

**CLASSIFICATION/RECOGNITION**

**Readings:**

**Szeliski: Chapter 14.1,14.3**

**Some Slides adapted from Univ. of Washington**

**SONKA et.al**

**Classification Chapter-9.2**

Horst Bischof Computer Vision II SS11

### What is pattern recognition?

"The assignment of a physical object or event to one of several prespecified categories" -  
- Duda & Hart

- A **pattern** is an object, process or event that can be given a name.
- A **pattern class** (or category) is a set of patterns sharing common attributes and usually originating from the same source.
- During **recognition** (or **classification**) given objects are assigned to prescribed classes.
- A **classifier** is a machine which performs classification.

Horst Bischof Computer Vision II SS11

### Examples of applications

- **Optical Character Recognition (OCR)**
  - Handwritten: sorting letters by postal code, input device for PDA's.
  - Printed texts: reading machines for blind people, digitalization of text documents.
- **Biometrics**
  - Face recognition, verification, retrieval.
  - Finger prints recognition.
  - Speech recognition.
- **Diagnostic systems**
  - Medical diagnosis: X-Ray, EKG analysis.
  - Machine diagnostics, waster detection.
- **Military applications**
  - Automated Target Recognition (ATR).
  - Image segmentation and analysis (recognition from aerial or satellite photographs).


Horst Bischof Computer Vision II SS11

### Approaches

- **Statistical PR:** based on underlying statistical model of patterns and pattern classes.
- **Structural (or syntactic) PR:** pattern classes represented by means of formal structures as grammars, automata, strings, etc.
- **Neural networks:** classifier is represented as a network of cells modeling neurons of the human brain (connectionist approach).

Horst Bischof Computer Vision II SS11

### Classification



```

    graph TD
      Input[Input Image] --> Preprocessing[Preprocessing]
      Preprocessing --> FeatureExtraction[Feature extraction]
      FeatureExtraction --> Classification[Classification]
      Classification --> Output["'salmon' 'sea bass'"]
  
```

Horst Bischof Computer Vision II SS11

### Pattern Recognition & IP

```

    graph TD
      Input[Input] --> Sensing[sensing]
      Sensing --> Segmentation[segmentation]
      Segmentation --> FeatureExtraction[feature extraction]
      FeatureExtraction --> Classification[classification]
      Classification --> PostProcessing[post-processing]
      PostProcessing --> Decision[decision]
      
      Costs[costs] --> PostProcessing
      Context[adjustments for context] --> PostProcessing
      Missing[adjustments for missing features] --> Classification
  
```

Horst Bischof Computer Vision II SS11

**Features, Patterns**

**Featurespaces**

Horst Bischof Computer Vision II SS11 7

**Merkmale (Features):**

Horst Bischof Computer Vision II SS11 8

**Featurespace**

Horst Bischof Computer Vision II SS11 9

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

Spectral band 4  
Spectral band 3  
Spectral band 2  
Spectral band 1

$X_1$

**Classification Principle**

Horst Bischof Computer Vision II SS11 11

**Classification**

Feature Vector  $\rightarrow$  Class Label

$$\Theta : X \rightarrow \Omega$$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \rightarrow \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_c \end{pmatrix} = \Omega$$

Horst Bischof Computer Vision II SS11 12

### Classifier

A classifier partitions feature space  $X$  into **class-labeled regions** such that  
 $X = X_1 \cup X_2 \cup \dots \cup X_{|Y|}$  and  $X_1 \cap X_2 \cap \dots \cap X_{|Y|} = \{0\}$

The classification consists of determining to which region a feature vector  $x$  belongs to.

Borders between **decision boundaries** are called decision regions.

Horst Bischof Computer Vision II SS11 13

### A priori probability

$P(\omega_1), P(\omega_2) \dots$  a priori probability

sea bass salmon

$P(\omega_1) + P(\omega_2) = 1$  (Only  $\omega_1$  and  $\omega_2$ )

Simple Classification:

$$S = \begin{cases} \omega_1 & \text{if } P(\omega_1) > P(\omega_2) \\ \omega_2 & \text{else} \end{cases} \quad \text{Without Features}$$

Horst Bischof Computer Vision II SS11 14

### Class-conditional probability density

$\int_{-\infty}^{\infty} p(x|\omega_1) dx = 1$

$\int_{-\infty}^{\infty} p(x|\omega_2) dx = 1$

Feature (e.g. Brightness)

Horst Bischof Computer Vision II SS11 15

### Posterior Probability

Bayes Theorem:

$$P(\omega_j|\bar{x}) = \frac{p(\bar{x}|\omega_j)P(\omega_j)}{p(\bar{x})} \quad \text{with} \quad p(\bar{x}) = \sum_{j=1}^c p(\bar{x}|\omega_j)P(\omega_j)$$

Classification:

$$S = \begin{cases} \omega_1 & \text{falls } P(\omega_1|\bar{x}) > P(\omega_2|\bar{x}) \\ \omega_2 & \text{sonst} \end{cases}$$

Error Probability:  $P(\text{error}|\bar{x}) = \begin{cases} P(\omega_1|\bar{x}) & \text{falls } S = \omega_2 \\ P(\omega_2|\bar{x}) & \text{falls } S = \omega_1 \end{cases}$

Horst Bischof Computer Vision II SS11 16

### Posterior Probability

$P(\omega_1|\bar{x}) + P(\omega_2|\bar{x}) = 1$

Horst Bischof Computer Vision II SS11 17

### Representation of classifier

A classifier is typically represented as a set of discriminant functions

$$f_i(\mathbf{x}) : X \rightarrow \mathcal{R}, i = 1, \dots, |Y|$$

The classifier assigns a feature vector  $\mathbf{x}$  to the  $i$ -th class if  $f_i(\mathbf{x}) > f_j(\mathbf{x}) \quad \forall j \neq i$

Feature vector  $\mathbf{x}$  →  $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_{|Y|}(\mathbf{x})$  → max → Class identifier  $y$

Discriminant function

Horst Bischof Computer Vision II SS11 18

### Decision Function:

Horst Bischof Computer Vision II SS11 19

### Bayes Classifier

Classifier:  $\Theta_{Bayes}(\vec{x}) = S_{\max}(\Phi_{Bayes}, \vec{x})$

$$S_{\max}(\Phi, \vec{x}) = \omega_i \text{ falls } \varphi_i(\vec{x}) = \max_{j=1,2,\dots,c} \varphi_j(\vec{x})$$

Always optimal decision  $\varphi_{Bayes}(\vec{x}) = P(\omega_i | \vec{x})$

not known!!

$$P(\omega_i | \vec{x})$$

Horst Bischof Computer Vision II SS11 20

### Bayes Classifier

Decision Function  $\varphi_{Bayes_i}(\vec{x}) = P(\omega_i | \vec{x})$

Bayes Theorem:

$$P(\omega_i | \vec{x}) = \frac{p(\vec{x} | \omega_i)P(\omega_i)}{\sum_{k=1}^c p(\vec{x} | \omega_k)P(\omega_k)}$$

Decision Function

$$\varphi_{Bayes_i}(\vec{x}) = p(\vec{x} | \omega_i)P(\omega_i)$$

$$\ln \varphi_{Bayes_i}(\vec{x}) = \ln p(\vec{x} | \omega_i) + \ln P(\omega_i)$$

Horst Bischof Computer Vision II SS11 21

### Parametric Classifier

Assumption that all

$$p(\vec{x} | \omega_i)$$

a distributed according to a parametric model

Learning: Estimating model parameters

Horst Bischof Computer Vision II SS11 22

### Multivariate Normaldistribution

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu})\right\}$$

$\mu$  Mean  
 $\Sigma$  Covariance  
 $|\Sigma|$  Determinant

$$\Sigma = E[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T]$$

$$\vec{\mu} = E[\vec{x}]$$

Horst Bischof Computer Vision II SS11 23

### Bivariate Normalverteilung

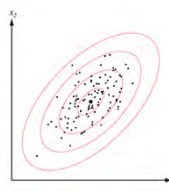
Horst Bischof Computer Vision II SS11 24

2D-Distribution

Equidistances are Ellipses (Mahalanobis Distance)

$$(\bar{x} - \bar{\mu})^T \Sigma^{-1} (\bar{x} - \bar{\mu}) = c^2$$

Center  $(\mu_1, \mu_2)$ :



Horst Bischof Computer Vision II SS11 25

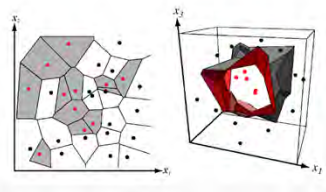
## Supervised Classification

We have labeled Training Data

Horst Bischof Computer Vision II SS11 26

### Nearest Neighbor Classification

Assign Pattern class of closest training pattern

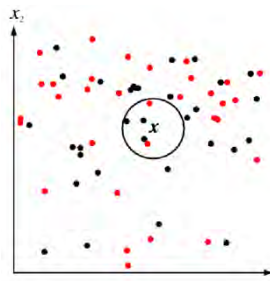


For  $n \rightarrow \infty$  at maximum twice Bayes error rate.

Horst Bischof Computer Vision II SS11 27

### K-Nearest Neighbor Classifier

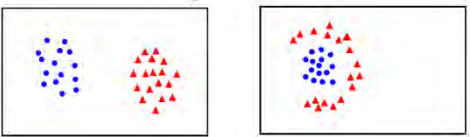
Assign Pattern x majority among k nearest neighbors



Horst Bischof Computer Vision II SS11 28

### Binary Classification

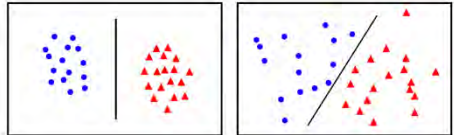
Given training data  $(x_i, y_i)$  for  $i = 1 \dots N$ , with  $x_i \in \mathbb{R}^d$  and  $y_i \in \{1, -1\}$ , learn a classifier  $f(x)$  such that

$$f(x_i) \begin{cases} \geq 0 & y_i = +1 \\ < 0 & y_i = -1 \end{cases}$$


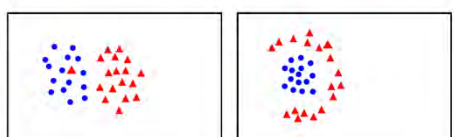
Horst Bischof Computer Vision II SS11 29

### Linear Separable

linearly separable



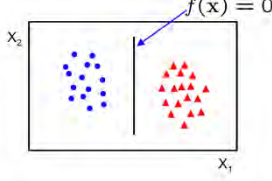
not linearly separable



Horst Bischof Computer Vision II SS11 30

### Linear Classifier

A linear classifier has the form

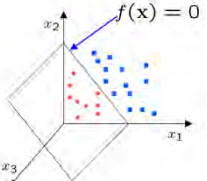
$$f(x) = w^T x + b$$


- in 2D the discriminant is a line
- $w$  is the normal to the plane, and  $b$  the bias
- $w$  is known as the weight vector

Horst Bischof Computer Vision II SS11 31

### Linear Classifier

A linear classifier has the form

$$f(x) = w^T x + b$$


- in 3D the discriminant is a plane, and in nD it is a hyperplane

For a K-NN classifier it was necessary to 'carry' the training data  
 For a linear classifier, the training data is used to learn  $w$  and then discarded  
 Only  $w$  is needed for classifying new data

Horst Bischof Computer Vision II SS11 32

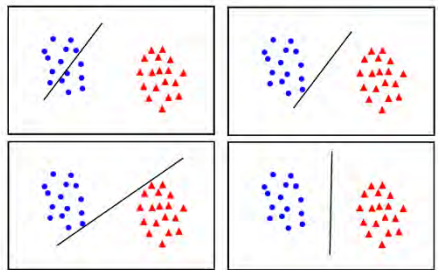
### How to find parameters

Various Algorithms to find parameters  $w$ :

- Perceptron and Variants
- LMS
- Linear Discriminant Analysis (LDA)
- Support Vector Machines SVMs

Horst Bischof Computer Vision II SS11 33

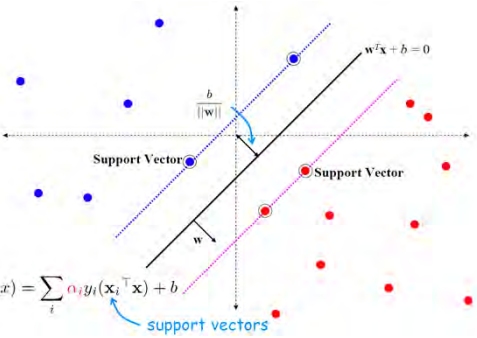
### Best $w$ ?



- maximum margin solution: most stable under perturbations of the inputs

Horst Bischof Computer Vision II SS11 34

### Support Vector Machines



$$f(x) = \sum_i \alpha_i y_i (x_i^T x) + b$$


support vectors

Horst Bischof Computer Vision II SS11 35

### Example Object Detection

Objective: detect (localize) standing humans in an image

- of face detection with a sliding window classifier



- reduces object detection to binary classification
- does an image window contain the object or not?

Horst Bischof Computer Vision II SS11 36

**Detection Phase**

- Scan image(s) at all scales and locations
- Extract features over windows
- Run window classifier at all locations
- Fuse multiple detections in 3-D position & scale space

Object detections with bounding boxes

Scale-space pyramid  
Detection window

Horst Bischof Computer Vision II SS11 37

Haar wavelet descriptors

Training set (2k positive / 10k negative)

training

1326-D descriptor

Support vector machine

results

Test image

Multi-scale search

[Papageorgiou & Poggio, 1998]

Horst Bischof Computer Vision II SS11 38

**Feature HOG: Histogram of Oriented Gradients**

image

dominant direction

HOG

- tile window into 8 x 8 pixel cells
- each cell represented by HOG

frequency

orientation

Feature vector dimension = 16 x 8 (for tiling) x 8 (orientations) = 1024

Horst Bischof Computer Vision II SS11 39

**Face Detection (Buffy)**

Horst Bischof Computer Vision II SS11 40

**Unsupervised Classification**

**Clustering**

We do not have labeled Training Data

Horst Bischof Computer Vision II SS11 41

**Clustering**

**N Patterns in K-Clusters**


Horst Bischof Computer Vision II SS11 42

**Clustering Tasks**

**Familiarity** Learn what is typical  
**Clustering** Which class a pattern belongs to  
**Prototyping** Find prototypes  
**Encoding** Data compression  
**Feature mapping** Topographic map of the input  
**Principal Component Analysis** Eigenvectors of Correlation Matrix

Horst Bischof Computer Vision II SS11 43

**What is Similarity?**



Similarity is hard to define, but...  
*"We know it when we see it"*

The real meaning of similarity is a philosophical question. We will take a more pragmatic approach.

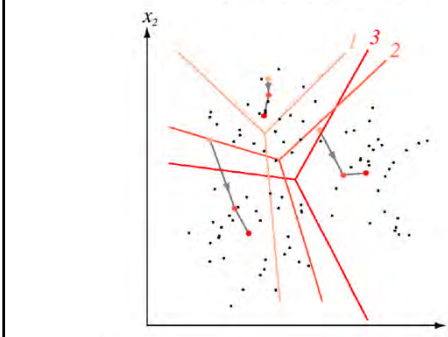
Horst Bischof Computer Vision II SS11 44

**K-Means Algorithm**

Select initial Cluster centers  $c_i$  from data points  $p_j$   
 Assign all data points to nearest cluster center  
 Update cluster center by calculating new mean value  
 Repeat until there are no more changes

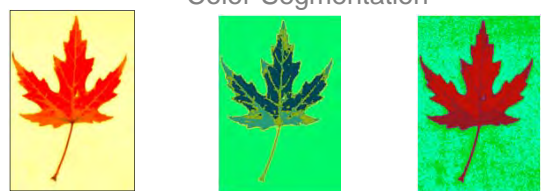
Horst Bischof Computer Vision II SS11 45

**K-Means Algorithm**



Horst Bischof Computer Vision II SS11 46

**Color-Segmentation**

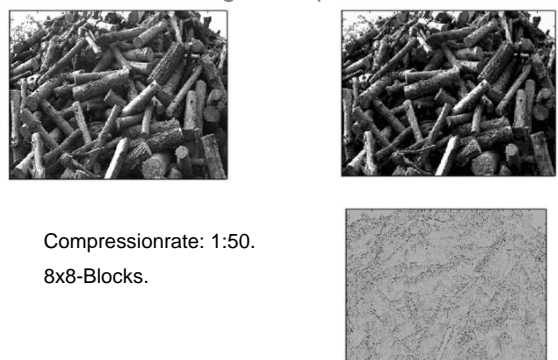


RGB Trippels are Datapoints.  
 Clustering using:

- k-means
- fuzzy k-means

Horst Bischof Computer Vision II SS11 47

**Image Compression**



Compressionrate: 1:50.  
 8x8-Blocks.

Horst Bischof Computer Vision II SS11 48

**Segmentation as clustering**

**Cluster together** (pixels, tokens, etc.) that belong together

**Agglomerative clustering**

- attach closest to cluster it is closest to
- repeat

**Divisive clustering**

- split cluster along best boundary
- repeat

**Point-Cluster distance**


- single-link clustering
- complete-link clustering
- group-average clustering

**Dendrograms**


- yield a picture of output as clustering process continues

Horst Bischof Computer Vision II SS11 49


Image



Clusters on intensity

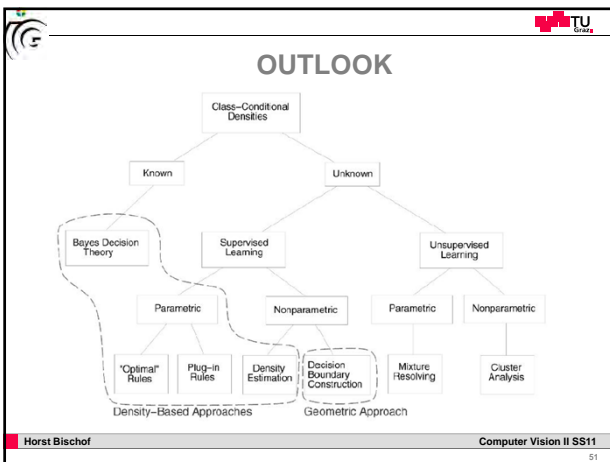


Clusters on color



K-means clustering using intensity alone and color alone

Horst Bischof Computer Vision II SS11 50

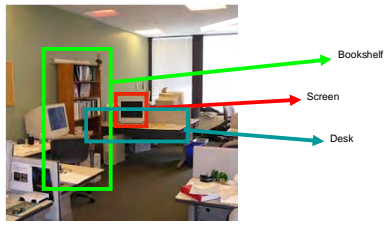


# Object Recognition

Horst Bischof Computer Vision II SS11 52

**A computer vision goal**

Recognize many different objects under many viewing conditions in unconstrained settings.



Horst Bischof Computer Vision II SS11 53

## Object Recognition

(D. Lowe 2003)

**Definition:** Identify an object and determine its pose and model parameters

**Commercial object recognition**

- Currently a \$4 billion/year industry for inspection and assembly
- Almost entirely based on template matching


**Upcoming applications**

- Mobile robots, toys, user interfaces
- Location recognition
- Digital camera panoramas, 3D scene modeling



Horst Bischof Computer Vision II SS11 54

**Where is the field of computer vision?**


There are efficient solutions for detecting a single object category and view:  
Viola & Jones 2001; Papageorgiou & Poggio 2000, ...



detecting particular objects:  
Lowe, 1999

Recognizing objects in isolation  
Leibe & Schiele, 2003;



Horst Bischof Computer Vision II SS11 55

**Computational photography**


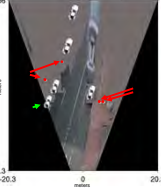




[Face priority AE] When a bright part of the face is too bright



Horst Bischof Computer Vision II SS11 56

**Assisted driving**

Pedestrian and car detection

Lane detection

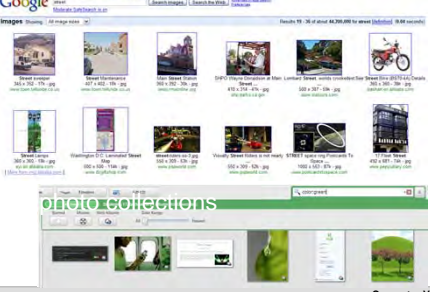
- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems.

Horst Bischof Computer Vision II SS11 57

**Improving online search**

flickr webshots Ask Images picsearch Google altavista Cydral

Query: STREET



Horst Bischof Computer Vision II SS11 58

**Recognition problems**

- What is it?
  - Object and scene recognition
- Who is it?
  - Identity recognition
- Where is it?
  - Object detection
- What are they doing?
  - Activities
- All of these are **classification** problems
  - Choose one class from a list of possible candidates

Horst Bischof Computer Vision II SS11 59

**What is recognition?**

- A different taxonomy from [Csurka *et al.* 2006]:
- Recognition
  - Where is *this* particular object?
- Categorization
  - What *kind* of object(s) is(are) present?
- Content-based image retrieval
  - Find me something that looks similar
- Detection
  - Locate *all* instances of a given class

Horst Bischof Computer Vision II SS11 60

**Issues in Object Recognition**

We want a machine to be able to identify thousands of different objects in their usual environment.

**Representation**

- Local vs. Global Appearance based
- Discriminative vs. Generative

**Recognition**

- Robustness

**Learning**

- Robustness
- Minimal samples

Horst Bischof Computer Vision II SS11

**Object Recognition**

**Hierarchy of problems**

- classify image according to objects present
- localize objects (position, scale, pose, ...)
- segment objects

**Research frontier: 100s or 1000s of categories**



Horst Bischof Computer Vision II SS11

**Issues**

**Variability due to changes in**

- location
- scale
- pose
- lighting
- non-rigid deformation
- intra-class variability

**High dimensionality of input space**

**Occlusion**

**Large number of object classes**

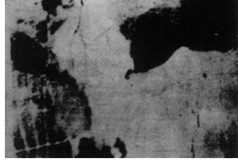
**Multiple occurrences of objects**

**Some categories semantic rather than visual e.g. chairs**


Horst Bischof Computer Vision II SS11

**Problems**

**Segmentation:**

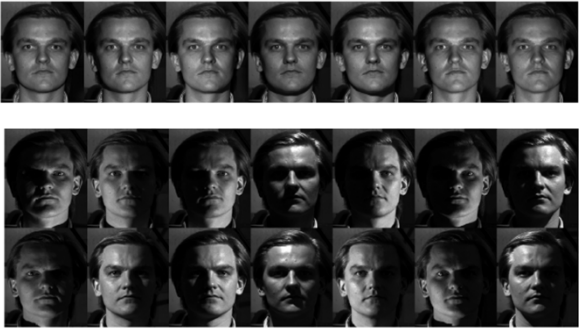


**Pose/Shape:**



Horst Bischof Computer Vision II SS11

**Illumination**



Horst Bischof Computer Vision II SS11

**What is a chair**





Image from H. Bülthoff MPI Tübingen

Horst Bischof Computer Vision II SS11

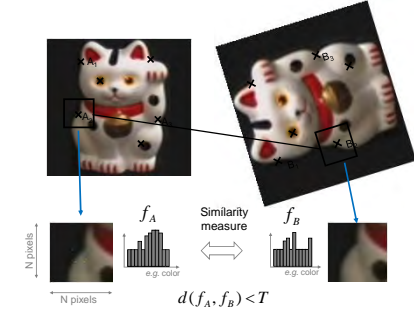
Invariance?



Horst Bischof Computer Vision II SS11

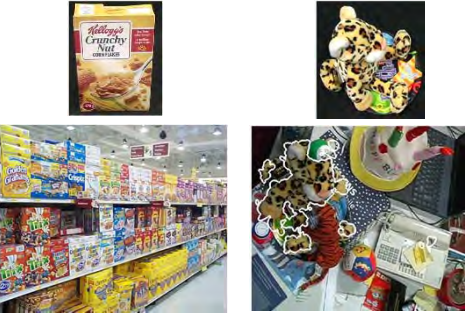
### Local Object Recognition

1. Find a set of distinctive keypoints
2. Define a region around each keypoint
3. Extract and normalize the region content
4. Compute a local descriptor from the normalized region
5. Match local descriptors



Horst Bischof Computer Vision II SS11

### Single object recognition



Horst Bischof Computer Vision II SS11


### Single object recognition

- Lowe, et al. 1999, 2003
- Mahamud and Herbert, 2000
- Ferrari, Tuytelaars, and Van Gool, 2004
- Rothganger, Lazebnik, and Ponce, 2004
- Moreels and Perona, 2005
- ...

Horst Bischof Computer Vision II SS11

### Planar object recognition [Lowe]


- Use SIFT features
- Verify affine (or homography) geometric alignment



Horst Bischof Computer Vision II SS11

### Planar object recognition [Lowe]


- Use SIFT features
- Verify affine (or homography) geometric alignment



Horst Bischof Computer Vision II SS11

3D object recognition [Lowe]

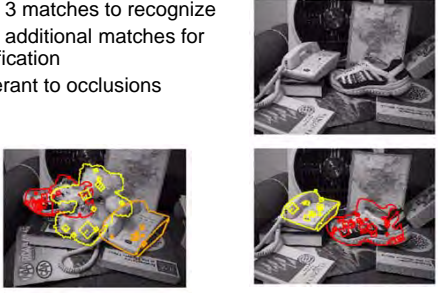
- Extract object outlines with background subtraction



Horst Bischof Computer Vision II SS11 73

3D object recognition [Lowe]

- Use 3 matches to recognize
- Use additional matches for verification
- Tolerant to occlusions



Horst Bischof Computer Vision II SS11 74

Feature-based recognition

- How can we scale to millions of objects?
- Comparison to *all* stored objects/features is infeasible.
- **Answer:**
- quantize features into *words* [Csurka *et al.* 04]
- use information retrieval (inverted index)
- use *metric tree* for faster quantization (NN) [Nister & Stewenius 05]

Horst Bischof Computer Vision II SS11 75