Computer Vision Systems Programming VO

Introduction to Computer Vision

Christopher Pramerdorfer

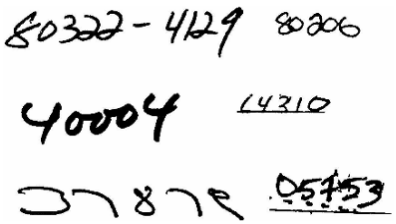
Computer Vision Lab, Vienna University of Technology

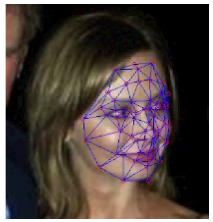
Topics

What is Computer Vision (CV) and why is it important?

CV past, present, future

Relation to other research fields





Images from LeCun et al. 1989, Shotton et al. 2011, Taigman et al. 2013

What Is CV and Why Is It Important?

Let's hear what Fei-Fei Li has to say

Image from ted.com

What Is CV and Why Is It Important?

CV is about making computers understand images like humans do

Key to novel autonomous systems (cars, security, data analysis)

Tremendous progress in last decades, but still unsolved

CV Past, Present, Future

CV research started around 50 years ago

Let's take a look at a few examples

CV Past, Present, Future

1963: Pose Estimation

Edge-based pose estimation of polyhedra

Among first CV applications

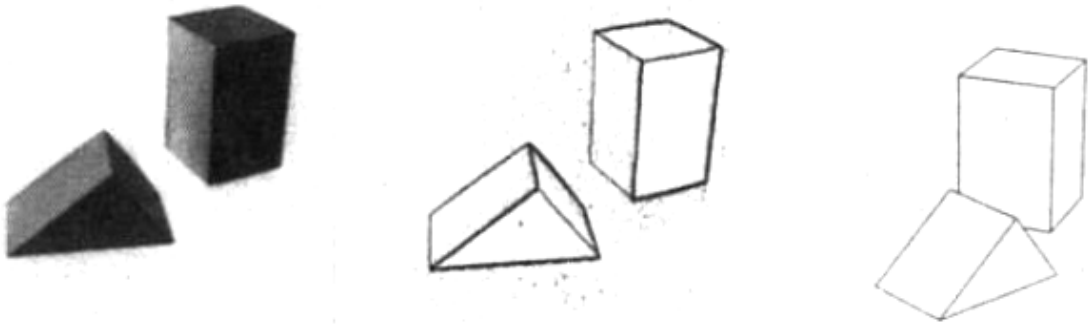


Image from Roberts 1963

CV Past, Present, Future
1973: Part-Based Object Detection

Object representation as parts connected by springs

Known as pictorial structures or constellation models

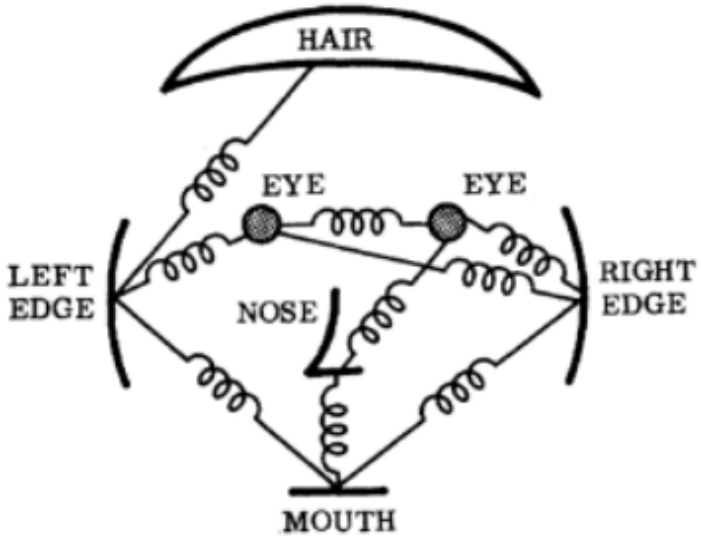


Image from Fischler and Elschlager 1973

CV Past, Present, Future

1989: OCR Using Convolutional Neural Networks

Zip code recognition from images

Among first applications using convolutional neural networks

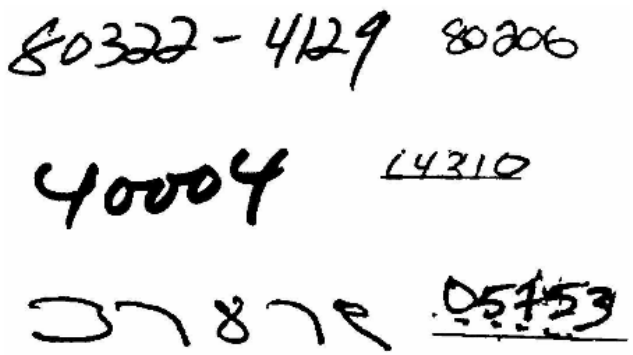


Image from LeCun et al. 1989

CV Past, Present, Future

1989: OCR Using Convolutional Neural Networks

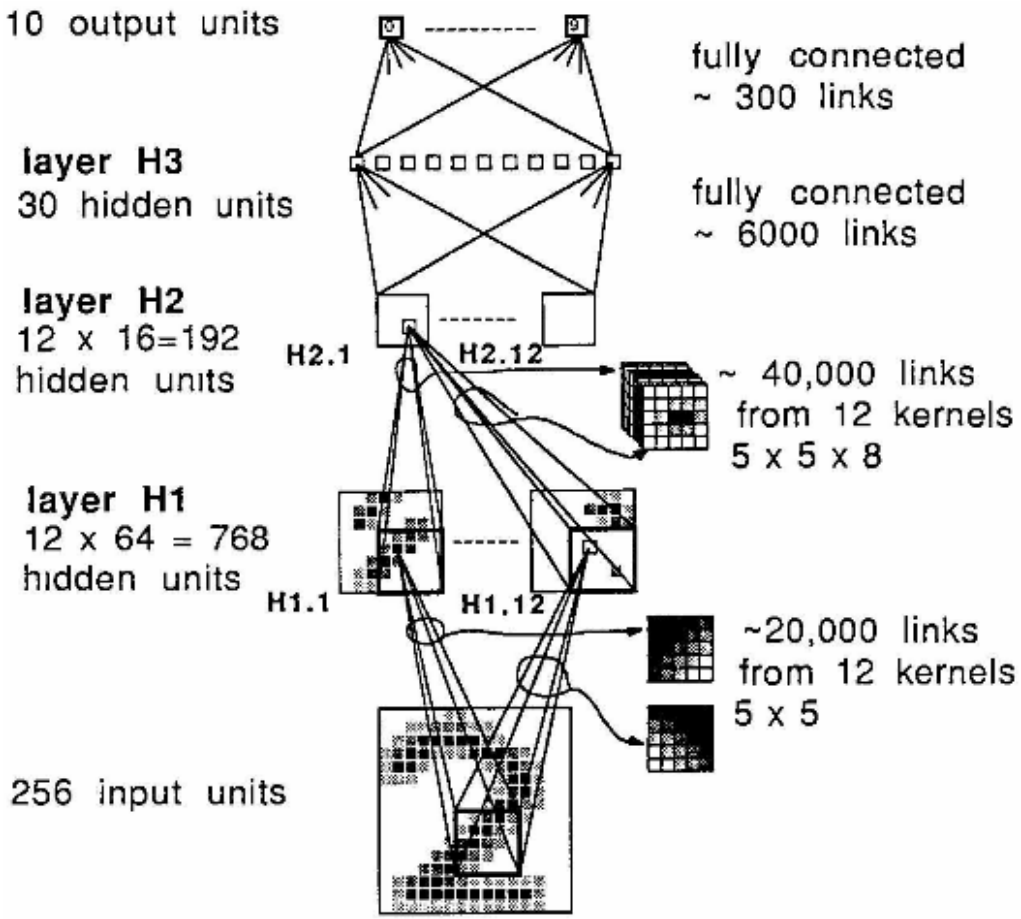


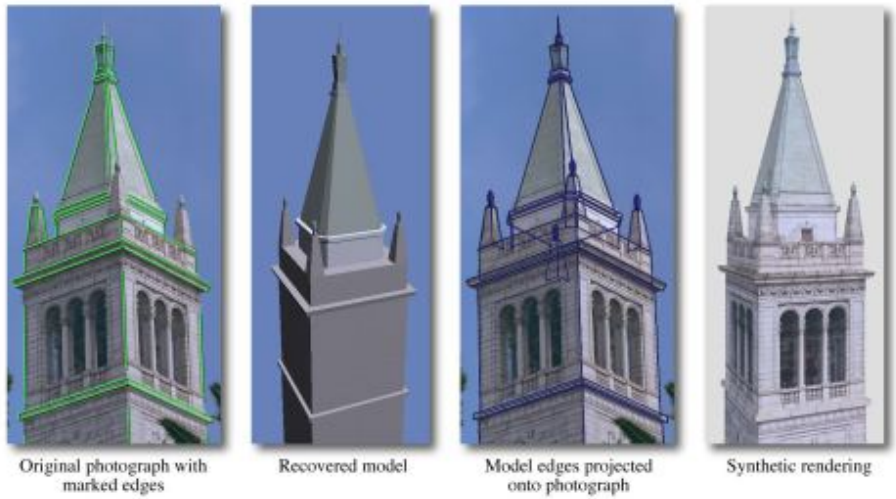
Image from LeCun et al. 1989

CV Past, Present, Future
1996: Image-Based Modeling

Generate a 3D model from a set of images

Use this model and input images to render new images

`https://www.youtube.com/watch?v=RPhGEiM_6lM`



Images from Debevec 1996

CV Past, Present, Future

2001: Real-Time Object Detection

Fast object detection using Haar features and boosting

Similar technologies used in smart cameras for auto focus

Image from olympus-europa.com

CV Past, Present, Future

2006: Photo Tourism

3D reconstruction from photo collections

Structure from Motion (SIFT + bundle adjustment)

Image from Snavely, Seitz, and Szeliski 2006

CV Past, Present, Future

2006: Photo Tourism – Microsoft Photosynth

Image from photosynth.net

CV Past, Present, Future

2006: Photo Tourism – Building Rome in a Day

Image from https://www.youtube.com/watch?v=sQegEro5Bfo

CV Past, Present, Future

2011: Kinect

Depth estimation via active stereo

Real-time pose estimation of multiple players

Image from wikipedia.org

Image from Shotton et al. 2011

CV Past, Present, Future
2012: Deep Learning and Big Data

Deep Learning on huge datasets for object recognition

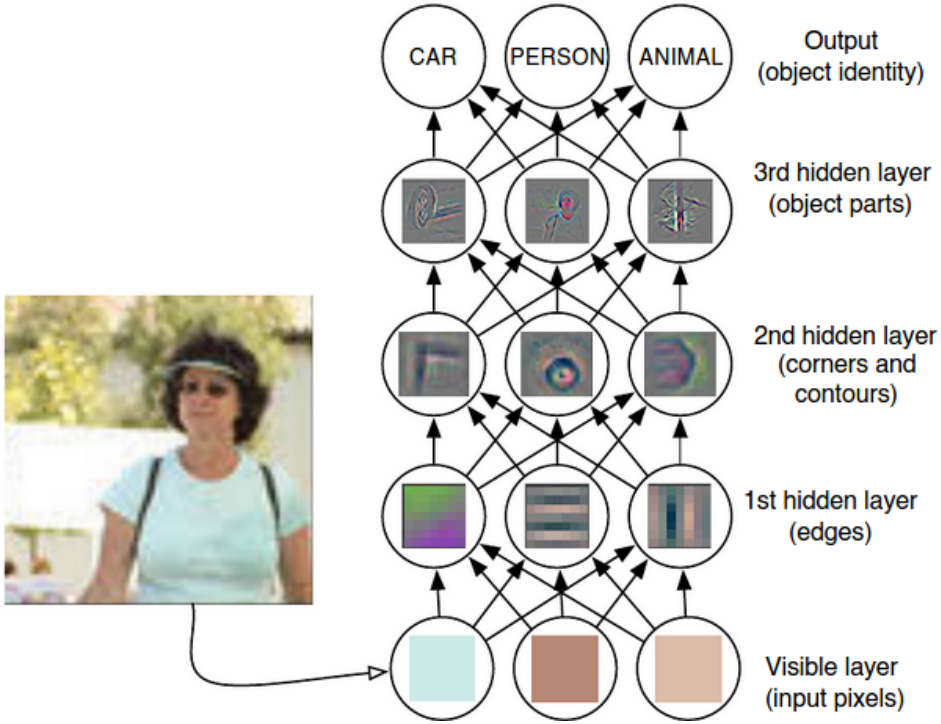


Image from Bengio, Goodfellow, and Courville 2015

CV Past, Present, Future

2012: Deep Learning and Big Data – Clarifai

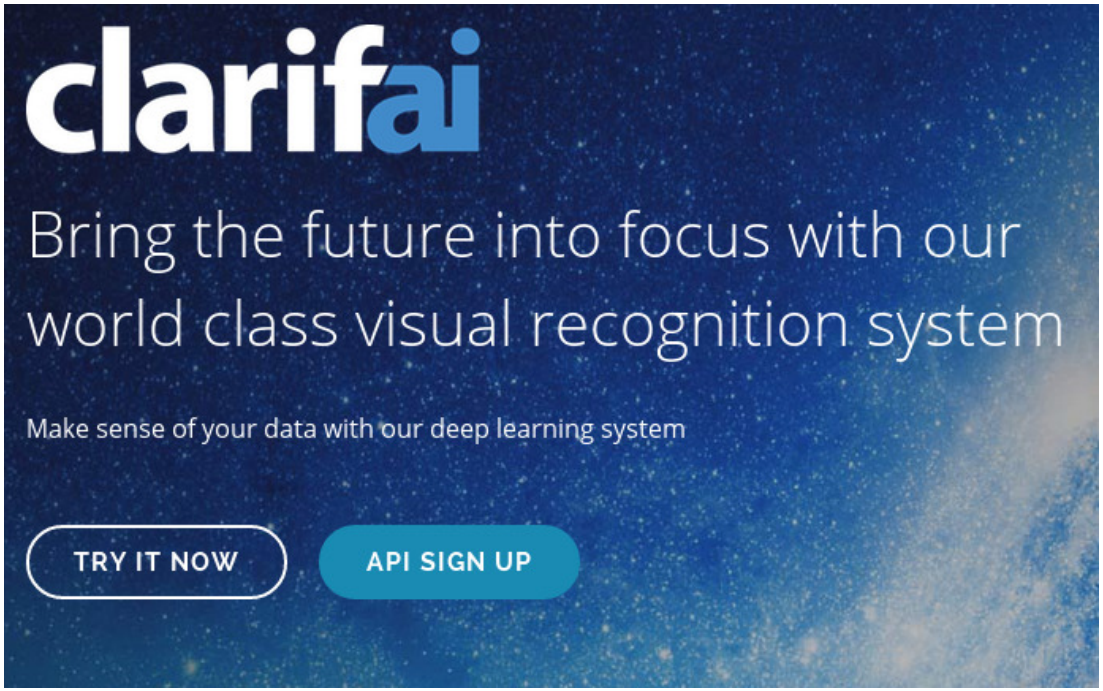


Image from clarifai.com

CV Past, Present, Future

2012: Deep Learning and Big Data

Image from ted.com

Object recognition without constraints

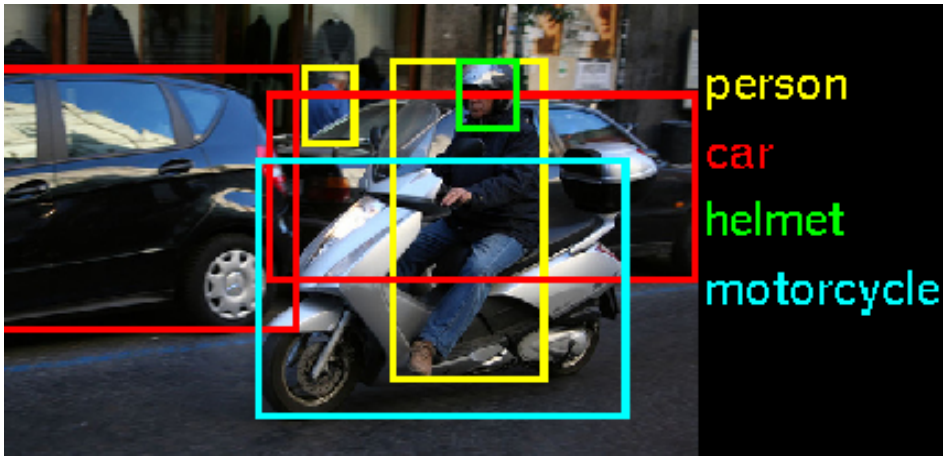


Image from image-net.org

CV Past, Present, Future

20xx: Autonomous Cars

Cars that drive autonomously

`https://www.youtube.com/watch?v=bDOnn0-4Nq8`

Image by Google

CV Past, Present, Future

20xx: Human-Level Vision

Segmentation, context, motion, emotions

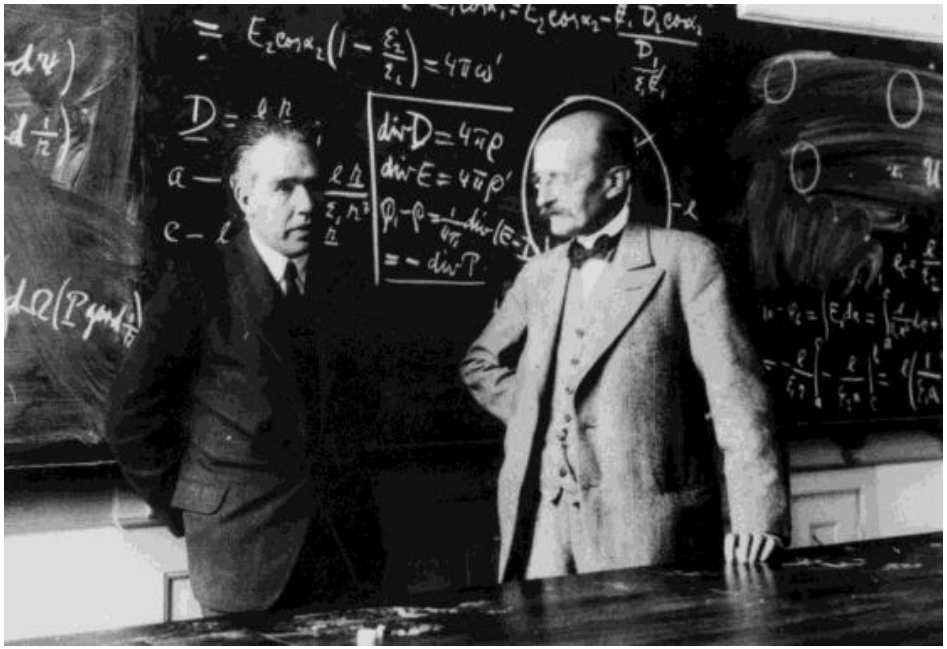


Image from Larry Zitnick's slides

CV Past, Present, Future

20xx: Human-Level Vision

Image from ted.com

CV and Related Fields

In other lectures you probably heard about

- Mathematics and statistics
- Image processing (e.g. linear filtering, SIFT)
- Machine learning (e.g. SVM)

Let's see how CV and these fields are related

- Using an example application for explanation purposes

We want to distinguish two kinds of fish in images

- Each image contains exactly one fish
- Only fishes of two known kinds (e.g. sea bass and salmon)
- Little background clutter

This is a binary (two-class) image classification problem

CV is about making computers understand images like humans do

So in mathematical terms CV is about

- Inferring some world state (a scalar w or vector $\mathbf{w}$)
- From measurements $\mathbf{x}$ (a feature vector)

w encodes what we care about, the fish type

- Let $w = 0$ if the image depicts a sea bass
- Let $w = 1$ if the image depicts a salmon

What are our measurements $\mathbf{x}$?

- We could simply use the whole image
- But we should be able to come up with something better
- This is where image processing comes in

CV and Related Fields
Image Processing

We use image processing to extract $\mathbf{x}$ from images

- Preprocessing step for CV

Different problems favor different features $\mathbf{x}$

Do sea basses and salmons look differently?

- $\mathbf{x}$ should encode fish appearance

Do they differ in size?

- $\mathbf{x}$ should encode fish size

CV and Related Fields
Statistics

How do we get from $\mathbf{x}$ to $\mathbf{w}$?

- This is what CV is all about!

CV is about describing the relationship between $\mathbf{x}$ and $\mathbf{w}$

- This relationship is called model

Put simply, there are two ways for modeling this relationship

- Statistically by looking at data and using domain knowledge
- Using a generic machine learning algorithm

More on models later

CV and Related Fields
Mathematics

Models are mathematical functions

Models have parameters that are learned from data

- This is a mathematical optimization problem

CV and Related Fields
Machine Learning

Machine Learning (ML) studies techniques for learning from data

- Namely algorithms for learning and inference
- So any CV model that involves learning is a ML technique

Generic ML algorithms (e.g. SVM) are often used as the model

Strictly speaking, models and algorithms are not the same

- More on this later

Bibliography I

Bengio, Yoshua, Ian Goodfellow, and Aaron Courville (2015). *Deep Learning (Draft)*. MIT Press.

Debevec, Paul E. (1996). *Modeling and Rendering Architecture from Photographs*. PhD thesis. Berkley.

Fischler, Martin A and Robert A Elschlager (1973). *The representation and matching of pictorial structures*. IEEE Transactions on Computers.

LeCun, Yann et al. (1989). *Backpropagation applied to handwritten zip code recognition*. Neural computation.

Roberts, Lawrence Gilman (1963). *Machine perception of three-dimensional solids*. PhD thesis. MIT.

Bibliography II

Shotton, Jamie et al. (2011). *Real-Time Human Pose Recognition in Parts from a Single Depth Image*. CVPR.

Snavely, Noah, Steven M. Seitz, and Richard Szeliski (2006). *Photo tourism: Exploring photo collections in 3D*. SIGGRAPH.

Taigman, Yaniv et al. (2013). *Deepface: Closing the gap to human-level performance in face verification*. CVPR.