Computer Vision Systems Programming VO

Object Recognition

Christopher Pramerdorfer

Computer Vision Lab, Vienna University of Technology

Topics

Taxonomy of recognition problems

Selection of popular applications involving object recognition

Taxonomy of Object Recognition

Instance vs. category recognition

Instance : my face, the Eiffel tower

Category : faces, buildings, people

Taxonomy of Object Recognition

Levels of Recognition – Classification

Predict class of main object in image

Top 5:
pencil sharpener
pool table
hand blower
oil filter
packet

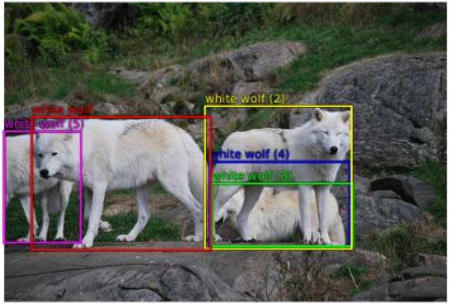
Groundtruth:
pencil sharpener

ILSVRC2012_val_00010000.JPEG

Image from Pierre Sermanet's slides

Taxonomy of Object Recognition
Levels of Recognition – Localization

Predict class and location(s) of main object in image



Groundtruth:
white wolf
white wolf (2)
white wolf (3)
white wolf (4)
white wolf (5)

Image from Pierre Sermanet's slides

Taxonomy of Object Recognition
Levels of Recognition – Detection

Any number of different objects

Groundtruth:
tv or monitor
tv or monitor (2)
tv or monitor (3)
person
remote control
remote control (2)

Image from Pierre Sermanet's slides

Challenges

Instances of same category can look very differently

- Illumination, pose, viewpoint, occlusions, background

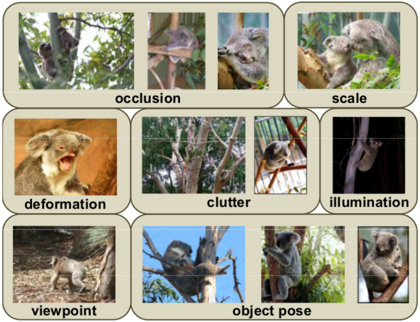


Image from Grauman and Leibe 2011

Instance Recognition

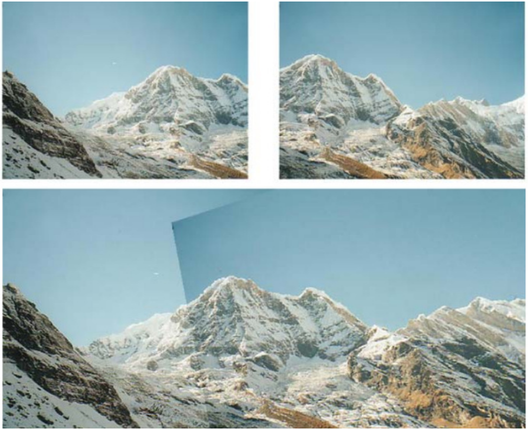


Image adapted from Brown and Lowe 2007

Instance Recognition

Assume that objects are rigid (nonrigid later)

Often accomplished via **local feature matching**

Given an image of the instance and a search image

- Compute local features in both images
- Match features between images to find correspondences
- Perform geometric verification

Local features form a sparse object representation

- Features capture characteristic structure
- Allow for efficient matching between images
- Representation robust to occlusion and clutter

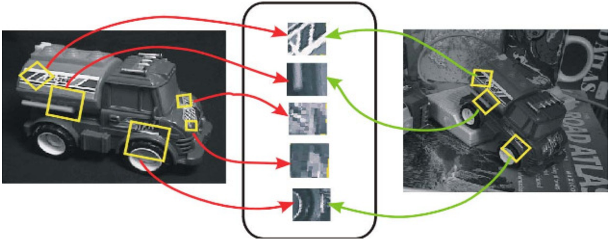


Image from Grauman and Leibe 2011

Many different feature extractors available

- SIFT, SURF, BRISK, FREAK, ...

Notes on feature extractor selection

- Features should be invariant to expected transformations
- But not to other transformations
- Extraction and matching speeds differ

Features are n-dimensional vectors

- Perform nearest neighbor matching in this feature space

Popular matching strategy

- Given feature $\mathbf{x}$ in first image
- Find two nearest neighbors $\mathbf{y}_1, \mathbf{y}_2$ from second image
- $\{\mathbf{x}, \mathbf{y}_1\}$ correspond if $\|\mathbf{x} - \mathbf{y}_1\| / \|\mathbf{x} - \mathbf{y}_2\| < 0.8$

Assume that the object in question is planar

- Images of planar objects are related by a homography
- Also applies to local feature locations

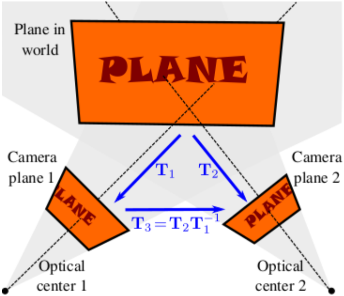


Image from Prince 2012

Relation allows for detecting erroneous correspondences

- Estimate homography T from correspondences
- Discard correspondences for which $\|\mathbf{x} - T(\mathbf{y}_1)\| > t$

Verification also possible for nonplanar scenes

- Epipolar geometry constraints (previous lecture)

Detection and pose estimation of rigid objects

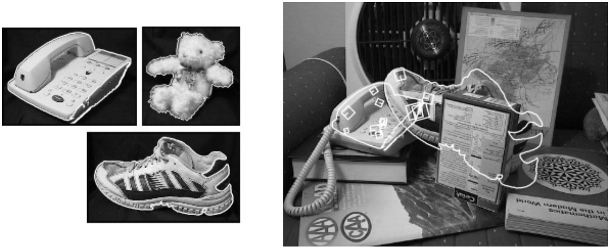
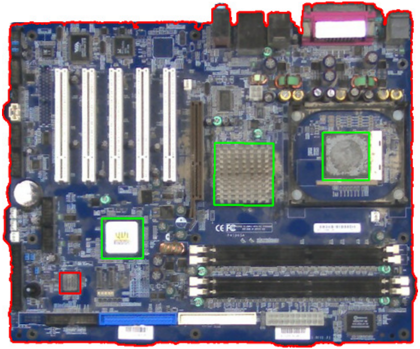


Image adapted from Lowe 2004

Instance Recognition
Applications – Object Detection

Industrial applications like PCB recycling



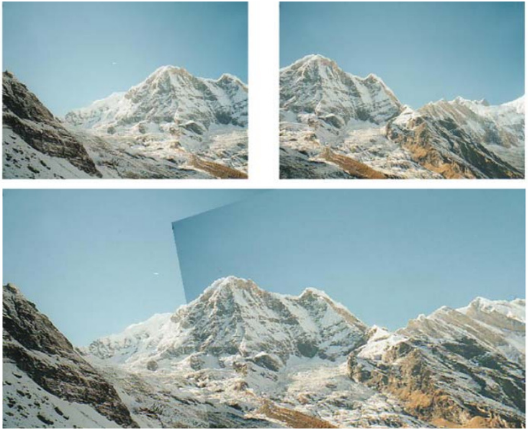


Image adapted from Brown and Lowe 2007

Face Localization

Image from `olympus-europa.com`

Face Localization

Many applications, such as

- Smart cameras (autofocus on faces)
- Security (preprocessing step to face recognition)
- Augmented reality & gaming

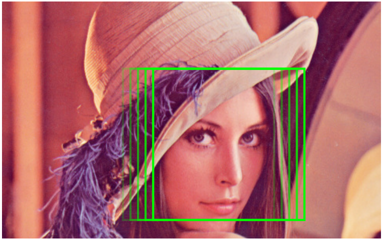
We focus on the popular method from [Viola and Jones 2001]

- Fast enough to run on e.g. cameras

Face Localization
Viola and Jones 2001 – Approach

Sliding window approach

- Slide window over image
- Infer label $w \in \{0, 1\}$ based on measurements $\mathbf{x}$
- Perform non-maximum suppression of confidence scores



Face Localization
Viola and Jones 2001 – Features

Simple features – difference d in rectangular subwindow of $\mathbf{x}$

- Can be computed in constant time using integral images
- Limited number of such features $\{f_t\}_{t=1}^{T}$

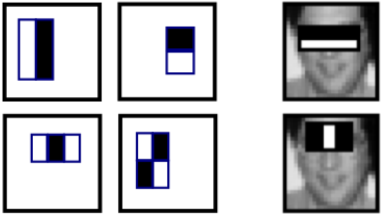


Image adapted from Prince 2012

Classification using a cascading classifier

- Cascade of $K \leq T$ weak but fast classifiers $c_k = f_k > t_k$
- Early rejection of non-face windows for speed
- Final classification is $C(\mathbf{x}) = \text{sign}(w_0 + \sum_k w_k c_k)$

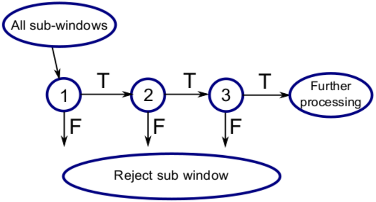


Image from Prince 2012

Face Localization
Viola and Jones 2001 – Classifier

Subset of K classifiers, order, and weights w are learned

Accomplished via **boosting** – for each $k = 1 \cdots K$

- Find best classifier according to training set, set as c_k
- Raise weights of samples misclassified by c_k

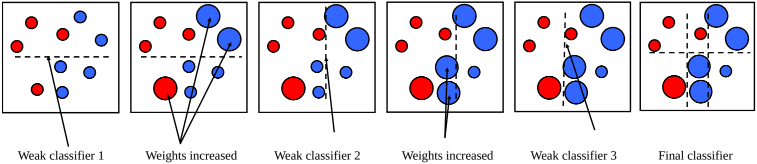


Image from Szeliski 2010

Bibliography I

Brown, Matthew and David G Lowe (2007). **Automatic panoramic image stitching using invariant features**.

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

Lowe, David G (2004). **Distinctive image features from scale-invariant keypoints**.

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

Szeliski, Richard (2010). **Computer vision: algorithms and applications**. Springer.

Bibliography II

Viola, Paul and Michael Jones (2001). **Rapid object detection using a boosted cascade of simple features**.