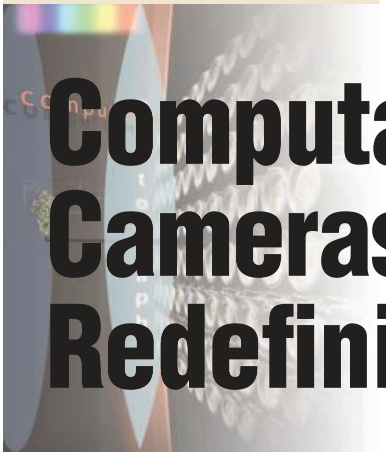
COVER FEATURE



Computational Cameras: Redefining the Image

Shree K. Nayar
Columbia University

Computational cameras use unconventional optics and software to produce new forms of visual information, including wide field-of-view images, high dynamic range images, multispectral images, and depth images. Using a controllable optical system to form the image and a programmable light source as the camera's flash can further enhance the capabilities of these cameras.

The camera's evolution over the past century has been truly remarkable. Throughout this evolutionary process, however, the principle underlying the camera has remained the same—namely, the camera obscura,[1] Latin for "dark room." As Figure 1a shows, the *traditional camera* has a detector—either film or solid-state—and a lens that essentially captures the light rays that pass through its center of projection, or effective pinhole. In other words, the traditional camera performs a special and restrictive sampling of the complete set of rays, or the light field,[2] that resides in a real scene.

Computational cameras sample the light field in radically different ways to create new and useful forms of visual information. A computational camera embodies the convergence of the camera and the computer. As Figure 1b shows, it uses new optics to map rays in the light field to pixels on the detector in an unconventional fashion. For example, the computational camera assigns the yellow ray, which would travel straight through to the detector in a traditional camera, to a different pixel. In addition, it can alter the ray's brightness and spectrum before the pixel receives it, as illustrated by the change in its color from yellow to red.

In all cases, because the captured image is optically coded, interpreting it in its raw form might be difficult. However, the computational module knows everything it needs to know about the optics. Hence, it can decode the captured image to produce new types of images that could benefit a vision system—either a human observing the images or a computer that analyzes the images to interpret the scene.

COMPUTATIONAL CAMERAS

At Columbia University's Computer Vision Laboratory, we have developed several types of computational cameras. As the "Related Research" sidebar describes, several research groups around the world are working on the development of computational cameras and related technologies.

Imaging can be viewed as having several dimensions, including spatial resolution, temporal resolution, spectral resolution, field of view, dynamic range, and depth. Each of the cameras presented here can be viewed as exploring one of these dimensions.

Field of view

The first imaging dimension we will look at is field of view. Most imaging systems, both biological and artificial, are rather limited in their fields of view. They can only capture a small fraction of the complete sphere around their location in space. Clearly, if a camera could capture the complete sphere or even a hemisphere, it would profoundly impact the capability of the vision system that uses it. French philosopher Michel Foucault explored at great length the psychological implications

Published by the IEEE Computer Society

0018-9162/06/$20.00 © 2006 IEEE

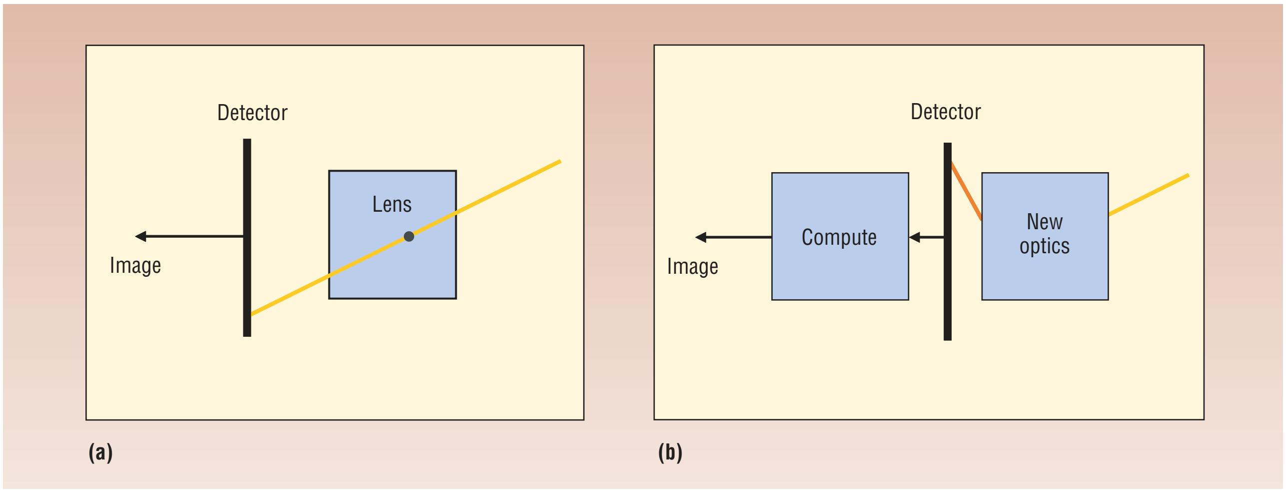


Figure 1. Traditional and computational cameras. (a) The traditional camera is based on the camera obscura principle and produces a linear perspective image. (b) A computational camera uses novel optics to capture a coded image and a computational module to decode the captured image to produce new types of visual information.

of being able to see everything at once in his discussion of the panopticon.[3]

First introduced about a century ago, the fish-eye lens[4] is a wide-angle imaging apparatus that uses meniscus (crescent-shaped) lenses to severely bend light rays into the camera—in particular, the rays that are in the periphery of the field of view. However, it is difficult to design a fish-eye lens with a field of view that is much larger than a hemisphere while maintaining high image quality. To address this limitation, we use *catadioptrics*, an approach that combines the use of lenses and mirrors. Catadioptrics has been used extensively to develop telescopes.[5] While a telescope captures a very small field of view, here we are interested in exactly the opposite: capturing an unusually large field of view.

In developing a wide-angle imaging system, ensuring that the camera captures principal rays of light that pass through a single viewpoint, or center of projection, is highly desirable. If the system meets this condition, regardless of how distorted the captured image is, software can map any part of it to a normal perspective image. For that matter, the user can emulate a rotating camera to freely explore the captured field of view. In our work, we have derived a complete class of mirror-lens combinations that capture wide-angle images while satisfying the single viewpoint constraint. This family of cameras uses ellipsoidal, hyperboloidal, or paraboloidal mirrors, some of which were implemented in the past. We have also shown that it is possible to use two mirrors to reduce the imaging system's packaging while maintaining a single viewpoint.

Related Research

Several academic and industrial research teams around the world are developing a variety of computational cameras. In addition, some well-established imaging techniques naturally fall within the definition of a computational camera. A few examples are integral imaging[1] for capturing a scene's 4D light field; coded aperture imaging[2] for enhancing an image's signal-to-noise ratio; and wavefront coded imaging[3] for increasing an imaging system's depth of field. Each of these techniques uses unconventional optics to capture a coded image of the scene, which is then computationally decoded to produce the final image. This approach is also used for medical and biological imaging, where it is referred to as *computational imaging*. Finally, significant technological advances are also being made with respect to image detectors.[4-6] In particular, several research teams are developing detectors that can perform image sensing as well as early visual processing.

References

1. G. Lippmann, "La Photographie Intégral," *Comptes-Rendus*, vol. 146, Académie des Sciences, 1908, pp. 446-551.
2. E.E. Fenimore and T.M. Cannon, "Coded Aperture Imaging with Uniformly Redundant Arrays," *Applied Optics*, vol. 17, 1978, pp. 337-347.
3. E.R. Dowski Jr. and W.T. Cathey, "Wavefront Coding for Detection and Estimation with a Single-Lens Incoherent Optical System," *Proc. ICASSP Conf.*, vol. 4, May 1995, pp. 2451-2454.
4. C. Mead, *Analog VLSI and Neural Systems*, Addison Wesley Longman, 1989.
5. J.L. Wyatt et al., "Analog VLSI Systems for Image Acquisition and Fast Early Vision Processing," *Int'l J. Computer Vision*, vol. 8, 1992, pp. 217-230.
6. T. Kanade and R. Bajcsy, "Computational Sensors," DARPA workshop report, May 1993; www.ri.cmu.edu/pubs/pub_1333.html.

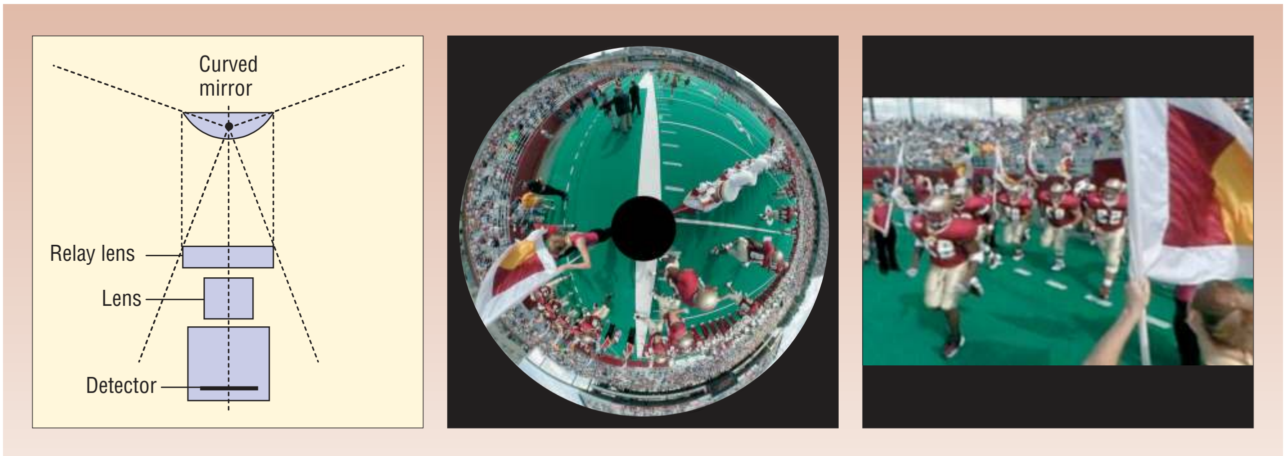


Figure 2. Wide-angle imaging using a catadioptric camera.

Figure 2 shows an example of this class of wide-angle catadioptric cameras. This implementation is an attachment to a conventional camera with a lens that includes a relay lens and a paraboloidal mirror. As the figure shows, this camera's field of view is significantly greater than a hemisphere. It has a 220-degree field of view in the vertical plane and a 360-degree field of view in the horizontal plane. The middle of the figure shows an image captured by the camera. The black spot in the center is the camera's blind spot where the mirror sees the relay lens. Although the image was captured from close to ground level, the sky is visible above the football stadium bleachers.

This image illustrates the power of a single-shot wide-angle camera over traditional methods that stitch a sequence of images taken by rotating a camera to obtain a wide-angle mosaic. While mosaicing methods require the scene to be static during the capture process, a single-shot camera can capture a wide view of even a highly dynamic scene.

Since the camera's computational module knows the optical compression of the catadioptric field of view, it can map any part of the captured image to a perspective image, such as the one shown on the right. This mapping is a simple operation that can be done at video rate using even a low-end computer. We have demonstrated the use of 360-degree cameras for videoconferencing and video surveillance.

Dynamic range

While digital cameras have improved dramatically with respect to spatial resolution, they remain limited in terms of the number of discrete brightness values they can measure. Consider a scene that includes a person indoors lit by room lamps and standing next to an open window where the sun brightly lights the scene outside. If the camera's exposure time is increased to ensure the person appears well lit in the image, the window would be washed out, or saturated. Conversely, if the exposure time is lowered to capture the bright outdoor scene, the person will appear dark in the image. This occurs because digital cameras typically measure 256 levels (8 bits) of brightness in each color channel, which is simply not enough to capture the rich brightness variations in most real scenes.

A popular way to increase a camera's dynamic range is to capture many images of the scene using different exposures and then use software to combine the best parts of the differently exposed images. Unfortunately, this method requires the scene to be more or less static as there is no reliable way to combine the different images if they include fast-moving objects. Ideally, we would like to have the benefits of combining multiple exposures of a scene with the capture of a single image.

In a conventional camera, all pixels on the image detector are equally sensitive to light. Our solution is to create a detector with an *assortment of pixels* with different sensitivities either by placing an optical mask with cells of different transmittances on the detector or by having interspersed sets of pixels on the detector exposed to the scene over different integration times. Most color cameras already come with an assortment of pixels: Neighboring pixels have different color filters attached to them. In our case, the assortment is more complex as a small neighborhood of pixels will not only be sensitive to different colors, but the pixels of the same color will also have different transmittances or integration times.

The left side of Figure 3 shows a camera with assorted pixels. Unlike a conventional camera, for every pixel that is saturated or too dark there will likely be a neighboring pixel that is not. Hence, even though the captured image may have bad data, it is interspersed with good data. The middle of the figure shows an image captured with this camera. The magnified inset image shows the image's expected checkerboard appearance. Applying image reconstruction software to this optically coded image creates a wide dynamic range image, as the

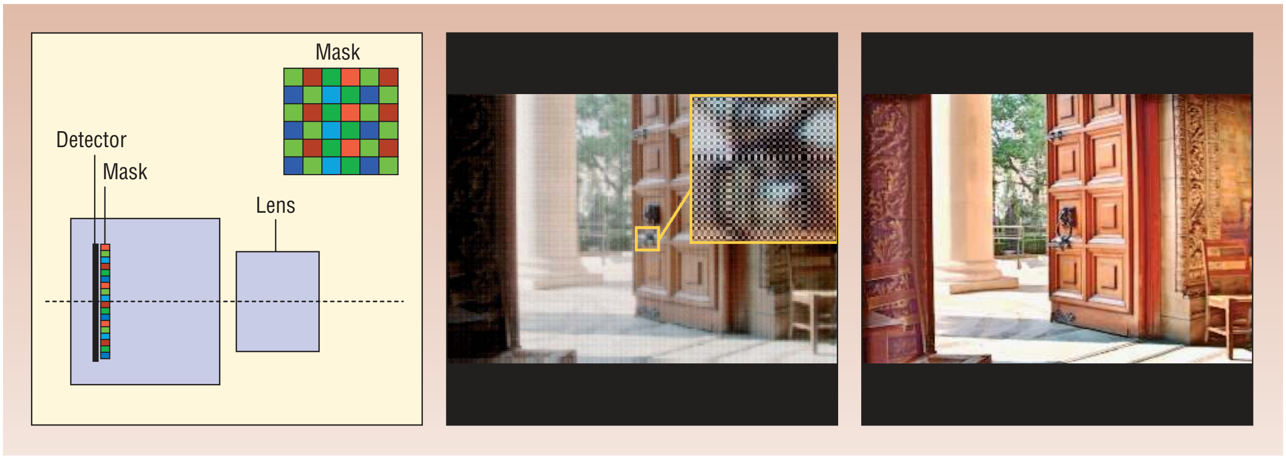


Figure 3. High dynamic range imaging using assorted pixels.

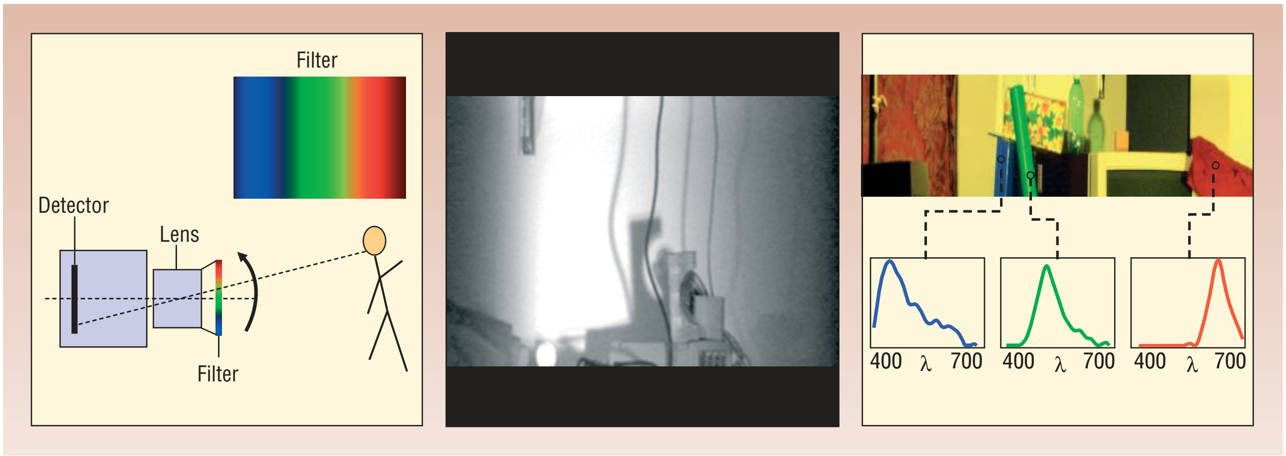


Figure 4. Multispectral imaging using generalized mosaicing.

right side of the figure shows. This image includes details on the dark walls lit by indoor lighting as well as the bright sunlit regions outside the door.

Spectrum

Figure 4 shows how the well-known method of image mosaicing can be extended to capture both a wide-angle image and additional scene information. The left side of the figure illustrates the key idea, showing a video camera with an optical filter with spatially varying properties attached to the front of the camera lens. In this example, a black-and-white video camera is used with a linear interference filter that passes a different wavelength of the visible light spectrum through each of its columns (inset image). The middle of the figure shows an image captured by the video camera. The camera is moved with respect to a stationary scene, and a registration algorithm aligns the acquired images. Registration provides multiple measurements of the radiance of each scene point that correspond to different wavelengths. Interpolation of these measurements determines the spectral distribution of each scene point. Instead of the three-color mosaic (red, green, blue) that traditional mosaicing provides, the result is the multispectral mosaic shown on the right side of Figure 4.

This *generalized mosaicing* approach can be used to explore various dimensions of imaging by simply using the appropriate optical filter. A spatially varying neutral density filter can be used to capture a wide dynamic range mosaic, and a filter with spatially varying polarization direction can be used to separate diffuse and specular reflections from the scene and detect material properties. When the filter is a wedge-shaped slab of glass, the scene points are measured under different focus settings to compute an all-focused mosaic. In fact, multiple imaging dimensions can be explored simultaneously by using more complex optical filters.

Depth

Figure 5 on the next page shows how a computational camera can be used to extract a scene's 3D structure from a single image. A hollow cone that is mirrored on the inside is placed in front of a conventional perspective camera. The cone's axis is aligned with the camera's opti-

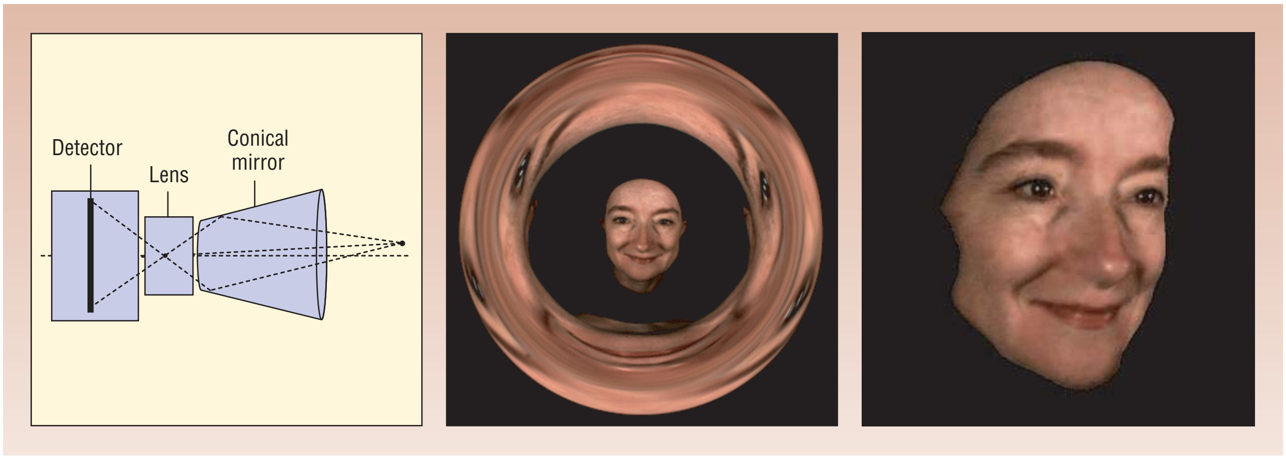


Figure 5. Depth imaging using multiview radial camera.

cal axis. Since the mirror is hollow, the camera lens sees a scene point directly. In addition, it is reflected by exactly two points on the conical mirror that lie on a plane that passes through the scene point and the camera's optical axis. As a result, each scene point is imaged from three different viewpoints: the center of projection of the camera lens and two virtual viewpoints that are equidistant and on opposite sides with respect to the optical axis. Consequently, the image includes three views of an entire scene: one from the center of projection of the lens and two additional views from a circular locus of viewpoints whose center lies on the optical axis.

The middle of Figure 5 shows an image of a face captured by this *radial imaging system*. Notice how the center of the image is just a regular perspective view of the face. Two additional views of the face are embedded in the annulus around this view. A stereo matching algorithm finds correspondences between the three views and computes the face's 3D geometry.

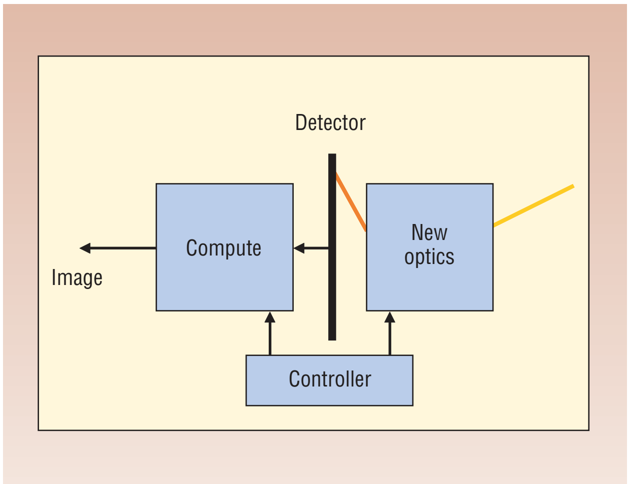


Figure 6. A programmable imaging system is a computational camera in which the optics and software can be varied to emulate different imaging functionalities.

The image on the right in this figure shows a new rotated view of the computed face geometry. While a conical mirror with specific parameters was used here, changing the mirror's parameters can create a variety of radial imaging systems with different imaging properties. We have used this approach to recover the fine geometry of a 3D texture, to capture complete texture maps of simple objects, and to measure the reflectance properties of real-world materials.

PROGRAMMABLE IMAGING

Although computational cameras produce images that are fundamentally different from the traditional perspective image, the hardware and software of each of these devices are designed to produce a particular type of image. The nature of this image cannot be altered without significant redesign of the device.

A *programmable imaging system* uses an optical system for forming the image that a controller can vary in terms of its radiometric or geometric properties as shown in Figure 6. When such a change is applied to the optics, the controller also changes the software in the computational module. The result is a single imaging system that can emulate the functionalities of several specialized systems. Such a flexible camera has two major benefits. First, a user is free to change the camera's role as needed. Second, we can begin to explore the notion of a purposive camera that, as time progresses, always produces the visual information that is most pertinent to the task.

The left side of Figure 7a shows a programmable imaging system that uses a two-dimensional array of *micromirrors*, which have controllable orientations. The image of the scene is first formed using a lens on the micromirror array. The plane on which the array resides is then reimaged onto an image detector using a second lens. While it would be ideal to have a micromirror array with mirror orientations that can be set to any desired value, such a device is not available at this time.

Figure 7. Programmable imaging systems that use controllable spatial light modulators to vary their radiometric and geometric properties based on the application's needs. (a) Adaptive dynamic-range imaging with a micromirror array. (b) Split field-of-view imaging with a volumetric aperture.

Our implementation uses the digital micromirror device (DMD) developed by Texas Instruments[6] that serves as the workhorse for many currently available digital projectors. This array's mirror can only be switched between two orientations: +10 degrees and −10 degrees. When a micromirror is oriented at +10 degrees, the corresponding image detector pixel is exposed to a scene point; when the micromirror is at −10 degrees, it receives no light. The DMD can switch between the two orientation states in a matter of microseconds.

This system can independently adapt the dynamic range of each of its pixels based on the brightness of the scene point it sees. In this case, each pixel's exposure on the image detector is determined by the fraction of the integration time of the detector for which the corresponding micromirror on the DMD is oriented at +10 degrees. A simple control algorithm updates each pixel's exposure duration based on the most recent captured image.

A conventional 8-bit video camera was used to capture the image in the middle of Figure 7a. The image on the right shows the programmable imaging system's output with adaptive dynamic range. Note how the pixels that are saturated in the conventional camera image are brought into the dynamic range of the 8-bit detector. The inset image on the left of Figure 7a shows the adaptive exposure pattern applied to the micromirror array. The system can use this image with the captured image on the right to compute an image with a very wide dynamic range. This imaging system can also perform other functions such as feature detection and object recognition.

In virtually any imaging system, the main reason to use a lens is to gather more light. However, this benefit of a lens comes with a price in that it severely restricts the geometric mapping of scene rays to image points. To address this limitation, we have been exploring lensless imaging systems.

Consider a bare image detector exposed to a scene. In this case, each pixel on the detector receives a 2D set of rays of different directions from the scene. The detector itself is a 2D set of pixels of different spatial locations arranged on a plane. Therefore, although the detector produces a 2D image, it receives a 4D set of light rays from the scene.

Now, consider what happens when a *3D* (*volumetric*) *aperture* is placed in front of the detector instead of a lens, as shown on the left of Figure 7b. If the aperture has a

3D transmittance function embedded within it, it will modulate the 4D set of light rays before the 2D detector receives them. If this transmittance function could be controlled, it would be possible to apply a variety of modulation operations on the 4D set of scene rays. Such a device could map scene rays to pixels in ways that would be difficult, if not impossible, using a lens-based camera.

Unfortunately, implementing a controllable volumetric aperture is not easy. Consequently, we have implemented the aperture as a stack of controllable 2D apertures. Each aperture is a liquid crystal (LC) sheet of the type used in displays. By simply applying an image to the LC sheet, we can control its modulation function and change it from one captured image to the next.

The inset image on the left of Figure 7b shows how appropriately selecting the open (full transmittance) and closed (zero transmittance) areas on two apertures projects three disconnected fields of view onto adjacent regions on the detector. Comparing the middle and right images in Figure 7b demonstrates the advantage of such a "split field-of-view" projection. The middle image was taken with a conventional camera. Although we are only interested in the three people in the scene, a large fraction of the detector's resolution is wasted on the scene regions in between the people. In the right image, taken using the lensless system, the three people are optically cropped out of the scene and imaged with higher resolution.

The time may have arrived for the flash to play a more sophisticated role in capturing images.

PROGRAMMABLE ILLUMINATION: A SMARTER FLASH

Since the dawn of photography, people have been trying to take pictures of dimly lit scenes. The only way to obtain a reasonably bright image of a dark scene was by using a very long exposure time, during which the scene had to remain stationary. The flashbulb was invented to overcome this limitation. Based on patents awarded to Johannes Ostermeier, a German inventor, the first commercial flashbulb became available around 1930. Today, the flashbulb, commonly referred to as the "flash," is an integral part of virtually every consumer camera.

The flash's basic capability has remained the same since its invention. Used to brightly illuminate the camera's field of view during the image detector's exposure time, the flash essentially serves as a point light source that illuminates everything within a reasonable distance from the camera.

In recent years, researchers have begun exploring ways to combine images taken with and without a flash to produce higher-quality images. Multiple flashes placed around the camera's lens have also been used to detect depth discontinuities and produce stylized renderings of the scene.

Given the enormous technological advancements made with respect to digital projectors, the time may have arrived for the flash to play a more sophisticated role in capturing images. Using a projector-like light source as a camera flash is a powerful alternative as it provides full control over the 2D set of rays it emits. The camera can project arbitrarily complex illumination patterns onto the scene, capture the corresponding images, and compute information regarding the scene that is not possible to obtain with the traditional flash. In this case, the captured images are optically coded due to the patterned illumination of the scene. Two examples illustrate the benefits of using a digital projector as a programmable camera flash.

In the left side of Figure 8a, a camera and projector are colocated by using a half-mirror. This configuration has the unique property that the projector can illuminate all the points that are visible to the camera. To maximize the brightness of the images they produce, projectors have large apertures and hence narrow depths of field. We have developed a method that exploits a projector's narrow depth of field to recover the geometry of the scene the camera views. The method uses a stripe pattern like the one shown in the inset image in Figure 8a. This pattern is shifted a minimum of three times, and the camera captures the corresponding images. The set of intensities measured at each camera pixel reveals the defocus of the shifted pattern, which in turn gives the depth of the scene point.

This *temporal defocus method* has two advantages. First, since depth is computed independently for each camera pixel, we can recover sharp depth discontinuities. Second, since it is based on defocus and not triangulation, we can colocate the projector and the camera and compute a depth map that is "image-complete"—that is, there are no holes in the depth map from the camera's perspective.

The middle of Figure 8a shows an image of a complex scene that includes a flower behind a wooden fence and its depth map (shown as a gray-scale image) computed using the temporal defocus method. The depth map can be used to blur the scene image spatially to render it as it would appear through a narrow depth-of-field camera lens. The right side of Figure 8a shows such a "refocused" image, in which the flower petals in the back are in focus while the fence in the front is blurred.

In short, a photographer can vary the image's depth of field after capturing it. We have also used depth maps computed using the temporal defocus method to insert synthetic objects within the captured image with all the desired occlusion effects.

Finally, consider a scene lit by a point light source and viewed by a camera. The brightness of each scene point

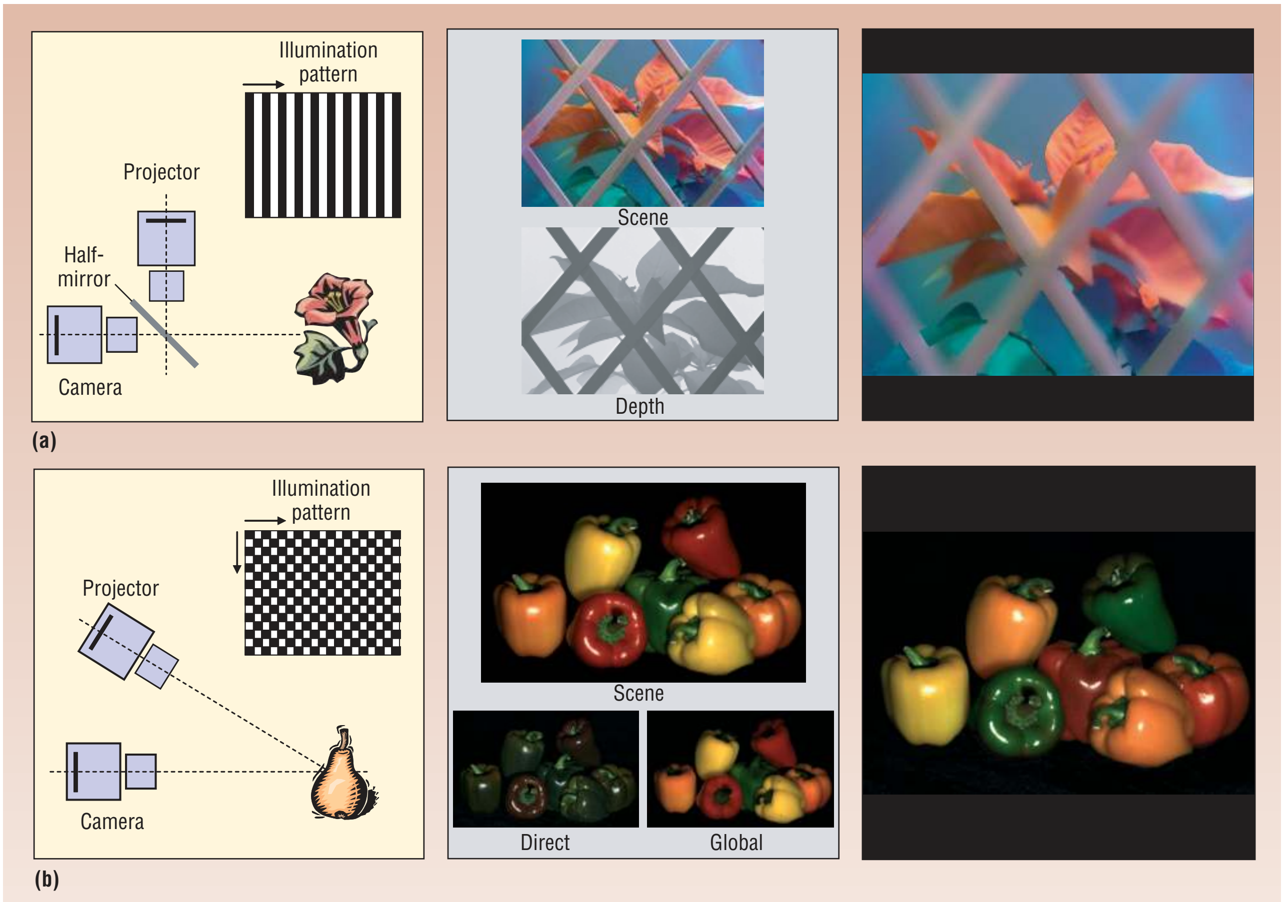


Figure 8. A projector can be used as a programmable camera flash to recover important scene information such as depth and illumination effects. Such information can be used to compute novel images of the scene. (a) Computing image-complete depth maps using projector defocus. (b) Separation of direct and global illumination using high-frequency illumination.

has two components: direct and global. The direct component results from light the point receives directly from the source, and the global component results from light the point receives from all other points in the scene. A programmable flash can be used to separate a scene into its direct and global components. The two components can then be used to edit the physical properties of objects in the scene and produce novel images.

The image on the left side of Figure 8b shows a scene captured using a checkerboard illumination pattern (inset image). If the checkerboard pattern's frequency is high, then the camera brightness of a point that is lit by one of the checkers includes the direct component and exactly half of the global component because the checkerboard pattern lights only half of the remaining scene points.

Now consider a second image captured using the complement of this illumination pattern. In this case, the point does not have a direct component but still produces exactly half of the global component. Since the above argument applies to all points in the scene, the direct and global components of all the scene points can be measured by projecting just two illumination patterns. In practice, to overcome the resolution limitations of the light source, it might be necessary to capture a larger set of images by shifting the checkerboard pattern in small steps.

The middle of Figure 8b shows separation results for a scene with peppers of different colors. The direct image includes mainly the specular reflections from the surfaces of the peppers. The colors of the peppers come from subsurface scattering effects that the global image captures. Altering the colors of the peppers in the global image and recombining it with the direct image yields a novel image, like the one shown on the right in Figure 8b. In addition to subsurface scattering, this separation method is applicable to a variety of global illumination effects, including interreflections between opaque surfaces and volumetric scattering from participating media.

Computational cameras use unconventional optics and software to produce new forms of visual information. This concept can be taken one step further by using controllable optics to realize programmable imaging systems that can change their functionalities

based on the needs of the user or the application. Finally, using a programmable illumination source as a camera flash offers many benefits. Ultimately, the success of these concepts will depend on technological advances made in imaging optics, image detectors, and digital projectors. If progress in these fields continues at the remarkable pace we have seen in the past decade, we can expect the camera to evolve into a more versatile device that could further impact the ways in which we communicate with each other and express ourselves. ■

Acknowledgments

The imaging systems described in this article were developed with support from NSF, ONR, DARPA, and the David and Lucile Packard Foundation. In particular, the author thanks Tom Strat, Tom McKenna, and Behzad Kamgar-Parsi for their support and encouragement. The systems described here were developed by the author in collaboration with Venkat Peri and Simon Baker (wide-angle imaging); Tomoo Mitsunaga (high dynamic range imaging); Yoav Schechner (generalized mosaicing); Sujit Kuthirummal (radial imaging); Vlad Branzoi and Terry Boult (programmable imaging with micromirror arrays); Assaf Zomet (programmable imaging with volumetric apertures); Li Zhang (projector defocus analysis); and Gurunandan Krishnan, Michael Grossberg, and Ramesh Raskar (separation of direct and global illumination). The wide-angle image was provided by RemoteReality Inc. The author thanks Anne Fleming, Gurunandan Krishnan, and Li Zhang for their useful comments. Technical details of the systems described here and several others can be found at www.cs.columbia.edu/CAVE.

References

1. B. Newhall, *The History of Photography*, The Museum of Modern Art, New York, 1982.
2. A. Gershun, "The Light Field," *J. Math. and Physics*, vol. 18, 1939.
3. M. Foucault, *Discipline and Punish: The Birth of the Prison*, Vintage, 1979.
4. K. Miyamoto, "Fish Eye Lens," *J. Optical Soc. of America*, vol. 54, no. 8, 1964, pp. 1060-1061.
5. P.L. Manly, *Unusual Telescopes*, Cambridge Univ. Press, 1991.
6. L.J. Hornbeck, "Deformable-Mirror Spatial Light Modulators," *Projection Displays III*, vol. 1150, SPIE, Aug. 1989, pp. 86-102.

***Shree K. Nayar** is the T.C. Chang Professor of Computer Science at Columbia University. His research interests include digital imaging, computer vision, computer graphics, human-computer interfaces, and robotics. Nayar received a PhD in electrical and computer engineering from the Robotics Institute at Carnegie Mellon University. Contact him at nayar@cs.columbia.edu.*