Introduction to Computer Vision

CSCE604133 Computer Vision

Fakultas Ilmu Komputer

Universitas Indonesia

Acknowledgements

- These slides are created with reference to:
  - Computer Vision: Algorithms and Applications, 2nd ed., Richard Szeliski https://szeliski.org/Book/
  - Digital Image Processing, Gonzales and Woods, 3rd ed, 2008.
  - Course slides for CSCE604133 Image Processing – Faculty of Computer Science, Universitas Indonesia
  - Introduction to Computer Vision, Cornell Tech https://www.cs.cornell.edu/courses/cs5670/2024sp/lectures/lectures.html
  - Computer Vision, University of Washington https://courses.cs.washington.edu/courses/cse576/08sp/

Human vision can understand this image effortlessly

Image courtesy of Tomas Anton Escobar, Unsplash.

Recognition

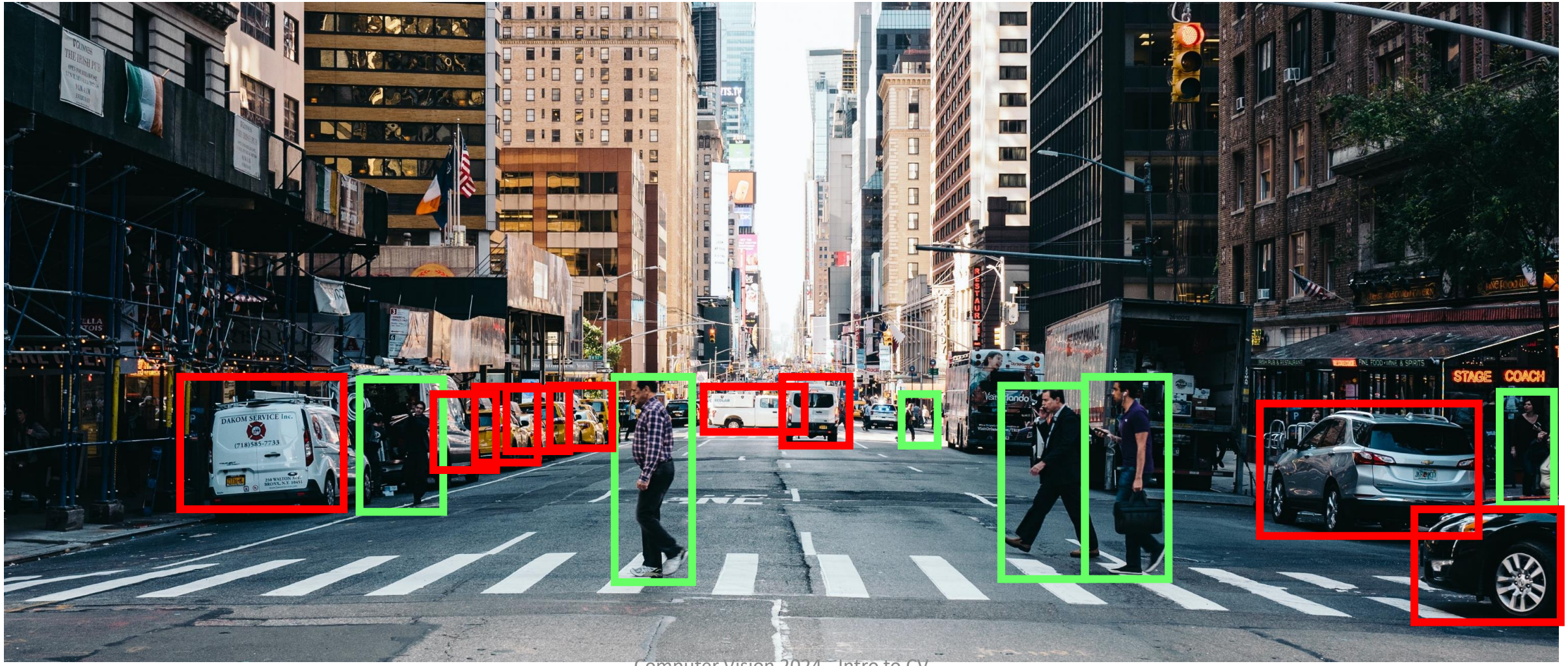


Image courtesy of Tomas Anton Escobar, Unsplash.

Segmentation

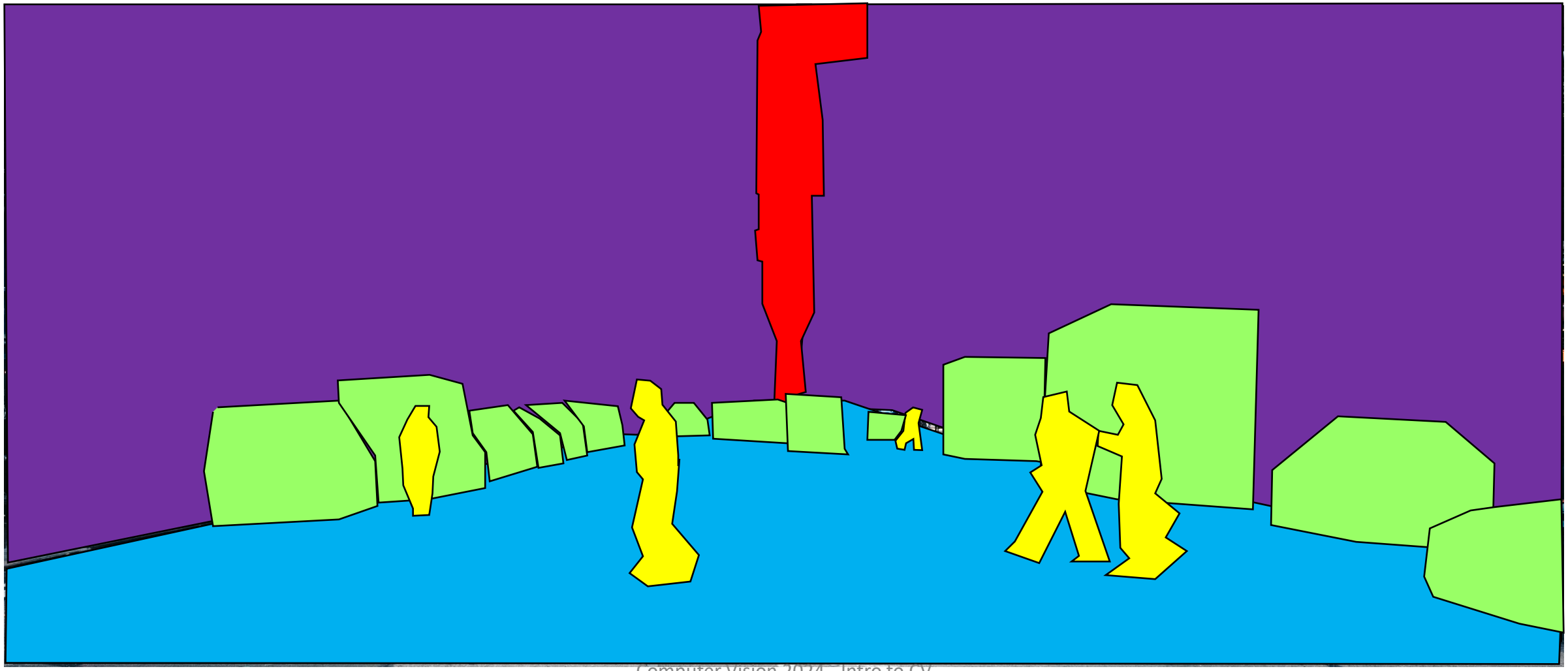


Image courtesy of Tomas Anton Escobar, Unsplash.

Reconstruction

ada proses flatting dari 2d jadiin 3d

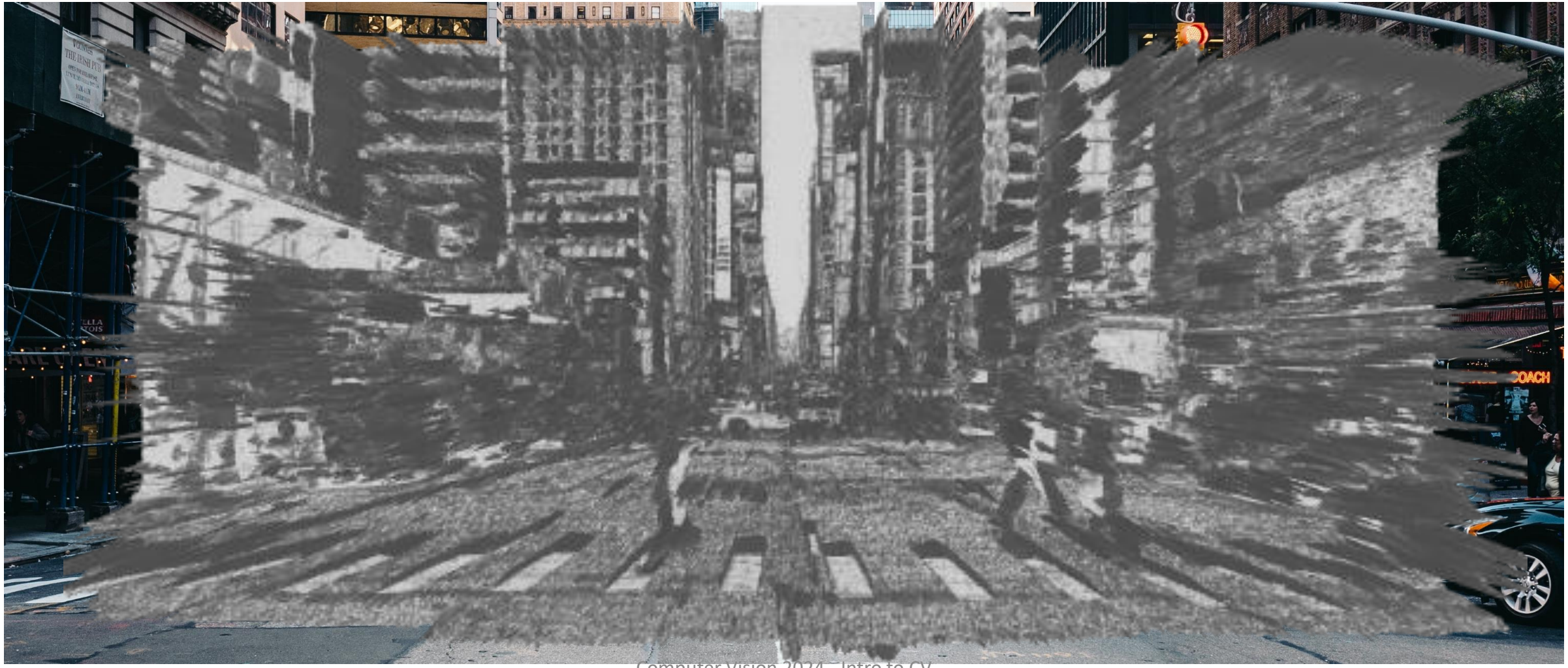
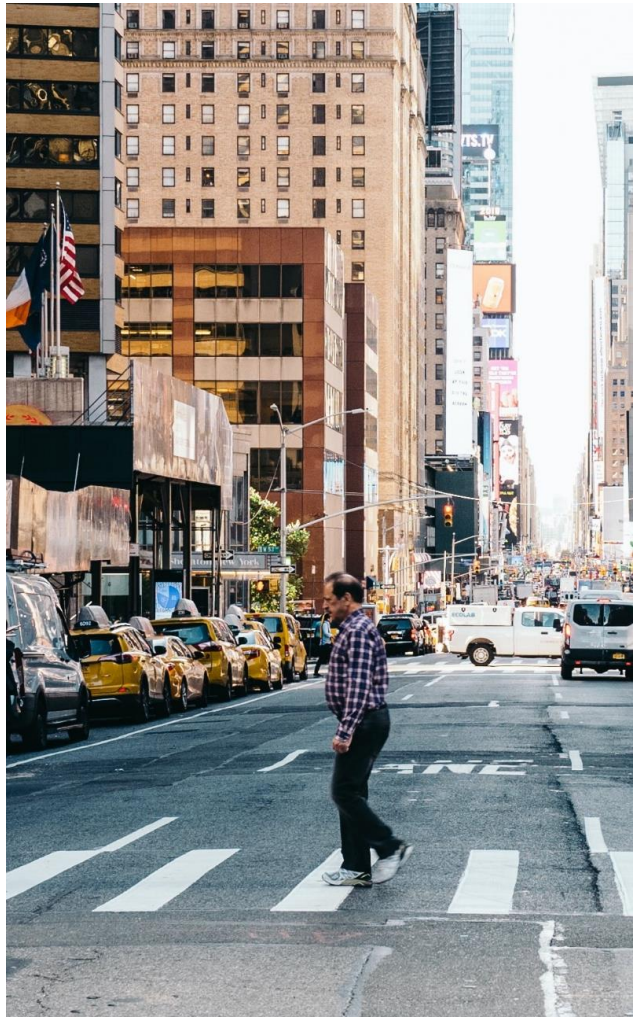
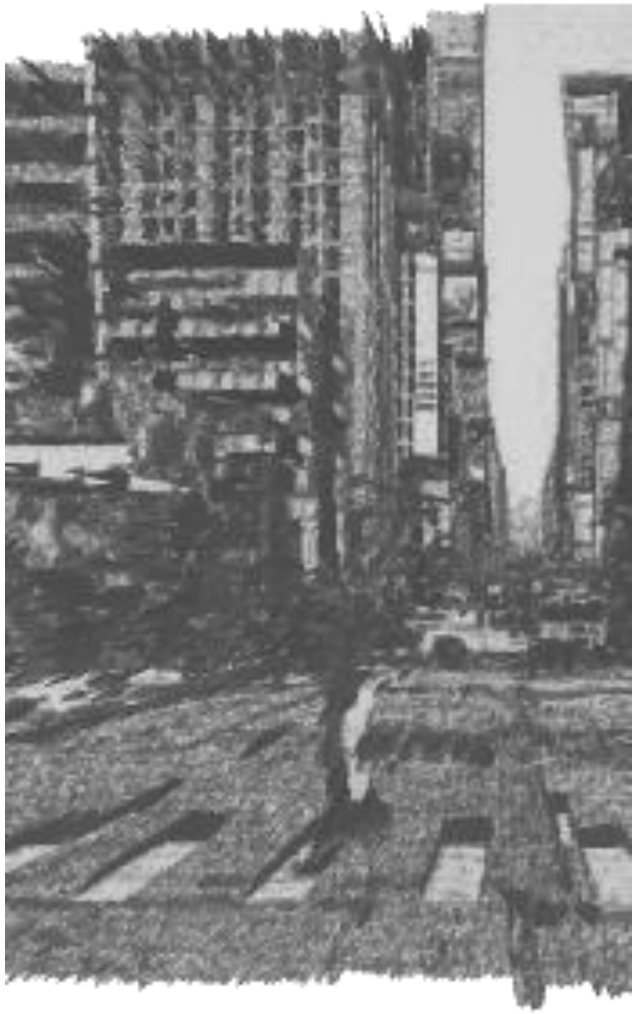


Image courtesy of Tomas Anton Escobar, Unsplash.

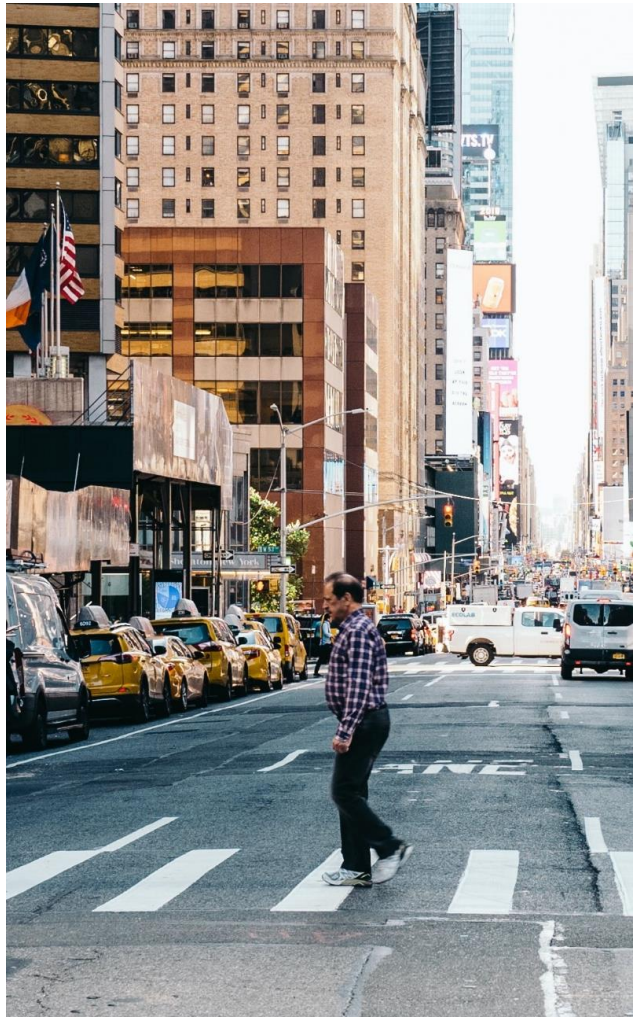


How does human vision do it?
How would a computer do it?



What is Computer Vision?

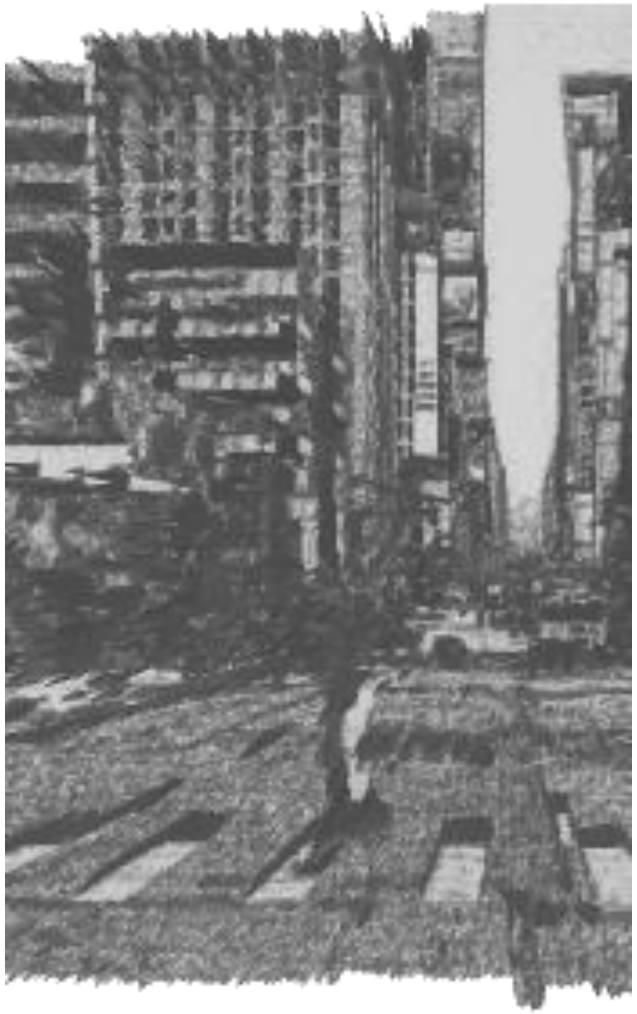
mau bailikkin ke image yang satunya



definition:

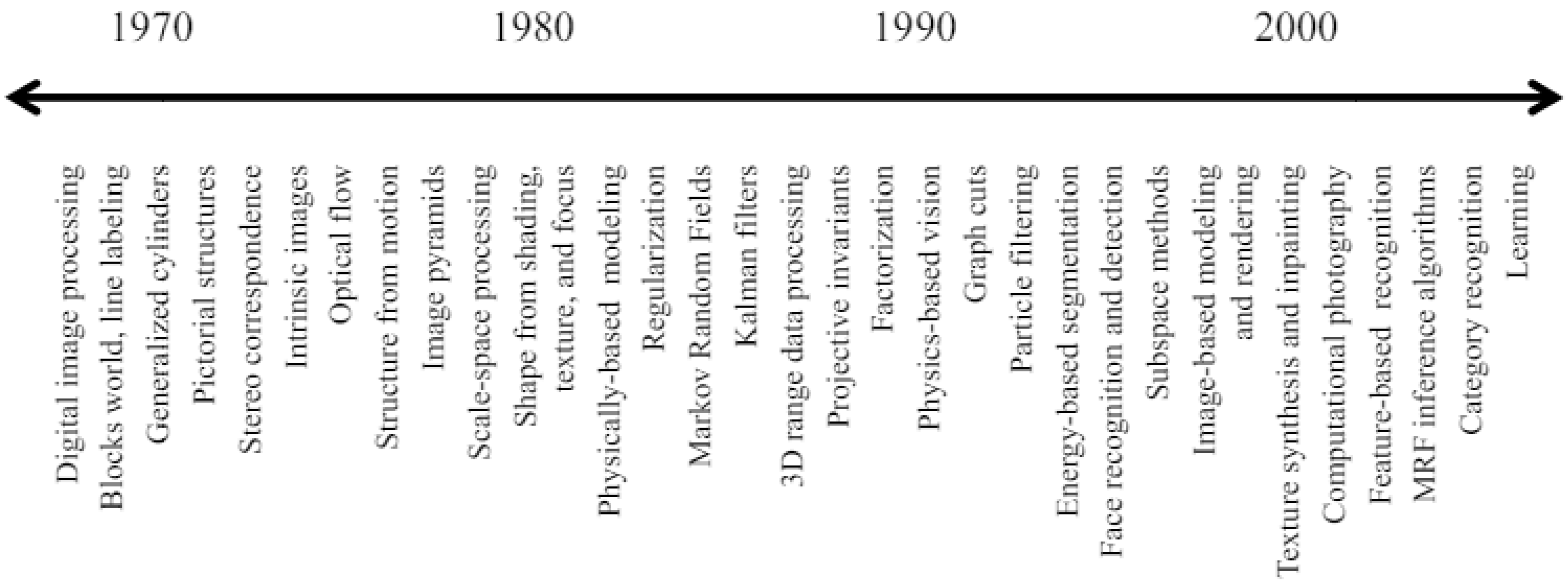
perlu kerjain step A, B, C, D, E, F
(ada 6 kotak makanya ada 6 cara)

mathematical techniques for recovering the three-dimensional shape and appearance of objects in imagery.



A Brief History of Computer Vision

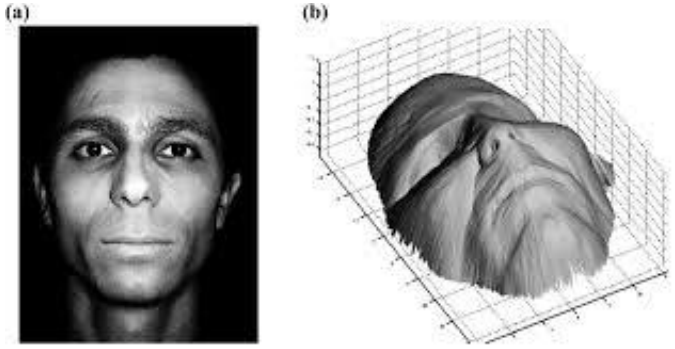
Szeliski, 2010.



A Brief History of Computer Vision (2)

dari 2d jadi 3d

- 1980s



Shape from Shading

(Freeman and Adelson, 1991; Image from Yu and Chen, 2018)

Edge Detection

(Freeman and Adelson, 1991; Image from Yu and Chen, 2018)

- 1990s

Menggunakan rekonstruksi untuk melihat jalan mana yang baik untuk diambil

Dense Stereo Matching

(Boykov, Veksler, and Zabih 2001); Image from Li - github, 2024)

3D Modelling

(Schaffalitzky, F. and Zisserman, A. (2002).)

Computer Vision Applications Now

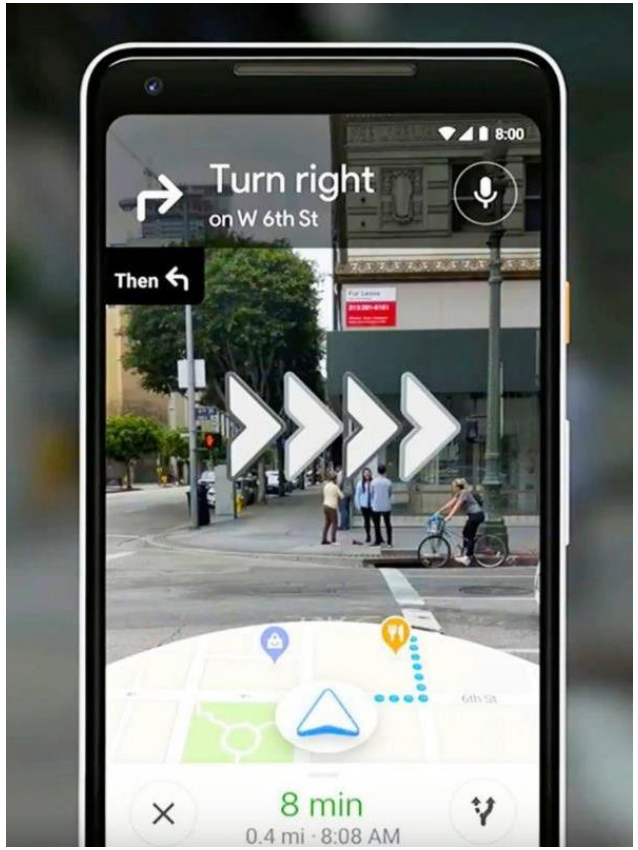
- Image Registration

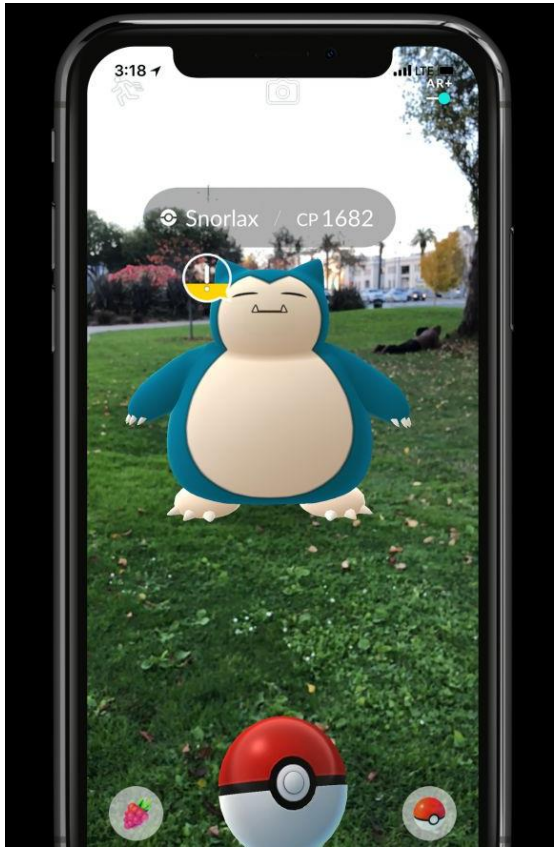
image has to be stitched because it uses camera to join each other

basically getting lots of image, how do we connect it (distinct points)/features.

Computer Vision Applications Now (2)

- Augmented and Virtual Reality





Computer Vision Applications Now (3)

- ETLE (Electronic Traffic Law Enforcement)

bisa deteksi gak pakai seatbelt

platnomor jadi warna putih

NTMC KORLANTAS POLRI

Computer Vision Applications Now (4)

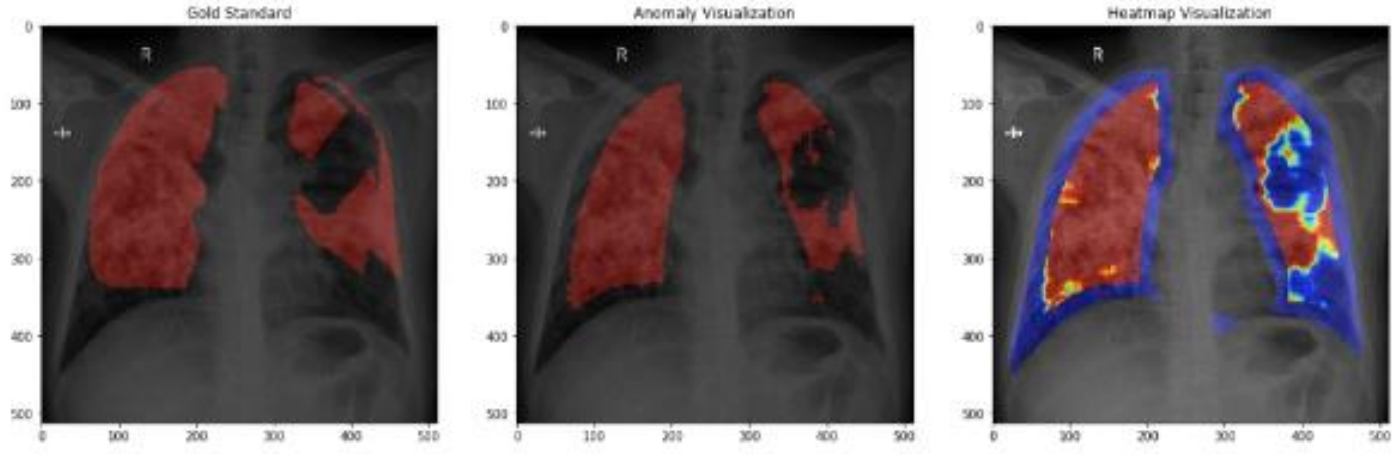
- Autonomous vehicles

lane detector juga pakai compvis

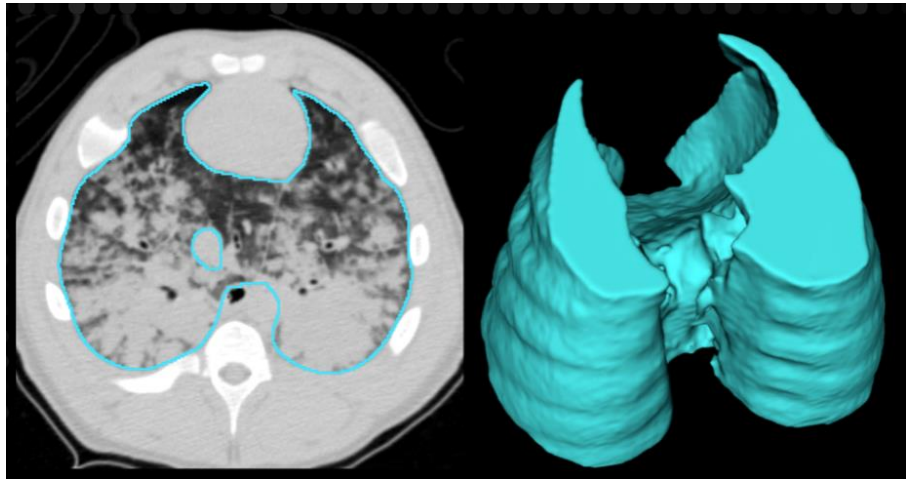
https://waymo.com/waymo-driver/

Computer Vision Applications Now (5)

- Medical Applications



Computer Aided Diagnosis (Nurhayati, et al, 2021)

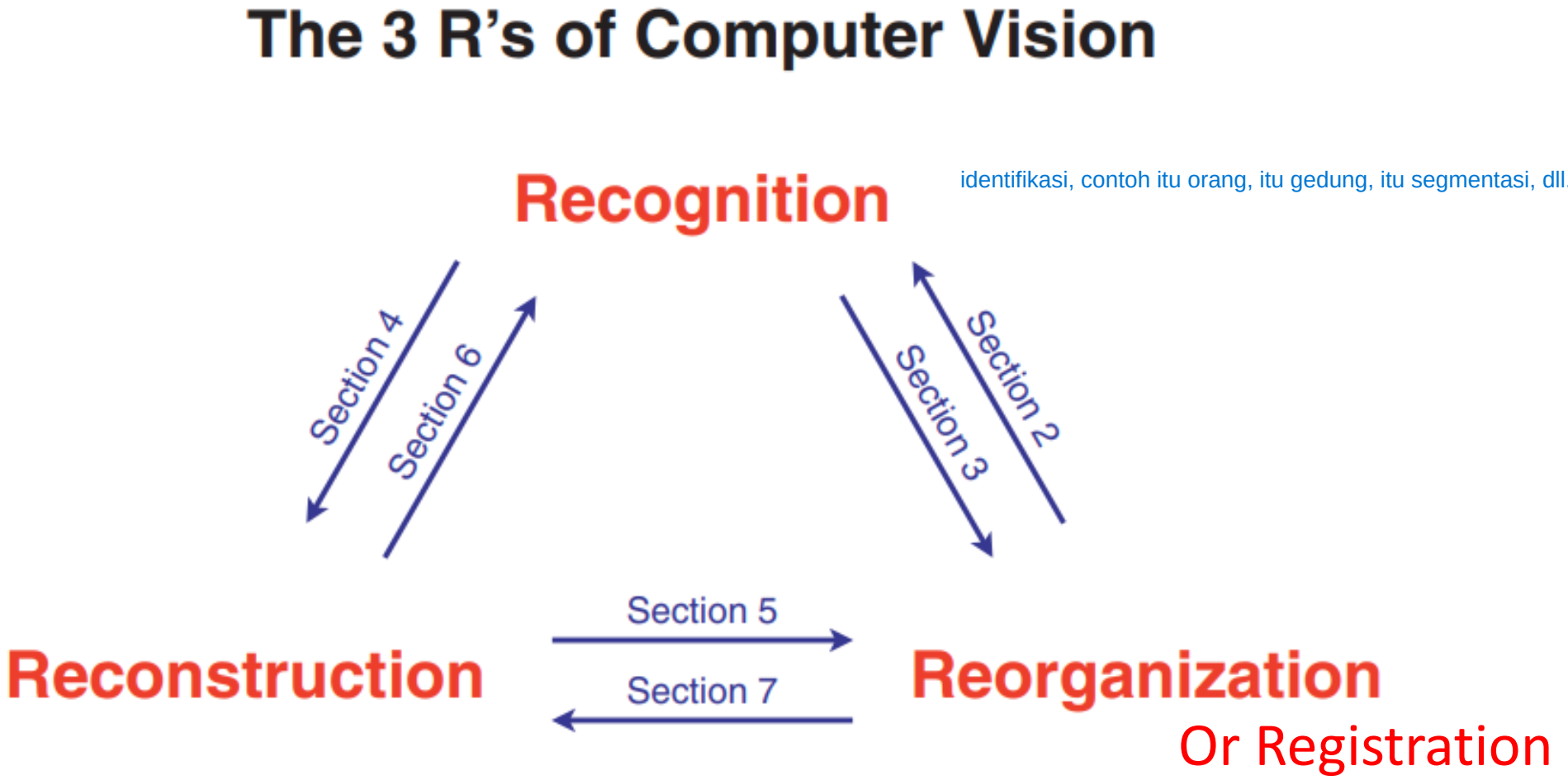


3D visualization (Noshadi, et al, 2017)

The 3 R’s of Computer Vision

Malik, Jitendra, et al. "The three R’s of computer vision: Recognition, reconstruction and reorganization." *Pattern Recognition Letters* 72 (2016): 4-14.

3 hal paling sering: classification, detection dan segmentation



identifikasi, contoh itu orang, itu gedung, itu segmentasi, dll.

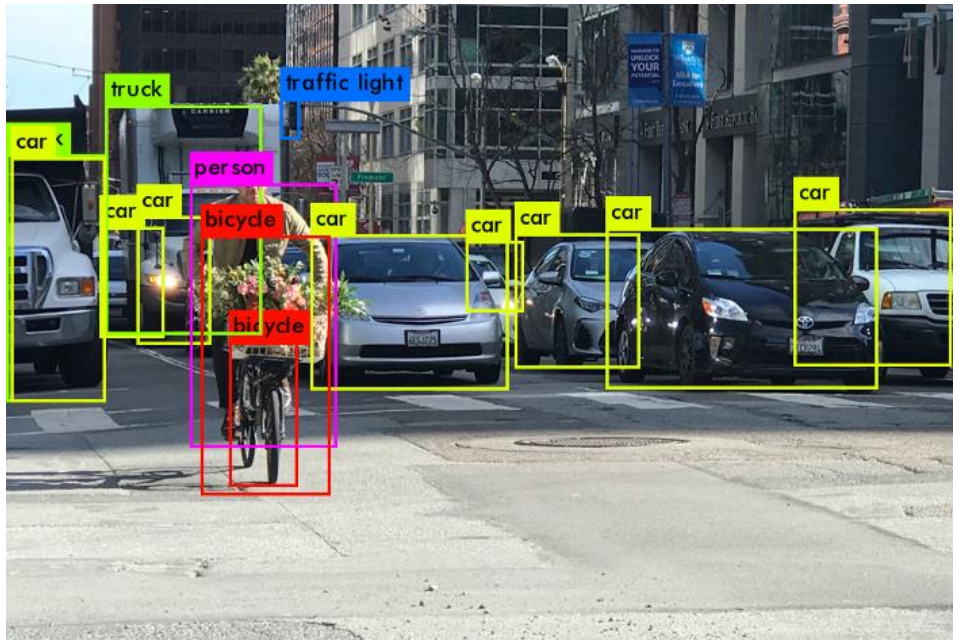
Or Registration

The 3 R’s of Computer Vision: Recognition

Malik, Jitendra, et al. "The three R’s of computer vision: Recognition, reconstruction and reorganization." *Pattern Recognition Letters* 72 (2016): 4-14.

- **Recognition** is about attaching semantic category labels to objects and scenes as well as to events and activities.

buat identifikasi itu apa



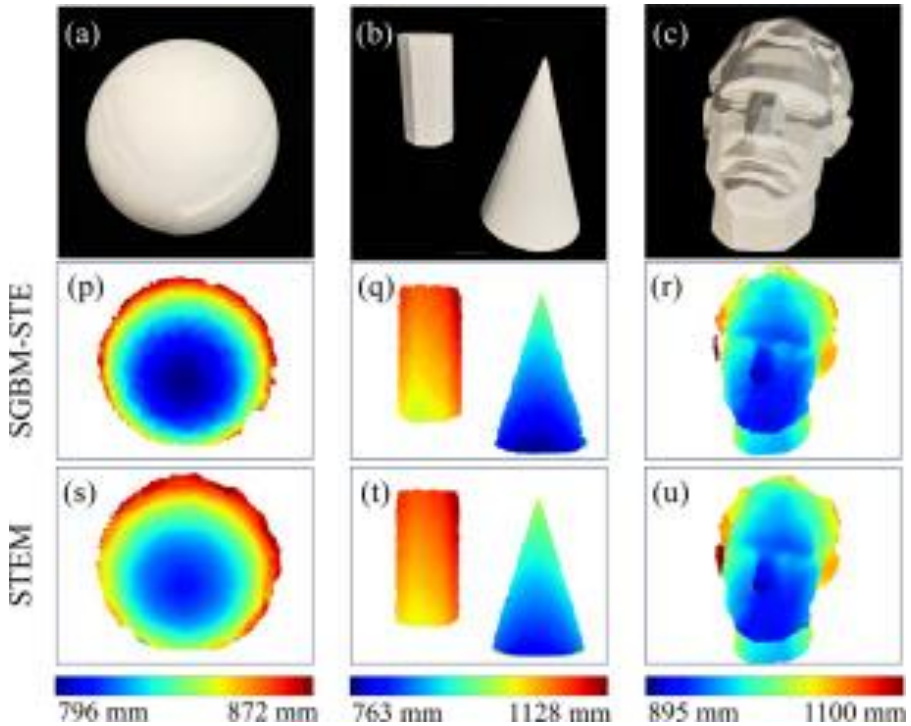
https://medium.com/analytics-vidhya/yolo-explained-5b6f4564f31

Zhao, Hengshuang, et al. "Icnet for real-time semantic segmentation on high-resolution images." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

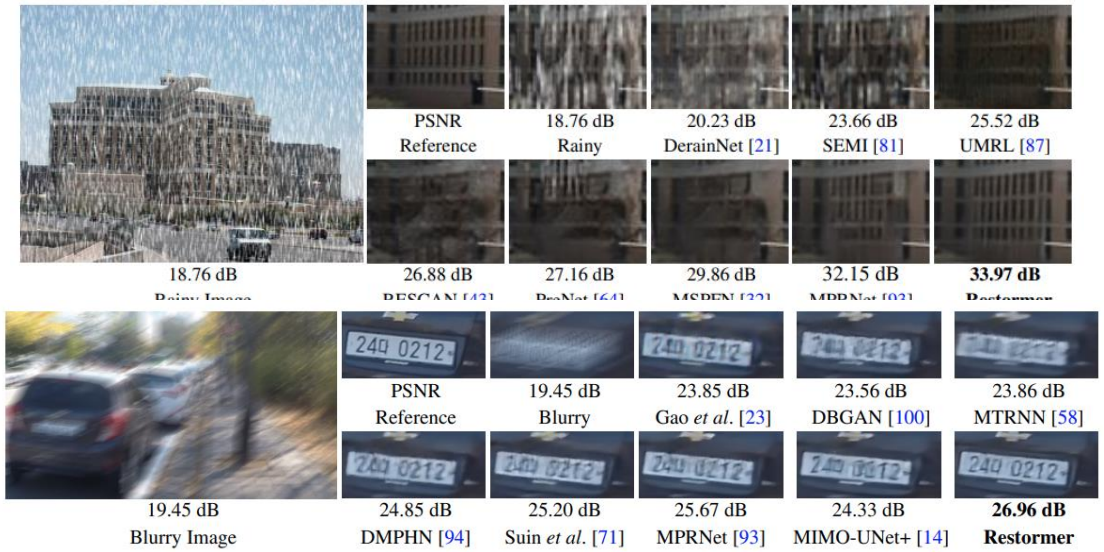
The 3 R's of Computer Vision: Reconstruction

Malik, Jitendra, et al. "The three R's of computer vision: Recognition, reconstruction and reorganization." *Pattern Recognition Letters* 72 (2016): 4-14.

- **Reconstruction** is traditionally about recovering information from images or reconstructing images to a different form



Fu, Jiacheng, et al. "Fast 3D reconstruction via event-based structured light with spatio-temporal coding." *Optics Express* 31.26 (2023).

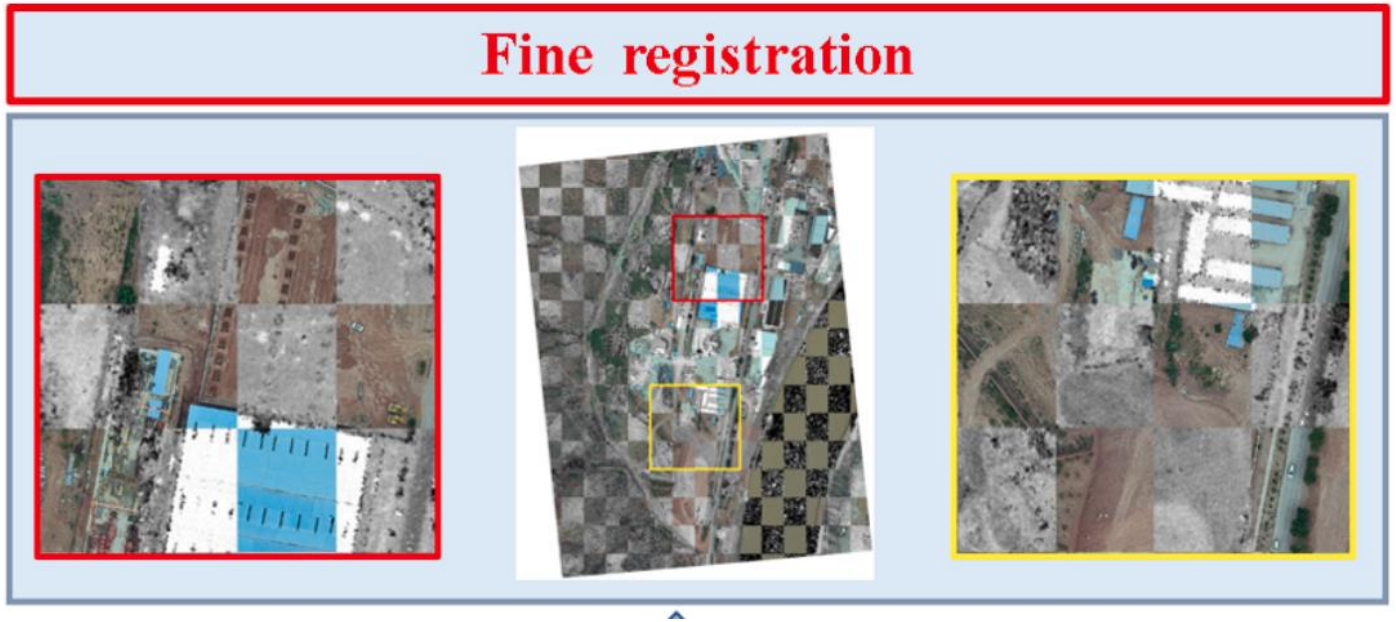


Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

The 3 R's of Computer Vision: Reorganization

Malik, Jitendra, et al. "The three R's of computer vision: Recognition, reconstruction and reorganization." *Pattern Recognition Letters* 72 (2016): 4-14.

- **Reorganization** – or Registration is usually called "perceptual organization" in human vision.



Zhu, Bai, et al. "Robust registration of aerial images and LiDAR data using spatial constraints and Gabor structural features." *ISPRS Journal of Photogrammetry and Remote Sensing* 181 (2021): 129-147.

3D Scene vs 2D Image

forward: dari scene ke gambar

contoh:

Forward Process

- Models the physical process from 3D scenes (movement, light) that is projected onto a 2D image.
- Studied in fields of physics/optics and computer graphics.

dari gambar, dapeting informasi

Inverse Process

- Taking the 2D image, and extracting the information, 3D scene, and other properties.
- This is usually what is called **computer vision**

usually computer vision

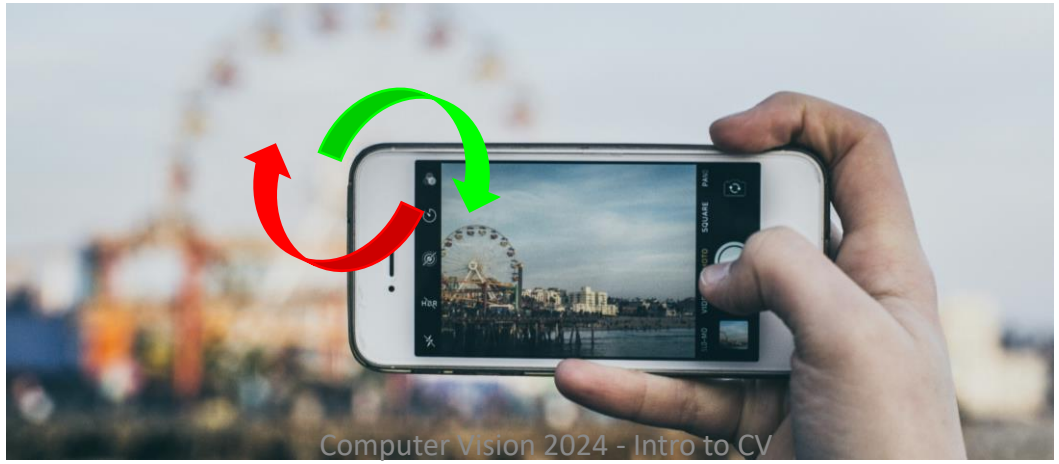
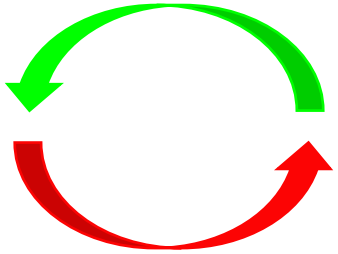


Image courtesy of Alina Fiene, Unsplash.

How does Human Vision do it?

Szeliski, 2010.

Images (2D)  Geometry (3D) + Photometry (*appearance*)

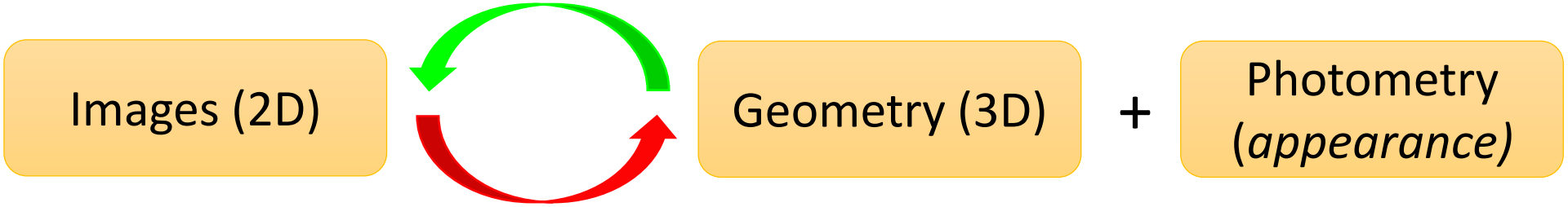
- Human vision can understand a lot from so little…

Torallba, 2008.

Karungaru, 2007.

Why is 2D to 3D difficult for computers?

Szeliski, 2010.

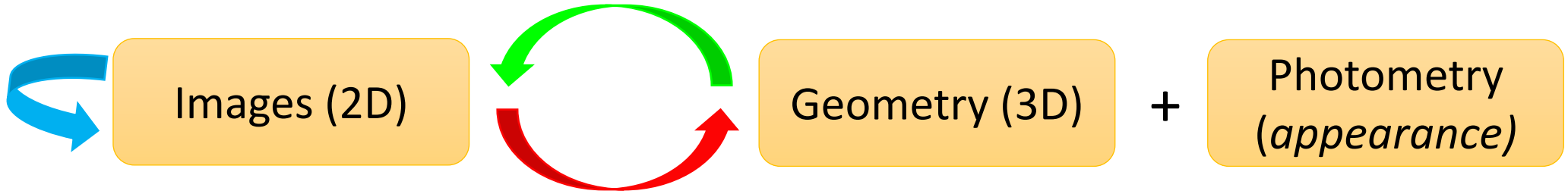


- Human vision does this effortlessly, how can computer vision do the same?
- Human vision is teamwork between **eye** and **brain.**
- Computer vision have cameras as **eyes** – but lack the **brain.**
- Computers can be better at “easy” things – humans are better at “hard” things
- The process from 2D to 3D is severely underdetermined, with so many unknowns

Image courtesy of Alina Fiene, Unsplash.

Computer Vision Topics

Szeliski, 2010.

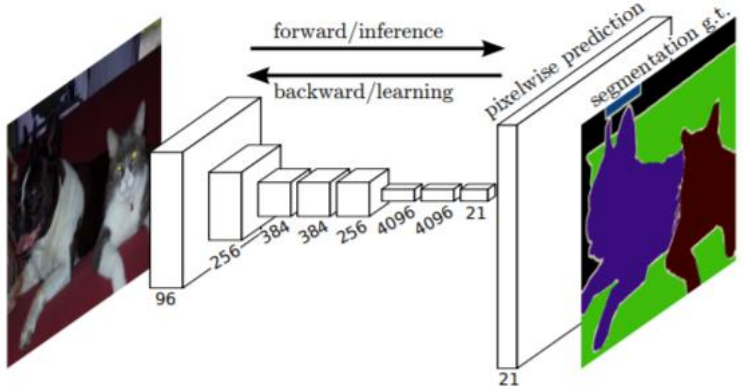


- Image Formation
  - Geometry and photometry
  - Image acquisition
  - Image anatomy
- Vision
  - Human Vision
  - Computer Vision
- Image Processing
  - Image enhancement
  - Image transform

Machine Learning for Computer Vision

Images (2D) → Geometry (3D) + Photometry (*appearance*)

- In recent years, computational capacity has enabled computer vision research to develop solutions using machine learning models



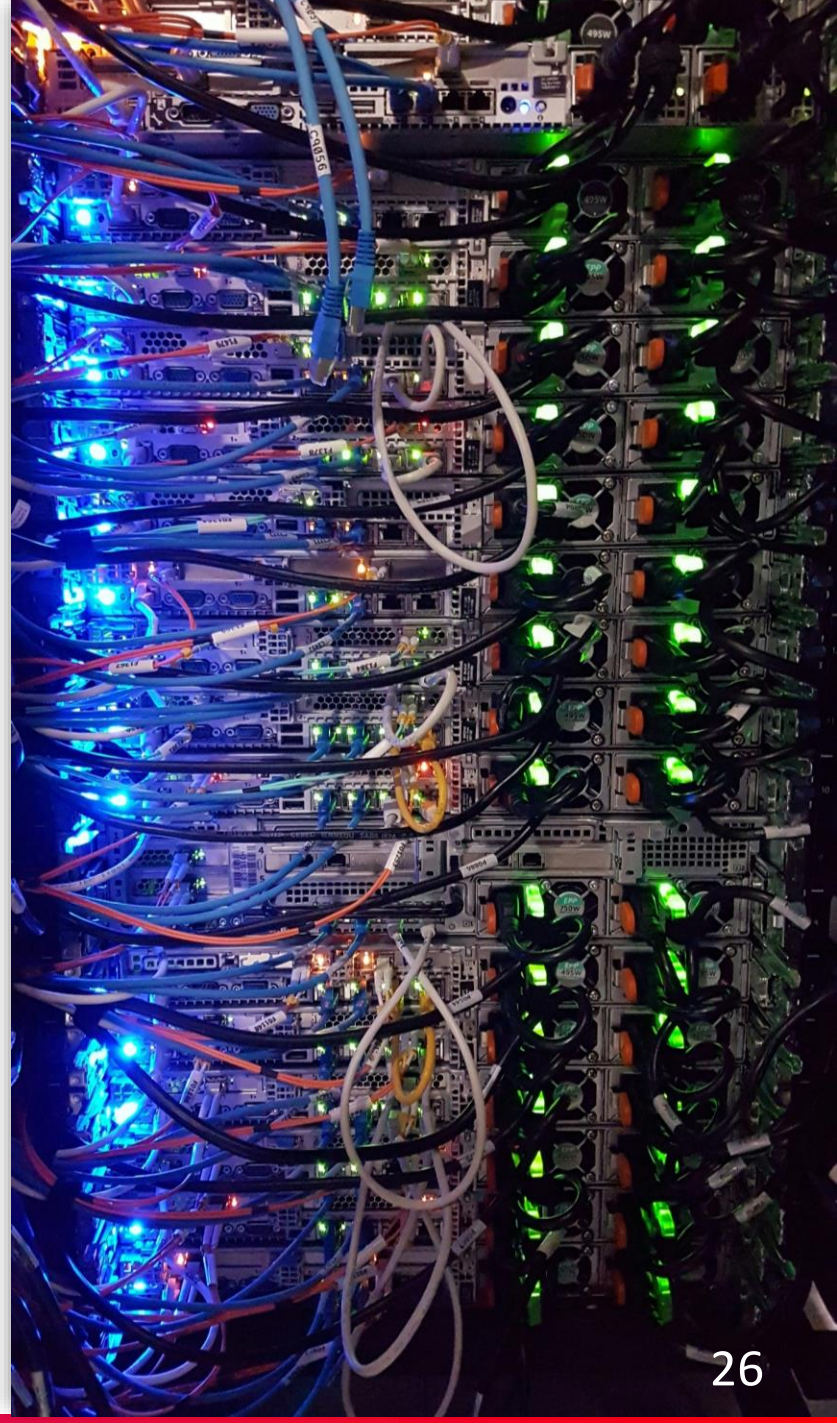
Long, et al, 2015.

Drivers and Challenges in Computer Vision

- Drivers
  - Growing demand for computer vision solutions
  - Increasing need for automation
  - Increasing demand for vision-guided robotic systems
  - Growing adoption of Industry 4.0
  - Growing demand for AI in computer vision
  - Increased manufacturing of autonomous vehicles
- Challenges
  - High costs
  - Lack of experienced professionals
  - Size of required data sets
  - Need for regular monitoring

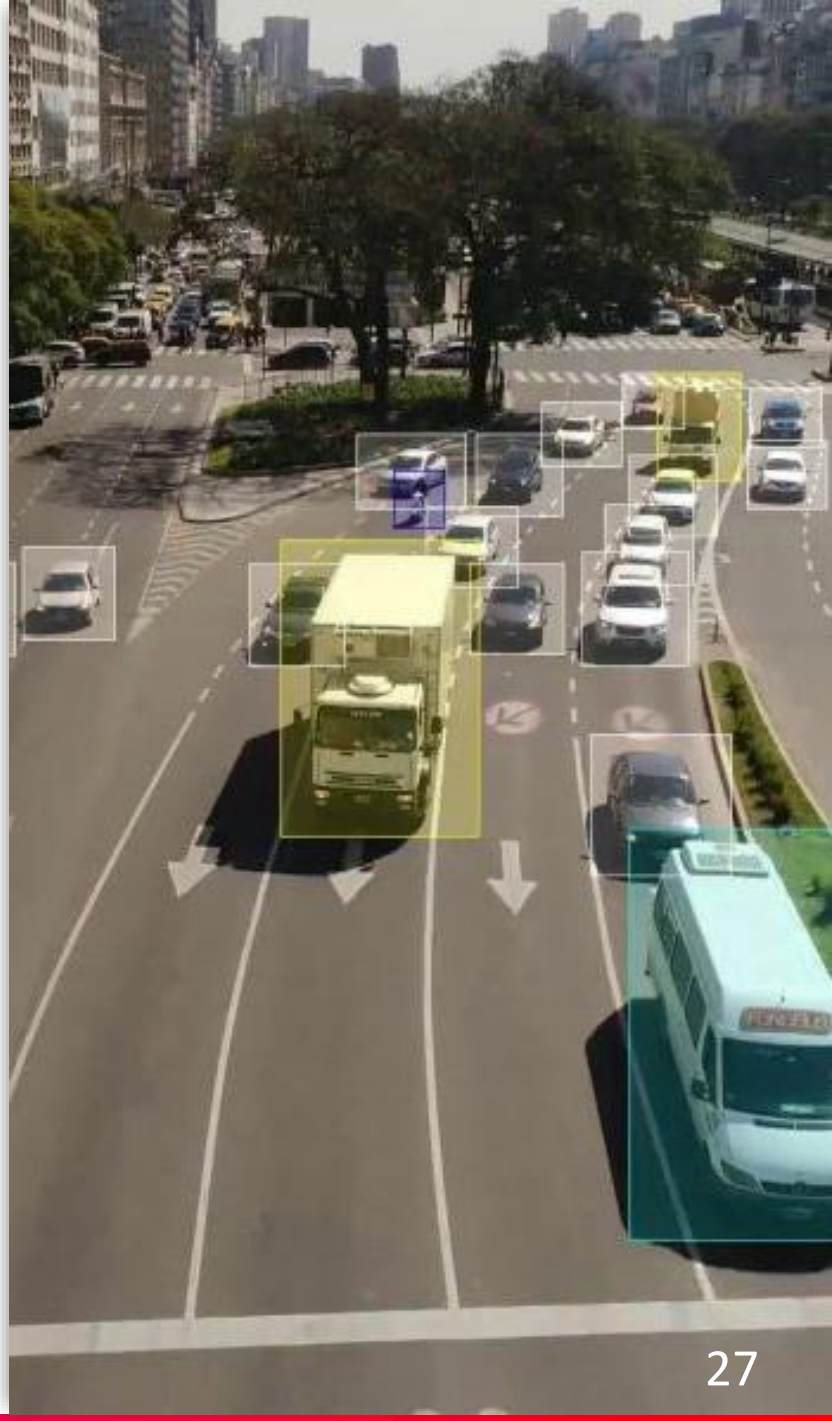
Computer Vision in Industry

- Processing power for real-time data-intensive applications
  - running computer vision in the cloud is heavily limiting real-time applications
  - computer vision solutions will need to be deployed on edge endpoints for most use cases
- Hardware Limitations
  - require hardware to run, cameras to provide the visual input, and computing hardware for AI inference.
- Scaling Problems
  - what works here may not work in bigger scales
  - too many variations of scenarios that can not be considered



Computer Vision in Real Life

- Highest accuracy is not always the goal
- Business driven objectives
- Data is not always available
- Data design and data preparation is limited
- Data **annotationss aree a nightmareeeeeeee**



Case Study: Count the motorcycles on Jalan Margonda

- What type of camera should I use?
- What about lighting and other mitigation factors?
- After obtaining the image, is it good enough?
- What do I need to find? How do I find them?
- Which ones are motorcycles?

Case Study: Count the motorcycles on Jalan Margonda

- Image Acquisition — cara dapetin gambarnya
- Image Enhancement — cara memperbagus gambarnya
- Image Segmentation — cara identifikasi gambarnya/identifikasi semua objek yang bukan background
- Feature Detection — fitur-fitur yang dimiliki objek yang mau di observasi, contohnya fitur yang dimiliki motor
- Recognition — after feature detection, recognizing which one is suited to the feature.