Computer Vision

Dr. Syed Faisal Bukhari
Associate Professor
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab

Textbooks

Multiple View Geometry in Computer Vision, Hartley, R., and Zisserman

Richard Szeliski, **Computer Vision: Algorithms and Applications**, 2nd edition, 2022

Reference books

Readings for these lecture notes:

- Hartley, R., and Zisserman, A. **Multiple View Geometry in Computer Vision**, Cambridge University Press, 2004, Chapters 1-3.

- Forsyth, D., and Ponce, J. **Computer Vision: A Modern Approach**, Prentice-Hall, 2003, Chapter 2.

These notes contain material c Hartley and Zisserman (2004) and Forsyth and Ponce (2003).

References

These notes are based

- Dr. Matthew N. Dailey's course: AT70.20: Machine Vision for Robotics and HCI
- Dr. Sohaib Ahmad Khan CS436 / CS5310 Computer Vision Fundamentals at LUMS
- https://www.photographytalk.com/beginner-photography-tips/7204-focal-length-and-field-of-view-explained-in-4-steps
- https://ipvm.com/reports/testing-wide-vs-telephoto-fov#:~:text=Testing%20Wide%20vs%20Narrow%20FoV&text=The%20first%2C%20the%20'wide',of%20the%20distant%20parking%20lot.
- http://rimstar.org/science_electronics_projects/pinhole_camera.htm
- https://www.lorextechnology.com/self-serve/guide-to-field-of-view-lens-types/R-sc2900041

Grading breakup

I. Midterm = 35 points
II. Final term = 40 points
III. Quizzes = 6 points (A total of 6 quizzes)
IV. Group project = 15 points
   a. Pitch your project idea = 2 points
   b. Research paper presentation relevant to your project = 3 points
   c. Project prototype and its presentation = 5 points
   d. Research paper in IEEE conference template = 5 points
V. OpenCV based on Python presentation = 2.5 points
VI. Matlab presentation = 2.5 points

Some top tier conferences of computer vision

I. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition **(CVPR).**
II. Proceedings of the European Conference on Computer Vision **(ECCV).**
III. Proceedings of the Asian Conference on Computer Vision **(ACCV).**
IV. Proceedings of the International Conference on Robotics and Automation **(ICRA).**
V. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems **(IROS).**

Some well known Journals

I. International Journal of Computer Vision **(IJCV).**
II. IEEE Transactions on Pattern Analysis and Machine Intelligence **(PAMI).**
III. Image and Vision Computing.
IV. Pattern Recognition.
V. Computer Vision and Image Understanding.
VI. IEEE Transactions on Robotics.
VII. Journal of Mathematical Imaging and Vision

Why do we like to use vectors or column vectors in particular?

- A **linear transformation** between **vector spaces** is represented using **matrices**, allowing us to express the transformation between one vector space and another as **matrix multiplication**.

  Example:

  $$\vec{x}' = A\vec{x}$$

- Here A is a matrix, and $\vec{x}$ is a **2D vector**. If we are considering a **2D transformation**, then $\vec{x}'$ will also be a **2D vector**.

- We can have **transformations** between **different vector spaces** with **varying dimensions**.

How you know the size of $\vec{x}$'?

$\vec{x}' = A\vec{x}$ e.g., $\vec{x}'_{2\times1} = A_{2\times2}\vec{x}_{2\times1}$

- If A is **2 × 2,** then we are transforming from **2D space** to another **2D space**.

  A: $\mathbb{R}^2 \mapsto \mathbb{R}^2$

- The number of **columns of A** must be equal to the **number of rows of** $\vec{x}$.

- The number of **rows in A** must be equal to number of rows in $\vec{x}'$.

- We will later discuss the **types of transformations** that can be modeled by a **2×2** matrix as **linear multiplications** between this matrix and a two-dimensional vector.

2D projective geometry
The 2D projective plane: lines in $\mathbb{R}^2$

- A **line** in the plane is typically represented by **an equation** such as

  ax + by + c = 0.

- The parameters **a, b**, and **c** determine different lines.

- This means we can represent a line in $\mathbb{R}^2$ as the vector $(a, b, c)^T$.

Lines in the plane

- **Slope-y intercept form or slope intercept form**:

$y = mx + c$

- We interpret this equation as follows: **x represents the x-coordinate** of a point, and **y represents the y-coordinate**. Therefore, we have a 2D point, which defines a relationship between x and y
- A line in a plane represents a linear relationship between x and y

$$y = \underbrace{m}_{\textbf{Slope}}x + \underbrace{c}_{\textbf{y-intercept}}$$

m is called the **slope** of a line and **c** is called the **y-intercept**.

Slope of a line

For example

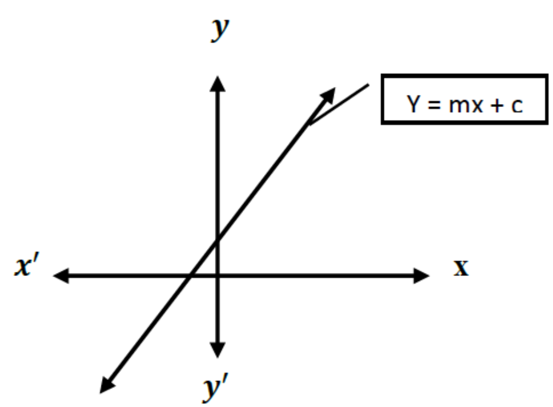
$y = 3x + 2$

Here **m =** $\mathbf{\frac{3}{1}}$ and **c = 2**

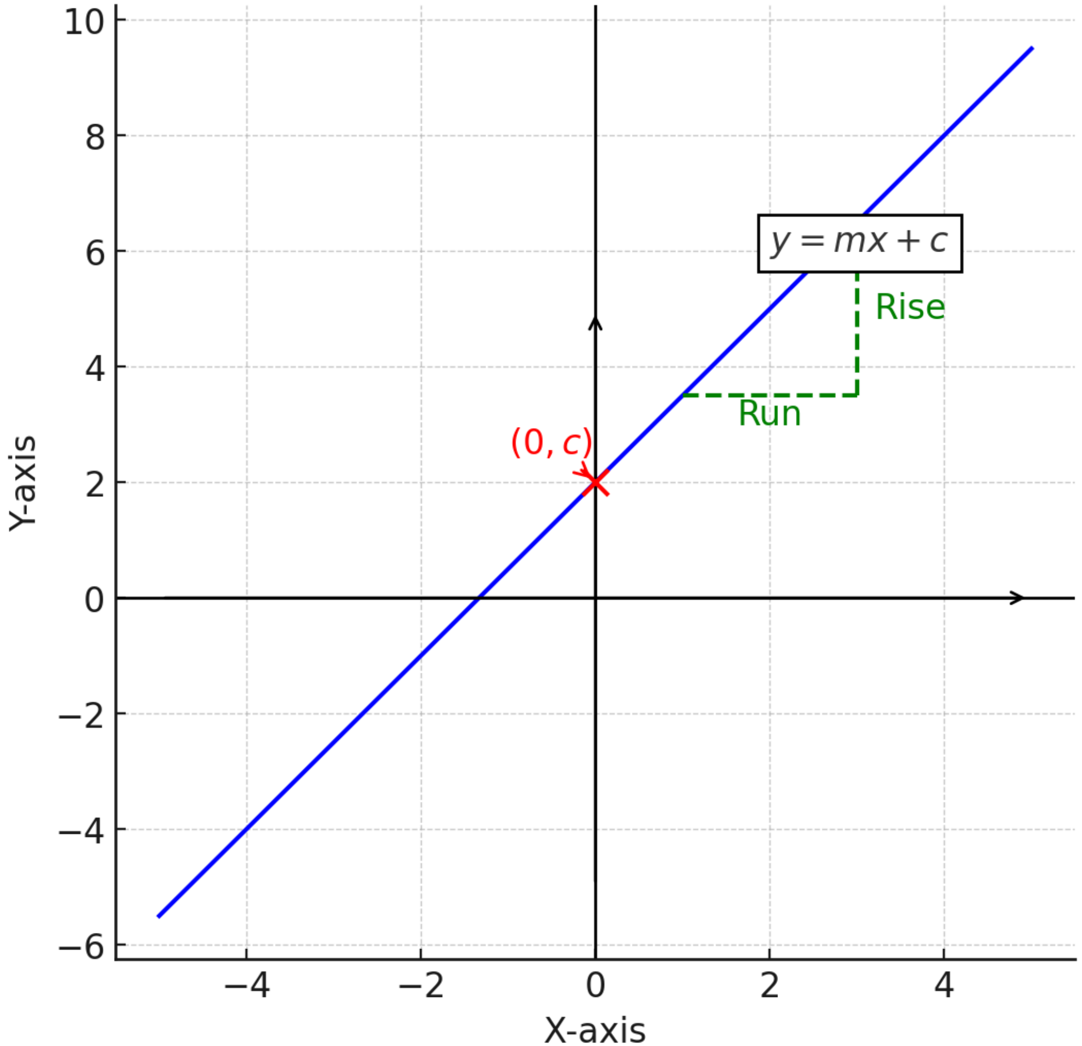
How do we interpret the **slope**? It is the **rise over the run**. The **rise** represents how far you move in **the y-direction**, while the **run** represents how far you move in **the x-direction.**

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\text{rise}}{\text{run}} = \frac{3}{1}$$

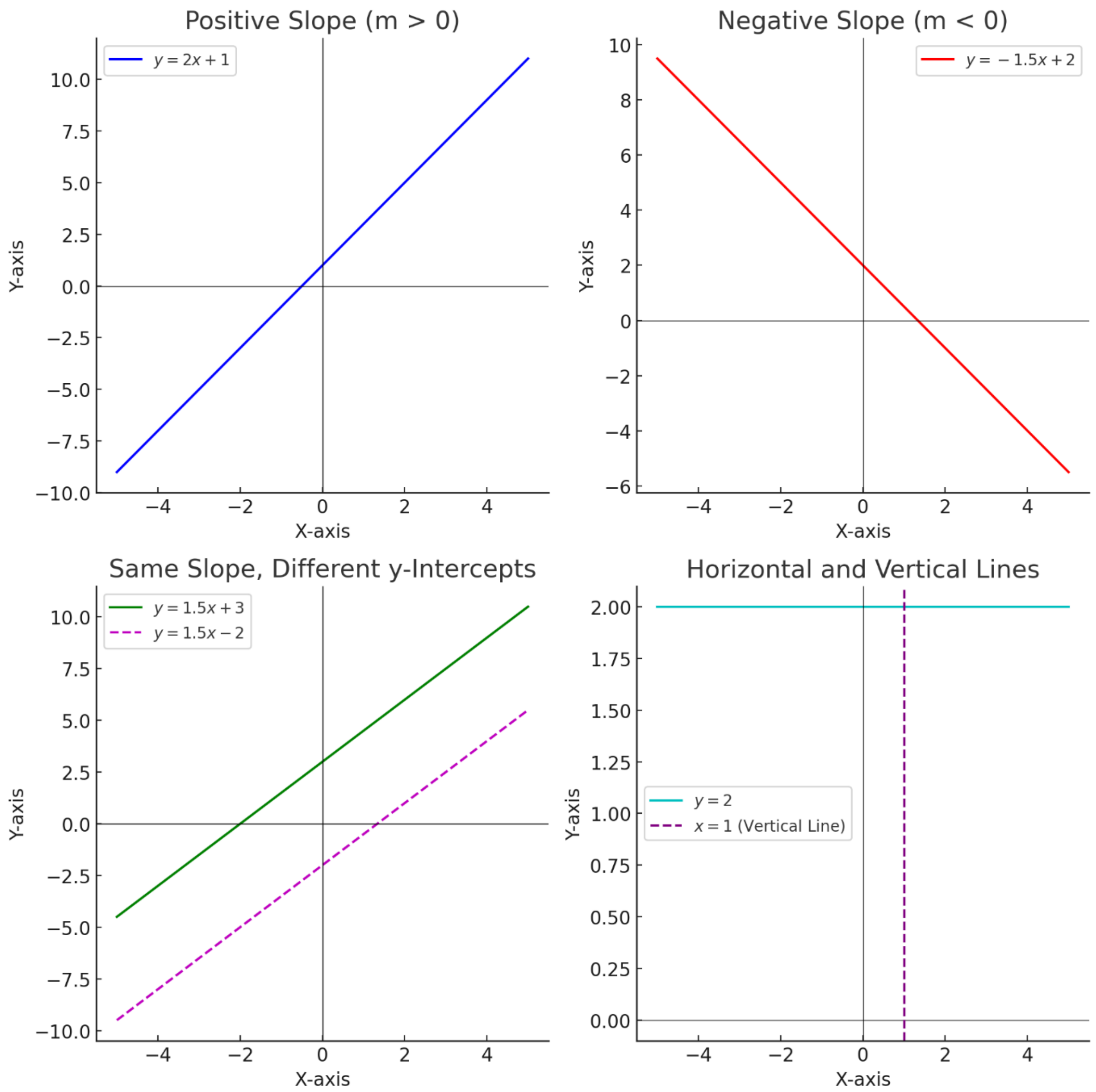
c = 2 **means that the line intercepts the y-axis at 2.**



Slope-Intercept Form: A Graphical Illustration



Slope-y intercept form



What's wrong with the slope-intercept form of a line?

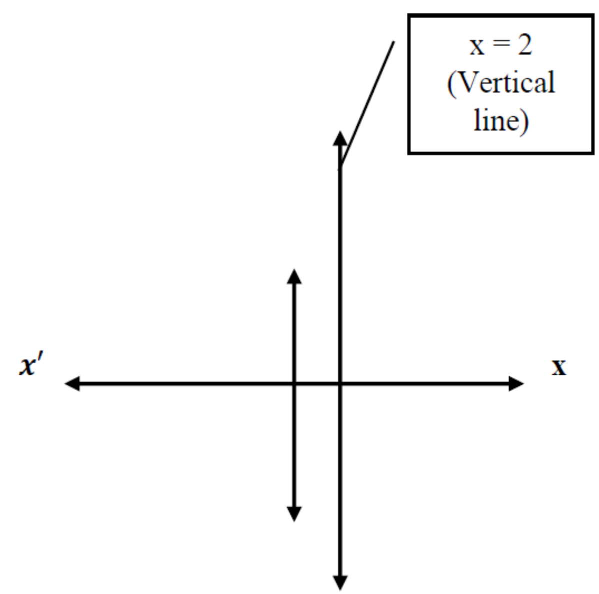
- **Vertical lines** cannot be represented in **slope-intercept form** because they have an **infinite slope**.

- This means we have a **rise** of **some amount** divided by a **run of zero**, resulting in an **infinite slope.**

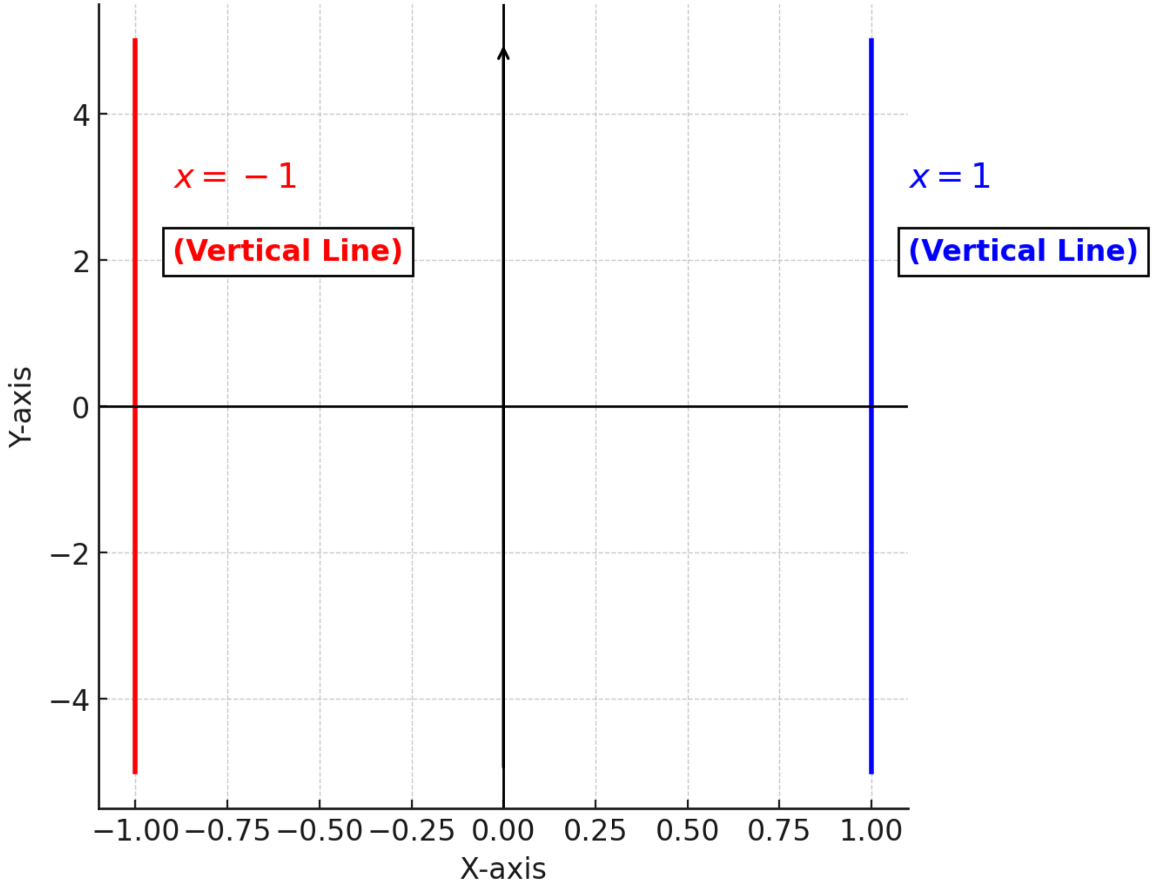
For example

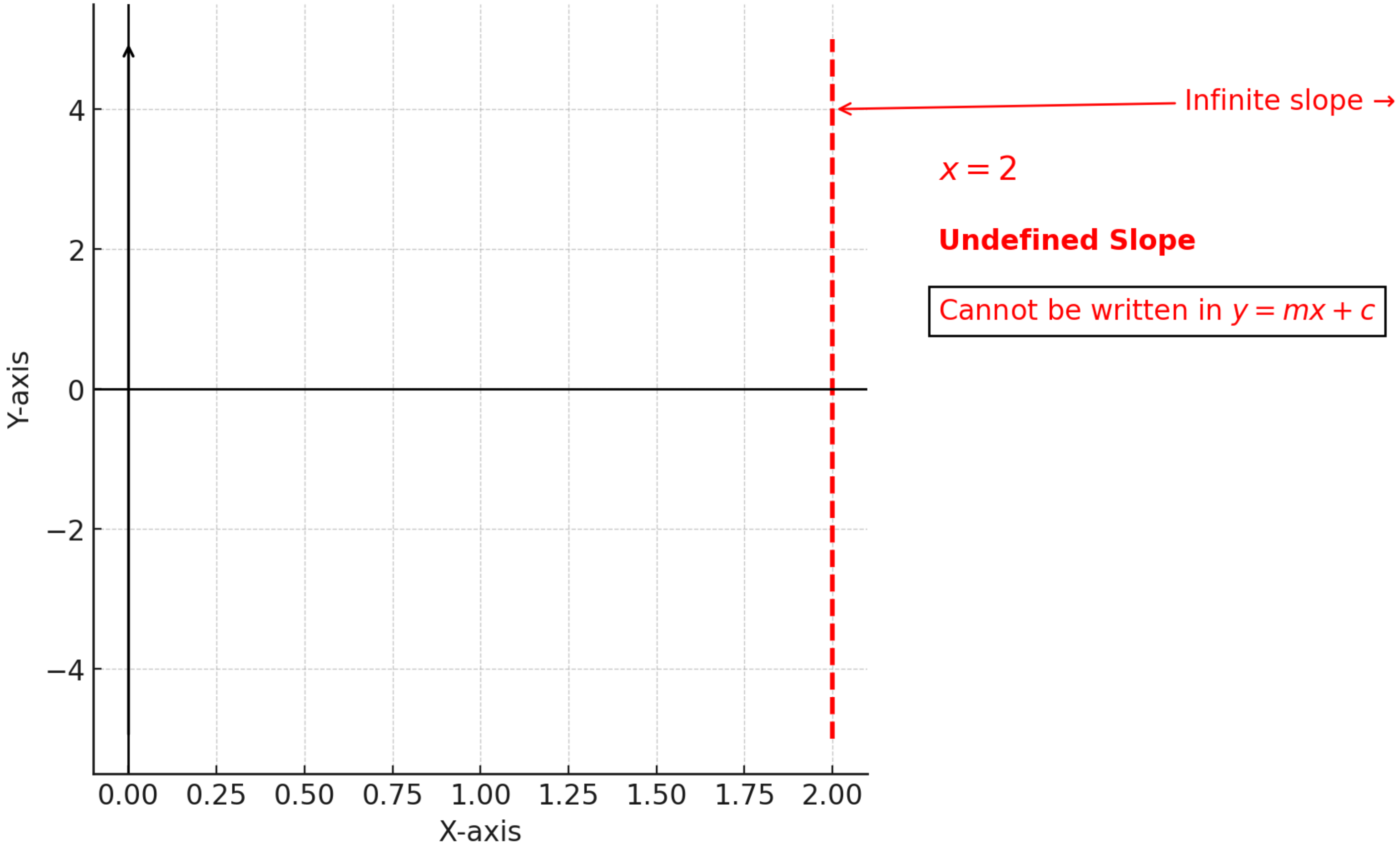
x = 2

$$m = \frac{\text{rise}}{\text{run}} = \frac{2}{0} = \infty$$



Vertical lines cannot be represented in slope-intercept form y = mx + c because they have an undefined (infinite) slope





Infinite slope →
$x = 2$
Undefined Slope
Cannot be written in $y = mx + c$
4
2
0
−2
−4
Y-axis
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
X-axis

General equation of a line

$ax + by + c = 0$

where **a, b,** and **c** are arbitrary scalars. Now, we can represent a **vertical line** using the **general equation of a line**.

For example:

$x = 2$

$\Rightarrow 1.x + 0.y - 2 = 0$

$a = 1$, $b = 0$ and $c = -2$

So, $\mathbf{ax + by + c = 0}$ can model any line in the plane.

Parametric vector from of a line [1]

We can write parameters a, b, and c of a general equation of a line in the vector form as

$\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ or $(\mathbf{a}, \mathbf{b}, \mathbf{c})^T$

For example:

mx – y + b = 0

If we divide above equation by m, then we still have the same line.

x – (1/m) y + b/m = 0

a = 1, b = -1/m, and c = b/m

Where a, b, and c are valid scalars.

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ -\mathbf{1}/\boldsymbol{m} \\ \boldsymbol{b}/\boldsymbol{m} \end{bmatrix}$$

Parametric vector from of a line [2]

y = 2x + 3

We can represent the **slope-intercept form** using the **parametric vector form**

2x – y + 3 = 0

a = 2, b = -1, and c = 3

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \mathbf{2} \\ \mathbf{-1} \\ \mathbf{3} \end{bmatrix} \text{ or } (\mathbf{a}, \mathbf{b}, \mathbf{c})^{\boldsymbol{T}} = (\mathbf{2}, \mathbf{-1}, \mathbf{3})^{\boldsymbol{T}}$$

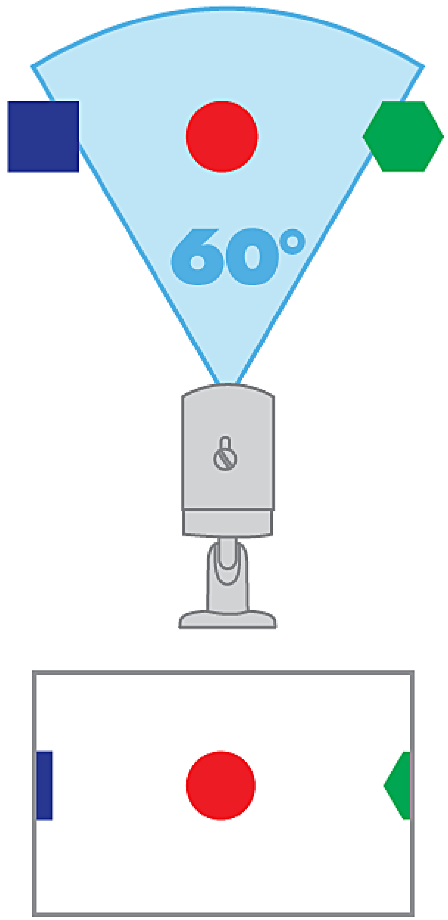
What is Field of View (FOV)?

- The **field of view** of a security camera, also called the **viewing angle**, is the **area that the camera can see**.

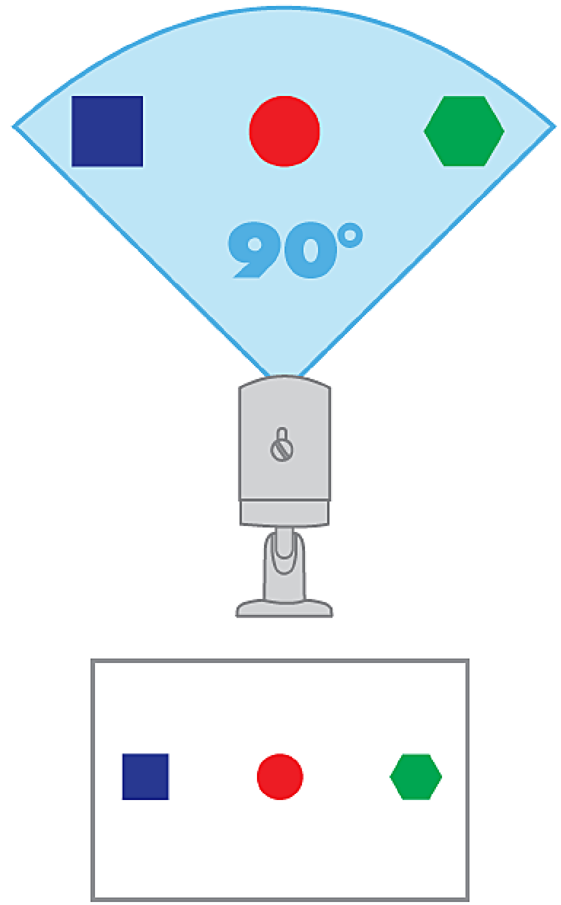
- On a specification sheet, you will see the **field of view measured in degrees**.

- Think of the field of view as the angle between the two horizontal edges of the camera image.

What is Field of View (FOV)?



The 60° field of view captures some of the objects but in greater detail



The 90° field of view captures all of the objects but in lesser detail

What is Field of View (FOV)?

- As you can see in the example images in the previous slides, the camera with **90° field of view captures all 3 objects** in the scene, though each object takes up a small part of the camera image.

- The camera with **60° field of view captures some of the objects** but in greater detail.

- Remember, a **wider field of view** isn't always better!

Wide-angle Lens

Wide-angle lenses

- **Smaller lenses** are known as **wide-angle lenses**, which produce a greater field of view than cameras with a larger lens. They capture a large area, though objects will appear smaller within the camera image. **Wide-angle lenses** are designed for **monitoring large areas**, such as:

Foyers

Warehouses

Back or Front Yards

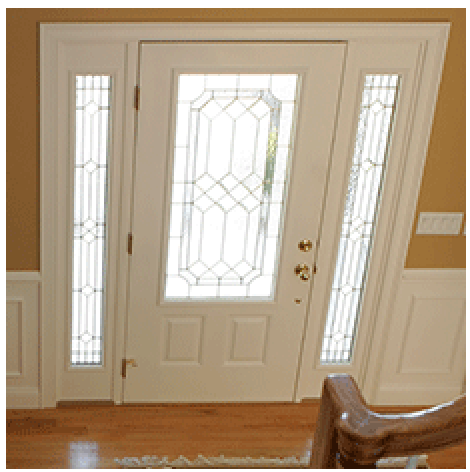
Parking Lots

Narrow-angle Lens

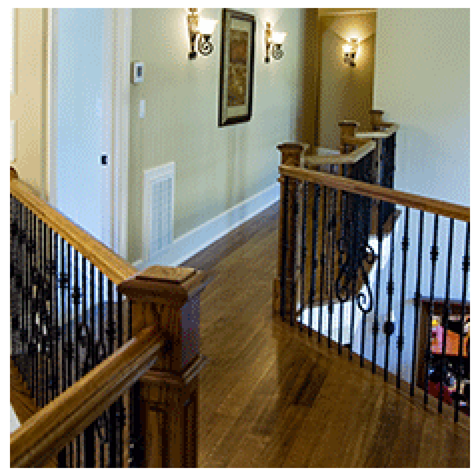
Narrow-angle Lens

- Larger lenses, or **narrow-angle lenses**, have a **smaller field of view**. They capture a **limited area**, but objects **will appear larger and more detailed** within the **camera image**. **Narrow-angle lenses** are designed for **monitoring a specific target**, such as:

Cash Registers



Doorways and Entrances



Hallways

Objects of Value

How is lens size related to field of view?

- The size of the camera lens, or **focal length**, is the main factor that determines **the field of view**.

- The example images shown in the next slides compare multiple focal lengths and the resulting fields of view:

How is lens size related to field of view?

Focal length: 3.6mm
Field of view: 78°

Focal length: 5.1mm
Field of view: 58°

Focal length: 6mm
Field of view: 51°

Focal length: 9mm
Field of view: 39°

Lens Field of View

- Just as our eyes serve as windows to the outer world, **lenses are the eyes of a camera**, allowing it to capture what we see.

- Similarly, just as each person has different eyes with varying capabilities, **lenses also differ in their characteristics**. This means that what one person perceives might not appear the same to another**—the same principle applies to camera lenses.**

- Some lenses have a **short focal length**, providing a **wide angle of view**, while others have a **long focal length**, resulting in a **narrow angle of view**.

Key Points on Field of View

1. Focal Length vs. Field of View:

- Focal length determines how long a lens is.
- Field of view (FoV) defines how much of the scene a lens captures.

2. Understanding Field of View:

- FoV tells us how much of a scene is visible through a lens.
- More useful than focal length alone.

Key Points on Field of View

3. Factors Affecting Field of View:

FoV changes based on:

- Focal length of the lens (shorter = wider, longer = narrower).
- Size of the camera sensor (larger sensor = wider FoV).

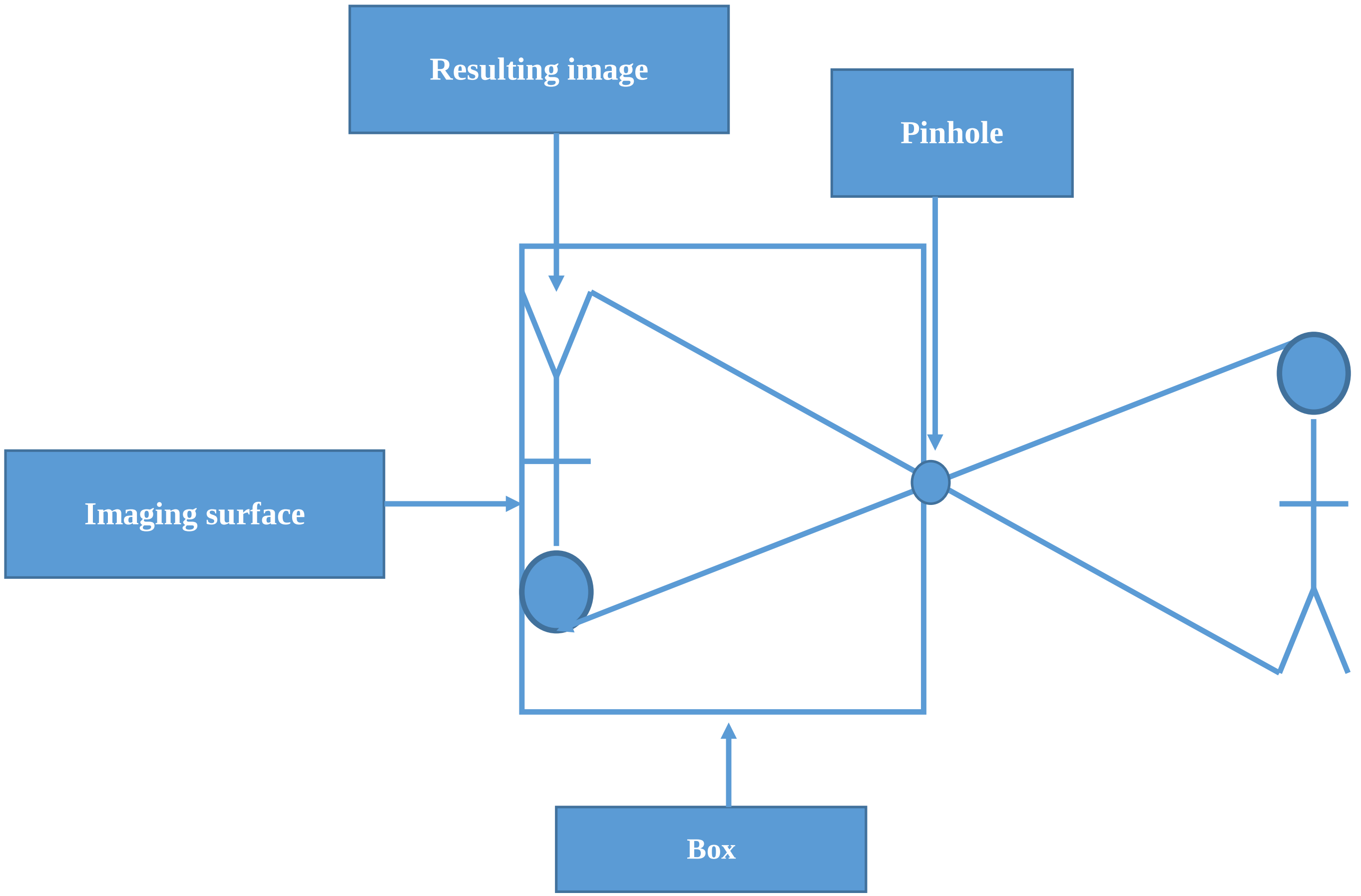
4. Challenges in Measuring Field of View:

- FoV varies with sensor size.
- Manufacturers often specify lenses by focal length rather than FoV.

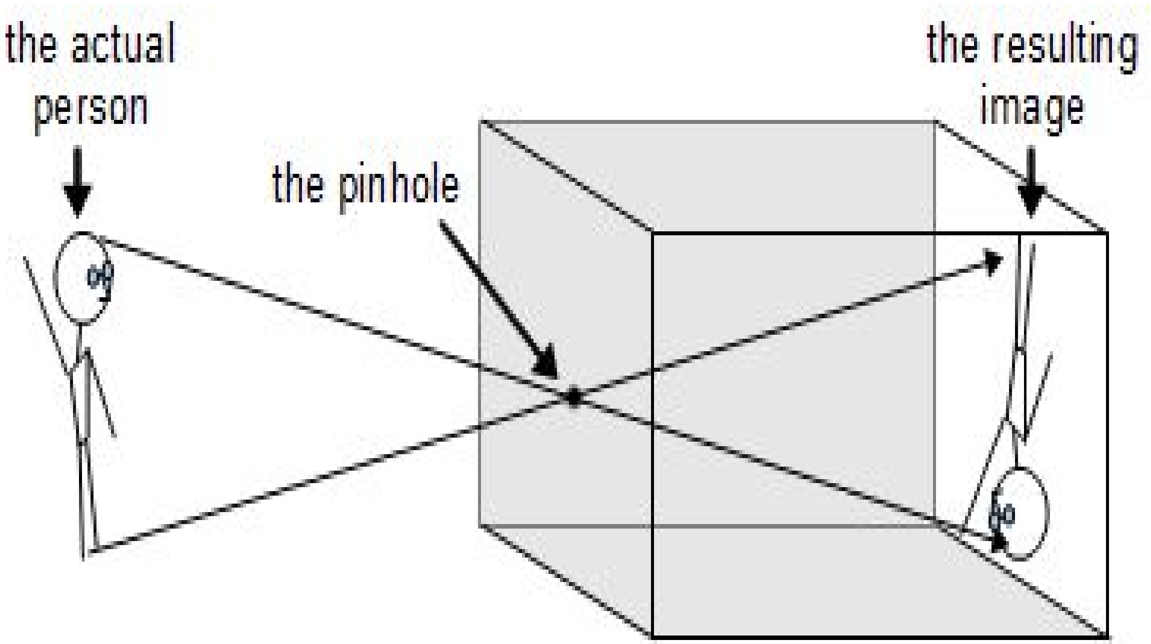

Wide FoV
Narrow FoV

Pinhole model

- The classical model of camera is called a pinhole model.



Pinhole model



Pinhole model

- In the **pinhole model of a camera**, what happens?
- We have an **infinite set of small points** emitting light into a box.
- Points in a **scene reflect light**.
- Some of these **light rays pass** through the **pinhole** in the box
- The image of the scene is then **reconstructed** on the back of the box.

How to build a pinhole camera?

- Simply take a box, poke a small hole in it, and point it in any direction. If you could open the box and look inside, you would see an image of the **scene projected onto the back of the box**.

- However, **opening the box** lets in too **much light**, which can **ruin the image**. This is how a pinhole camera works.

- In a pinhole camera, the image is **inverted**.

How to build a pinhole camera? (cont.)

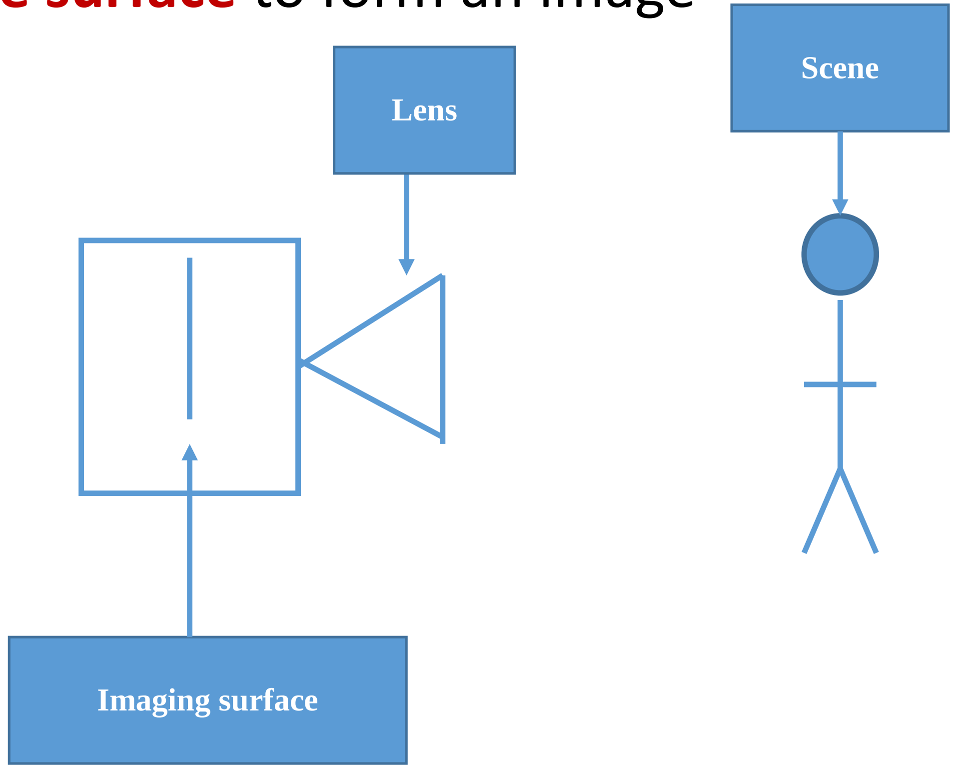
- Light reflected from the **top of a person's head** passes through the pinhole and appears **lower** on the back of the box.

- Similarly, light reflecting from the **person's foot** passes through the pinhole and appears **higher** on the back of the box.

- As a result, the **image of the scene appears upside down** on the back of the box.

Camera model

- A **modern camera** consists of a **lens** and an **imaging surface**.
- When we examine a camera, we see a **lens** on the front, and inside, there is an **image surface** where the scene is captured.
- If there is a **scene**, the camera's role is to **project the scene** onto the **image surface** to form an image



Resemblance Between a Pinhole and a Modern Camera [1]

- The **pinhole camera** serves as the **fundamental model**, and **modern cameras** are designed to emulate its principles. But why?

- Since this **principle applies to real cameras**, they share **similarities with the pinhole model**.

- However, modern cameras use a **lens to focus light** onto a specific point, improving image clarity and brightness.

Resemblance Between a Pinhole and a Modern Camera [2]

- **How Light is Processed in a Camera**

  - The **lens collects light** and focuses it on a **focal point**.

  - The light is then **spread out and captured** by a **sensor array** or an **imaging surface**.

Resemblance Between a Pinhole and a Modern Camera [3]

- **Camera Sensors**

o The imaging surface could be a **CCD (Charge-Coupled Device)** or a **CMOS (Complementary Metal-Oxide-Semiconductor) sensor**.

o A variety of **semiconductor technologies** are used to construct these cameras.

Resemblance in a pinhole and a modern camera [4]

- **How Light is Captured in a Camera**

  - The **lens** collects light and focuses it on a **focal point**.

  - This light is then captured by the **sensor array,** which is structured as a **2D grid of picture elements (pixels)**.

- **Similarity to the Pinhole Model**

  - The **pinhole model** closely resembles how **modern cameras** function today.

Note: "Lens" is singular. **"Lenses"** is plural. "Lense" does not exist as an accepted word in English

Why do we use lens instead of pinhole? [1]

- If the pinhole is not an **infinitesimally small point**, some **scattering of light may occur** as it passes through the hole.

- **The smaller the hole**, the less light that passes through, resulting in a dimmer image.

Why do we use lens instead of pinhole? [2]

- The **main role of a lens** in **computer vision** or **optics** is to:
  - **Collect more light**
  - **Focus light more accurately** on a particular point
  - **Produce sharper images** with **better contrast**

- Thus, achieving **high contrast** is a key advantage of using a **lens** over a pinhole.

Why do we use lens instead of pinhole? [3]

- In fact, from a machine vision point of view, we're going to see that **the pinhole model** is **mathematically attractive**. However, building a **perfect pinhole camera** is **physically impossible**.

Limitations of the Pinhole Model and the Use of Lenses [1]

- **Why do we use cameras with lenses?**
  - We aim to use **high-quality lenses** to improve image capture.
  - When the **field of view is narrow**, the resulting image is **mathematically similar** to what we get from a **pinhole camera**.

- **Challenges with a Wide Field of View:**
  - If the **field of view is very wide**, it becomes **difficult** to construct a proper pinhole image.
  - **Distortions** may occur, which need **corrections** in real-world imaging.

Limitations of the Pinhole Model and the Use of Lenses [2]

- **Limitations of the Pinhole Model:**

  - The **pinhole model** does not perfectly represent how cameras function in the real world.

  - However, it serves as a valuable mathematical tool for understanding imaging principles

2D projective geometry

- We talk out about 2D projective geometry, which is essentially the **mathematics of this projections**, i.e., the projections of **world scene** onto a **2D surface**.

- The **projective geometry** is beneficial for the **analysis of the images**, as we see in the following lectures.

Recall: Parametric vector from of a line [1]

We can write parameters a, b, and c of a **general equation** of a line in the **vector form** as

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

For example:

2x – 3y + 5 = 0

If we divide above equation by 2, then we still have the same line.

x – (3/2) y + 5/2 = 0

a = 1, b = -3/2, and c = 5/2

These (a, b, and c) are perfectly legitimate scalars.

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{-3/2} \\ \mathbf{5/2} \end{bmatrix}$$

Recall: Parametric vector from of a line [2]

y = -8x + 5

We can represent a **slope intercept** form in a parametric vector form

8x – y + 5 = 0

a = 8, b = -1, and c = 5

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 8 \\ -1 \\ 5 \end{bmatrix}$$

Homogeneous Representation of a Line in 2D

How do we derive the vector form from the coefficient form?

1. ax + by + c = 0

2. $\begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ = 0 or (x, y, 1) $(a, b, c)^{T}$ = 0

Recall: Dot product is the sum of the product of each element.

In projective geometry, we can express a line equation using the **dot product** of the **coefficient vector** $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ with this very **special vector**, $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ representing a **point in homogeneous coordinates**

Homogeneous Representation of a Line in 2D

How do we derive the vector form from the coefficient form?

Mathematically, this can be written as:

$(x, y, 1)\ (a, b, c)^T = 0$

or equivalently, using matrix notation:

$$\begin{bmatrix} a & b & c \end{bmatrix}_{1\times 3} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}_{3\times 1} = 0$$

$\Rightarrow ax + by + c = 0$

This equation represents a **homogeneous line equation**, a key concept in projective geometry and computer vision.

Note: These are two different ways of saying the same thing.

Homogeneous Objects in Mathematics

If we take the coefficient form of the equation of a line (ax + by + c = 0):

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

and multiply it by any scalar k, we still have the same line:

$$\begin{bmatrix} ka \\ kb \\ kc \end{bmatrix}$$

$\begin{bmatrix} ka \\ kb \\ kc \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$ **remains unchanged**, demonstrating that it is a **homogeneous object**.

Definition of Homogeneous Objects:

- Homogeneous objects are mathematical entities that are determined **only up to scale**. This means that scaling them by a **nonzero factor** does not change their fundamental nature.

- In this case, the vector $(\mathbf{a}, \mathbf{b}, \mathbf{c})^{T}$ represents the **same line** as $\mathbf{k}(\mathbf{a}, \mathbf{b}, \mathbf{c})^{T}$ for any **nonzero constant k**.

What are Homogeneous Coordinates?

- In **projective geometry**, we extend 2D coordinates to 3D for easier transformations.
- A **point (x, y) in Cartesian coordinates** is represented **as (x, y, 1) in homogeneous coordinates**.
- This allows for **uniform representation** of transformations like scaling and translation.