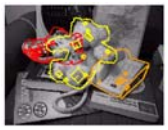
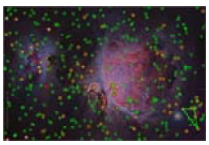


Introduction to recognition
Alignment-based approaches

Tuesday, Nov 4

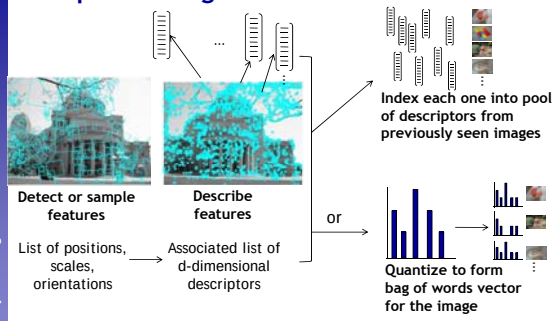


Kristen Grauman
UT-Austin

Today

- Brief recap of visual words
- Introduction to recognition problem
- Recognition by alignment, pose clustering

Recap: indexing features



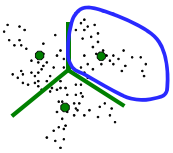
Visual Object Recognition Tutorial

K. Grauman, B. Leibe

3

Visual words

- Example: each group of patches belongs to the same visual word



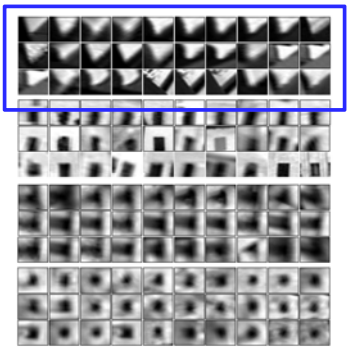


Figure from Sivic & Zisserman, ICCV 2003

Visual Object Recognition Tutorial

Inverted file index for images comprised of visual words

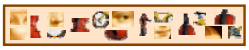
Word number	List of image numbers
1	5,10, …
2	10,…
…	…

Image credit: A. Zisserman

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

Bags of visual words



- Summarize entire image based on its distribution (histogram) of word occurrences.
- Analogous to bag of words representation commonly used for documents.

Image credit: Fei-Fei Li

K. Grauman, B. Leibe

6

Visual Object Recognition Tutorial

Bags of words: pros and cons

+ flexible to geometry / deformations / viewpoint
+ compact summary of image content
+ provides vector representation for sets
+ has yielded good recognition results in practice

- basic model ignores geometry – must verify afterwards, or encode via features
- background and foreground mixed when bag covers whole image
- interest points or sampling: no guarantee to capture object-level parts
- optimal vocabulary formation remains unclear

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

7

Review

- What are the tradeoffs related to the visual vocabulary size (number of words)?
- What is the role of tf-idf weighting for a bag-of-words representation?
- If we have established a vocabulary, and get a new image with some SIFT descriptors, how do we assign its features to words?

What does object recognition involve?

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

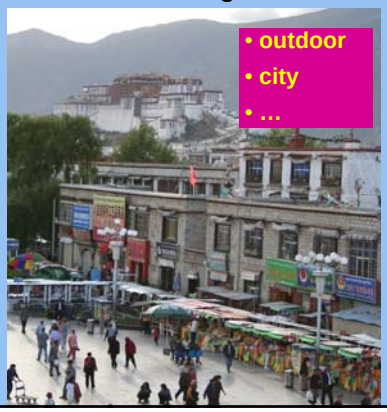
Verification: is that a lamp?

Detection: are there people?

Identification: is that Potala Palace?

Object categorization

Scene and context categorization



Object Categorization

- How to recognize ANY car

- How to recognize ANY cow

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

15

What could be done with recognition algorithms?

There is a wide range of applications, including…

Autonomous robots

Navigation, driver safety

Situated search

Content-based retrieval and analysis for images and videos

Medical image analysis

Visual Object Recognition Tutorial

Object Categorization

- Task Description
  - "Given a small number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label."
- Which categories are feasible visually?
  - Extensively studied in Cognitive Psychology, e.g. [Brown'58]

"Fido"

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Visual Object Categories

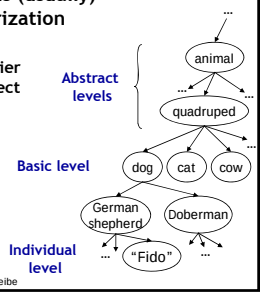
- Basic Level Categories in human categorization [Rosch 76, Lakoff 87]
  - The highest level at which category members have similar perceived shape
  - The highest level at which a single mental image reflects the entire category
  - The level at which human subjects are usually fastest at identifying category members
  - The first level named and understood by children
  - The highest level at which a person uses similar motor actions for interaction with category members

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Visual Object Categories

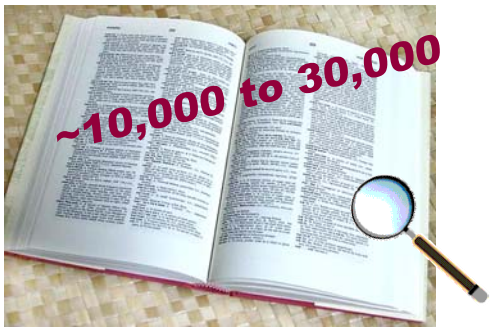
- Basic-level categories in humans seem to be defined predominantly visually.
- There is evidence that humans (usually) start with basic-level categorization *before* doing identification.
  - ⇒ Basic-level categorization is easier and faster for humans than object identification!
  - ⇒ Most promising starting point for visual classification



Visual Object Recognition Tutorial

K. Grauman, B. Leibe

How many object categories are there?



Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Biederman 1987

Other Types of Categories

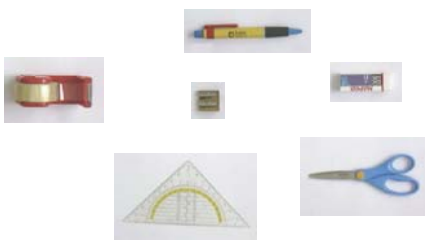
- Functional Categories
  - e.g. chairs = *"something you can sit on"*

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Other Types of Categories

- Ad-hoc categories
  - e.g. *"something you can find in an office environment"*



Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Challenges: robustness

Illumination | Object pose | Clutter

Occlusions | Intra-class appearance | Viewpoint

Visual Object Recognition Tutorial

Challenges: robustness

- Detection in Crowded Scenes
  - Learn object variability
    - Changes in appearance, scale, and articulation
  - Compensate for clutter, overlap, and occlusion

K. Grauman, B. Leibe

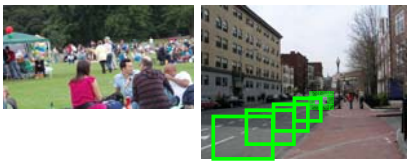
Visual Object Recognition Tutorial

Challenges: context and human experience

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

Challenges: context and human experience

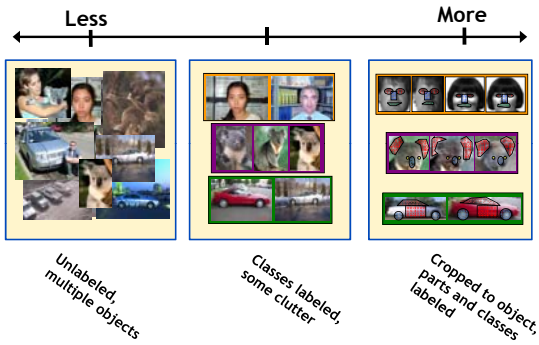


Context cues

Image credit: D. Hoeim

Visual Object Recognition Tutorial

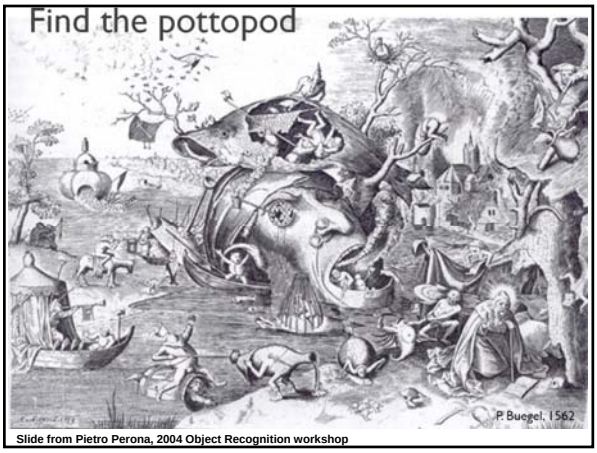
Challenges: learning with minimal supervision



K. Grauman, B. Leibe

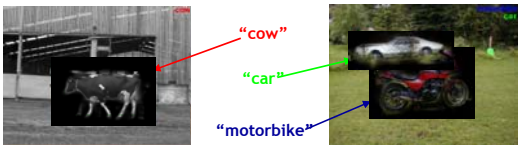
Visual Object Recognition Tutorial

Slide from Pietro Perona, 2004 Object Recognition workshop



Slide from Pietro Perona, 2004 Object Recognition workshop

Levels of Object Categorization



- **Different levels of recognition**
  - *Which* **object class is in the image?** ⇒ **Obj/Img classification**
  - *Where* **is it in the image?** ⇒ **Detection/Localization**
  - *Where exactly* **– which pixels?** ⇒ **Figure/Ground segmentation**

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

31

Inputs/outputs/assumptions

- What is the **goal**?
  - Say yes/no as to whether an object present in image
  - Determine pose of an object, e.g. for robot to grasp
  - Categorize all objects
  - Forced choice from pool of categories
  - Bounding box on object
  - Full segmentation
  - Build a model of an object category

Primary issues

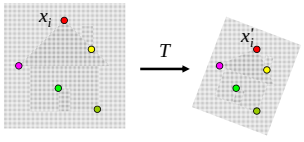
- How to **represent** a category or object
- How to perform **recognition** (classification, detection) with that representation
- How to **learn** models, new categories/objects

Genres of approaches

- Alignment
  - Pose clustering with object instances
- Global appearance
  - With or without a sliding window
- Local features
  - Indexing
  - Part-based models
    - Constellation models
    - Voting
  - Bags of words models

Recall: Alignment

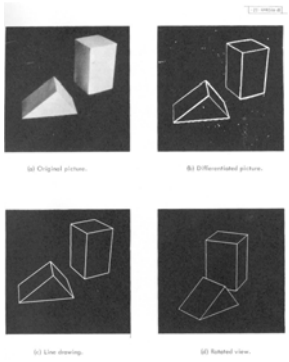
- Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images



Find transformation T that minimizes

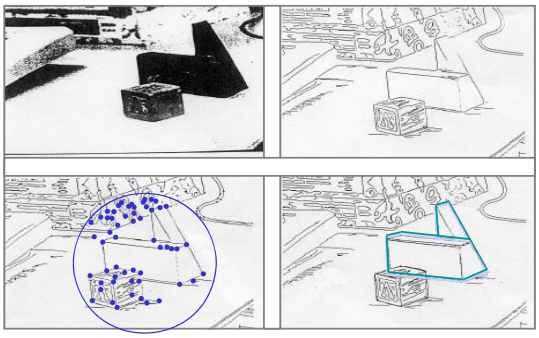
$$\sum_i \text{residual}(T(x_i), x_i')$$

Alignment-based



L. G. Roberts, Machine Perception of Three Dimensional Solids, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

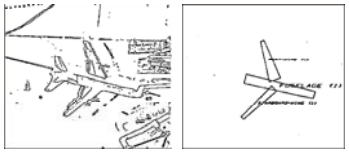
Alignment-based



Huttenlocher & Ullman (1987)

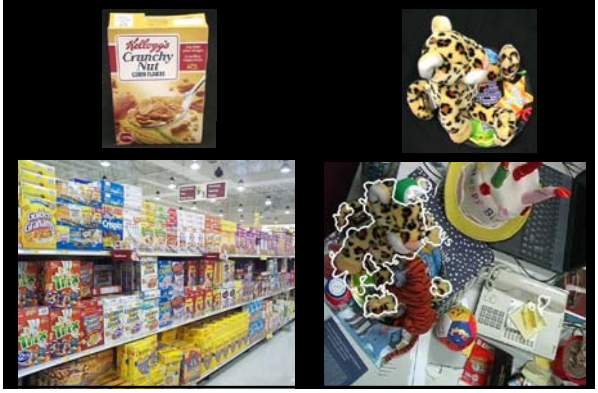
Source: Lana Lazebnik

Alignment-based



ACRONYM (Brooks and Binford, 1981)

Sparser patch matches : for object instances



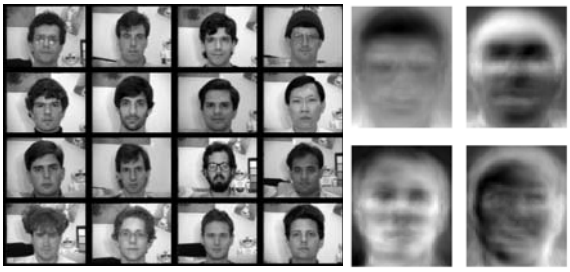
Genres of approaches

- Alignment
  - Pose clustering with object instances
- Global appearance
  - With or without a sliding window
- Local features
  - Indexing
  - Part-based models
    - Constellation models
    - Voting
  - Bags of words models

Global appearance-based

Swain and Ballard, Color Indexing, IJCV 1991.

Global appearance-based



Eigenfaces (Turk & Pentland, 1991)

Global appearance-based

Scene recognition based on global texture pattern.
[Oliva & Torralba (2001)]

Global appearance-based: sliding windows

Given a binary classifier that makes a decision based on global appearance, can slide a window around

Car/non-car Classifier

No, not a car.

Yes, a car.

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Global appearance-based: sliding windows

Given a binary classifier that makes a decision based on global appearance, can slide a window around

Car/non-car Classifier

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

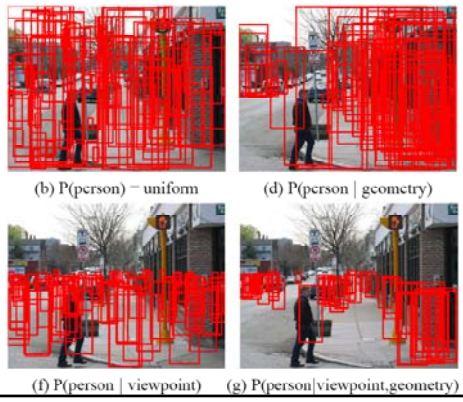
Sliding window approaches

- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000

- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993

Source: Fei-Fei, Fergus, & Torralba

Context can constrain a sliding window search



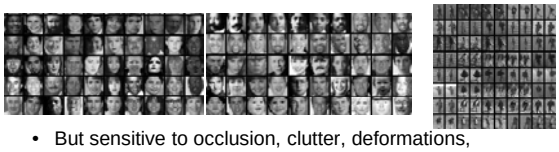
(b) P(person) ~ uniform (d) P(person | geometry)

(f) P(person | viewpoint) (g) P(person|viewpoint,geometry)

Hoiem, Efros, Herbert, 2006

Global appearance-based

- Appropriate for classes with more rigid structure, and when good training examples available



- But sensitive to occlusion, clutter, deformations, larger variability within the class.

Genres of approaches

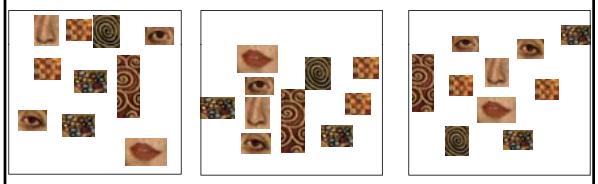
- Alignment
  - Pose clustering with object instances
- Global appearance
  - With or without a sliding window
- Local features
  - Indexing
  - Part-based models
    - Constellation models
    - Voting
  - Bags of words models

Local feature-based: indexing

Match examples by searching for similar local features within a database.

Local feature-based: bag of words models

- Remove spatial information, treat object as a collection of local appearance regions.



Local feature-based: constellation models

Visual Object Recognition Tutorial

- In categorization problem, we no longer have exact correspondences…
- On a local level, we can still detect similar parts.
- Represent objects by their parts
  ⇒ Bag-of-features
- How can we improve on this?
  - Encode structure

T. Tuytelaars, B. Leibe

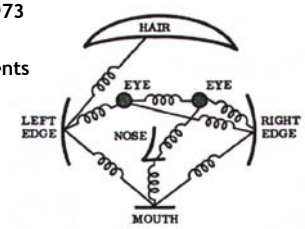
Slide credit: Rob Fergus

52

Local feature-based: constellation models

Visual Object Recognition Tutorial

- Fischler & Elschlager 1973
- Model has two components
  - parts (2D image fragments)
  - structure (configuration of parts)



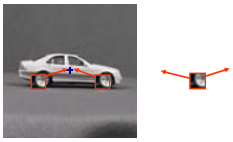
K. Grauman, B. Leibe

53

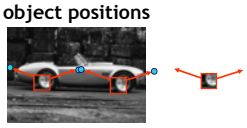
Local feature-based: voting

Visual Object Recognition Tutorial

- For every feature, store possible "occurrences"



- For new image, let the matched features vote for possible object positions



K. Grauman, B. Leibe

54

Local feature-based: voting

- **Backproject for segmentation estimate**

Leibe et al. 2004 Implicit Shape Model

Visual Object Recognition Tutorial

What "works" today

- Reading license plates, zip codes, checks

Source: Lana Lazebnik

What "works" today

- Reading license plates, zip codes, checks
- Fingerprint recognition

Source: Lana Lazebnik

What "works" today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

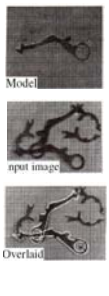
Source: Lana Lazebnik

What "works" today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)

Source: Lana Lazebnik

Rough evolution of focus in recognition research

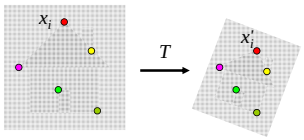


1980s | 1990s to early 2000s | Currently

Visual Object Recognition Tutorial

Recall: Alignment

- Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images
- We can use this idea to recognize / verify **instances** of an object.

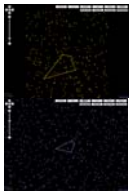


Find transformation T that minimizes

$$\sum_i \text{residual}(T(x_i), x_i')$$

Recall: Alignment

- Alignment: fitting a model to a transformation between pairs of features (*matches*) in two images
- We can use this idea to recognize / verify **instances** of an object.
- First we'll look at an interesting application doing this with simple image features, then for objects.



CCPP

Astrometry.net

Making the Sky Searchable: Fast Geometric Hashing for Automated Astrometry

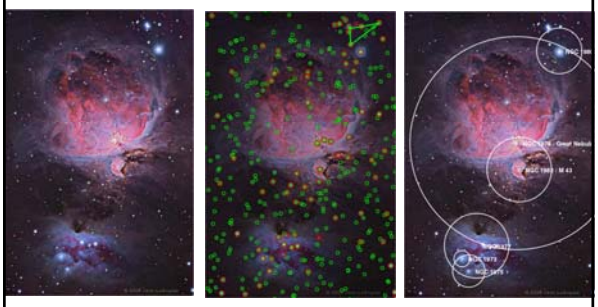
Sam Roweis, Dustin Lang & Keir Mierle
University of Toronto

David Hogg & Michael Blanton
New York University

SLIDES:
http://cosmo.nyu.edu/hogg/research/2006/09/28/astrometry_google.ppt

http://astrometry.net roweis@cs.toronto.edu

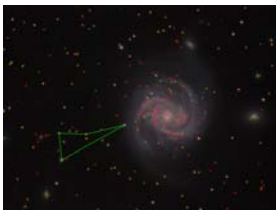
Example



A shot of the Great Nebula, by Jerry Lodriguss (c.2006), from astropix.com
http://astrometry.net/gallery.html

Roweis et al.

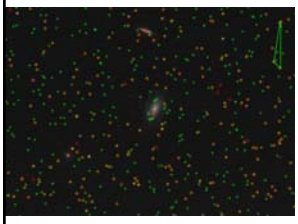
Example

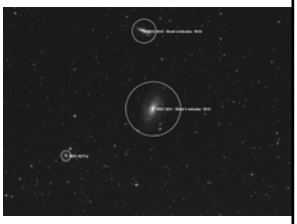


An amateur shot of M100, by Filippo Ciferri (c.2007) from flickr.com
http://astrometry.net/gallery.html

Roweis et al.

Example





A beautiful image of Bode's nebula (c.2007) by Peter Bresseler, from starlightfriend.de
http://astrometry.net/gallery.html

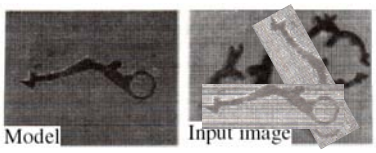
Roweis et al.

Model-based recognition

- Which image features correspond to which features on which object model in the "modelbase"?

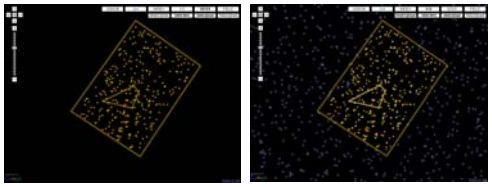
Hypothesize and test: main idea

- Given model of object
- New image: hypothesize object identity and pose
- Render object in camera
- Compare rendering to actual image: if close, good hypothesis.



Hypothesize and test: main idea

- Given model of object
- New image: hypothesize object identity and pose
- Render object in camera
- Compare rendering to actual image: if close, good hypothesis.



How to form a hypothesis?

Given a particular model object, we can estimate the *correspondences* between image and model features

Use correspondence to estimate model pose relative to object coordinate frame

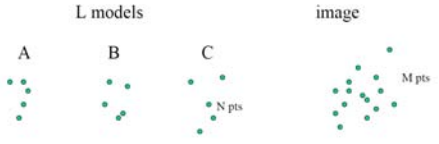
Generating hypotheses

We want a good correspondence between model features and image features.

- Brute force?

Brute force hypothesis generation

- For every possible model, try every possible subset of image points as matches for that model's points.
- Say we have L objects with N features, M features in image



Generating hypotheses

We want a good correspondence between model features and image features.

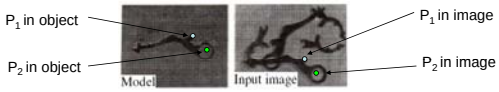
- Brute force?
- **Pose consistency**, alignment: use subsets of features to estimate larger correspondence
- **Voting**, pose clustering

Pose consistency / alignment

- Key idea:
  - If we find good correspondences for a small set of features, it is easy to obtain correspondences for a much larger set.
- Strategy:
  - Generate hypotheses using small numbers of correspondences
  - Backproject: transform *all* model features to image features
  - Verify

Example: 2d affine mappings

- Say camera is looking down perpendicularly on planar surface



- We have two coordinate systems (object and image), and they are related by some affine mapping (rotation, scale, translation, shear).

Example: 2d affine mappings

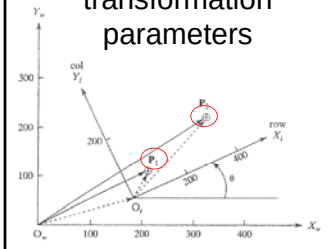
In non-homogenous coordinates

$$\underbrace{\begin{bmatrix} u \\ v \end{bmatrix}}_{\text{[image point]}} = \underbrace{\begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix}}_{\text{[scale, rotation, shear]}} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{[model point]}} + \underbrace{\begin{bmatrix} t_x \\ t_y \end{bmatrix}}_{\text{[translation]}}$$

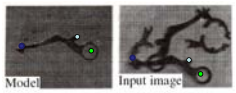
In homogenous coordinates

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}}_{\text{[translation, scale, rotation, shear]}} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Solving for the transformation parameters



$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$



$\mathbf{P}_1^{(model)} = [200,100]$ $\mathbf{P}_1^{(image)} = [100,60]$

$\mathbf{P}_2^{(model)} = [300,200]$ $\mathbf{P}_2^{(image)} = [380,120]$

$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ \cdots & & & & & \\ \cdots & & & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u \\ v \\ \vdots \end{bmatrix}$$

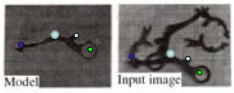
Rewrite in terms of unknown parameters

Similar ideas for camera models (3d->2d)

Alignment: backprojection

- Having solved for this transformation from some number of detected matches (3+ here), can compute (hypothesized) location of any *other* model points in the image space.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$



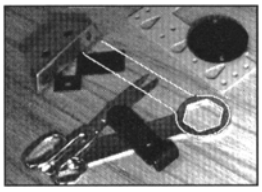
Alignment: verification

- Given the back-projected model in the image:
  - Check if image edges coincide with predicted model edges
  - May be more robust if also require edges to have the same orientation
  - Consider texture in corresponding regions
- Possible issues?

Alignment: verification

Figure from "Object recognition using alignment," D.P. Huttenlocher and S. Ullman, Proc. Int. Conf. Computer Vision, 1986, copyright IEEE, 1986

Alignment: verification



Issue with hypothesis & test approach

- May have false matches
  - We want *reliable* features to form the matches
    - **Local invariant features** useful to find matches, and to verify hypothesis
- May be too many hypotheses to consider
  - We want to look at the *most likely* hypotheses first
    - **Pose clustering (voting):** Narrow down number of hypotheses to verify by letting features *vote* on model parameters.

Pose clustering (voting)

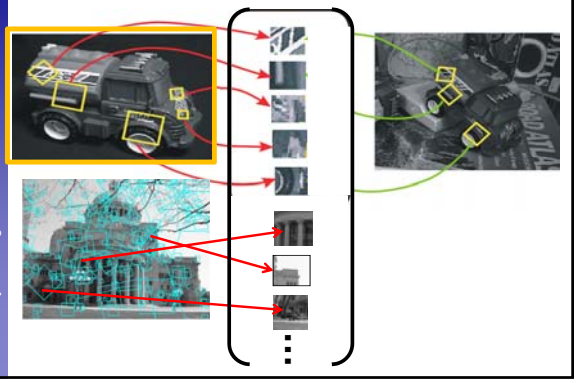
- Narrow down the number of hypotheses to verify: identify those model poses that a lot of features agree on.
  - Use each group's correspondence to estimate pose
  - Vote for that object pose in accumulator array (one array per object if we have multiple models)
- Local invariant features can give more reliable matches and means of verification

Pose clustering and verification with SIFT [Lowe]

To detect **instances** of objects from a model base:

1) Index descriptors (distinctive features narrow possible matches)

Indexing local features



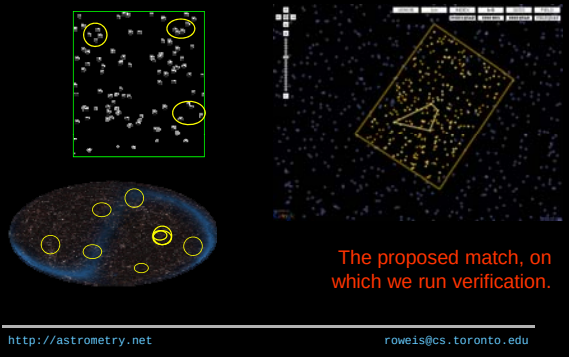
Visual Object Recognition Tutorial

Pose clustering and verification with SIFT [Lowe]

To detect **instances** of objects from a model base:

1) Index descriptors (distinctive features narrow possible matches)
2) Generalized Hough transform to vote for poses (keypoints have record of parameters relative to model coordinate system)
3) Affine fit to check for agreement between model and image features (approximates perspective projection for planar objects)

Matching a test image



The proposed match, on which we run verification.

http://astrometry.net roweis@cs.toronto.edu

Planar objects

Model images and their SIFT keypoints

Input image

Model keypoints that were used to recognize, get least squares solution.

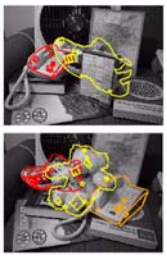
Recognition result

[Lowe]

3d objects

Background subtract for model boundaries

Objects recognized, though affine model not as accurate.



Recognition in spite of occlusion

[Lowe]

Recall: difficulties of voting

- Noise/clutter can lead to as many votes as true target
- Bin size for the accumulator array must be chosen carefully
- (Recall Hough Transform)

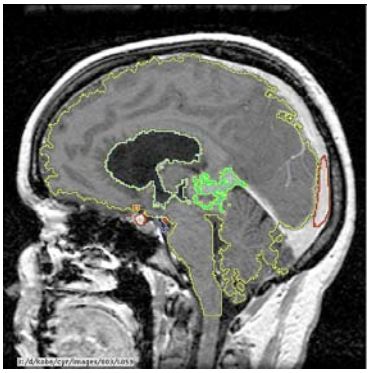
- In practice, good idea to make broad bins and spread votes to nearby bins, since verification stage can prune bad vote peaks.

Application: Surgery

- To minimize damage by operation planning
- To reduce number of operations by planning surgery
- To remove only affected tissue
- Problem
  - ensure that the model with the operations planned on it and the information about the affected tissue lines up with the patient
  - display model information superimposed on view of patient
  - **Big Issue**: coordinate alignment, as above

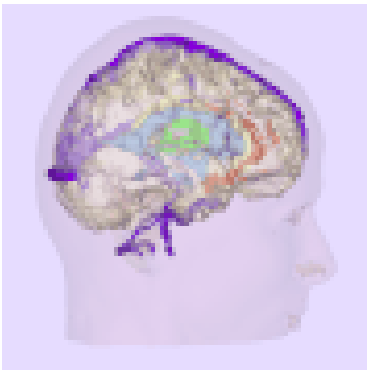
Computer Vision - A Modern Approach
Set: Model-based Vision
Slide by D.A. Forsyth

Segmentation used to break single MRI slice into regions.

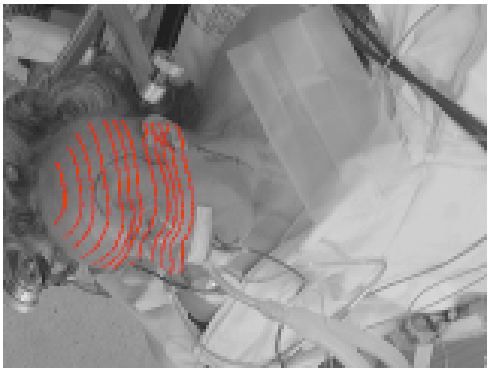


Figures by kind permission of Eric Grimson;
http://www.ai.mit.edu/people/welg/welg.html.

Regions assembled into 3d model

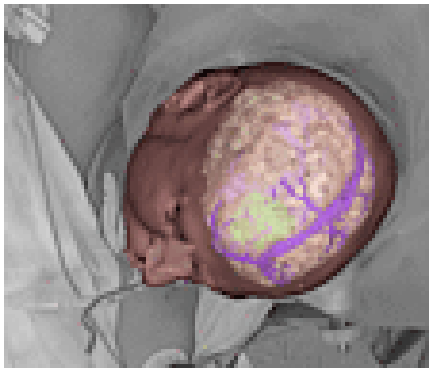


Figures by kind permission of Eric Grimson;
http://www.ai.mit.edu/people/welg/welg.html.



Figures by kind permission of Eric Grimson;
http://www.ai.mit.edu/people/welg/welg.html.

Patient with model superimposed. Note that view of model is registered to patient's pose here.



Figures by kind permission of Eric Grimson;
http://www.ai.mit.edu/people/welg/welg.html.

Figures by kind permission of Eric Grimson;
http://www.ai.mit.edu/people/welg/welg.html.

Summary

- Recognition by alignment: looking for object and pose that fits well with image
  - Use good correspondences to designate hypotheses
    - Invariant local features offer more reliable matches
  - Fast lookup with inverted file (sky app)
  - Limit verifications performed by voting (SIFT app)
- Alignment approach to recognition can be effective if we find reliable features within clutter, but does not scale well with the number of models, and is intended for specific instances of objects (vs. categorization).

Next

- Global appearance models
- Read F&P Chapter 22.1-22.3, 22.5