Tracking wrapup
Course recap

Tuesday, Dec 1

Announcements

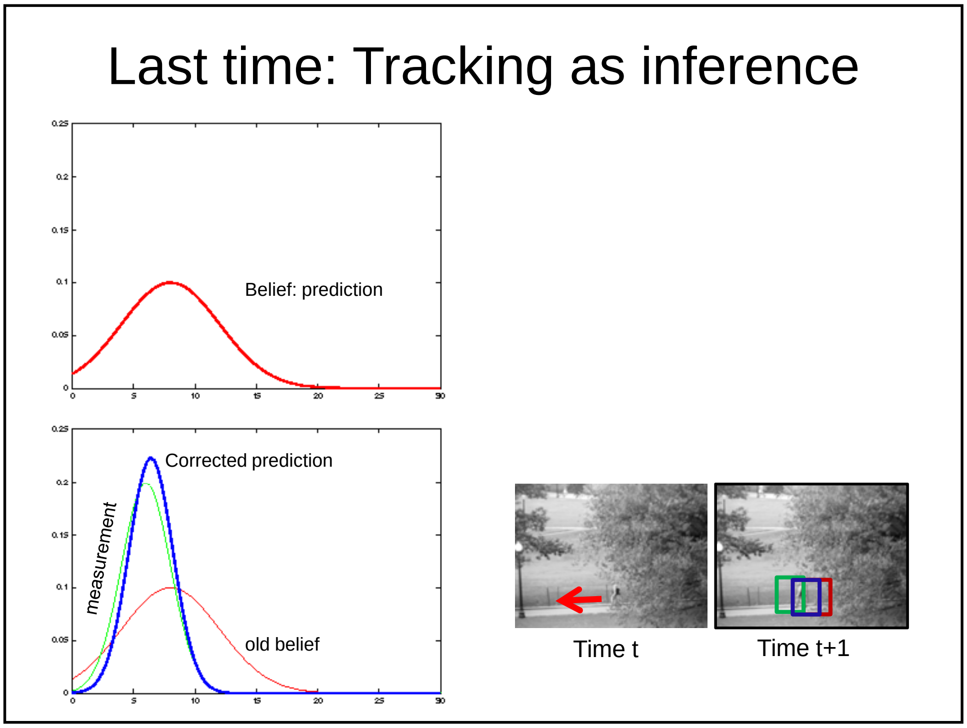
- Pset 4 grades and solutions available today
- Reminder: Pset 5 due 12/4, extended to 12/8 if needed
  - Choose between Section I (short answers) and II (program)
  - Extra credit only given for Section III
- Final exam is 12/14 Monday
  - Today's handout has example final exams
- Thursday in class: exam review

Previously

- Tracking as inference
  - Goal: estimate posterior of object position given measurement
- Linear models of dynamics
  - Represent state evolution and measurement models
- Kalman filters
  - Recursive prediction/correction updates to refine measurement
- General tracking challenges

Last time: Tracking as inference

- The *hidden state* consists of the true parameters we care about, denoted *X*.
- The *measurement* is our noisy observation that results from the underlying state, denoted *Y*.
- At each time step, state changes (from X_{t-1} to X_t) and we get a new observation Y_t.
- Our goal: recover most likely state X_t given
  - All observations seen so far.
  - Knowledge about dynamics of state transitions.



Last time: Tracking as inference

Time t

Time t+1

Last time: Linear dynamic model

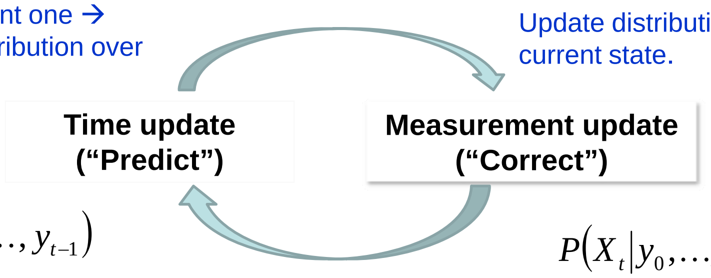
- Describe the *a priori* knowledge about
  - System dynamics model: represents evolution of state over time, with noise.

$$\mathbf{x}_t \sim N(\mathbf{D}\mathbf{x}_{t-1}; \mathbf{\Sigma}_d)$$

  - Measurement model: at every time step we get a noisy measurement of the state.

$$\mathbf{y}_t \sim N(\mathbf{M}\mathbf{x}_t; \mathbf{\Sigma}_m)$$

Last time: Kalman filter



Know corrected state from previous time step, and all measurements up to the current one → Predict distribution over next state.

Receive measurement

Know prediction of state, and next measurement → Update distribution over current state.

Time update (“Predict”)

Measurement update (“Correct”)

$P\left(X_t \middle| y_0, \ldots, y_{t-1}\right)$

Time advances: t++

$P\left(X_t \middle| y_0, \ldots, y_t\right)$

Mean and std. dev. of predicted state:

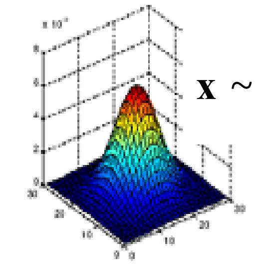
μ_t^-, σ_t^-

Mean and std. dev. of corrected state:

μ_t^+, σ_t^+

Kalman filter: pros and cons

- Gaussian densities, linear dynamic model:
  - + Simple updates, compact and efficient
  - – But, restricted class of motions defined by linear model
  - – *Unimodal* distribution = only single hypothesis



$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

When is a single hypothesis too limiting?

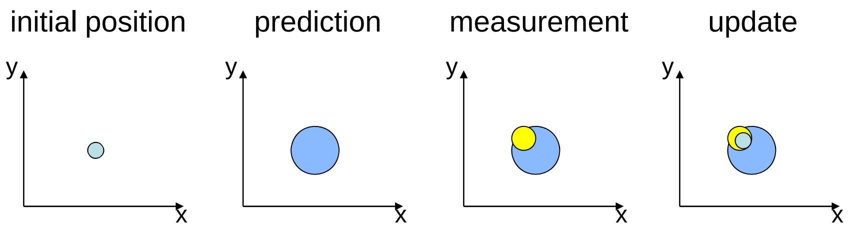
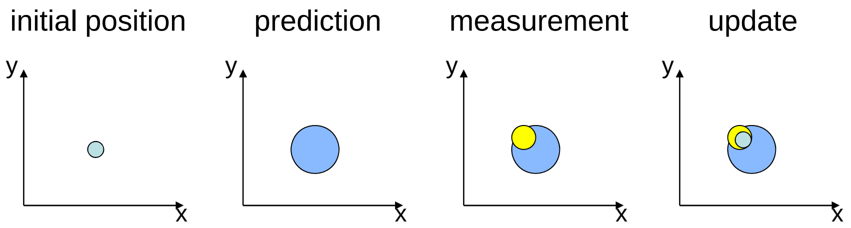


Figure from Thrun & Kosecka

When is a single hypothesis too limiting?

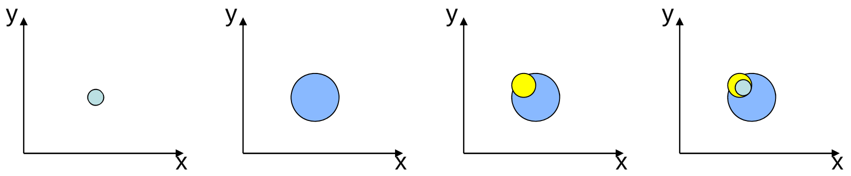


Consider this example: say we are tracking the face on the right using a skin color blob to get our measurement.

Video from Jojic & Frey

When is a single hypothesis too limiting?

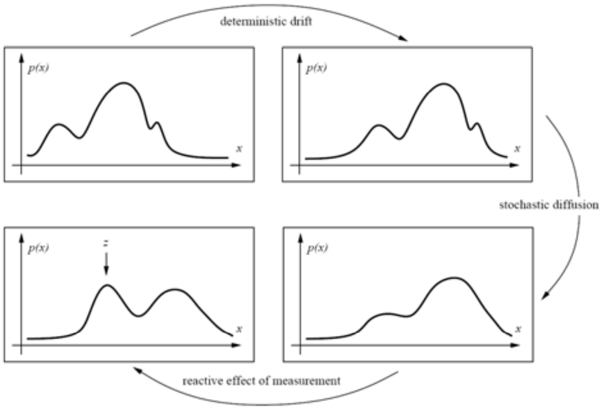
initial position prediction measurement update

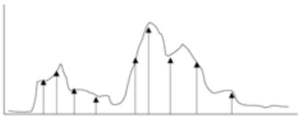


Consider this example: say we are tracking the face on the right using a skin color blob to get our measurement.

Video from Jojic & Frey

Alternative: particle-filtering and non-Gaussian densities

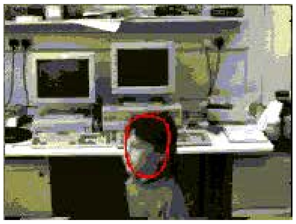


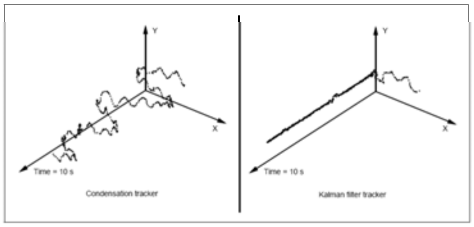


- Can represent distribution with set of weighted samples ("*particles*")
- Allows us to maintain **multiple hypotheses.**

For details: *CONDENSATION -- conditional density propagation for visual tracking*, by Michael Isard and Andrew Blake, Int. J. Computer Vision, 29, 1, 5--28, (1998)

Alternative: particle-filtering and non-Gaussian densities



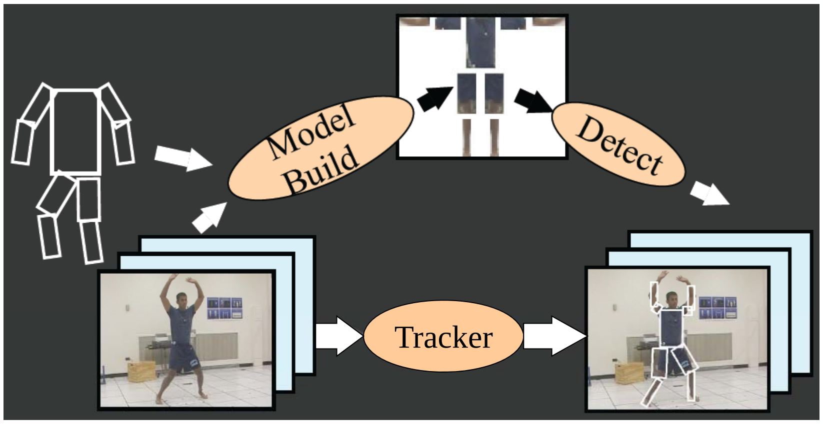


Monitor is a distractor, multiple hypotheses necessary.

Kalman filter fails once it starts tracking the monitor.

http://www.robots.ox.ac.uk/~vdg/dynamics.html
Visual Dynamics Group, Dept. Engineering Science, University of Oxford, 1998

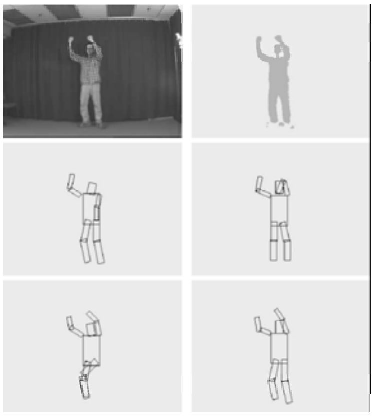
Tracking people by learning their appearance

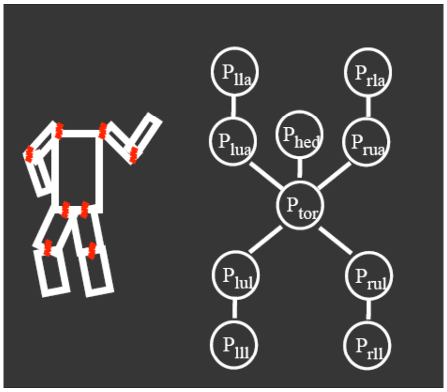


D. Ramanan, D. Forsyth, and A. Zisserman. Tracking People by Learning their Appearance. PAMI 2007.

Source: Lana Lazebnik

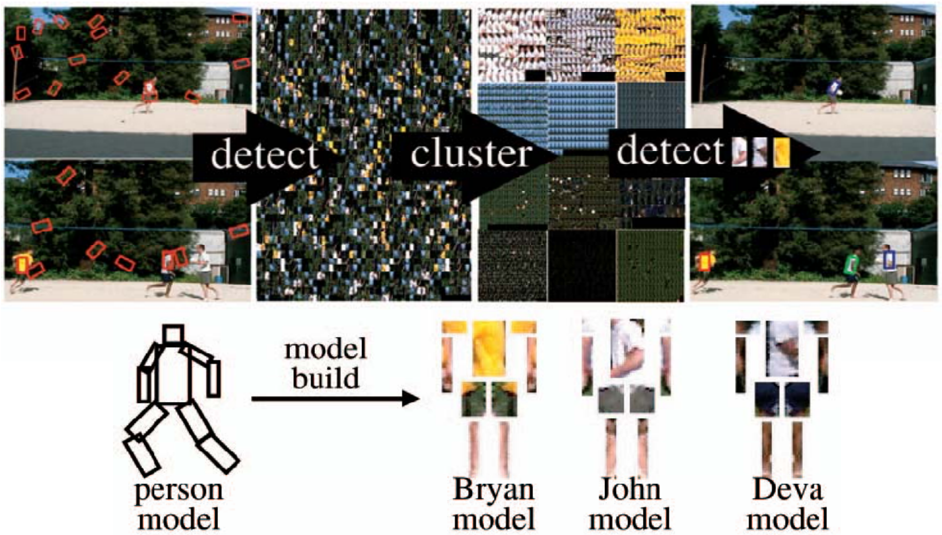
Tracking people by learning their appearance





Use a **part-based model** to encode part appearance + relative geometry.

Bottom-up initialization: Clustering



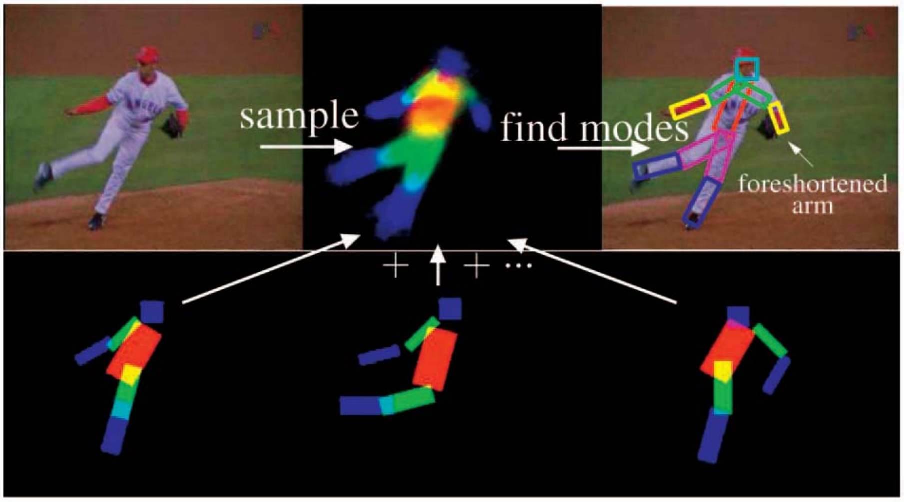
D. Ramanan, D. Forsyth, and A. Zisserman. Tracking People by Learning their Appearance. PAMI 2007.

Source: Lana Lazebnik

Top-down initialization: Exploit "easy" poses

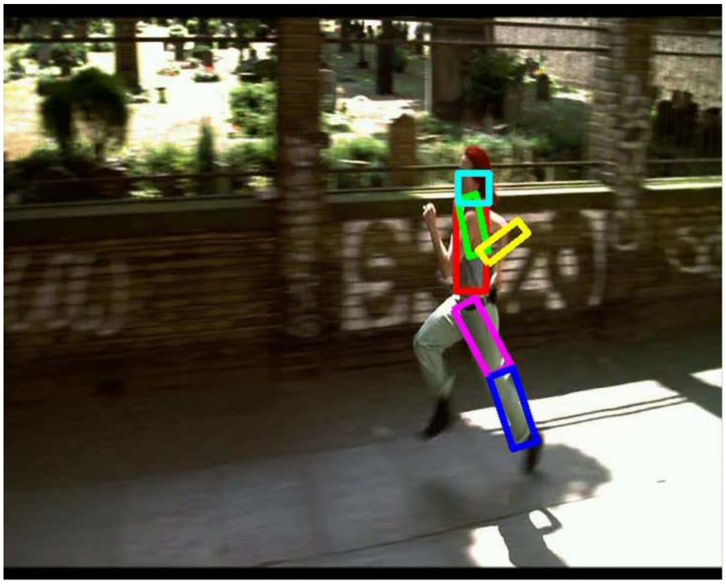
D. Ramanan, D. Forsyth, and A. Zisserman. Tracking People by Learning their Appearance. PAMI 2007.

Tracking by model detection



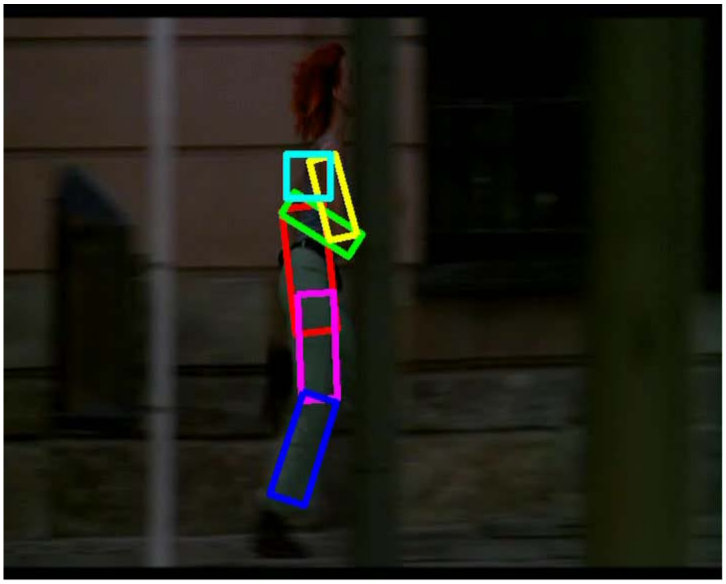
D. Ramanan, D. Forsyth, and A. Zisserman. Tracking People by Learning their Appearance. PAMI 2007.

Example results



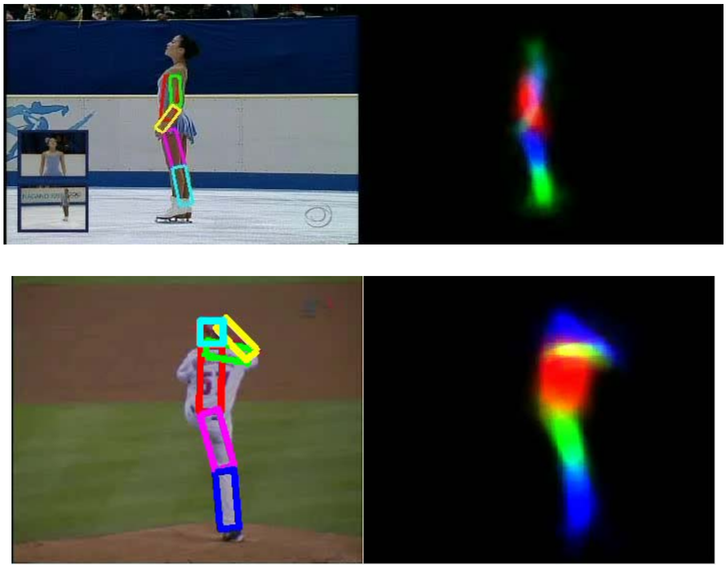
http://www.ics.uci.edu/~dramanan/papers/pose/index.html

Example results



Example results

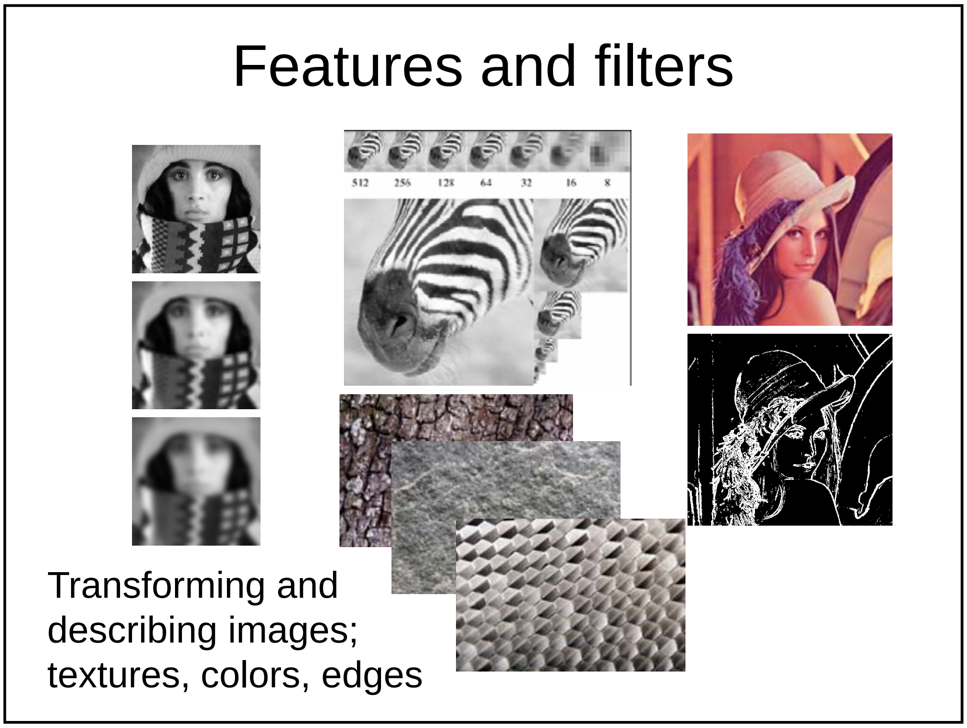
Example results



Tracking : summary

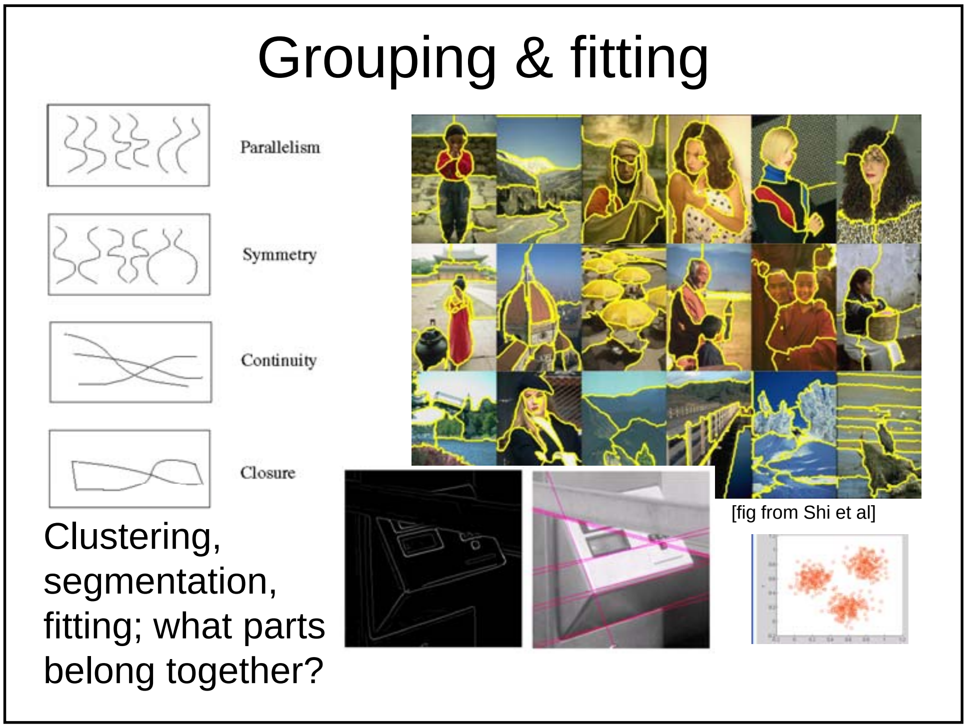
- Tracking as inference
  - Goal: estimate posterior of object position given measurement
- Linear models of dynamics
  - Represent state evolution and measurement models
- Kalman filters
  - Recursive prediction/correction updates to refine measurement
  - Single hypothesis can be limiting
- General tracking challenges
- Tracking via detection one way to mitigate drift (though means losing out on prediction help).

Course recap



Features and filters

Transforming and describing images; textures, colors, edges



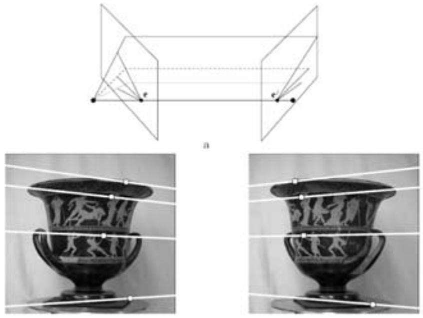
Grouping & fitting

[fig from Shi et al]

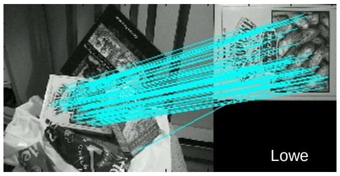
Clustering, segmentation, fitting; what parts belong together?

Multiple views

Multi-view geometry, matching, invariant features, stereo vision



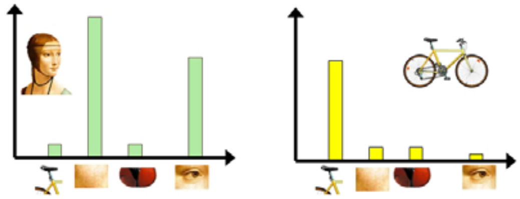
Hartley and Zisserman



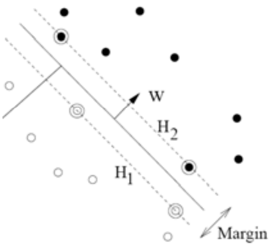
Lowe

Fei-Fei Li

Recognition and learning

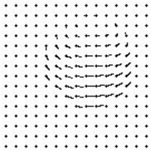


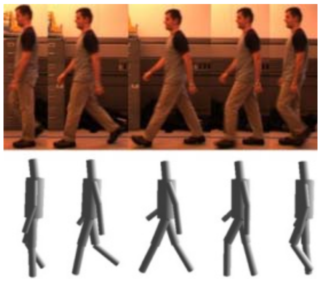
Recognizing objects and categories, learning techniques



Motion and tracking

Tracking objects, video analysis, low level motion, optical flow





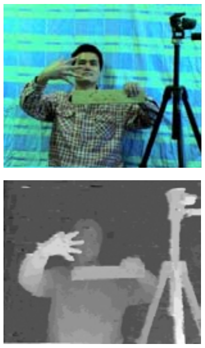
Tomas Izo

Computer Vision

- Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data *(measurement)*

1. Vision for measurement

Real-time stereo



Wang et al.

Structure from motion

Snavely et al.

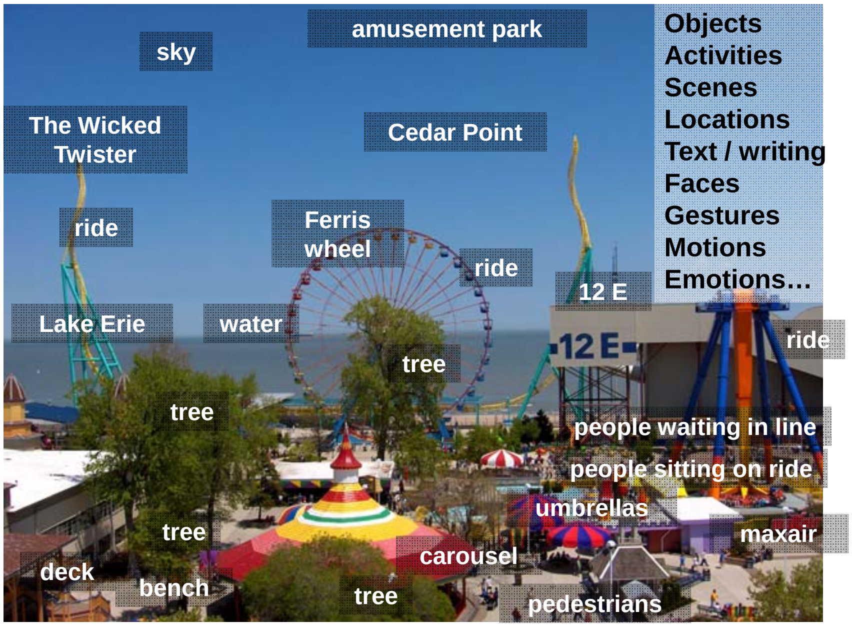
Tracking

Demirdjian et al.

Computer Vision

- Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data *(measurement)*
  2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. *(perception and interpretation)*

2. Vision for perception, interpretation

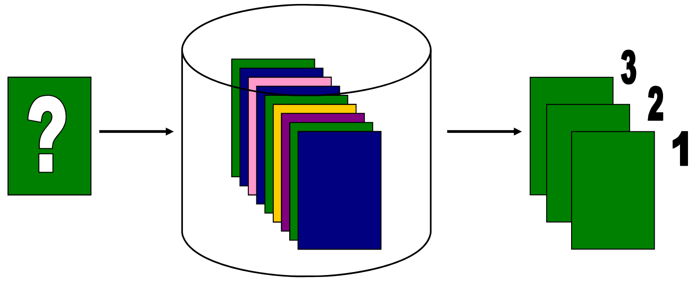


Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions…

Computer Vision

- Automatic understanding of images and video
    1. Computing properties of the 3D world from visual data *(measurement)*
    2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. *(perception and interpretation)*
    3. Algorithms to mine, search, and interact with visual data (*search and organization*)

3. Visual search, organization

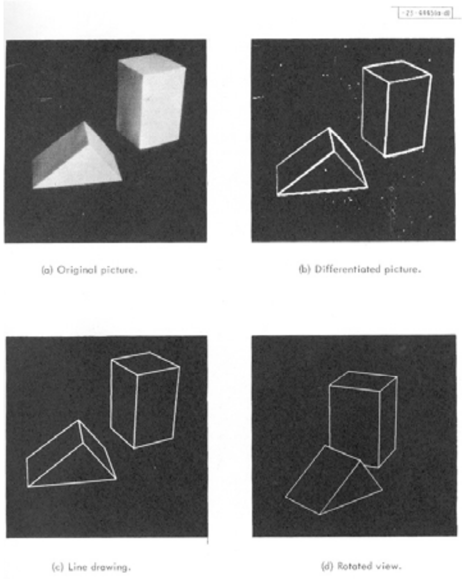


Query

Image or video archives

Relevant content

Visual data in 1963



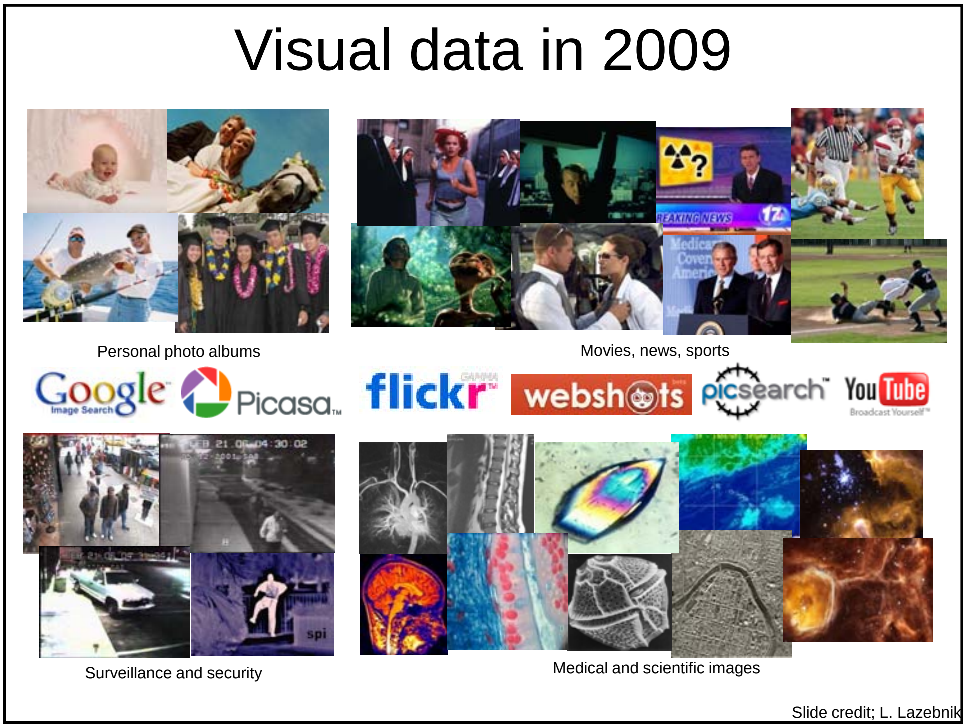
(a) Original picture.

(b) Differentiated picture.

(c) Line drawing.

(d) Rotated view.

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



Visual data in 2009

Personal photo albums

Movies, news, sports

Surveillance and security

Medical and scientific images

Slide credit; L. Lazebnik

Why vision?

- As image sources multiply, so do applications
  - – Relieve humans of boring, easy tasks
  - – Enhance human abilities
  - – Advance human-computer interaction, visualization
  - – Perception for robotics / autonomous agents
  - – Organize and give access to visual content

Faces and digital cameras

Camera waits for everyone to smile to take a photo [Canon]

Setting camera focus via face detection

Linking to info with a mobile device

Situated search
Yeh et al., MIT

MSR Lincoln

kooaba

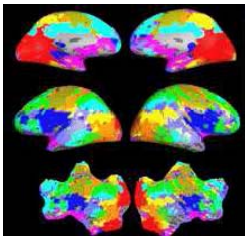
Video-based interfaces

Human joystick
NewsBreaker Live

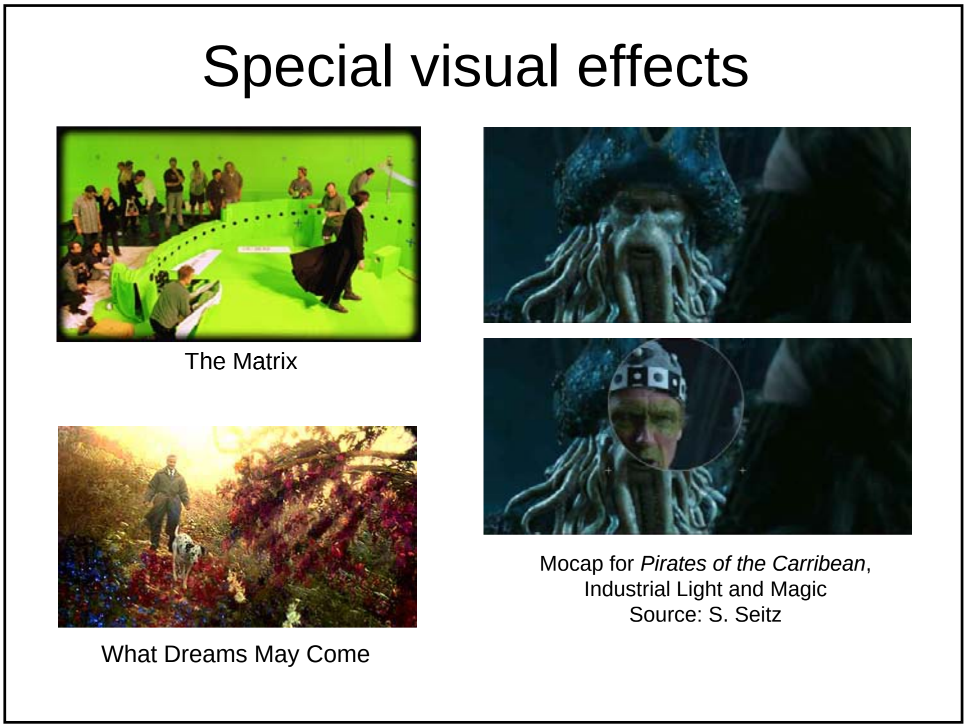
Assistive technology systems
Camera Mouse
Boston College

Vision for medical & neuroimages

Image guided surgery
MIT AI Vision Group



fMRI data
Golland et al.



Special visual effects

The Matrix

What Dreams May Come

Mocap for *Pirates of the Carribean*,
Industrial Light and Magic
Source: S. Seitz

Safety & security

Navigation,
driver safety

Monitoring pool
(Poseidon)

Pedestrian detection
MERL, Viola et al.

Surveillance