Scene Classification of Images and Video via Semantic Segmentation

by
Heather Dunlop
Digitalsmiths Corporation

Workshop on Perceptual Organization in Computer Vision
CVPR

June 13, 2010

Digitalsmiths®
Unleash Media. Build Audience.™

Goal

- Identify scene types in video

mountain

beach

indoor

urban

Challenges

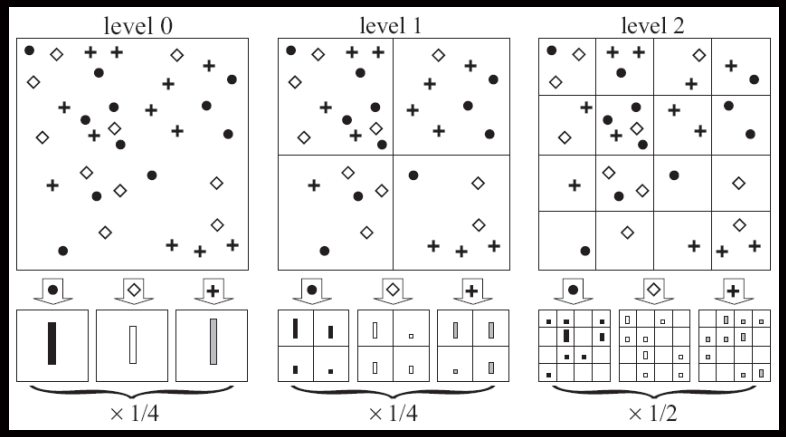
- It's not as easy as it sounds...
  - Viewpoint
  - Lighting
  - Spatial arrangement
  - Close-up shots



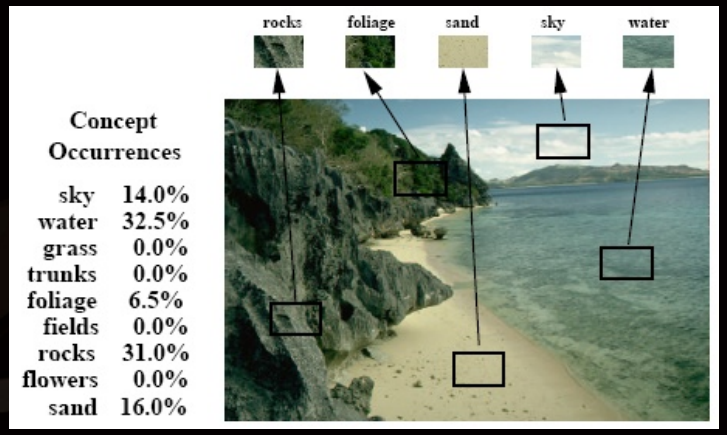



Related Work

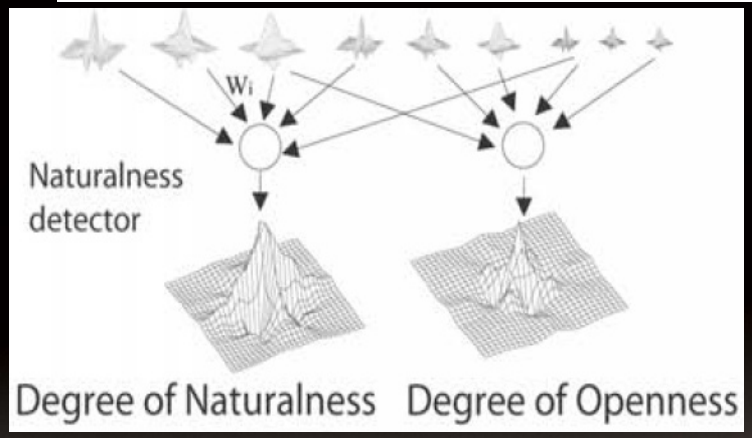
- Lots of prior work on images:
  - Lazebnik, et al. (2006)
    Spatial pyramid matching



  - Vogel & Schiele (2004)
    Semantic modeling



  - Oliva & Torralba (2001)
    Spatial envelope



Algorithm Overview

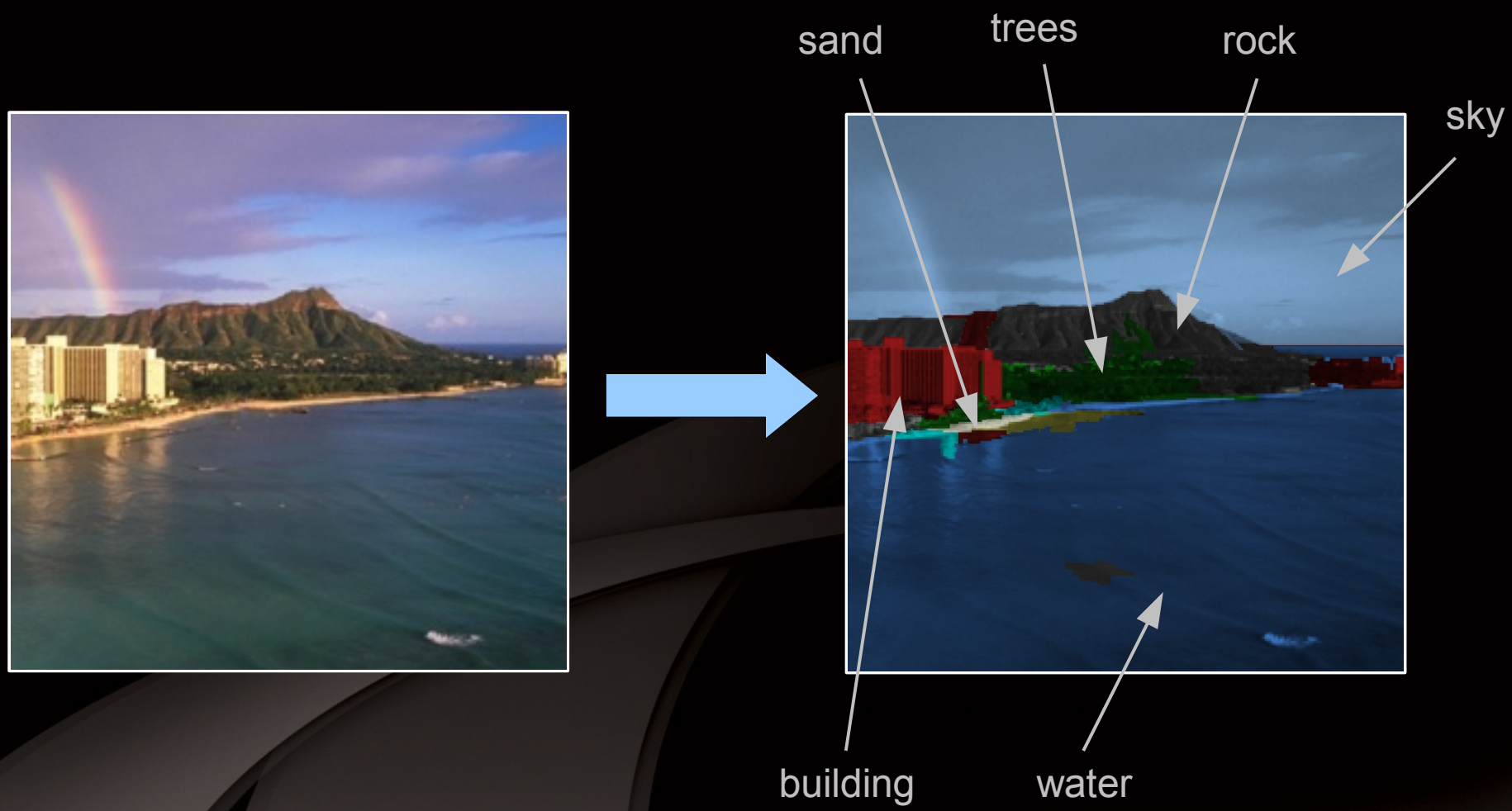
- Segment video into shots and scenes
- Select key frames
- On each key frame:
  - Classify scene as indoor or outdoor
  - If outdoor:
    - Semantic segmentation
    - Classify outdoor scene type with spatial pyramid
- Aggregate results across shots and scenes

Outline

- Semantic segmentation
- Scene classification of images
- Scene classification of video

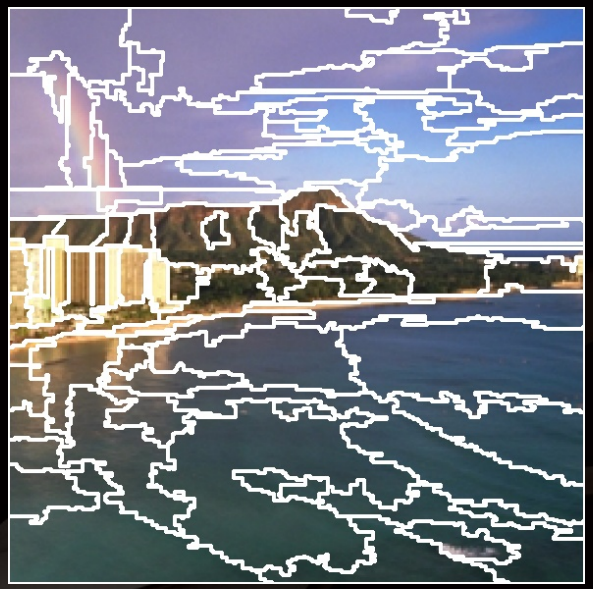
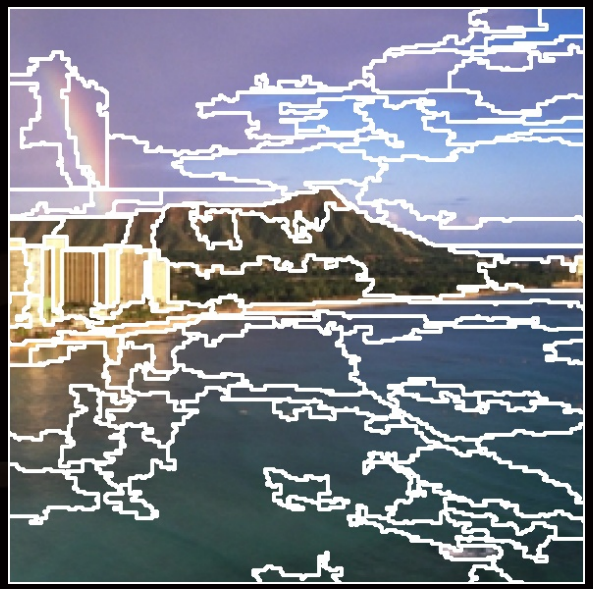
Semantic Segmentation

- Goal: predict a material label for each image pixel



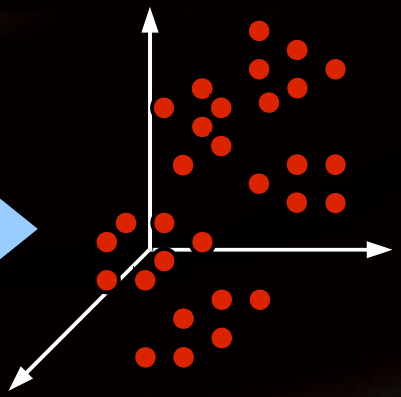
Segmentation

- Generate multiple segmentations

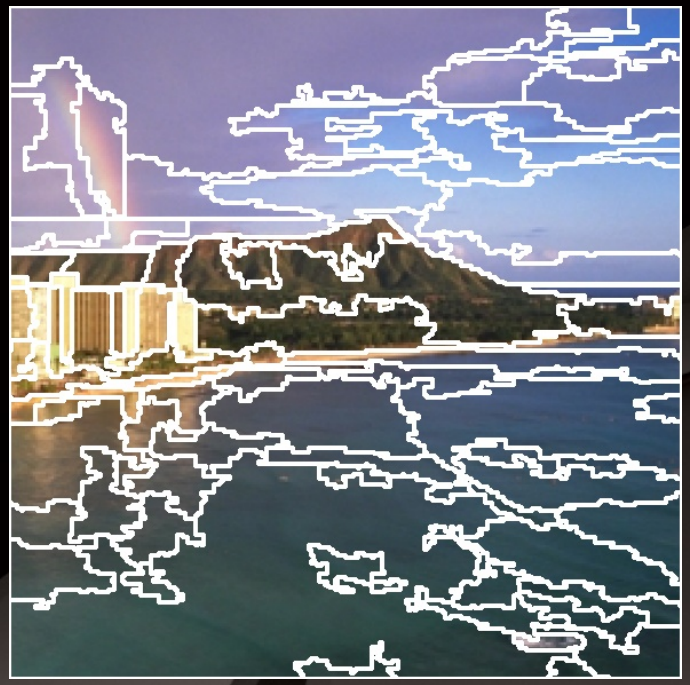
Feature Extraction

- For each segment, extract:
  - Color histogram
  - Edge strength and direction histograms
  - Line length histogram
  - Texton histogram
  - Shape metrics



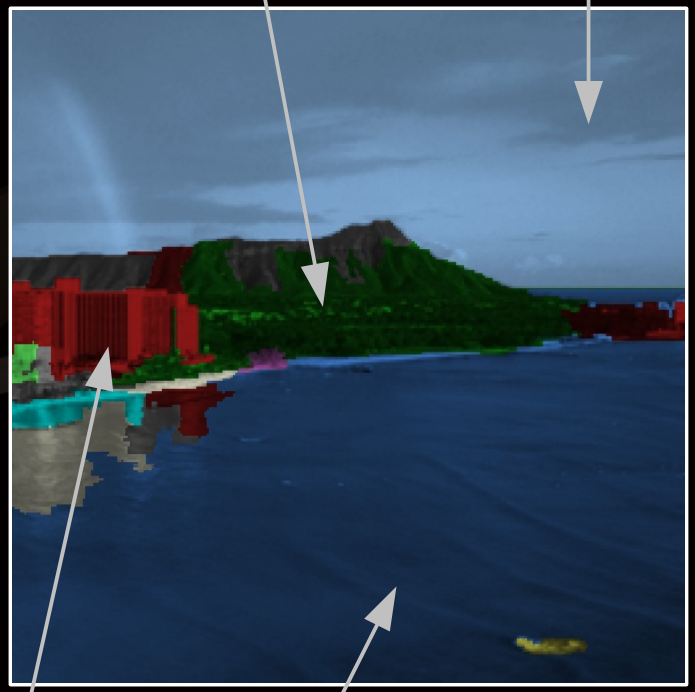
Segment Merging

- Compute feature vector for each segment
- Compute difference of feature vectors for each adjacent pair
- Using Random Forest classifier, merge those most likely to belong to same material class



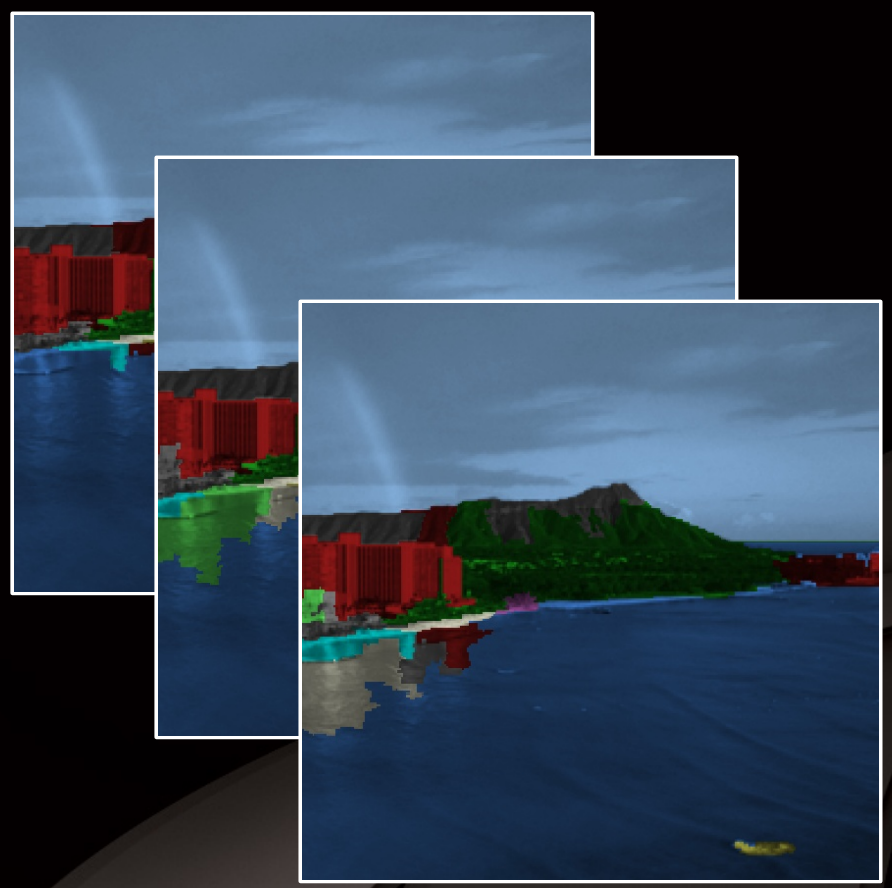
Material Classification

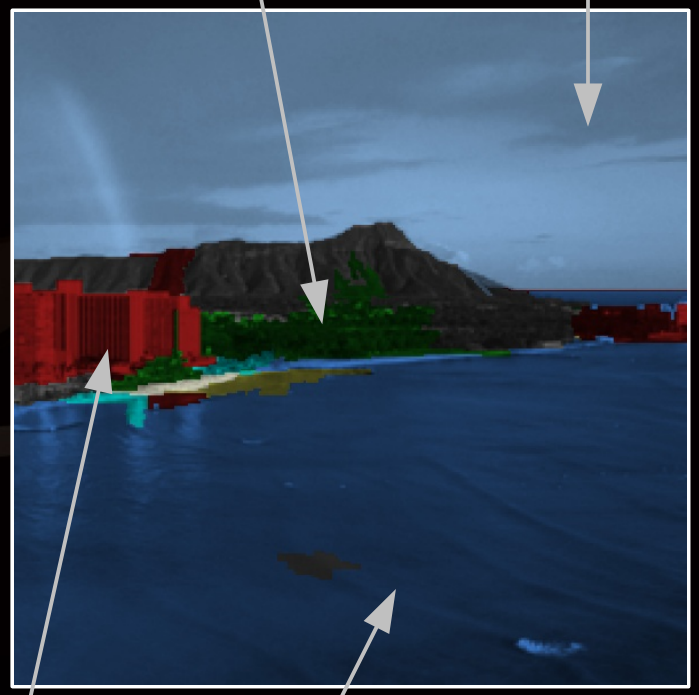
- Extract features for each region
- SVM for material classification



Semantic Segmentation Result

- Merge results across multiple segmentations





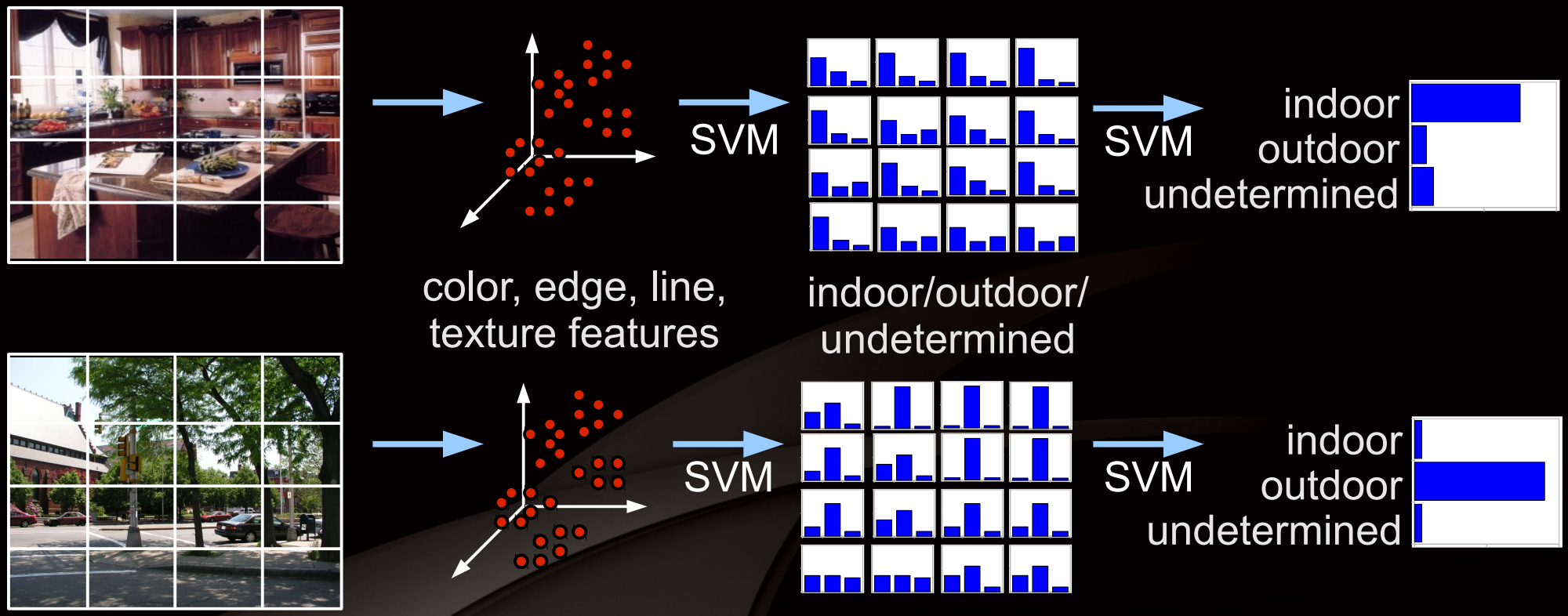
Scene Classification

- Goal:
  - Indoor

  - Outdoor

  - Undetermined

Indoor/Outdoor Classification



Outdoor Classification

- Semantic segmentation
- Spatial pyramid
- SVMs for multi-label classification

Video

- Goal: extract scene types from a sequence of frames

open water

urban

Segmenting Video

- Shot and scene boundary detection: Rasheed and Shah (2003)

frames

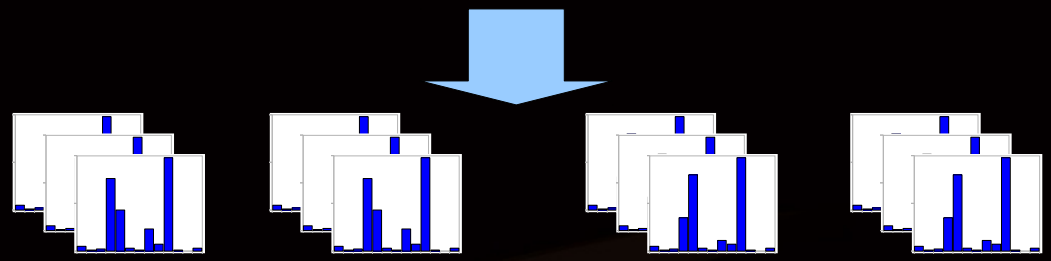
shots &
key frames

scenes

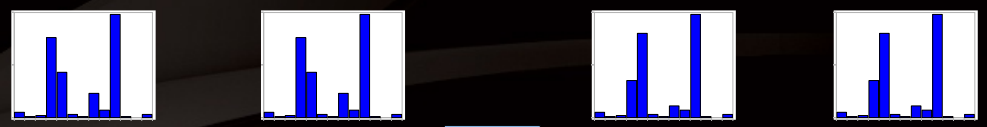
From Frames to Shots to Scenes

key frames

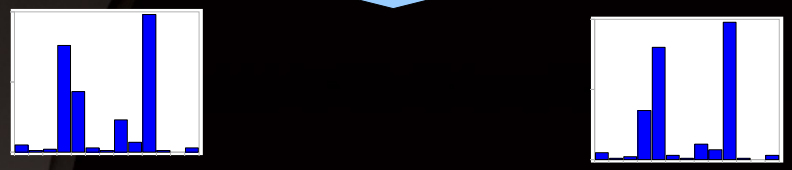
classified key frames



average across shot



95th percentile across scene



open water

urban

Experiments

- Image data set:
  - LabelMe, Google Images, movie frames
  - For semantic segmentation: 1019 images
  - For scene classification: 9855 images
- Video data set:
  - 281 videos from 49 TV shows and 6 movies (110 hours of content)
  - Each scene labeled

Sample Results

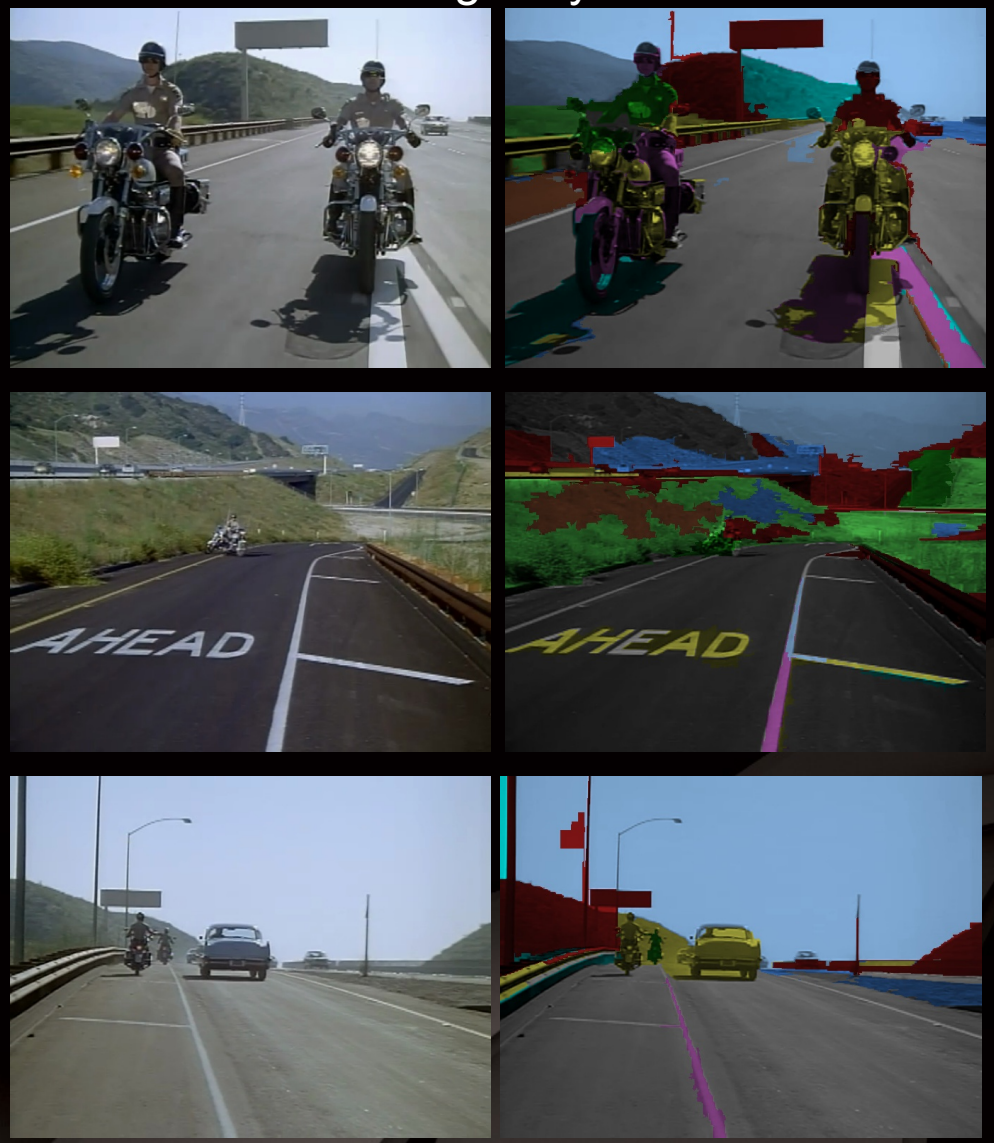
desert

forest

building
grass
person
road/sidewalk
rock
sand/gravel
sky/clouds
snow/ice
trees/bushes
vehicle
water
miscellaneous

Sample Results

highway



snow

building
grass
person
road/sidewalk
rock
sand/gravel
sky/clouds
snow/ice
trees/bushes
vehicle
water
miscellaneous

Sample Results

open water

urban

building
grass
person
road/sidewalk
rock
sand/gravel
sky/clouds
snow/ice
trees/bushes
vehicle
water
miscellaneous

Sample Results

indoor

undetermined

Semantic Segmentation

True Class \ Predicted Class	building	grass	person	road/sidewalk	rock	sand/gravel	sky/clouds	snow/ice	trees/bushes	vehicle	water	miscellaneous
building	71	1	6	1	2	2	5	0	2	7	1	3
grass	1	71	1	0	3	7	0	0	11	0	3	2
person	5	1	75	2	1	0	2	0	4	3	1	6
road/sidewalk	7	1	3	52	2	12	0	3	0	4	15	1
rock	4	4	7	3	47	16	0	0	13	1	1	4
sand/gravel	4	9	3	3	14	48	2	6	3	1	2	6
sky/clouds	2	0	0	2	1	1	91	1	0	0	3	0
snow/ice	4	0	2	6	2	2	13	41	0	2	28	1
trees/bushes	4	3	2	0	3	1	2	1	77	1	3	4
vehicle	22	0	15	2	4	0	0	1	1	51	0	3
water	1	3	1	3	1	11	8	4	1	1	66	1
miscellaneous	13	6	26	1	7	2	3	0	4	13	1	23

Scene Classification on Photographs

	Our Method	Lazebnik et al.
Coast/beach	**.60**	.44
Desert	**.76**	.48
Forest	.71	**.84**
Grassland	**.79**	.56
Highway	.67	**.79**
Lake/River	**.44**	.42
Mountainous	.73	**.81**
Open Water	**.70**	.67
Sky	.82	**.83**
Snow	**.75**	.69
Urban	**.90**	.87
Outdoor	.94	**.99**
Indoor	.73	**.87**
Average	**.73**	.71

Up to 28% per-category improvement.

Scene Classification on Video

	Keyframes	Scenes
Coast/beach	.13	.34
Desert	.04	.09
Forest	.29	.45
Grassland	.32	.47
Highway	.16	.33
Lake/River	.02	.07
Mountainous	.05	.11
Open Water	.33	.52
Sky	.24	.34
Snow	.04	.08
Urban	.33	.62
Outdoor	.67	.86
Indoor	.72	.82
Average	**.26**	.39

Method for aggregating across shots and scenes produces 13% improvement.

Conclusions

- We have developed a system that integrates:
  - segmentation
  - recognition of scene components
  - classification of whole images and video sequences
- Techniques that address the unique properties of video are a necessity

Future Work

- Incorporating face and body detection to identify when background is obstructed
- Background/motion segmentation
- Bag of features techniques for classifying material concepts

Thank you!

Questions?

Digitalsmiths®
Unleash Media. Build Audience.™