hidden behind it. Figure 9 demonstrates this idea, using a toy soldier positioned behind the plant.

Since our projectors and cameras are coaxial, the mattes computed in figure 8 can be loaded back into the projectors, allowing us to selectively illuminate the plant or the soldier, as shown in figure 10. It is possible to produce a matte for figure 10(b) using only the plant and the soldier - without the diffuse screen - by thresholding or contrast-stretching figure 9(c). Loading this matte into the projectors would produce a visual effect similar to the one shown here, but since confocal imaging has a shallow depth of field, only the soldier's chest would be illuminated.

5.2. Weakly scattering environments

Scattering in a participating medium is a well-studied problem. Its equilibrium solution is an integro-differential equation relating the change in radiance per unit distance in the medium to the physical mechanisms of emission, attenuation and scattering. The impact of these mechanisms on visibility through the medium is loss of contrast and blurring. For weakly scattering media such as atmospheric aerosols [Middleton 1952] and non-turbid ocean waters [Mobley 1994], loss of contrast dominates. This suggests that we can enhance visibility in these media by capturing images digitally and stretching their contrast, subject to the limits imposed by imaging noise.

In shallow waters, sunlight contributes greatly to scattering. Fortunately, this effect can be reduced using polarization [Schechner 2004]. In deep waters where the scene must be artificially illuminated, backscatter from the floodlights to the camera creates strong backscattering near the camera, sharply limiting visibility. To reduce this effect, oceanic engineers typically place their floodlights well to the side of the camera [Jaffe 1990]. Alternatively, one can restrict illumination to a scanned sheet whose intersection with the target is recorded by a synchronously scanning camera [Jaffe 2001]. We take this idea further, distributing the illumination across a wide aperture and restricting its intersection with the target to a single beam or set of beams.

Our results for this investigation are summarized in figure 11. Each row of the figure demonstrates an improvement over the row above it. For confocal imaging (c-e), since our algorithms record intra-tile images, they are well suited to the task of examining opaque objects submersed in a scattering medium, such as the AT&T calling card used here. In (e) we combine confocal imaging with synthetic aperture photography and use it to make a block of coral disappear. In (d-e) our ability to perform coaxial imaging permits us to derive and use mattes. We have tried this, but the improvement is modest at these concentrations. Using a matte would also reduce backscatter from the coral as viewed by the naked eye. However, the water in these experiments was too murky to read the text without synthetic imaging.

Although the results in this figure look good, they degrade as we add more milk. At double the concentrations listed in the figure, multiple scattering begins to dominate, and the confocal effect disappears. Also, if we use a larger tank and place our target farther away, attenuation plays a larger role, reducing the relative strength of the reflection from our target.

5.3. Cross-sectional imaging

In our final experiment, we demonstrate how confocal imaging using a synthetic aperture can be used to generate cross-sectional images (up to occlusion) of opaque objects. In section 4, we used one position of a target to determine homographies for the virtual projectors and cameras. If we instead perform a plane + parallax calibration [Vaish 2004], then by shifting the illumination patterns in a manner determined by this calibration, we can translate the synthetic focal plane forward and backward.

Figure 12 demonstrates this idea. By thresholding these confocal images to produce mattes for each depth and stacking the mattes together, we could create a 3D volumetric model. We have not done this, since the depth of field in our current system is too large to produce a good model.

In this experiment we used structured illumination to determine object shape. It is therefore natural to compare our approach with triangulation rangefinding (for example [Rusinkiewicz 2002]). In the latter case, if an occlusion blocks either the line of sight from the projector to the object or from the object to the camera, then the range image will contain holes. In confocal imaging, if any part of the aperture remains unoccluded as seen from a point on the object, an image will be formed of that point. This makes confocal imaging more robust, at the cost of requiring more projectors and cameras. One could add more cameras to a triangulation rangefinder to improve its robustness, and one could add more projectors to a stripe-based system if they lie in a line parallel to the stripes, but we know of no way to use a 2D array of projectors operating simultaneously at the same wavelength.

(a) floodlit scene   (b) confocal image   (c) holdout matte   (d) inverse matte
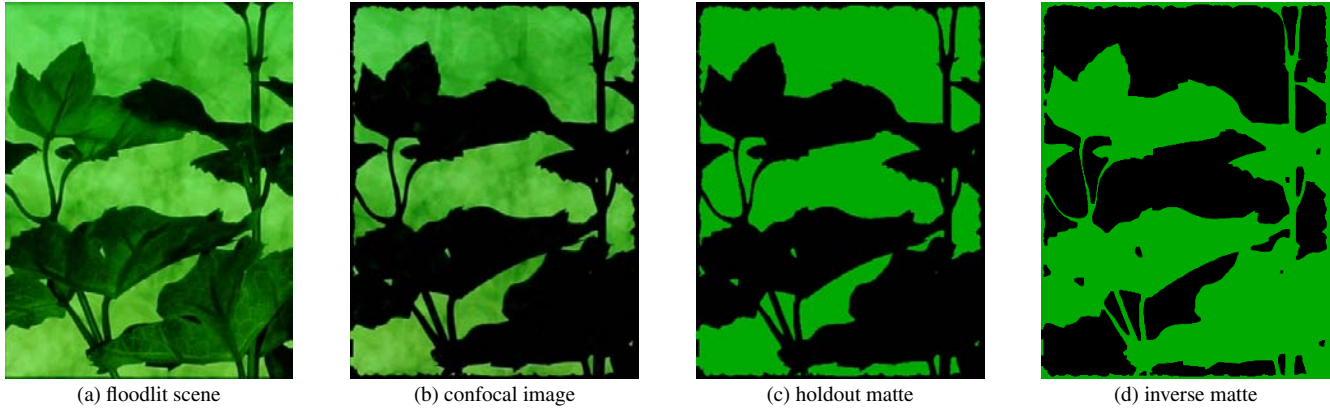
Figure 8: Synthetic aperture confocal imaging. (a) shows a plant in front of a diffuse white screen. Coded aperture imaging was used, with 32 trials of sinuous patterns. However, to reduce chromatic aberration only the green channel was used. Since the focal plane coincided with the screen, only it remains bright in the confocal image; the plant is nearly black. Dividing (b) by (a) to eliminate shadows, then contrast-stretching the result, produces a holdout matte (c) and its inverse (d). These mattes are used in figures 10(b) and (c).

(a) single viewpoint   (b) synthetic aperture photograph   (c) confocal image   (d) combining (b) and (c)
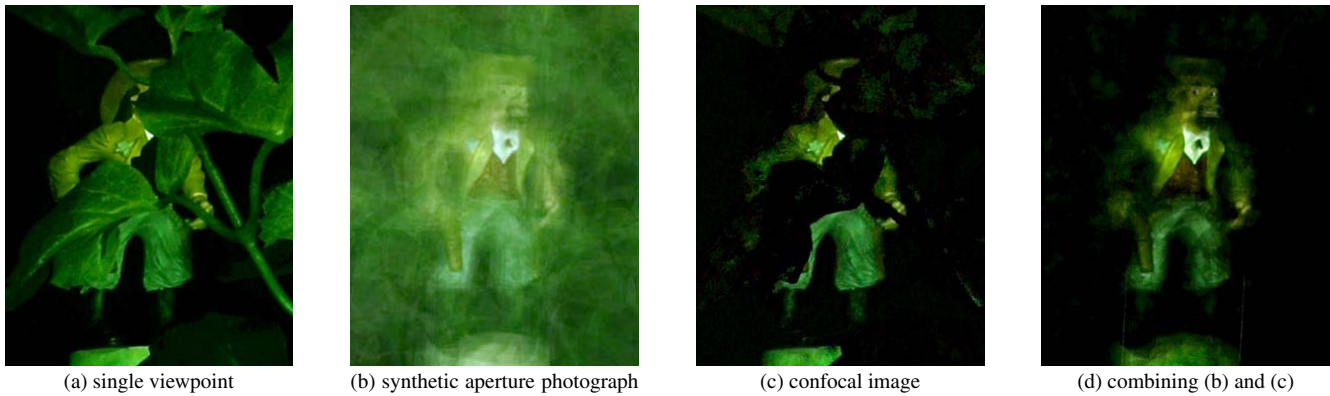
Figure 9: Combining synthetic aperture photography and confocal imaging. (a) shows a view similar to 8(a), but with the diffuse screen replaced by a toy soldier. Adding together views from all 16 virtual cameras produces a synthetic view (b) with an extremely shallow depth of field; the soldier's chest - which lies astride the focal plane - is sharp, but his arms and the plant are blurry. Performing 32 trials of sinuous-pattern confocal imaging produces (c), in which only surfaces near the focal plane are bright, leaving his arms and the plant dark. Computing and adding together 16 such views produces (d), in which the plant becomes both dark and blurry, effectively disappearing.

(a) floodlit   (b) with holdout matte   (c) with inverse matte

Figure 10: Illumination using confocally derived mattes. (a) is an oblique view of figure 9(a). (b) shows the visual effect of loading the holdout mattes of figure 8(c) into the virtual projectors; the plant turns dark, but the soldier remains bright. It should be remembered that the illumination falling on the soldier is coming *through* the plant (actually through gaps between its leaves), an eerie effect when seen in person. (c) shows the effect of using the inverse mattes. Although almost no light directly reaches the soldier, he is slightly illuminated by light scattered from the plant.

832

(a) Base case: side lighting using a single video projector. By moving the projector to one side, backscatter in the viewing column is reduced relative to near-coaxial illumination. However, the resulting asymmetry produces a non-uniform image, leaving some areas too dark and others saturated.
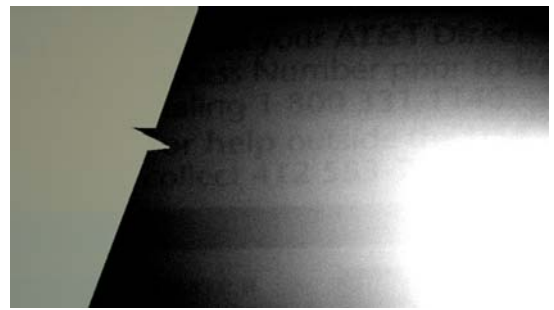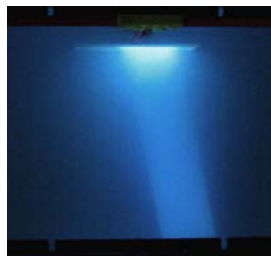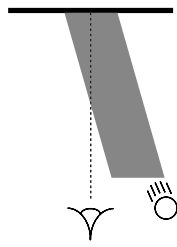
(b) Synthetic aperture illumination. 14 virtual projectors were used. Less light passes through the viewing column, as the diagram shows, thereby improving uniformity. However, a hot spot remains, making the center illegible. As the number of projectors rises, this hot spot approaches an ellipse in shape.
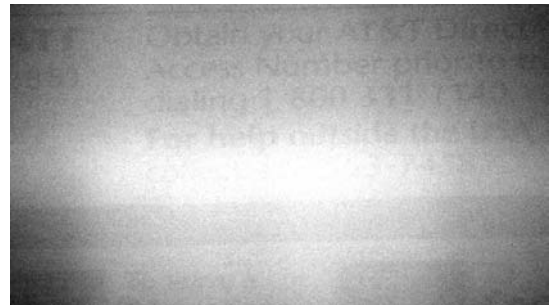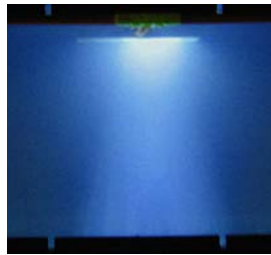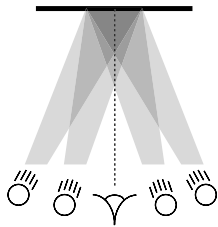
(c) Scanned aperture confocal imaging as described in section 3. The use of narrow beams shrinks the hot spot, which improves contrast. The use of scanning improves uniformity. The entire card is now legible. However, scanning is slow, and total illumination is low, leading to a low signal-to-noise (SNR) ratio.
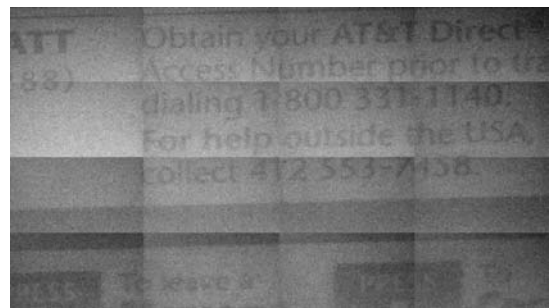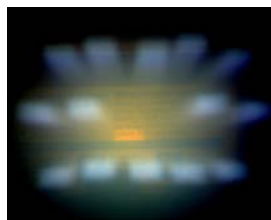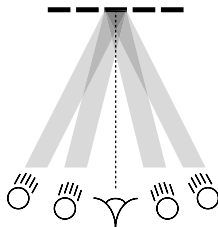
(d) Coded aperture confocal imaging. 16 virtual projectors were used. The camera was coaxial with one projector, although this was not necessary. Compared to (c), the milk concentration is lower, but the tank is larger, so the optical densities are similar. More importantly, total illumination is higher, so SNR is better.
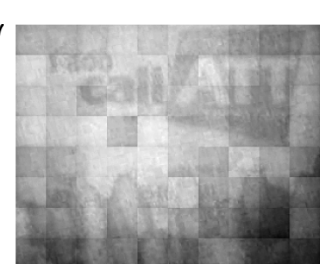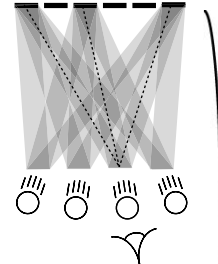
(e) Coded aperture confocal imaging combined with synthetic aperture photography (SAP). 16 virtual cameras were used, coaxial with the 16 virtual projectors. Confocal imaging alone (d) darkens the coral; confocal imaging with SAP darkens and blurs it. The text behind the coral is now legible, and SNR improves again.
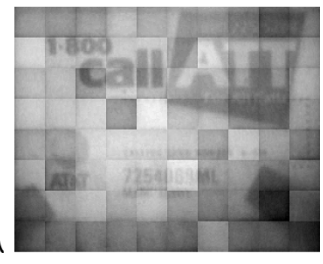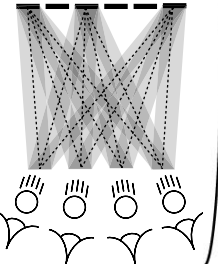
Figure 11: Using synthetic aperture illumination and photography to see through murky water. The target is the front or back of an AT&T calling card. The medium is 15 ml of 2% milk in a 10-liter tank (a-c) or 20 ml in a 20-liter tank (d-e). The diagrams depict each experiment: the target is at the top, the virtual projector(s) and camera(s) are at the bottom, illumination beams are gray polygons, and lines of sight are dashes. These diagrams are not drawn to scale; see figure 4 for the actual arrangement used in (d-e); (a-c) was similar. The next column shows the tank as seen from the top (a-b), from near the camera (c), or obliquely (d-e). This column of images was taken with less milk than the actual experiments, to make them clearer. The rightmost column shows grayscale contrast-stretched images, shot directly or synthesized using our algorithms. For comparison, part of an original (unstretched) color image is spliced into the left side of the top-right image. Note that only the magnetic stripe is visible in this image, and this is only barely visible. This is what the calling card looks like to the naked eye as seen through the tank during the actual experiments.

833

6. Conclusions and future work

Synthetic aperture illumination and synthetic aperture photography represent powerful but relatively unexplored imaging techniques. As these techniques mature, we expect applications to arise in surveillance, military reconnaissance, remote sensing and mapping, scientific and medical imaging, illumination engineering, and possibly stage and movie lighting.

Although their potential is great, the algorithms we propose here have a number of limitations. We cannot image partially occluded environments that are too dense, although more cameras and a wider aperture helps, and we cannot image scattering environments that are too opaque. We also assume that our scenes are diffuse rather than specular. Finally, our techniques are active; we must illuminate the scene. Thus, our techniques are not stealthy, and implementing them at very long distances would require very bright illumination.

The most important limitation of our current implementation is that the spatial resolution of our virtual projectors is low. (Camera resolution is also an issue, although less so.) This limits how small we can make our tiles, which leads to a larger depth of field. Underwater, it leads to larger hot spots, increasing backscattering and degrading contrast. Limited resolution also prevents us from adding more mirrors to reduce the statistical variability discussed earlier. To address this limitation, we envision replacing our array of mirrors by an interleaved array of cameras and projectors.

Many aspects of our techniques would benefit from further study. Unresolved theoretical questions include finding good illumination patterns, exploring the properties of aperture shapes and sampling patterns, and developing an aberration theory for synthetic apertures. Unaddressed empirical problems include quantitatively evaluating the performance of synthetic aperture photography and confocal imaging on large scenes and underwater.

7. Acknowledgments

8. References

CHAI, J.-X., TONG, X., CHAN, S.-C. SHUM, H.-Y. "Plenoptic Sampling," *Proc. SIGGRAPH 2000*.

CHAM, T.-J., REHG, J.M., SUKTHANKAR, R., SUKTHANKAR, G., "Shadow Elimination and Occluder Light Suppression for Multi-Projector Displays," *Proc. PROCAMS 2003*.

CHUANG, Y.-Y., CURLESS, B., SALESIN, D.H., SZELISKI, R., "A Bayesian Approach to Digital Matting," *Proc. CVPR 2001*.

COORG, S., TELLER, S., "Extracting Textured Vertical Facades from

(a) $Z = 0$ cm   (b) $Z = 3$ cm   (c) $Z = 6$ cm
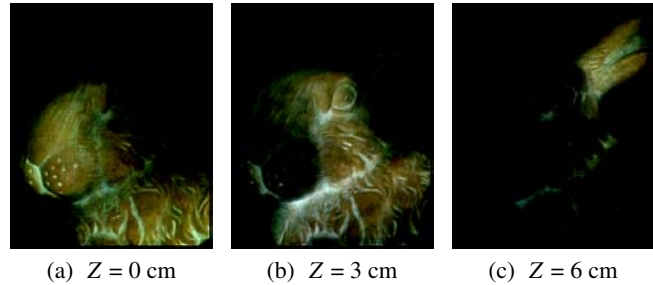
Figure 12: Using confocal imaging to compute cross-sectional images of the Stanford bunny. The synthetic focal plane was translated in software to three depths. At each depth confocal imaging was performed using 64 trials of sinuous patterns. In the resulting images, only a narrow band of surface straddling the focal plane is present. The depth of this band is about 2 cm.

Controlled Close-Range Imagery," *Proc. CVPR 1999*.

CORLE, T.R.. KINO, G.S., *Confocal Scanning Optical Microscopy and Related Imaging Systems*, Academic Press, 1996.

DEBEVEC, P., "Image-Based Lighting," *IEEE Computer Graphics and Applications*, Vol. 22, No. 2, March/April, 2002.

HAN, J.Y., PERLIN, K., "Measuring Bidirectional Texture Reflectance With a Kaleidoscope," *Proc. SIGGRAPH 2003*.

HARWIT, M., SLOANE, N.J.A., "Hadamard Transform Optics, " Academic Press, 1979.

ISAKSEN, A., MCMILLAN, L., GORTLER, S.J., "Dynamically Reparameterized Light Fields," *Proc. SIGGRAPH 2000*.

JAFFE, J.S., "Computer Modeling and the Design of Optimal Underwater Imaging Systems," *J. Oceanic Eng.*, Vol. 15, No. 2, 1990.

JAFFE, J.S., MCLEAN, J., STRAND, M.P., MOORE, K.D., "Underwater Optical Imaging: Status and Prospects," *Oceanography,* Vol. 14, No. 3, 2001, pp. 66-76.

LEVOY, M., HANRAHAN, P., "Light Field Rendering," *Proc. SIGGRAPH 1996*.

MALZBENDER, T., GELB, D., WOLTERS, H., "Polynomial Texture Maps," *Proc. SIGGRAPH 2001*.

MASSELUS, V., PEERS, P., DUTRE, P., WILLEMS, Y.D., "Relighting with 4D incident light fields," *Proc. SIGGRAPH 2003*.

MIDDLETON, W., *Vision Through the Atmosphere*, University of Toronto Press, 1952.

MOBLEY, C., *Light and Water: Radiative Transfer in Natural Waters*, Academic Press, 1994.

OKASHI, T., *Three-Dimensional Imaging Techniques*, Academic Press, 1976.

RUSINKIEWICZ, S., HALL-HOLT, O., LEVOY, M., "Real-Time 3D Model Acquisition," *Proc. SIGGRAPH 2002*.

SCHECHNER, Y.Y., KARPEL, N., "Clear Underwater Vision," *Proc. CVPR 2004*.

SCHECHNER, Y., NAYAR., S., BELHUMEUR, P., "A Theory of Multiplexed Illumination," *Proc. ICCV 2003*.

SMITH, A., BLINN, J., "Blue Screen Matting," *Proc. SIGGRAPH 1996*.

VAISH, V., WILBURN, B., JOSHI, N., LEVOY, M., "Using Plane + Parallax for Calibrating Dense Camera Arrays." *Proc. CVPR 2004*.

WILBURN, B., SMULSKI, M., LEE, K., HOROWITZ, M.A., "The Light Field Video Camera," *Proc. SPIE Electronic Imaging 2002*.

WILSON, T., JUSKAITIS, R., NEIL, M., KOZUBEK, M., "Confocal Microscopy by Aperture Correlation," *Optics Letters*, 21(3), 1996.

ZAND, J., "Coded aperture imaging in high energy astronomy," NASA Laboratory for High Energy Astrophysics (LHEA) at NASA's GSFC, 1996. URL: *http://lheawww.gsfc.nasa.gov/docs/cai/*.

834