DEPTH SENSING

ABISHEK M (CB.EN.U4CSE20601)
SHANKARA NARAYANAN V(CB.EN.U4CSE20656)
SNEHA VARSHA M (CB.EN.U4CSE20659)
SYED ASHFAQ AHMED (CB.EN.U4CSE20665)
TARUN RAMASWAMY (CB.EN.U4CSE20666)

CONTENTS

PROBLEM STATEMENT

Given an image and camera calibration details, the goal is to find the relative depth of various objects in the image using stereo vision.

Objective : To apply several filters and image processing techniques and determine the relevance of said filters and processing techniques towards the end goal of extracting depth information from a given stereo image set.

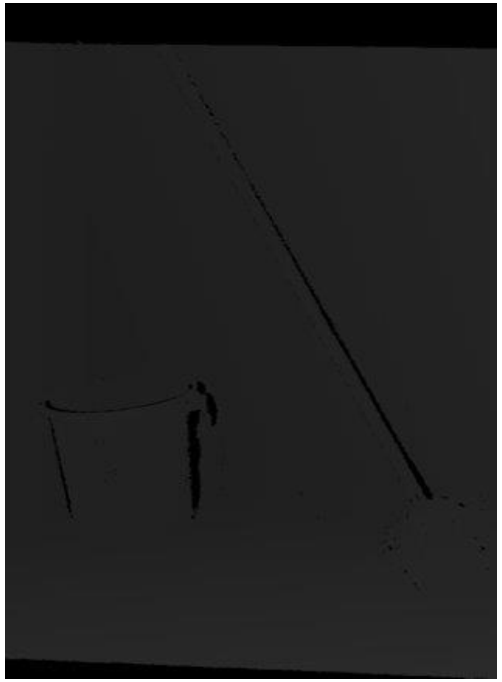
THE DATASET

OUTDOOR DATASET

Driving Stereo is the name of the dataset. This dataset contains

a. Left angle images
b. Right angle images
c. Rainy images
d. Foggy images
e. Respective depth for all types of images


LEFT IMAGE


LEFT DISPARITY MAP


RIGHT IMAGE


RIGHT DISPARITY MAP

THE DATASET

INDOOR DATASET

- The dataset contains indoor scenes captured using a stereo camera with left and right images along with their corresponding disparity maps.
- This dataset was chosen as this is the closest to a dataset of museum articles. Depth sensing can be extended on a museum dataset for 3D reconstruction.

TOPICS IMPLEMENTED

HISTOGRAM EQUALIZATION

Histogram equalization is a technique used in image processing to adjust the contrast of an image by spreading out the intensity values of the pixels. It works by transforming the intensity values of the pixels in the image so that the histogram of the image is flattened.

The histogram of an image is a graph that shows the number of pixels in the image at each intensity value. An image with a narrow range of intensity values will have a histogram that is peaked, while an image with a wide range of intensity values will have a flatter histogram.

Histogram equalization works by transforming the intensity values of the pixels in the image so that the histogram becomes more evenly distributed. This can help to improve the contrast of the image and make it easier to see details in both the dark and light areas of the image.

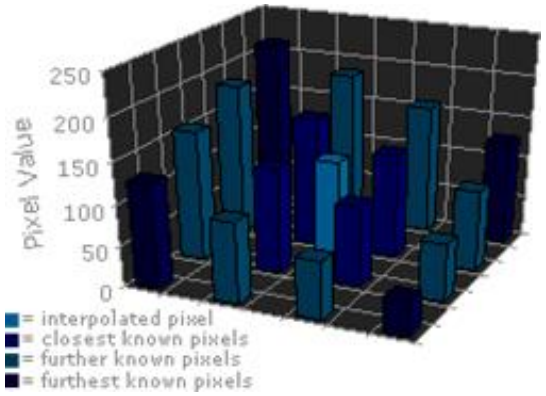
INTERPOLATION

Interpolation refers to the process of estimating the values of pixels in an image based on the values of surrounding pixels. This can be used to increase the resolution of an image, to change the aspect ratio of an image, or to correct for distortion. There are several different methods that can be used for interpolation, including nearest neighbor interpolation, bilinear interpolation, and bicubic interpolation. The choice of interpolation method will depend on the specific requirements of the application and the trade-off between accuracy and computational complexity.

CUBIC INTERPOLATION

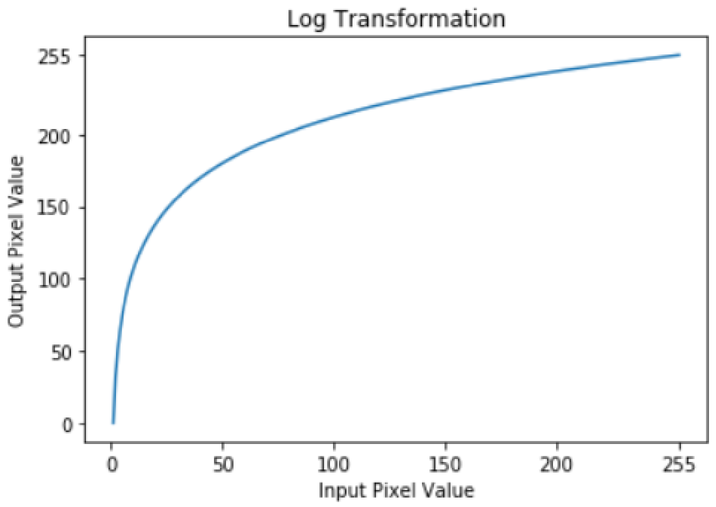
Cubic interpolation is a method for image interpolation that estimates the value of a pixel based on a weighted average of the surrounding pixels in the original image. It works by dividing the region around the pixel of interest into a grid of 16 pixels, and using the values of these pixels to estimate the value of the pixel. The weights for the average are calculated based on the distance of the pixel of interest from each of the surrounding pixels, using a cubic function.

Cubic interpolation is more accurate and produces higher-quality images than linear interpolation methods such as bilinear or nearest neighbor interpolation. It is often used in image resizing, image rotation, and other image processing operations that require high-quality contrast or to correct for non-uniform illumination. This type of transformation is based on the logarithmic function, which is a mathematical function that is used to transform data in order to make it more easily interpretable or to bring out certain features that may not be immediately apparent in the original data.

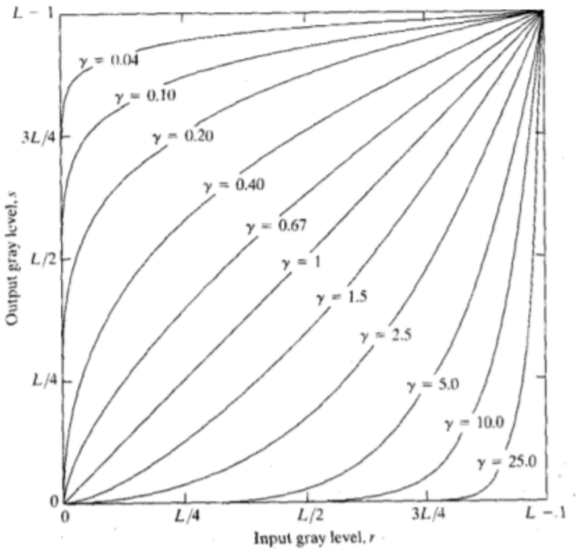


LOG TRANSFORM

In image processing, a logarithmic transformation is often used to stretch the intensity values of an image in order to bring out detail in the shadows or highlight areas of an image that may be too dark or too light. This is achieved by applying the logarithmic function to the pixel values of the image, which has the effect of compressing the values at one end of the intensity range and expanding them at the other end.



GAMMA CORRECTION



$$S = C \cdot R^{\gamma}$$

Is a method used in image processing to adjust the luminance of an image. It is often used to correct for the non-linear response of a display device, such as a computer monitor or television, to the intensity of the input signal.

Gamma correction is based on the concept of gamma, which is a measure of the non-linear relationship between the input signal and the output luminance of a display device. Different display devices have different gamma values, which can affect the way that an image is displayed.To correct for these differences in gamma, the input signal is adjusted using a gamma correction curve. This curve maps the input signal to the desired output luminance in a way that is designed to produce a more accurate and pleasing image.

INTENSITY RESOLUTION

Intensity resolution, also known as gray level resolution or bit depth, is an important aspect of image processing. It refers to the number of distinct intensity values that can be represented in an image. The intensity resolution of an image is determined by the number of bits used to represent each pixel.

In general, images with a higher intensity resolution will appear more realistic and lifelike because they are able to capture and represent a greater range of intensity values. However, increasing the intensity resolution of an image also increases the file size of the image and the amount of storage and processing required to handle it. It is important to balance the need for high intensity resolution with the constraints of available storage and processing resources.

It refers to the number of intensity levels used to represent an image. The more intensity levels used, the finer the level of detail discernible in an image. Intensity level resolution is usually given in terms of the number of bits used to store each intensity level.

FILTERING FOR NOISE REDUCTION

The presence of noise can heavily distort the depth and disparity map calculations. For this, the de noising processes is extremely vital.

SALT AND PEPPER NOISE REMOVAL

Salt-and-pepper noise is also called impulse noise. It can be caused by several reasons like dead pixels, analog-to-digital conversion error, bit transmission error, etc. The dataset can be cleaned from images containing salt and pepper noise using the following method.

In image processing, salt and pepper noise is a type of noise that is typically characterized by white and black pixels scattered randomly throughout the image. This type of noise is usually caused by electronic interference or other types of damage to the image. It can be difficult to remove salt and pepper noise from an image, because the pixels affected by the noise are randomly distributed and do not follow any particular pattern. One way to remove salt and pepper noise is to use a median filter, which replaces the value of each pixel with the median value of the pixels in its neighborhood. This can help to smooth out the noise and restore the image to its original appearance. Other techniques, such as Gaussian filtering and Wiener filtering, can also be used to reduce the effects of salt and pepper noise.

A median filter is used to clean the images from salt and pepper noise as it is the most efficient among non linear filters. The implementation of both median and average filter is done and plotted side by side for comparison.

GAUSSIAN NOISE REMOVAL

Gaussian noise is a type of noise that is commonly used to add random variations to an image in image processing. It is named after the Gaussian distribution, which is a probability distribution that is characterized by a bell-shaped curve. In image processing, Gaussian noise is typically added to an image to simulate the effects of noise that may be present in real-world images. This can be useful for testing image processing algorithms and for training machine learning models to recognize patterns in noisy images.

The removal of Gaussian noise can be done using any non-linear filters (average, median etc) but Wiener's filter is a type of filter that can do both filtering of Gaussian noise and also fix the motion blur in an image. In the project, there is a high possibility of having images with both Gaussian blur and most importantly motion blur hence correcting them becomes a crucial step towards finding the depth of objects in the image.

ARITHMETIC MEAN FILTER

In image processing, an arithmetic mean filter is a simple sliding window spatial filter that replaces the center pixel in the window with the arithmetic mean of all the pixel values in the window. It is often used to smooth images, reduce noise, and reduce detail.

The size of the window, or kernel, determines the extent to which the image is smoothed. A larger kernel size will result in more smoothing, while a smaller kernel size will result in less smoothing.

$$\hat{f}(x,y) = \frac{1}{mn} \sum_{(s,t)\in S_{xy}} g(s,t)$$

GEOMETRIC MEAN FILTER

In image processing, a geometric mean filter is a spatial filter that replaces the value of each pixel in an image with the geometric mean of the values of the pixels in a local neighborhood around that pixel. The geometric mean is defined as the nth root of the product of the values of the pixels in the neighborhood, where n is the number of pixels in the neighborhood.

$$\hat{f}(x,y) = \left[\prod_{(s,t)\in S_{xy}} g(s,t) \right]^{\frac{1}{mn}}$$

HARMONIC MEAN FILTER

In image processing, a contraharmonic mean filter is a spatial filter that replaces the value of each pixel in an image with the weighted average of the values of the pixels in a local neighborhood around that pixel. The weights are determined by a parameter called the "order" of the filter, which can be either positive or negative.

A contraharmonic mean filter reduces or virtually eliminates the effects of salt-and-pepper noise. For positive values of Q, the filter eliminates pepper noise. For negative values of Q it eliminates salt noise. It cannot do both simultaneously.

Note that the contraharmonic filter is simply the arithmetic mean filter if Q = 0, and the harmonic mean filter if Q = −1.

A larger region (filter size) yields a stronger filter effect with the drawback of some blurring.

$$\hat{f}(x,y) = \frac{\sum_{(s,t)\in S_{xy}} g(s,t)^{Q+1}}{\sum_{(s,t)\in S_{xy}} g(s,t)^{Q}}$$

ADAPTIVE MEDIAN FILTER

The adaptive median filter is a non-linear digital filter used for image processing and computer vision. It is designed to remove noise from images while preserving edges and other important image features.

The adaptive median filter works by comparing the intensity of each pixel in the image to the median intensity of the surrounding pixels. If the intensity of the pixel is significantly different from the median intensity, it is considered to be noise and is replaced with the median intensity. This process is repeated for all pixels in the image, resulting in a smoothed and denoised image.

One of the key advantages of the adaptive median filter is its ability to effectively remove salt and pepper noise (impulse noise). It is also effective at preserving edges and other important features in the image, making it a popular choice for image processing applications.

Stage A: $A1 = z_{med} - z_{min}$

$A2 = z_{med} - z_{max}$

If $A1 > 0$ and $A2 < 0$, Go to level B

Else increase the window size

If window size $\leq S_{max}$ repeat level A

Else output z_{med}

Stage B: $B1 = z_{xy} - z_{min}$

$B2 = z_{xy} - z_{max}$

If $B1 > 0$ and $B2 < 0$, output z_{xy}

Else output z_{med}

- z_{min} = minimum grey level in S_{xy}
- z_{max} = maximum grey level in S_{xy}
- z_{med} = median of grey levels in S_{xy}
- z_{xy} = grey level at coordinates (x, y)
- S_{max} =maximum allowed size of S_{xy}

ADAPTIVE NOISE REDUCTION

Adaptive noise reduction is a technique used to reduce the level of noise in a signal or image. It involves analyzing the characteristics of the noise in the signal and adapting the noise reduction algorithm to better fit the noise profile.

The technique used here involves using the mean of a neighborhood and variance of both the neighborhood and of the overall image.

One of the key benefits of adaptive noise reduction in image processing is that it can effectively remove noise while preserving edges and other important features in the image. This is important because traditional noise reduction techniques can often blur or distort these features, leading to a loss of detail and clarity in the image.

Adaptive noise reduction is commonly used in a wide range of image processing applications, including medical imaging, surveillance, and industrial inspection. It is also used in digital photography to improve the quality of images captured in low light or high-noise environments.

$$\hat{f}(x,y) = g(x,y) - \frac{\sigma_\eta^2}{\sigma_L^2}\left[g(x,y) - m_L\right]$$

GAUSSIAN AND UNIFORM NOISE REMOVAL USING MIDPOINT FILTER

A midpoint filter is a type of digital filter that is used to smooth out data by removing noise or outliers. It works by replacing each point in a dataset with the average of that point and its two nearest neighbors. This has the effect of smoothing out fluctuations in the data and making it more predictable.

The midpoint filter is typically used to filter images containing short tailed noise such as Gaussian and uniform type noises. The midpoint filter is defined as :

$$MidPoint(A) = \frac{\min[A(x+i, y+j)] + \max[A(x+i, y+j)]}{2}$$

where the coordinate (x+i, y+j) is defined over the image A and the coordinate (i, j) is defined over the N x N size square mask.

SALT NOISE REMOVAL USING MIN FILTER

When the minimum filter is applied to a digital image it picks up the minimum value of the neighborhood pixel window and assigns it to the current pixel. A pixel with the minimum value is the darkest among the pixels present in the pixel window. The dark values present in an image are enhanced by the minimum filter.

Minimum filter is also called a dilation filter. When a minimum filter is applied the object boundaries present in an image are extended. The minimum filter is typically applied to an image to remove positive outlier noise(salt noise).

PEPPER NOISE REMOVAL USING MAX FILTER

The maximum filter replaces each pixel value of a Digital Image with the maximum value(i.e., the value of the brightest pixel) of its neighborhood pixel window. It is the opposite of what the minimum filter does to an Image.

Applying the maximum filter removes the negative outlier noise(pepper noise) present in a Digital Image.

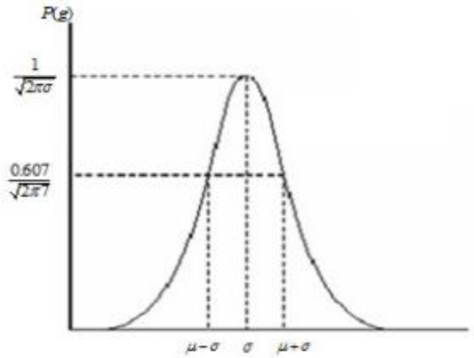
INDUCING SALT AND PEPPER NOISE AND REMOVING IT WITH MEDIAN FILTER

A median filter is a type of digital filter that is used to smooth out data by removing noise or outliers. It works by replacing each point in a dataset with the median of that point and its neighbors. In the context of image processing, this has the effect of smoothing out noisy pixels and reducing the appearance of "salt and pepper" noise in the image.

To implement a median filter in image processing, you would first define the size of the filter window. This is the number of neighboring pixels that will be included in the median calculation for each pixel. A larger window size will result in more smoothing, but may also blur important details in the image.

ESTIMATING PROBABILITY DENSITY FUNCTION OF NOISE THROUGH HISTOGRAMS

Gaussian Noise is a statistical noise with a Gaussian (normal) distribution. It means that the noise values are distributed in a normal Gaussian way.



The Gaussian noise is added to the original image. The probability density function p of a Gaussian random variable z is calculated by the following formula:

$$p_G(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

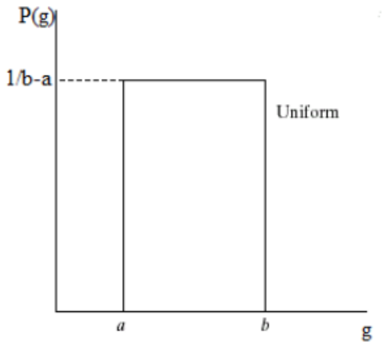
UNIFORM NOISE

Uniform noise is not often encountered in real-world imaging systems, but provides a useful comparison with Gaussian noise. The linear average is a comparatively poor estimator for the mean of a uniform distribution. This implies that nonlinear filters should be better at removing uniform noise than Gaussian noise.

The Uniform pdf is given by:

$$P(g) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le g \le b \\ 0 & \text{otherwise} \end{cases}$$

The uniform distribution is given below:



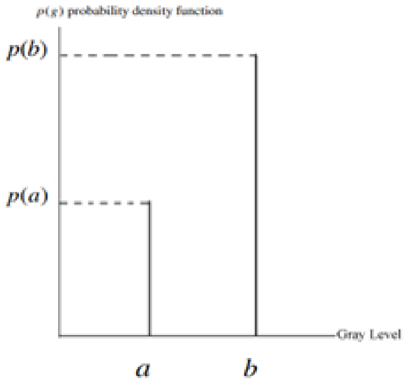
SALT AND PEPPER NOISE

Salt-and-pepper noise, also known as impulse noise, is a form of noise sometimes seen on digital images. This noise can be caused by sharp and sudden disturbances in the image signal. It presents itself as sparsely occurring white and black pixels.

The Uniform pdf is given by

$$P(g) = \begin{cases} Pa \text{ for } g = a \\ Pb \text{ for } g = b \\ 0 \text{ otherwise} \end{cases}$$

The salt and pepper noise distribution is given below:



PSNR (METRICS FOR QUALITY OF DENOISING EFFECT)

The Peak Signal to Noise Ratio (PSNR) is a measure of the quality of a denoised image in comparison to the original image. This metric is used to measure the noise present in an image that has been subjected to noise induction, as compared to a denoised version of the same image.

In general, a higher PSNR value indicates a higher quality reconstructed signal and a lower PSNR value indicates a lower quality reconstructed signal. There is no one optimal PSNR value that applies to all images for this data.

The Contra-harmonic mean filter is more effective at denoising the Salt Noise induced image, as indicated by its higher PSNR value compared to the Mean filter removing Salt-Pepper noise induced image.

MOTION BLUR AND DE-BLURRING

Motion blur is a common issue in image processing, and it occurs when the camera or subject is moving during the exposure time of the image. This results in an image where objects appear blurred or smeared, as the image captures the movement of the object rather than a static image.

The noise is cleaned using the Wiener filtering method. A gaussian kernel is created (default 3x3 filter) using normal random numbers. That kernel is passed into the Wiener filter method. The constant K is the ratio of Spectral Power Density of the noisy image and the Spectral Power Density of the filter kernel used. We pass it as 10 instead of calculating it manually in the code each time.

Finally after the Fast Fourier Transform function is used, the output must be inverted using Inverse Fast Fourier Transform Function.

$$W(u,v) = \frac{H^*(u,v)}{|H(u,v)|^2 + S_\epsilon(u,v)/S_f(u,v)}$$

PERIODIC NOISE

Periodic noise is an unwanted signal that interferes with the source image or signal at a random frequency, depending on its source. Generally, this interference can be added to the image from nature, the electricity network, or electronics devices.

BAND REJECT FILTER

A band reject filter is useful when the general location of the noise in the frequency domain is known. A band reject filter blocks frequencies within the chosen range and lets frequencies outside of the range pass through.

Band reject filters are ideally suited for filtering out periodic interference. The Fourier transform of a pure sine or cosine function is just a pair of impulses. Therefore the interference is "localized" in the spectral domain and one can easily identify this region and filter it out.

BAND PASS FILTER

A band-pass filter is a type of image processing filter that is used to remove any image content that falls outside a specified frequency range. It is called a "band-pass" filter because it allows a specific range of frequencies to pass through, while blocking frequencies outside of this range. This can be useful for isolating specific features or patterns in an image, or for removing unwanted noise or clutter.

To use a band-pass filter, you first need to specify the range of frequencies that you want to allow through the filter. This can be done using a lower and upper cutoff frequency. The filter will then remove any image content that falls outside of this range.

NOTCH FILTERING

Notch filtering is a type of frequency-domain filtering that is used to remove or reduce certain frequency components in a signal or image. It is called "notch" filtering because it creates a "notch" or a dip in the frequency spectrum at a specific frequency or set of frequencies. This can be useful for removing specific types of noise or interference from an image, or for emphasizing certain features in the image.

Notch filters are often used in image processing applications to remove periodic noise, such as noise introduced by a camera's rolling shutter effect or by interference from electrical power sources. They can also be used to remove other types of noise, such as impulse noise or Gaussian noise.

EDGE DETECTION

Edges are the set of pixel positions present in locations of abrupt intensity changes. They represent boundaries between objects with other objects and backgrounds. For detecting edges, there are several filters we can apply to our image.

LAPLACIAN FILTER

This filter uses the property of second derivatives, that at points of onset of ramps / steps, the second derivative value is non zero, while it is zero in regions of constant intensity and on ramps / steps.

UNSHARP MASKING SHARPENING FILTER

This sharpening filter works in two steps. First, we blur the image, and then subtract the original image from the blurred image. The resultant is an image of the same dimensions as that of the original image, but with edge details highlighted.

SOBEL EDGE DETECTOR

The Sorbel edge detector approximates the Roberts cross gradient operator (which works based on pixel differentiation/ the gradient).

MARR HILDRETH EDGE DETECTION

The Marr-Hildreth edge detection method is a way to identify the edges of objects in an image. It is based on the idea that the intensity gradient of an image is highest at the edges of objects.

CANNY EDGE DETECTOR

The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. The Canny edge detection algorithm is composed of 5 steps:

- Noise reduction
- Gradient calculation
- Non-maximum suppression
- Double threshold
- Edge Tracking by Hysteresis

SOBEL AND PREWITT EDGE DETECTION

The Sobel is one of the most commonly used edge detectors. It is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of computations. The Sobel edge enhancement filter has the advantage of providing differentiating and smoothing concurrently.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Prewitt operator is similar to the Sobel operator and is used for detecting vertical and horizontal edges in images. However, unlike the Sobel, this operator does not place any emphasis on the pixels that are closer to the center of the mask.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

MORPHOLOGICAL GRADIENT OPERATOR

A simple morphological image processing technique used to find the borders / outline of images.

Consist of two steps:

(i) Perform erosion to shrink the image based on the structuring element.

(ii) Subtract eroded image from the dilated image.

GAUSSIAN LOW PASS AND HIGH PASS FILTER

frequency signals to pass through it while attenuating high frequency signals. This type of filter is often used to smooth images or remove noise from signals. It works by convolving the input signal with a Gaussian kernel, which is a bell-shaped curve that decays rapidly as the distance from the center increases.

A Gaussian high pass filter is the opposite of a low pass filter, in that it allows high frequency signals to pass through while attenuating low frequency signals. It is often used to sharpen images or to highlight features in a signal. Like the low pass filter, it works by convolving the input signal with a Gaussian kernel, but in this case the kernel has a negative value in the center and positive values at the edges.

Both types of filters are named after the Gaussian function, which is a continuous, smooth curve that is used to model the probability distribution of many real-world phenomena.

UNSHARP MASKING

In unsharp masking, the blurred version of the image is taken as a mask which is then subtracted from the original image. This gives a clearer image as the initial blur is removed.

HIGH-BOOST FILTERING

In image processing, it is often desirable to emphasize high frequency components representing the image details without eliminating low frequency components (such as sharpening). The high–boost filter can be used to enhance high frequency components.

LAPLACIAN FILTER

Laplacian filter is a second-order derivative filter used in edge detection, in digital image processing. In 1st order derivative filters, we detect the edge along with horizontal and vertical directions separately and then combine both. But using the Laplacian filter we detect the edges in the whole image at once.

FREQUENCY DOMAIN FILTERS

In the frequency domain, once the image has been centered, it must be noticed that the low frequency components are present close to the center, and as we go away from it, the high frequency components dominate. Since the edges are associated with high frequency components, the frequency domain can be used to efficiently enhance / cull certain parts of an image by frequency.

IDEAL HIGH PASS FILTER

This filter works by attenuating low frequency components around the image center. The filter is mathematically defined as:

$$H(u,v) = \begin{cases} 1, & if\ D(u,v) \geq D_0 \\ 0, & if\ D(u,v) < D_0 \end{cases}$$

Here, D(u, v) is given by: $\sqrt{(u - M/2)^2 + (v - N/2)^2}$

where,

u, v : coordinate of image in frequency domain, M, N : width and height of the source image (in pixels), D_0 is the cutoff frequency of the image (or the radius of the circle encompassing the low frequency components of frequency domain image)

IDEAL LOW PASS FILTER

The Ideal Low pass filter is simply obtained by performing 1 - IHPF, which means:

The high frequency components in the frequency domain of source image are attenuated, leaving behind only the low frequency components (I.e. components associated with intensity / contrast).

The filter is obtained by the computation: $H(u,v) = \begin{cases} 0, \ if D(u,v) < D_0 \\ 1, \ if D(u,v) < D_0 \end{cases}$

Here, D(u, v) is given by: $\sqrt{(u - M/2)^2 + (v - N/2)^2}$

Where,

u, v : coordinate of image in frequency domain, M, N : width and height of the source image (in pixels), D_0 is the cutoff frequency of the image (or the radius of the circle encompassing the low frequency components of frequency domain image)

BUTTERWORTH LOW PASS FILTER

This is a frequency domain low pass filter which gives us more control and flexibility compared to the ideal low pass filter.

Mathematically, this filter is given by: $H(u,v) = 1/(1 + (D/D_0)^{2n})$

Where, D is given by: $\sqrt{(u - M/2)^2 + (v - N/2)^2}$

D_0 = cut off frequency, n = order of filter, U, v : coordinate of image in frequency domain, M, N : width and height of the source image (in pixels)

BUTTERWORTH HIGH PASS FILTER

This is a frequency domain low pass filter which gives us more control and flexibility compared to the ideal high pass filter. It is obtained by performing 1 - Butterworth Low Pass Filter.

Mathematically, this filter is given by: $H(u,v) = 1/(1 + (D0/D)^{2n})$

Where D is given by: $\sqrt{(u-M/2)^2 + (v-N/2)^2}$

D0 = cut off frequency, n = order of filter, U, v : coordinate of image in frequency domain, M, N : width and height of the source image (in pixels)

SMOOTHING AND SEGMENTATION

Smoothing is a process that aims to reduce the level of detail in an image, often by removing noise or reducing the contrast between neighboring pixels. This can be useful for eliminating distractions or making an image easier to interpret. Segmentation, on the other hand, is the process of dividing an image into distinct regions or segments, each of which corresponds to a different object or background in the image. This can be useful for identifying and analyzing specific objects or features in an image. Segmentation can be performed using a variety of methods, including thresholding, clustering, and edge detection.

Here thresholding has been used as a form of segmentation to segment pixels with values greater than 127 and pixels with values lesser than 127.

Together, smoothing and segmentation can be used to improve the clarity and interpretability of an image, as well as to extract useful information or features from it.

LOW AND HIGH THRESHOLDING

Low thresholding is a technique in which pixels with intensity values below a certain threshold are set to black, and pixels with intensity values above the threshold are set to white. This results in a binary image in which the objects of interest are represented as white pixels, and the background is represented as black pixels.

High thresholding is similar to low thresholding, but the roles of the black and white pixels are reversed. In high thresholding, pixels with intensity values above a certain threshold are set to white, and pixels with intensity values below the threshold are set to black. This results in a binary image in which the background is represented as white pixels, and the objects of interest are represented as black pixels.

SMOOTH AND HARD THRESHOLDING

In smooth thresholding, the transition from black to white pixels is not abrupt, but rather occurs gradually over a range of intensity values. This results in a smoother, more gradual transition between the foreground and background pixels in the binary image. Smooth thresholding is often used when the intensity values of the pixels in the image are not clearly separated, and a more gradual transition is needed to accurately segment the image.

Hard thresholding, on the other hand, involves an abrupt transition from black to white pixels at a certain intensity value. Pixels with intensity values below the threshold are set to black, and pixels with intensity values above the threshold are set to white. This results in a binary image with a clear separation between the foreground and background pixels. Hard thresholding is often used when the intensity values of the pixels in the image are well-separated and an abrupt transition is sufficient to accurately segment the image.

OTSU GLOBAL THRESHOLD

The image is plotted as a histogram and the two separable components in such a histogram will be the foreground pixels and the background pixels. Otsu's algorithm deals with finding a global threshold K from the histogram of a gray scale image. The algorithm uses two statistical measures: within class variance and between class variance to find the optimal global threshold.

IMPROVING OTSU'S GLOBAL THRESHOLDING BY IMAGE SMOOTHENING

Using any sort of image smoothing technique (Gaussian filtering, Average filtering, etc.) then applying Otsu results in a better segmentation of foreground and background pixels as noise will be either removed or suppressed in the smoothened image.

IMPROVING OTSU'S GLOBAL THRESHOLDING BY EDGE ENHANCEMENT

Using any edge detection techniques (Canny, Laplacian, Maar Hildreth, etc.) the edges in the image can be detected and enhanced. Applying Otsu's Global Thresholding on edge enhanced image results in stronger boundaries in the threshold image.

MULTILEVEL THRESHOLDING USING OTSU'S THRESHOLDING

Multilevel thresholding is a process that segments a gray level image into several distinct regions. A value "n" can be fed into the algorithm and "n" number of gray levels will be present in the threshold image

VARIABLE THRESHOLDING

Applying thresholding techniques on a local neighborhood using a local threshold value is the concept behind variable thresholding. Applying local thresholding on images results in segmentation of both foreground and background images along with in-scene objects based on their relative depth.

GLOBAL THRESHOLDING USING HOUGH TRANSFORM

The Hough transform is a technique which can be used to isolate features of a particular shape within an image. Because it requires that the desired features be specified in some parametric form, the classical Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, etc

COLOR SPACE SEGMENTATION (HSV)

Given an RGB image, we first convert the color space to HSV (Hue Saturation Value). The reason for this is :

(i) Hue: Gives only color information (based on degrees, where red falls between 0 to 60, yellow falls between 61 to 120, etc).

(ii) Value: The darkness / brightness of the color.

(iii) Saturation: The amount of gray in the image. If this value is 0%, it produces a faded effect and produces whiter. On the other hand, if the value is 100%, we get the primary color itself.

If we are able to retrieve the Hue value alone from a given HSV image, we can segment the objects based on just the color information, disregarding the shade / brightness of the *same* color present in several objects over the scene.

WATERSHED SEGMENTATION

In geographical sense: An array where water accumulates. It is a land area that channels / drains rain and snow to creeks, streams, etc. Watersheds can be segmented as topographical maps with boundaries (lower the altitude of the land, more water it accumulates). I.E, the gray scale images can be considered to have valleys and peaks, based on the intensity. There is also a threshold that segregates between the two. The brightness determines the high (peak) and low (valley) in this algorithm.

It works by filling values (i.e., local minima), with pixels belonging to the same predefined label. High intensity denotes peaks and hills, while low intensity denotes valley. We fill every local isolated minimum and mark it as a different segment. We also create a boundary / barrier between two isolated valleys.

MORPHOLOGICAL TRANSFORMATIONS

Broad set of image processing operations that processes binary images based on SE (Structuring elements) which dictates the nature of the operation.

Each pixel in the image is adjusted based on the value of pixels in its neighborhood.

It is not subjective. Used for getting the ground truth of the description of region shapes, boundaries, and for extracting the various components from the image.

EROSION

It is a morphological image processing technique that is used to reduce the size of bright regions in an image. This process erodes the boundaries of these regions in the image. It is typically used to remove noise or small objects from an image.

Pixel will be considered 255 or 0 only if ALL pixels under the kernel are 255 or 0. Otherwise it is eroded (i.e color changes to 0 or 255)

DILATION IN IMAGE PROCESSING

Used for 'expanding an element A by some structural element B'. Adds pixels to object boundaries.

Dilation is a morphological image processing technique that is used to increase the size of the bright regions in an image. This process is used to add pixels to the boundaries of objects in an image. It is used to connect or join the disconnected objects.

Even if a single pixel of the image is covered by the structural element, we set the entire area covered by the structural element to 1, else 0.

OPENING

The Opening is a process of removing small objects from the foreground of an image and placing them in the background. This process, which consists of dilation followed by erosion, modifies the image. $(A \bullet B) = (A \oplus B) \ominus B$

CLOSING

Closing is a process of removing small holes from the foreground of an image and placing them in the background. This process, which consists of erosion followed by dilation, modifies the image.

$$(A \bullet B) = (A \oplus B) \ominus B$$

CONTOUR DETECTION

A contour is a closed curve joining all the continuous points having some color or intensity, they represent the shapes of objects found in an image. Contour detection is a useful technique for shape analysis and object detection and recognition.

MORPHOLOGICAL TOP HAT OPERATOR

A simple morphological image processing technique used to extract the small and fine details from an image.

It consist of two steps :

(i) Obtain the 'opening' of the image.

(ii) Subtract the opening of image from the original image.

CLUSTERING

K-means clustering is a technique for partitioning a set of data points into K clusters, where each data point belongs to the cluster with the nearest mean. In image processing, K-means clustering can be used to classify the pixels in an image into K different clusters, based on the pixel values (e.g., intensity values).

K-means clustering is an iterative process that aims to minimize the sum of the distances between each pixel and its assigned centroid. It is a simple and fast technique that is widely used in image processing and computer vision.

One advantage of K-means clustering is that it can be used to reduce the dimensionality of an image by compressing it into K clusters. This can be useful for tasks such as image compression, image segmentation, and feature extraction.

HIT OR MISS TRANSFORM (SHAPE DETECTION)

An morphological compound transformation that can be used in finding a given configuration / pattern in a binary image. This technique uses basic morphological image processing as a basis (i.e erosion and dilation).

This morphological transform finds those pixels whose neighborhood matches the shape of a first structuring element B1 while *not* matching the shape of a second structuring element B2 simultaneously.

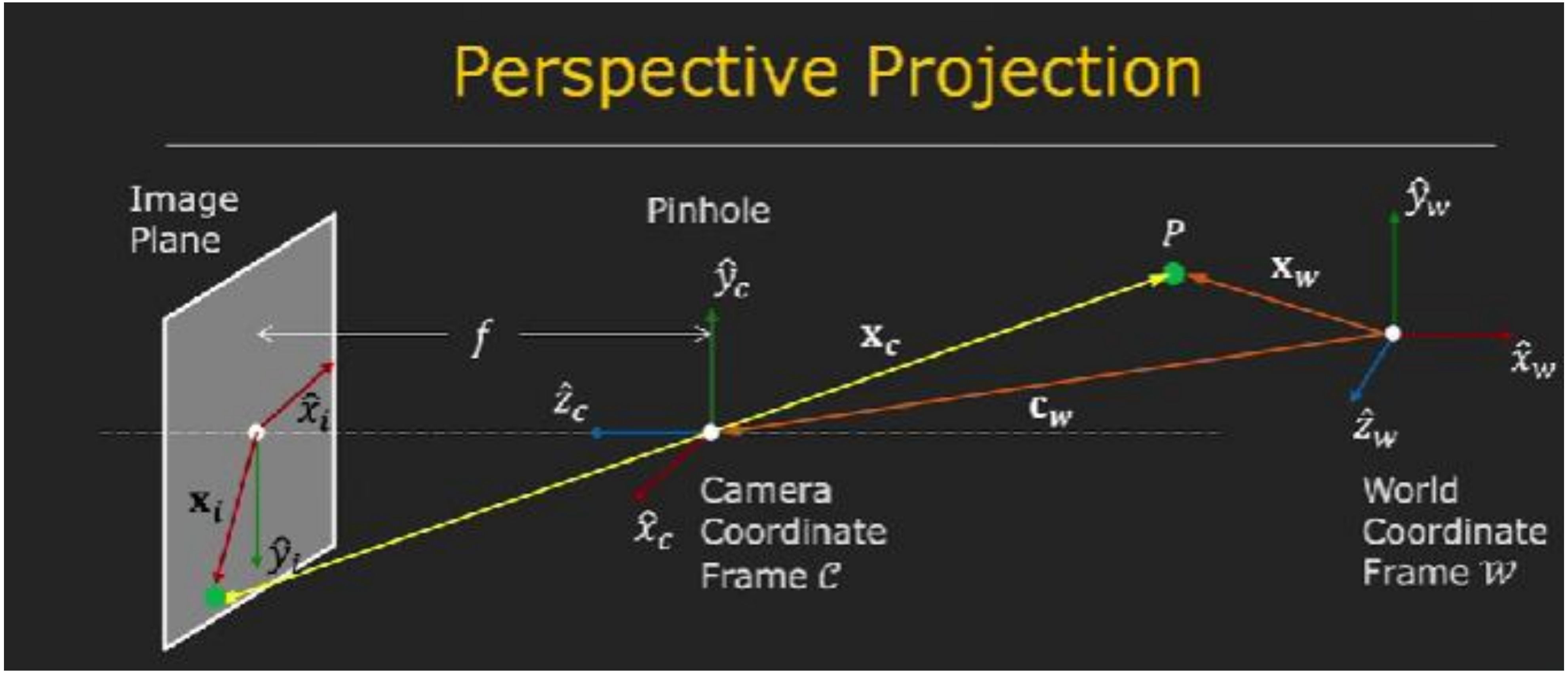
$$A(*)B = ((A \ominus B1) \cap (\sim A \ominus B2)).$$

DEPTH / DISPARITY MAP GENERATION

When we take a photo on a camera, we 'convert' the scene's world coordinate into the view / camera coordinate frame, where the camera is at the center, looking down on either the +z or -z axis. Then, we apply the projection matrix and perspective divide to go from the view space onto pixel space.

Our goal is the estimate the relative objects in the image, given camera calibration parameters and a pair of stereo (i.e left and right images).

Essentially, we try to estimate the world coordinates of different pixels in the image.



CAMERA CALIBRATION AND PARAMETERS

DISPARITY CALCULATION USING TEMPLATE MATCHING

Disparity map refers to the apparent pixel difference or motion between a pair of stereo images.

Since the displacement in the vertical y axis is 0, we can have a moving window in scanline order (i.e x axis, going from left to right). In the left stereo image, we place a template : usually of a rectangle and try to find the template's corresponding position in the right stereo image. The horizontal position (x) would vary, while the vertical coordinate (y) remains same as the templates y coordinate. We will use the Sum of absolute differences metric to determine the values of ul and ur in the above explanatory image.

$$SAD = \sum_{i=0}^{k} \sum_{j=0}^{k} (abs(image_{right}[i,j] - image_{left}[i,j]))$$

DEPTH MAP CALCULATION

For Generation of Depth map, we need to use the internal camera geometry parameters (ie. baseline and focal length)and disparity map as follows:

$$z = f * b(xl - xr) = fb / d$$

Here, f is the focal length (obtainable from the camera matrix, b is the baseline values corresponding from the translation vectors t, d is the disparity map.

THANK YOU!