Improving underwater vision using confocal imaging

Stanford Computer Graphics Laboratory Technical Memo 2009-001

June 28, 2009
(written March, 2005)

Marc Levoy
Computer Science Department
Stanford University

Hanumant Singh
Deep Submergence Laboratory
Woods Hole Oceanographic Institution

Abstract

Confocal microscopy is a family of techniques that employ patterned illumination and synchronized imaging to create cross-sectional views of biological specimens. In this paper, we adapt confocal imaging to improving visibility underwater by replacing the optical apertures used in microscopy with arrays of video projectors and cameras. In our principal experiment, we show that using one projector and one synchronized camera we can see further through turbid water than is possible using floodlighting. Drawing on experiments reported by us elsewhere, we also show that using multiple projectors and multiple cameras we can selectively image any plane in a partially occluded environment. Among underwater occluders we can potentially disregard this way are coral, fish, and marine snow, but we have not verified this underwater. Finally, we describe a large array of cameras we have built to simulate a single video camera of unusually high resolution and dynamic range. Such arrays might be useful in underwater imaging.

1. Introduction

Seeing through turbid water is a long-standing problem in ocean science. Its cause, of course, is the scattering of light as it passes through a participating medium. The mathematics of this problem have been well studied. 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 is loss of contrast and blurring. For weakly scattering media such as atmospheric aerosols and ocean waters [Mobley 1994], loss of contrast dominates. This suggests that we can enhance visibility by capturing images digitally and stretching their contrast, subject to the limits imposed by imaging noise. However, if the scene is artificially illuminated as happens in deep water, then the lighting will be backscattered to the camera, severely degrading contrast. To avoid this problem, oceanic engineers typically place floodlights well to the side of their camera [Jaffe 1990].

Alternatively, one can restrict illumination to a scanned stripe whose intersection with the target is recorded by a synchronously scanning camera. Although proposed in [Jaffe 2001], this method has not to our knowledge (as of March 2005) been tried underwater. The most similar existing system is a stationary underwater laser line scanner [Moore 2002]. However, the goal of that system is rangefinding, not photographic imaging, and the quality of the reflectance maps it returns are limited by geometrical and physical optics effects, including laser speckle. In this paper, we test Jaffe's proposal - using a video projector to create and move the stripe and a digital camera to record it. Our experiment was performed in a water tank at the Woods Hole Oceanographic Institution, using tap water laced with titanium dioxide to simulate turbid sea water.

In his 2001 paper, Jaffe relates synchronous scanning to confocal imaging techniques in microscopy [Wilson 1984]. One of us (Levoy) independently realized that confocal microscopy techniques could be applied to underwater vision [Levoy 2004]. Let us therefore begin by placing our present work in the context of these techniques.

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 employs the optical principle described in figure 1 (adapted from [Corle 1996]) to reduce the amount of light falling on, and recorded from, points off the focal plane. This principle holds only if the specimen scatters light diffusely and single scattering dominates over multiple scattering. However, useful images can often be obtained even if these properties are not strictly

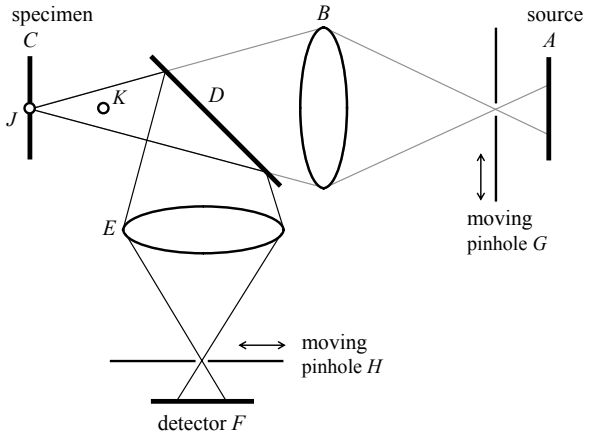


Figure 1: The principle of confocal microscopy (from [Levoy 2004]). An illumination source at A is imaged by an optical system B onto a 3D specimen that sits astride focal plane C. The specimen is imaged through a beamsplitter D and a second optical system E onto a detector F. A pinhole at G focuses 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. A second pinhole at H masks out everything but that portion of the image that is focused on J - hence the term confocal. The light gathered from K also falls off as the square of the distance from the plane. Hence, the contribution to the image by points off the focal plane falls off as the fourth power of distance from this plane.

satisfied. By moving the two pinholes in tandem, the specimen can be scanned. This yields a cross-sectional image of the specimen where it intersects the focal plane.

In previous work [Levoy 2004], we proposed replacing the optical apertures in confocal microscopy with synthetic apertures formed by arrays of projectors and cameras. In so doing, we obtained a discrete approximation of confocal imaging. 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 impractical to build a lens. In that paper, we applied our synthetic confocal imaging techniques to two problems: seeing through partially occluded environments such as foliage and crowds, and seeing through turbid water. However, in the latter application we tested our techniques only in a 10-gallon tank, with illumination and viewing distances of 10-30 cm. At that scale, backscatter dominated over attenuation. In typical ROV operations illumination and imaging distances are an order of magnitude larger, making attenuation more important.

In the present paper we describe a new test of confocal imaging in turbid water, performed at realistic scale and with better control over turbidity. Also, in our previous work we employed an array of 16 projectors and cameras. For the purpose of viewing a target (such as the seabed) through water that contains turbidity but no foreground occluders, it suffices to use one projector and one camera. Similar systems have been used for in-vivo imaging of cellular structures [Wang 2003]. In the next section, we describe such an implementation. In section 3, we revert to multiple projectors and cameras and revisit the problem of seeing through partial occluders.

2. Visibility in turbid water

For seeing through turbid water, we employ a single video projector and a synchronously scanning camera. With computer control over the projector, we scan over a set of finite-sized tiles that cover a nominal focal surface in 3-space, which may lie at any depth and may be tilted or even curved. For each tile position we capture an image with the camera, extract the pixels corresponding to that tile, and insert these into the output image. (We know which pixels to extract because our projector and camera have been calibrated to the focal surface, as described later.) By discarding pixels outside the tile, we effectively focus our image at the same place the light is focused, making the system confocal. Since only one tile is illuminated at once, the total time required to cover the target is proportional to the number of tiles. In the resulting image objects lying astride the focal surface are brightly illuminated, and objects lying off this surface are dark.

Experimental setup. To characterize the performance of two-beam confocal imaging in turbid water, we blocked off a roughly 6-foot section of a 50-foot water tank, blackened the walls, floor, and water surface to kill stray reflections, and arranged five projectors and one camera outside the tank as shown in figure 2. The projectors were Compaq MP1800 (1024 x 768 pixels) with an advertised white:black contrast ratio of 400:1. The camera was a Canon 10D (3072 x 2048 pixels) with about 12 bits of usable dynamic range. The projectors were placed in a row beside the camera, and all six devices were fixed to a platform. In the experiments whose results are shown in figures 3 and 4, we used only the fourth projector from the left or the leftmost three projectors, respectively. (Three projectors were used in the latter experiment to increase total light; since their beams coincided at the target, and the target was flat, they can be treated as a single projector.)

The target was an aluminum plate placed 132 cm from the camera and centered in its field of view. On this plate was affixed an EIA 1956 video resolution test chart, printed on white paper, laminated with plastic for waterproofing, and sanded to reduce specular reflections from the plastic.

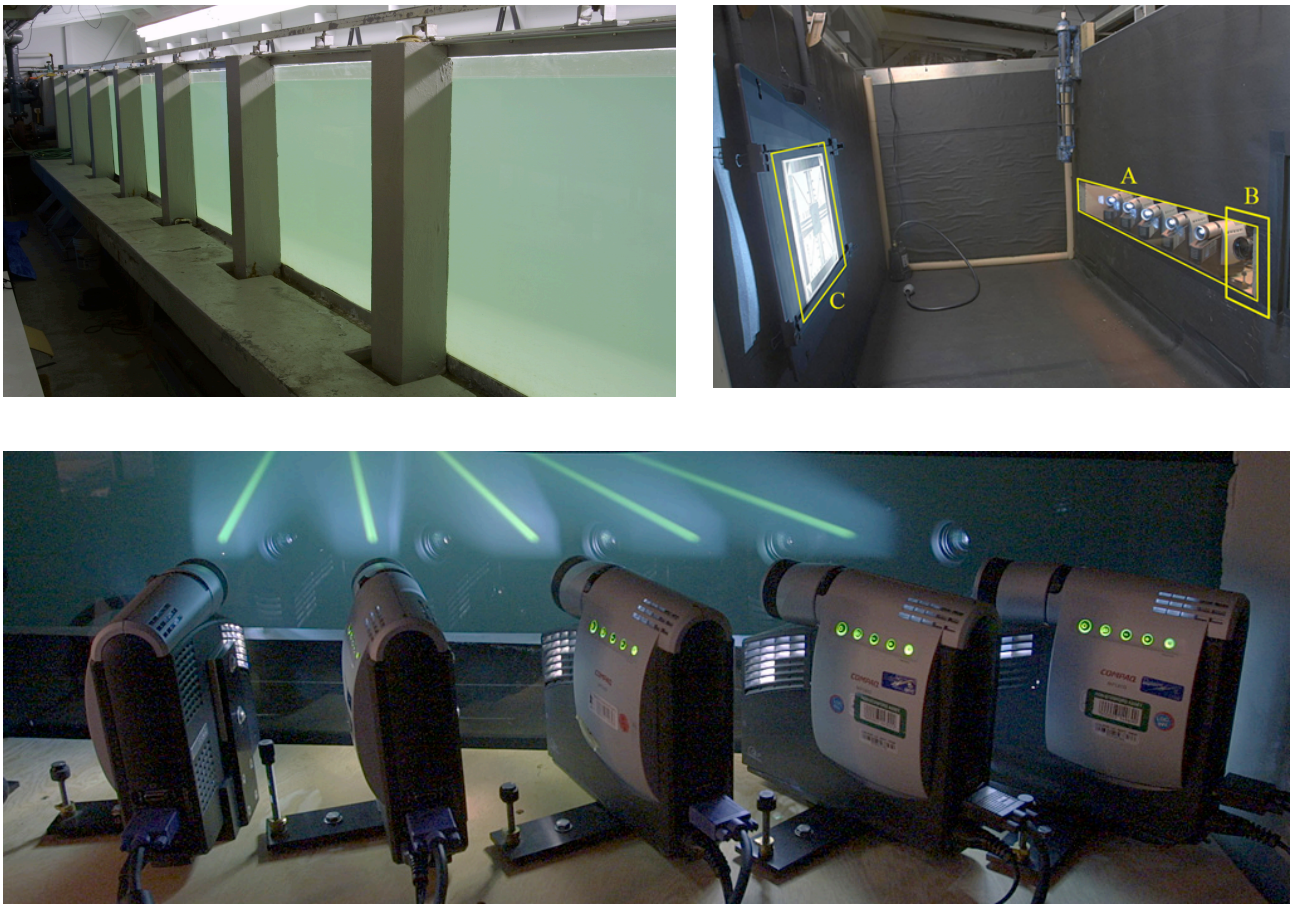
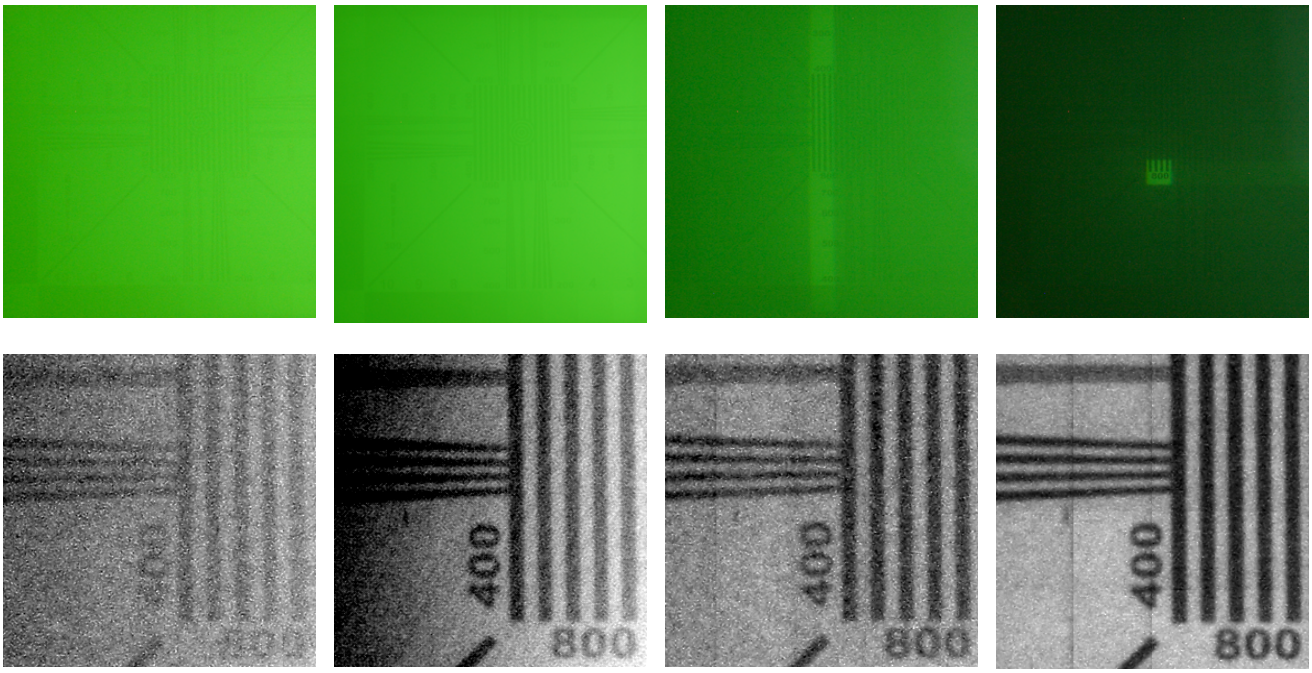


Figure 2: To test the ability of confocal imaging to see through turbid water, we arranged projectors and a camera outside a water tank and aligned them to image a common focal plane inside the tank. The image at top left shows the 50-foot flume at the Woods Hole Oceanographic Institution. We blocked off one of the 6-foot sections lying between two pillars, and on the low concrete shelf in the foreground we built a platform to hold our projectors and camera. The image at top right shows the interior of the section before filling with water. Visible at right are five video projectors (at A) and one camera (at B). A video test chart (at C) lies at the focal plane. The walls of the section were blackened to kill reflections, and a pump and transmissometer were used to maintain and measure turbidity. In the image at bottom the tank has been filled with turbid water, and the projectors are shown illuminating a small rectangular region of the test chart with green light. The white haze surrounding each green beam is caused by backscattering due to stray light from the black pixels in the projectors.

We calibrated the projectors and camera to the target using standard computer vision techniques [Vaish 2004], with a checkerboard affixed to the aluminum plate in place of the test chart. This procedure yielded a set of six 4 x 3 matrices giving homographies from points on the target plane to pixels on the camera and five projectors. Using these mappings, any point on the test chart could be addressed by any projector and by the camera.

The challenges of using projectors as light sources. While this physical arrangement sufficed for our experiment, it suffered from several problems. First, the projectors' images suffered chromatic dispersion as they passed through the glass wall of the tank. To ameliorate this problem we used only the green channel. Second, the projectors' images suffered misfocus due their angle of incidence on the test chart. To ameliorate the last two problems we decided not to use the rightmost projector. Third, the spatial resolution of these projectors is lower by a factor of three than the camera. These three problems, combined with the residual effects of calibration errors (although these were small by comparison), caused the boundaries of shaped beams to appear slightly blurry on the target.

The impact of this blur was that, when scanning an illumination beam across the target during confocal imaging, that portion of the illuminated area that could be extracted



(a) floodlit, single exposure SNR = 0.21 (b) floodlit, 20 exposures 4.8 (c) scanned vertical strip 3.7 (d) scanned square tile 8.0

Figure 3: Images of a video test chart seen through 118 cm of turbid water (attenuation length = 26 cm) as illuminated by the green channel of one video projector. The angle between the centerlines of illumination and imaging was 32 degrees. The top row shows raw images returned by the camera, and the bottom row shows grayscale versions after linear contrast stretching and, in the case of (c) and (d), confocal scanning as described in the text.

at each position of the beam needed to be inset slightly from the nominal beam width. Failure to do this would have led to objectionable seams in the confocal image. Unfortunately, this solution slows scanning and increases the amount of illumination contributing to backscatter but not to the final image. For example, the beam size used to produce figure 3(d) was 90 x 90 pixels, of which only the innermost 70 x 70 pixels were extracted at each beam position. Thus, 40% of the illumination was wasted. This blur could be reduced in the future by employing appropriate corrective optics on the projectors.

A fourth problem was the relatively poor contrast of our 2001-era video projectors. For any beam position during confocal scanning, most of the projector pixels are black. However, due to stray light in the optics, these pixels are not truly dark, as evidenced by white halos around each green beam in figure 2. This unwanted light contributed to backscattering and thus to loss of contrast. Moreover, since the number of black pixels was greater than the number of white pixels by a large factor, the amount of light that contributed to backscattering but did not directly illuminate the tile to be extracted was greater than the amount that directly illuminated the tile. In our experiment, there were 90 x 90 = 8100 white pixels and 1024 x 768 - 8100 = 778,332 black pixels. Assuming a contrast ratio of 400:1, this means that 25% of the light entering the tank was from nominally black pixels.

Fortunately, the contrast of commercial video projectors is steadily rising. As of March 2004 the highest contrast ratio available in a moderately-priced micromirror-based projector is 2000:1. Alternatively, this problem could be eliminated entirely by employing a scanned beam instead of a micromirror array to modulate the light. In a follow-on experiment (not shown here), we tested the likely effect of this improvement by blocking the black pixels from entering the tank using a physical mask affixed to the tank wall. The improvement in SNR was about 30%.

Creating and measuring turbidity. To create a scattering medium, we filled the tank with filtered tap water, allowed it to outgas for 48 hours, and introduced rutile titanium dioxide (99.5% TiO_2, 1-2 micron particles). Since the TiO_2 settled out of solution throughout the experiment, we monitored water opacity continuously using a Sea Tech 25 cm transmissometer. This device directly reports the transparency T of the water column lying between its light

source and sensor. Using standard formulae, this can be converted to an extinction coefficient $C = ln(T)/z$ where z is the path length (25 cm for this device) or to an attenuation length $L = -1/C$. (When light travels one attenuation length in a collimated beam, the irradiance due to unscattered photons is reduced to $1/e$ or approximately 37% of its original value [Mertens 1970].)

For the images in figure 3, measured turbidity was $T = 38\%$, which is equivalent to $C = -3.8/m$ or $L = 26cm$. Since the path length in water from the target to the camera was 118 cm, this corresponds to 4.5 attenuation lengths. Thus, about 1% of the light leaving the target in this experiment reached the camera without being absorbed or outscattered. To this signal must be added the relatively large quantity of light backscattered into the viewing column. This backscattered light produces noise, which we could directly measure in our images.

Experiment #1: different illumination protocols. In this experiment, we used the fourth projector from the left in figure 2, controlled by a PC to implement each of the following four protocols:

(a) Floodlighting of the target, by turning all pixels to full green. Exposure time was 3 seconds.

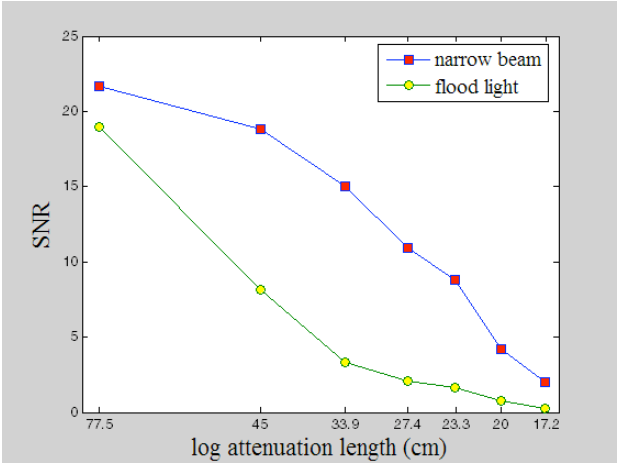
(b) A sum of 20 floodlit exposures. Total capture time was therefore 60 seconds.

(c) Confocal scanning using a horizontally scanned vertical stripe. Exposure time was 20 seconds per stripe. The target was divided into 21 stripes, so total capture time was 420 seconds.

(d) Confocal scanning using a horizontally and vertically scanned square tile. Exposure time was 30 seconds per tile. At this rate, more than three hours would be required to scan the 441 tiles in the target.

In each of these protocols, exposure time was set to the maximum that did not produce saturation of the sensor. Thus, we optimized each case for maximum performance given a projector of fixed brightness and and a camera of fixed sensitivity. We could instead have held exposure time constant; however, the comparison would have been less useful, since the confocal cases (c) and (d) would have been very dark.

Example camera images are shown in the top row of figures 3(a-d), with contrast-stretched versions below each one. For each stripe or tile position in cases (c) and (d), we extracted from each image only the illuminated region of pixels as explained earlier, inserting these pixels into the output image. At this high level of turbidity the floodlit image (a) has a signal-to-noise ratio less than 1.0 and is not legible. The sum of floodlit images (b) is more legible, due



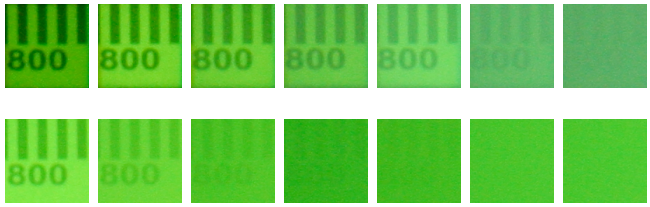


Figure 4: Images of a video test chart seen through water of varying turbidity and illuminated by three video projectors (angled at 14, 25, and 35 degrees). The top row of images were illuminated by narrow beams of light as in figure 3(d). The beams were aimed to converge on the test chart, where they formed a square measuring 2.5 cm on a side. The bottom row were floodlit as in figure 3(a). Above the images is a plot of SNR as a function of turbidity. Each data point corresponds to one image. Turbidity varies from clear tap water at left (extinction coefficient = -1.3/m, attenuation length = 77 cm) to very opaque at right (extinction coefficient = -5.8/m, attenuation length = 17 cm).

to the expected linear increase in SNR with the number of images summed. The confocally scanned stripe (c) is about the same, and the scanned square tile (d) is best. Note also that the floodlit images are more spatially nonuniform.

Experiment #2: increasing the turbidity. In this experiment, we varied the turbidity from clear to very opaque, using the leftmost three projectors in figure 2 and two illumination protocols: flood lighting and a confocally scanned tile. Exposure times for the floodlit case ranged from 0.1 second for clear water to 0.3 seconds for opaque water, and for the narrow beam from 0.1 second to 15 seconds per tile. As the images and plot in figure 4 show, confocal imaging provides consistently superior visibility compared to floodlighting. In fact, the improvement (ratios of SNRs) increases with increasing turbidity.

(a) single viewpoint (b) synthetic aperture image (c) confocal image (d) combining (b) and (c)

Figure 5: A demonstration (from [Levoy 2004]) of synthetic aperture imaging and confocal imaging. The scene is a toy soldier standing behind a plant. Adding together views from an array of 16 cameras produces a synthetic view (b) with a 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 confocal imaging using an array of 16 projectors produces (c), in which only surfaces near the focal plane are bright, leaving the soldier's arms and the plant dark. Computing and adding together 16 such views produces (d), in which the plant is both dark and blurry, effectively disappearing.

Discussion. From these experiments, it is evident that confocal imaging provides better visibility than floodlighting in turbid water. The likely cause of this improvement is reduced backscatter in the viewing column. The falloff in this improvement at very high turbidities (right side of the plot in figure 4) probably arises from attenuation of the incident and reflected beams, rather than from excessive backscatter. However, a conclusive explanation of this phenomenon awaits development of an analytical model and scattering simulation for underwater confocal imaging. This remains as future work.

Despite these promising results, deploying confocal imaging on an underwater platform or vehicle would require addressing several difficult practical issues. One such issue is the requirement in our experiments of knowing the distance between the imaging platform and the target, so that beams from the projector and camera can be confocally imaged. However, this knowledge is not strictly necessary. Instead, one can search along epipolar lines in the camera for the illuminated interval of pixels. (The projection on the camera's sensor of the ray associated with a projector pixel is called an epipolar line.)

A second practical issue is the long acquisition time required to perform confocal scanning. However, three factors mitigate this disadvantage. First, in this paper we employed only single-stripe or single-beam illumination protocols. In our earlier paper [Levoy 2004], we also explored protocols employing sequences of random binary patterns. Although total illumination of the tank is higher, thereby lowering SNR, the number of patterns is smaller (tens of exposures instead of hundreds). Second, if the projector were replaced by a light source of fixed brightness and condenser optics optimized for each illumination protocol, then total capture times for (a), (c), and (d) would be equal, but the superior SNR of (c) and (d) would remain.

Lastly, if the scanning direction in protocol (c) were made perpendicular to the motion of an underwater vehicle, it might be possible to employ the motion of the vehicle to perform the scan (as envisioned by Jaffe [2001]). The output of this process would be a pushbroom panorama [Gupta 1997] - perspective parallel to the stripe and orthographic in the direction of motion. Protocol (c) has a further advantage if deployed in this way; since the illumination is a stripe being viewed from the side, one can triangulate from its leading or trailing to compute bathymetry. In other words, illumination using a fixed stripe on a vehicle moving perpendicular to the stripe is both a confocal imaging method and a rangefinding method.

3. Visibility in occluded environments

In section 2 we considered the problem of viewing a target through water that contains only turbidity. In this section, we apply confocal imaging to seeing through partial foreground occlusions such as coral, fish, or marine snow. To make these occluders disappear, one must employ the full power of confocal imaging - illuminating the target with multiple projectors and imaging it with multiple cameras. The procedure is the same as before, except all projectors and cameras must be calibrated to the same focal plane. As long as at least one pair of illumination and imaging lines of sight remain unoccluded, the target

remains visible, although possibly at reduced contrast. A theoretical analysis of the performance of multi-projector multi-camera confocal imaging is beyond the scope of this paper; the interested reader is referred to [Levoy 2004].

Figure 5, taken from [Levoy 2004], demonstrates this method, using a non-underwater scene. The scene was illuminated by a single Compaq MP1800 video projector reflected through a tightly packed array of planar mirrors, thereby creating 16 virtual projectors, each with lower spatial resolution but a different effective viewpoint. Imaging was performed by a single Canon 10D digital camera reflected through the same array of planar mirrors using a beamsplitter. This produced 16 virtual cameras, again with lower spatial resolution and differing viewpoints. The locus of these illumination and imaging viewpoints becomes the aperture of this synthetic confocal imaging system. The size of this aperture was 27 degrees as viewed from the center of the scene. As the figure shows, we can readily produce images in which the plant has become blurry, dark, or both.

Although we have not tried this method underwater, we can think of no reason why it should not work. Depending on the application, it may suffice to use one projector and multiple cameras - for example if one only needs to disregard marine snow.

4. Arrays of cameras and projectors

In sections 2 and 3, we used conference room video projectors and a high-quality digital SLR camera. Due to cost, physical size, and power requirements, this choice of equipment limits our ability to scale confocal imaging to much larger apertures while maintaining a dense sampling of the aperture. To address this scaling problem, one of us (Levoy) has for several years been building arrays of inexpensive cameras and projectors [Wilburn 2005]. In experiments reported elsewhere [Vaish 2004; Vaish 2005], we used an array of 100 continuously streaming, VGA-resolution video cameras to successfully isolate one person from a crowd, even as they walk around. In that system, the selected person is imaged sharply, while the crowd disappears into a blur. We imagine that this capability could be exploited underwater, for example to track and image one fish in a school.

In these experiments, the fields of view of these 100 cameras overlapped at some depth in front of the array. If the fields of view instead abut one another with a small overlap, we then obtain a virtual video camera of very high spatial resolution - more than 6,000 pixels on a side. Each camera in this arrangement can meter independently on that portion of the scene that it sees, giving us the ability to capture a scene having widely varying brightnesses across the field of view. In the arrangement shown in figure 6, the fields of view of the cameras overlap by 50%, creating a virtual camera 3,000 pixels wide. Due to this overlap each viewing ray is seen by four cameras. This gives us the ability to capture a scene with high dynamic range within the tile seen by each camera as well varying brightnesses across the field of view. See [Wilburn 2005] for details.

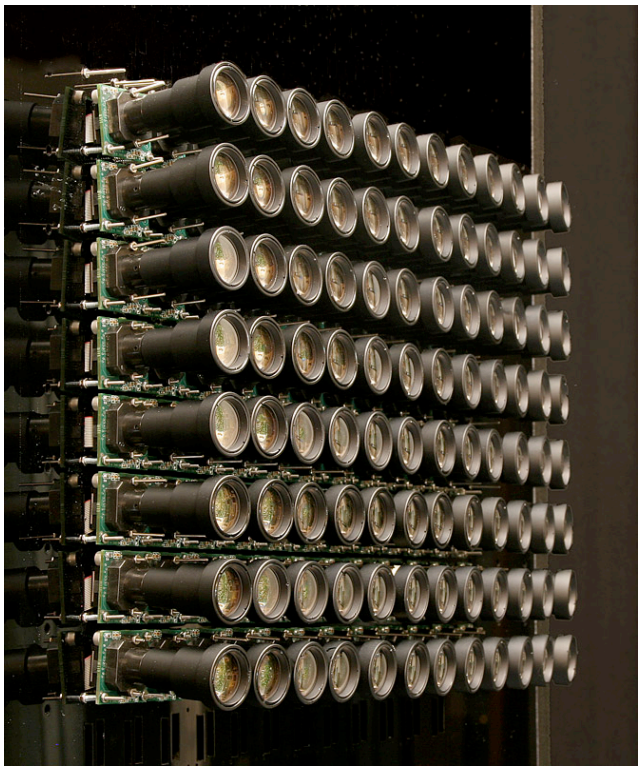


Figure 6: An array of inexpensive cameras is used to create a single virtual camera of unusually high spatial resolution. The cameras are fitted with telephoto lenses, and the cameras are splayed outwards, creating a video camera with a normal field of view (about 30 degrees) and very high spatial resolution (6,000 pixels wide). Each camera adjusts its exposure independently, allowing us to record high dynamic range environments.

Finally, one can envision building a matching array of miniature video projectors - possibly based on LED illumination sources, thereby permitting confocal imaging over a large synthetic aperture using inexpensive, energy-efficient components. Such a "hybrid array" could be deployed on stationary platforms or larger underwater vehicles.

Acknowledgements

We thank Vaibhav Vaish and Billy Chen of Stanford University for calibration software, and Megan Carroll, Donald Peters, David Gray, Sandy Williams, and Chris Roman of the Woods Hole Oceanographic Institution for assistance building our experimental apparatus. This work was supported by the NSF under contract IIS-0219856-001 and DARPA under contract NBCH-1030009.

References

[Corle 1996] Corle, T.R.. Kino, G.S., *Confocal Scanning Optical Microscopy and Related Imaging Systems*, Academic Press, 1996.

[Gupta 1997] Gupta, R., Hartley, R.I., "Linear Pushbroom Cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 19, No. 9, September, 1997.

[Jaffe 1990] Jaffe, J.S., "Computer modeling and the design of optimal underwater imaging systems," *J. Oceanic Eng*., Vol. 15, No. 2, 1990.

[Jaffe 2001] 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 2004] Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., Bolas, M., "Synthetic aperture confocal imaging," *Proc. SIGGRAPH 2004*.

[Mertens 1970] Mertens, L.E., *In-Water Photography*, Wiley-Interscience, 1970.

[Moore 2002] Moore, K.D., Jaffe, J.S., "Time-evolution of high-resolution topographic measurements of the sea floor using a 3D laser line scan mapping system," *IEEE Journal of Oceanic Engineering*, Vol. 27, No. 3, July, 2002.

[Schechner 2004] Schechner, Y.Y., Karpel, N., "Clear Underwater Vision," *Proc. CVPR 2004*.

[Vaish 2004] Vaish, V., Wilburn, B., Joshi, N., Levoy, M., "Using Plane + Parallax for Calibrating Dense Camera Arrays." *Proc. CVPR 2004*.

[Vaish 2005] Vaish, V., Garg, G., Talvala, E., Antunez, E., Wilburn, B., Horowitz, M., Levoy, M., "Synthetic Aperture Focusing using a Shear-Warp Factorization of the Viewing Transform," *Proc. Workshop on Advanced 3D Imaging for Safety and Security*, in conjunction with CVPR 2005.

[Wang 2003] Wang, T.D., Mandella, M.J., Contag, C.H., Kino, G.S., Dual-axis confocal microscope for high-resolution in vivo imaging, Optics Letters, Vol. 28, No. 6, March 15, 2003.

[Wilburn 2005] Wilburn, B., Joshi, N., Vaish, V., Talvala, E.-V., Antunez, E., Barth, A., Adams, A., Horowitz, M., Levoy, M., "High Performance Imaging Using Large Camera Arrays," *ACM Transactions on Graphics* (Proc. SIGGRAPH), Vol. 24, No. 3, 2005.

[Wilson 1984] Wilson, ,T., Sheppard, C., *Theory and Practice of Scanning Optical Microscopy*, Academic Press, 1984.