Computer Vision Systems Programming VO

Specific Object Recognition

Christopher Pramerdorfer

Computer Vision Lab, Vienna University of Technology

Topics

Introduction to object recognition

Specific object recognition (rigid, nonrigid)

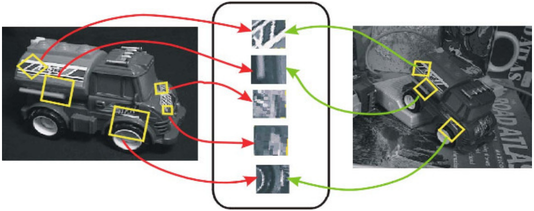


Image from Grauman and Leibe 2011

Object Recognition

Fundamental problem in Computer Vision

Many applications

- Panorama stitching, 3D reconstruction
- HCI and surveillance (face recognition)
- Image understanding (recall Fei-Fei Li's TED talk)

Object Recognition
Taxonomy – Instance vs. Category

Instance recognition (specific object recognition)

- Recognize a specific, uniquely looking object
- Face of a certain person, the Eiffel tower

Object category recognition

- Recognize objects of a certain category
- Human faces, buildings

Object Recognition
Taxonomy – Instance vs. Category

Object Recognition

Taxonomy – Classification vs. Detection

Object classification

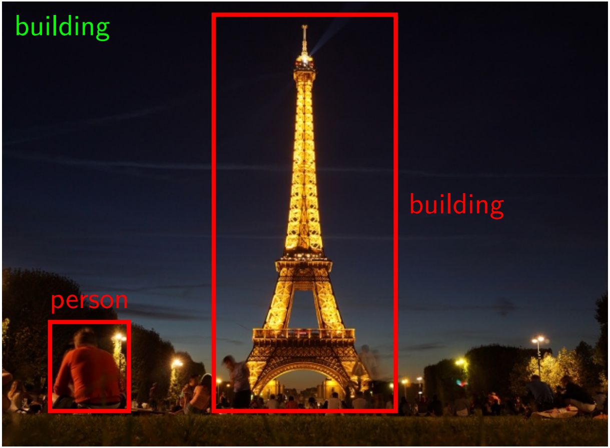
- Recognize main object in image
- Location and other objects not relevant

Object detection

- Recognize multiple objects, possibly of different category

Object Recognition

Taxonomy – Classification vs. Detection



Instances of same category can look very differently

- Illumination, pose, viewpoint, occlusions, background

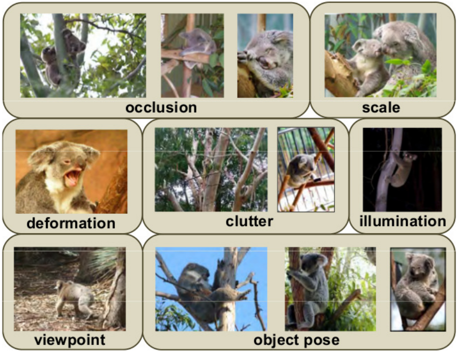


Image from Grauman and Leibe 2011

Specific Planar Object Detection

We want to detect specific planar objects (e.g. markers, books)

- Comparatively easy problem

Challenges

- Unknown object pose and scale
- Partial occlusions
- Illumination

Specific Planar Object Detection

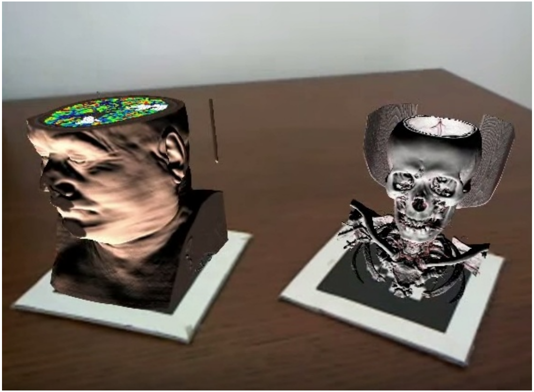


Image from youtube.com

Our problem formulation is

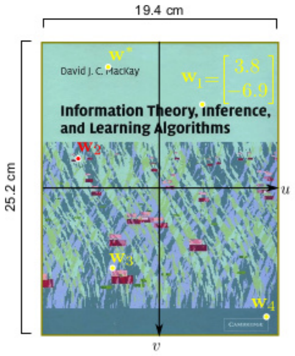
- Given a pixel location in a query image
- Predict location in reference image of sought object

So we know how to select $\mathbf{x}$ and $\mathbf{w}$

- $\mathbf{x} = (x, y)$: location in query image
- $\mathbf{w} = (u, v)$: corresponding location in reference image

Specific Planar Object Detection

Selecting $\mathbf{x}$ and $\mathbf{w}$



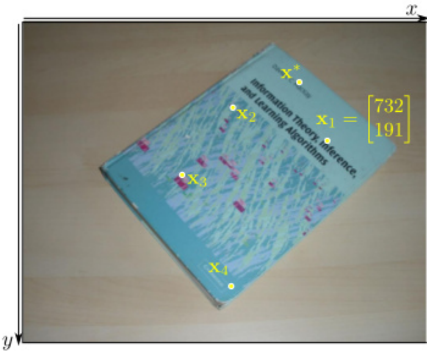


Image from Prince 2012

Images of planar objects are always related by a homography $\mathbf{\Phi}$

- 3×3 matrix mapping between corresponding points

In homogeneous coordinates this means that

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{\Phi} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

The model of choice is thus (disregarding noise)

$$\mathbf{w} = \Gamma(\mathbf{x}) = \begin{pmatrix} u \\ v \end{pmatrix} \quad , \quad \lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{\Phi} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

We again learn parameters $\boldsymbol{\theta}$ from samples $\{\mathbf{x}_i, \mathbf{w}_i\}_{i=1}^{n}$

- $\boldsymbol{\theta}$ contains 9 parameters comprising $\boldsymbol{\Phi}$

Usually no exact solution because of noisy $\mathbf{x}_i$

- Formulate as a least squares problem instead

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\min} \left[\sum_{i=1}^{n} (\mathbf{w}_i - \Gamma(\mathbf{x}_i))^\top (\mathbf{w}_i - \Gamma(\mathbf{x}_i)) \right]$$

This least squares approach is optimal

- If noise is distributed normally with spherical covariance

This is a nonlinear optimization problem

- Solvable using any general nonlinear least squares solver
- OpenCV has an own function `findHomography`

Specific Planar Object Detection

But how can we compute $\{\mathbf{x}_i, \mathbf{w}_i\}_{i=1}^{n}$ automatically?

- Next lecture

Bibliography

Grauman, Kristen and Bastian Leibe (2011). *Visual object recognition*. Morgan & Claypool.

Prince, S.J.D. (2012). *Computer Vision: Models Learning and Inference*. Cambridge University Press.