Computer Vision

The goal of computer vision

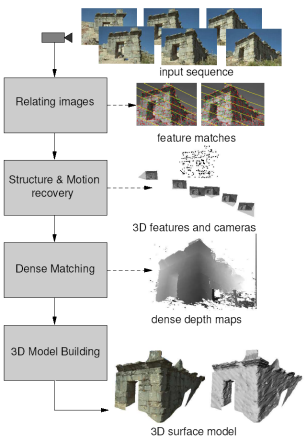
bet\

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What kind of information can we extract from an image?

- Metric 3D information
- Semantic information

Vision as measurement device



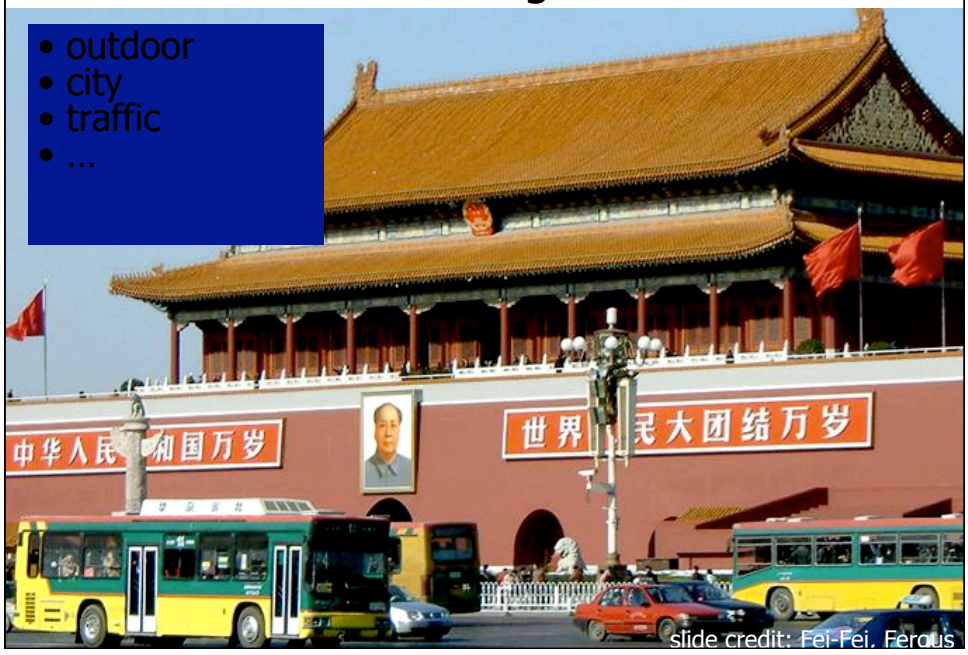
Vision as a source of semantic information

slide credit: Fei-Fei, Fergus

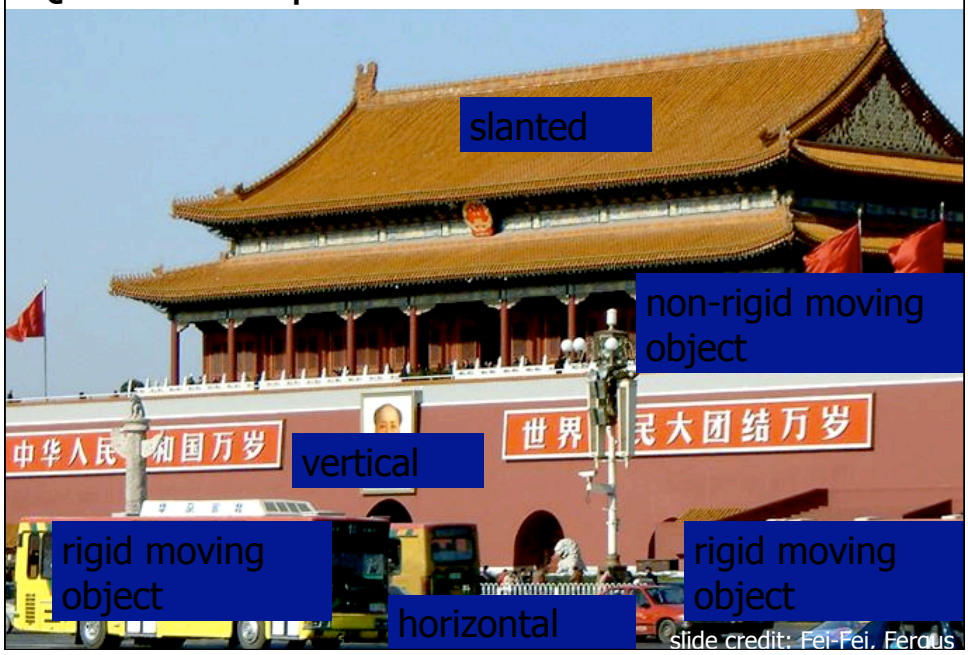
Object categorization

slide credit: Fei-Fei, Fergus

Scene and context categorization



Qualitative spatial information



Why study computer vision?

- Vision is useful: Images and video are everywhere!

Why study computer vision?

- Vision is useful
- Vision is interesting
- Vision is difficult
  - Half of primate cerebral cortex is devoted to visual processing

Why is computer vision difficult?
Challenges of recognizing objects in images

Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus

Challenges: illumination

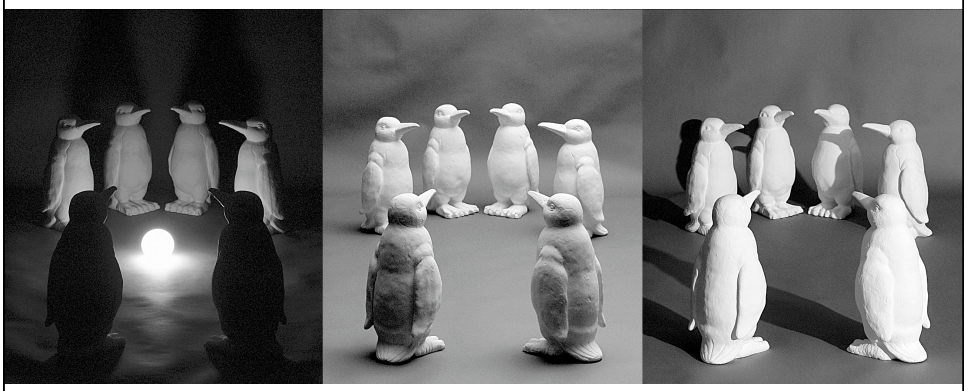
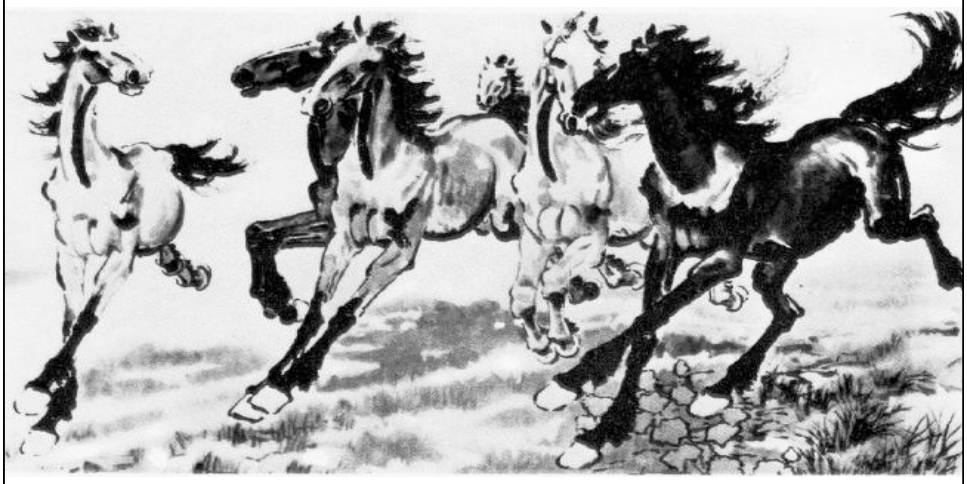


image credit: J. Koenderi

Challenges: scale

slide credit: Fei-Fei, Fergus

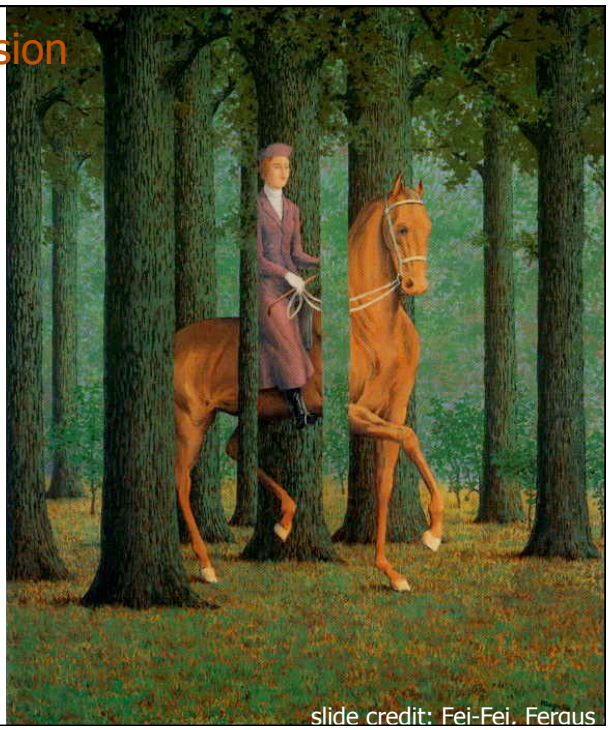
Challenges: deformation



Xu, Beihong 1943

slide credit: Fei-Fei, Fergus

Challenges: occlusion



Magritte, 1957

slide credit: Fei-Fei, Fergus

Challenges: background clutter

Challenges: Motion

Challenges: object intra-class variation






slide credit: Fei-Fei, Fergus

Challenges: local ambiguity

slide credit: Fei-Fei, Fergus

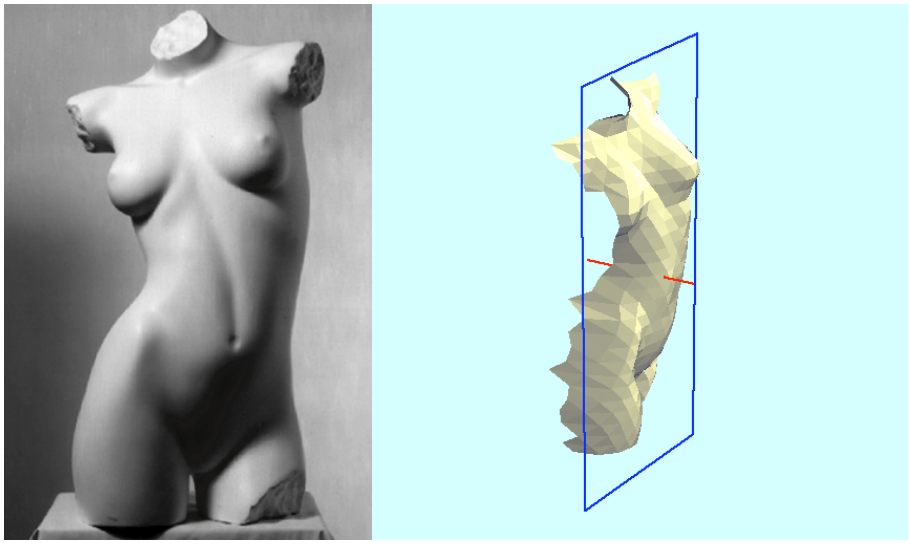
Challenges or opportunities?

- Images are confusing, but they also reveal the structure of the world through numerous cues
- Our job is to interpret the cues!

Depth cues: Linear perspective

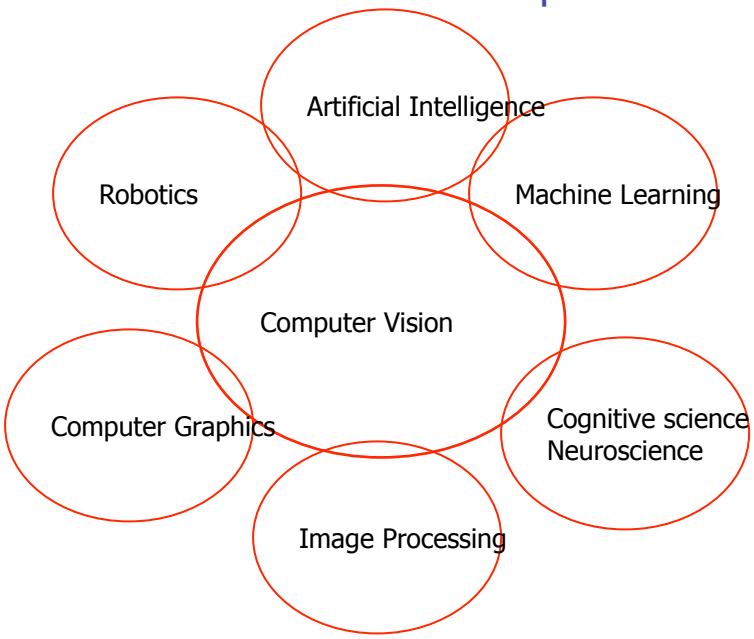
Depth cues: Aerial perspective

Shape and lighting cues: Shading

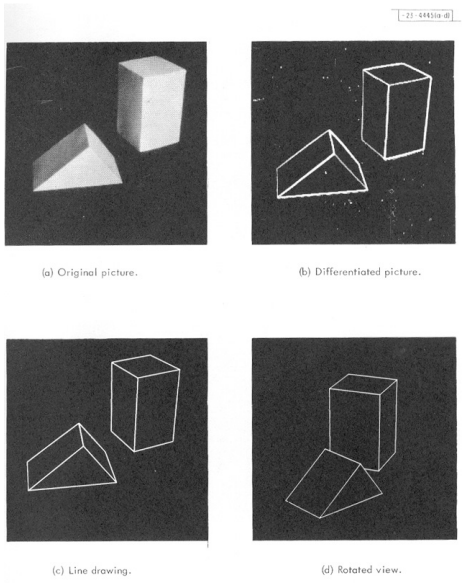


Grouping cues: "Common fate"

Connections to other disciplines

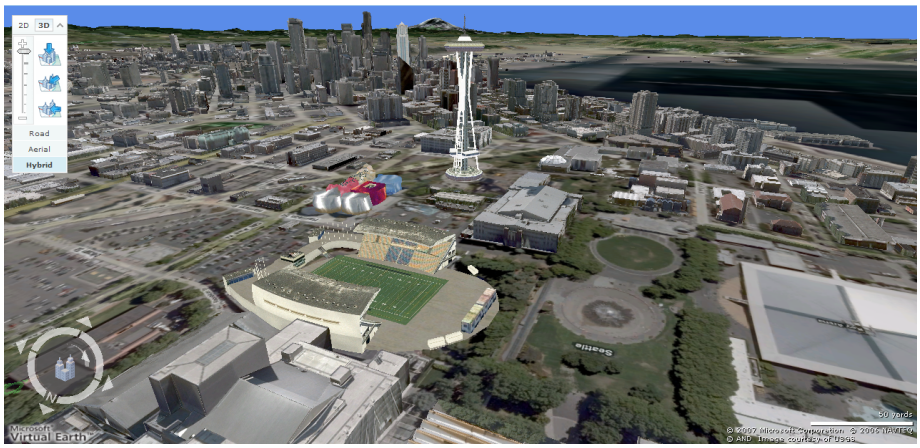


Origins of computer vision



Computer Vision in the Real World

3D urban modeling



Bing maps, Google Streetview

Source: S. Se

3D urban modeling: Microsoft Photosynth

http://labs.live.com/photosynth/

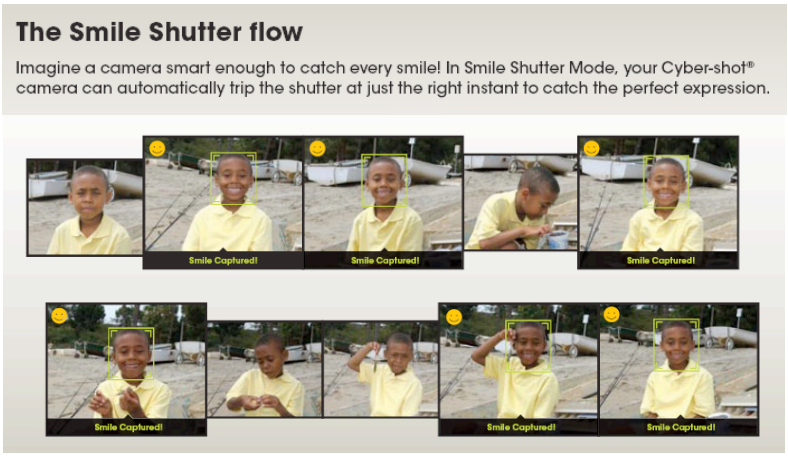
Source: S. Se

Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Source: S. Se

Smile detection



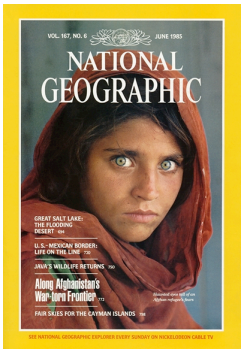
Sony Cyber-shot® T70 Digital Still Camera

Source: S. Se

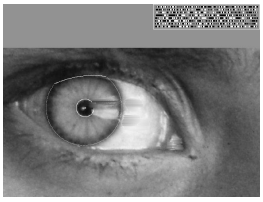
Face recognition: Apple iPhoto software

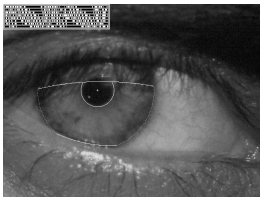
http://www.apple.com/ilife/iphoto/

Biometrics



How the Afghan Girl was Identified by Her Iris Patterns

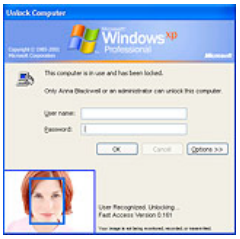




Source: S. Seitz

Biometrics

Fingerprint scanners on many new laptops, other devices



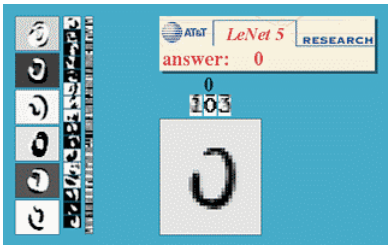
Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Source: S. Seitz

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software

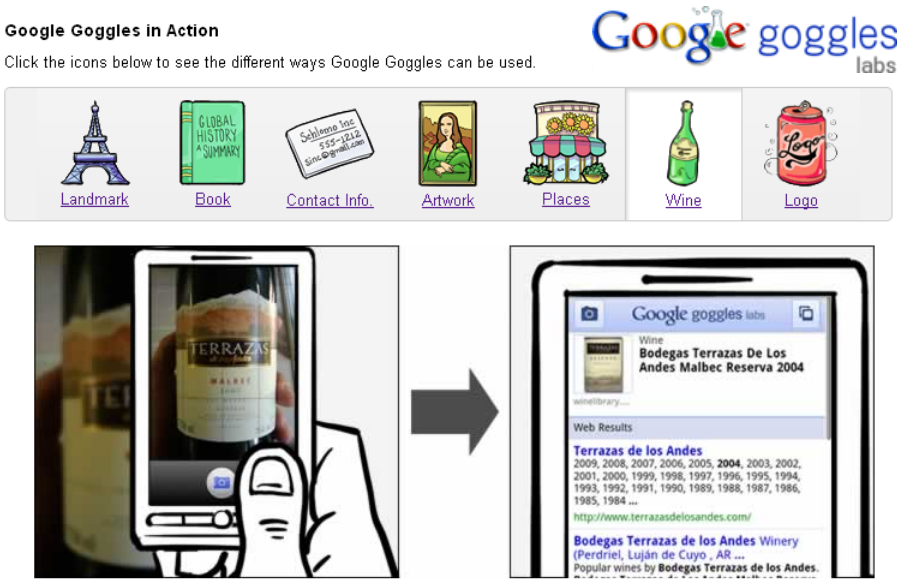


Digit recognition, AT&T labs

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

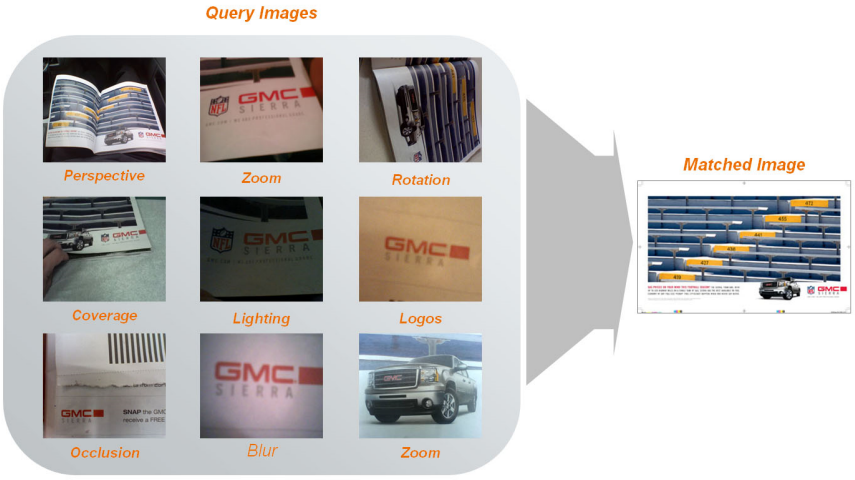
Source: S. Seitz

Mobile visual search: Google Goggles



Mobile visual search: iPhone Apps

snaptell kooaba



Automotive safety

- Mobileye: Vision systems in high-end BMW, GM, Volvo models
  - "In mid 2010 Mobileye will launch a world's first application of full emergency braking for collision mitigation for pedestrians where vision is the key technology for detecting pedestrians."

Source: A. Shashua, S. Seitz

Vision in supermarkets

LaneHawk by EvolutionRobotics

"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… "

Source: S. Seitz

Vision-based interaction (and games)

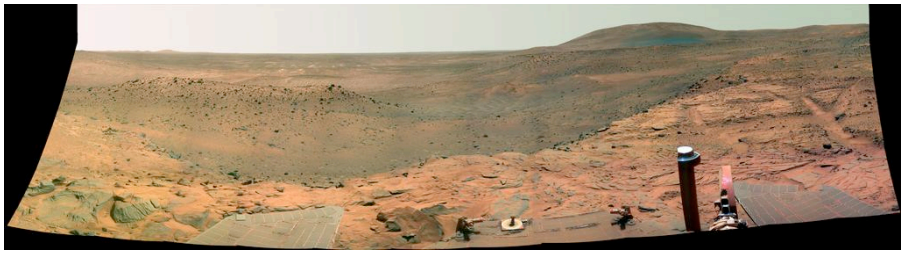
Nintendo Wii has camera-based IR tracking built in. See Lee's work at CMU on clever tricks on using it to create a multi-touch display!

Sony EyeToy

Assistive technologies

Source: S. Seitz

Vision for robotics, space exploration



NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

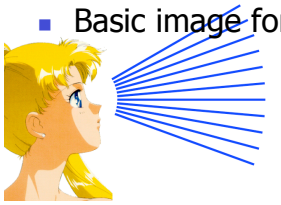
Vision systems (JPL) used for several tasks

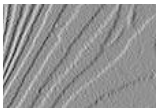
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

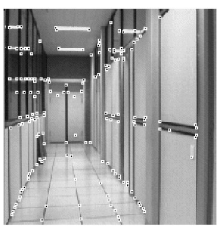
Source: S. Seitz

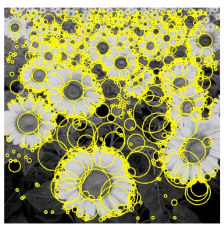
I. Early vision

- Basic image formation and processing



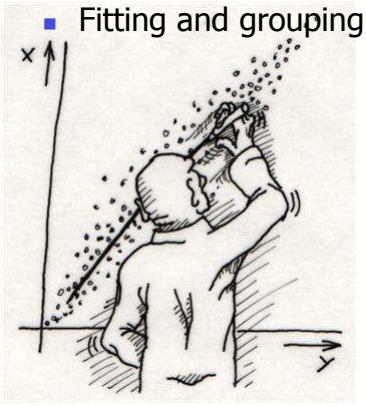




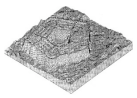


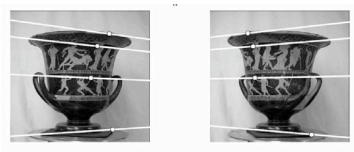
II. "Mid-level vision"

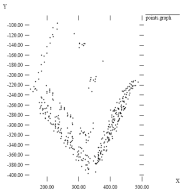
- Fitting and grouping

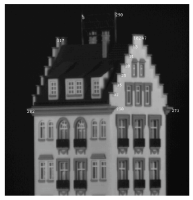


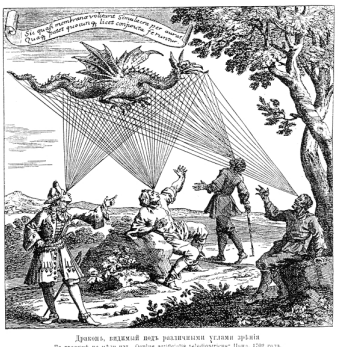
III. Multi-view geometry



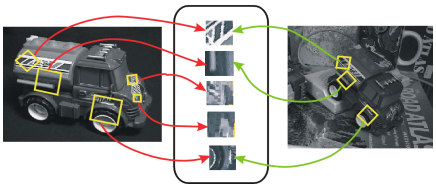


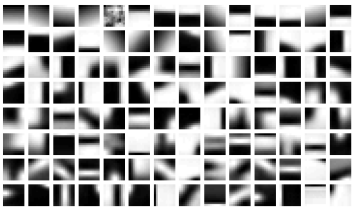


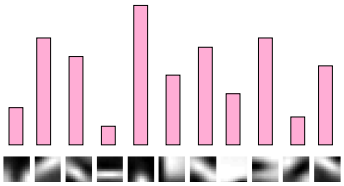


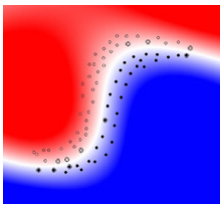


IV. Recognition

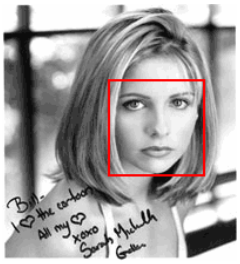


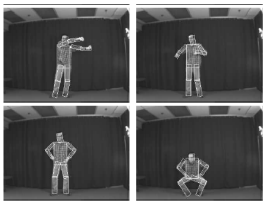


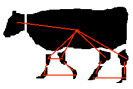




V. Advanced Topics







Semantic Labeling of Street Scenes

- Using image segmentation and scene and object recognition for automated image labeling

Role of Perception in Robotics

- Where am I relative to the world?
  - sensors: vision, stereo, range sensors, acoustics
  - problems: scene modeling/classification/recognition
  - integration: localization/mapping algorithms (e.g. SLAM)
- What is around me?
  - sensors: vision, stereo, range sensors, acoustics, sounds, smell
  - problems: object recognition, structure from x, qualitative modeling
  - integration: collision avoidance/navigation, learning

Jana Kosecka

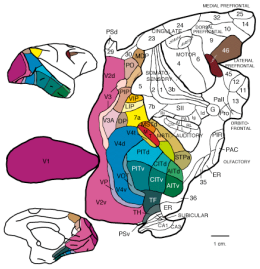
Role of Perception in Robotics

- How can I safely interact with environment (including people!)?
  - sensors: vision, range, haptics (force+tactile)
  - problems: structure/range estimation, modeling, tracking, materials, size, weight, inference
  - integration: navigation, manipulation, control, learning
- How can I solve "new" problems (generalization)?
  - sensors: vision, range, haptics, undefined new sensor
  - problems: categorization by function/shape/context/??
  - integrate: inference, navigation, manipulation, control, learning

Jana Kosecka

Challenges/Issues

- About 60% of our brain is devoted to vision
- We see immediately and can form and understand images instantly



- Detailed representations are often not necessary
- Different approaches in the past Animate Vision (biologically inspired), Purposive Vision (depending on the task/purpose)

Jana Kosecka

Visual Perception Topics

Techniques

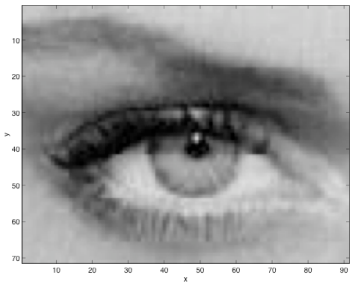
- Computational Stereo
- Feature detection and matching
- Motion tracking and visual feedback

Applications in Robotics:

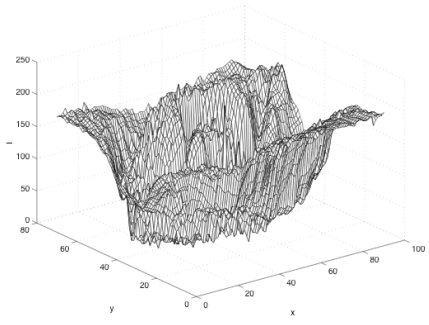
- range sensing, Obstacle detection, environment interaction
- Mapping, registration, localization, recognition
- Manipulation

Image - Apperance

Image



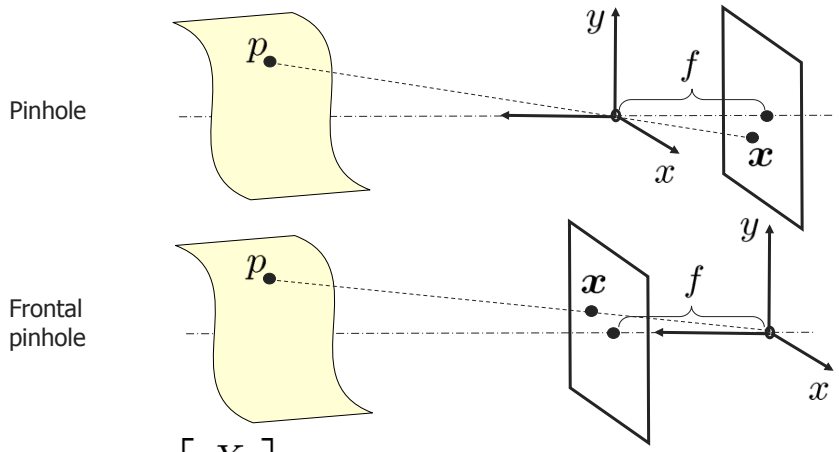
Brightness values



I(x,y)

Jana Kosecka

Image Formation



$$\boldsymbol{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \rightarrow \boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$$

J. Kosecka, GMU

Pinhole Camera Model

- Image coordinates are nonlinear function of world coordinates
- Relationship between coordinates in the camera frame and sensor plane

2-D coordinates $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$

Homogeneous coordinates

$$\boldsymbol{x} \rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix}, \quad \boldsymbol{X} \rightarrow \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

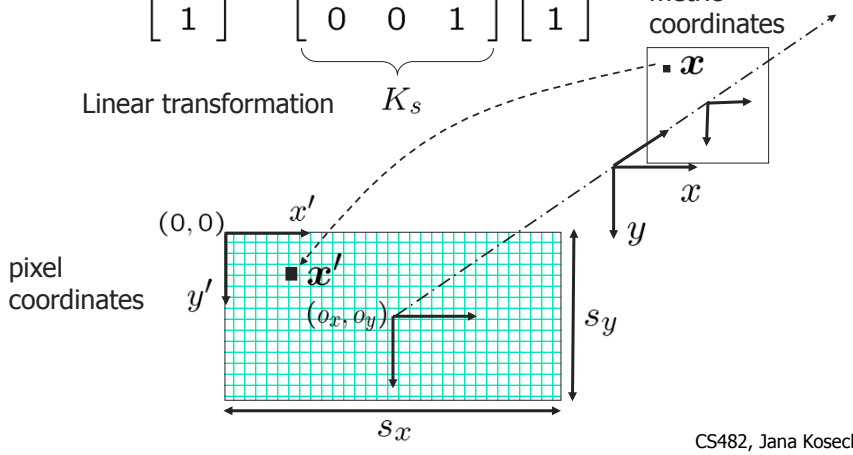
$$Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{K_f} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Pi_0} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Image Coordinates

- Relationship between coordinates in the sensor plane and image

$$\boldsymbol{x}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{K_s} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Linear transformation



CS482, Jana Kosecka

Calibration Matrix and Camera Model

- Relationship between coordinates in the world frame and image
- Intrinsic parameters

Pinhole camera

$$\lambda \boldsymbol{x} = K_f \Pi_0 \boldsymbol{X}$$

Pixel coordinates

$$\boldsymbol{x}' = K_s \boldsymbol{x}$$

- Adding transformation between camera coordinate systems and world coordinate system
- Extrinsic Parameters

$$\lambda \boldsymbol{x}' = \begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$\lambda \boldsymbol{x} = K_f \Pi_0 g \boldsymbol{X} = \Pi \boldsymbol{X}$$

Image of a Point

Homogeneous coordinates of a 3-D point p

$$\boldsymbol{X} = [X, Y, Z, W]^T \in \mathbb{R}^4, \quad (W = 1)$$

Homogeneous coordinates of its 2-D image

$$\boldsymbol{x} = [x, y, z]^T \in \mathbb{R}^3, \quad (z = 1)$$

Projection of a 3-D point to an image plane

$$\lambda \boldsymbol{x} = \Pi \boldsymbol{X}$$

$$\lambda \in \mathbb{R}, \ \Pi = [R, T] \in \mathbb{R}^{3\times 4}$$

$$\lambda \boldsymbol{x}' = \Pi \boldsymbol{X}$$

$$\lambda \in \mathbb{R}, \ \Pi = [KR, KT] \in \mathbb{R}^{3\times 4}$$

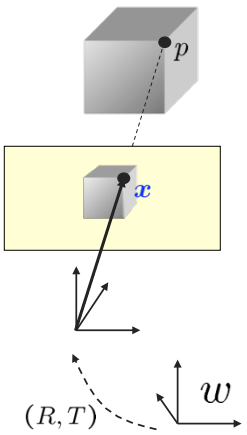


Image of a Line

Homogeneous representation of a 3-D line L

$$\boldsymbol{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{bmatrix} + \mu \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ 0 \end{bmatrix}, \quad \mu \in \mathbb{R}$$

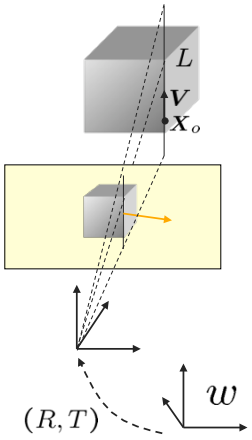
Homogeneous representation of its 2-D image

$$\boldsymbol{l} = [a, b, c]^T \in \mathbb{R}^3$$

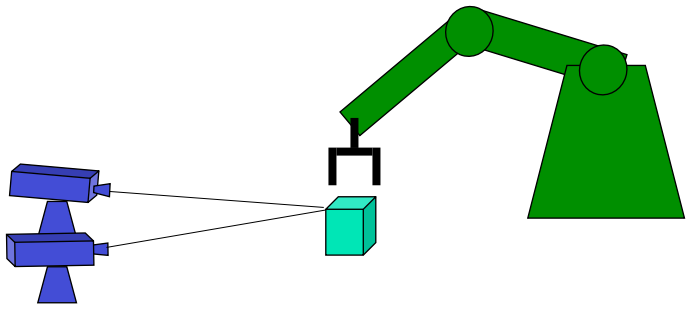
Projection of a 3-D line to an image plane

$$\boldsymbol{l}^T \boldsymbol{x} = \boldsymbol{l}^T \Pi \boldsymbol{X} = 0$$

$$\Pi = [KR, KT] \in \mathbb{R}^{3\times 4}$$



What is Computational Stereo?



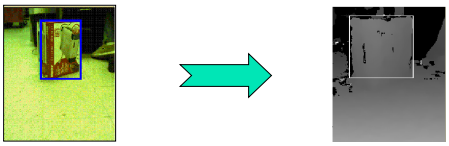
Viewing the same physical point from two different viewpoints allows depth from triangulation

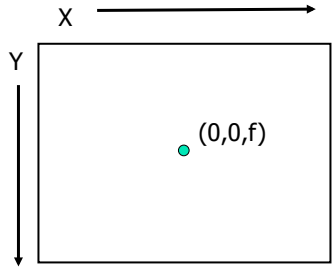
Computational Stereo

- Much of geometric vision is based on information from 2 (or more) camera locations
- Hard to recover 3D information from a single 2D image without extra knowledge
- Motion and stereo (multiple cameras) are both common in the world
- Stereo vision is ubiquitous in nature
  (oddly, nearly 10% of people are stereo blind)
- Stereo involves the following *three problems*:
  1. calibration
  2. matching (correspondence problem)
  3. reconstruction (reconstruction problem)

Binocular Stereo System: Geometry

- **GOAL:** Passive 2-camera system using triangulation to generate a depth map of a world scene.
- **Depth map:** z=f(x,y) where x,y are coordinates one of the image planes and z is the height above the respective image plane.
  - Note that for stereo systems which differ only by an offset in x, the v coordinates (projection of y) is the same in both images!
  - Note we must convert from image (pixel) coordinates to external coordinates -- **requires calibration**

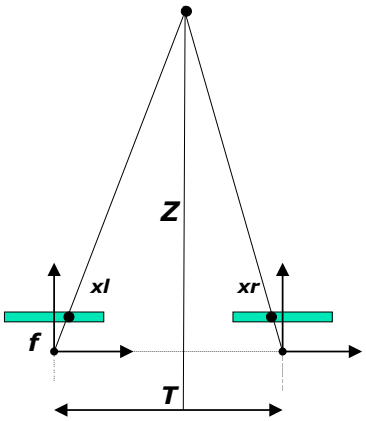




4 intrinsic parameters convert from pixel to metric values
s_x s_y c_x c_y

Stereo Configuration

- Images are scan-aligned
- Disparity between two images – inversely proportional to depth
- Disparity – difference between x-coordinates of a feature
- Triangle similarity



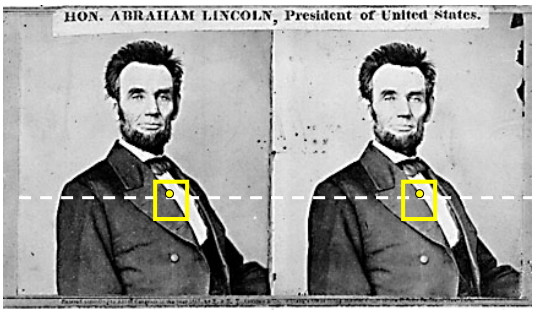
$$\frac{Z}{T} = \frac{Z-f}{T-\mathrm{x}_l-\mathrm{x}_r}$$

$$Z = \frac{fT}{\text{disparity}}$$

Stereo Vision

- Distance is inversely proportional to disparity
  - closer objects can be measured more accurately
- Disparity is proportional to baseline
  - For a given disparity error, the accuracy of the depth estimate increases with increasing baseline baseline
  - However, as baseline is increased, some objects may appear in one camera, but not in the other.
  - Image resolution is also a factor

Stereo Matching – Stereo Correspondence



For each epipolar line (scanline)

For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost
- This will never work, so:
- Match Windows

Region based Similarity Metric

- Sum of squared differences

$$SSD(h) = \sum_{\tilde{\mathbf{x}} \in W(\mathbf{x})} \|I_1(\tilde{\mathbf{x}}) - I_2(h(\tilde{\mathbf{x}})\|^2$$

- Normalize cross-correlation

$$NCC(h) = \frac{\sum_{W(\mathbf{x})}(I_1(\tilde{\mathbf{x}}) - \bar{I}_1)(I_2(h(\tilde{\mathbf{x}})) - \bar{I}_2))}{\sqrt{\sum_{W(\mathbf{x})}(I_1(\tilde{\mathbf{x}}) - \bar{I}_1)^2 \sum_{W(\mathbf{x})}(I_2(h(\tilde{\mathbf{x}})) - \bar{I}_2)^2)}}$$

- Sum of absolute differences

$$SAD(h) = \sum_{\tilde{\mathbf{x}} \in W(\mathbf{x})} |I_1(\tilde{\mathbf{x}}) - I_2(h(\tilde{\mathbf{x}}))|$$

66

Window size

W = 3 W = 20

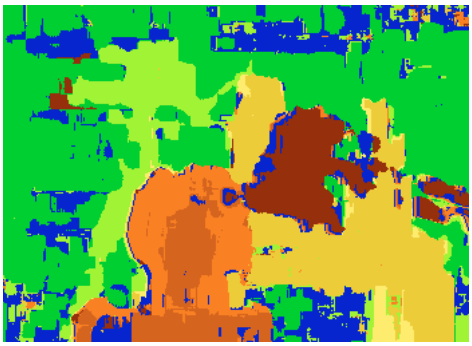
- Effect of window size

With adaptive window

- T. Kanade and M. Okutomi, *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*,, Proc. International Conference on Robotics and Automation, 1991.
- D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. International Journal of Computer Vision, 28(2): 155-174, July 1998

(S. Seitz) Jana Kosecka

Results with window correlation



Window-based matching
(best window size)

Ground truth

(slide courtesy S. Seitz)

Jana Kosecka

Results with better method

State of the art method

Boykov et al., Fast Approximate Energy Minimization via Graph Cuts, International Conference on Computer Vision, September 1999.

Ground truth

(slide courtesy S. Seitz)

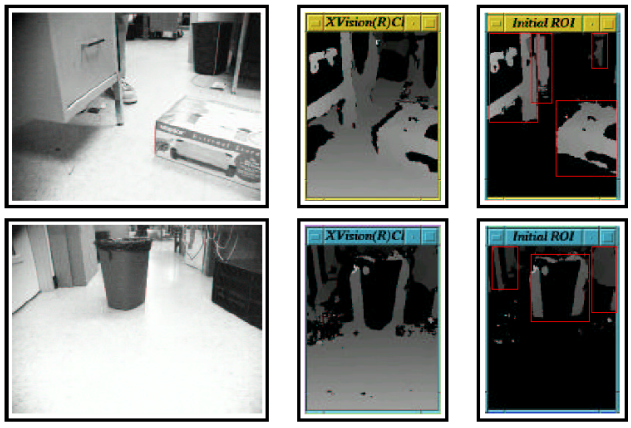
Jana Kosecka

Applications of Real-Time Stereo

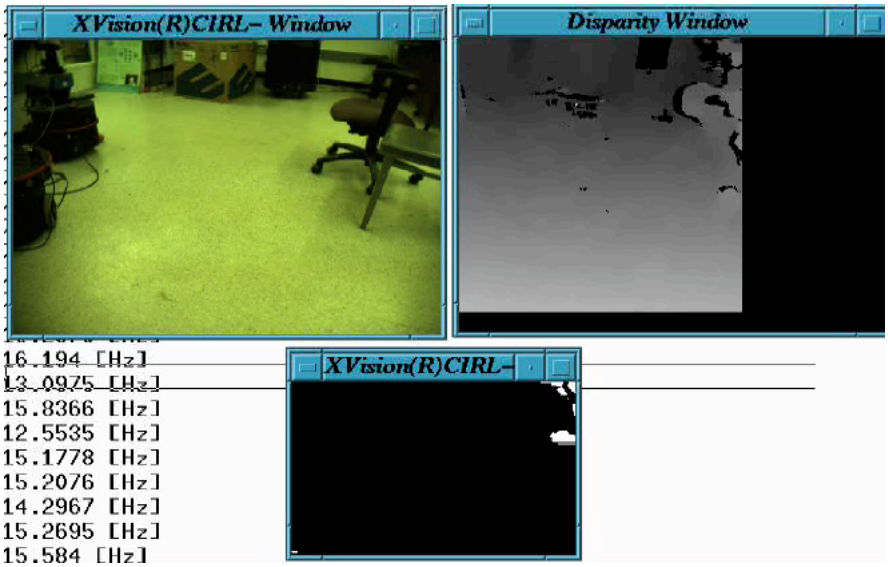
- Mobile robotics
  - Detect the structure of ground; detect obstacles; convoying
- Graphics/video
  - Detect foreground objects and matte in other objects (super-matrix effect)
- Surveillance
  - Detect and classify vehicles on a street or in a parking garage
- Medical
  - Measurement (e.g. sizing tumors)
  - Visualization (e.g. register with pre-operative CT)

Obstacle Detection (cont'd)

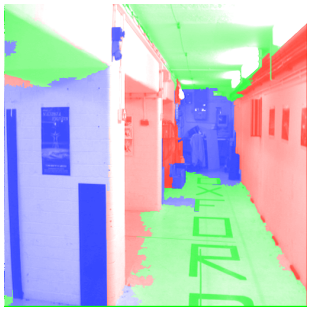
Observation: Removing the ground plane immediately exposes obstacles



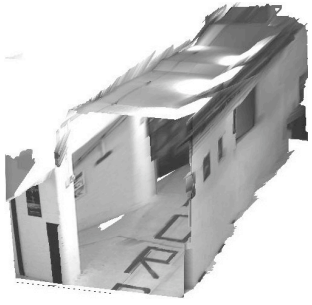
Applications of Real-Time Stereo



Oxford corridor



using 6 images



3D model

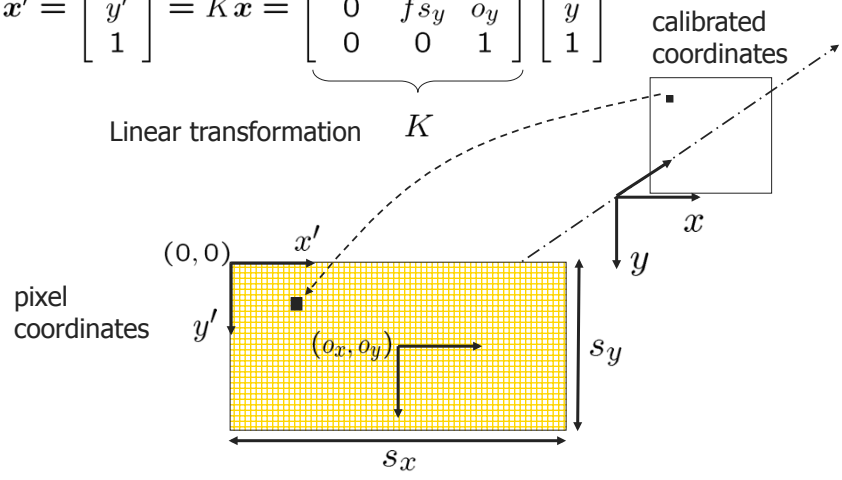
Feature based stereo

- Instead of matching each pixel
- Match features in the image
- What are good features ? – next lecture
- Examples of features – line matching, point matching region matching

Uncalibrated Camera

$$\boldsymbol{x}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = K\boldsymbol{x} = \underbrace{\begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{K} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Linear transformation

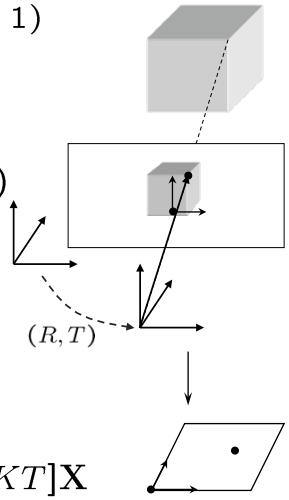


Uncalibrated Camera

$$\mathbf{X} = [X, Y, Z, W]^T \in \mathbb{R}^4, \quad (W = 1)$$

Calibrated camera

- Image plane coordinates $\mathbf{x} = [x, y, 1]^T$
- Camera extrinsic parameters $g = (R, T)$
- Perspective projection $\lambda \mathbf{x} = [R, T]\mathbf{X}$



Uncalibrated camera

- Pixel coordinates $\mathbf{x}' = K\mathbf{x}$
- Projection matrix $\lambda \mathbf{x}' = \Pi \mathbf{X} = [KR, KT]\mathbf{X}$

Jana Kosecka, CS 685 76

Calibration with a Rig

Use the fact that both 3-D and 2-D coordinates of feature points on a pre-fabricated object (e.g., a cube) are known.

Jana Kosecka, CS 685 77

Calibration with a Rig

- Given 3-D coordinates on known object

$$\lambda \mathbf{x}' = [KR, KT]\mathbf{X} \Rightarrow \lambda \mathbf{x}' = \Pi \mathbf{X}$$

$$\lambda \begin{bmatrix} x^i \\ y^i \\ 1 \end{bmatrix} = \begin{bmatrix} \pi_1^T \\ \pi_2^T \\ \pi_3^T \end{bmatrix} \begin{bmatrix} X^i \\ Y^i \\ Z^i \\ 1 \end{bmatrix}$$

- Eliminate unknown scales

$$\begin{aligned} x^i(\pi_3^T \mathbf{X}) &= \pi_1^T \mathbf{X}, \\ y^i(\pi_3^T \mathbf{X}) &= \pi_2^T \mathbf{X} \end{aligned}$$

- Recover projection matrix $\Pi = [KR, KT] = [R', T']$

$$\min \|M\Pi^s\|^2 \quad \text{subject to} \quad \|\Pi^s\|^2 = 1$$

$$\Pi^s = [\pi_{11}, \pi_{21}, \pi_{31}, \pi_{12}, \pi_{22}, \pi_{32}, \pi_{13}, \pi_{23}, \pi_{33}, \pi_{14}, \pi_{24}, \pi_{34}]^T$$

- Factor the into $R \in SO(3)$ and using QR decomposition
- Solve for translation $T = K^{-1}T'$

78

Alternative camera models/projections

Orthographic projection

$$\mathbf{x}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Scaled orthographic projection

$$\mathbf{x}' = s \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Affine camera model

$$\mathbf{x}' = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{33} & a_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

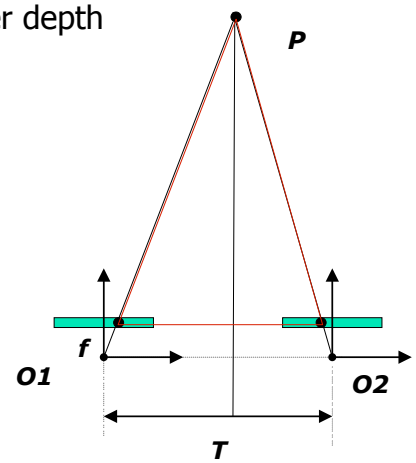
83

Stereo

- What if the motion between cameras is not known ?

Canonical Stereo Configuration

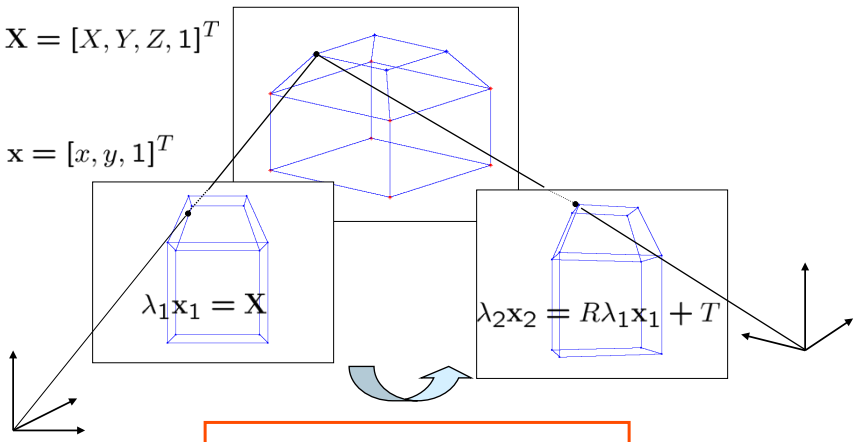
- Assumes (two) cameras
- Known positions and focal lengths
- Recover depth



$$\frac{Z}{T} = \frac{Z-f}{T-\mathbf{x}_l-\mathbf{x}_r}$$

$$Z = \frac{fT}{\text{disparity}}$$

Rigid Body Motion – Two Views



$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T$$

3D Structure and Motion Recovery

Euclidean transformation

$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T$$

measurements unknowns

$$\sum_{j=1}^{n} \|\mathbf{x}_1^j - \pi(R_1, T_1, \mathbf{X})\|^2 + \|\mathbf{x}_2^j - \pi(R_2, T_2, \mathbf{X})\|^2$$

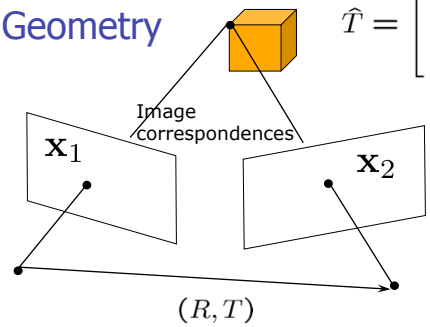
Find such Rotation and Translation and Depth that the reprojection error is minimized

Two views ~ 200 points
6 unknowns – Motion 3 Rotation, 3 Translation
- Structure 200x3 coordinates
- (-) universal scale

Difficult optimization problem

Epipolar Geometry

$$\hat{T} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$



$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T \quad / \widehat{\mathbf{x}_2} T$$

- Algebraic Elimination of Depth [Longuet-Higgins '81]:

$$\mathbf{x}_2^T \underbrace{\hat{T}R}_{E} \mathbf{x}_1 = 0$$

- Essential matrix $E = \hat{T}R$

Characterization of Essential Matrix

$$\mathbf{x}_2^T \hat{T} R \mathbf{x}_1 = 0$$

Essential matrix $E = \hat{T}R$ special 3x3 matrix

$$\mathbf{x}_2^T \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \mathbf{x}_1 = 0$$

(Essential Matrix Characterization)
A non-zero matrix E is an essential matrix iff its SVD: $E = U\Sigma V^T$ satisfies: $\Sigma = diag([\sigma_1, \sigma_2, \sigma_3])$ with $\sigma_1 = \sigma_2 \neq 0$ and $\sigma_3 = 0$ and $U, V \in SO(3)$

91

Pose Recovery

- There are two pairs (R, T) corresponding to essential matrix E .
- There are two pairs (R, T) corresponding to essential matrix $-E$.
- Positive depth constraint disambiguates the impossible solutions
- Translation has to be non-zero, can be recovered up to scale
- Points have to be in general position
  - degenerate configurations – planar points
  - quadratic surface
- Linear 8-point algorithm
- Nonlinear 5-point algorithms yields up to 10 solutions

95

3D Structure Recovery

$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + \gamma T$$

unknowns

- Eliminate one of the scale's

$$\lambda_1^j \widehat{\mathbf{x}_2^j} R \mathbf{x}_1^j + \gamma \widehat{\mathbf{x}_2^j} T = 0, \quad j = 1, 2, \ldots, n$$

- Solve LLSE problem

$$M^j \bar{\lambda}^j \doteq \left[\widehat{\mathbf{x}_2^j} R \mathbf{x}_1^j, \ \widehat{\mathbf{x}_2^j} T \right] \begin{bmatrix} \lambda_1^j \\ \gamma \end{bmatrix} = 0$$

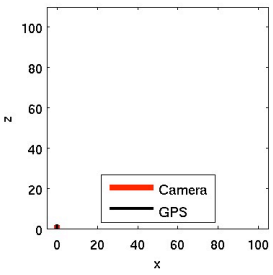
If the configuration is non-critical, the Euclidean structure of the points and motion of the camera can be reconstructed up to a universal scale.

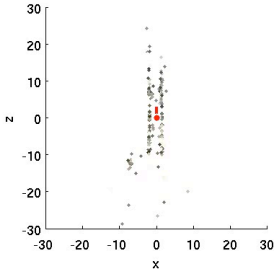
- Alternatively recover each point depth separately

96

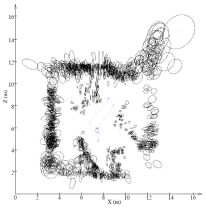
Visual Odometry

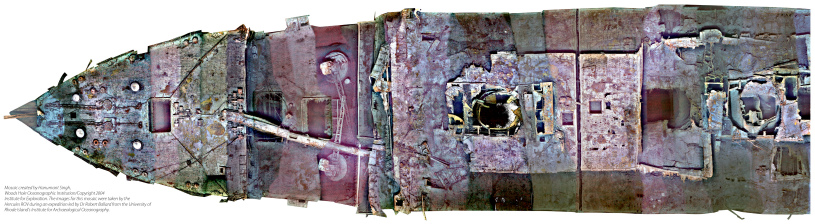
estimate motion from image correspondences





Mapping, Localization, Recognition





Dealing with correspondences

- Previous methods assumed that we have exact correspondences
- Followed by linear least squares estimation
- Correspondences established either by tracking (using affine or translational flow models)
- Or wide-baseline matching (using scale/rotation invariant features and their descriptors)
- In many cases we get incorrect matches/tracks

102

The RANSAC algorithm

- Generate *M* (a predetermined number) model hypotheses, each of them is computed using a minimal subset of points
- Evaluate each hypothesis
- Compute its residuals with respect to all data points.
- Points with residuals less than some threshold are classified as its inliers
- The hypothesis with the maximal number of inliers is chosen. Then re-estimate the model parameter using its identified inliers.

103

RANSAC – Practice

- The theoretical number of samples needed to ensure 95% confidence that at least one outlier free sample could be obtained.

$$\rho = 1 - (1 - (1 - \epsilon)^k)^s$$

- Probability that a point is an outlier $1 - \epsilon$
- Number of points per sample k
- Probability of at least one outlier free sample ρ

- Then number of samples needed to get an outlier free sample with probability ρ

$$s = \frac{\log(1 - \rho)}{\log(1 - (1 - \epsilon)^k)}$$

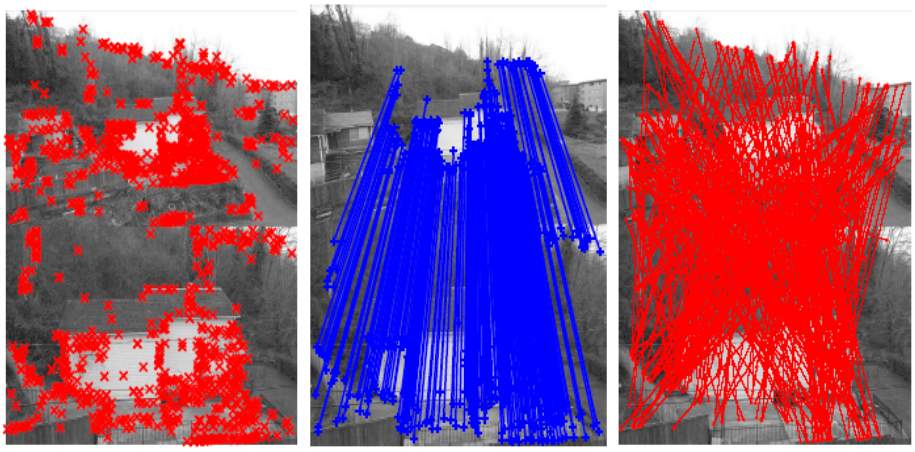
104

Adaptive RANSAC

- s = infinity, sample_count = 0;
- While s > sample_count repeat
  - choose a sample and count the number of inliers
  - set $\epsilon = 1 -$ (number_of_inliers/total_number_of_points)
  - set s from ϵ and $\rho = 0.99$
  - increment sample_count by 1
- terminate

107

Robust technique



(a) correspondences. (b) identified inliers. (c) identified outliers.

108

Robust matching

- Select set of putative correspondences $\mathbf{x}_1^j, \mathbf{x}_2^j$

$$\mathbf{x}_2^T F \mathbf{x}_1 = 0$$

- Repeat
  1. Select at random a set of 8 successful matches
  2. Compute fundamental matrix
  3. Determine the subset of inliers, compute distance to epipolar line

$$\hat{d}^j \doteq \frac{(\mathbf{x}_2^{jT} F_k \mathbf{x}_1^j)^2}{\|\hat{e}_3 F \mathbf{x}_1^j\|^2 + \|\mathbf{x}_2^{jT} F \hat{e}_3\|^2} \qquad d_j \leq \tau_d$$

  4. Count the number of points in the consensus set

109