ET3112 Image Processing and Machine Vision: Introduction

Ranga Rodrigo
ranga@uom.lk

The University of Moratuwa, Sri Lanka

January 12, 2024

Section 1

Introduction to the Course

What Is Image Processing?

- In digital image processing, we manipulate a digital image to produce another digital image, which is an enhanced version of the original image.
- E.g., blurring, noise filtering, color enhancement, segmentation (?).

(a) Noisy image

(b) Gaussian filtered image

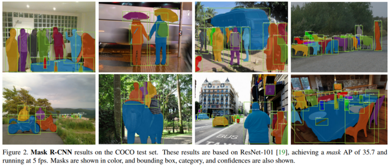
(c) Image with salt and pepper noise

(d) Median filtered image

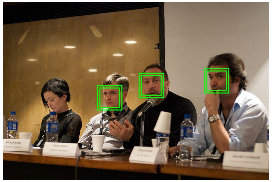
Figure: Filtering

What is computer vision?

- In computer vision, we analyze a digital images or videos to make a decision.
- E.g., face detection, object detection, semantic segmentation.



(a) Mask RCNN Object detection and semantic segmentation

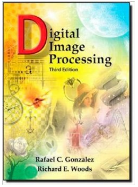


(b) Face detection (from Wikipedia)

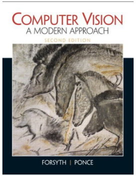
Figure: Examples of computer vision.

Text Books

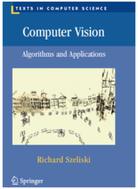
1. Gonzalez and Woods, Digital Image Processing
2. Forsyth and Ponce, Computer Vision: A Modern Approach
3. Richard Szeliski, Computer Vision: Algorithms and Applications (available online)
4. Milan Sonka, Image Processing, Analysis, and Machine Vision



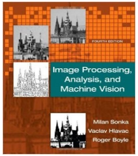
(a)



(b)



(c)



(d)

Figure

Course Requirements

1. Programming assignments: 20%
2. Four assignments
   - Expect the first one in the next couple of classes.
   - We use OpenCV and Python. We will do a tutorial on Python and NumPy.
   - Other platforms, e.g., MATLAB or OpenCV withe C++, are fine too.
3. Grading

Item	Date	Weight	Minimum	Comments
In-class quizzes	Everyday	10%	50%	Easy
Assignment 1		5%	50%	Easy
Assignment 2		5%	50%	Easy
Assignment 3		10%	50%	Moderately difficult
Final examination		70%	50%	?

Tips for Successful Completion

To extract meaning from pixels.

1. Mindset: “I want to solve this vision problem. What are the tools available? How have others solved it? How should I go about solving it?”
2. Attend every single lecture.
3. Go through the material the night before and engage in a good discussion in class.
4. Implement at least one algorithm discussed in class before the next class. This is in addition to the assignments.

Academic Integrity Policy

- You may discuss the assignments with each other. However, you must code on your own, and make the submissions on your own.
- You may benefit from the code and information on the internet, as long as you do not borrow code for the main theme of the assignment.
- Acknowledge your sources.

Source: S. Lazebnik

The Goal of Computer Vision

(a) What we see.

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

(b) What computer sees.

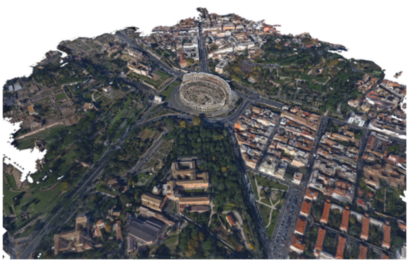
Source: S. Narasimhan

Section 2

Sate-of-the-Art in Vision

Reconstruction: 3D from Photo Collections

Colosseum, Rome, Italy



San Marco Square, Venice, Italy

Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. Seitz, The Visual Turing Test for Scene Reconstruction, 3DV 2013. `https://www.youtube.com/watch?v=NdeD4cjLI0c`

Source: S. Lazebnik

Reconstruction: 4D from Photo Collections

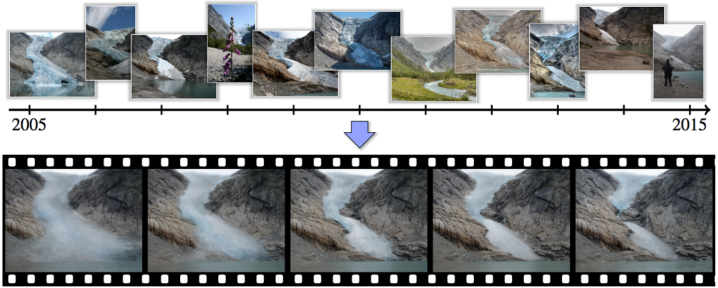


Figure 1: *We mine Internet photo collections to generate time-lapse videos of locations all over the world. Our time-lapses visualize a multitude of changes, like the retreat of the Briksdalsbreen Glacier in Norway shown above. The continuous time-lapse (bottom) is computed from hundreds of Internet photos (samples on top).* *Photo credits:* `Aliento Más Allá`, `jirihnidek`, `mcxurxo`, `elka_cz`, *Juan Jesús Orío, Klaus Wißkirchen,* `Daikrieg`, `Free the image`, `dration` *and Nadav Tobias.*

R. Martin-Brualla, D. Gallup, and S. Seitz, Time-Lapse Mining from Internet Photos, SIGGRAPH 2015. `https://www.youtube.com/watch?v=wptzVm0tngc&feature=youtu.be`

Source: S. Lazebnik

Reconstruction: 4D from Depth Cameras

Figure 1: Real-time reconstructions of a moving scene with DynamicFusion; both the person and the camera are moving. The initially noisy and incomplete model is progressively denoised and completed over time (left to right).

R. Newcombe, D. Fox, and S. Seitz, DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time, CVPR 2015.

`https://www.youtube.com/watch?v=i1eZekcc_lM&feature=youtu.bee`

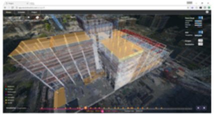
Source: S. Lazebnik

Reconstruction in Construction Industry

RECONSTRUCT INTEGRATES REALITY AND PLAN

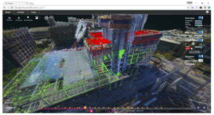
Visual Asset Management

Reconstruct 4D point clouds and organize images and videos from smartphones, time-lapse cameras, and drones around the project schedule. View, annotate, and share anywhere with a web interface.



4D Visual Production Models

Integrate 4D point clouds with 4D BIM, review "who does what work at what location" on a daily basis and improve coordination and communication among project teams.



Predictive Visual Data Analytics

Analyze actual progress deviations by comparing Reality and Plan and predict risk with respect to the execution of the look-ahead schedule for each project location, to offer your project team with an opportunity to tap off potential delays before they surface on your jobsite.

`https://www.reconstructinc.com/#product-demo`

Source: D. Hoiem

Recognition: “Simple Patterns”

(a)

(b)

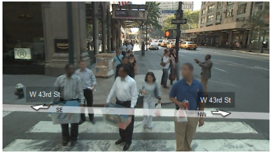
(c)

(d)

Recognition: Faces

(a)

(b)



(c)

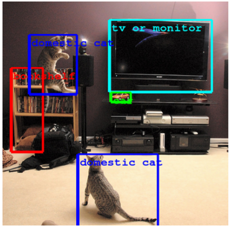
Figure

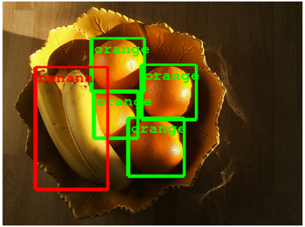
Concerns about Face Recognition

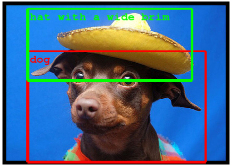
Beijing bets on facial recognition in a big drive for total surveillance — Washington Post, 1/8/2018

Source: Lazebnik

Recognition: General Categories



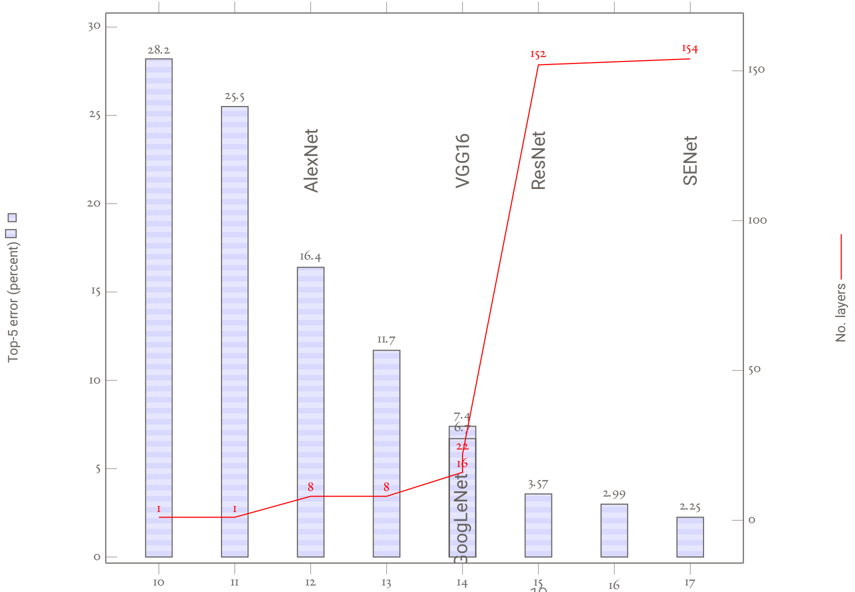




Computer Eyesight Gets a Lot More Accurate, NY Times Bits blog, August 18, 2014
Building A Deeper Understanding of Images, Google Research Blog, September 5, 2014

Source: Lazebnik

ImageNet Challenge



Bengio, Hinton and LeCun Win the Turing Award

Fathers of the Deep Learning Revolution Receive ACM A. M. Turing Award

Recognition: Instance Segmentation

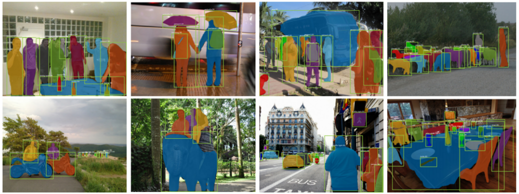


Figure 2. **Mask R-CNN** results on the COCO test set. These results are based on ResNet-101 [19], achieving a *mask* AP of 35.7 and running at 5 fps. Masks are shown in color, and bounding box, category, and confidences are also shown.

(a)

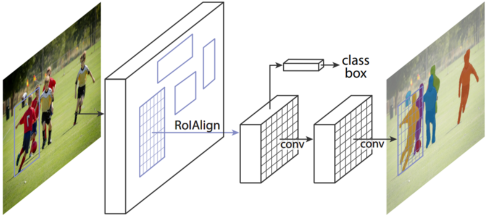


Figure 1. The **Mask R-CNN** framework for instance segmentation.

(b)

Figure

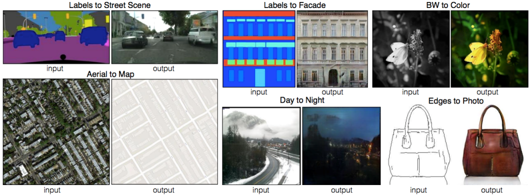
K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in IEEE International Conference on Computer Vision, Venice, Italy, 2017, pp. 2980–2988.

Image Generation

T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, ICLR 2018.

Source: Lazebnik

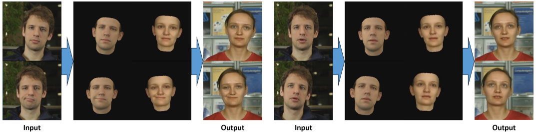
Image Generation



P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017.

Source: Lazebnik

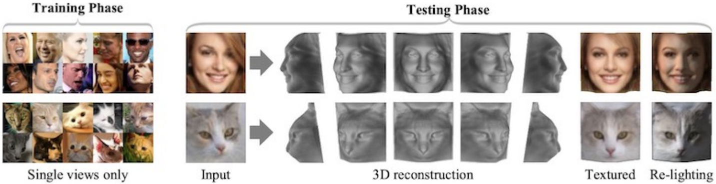
Deep Video Portraits (DeepFakes)



Kim, H. and Garrido, P. and Tewari, A. and Xu, W. and Thies, J. and Nießner, N. and Pérez, P. and Richardt, C. and Zollhöfer, M. and Theobalt, C. Deep Video Portraits, Siggraph 2018.

Source: Lazebnik

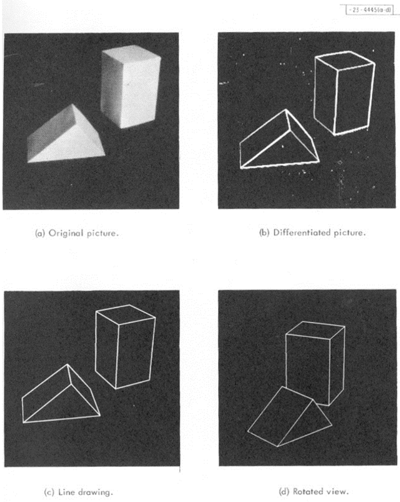
SinGAN: Learning a Generative Model From a Single Natural Image[1]



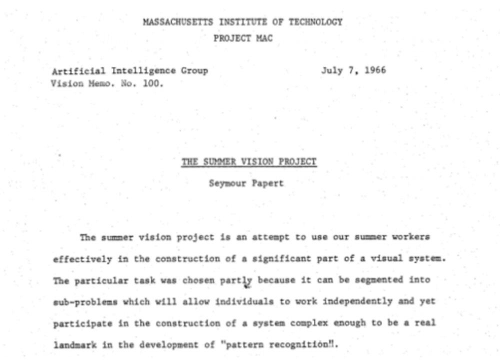
Unsupervised Learning of Probably Symmetric Deformable 3D Objects From Images in the Wild[2]

12

Origins of Computer Vision



(a) Machine perception of three-dimensional solids by Lawrence G. Roberts, MIT 1937



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

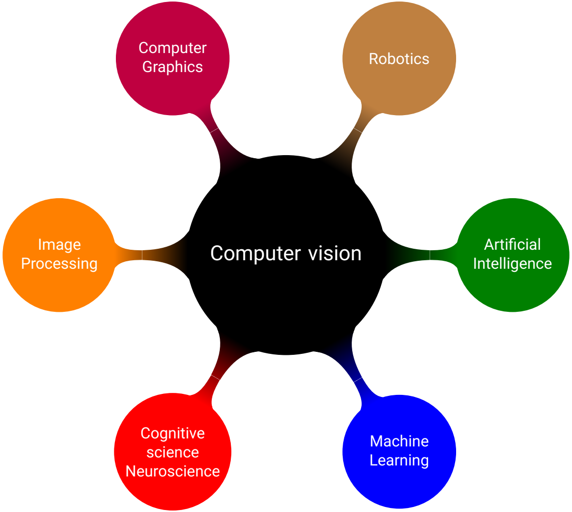
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

(b) The Summer Vision Project at MIT, 1966.

Connections to Other Disciplines



Source: Lazebnik

Module Outline

1. Early vision
   - Point operations, liner filtering, and edge detection
   - Cameras, light, and color
   - Feature extraction
   - Optical flow
   - Morphological processing
   - Frequency domain processing.
2. Mid-level vision
   - Fitting: least squares, total squares, RANSAC, Hough lines
   - Alignment
   - Image stitching

3. Multiple-View Geometry
   - Cameras
   - Epipolar geometry
   - Structure from motion

5. Recognition
   - Basic classification
   - Object detection
   - Deep learning in classification and detection

4. Segmentation

1. Thresholds
2. Region growing
3. Deep learning based segmentation