Editors: Lawrence Rosenblum and Michael Macedonia

Multiperspective Imaging

Steven M. Seitz and Jiwon Kim

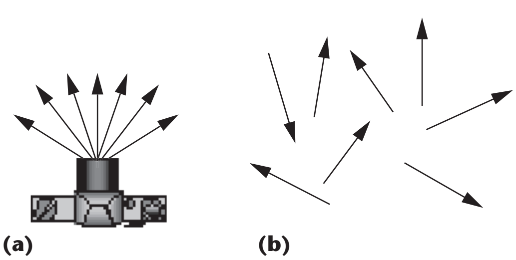
University of Washington

Our eyes have evolved with perspective optics. Because of this, perspective images seem somewhat natural to our eyes; they're well tailored for human vision. In a perspective image, the objects close to us appear large and in detail, yet we enjoy sweeping wide-range views of distant scenery.

Cameras have also evolved with perspective optics. It's natural for the optics of cameras to mimic the human eye—after all, a camera's primary function is to produce images that humans can interpret and enjoy.

However, our perspective has some unfortunate shortcomings. In particular, our eyes have a limited field of view, and we can only see the world in front of us. Ideally, we could see in all directions at once. Additionally, we can only see one side of an object at a time—for example, the front or the back. But suppose you could see all sides at the same time?

In the last several years, some researchers (including ourselves) have investigated techniques that capture multiple perspectives into a single image—a problem known as *multiperspective imaging*. Multiperspective images are useful for several reasons. The ability to capture a panoramic field of view or both the front and back of an object leads to richer and more complete visualizations. At the same time, these images are well suited for processing in computer vision problems such as stereo reconstruction and motion analysis. This article presents an overview of our work in this area, and our view of multiperspective imaging in general. References to additional research are available at http://grail.cs.washington.edu/projects/stereo/cga.htm and elsewhere.[1]



1 Set of rays corresponding to (a) a perspective image and (b) a multiperspective image.

Beginnings

Multiperspective imaging has a long and interesting background. Indeed, before the Italian Renaissance, virtually all paintings were multiperspective. Purposefully bending the laws of perspective is a common theme in modern art as well, for instance in the work of Picasso and Cezanne. A particularly striking example is M.C. Escher's *Print Gallery* (see http://escherdroste.math.leidenuniv.nl/).

Outside of art, multiperspective projections are common in cartography and in aerial and satellite-sensing applications. You can find a fascinating range of multiperspective optics in biological systems; perhaps the best-known example is the common house fly's compound eye. Studying these biological systems has inspired man-made devices, including a cosmic ray detector known as "The Fly's Eye" (see http://www.cosmic-ray.org/reading/flyseye.html).

Plenoptic function

An image captures light emanating from a scene in certain directions—that is, along a distribution of light rays. We may characterize an image based on which distribution of light rays it captures. In particular, a perspective image captures only the light in the scene that hits the focal point, as Figure 1 shows.

Other light ray distributions give rise to multiperspective images. Generally, we can define an image to be any 2D distribution of rays in space. A 5D function known as the plenoptic function $p(x, y, z, \theta, \phi)$ describes the set of all light rays. This function specifies each ray's origin (x, y, z) and direction (θ, ϕ).[2] The light along each ray is defined by additional parameters of wavelength λ and the time t at which point the light was sensed. The plenoptic function provides a mathematical framework for categorizing different varieties of images. In particular, we can represent any image as a 2D subset, or slice, of the plenoptic function.

Path images

How do you actually produce multiperspective images? Unlike perspective images captured with conventional cameras, producing multiperspective images requires specialized optical devices, arrays of conventional cameras, or moving cameras in special ways.

The easiest way to capture multiperspective images

Published by the IEEE Computer Society

is to move a regular video camera along a path and assemble the resulting image sequence into an x-y-t block of pixel data. The resulting pixel data is known as the *spatiotemporal volume*, or simply, the *video cube*. Once assembled, you can slice the video cube to produce different types of multiperspective images, as Figure 2 shows. We call these slices *path images*.

As a concrete example, Figure 3 shows a video cube created by pointing a camcorder out a car window and driving slowly down a residential street. The cube's left face is the last image of the input sequence, an x-y slice with a constant value of t.

The video cube's top face is an x-t slice, corresponding to $y = 1$. This image contains the first row of all of the input images, stacked one on top of the next. We generally refer to this as an *epipolar plane image* (EPI) in computer vision literature. Each scene point traces out a linear path in the EPI. Furthermore, the line's slope is proportional to scene depth, a useful property for image analysis.

Notice the cube's front face, a y-t slice containing the last column of all of the input images. This image—known as a *pushbroom image*—provides a panoramic view of the street. Although it looks similar to a perspective image, each column of a pushbroom image is acquired from a different point along the camera's trajectory. It therefore depicts a continuum of camera viewpoints.

We can create a pushbroom image from any column of the image—that is, any y-t slice. We can achieve an interesting effect by viewing all the y-t slice images as a movie sequence, in order of increasing x. The street scene appears to rotate in place from left to right. To see this movie, visit http://grail.cs.washington.edu/projects/stereo/cga.htm.

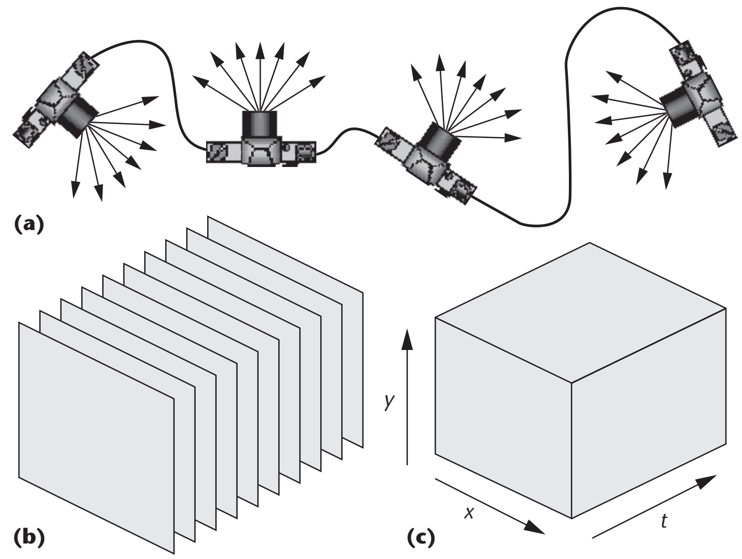
Pushbroom images yield superior visualizations of streets, landscapes, and other long linear scenes.

We can produce different types of video cubes and multiperspective images by moving a camera on a curved path instead of a line. For example, consider moving a camera in a circle around an object of interest, with the camera facing in toward the center of the circle. If the image is assembled into a video cube, y-t slices capture an inward-facing panorama of the object or scene within the circle.

Archeologists sometimes use these images (known as *cyclographs*) to create unwrapped views of ancient pottery. Traditional cyclographs are produced by photographing a rotating object through a narrow slit placed in front of a length of moving film—a technique that dates back to the late 19th century. We can simulate the same effect with a regular video camera, as we show in the "Multiperspective stereo" section.

Multiperspective stereo

Sometimes we can view two perspective images with the right characteristics stereoscopically. Our brain fuses the two images to produce a sensation of depth. Interestingly, the same is true for certain types of multiperspective images. For example, any two pushbroom images created from different y-t slices of the same video cube may be fused stereoscopically. Figure 4a (next page) shows an example of a stereo pushbroom image created in this manner from a longer version of the same sequence shown in Figure 3. It's displayed as an anaglyph, viewable using red-blue glasses.

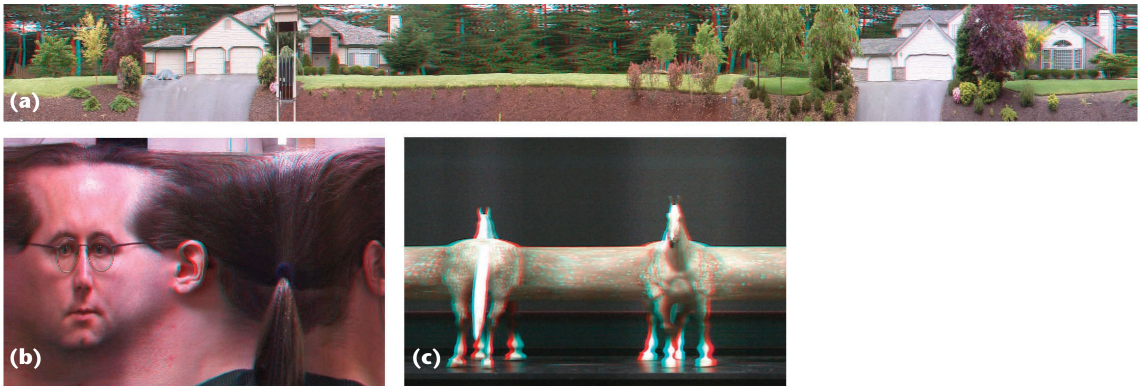


2 **(a) A camera moves along a path and captures light rays. (b) Stacking the images one on top of another yields (c) an *x-y-t* video cube. Each slice of the video cube produces a path image and represents a subset of the captured light rays (shown figuratively in red).**

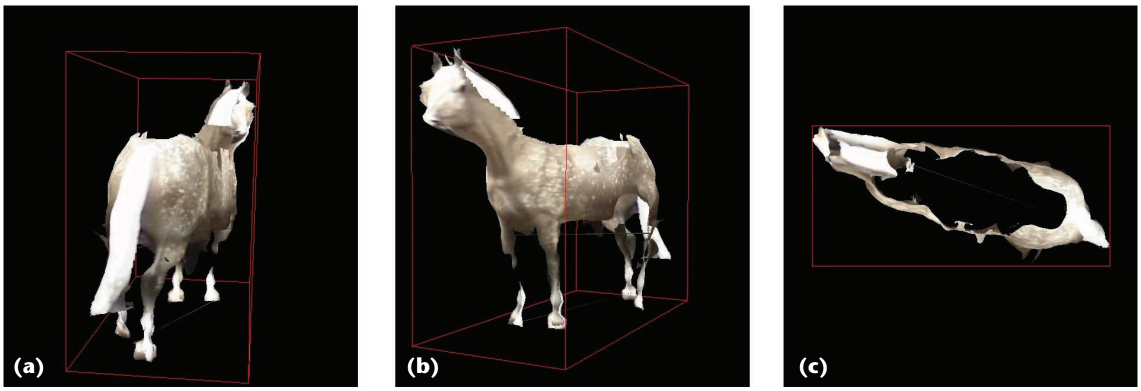
3 **Video cube captured by driving a car down a residential street with a camera pointed out the window.**

Stereo images may also be created by moving a camera on a circle instead of a line. If the camera is facing outward, the resulting images are often referred to as *stereo panoramas*. If the camera is facing inward, the results are *stereo cyclographs*. Figure 4 shows a stereo cyclograph anaglyph image created by moving a camera on a rotary arm around a person's head. Also shown is a stereo cyclograph of a toy horse, generated by rotating the horse on a turntable, in front of a stationary video camera. Note that the head and horse stereo cyclographs let you see both the front and back of the subject in the same image.

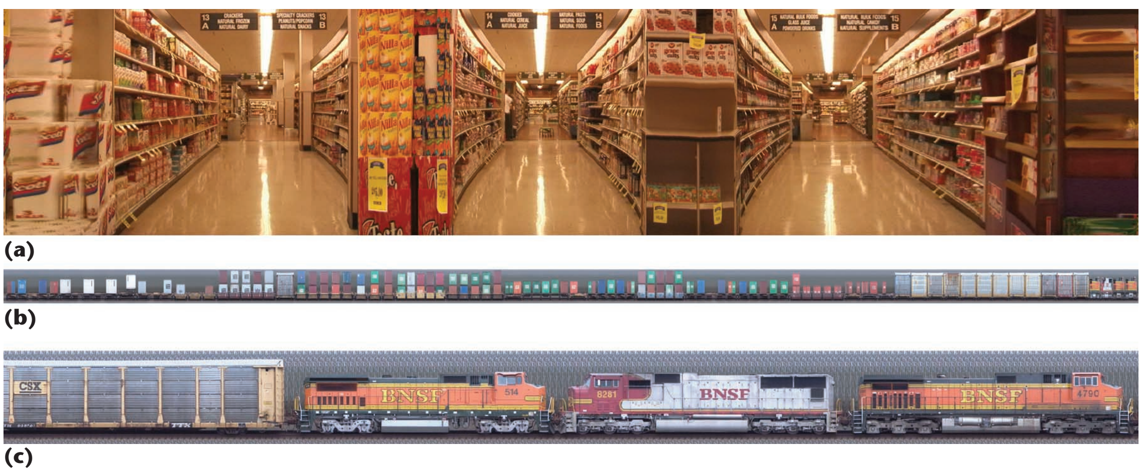
We can usually generate a stereo pair by moving a camera along any conic path—for example, a line, circle, ellipse, hyperbola, or parabola. For more informa-



4 (a) Stereo pushbroom of a residential street. Stereo cyclographs of (b) a human head and (c) a toy horse. All are 3D viewable with red-blue glasses.



5 Renderings of a 3D model reconstructed from the horse cyclograph stereo pair in Figure 4: (a) front, (b) back, and (c) top-down view.



6 (a) Multi-perspective view showing three aisles of a supermarket at once. (b) Strip image of a train, horizontally compressed to fit on this page. (c) Expanded view of four train cars.

tion on multiperspective stereo images and how to create them, see our related article.[1]

Beyond their use for 3D visualization, stereo images also enable 3D measurement and reconstruction using computer vision algorithms. Traditional stereo matching algorithms operate on perspective images. However, we can easily adapt and apply the same techniques for multiperspective stereo pairs. Figure 5 shows a texture-mapped mesh model reconstructed from the horse stereo pair in Figure 4. Observe how the front, back, and both sides of the horse are reconstructed from a single stereo pair—a capability not possible with perspective images. The top-down view (see Figure 5c) is hollow, since the top of the horse wasn't visible.

Looking ahead and all around

So far, we've only considered axis-aligned planar slices of the video cube—that is, *x-y*, *y-t*, or *x-t* slices. To see the effects of other planar slices, we recommend downloading the video cube application (available at http://research.microsoft.com/downloads/VideoCube/VideoCube.asp). The application lets you view any video as an *x-y-t* cube and slice it interactively.

Nonplanar slices enable other visualization types. We've developed an interactive tool that lets users specify any vertical (composed of columns from the input images) video cube slice and display the result as a multiperspective image. Users specify slices through two mechanisms. The first option is to draw a curve in the *x-t* plane, specifying what the slice looks like from the top down. The second option is to click on regions from a set of input images that should be included in the panorama. The tool interpolates these samples via an optimization procedure to produce a smooth slice through the video cube. Figure 6a shows an image of a supermarket (created using this tool) in which the contents of three aisles are visible at once. We captured the input sequence by mounting the camera on a shopping cart and rolling it in a straight line in front of the aisles.

We can also apply these techniques to moving scenes. For example, Figure 6b shows a pushbroom-like image of a moving train, captured from a stationary video camera. David Dewey created this image by taking a narrow vertical strip from the center of each image and compositing them. Because the scene background doesn't move, it's repeated in each image and gives rise to the texture pattern seen in Figure 6c's background.

There's still much room for improvement and growth in the area of multiperspective imaging. While it's better suited than perspective images for stereo processing, problems still exist. For example, the best way to efficiently capture multiperspective images using specially designed sensors or arrays of cameras is under debate and remains an important and active topic of research.

The images shown in this article are only examples and not representative of the full range of image varieties. Researchers are still investigating the range of images we can create as well as identifying their practical uses. We believe that multiperspective images will have promising applications to a wide range of computer vision and visualization problems. ■

Acknowledgments

Kiera Henning and David Salesin helped us develop the tool used to create the supermarket image shown in Figure 6a. We thank David Dewey for providing the train images shown in Figures 6b and 6c.

References

1. S.M. Seitz and J. Kim, "The Space of All Stereo Images," *Int'l J. Computer Vision,* vol. 48, no. 1, 2002, pp. 21-38.
2. E.H. Adelson and E.H. Bergen, "The Plenoptic Function and the Elements of Early Vision," *Computation Models of Visual Processing*, M. Landy and J.A. Movshon eds., MIT Press, 1991.

Readers may contact Steven M. Seitz and Jiwon Kim at {seitz, jwkim}@cs.washington.edu.

Readers may contact the department editors by email at rosenblu@ait.nrl.navy.mil or michael_macedonia@ stricom.army.mil.