Computer Vision

L01:Introduction

Prof. Dr. Mohammed Elmogy

Mansoura University
Faculty of Computers and Information
Information Technology Department

melmogy@mans.edu.eg

21. September 2025

Outline

What is Computer Vision?
The Goal of Computer Vision
Why is Vision So Difficult?
Computer Vision Applications
CV Evolution

Outline

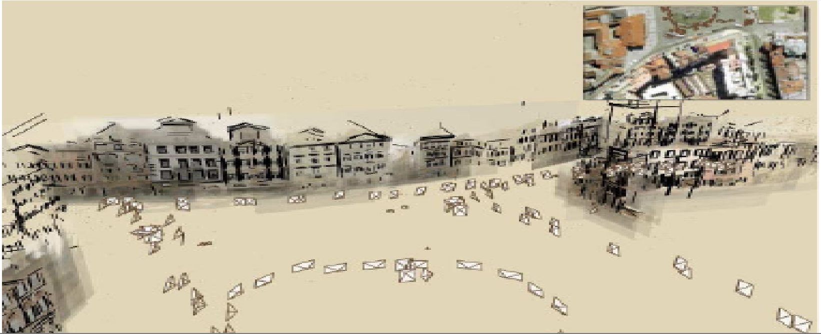
What is Computer Vision?
The Goal of Computer Vision
Why is Vision So Difficult?
Computer Vision Applications
CV Evolution

What is Computer Vision (CV)?

The human visual system has no problem interpreting the subtle variations in translucency and shading in this photograph and correctly segmenting the object from its background.

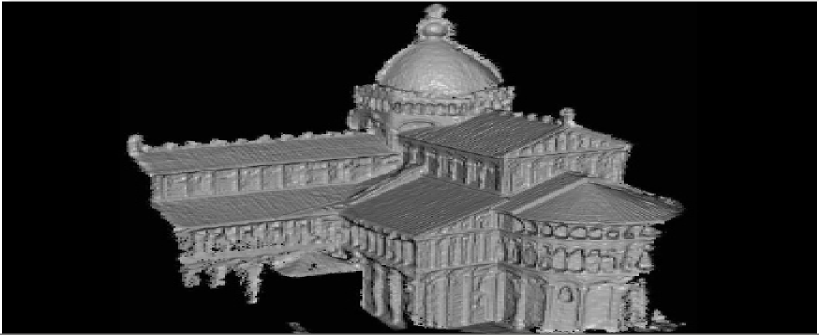
What is Computer Vision? (cont.)

Reliable techniques for accurately computing a partial 3D model of an environment from thousands of partially overlapping photographs.



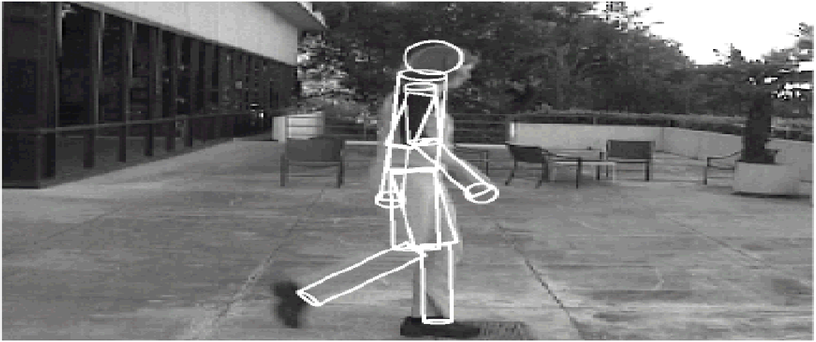
What is Computer Vision? (cont.)

Given a large enough set of views of a particular object or facade, we can create accurate dense 3D surface models using stereo matching.



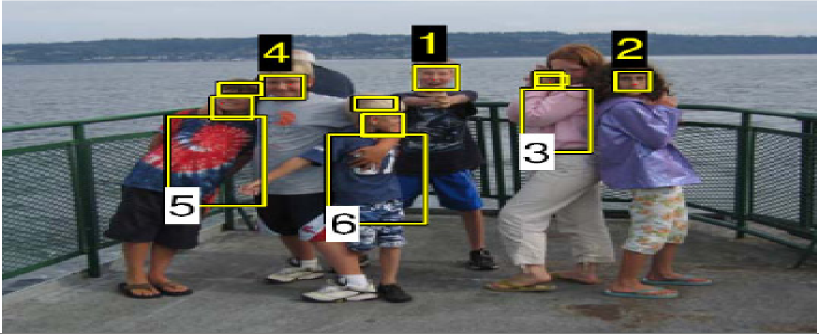
What is Computer Vision? (cont.)

We can track a person moving against a complex background.



What is Computer Vision? (cont.)

We can even, with moderate success, attempt to find and name all of the people in a photograph using a combination of face, clothing, and hair detection and recognition.



CV Definition

Definitions

- It can be defined as the ability of computers to see.
- Building artificial systems that **perceive**, **process**, and **reason** about visual data.
- A field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from **digital images**, **videos**, and other **visual inputs** — and take actions or make recommendations based on that information.

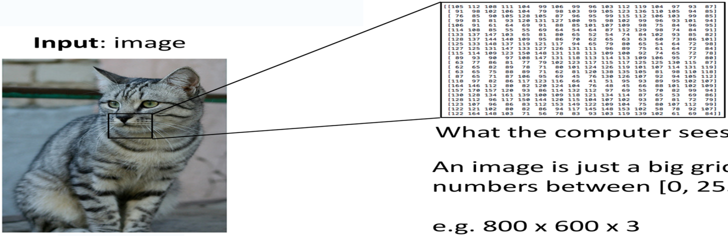
- CV has many names, such as Image Understanding, Machine Vision, Robot Vision, Image Analysis, and Video Understanding.

CV Definition (cont'd)

- Analogy: If **AI is about building brains**, **CV is about giving those brains eyes** and the **visual cortex to understand what they se**e.
- Goal: To **replicate** and **surpass** human visual capabilities.

CV Definition (cont'd)

Input: image



What the computer sees

An image is just a big grid of numbers between [0, 255]:

e.g. 800 x 600 x 3
(3 channels RGB)

Cat
Dog
Deer
Bird
Horse

Output: Assign image to one of a fixed set of categories

Every Image Tells a Story

A Picture Is Worth a Thousand Words

- Goal of computer vision: perceive the story behind the picture
- Compute properties of the world
  - 3D shape and appearance
  - Names of people or objects
  - Track a person moving against a complex background
  - What happened?

Related Disciplines

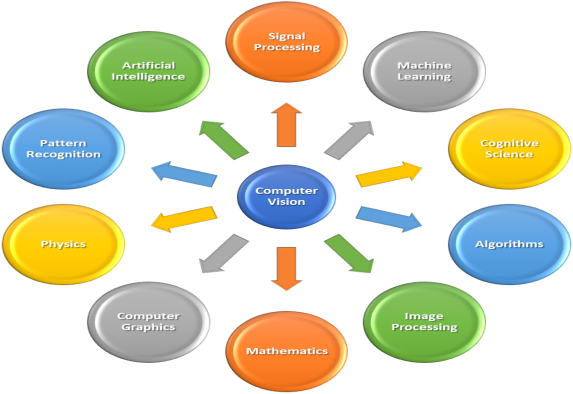
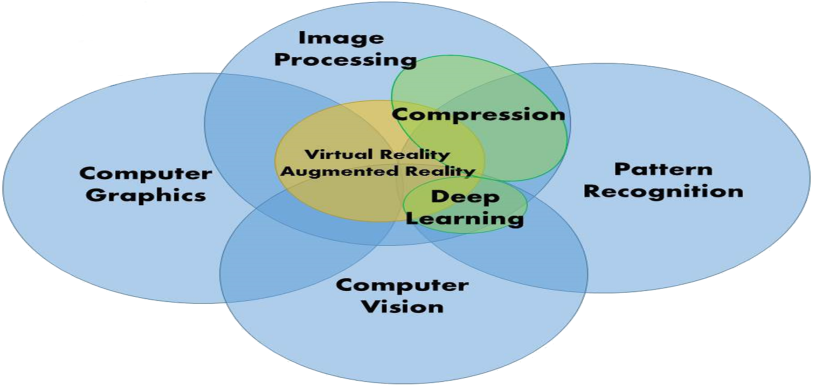
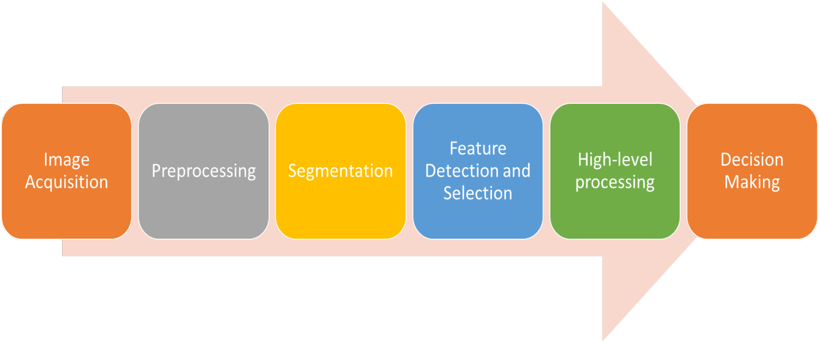


Image Processing, Computer Graphics, and Computer Vision



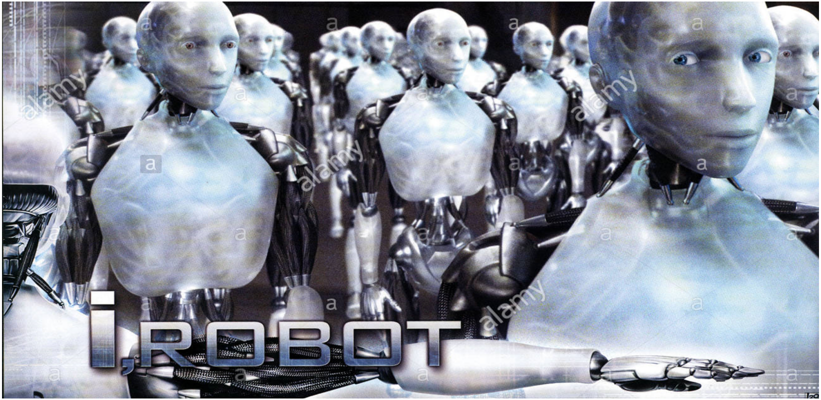
Common CV Pipeline



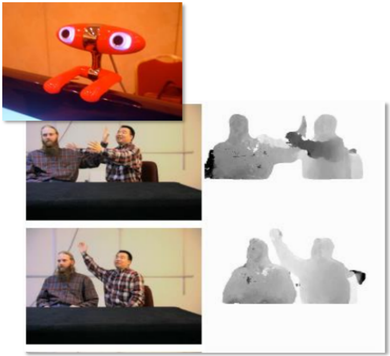
Outline

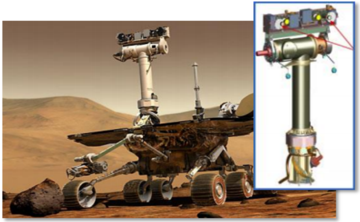
What is Computer Vision?
The Goal of Computer Vision
Why is Vision So Difficult?
Computer Vision Applications
CV Evolution

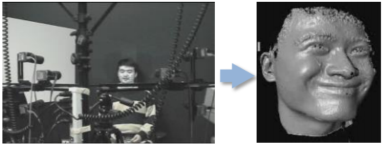
The Goal of Computer Vision



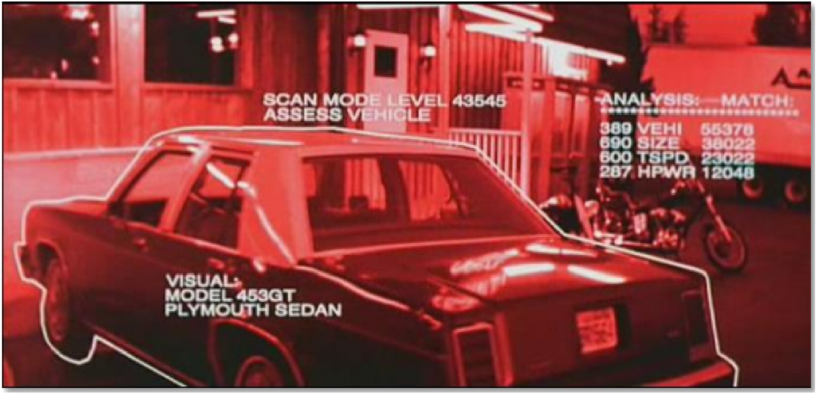
Computing the 3D Shape of the World







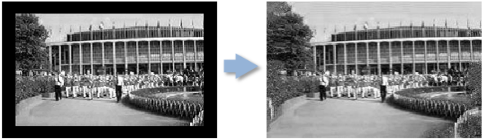
Recognizing Objects and People



Recognizing Objects and People (cont'd)

Enhancing Images (c.f. Computational Photography)

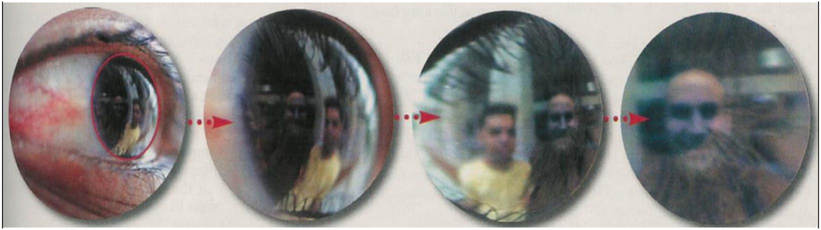
Super-resolution / denoising
(source: 2d3)



Texture synthesis / increased field of view (uncropping)
(image credit: Efros and Leung)

Inpainting / image completion
(image credit: Hays and Efros)

Forensics



Outline

What is Computer Vision?
The Goal of Computer Vision
Why is Vision So Difficult?
Computer Vision Applications
CV Evolution

Why is Vision So Difficult?

- Despite all of these advances, the dream of having a computer interpret an image at the same level as a two-year old remains elusive.
- It is because vision is an inverse problem, in which we seek to recover some unknowns given insufficient information to fully specify the solution.
- However, modeling the visual world in all of its rich complexity is far more difficult than, say, modeling the vocal tract that produces spoken sounds.

The Forward Models

- The forward models that we use in computer vision are usually developed in physics (**radiometry**, **optics**, and **sensor** design) and in **computer graphics**.
- **Both of these fields model**:
  - How objects move and animate.
  - How light reflects off their surfaces, is scattered by the atmosphere, refracted through camera lenses (or human eyes), and finally projected onto a flat (or curved) image plane.
- Computer graphics are not yet perfect.

The Inverse Models

- In computer vision, we are trying to do the inverse, i.e., to describe the world that we see in one or more images and to reconstruct its properties, such as **shape**, **illumination**, and **color distributions**.

Classic Problems of Computational Vision: 3R's

- Reconstruction: It focuses on recovering the 3D geometry and properties of the world from images.
  - 3D model generation of environments, medical imaging (like CAT/MRI scans), and sensing for robots.
- Recognition: It involves assigning semantic labels to objects and scenes.
  - Face recognition, object detection, and instance-level identification (e.g., recognizing a specific person).
- Reorganization: It refers to perceptual organization, or the process of grouping and segmenting visual information.
  - Grouping and segmentation.

Why is CV So Challenging?

- The Semantic Gap: The huge gap between pixel data and meaningful information.
- Viewpoint Variation: An object looks different from different angles.
- Illumination Conditions: Light changes everything.
- Occlusion: Objects hide behind each other.
- Background Clutter: Distinguishing object from background.
- Intra-class Variation: Many types of "chairs" or "cats".

Formulating and Solving Computer Vision Problems

We have four high-level approaches:

- Scientific: build **detailed models of the image formation process** and **develop mathematical techniques** to invert these in order to recover the quantities of interest.

- Statistical: use **probabilistic models to quantify the prior likelihood of your unknowns** and the **noisy measurement** processes that produce the input images, then infer the best possible estimates of your desired quantities and analyze their resulting uncertainties.

Formulating and Solving Computer Vision Problems (cont'd)

- Engineering: develop techniques that are simple to describe and implement but that are also known to work well in practice. Test these techniques to understand their limitation and failure modes, as well as their expected computational costs.
- Data-driven: **Collect a representative set of test data** (ideally, with labels or ground-truth answers) and use this data to either tune or learn your model parameters, or at least to validate and quantify its performance.

Validate Your CV Algorithms

Use a three-part strategy:

1. Test your algorithm on clean synthetic data, for which the exact results are known.

2. Add noise to the data and evaluate how the performance degrades as a function of noise level.

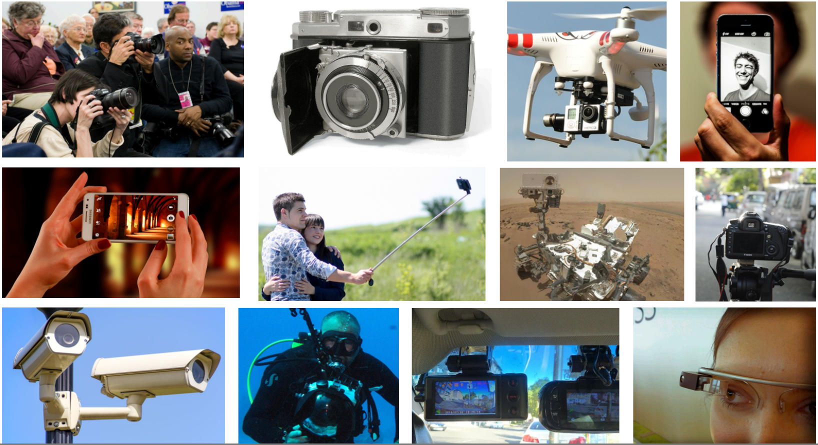
3. Test the algorithm on real-world data, preferably drawn from a wide variety of sources, such as photos found on the Web.

Outline

- What is Computer Vision?
- The Goal of Computer Vision
- Why is Vision So Difficult?
- Computer Vision Applications
- CV Evolution

Computer Vision Applications

Computer Vision Is Everywhere

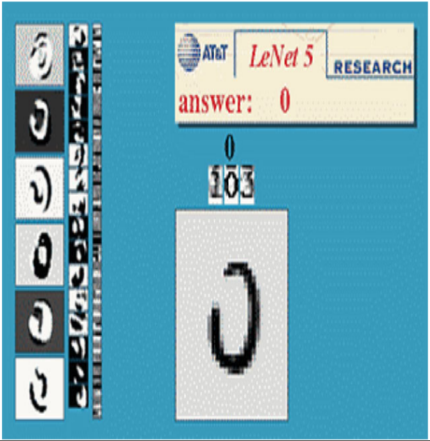


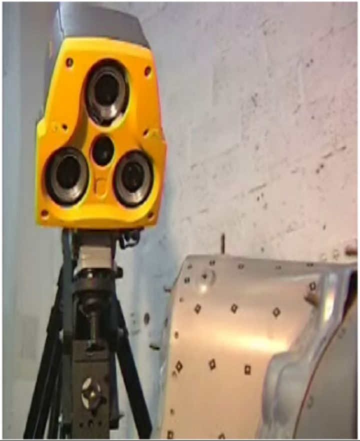
Key Applications

- Healthcare: Medical imaging (tumor detection in MRI/X-rays), surgery assistance.
- Automotive: Self-driving cars (lane detection, pedestrian avoidance).
- Security & Surveillance: Facial recognition, anomaly detection.
- Retail: Automated checkout (Amazon Go), inventory management.
- Augmented Reality (AR): Snapchat filters, Pokemon Go, IKEA Place app.
- Agriculture: Crop monitoring, automated harvesting.
- Social Media: Photo tagging, content moderation.

Computer Vision Applications

Optical Character Recognition (OCR) & Machine Inspection



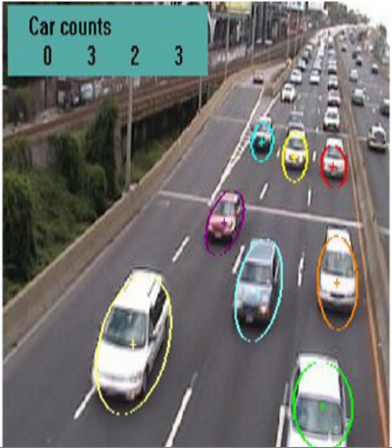


Computer Vision Applications (cont.)

Retail & Medical Imaging

Computer Vision Applications (cont.)

Automotive Safety & Surveillance

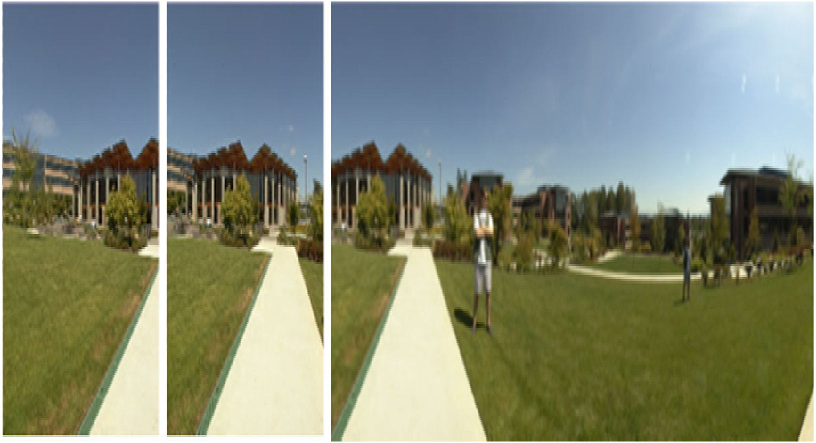


Computer Vision Applications (cont.)

- 3D model building (photogrammetry).
- Match move: merging computer-generated imagery (CGI) with live action footage by tracking feature points in the source video to estimate the 3D camera motion and shape of the environment.
- Motion capture (mocap): using retro-reflective markers viewed from multiple cameras or other vision-based techniques to capture actors for computer animation.
- Fingerprint recognition and biometrics.

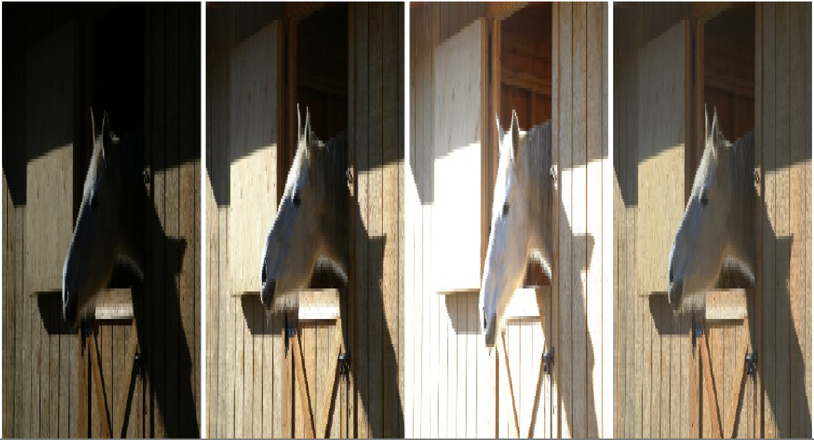
Some Consumer Applications of Computer Vision

Image Stitching: merging different views



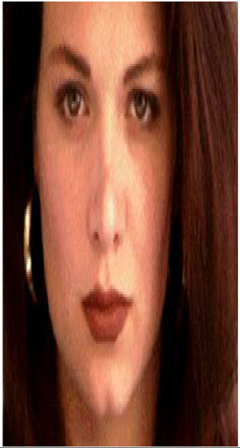
Some Consumer Applications of Computer Vision

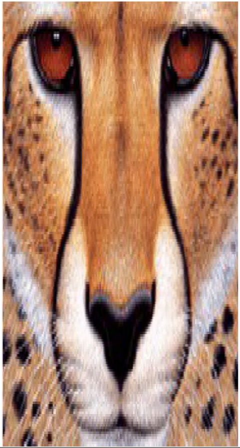
Exposure Bracketing: merging different exposures

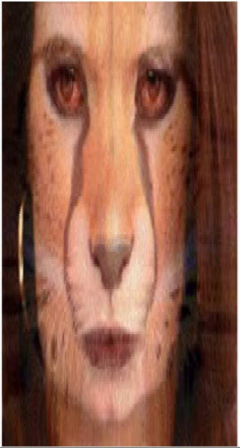


Some Consumer Applications of Computer Vision

Morphing: blending between two photographs

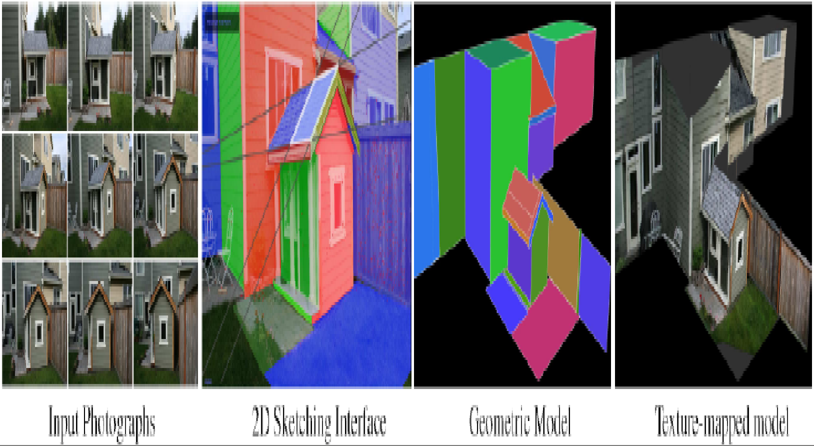






Some Consumer Applications of Computer Vision

Turning a collection of photographs into a 3D model



Some Consumer Applications of Computer Vision

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz

Some Consumer Applications of Computer Vision

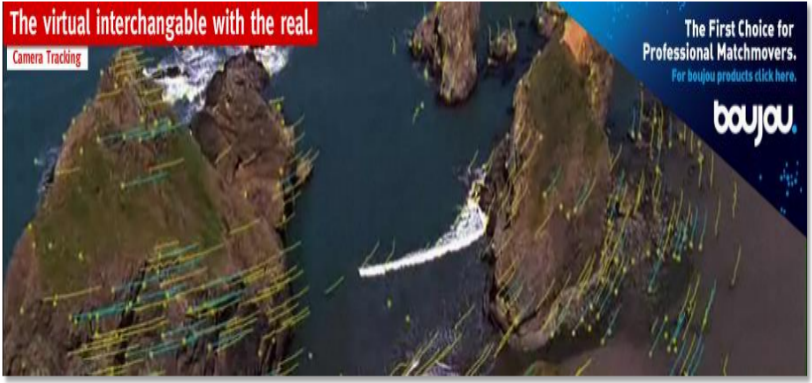
Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Source: S. Seitz

Some Consumer Applications of Computer Vision

Special effects: camera tracking



Boujou, 2d3

Some Consumer Applications of Computer Vision

Sports

Sportvision first down line

Nice explanation on www.howstuffworks.com

Some Consumer Applications of Computer Vision

Vision-based interaction (and games)

Nintendo Wii has camera-based IR tracking built in. See Lee's work at CMU on clever tricks on using it to create a multi-touch display!

Assistive technologies

Some Consumer Applications of Computer Vision

Robotics

NASA's Mars Curiosity Rover (Mars Science Laboratory)
http://en.wikipedia.org/wiki/Spirit_rover

Autonomous RC Car
http://www.cs.cornell.edu/~asaxena/rccar/

Outline

What is Computer Vision?
The Goal of Computer Vision
Why is Vision So Difficult?
Computer Vision Applications
CV Evolution

From Traditional Methods to Deep Learning

- Traditional (Pre-2012): **Feature Engineering**
  - **Idea**: Humans design "features" (descriptors) that machines look for.
  - **Examples**: SIFT, HOG, Haar Cascades.
  - **Limitation**: Requires expert knowledge, brittle, doesn't generalize well.
- Deep Learning Revolution (Post-2012): **Feature Learning**
  - **Idea**: Let the machine learn the optimal features directly from the data.
  - **Key Enabler**: Convolutional Neural Networks (CNNs) + Big Data (ImageNet) + GPUs.
  - **Advantage**: Highly accurate, scalable, and generalizable

.

The Building Block: Convolutional Neural Networks (CNNs)

- **Inspired by**: The human visual cortex.
- **Core Layers**:
  1. **Convolutional Layer**: Detects features (edges, textures, patterns) using filters.
  2. **Pooling Layer** (e.g., Max Pooling): Reduces dimensionality, makes detection invariant to scale and small shifts.
  3. **Fully Connected Layer**: Final classification (e.g., "cat", "dog").

Landmark CNN Architectures: A Timeline

- 1998: LeNet-5 (Yann LeCun): Recognized handwritten digits (MNIST). Proof of concept.
- 2012: AlexNet (Krizhevsky et al.): The breakthrough. Won ImageNet by a huge margin, starting the DL boom.
- 2014: VGGNet (Simonyan & Zisserman): Very deep (16-19 layers), simple architecture. Showed that depth is key.
- 2015: ResNet (He et al.): Introduced "skip connections" to solve the vanishing gradient problem, enabling networks of 100+ layers. Still widely used today.

Beyond Classification: Object Detection (YOLO/SSD)

- Task: Not just what is in the image, but where is it? (Bounding Box + Class)
- Key Idea: Look at the image once in a grid and predict boxes and classes simultaneously.
- Frameworks: YOLO (You Only Look Once) and SSD (Single Shot Detector) are fast and accurate.

Beyond Classification: Image Segmentation

- Semantic Segmentation: Labels each pixel with a class (e.g., road, car, sky). "What is what?"
- Instance Segmentation: Labels each pixel and differentiates between separate objects of the same class. "What is what and which is which?"
- Application: Medical imaging (precise tumor boundaries), autonomous driving (precise drivable area).

Modern Trends & The Future

- Vision Transformers (ViTs): Applying the powerful Transformer architecture (from NLP) to images, challenging CNNs.
- Generative Models:
  - **GANs (Generative Adversarial Networks)**: Create realistic fake images, style transfer.
  - **Diffusion Models**: (e.g., DALL-E, Midjourney, Stable Diffusion) The new state-of-the-art in generative AI.
- Video Understanding: Moving from static images to dynamic video for action recognition.
- Efficiency: Making models smaller and faster for mobile devices (MobileNet, SqueezeNet).

Conclusion

- Computer Vision enables machines to interpret the visual world.
- Deep Learning, specifically CNNs, was a paradigm shift that unlocked modern CV.
- It powers a vast array of real-world applications.
- The field is rapidly evolving with Transformers and Generative AI.
- Final Thought: "We are teaching machines to see, and in doing so, we are gaining a deeper understanding of our own visual perception."

References

- Richard Szeliski, Computer Vision: Algorithms and Applications, 2nd Ed., Springer-Verlag London, 2020.
- Reinhard Klette, Concise Computer Vision: An Introduction into Theory and Algorithms, Springer-Verlag London 2014.
- Mubarak Shah, CV Course: Lecture 1 (http://www.youtube.com/watch?v=715uLCHt4jE&feature=plcp).
- CSCI 1430: Introduction to Computer Vision, (https://cs.brown.edu/courses/csci1430/)
- Deep Learning for Computer Vision, University of Michigan (https://web.eecs.umich.edu/ justincj/teaching/eecs498/FA2020/sylla