CMSC 491/691: Computer Vision **Spring 2024**

Lecture 2: 2024-01-31 Image Formation and Acquisition

Lecturer: Tejas Gokhale *Scribe: Bryan Yang*

2.1 Recap of Last Lecture

In the last lecture, we discussed the capabilities of pinhole imaging. Key components involved are the object to capture, the barrier (or diaphragm), the pinhole (or aperture), and the sensor (or image plane). The resulting captured image can be different depending on the attributes of these components. For example, changing the pinhole diameter will change the level of sharpness of the resulting image (as it alters the amount of light that passes through the diaphragm), while changing the distance (or focal length) between the diaphragm and the sensor will change the magnification of the resulting image.

While pinhole imaging carries the fundamental properties that cameras have, its shortcomings lie in the fact that the diameter of the aperture must be small enough to acquire a clear image, while also being large enough to prevent diffraction.

The optimal image is obtained when the diameter of the pinhole, d, satisfies the equation:

$$d = 2\sqrt{f'\lambda} \tag{2.1}$$

where:

- d represents the diameter of the pinhole,
- f' is the effective focal length of the imaging system,
- λ stands for the wavelength of light used to capture the image.

This relationship balances the effects of diffraction and sharpness, resulting in the sharpest possible image under given conditions.

2.2 The Thin Lens Model

When we introduce a lens into optics, we can have more precise manipulation of light. Lenses map "bundles" of rays from points on the scene to the sensor, focusing them to form clear images. However, due to the complexity of this process, the thin lens model is used as a simplification.

By assuming the lens is infinitely thin, light rays entering the lens from a point source are refracted (bent) and converge at a point on the other side of the lens. The rays passing through the lens center are unaffected, traveling straight without deviation, while the rays entering the lens parallel to its axis converge to a single

point on the focal plane, which is determined by the lens's focal length. The model is described by the Gauss' lens equation:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} \tag{2.2}$$

where:

- f is the focal length of the lens,
- d_o is the distance from the object to the lens,
- d_i is the distance from the lens to the image formed by the lens.

Depth of field (DoF) ties closely to this model, referring to the zone within which objects appear sharply focused. The DoF is significantly affected by the aperture size; a smaller aperture increases the DoF, bringing a wider range of the scene into focus. In photography, this principle is used to allow for either isolated subject focus or broad scene clarity.

Similarly, the field of view (FoV) is another critical concept influenced by lens properties, specifically the focal length. A shorter focal length enlarges the FoV, offering a wider perspective of the scene. In photography and cinematography, the choice of focal length can impact composition and storytelling.

2.3 Color Theory

Color theory provides a framework for how colors are perceived, created, and applied in digital imaging and computer vision.

Color can be characterized by three main properties:

- **Hue**: The name or type of color (e.g., red, blue, green) as perceived in the color spectrum.
- **Saturation** (Chroma): The intensity or purity of a color. High saturation colors are vivid, while low saturation colors appear more muted.
- **Value** (Brightness/Lightness): The lightness or darkness of a color, ranging from light (near white) to dark (near black).

2.3.1 Color Perception

For humans, color perception is limited to a narrow band of the electromagnetic spectrum, with wavelengths from approximately 380 to 720 nanometers, encompassing all the colors visible to the average human eye. From a biological perspective, the human eye possesses three types of cone cells, each sensitive to different segments of the visible light spectrum (short, medium, and long wavelengths). These cells enable the perception of a wide range of colors. From a cognitive perspective, light is converted into signals by the retina upon reaching the eye and transmitted to the brain through the optical nerve, resulting in the sensation of color.

2.3.2 Spectral Power Distribution

The Spectral Power Distribution (SPD) of a light source is a concept that provides insight into the composition of light that the source emits. The reality is that most light sources in our environment emit light that comprises a complex mixture of various wavelengths and it is rarely the case that monochromatic light (which consists of a single wavelength) is ever seen in real-world scenes. This mixture is what gives light its characteristic color and intensity. The SPD represents this complexity by detailing the distribution and intensity of these wavelengths across the visible spectrum.

A SPD plot maps out how much of each wavelength is present in the light emitted by a source, from the shortest wavelengths at one end of the visible spectrum to the longest at the other. This representation helps in understanding the characteristics of various sources of light. For instance, sunlight has a broad and relatively uniform SPD (reflecting a wide range of visible wavelengths), which is why it appears white to the human eye. In contrast, the SPD of a fluorescent lamp might show noticeable spikes at certain wavelengths, indicating a different composition of light that can influence color perception.

2.3.3 Tristimulus color theory

It turns out that all visible colors can be matched by a linear combination of three independent "primaries", which allows for the reproduction of colors digitally. This is known as the Tristimulus color theory (or Grassman's Law).

To illustrate, let's define two source lights with their respective primary color intensities as follows:

- Light 1: R_1, G_1, B_1
- Light 2: R_2, G_2, B_2

When these two lights are combined to create a new source light (Light 3), the primary color intensities of this new light can be represented by the sum of the corresponding primary intensities of Lights 1 and 2. Mathematically, this can be denoted as:

- For the Red component of Light 3: $R_3 = R_1 + R_2$
- For the Green component of Light 3: $G_3 = G_1 + G_2$
- For the Blue component of Light 3: $B_3 = B_1 + B_2$

Therefore, the equation representing the new source light (Light 3) in terms of its primary color intensities can be expressed as:

$$\text{Light } 3 = (R_1 + R_2) + (G_1 + G_2) + (B_1 + B_2)$$

This equation demonstrates how the combination of different light sources with known primary color intensities can produce a new light source with its unique set of primary color intensities which can be applied for any color reproduction task in digital display/imaging.

2.4 Digital Imaging

Now that there is a foundation for understanding Digital Imaging, this section explores how digital devices capture and represent the interaction between light and color in a way that can be processed, displayed, and

interpreted electronically.

2.4.1 RGB in Cameras

Within the human retina, approximately 60–64% of these cones are attuned to red light (L-cones), around 30–32% to green (M-cones), and a smaller fraction, between 2–7%, to blue light (S-cones). At moderate to bright light levels, the eye is more sensitive to yellowish-green light than other colors because this stimulates the two most common (M and L) of the three kinds of cones almost equally.

Digital camera sensors, on the other hand, employ a strategic allocation of pixels on a grid to mimic this color sensitivity: 25% of the pixels are dedicated to capturing red light, another 25% to blue, and the remaining 50% to green light. Known as the Bayer pattern, this configuration exploits the human eye's greater sensitivity to green light to provide a higher resolution for green. From this, digital cameras are able to produce images that better approximate the way we see the world around us.

2.4.2 Debayering/Demosaicing

Since each pixel in a sensor covered by a Bayer filter captures only one of the three RGB colors, debayering (or demosaicing) is the process used to reconstruct a full-color image from the incomplete color samples captured by the camera sensor. This involves interpolating the missing color information for each pixel based on the values of adjacent pixels. Here are some examples of the types of interpolation methods that can be used to achieve this:

- **Nearest-Neighbor Interpolation:** This method assigns the value of the closest pixel with the missing color information to a pixel lacking that color. It is easy to implement but may result in blocky or pixelated images due to its simplistic assumption about color uniformity.
- **Bi-Linear Interpolation:** A more sophisticated approach that averages the values of adjacent pixels for the missing colors. This method smooths transitions and produces more natural-looking images by considering the color information of two or four neighboring pixels, depending on the color being interpolated, to fill in missing data.
- **Advanced Methods:** These involve more complex interpolation algorithms that consider a larger neighborhood around each pixel or apply machine learning techniques. The goal is to predict missing colors with higher accuracy, resulting in images of higher quality with more accurate color reproduction and fewer visual artifacts.

The result is a full RGB image that closely represents the colors and details of the scene as perceived by the human eye.

2.5 Image Processing

The ability to manipulate and understand images through computational methods is fundamental in computer vision and image processing. This section introduces several concepts that form the basis of how computers can interpret/manipulate digital images in a meaningful manner.

2.5.1 Point Processing and Image Filtering

When working with images, there are two main ways to make changes to them. The first is by manipulating individual pixels, which involves changing the values of each pixel individually (called point processing). The second option is to apply transformations that consider the pixel in relation to its surrounding pixels (called image filtering).

2.5.1.1 Point Processing

At its core, Point Processing refers to the technique of modifying individual pixels in an image without regard to the surrounding pixels. Thus, the result of a point operation does not change the structure, texture, or introduce new objects within the image; it only alters the existing pixel values. Essentially, the spatial relationship between pixels remains unchanged. This method is important for tasks such as adjusting the brightness and contrast of an image or applying more complex transformations that re-map pixel values according to a specific algorithm – all of which make point processing a powerful tool for preparing image data for subsequent analysis or as intermediary processing stages.

Here are some examples of some Point Processing Operations:

- **Identity/Inversion**: Keeps the pixel value as is (identity) or inverts it $255 - x$, effectively turning the image into its photographic negative.
- **Darken/Lighten**: Decreases $x - 128$ or increases $x + 128$ the pixel values, making the image darker or lighter, respectively.
- **Lower/Raise Contrast**: Reduces $x/2$ or amplifies $2x$ the contrast by adjusting the range of pixel intensity values.
- **Non-linear Lower/Raise Contrast**: Applies a non-linear transformation to reduce $\left(\frac{x}{255}\right)^{\frac{1}{3}} \times 255$ or increase $\left(\frac{x}{255}\right)^{3} \times 255$ contrast, which can help in highlighting features or reducing detail in an image.

2.5.1.2 Image Filtering

Contrasting with point processing, Image Filtering involves examining a pixel in the context of its neighbors. Thus, the result of an image filtering operation can oftentimes modify the structure of the image because it integrates information from a group of pixels. By applying filters that aggregate the values of a pixel and its immediate surroundings, image filtering can dramatically alter the image's characteristics, whether it's smoothing out noise, emphasizing edges, or preparing the image for higher-level analysis. An approach like this is useful for tasks that aim to enhance or suppress certain features within an image.

Here are some examples of some Image Filtering Operations:

- **Convolution**: This process involves sliding a small, predefined matrix known as a kernel over the image. At each position, the kernel is applied to the corresponding window of pixels, computing a weighted sum that blends the pixel values. This method enables selective emphasis on certain image features—such as textures, edges, or patterns—determined by the specific configuration of the kernel.

  The mathematical definition of a 2D discrete convolution is given by:

$$(f * g)(x, y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f(i, j) \cdot I(x - i, y - j) \tag{2.3}$$

Notice the flip in the kernel coordinates ($x-i$ and $y-j$), which is inherent in the convolution operation.

- **Correlation**: Like a convolution, this method also uses a predefined matrix or kernel which is then systematically aligned with and moved across different regions of an image. However, unlike convolution, correlation does not involve flipping the kernel. Correlation is particularly effective for template matching and recognizing specific features or patterns within a given image.

  The mathematical definition of a discrete 2D correlation is given by:

$$(f \star g)(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f(i,j) \cdot I(x+i, y+j) \tag{2.4}$$

  Notice the absence of a flip in the function arguments ($x+i$ and $y+j$), as indicated by the positive signs.

- **Gaussian Filter**: This filter uses a kernel that is defined by a 2-dimensional Gaussian distribution to effectively blur images smoothly. The filter weights each pixel in the kernel based on their distance from the center and causes a fall-off in influence with increasing distance, which helps in reducing noise and fine details. The kernel is theoretically infinite but is practically truncated to a finite window (the kernel) for computation.

  The mathematical expression for a Gaussian function in two dimensions is:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.5}$$

  where x and y are the distances from the center of the kernel, and σ is the standard deviation of the Gaussian distribution. The filter's ability to smooth and apply noise reduction to images makes it a valuable tool for image processing tasks, including the computation of gradients. The commonly used 3x3 Gaussian kernel is a discrete approximation of this function, which is scaled and represented as:

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{2.6}$$

  The 3x3 matrix provides a computationally efficient approximation for practical image processing applications.

- **Box Filter**: Like the Gaussian filter, the Box filter also provides blurring to an image. However, this filter operates on a simpler principle, where it makes each pixel set to the average of its neighbors in a given kernel. The resulting effect is a uniform smoothing, which can efficiently reduce noise and soften the image's details. A standard form of the Box filter kernel is often a square matrix where all the values are equal, and the sum of the values is typically normalized to 1 to maintain the overall brightness of the image. A basic example of a 3x3 Box filter kernel is:

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{2.7}$$

  This uniformity means that each neighbor has an equal contribution to the new pixel value, leading to an evenly distributed blurring effect across the image.