AI VIETNAM
All-in-One Course

How to do Research

(Research Plan: Study and Implement Existing Algorithms)

Vinh Dinh Nguyen
PhD in Computer Science

Year 2023

Outline

- How to do Research
- Background of the Research Topic
- Implement Existing Methods
- How to Write Research Plan
- Assignment

Outline

- How to do Research
- Background of the Research Topic
- Implement Existing Methods
- How to Write Research Plan
- Assignment

How to do Research

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title:
   Provide a clear and concise title that reflects the main focus of your research
2. Introduction
   Briefly introduce your research topic, explaining its significance and relevance.
3. Background and Literature Review
   Summarize the existing knowledge and research related to your topic. Highlight key studies, theories, or gaps in the literature that your research aims to address.
4. Research Objectives
   What do you aim to achieve or discover through your study?
5. Research Questions
   Formulate the research questions or hypotheses that will guide your investigation
6. Methodology
   Detail the research design, methods, and procedures you plan to employ
7. Timeline
   Provide a timeline for your research activities. Include milestones, deadlines, and key events.
8. Resources
   Outline the resources you'll need for your research, such as access to specific databases, equipment, or funding.
9. Expected Result
   Discuss the anticipated outcomes of your research.
10. Significant and Contribution
    Explain the potential significance of your research.

Research Topic: Stereo Matching

Stereo Matching – Depth Estimation

Define Your Research Question or Topic

⇩

Background Research

⇩

Study and Implement Existing Methods

⇩

Develop a Research Plan

⇩

Implement Your Proposed Methods

⇩

Evaluate Performance of Compared Methods

RQ1: How to improve the performance of existing stereo matching methods under various lighting conditions and complex material properties?

RQ2: How to obtain a real-time processing for current state-of-the-art stereo matching-based deep learning methods?

RQ3: How to design an automatically learning the network architecture, its activation functions, and its parameters from data ?

Last week...

Outline

- How to do Research
- Background of the Research Topic
- Implement Existing Methods
- How to Write Research Plan
- Assignment

Stereo Matching: Background

Define Your Research Question or Topic

⇩

Background Research

⇩

Study and Implement Existing Methods

⇩

Develop a Research Plan

⇩

Implement Your Proposed Methods

⇩

Evaluate Performance of Compared Methods

1. Identify Keywords and Concepts
2. Conduct a Literature Review
3. Review Relevant Journals and Conferences
4. Identify Key Theories and Frameworks
5. Explore Citations
6. Check for Recent Developments

Some popular databases include PubMed, Google Scholar, IEEE Xplore, Springer, Elsiver, and JSTOR, ….

Stereo Matching: Background

Define Your Research Question or Topic

⇩

Background Research

⇩

Study and Implement Existing Methods

⇩

Develop a Research Plan

⇩

Implement Your Proposed Methods

⇩

Evaluate Performance of Compared Methods

A Survey on Deep Learning Techniques for Stereo-based Depth Estimation

Hamid Laga, Laurent Valentin Jospin, Farid Boussaid, Mohammed Bennamoun *Senior Member, IEEE*

Abstract—Estimating depth from RGB images is a long-standing ill-posed problem, which has been explored for decades by the computer vision, graphics, and machine learning communities. Among the existing techniques, stereo matching remains one of the most widely used in the literature due to its strong connection to the human binocular system. Traditionally, stereo-based depth estimation has been addressed through matching hand-crafted features across multiple images. Despite the extensive amount of research, these traditional techniques still suffer in the presence of highly textured areas, large uniform regions, and occlusions. Motivated by their growing success in solving various 2D and 3D vision problems, deep learning for stereo-based depth estimation has attracted a growing interest from the community, with more than 150 papers published in this area between 2014 and 2019. This new generation of methods has demonstrated a significant leap in performance, enabling applications such as autonomous driving and augmented reality. In this article, we provide a comprehensive survey of this new and continuously growing field of research, summarize the most commonly used pipelines, and discuss their benefits and limitations. In retrospect of what has been achieved so far, we also conjecture what the future may hold for deep learning-based stereo for depth estimation research.

Index Terms—CNN, Deep Learning, 3D Reconstruction, Stereo Matching, Multi-view Stereo, Disparity Estimation, Feature Leaning, Feature Matching.

1 INTRODUCTION

DEPTH estimation from one or multiple RGB images is a long standing ill-posed problem, with applications in various domains such as robotics, autonomous driving, object recognition and scene understanding, 3D modeling and animation, augmented reality, industrial control, and medical diagnosis. This problem has been extensively investigated for many decades. Among all the techniques that have been proposed in the literature, stereo matching is traditionally the most explored one due to its strong connection to the human binocular system.

previously seen objects and scenes have enabled us to build prior knowledge and develop mental models of how the 3D world looks like. The **second** generation of methods tries to leverage this prior knowledge by formulating the problem as a learning task. The advent of deep learning techniques in computer vision [1] coupled with the increasing availability of large training datasets, have led to a **third** generation of methods that are able to recover the lost dimension. Despite being recent, these methods have demonstrated exciting and promising results on various tasks related to computer vision and graphics.

Stereo Matching: Background

Define Your Research Question or Topic

⇩

Background Research

⇩

Study and Implement Existing Methods

⇩

Develop a Research Plan

⇩

Implement Your Proposed Methods

⇩

Evaluate Performance of Compared Methods

Survey on Semantic Stereo Matching / Semantic Depth Estimation

Viny Saajan Victor[1] and Peter Neigel[2]

[1] victor@rhrk.uni-kl.de
[2] peter.neigel@dfki.de

Abstract. Stereo matching is one of the widely used techniques for inferring depth from stereo images owing to its robustness and speed. It has become one of the major topics of research since it finds its applications in autonomous driving, robotic navigation, 3D reconstruction, and many other fields. Finding pixel correspondences in non-textured, occluded and reflective areas is the major challenge in stereo matching. Recent developments have shown that semantic cues from image segmentation can be used to improve the results of stereo matching. Many deep neural network architectures have been proposed to leverage the advantages of semantic segmentation in stereo matching. This paper aims to give a comparison among the state of art networks both in terms of accuracy and in terms of speed which are of higher importance in real-time applications.

Keywords: Stereo Matching, Semantic Segmentation, Depth Estimation, Conditional Random Fields, Warping Error, Multiscale Context Intertwining

1 Introduction

1.1 Stereo Matching: Stereo Matching is a technique used to find point correspondences between two images of a scene acquired by known cameras. The traditional approaches such as Semi-Global Matching [5] includes well-defined steps such as cost matching (calculating matching cost for each pixel in one of the stereo images depending on the corresponding pixel in the other image), cost aggregation (calculating matching cost for range of disparity values to form cost volumes), disparity estimation (calculating the disparity from cost volume by minimizing the cost function) and disparity post-processing (improving the disparity). These approaches make assumptions that the scene is uniformly illuminated and textured.

Stereo Matching: Background

Define Your Research Question or Topic

⇓

Background Research

⇓

Study and Implement Existing Methods

⇓

Develop a Research Plan

⇓

Implement Your Proposed Methods

⇓

Evaluate Performance of Compared Methods

1582 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 31, NO. 9, SEPTEMBER 2009

Evaluation of Stereo Matching Costs on Images with Radiometric Differences

Heiko Hirschmüller and Daniel Scharstein, *Member*, *IEEE*

Abstract—Stereo correspondence methods rely on matching costs for computing the similarity of image locations. We evaluate the insensitivity of different costs for passive binocular stereo methods with respect to radiometric variations of the input images. We consider both pixel-based and window-based variants like the absolute difference, the sampling-insensitive absolute difference, and normalized cross correlation, as well as their zero-mean versions. We also consider filters like LoG, mean, and bilateral background subtraction (BilSub) and nonparametric measures like Rank, SoftRank, Census, and Ordinal. Finally, hierarchical mutual information (HMI) is considered as pixelwise cost. Using stereo data sets with ground-truth disparities taken under controlled changes of exposure and lighting, we evaluate the costs with a local, a semiglobal, and a global stereo method. We measure the performance of all costs in the presence of simulated and real radiometric differences, including exposure differences, vignetting, varying lighting, and noise. Overall, the ranking of methods across all data sets and experiments appears to be consistent. Among the best costs are BilSub, which performs consistently very well for low radiometric differences; HMI, which is slightly better as pixelwise matching cost in some cases and for strong image noise; and Census, which showed the best and most robust overall performance.

Index Terms—Stereo, matching cost, performance evaluation, radiometric differences.

✦

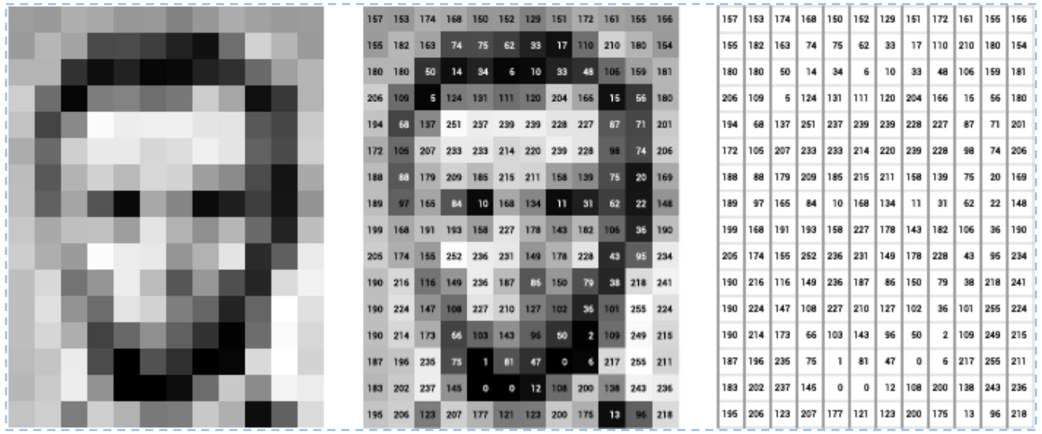
1 Introduction

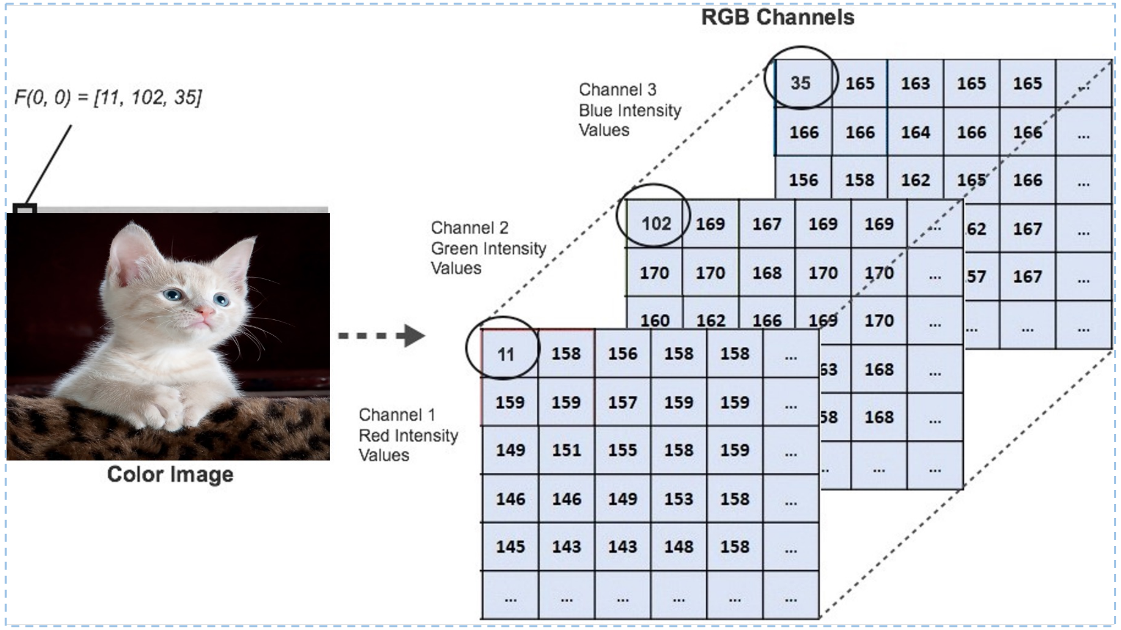
All passive stereo correspondence algorithms have a way of measuring the similarity of image locations. Typically, a *matching cost* is computed at each pixel for all disparities under consideration. The simplest matching costs assume constant intensities at matching image outdoors). Similar situations arise when matching aerial or satellite images.

Due to all of the above reasons, it is safe to say that any real-world stereo application requires radiometric robustness. This includes existing commercial systems, which

Stereo Matching: Background

What is a digital image?



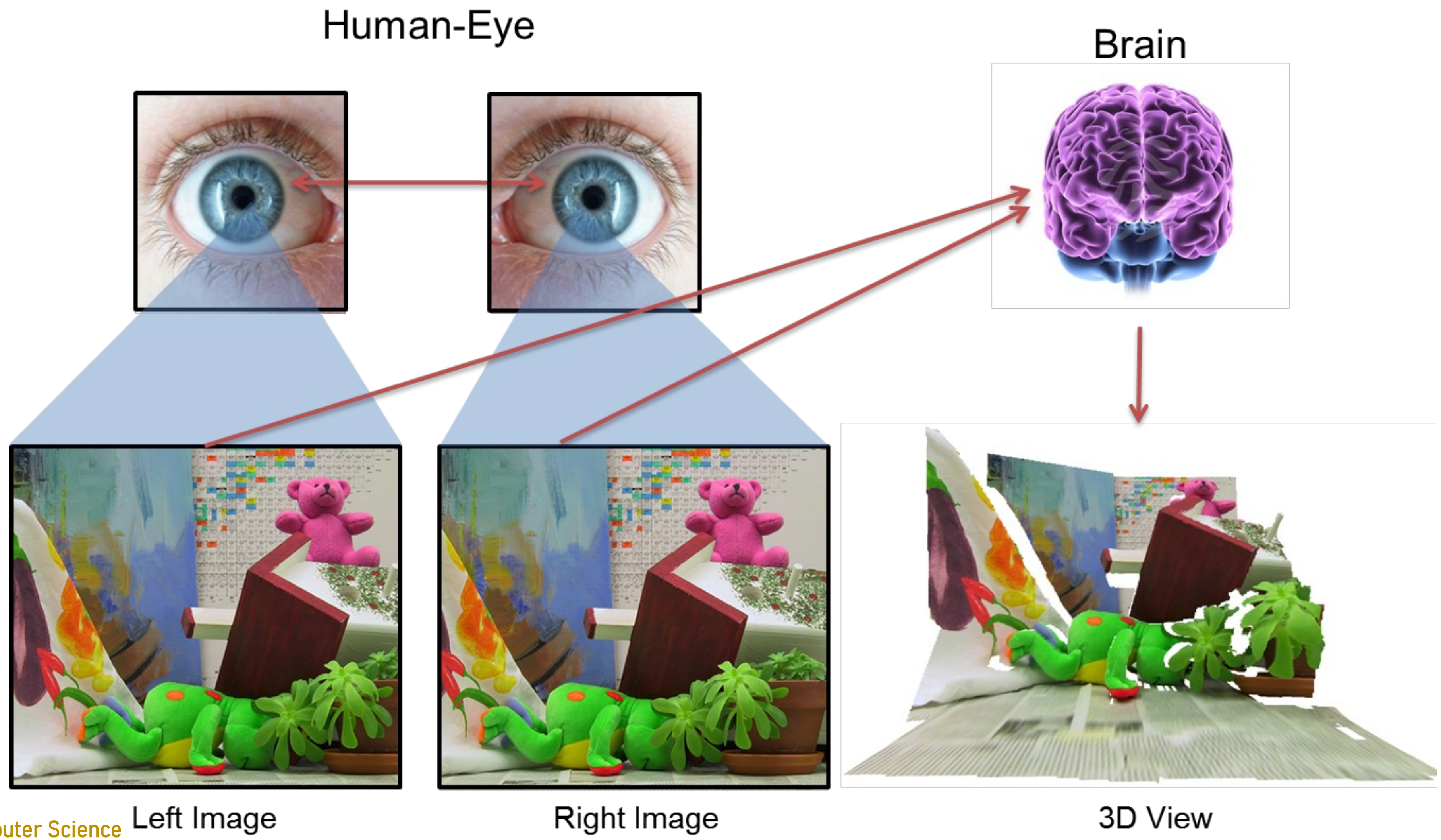


Color Image

Graycalse Image

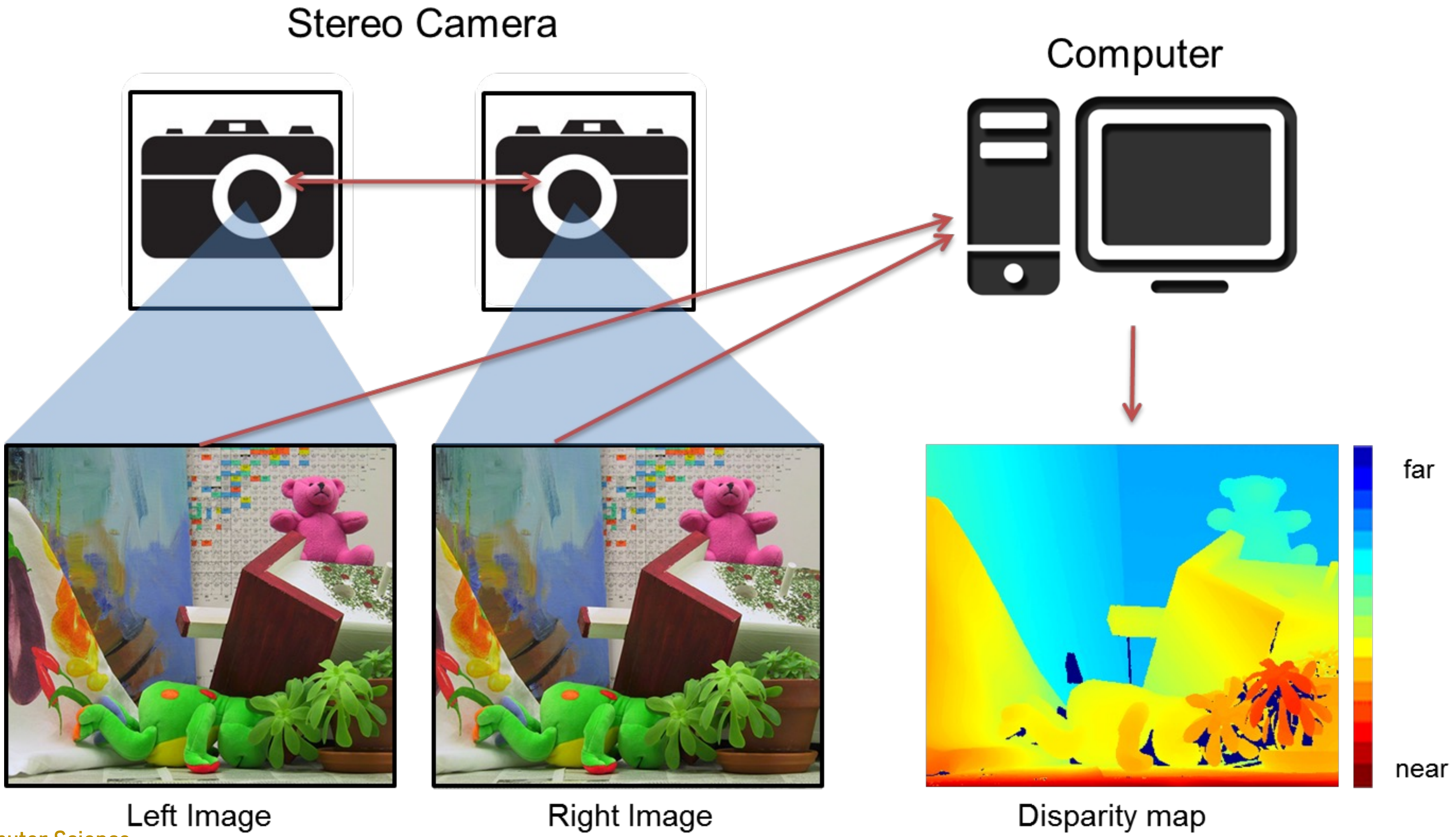
Stereo Matching: Background

Human perception for 3D information



Stereo Matching: Background

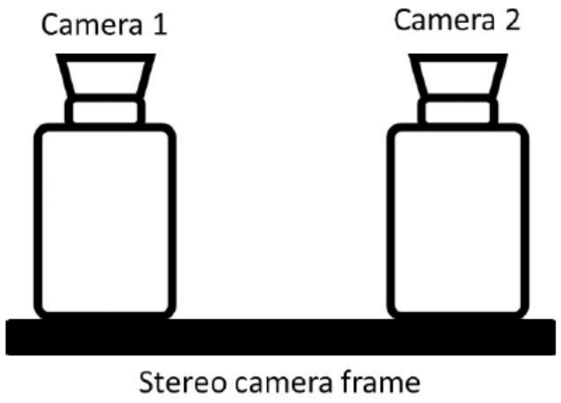
We generate 3D information using disparity maps obtained from stereo matching algorithm

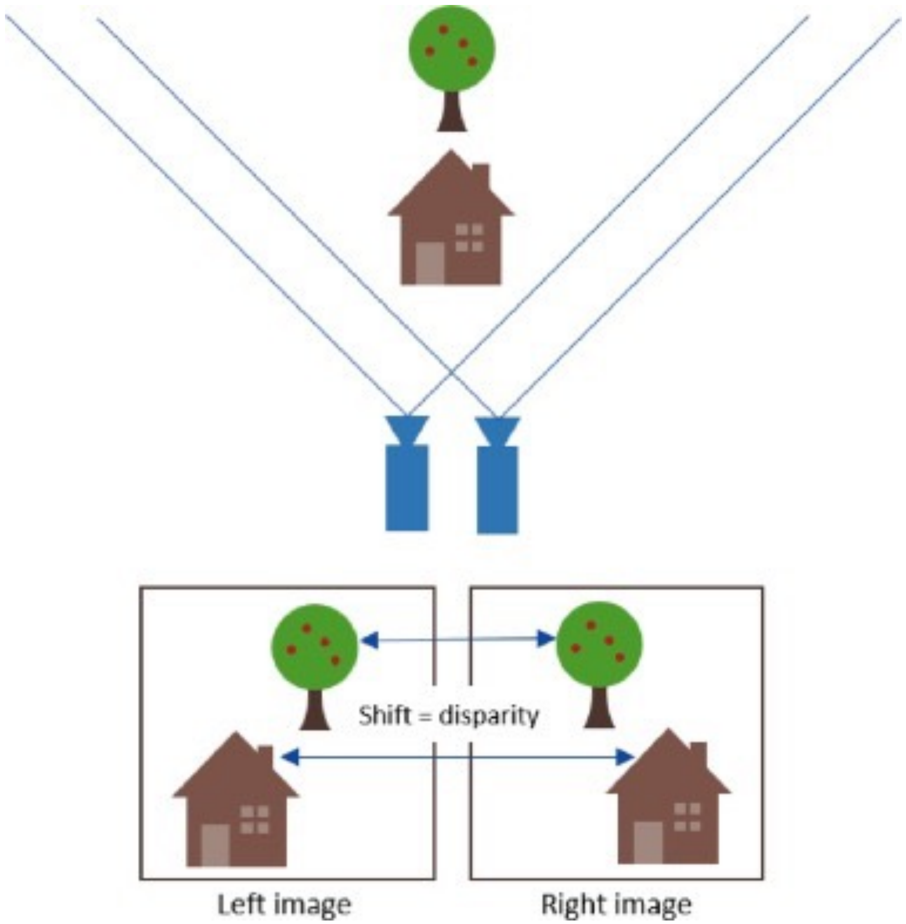


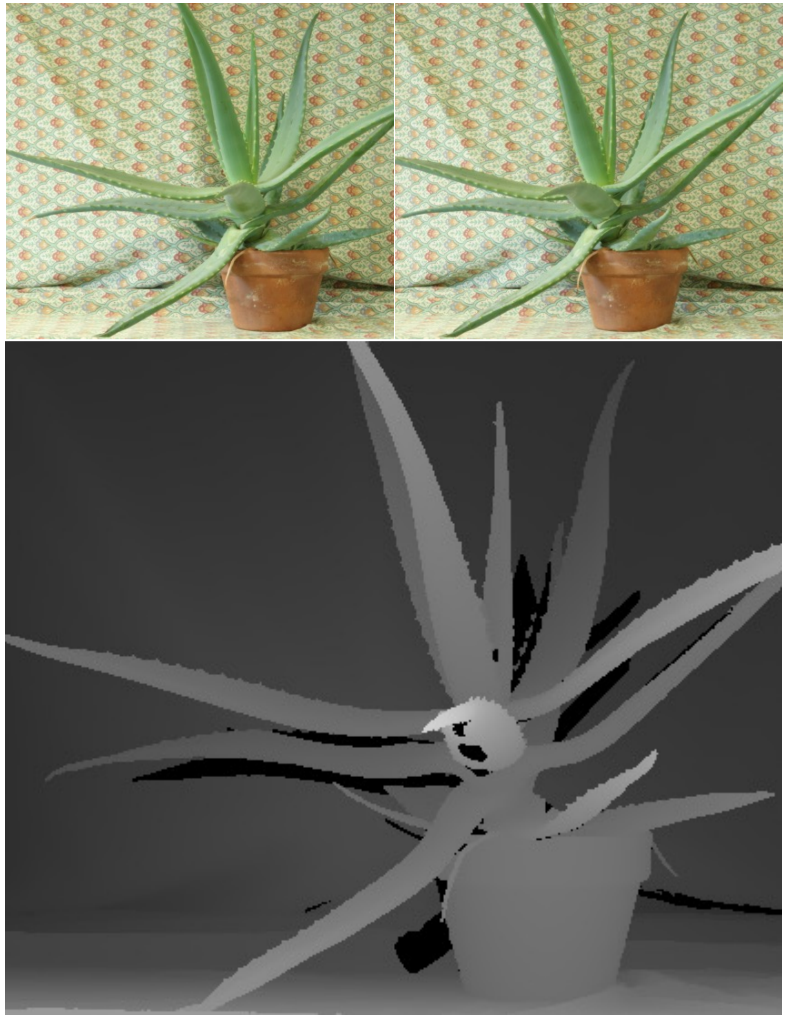
Stereo Matching: Background

❖ Computer stereo vision is the extraction of 3D information from digital images

❖ Stereo cameras

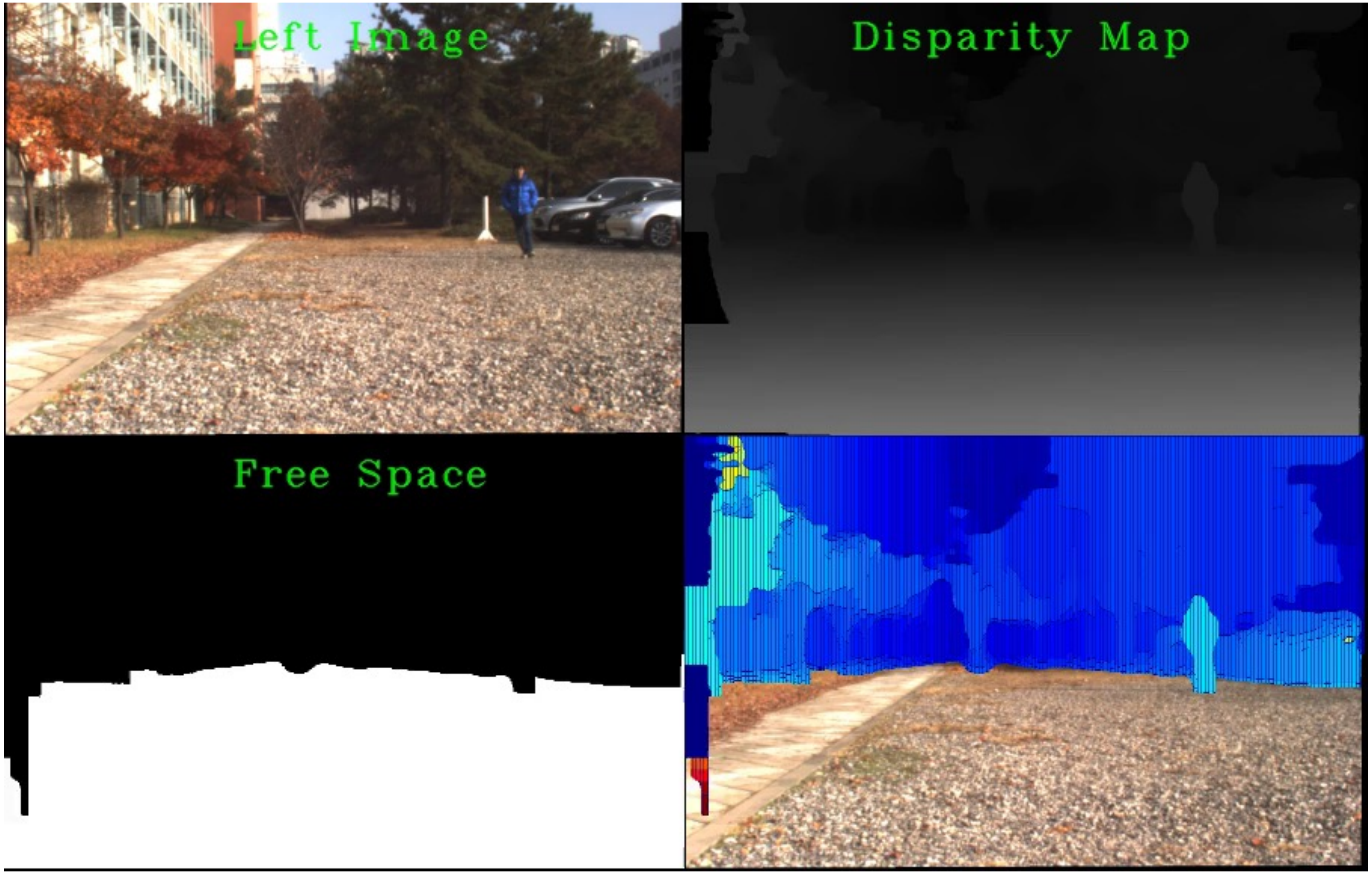






Stereo Matching Algorithm

Applications



Stereo Matching: Background

Challenges in stereo matching

Normal stereo pair ADCensus SAD

Stereo pair with different illumination ADCensus SAD

Stereo Matching: Background

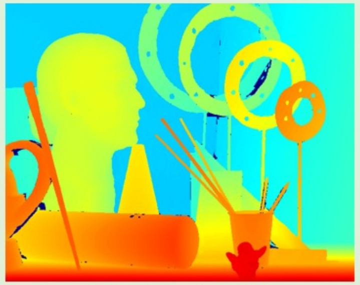
Stereo Matching / Binocular Depth Estimation

Hình từ camera trái

Hình từ camera phải

Thuật toán stereo matching

Hình chiều sâu
cho hình trái



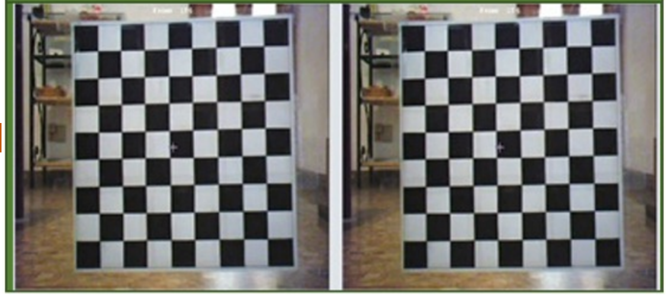
Hình chiều sâu
cho hình phải

Stereo camera

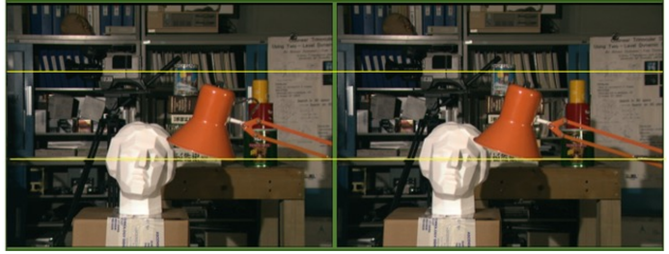
<u>Ứng dụng</u>

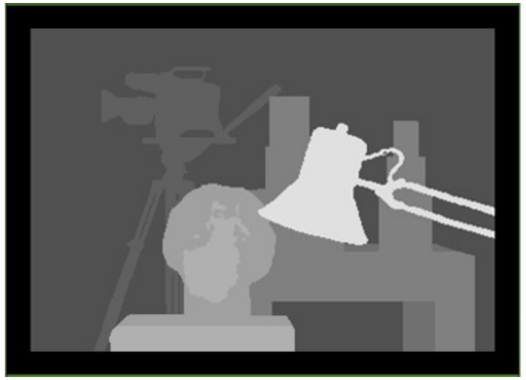
- Đo khoảng các tới các object trong hình
- Xây dựng mô hình 3D
- Thông tin cho các ứng dụng khác: detection, tracking, …

Calibration



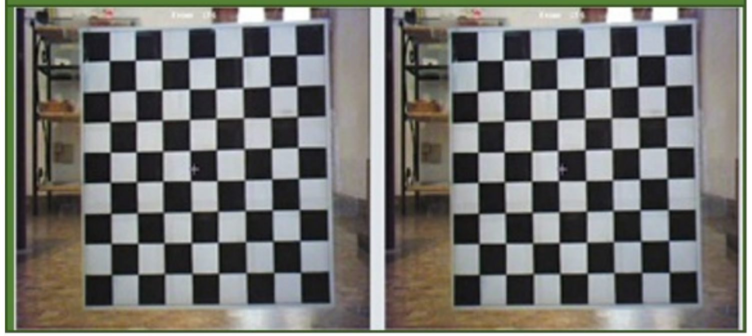
Rectification



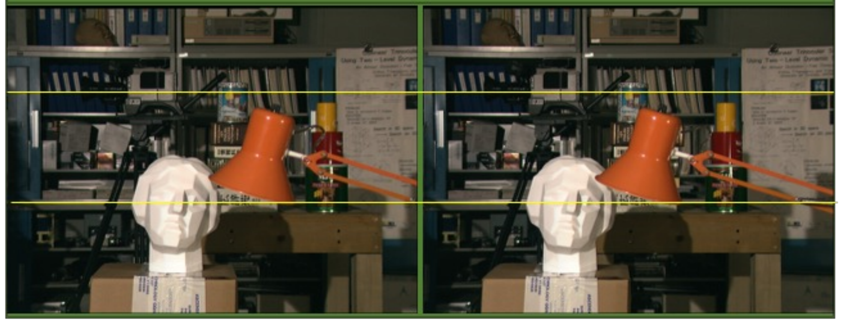


Stereo Matching: Background

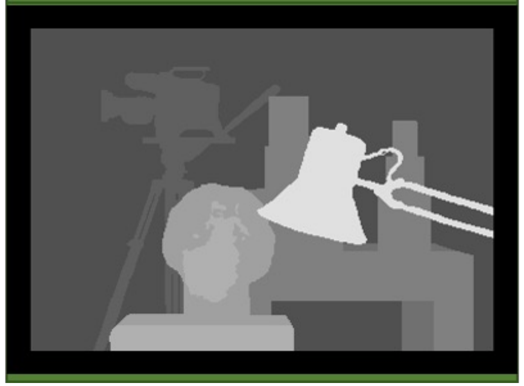
Calibration

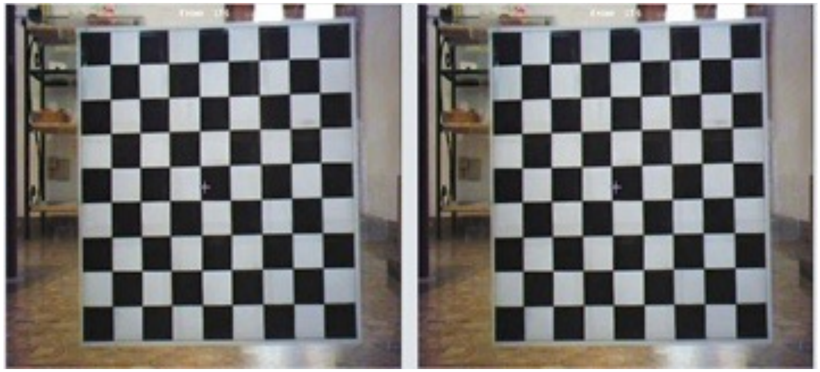


Rectification



Stereo Matching

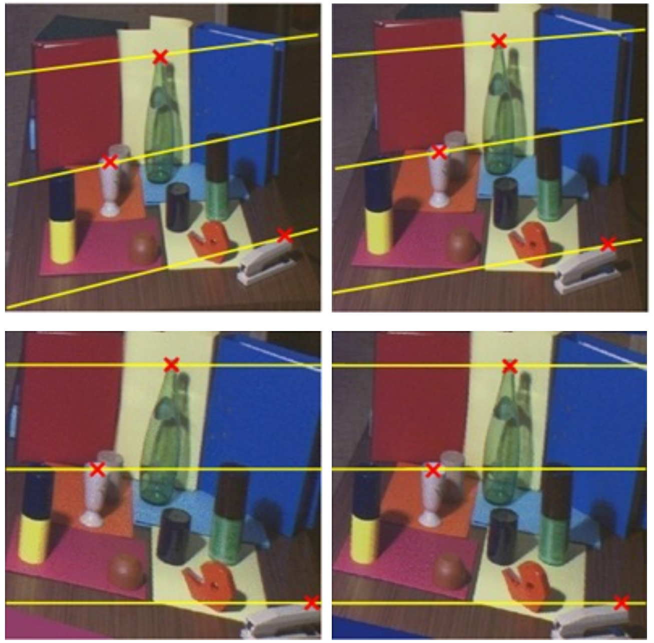


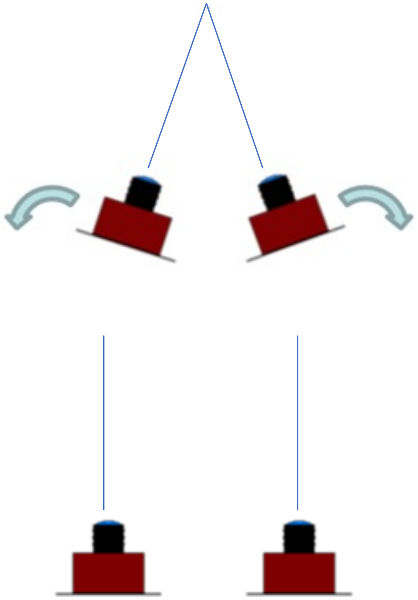


$$K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}$$

Essential matrix $P = K[R \mid t]$

- intrinsic camera parameters
- extrinsic camera parameters





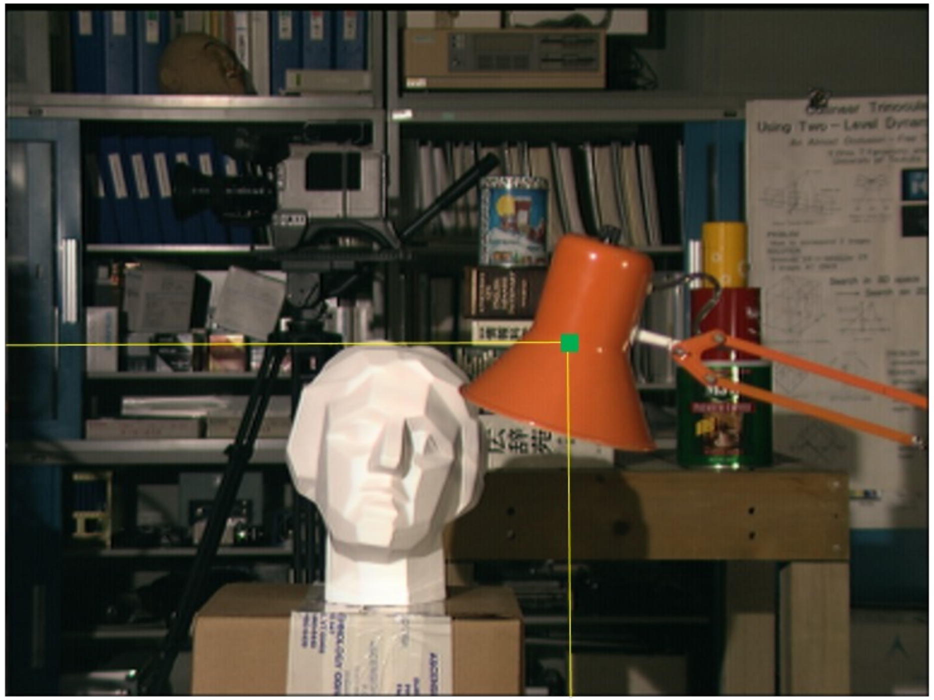
Stereo camera in standard form

Stereo Matching: Background

Another Approach

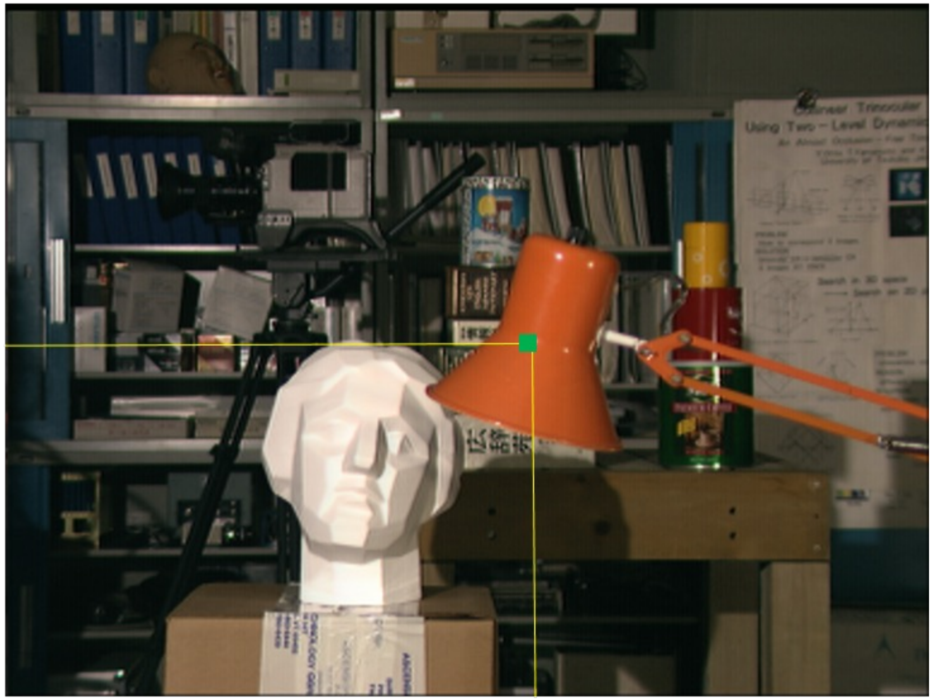
p = (234, 140) **q** = (220, 140)

y_p=140



x_p = 234

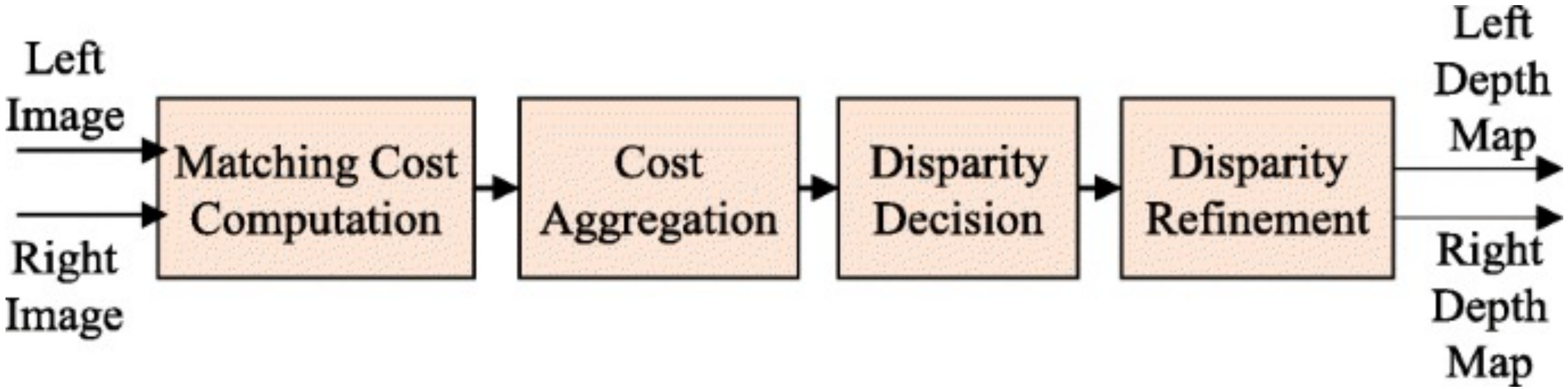
y_q=140



x_q = 220

disparity_q = x_p – x_q = 14

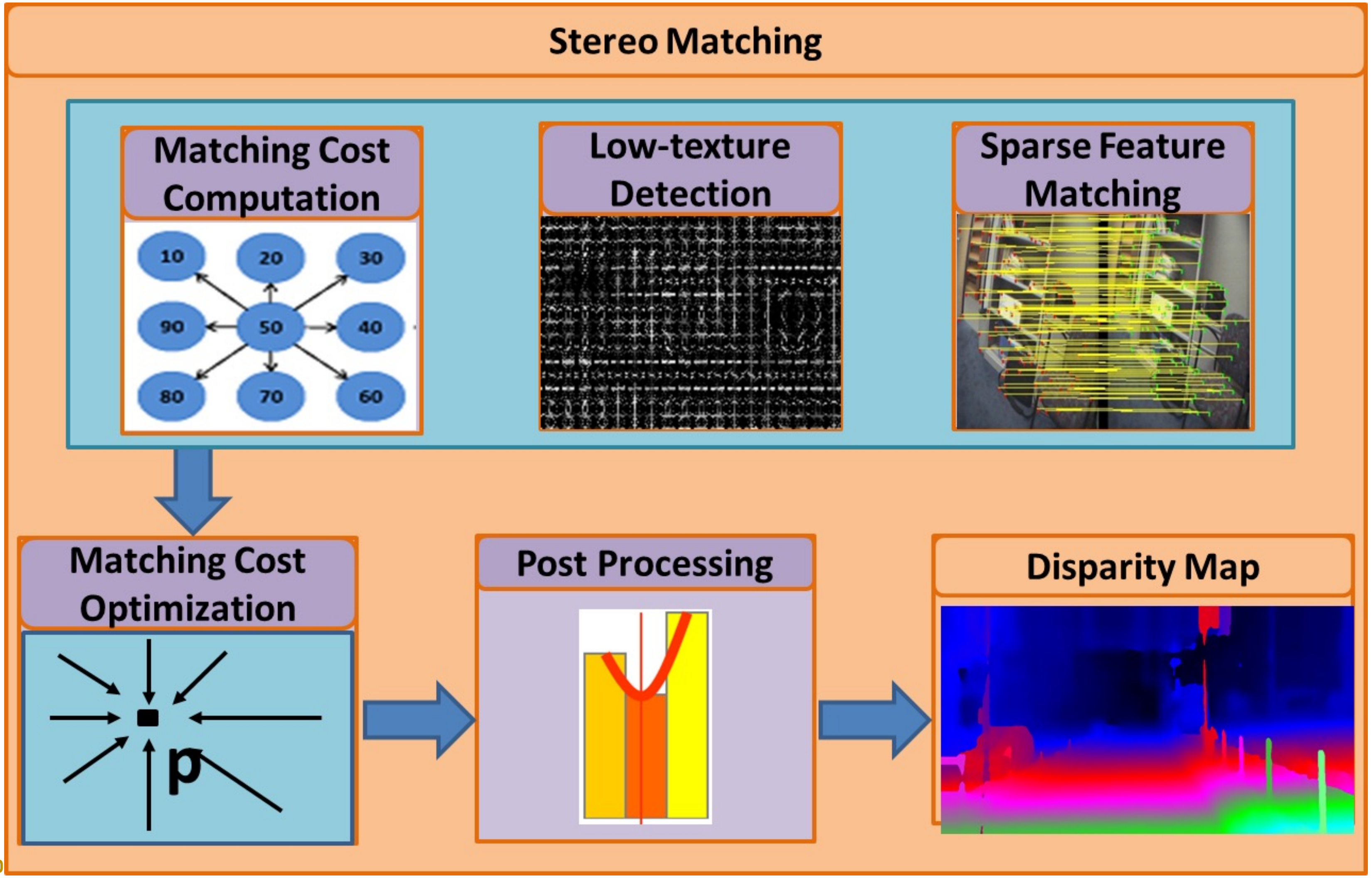
Stereo Matching: Background



AI VIETNAM
All-in-One Course

Stereo Matching: Background

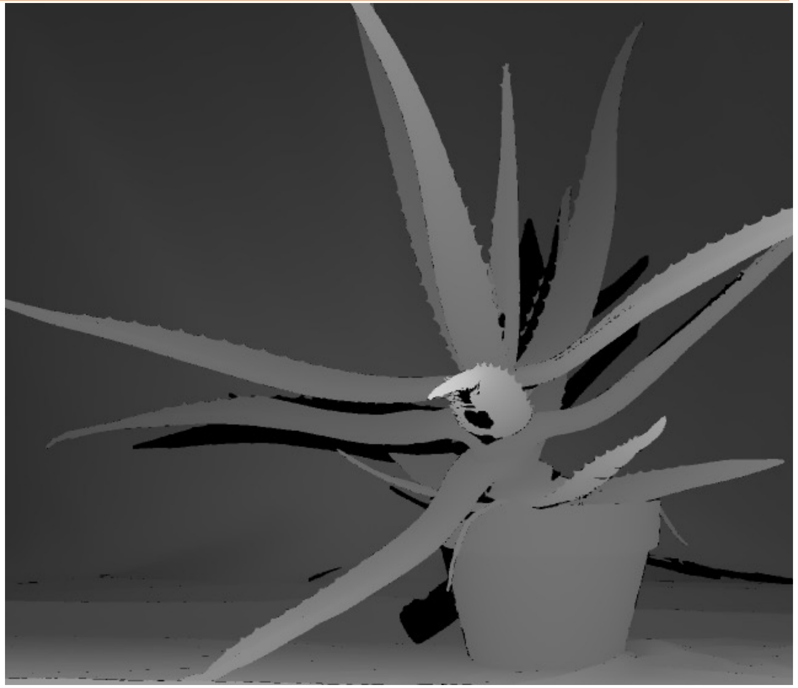
Traditional methods



Stereo Matching: Background

Left Image

Right Image



Disparity Map
(for the left image)

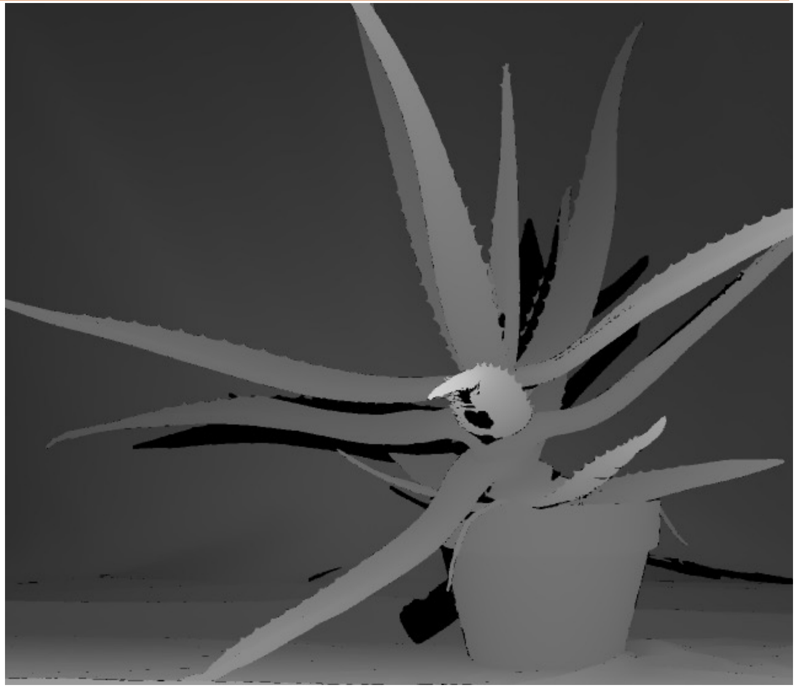
From the left and right images, a disparity map is computed.

Each pixel in the map is in an interval [0, D], where D is disparity range.

Stereo Matching: Background

Left Image

Right Image



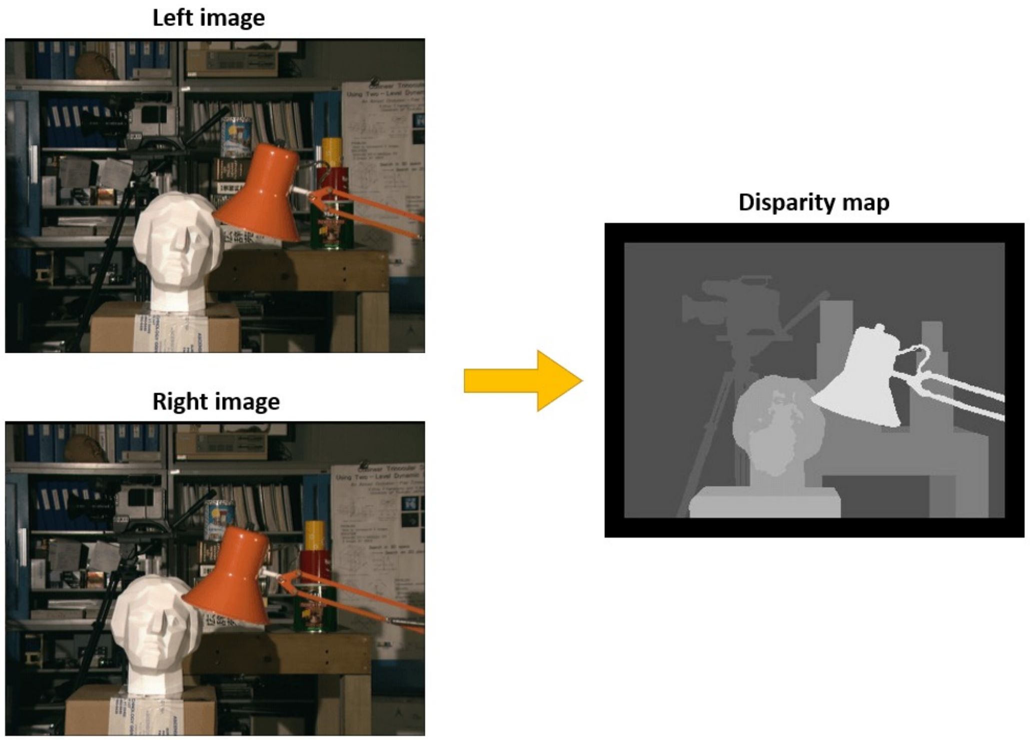
Disparity Map
(for the left image)

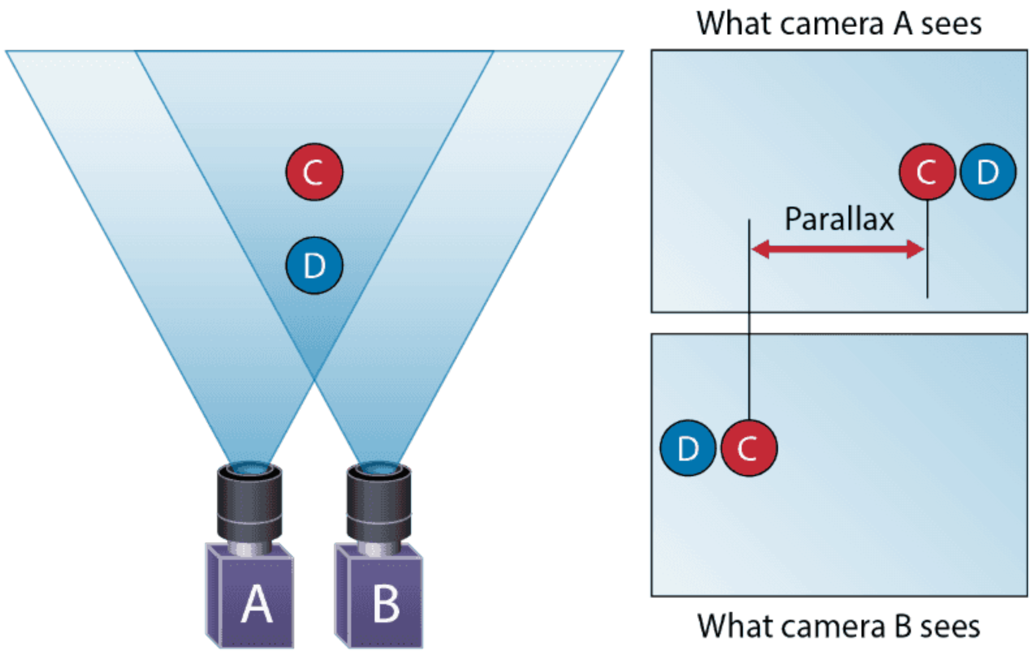
Constraints

1) A pixel p in the left image and its corresponding pixel q in the right image are in the same line ($y_p = y_q$)

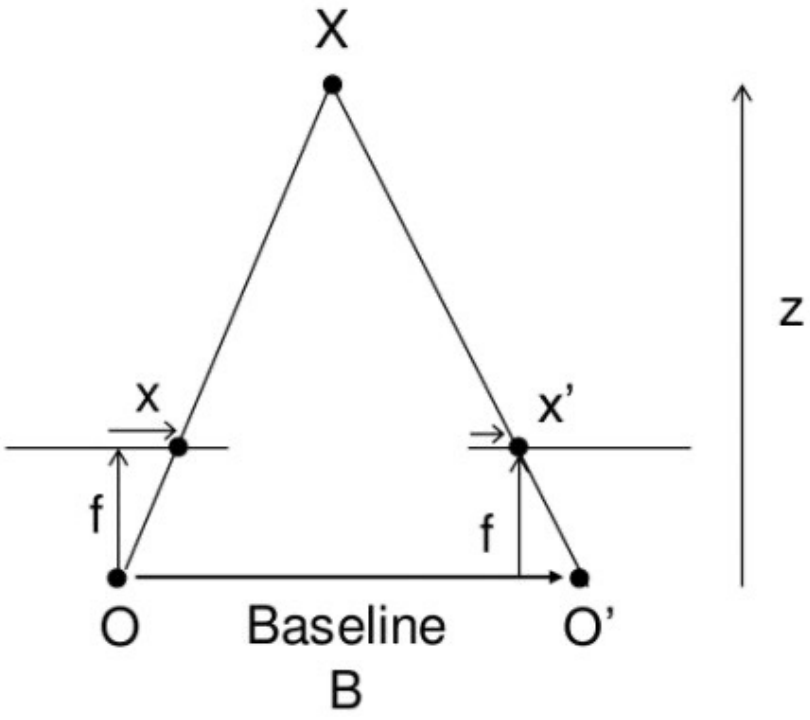
2) Typically, $x_p - D \leq x_q \leq x_p$

Stereo Matching: Background

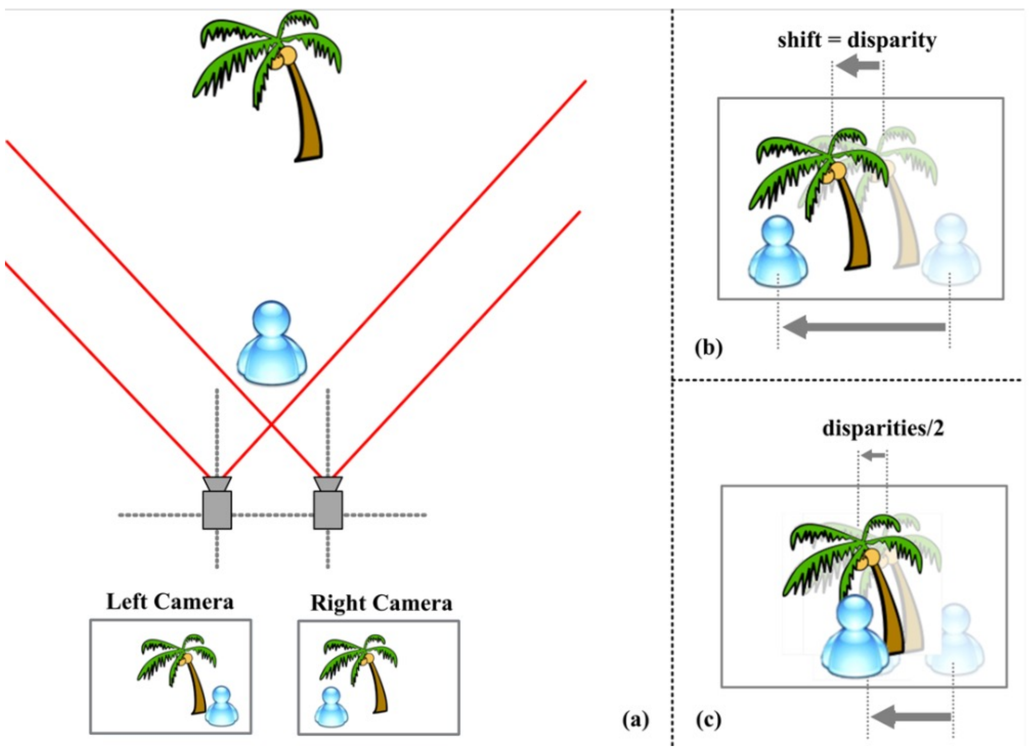




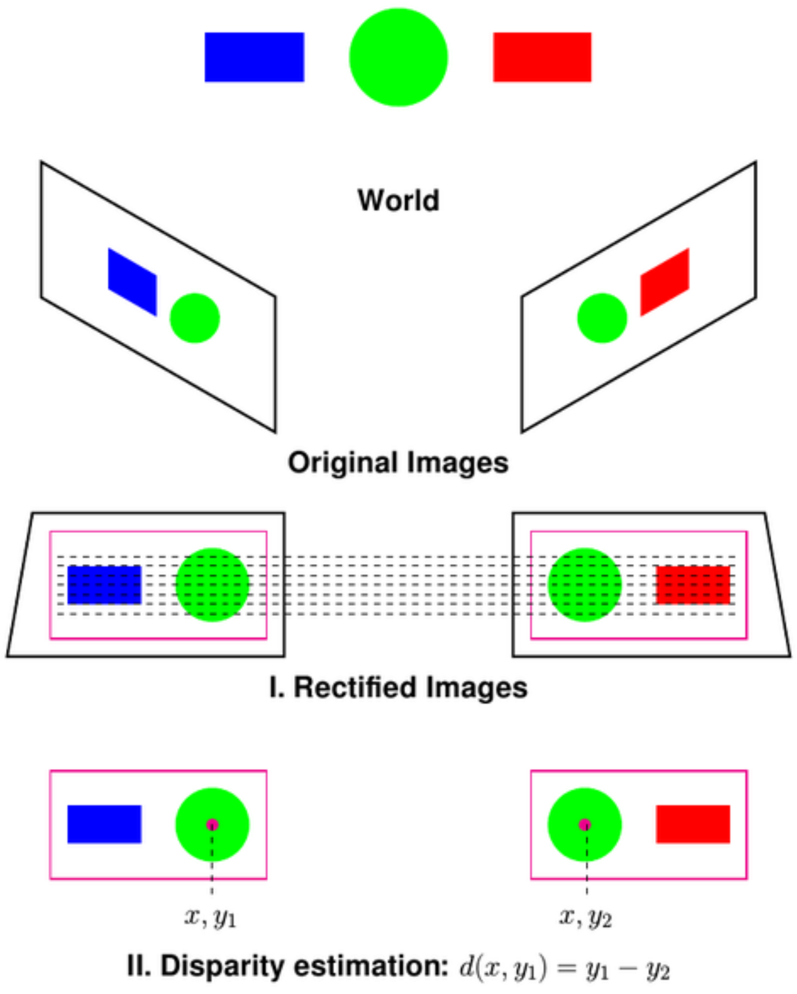
Stereo Matching: Background

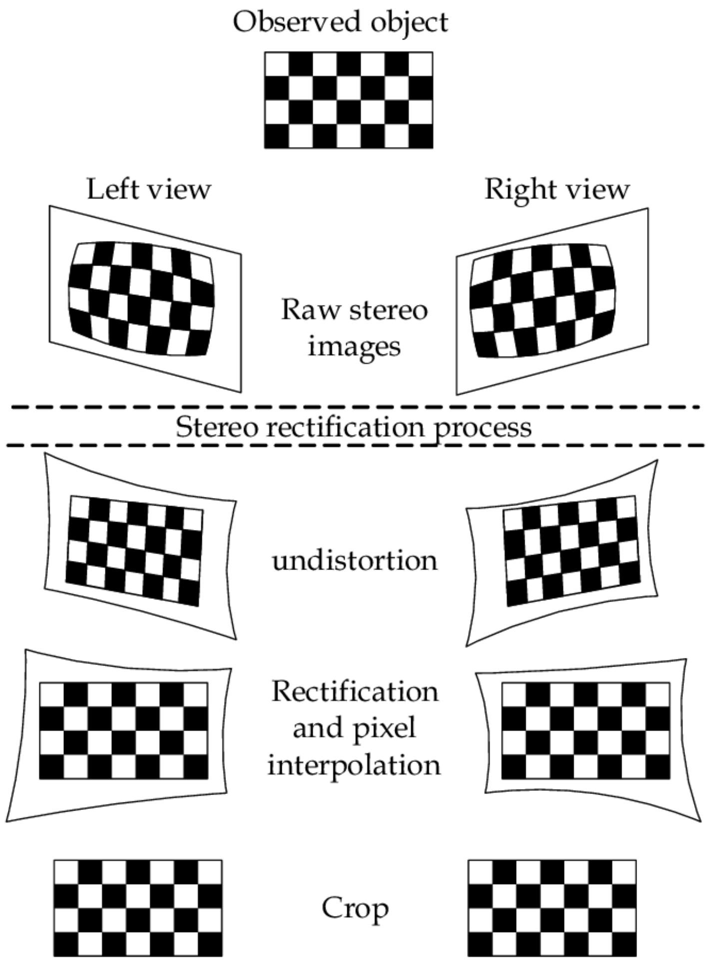


$$disparity = x - x' = \frac{Bf}{Z}$$

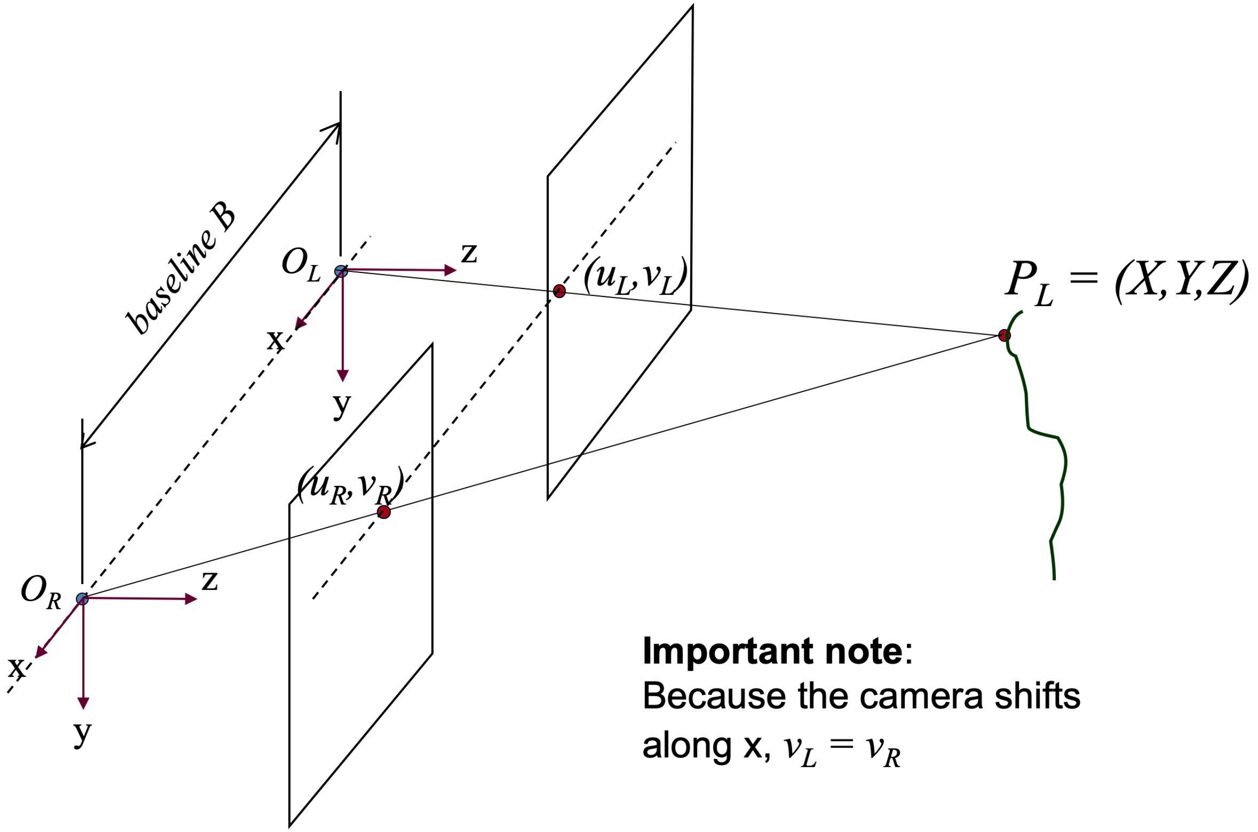


Stereo Matching: Background

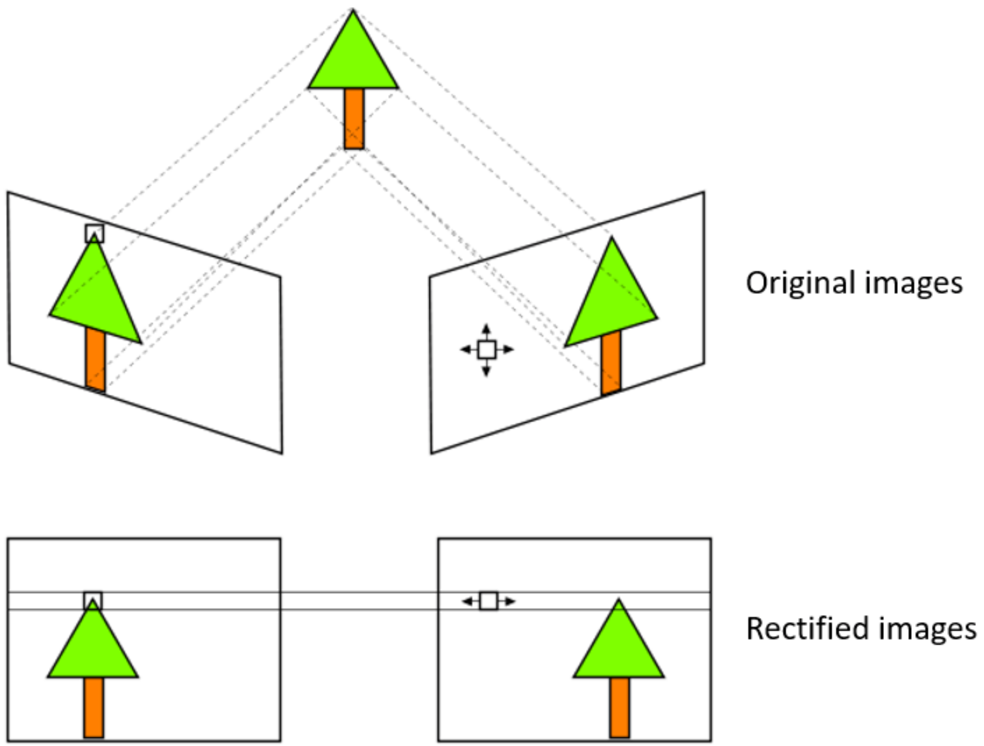




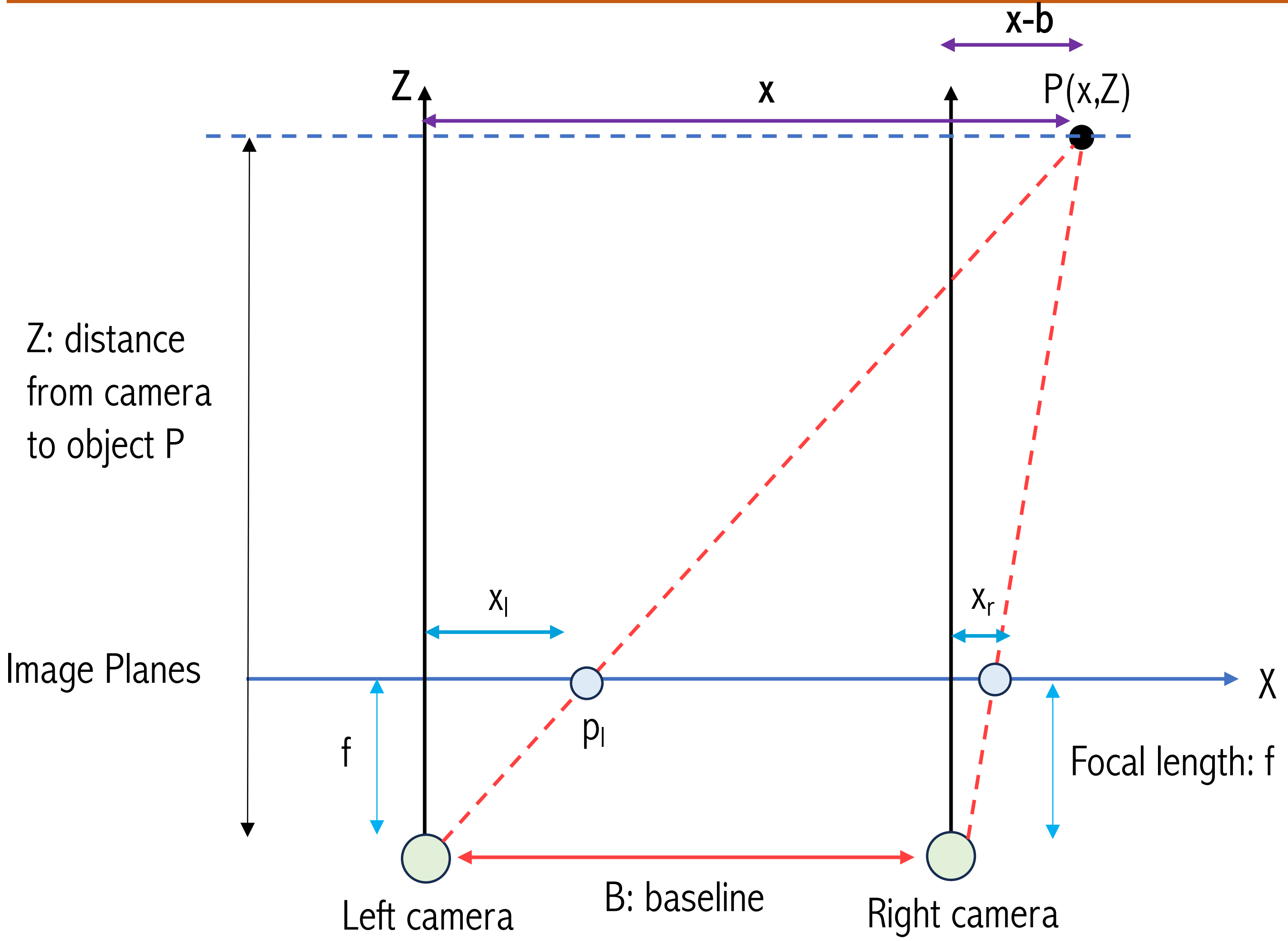
Stereo Matching: Background



Important note:
Because the camera shifts along x, $v_L = v_R$



Stereo Matching: Background



$$\frac{z}{f} = \frac{x}{x_l}$$

$$\frac{z}{f} = \frac{x-b}{x_r}$$

Vinh Dinh Nguyen- PhD in Computer Science

AI VIETNAM
All-in-One Course

AI VIETNAM
All-in-One Course

Stereo Matching: Background

$$\frac{x}{x_l} = \frac{x-b}{x_r}$$

$$x x_r = x_l(x-b)$$

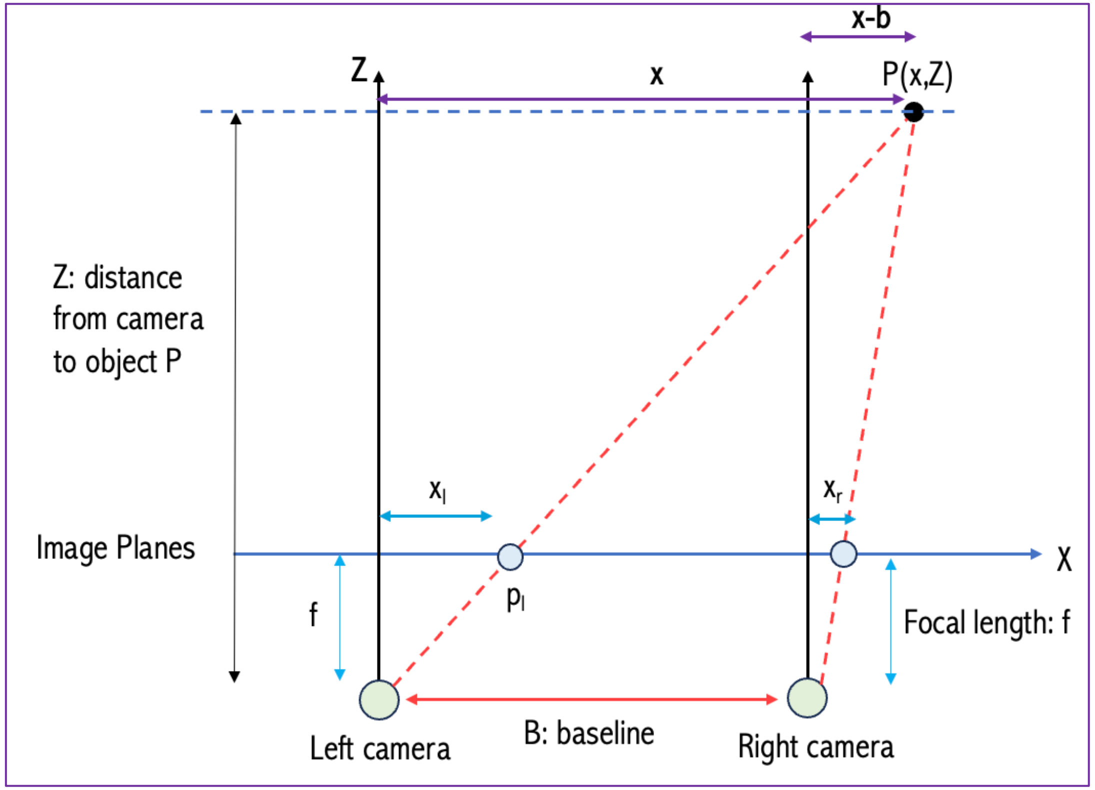
$$x x_r = x_l x - x_l b$$

$$x_r = \frac{x_l x}{x} - b\left(\frac{x_l}{x}\right)$$

$$b\frac{x_l}{x} = x_l - x_r$$

$$b\left(\frac{x_l}{x}\right)\left(\frac{x}{x_l}\right) = (x_l - x_r)\left(\frac{x}{x_l}\right)$$

$$\frac{b}{x_l - x_r} = \frac{x}{x_l}$$



$$\frac{z}{f} = \frac{x}{x_l}$$

$$z = f\left(\frac{x}{x_l}\right)$$

$$\boxed{z = f\left(\frac{x}{x_l}\right) = f\left(\frac{b}{x_l - x_r}\right)}$$

Object

Left image

255	254	253	252	251
15	16	17	255	255
18	18	17	255	255
15	17	18	255	255
255	255	255	255	255

Right image

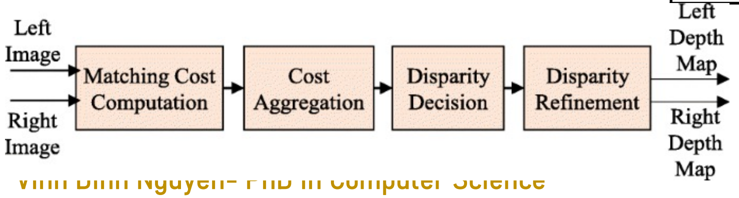
255	254	253	252	251
255	15	18	18	255
255	16	17	20	255
255	14	14	19	255
255	255	255	255	255

1	1	1		
1	1	1		
1	1	1		

Left Disparity

	1	1	1	
	1	1	1	
	1	1	1	

Right Disparity



CASE STUDY

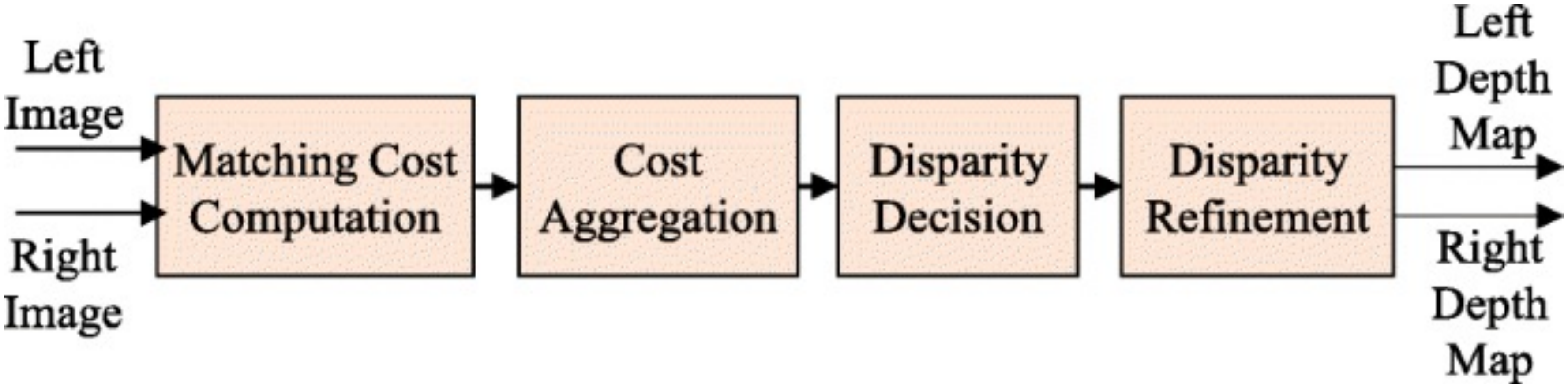
Object

Left image

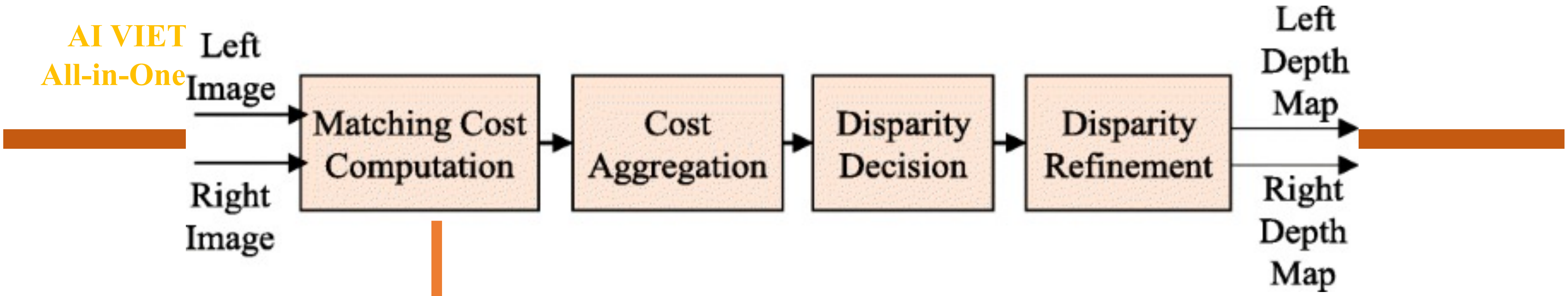
255	254	253	252	251
15	16	17	255	255
18	18	17	255	255
15	17	18	255	255
255	255	255	255	255

Right image

255	254	253	252	251
255	15	18	18	255
255	16	17	20	255
255	14	14	19	255
255	255	255	255	255



AI VIET
All-in-One



Left image

255	254	253	252	251
15	16	17	255	255
18	18	17	255	255
15	17	18	255	255
255	255	255	255	255

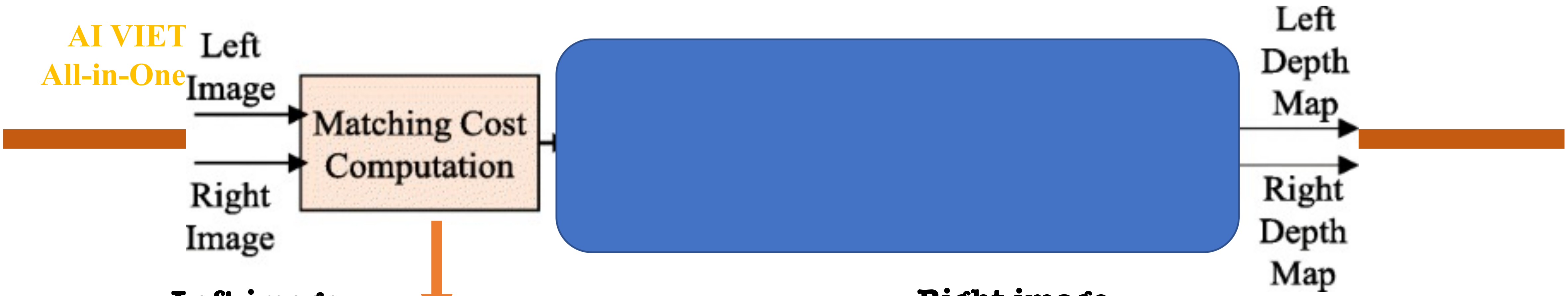
Right image

255	254	253	252	251
255	15	18	18	255
255	16	17	20	255
255	14	14	19	255
255	255	255	255	255

Maximum Disparity/ Maximum Search Region: 3

Disparity	Cost
1	|15-255|
2	|15-15|
3	|15-8|

Cost
|15-255|
|15-15|
|15-8|



Left image

255	254	253	252	251
15	16	17	255	255
18	18	17	255	255
15	17	18	255	255
255	255	255	255	255

Right image

255	254	253	252	251
255	15	18	18	255
255	16	17	20	255
255	14	14	19	255
255	255	255	255	255

Cost matrix size: 5x5x3

0	0	0	0	0
2	1	1	1	0
0	1	1	1	0
0	1	1	1	0
0	0	0	0	0

Maximum Disparity/
Maximum Search Region: 3

Stereo Matching

❖ Method 1

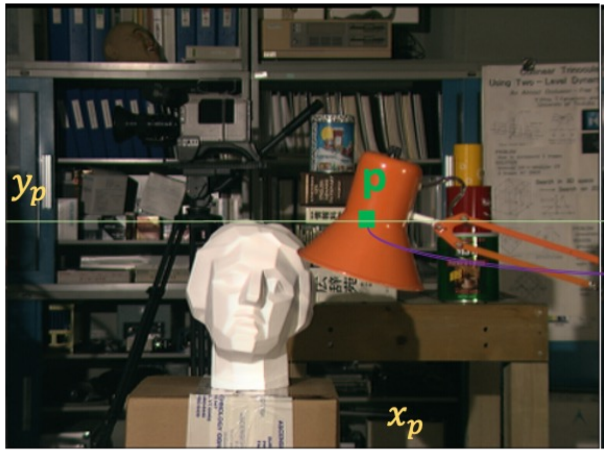
L is the left image

R is the right image

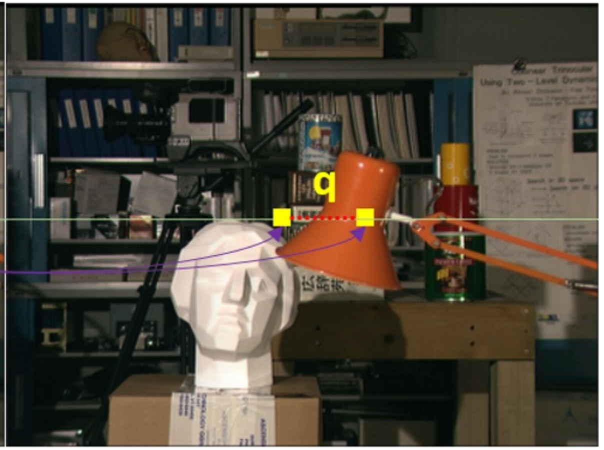
$L(\boldsymbol{p})$ is the (vector) value of $\mathbf{p}$

$$\boldsymbol{p} = \begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} 234 \\ 140 \end{bmatrix}$$

$$D = 16$$



Left Image



Right Image

$\boldsymbol{q}$

$$\begin{bmatrix} x_p - 0 \\ y_p \end{bmatrix} = \begin{bmatrix} 234 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 8 \\ y_p \end{bmatrix} = \begin{bmatrix} 226 \\ 140 \end{bmatrix}$$

$$\begin{bmatrix} x_p - 1 \\ y_p \end{bmatrix} = \begin{bmatrix} 233 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 9 \\ y_p \end{bmatrix} = \begin{bmatrix} 225 \\ 140 \end{bmatrix}$$

$$\begin{bmatrix} x_p - 2 \\ y_p \end{bmatrix} = \begin{bmatrix} 232 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 10 \\ y_p \end{bmatrix} = \begin{bmatrix} 224 \\ 140 \end{bmatrix}$$

$$\begin{bmatrix} x_p - 3 \\ y_p \end{bmatrix} = \begin{bmatrix} 231 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 11 \\ y_p \end{bmatrix} = \begin{bmatrix} 223 \\ 140 \end{bmatrix}$$

$$\begin{bmatrix} x_p - 4 \\ y_p \end{bmatrix} = \begin{bmatrix} 230 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 12 \\ y_p \end{bmatrix} = \begin{bmatrix} 222 \\ 140 \end{bmatrix}$$

$$\begin{bmatrix} x_p - 5 \\ y_p \end{bmatrix} = \begin{bmatrix} 229 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 13 \\ y_p \end{bmatrix} = \begin{bmatrix} 221 \\ 140 \end{bmatrix}$$

$$\begin{bmatrix} x_p - 6 \\ y_p \end{bmatrix} = \begin{bmatrix} 228 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 14 \\ y_p \end{bmatrix} = \begin{bmatrix} 220 \\ 140 \end{bmatrix}$$

$$\begin{bmatrix} x_p - 7 \\ y_p \end{bmatrix} = \begin{bmatrix} 227 \\ 140 \end{bmatrix} \quad \begin{bmatrix} x_p - 15 \\ y_p \end{bmatrix} = \begin{bmatrix} 219 \\ 140 \end{bmatrix}$$

Stereo Matching

❖ Method 1

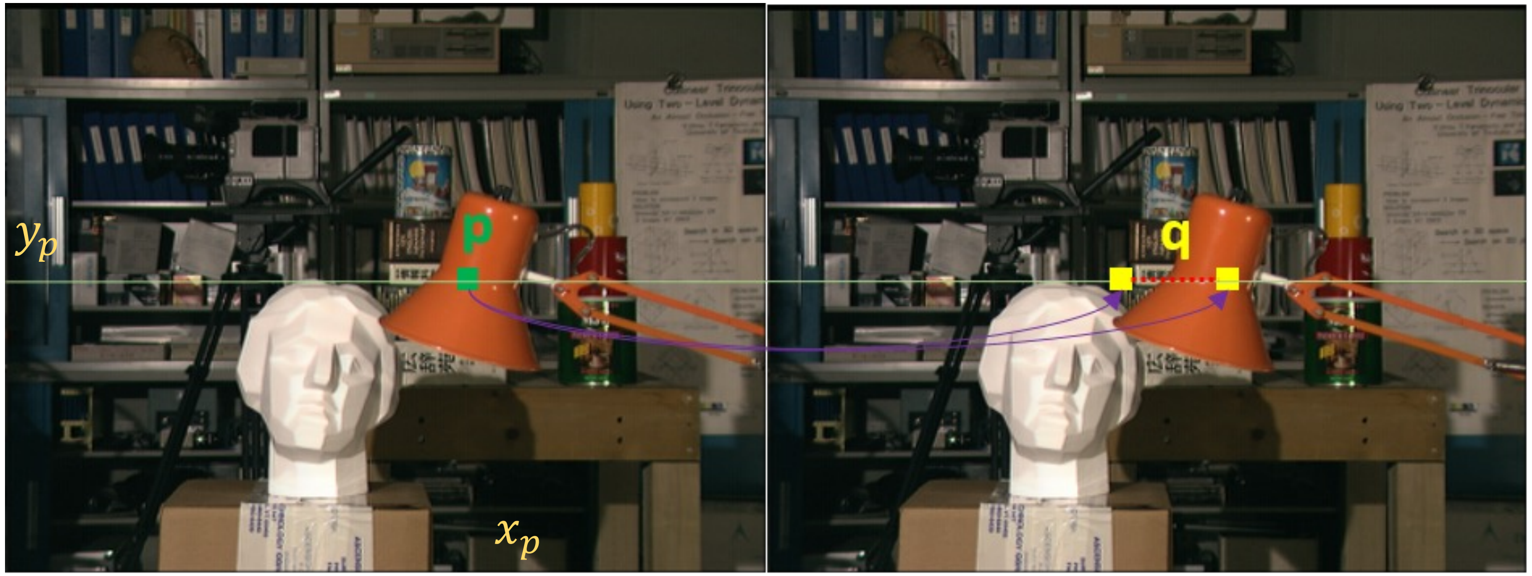
L is the left image

R is the right image

$L(\mathbf{p})$ is the (vector) value of **p**

$$\boldsymbol{p} = \begin{bmatrix} x_p \\ y_p \end{bmatrix} \qquad d \in D$$

$$\boldsymbol{d} = \begin{bmatrix} d \\ 0 \end{bmatrix} \qquad C_1(\boldsymbol{p}, \boldsymbol{d}) = |L(\boldsymbol{p}) - R(\boldsymbol{p} - \boldsymbol{d})|$$



Left Image

Right Image

Finding d so that $C_1(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{d})$ is minimum.

$$d = \underset{d \in D}{\operatorname{argmin}}\left(C_1(\boldsymbol{p}, \boldsymbol{d})\right)$$

Then, d is the value for the pixel **p** in disparity map

Stereo Matching

❖ Method 1: Result

$$\boldsymbol{d} = \begin{bmatrix} d \\ 0 \end{bmatrix} \qquad d \in D$$

$$C_1(\boldsymbol{p}, \boldsymbol{d}) = |L(\boldsymbol{p}) - R(\boldsymbol{p} - \boldsymbol{d})|$$

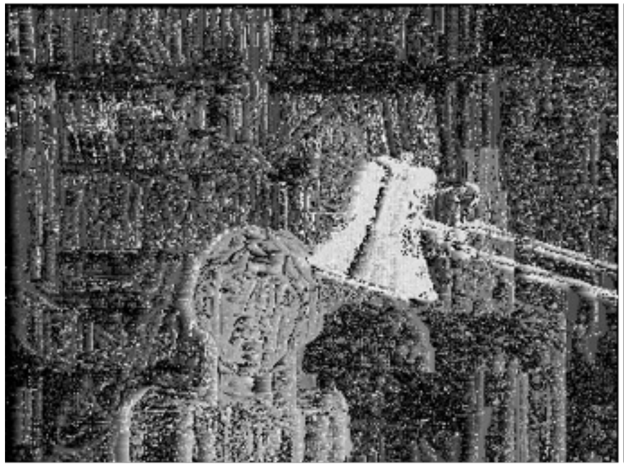
Finding d so that $C_1(\boldsymbol{p}, \boldsymbol{d})$ is minimum.

$$d = \underset{d \in D}{\operatorname{argmin}}\left(C_1(\boldsymbol{p}, \boldsymbol{d})\right)$$

Then, d is the value for the pixel **p** in disparity map

Left Image

Right Image



Disparity Map

Ground Truth

How to do Research

Define Your Research Question or Topic

⇩

Background Research

⇩

Study and Implement Existing Methods

⇩

Develop a Research Plan

⇩

Implement Your Proposed Methods

⇩

Evaluate Performance of Compared Methods

AI VIETNAM
All-in-One Course

Stereo Matching

❖ Method 1: Implementation

$$\boldsymbol{d} = \begin{bmatrix} d \\ 0 \end{bmatrix} \qquad d \in D$$

$$C_1(\boldsymbol{p}, \boldsymbol{d}) = |L(\boldsymbol{p}) - R(\boldsymbol{p} - \boldsymbol{d})|$$

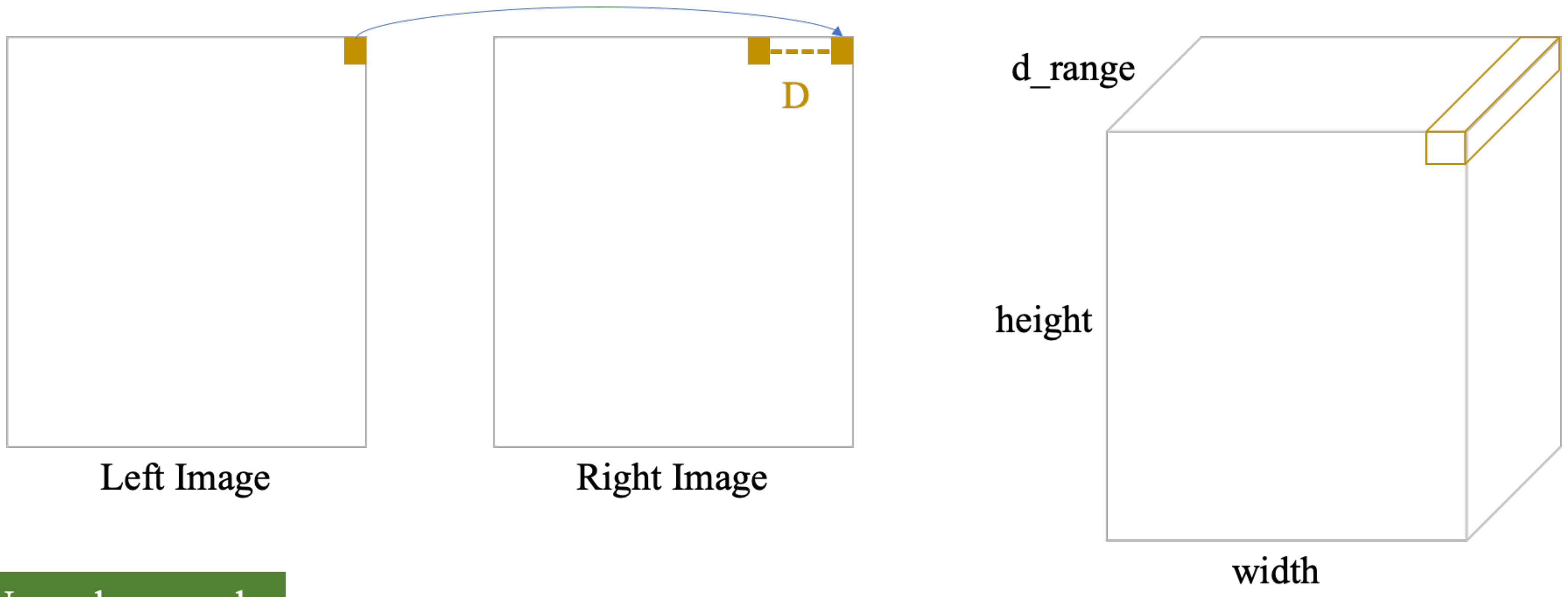
Finding d so that $C_1(\boldsymbol{p}, \boldsymbol{d})$ is minimum.

$$d = \underset{d \in D}{\operatorname{argmin}}\big(C_1(\boldsymbol{p}, \boldsymbol{d})\big)$$

Then, d is the value for the pixel **p** in disparity map

```python
# disparity map
depth = np.zeros((height, width), np.uint8)
scale = 255 / disparity_range

for y in range(height):
    for x in range(width):

        # find argmin_d
        disparity = 0
        cost_min  = abs(left[y, x] - right[y, x])
        for d in range(disparity_range):
            if (x - d) < 0:
                cost = 255
            else:
                cost = abs(left[y, x] - right[y, x - d])

            # update cost_min
            if cost < cost_min:
                cost_min  = cost
                disparity = d

        # set to the disparity map
        depth[y, x] = disparity*scale

# save
cv2.imwrite('images/disparity_ad.png', depth)
```

Stereo Matching



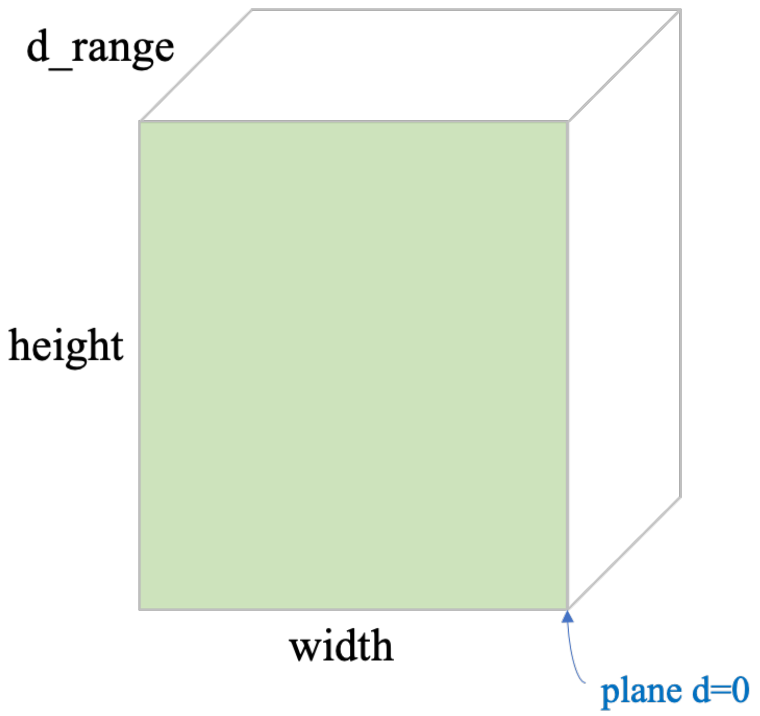
Normal approach

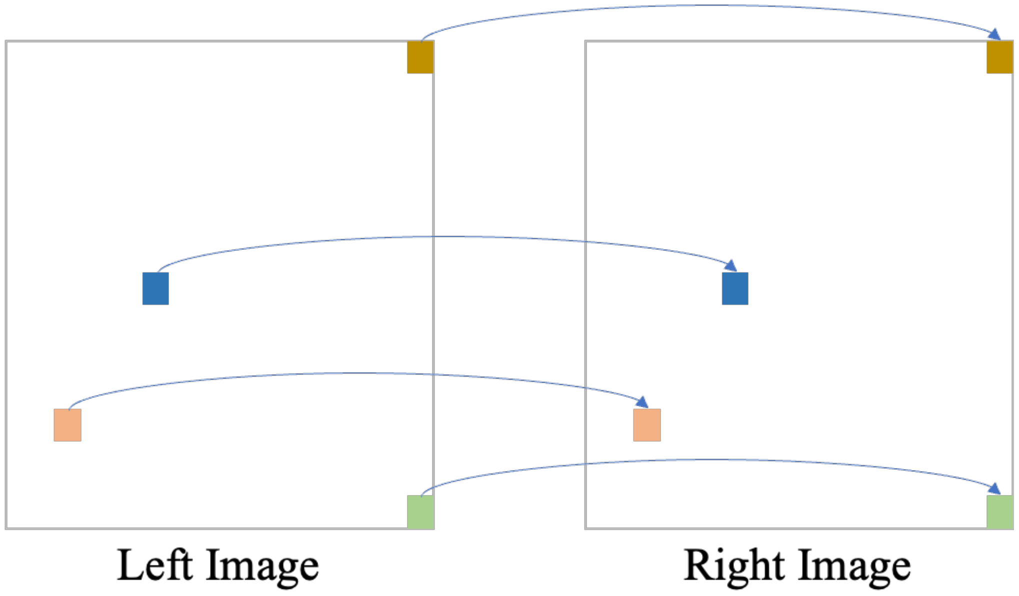
Each pixel p in the left image has D candidate pixel q in the right image.

Then, D cost values are computed from D pairs (**p**, **q**)

Stereo Matching

Vectorization





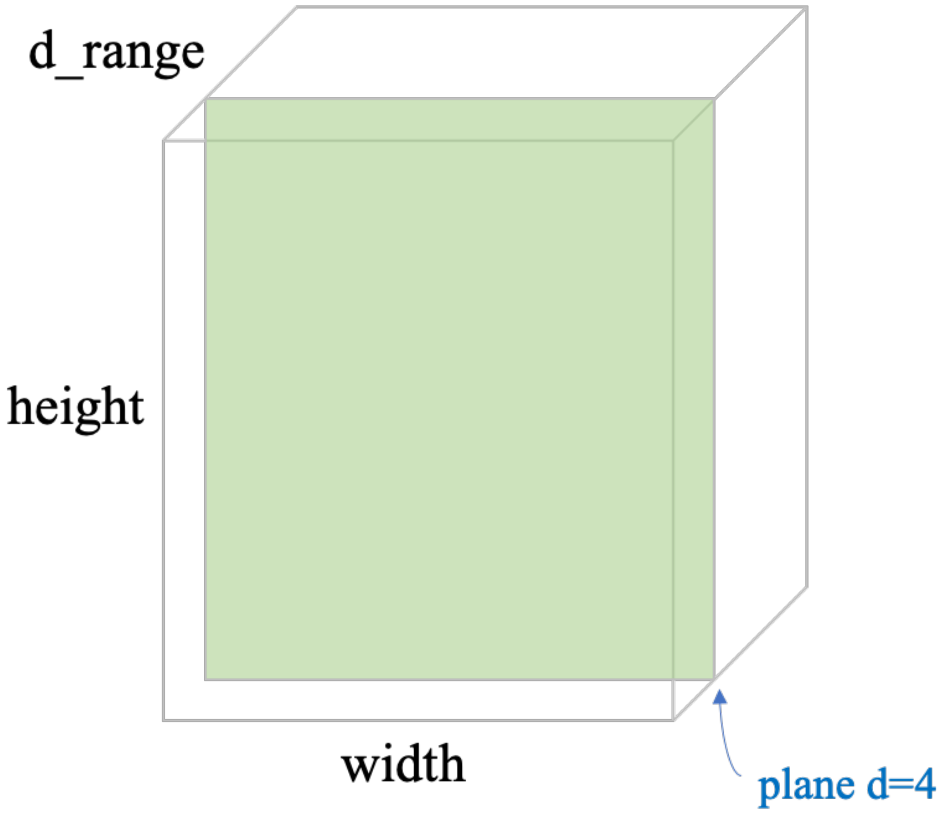
Cost values for a disparity d=0

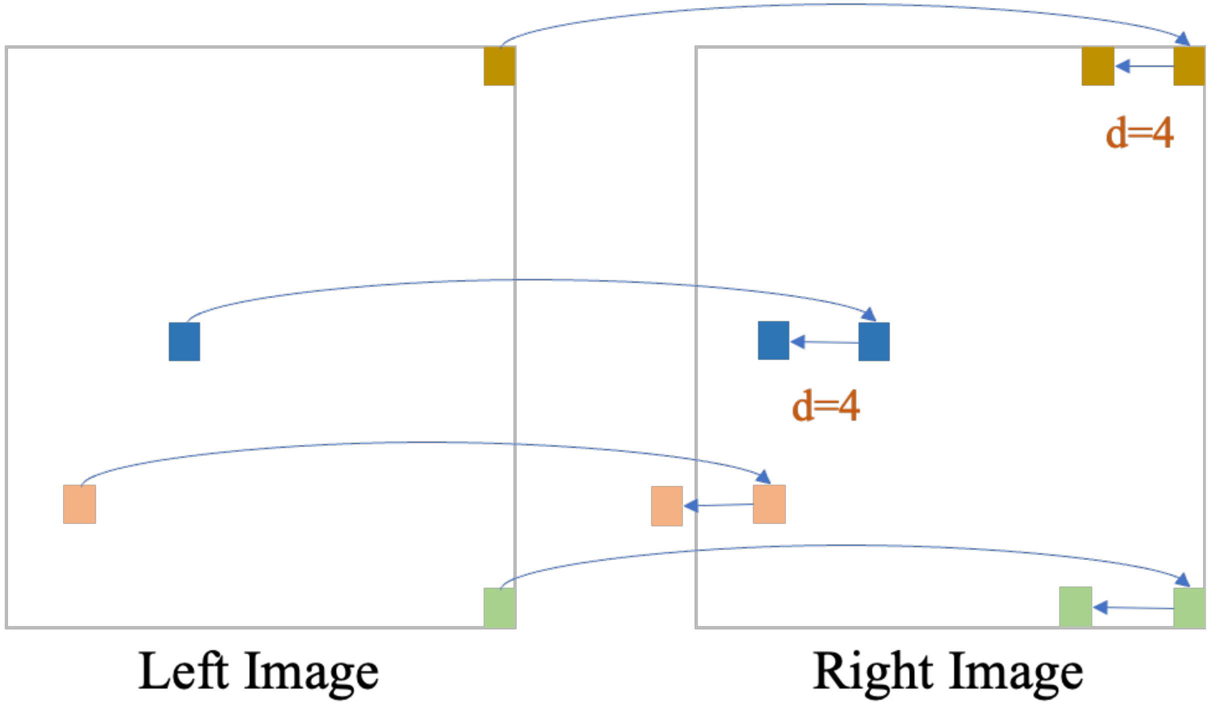
Given C_d is a cost plane for a disparity d

$$C_0 = |L - R|$$

Stereo Matching

Vectorization

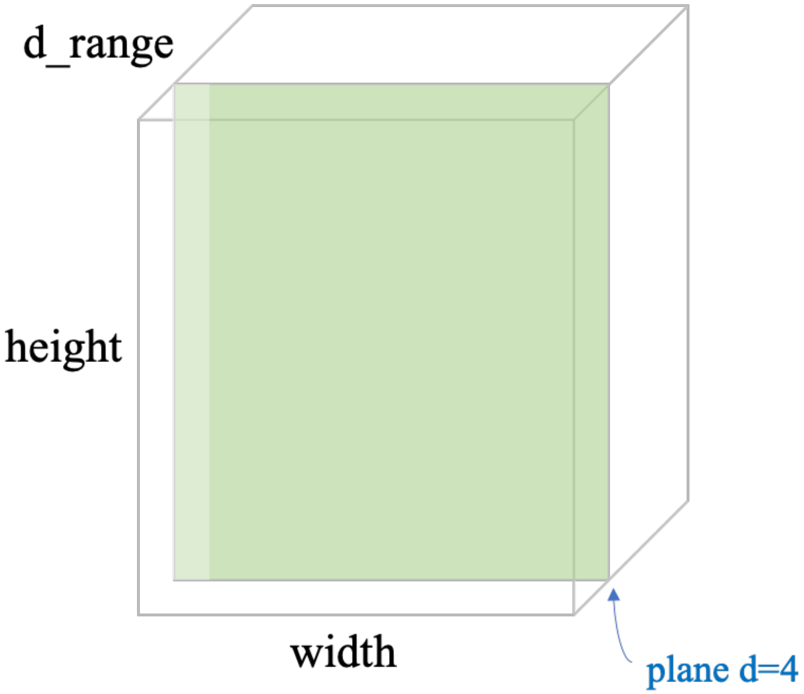


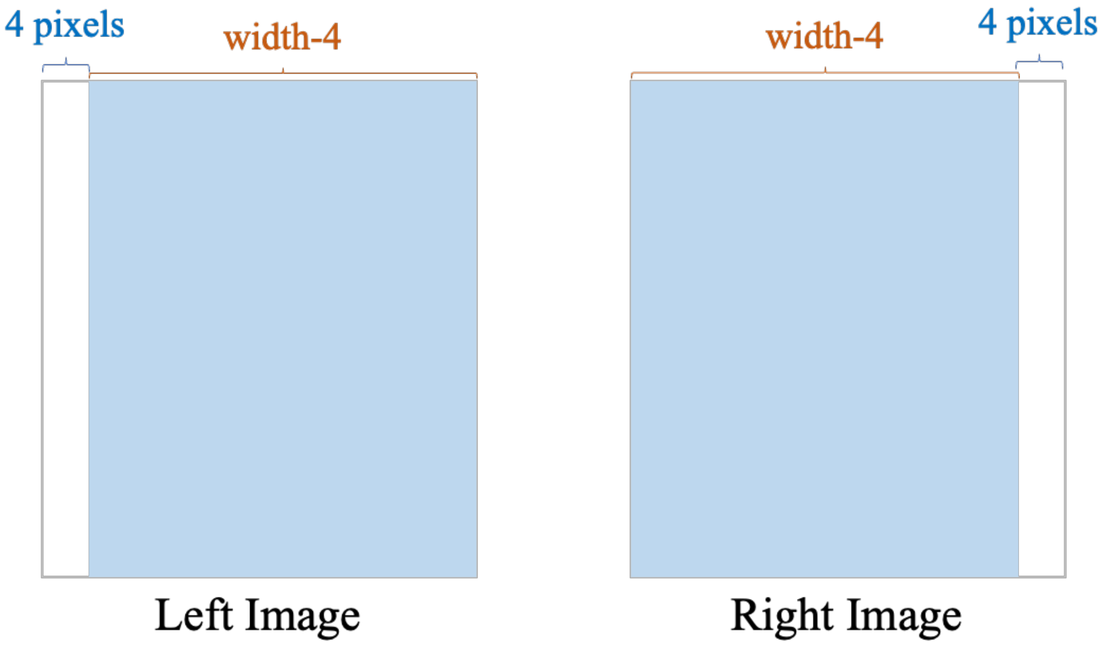


Cost values for a disparity d=4

Stereo Matching

Vectorization





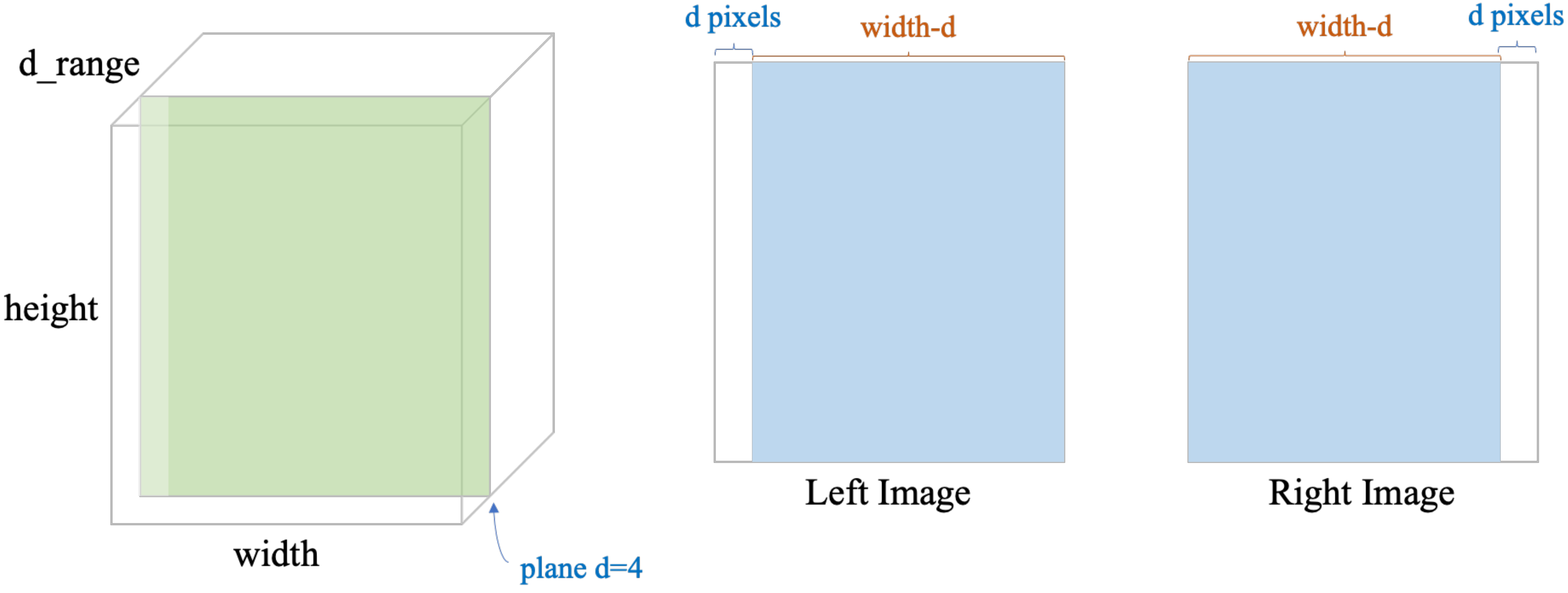
Given C_d is a cost plane for a disparity d

$$C_4 = |L[:, 4{:}W] - R[:, 0{:}W-4]|$$

Cost values for a disparity d=4

Stereo Matching

Vectorization



Given C_d is a cost plane for a disparity d

$$C_d = |L[:, d: W] - R[:, 0: W - d]|$$

Cost values for a disparity d

Outline

- How to do Research
- Background of the Research Topic
- Implement Existing Methods
- How to Write Research Plan
- Assignment

Stereo Matching

❖ Method 1: Vectorization

$$\boldsymbol{d} = \begin{bmatrix} d \\ 0 \end{bmatrix} \qquad d \in D$$

$$C_1(\boldsymbol{p}, \boldsymbol{d}) = |L(\boldsymbol{p}) - R(\boldsymbol{p} - \boldsymbol{d})|$$

Finding d so that $C_1(\boldsymbol{p}, \boldsymbol{d})$ is minimum.

$$d = \underset{d \in D}{\operatorname{argmin}}\left(C_1(\boldsymbol{p}, \boldsymbol{d})\right)$$

Then, d is the value for the pixel **p** in disparity map

```python
left  = cv2.imread(left_img, 0)
right = cv2.imread(right_img, 0)

left  = left.astype(np.float32)
right = right.astype(np.float32)

# compute cost space
cost_space = np.zeros((height, width, d_range))
for d in range(disparity_range):
    left_d  = left[:,d:width]
    right_d = right[:, 0:width-d]
    cost_d  = np.abs(left_d-right_d)
    cost_space[:, d:width, d] = cost_d

# compute disparity from cost space
scale = 255 / disparity_range
depth = np.argmin(cost_space, axis=2)
depth = depth*scale
depth = depth.astype(np.uint8)

# save
cv2.imwrite('images/disparity_ad.png', depth)
```

Stereo Matching

❖ Method 2

L is the left image

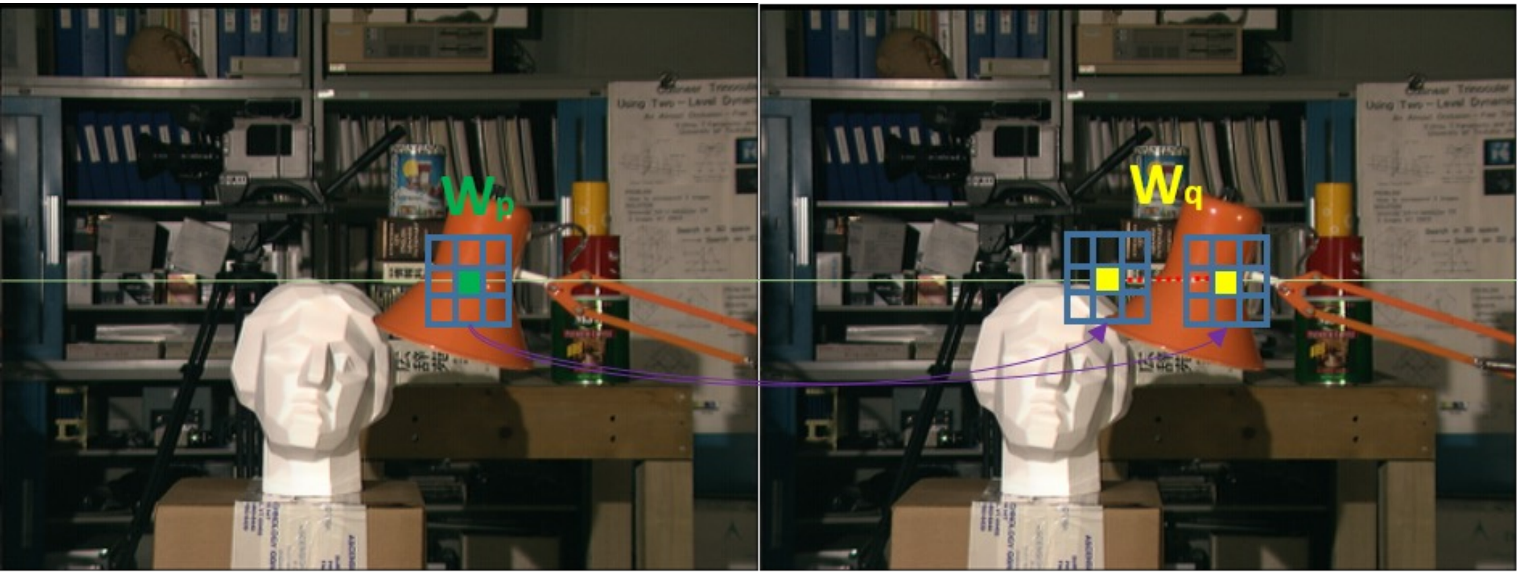
R is the right image

$L(\boldsymbol{p})$ is the (vector) value of **p**

$\boldsymbol{p} = \begin{bmatrix} x_p \\ y_p \end{bmatrix}$ $W_{\boldsymbol{p}}$ is a window centered at p

$\boldsymbol{d} = \begin{bmatrix} d \\ 0 \end{bmatrix}$ $C_2(\boldsymbol{p}, \boldsymbol{d}) = \sum_{\boldsymbol{u} \in W_{\boldsymbol{p}}} |L(\boldsymbol{u}) - R(\boldsymbol{u} - \boldsymbol{d})|$

$d \in D$



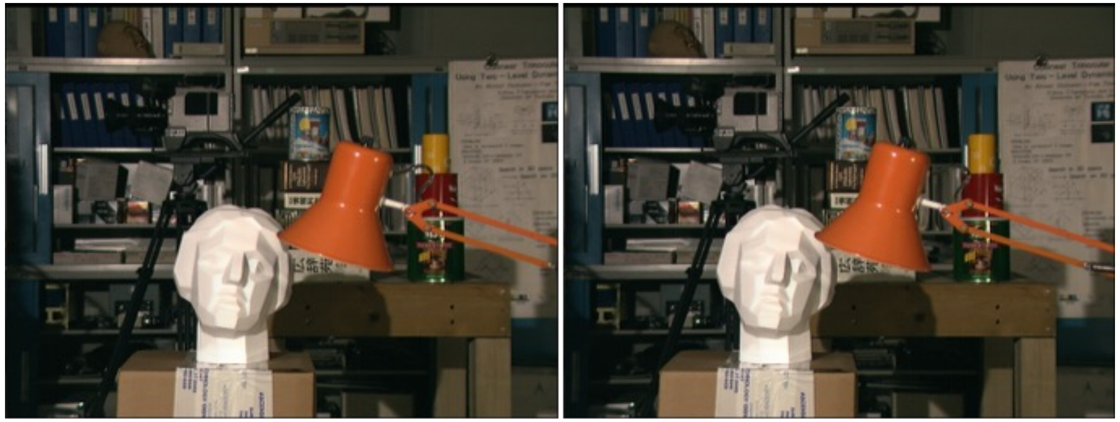
Left Image Right Image

Finding d so that $C_2(\boldsymbol{p}, \boldsymbol{d})$ is minimum.

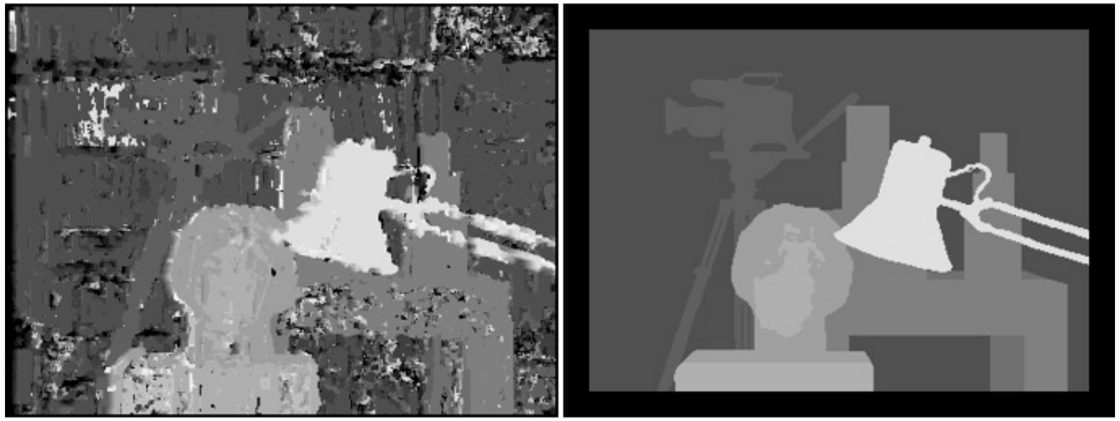
$$d = \underset{d \in D}{\operatorname{argmin}}\left(C_2(\boldsymbol{p}, \boldsymbol{d})\right)$$

Then, d is the value for the pixel **p** in disparity map

Stereo Matching

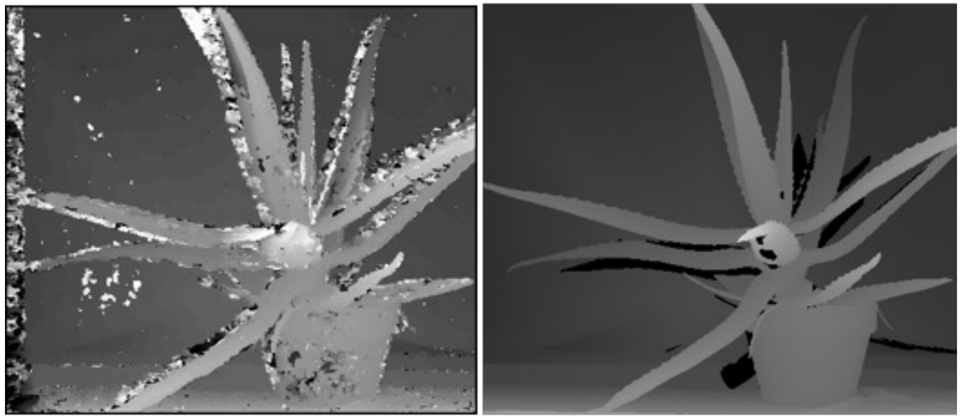


Left Image Right Image



Disparity Map Ground Truth

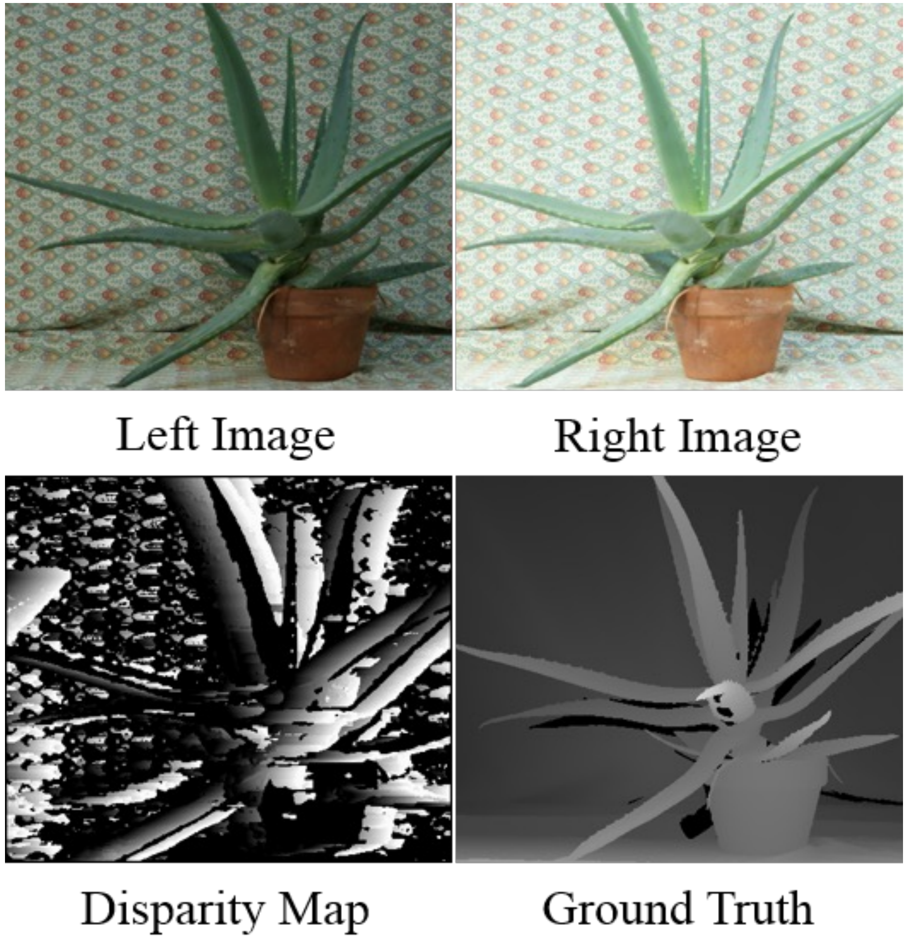
Left Image Right Image

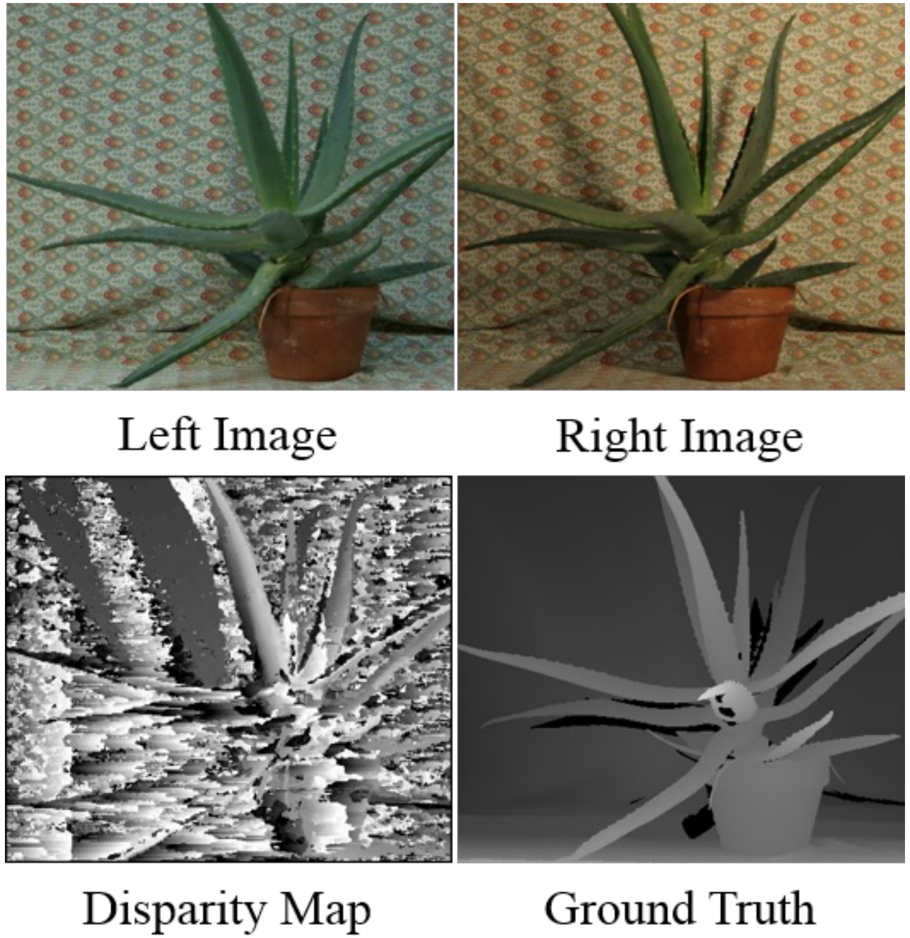


Disparity Map Ground Truth

Stereo Matching

❖ Aloe stereo pair





Stereo Matching

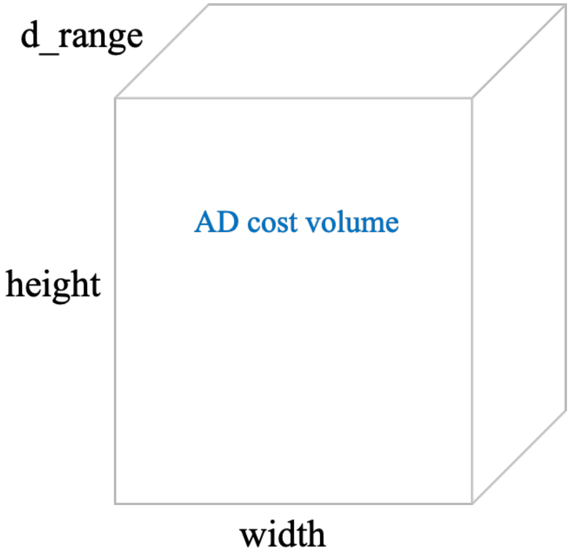
$$\boldsymbol{p} = \begin{bmatrix} x_p \\ y_p \end{bmatrix} \qquad W_{\boldsymbol{p}} \text{ and } W_{\boldsymbol{q}} \text{ are two windows centered at p and q}$$

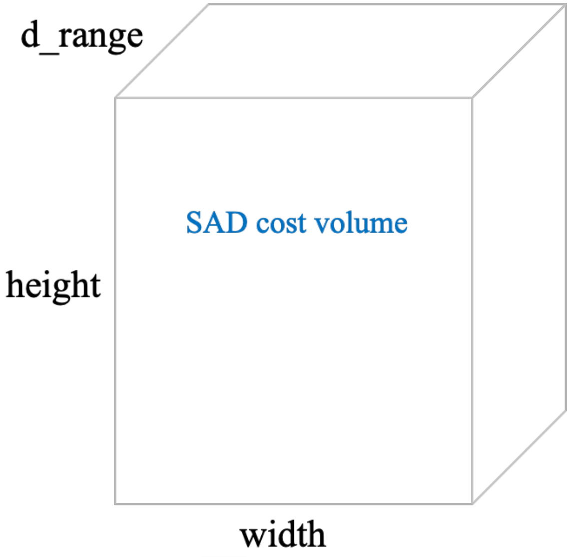
$$\boldsymbol{d} = \begin{bmatrix} d \\ 0 \end{bmatrix} \qquad C_2(\boldsymbol{p}, \boldsymbol{d}) = \sum_{\boldsymbol{u} \in W_{\boldsymbol{p}}} |L(\boldsymbol{u}) - R(\boldsymbol{u} - \boldsymbol{d})|$$

$$d \in D$$

❖ **Method 2: Implementation**

$$C_1(\boldsymbol{p}, \boldsymbol{d}) = |L(\boldsymbol{p}) - R(\boldsymbol{p} - \boldsymbol{d})|$$





$$C_2(\boldsymbol{p}, \boldsymbol{d}) = \sum_{\boldsymbol{u} \in W_{\boldsymbol{p}}} |L(\boldsymbol{u}) - R(\boldsymbol{u} - \boldsymbol{d})| = \sum_{\boldsymbol{u} \in W_{\boldsymbol{p}}} C_1(\boldsymbol{u}, \boldsymbol{d})$$

Cosine similarity

Cosine similarity (cs) được dùng để đo mức độ giống nhau/tương đồng giữa hai vector

Gọi $\vec{x}$ và $\vec{y}$ là hai vector, cs được tính như sau

$$\text{cs}(\vec{x},\vec{y}) = \frac{\vec{x}.\vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \frac{\sum_1^n x_i y_i}{\sqrt{\sum_1^n x_i^2}\sqrt{\sum_1^n y_i^2}}$$

Tính chất 1: $\text{cs}(\vec{x},\vec{y}) = \text{cs}(a\vec{x}, b\vec{y})$

$$\text{cs}(a\vec{x}, b\vec{y}) = \frac{a\vec{x}.b\vec{y}}{\|a\vec{x}\|\|b\vec{y}\|} = \frac{\sum_1^n a x_i b y_i}{\sqrt{\sum_1^n a^2 x_i^2}\sqrt{\sum_1^n b^2 y_i^2}}$$

$$= \frac{ab\sum_1^n x_i y_i}{\sqrt{a^2\sum_1^n x_i^2}\sqrt{b^2\sum_1^n y_i^2}}$$

$$= \frac{\sum_1^n x_i y_i}{\sqrt{\sum_1^n x_i^2}\sqrt{\sum_1^n y_i^2}} = \text{cs}(\vec{x},\vec{y})$$

Tính chất 2: $\text{cs}(\vec{x},\vec{y}) \neq \text{cs}(\vec{x}+c, \vec{y}+d)$

Ví dụ:

$$\vec{x} = [4, 2, 1, 2]^T$$

$$\vec{y} = [1, 2, 2, 0]^T$$

$$\vec{u} = 2\vec{x} = [8, 4, 2, 4]^T$$

$$\vec{v} = 3\vec{y} = [3, 6, 6, 0]^T$$

$$\text{cs}(\vec{x},\vec{y}) = \frac{4*1+2*2+1*2+2*0}{\sqrt{4^2+2^2+1^2+2^2}\sqrt{1^2+2^2+2^2+0}}$$

$$= \frac{10}{\sqrt{25}\sqrt{9}} = \frac{10}{15} = 0.67$$

$$\text{cs}(\vec{u},\vec{v}) = \frac{8*3+4*6+2*6+4*0}{\sqrt{8^2+4^2+2^2+4^2}\sqrt{3^2+6^2+6^2+0}}$$

$$= \frac{60}{\sqrt{100}\sqrt{81}} = \frac{60}{90} = 0.67$$

$$= \text{cs}(\vec{x},\vec{y})$$

Stereo Matching

❖ Method 3

L is the left image

R is the right image

$L(\boldsymbol{p})$ is the (vector) value of **p**

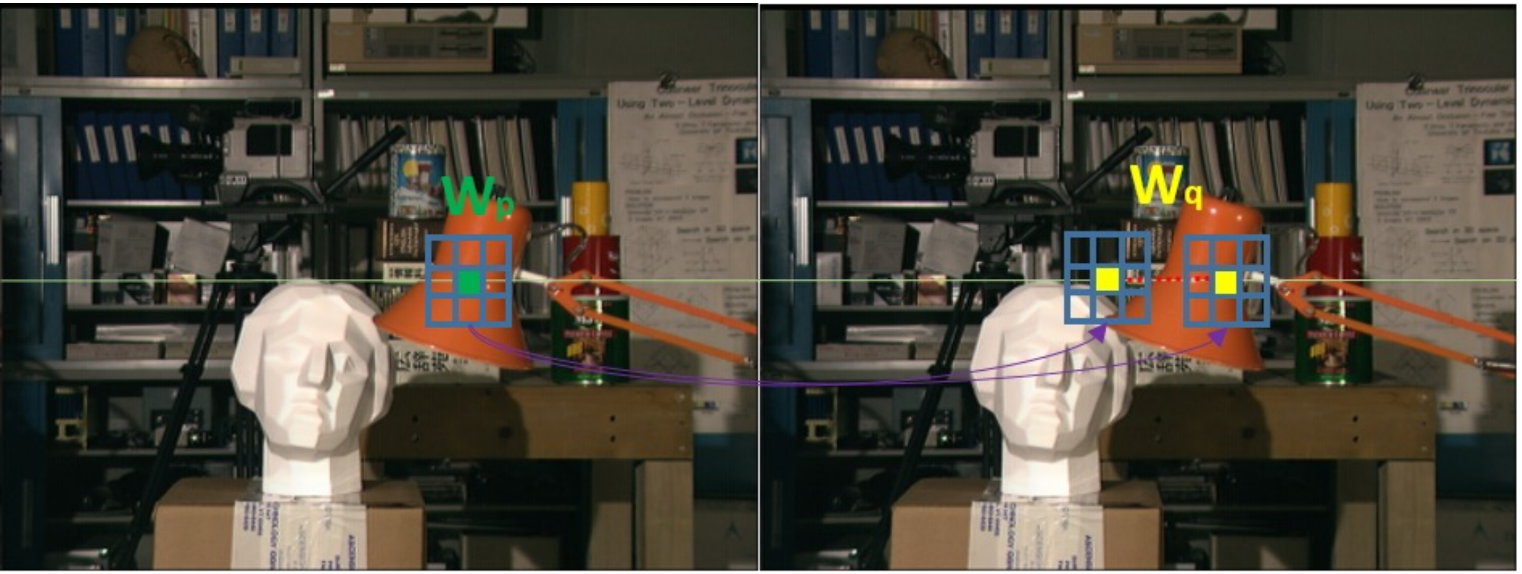
$$\boldsymbol{p} = \begin{bmatrix} x_p \\ y_p \end{bmatrix}$$

$W_{\boldsymbol{p}}$ is a window centered at p

$$\boldsymbol{d} = \begin{bmatrix} d \\ 0 \end{bmatrix}$$

$$C_3(\boldsymbol{p}, \boldsymbol{d}) = CS\left(L(W_{\boldsymbol{p}}), R(W_{\boldsymbol{p}-\boldsymbol{d}})\right)$$

$d \in D$



Left Image Right Image

Finding d so that $C_3(\boldsymbol{p}, \boldsymbol{d})$ is maximum.

$$d = \underset{d \in D}{\operatorname{argmax}}\left(C_3(\boldsymbol{p}, \boldsymbol{d})\right)$$

Then, d is the value for the pixel **p** in disparity map

Cosine Similarity for Matching Cost

Cosine similarity (cs) is used to measure the similarity between two vectors

Let $\vec{x}$ and $\vec{y}$ be two vectors, cs is defined as

$$cs(\vec{x},\vec{y}) = \frac{\vec{x}.\vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \frac{\sum_1^n x_i y_i}{\sqrt{\sum_1^n x_i^2}\sqrt{\sum_1^n y_i^2}}$$

Property 1: $cs(\vec{x},\vec{y}) = cs(a\vec{x}, b\vec{y})$; $ab > 0$

$$cs(a\vec{x}, b\vec{y}) = \frac{a\vec{x}.b\vec{y}}{\|a\vec{x}\|\|b\vec{y}\|} = \frac{\sum_1^n ax_i by_i}{\sqrt{\sum_1^n a^2x_i^2}\sqrt{\sum_1^n b^2y_i^2}}$$

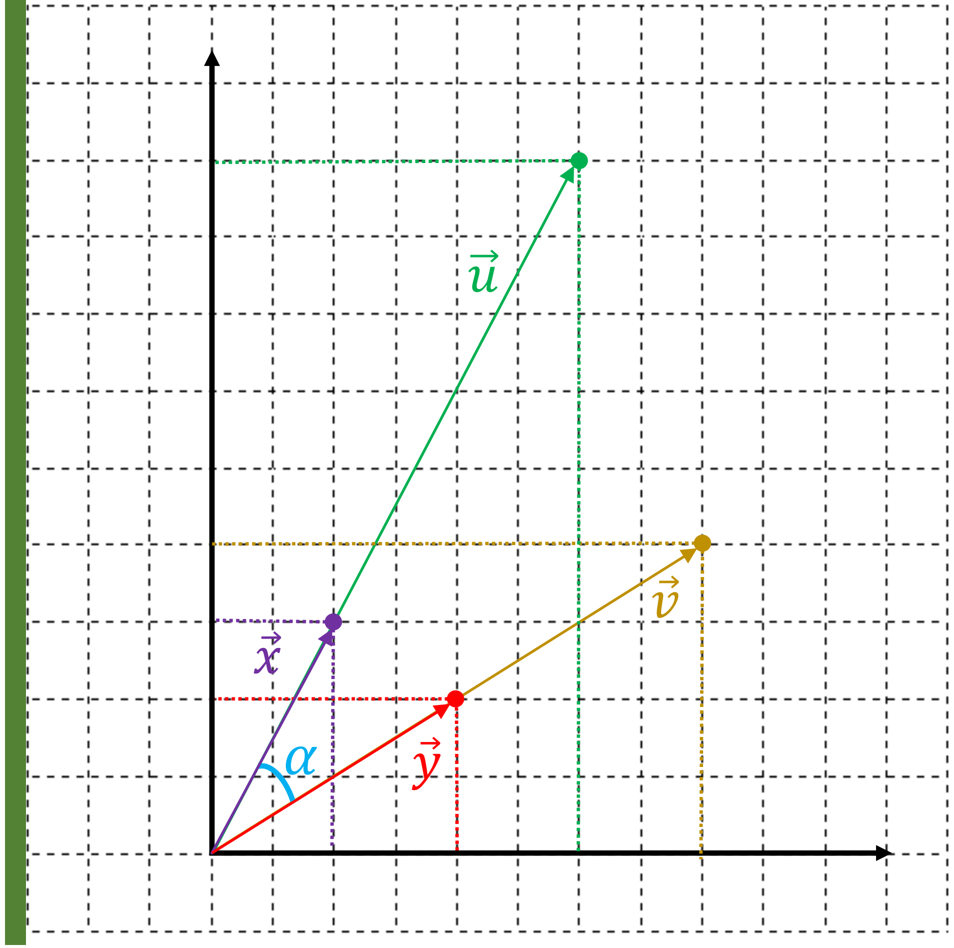
$$= \frac{ab\sum_1^n x_i y_i}{\sqrt{a^2\sum_1^n x_i^2}\sqrt{b^2\sum_1^n y_i^2}}$$

$$= \frac{\sum_1^n x_i y_i}{\sqrt{\sum_1^n x_i^2}\sqrt{\sum_1^n y_i^2}} = cs(\vec{x},\vec{y})$$

Property 2: $cs(\vec{x},\vec{y}) \neq cs(\vec{x}+c, \vec{y}+d)$

$\vec{x} = \begin{pmatrix}2\\3\end{pmatrix}$ $\qquad$ $\vec{u} = 3 * \vec{x} = \begin{pmatrix}6\\9\end{pmatrix}$

$\vec{y} = \begin{pmatrix}4\\2\end{pmatrix}$ $\qquad$ $\vec{v} = 2 * \vec{y} = \begin{pmatrix}8\\4\end{pmatrix}$



Cosine similarity

Cosine similarity (cs) được dùng để đo mức độ giống nhau/tương đồng giữa hai vector

Gọi $\vec{x}$ và $\vec{y}$ là hai vector, cs được tính như sau

$$\mathrm{cs}(\vec{x},\vec{y}) = \frac{\vec{x}.\vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \frac{\sum_1^n x_i y_i}{\sqrt{\sum_1^n x_i^2}\sqrt{\sum_1^n y_i^2}}$$

Tính chất 1: $\mathrm{cs}(\vec{x},\vec{y}) = \mathrm{cs}(a\vec{x}, b\vec{y})$

$$\mathrm{cs}(a\vec{x}, b\vec{y}) = \frac{a\vec{x}.b\vec{y}}{\|a\vec{x}\|\|b\vec{y}\|} = \frac{\sum_1^n a x_i b y_i}{\sqrt{\sum_1^n a^2 x_i^2}\sqrt{\sum_1^n b^2 y_i^2}}$$

$$= \frac{ab\sum_1^n x_i y_i}{\sqrt{a^2\sum_1^n x_i^2}\sqrt{b^2\sum_1^n y_i^2}}$$

$$= \frac{\sum_1^n x_i y_i}{\sqrt{\sum_1^n x_i^2}\sqrt{\sum_1^n y_i^2}} = \mathrm{cs}(\vec{x},\vec{y})$$

Tính chất 2: $\mathrm{cs}(\vec{x},\vec{y}) \neq \mathrm{cs}(\vec{x}+c, \vec{y}+d)$

Ví dụ:

$\vec{x} = [4, 2, 1, 2]^T$

$\vec{y} = [1, 2, 2, 0]^T$

$\vec{u} = 2\vec{x} = [8, 4, 2, 4]^T$

$\vec{v} = 3\vec{y} = [3, 6, 6, 0]^T$

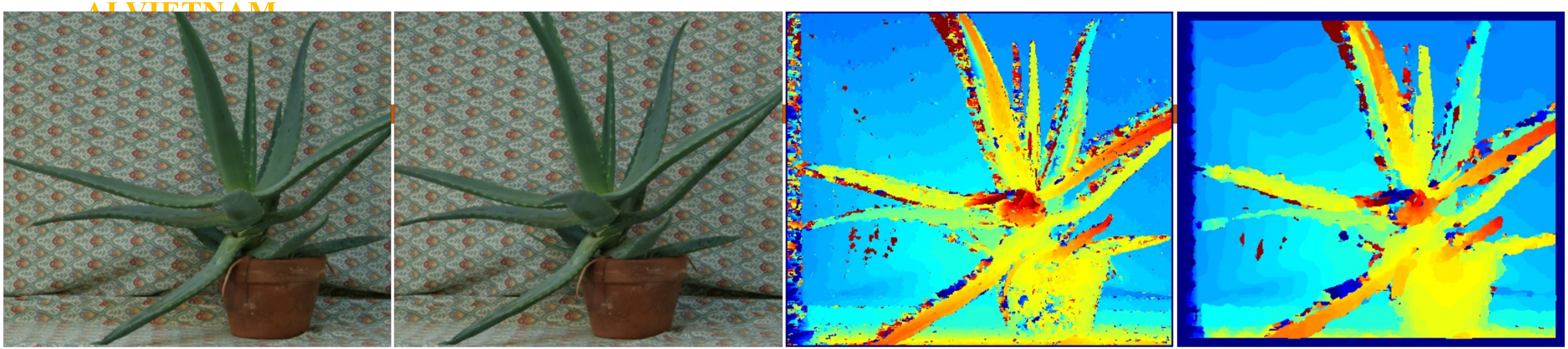
$$\mathrm{cs}(\vec{x},\vec{y}) = \frac{4*1+2*2+1*2+2*0}{\sqrt{4^2+2^2+1^2+2^2}\sqrt{1^2+2^2+2^2+0}}$$

$$= \frac{10}{\sqrt{25}\sqrt{9}} = \frac{10}{15} = 0.67$$

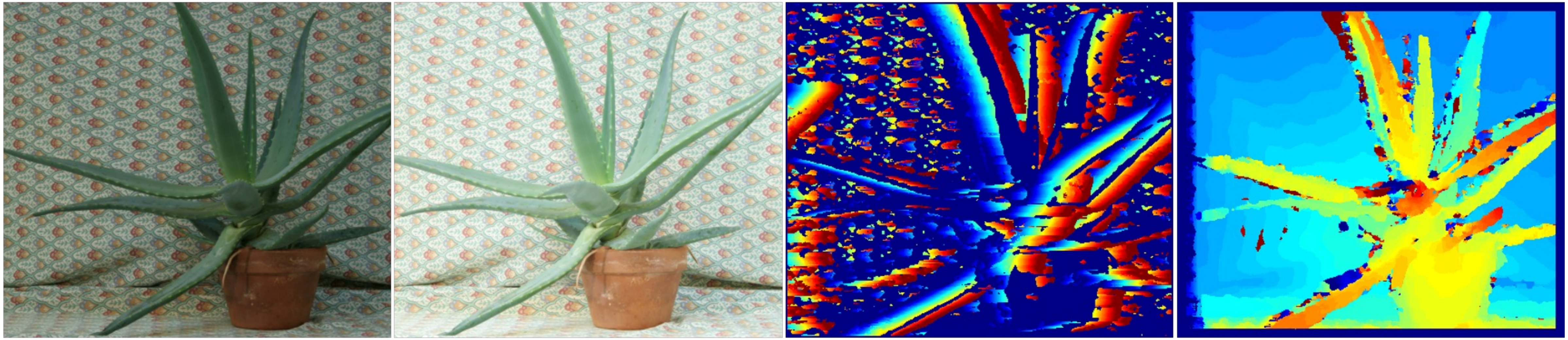
$$\mathrm{cs}(\vec{u},\vec{v}) = \frac{8*3+4*6+2*6+4*0}{\sqrt{8^2+4^2+2^2+4^2}\sqrt{3^2+6^2+6^2+0}}$$

$$= \frac{60}{\sqrt{100}\sqrt{81}} = \frac{60}{90} = 0.67$$

$$= \mathrm{cs}(\vec{x},\vec{y})$$



Ảnh stereo có cùng độ sáng | Absolute difference | Cosine Similarity



Ảnh stereo khác độ sáng | Absolute difference | Cosine Similarity

Cosine similarity hoạt động ổn định khi ảnh stereo thay đổi độ sáng

Correlation Coefficient Matching Cost

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

ignore the differences between population and sample

❖ Definition

Công thức: Gọi x,y là hai biến ngẫu nhiên

$$\rho_{xy} = \frac{E[(x-\mu_x)(y-\mu_y)]}{\sqrt{var(x)}\sqrt{var(y)}}$$

$$= \frac{n(\sum_i x_i y_i) - (\sum_i x_i)(\sum_i y_i)}{\sqrt{n\sum_i x_i^2 - (\sum_i x_i)^2}\sqrt{n\sum_i y_i^2 - (\sum_i y_i)^2}}$$

Tính chất 1

$$-1 \leq \rho_{xy} \leq 1$$

Tương quan nghịch ← → Tương quan thuận

Tính chất 2

$$\rho_{xy} = \rho_{uv}$$

trong đó

$$u = ax + b$$

$$v = cy + d$$

Ví dụ 1

x = [7, 18, 29, 2, 10, 9, 9]
y = [1, 6, 12, 8, 6, 21, 10]

$$\rho_{xy} = \frac{E[(x-\mu_x)(y-\mu_y)]}{\sqrt{var(x)}\sqrt{var(y)}}$$

$$= \frac{n*818 - 84*64}{\sqrt{n*1480 - 7056}\sqrt{n*822 - 4096}} = 0.149$$

Ví dụ 2

u=2*x-14 = [0, 22, 44, -10, 6, 4, 4]
v=y+2 = [3, 8, 14, 10, 8, 23, 12]

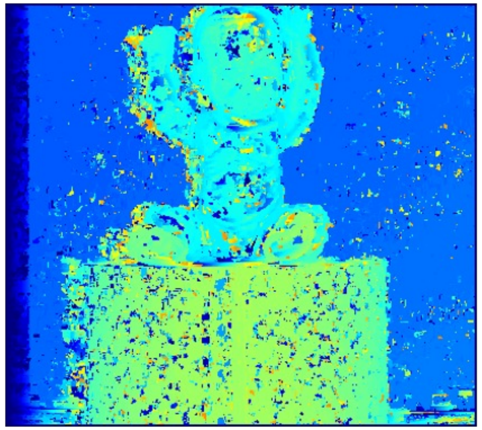
$$\rho_{uv} = \frac{E[(u-\mu_u)(v-\mu_v)]}{\sqrt{var(u)}\sqrt{var(v)}}$$

$$= \frac{n*880 - 70*78}{\sqrt{n*2588 - 4900}\sqrt{n*1106 - 6084}} = 0.149$$

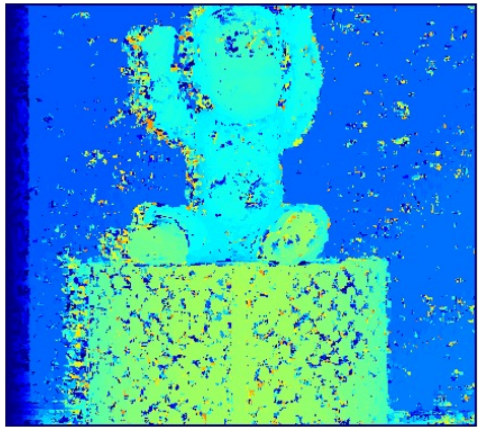
Correlation for Stereo Matching Cost

Left image

Right image



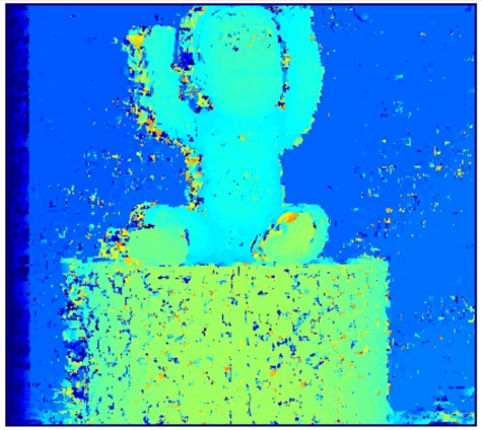
Cosine Result/ RMS = 5.6



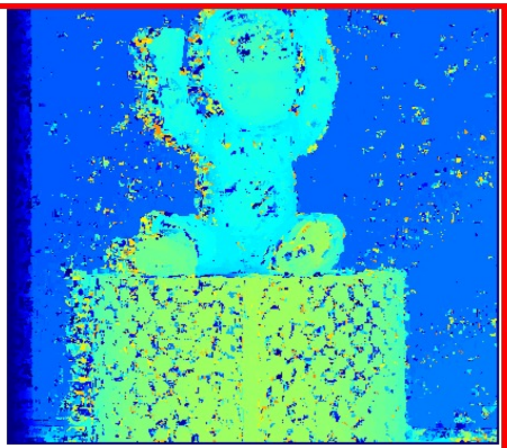
Correlation Result / RMS = 5.38

Left image

Right image



Cosine Result / RMS = 4.98



Correlation Result / RMS = 5.36

Outline

- How to do Research
- Background of the Research Topic
- Implement Existing Methods
- How to Write Research Plan
- Assignment

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title:
   Provide a clear and concise title that reflects the main focus of your research
2. Introduction
   Briefly introduce your research topic, explaining its significance and relevance.
3. Background and Literature Review
   Summarize the existing knowledge and research related to your topic. Highlight key studies, theories, or gaps in the literature that your research aims to address.
4. Research Objectives
   What do you aim to achieve or discover through your study?
5. Research Questions
   Formulate the research questions or hypotheses that will guide your investigation
6. Methodology
   Detail the research design, methods, and procedures you plan to employ
7. Timeline
   Provide a timeline for your research activities. Include milestones, deadlines, and key events.
8. Resources
   Outline the resources you'll need for your research, such as access to specific databases, equipment, or funding.
9. Expected Result
   Discuss the anticipated outcomes of your research.
10. Significant and Contribution
    Explain the potential significance of your research.

Research Plan

Define Your Research Question or Topic

Background Research

Study and Implement Existing Methods

Develop a Research Plan

Implement Your Proposed Methods

Evaluate Performance of Compared Methods

1. Research Title:

Robust Stereo Matching Various Weather Conditions

2. Introduction

Stereo matching is a fundamental computer vision task that involves the computation of dense disparity maps from stereo image pairs. The goal is to establish correspondences between pixels in the left and right images, providing depth information crucial for 3D perception. However, the accuracy and reliability of stereo matching algorithms are often challenged by various real-world factors such as occlusions, varying lighting conditions, and textureless regions.

In response to these challenges, the concept of robust stereo matching has emerged as a critical area of research within computer vision. Robust stereo matching aims to enhance the accuracy and efficiency of disparity estimation under adverse conditions, ensuring the reliability of depth information in a wide range of scenarios.

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title: (done)

2. Introduction

…

Robust stereo matching techniques incorporate advanced algorithms and methodologies to address common issues encountered in stereo vision. These may include the development of adaptive matching cost functions, the integration of machine learning approaches for robust feature extraction, and the incorporation of strategies to handle occlusions and textureless regions effectively.

This field plays a pivotal role in applications such as autonomous vehicles, robotics, and augmented reality, where precise depth perception is essential for making informed decisions. As researchers delve into the complexities of real-world stereo vision challenges, the development of robust stereo matching algorithms becomes increasingly crucial to unlocking the full potential of stereo imaging systems in diverse and dynamic environments.

Research Plan

Define Your Research Question or Topic

⬇

Background Research

⬇

Study and Implement Existing Methods

⬇

Develop a Research Plan

⬇

Implement Your Proposed Methods

⬇

Evaluate Performance of Compared Methods

1. Research Title (done)
2. Introduction (done)
3. Background and Literature Review

3.a Introduction

Stereo matching, a vital component of computer vision, aims to extract depth information from stereo image pairs. This literature review explores the historical evolution, foundational concepts, and recent advancements in stereo matching algorithms.

3.b Historical of Stereo Matching

Early stereo matching methods, such as block matching and dynamic programming, laid the groundwork for depth perception. Despite their simplicity, these approaches struggled with challenges like occlusions and varying lighting conditions.

3.Foundational Concepts in Stereo Matching

Key concepts, including epipolar geometry and the stereo correspondence problem, form the basis of stereo matching. Understanding these principles is essential for developing and evaluating stereo matching algorithms.

Research Plan

Define Your Research Question or Topic

Background Research

Study and Implement Existing Methods

Develop a Research Plan

Implement Your Proposed Methods

Evaluate Performance of Compared Methods

1. Research Title (done)
2. Introduction (done)
3. Background and Literature Review

3.d Traditional Stereo Matching Techniques
Classical methods like block-based matching and graph cuts provided initial solutions to stereo correspondence. However, they faced limitations in handling textureless regions and were sensitive to noise.

3.e Historical of Stereo Matching
Early stereo matching methods, such as block matching and dynamic programming, laid the groundwork for depth perception. Despite their simplicity, these approaches struggled with challenges like occlusions and varying lighting conditions.

3.f Challenges in Traditional Stereo Matching
Discuss challenges faced by traditional stereo matching techniques, such as sensitivity to noise, computational inefficiency, and difficulties in handling real-world scenarios with dynamic lighting conditions.

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title (done)
2. Introduction (done)
3. Background and Literature Review

3.g State-of-the-Art Stereo Matching Approaches

Highlight recent advancements in stereo matching algorithms. State-of-the-art approaches leverage machine learning, advanced matching cost functions, and efficient optimization techniques to address challenges and achieve high accuracy.

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title (done)
2. Introduction (done)
3. Background and Literature Review (done)
4. Research Objectives

Investigate and address challenges such as occlusions, varying lighting conditions, and textureless regions. Research could focus on developing algorithms that are robust in real-world scenarios, ensuring reliable depth estimation even in adverse conditions.

Explore the integration of machine learning, including deep learning techniques, to enhance feature extraction and matching cost computation. This objective aims to leverage the power of neural networks to improve the overall performance of stereo matching algorithms.

Develop real-time stereo matching algorithms for applications that require fast and efficient processing, such as robotics or autonomous vehicles. This objective involves optimizing algorithms for speed without compromising accuracy.

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title
2. Introduction
3. Background and Literature Review
4. Research Objectives
5. Research Questions

   RQ1: How to improve the performance of existing stereo matching methods under various lighting conditions and complex material properties?

   RQ2: How to obtain a real-time processing for current state-of-the-art stereo matching-based deep learning methods?

   RQ3: How to design an automatically learning the network architecture, its activation functions, and its parameters from data ?

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title:
2. Introduction
3. Background and Literature Review
4. Research Objectives
5. Research Questions
6. Methodology

a. Data Preprocessing:

Noise Reduction: Apply preprocessing techniques to reduce noise in the stereo images. This may include Gaussian blurring, median filtering, or other denoising methods to improve the quality of input data.

Color Correction: Normalize lighting conditions and correct color disparities between the stereo images to ensure consistency in feature extraction.

b. Feature Extraction:

Use advanced feature extraction techniques to capture meaningful information from the stereo images.

Explore local features (e.g., SIFT, SURF) and global features (e.g., color histograms) that are robust to variations in lighting and texture.

c. Matching Cost Computation:

Employ adaptive matching cost functions that dynamically adjust to local image characteristics.

Experiment with different cost aggregation methods, such as cross-based or semi-global aggregation, to improve the accuracy of disparity maps in challenging regions.

Research Plan

Define Your Research Question or Topic

Background Research

Study and Implement Existing Methods

Develop a Research Plan

Implement Your Proposed Methods

Evaluate Performance of Compared Methods

1. Research Title:
2. Introduction
3. Background and Literature Review
4. Research Objectives
5. Research Questions
6. Methodology

d. Machine Learning Integration:

Explore the integration of machine learning techniques, such as Convolutional Neural Networks (CNNs), to enhance feature extraction and matching cost computation.
Train models on diverse datasets to improve the network's ability to handle variations in lighting, occlusions, and textureless regions.

e. Occlusion Handling:

Develop strategies to handle occlusions effectively. Investigate methods for detecting and managing occluded regions to prevent them from affecting disparity estimation.
Use confidence measures or segmentation techniques to identify unreliable disparities in occluded areas.

f. Adaptive Window Sizes:

Implement adaptive window sizes based on local image characteristics. This helps in focusing computational efforts where detailed disparities are needed and using larger windows in textureless or uniform areas.

Research Plan

Define Your Research Question or Topic

Background Research

Study and Implement Existing Methods

Develop a Research Plan

Implement Your Proposed Methods

Evaluate Performance of Compared Methods

1. Research Title:
2. Introduction
3. Background and Literature Review
4. Research Objectives
5. Research Questions
6. Methodology
7. Timeline

   Study and implement existing methods: 1 week

   Implement the proposed method: 1 week

   Evaluate and analysis the performance of compared methods: 1 week

How to do Research

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title:
2. Introduction
3. Background and Literature Review
4. Research Objectives
5. Research Questions
6. Methodology
7. Timeline
8. Resources
   - Dataset
   - Hardware
   - Software
   - Human resource

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title:
2. Introduction
3. Background and Literature Review
4. Research Objectives
5. Research Questions
6. Methodology
7. Timeline
8. Resources
9. Expected Result

   Improve the performance of existing methods under difficult conditions by 1%.

   Reduce the processing time of the-state-of-the-art method to 30ms/frame

Research Plan

Define Your Research Question or Topic

↓

Background Research

↓

Study and Implement Existing Methods

↓

Develop a Research Plan

↓

Implement Your Proposed Methods

↓

Evaluate Performance of Compared Methods

1. Research Title:
2. Introduction
3. Background and Literature Review
4. Research Objectives
5. Research Questions
6. Methodology
7. Timeline
8. Resources
9. Expected Result
10. Significant and Contribution

    Significance: The proposed stereo matching methodology addresses the limitations of existing techniques by specifically targeting challenging conditions, including occlusions, varying lighting conditions, and textureless regions.

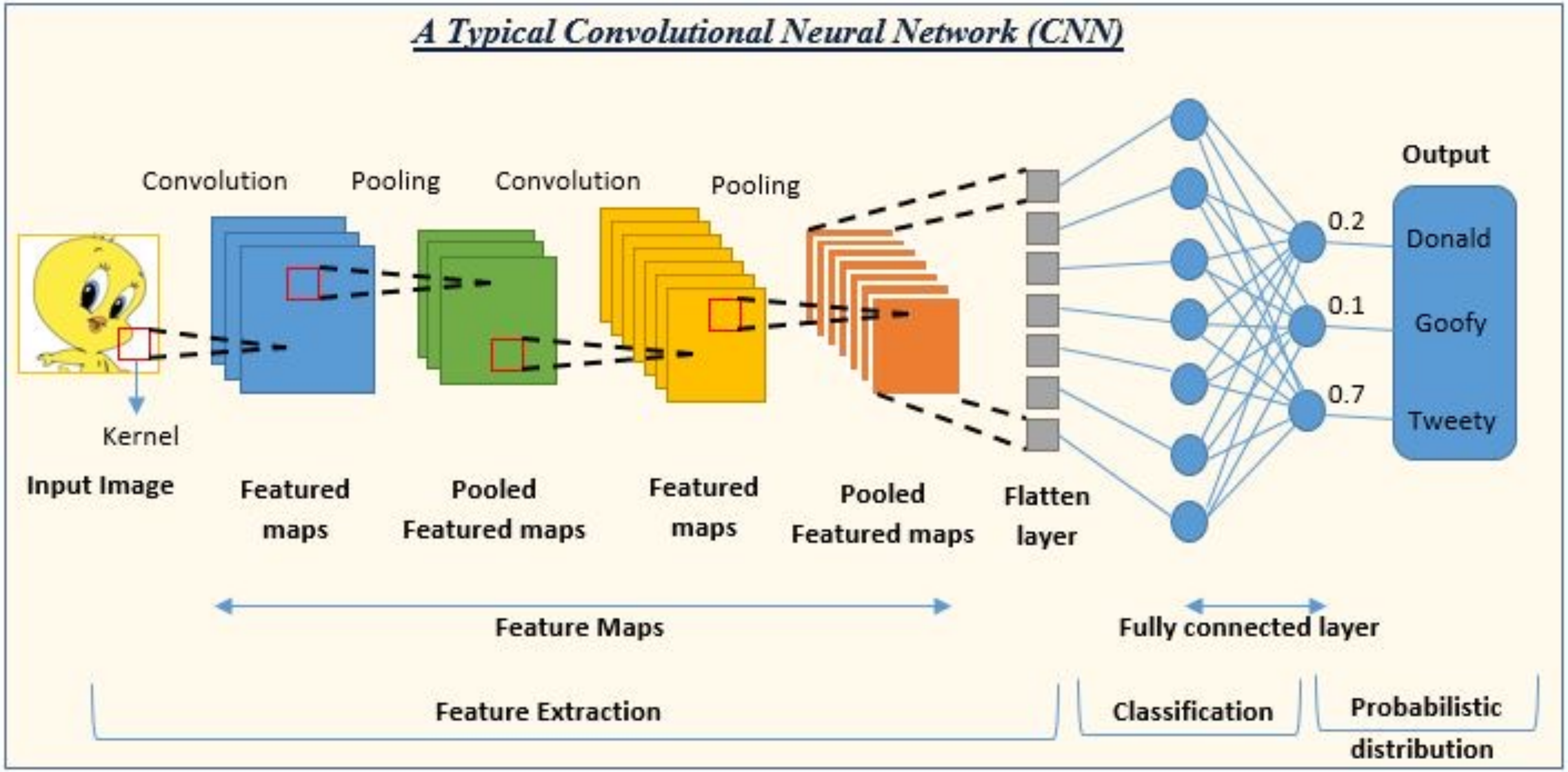
    Contribution: The methodology introduces novel strategies, such as adaptive matching cost functions and machine learning integration, to enhance the accuracy and robustness of disparity estimation under difficult scenarios.

Brainstorming Idea

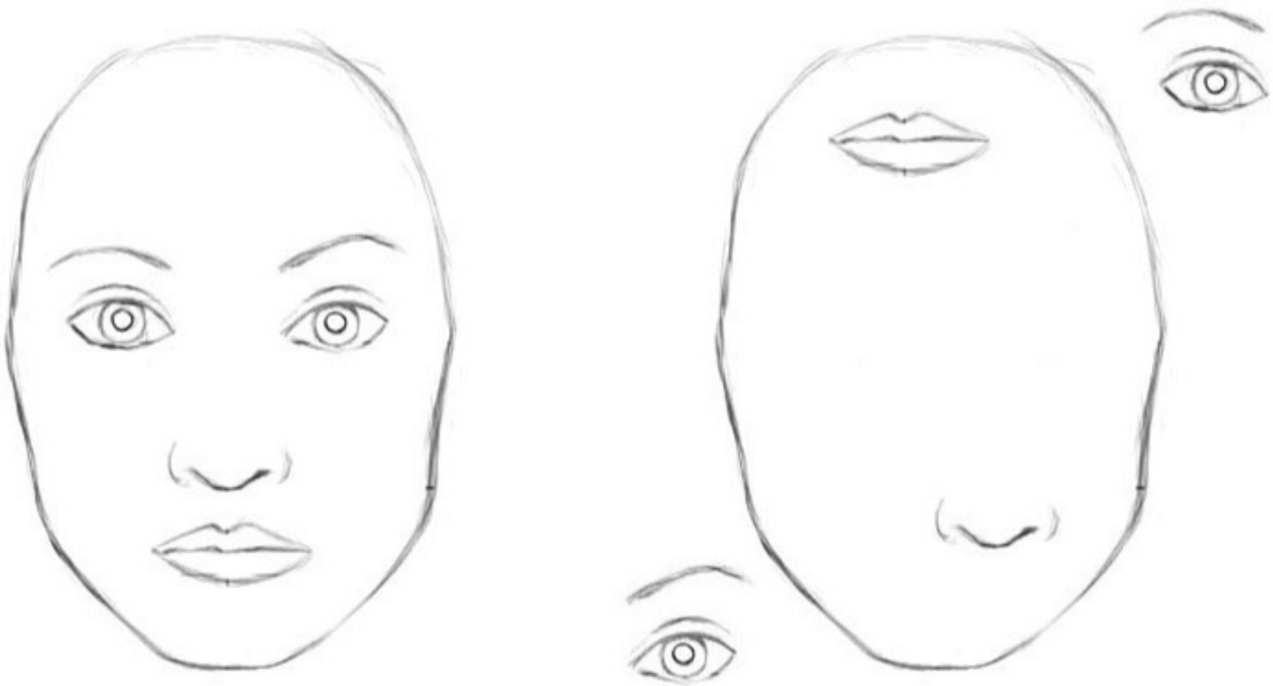
LeNet → AlexNet → ZFNet → VGGNet → InceptionNet → ResNet → MobileNet → ConvNext

CapsuleNet

Brainstorming Idea



CapsuleNet



To a CNN, both pictures are similar, since they both contain similar elements

CapsuleNet

Your brain can easily recognize this is the same object, even though all photos are taken from different angles. CNNs do not have this capability.

The paper that uses this approach was able to **cut error rate by 45%** as compared to the previous state of the art, which is a huge improvement.

Assignment 3

- Requirement:
  - Study background of the research topic
  - Implement existing methods related to your research topics/questions
  - Finish the “Research Plan”
  - Submit to: aivnresearch@gmail.com
- Deadline: 12:00 02/12/2023
- Receiving Feedback: 1 week from TA Research Team of AIVN

How to do Research

(Research Plan: Brainstorming Idea and Implementation)

Vinh Dinh Nguyen
PhD in Computer Science


Thank you