

EYES, CAMERA, ACTION!

TO IMPROVE PHOTOGRAPHY, ENGINEERS ARE DESIGNING CAMERAS THAT



BEHAVE MORE LIKE EYES BY KLAUS M. STIEFEL AND ALEX O. HOLCOMBE

In a clearing in a subtropical rain forest in northern Australia, you can watch the light dance as it filters through the rustling canopy. Below, the leaves of the bushes form an intricate pattern of shadows on the trunks of trees. A wallaby grazes in the open space. You raise your smartphone and aim it at the tranquil marsupial. Just as you tap the button to take its picture, the wallaby notices you and hops away. In the image on your screen, half of the snapshot is too dark to make out details, and the sky between the treetops looks bleached white. The hopping wallaby is a blurry, small blob near the center of the photograph. Zooming in on the animal exposes an almost Cubist field of pixels, his outline visibly broken up into the smallest squares of the camera's sensor.

For any of us who snap photos, whether with a tap of the screen or by holding up a professional-grade piece of equipment, the experience described above—if perhaps not the wallaby—will be a familiar one. The proliferation of smartphones has turned nearly all of us into amateur shutterbugs. According to a Pew Research Center survey, more

motion blur *ever* happens when you use your eyes. So where is the app that turns your smartphone camera into the equivalent of your eye? Engineers are now working on just that. By designing cameras that mimic the ways in which evolution has solved the image-creation problem in the brain, they hope to improve the quality of our personal photos. But there is more. With better cameras, we will have robots that can independently and smartly navigate the world and security cameras that recognize, as a human can, when a person is in trouble and can swiftly dispatch help. As we view things more and more through the eyes of computers, so, too, will our computers learn to see more like humans do.

To understand how this technological innovation is coming about, we have to first understand how the eye does its inimitable job—and where cameras fall short.

The Nature of Exposure

A glaring weakness of cameras is their inability to handle high and low lighting conditions in a single shot. In rare circumstances, our eyes also encounter this problem. When emerging from a dark basement into full sun, for example, we speak of being “blinded by the light.” This transient moment, from which our eyes quickly recover, is one of the few instances in which our eyes can be said to suf-

AS WE VIEW THINGS MORE AND MORE THROUGH THE EYES OF COMPUTERS,

than half of all U.S. Internet users post original photos online. Instagram, the popular sharing service, reports that some 55 million pictures are posted to its network daily—that's 38,000 a minute. Yet not a single one of those millions on millions of images comes anywhere close to capturing the vivid, rich world we experience with our eyes.

None of the problems of exposure, pixelation or

fer from overexposure. Historically, English did not even have a word for overexposure, because our vision has been peerless in its ability to avoid the problem. It took the invention of cameras for the concept of an inappropriately lit image to emerge.

The reason is dynamic range. It is the difference between the lowest and highest light intensity that our eyes or a camera can register. Light comes in tiny packages, called photons, that race around the universe at—you guessed it—light speed. But they do so at different energy levels. High-energy photons are perceived as blue, and those with much less energy look red. When photons collide with matter, they can get rerouted or absorbed. For example, water molecules selectively absorb low-energy photons, which is why water appears blue. A solid dark wall absorbs nearly all the photons hitting it and turns their energy into minuscule bits of heat, which explains why a wall can sometimes feel warm to the touch. More exotic materials absorb photons, and

FAST FACTS

PICTURE PERFECT

- 1 Cameras encounter problems that eyes never face in producing images, such as incorrect exposure, motion blur and insufficient resolution.
- 2 To address these shortcomings, engineers are now designing camera technologies that mimic the human eye.
- 3 A retina-inspired chip, for example, could pave the way not only to better cameras but also to smarter robots.



Emerging from darkness into full sun is one of the rare cases when our eyes struggle with light exposure. Cameras, however, frequently produce underexposed and overexposed images.

instead of emitting heat, they amplify that energy into signals that are useful to cameras and brains.

In a digital camera, the photon-absorbing objects are called photodiodes. A photodiode is equivalent to a pixel, so the more photodiodes a camera has, the higher the picture's quality. This device, often made of silicon, is simply a light detector. When a photon hits it, the particle knocks an electron in the silicon to a higher energy level. The resulting

ways in cell membranes slamming shut, and the slowing of the flow of glutamate, an amino acid. All this squishy biological machinery amplifies the infinitesimal energy of a photon enormously, producing a signal strong enough to drive neurons.

In fact, the amplifying power of the retina is so immense that in a completely dark room, a light source need only emit about five photons for you to perceive it. To achieve this level of sensitivity, our

SO, TOO, WILL OUR COMPUTERS LEARN TO SEE MORE LIKE HUMANS DO

charge excites the electron, causing electricity to flow. A semiconductor chip amplifies the electrical signal from every photodiode.

The brightest light a Canon 5D II—a top-of-the-line single-lens reflex camera—can discriminate is 2,000 times stronger than the weakest light it can sense. If a scene's luminance exceeds this range, overexposed and underexposed image regions occur, and photographic shame ensues. But if you had looked with your eyes instead, the same photon would have hit your retina. More precisely, it would enter a cell in your retina called a photoreceptor and excite an electron. The particle in question sits inside a retinal molecule (a form of vitamin A), which is part of a protein in the photoreceptor cell.

Tickled by the excited electron, the retinal molecule starts to twist, which in turn triggers its encompassing protein to change its configuration. This shape shifting kicks off a chain of downstream effects, involving other proteins morphing, gate-

eyes have evolved a special type of supersensitive photoreceptor dedicated to dark, nightlike conditions. These so-called rods, although they are used only in the dark, are 20 times more numerous than the cone-shaped photoreceptors we use during the day. Vision at night was apparently very important in our evolutionary history because including all those rods does not leave much room for our cone-shaped, daytime receptors.

The two kinds of photoreceptors together allow us to register an enormous range of light levels. Yet even without the nighttime receptors, our eyes operate over an incredible range. If you work late in a brightly lit office, you may look out the window wistfully as the sun sets and the trees become dark silhouettes. Yet still you can see objects outside and things inside your brightly lit office simultaneously. The range of light levels to which your eye is sensitive is so vast that it can differentiate between two objects, one of which is a million times brighter than the other.

YOUR EYE IS SO SENSITIVE TO LIGHT THAT IT CAN DIFFERENTIATE BETWEEN

During the fraction of a second that it takes for a camera's shutter to snap, this biker will have moved several centimeters. The camera's photodiodes will capture light during that entire span of time.



The advantage lies in the fact that every photoreceptor has its own exposure setting, which is constantly changing in response to the level of light received. To mimic the range of the eye, some cameras can now combine several exposures taken in quick succession. An overexposed shot provides a properly lit view of the dark parts of a scene, and an underexposed shot captures bright parts, such as the sky. Fused together, these too bright and too dark photographs produce an image with a range larger than what is possible with any individual shot. The trick fails when photographing fast-moving objects because they change position between the different exposures, but it works well for landscape photography. Even if your camera does not have a built-in high-dynamic range function, you can fuse several images post hoc on your laptop to achieve a compound image devoid of overexposed and underexposed areas.

Caught in the Act

Let's return to the hopping wallaby and why it turned out blurry. One of the problems is that a cam-

era's shutter speed is only so fast (say, one fiftieth of a second), so a photograph will capture the light during that entire span of time, during which the wallaby's body traveled several centimeters. Our visual system is no quicker, so the image created by our photoreceptors also is blurred. Yet somehow we do not perceive much blur.

After light hits the retina, several specialized types of neurons, which connect neighboring photoreceptors, modify the light signals before sending them on to the brain. Some of these neurons react to movement in a certain direction, others to bright signals surrounded by darkness, and so on. Together they allow the eye to adjust its sensitivity.

Ultimately your visual system is most interested in change. The eyes move constantly, altering the amount of light impinging on your photoreceptors and maintaining your image of the world. If your eyes are kept still, the lack of change in a scene will cause the retina to stop signaling, and objects will begin to fade away. Swiss physician Ignaz Troxler first noticed this phenomenon in 1804. A bias toward change helps to emphasize new data over old. And it is a neat trick for overcoming the imperfections of the optical apparatus. For example, this change bias is the reason we never see the blood vessels in the eye, which sit between the outside world and our photoreceptors.

Although this trick has yet to be incorporated into consumer cameras, an experimental camera that has been developed by Tobi Delbrück of the In-

THE AUTHORS

KLAUS M. STIEFEL is an underwater photographer and a researcher at the University of Western Sydney, where he studies brain cells and their networks. He is also author of *Sex, Drugs and Scuba Diving*, a popular science account of marine biology, underwater photography and diving.

ALEX O. HOLCOMBE, an associate professor of psychology at the University of Sydney, investigates perception.

TWO OBJECTS, ONE OF WHICH IS A MILLION TIMES BRIGHTER THAN THE OTHER

stitute for Neuroinformatics in Zurich illustrates an extreme form of change bias. This camera's chip does not simply record the amount of light hitting every pixel, as a standard camera does, but relies on changes in the light intensity. The image that this camera creates is essentially a record of the movement and change that occurred while the picture was being taken. Pixels that increase in intensity appear white, whereas lessening intensity shows up as a black pixel. If a pixel does not change from moment to moment, the image shows only a bland gray pixel. This emphasis on change ignores stationary, unchanging objects to help isolate moving ones.

Graduate student Greg Cohen of the University of Western Sydney (a colleague of Stiefel) is working with this retina-inspired camera chip to create a robot that can play Ping-Pong, a game that is all about change and motion. In Ping-Pong, the opponent, his paddle and especially the ball move at astonishing speeds. Not all the information in a Ping-Pong scene helps in hitting the ball back across the table, such as the window behind an opponent or the patterns on the floor. The retina-inspired camera's feature of ignoring static objects helps with the task, allowing the robot to concentrate on detecting and responding to motion. Playing Ping-Pong requires such brilliant hand-eye coordination that success at this task may lead to solutions useful for a variety of applications, such as care for the elderly or search-and-rescue operations.

Saving Face

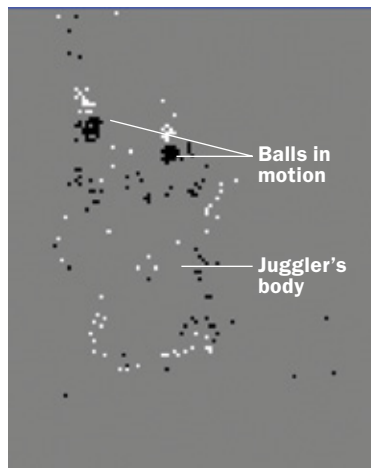
Although the retina takes care of the first steps in seeing, much more processing occurs in the

brain. For example, we rapidly appreciate a photo when our brain can easily separate the main subject from its background. Skilled photographers know how to make that task easy for the brain, for instance, by putting one person's face in focus while limiting the depth of field so that the background is blurred. Faces are a special class of objects for us. In a busy visual scene, the human gaze will preferentially seek them out. A photograph in which they are blurred is almost always considered a ruined shot.

Several brain areas contribute to our ability to process faces. When a visual signal leaves the retina, it travels to a part of the brain called the thalamus. The thalamus is a sophisticated relay station en route to the cortex, the tightly folded mantle that makes up the brain's surface. A number of patches of cortex help us process what we see. The primary visual cortex is a large piece of real estate at the back of the brain where most signals leaving the thalamus end up. From there the information about our visual world travels to several additional visual regions of the cortex. Of these, various small areas in the temporal cortex (located on the sides of the brain) react very specifically to seeing faces.

Camera makers have begun to implement something akin to our brain's ability to recognize and prioritize faces. Many of today's cameras, even simple point-and-shoot ones, recognize faces in their field of view. This is typically done with an advanced statistical method known as the Viola-Jones algorithm. In brief, the camera's chip filters the image for basic features such as edges and corners. Region by region, it then runs a series of tests to look for facial

The Troxler illusion (seen at left) illustrates the way neurons cease to respond to an unchanging stimulus. Focus on the dot in the middle the circle. After several seconds, the circle will fade away. At the right, a retina-inspired camera was used to create an image of a man juggling. The dark spots in the upper part of the image are the locations that the balls are moving away from, and the white dots are the locations they are moving to.



To help put a face in focus, a photographer might limit the depth of field, rendering the background blurry. Cameras replicate that capability by using algorithms that search a scene for facelike features.



features. For example, it would look to see if a bright spot (a nose) occurs between two darker spots (the eyes). Only if part of the image passes all these tests does the algorithm decide that it is seeing a face. Now the camera can make sure to keep that visage in focus.

Most likely, the brain's method of processing faces differs considerably from the Viola-Jones algorithm. Hence, the face-recognition algorithm in modern cameras is not a software implementation of the brain's way of recognizing faces but rather a different solution to the same problem. By pairing such advances in image processing with knowledge about human visual preferences, we can greatly improve the photographs we produce.

Megapixels on the Mind

The face-selective areas of the cortex are only a small subset of the brain territory devoted to vision. Other sections of it react to different aspects of the visual scene, such as color, motion and orientation. This hubbub of activity culminates in the visual world we perceive around us.

The coordinated efforts of these brain areas are the reason why in real life you never see anything coarse-grained the way you do when you zoom in on a photo. Increasing the number of megapixels (MP) in a camera cannot solve this problem. The first digital camera Stiefel proudly owned had a 2-MP sensor, yet these days even most smartphones have at least double that. We can continue cramming in more pixels—advances in manufacturing will very likely miniaturize the hardware even further—yet it will remain the case that blowing up a smooth-seeming image will even-

tually transform it into a mess of boxy colors.

This limitation arises when two neighboring photons strike the same photodiode, meaning their energy will be combined into a single pixel. At that point, the information about their exact original locations is lost forever. Unfortunately, no image-processing software can create more meaningful pixels. You can scale up the size of your digital photograph, but the newly created pixels will have no new information about the light entering your camera when you pressed the shutter. Further, the scaling is not as big as you might think. The pixels of a 16-MP camera are only twice as small as a 4-MP camera. The human retina, in contrast, contains only about 6 million functioning daylight photoreceptors (cones)—just 6 MP.

In essence, our brain constructs a percept of what it evolved to regard as reality—and the human brain does not consider the graininess of the human retina a feature of external reality. What we perceive is a construction, a masterful portrait that involves much filling in between our individual sensors. There are no such things as pixels in our percepts—our brain does not to reproduce an image of light piece by piece, as if it were a biological super-camera. Rather the brain synthesizes a coherent impression for a specific purpose—that of allowing us to find our way through the world. The principle of the eye and that of the camera are fundamentally different. Unless, in a far-off future, we develop truly intelligent machines and put one in a camera body, that difference will not be bridged.

Nevertheless, the possibilities available to engineers continue to increase, together with better understanding of the eye and brain. Combining these with a little creative thinking should yield many more exciting advances in camera technology. **M**

FURTHER READING

- **The First Steps in Seeing.** Robert W. Rodieck. Sinauer Associates, 1998.
 - **Science for the Curious Photographer: An Introduction to the Science of Photography.** Charles S. Johnson, Jr. A. K. Peters/CRC Press, 2010.
 - **Basic Vision: An Introduction to Visual Perception.** Revised edition. Robert Snowden, Peter Thompson and Tom Troscianko. Oxford University Press, 2012.
- From Our Archives*
- **Shifting Focus.** Susana Martinez-Conde and Stephen L. Macknik; November/December 2011.
 - **Cracking the Retinal Code.** Christof Koch; January/February 2013.
 - **To See or Not to See.** Michael C. Hout and Stephen D. Goldinger; July/August 2013.

Zooming in on the image on page 53 reveals the pixelation that inevitably afflicts photographs. Our visual system does not create pixelated images, because our mind has evolved to fill in the gaps.

