A Walk through "Bildverarbeitung WS 2007/08"

1

Definition of Image Understanding

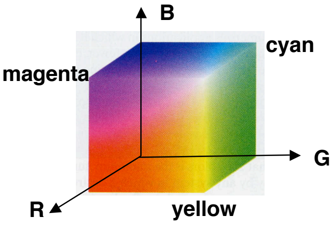
Image understanding is the task-oriented reconstruction and interpretation of a scene by means of images

scene:	section of the real world stationary (3D) or moving (4D)
image:	view of a scene projection, density image (2D) depth image (2 1/2D) image sequence (3D)
reconstruction and interpretation:	computer-internal scene description quantitative + qualitative + symbolic
task-oriented:	for a purpose, to fulfill a particular task context-dependent, supporting actions of an agent

2

Technical Colour Models

RGB colour model



Typical discretization:

8 bits per colour dimension
=> 16,77,216 colours

CMY colour model

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

HSI colour model

Hue:

$$H = \begin{cases} \Theta & \text{if } B \leq G \\ 360 - \Theta & \text{if } B > G \end{cases}$$

$$\Theta = \cos^{-1} \frac{1/2\,[(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}}$$

Saturation:

$$S = 1 - \frac{3}{(R + G + B)} [\min (R, G, B)]$$

Intensity:

$$I = 1/3\,(R + G + B)$$

3

Sampling Theorem

Shannon´s Sampling Theorem:

A bandlimited function with bandwidth W can be exactly reconstructed from equally spaced samples, if the sampling distance is not larger than $\frac{1}{2W}$

bandwidth = largest frequency contained in signal

(=> Fourier decomposition of a signal)

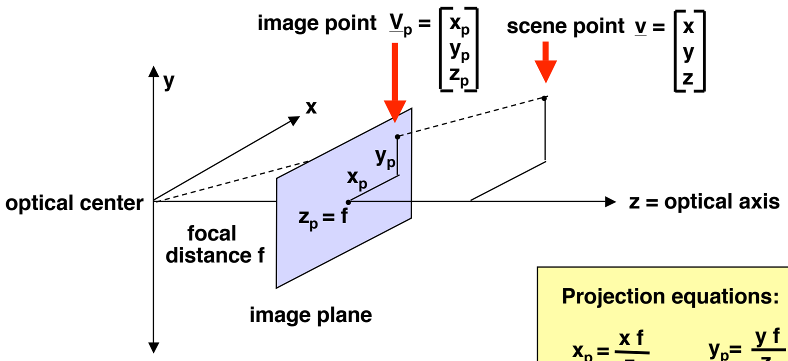
Analogous theorem holds for 2D signals with limited spatial frequencies W_x and W_y

4

Perspective Projection Geometry

Projective geometry relates the coordinates of a point in a scene to the coordinates of its projection onto an image plane.

<u>Perspective projection</u> is an adequate model for most cameras.



Projection equations:

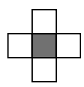
$$x_p = \frac{x\,f}{z} \qquad y_p = \frac{y\,f}{z}$$

5

Connectivity in Digital Images

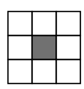
Connectivity is an important property of subsets of pixels. It is based on <u>adjacency</u> (or neighbourhood):

Pixels are 4-neighbours if their distance is $D_4 = 1$



all 4-neighbours of center pixel

Pixels are 8-neighbours if their distance is $D_8 = 1$



all 8-neighbours of center pixel

A <u>path</u> from pixel P to pixel Q is a sequence of pixels beginning at Q and ending at P, where consecutive pixels are neighbours.

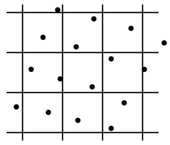
In a set of pixels, two pixels P and Q are <u>connected</u>, if there is a path between P and Q with pixels belonging to the set.

A <u>region</u> is a set of pixels where each pair of pixels is connected.

6

Principle of Greyvalue Interpolation

Greyvalue interpolation = computation of unknown greyvalues at locations (u´v´) from known greyvalues at locations (x´y´)



Two ways of viewing interpolation in the context of geometric transformations:

A Greyvalues at grid locations (x y) in old image are placed at corresponding locations (x´y´) in new image: g(x´y´) = g(T(x y))

=> interpolation in new image

B Grid locations (u´v´) in new image are transformed into corresponding locations (u v) in old image: g(u v) = g(T^{-1}(u´v´))

=> interpolation in old image

We will take view B:

Compute greyvalues between grid from greyvalues at grid locations.

Global Image Properties

Global image properties refer to an image as a whole rather than components. Computation of global image properties is often required for image enhancement, preceding image analysis.

We treat

- empirical mean and variance
- histograms
- projections
- cross-sections
- frequency spectrum

Median Filter

Median of a distribution P(x): x_m such that $P(x < x_m) = 1/2$

<u>Median Filter</u>:

$$\hat{g}_{ij} = \max a \text{ with } g_k \in D \text{ and } |\{g_k < a\}| < \frac{|D|}{2}$$

1. Sort pixels in D according to greyvalue
2. Choose greyvalue in middle position

<u>Example</u>:

11	14	15
13	12	25
15	19	26

11
12
13
14
15
15
19
25
26

greyvalue of center pixel of region is set to 15

Median Filter reduces influence of outliers in either direction!

9

Discrete Fourier Transform (DFT)

Computes image representation as a sum of sinusoidals.

Discrete Fourier Transform:

$$G_{uv} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} g_{mn} e^{-2\pi i \left(\frac{mu}{M} + \frac{nv}{N}\right)}$$

for $u = 0 \ldots M-1$, $v = 0 \ldots N-1$

Inverse Discrete Fourier Transform:

$$g_{mn} = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} G_{uv} e^{2\pi i \left(\frac{mu}{M} + \frac{nv}{N}\right)}$$

for $m = 0 \ldots M-1$, $n = 0 \ldots N-1$

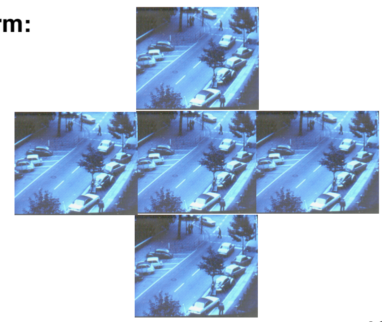
Notation for computing the Fourier Transform:

$G_{uv} = F\{ g_{mn} \}$

$g_{mn} = F^{-1}\{ G_{uv} \}$

Transform is based on periodicity assumption

=> periodic continuation may cause boundary effects



10

Filtering in the Frequency Domain

A filter transforms a signal by modifying its spectrum.

$G(u, v) = F(u, v)\,H(u, v)$

F Fourier transform of the signal
H frequency transfer function of the filter
G modified Fourier transform of signal

- low-pass filter — *low frequencies pass, high frequencies are attenuated or removed*
- high-pass filter — *high frequencies pass, low frequencies are attenuated or removed*
- band-pass filter — *frequencies within a frequency band pass, other frequencies below or above are attenuated or removed*

Often (but not always) the noise part of an image is high-frequency and the signal part is low-frequency. Low-pass filtering then improves the signal-to-noise ratio.

Convolution Using the FFT

Convolution in the spatial domain may be performed more efficiently using the FFT.

$$g'_{ij} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} g_{mn} h_{i-m,j-n} \qquad (MN)^2 \text{ operations needed}$$

Using the FFT and filtering in the frequency domain:

$$g_{mn} \xrightarrow{FFT} G_{uv} \xrightarrow{H_{uv}} G'_{uv} \xrightarrow{FFT^{-1}} g'_{mn}$$

MN log(MN)   MN   MN log(MN)   # of operations

Example with M = N = 512:

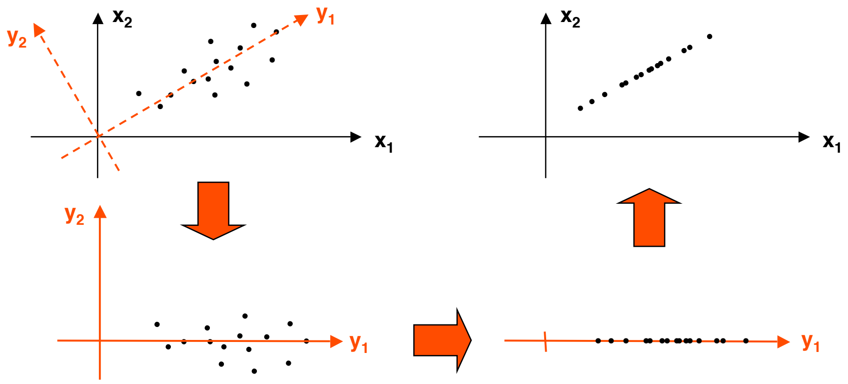
- straight convolution needs ~ 10^{10} operations
- convolution using the FFT needs ~10^{7} operations

Illustration of Minimum-loss Dimension Reduction

Using the Karhunen-Loève transform data compression is achieved by

- changing (rotating) the coordinate system
- omitting the least informative dimension(s) in the new coodinate system

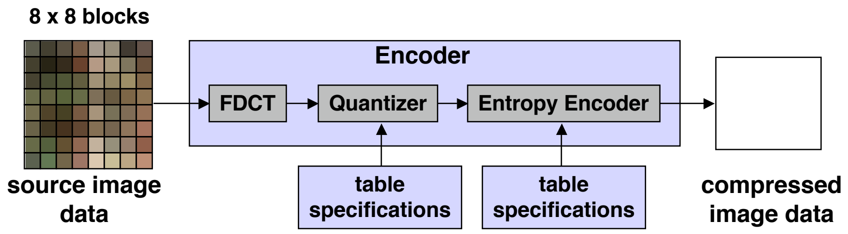
Example:



13

Principle of Baseline JPEG

(Source: Gibson et al., Digital Compression for Multimedia, Morgan Kaufmann 98)

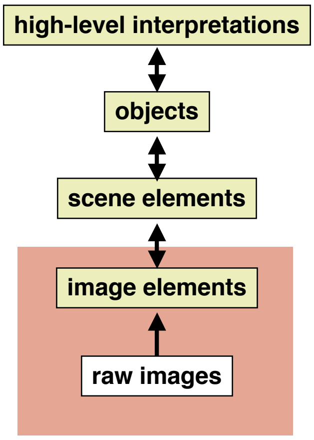


- transform RGB into YUV coding, subsample color information
- partition image into 8 x 8 blocks, left-to-right, top-to-bottom
- compute Discrete Cosine Transform (DCT) of each block
- quantize coefficients according to psychovisual quantization tables
- order DCT coefficients in zigzag order
- perform runlength coding of bitstream of all coefficients of a block
- perform Huffman coding for symbols formed by bit patterns of a block

14

Segmentation

Segmenting the image into image elements which may correspond to meaningful scene elements



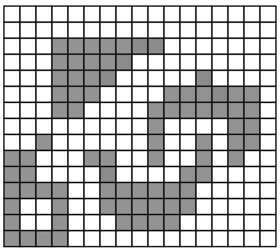
Example: Partitioning an image into regions which may correspond to objects

Typical results of first segmentation steps

15

Representing Regions

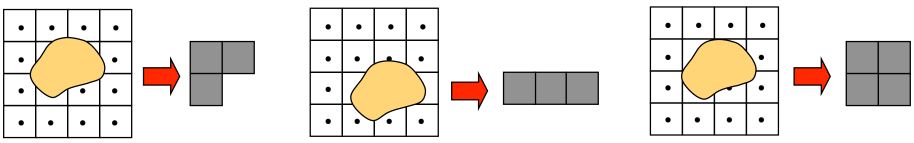
A region is a maximal 4- (or 8-) connected set of pixels.



Methods for digital region representation:

- grid occupancy
  - labelling
  - run-length coding
  - quadtree coding
  - cell sets
- boundary description
  - chain code
  - straight-line segments, polygons
  - higher-order polynomials

Note that discretizations of an analog region are not shift or rotation invariant:

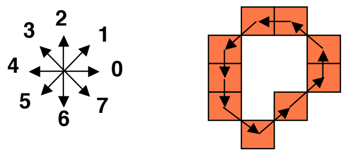


16

Chain Code

Chain code represents boundaries by "chaining" direction arrows between successive boundary elements.

Chain code for 8-connectivity:



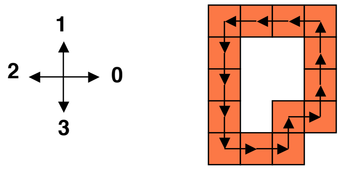
Arbitrary choice of starting point, chain code can be represented e.g. by

{456671123}

Normalization by circular shift until the smallest integer is obtained:

{112345667}

Chain code for 4-connectivity:



Arbitrary starting point:

{22233330010111}

Normalized:

{00101112223333}

Canny Edge Detector (1)

Optimal edge detector for step edges corrupted by white noise.

Optimality criteria:

- **Detection of all important edges and no spurious responses**
- **Minimal distance between location of edge and actual edge**
- **One response per edge only**

1. **Derivation for 1D results in edge detection filter which can be effectively approximated (< 20% error) by the 1rst derivative of a Gaussian smoothing filter.**

2. **Generalization to 2D requires estimation of edge orientation:**

$$n = \frac{\nabla(f \bullet g)}{|\nabla(f \bullet g)|}$$

n normal perpendicular to edge
f Gaussian smoothing filter
g greyvalue image

Edge is located at local maximum of g convolved with f in direction n:

$$\frac{\partial^2}{\partial n^2} f \bullet g = 0$$

"non-maximal suppression"

Canny Edge Detector (2)

Algorithm includes

- choice of scale σ
- hysteresis thresholding to avoid streaking (breaking up edges)
- "feature synthesis" by selecting large-scale edges dependent on lower-scale support

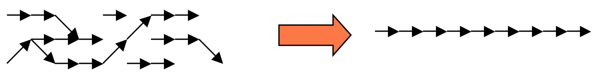
1. Convolve image g with Gaussian filter f of scale σ
2. Estimate local edge normal direction n for each point in the image
3. Find edge locations using non-maximal suppression
4. Compute magnitude of edges by $|\nabla(f \bullet g)|$
5. Threshold edges with hysteresis to eliminate spurious edges
6. Repeat steps (1) through (5) for increasing values of σ
7. Aggregate edges at multiple scales using feature synthesis

19

Grouping

To make sense of image elements, they first have to be grouped into larger structures.

Example: Grouping noisy edge elements into a straight edge

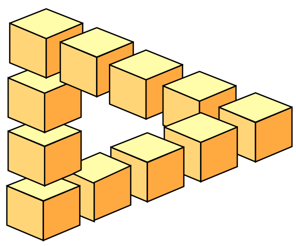


Essential problem:

Obtaining globally valid results by local decisions

Important methods:

- Fitting
- Clustering
- Hough Transform
- Relaxation

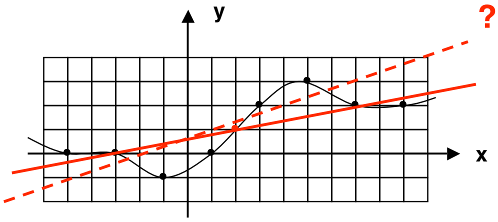


- locally compatible
- globally incompatible

20

Example for Straight Line Fitting by Eigenvector Analysis

What is the best straight-line approximation of the contour?



Given points: { (-5 0) (-3 0) (-1 -1) (1 0) (3 2) (5 3) (7 2) (9 2) }

Center of gravity: $m_x = 2$ $m_y = 1$

Scatter matrix: $S_{11} = 168$ $S_{12} = S_{21} = 38$ $S_{22} = 14$

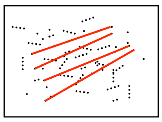
Eigenvalues: $\lambda_1 = 176{,}87$ $\lambda_2 = 5{,}13$

Direction of straight line: $r_y/r_x = 0{,}23$

Straight line equation: $y = 0{,}23\ x + 0{,}54$

Hough Transform (1)

Robust method for fitting straight lines, circles or other geometric figures which can be described analytically.

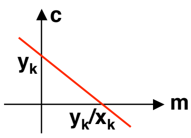


Given: Edge points in an image

Wanted: Straight lines supported by the edge points

An edge point (x_k, y_k) supports all straight lines $y = mx + c$ with parameters m and c such that $y_k = mx_k + c$.

The locus of the parameter combinations for straight lines through (x_k, y_k) is a straight line in parameter space.



Principle of Hough transform for straight line fitting:

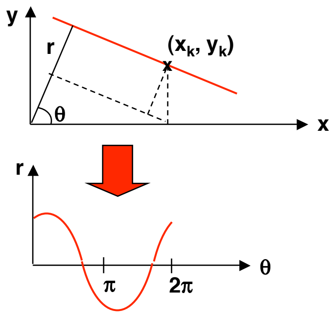
- Provide accumulator array for quantized straight line parameter combinations
- For each edge point, increase accumulator cells for all parameter combinations supported by the edge point
- Maxima in accumulator array correspond to straight lines in the image

Hough Transform (2)

For straight line finding, the parameter pair (r, θ) is commonly used because it avoids infinite parameter values:

$$x_k\cos\theta + y_k\sin\theta = r$$

Each edge point (x_k, y_k) corresponds to a sinusoidal in parameter space:



Important improvement by exploiting direction information at edge points:

(x_k, y_k, φ) ➡ $x_k\cos\theta + y_k\sin\theta = r$ restricted to $\varphi-\delta \le \theta \le \varphi+\delta$

gradient direction

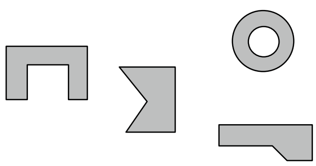
direction tolerance

23

Simple 2D Shape Features

For industrial recognition tasks it is often required to distinguish
- a small number of different shapes
- viewed from a small number of different view points
- with a small computational effort.

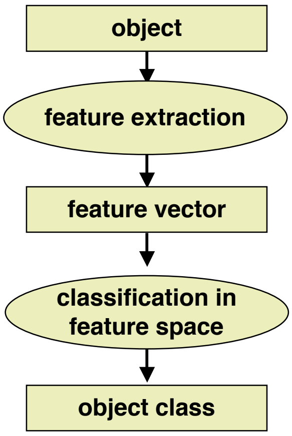
In such cases simple 2D shape features may be useful, such as:
- area
- boxing rectangle
- boundary length
- compactness
- second-order momentums
- polar signature
- templates



Features may or may not have <u>invariance properties</u>:
- 2D translation invariance
- 2D rotation invariance
- scale invariance

24

Basic Terminology for Pattern Recognition



K classes $\omega_1 \dots \omega_K$

N dimension of feature space

$\underline{x}^T = [x_1\ x_2 \dots\ x_N]$ feature vector

$\underline{y}^T = [y_1\ y_2 \dots\ y_N]$ prototype
(feature vector with known class membership)

$\underline{y}_i^{(k)}$ i-th prototyp of class k

M_k number of prototypes for class k

$g_k(\underline{x})$ discriminant function for class k

Problem:

Determine $g_k(\underline{x})$ such that

$g_k(\underline{x}) > g_j(\underline{x}), \quad \forall\, \underline{x} \in \omega_K \quad \forall\, k \neq j$

25

Minimum Squared Error

New criterion function for all samples:

Find $\underline{a}$ such that $\underline{a}^T\underline{y}_i = b_i$ with b_i = some positive constant

In matrix notation: $Y\underline{a} = \underline{b}$ with $Y = \begin{bmatrix} \underline{y}_1^T \\ \underline{y}_2^T \\ \dots \\ \underline{y}_M^T \end{bmatrix}$ and $\underline{y}_i^T = [\, y_{i1} \dots y_{iN} \,]$

In general, $M >> N$ and Y^{-1} does not exist, hence $\underline{a} = Y^{-1}\underline{b}$ is no solution.

Classical solution technique: Minimize squared error criterion:

$$J_s(\underline{a}) = \|Y\underline{a} - \underline{b}\|^2 = \Sigma\,(\underline{a}^T\underline{y}_i - b_i)^2$$

Closed-form solution by setting the gradient equal to 0.

$\nabla J_s = 2Y^T(Y\underline{a} - \underline{b}) = \underline{0} \quad \Rightarrow \quad \boxed{\underline{a} = (Y^TY)^{-1}Y^T\underline{b}}$ if $\underbrace{(Y^TY)^{-1}Y^T}_{\text{pseudoinverse of } Y}$ is nonsingular

26

Statistical Decision Theory

Generating decision functions from a statistical characterization of classes (as opposed to a characterization by prototypes)

Advantages:

1. The classification scheme may be designed to satisfy an objective optimality criterion:

   Optimal decisions minimize the probability of error.

2. Statistical descriptions may be much more compact than a collection of prototypes.

3. Some phenomena may only be adequately described using statistics, e.g. noise.

27

General Framework for Bayes Classification

Statistical decision theory which minimizes the probability of error for classifications based on uncertain evidence

$\omega_1 \ldots \omega_K$	K classes
$P(\omega_k)$	prior probability that an object of class k will be observed
$\underline{x} = [x_1 \ldots x_N]$	N-dimensional feature vector of an object
$p(\underline{x} \mid \omega_k)$	conditional probability ("likelihood") of observing $\underline{x}$ given that the object belongs to class ω_K
$P(\omega_k \mid \underline{x})$	conditional probability ("posterior probability") that an object belongs to class ω_K given $\underline{x}$ is observed

Bayes decision rule:

Classify given evidence $\underline{x}$ as class ω' such that ω' minimizes the probability of error $P(\omega \neq \omega' \mid \underline{x})$

=> Choose ω' which maximizes the posterior probability $P(\omega \mid \underline{x})$

$g_i(\underline{x}) = P(\omega_i \mid \underline{x})$ are discriminant functions.

28

Motion Analysis

Motion analysis of digital images is based on a temporal sequence of image frames of a coherent scene.

"sparse sequence"	**=>**	**few frames, temporally spaced apart, considerable differences between frames**
"dense sequence"	**=>**	**many frames, incremental time steps, incremental differences between frames**
video	**=>**	**50 half frames per sec, interleaving, line-by-line sampling**

Motion detection
Register locations in an image sequence which have change due to motion

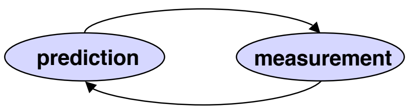
Moving object detection and tracking
Detect individual moving objects, determine and predict object trajectories, track objects with a moving camera

Derivation of 3D object properties
Determine 3D object shape from multiple views ("shape from motion")

29

Kalman Filters (1)

A Kalman filter provides an iterative scheme for (i) predicting an event and (ii) incorporating new measurements.



Assume a linear system with observations depending linearly on the system state, and white Gaussian noise disturbing the system evolution and the observations:

$$\underline{x}_{k+1} = A_k \underline{x}_k + \underline{w}_k$$
$$\underline{z}_k = H_k \underline{x}_k + \underline{v}_k$$

What is the best estimate of $\underline{x}_k$ based on the previous estimate $\underline{x}_{k-1}$ and the observation $\underline{z}_k$?

$\underline{x}_k$	**quantity of interest ("state") at time k**
A_k	**model for evolution of $\underline{x}_k$**
$\underline{w}_k$	**zero mean Gaussian noise with covariance Q_k**
$\underline{z}_k$	**observations at time k**
H_k	**relation of observations to state**
$\underline{v}_k$	**zero mean Gaussian noise with covariance R_k**

Often, A_k, Q_k, H_k and R_k are constant.

30

Optical Flow Constraint Equation

Optical flow is the displacement field of surface elements of a scene during an incremental time interval dt ("velocity field").

Assumptions:

- Observed brightness is constant over time (no illumination changes)
- Nearby image points move similarly (velocity smoothness constraint)

For a continuous image g(x, y, t) a linear Taylor series approximation gives

$$g(x+dx, y+dy, t+dt) \approx g(x, y, t) + g_x dx + g_y dy + g_t dt$$

For motion without illumination change we have

$$g(x+dx, y+dy, t+dt) = g(x, y, t)$$

Hence $g_x dx/dt + g_y dy/dt = g_x u + g_y v = -g_t$ u, v velocity components

$g_x u + g_y v = -g_t$ optical flow constraint equation

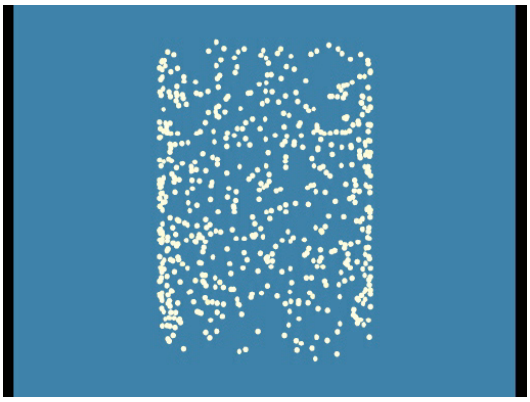
$g_x \approx \Delta g/\Delta x$, $g_y \approx \Delta g/\Delta y$, $g_t \approx \Delta g/\Delta t$ may be estimated from the spatial and temporal surround of a location (x, y), hence the optical flow constraint equation provides one equation for the two unknowns u and v.

31

3D Motion Analysis Based on 2D Point Displacements

2D displacements of points observed on an unknown moving rigid body may provide information about

- the 3D structure of the points
- the 3D motion parameters



Rotating cylinder experiment by S. Ullman (1981)

Cases of interest:

- stationary camera, moving object(s)
- moving camera, stationary object(s)
- moving camera, moving object(s)

camera motion parameters may be known

32

Essential Matrix

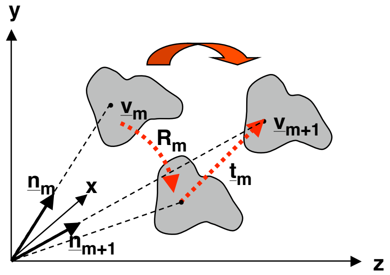
Geometrical constraints derived from 2 views of a point in motion

- motion between image m and m+1 may be decomposed into

  1) rotation R_m about origin of coordinate system (= optical center)

  2) translation $\underline{t}_m$

- observations are given by direction vectors $\underline{n}_m$ and $\underline{n}_{m+1}$ along projection rays



$R_m\underline{n}_m$, $\underline{t}_m$ and $\underline{n}_m$ are coplanar: $[\underline{t}_m \times R_m\underline{n}_m]^T \underline{n}_{m+1} = 0$

After some manipulation: $\underline{n}_m^T E_m \underline{n}_{m+1} = 0$ E = essential matrix

with $E_m = \begin{bmatrix} | & | & | \\ \underline{t}_m x \underline{r}_1 & \underline{t}_m x \underline{r}_2 & \underline{t}_m x \underline{r}_3 \\ | & | & | \end{bmatrix}$ and $R_m = \begin{bmatrix} | & | & | \\ \underline{r}_1 & \underline{r}_2 & \underline{r}_3 \\ | & | & | \end{bmatrix}$

33

Principle of Shape from Shading

See "Shape from Shading" (B.K.P. Horn, M.J. Brooks, eds.), MIT Press 1989

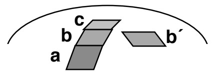
Physical surface properties, surface orientation, illumination and viewing direction determine the greyvalue of a surface patch in a sensor signal.

For a single object surface viewed in one image, greyvalue changes are mainly caused by surface orientation changes.

The reconstruction of arbitrary surface shapes is not possible because different surface orientations may give rise to identical greyvalues.

Surface shapes may be uniquely reconstructed from shading information if possible surface shapes are constrained by smoothness assumptions.

Principle of incremental procedure for surface shape reconstruction:



a: patch with known orientation

b, c: neighbouring patches with similar orientations

b´: radical different orientation may not be neighbour of a

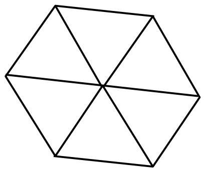
34

Generality Assumption

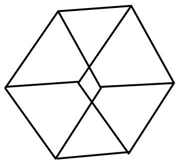
Assume that
- viewpoint
- illumination
- physical surface properties

are general, i.e. do not produce coincidental structures in the image.

Example: Do not interpret this figure as a 3D wireframe cube, because this view is not general.



General view:



The generality assumption is the basis for several specialized interpretation methods, e.g.
- shape from texture
- shape from shading

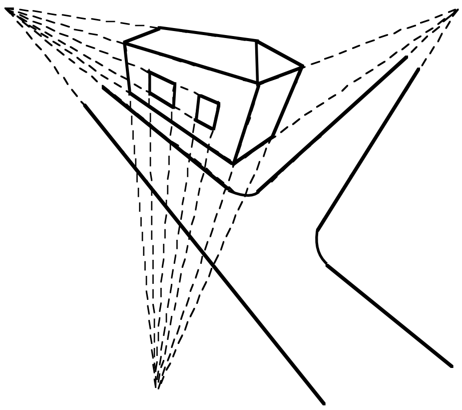
...
- "shape from X"

35

3D Line Orientation from Vanishing Points

From the laws of perspective projection:

The projections of 3D parallel straight lines intersect in a single point, the vanishing point.



Assume that more than 2 straight lines do not intersect in a single point by coincidence

⬇

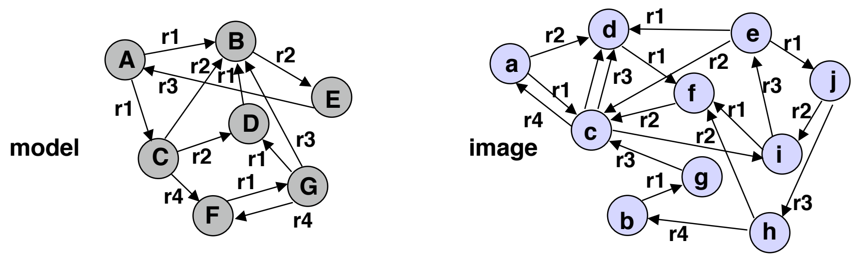
If more than 2 straight lines intersect, assume that they are parallel in 3D

36

Object Recognition by Relational Matching

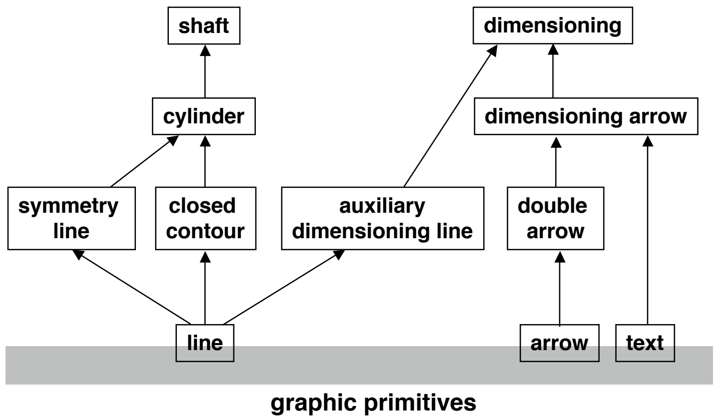
Principle:

- construct relational model(s) for object class(es)
- construct relational image description
- compute R-morphism (best partial match) between image and model(s)
- top-down verification with extended model



37

Partonomy of Object Parts



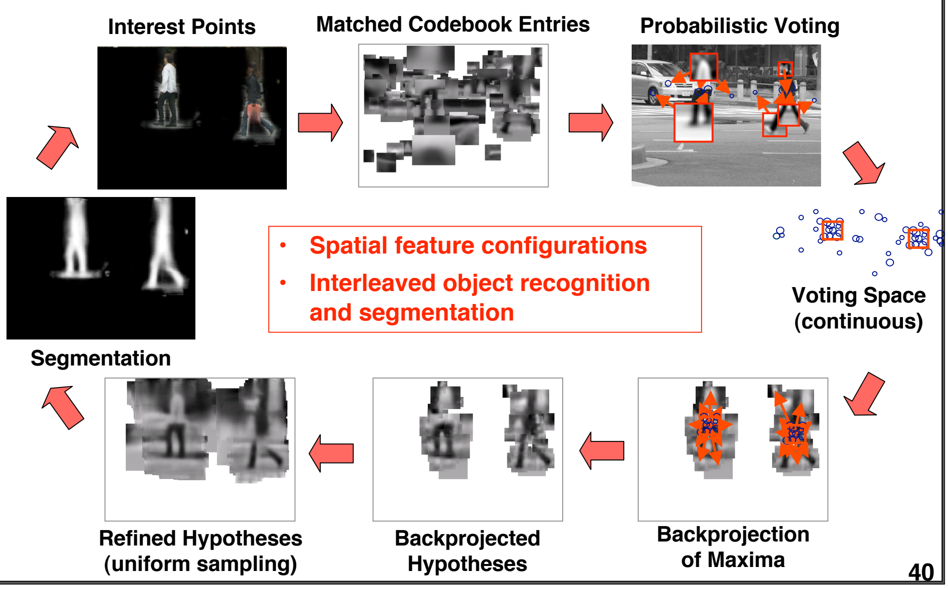
38

SIFT Features Summary

- SIFT features are reasonably invariant to rotation, scaling, and illumination changes.
- They can be used for matching and object recognition (among other things).
- Robust to occlusion: as long as we can see at least 3 features from the object we can compute the location and pose.
- Efficient on-line matching: recognition can be performed in close-to-real time (at least for small object databases).

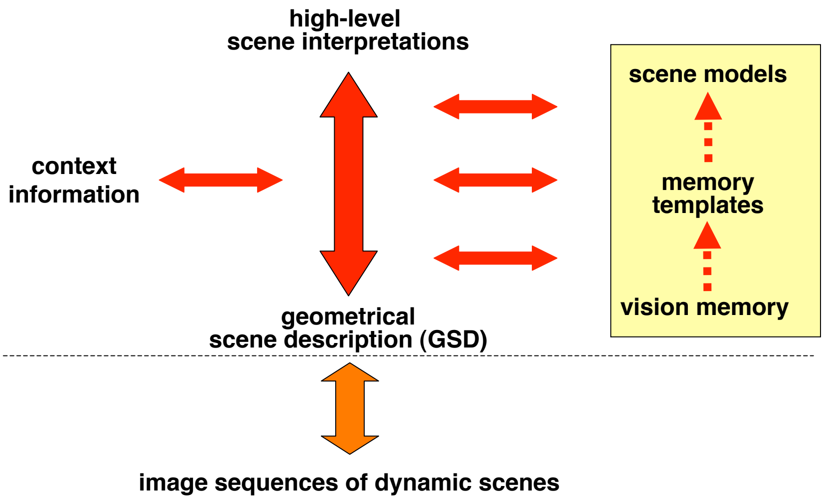
39

Implicit Shape Model - Recognition (2)



40

Basic Building Blocks for High-level Scene Interpretation



41

Occurrence Models

Basic ingredients:	•	relational structure
	•	taxonomy
	•	partonomy
	•	spatial relational language
	•	temporal relational language
	•	object appearance models

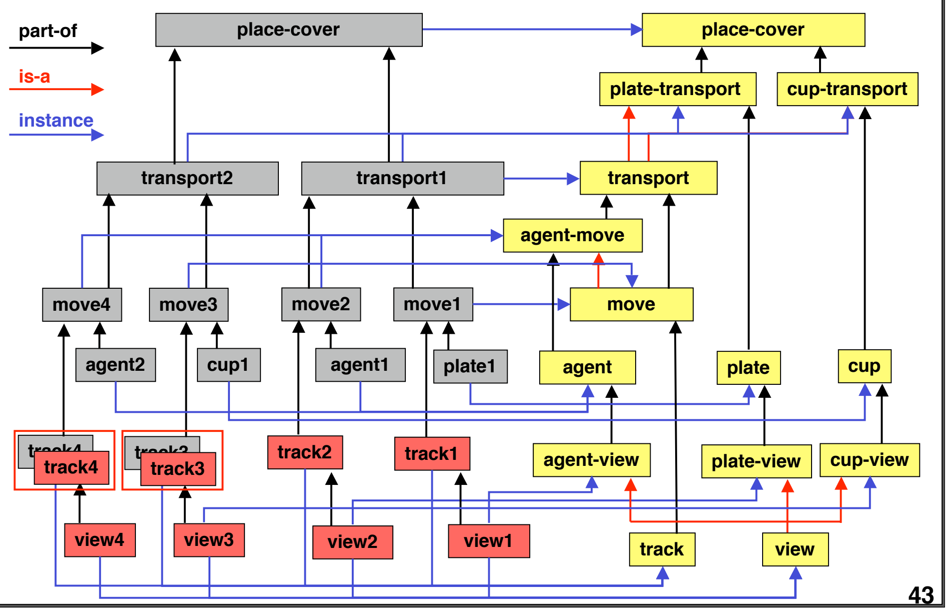
- An occurrence model describes a class of occurrences by
  - properties
  - sub-occurrences (= components of the occurrence)
  - relations between sub-occurrences

- A primitive occurrence model consists of
  - properties
  - a qualitative predicate

- Each occurrence has a begin and end time point

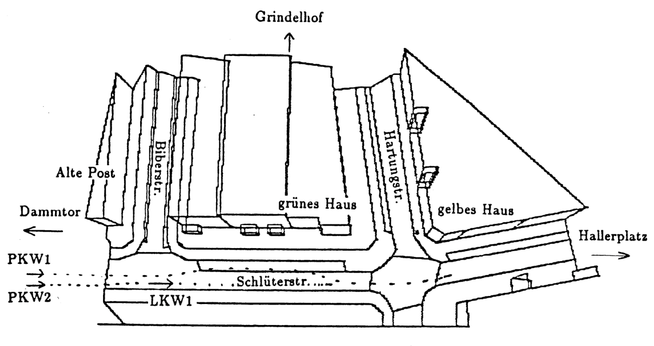
42

Model-based Interpretation



43

Example of an Automatically Generated Traffic Scene Description



DIE SZENE ENTHAELT DREI BEWEGTE OBJEKTE: ZWEI PKWS UND EINEN LKW.

EIN GELBER PKW FAEHRT IN RICHTUNG HALLERPLATZ. DABEI UEBERHOLT ER DEN LKW AUF DER SCHLUETERSTRASSE. DER GELBE PKW RAST VON DER ALTEN POST VOR DAS GELBE HAUS. ER ERREICHT DIE HARTUNGSTRASSE. ER HAELT AN. ER HAELT.

EIN SCHWARZER PKW ERREICHT DIE SCHLUETERSTRASSE. ER NAEHERT SICH DEM LKW VON DER ALTEN POST. DER SCHWARZE PKW FAEHRT IN RICHTUNG HALLERSTRASSE.

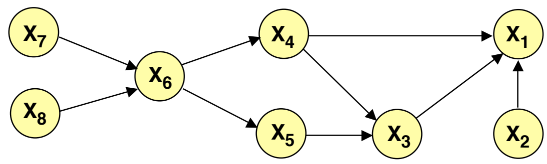
DER LKW FAEHRT VON DER ALTEN POST VOR DAS GRUENE HAUS. DABEI STOPPT ER VOR IHM. ER HAELT. ER FAEHRT IN RICHTUNG DAMMTOR WEITER. ER ENTFERNT SICH VON DEM GELBEN PKW. DER LKW HAELT AN. ER HAELT.

44

Causality Graph

Conditional dependencies (causality relations) of random variables define partial order.

Representation as a directed acyclic graph (DAG):



$P(X_1, X_2, X_3, \ldots, X_8) =$

$P(X_1 \mid X_2, X_3, X_4) \cdot P(X_2) \cdot P(X_3 \mid X_4, X_5) \cdot P(X_4 \mid X_6) \cdot P(X_5 \mid X_6) \cdot P(X_6 \mid X_7 X_8) \cdot P(X_7) \cdot P(X_8)$

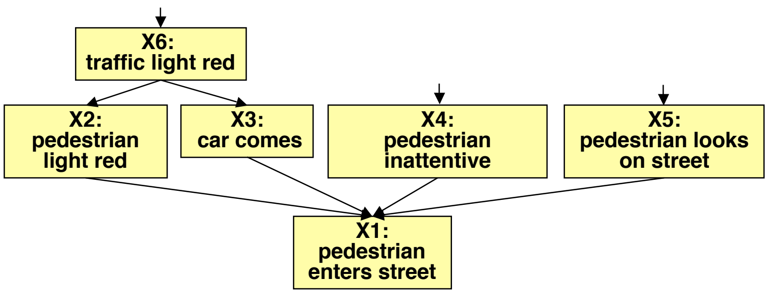
For any DAG, we obtain the JPD as follows:

$Pa(X_i)$ parents of node X_i

$$P(X_1 \ldots X_N) = \prod_i P(X_i \mid Pa(X_i))$$

45

Example: Traffic Behaviour of Pedestrians



Conditional probability table for each node must be known

P(X1 | X2, X3, X4, X5)

X1	X2	X3	X4	X5	P
T	T	T	T	T	0.3
F	T	T	T	T	0.7
T	F	T	T	T	0.9
F	F	T	T	T	0.1
⋮	⋮	⋮	⋮	⋮	⋮

P(X2 | X6)

X2	X6	P
T	T	0.2
F	T	0.8
T	F	1.0
F	F	0.0

P(X3 | X6)

X3	X6	P
T	T	0.01
F	T	0.99
T	F	0.6
F	F	0.4

P(X4)

X4	P
T	0.1
F	0.9

P(X5)

X5	P
T	0.7
F	0.3

P(X6)

X6	P
T	0.7
F	0.3

46

Computing Inferences

We want to use a Bayes Net for probabilistic inferences of the following kind:

Given a joint probability $P(X_1, \ldots, X_N)$ represented by a Bayes Net, and evidence $X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}$ for some of the variables, what is the probability $P(X_n = a_i \mid X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K})$ of an unobserved variable to take on a value a_i ?

In general this requires

- expressing a conditional probability by a quotient of joint probabilities

$$P(X_n = a_i \mid X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}) = \frac{P(X_n = a_i, X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K})}{P(X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K})}$$

- determining partial joint probabilities from the given total joint probability by summing out unwanted variables

$$P(X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}) = \sum_{X_{n_1}, \ldots, X_{n_K}} P(X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}, X_{n_1}, \ldots, X_{n_K})$$

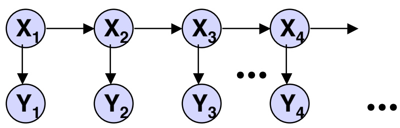
47

What Kind of Bayes Net is a HMM?

Bayes Net structure:

states

observations

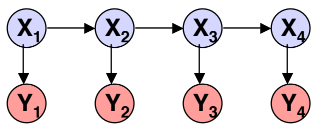


Finding most probable paths:

$P(\underline{X} = \underline{a} \mid \underline{Y} = \underline{b}) = ?$

hidden states

given observations

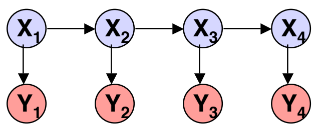


Evaluating likelihood of model:

$P(\underline{Y} = \underline{b} \mid \text{model}) = ?$

hidden states

given observations



48

How come, you see
what you see?

49