



TSBB15 Computer Vision

Examiner:
Per-Erik Forssén



Course Information

- **Website:** <http://www.cvl.isy.liu.se/education/undergraduate/tsbb15>
 - Updated continuously during the course
 - Schedule, lab sheets, pointers to code, old exams etc.



Course Information

- **Website:** <http://www.cvl.isy.liu.se/education/undergraduate/tsbb15>
- **Lectures**



Klas Nordberg



Michael Felsberg



Per-Erik Forssén

+ two special guests



Course Information

- **Website:** <http://www.cvl.isy.liu.se/education/undergraduate/tsbb15>

- **Lectures**



Klas Nordberg



Michael Felsberg



Per-Erik Forssén

+ two special guests

- **Computer Exercises and Project Guides**



Martin Danelljan
CE + PG



Gustav Häger
CE + PG



Hannes Ovrén
CE + PG

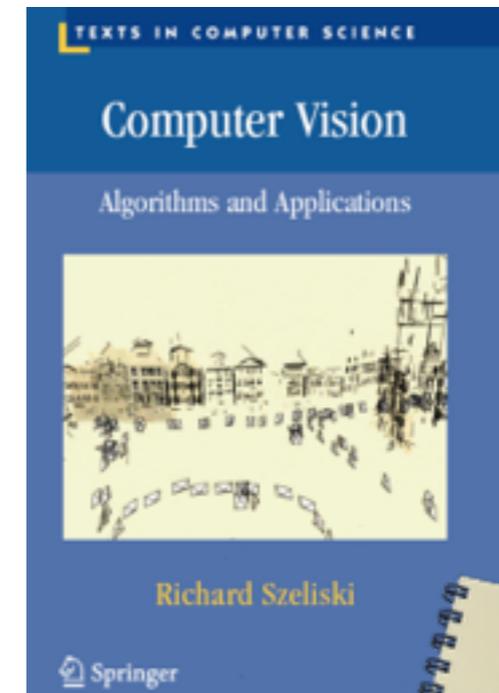


Marcus Wallenberg
CE



Course Book

- Main book: Computer Vision— Algorithms and Applications, By Richard Szeliski 2011. Electronic version - see webpage.
- For geometry we will also use the IREG compendium by Klas Nordberg.
- Other sources as linked in the Materials column on the webpage





What is Computer Vision?

- **Course aims and structure**
methods, theory, exercises, projects
- **Applications**
- **Research**
Examples from CVL research projects
- **Examples from CDIO student projects**
- **Methodology**



Course Aims

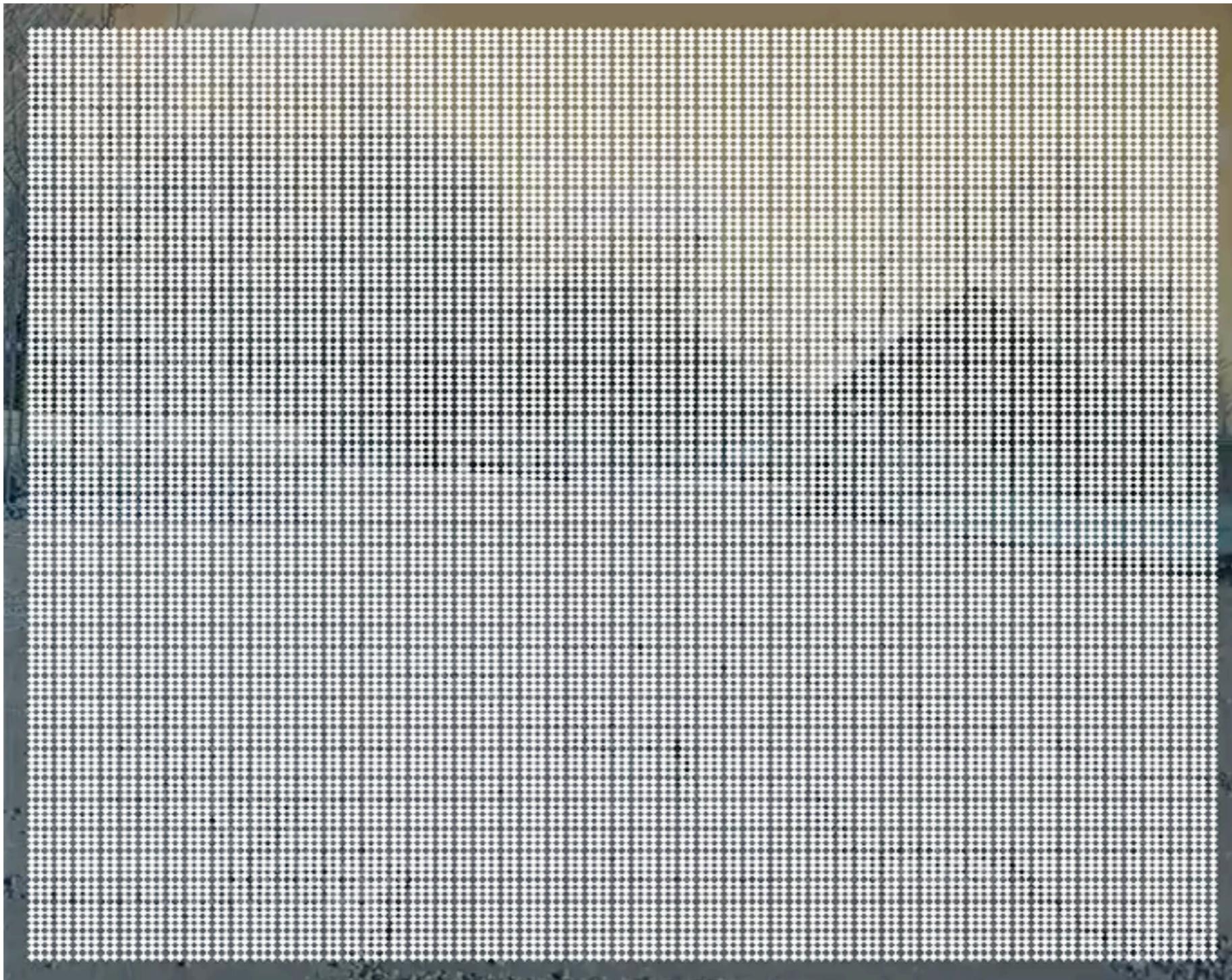
- From the LiTH Study Guide:

After having passed this course, the student is expected to be able to **describe problems and algorithms** for the following basic computer vision and image processing tasks:

- tracking of image regions, triangulation from stereo images, estimation of optical flow, detection of several image features, matching of image features, graph and tree structures and other image representations, generative image models, segmentation of image regions, enhancement of images, debugging and visualisation



Tracking





Motion Estimation





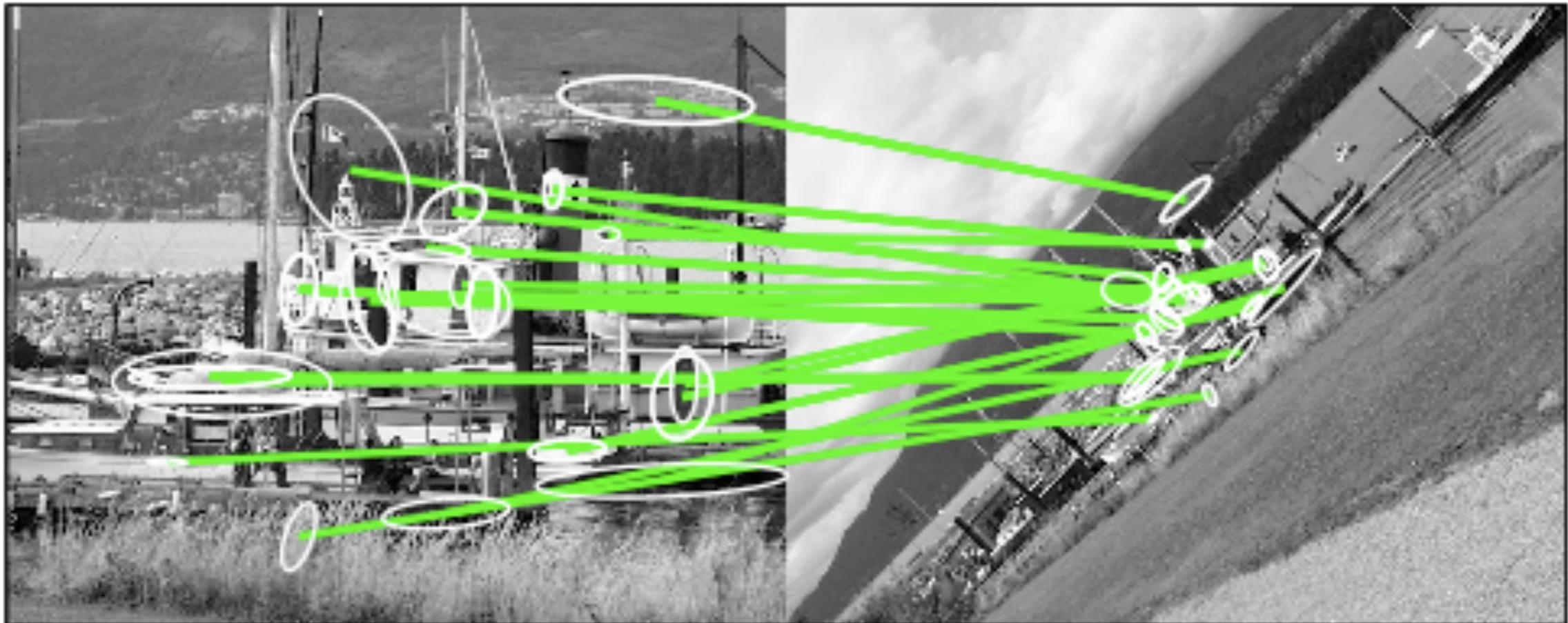
Feature Matching



Find out which image regions are corresponding



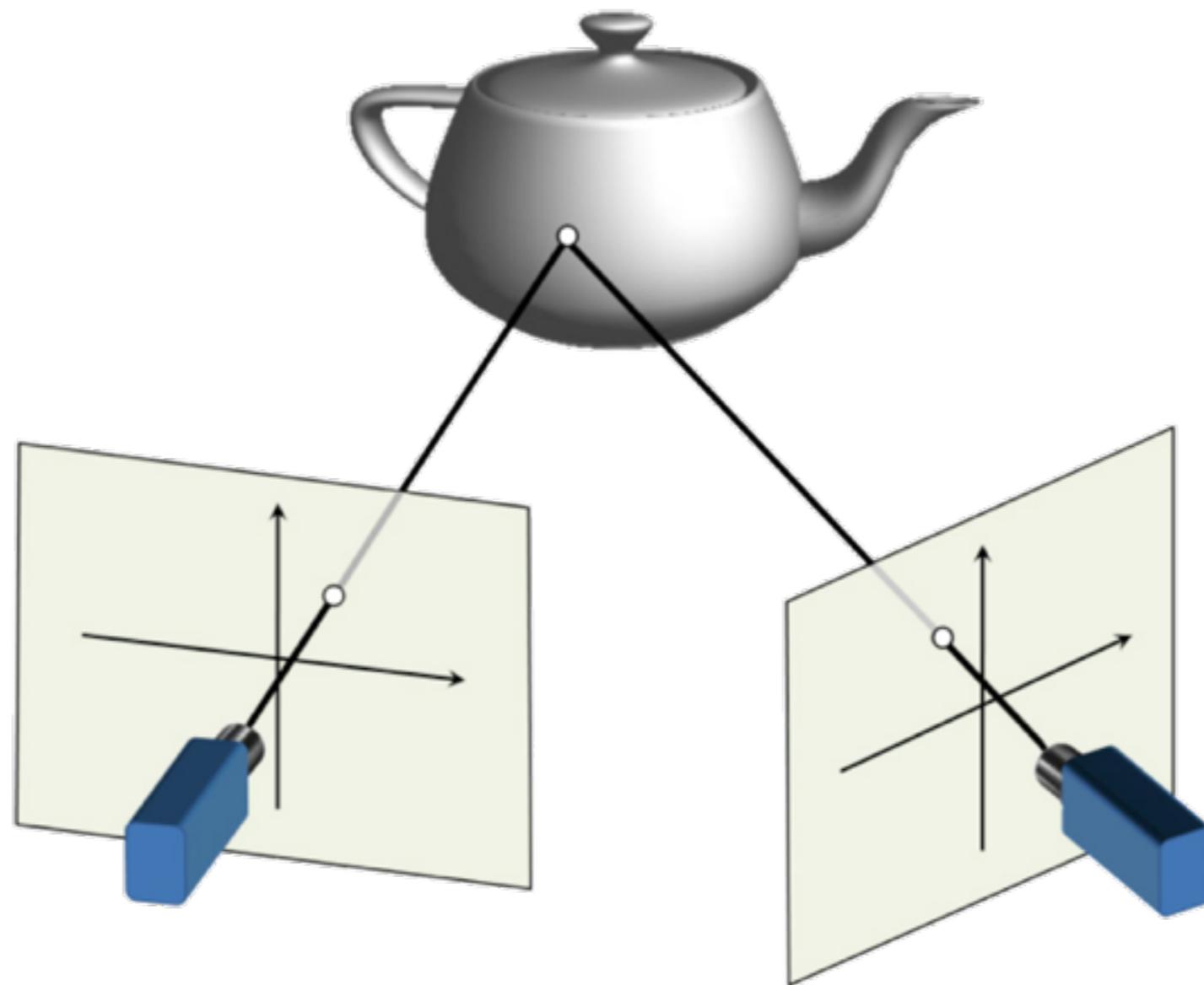
Feature Matching



Find out which image regions are corresponding (LE6)



Triangulation



T. Moons et al. FTGV 2008



Structure from motion

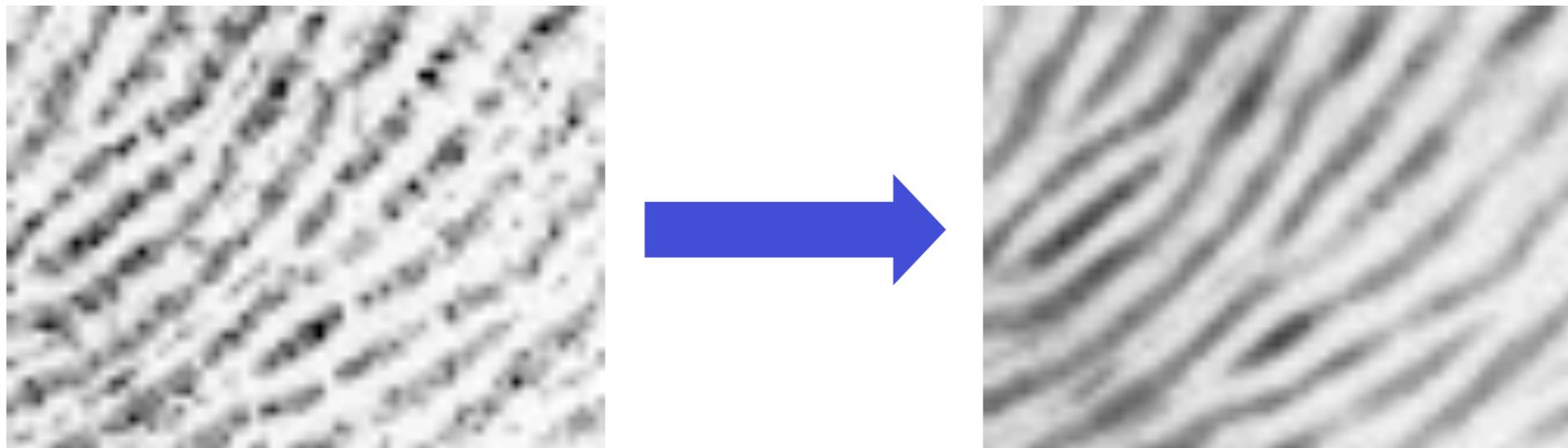


Image credit: Bundler home page <http://www.cs.cornell.edu/~snavely/bundler/>

- Project 2: Start with two views, and successively add more.
- Finally perform simultaneous refinement of 3D point positions and camera positions (Bundle adjustment).

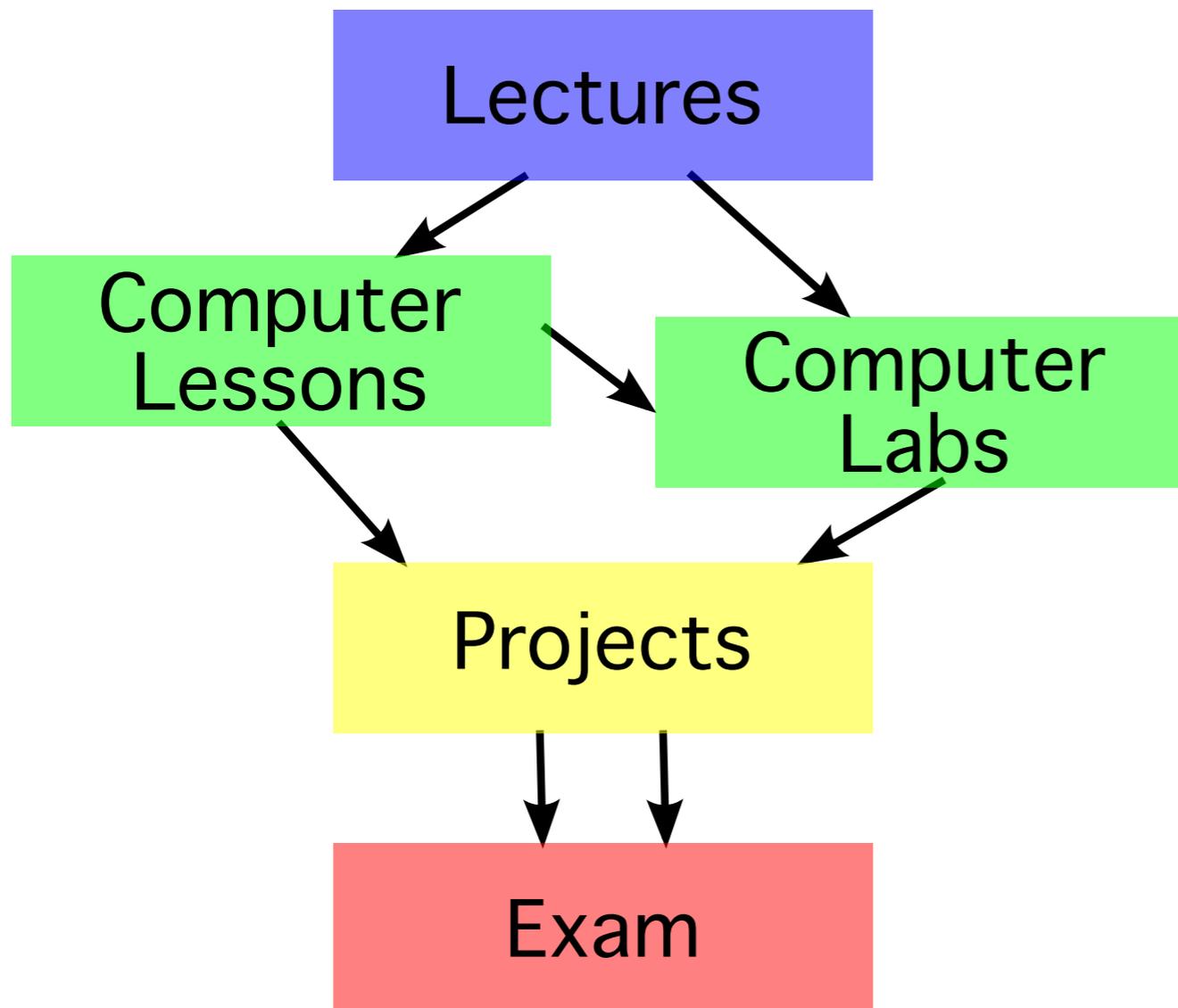


Image Denoising, Enhancement and Inpainting





Course contents



- Theory and two levels of practise:
- **Computer exercises** are designed to be useful for completing the projects. If you prepare well and understand them, the projects will be easier.
- **Projects** will give you plenty of time to try things out in practise, and gain better understanding for the exam.



Course contents

- **16 lectures** (2h each)
New theory, and project introductions
- **Two computer lessons** (2h each)
How to use Matlab and Python for computer vision.
Hints for the exercises.
- **Four computer exercises** (4h each)
Practical experience of critical skills needed for the projects.
- **Written mid-term exam** (4h)
Voluntary, but very useful
- **Written exam** (4h)
Only half if you do well on the mid-term exam
- **Two group projects** (50% of course)



Warning: Labs are hard!

- The computer labs are more difficult than in other courses.
- Each lab sheet has an estimate of preparation time. This is **assuming that you have already done all parts of the computer lessons** and understood them fully. If not expect longer preparation.
- Labs are however very useful for the projects, for the project course, masters thesis projects etc.



How do I pass the course?

- **Pass the four computer exercises**
come prepared (read lab document and solve preparatory questions)
- **Pass the two group projects**
code in svn, report, oral presentation
- **Pass the written exam**



What grade will I get?

- If you have passed the course, the grade is a weighted average of the project grades and the exam grade:

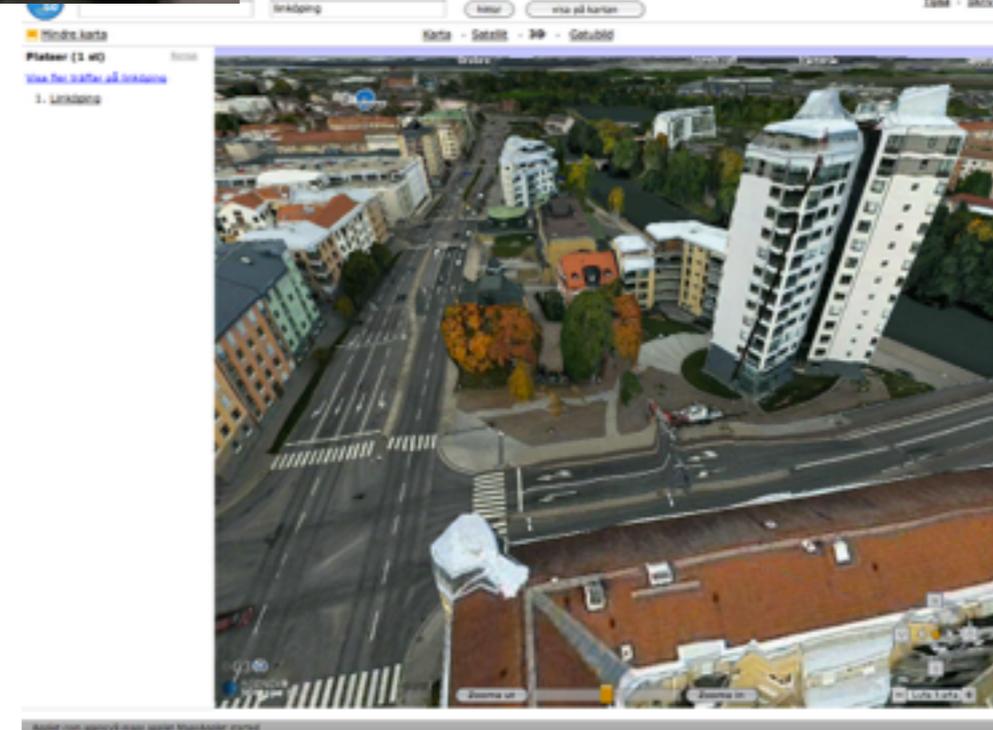
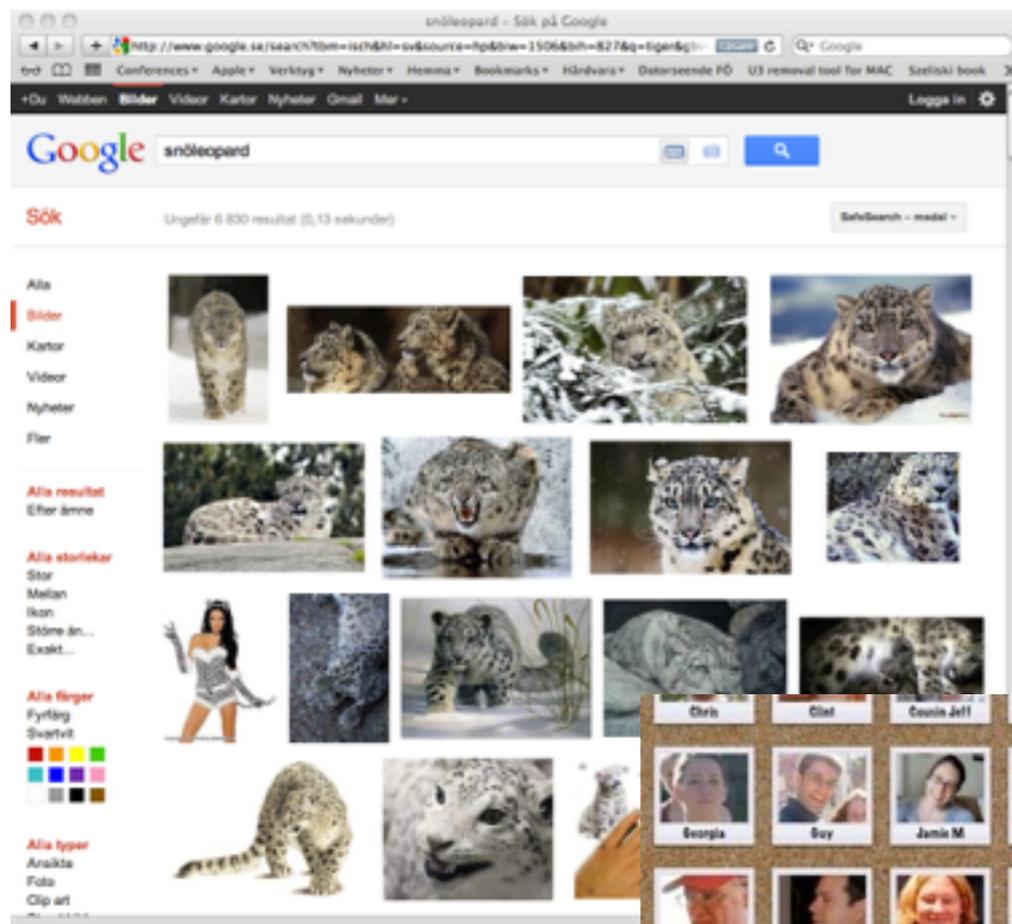
```
course_grade = round(0.48*exam_grade+  
                    0.26*project1_grade+  
                    0.26*project2_grade);
```

- Project grade 4 if all requirements are satisfied
Project grade 3 if minor errors
Project grade 5 if an individual extra assignment is completed



Applications

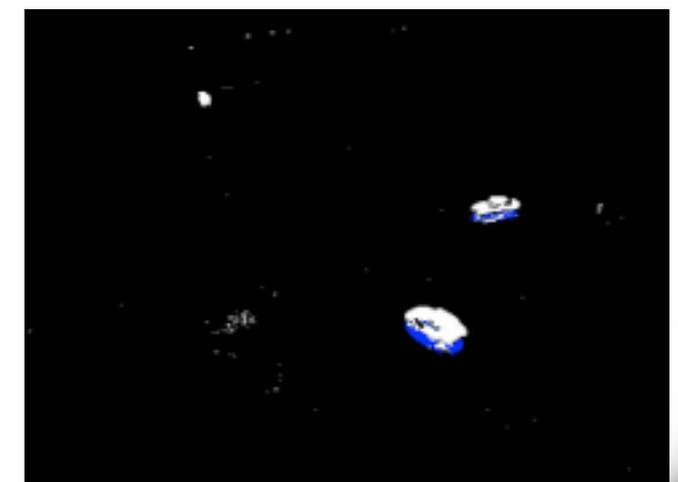
- Google, Apple (w. C3 and Polar rose), Facebook, MS





Applications

- Hollywood, Autoliv/Bosch, Volvo/BMW/Google/Tesla, IVSS (e.g. Vägverket)





MATRIS (2004-2007)

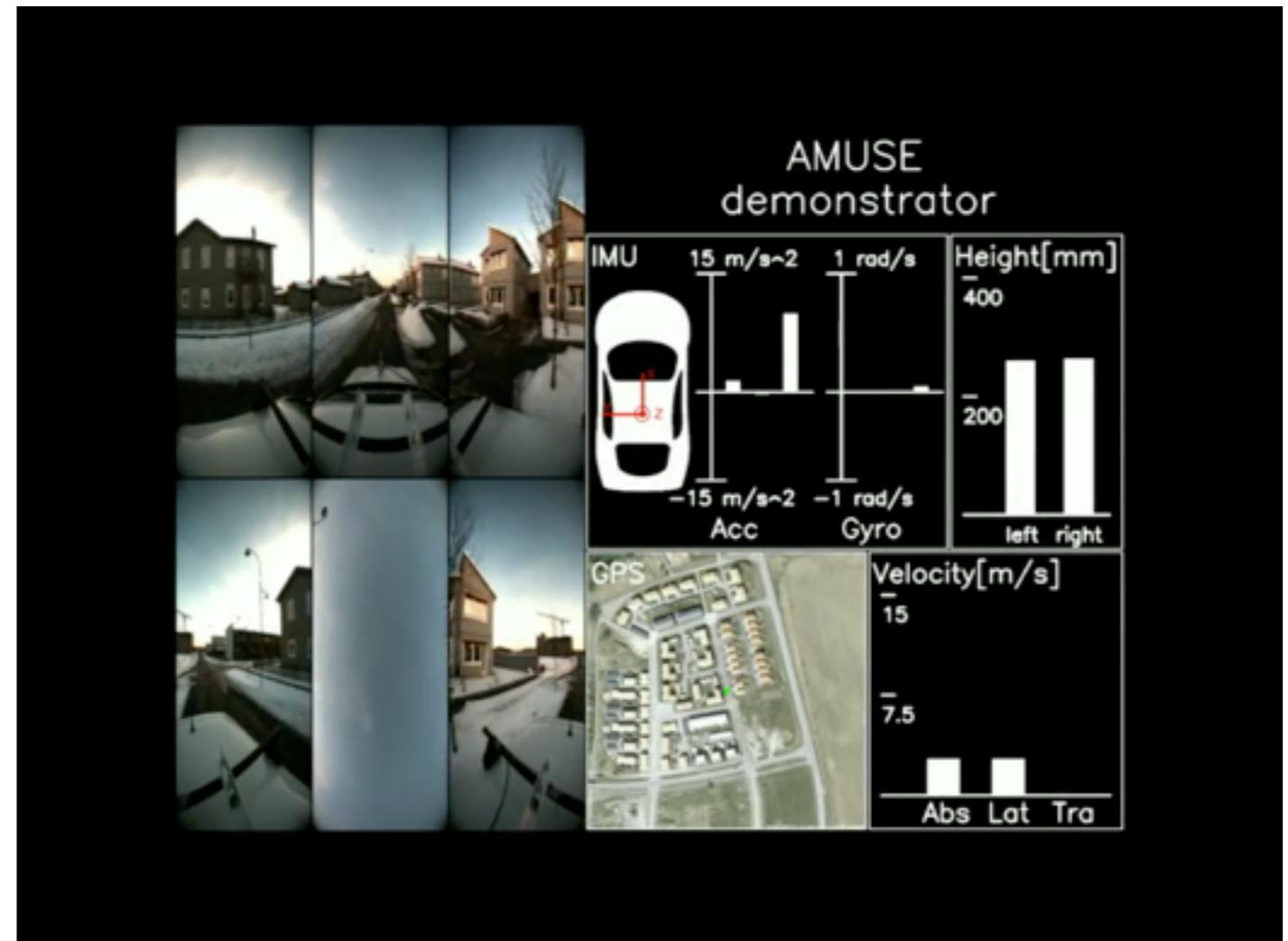


Markerless Real-Time Tracking for
Augmented Reality Image Synthesis



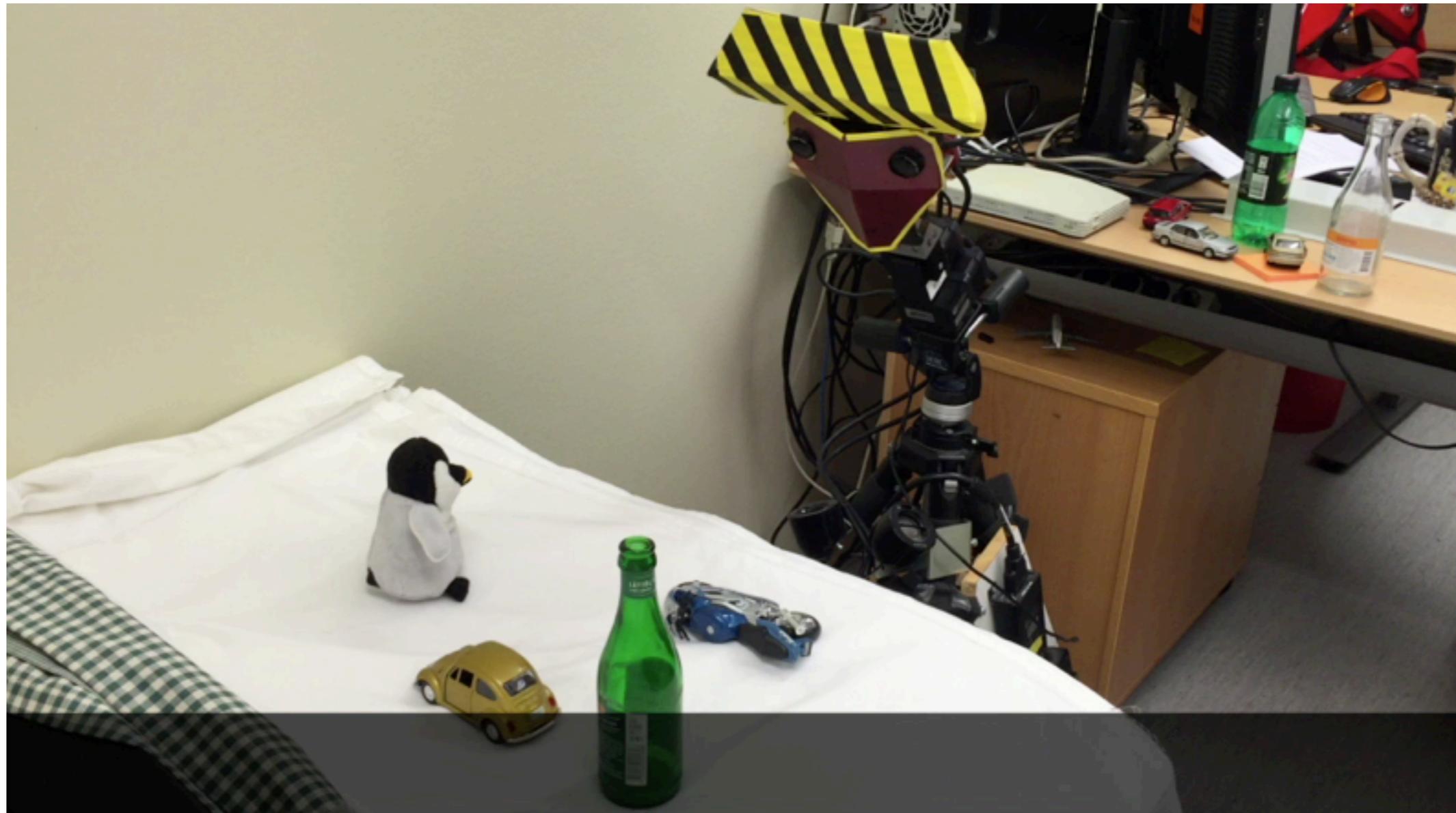
Visual perception in cars

- EU-project DIPLECS (2007-2010)
- VINNOVA IQmatic (2013-2016)
- Daimler collaboration





Embodied Object Recognition 2009-this Friday





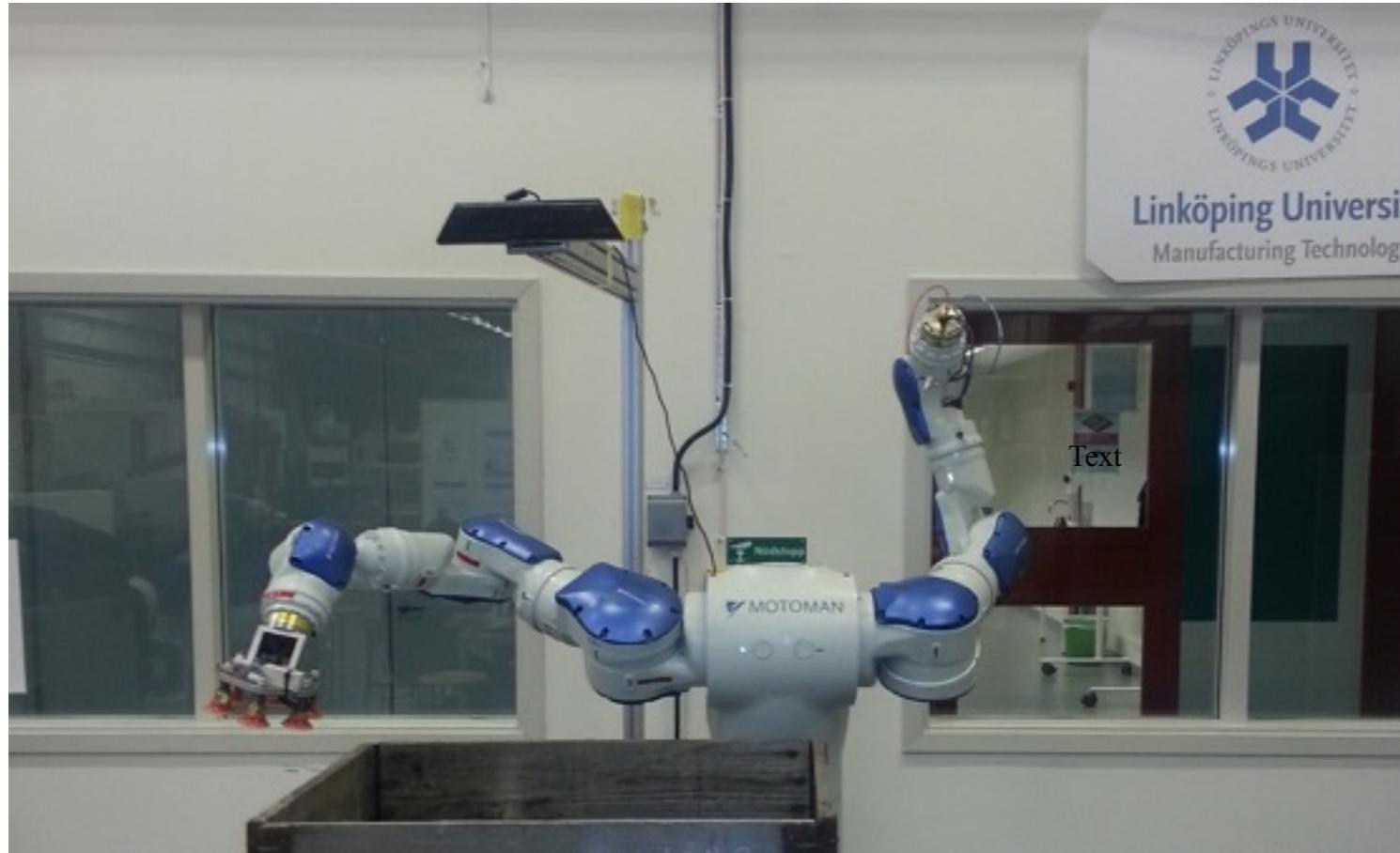
Rolling Shutter Correction 2009-



- LCMM (Learnable Camera Motion Models) project



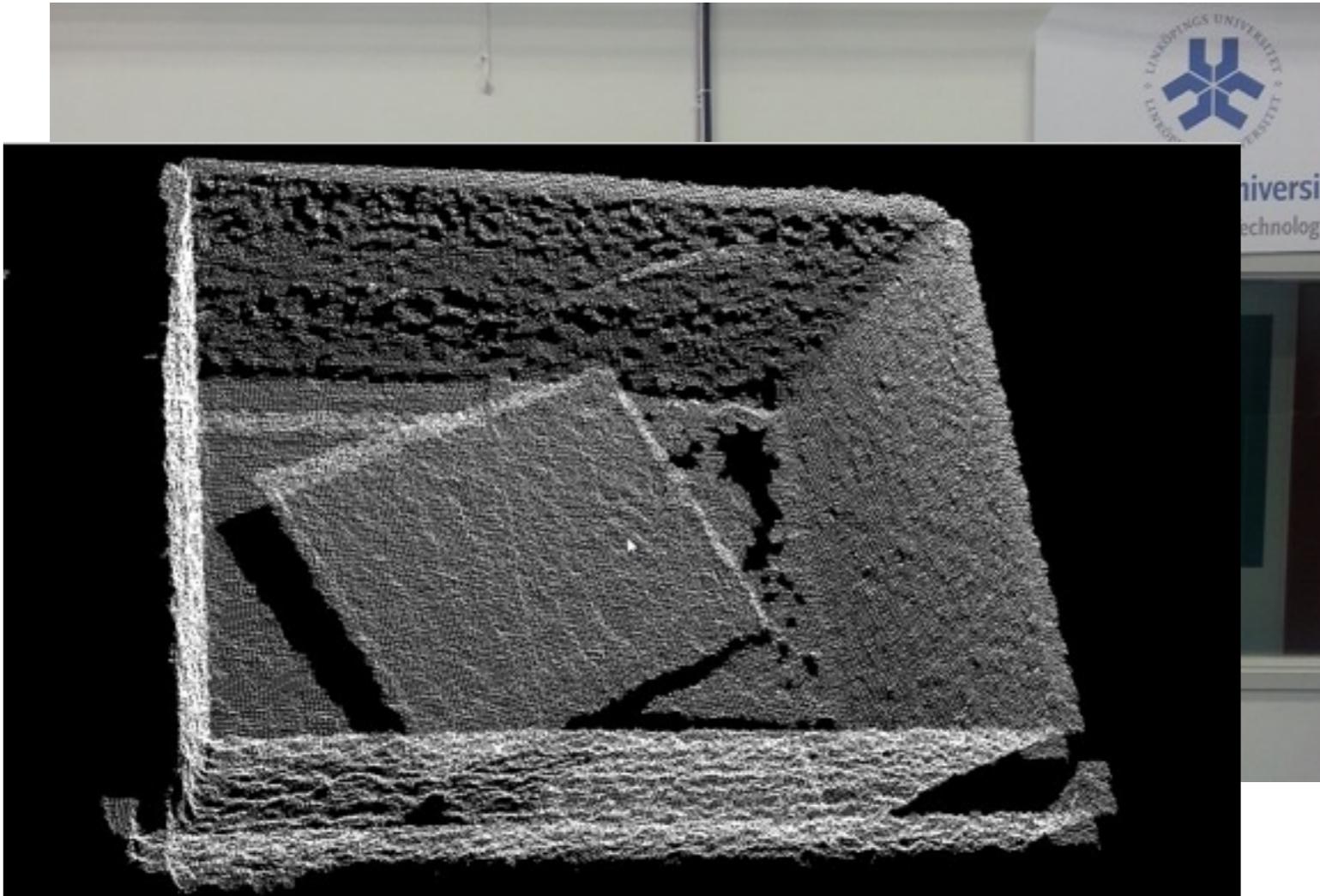
CDIO Project Example (2012)



- Robot should pick up used computer monitors from a crate.
- Perception using a Microsoft Kinect depth sensor.



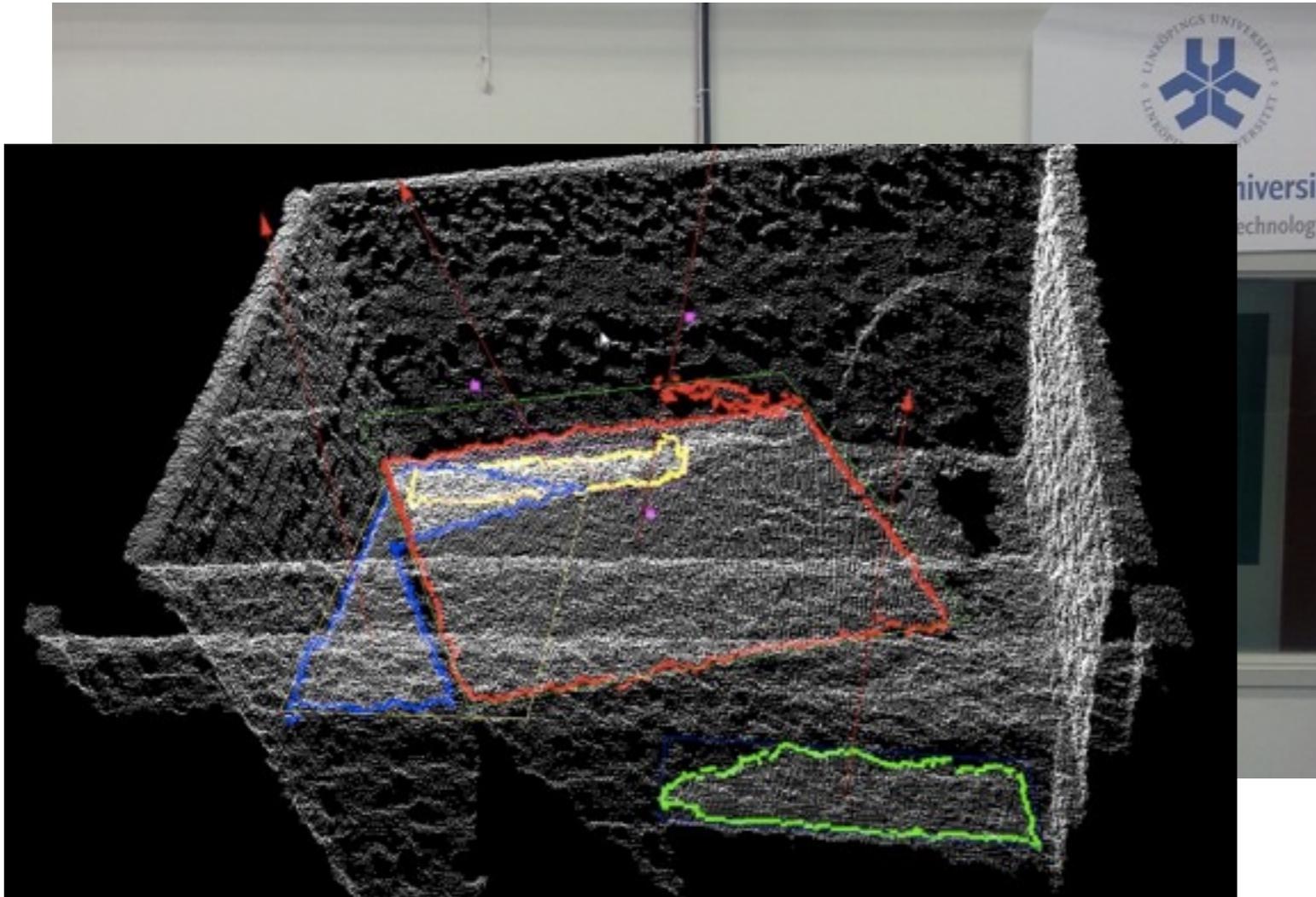
CDIO Project Example (2012)



- Robot should pick up used computer monitors from a crate.
- Perception using a Microsoft Kinect depth sensor.



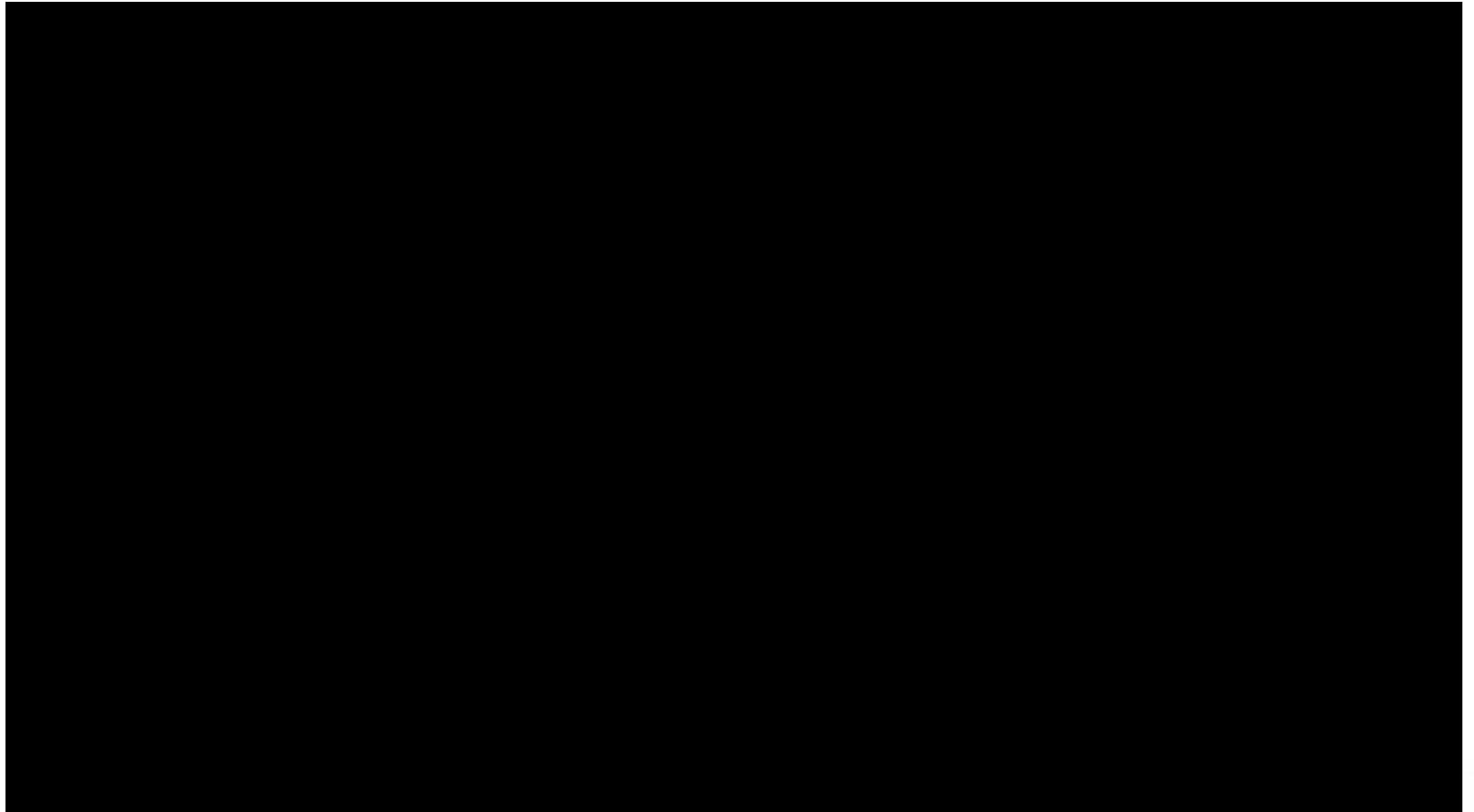
CDIO Project Example (2012)



- Robot should pick up used computer monitors from a crate.
- Perception using a Microsoft Kinect depth sensor.



CDIO Project Example (2012)





CDIO Project Example (2012)



- Robot Dog
the robot should be able to recognize and to follow a specific person.



CDIO Project Example (2012)





Methodology

- Local models
- Local features (detector+descriptor)
- Representations
- Estimation



Local models

- Here is an image that we want to analyze:



- How do we begin?



Local models

- Global approach: $\mathbf{I}(\mathbf{x}) \quad \mathbf{x} \in \Omega \subset \mathbb{Z}^2$

stack all pixels in a vector:

$$\mathbf{v} = [\mathbf{I}(\mathbf{x}_1) \dots \mathbf{I}(\mathbf{x}_k) \dots \mathbf{I}(\mathbf{x}_K)] \quad \mathbf{x}_k \in \Omega$$





Local models

- Global approach: $\mathbf{I}(\mathbf{x}) \quad \mathbf{x} \in \Omega \subset \mathbb{Z}^2$

stack all pixels in a vector:

$$\mathbf{v} = [\mathbf{I}(\mathbf{x}_1) \dots \mathbf{I}(\mathbf{x}_k) \dots \mathbf{I}(\mathbf{x}_K)] \quad \mathbf{x}_k \in \Omega$$



- A local model instead looks at limited neighbourhoods.
- These may be approximated by a simple model. (e.g. the signal is locally 1D)
- Each model tells us a little bit about the image.





Locality

- **Local appearance**
describe the appearance for relatively small image regions
- **Local processing**
processing of images is done locally, i.e. the result at a certain image point depends only on the input data in a neighborhood of the point.
⇒ Suitable for implementation on graphics cards.



Locality

- This is also how the brain does it (LE9).
- Locality in a resolution pyramid (LE2).
- CNN (Convolutional Neural Networks)

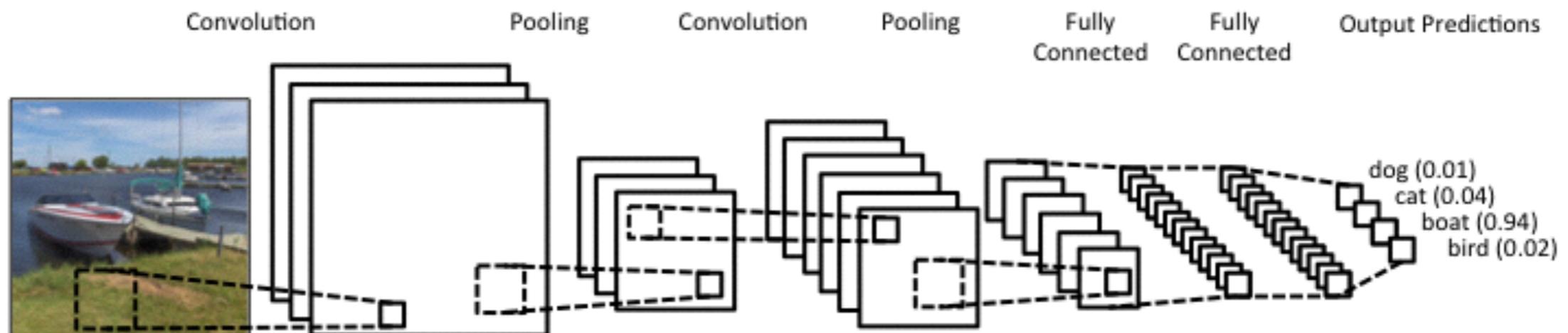
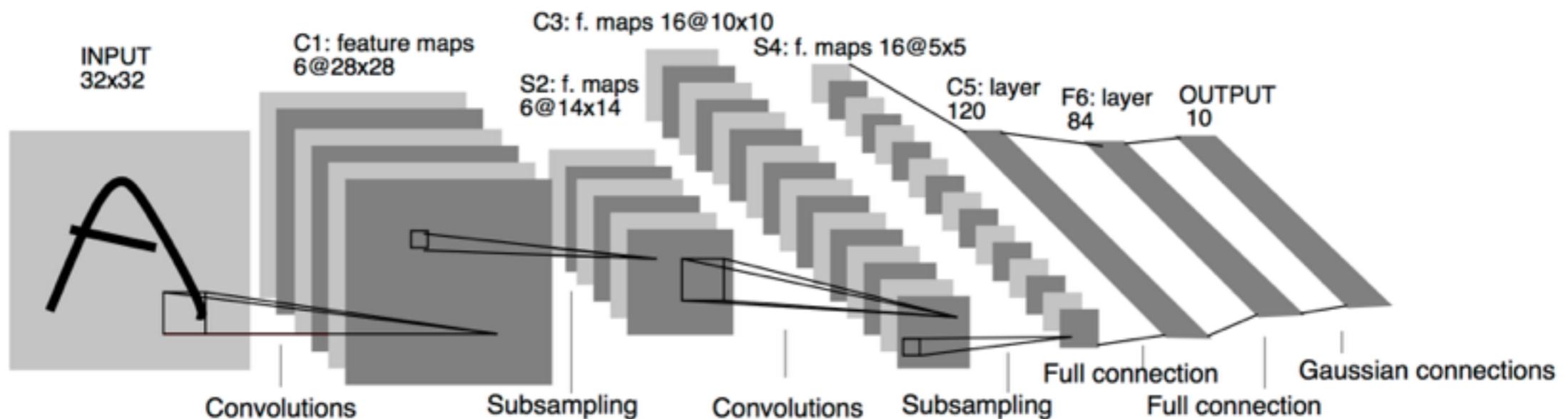


Image source: <http://www.clarifai.com/technology>



Locality

- This is also how the brain does it (LE9).
- Locality in a resolution pyramid (LE2).
- CNN (Convolutional Neural Networks)



- LeNet-5 Y. LeCun et al. 1988

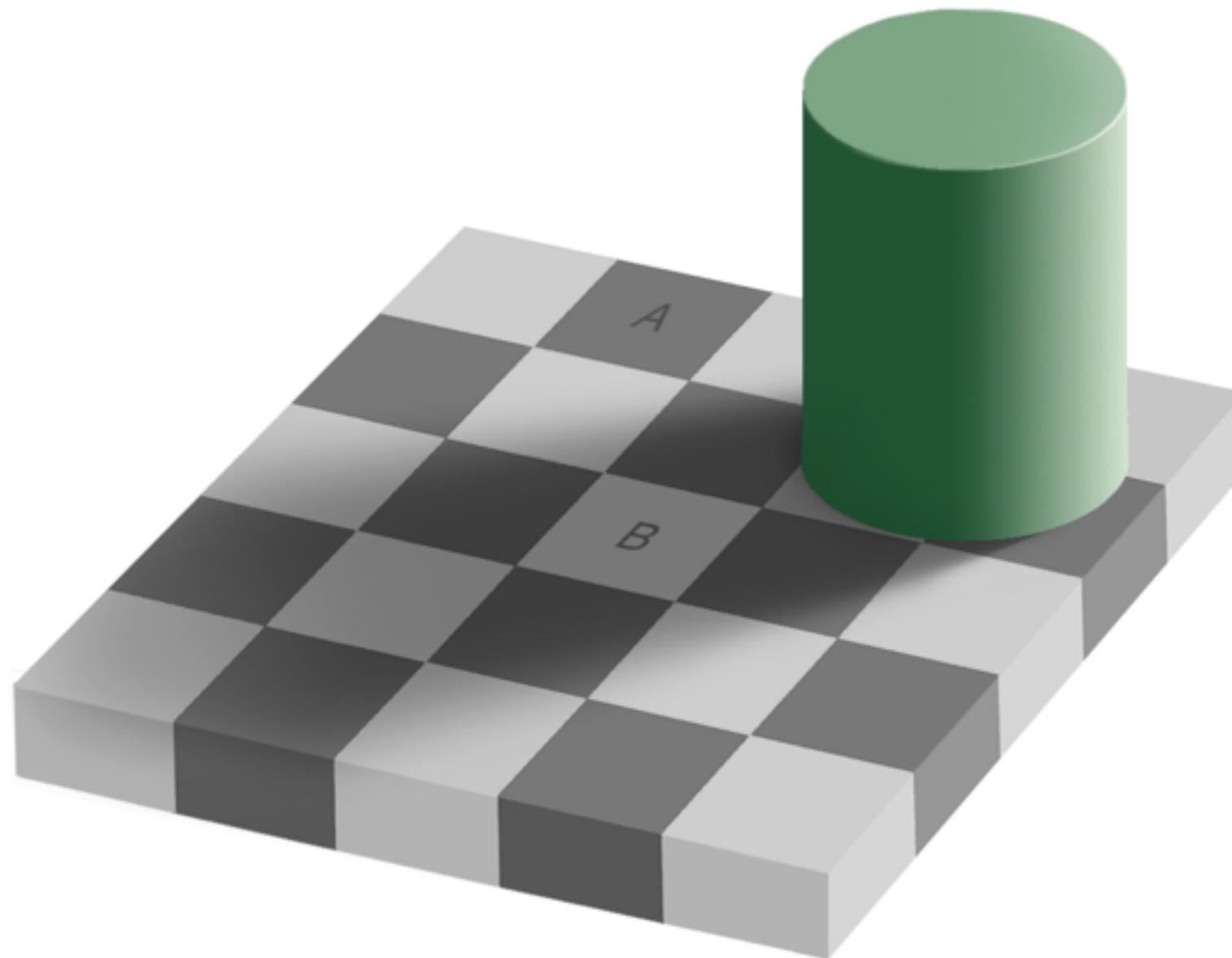


Locality & Orientation (illusion)



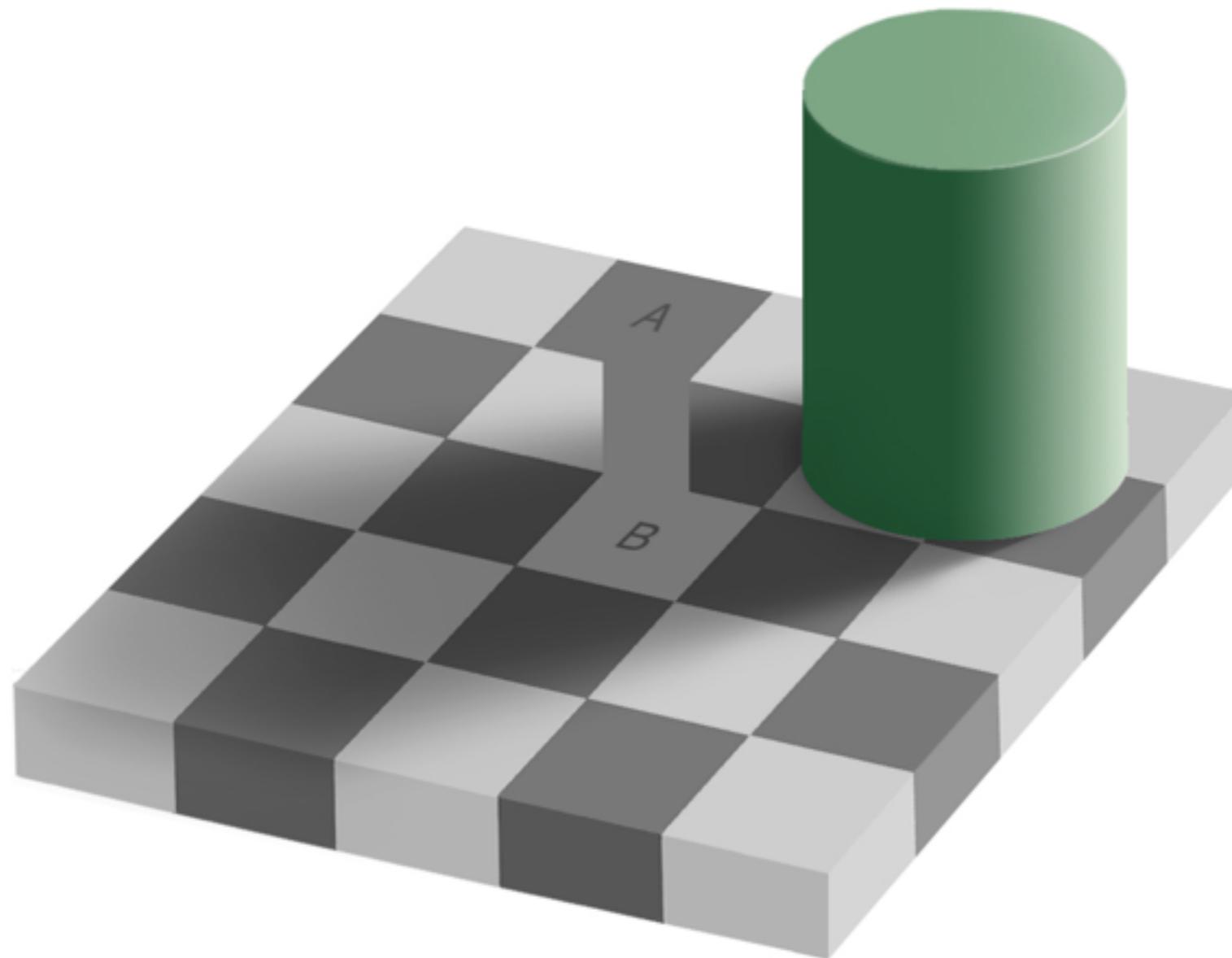


Locality & Intensity (illusion)





Locality & Intensity (illusion)





Locality & Motion – Aperture





Methodology

- Local models
 - Model based: Signal model: intrinsically 1D, band-limited, ergodic
 - Data driven: local model learned from data
- Camera models
 - pinhole model
 - multiple views will be used



Example: i1D-signals

- Local neighbourhoods f which contain edges or lines can be described by

$$f(\mathbf{x}) = g(\mathbf{x}^T \hat{\mathbf{n}})$$

- where g is a function of a single variable
- \mathbf{n} is a normal vector
- f is called *simple* signal or **intrinsically one-dimensional signal** (i1D)



Example: i1D-signals

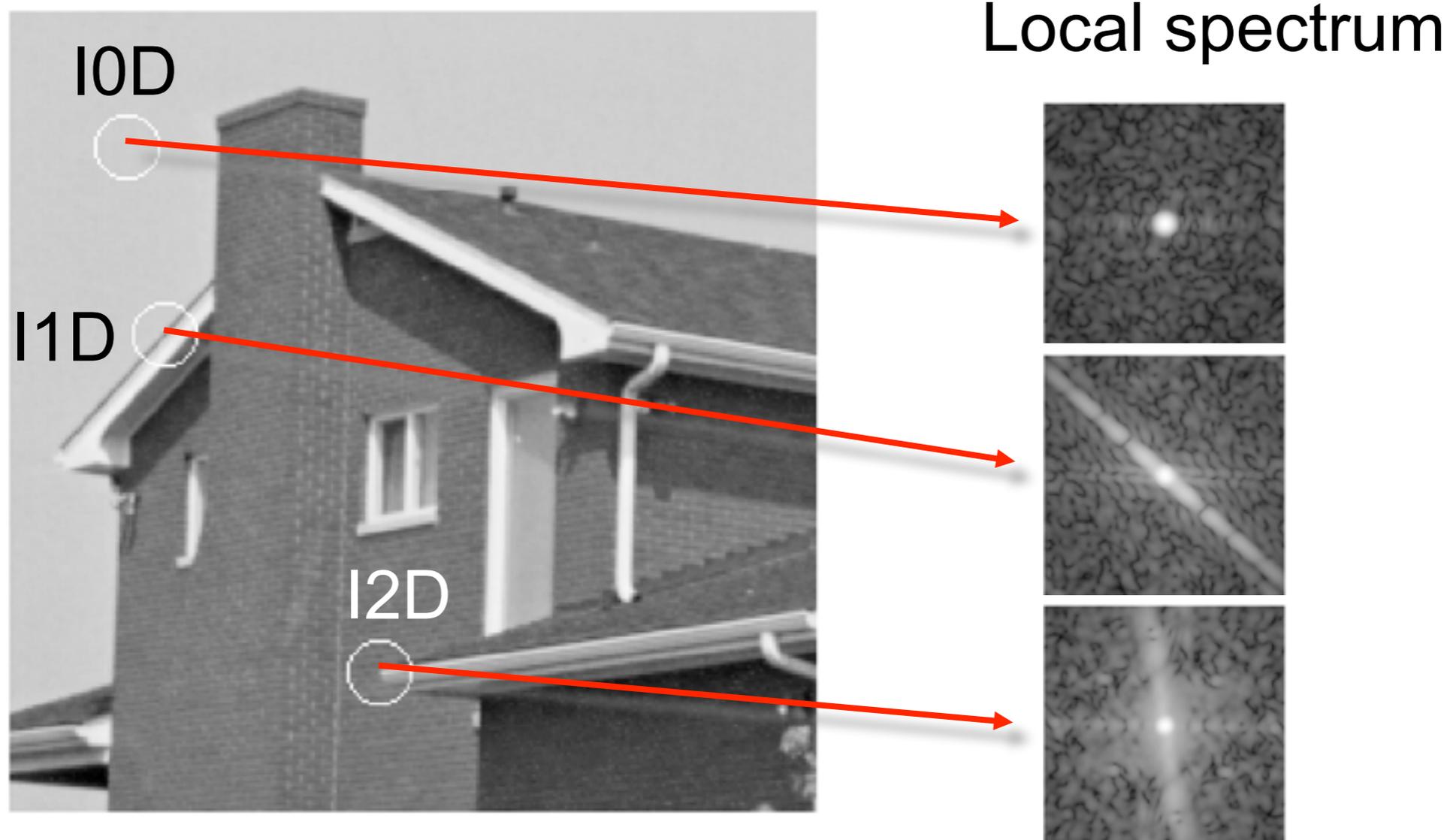
- For an i1D-signal we know that its *Fourier transform* reads

$$F(\mathbf{u}) = (2\pi)^{d-1} \delta_{\hat{\mathbf{n}}}^{\text{line}}(\mathbf{u}) G(\mathbf{u}^T \hat{\mathbf{n}})$$

- an impulse line which goes through the origin with orientation $\hat{\mathbf{n}}$ and which varies as G , the Fourier transform of g (**see EDUPACK**)
- d is the dimensionality of the signal

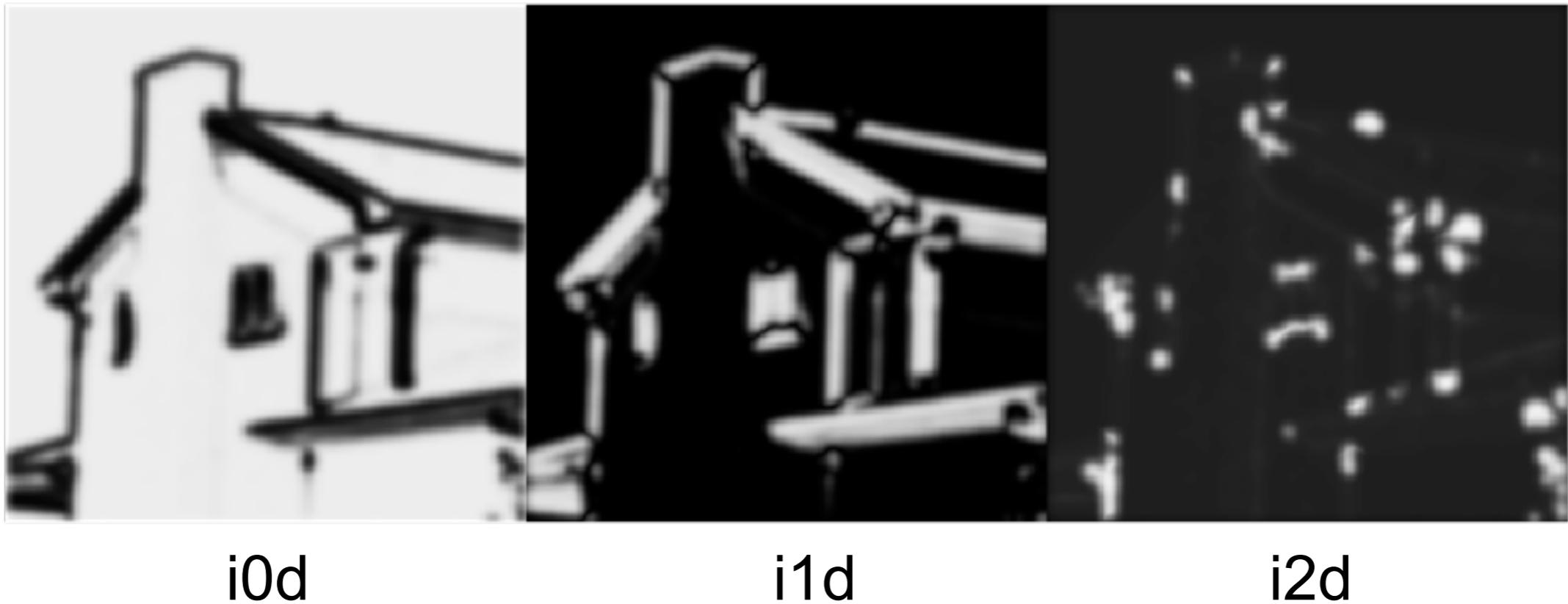


The Intrinsic Dimension



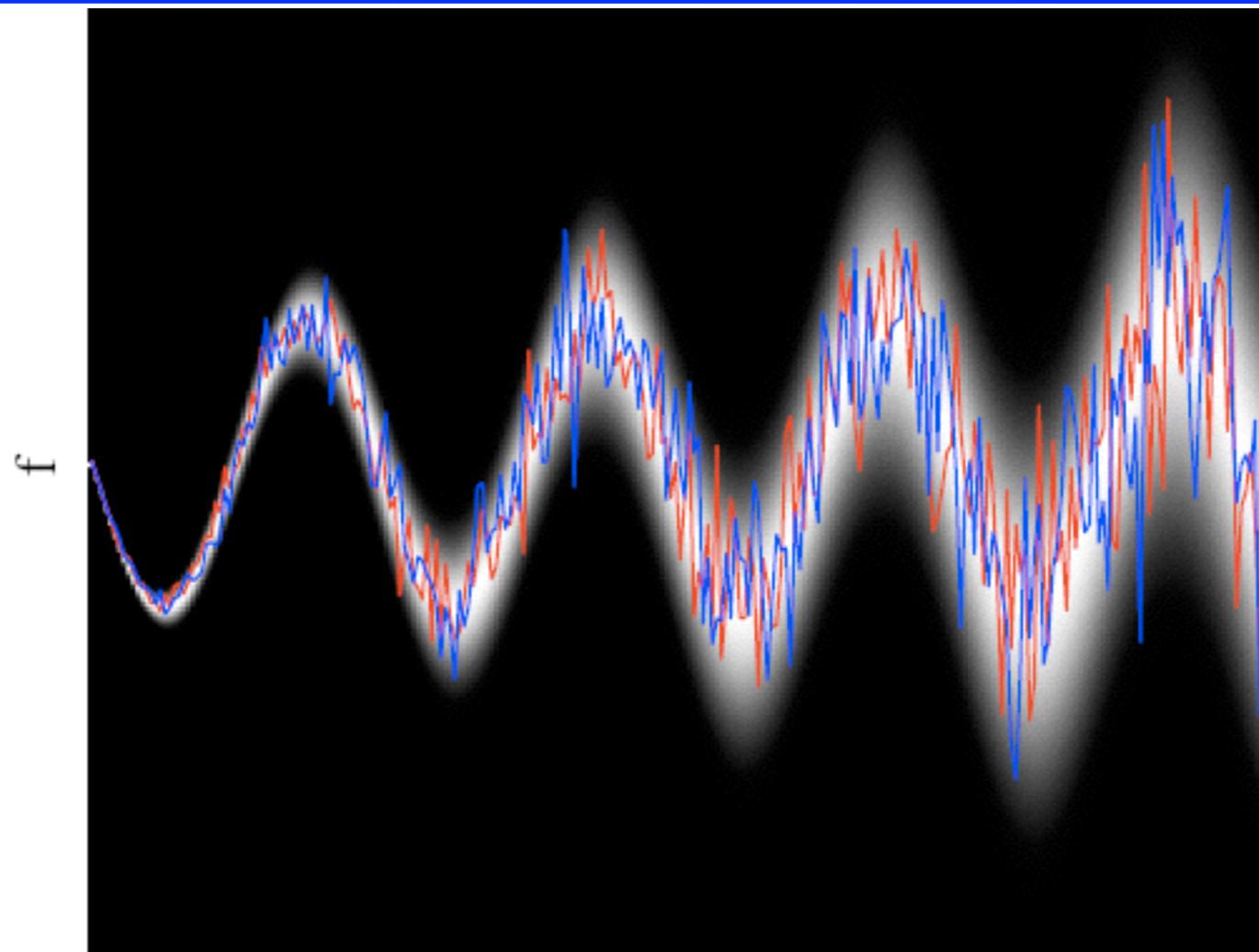


The Intrinsic Dimension





Ergodic Signal



- In the absence of video, we may assume that a local spatial average can approximate instance average.



Local Features

- Examples of local features:
 - Local orientation, motion, curvature, color, symmetry, phase, and their histograms.



Local Features

- Examples of local features:
 - Local orientation, motion, curvature, color, symmetry, phase, and their histograms.
- How local is local?
 - What locality is best depends on the signal (E.g. brick wall at coarse scale and fine scale)
 - For better generality one needs to try several choices of locality in parallel (LE2)



Example: Local Orientation

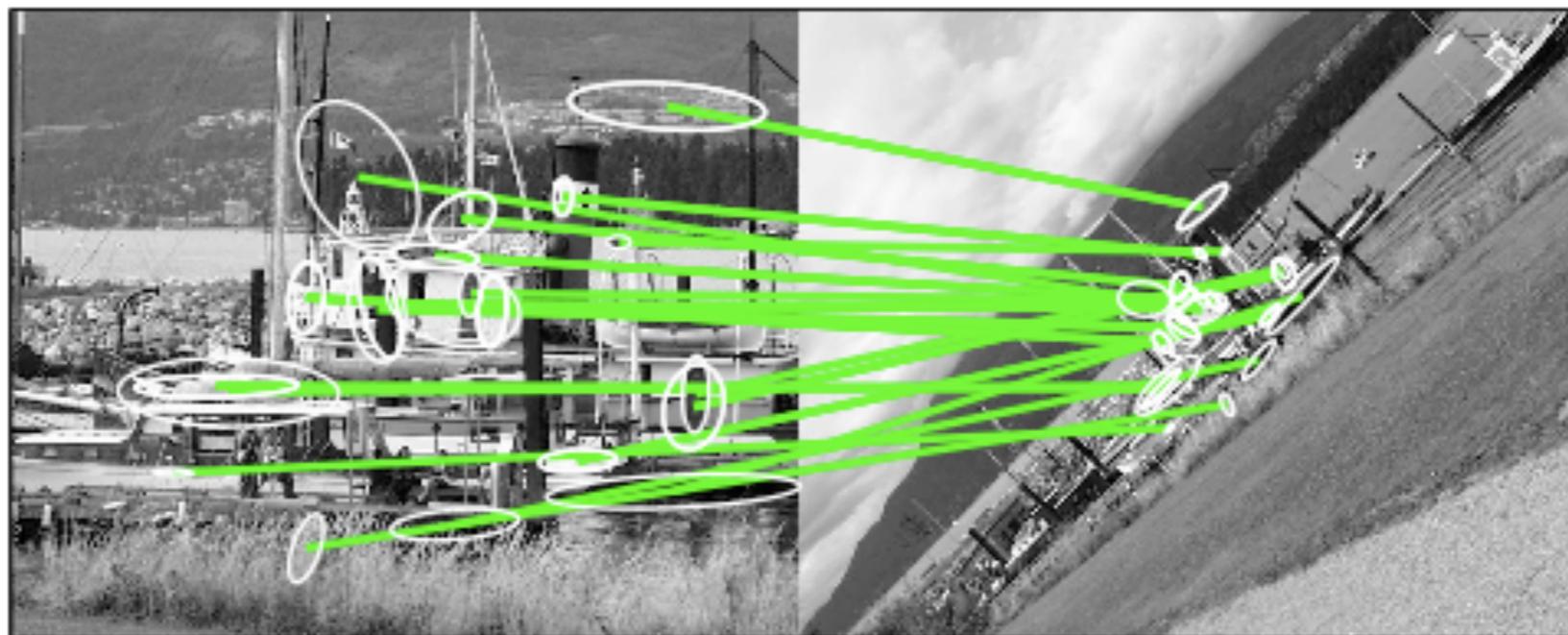
- A local signal which is 1D has a natural notion of orientation
- The signal is constant in all directions which are orthogonal to the vector \mathbf{n}
- Histogram is called HOG (histogram of oriented gradients)





Local Features

- Detector+descriptor paradigm



- Detect regions in a scale invariant manner
- Describe their contents [LE8]



Representations

- A **representation** is a mathematical description of a feature
- A feature can have one or several representations, for instance
 - Real or complex numbers (scalars)
 - Vectors
 - Matrices
- A collection of one or more features for a particular image object or local region is called a **descriptor**



Example: Local Orientation

- For the feature local orientation, the vector \mathbf{n} is a possible representation.
- BUT: for an i1D-signal, \mathbf{n} is not unique, also $-\mathbf{n}$ can be used!
- **Unique** representations are normally preferred





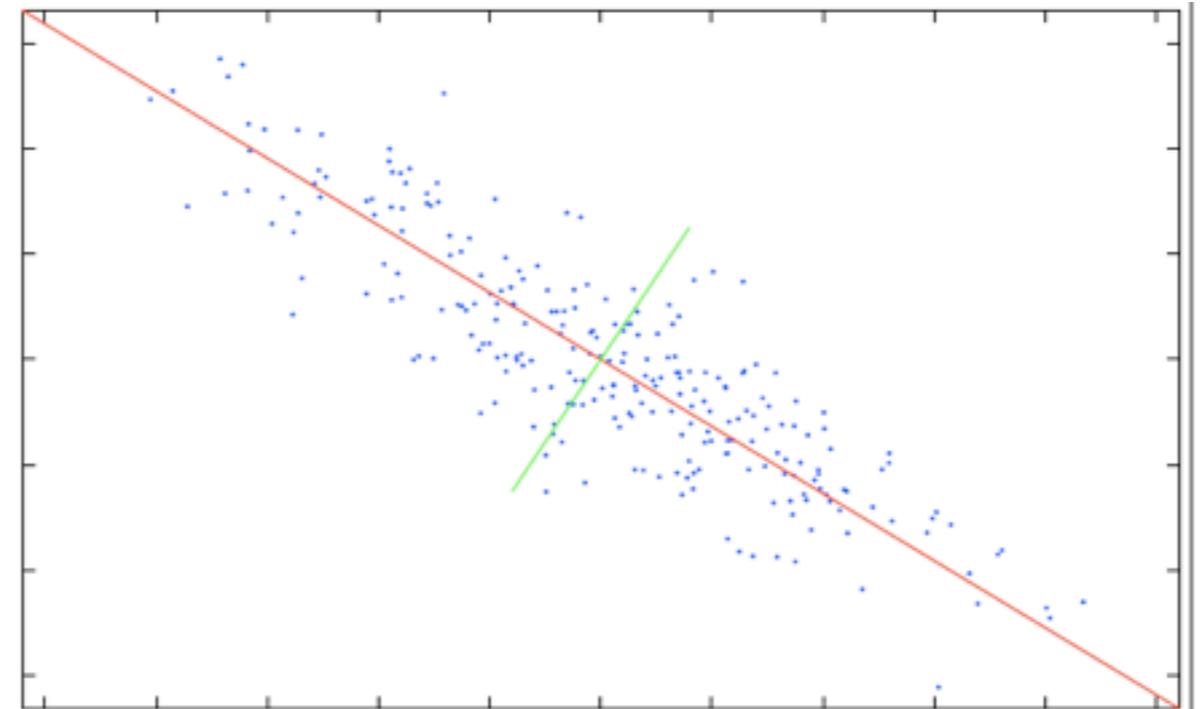
Example: Local Orientation

- In the 2D case, an angle in the interval 0° - 180° between the vector \mathbf{n} and the horizontal axis would be unique.
- BUT: if \mathbf{n} is approximately horizontal, very small changes of orientation result in the angle jumping between 0° and 180° .



Representations

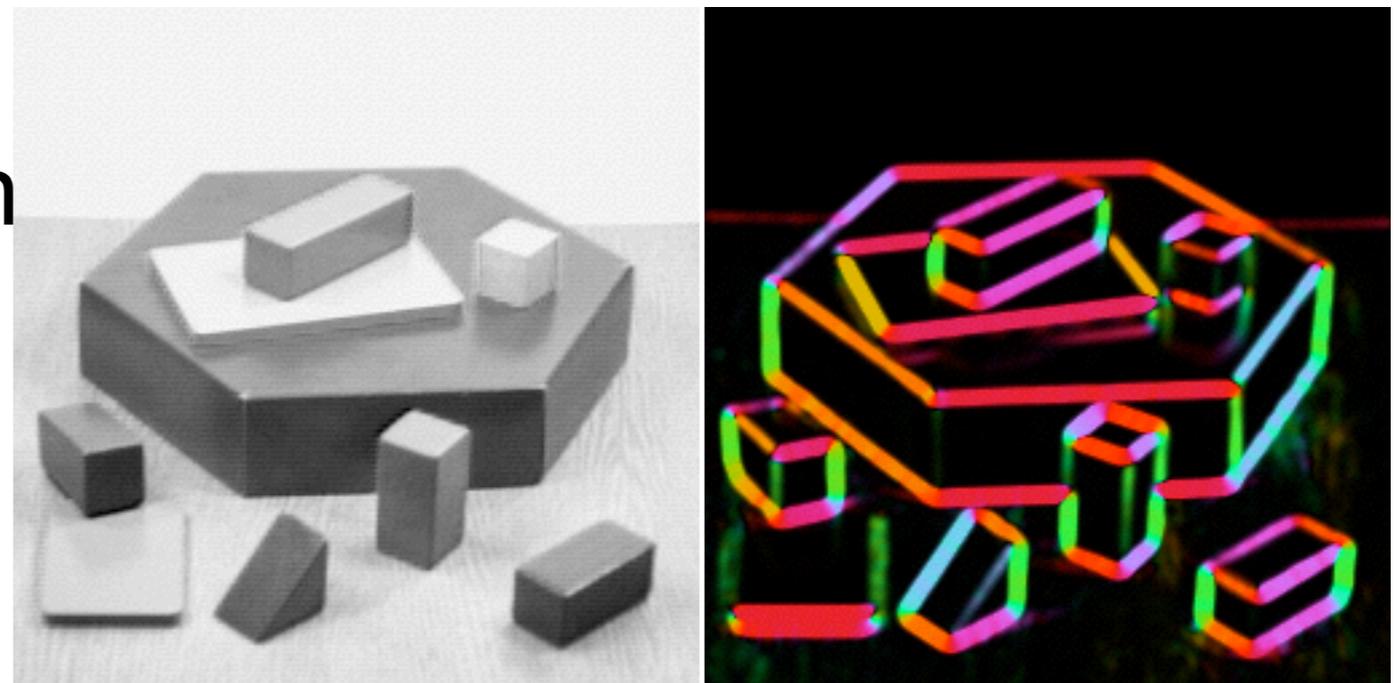
- The **confidence** of a signal is a measure of how well it fits the assumed model. It does not depend on the feature value
- E.g. for a fitted line, the inverse variance orthogonal to the line can be used
- The confidence can be used for weighted averaging (maximum likelihood estimation)





Example: Local Orientation

- For representation of orientation (i.e. the vector \mathbf{n}), we need a confidence measure that is related to the fitness of the corresponding local signal to the i1D model. Because some regions do not have a unique orientation.
- e.g. the local orientation variance can be used
- We will later introduce several such representations





Estimation

- For a given feature representation, we want to compute the numerical descriptor value for the local image region
- This step is called **estimation**
- The same representation can often be estimated in several ways, with different **algorithms**



Estimation

- Estimation examples:
 - **linear filtering** convolution in the spatial domain and multiplication in the Fourier domain (e.g. gradient)
 - **histogramming** (e.g. colour distributions, and HoG - gradient orientation histograms)
 - solving a **system of equations** (e.g. fundamental matrix estimation)
 - **gradient descent** (e.g. tracking of image regions)



Estimation

- Which algorithm to choose depends on
 - Available computational power (how many operations can be computed for each image)
 - Required accuracy (acceptable measurement error)
 - Required robustness (e.g. handling of outliers in the data)
 - Required dynamics of the feature space (estimation of very large and very small values)



Example: Local Orientation

- Convolve the image with two (or more) filters
- In each image point:
 - Combine the filter responses in a non-linear way to obtain the orientation estimate
 - Combine the filter responses in a non-linear way to obtain a confidence value
- During the course, we will introduce several estimation methods of this type. e.g. in the computer lesson on Thursday.



Summary

- You should now know about:
- **Course structure** and philosophy
- What **Computer Vision** is used for
- Methodology concepts: **local models**, **local features**, **representations** and **estimation**