University of Osnabrück, Institute of Cognitive Science

Computer Vision

1 – Introduction

WS 2017 / 2018

Gunther Heidemann

Survey

1. Organization of the course
2. Survey of computer vision:
   - What are
     - image processing?
     - computer vision?
   - Why are we doing computer vision?
   - Basic processing strategies
   - Challenges
   - Typical applications

Organization of the course

4h lecture + 2h practice

Time and location:

Practice:	Tuesday	14.00h – 16.00h	(c.t.)	32/110
Lecture:	Wednesday	10.00h – 12.00h	(c.t.)	93/E31
Lecture:	Thursday	10.00h – 12.00h	(c.t.)	93/E31

There may be exceptions, lecture and practice may be switched.

See Stud.IP for the schedule!

Organization of the course

- Slides and practice materials will be available at Studip.
- Practice:
  - Work in groups of 3 people
  - Explain your solutions to the tutors (feedback meeting)
- Requirements to participate in the final exam:
  - More than 50% of the points in each of n-2 of n assignments
  - Details in the first practice session
- Written exam: Thursday February 8th

Tentative schedule

1. Introduction, motivation, examples
2. Image acquisition
3. Basic operations
4. Morphological operations
5. Color
6. Segmentation
7. Hough transform
8. Fourier transform
9. Sampling theorem and image enhancement
10. Machine learning techniques
11. Template matching
12. Pattern recognition
13. Local Features
14. Cosine and wavelet transform
15. Compression
16. Motion
17. Image retrieval

Literature and image sources

Most of the presented images are from the accompanying materials of the following books (sorted by relevance), which are also recommended for reading:

1. Klaus D. Tönnies, *Grundlagen der Bildverarbeitung*, Pearson Studium, 2005 [T]
2. David Forsyth, Jean Ponce, *Computer Vision: A Modern Approach,* Prentice Hall [FP]
3. Bernd Jähne, *Digital Image Processing,* Springer, 2011 [J]
4. Linda G. Shapiro, George C. Stockman, *Computer Vision,* Prentice Hall, 2001 [SS]
5. Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing Using MATLAB,* Prentice Hall, 2004 [GW]
6. Henning Bässmann, Jutta Kreyss, *Bildverarbeitung Ad Oculos*, Springer, 2004 [BK]

Other image sources

- *Artexplosion Explosion® Photo Gallery*, Nova Development Corporation, 23801 Calabasas Road, Suite 2005 Calabasas, California 91302-1547, USA [A]
- Corel GALLERY™ Magic 65000, Corel Corporation, 1600 Carling Ave., Ottawa, Ontario, Canada K1Z 8R7 [C]
- David Lowe, Slides [L]
- Copyright Gunther Heidemann [H]

Images from other sources are named explicitly [...].

Motivation

Three lines of research and development:

1. Improving / enhancing images to **facilitate analysis by a human**. This is the task of **image processing.**
   - Repair corrupted images
   - Compensation of bad acquisition conditions
   - Improve perceptibility
   - More generally: Highlight “information” in images

2. **Computer vision:** Recognition of (parts of) the image **by the computer**. This is the main focus of the course.
   - Important sub-tasks:
     - Detection of regions of interest
     - Boundary detection
     - Feature extraction

Motivation

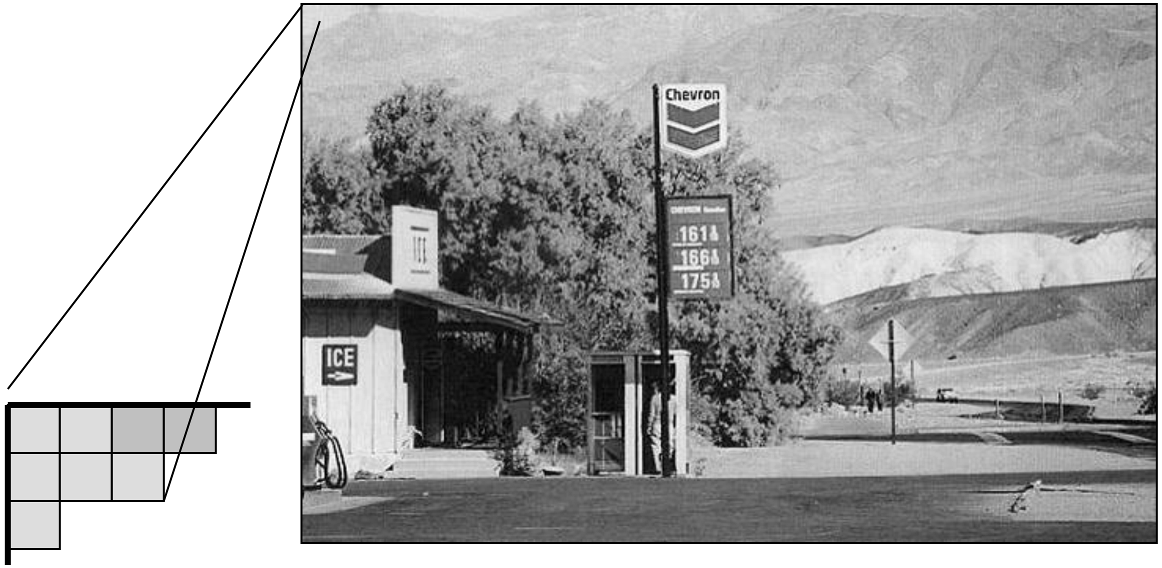
- Classification of colors, shapes, objects
- 3D-representations of real scenes
- Reconstruction of 3D-surfaces
- Motion detection: Object / background separation, direction and velocity computation, object tracking

- Areas of application:
  - Industrial quality control
  - Character recognition
  - Person recognition and tracking
  - Medical image processing
  - Surveillance
  - Driver assistance
  - Image search in databases / internet
  - Robotics (e.g. autonomous vehicles, underwater robotics)
- Advanced: Recognition of meaning
- Advanced image processing often requires computer vision → no clear boundary between 1. and 2.

3. **Understanding of human vision** and pattern recognition in general
   - About 25% of the human brain deals with vision
   - Basic problems such as object recognition not yet understood:
     - No technical solutions (except for special cases)
     - Biological solutions (brain) only understood in certain aspects
   - Understanding vision is both the foundation and the result of computer vision research!

What is a digital image?

Image is represented as 2d-array of pixels:



[T]

Examples of Image Processing: Restoration

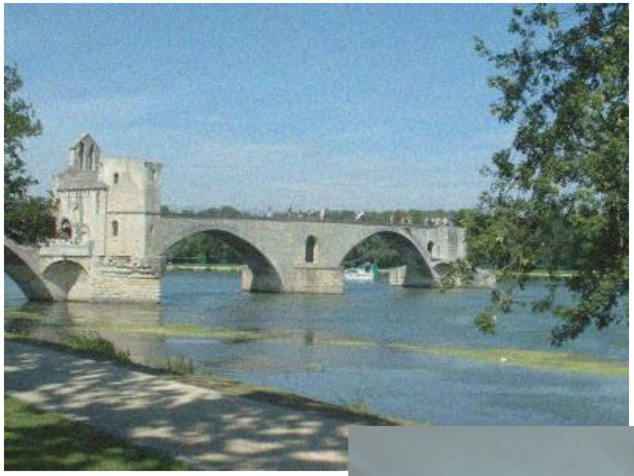
[T]

Restoration and Enhancement

Newspaper Rock, Utah

[T]

Restoration and Enhancement



[T]

Problem:

Infer local patterns from pixels, infer scene from local patterns!

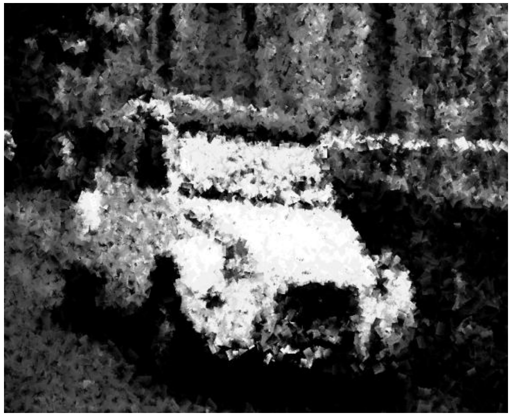
Pixel:
- Luminance
- Color
- Position

→

Interpretation:
E.g. pattern, texture, edge, corner, highlight

→

Interpretation:
E.g. object category, properties, scene geometry

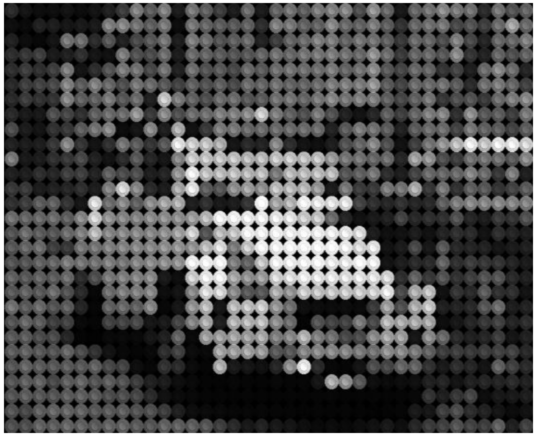


[T]

→ **2 CV**

Computer Vision: Context

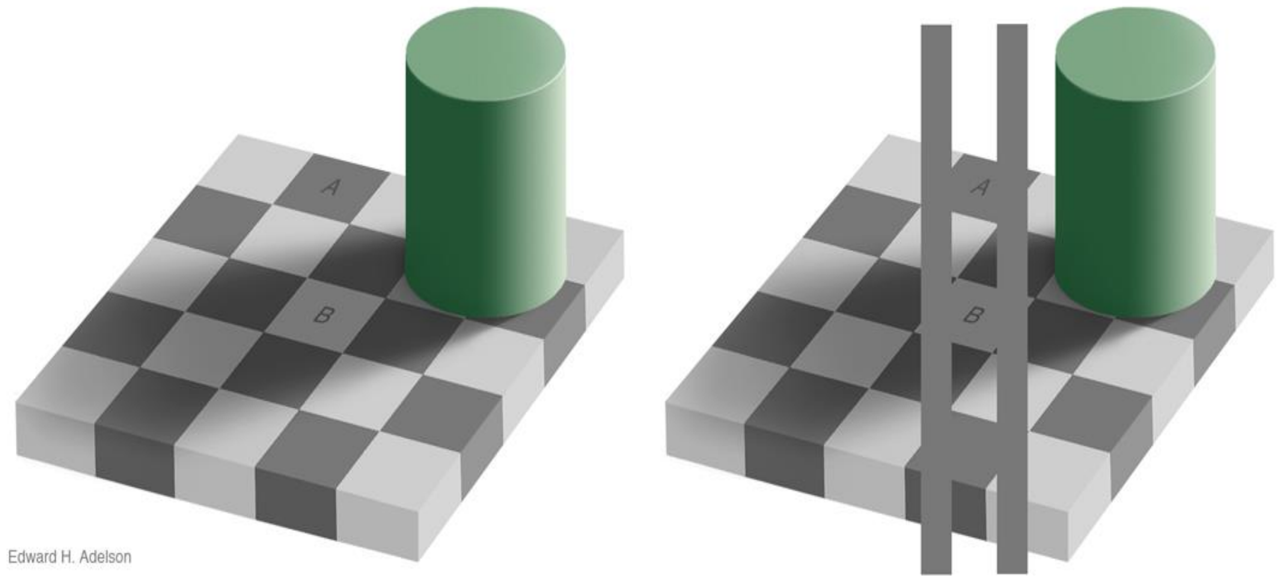
- Interpretation often impossible based on pixels only
- Image interpretation requires context
  - Spatial context (of the pixel or region)
  - Context of meaning (type of image)
  - Context of task
  - Temporal context (image sequence)
- Image + context provide sufficient information for interpretation, even in the presence of severe disruption!



[T]

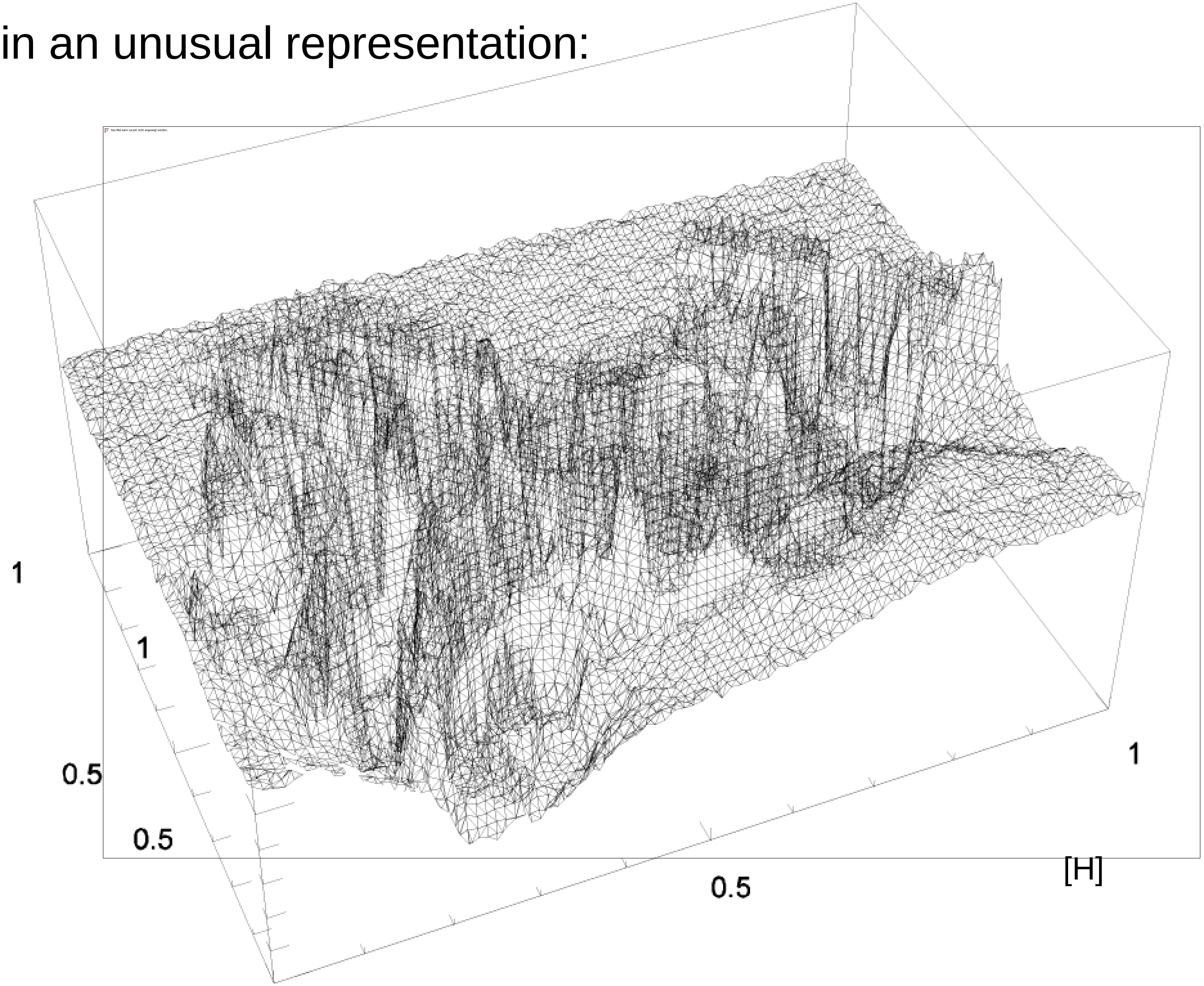
Context

Interpretation of an isolated pixel is context-sensitive:



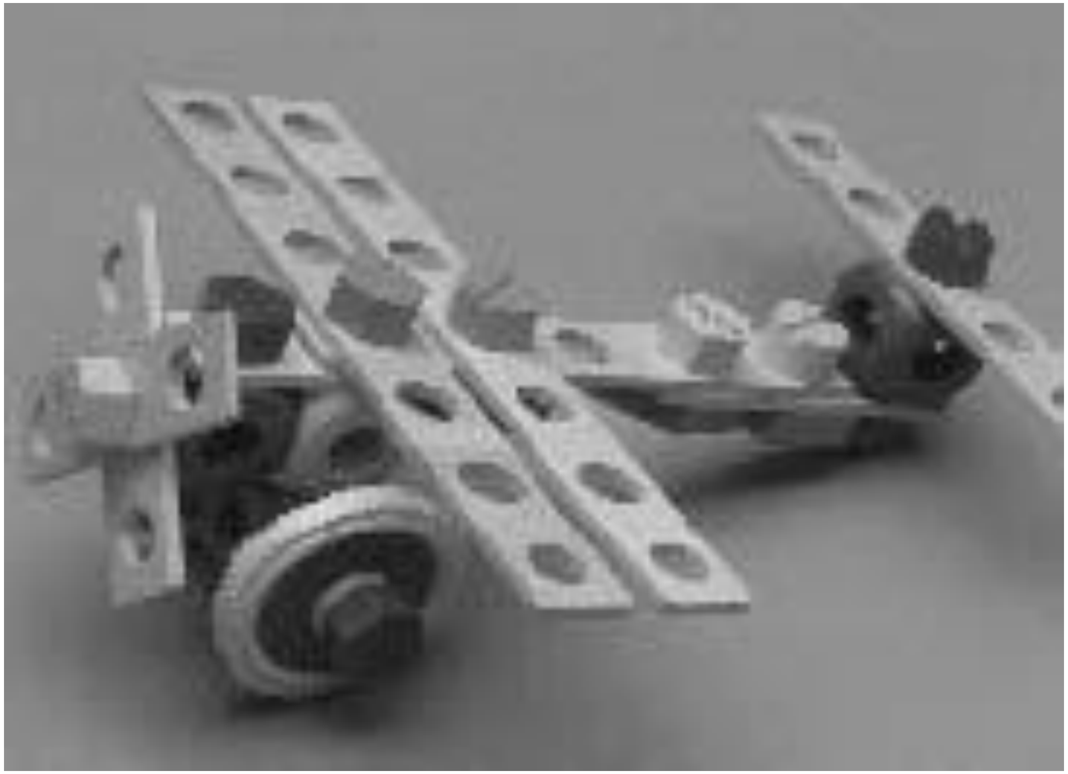
http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html

Image in an unusual representation:



Why is computer vision difficult?

Image of the previous slide:

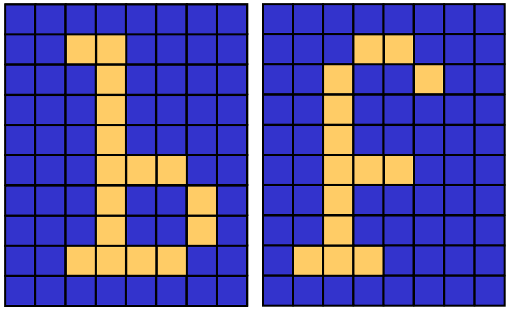


[H]

Ideas for computer vision

- Idea: Assign a hash-code to each image !
- Hash-table holds meaning of the images.
- Hash-code: $h(f) = \Sigma_{x=0,X-1} \Sigma_{y=0,Y-1} f(x,y) \cdot 256^{y \cdot X+x}$

  where pixel $f(x,y)$ denotes the luminance at (x,y) and image dimensions are (X, Y).
- Example: Character recognition in binary 8x10-segments.

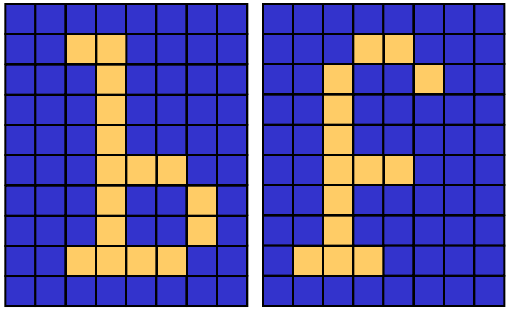


[H]

How big is the hash table?

Ideas for computer vision

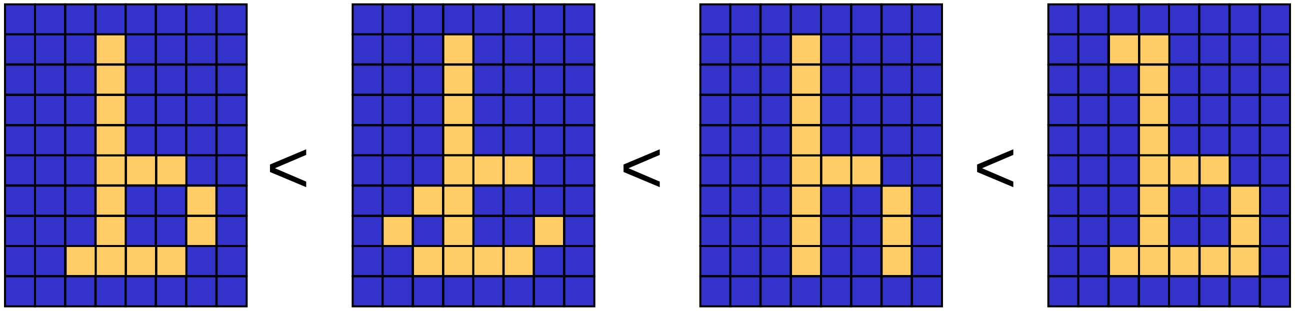
- Hash-code: $h(f) = \Sigma_{y=0,9} \Sigma_{x=0,7} f(x,y) \cdot 2^{y \cdot 8+x}$
- # hash codes: $2^{10 \cdot 8} = 2^{80} \approx 10^{24}$
- Making 1.000.000 entries to the hash table per second we will need 31 billion years
- No generalization (different resolution, different angle …)



[H]

Ideas for computer vision

- Idea: Compressed hash-code
- Find a hash function where the number of codes corresponds to the number of different meanings!
- Problem: We don’t have a hash function that maps similar meaning to similar codes!



[H]

(0,0) is upper left.

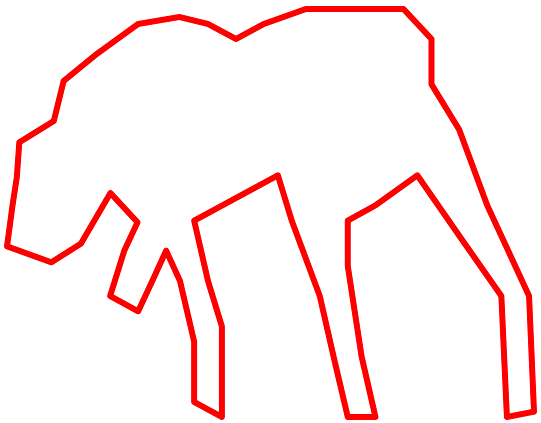
The following questions arise:

- What are „invariant carriers of **information**“ ?
- Connection to the image domain (such as object type, acquisition conditions …) ?

Model based analysis

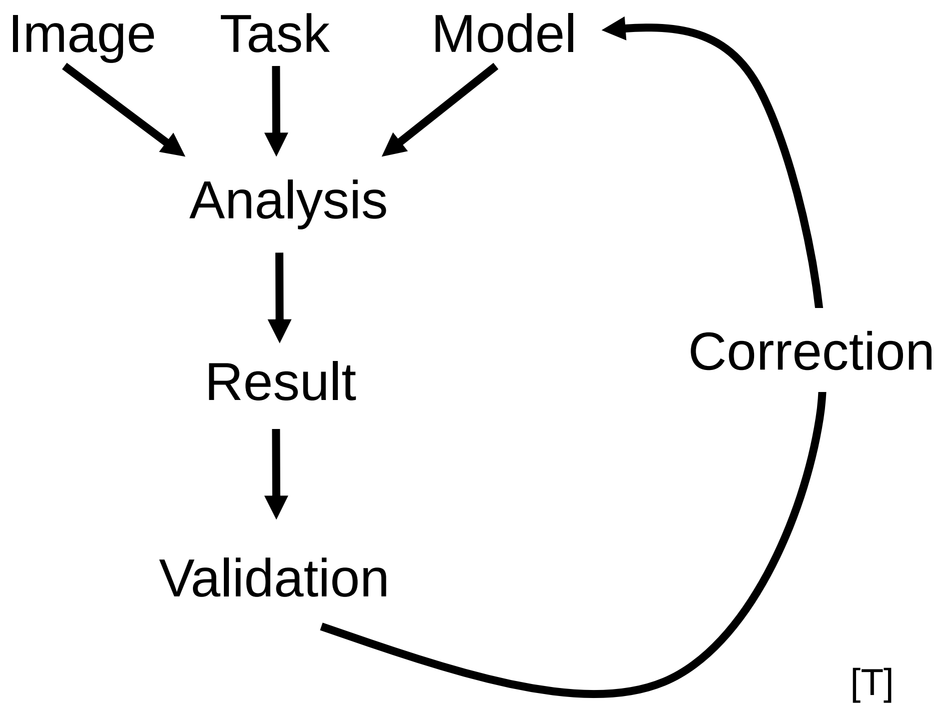
- What kind of information are we looking for? – Task dependent, e.g.
    - existence or identity of objects
    - distribution of intensity
    - direction of illumination
    - motion
- Image interpretation is impossible without some kind of "expectation" (context, task, assumption) !
- Hence we need a model that
    - Provides knowledge about scene, image acquisition, objects …
    - is appropriate for the given task,
    - provides sufficient (but not too much) "degrees of freedom" for adaptation to the observed scene.
- Thus *vision* means: Fit a model to the data such that we get the most likely explanation !

Model based analysis



[T]

Model based analysis



Procedure:

- Fit model to image
- Validation of result
- Correction of model

Model based analysis

How can the model be “fitted” to the data?

Two processing strategies:

1. Bottom up: Starting from the data we are looking for increasingly complex features and connections until they match the model.
2. Top down: Try to “find the model within the data”

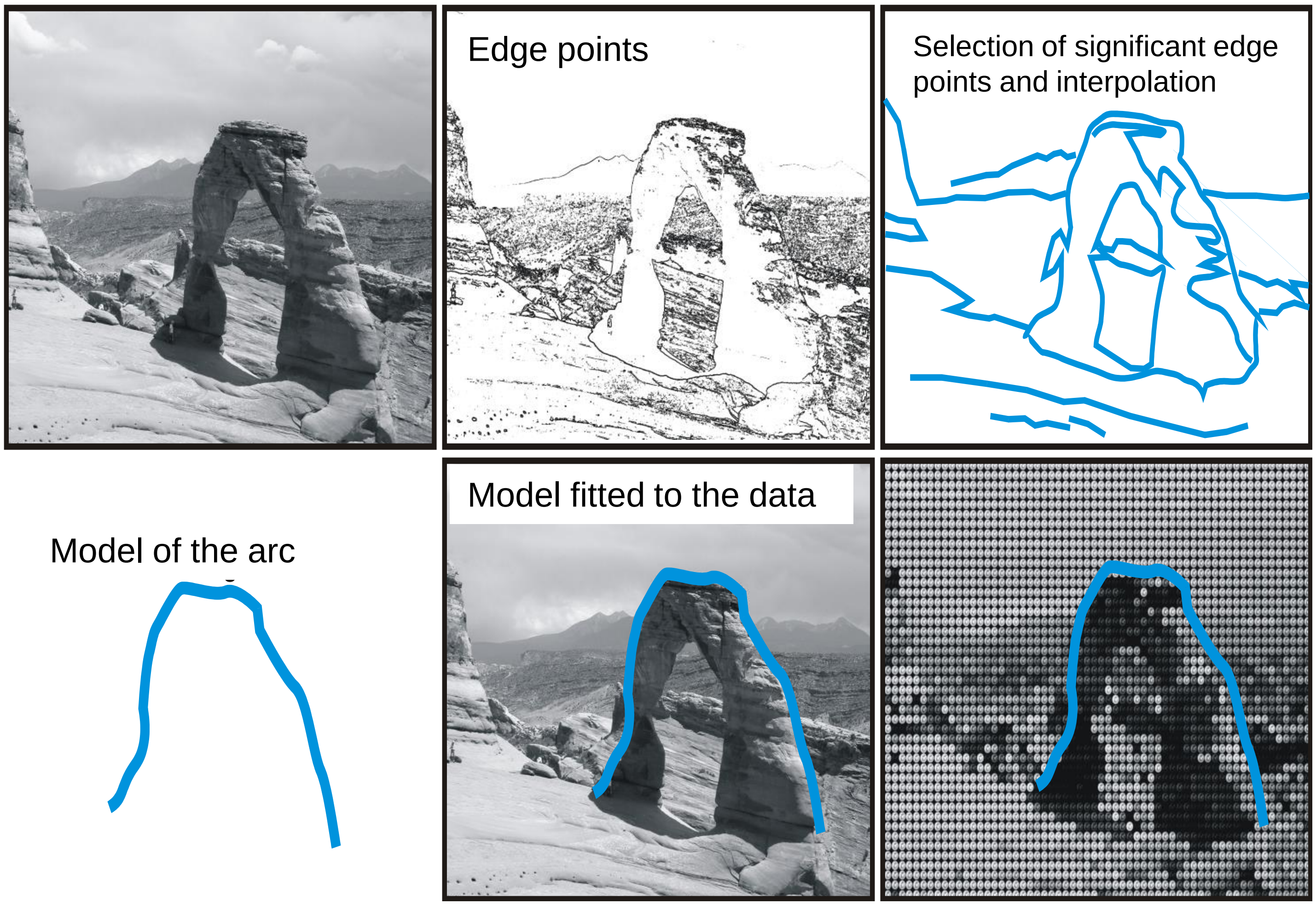
From another point of view, these processing strategies are also called

1. Data driven
2. Model driven

Commonly a mixture of both strategies is used.

Bottom-up and top-down processing

Data driven



Model driven

[T]

Modeling

- To explain an image entirely from the scene, a model must comprise
  - A physical model of the entire scene, i.e., objects, persons, liquids, air (including humidity, mist, dust etc.)
  - All light sources (geometry, location, direction, spectrum)
  - Reflection properties of all surfaces (particularly difficult for skin, liquids, hair)
- Using this model, we could perform the mapping *scene → image*, but still not its inversion *image → scene* !

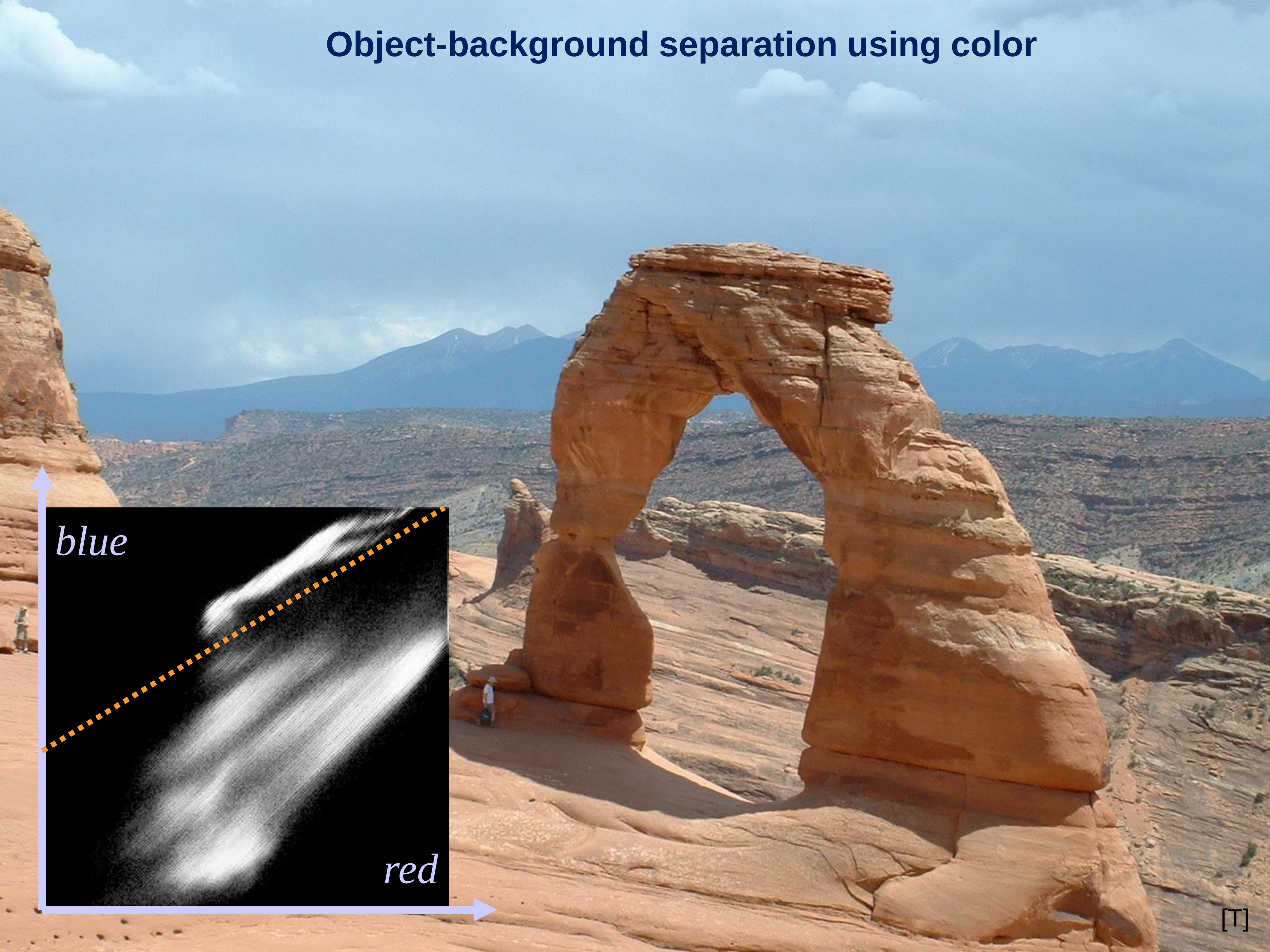
→ Bad idea

→ A model covers only those aspects which are relevant for the task

Modeling

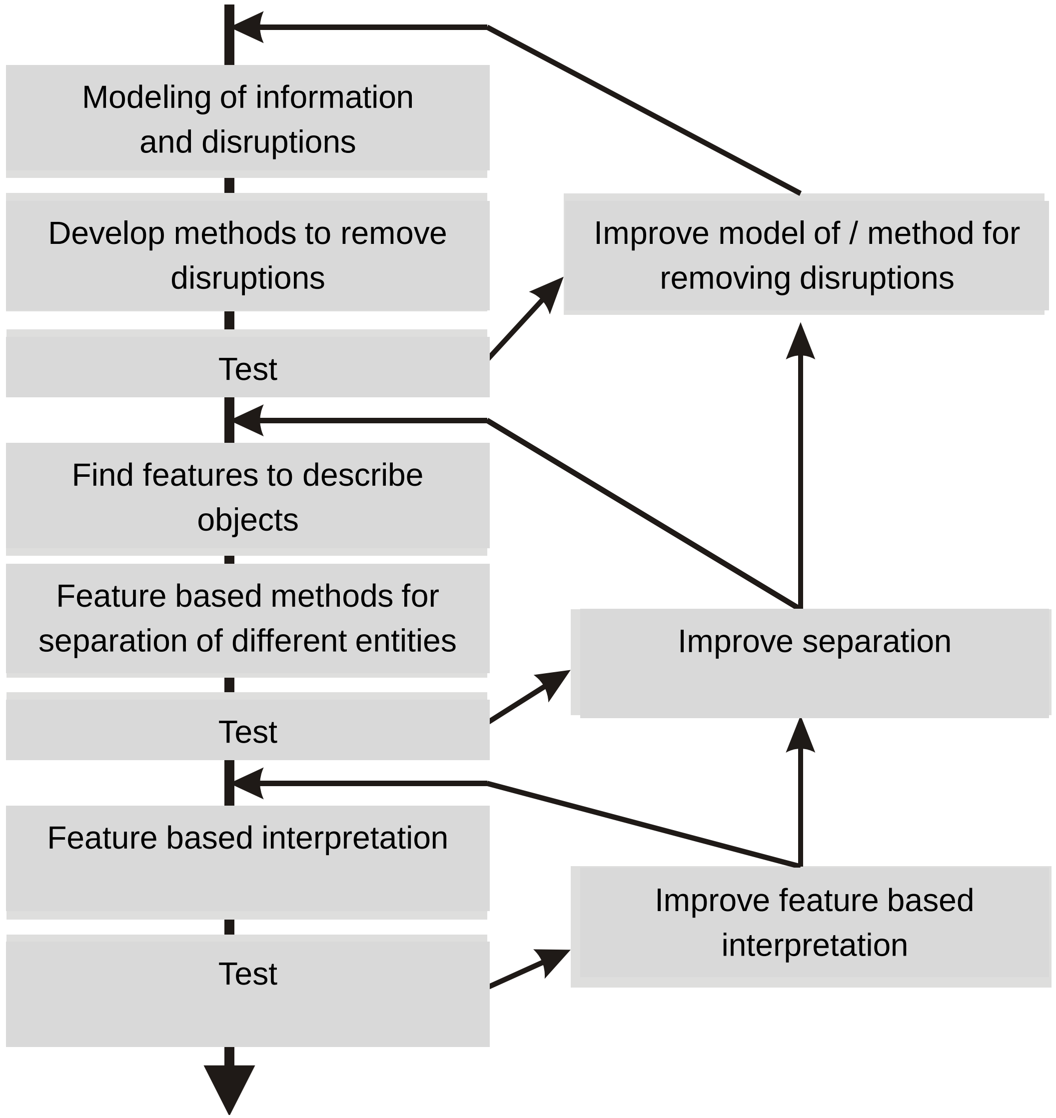
- Examples:
  - Image restoration: Model of image generation
  - Image enhancement: Model of perception / perceptibility
  - Recognition: Models of objects, persons etc.
- To date models often refer only to close-to-signal features (low level features), not to the „high level“ concepts used by humans.
- Example: Object-background separation using the distribution of colors

Object-background separation using color



Object-background separation using color

Typical procedure of building a computer vision system



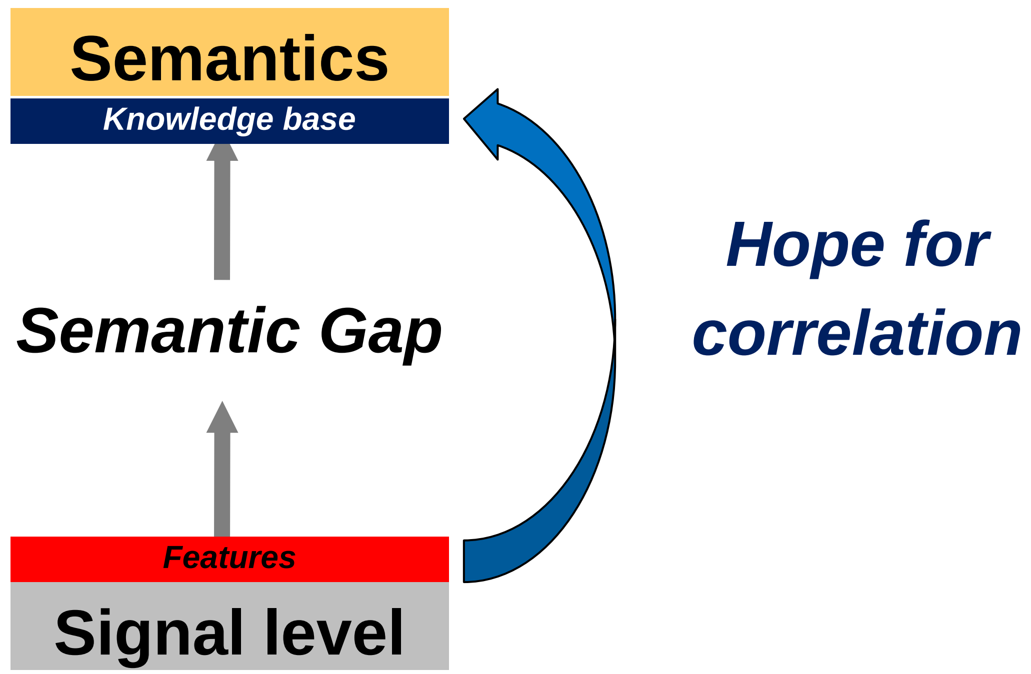
[T]

Problems of computer vision

Todays problems:

- Tasks are specified using high-level concepts of humans.
- But computer vision provides only close-to-signal features.

Semantic Gap



Todays vision systems rely on the correlation between high level concepts and low level features, such as a red spot indicating a traffic light, regardless of other concepts that might exhibit the same features.

Some vision systems

Application examples

Football:

First down recognition from video frames

Important subtasks:

- Camera registration
- Object / background segmentation

[L]

Application examples

[L]

Augmented Reality

Application examples

[L]

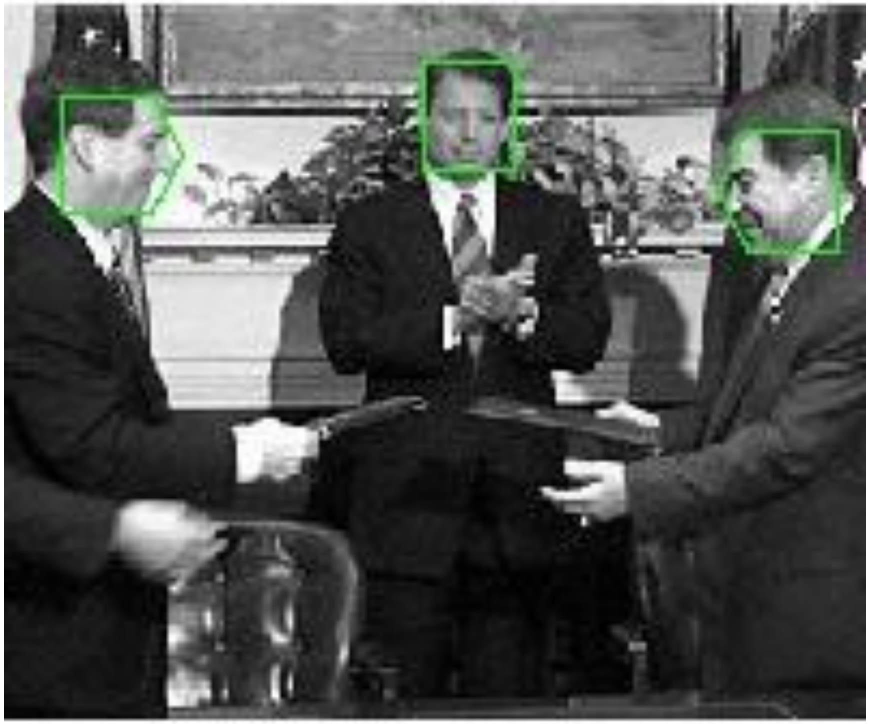
Driver assistance systems

Application examples

http://www.ri.cmu.edu/projects/project_271.html

Face recognition

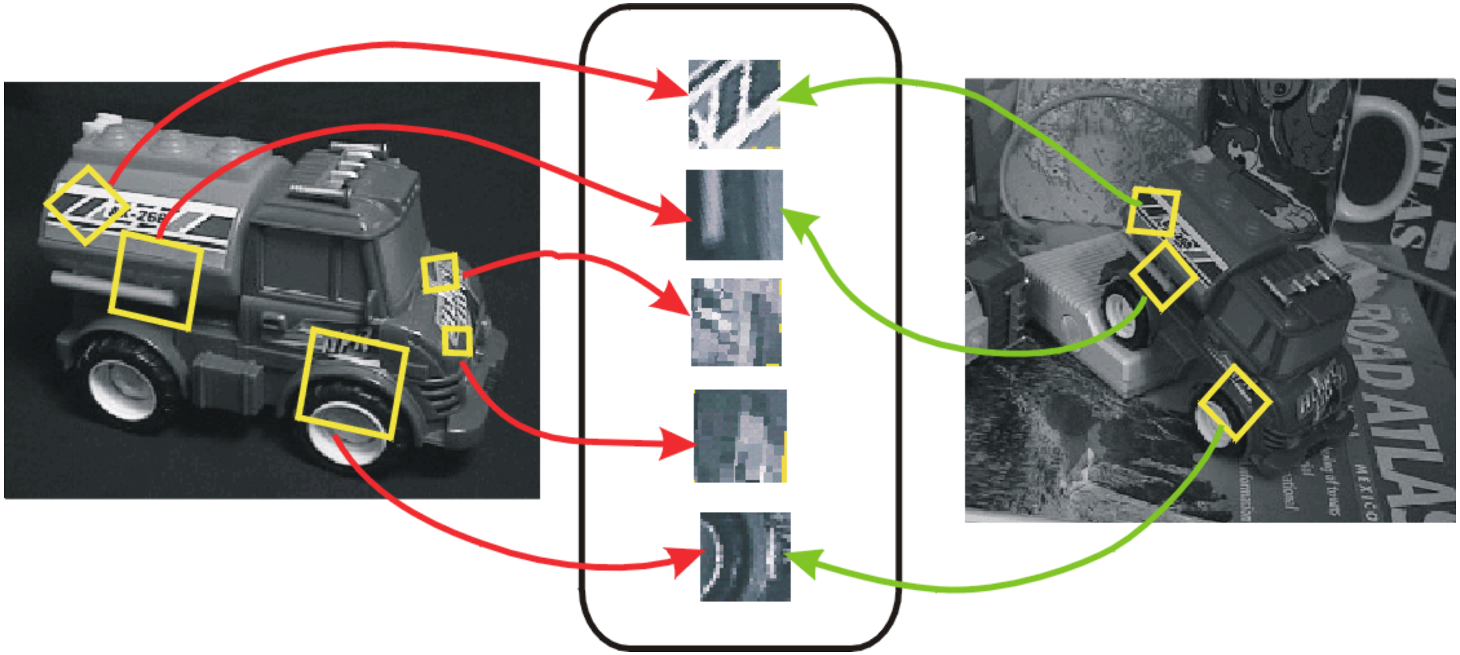
Application examples



http://www.ri.cmu.edu/projects/project_320.html

Face recognition

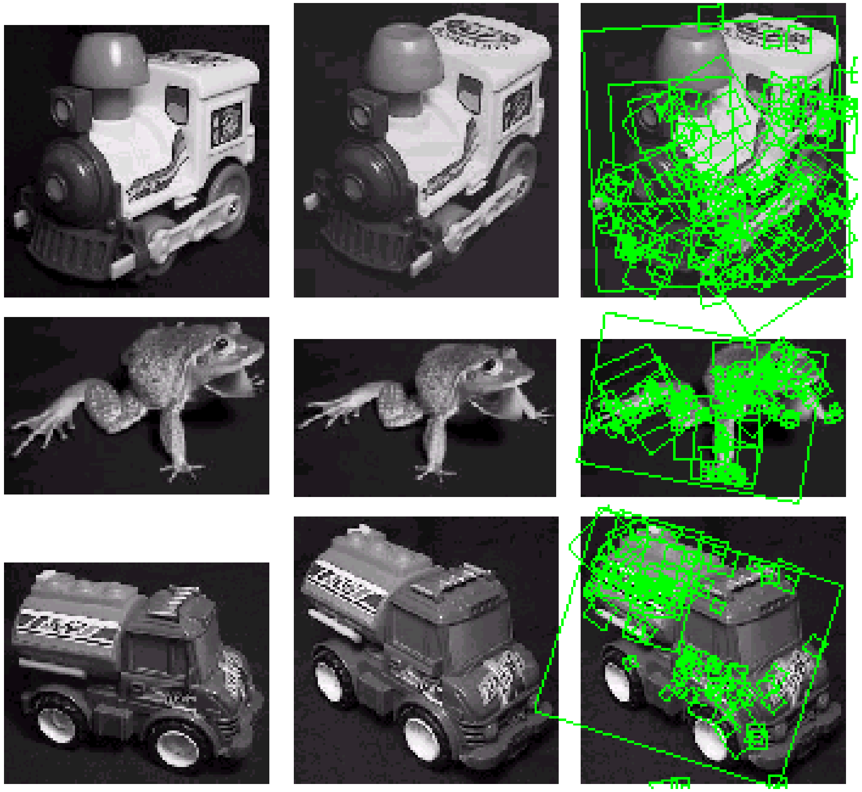
Object recognition



[L]

Object recognition from local features allows robustness against changes of viewpoint and occlusions.

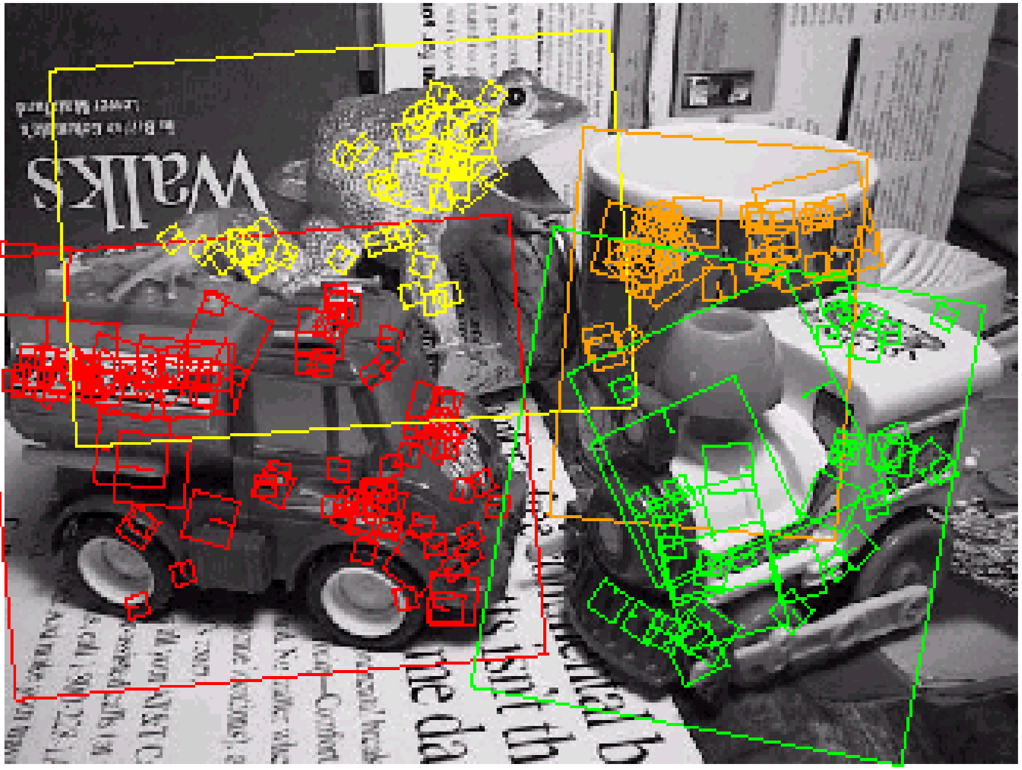
Object recognition



[L]

Interpolation between views

Object recognition



[L]

Recognition in the presence of partial occlusions.

Literature and sources

[T] Klaus D. Tönnies, *Grundlagen der Bildverarbeitung*, Pearson Studium, 2005.

[J] Bernd Jähne, *Digitale Bildverarbeitung,* Springer, 2005.

[FP] David Forsyth, Jean Ponce, *Computer Vision: A Modern Approach,* Prentice Hall, 2002.

[SS] Linda G. Shapiro, George C. Stockman, *Computer Vision,* Prentice Hall, 2001.

[BK] Henning Bässmann, Jutta Kreyss, *Bildverarbeitung Ad Oculos*, Springer, 2004.

[He] Hecht, *Optics*, Addison-Wesley, 1987.

[L] David Lowe, Slides, http://www.cs.ubc.ca/~lowe/425/.

[A] *Artexplosion Explosion® Photo Gallery*, Nova Development Corporation, 23801 Calabasas Road, Suite 2005 Calabasas, California 91302-1547, USA.

[C] Corel GALLERY™ Magic 65000, Corel Corporation, 1600 Carling Ave., Ottawa, Ontario, Canada K1Z 8R7.

[V] Vision Texture Database (VisTex). R. Picard, C. Graczyk, S. Mann, J. Wachman, L. Picard and L. Campbell. Media Laboratory, MIT. Copyright 1995 by the Massachusetts Institute of Technology. http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html

[COIL] S.A. Nene, S.K. Nayar, H. Murase: Columbia Object Image Library: COIL-100, Technical Report, *Dept. of Computer Science, Columbia Univ.,* CUCS-006-96, 1996.

[H] Copyright Gunther Heidemann, 2008.

Literature and sources

[MQ] J. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations. In *Proc. 5th Berkeley Symp. Math. Stat. Probab.,* volume 1, pp. 281-297, 1965.

[CM] D. Comaniciu, P. Meer. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24(5): 603-619, 2002.

[HIm] G. Heidemann. Region Saliency as a Measure for Colour Segmentation Stability. *Image and Vision Computing,* Vol. 23, p. 861-876, 2005.

[TP] M. Turk, A. Pentland: Eigenfaces for Recognition, *J. Cognitive Neuroscience*, Vol. 3, p. 71-86, 1991.

[MN] H. Murase, S.K. Nayar: Visual Learning and Recognition of 3-D Objects from Appearance, *Int. J. of Computer Vision,* Vol. 14, p. 5-24, 1995.

[DL] D. Lowe: Distinctive image features from scale-invariant keypoints, *Int. J. of Computer Vision,* Vol. 60(2), p. 91-110, 2004.

[HCv] G. Heidemann: Combining spatial and colour information for content based image retrieval, *Computer Vision and Image Understanding*, Vol. 94(1-3), p. 234-270, 2004.

[IKH] J. Imo, S. Klenk, G. Heidemann: Interactive Feature Visualization for Image Retrieval. *Proc. 19th Int. Conf. on Pattern Recognition ICPR 2008,* 2008.