Image Sensors

Suren Jayasuriya
School of Arts, Media and Engineering and
School of Electrical, Computer and Energy
Engineering, Arizona State University, Tempe,
AZ, USA

Synonyms

Camera sensor

Related Concepts

▶ Camera
▶ High Dynamic Range Imaging
▶ Photon Shot Noise
▶ Rolling Shutter
▶ Saturation

Definition

Image sensors are devices that produce an electronic signal proportional to the amount of light impinging on the device, typically arranged in two-dimensional pixel arrays. They can read out the image using on-chip amplification and analog-to-digital conversion and are typically controlled using electronic shutters.

Background

Materials and devices to sense light have a long history from antiquity to the present day. While camera obscura or pinhole cameras were being developed for centuries to focus light onto the image plane, only with the invention of the daguerreotype were images captured directly on plates in the late 1800s (Note: there were some camera precursors like heliography, but daguerreotypes were the first popular cameras). These plates would consist of polished metal such as silver exposed to photosensitive chemicals such as Bitumen of Judea, iodine, or halogen fumes, and after exposure with the scene, the resulting latent image was developed using mercury vapors. The resulting images had high spatial resolution but required dangerous chemicals and bulky equipment to be carried around with the camera. Later research was able to transfer image capture from plates to films containing photographic emulsions which are light-sensitive. Analog photography, utilizing film, emerged as the popular camera technology for most of the twentieth century. We refer the reader to Chapter 16 of [1] for more details about the history of photography.

The rise of digital electronics led to the use of silicon-based sensors for capturing images. CCDs (charge-coupled devices) were invented in 1969 at Bell Labs and featured an array of gate capacitors which are doped in silicon to be photoactive when a positive voltage is...
Image Sensors, Fig. 1 (a) The structure of a back-side illuminated photodiode consisting of microlens, color filter, photodiode, and circuitry underneath to read out the signals and (b) a 4T pixel architecture consisting of a reset transistor for the photodiode, a transfer gate to transfer charge to the pixel amplifier (source follower), and a read transistor that selects the pixel voltage to travel down the column for amplification and digitization.

applied [2]. Incoming light accumulates electric charge proportional to the light irradiance, which is then transferred using a bucket-brigade readout scheme to the edges of the array where it is either transmitted in analog or digitized. CCDs become widely available in cameras, particularly for scientific applications where their low noise properties were beneficial.

However, the majority of image sensors used presently are made from CMOS technology used for fabricated integrated circuits. Work in the Jet Propulsion Laboratory in the early 1990s led to the development of the CMOS image sensor, which consisted of an array of photodiodes that contained a small amplifier per pixel, and column electronics including further amplification and ADCs [3]. CMOS image sensors quickly grew in market share due to their ability to integrate electronics on the same silicon die as the photodiode array, and was mass-manufacturable in the CMOS process. Today, most cameras from DSLRs to cellphones utilize CMOS image sensors as their main image sensor of choice.

Theory

The theory behind image sensors is rooted in semiconductor physics, namely, optoelectronics, as well as mixed-signal circuit design for the readout, amplification, and analog-to-digital conversion. In this section, we briefly discuss these fundamentals and identify commonly used metrics for image sensor performance. We refer the reader to [4] for a more comprehensive introduction to image sensors.

After light arrives at a photodiode, it typically goes through a microlens and color filter before reaching the photodiode, as shown in Fig. 1a. Microlenses are small lenses on the order of microns (μm) that are fixed on top of the silicon photodiode to focus light. After passing through the microlens, light is sometimes filtered by a color filter, typically for either red, green, or blue visible wavelengths which are spatially arranged over the sensor in a Bayer pattern. This allows color spatial subsampling of the image with each pixel sensing only one of those three primary wavelengths, and this subsampling is corrected in a process known as demosaicing in image processing. Finally the light reaches the photoactive silicon region, which is typically a p-n junction of doped silicon.

The photoactive region itself exploits the photoelectric effect (which Albert Einstein discovered in 1905 and for which he received the Nobel Prize in physics in 1921) where light displaces electrons in the semiconductor lattice and the corresponding charge carriers are swept across a diode junction to generate a photocurrent. Photodiodes can operate in photovoltaic mode with no bias across their diode, or more typically in photoconductive mode where the diode is reverse biased to increase the responsivity to light. The
photocurrent is accumulated on a capacitor to aggregate charge proportional to the number of photons per second per meter squared (equivalent to irradiance). To expose the photodiode, typically one needs to reset the photodiode to its initial reverse bias by applying an external voltage across its terminals before floating the diode and allowing the photocurrent to accumulate charge. After the charge has been accumulated, it is typically read out of the pixel using an in-pixel amplifier (typically known as a source follower in CMOS) or using a bucket-brigade readout in CCDs. A 2D spatial array of pixels are arranged in row-column format, and commonly one row is read out at a time in column-parallel fashion. Common pixel architectures include the 3T or 4T transistor topology shown in Fig. 1b. To expose and read out the entire array, either a rolling or global electronic shutter is used. In rolling shutter, each row is reset, then exposed, and read out in the timing of the sensor, which means that the entire image is exposed at slightly staggered times (corresponding to different rows), as shown in Fig. 2. Global shutter exposes the entire image sensor at the same time but requires in-pixel memories to store the charge while readout occurs or specialized circuits to do full array readout simultaneously.

After reading out each column signal in parallel (corresponding to one row), the voltage is usually amplified with a column amplifier and then passes through an analog-to-digital converter (ADC). This results in 8–16 bit images depending on the image quantization used, and the image readout is known as the RAW image. After this, the image is usually sent to an image signal/sensor processor (ISP) to perform image processing such as demosaicing, denoising, white balancing, color correction, tone mapping, and compression.

Image Sensor Metrics

There are several key metrics that are important for image sensor performance. Quantum efficiency, $QE(\lambda) = N_e(\lambda)/N_p(\lambda)$ where $N_p(\lambda)$ is the number of incident photons for a specific wavelength and $N_e(\lambda)$ is the number of electrons produced, determines the sensitivity of the image sensor with higher efficiency meaning less wasted photons.

Various noise sources for the sensor include photon shot noise due to the physics of light collection and read noise of the sensor. Following a model formulated in [5], photon shot noise average power is given by $\sigma_{\text{shot}}^2 = q(i_{ph} + i_{dc}) \cdot t_{\text{int}}$ where $i_{ph}$ is the photocurrent, $i_{dc}$ is dark current (a type of noise), $t_{\text{int}}$ is the integration time, and $q$ is the charge of an electron. Read noise power is given by $\sigma_r^2 = \sigma_{\text{reset}}^2 + \sigma_{\text{readout}}^2 + \sigma_{\text{FPN}}^2$ where reset is the noise of resetting the photodiode, readout is the noise of all the electronics, and FPN is fixed pattern noise due to non-uniformities in the sensor. From these quantities, the signal-to-noise ratio is given by the following equation [5]:

\[
\frac{S}{N} = \frac{\int \sigma_{\text{shot}}^2 \cdot \gamma^2 \cdot \lambda}{\int \sigma_r^2 \cdot \gamma^2 \cdot \lambda}
\]
Image Sensors

\[ SNR(i_{ph}) = \frac{(i_{ph} t_{int})^2}{q(i_{ph} + i_{dc} t_{int} + \sigma_r^2)}. \]  

(1)

The dynamic range of the pixel is the range of light intensities that can be represented by a single pixel (or equivalently the amount of charge that can be held by the photodiode). Dynamic range is defined as the ratio of the largest non-saturating input signal to the smallest detectable image signal. This is given by the following formula [5]:

\[ DR = \frac{i_{max}}{i_{min}} = \frac{q_{max} - i_{dc} t_{int}}{\sqrt{q i_{dc} t_{int} + \sigma_r^2}}. \]  

(2)

where \( q_{max} \) is the maximum charge the photodiode can hold. Note that \( i_{max} \) increases as the integration time decreases and \( i_{min} \) decreases as integration time increases. Thus, high dynamic range imaging needs short integration times for bright illumination, and long integration times for dark illumination.

Spatial resolution of the image sensor is partially governed by the pixel pitch of the sensor, with most modern CMOS image sensors being less than 1 μm in pitch (spatial resolution of the image is a complex formula of both pixel pitch but also the entire optical system’s modulation transfer function (MTF)). Finally, the frame rate of the sensor (fps) is typically set by the exposure time and the readout timing and synchronization of the array. All these metrics are optimized for performance in modern image sensors.

Application

Images and video have become ubiquitous in the twenty-first century, generating content that is uploaded daily to the Internet, shared on social media, and used in a host of different applications. Image sensors are currently reaching tens of megapixels in resolution, frame rates that far exceed 60 frames per second, and are embedded on many different platforms. For photographers, high-end image sensors are found in digital single-lens reflex (DSLR) cameras, yet most typical pictures that are uploaded online are taken using image sensors on mobile devices and smartphones.

In fact, one main application of image sensors is the entire field of computer vision, the focus of this reference guide. Machine vision cameras have been developed to optimize image sensors for application-specific demands. Machine vision cameras typically offer both global and rolling shutter, fast interfaces for data transfer including Ethernet and camera-specific protocols, and have the ability to output RAW images or configure the ISP. These cameras are commonly used for computer vision methods that require accurate photometric measurements from the scene.

In addition to traditional CMOS and CCD image sensors, image sensors have been customized for various other imaging regimes. Image sensors for non-visible wavelengths including infrared and terahertz domains have been developed. These typically use semiconductor materials other than silicon (as silicon is only sensitive to visible and near infrared), such as III–V materials. Such image sensors have applications in remote sensing, surveillance, and biomedical imaging where these non-visible wavelengths yield additional spectral information from the scene.

Open Problems

There are several new frontiers for image sensors and their applications. Computational imaging and photography is a new field of research which co-designs optics, sensor hardware, and algorithms to capture new visual information. New sensors for computational photography include pixels for high dynamic range imaging [6], light field imaging [7, 8], polarization imaging [9], and compressive sensing [10]. Sensors have been developed to support coded exposure and readout [11] for motion deblurring, as well as support arbitrary region-of-interest readout [12] for selective imaging. Other computational image sensors include on-board image gradient calculation [13].

New pixel designs can expand the possibilities of visual data that can be captured on board
the sensor. Neuromorphic silicon retinas are able to sense edge gradients and motion information directly through novel silicon detectors that emulate neurons [14]. Event-based sensors (also known as dynamic vision sensors) utilize pixels which output binary or trinary pixels when the pixel changes value temporally and can operate at extremely low powers and extremely fast frame rates [15]. These have been used for a host of applications in robotics, optical flow, and simultaneous localization and mapping.

Time-of-flight sensors have seen widespread application for depth sensing for computer vision. These typically feature either photogates [16] or single-photon avalanche diodes (SPADs) which operate the photodiode in avalanche mode to be sensitive to a few photons at a time [17]. SPADs have been used widely in LIDAR systems and for emerging applications in non-line-of-sight and transient imaging. Finally, as CMOS technology shrinks to smaller sizes, the ability to sense individual photons has become possible. Quanta image sensors using jot pixels are able to achieve single-photon sensing without the use of avalanche gain, which allows for megapixel resolutions (1.1 μm pixel pitch) and 1000+ fps frame rates [18]. These new image sensors will open up new avenues for research in computer vision as these devices become more available and widespread.

References