1. The primary goal of convolution in computer vision is to perform a mathematical operation that combines two functions, typically an image and a filter or kernel, to produce a third function, which highlights specific features such as edges, textures, or patterns within the image.
2. Computer Vision helps self-driving cars by analyzing data from cameras to accurately detect and identify lane markings, recognize obstacles, and make real-time decisions for safe navigation and path planning.
3. A significant benefit of using computer vision in video surveillance is the ability to automatically recognize and classify objects, enabling quicker and more efficient searches through video footage for specific events, individuals, or activities without the need for manual tagging.
4. The primary goal of convolutional neural networks (CNNs) in computer vision is to learn and extract hierarchical features from images, starting with simple edges and textures at lower layers and progressing to more complex structures and objects at higher layers, to accurately perform tasks such as classification, detection, and segmentation.
5. Deep Learning is based on learning data representations through multiple layers of neural networks, which automatically discover intricate patterns and features from large datasets, enabling the model to make accurate predictions and decisions.
6. Machine Learning is crucial for handling dynamic and unpredictable environments, as it enables systems to adapt by learning from data and identifying patterns, thus effectively managing unknown parameters and improving performance over time.
7. The main goal of computer vision is to interpret and understand the visual content of images and videos in a way that allows machines to perceive and infer the context or "story" behind the visual data, similar to human perception.

8. Some challenges in computer vision include local ambiguity, where individual image regions may have multiple plausible interpretations, making it difficult for algorithms to consistently and accurately discern the correct information.
9. In image processing, a Mean Filter works by averaging the pixel values within a defined neighborhood around each pixel, resulting in a blurred effect that reduces noise and smooths out variations in the image.
10. Some subareas of artificial intelligence mentioned in the text include Machine Learning, which focuses on developing algorithms that learn from data; Computer Vision, which aims to enable machines to interpret visual information; Speech Recognition, which involves converting spoken language into text; and other areas such as Natural Language Processing and Robotics.

AI is the study and design of intelligent agents.
It deals with the machines which are designed and programmed in such a manner that they think and act like humans.
They perceives the environment and takes action that maximizes the chances of the success.

Significance of AI
AI brings the idea of error free world.
Automated machines can perform repetitive tasks involving complex computations precisely than humans.
Accurate and faster results.
Used in important areas like Banking and Financial systems, Medical sciences, Air transport, Gaming Zone.

Machines Versus Humans

- Machines can do certain things better than humans can
  for example,
  Adding a thousand four-digit numbers
  Drawing complex, 3D images
  Store and retrieve massive amounts of data

- However, humans possess natural creativity, common sense.

Deep Learning & Artificial Neural Networks

Deep learning is a machine learning technique that teaches computers to learn and adapt through experience.

It is based on learning data representations.

For example, Deep learning is a key technology behind driverless cars.

An Artificial Neuron network (ANN) is a computational model that simulates structure and functions of biological neural networks.

Information that flows through the network affects the structure of the ANN .

A neural network changes or learns, in a sense based on that input and output.

The basic component of ANN is artificial neuron.

Each of the connections has a number associated with it called the connection weight.

Each of the neurons has a number and a special formula associated with them called a threshold value and an activation function respectively.

Neural networks which consist of more than three layers of neurons (including the input and output layer) are called as *Deep Neural Networks*.

And training them is called as *Deep Learning*.

Speech Recognition

Speech recognition includes capturing and digitizing the soundwaves and converting into text.

To make voice assistants speak and reply with greater accuracy.

Technological development of AI can have a crucial impact on Labour market.

The middle level jobs that require routine manual skills are the ones that are at more risk.

Jobs of Factory workers, drivers, cashiers, clerks, analysts etc are being automated by robots and computers.

Jobs requiring high skilled set like Data analyst, computer vision engineer, Machine learning engineer are at high demand.

Applications of AI

Autonomous planning and scheduling of tasks aboard in a spacecraft.

Game playing: Specially designed algorithms are programmed into computers to play games like chess.

Monitoring trade in the stock market: Detecting and preventing unfair trading
Role of AI in air transport: Aircraft diagnosis, weather analysis, predictive risk management and flight planning.

Understanding Natural languages: Making machines to perform as per your instructions and to hear decision from it.

Robotic assistants in surgery: Assisting remote microsurgical procedures

Advantages

- Replace humans in repetitive, tedious tasks and in many laborious places of work.
- Precision and accuracy and speed.
- Can detect fraud in card based systems(cyber security).
- They wont be affected by hostile environments, thus able to performs dangerous tasks.
- Work in mining, digging fuels, explore in space.
- Predict what user will type , ask, search and assists user.

Disadvantages

- Initial Cost incurred in the setup and maintenance is high.
- Machines lack the human touch
- They may not be as efficient as humans in altering their responses depending on the changing situations.
- If the control of machines goes in the wrong hands, it may cause destruction.
- Creates economical gap between highly skilled people and mediocre skilled people.

Future scope

- Cyber security
- Data pattern Analysis
- Face recognisation
- Humanoids
- Autonomous driving vehicles

Medical & health care

Machine Learning

Machine Learning: A Definition

Definition: A computer program is said to *learn* from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

What is Machine Learning?

It is very hard to write programs that solve problems like recognizing a face.

We don't know what program to write because we don't know how our brain does it.

Even if we had a good idea about how to do it, the program might be horrendously complicated.

Instead of writing a program by hand, we collect lots of examples that specify the correct output for a given input.

A machine learning algorithm then takes these examples and produces a program that does the job.

The program produced by the learning algorithm may look very different from a typical hand-written program. It may contain millions of numbers.

If we do it right, the program works for new cases as well as the ones we trained it on.

A classic example of a task that requires machine learning: It is very hard to say what makes a 2 Some more examples of tasks that are best solved by using a learning algorithm

Recognizing patterns:

- Facial identities or facial expressions
- Handwritten or spoken words
- Medical images

Generating patterns:

- Generating images or motion sequences

Recognizing anomalies:

- Unusual sequences of credit card transactions
- Unusual patterns of sensor readings in a nuclear power plant or unusual sound in your car engine.

Prediction:

- Future stock prices or currency exchange rates

The web contains a lot of data. Tasks with very big datasets often use machine learning especially if the data is noisy or non-stationary.

Spam filtering, fraud detection:
The enemy adapts so we must adapt too.
Recommendation systems:
Lots of noisy data. Million dollar prize!
Information retrieval:
Find documents or images with similar content.
Data Visualization:
Display a huge database in a revealing way

Why is Machine Learning Important?
Some tasks cannot be defined well, except by examples (e.g., recognizing people).
Relationships and correlations can be hidden within large amounts of data. Machine Learning/Data Mining may be able to find these relationships.
Human designers often produce machines that do not work as well as desired in the environments in which they are used.

The amount of knowledge available about certain tasks might be too large for explicit encoding by humans (e.g., medical diagnostic).
Environments change over time.
New knowledge about tasks is constantly being discovered by humans. It may be difficult to continuously re-design systems "by hand".
Areas of Influence for Machine Learning
Statistics: How best to use samples drawn from unknown probability distributions to help decide from which distribution some new sample is drawn?
Brain Models: Non-linear elements with weighted inputs (Artificial Neural Networks) have been suggested as simple models of biological neurons.
Adaptive Control Theory: How to deal with controlling a process having unknown parameters that must be estimated during operation?
Psychology: How to model human performance on various learning tasks?

Convolution Proporties

City-scale 3D reconstruction
Depth from a single image
Visualizing scenes from tourist photos

Reconstructing dynamic 3D scenes
Teaching assistants

Every image tells a story
Goal of computer vision: perceive the “story” behind the picture
Compute properties of the world

- 3D shape
- Names of people or objects
- What happened?

The goal of computer vision
Can computers match human perception?
Yes and no (mainly no)

- computers can be better at “easy” things
- humans are better at “hard” things

But huge progress

- Accelerating in the last five years due to deep learning
- What is considered “hard” keeps changing

Human perception has its shortcomings
But humans can tell a lot about a scene from a little information…
The goal of computer vision
Compute the 3D shape of the world

The goal of computer vision Recognize objects and people

The goal of computer vision “Enhance” images

The goal of computer vision Forensics
The goal of computer vision Improve photos (“Computational Photography”)

Why study computer vision?
Billions of images/videos captured per day
Optical character recognition (OCR)
Face detection
Nearly all cameras detect faces in real time

- (Why?)

Face analysis and recognition
Vision-based biometrics
Bird identification
Special effects: shape capture
Special effects: motion capture
3D face tracking w/ consumer cameras

Image synthesis
Which face is real?
Image synthesis
Sports
Smart cars
Mobileye
Tesla Autopilot
Safety features in many cars
Self-driving cars

Robotics
Medical imaging
Virtual & Augmented Reality
Current state of the art
You just saw many examples of current systems.
    Many of these are less than 5 years old

Computer vision is an active research area, and rapidly changing
    Many new apps in the next 5 years
    Deep learning and generative methods powering many modern applications

Many startups across a dizzying array of areas
    Generative AI, robotics, autonomous vehicles, medical imaging, construction, inspection, VR/AR, …
Why is computer vision difficult?
Challenges: local ambiguity
But there are lots of visual cues we can use…
Bottom line
Perception is an inherently ambiguous problem
    Many different 3D scenes could have given rise to a given 2D image

    We often must use prior knowledge about the world’s structure
2. Geometry & appearance
Project: Creating panoramas

Project: 3D reconstruction
3. Recognition, Deep Learning & Generative Models
Project: Neural Radiance Fields (NeRFs)

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs, and take actions or make recommendations based on that information.

The subareas of artificial intelligence include:

- ***Machine Learning***
- ***Computer Vision***
- ***Speech Recognition, Synthesis, and Understanding***
- ***Natural language processing***
- ***Robotics & Control***

Overview:

Goals of computer vision; why they are so difficult.

Computer vision seeks to generate intelligent and useful descriptions of visual scenes and sequences, and of the objects that populate them, by performing operations on the signals received from video cameras.

Examples of computer vision applications and goals:

- ***automatic face recognition, and interpretation of expression***
- ***visual guidance of autonomous vehicles***
- ***automated medical image analysis, interpretation, and diagnosis***
- ***robotic manufacturing: manipulation, grading, and assembly of parts***
- ***OCR: recognition of printed or handwritten characters and words***
- ***agricultural robots: visual grading and harvesting of produce***
- ***smart offices: tracking of persons and objects; understanding gestures***
- ***biometric-based visual identification of persons***
- ***visually endowed robotic helpers***

- *security monitoring and alerting; detection of anomaly*
- *intelligent interpretive prostheses for the blind*
- *tracking of moving objects; collision avoidance; stereoscopic depth*
- *object-based (model-based) compression of video streams*
- *general scene understanding*
- *The idea to impart human intelligence and instincts to a computer seems rather effortless. Conceivably, because it is solved by very young children too, but we often tend to forget the limitations of computers as compared to our biological capabilities. The complexity of vision perception infinitely varies and is ever dynamic in the case of human beings itself, let alone Computer intelligence.*
- *Our brain has the ability to identify the object, process data and decide what to do, thus completing a complex task in a split second. The aim is to enable Computers to be able to do the same. Hence, it is a field that can be referred to as an amalgamation of Artificial Intelligence and Machine Learning, which involves learning algorithms and specialized methods to interpret what the Computer sees.*
- *The Beginning*
- *Initially, the puzzling idea that tech giants still brainstorm about, was thought to be simple enough for an undergraduate summer project by the very people who pioneered Artificial Intelligence. Taking you back to 1966, when Seymour Papert and Marvin Minsky at MIT Artificial Intelligence group started a project in which the goal was to build a system that can analyze a scene and identify the objects in it.*
- *Deep Learning*
- *The Science behind Computer Vision revolves around artificial neural networks. In*
- *simple words? The algorithms inspired by the human brain that learn using large amounts of data sets so as to clone the human instincts as close as possible. These algorithms have superior accuracy, even surpassing human level in some tasks. Merely a subset of Deep Learning, Deep Vision is what drives Computer Vision.*

- *Pixel Extraction*
- *OpenCV (Open Source Computer Vision), a cross- platform and free to use library of functions is based on real time Computer Vision which supports Deep Learning*

frameworks that aids in image and video processing. In Computer Vision, the principal element is to extract the pixels from the image so as to study the objects and thus understand what it contains. Below are a few key aspects that Computer Vision seeks to recognize in the photographs:

- ***Object Detection: The location of the object.***
- ***Object Recognition: The objects in the image, and their positions.***
- ***Object Classification: The broad category that the object lies in.***
- ***Object Segmentation: The pixels belonging to that object.***

Applications and Future

Computer Vision covers a huge ground as its applications know no bounds. It often escapes our minds as we fail to notice the role Computer Vision plays in the gadgets, we use day in and day out.

- ***Smartphones and Web: Google Lens, QR Codes, Snapchat filters (face tracking), Night Sight, Face and Expression Detection, Lens Blur, Portrait mode, Google Photos (Face, Object and scene recognition), Google Maps (Image Stitching).***
- ***Medical Imaging: CAT/MRI***
- ***Insurance: Property Inspection and Damage analysis***
- ***Optical Character Recognition (OCR)***
- ***3D Model Building (Photogrammetry)***
- ***Merging CGI with live actors in movies***

Computer Vision is an ever-evolving area of study, with specialized custom tasks and techniques to target application domains. I visualize its market value growing as fast as its capabilities. With our intelligence and interest, we will soon be able to blend

our abilities with Computer Vision and achieve new heights.

Applications of Computer Vision

***Computer vision** or machine vision is a field of science that enables computers or devices to recognize different objects just like human beings. the computers need to be trained to detect objects and also some patterns just like you teach a kid to identify the objects but the computers are more efficient as it takes very little time to be trained. Computer vision has applications in all industries and sectors and they are as follows:*

1. ***Oil and natural gas: The oil and natural gas companies produce millions of barrels of oil and billions of cubic feet of gas every day but for this to happen, first, the geologists have to find a feasible location from where oil and gas can be extracted. To find these locations they have to analyze thousands of different locations using images taken on the spot. Suppose if geologists had to analyze each image manually how long will it take to find the best location? Maybe months or even a year but due to the introduction of computer vision the period of analyzing can be brought down to a few days or even a few hours. You just need to feed in the images taken to the pre-trained model and it will get the work done.***
2. ***Hiring process: In the HR world, computer vision is changing how candidates get hired in the interview process. By using computer vision, machine learning, and data science, they're able to quantify soft skills and conduct early candidate assessments to help large companies shortlist the candidates.***

Applications of Computer Vision

3. Video surveillance: The Concept of video tagging is used to tag videos with keywords based on the objects that appear in each scene. Now imagine being that security company who's asking to look for a suspect in a blue van amongst hours and hours of footage. You will just have to feed the video to the algorithm. With computer vision and object recognition, searching through videos has become a thing of the past.

4. ***Construction: Take for example the electric towers or buildings, which require some degree of maintenance to check for degrees of rust and other structural defects. Certainly, manually climbing up the tower to look at every inch and corner would be extremely time-consuming, costly, and***

dangerous. Flying a drone with wires around the electric tower doesn't sound particularly safe either. So how could you apply computer vision here? Imagine that if a person on the ground took high- resolution images from different angles. Then the computer vision specialist could create a custom classifier and use it to detect the flaws and amount of rust or cracks present.

5. ***Healthcare: From the past few years, the healthcare industry has adopted many next-generation technologies that include artificial intelligence and machine learning concept. One of them is computer vision which helps determine or diagnose disease in humans or any living creatures.***
6. ***Agriculture: The agricultural farms have started using computer vision technologies in various forms such as smart tractors, smart farming equipment, and drones to help monitor and maintain the fields efficiently and easily. It also helps improve yield and the quality of crops.***
7. ***Military: For modern armies, Computer Vision is an important technology that helps them to detect enemy troops and it also enhances the targeting capabilities of guided missile systems. It uses image sensors to deliver battlefield intelligence used for tactical decision-making. One more important Computer Vision application in the areas of autonomous*** *vehicles like UAV's and remote-****controlled semi- automatic vehicles, which need to navigate challenging terrain.***
8. ***Industry: In manufacturing or assembly line, computer vision is being used for automated inspections, identifying defective products on the production line, and for remote inspections of machinery. The technology is also used to increase the efficiency of the production line.***
9. ***Automotive: This is one of the best examples of computer vision technologies, which is a dream come true for humans. Self-driving AI analyzes data from a camera mounted on the vehicle to automate lane finding, detect obstacles, and recognize traffic signs and signals.***
10. ***Automated Lip Reading: This is one of the practical implementations of computer vision to help people with disabilities or who cannot speak, it reads the movement of lips and compares it to already known movements that were recorded and used to create the model.***

Properties of Convolution:

- Convolution is linear
- Convolution is shift-invariant
- Convolution is commutative (w*f = f*w)
- Convolution is *associative* (v*(w*f) = (v*w)*f)
- Every linear shift-invariant operation is a convolution

Convolution in computer vision and image processing is a fundamental operation used for various tasks such as image filtering, feature extraction, and neural network operations. Here's a detailed description of convolution in these contexts:

1. **Basic Concept**: At its core, convolution involves combining two functions to produce a third function. In the context of image processing, it's about applying a filter (also known as a kernel or mask) to an image. The filter is a small matrix of numerical values that is applied to each pixel in the image.

2. **Filtering Process**: When performing convolution on an image, the filter is centered on each pixel in the image, and a mathematical operation is applied to calculate the new value of that pixel based on its neighbors. This process moves across the entire image, pixel by pixel, applying the filter operation each time.

3. **Mathematical Operation**: The mathematical operation at each pixel is the dot product between the filter and the corresponding neighborhood of the image. This neighborhood is determined by the size of the filter (e.g., a 3x3 filter considers a 3x3 neighborhood around each pixel).

4. Image Filtering: Convolution is commonly used for image filtering, where different types of filters are applied to achieve specific effects. For example:

- Blur Filters: These filters average the pixel values in the neighborhood, resulting in a blurring effect that smoothens the image.

- Edge Detection Filters: Filters like the Sobel or Prewitt operators highlight edges in the image by emphasizing changes in pixel intensity.

- Sharpening Filters: These filters enhance edges and details in the image, making it appear sharper.

- Custom Filters: Users can design custom filters with specific numerical values to achieve desired effects like noise reduction, embossing, or artistic transformations.

5. Feature Extraction: In computer vision tasks such as object detection or image recognition, convolutional filters are used as feature extractors. Deep learning models, particularly Convolutional Neural Networks (CNNs), use layers of convolutional filters to learn hierarchical features from images, enabling tasks like classification, segmentation, and object localization.

6. Padding and Stride: When applying convolution, padding (adding extra pixels around the image) and stride (the step size of the convolution operation) can be adjusted to control the output size and the information captured from the input image.

7. Convolution vs. Cross-Correlation: In practice, convolution and cross-correlation are used interchangeably in image processing, although they have a mathematical difference in the way the filter is applied. In convolution, the filter is flipped horizontally and vertically before applying it to the image, while in cross-correlation, the filter is not flipped. However, in most image processing libraries and frameworks, the term "convolution" is commonly used to refer to both operations.

Overall, convolution plays a crucial role in computer vision and image processing by enabling various operations that enhance, extract features, and analyze images for a wide range of applications.

Let's work through a simple numerical problem to apply a mean filter (also known as a blur filter) on a small grayscale image. We'll use a 3x3 kernel for the mean filter.

Given The original sample of the image:

Original Image:

[10 20 30

40 50 60

70 80 90]

Apply the Mean Filter Kernel

[1/9 1/9 1/9

1/9 1/9 1/9

1/9 1/9 1/9]

Find the output result image?

Solution:

Let's apply the mean filter step by step:

1. First Pixel (10):

   - Place the filter over the pixel (10).

   - Calculate the average value using the filter:

    = (10+20+40+50)/9 = 120/9 approximately = 13.33

   - Replace the original pixel value with the new calculated value:

Image[0,0]=13.33

Second Pixel (20)

- Place the filter over the pixel (20).
- Calculate the average value using the filter:

= (10+20+30+40+50+60)/9=210/9=23.33

- Replace the original pixel value with the new calculated value: \(\text{New Image[0,1]=23.33

Continue this process for all pixels until the entire image is processed.

After applying the mean filter, the new image (resulting from the convolution) would look like this:

Blurred Image (Result):

[13.33	23.33	33.33
43.33	53.33	63.33
73.33	83.33	93.33]

In this example, each pixel in the blurred image is the average of its surrounding pixels in a 3x3 neighborhood, as calculated using the mean filter kernel.

Computer Vision - Image Formation

1. **What is the basic unit of a digital image?**

Answer:

- A **pixel** (picture element) represents a single point in the image with a specific color or intensity value.

2. **How does light interact with an object to create an image?**

   Answer:

- Light reflects off the object's surface and enters the camera lens.

3. **What component in a camera focuses the light?**

   Answer:

- The **lens** focuses the incoming light rays onto the image sensor.

4. **What captures the light information in a camera?**

   Answer:

- The **image sensor** (often a CMOS or CCD sensor) converts the incoming light into electrical signals.

5. **What are the different color channels used in digital images?**

   Answer:

- Common color channels include **Red (R), Green (G), and Blue (B)** (RGB) for color images, and a single channel for grayscale images.

6. **What process converts the analog electrical signals from the sensor into digital data?**

   Answer:

- **Analog-to-digital conversion (ADC)** transforms the continuous electrical signal into discrete digital values.

7. **What determines the resolution (sharpness) of a digital image?**

   Answer:

- The resolution depends on the total number of pixels in the image sensor.

8. **How does focal length affect the image captured?**

   Answer:

   - Focal length influences magnification – a shorter focal length captures a wider field of view, while a longer focal length zooms in on the scene.

9. **What is the difference between perspective and orthographic projection in image formation?**

   Answer:

   - **Perspective projection** creates a natural, diminishing effect with distance, like how we see the world. **Orthographic projection** shows objects with parallel lines regardless of depth, often used in technical drawings.

10. **What is the role of image formation in computer vision?**

   Answer:

   - Understanding image formation is crucial for computer vision tasks like object recognition and scene analysis. It helps interpret the captured data and extract meaningful information.

11. **What factors affect the amount of light captured by the sensor?**

   Answer:

   - Aperture size, shutter speed, and scene illumination all influence the amount of light reaching the sensor.

12. **What is the relationship between pixel size and image quality?**

   Answer:

   - Larger pixels generally capture more light, leading to better low-light performance and higher signal-to-noise ratio.

13. **What is image noise?**

Answer:

- Image noise refers to unwanted variations in pixel values that can appear as grain or speckles in the image.

14. **What is depth information in an image?**

    Answer:

- Depth information captures how far objects are from the camera, which is not directly available in a standard 2D image but can be estimated using techniques like stereo vision.

15. **What is image distortion?**

    Answer:

- Image distortion occurs when the lens bends light unevenly, causing straight lines to appear curved or objects to appear warped.

16. **What is the role of lighting in image formation?**

    Answer:

- Lighting significantly affects object appearance in the image. Consistent lighting conditions are preferred for robust computer vision algorithms.

1. Converting Cartesian to Homogeneous Coordinates:

A point in 2D space with coordinates (x, y) can be represented in homogeneous coordinates as (X, Y, W) where:

- X = x * W
- Y = y * W
- W can be any non-zero value (often set to 1 for simplicity)

Example:

Convert (2, 3) to homogeneous coordinates.

Solution:

(2, 3, 1)

3. Translating a Point:

To translate a point (X, Y, W) by a vector (tx, ty):

- X' = X + tx * W
- Y' = Y + ty * W
- W' = W

Example:

Translate point (1, 2, 1) by (3, 1).

Solution:

(4, 3, 1)

4. Rotation in the Image Plane:

To rotate a point (X, Y, W) around the origin by angle θ:

- X' = X * cos(θ) - Y * sin(θ)
- Y' = X * sin(θ) + Y * cos(θ)
- W' = W

Example:

Rotate point (3, 4, 1) by 45 degrees.

Solution:

(≈ 2.12, ≈ 5.30, 1) (due to rounding from cosine and sine calculations)

5. Line Representation in Homogeneous Coordinates:

A line can be represented by the equation:

aX + bY + cW = 0

Example:

Find the equation of the line passing through points (1, 2, 1) and (3, 1, 1).

Solution:

Solve for a, b, and c using the point coordinates. (This might involve linear algebra)

Problem 1: Scaling and Translating an Object

A rectangle in an image has corner points A (2, 3, 1), B (5, 3, 1), C (5, 1, 1), and D (2, 1, 1) in homogeneous coordinates.

You want to:

- Scale the rectangle by a factor of 2 in both width and height.
- Translate the scaled rectangle 3 units to the right and 2 units down.

Solution:

1. **Scaling:**
   - For each point (X, Y, W):
     - X' = X * scale_factor (here, 2)
     - Y' = Y * scale_factor
     - W' = W (since scaling doesn't affect the homogeneous component)
   - Apply scaling to all points:
     - A' = (4, 6, 1)
     - B' = (10, 6, 1)
     - C' = (10, 2, 1)
     - D' = (4, 2, 1)
2. **Translation:**
   - Define a translation vector (tx, ty). Here, tx = 3 and ty = -2.
   - For each scaled point (X', Y', W'):
     - X'' = X' + tx * W'
     - Y'' = Y' + ty * W'
     - W'' = W' (translation doesn't change the homogeneous component)
   - Apply translation to all scaled points:

- A'' = (4 + 3 * 1, 6 - 2 * 1, 1) = (7, 4, 1)
- B'' = (10 + 3 * 1, 6 - 2 * 1, 1) = (13, 4, 1)
- C'' = (10 + 3 * 1, 2 - 2 * 1, 1) = (13, 0, 1)
- D'' = (4 + 3 * 1, 2 - 2 * 1, 1) = (7, 0, 1)

Therefore, the final corner points of the transformed rectangle in homogeneous coordinates are A'' (7, 4, 1), B'' (13, 4, 1), C'' (13, 0, 1), and D'' (7, 0, 1).

Numerical Problems in Computer Vision - Image Formation

1. **Pinhole Camera Model:**

- A camera has a focal length of 50mm and a sensor size of 36mm x 24mm. What is the field of view (FOV) for an object 1 meter away from the camera?
  Answer:
  - Use the formula: FOV = 2 * arctan(sensor_size / (2 * focal_length))
  - Convert sensor size to mm: FOV = 2 * arctan((36 * 24) / (2 * 50)) = 73.9 degrees

2. **Depth of Field:**

- A camera with an aperture of f/8 is focused on an object 2 meters away. What is the hyper focal distance and the near and far depth of field for a desired circle of confusion of 0.01mm?

- **Answer:**
  - Use the hyper focal distance formula: H = (f^2 * N) / (d_o * c)
    - f (aperture) = 8, N (circle of confusion) = 0.01mm, d_o (object distance) = 2m
    - H = (8^2 * 0.01) / (2 * 0.01) = 32 meters
  - Near and far depth of field can be calculated using hyperfocal distance and object distance:
    - Near = H * (d_o^2) / (H + d_o) = 1.58 meters

    - Far = H * (H + d_o) / d_o^2 = 48.42 meters

3. **Lens Distortion:**

- A camera exhibits barrel distortion with a distortion coefficient of k1 = 0.1. What is the corrected image position for a point originally located 5% away from the image center?
  Answer:
  - Use the barrel distortion correction formula: x_corrected = x * (1 + k1 * d^2)
    - x (original position) = 0.05 * image_width
    - d (normalized distance from center) = x / (image_width / 2)
  - Solve for x_corrected to find the corrected position.

4. **Radiometric Calibration:**

- A grayscale image has a pixel value of 128 at a known reflectance of 0.5. What is the scaling factor needed to convert pixel values to actual reflectance values?
- **Answer:**
  - Scaling factor = Reflectance / Pixel value = 0.5 / 128 = 0.0039

5. **Color Image Formation (Bayer Filter):**

- A camera uses a Bayer filter (RGGB pattern). How are red, green, and blue channel values reconstructed for a specific pixel in the image?
- **Answer:**
  - The pixel value itself represents the green channel.
  - Red and blue values are interpolated from neighboring pixels based on the Bayer filter pattern.
  - Different interpolation techniques like bilinear or nearest neighbor can be used.

6. **Image Sensor Noise:**

- An image sensor has a read noise of 5 electrons and a shot noise standard deviation proportional to the square root of the signal level. If a pixel has a signal level of 10,000 electrons, what is the total noise level (standard deviation)?
- **Answer:**
  - Shot noise standard deviation = sqrt(signal_level) = sqrt(10,000) = 100 electrons
  - Total noise (standard deviation) = sqrt(read_noise^2 + shot_noise^2) = sqrt(5^2 + 100^2) = 100.49 electrons

7. **Exposure and Image Quantization:**

- A scene with a brightness range of 100:1 needs to be captured on a camera with a dynamic range of 8 stops. How many bits per pixel are needed for proper quantization? **Answer:**
  - Each stop represents a doubling of brightness. 8 stops correspond to 2^8 = 256 levels.
  - To represent the full scene range within these levels, need enough bits to express 256 values: 8 bits/pixel.

8. **Image Compression and Quantization Error:**

- An image is compressed using JPEG with a quality factor of 80%. How does the quantization error affect the image quality, and how can it be measured?
- **Answer:**
  - Lower quality factors lead to higher quantization errors, resulting in loss of detail and artifacts.
  - Quantization error can be measured using metrics like Peak Signal-to-Noise Ratio (PSNR) or Mean Squared Error (MSE) between the original and compressed image.

9. **Gamma Correction:**

- A monitor has a gamma value of 2.2. How is the displayed intensity related to the actual intensity stored in the image?
- **Answer:**

- Gamma correction applies a power-law transformation to the image data.
- Displayed intensity = (Actual intensity)^(1/gamma)
- In this case, displayed intensity = (Actual intensity)^(1/2.2)

10. Color Space Conversion:

- An image is stored in RGB color space. How can it be converted to HSV color space for analysis?
- **Answer:**
  - Use conversion formulas involving RGB values to calculate Hue, Saturation, and Value components of HSV.
  - Various libraries and tools provide functions for color space conversion.

11. Illumination Effects:

- An image is captured under uneven lighting conditions. How can image normalization techniques be used to compensate for these variations?
- **Answer:**
  - Techniques like histogram equalization or normalization based on reference regions can be applied.
  - The goal is to adjust pixel values to achieve a more consistent intensity distribution across the image.

12. Focus Measure Functions:

- How can Laplacian or Sobel filter be used as focus measure functions to determine if an image is in focus?
- **Answer:**
  - In-focus images have sharper edges, resulting in higher responses from filters like Laplacian or Sobel that detect image gradients.
  - The focus measure function will have a higher value for a well-focused image compared to a blurry one.

13. **Camera Calibration:**

- A camera calibration process estimates intrinsic parameters like focal length and distortion coefficients. How is this information used in computer vision tasks?
- **Answer:**
  - Camera calibration parameters are used to correct image distortions and perform accurate 3D reconstruction from images.
  - This is crucial for tasks like object recognition, pose estimation, and augmented reality.

14. **Image Denoising:**

- An image captured in low-light conditions exhibits significant noise. How can techniques like median filtering or non-local means filtering be used for denoising?
- **Answer:**
  - Median filtering replaces a pixel value with the median of its neighborhood, suppressing noise while preserving edges.
  - Non-local means filtering utilizes similar image patches to denoise a pixel, achieving better preservation of details.

15. **Super-Resolution:**

- A low-resolution image needs to be enhanced to improve its resolution. How can techniques like bicubic interpolation or deep learning-based approaches be used for super-resolution?
- **Answer:**
  - Bicubic interpolation estimates missing pixels based on surrounding pixels in the low-resolution image.
  - Deep learning approaches learn to generate a high-resolution image from a low-resolution one by leveraging largae training datasets.

16. **Image Segmentation:**

- An image contains multiple objects that need to be segmented (identified and separated). How can thresholding or graph-based segmentation techniques be applied?
  Answer:
  - *Thresholding segments the image based on pixel intensity values.*
  - *Graph-based segmentation treats pixels as nodes and edges based on similarity, then segments based on connected*

<u>The cross product of (1, 2, -3) and (3, -1, 1)</u>

To find the cross product of two vectors, let's denote the first vector as

$v=(1,2,-3)$ and the second vector as $u=(3,-1,1)$. The cross product $v\times u$ is calculated as follows: $v\times u$ is

$$\mathbf{v} \times \mathbf{u} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \mathbf{1} & \mathbf{2} & \mathbf{-3} \\ \mathbf{3} & \mathbf{-1} & \mathbf{1} \end{vmatrix}$$

$$= i(2-3) - j(1+9) + k(-7)$$

$$(-1, -10, -7)$$

Field of view calculation

The relationship among AFOV, focal length and sensor size can be represented by the following equation:

AFOV = 2 x tan-1 (H/2f)

Here:

- H = sensor size (horizontal dimension)
- f = focal length of the lens

How to Calculate the FOV

To calculate the FOV requires the sensor size and the focal length of the lens: h = Sensor Size F = Focal Length of the Lens

FOV is represented by this equation: FOV = 2tan-1(*h*) / 2F

Example: h = 4.7mm F = 6mm

FOV = $2\tan^{-1}((h) / 2F) = 2\tan^{-1}(4.7/12) = 2\tan^{-1}(0.39) = 2(21.4°) = 42.8°$

For the following standard transformations, Rotation around X, Translation, and Scaling, Derive homogeneous coordinate transformation matrices?

Translation

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & d_x \\ 0 & 1 & 0 & d_y \\ 0 & 0 & 1 & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$x' = x + d_x$$
$$y' = y + d_y$$
$$z' = z + d_z$$

Rotation Matrix

$$\mathbf{R} = \mathbf{R}_Z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Rotation Matrix

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$x' = x\cos\theta - y\sin\theta$$
$$y' = x\sin\theta + y\cos\theta$$
$$z' = z$$

$$\mathbf{R} = \mathbf{R}_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R} = \mathbf{R}_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Scaling

or contract along each axis (fixed point of origin)

$$x' = s_x x$$
$$y' = s_y y$$
$$z' = s_z z$$

$$\mathbf{p}' = \mathbf{S}\mathbf{p}$$

$$\mathbf{S} = \mathbf{S}(s_x, s_y, s_z) = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$