Machine Learning Problems

	Supervised Learning	***Unsupervised Learning***
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

1

Machine Learning Problems

	Supervised Learning	***Unsupervised Learning***
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

2

The machine learning framework

- Apply a prediction function to a feature representation of the image to get the desired output:

f() = "apple"

f() = "tomato"

f() = "cow"

Slide credit: L. Lazebnik

3

The machine learning framework

$$y = f(\mathbf{x})$$

output | prediction function | Image feature

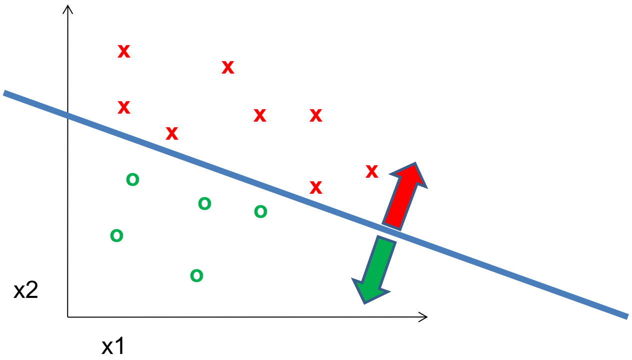
- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- **Testing:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Slide credit: L. Lazebnik

4

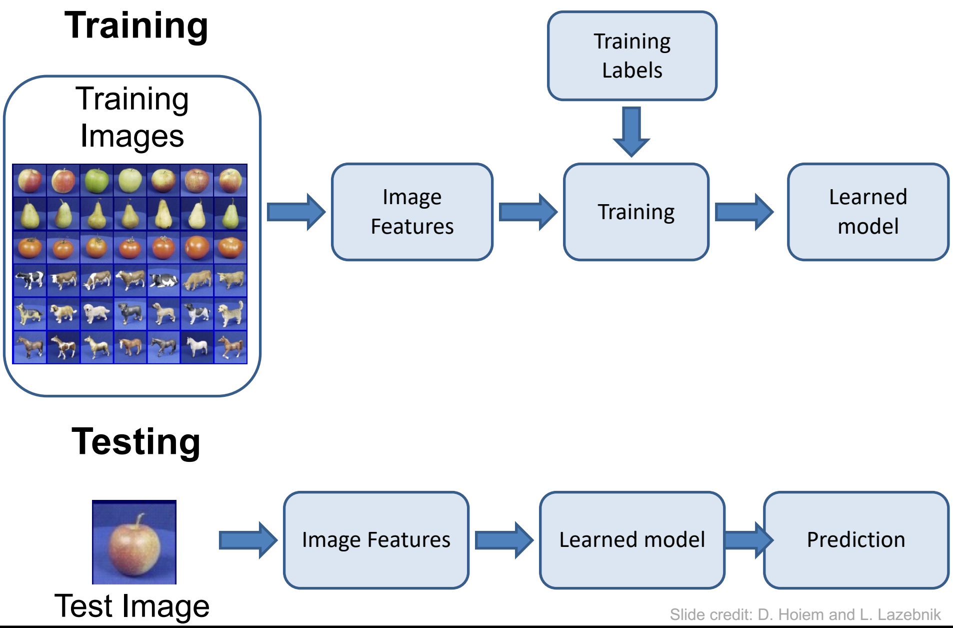
Learning a classifier

Given some set of features with corresponding labels, learn a function to predict the labels from the features



5

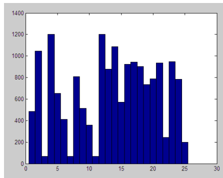
Steps

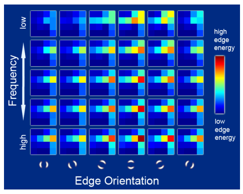


6

Features

- Raw pixels
- Histograms
- GIST descriptors



Slide credit: L. Lazebnik

7

One way to think about it…

- Training labels dictate that two examples are the same or different, in some sense
- Features and distance measures define visual similarity
- Classifiers try to learn weights or parameters for features and distance measures so that visual similarity predicts label similarity

8

Many classifiers to choose from

- SVM
- Neural networks
- Naïve Bayes
- Bayesian network
- Logistic regression
- Randomized Forests
- Boosted Decision Trees
- K-nearest neighbor
- RBMs
- Deep Convolutional Network
- Etc.

Which is the best one?

9

Claim:

The decision to *use* machine learning is more important than the choice of a *particular* learning method.

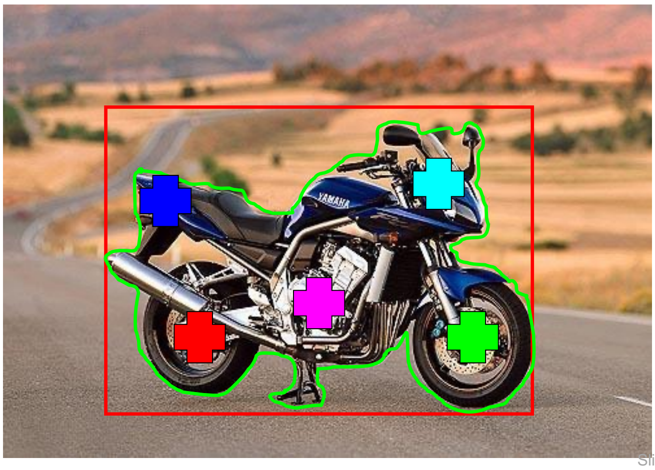
*Deep learning seems to be an exception to this, at the moment, probably because it is learning the feature representation.

10

Recognition task and supervision

- Images in the training set must be annotated with the "correct answer" that the model is expected to produce

Contains a motorbike



Slide credit: L. Lazebnik

11

Spectrum of supervision

Less **More**

Unsupervised

"Weakly" supervised

Fully supervised

Definition depends on task

Slide credit: L. Lazebnik

12

Generalization

Training set (labels known)

Test set (labels unknown)

- How well does a learned model generalize from the data it was trained on to a new test set?

Slide credit: L. Lazebnik

13

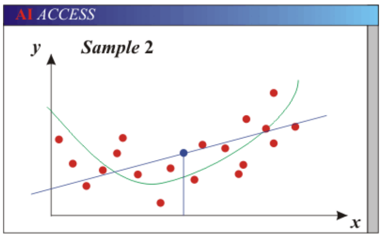
Generalization

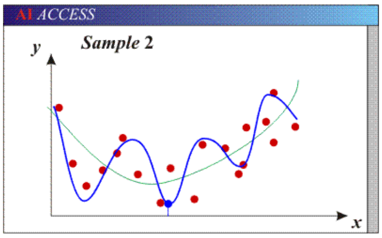
- Components of generalization error
  - **Bias:** how much the average model over all training sets differ from the true model?
    - Error due to inaccurate assumptions/simplifications made by the model.
  - **Variance:** how much models estimated from different training sets differ from each other.
- **Underfitting:** model is too "simple" to represent all the relevant class characteristics
  - High bias (few degrees of freedom) and low variance
  - High training error and high test error
- **Overfitting:** model is too "complex" and fits irrelevant characteristics (noise) in the data
  - Low bias (many degrees of freedom) and high variance
  - Low training error and high test error

Slide credit: L. Lazebnik

14

Bias-Variance Trade-off





- Models with too few parameters are inaccurate because of a large bias (not enough flexibility).
- Models with too many parameters are inaccurate because of a large variance (too much sensitivity to the sample).

Slide credit: D. Hoiem

15

Bias-Variance Trade-off

E(MSE) = noise^2 + bias^2 + variance

Unavoidable error

Error due to incorrect assumptions

Error due to variance of training samples

See the following for explanations of bias-variance (also Bishop's "Neural Networks" book):

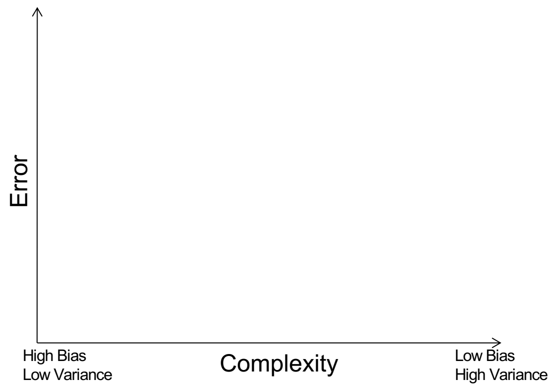
•http://www.inf.ed.ac.uk/teaching/courses/mlsc/Notes/Lecture4/BiasVariance.pdf

Slide credit: D. Hoiem

16

Bias-variance tradeoff

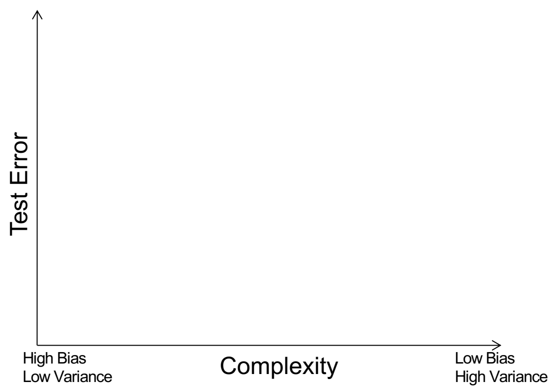
Underfitting **Overfitting**



Slide credit: D. Hoiem

17

Bias-variance tradeoff

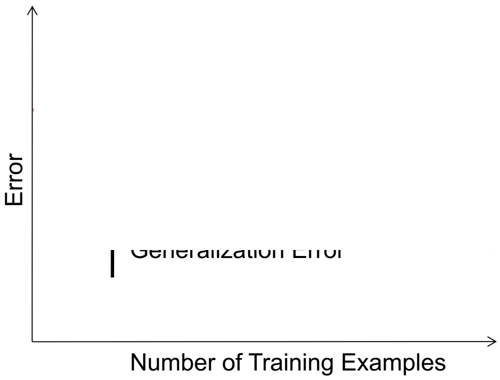


Slide credit: D. Hoiem

18

Effect of Training Size

Fixed prediction model



Slide credit: D. Hoiem

19

Remember…

- No classifier is inherently better than any other: you need to make assumptions to generalize

- Three kinds of error
  - Inherent: unavoidable
  - Bias: due to over-simplifications
  - Variance: due to inability to perfectly estimate parameters from limited data

Slide credit: D. Hoiem

20

How to reduce variance?

- Choose a simpler classifier
- Regularize the parameters
- Get more training data

Slide credit: D. Hoiem

21

Supervised learning

$$y = f(\mathbf{x})$$

output

prediction function

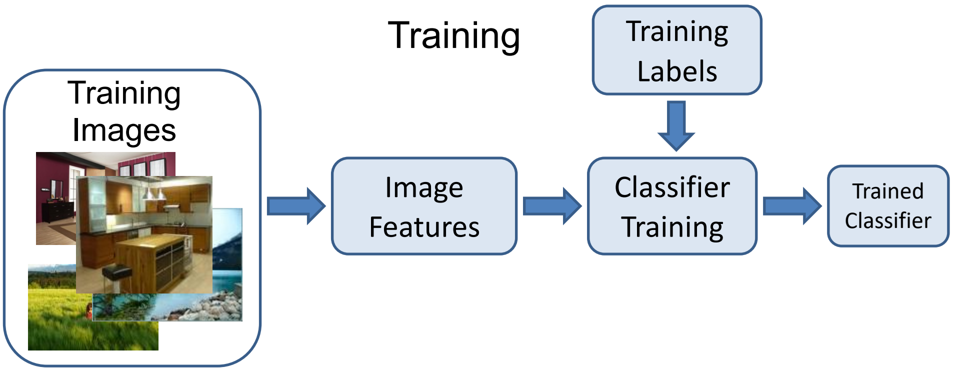
Image feature

- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- **Testing:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Slide credit: L. Lazebnik

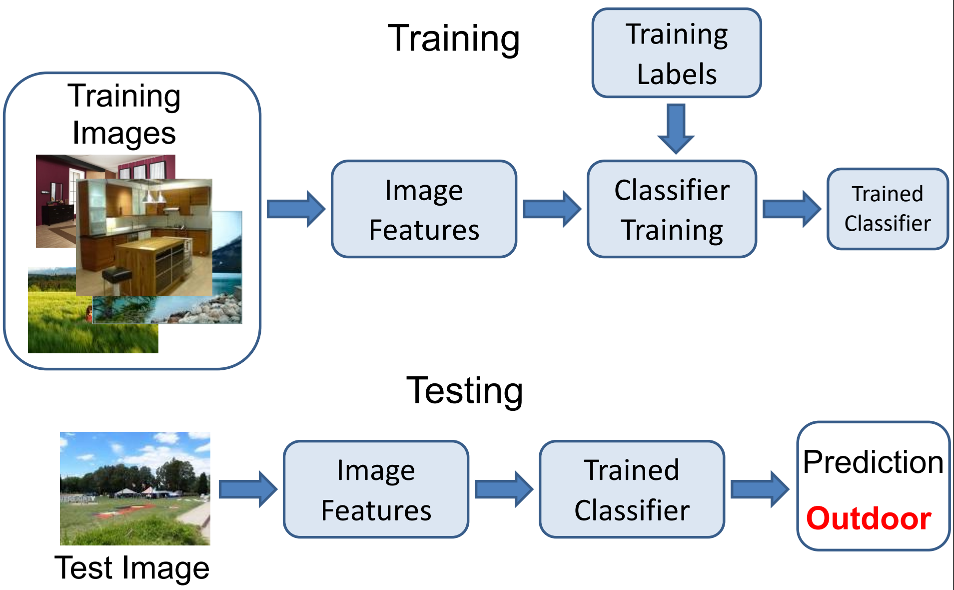
22

Image Categorization



23

Image Categorization



24

Example: Scene Categorization

- Is this a kitchen?



25

Image features

Training

Training Labels

Training Images

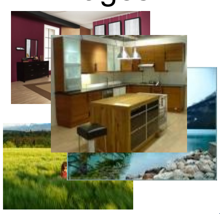


Image Features

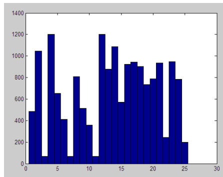
Classifier Training

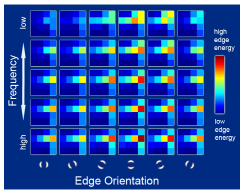
Trained Classifier

26

Features

- Raw pixels
- Histograms
- GIST descriptors
- …

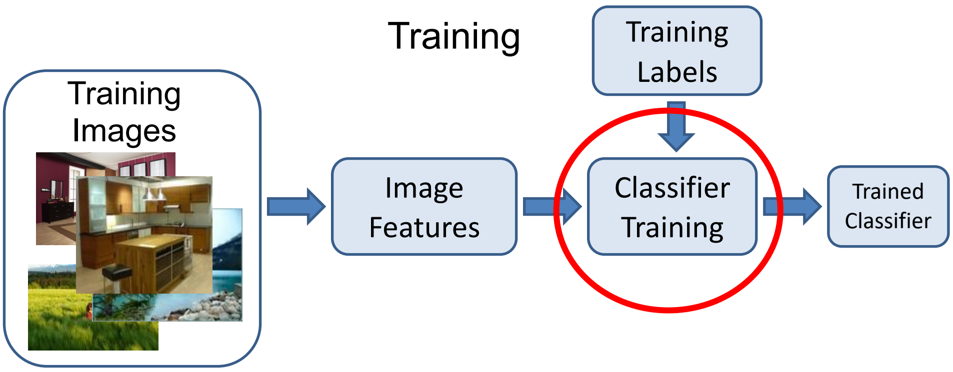





Slide credit: L. Lazebnik

27

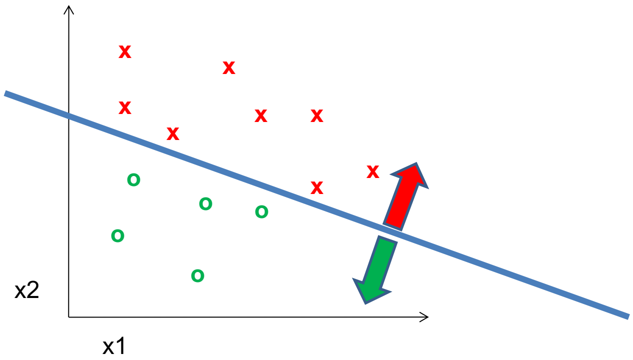
Classifiers



28

Learning a classifier

Given some set of features with corresponding labels, learn a function to predict the labels from the features



29

Very brief tour of some classifiers

- **K-nearest neighbor**
- **SVM**
- Boosted Decision Trees
- Neural networks
- Naïve Bayes
- Bayesian network
- Logistic regression
- Randomized Forests
- RBMs
- Etc.

30

Generative vs. Discriminative Classifiers

Generative Models

- Represent both the data and the labels
- Often, makes use of conditional independence and priors
- Examples
  - Naïve Bayes classifier
  - Bayesian network
- Models of data may apply to future prediction problems

Discriminative Models

- Learn to directly predict the labels from the data
- Often, assume a simple boundary (e.g., linear)
- Examples
  - Logistic regression
  - SVM
  - Boosted decision trees
- Often easier to predict a label from the data than to model the data

Slide credit: D. Hoiem

31

What to remember about classifiers

- No free lunch: machine learning algorithms are tools, not dogmas
- Try simple classifiers first
- Better to have smart features and simple classifiers than simple features and smart classifiers
- Use increasingly powerful classifiers with more training data (bias-variance tradeoff)

Slide credit: D. Hoiem

32

Making decisions about data

- 3 important design decisions:
  1) What data do I use?
  2) How do I represent my data (what feature)?
  3) What classifier / regressor / machine learning tool do I use?
- These are in decreasing order of importance
- Deep learning addresses 2 and 3 simultaneously (and blurs the boundary between them).
- You can take the representation from deep learning and use it with any classifier.

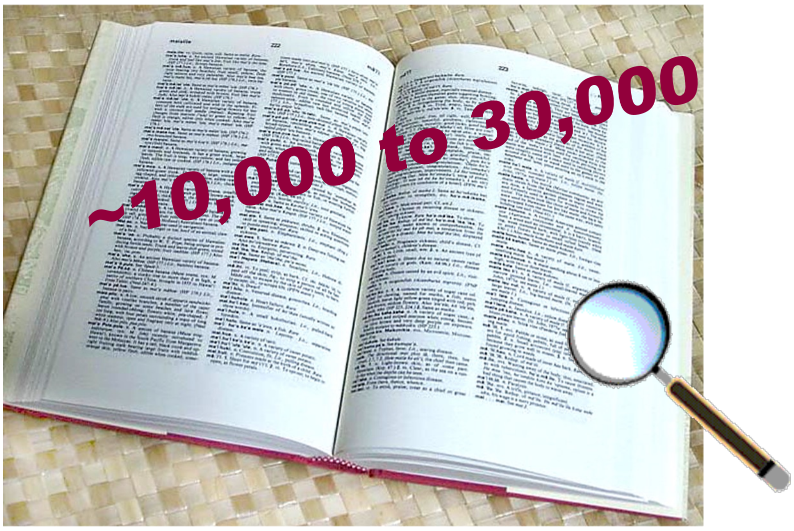
33

Recognition: Overview and History

Slides from Lana Lazebnik, Fei-Fei Li, Rob Fergus, Antonio Torralba, and Jean Ponce

34

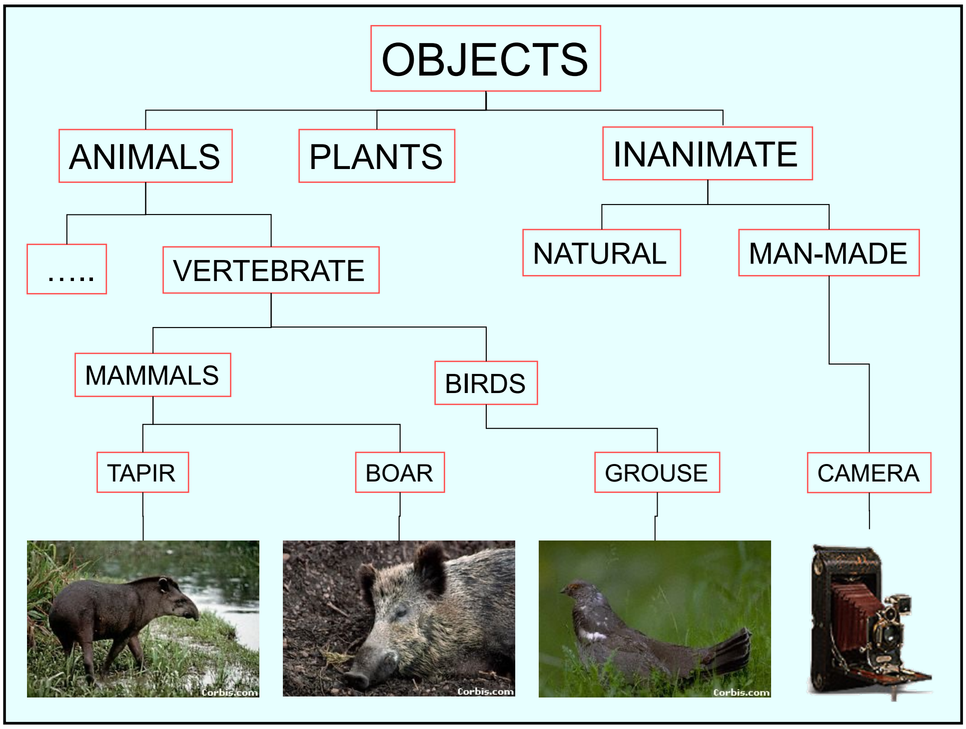
How many visual object categories are there?



Biederman 1987

35

36



37

Specific recognition tasks

Svetlana Lazebnik

38

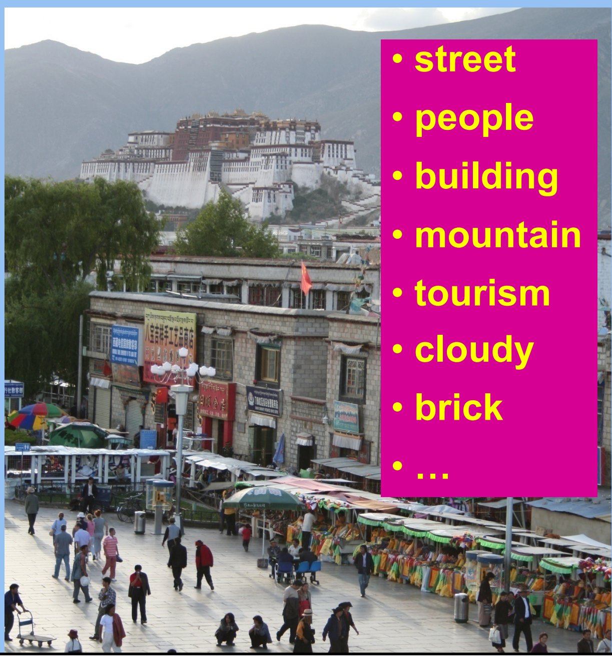
Scene categorization or classification

- outdoor/indoor
- city/forest/factory/etc.

Svetlana Lazebnik

39

Image annotation / tagging / attributes

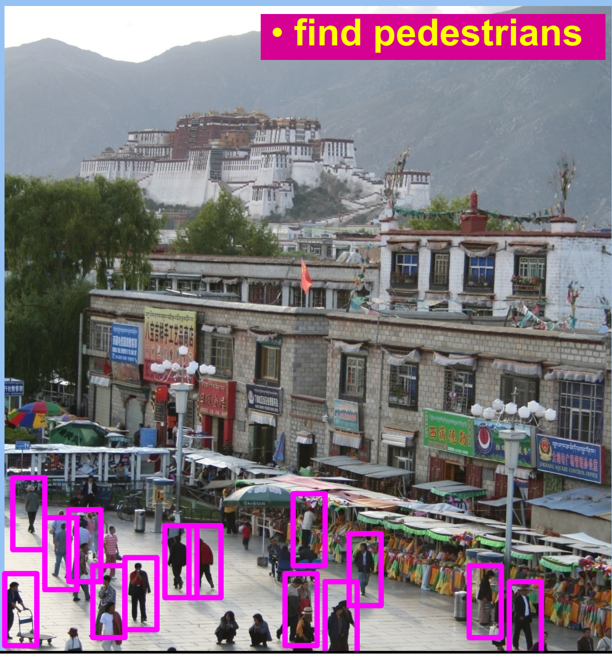


- street
- people
- building
- mountain
- tourism
- cloudy
- brick
- …

Svetlana Lazebnik

40

Object detection



Svetlana Lazebnik

41

Image parsing / semantic segmentation

Svetlana Lazebnik

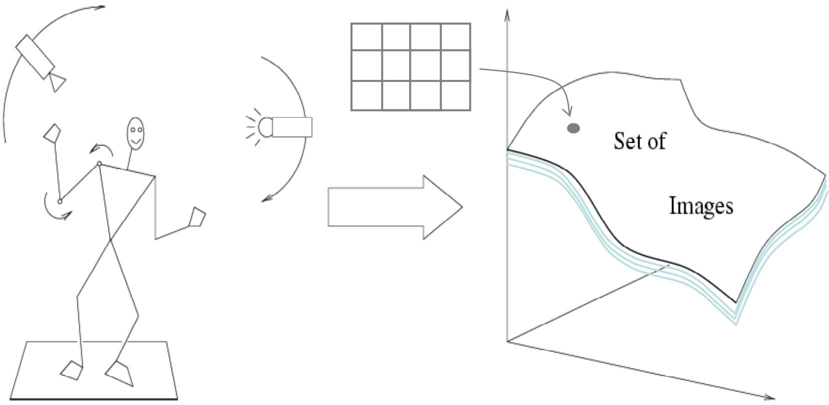
42

Scene understanding?

Svetlana Lazebnik

43

Recognition is all about modeling variability



Variability: Camera position
Illumination
Shape parameters

Within-class variations?

Svetlana Lazebnik

44

Within-class variations

Svetlana Lazebnik

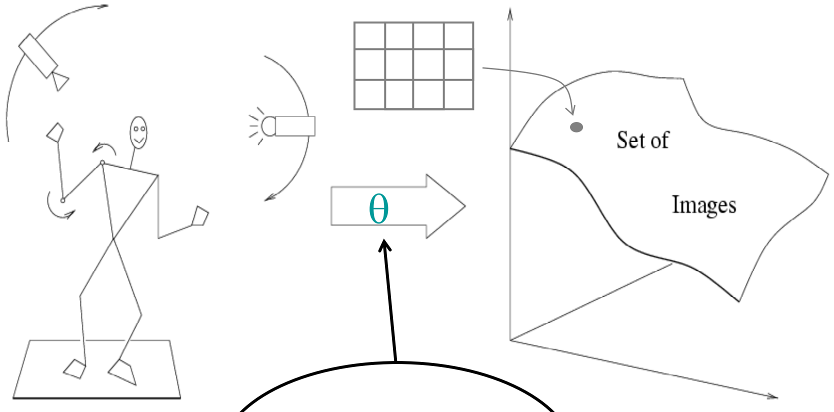
45

History of ideas in recognition

- 1960s – early 1990s: the geometric era

Svetlana Lazebnik

46



Variability: Camera position
Illumination

Alignment

Shape: assumed known

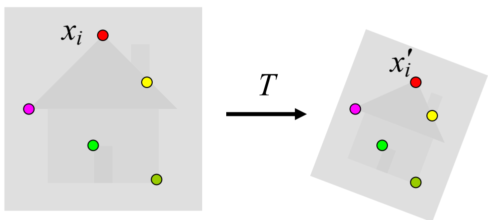
Roberts (1965); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)

Svetlana Lazebnik

47

Recall: Alignment

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



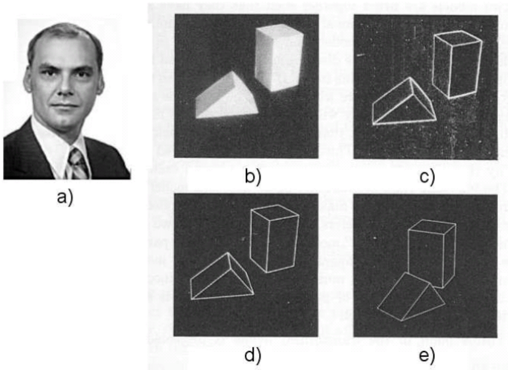
Find transformation T that minimizes

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

Svetlana Lazebnik

48

Recognition as an alignment problem: Block world



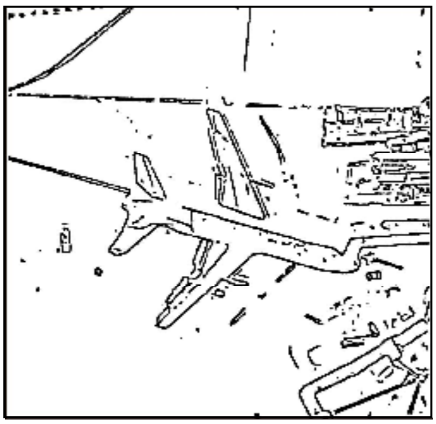
L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

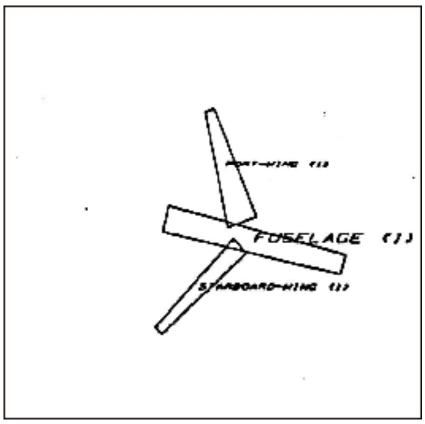
Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

J. Mundy, Object Recognition in the Geometric Era: a Retrospective, 2006

49

Representing and recognizing object categories is harder...





ACRONYM (Brooks and Binford, 1981)

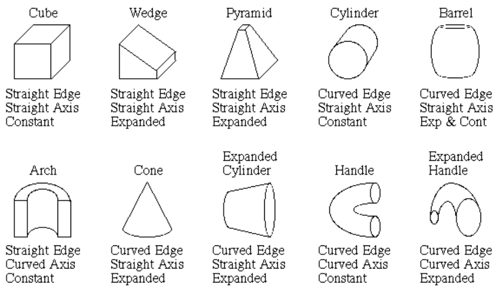
Binford (1971), Nevatia & Binford (1972), Marr & Nishihara (1978)

53

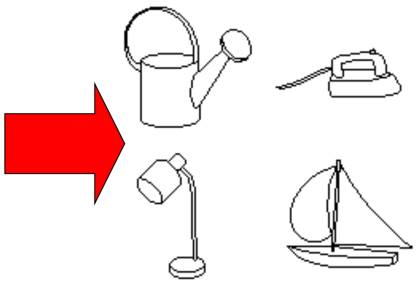
Recognition by components

Biederman (1987)

Primitives (geons)



Objects

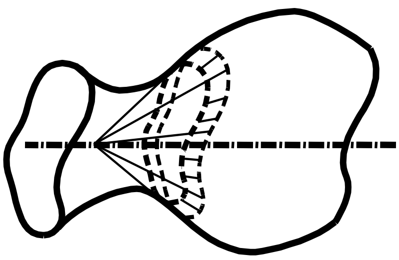


http://en.wikipedia.org/wiki/Recognition_by_Components_Theory

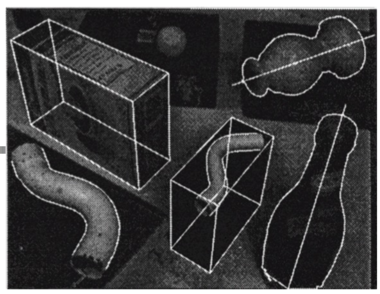
Svetlana Lazebnik

54

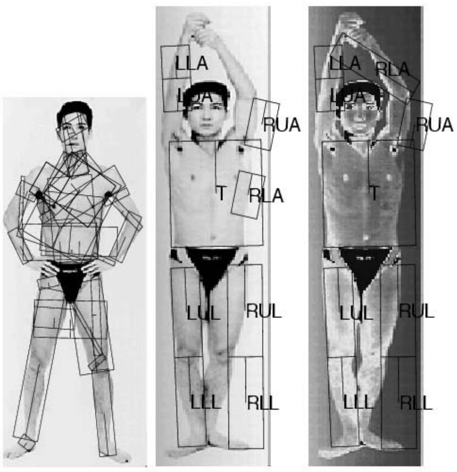
General shape primitives?



Generalized cylinders
Ponce et al. (1989)



Zisserman et al. (1995)



Forsyth (2000)

Svetlana Lazebnik

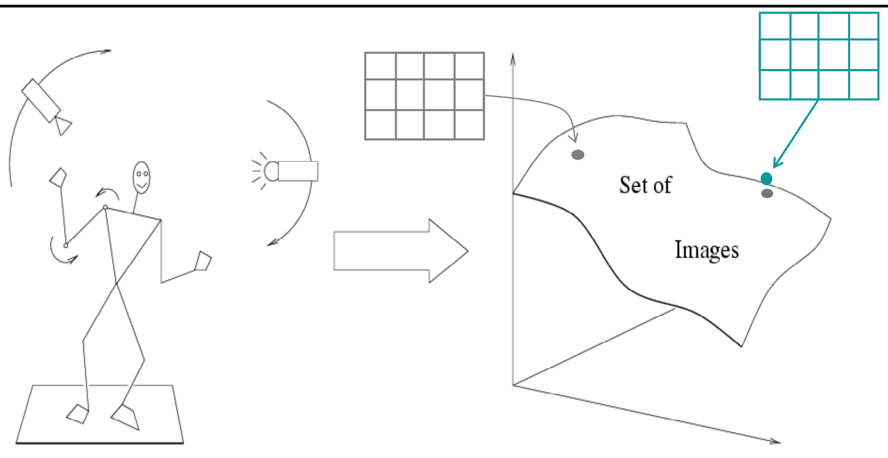
55

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models

Svetlana Lazebnik

56



Empirical models of image variability

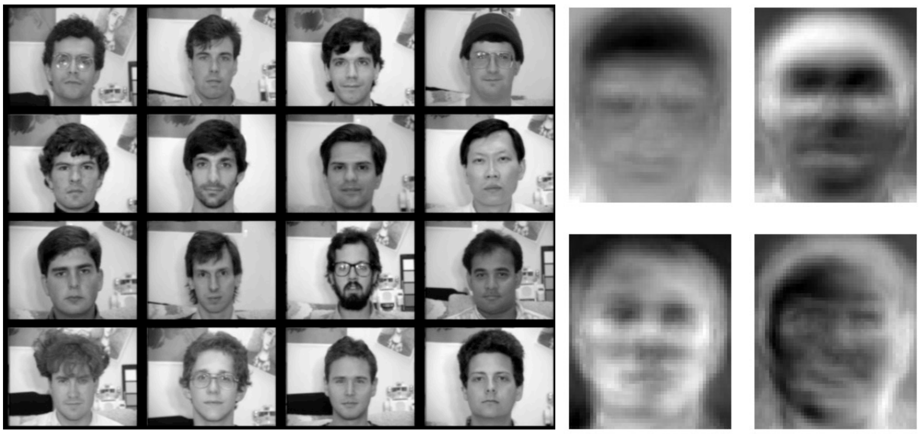
Appearance-based techniques

Turk & Pentland (1991); Murase & Nayar (1995); etc.

Svetlana Lazebnik

57

Eigenfaces (Turk & Pentland, 1991)

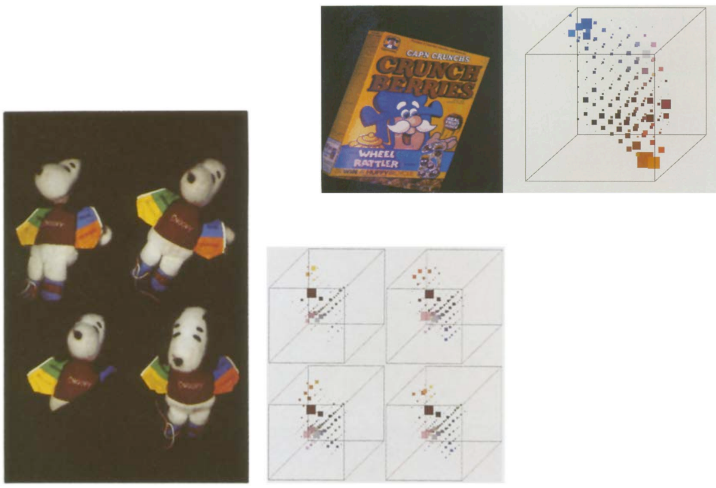


Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20

Svetlana Lazebnik

58

Color Histograms



Swain and Ballard, Color Indexing, IJCV 1991.

Svetlana Lazebnik

59

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- 1990s – present: sliding window approaches

Svetlana Lazebnik

62

Sliding window approaches

63

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

64

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features

Svetlana Lazebnik

65

Local features for object instance recognition

D. Lowe (1999, 2004)

66

Large-scale image search

Combining local features, indexing, and spatial constraints

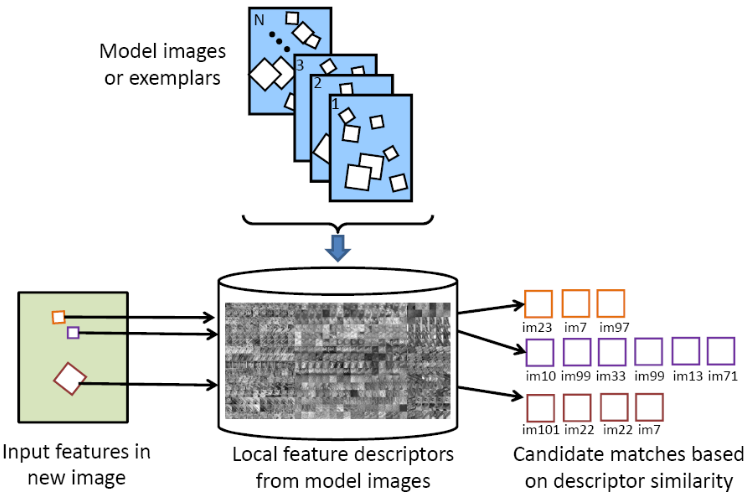


Image credit: K. Grauman and B. Leibe

67

Large-scale image search

Combining local features, indexing, and spatial constraints

Philbin et al. '07

68

Large-scale image search

Combining local features, indexing, and spatial constraints

Svetlana Lazebnik

69

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models

70

Parts-and-shape models

- Model:
  - Object as a set of parts
  - Relative locations between parts
  - Appearance of part

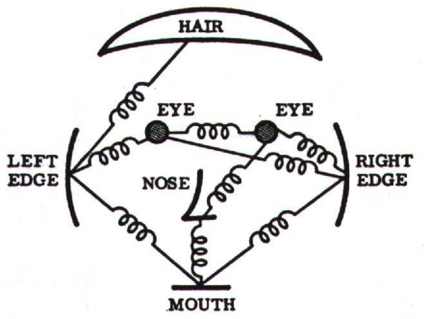


Figure from [Fischler & Elschlager 73]

71

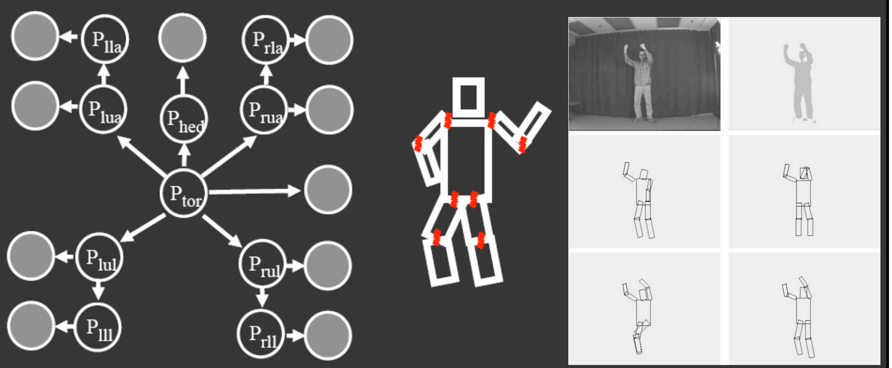
Constellation models

Weber, Welling & Perona (2000), Fergus, Perona & Zisserman (2003)

72

Pictorial structure model

Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)



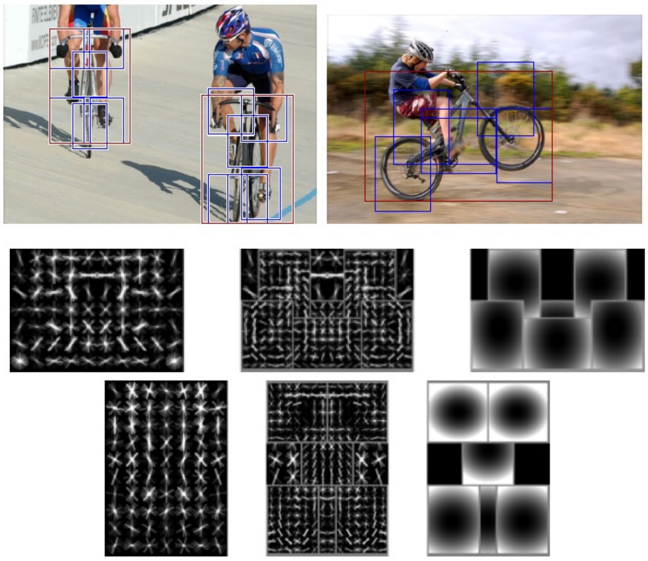
$$\Pr(P_{tor}, P_{arm}, \ldots | Im) \propto \prod_{i,j} \Pr(P_i \mid P_j) \prod_i \Pr(Im(P_i))$$

part geometry

part appearance

73

Discriminatively trained part-based models



P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," PAMI 2009

74

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features

Svetlana Lazebnik

75

Bag-of-features models

Svetlana Lazebnik

76

Bag-of-features models

Object → **Bag of 'words'**

Svetlana Lazebnik

77

Objects as texture

- All of these are treated as being the same

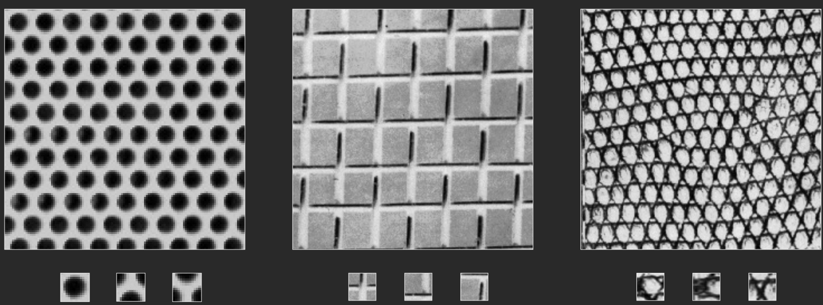
- No distinction between foreground and background: scene recognition?

Svetlana Lazebnik

78

Origin 1: Texture recognition

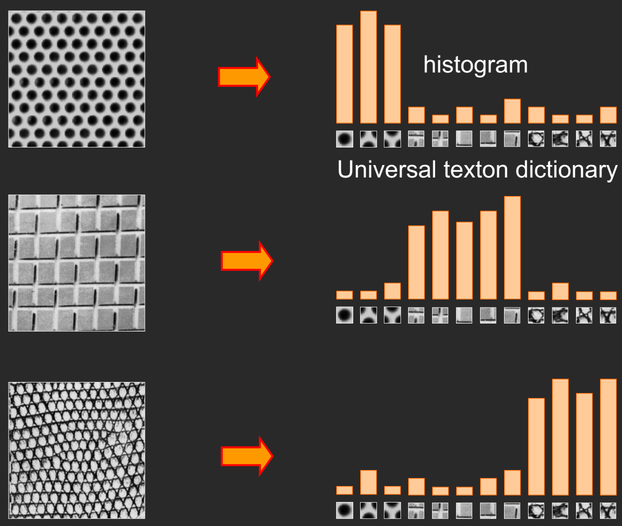
- Texture is characterized by the repetition of basic elements or *textons*
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

79

Origin 1: Texture recognition



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

80

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

81

Origin 2: Bag-of-words models

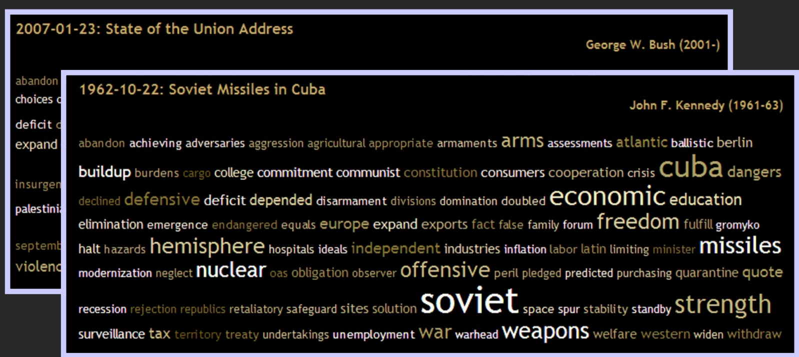
- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

82

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)



US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

83

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

US Presidential Speeches Tag Cloud
`http://chir.ag/phernalia/preztags/`

84

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Early-2010s: CNNs

Svetlana Lazebnik

85