# CS 6550 List of Project Topics

- Stereo matching
- Dynamic background subtraction
- Action recognition
- Face verification
- Facial expression recognition
- Image categorization
- Image stitching
- image segmentation
- Object tracking
- Object detection
- Deep learning related topics

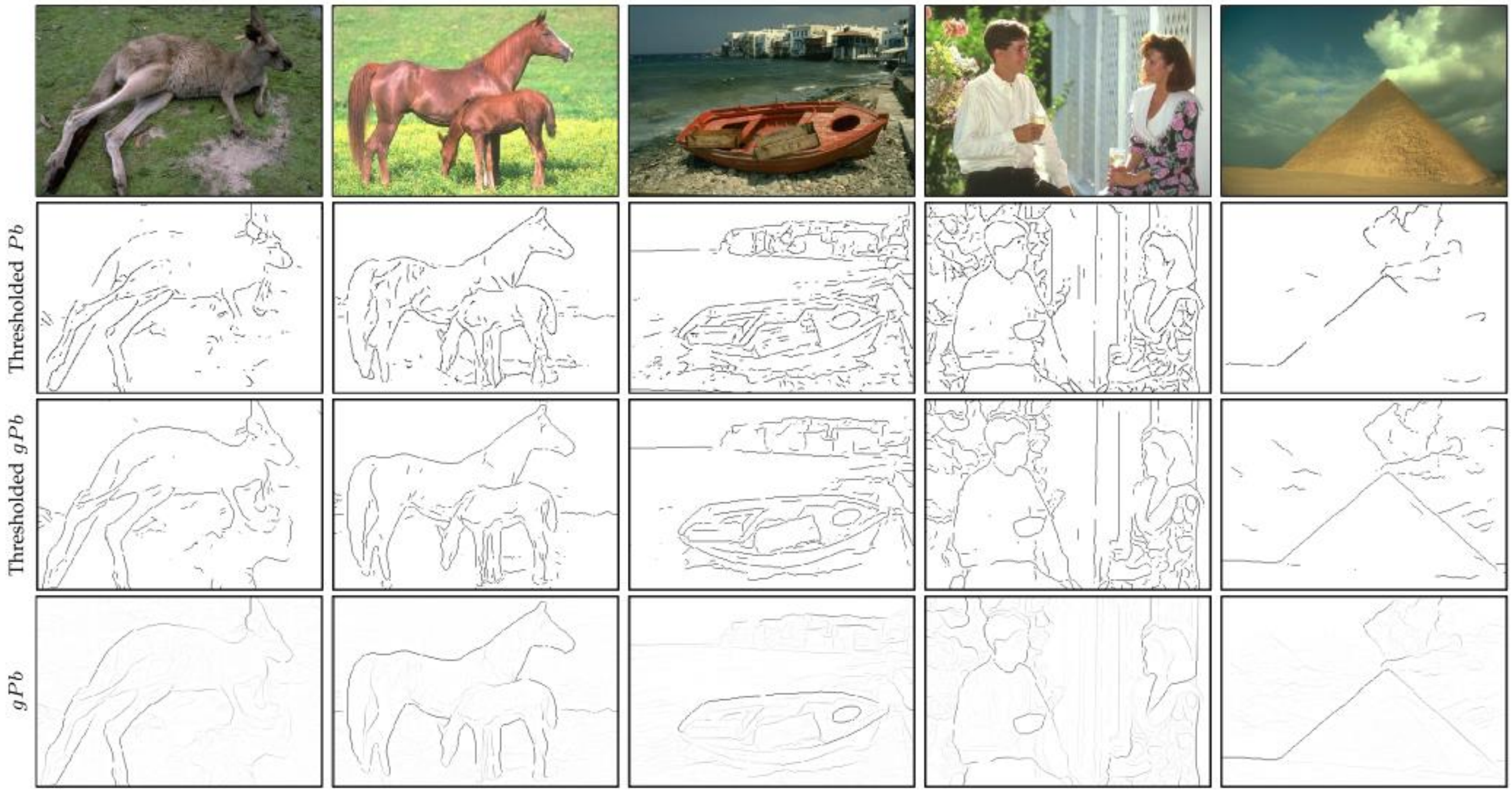
# CS 6550 Final Project: Image Segmentation & Boundary Finding

# Segmentation and Boundary Finding

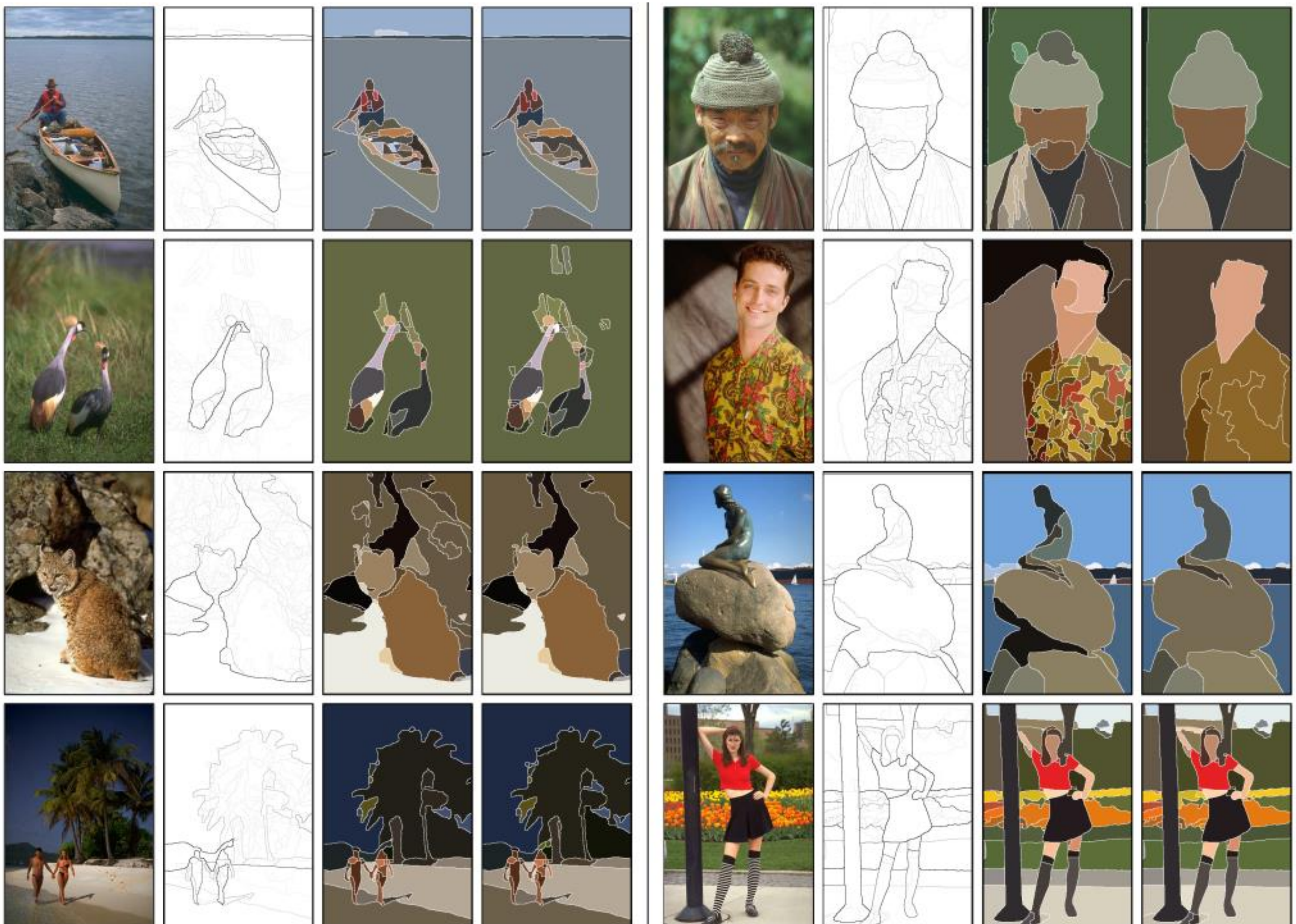
- **A large dataset** of natural images that have been segmented by human observers. This dataset serves as ground truth for learning grouping cues as well as a benchmark for comparing different segmentation and boundary finding algorithms.
- The normalized cut algorithm provides a mechanism for going from pairwise pixel affinities based on local cues to partitions that maximize the ratio of affinities within a group to that across groups.
- By computing oriented gradients on the eigenvectors of the normalized Laplacian, a robust signal for marking image contours can be obtained.
- The Berkeley group used Conditional Random Fields as a formalism for combing low-, mid- and high- level cues for grouping and figure-ground discrimination.

P. Arbelaez, M. Maire, C. Fowlkes and J. Malik., *Contour Detection and Hierarchical Image Segmentation.,* IEEE TPAMI, Vol. 33, No. 5, pp. 898-916, May 2011.

# Contour Detection and Hierarchical Image Segmentation



# gPb Segmentation Results



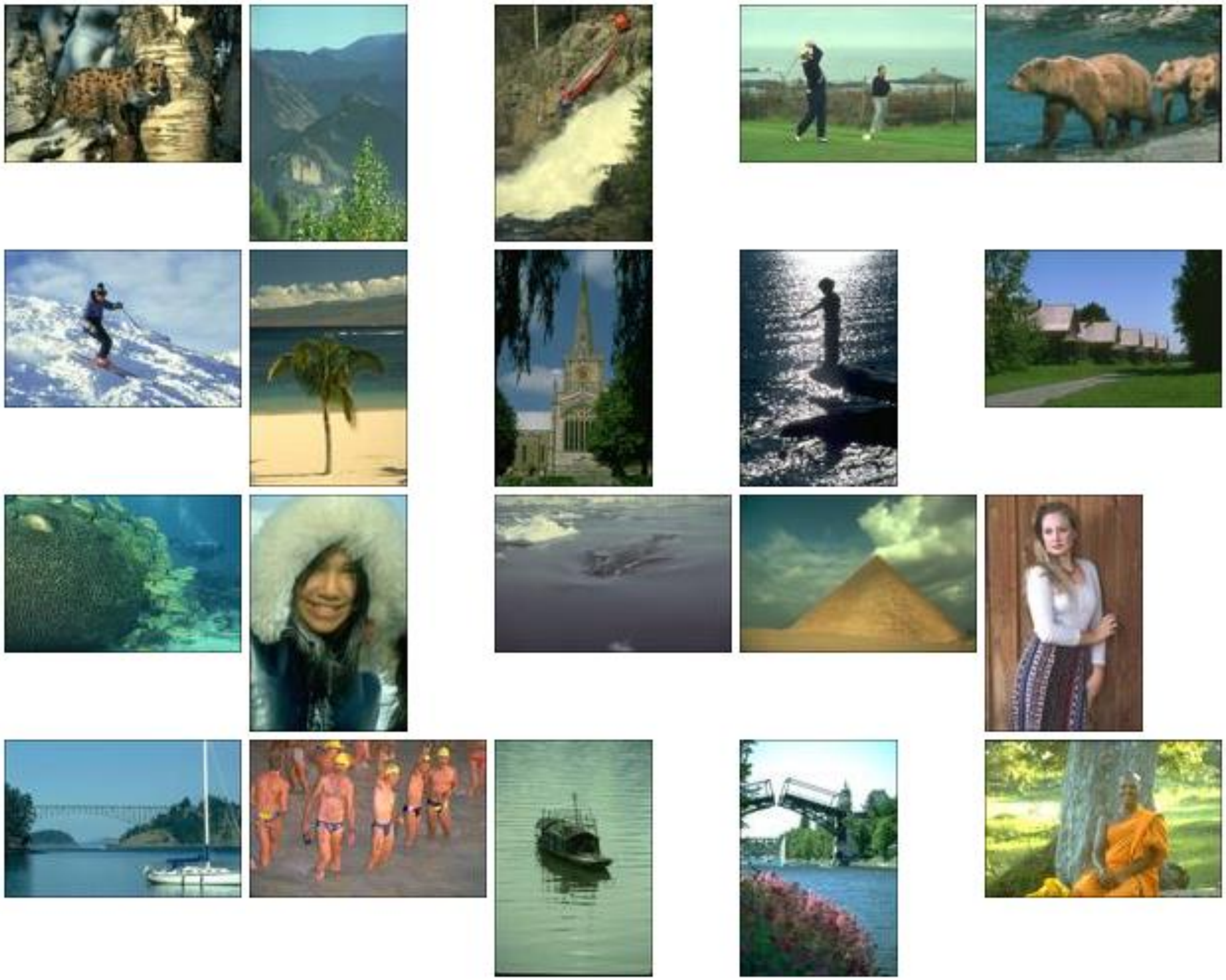
# Resources

- The Berkeley Segmentation Dataset and Benchmark
  - http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/
- Reference paper
  - L. Grady, Random Walks for Image Segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 28, No. 11, pp. 1768–1783, Nov., 2006.
  - P. Arbelaez, M. Maire, C. Fowlkes and J. Malik., *Contour Detection and Hierarchical Image Segmentation.,* IEEE TPAMI, Vol. 33, No. 5, pp. 898-916, May 2011. [resources]
  - Xiaofeng Ren and Liefeng Bo, "Discriminatively Trained Sparse Code Gradients for Contour Detection.", NIPS 2012.

# Resources for Image Segmentation

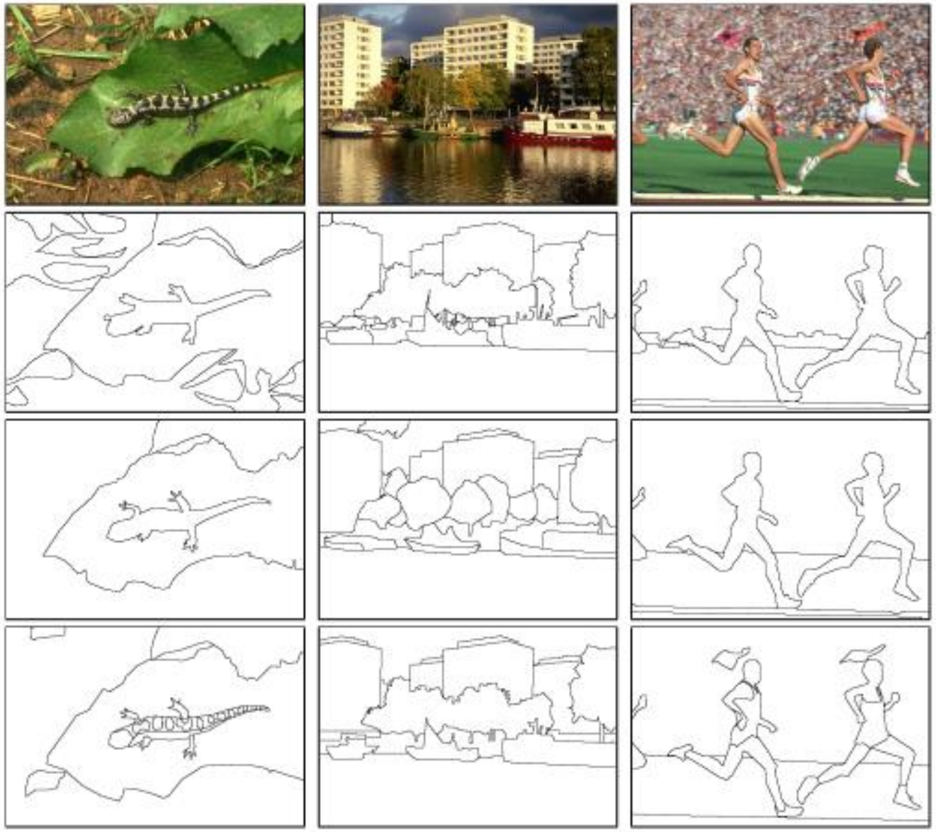
- **Multiscale Combinatorial Grouping (MCG)** is available.
- Resources for contour detection and image segmentation, including the **Berkeley Segmentation Data Set 500 (BSDS500)**, are available.
- The **Berkeley Segmentation Data Set 300 (BSDS300)** is still available.
- The **Berkeley Semantic Boundaries Dataset and Benchmark (SBD)** is available.
- The **Berkeley Video Segmentation Dataset (BVSD)** is available.

# Berkeley Segmentation Dataset



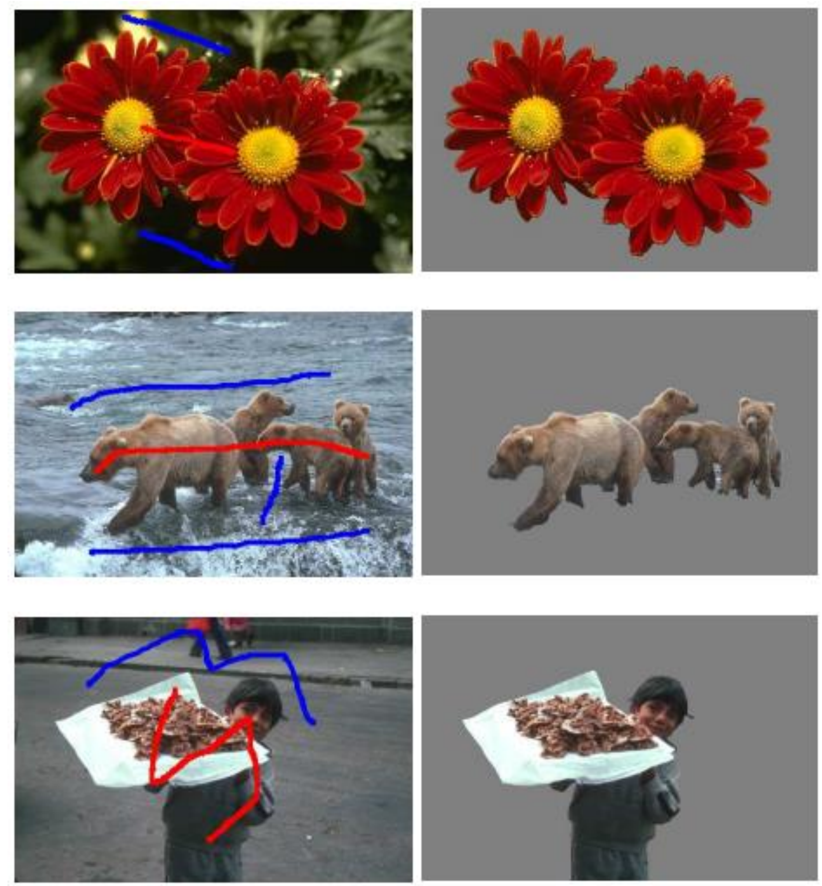
# Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500)

This new dataset is an extension of the BSDS300, where the original 300 images are used for training / validation and 200 fresh images, together with human annotations, are added for testing. Each image was segmented by five different subjects on average. Performance is evaluated by measuring Precision / Recall on detected boundaries and three additional region-based metrics.



# Interactive Image Segmentation

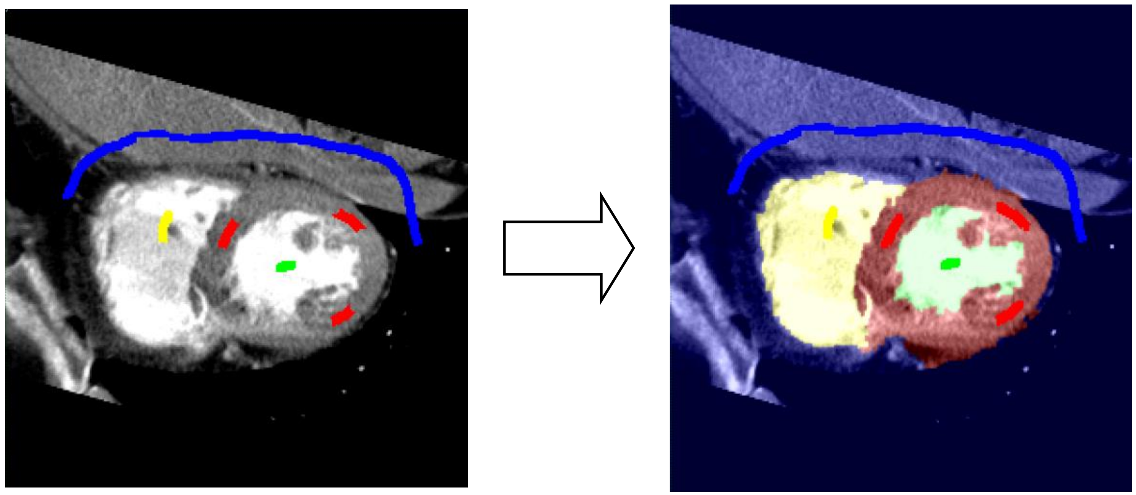
- Problem
  - Interactive segmentation with user defined labels
  - Two label interactive segmentation
  - Multi-label interactive segmentation
- Approaches
  - Random walk
  - Graph Cut



Input Image | Segmentation foreground

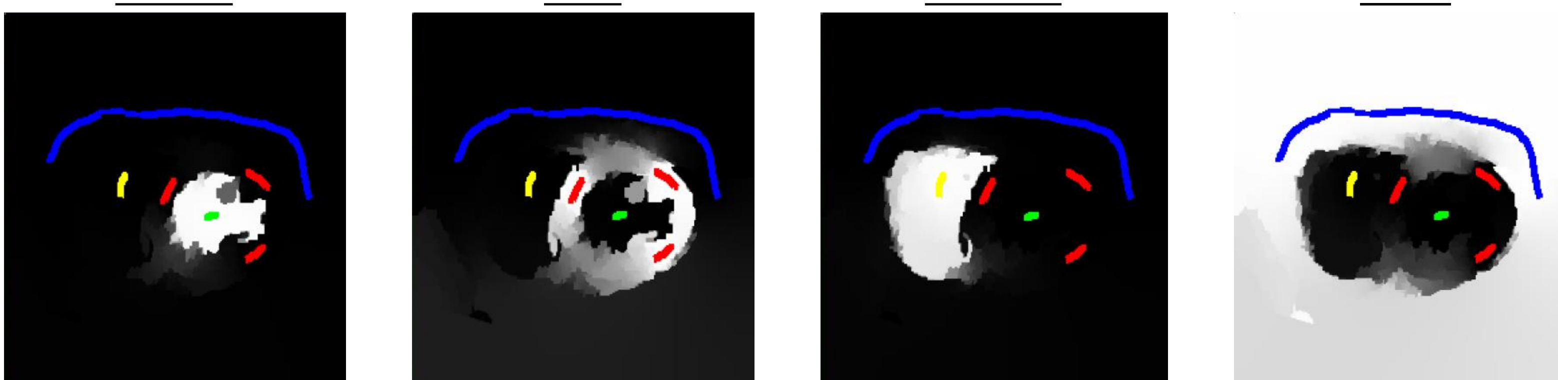
# Random walker segmentation

Partially labeled image Segmented image



Probabilities

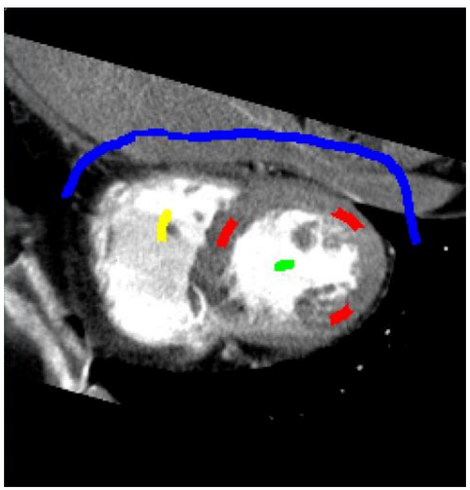
Green Red Yellow Blue

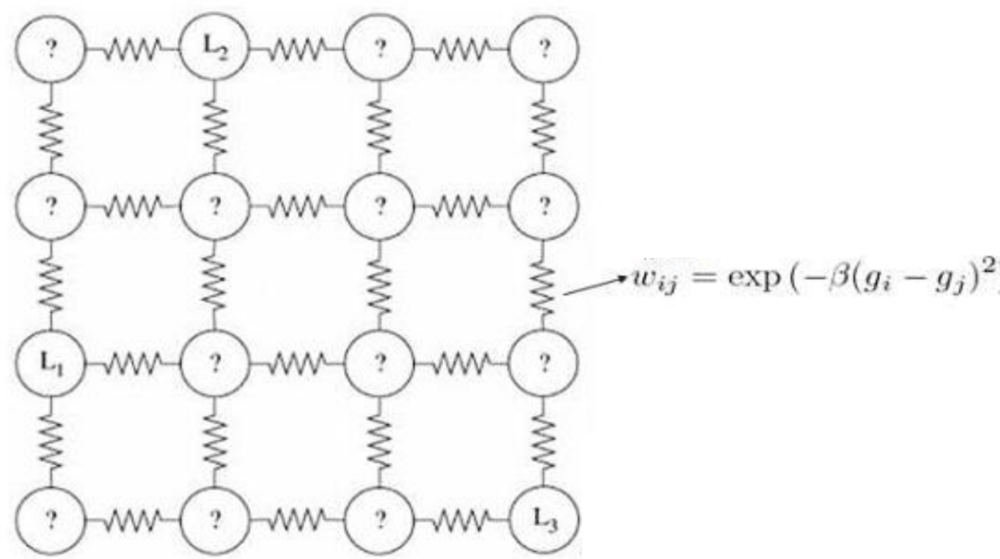


L. Grady, Random Walks for Image Segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 28, No. 11, pp. 1768–1783, Nov., 2006.

# Random walk image segmentation

- Graph theory
  - G=(V,E)





# Random walk image segmentation

- Given region $x_M$ and unknown region $x_U$

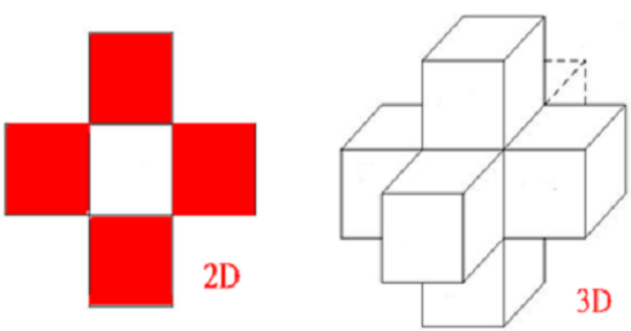
$$D[x] = \frac{1}{2} x^T L x \qquad L_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -w_{ij} & \text{if } v_i \text{ and } v_j \text{ are adjacent nodes} \\ 0 & \text{otherwise,} \end{cases}$$

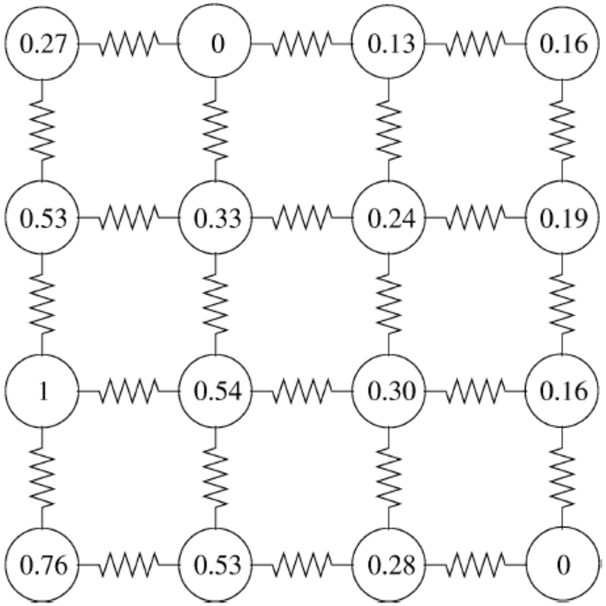
$$D[x_U] = \frac{1}{2} \begin{bmatrix} x_M^T x_U^T \end{bmatrix} \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix} \begin{bmatrix} x_M \\ x_U \end{bmatrix} = \frac{1}{2} \left( x_M^T L_M x_M + 2 x_U^T B^T x_M + x_U^T L_U x_U \right)$$

- $D[u] = 0$

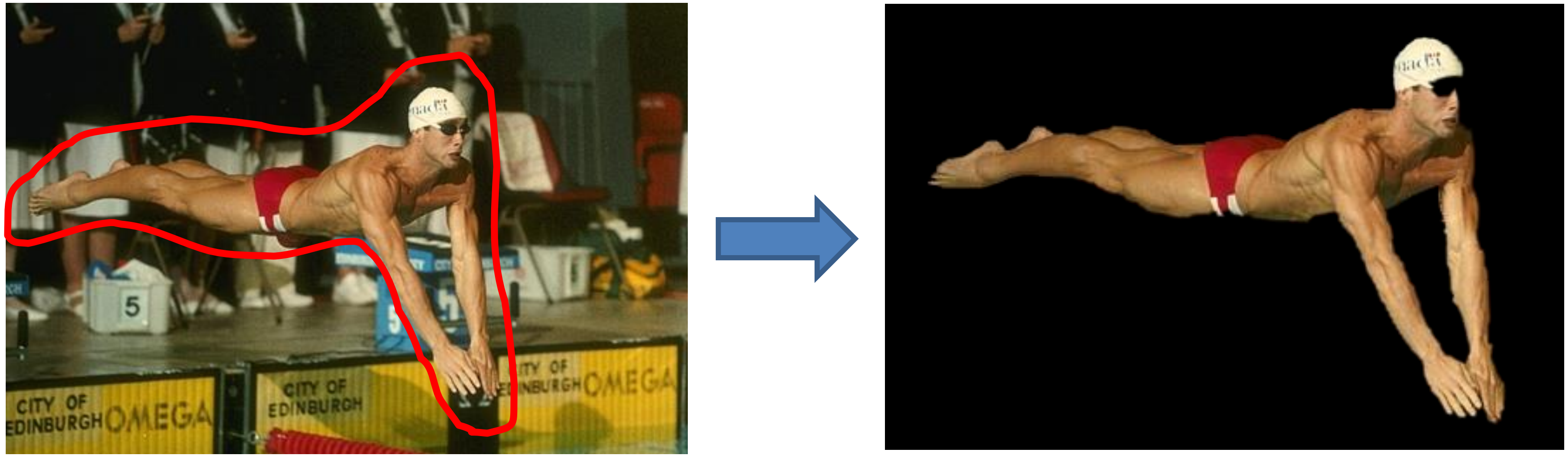
$$L_U x_U = -B^T x_M$$

- 2D and 3D



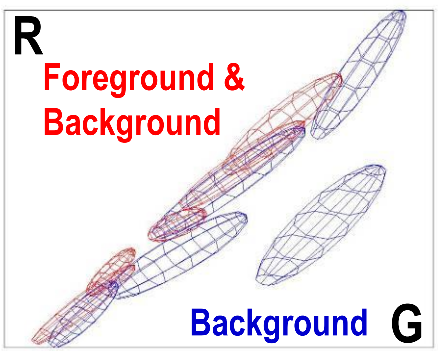
# Grabcut segmentation



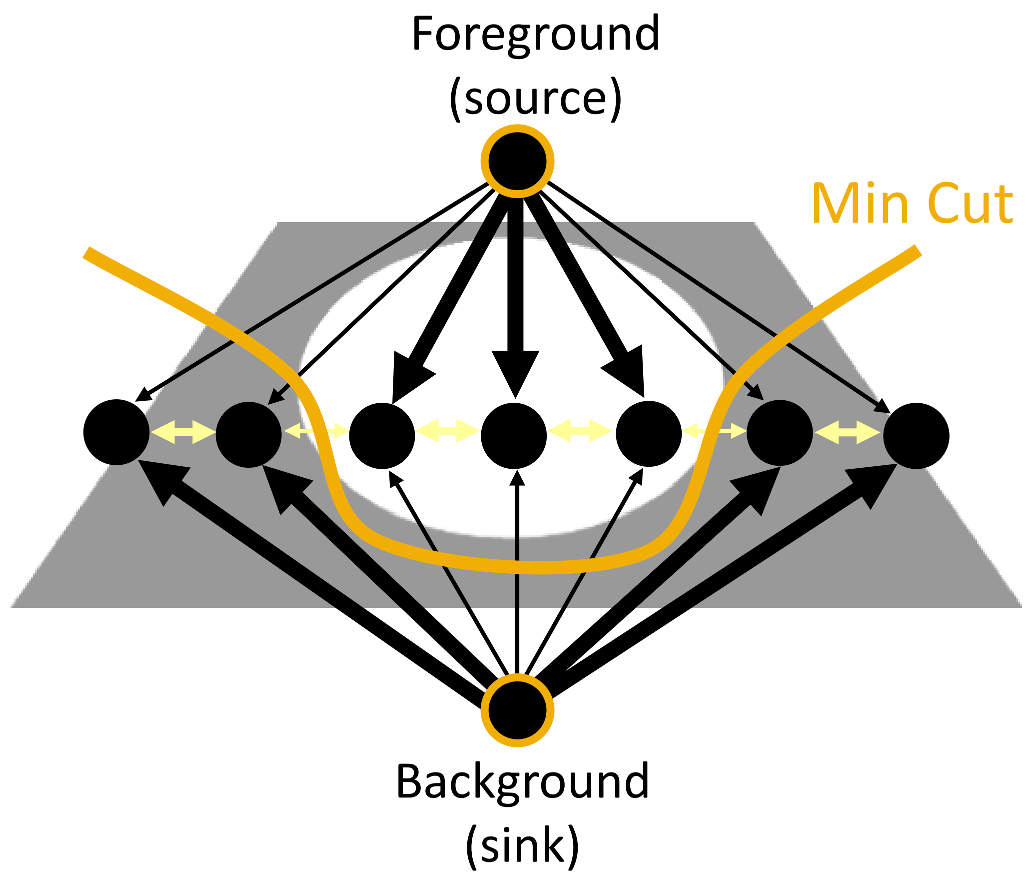

User provides rough indication of foreground region.

Goal: Automatically provide a pixel-level segmentation.

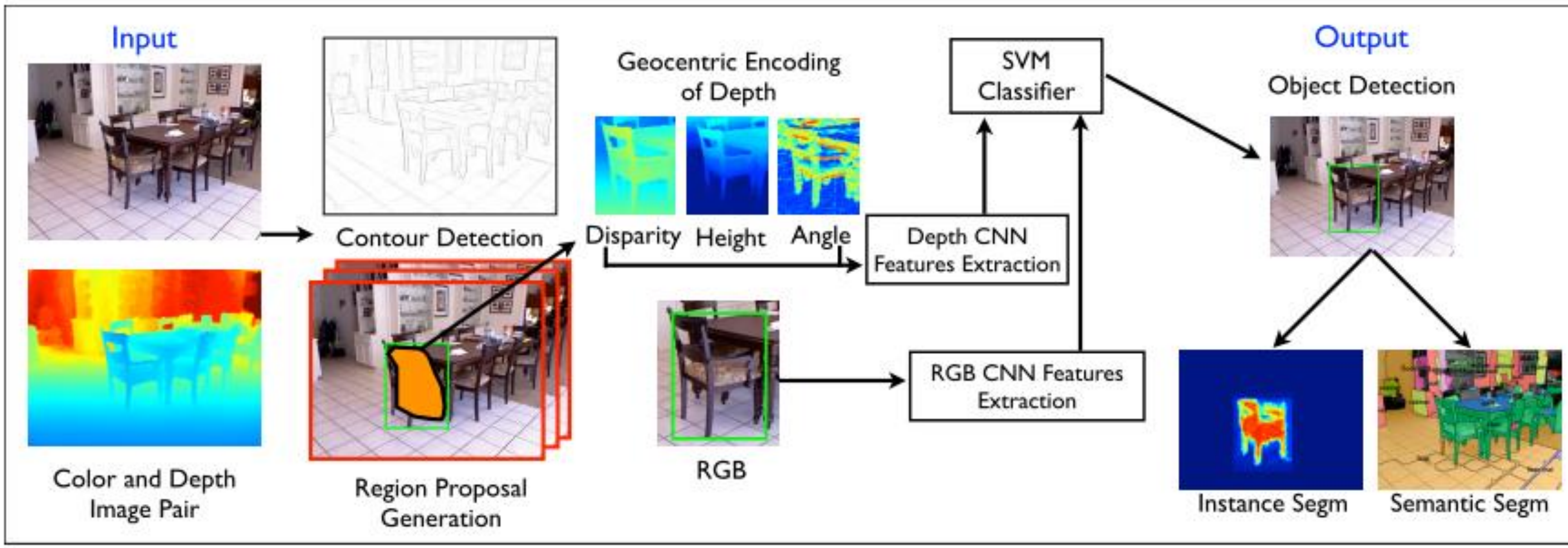
# Grabcut segmentation




Gaussian Mixture Model
(typically 5-8 components)

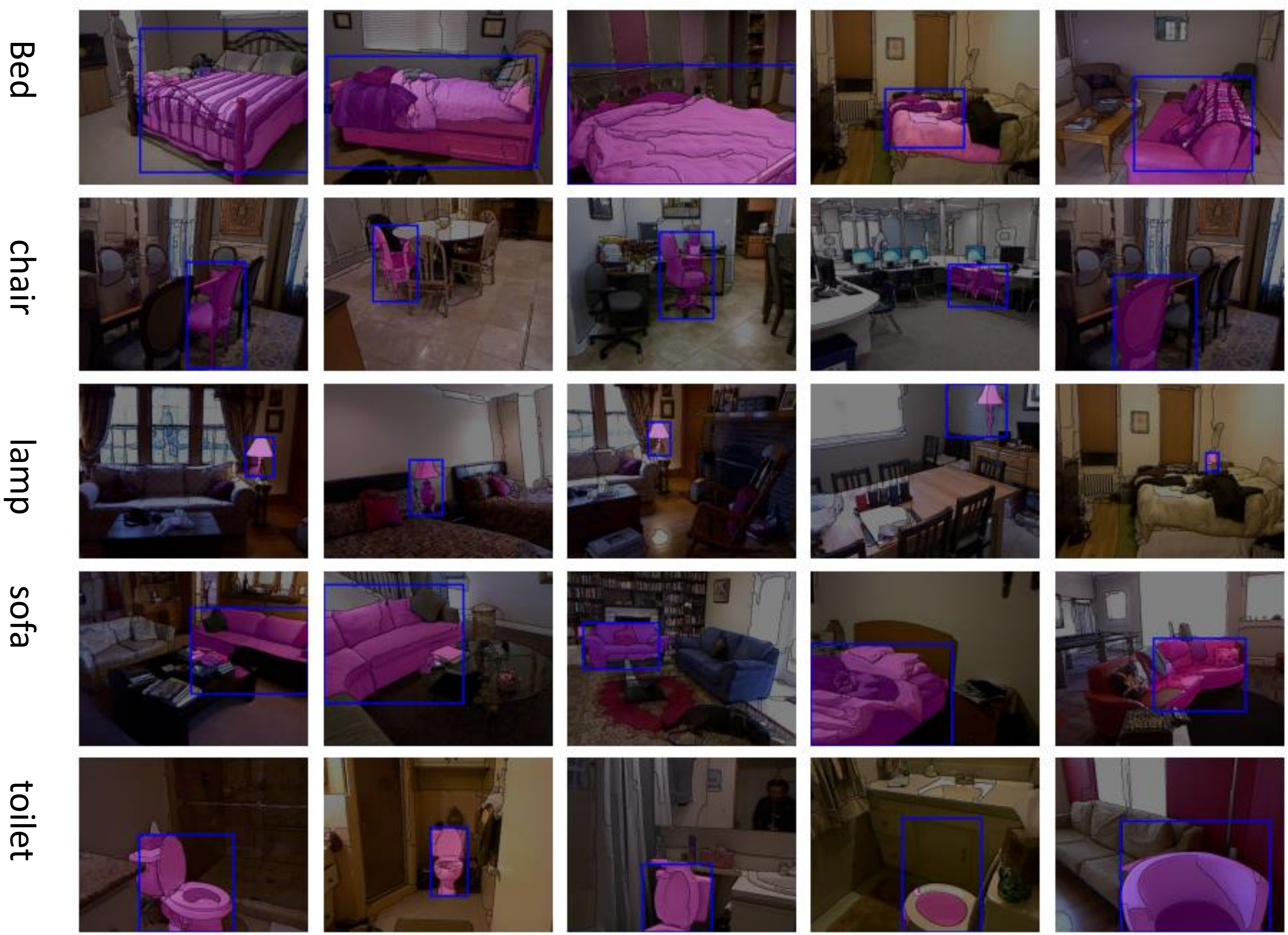


# Object Detection and Segmentation from RGBD Images




S. Gupta, R. Girshick, P. Arbelaez and J. Malik., "Learning Rich Features from RGB-D Images for Object Detection and Segmentation" , ECCV 2014
https://github.com/s-gupta/rcnn-depth
http://www.eecs.berkeley.edu/Research/Projects/CS/vision/papers/SaurabhECCV2014.pdf

# Examples of Object Detection Results

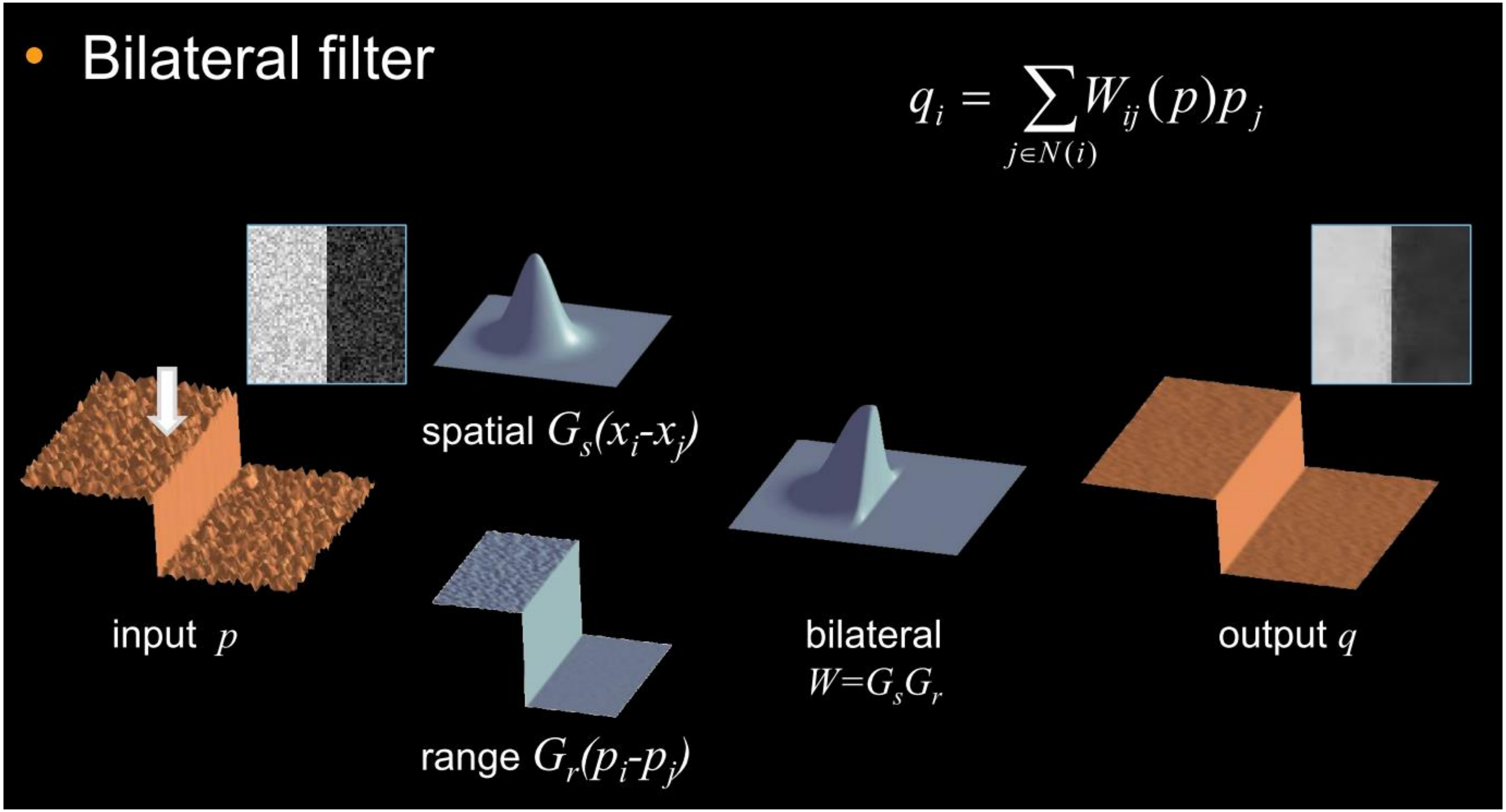


# CS 6550 Final Project: Guided Image Filter

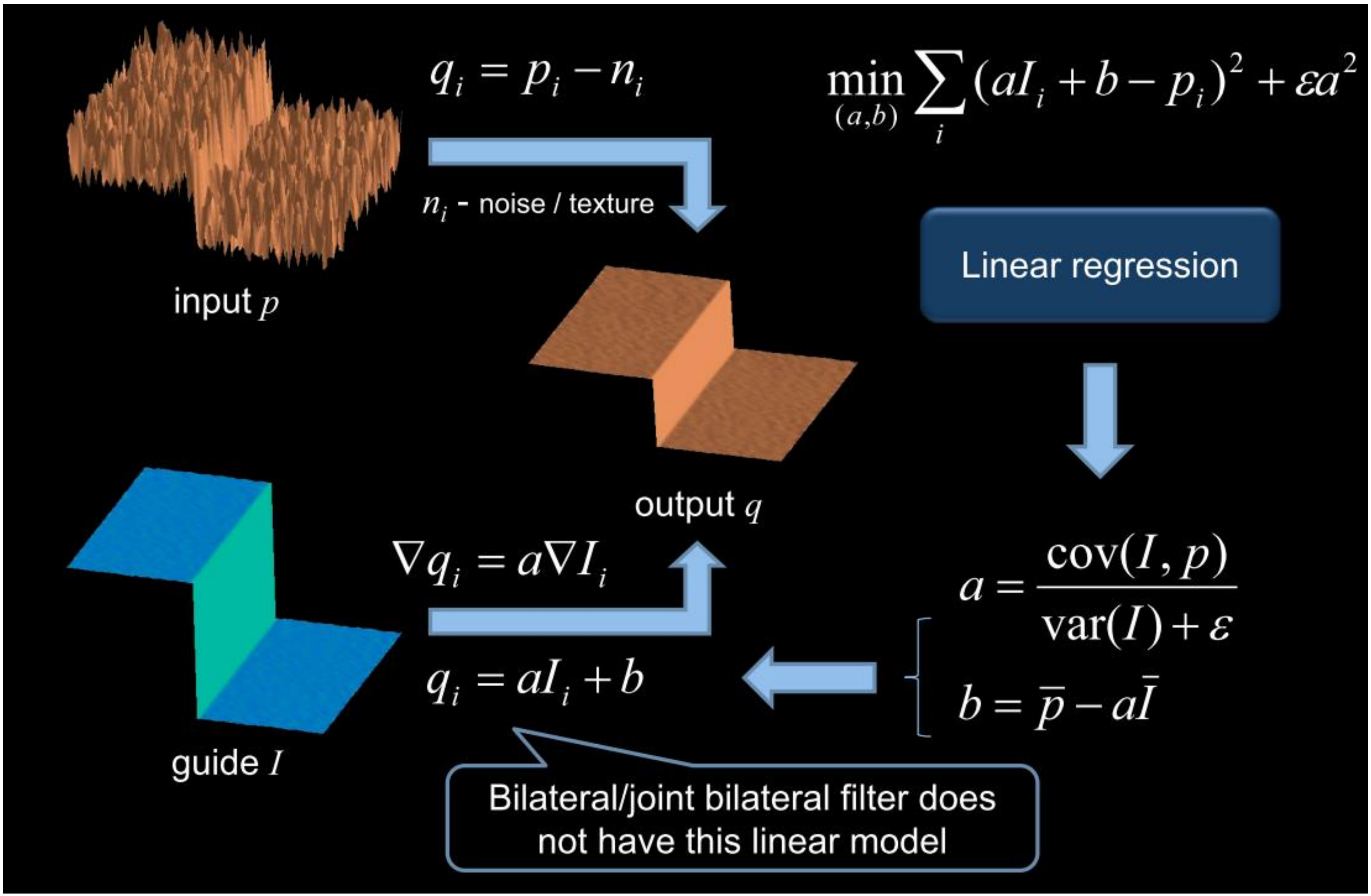
# Edge-Preserving Filtering

- An important topic in computer vision
  - Denoising, image smoothing/sharpening, texture decomposition, HDR, compression, image abstraction, optical flow estimation, image super-resolution, feature smoothing…
- Existing methods
  - Weighted Least Square [Lagendijk et al. 1988]
  - Anisotropic diffusion [Perona and Malik 1990]
  - Bilateral filter [Aurich and Weule 95], [Tomasi and Manduchi 98]
  - Digital TV (Total Variation) filter [Chan et al. 2001]

# Bilaterial Filter

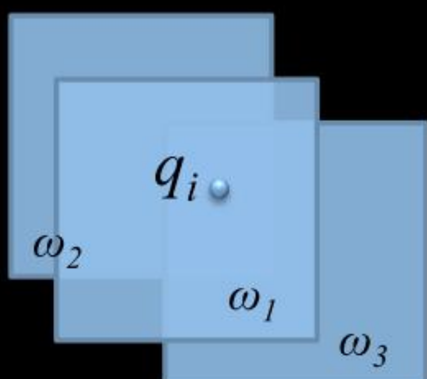


# Guided Filter



# Guided Filter

- Extend to the entire image
  - In all local windows $\omega_k$ ,compute the linear coefficients
  - Compute the average of $a_k I_i + b_k$ in all $\omega_k$ that covers pixel $q_i$




$$a_k = \frac{\mathrm{cov}_k(I, p)}{\mathrm{var}_k(I) + \varepsilon}$$

$$b_k = \bar{p}_k - a\bar{I}_k$$

$$q_i = \frac{1}{|\omega|} \sum_{k|i \in \omega_k} (a_k I_i + b_k)$$

$$= \bar{a}_i I_i + \bar{b}_i$$

# Example – detail enhancement


input ($I$=$p$)

bilateral filter
$\sigma_s=16,\ \sigma_r=0.1$

guided filter
$r=16,\ \varepsilon=0.1^2$

# Example – haze removal

guide $I$

guided filter
(<0.1s, 600x400p)

global optimization
(10s)

# CS 6550 Final Project: Image Stitch

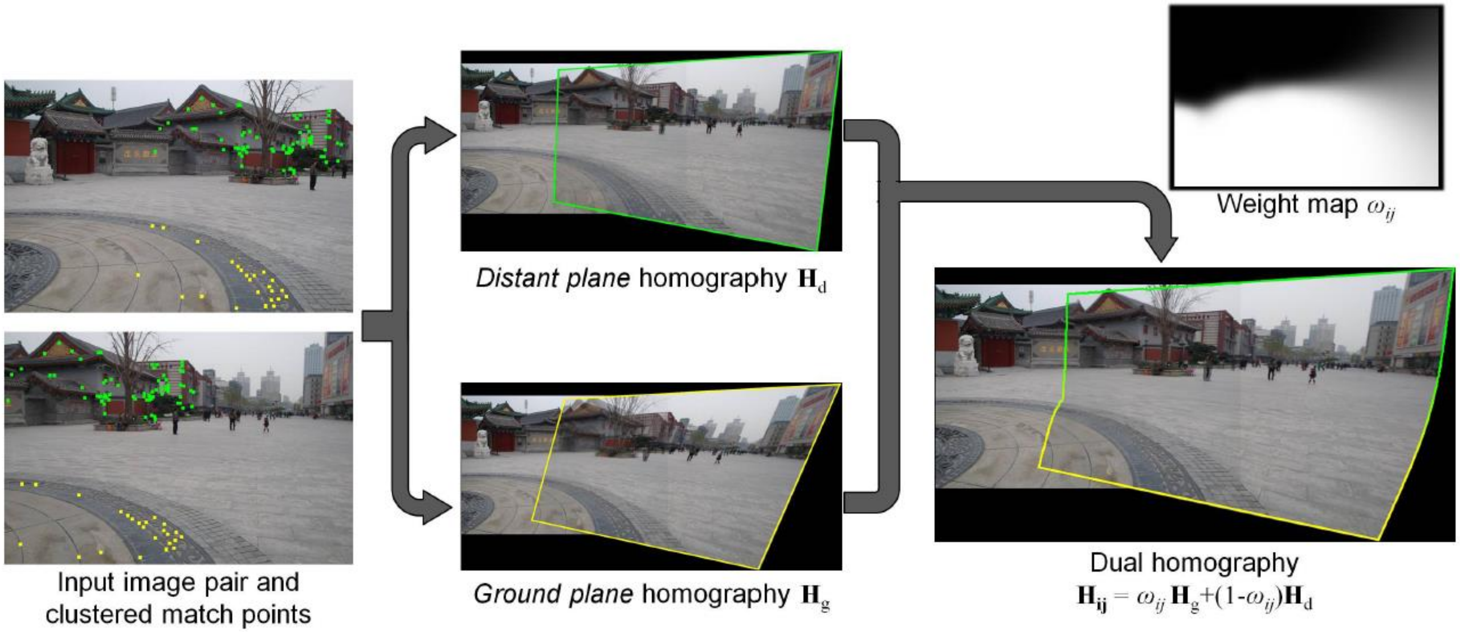
# Image stitching/paronoma

- Problem

# Image stitching/paronoma

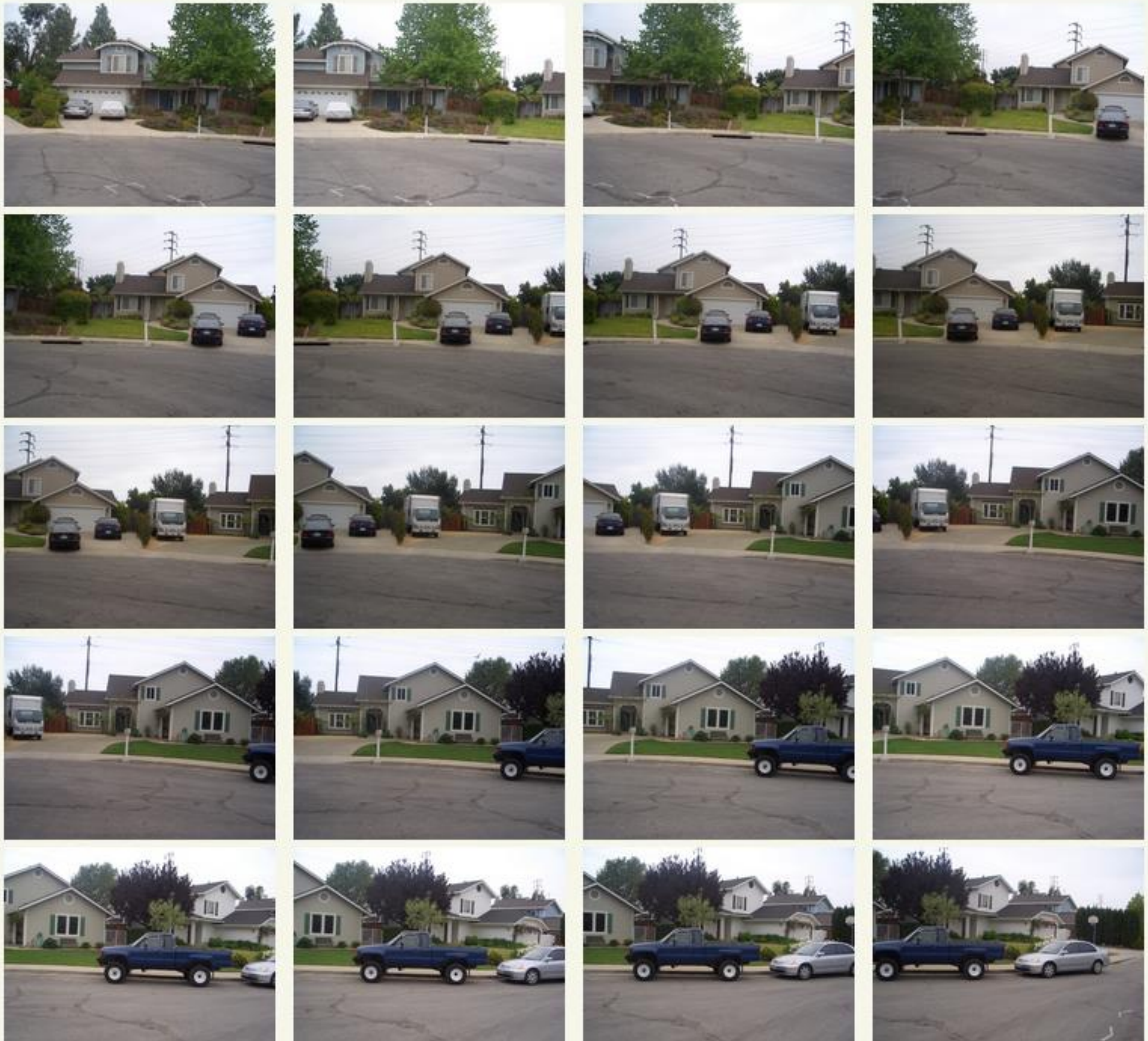
- Data set
  - http://www.visualsize.com/mosaic3d/index.php
- References
  - J. Jia and C.-K. Tang, Image stitching using structure deformation, , IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2008.
  - Gao et al., Constructing Image Panoramas using Dual-Homography Warping, CVPR 2011.
  - …

# Constructing Image Panoramas using Dual-Homography Warping




Work flow of our dual-homography computation. Candidate points are first clustered into groups from which two homographies are estimated using RANSAC. A per-pixel weight map $\omega_{ij}$ is computed to control the blending of the two homographies.

## Input Images Of The Cul de Sac Data Set

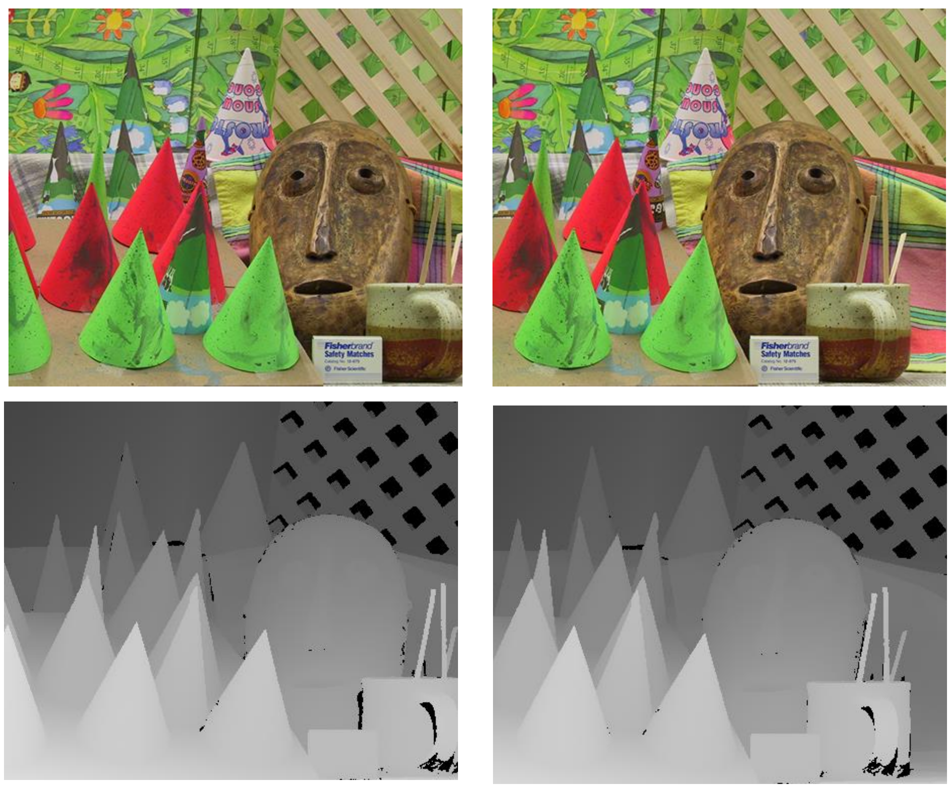


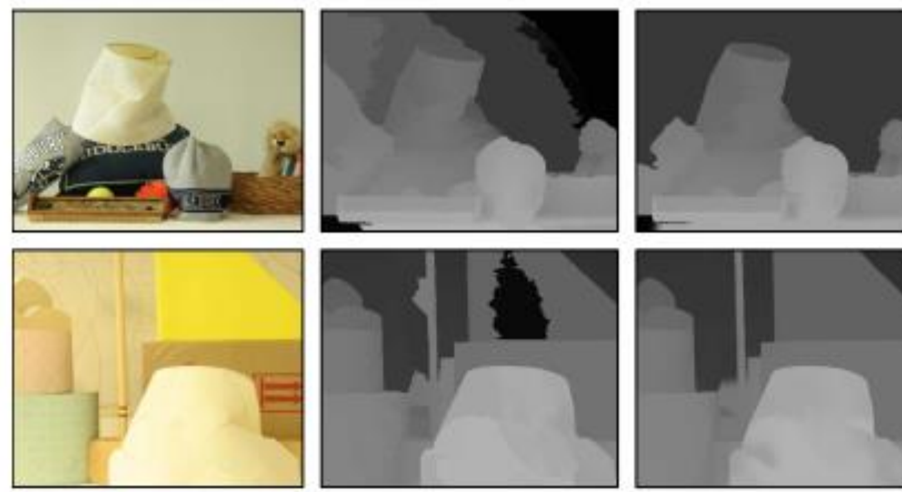
# Input Images Of The University of California, Santa Barbara Data Set

# CS 6550 Final Project: Stereo Matching

# Stereo Matching

- Assume that pairs of images are rectified.
- Given left and right image, compute left disparity and right disparity

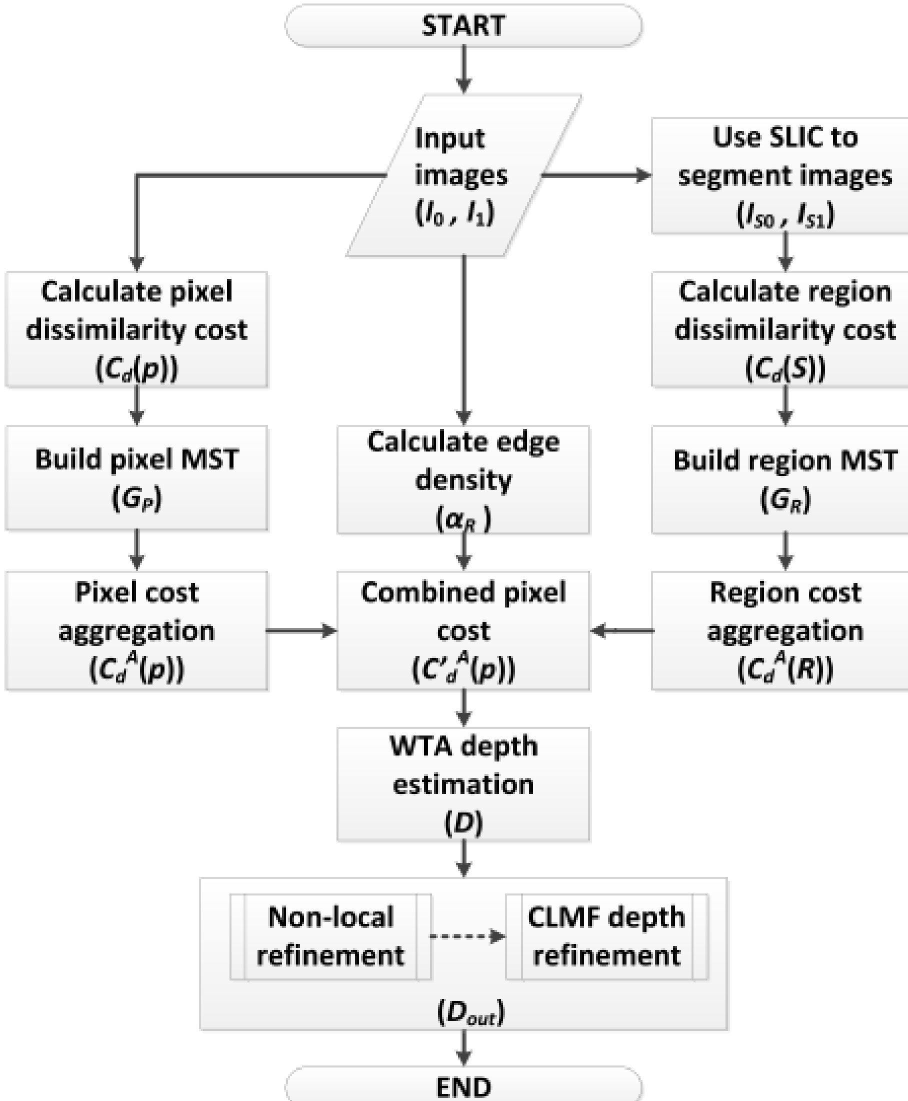




Traditional method

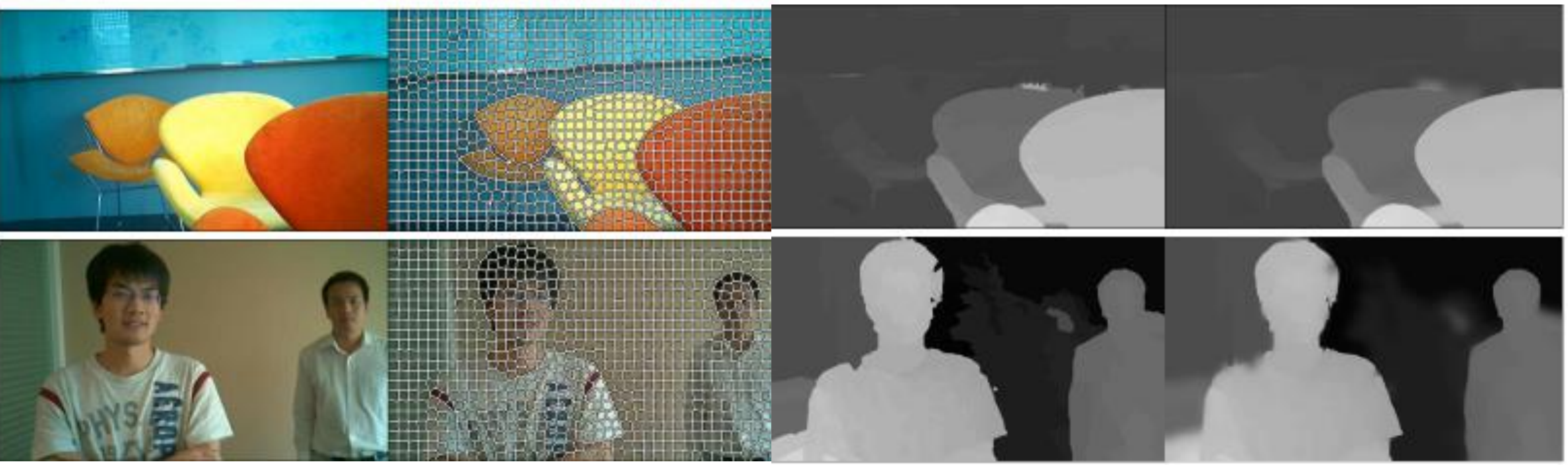
Global optimization method

# Tree-Based Stereo Matching




**Efficient hybrid tree-based stereo matching with applications to postcapture image refocusing,** Vu DT, Chidester B, Yang H, Do MN, Lu J., IEEE IP. 23(8):3428-42, 2014

# Stereo matching results

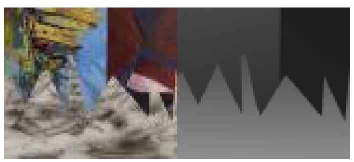


# Evaluation & Datasets

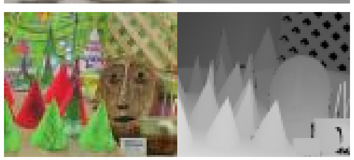
- Most popular datasets

**http://vision.middlebury.edu/stereo/data/**

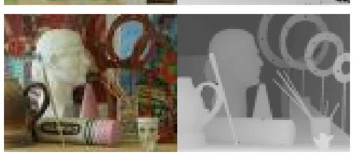
**Middlebury Stereo Datasets**



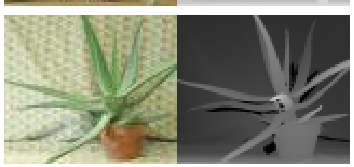
2001 datasets - 6 datasets of piecewise planar scenes [1]
(Sawtooth, Venus, Bull, Poster, Barn1, Barn2)



2003 datasets - 2 datasets with ground truth obtained using structured light [2]
(Cones, Teddy)



2005 datasets - 9 datasets obtained using the technique of [2], published in [3, 4]
(Art, Books, Dolls, Laundry, Moebius, Reindeer, Computer, Drumsticks, Dwarves)



2006 datasets - 21 datasets obtained using the technique of [2], published in [3, 4]
(Aloe, Baby1-3, Bowling1-2, Cloth1-4, Flowerpots, Lampshade1-2, Midd1-2, Monopoly, Plastic, Rocks1-2, Wood1-2)

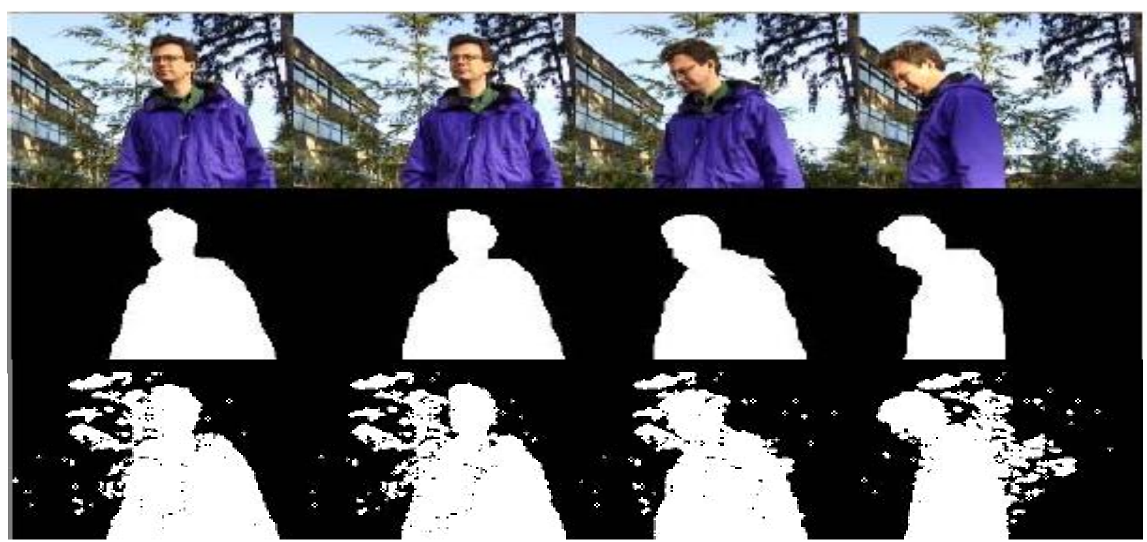
# References


- Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. PAMI 2008.
- X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. On building an accurate stereo matching system on graphics hardware. GPUCV 2011.
- Z. Wang and Z. Zheng. A region based stereo matching algorithm using cooperative optimization. CVPR 2008.

# CS 6550 Final Project: Background subtraction for dynamic scene

# Background

- The detection moving objects from a video sequence is a fundamental task in many computer vision applications
- In general environment, the background consists of stationary and dynamic object
  - Traditional background subtraction algorithm (e.g.  GMM) work well for stationary background. However,  the performance usually increase when the background is dynamic.



Ground truth

GMM based method

# Problem description & dataset

- You should develop an algorithm to extract foreground object in dynamic scenes .
- You should test some public datasets
  - Dataset1 includes fountain, water surfaces, moving escalators etc. (http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html)



  - Dataset2 includes boat, surf, ocean, rain etc. (http://www.svcl.ucsd.edu/projects/background_subtraction/ucsdbgsub_dataset.htm)

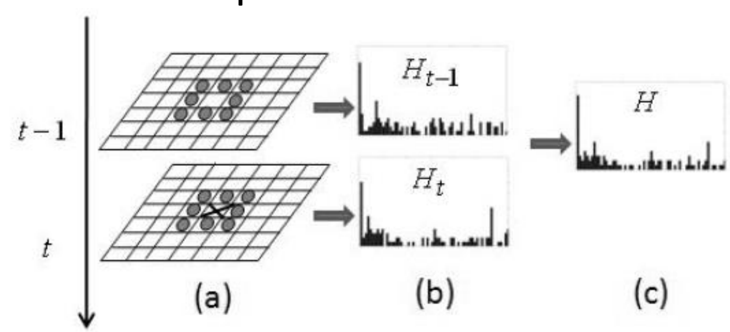



# Approaches

- Local spatial information based methods
  - L. Li, W. Huang, I.-Y.H. Gu, and Q. Tian, “Foreground object detection from videos containing complex background,” ACM MM’03
    - A Bayes decision rule for classification of background and foreground from selected feature vectors is formulated
    - The stationary background object is described by the color feature, and the moving background object is represented by the color co-occurrence feature
  - S.-P. Zhang, H.-X. Yao, and S.-H. Liu, “Dynamic background modeling and subtraction using spatiotemporal local binary patterns,” ICIP’08
    - Extend ordinary local binary patterns from spatial domain to spatiotemporal domain

$$LBP^{t}_{P,R}(x_{t,c}, y_{t,c}) = \sum_{p=0}^{P-1} s(g_{t,p} - g_{t,c})2^{p}$$

$$LBP^{t-1}_{P,R}(x_{t,c}, y_{t,c}) = \sum_{p=0}^{P-1} s(g_{t-1,p} - g_{t,c})2^{p}$$



# References

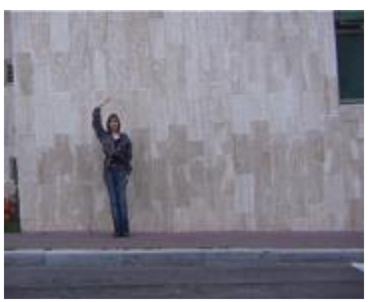

- L. Li, W. Huang, I.-Y.H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," ACM MM'03
- Y. Sheikh, M. Shah, "Bayesian object detection in dynamic scenes," CVPR'05
- S.-P. Zhang, H.-X. Yao, S.-H. Liu, X.-L. Chen and W. Gao, "A Covariance-based Method for Dynamic Background Subtraction," ICPR'08
- G. Dalley, J. Migdal, W.E.L. Grimson, "Background Subtraction for Temporally Irregular Dynamic Textures," WACV'08
- V. Mahadevan, N. Vasconcelos, "Spatiotemporal Saliency in Dynamic Scenes",PAMI'10

# CS 6550 Final Project
# Action Recognition

# background

- Automatically recognizing human actions is receiving increasing attention due to its wide range of applications
- Recent towards the use of Spatio-Temporal Interest Points (STIPs) for local descriptor-based recognition strategies
- Simple scenes V.S. complex scenes

Weizmann dataset (simple) – wave hand



hollywood2 dataset (complex) - Kiss

# Problem description & dataset

- You should develop an algorithm to classify actions effectively from actions datasets.
- Simple datasets
  - **KTH** dataset : 6 different actions performed in 4 different but well-controlled environments by 25 different actors

  http://www.nada.kth.se/cvap/actions/
  - **Weizmann** dataset : contains 90 videos separated into 10 actions performed by 9 persons

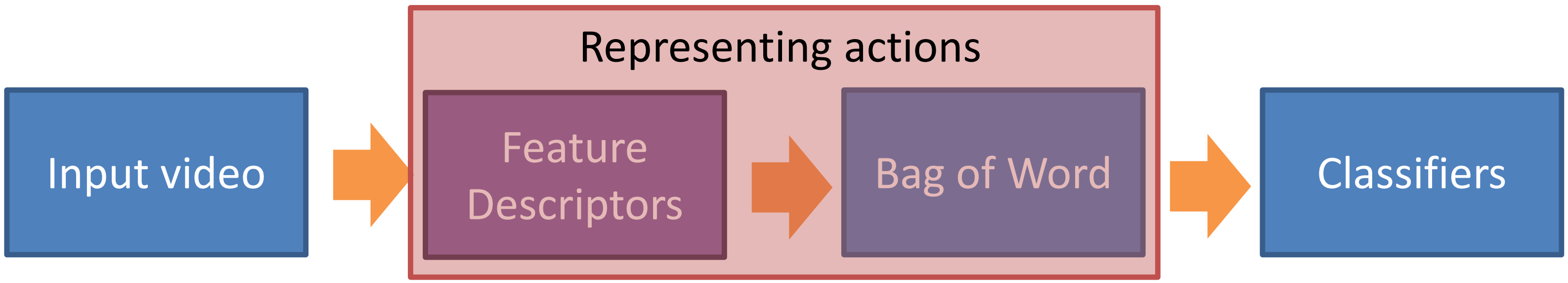
  http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html
- Complex dataset (including movie and YouTube video clips)
  - **Hollywood2** dataset : it composed of video clips extracted from 69 Hollywood movies, and contains 12 classes actions. Totally, there are 1707 action sample videos

  http://www.di.ens.fr/~laptev/actions/hollywood2/
  - **YouTube** action dataset : 1168 complex and challenging YouTube videos of 11 human action categories. (a mix of steady and shaky cameras, cluttered background, low resolution variation in viewpoint and illumination)

  http://www.cs.ucf.edu/~liujg/YouTube_Action_dataset.html

# Approaches

- General action recognition flowchart




- In the past decade, some approaches based on variants of features descriptors have developed for recognizing actions.
  - Spatio-Temporal Interest Points
  - Space time volume
  - Shape based representation
  - pose-based representation

# Reference


- P. Dollar, V. Rabaud, G. Cottrell, S. Belongie, ” Behavior recognition via sparse spatio-temporal features,” PETS’05
- I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, “Learning realistic human actions from movies,” CVPR’08
- J. Liu, S. Ali, M. Shah, “Recognizing human actions using multiple features”, CVPR’08
- M. Marszalek, I. Laptev, C. Schmid, ”Actions in context”, CVPR’09
- B. Chakraborty, M. B. Holte, T. B. Moeslund, J. Gonzalez, F. S. Roca, “A selective spatio-temporal interest point detector for human action recognition in complex scenes”, ICCV’11

# CS 6550 Final Project
# Face Verification

# Face Verification

- **Face recognition**: querying the input image from the stored database to recognize who the individual is.
- **Face verification**: to verify if two given face images are the same person or not.

# Dataset

- CMU PIE dataset
  - 68 people, 23 illumination conditions, 4 expressions

- LFW (Labeled Face in the Wild) Dataset
  - 13233 face images of 5749 people

(a) Same person (b) Different person

# CS 6550 Final Project
# Visual Tracking