Computer Vision II

Introduction - 16.04.2014

TECHNISCHE UNIVERSITÄT DARMSTADT

visual inference

Stefan Roth, 16.04.2014

Computer Vision II

TECHNISCHE UNIVERSITÄT DARMSTADT

- Lecturer:
  - Stefan Roth <sroth AT cs.tu-darmstadt...>
  - Office hours: Wednesdays, 13:45 - 14:45, S3|05, room 317
- Teaching Assistants:
  - Uwe Schmidt <uwe.schmidt AT gris.tu-...>
  - Tobias Plötz <tobias.ploetz AT gris.tu-...>
  - Office hours: TBA
- Course staff email: <cv2staff AT gris.tu-...>
  - Please use it for any questions, concerns etc.

Course Material

TECHNISCHE UNIVERSITÄT DARMSTADT

- Course web page: goo.gl/glQ8Kh
  - Contains pointers to readings
- Moodle ("HRZ Moodle"): goo.gl/l028kx
  - Please make sure you are signed up!
  - Contains slides and homework assignments
- Mailing list (see webpage):
  - We will sign you up, if you leave/left us your email during the first class.
  - Otherwise, please sign up yourself!

Forum

TECHNISCHE UNIVERSITÄT DARMSTADT

- There is a forum set up at the Fachschaft's website:
  http://www.fachschaft.informatik.tu-darmstadt.de/forum/viewforum.php?f=311
  - Please use it to ask questions of public interest.
  - You are encouraged to discuss with each other.
  - However: Please do not share solutions or give strong hints about the solutions to the homework problems.

Course language

TECHNISCHE UNIVERSITÄT DARMSTADT

- will be English.
  - This applies to lectures, exercises, announcements, etc.
- Why?
  - Essentially all computer vision publications and books are written in English.
  - Knowing the original terms is crucial.
- If strongly preferred, you may contact the course staff in German.
  - English is encouraged though, because we may use your (anonymized) question to clarify points to the entire class.

Organization

TECHNISCHE UNIVERSITÄT DARMSTADT

- Class type: IV4
- Lecture:
  - Wednesdays, 9:50 - 11:30, S3|05, room 074
  - We will cover the foundational aspects of each topic.
- Exercises, homework, etc.:
  - Wednesdays, 11:40 - 13:20, S3|05, room 074
  - irregularly
  - We will cover some practical aspects, and discuss the homework assignments.
  - Also: Time to work on your homework assignments.

Exam & Exercises

TECHNISCHE UNIVERSITÄT DARMSTADT

- Exam:
  - Written (possibly oral) exam at the end of the semester.
  - Can be taken in English or German.
  - Exam date will be set and announced soon.
- Exercises:
  - Regular homework assignments; exercises will be graded.
  - Excercise points are part of your final grade (vorlesungsbleitende Prüfung)!
- Grading: 2/3 exam, 1/3 exercise

Homework assignments

TECHNISCHE UNIVERSITÄT DARMSTADT

- Mostly programming assignments.
  - MATLAB, standard environment for scientific computing.
- Sometimes smaller pen and paper exercises
- The homework exercises are crucial for actually "digesting" the material from the lecture.
  - Hence your points count...
- Time commitment:
  - Solving the homework problems will require a substantial time commitment.
  - Keep in mind that 6CP equal a workload of approximately 180h / semester.

Collaboration policy

TECHNISCHE UNIVERSITÄT DARMSTADT

- I do not tolerate plagiarism in any way!
- You may (and are encouraged to) discuss general class topics.
- But each handed in homework solution must be your own!
  - Any sources you used (other than provided by us) must be cited.
  - Details: TBA on the first homework assignment sheet.

- Questions? Problem cases?
  - Talk to us.

Readings

TECHNISCHE UNIVERSITÄT DARMSTADT

- Good news: Additional book
  - Computer Vision: Models, Learning, and Inference by Simon Prince (Cambridge University Press, 2012)
  - Even better news: PDF online - http://www.computervisionmodels.com/
- Second book:
  - Computer Vision: Algorithms and Applications by Richard Szeliski (Springer, 2011)
  - Also good news: PDF available online - http://szeliski.org/Book
- Additional readings:
  - Papers and tutorials (on the web or course page)

How does it fit into your course plan?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Elective:
  - Part of Human Computer Systems (HCS) track.
- Related classes:
  - Human Computer Systems: prerequisite
  - Computer Vision I (Roth, WS): prerequisite
  - Maschinelles Lernen - Statistische Verfahren I (Peters, SS)
  - Maschinelles Lernen - Statistische Verfahren II (Roth, WS)
  - Bildverarbeitung (Sakas, SS)
- Theses and projects:
  - Topics in computer vision and machine learning.

Goal of today's lecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- Recap:
  - What is computer vision?
  - Why is it useful?
- CV2 - Probabilistic approaches to dense vision
  - Beyond sparse points
  - Beyond local processing
  - Focus on probabilistic approaches
- Tutorial example:
  - Stereo / disparity estimation
  - Optical flow

What is computer vision?

TECHNISCHE UNIVERSITÄT DARMSTADT

[from Steve Seitz]

What does it mean to "see"?

TECHNISCHE UNIVERSITÄT DARMSTADT

see [1] |sē|

verb (**sees** |sēz|, **seeing** |sē-i ng |; past **saw** |sô|; past part. **seen** |sēn|)

[trans.]
1 perceive with the eyes; discern visually *: in the distance she could see the blue sea* |
[intrans.] *Andrew couldn't see out of his left eye* figurative *I can't* ***see into*** *the future.*
• [with clause] be or become aware of something from observation or from a written or other visual source *: I see from your appraisal report that you have asked for training.*

...

[Oxford English dictionary, slide from Michael Black]

What does it mean to perceive"?

TECHNISCHE UNIVERSITÄT DARMSTADT

perceive |pərˈsēv|

verb [trans.]
1 become aware or conscious of (something); come to realize or understand *: his mouth fell open as he perceived the truth* | [with clause] *he was quick to perceive that there was little future in such arguments.*
• become aware of (something) by the use of one of the senses, esp. that of sight *: he perceived the faintest of flushes creeping up her neck.*
2 interpret or look on (someone or something) in a particular way; regard as *: if Guy does not* ***perceive*** *himself* ***as*** *disabled, nobody else should* | [trans.] *some geographers perceive hydrology to be a separate field of scientific inquiry.*

[Oxford English dictionary, slide from Michael Black]

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- What is computer vision (or machine vision)?
  - Developing computational models and algorithms to interpret digital images / understand the visual world we live in.

[from Steve Seitz]

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- What is computer vision (or machine vision)?
  - Developing computational models and algorithms to interpret digital images / understand the visual world we live in.
- Is that important?
- What can we (already) do with it?

Face detection

e.g. Canon [powershot.com], etc.

Human pose estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

Microsoft Kinect

[www.xbox.com/kinect]

Earth viewers

TECHNISCHE UNIVERSITÄT DARMSTADT

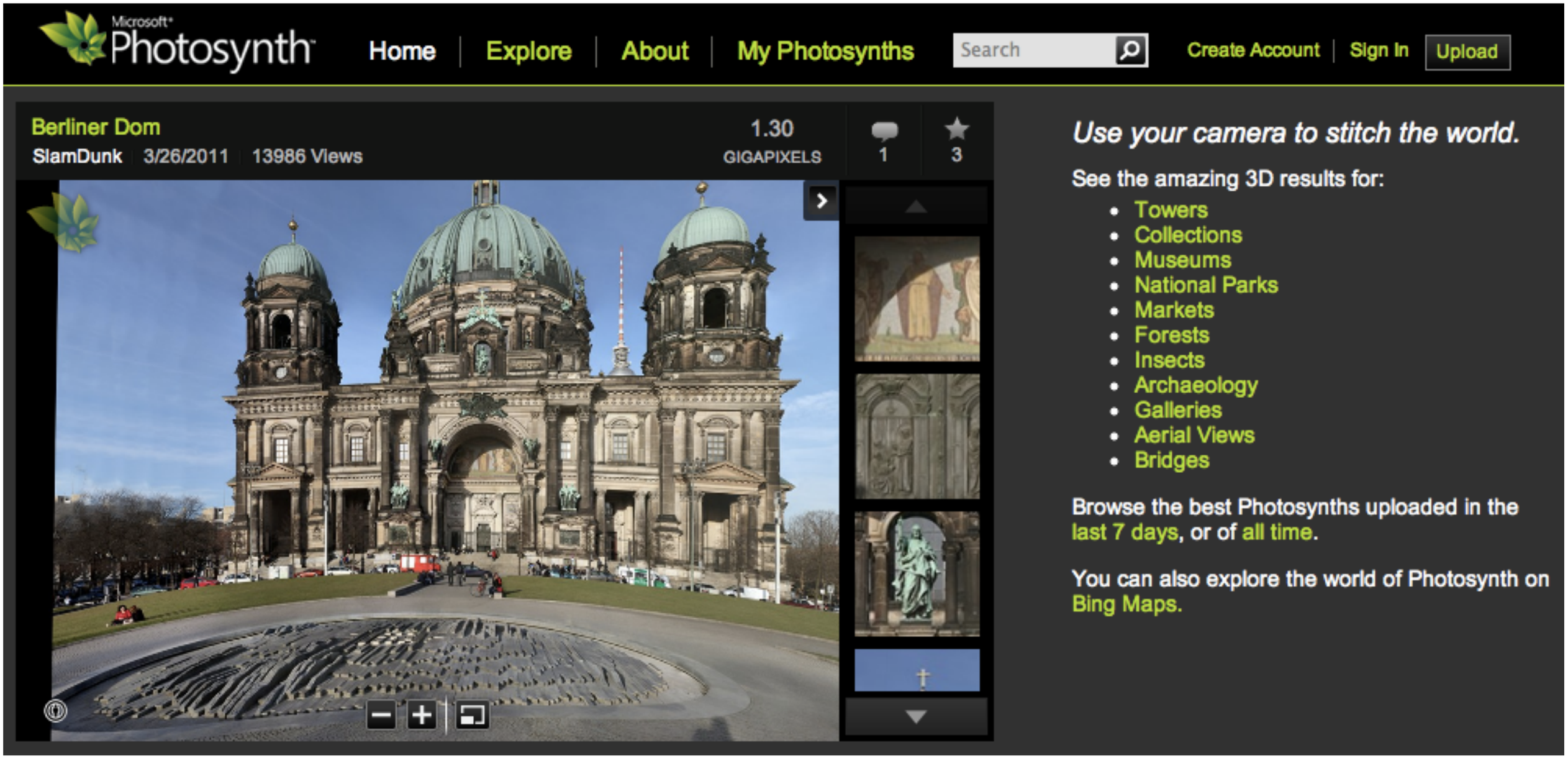
Google Street View

[www.google.com]

[www.mobile3dcity.com]

Photosynth

TECHNISCHE UNIVERSITÄT DARMSTADT

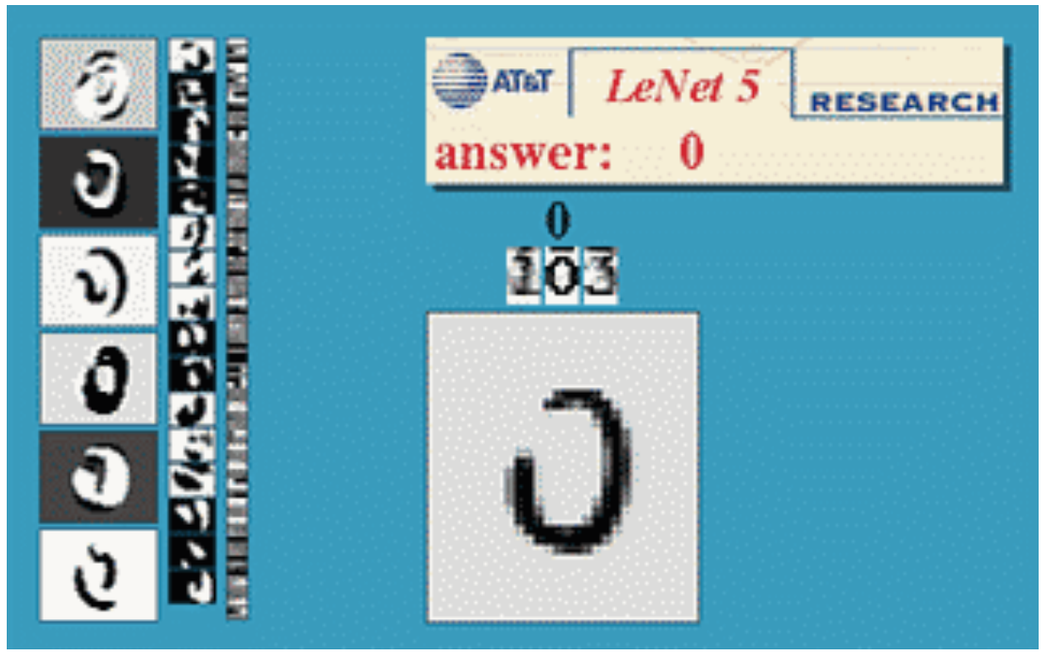


[photosynth.net]

Optical character recognition (OCR)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Convert scanned documents to text



Digit recognition, AT&T labs
[www.research.att.com/~yann/]

License plate readers
[en.wikipedia.org/wiki/Automatic_number_plate_recognition]

[from Steve Seitz]

Traffic sign recognition

TECHNISCHE UNIVERSITÄT DARMSTADT

Opel

VDO

Special effects: Shape capture

TECHNISCHE UNIVERSITÄT DARMSTADT

The Matrix movies, ...

[from Steve Seitz]

Special effects: Motion capture

TECHNISCHE UNIVERSITÄT DARMSTADT

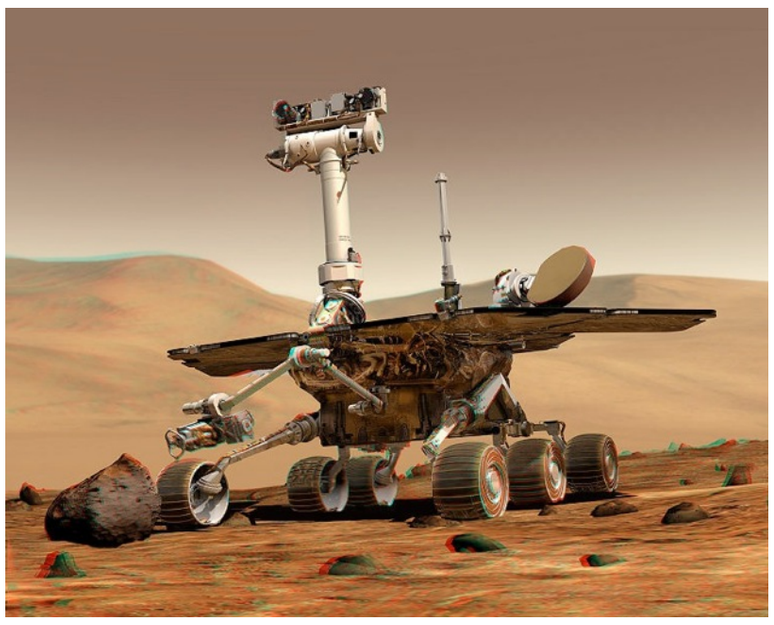
Pirates of the Carribean, Industrial Light and Magic [from Steve Seitz]

Sports

3D soccer analysis

[www.kicker.de]

Vision in space



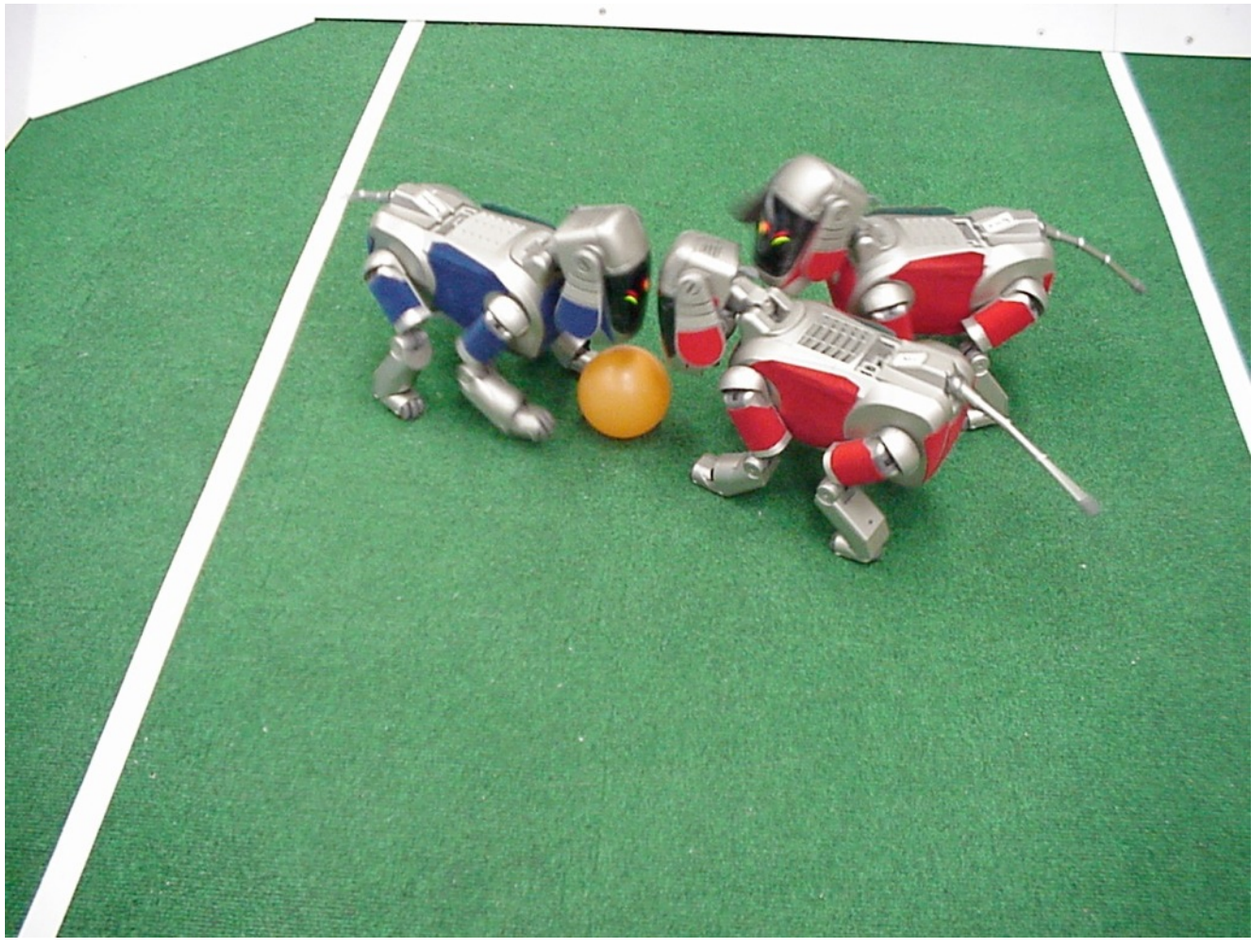
NASA's mars exploration rover Spirit

[en.wikipedia.org/wiki/Spirit_rover]

[from Steve Seitz]

Robotics

TECHNISCHE UNIVERSITÄT DARMSTADT



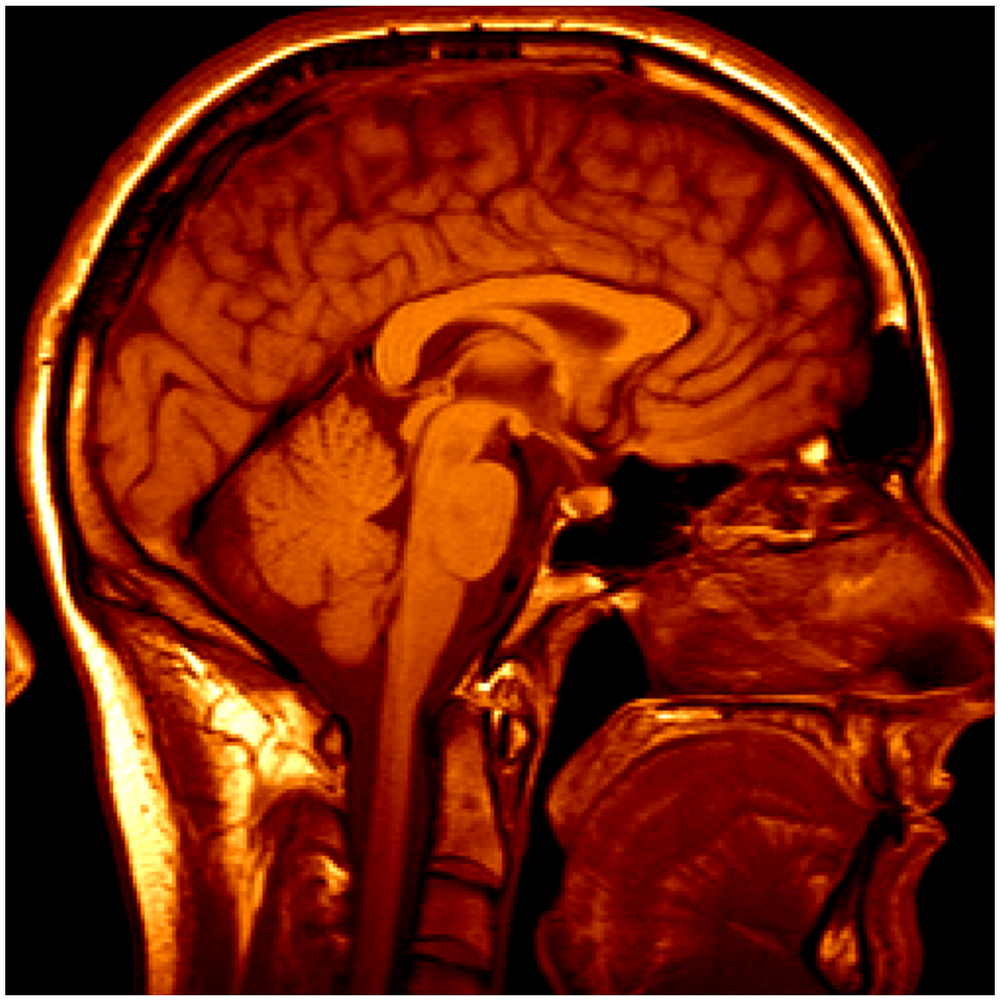
Robocup

[www.robocup.org]

Darmstadt Dribblers [FG SIM]

Medical imaging

TECHNISCHE UNIVERSITÄT DARMSTADT



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

[from Steve Seitz]

Computer Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

- We need a formal model that describes our problem *as well as* an algorithm that realizes (i.e. implements) it.
  - Neither the model alone, nor the algorithm alone suffices (in the long run).
  - Both mathematical and computational.

- What properties / cues of the visual world can we exploit or measure?
- What general (prior) knowledge of the world (not necessarily visual) should we exploit?

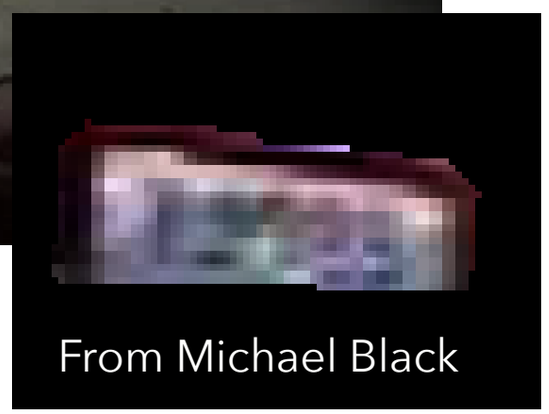
[adapted from Michael Black]

Ambiguity of Data

TECHNISCHE UNIVERSITÄT DARMSTADT

- Our image data is not only too little to fully recover and understand the “state of the visible world”.
- It may even be of poor quality:
  - Low resolution
  - (Sensor) noise
  - Etc.


Courtesy Criterion Collection


From Michael Black

- Our image data is always ambiguous.

Computer Vision

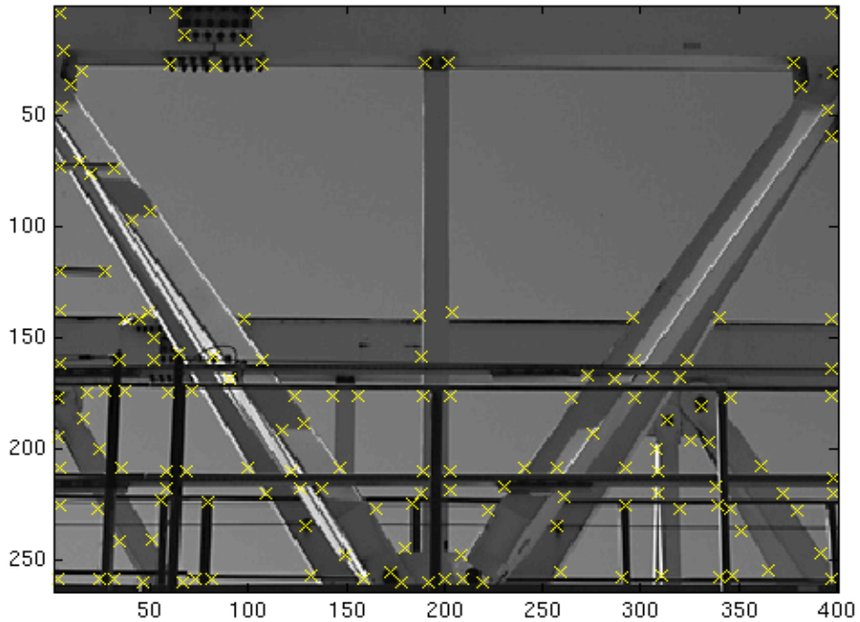
TECHNISCHE UNIVERSITÄT DARMSTADT

- We want to devise computer algorithms to understand the visual world much like we do.
- This means:
  - We have to use a lot of different cues...
  - We have to deal with ambiguity...
  - We have to exploit what is a plausible and meaningful interpretation...
  - ... in order to extract information about the 3D visual world from a small amount of data.
- CV is an inverse problem.

What have we learned in CV1?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Fundamentals of cameras and digital imaging
- Filtering and pyramid representations
- Template-based recognition
  - PCA, eigenfaces...
- Object recognition with histograms
  - ... with sparse points (BoW)
- Motion estimation
- Epipolar geometry
- Disparity estimation
- Segmentation



focus on sparse points or local methods

From CV1 to CV2

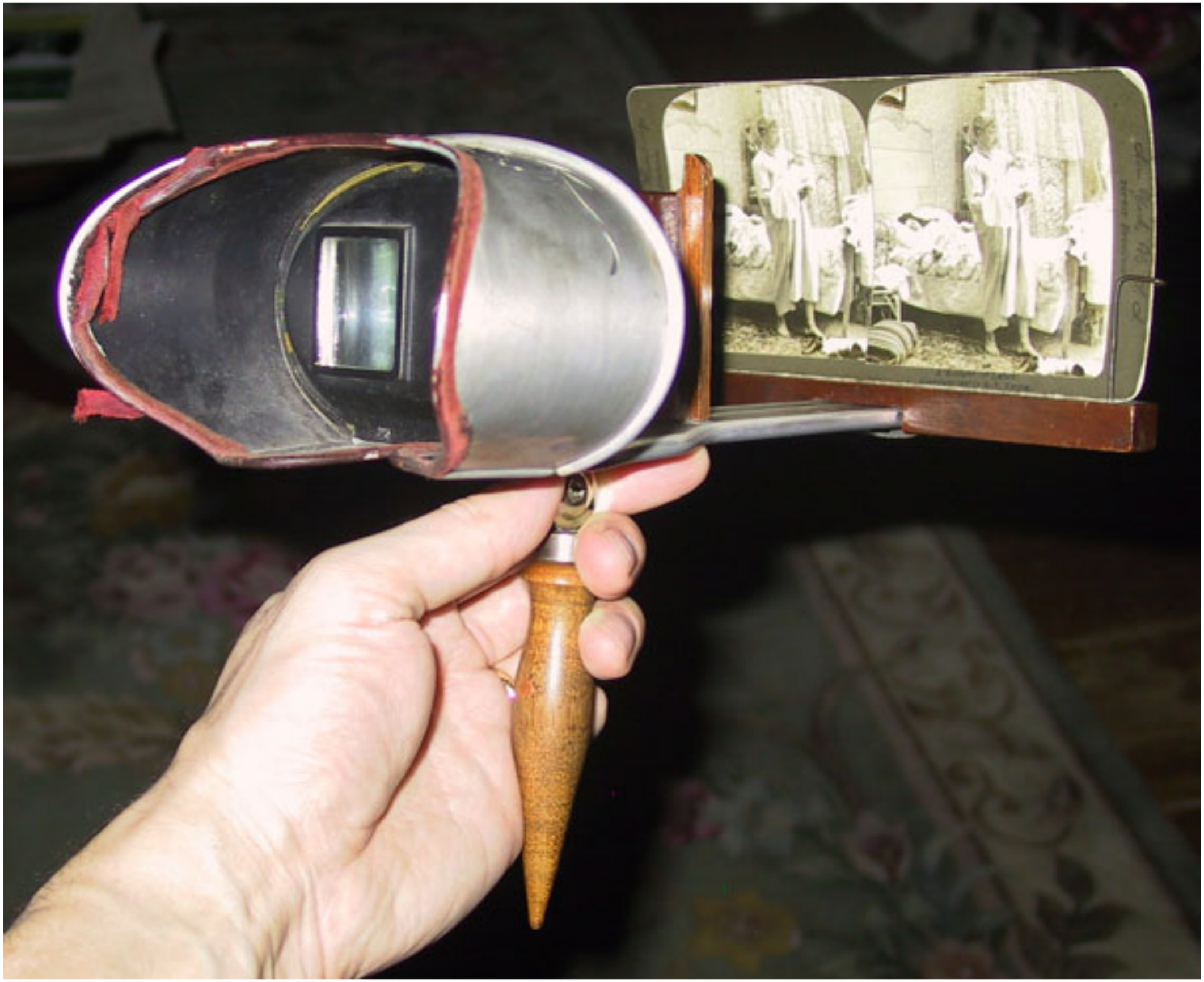
TECHNISCHE UNIVERSITÄT DARMSTADT

- Moving from
  - sparse to **dense**
  - local to **global**
- But why?
  - Haven't we solved these problems already?
- Two case studies:
  - Stereo
  - Optical flow

Stereo

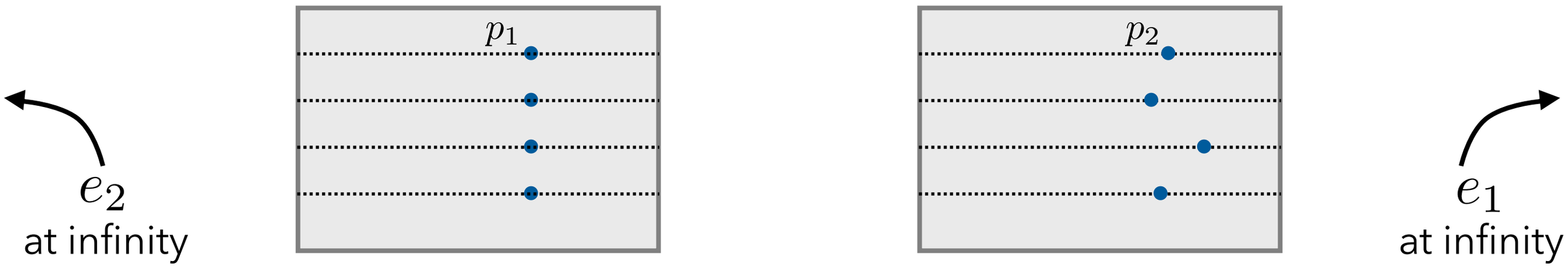
TECHNISCHE UNIVERSITÄT DARMSTADT

[Black]



[Black]

Binocular Stereo

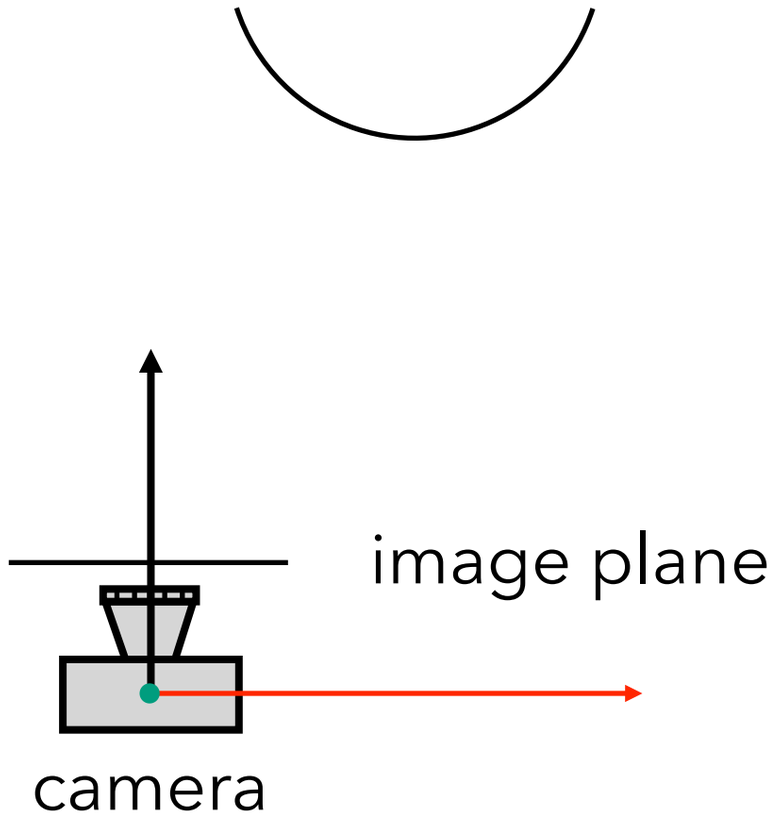


- Special case of epipolar geometry for stereo cameras with a standard binocular setup:
  - Epipoles are at infinity.
  - Epipolar lines are parallel.
  - Points correspond along the scanlines of the image.

Binocular Stereo

TECHNISCHE UNIVERSITÄT DARMSTADT

Left



Binocular Stereo

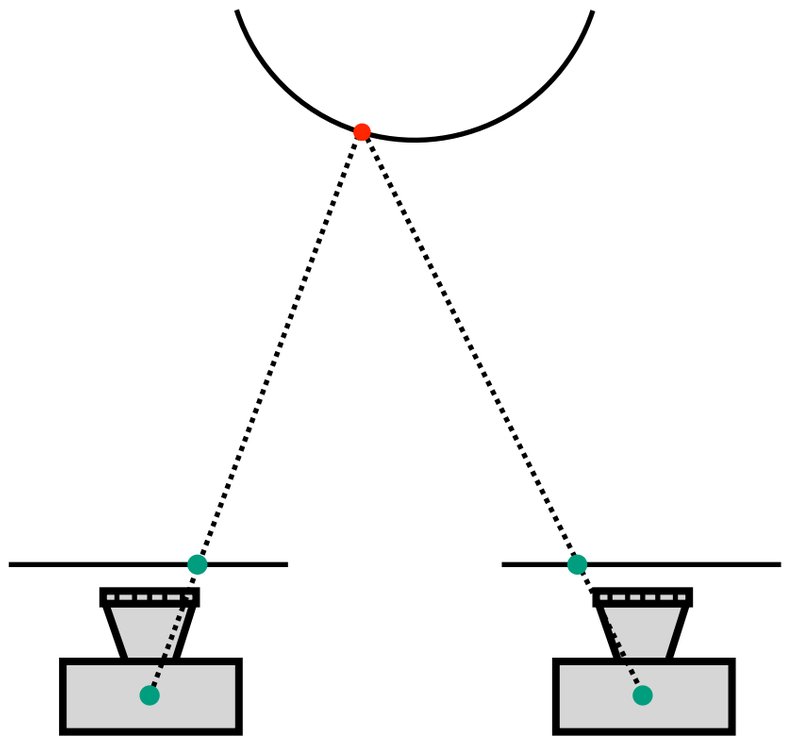
TECHNISCHE UNIVERSITÄT DARMSTADT

Left

Right

Binocular Stereo

TECHNISCHE UNIVERSITÄT DARMSTADT

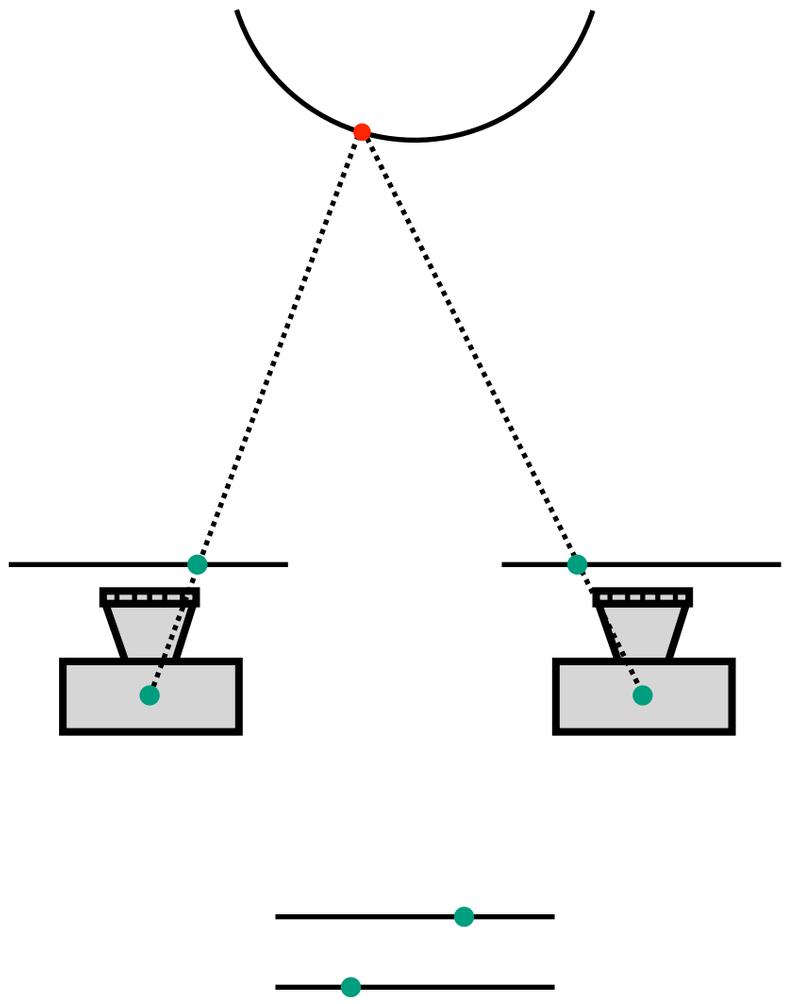


Left

Right

Binocular Stereo

TECHNISCHE UNIVERSITÄT DARMSTADT

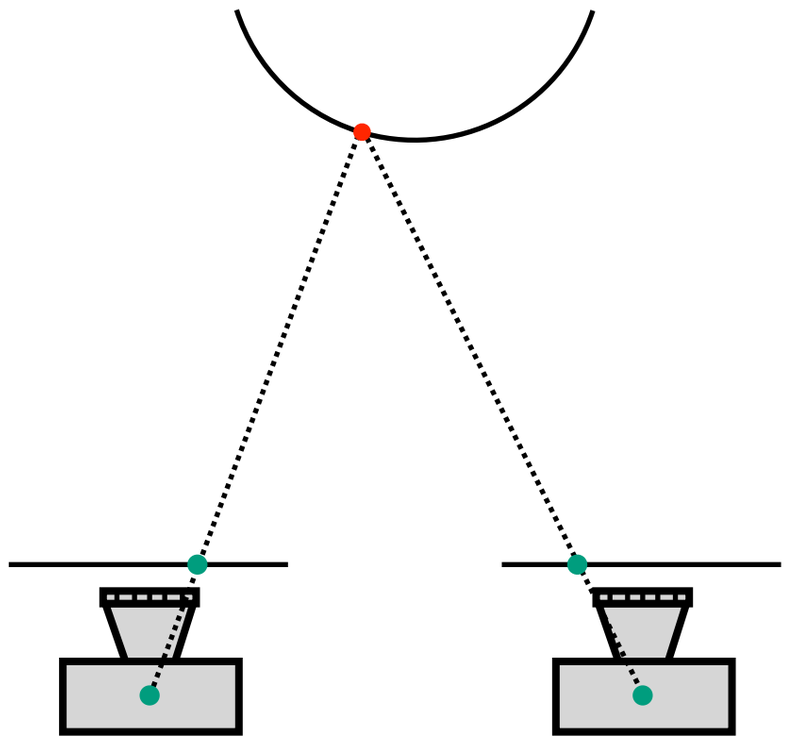


Left

Right

Binocular Stereo

TECHNISCHE UNIVERSITÄT DARMSTADT

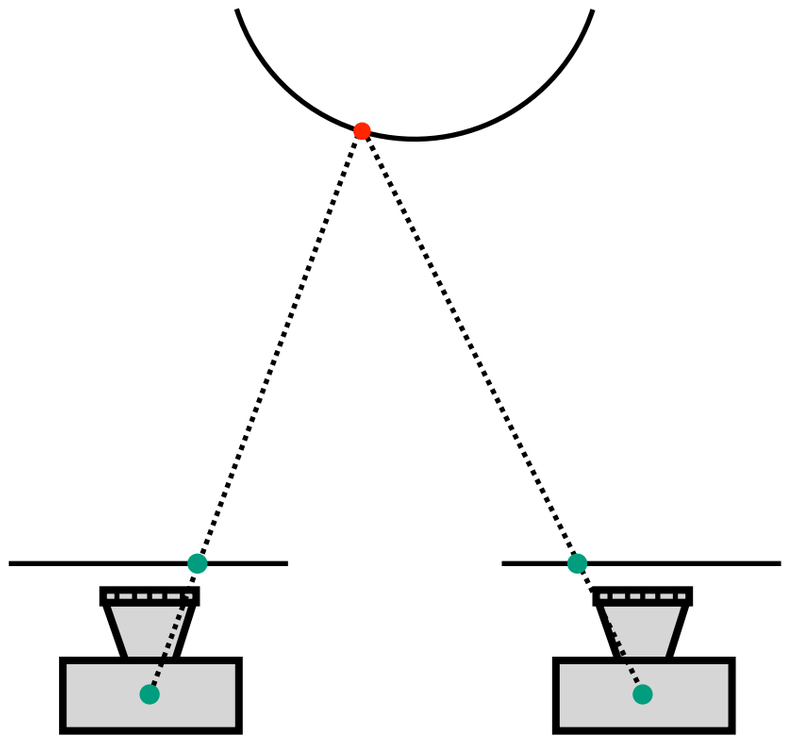


Left

Right

binocular disparity

Binocular Stereo



From known geometry of the cameras and estimated disparity, recover depth in the scene

Left

Right

binocular disparity

Triangulation

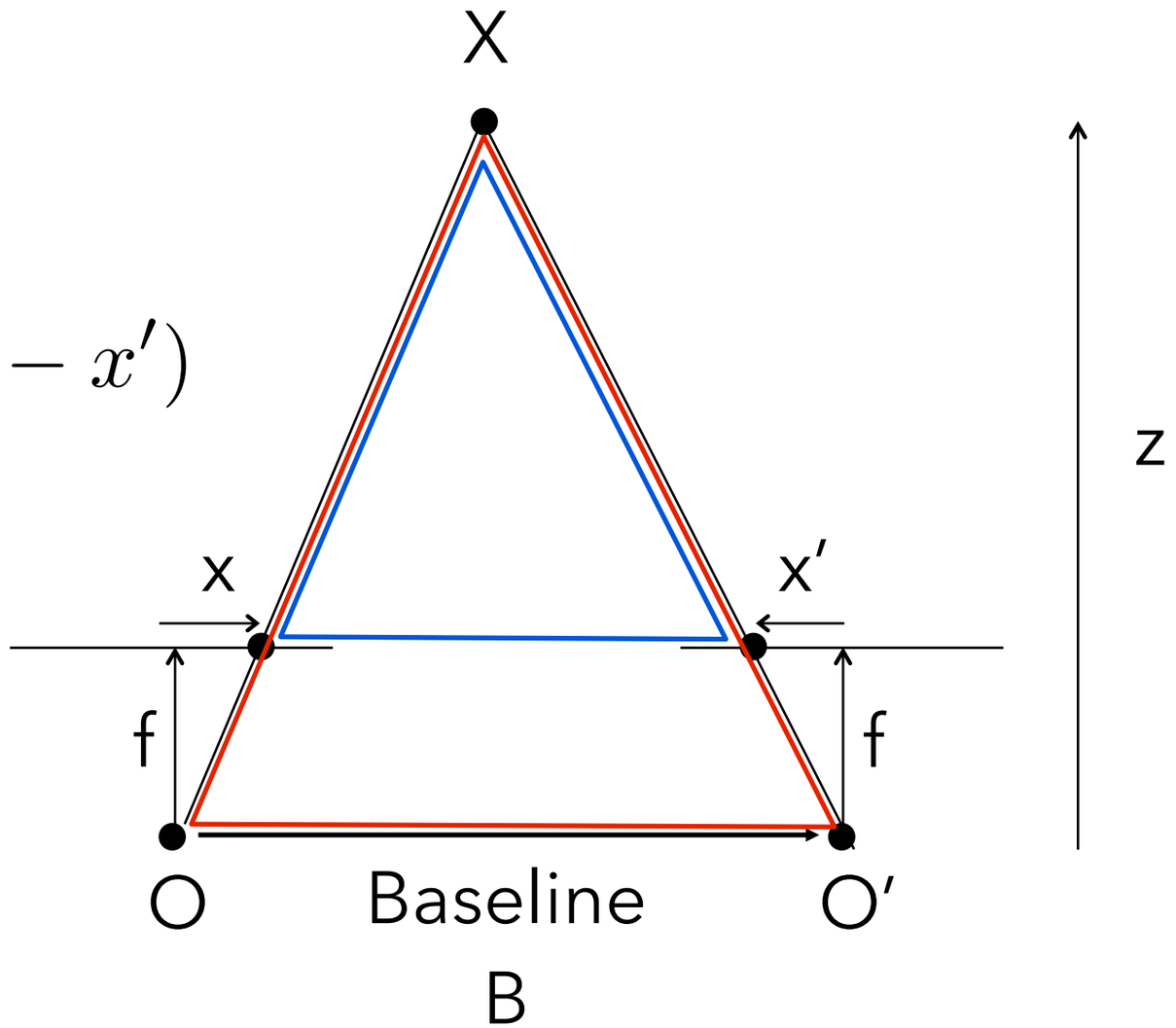
$$\frac{Z-f}{B-(x-x')} = \frac{Z}{B}$$

$$B \cdot Z - B \cdot f = B \cdot Z - Z \cdot (x-x')$$

$$Z = \frac{B \cdot f}{x-x'}$$

$$d = x - x' = \frac{B \cdot f}{Z}$$

disparity



- Parallel image planes: disparity from equal triangles

Binocular Disparity

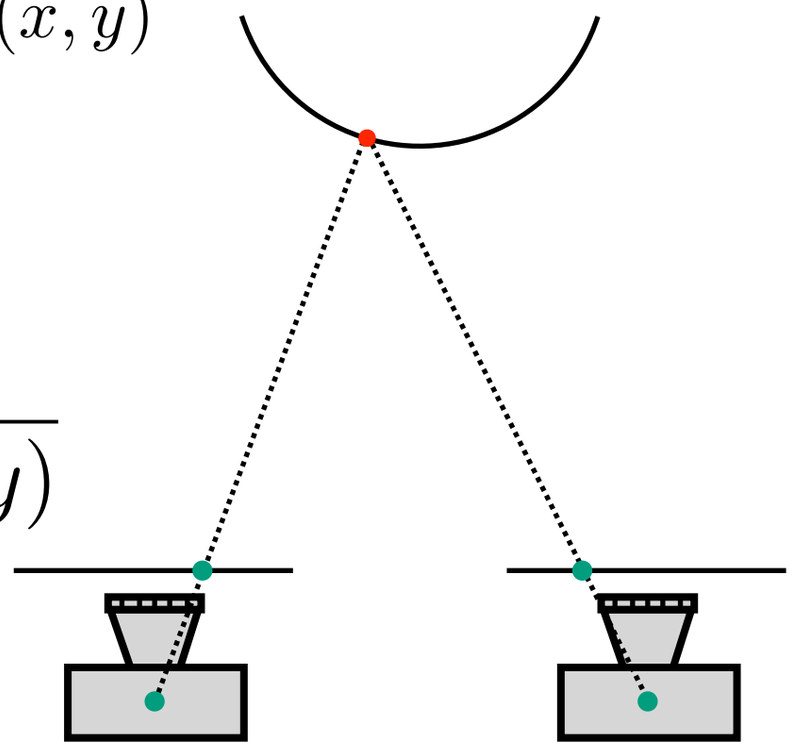
TECHNISCHE UNIVERSITÄT DARMSTADT

$Z(x,y)$ is depth at pixel (x,y)

$d(x,y)$ is disparity

Estimate:

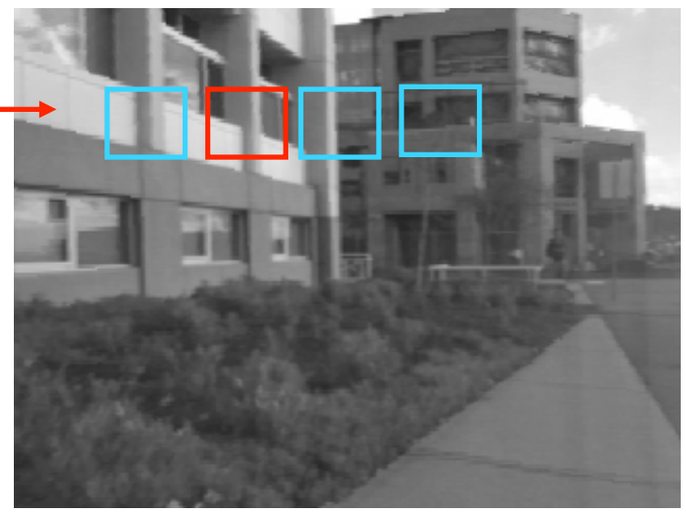
$$Z(x,y) = \frac{fB}{d(x,y)}$$



Left

Search for best match

Right

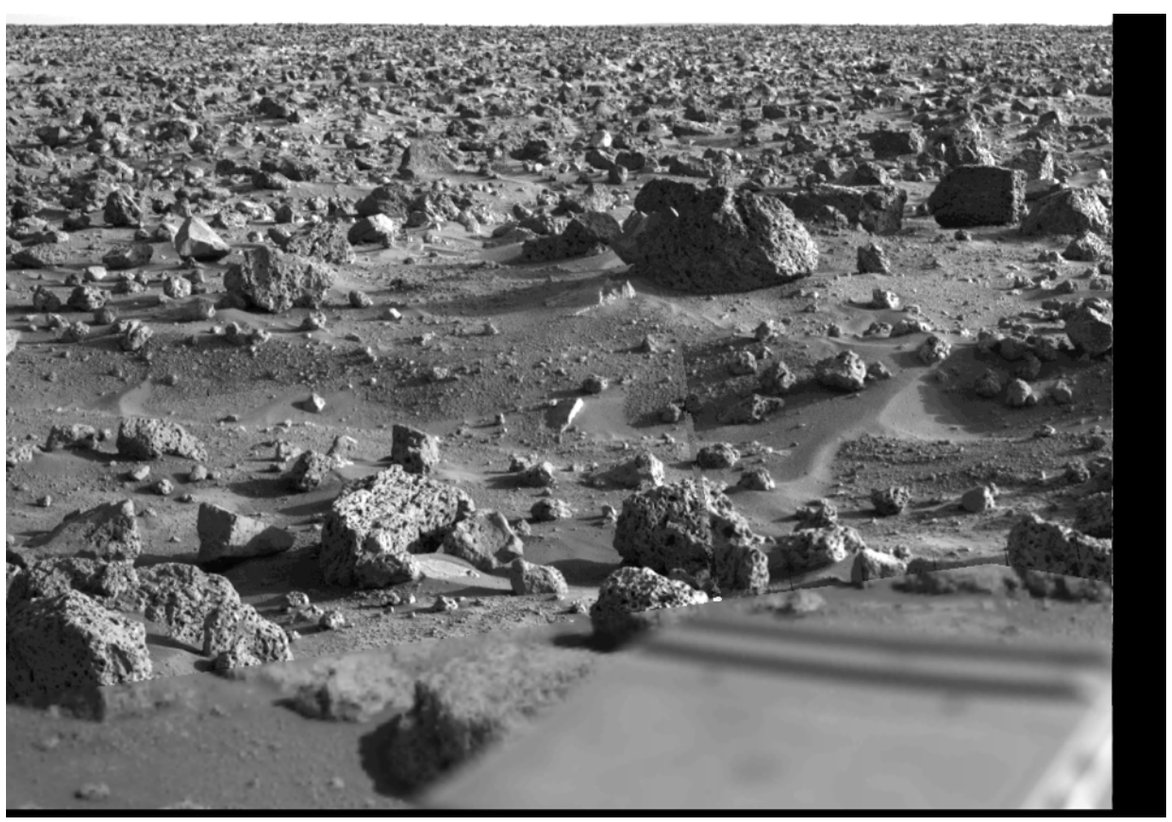


[Black]

Stereo Correspondence

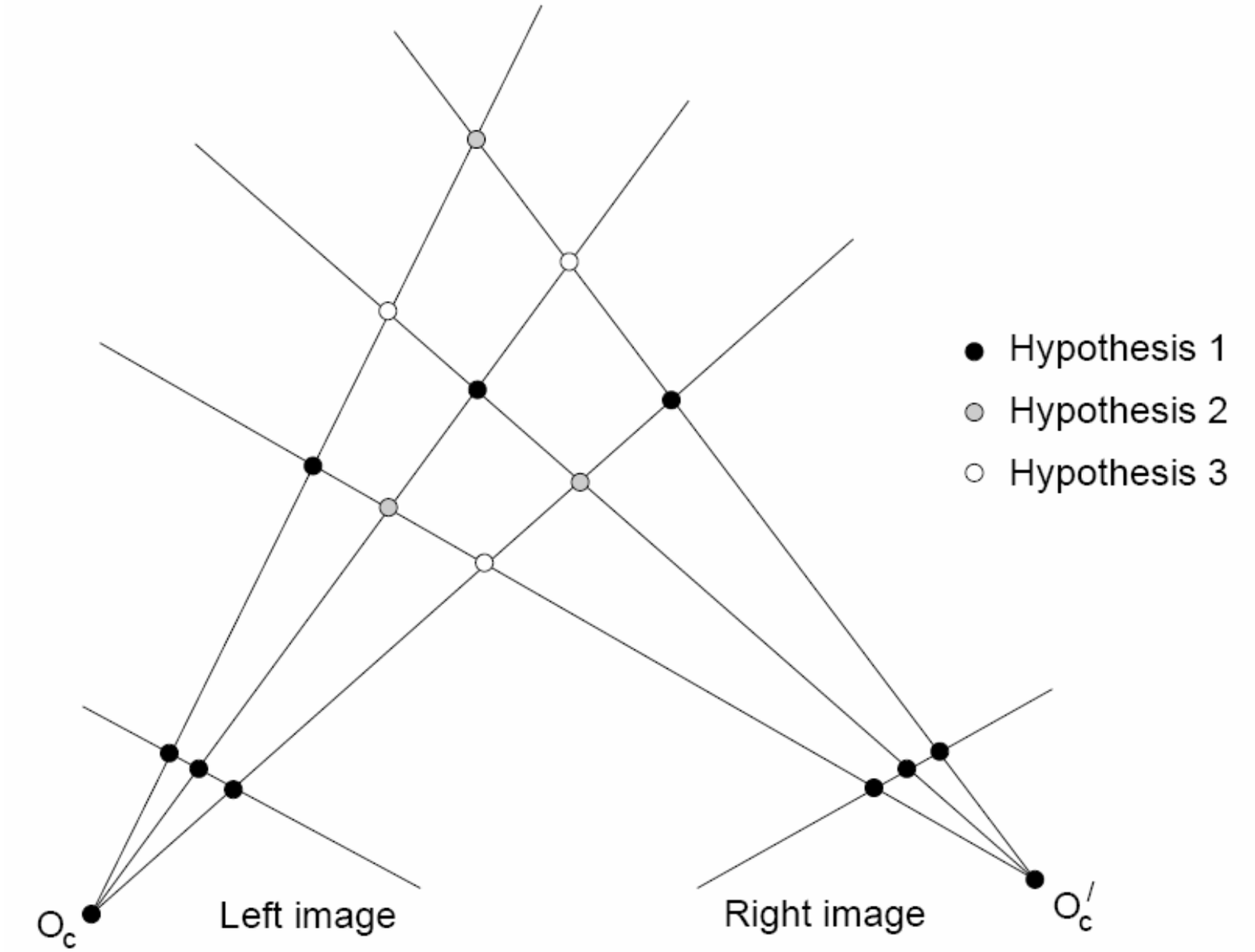
TECHNISCHE UNIVERSITÄT DARMSTADT

- Search over disparity to find correspondences
- Range of disparities to search over can change dramatically within a single image pair.



[Black]

Correspondence problem



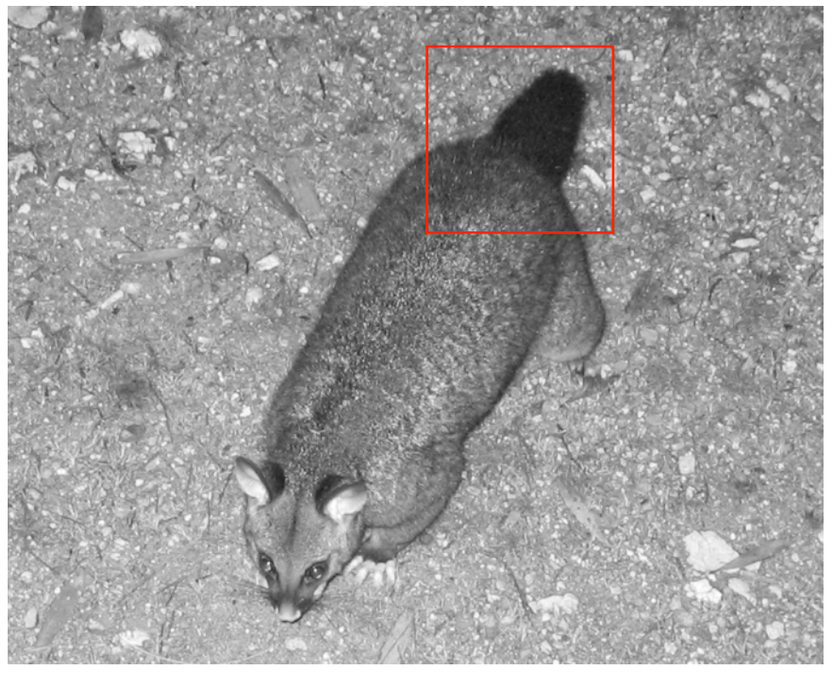
Even when constrained to 1D, there are multiple matching hypotheses - which one is correct?

Correspondence problem

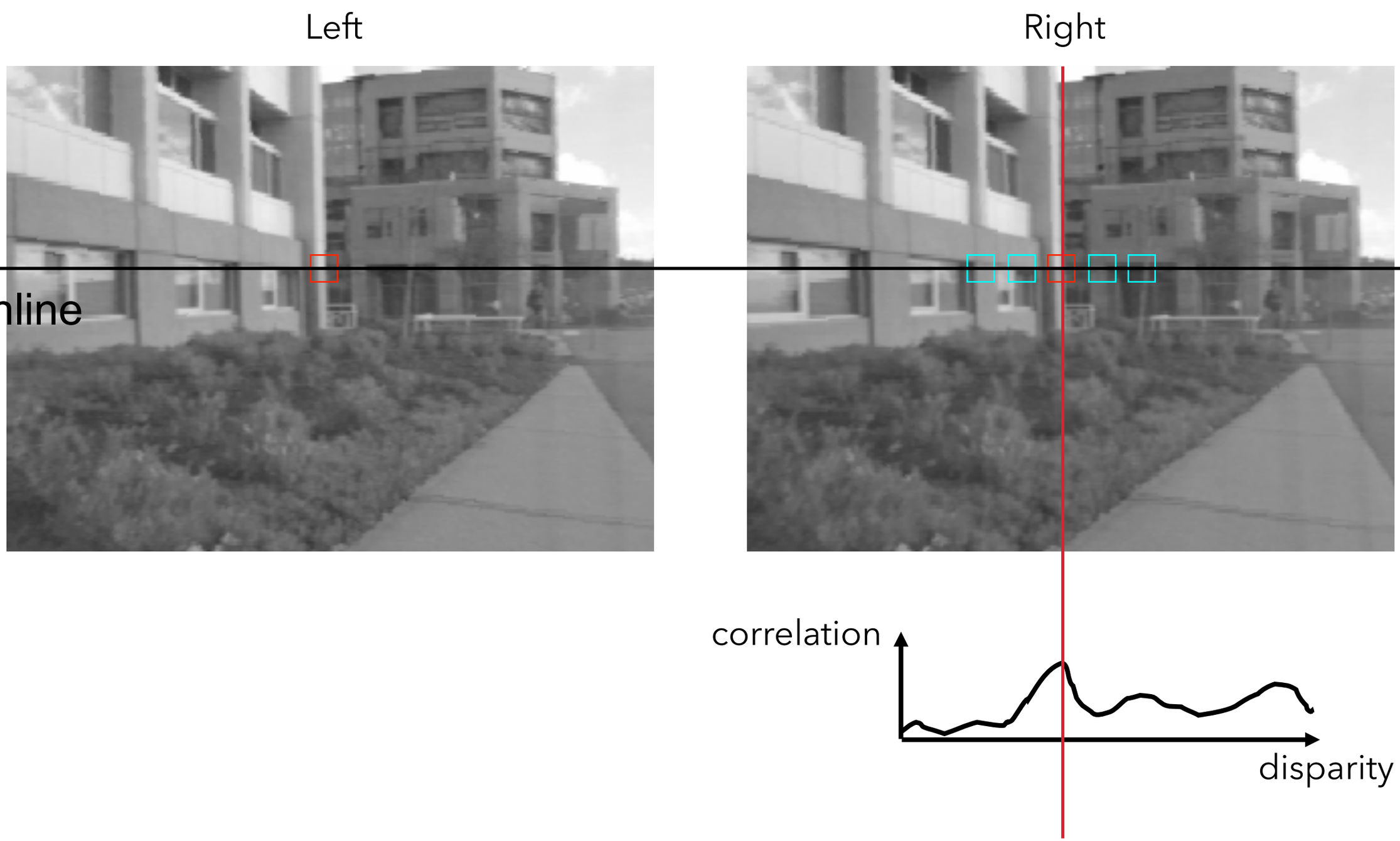
TECHNISCHE UNIVERSITÄT DARMSTADT

- Matching with narrow baseline
  - What are two “similar pixels”?
  - An ill-posed question, a single intensity sample does not reveal the local image structure
  - Measure similarity of the surrounding region
  - **Area-based similarity measures**



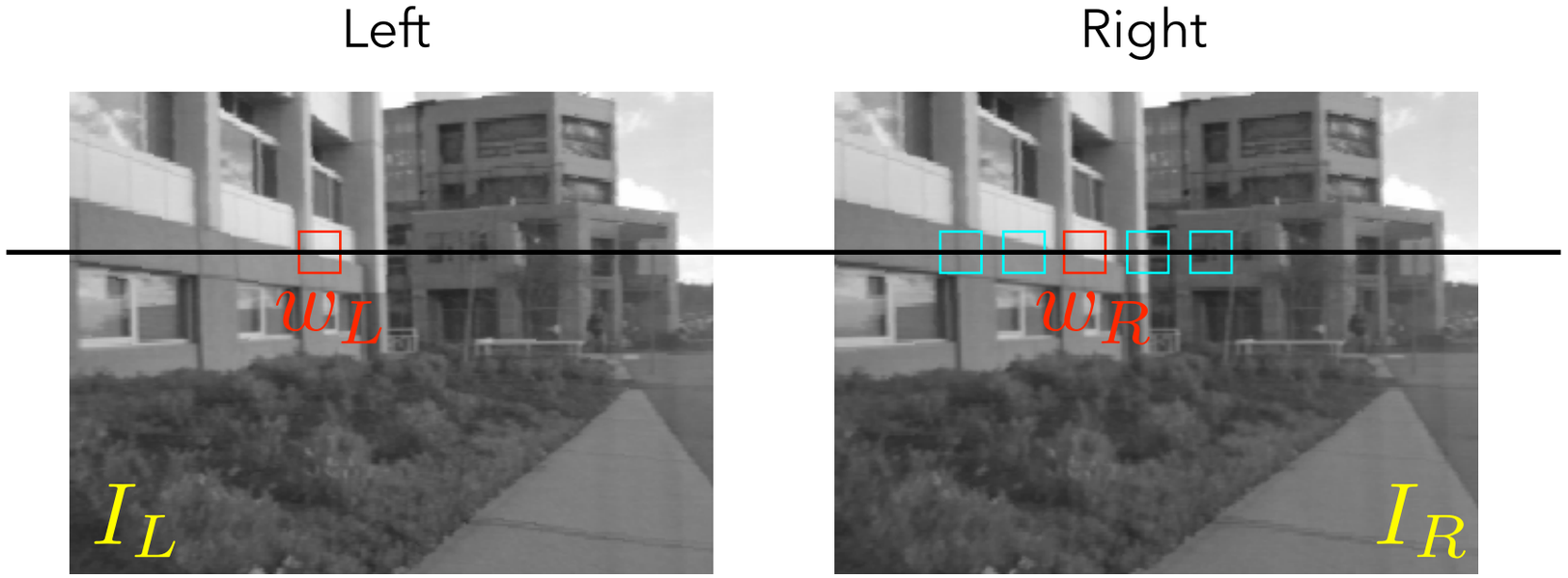


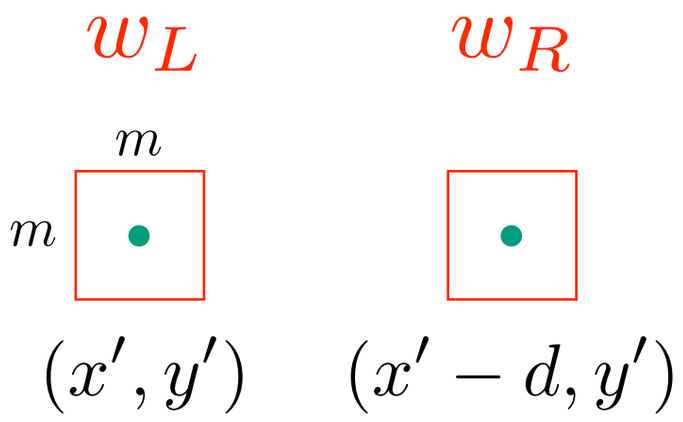
Correspondence Using Correlation



[Black]

Normalized Correlation



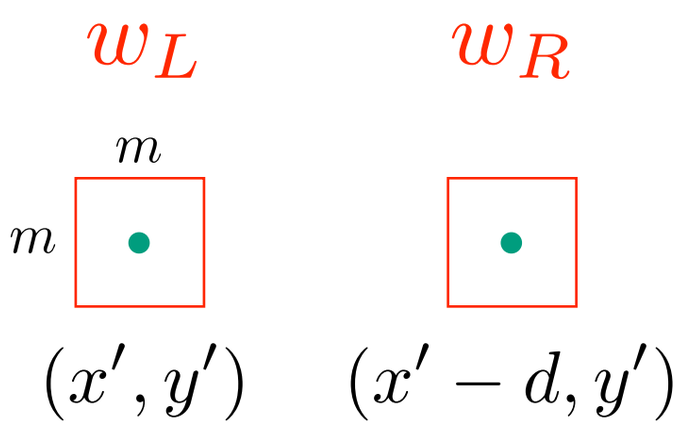


- w_L and w_R are corresponding $m \times m$ windows of pixels. We also write them as vectors $\mathbf{w}_L$ and $\mathbf{w}_R$.
- The normalized correlation computes the cosine of the angle between the patches:

$$\mathrm{NC}(x, y, d) = \frac{(\mathbf{w}_L(x, y) - \bar{\mathbf{w}}_L(x, y))^{\mathrm{T}} (\mathbf{w}_R(x - d, y) - \bar{\mathbf{w}}_R(x - d, y))}{|\mathbf{w}_L(x, y) - \bar{\mathbf{w}}_L(x, y)| \, |\mathbf{w}_R(x - d, y) - \bar{\mathbf{w}}_R(x - d, y)|}$$

patch mean

Even simpler: Sum of Squared (Pixel) Differences



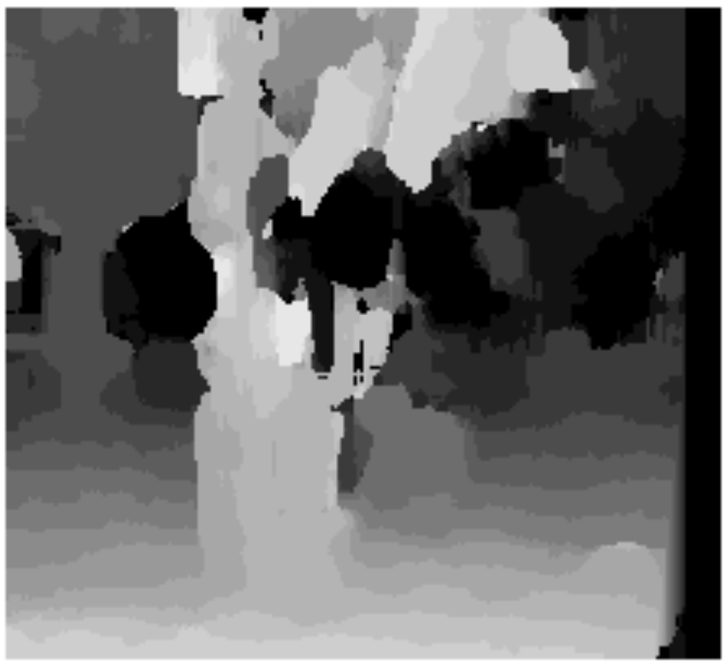
- w_L and w_R are corresponding $m \times m$ windows of pixels.
- The SSD cost measures the intensity difference as a function of disparity:

$$\mathrm{SSD}_R(x, y, d) = \sum_{(x',y') \in w_L(x,y)} (I_L(x', y') - I_R(x' - d, y'))^2$$

[Black]

Influence of window size

$m = 3$



$m = 20$

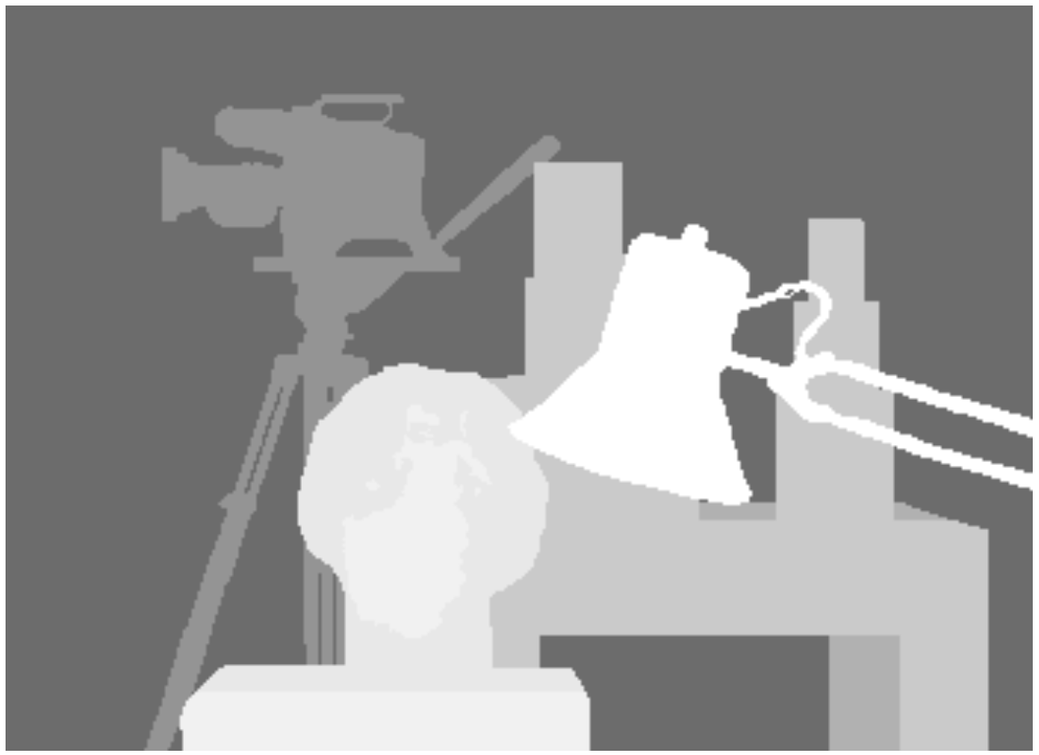
- Challenges and Problems:
  - How do we choose the right window size m?
  - Mismatches often lead to relatively poor results quality.

Preview: Stereo results

TECHNISCHE UNIVERSITÄT DARMSTADT

- Data from University of Tsukuba:

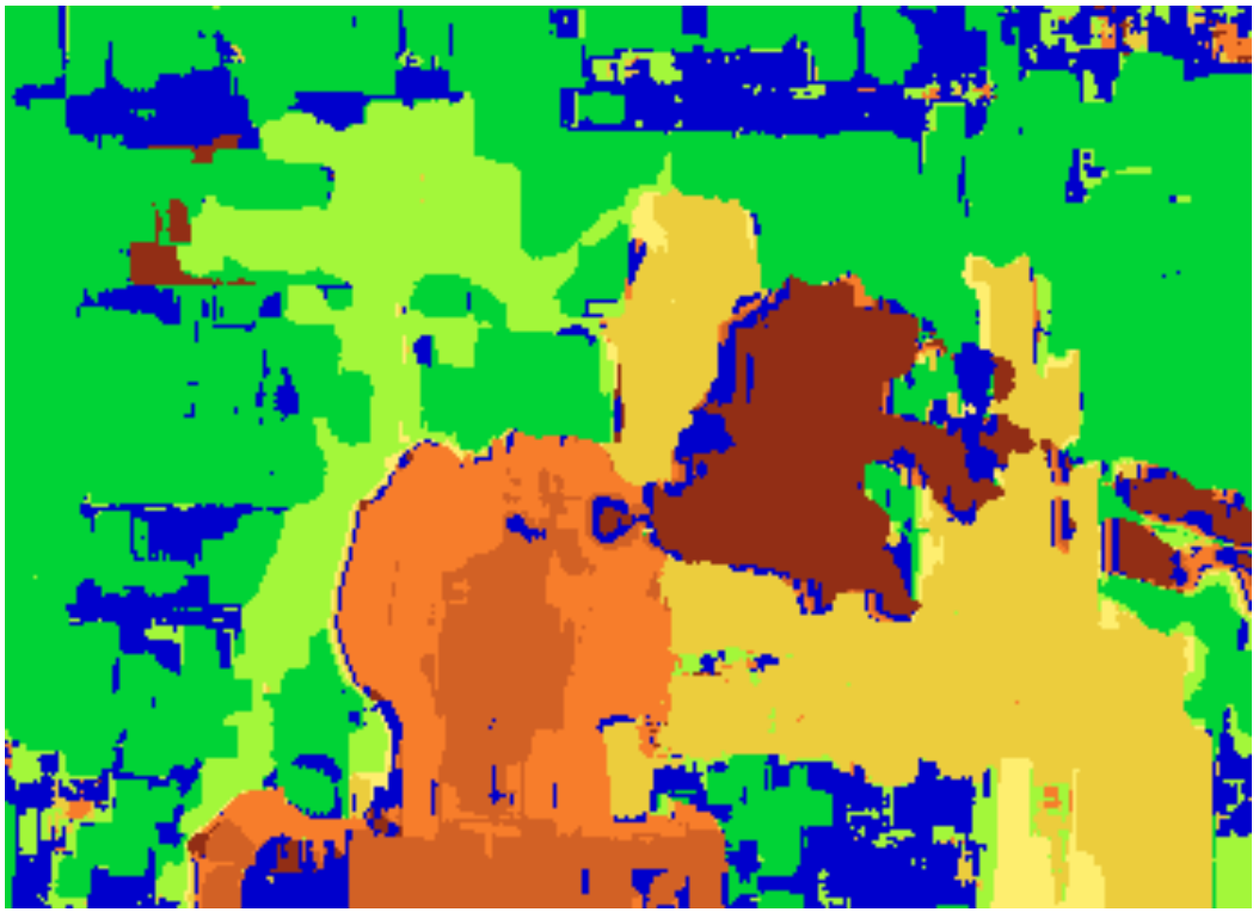
Scene



Ground truth

[Seitz]

Results with window correlation



Window-based matching
(best window size)

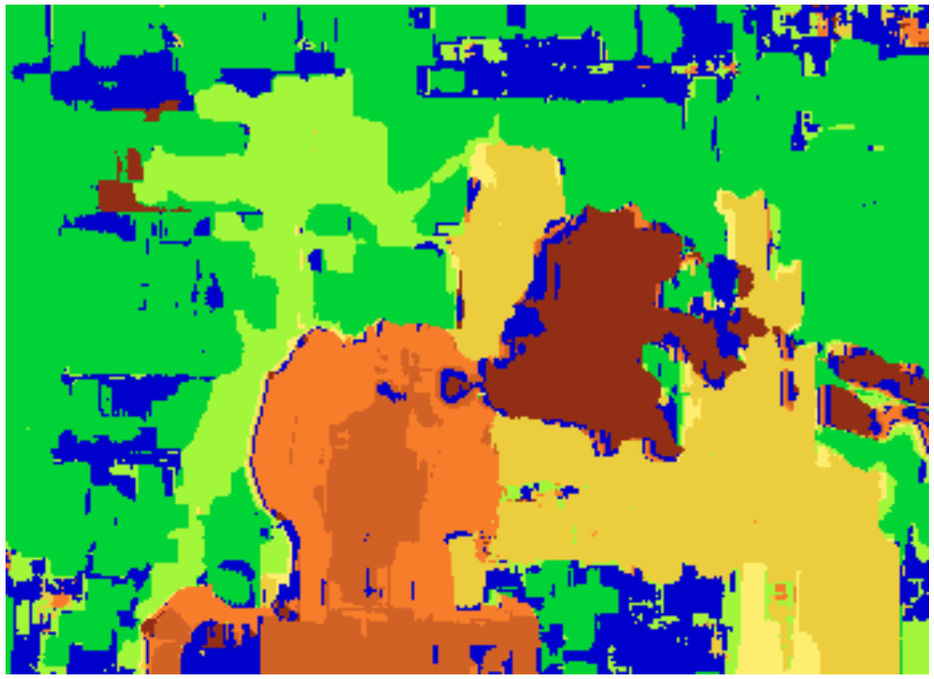
Ground truth

[Seitz]

How can we improve window-based matching?

TECHNISCHE UNIVERSITÄT DARMSTADT

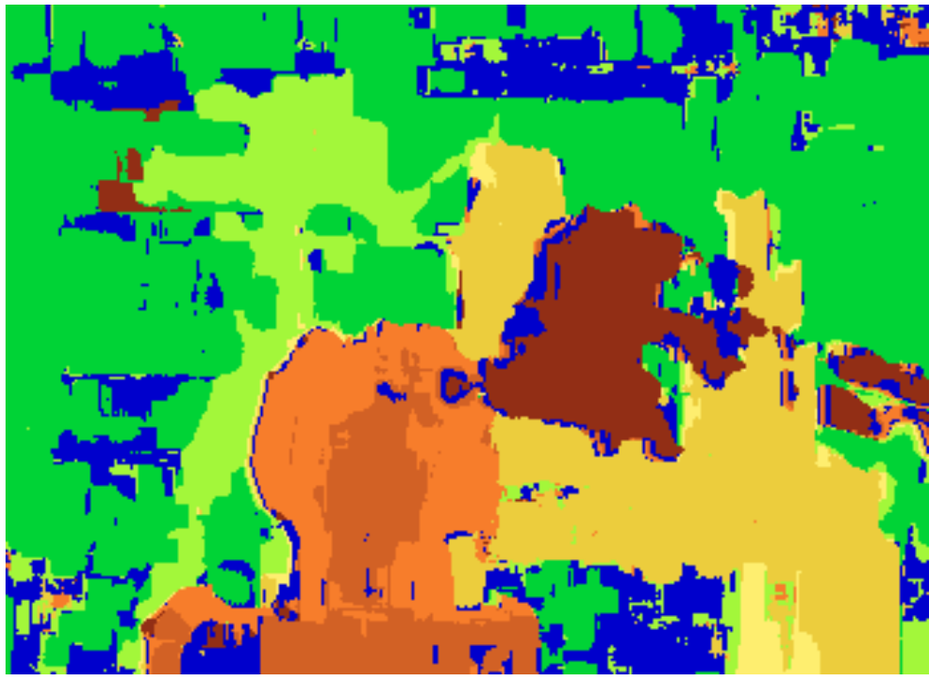
- The similarity constraint is local
  - each reference window is matched independently
  - other points do not influence the result



How can we improve window-based matching?

TECHNISCHE UNIVERSITÄT DARMSTADT

- The similarity constraint is local
  - each reference window is matched independently
  - other points do not influence the result
- Need to enforce non-local correspondence constraints



Slide by Lana Lazebnik

Where do we go from here?

TECHNISCHE UNIVERSITÄT DARMSTADT

- How can we impose regularity constraints that impose global consistency?
  - This is also called **regularization**.
- Goals:
  - We want to go beyond matching windows.
  - We want consistency within and between scanlines.
  - We want a model of consistency that is well supported by the properties of the real world, e.g. by real scene depth or real motion.
  - We want a model that is computationally manageable.
  - We would like to find a model of consistency that does not only work for stereo or flow, but also for other applications.

What is consistency?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Before we can do anything, we need to ask ourselves what it means to have spatial consistency or regularity.
- Let's look at some data to get inspiration:

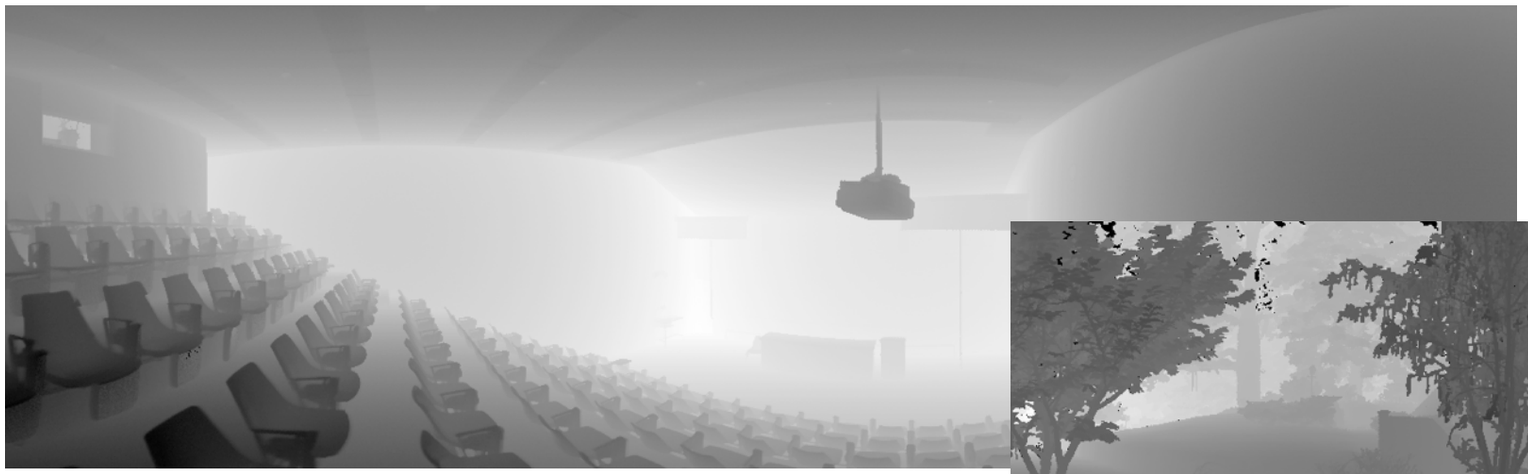
Range image - Scene depth from a range scanner

Range Scanning

TECHNISCHE UNIVERSITÄT DARMSTADT

- This image was captured using a laser range finder.
  - Measures the distance to each point in the scene.
  - Based on the time of flight of a laser beam.
  - Quite accurate: ~15mm
  - Angular resolution: 0.005°
- Brown Range Image Database:
  - David Mumford and (former) students

From Riegl



[Lee, Huang, & Mumford]

What can we conclude from this data?

- It helps us see more clearly what we know from everyday life:
  - The depth of nearby points in the scene is (almost) the same.
  - But sometimes, there are depth discontinuities, for example at object boundaries.

What can we conclude from this data?

- In other terms:
  - We (as humans) have **a-priori knowledge** about how 3D scenes typically look like, even if we have never seen the particular scene in question before.
- How do we exploit this a-priori knowledge for computer vision?

Vision as Probabilistic Inference

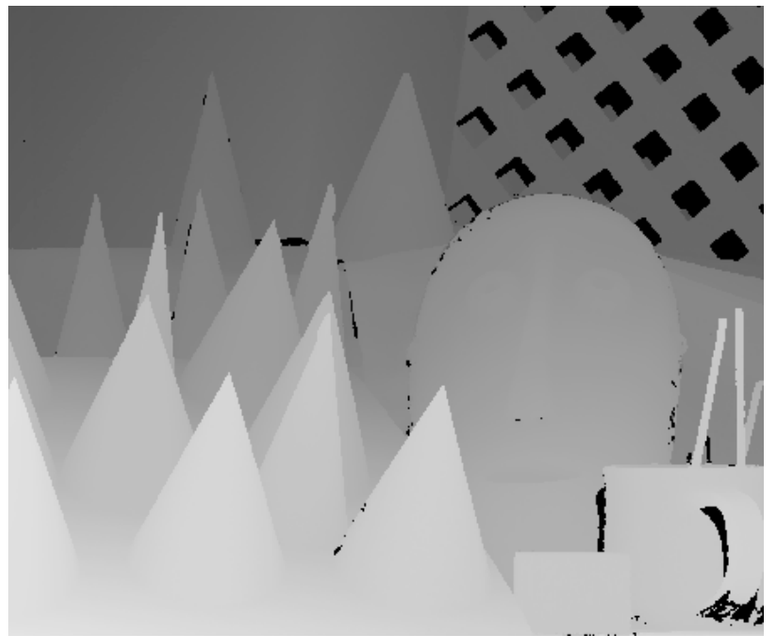
TECHNISCHE UNIVERSITÄT DARMSTADT

- In the introduction we discussed that:
  - Many vision problems are underconstrained and require prior knowledge to be solved (as is the case here).
  - We almost always have to deal with uncertain ("noisy") data.
- Both of these are key motivations for using **probabilistic approaches** to computer vision:
  - We regard both the measurement and the interpretation as uncertain (Bayesian approach).
  - We model the problem using probabilities (or prob. densities).
  - We obtain the solution using methods of probabilistic inference.
- What does all this mean?

Stereo using Probabilistic Methods

TECHNISCHE UNIVERSITÄT DARMSTADT

Uncertain image measurements



Uncertain state of the world

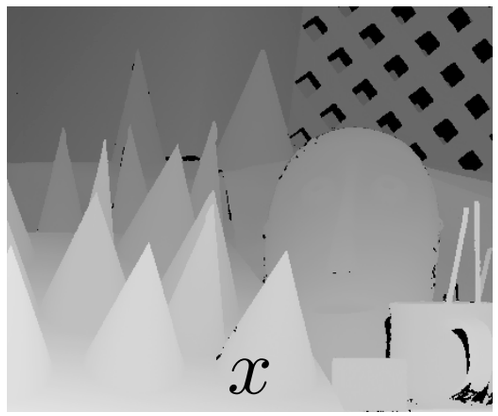
We're given

Want to know

Stereo using Probabilistic Methods

TECHNISCHE UNIVERSITÄT DARMSTADT

Uncertain image measurements



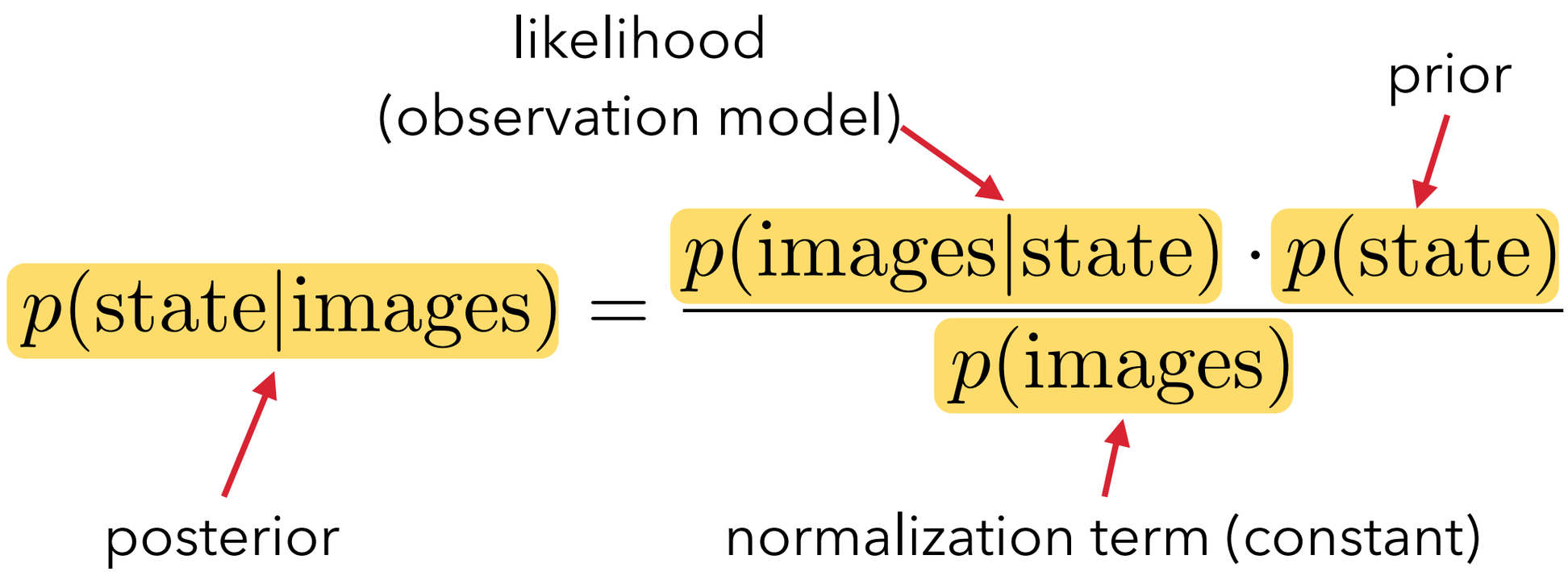
Uncertain state of the world

- Model using posterior distribution: $p(\text{state}|\text{images})$
  - Describe the probability of the state of the world given the image measurements.
- How do we find the “best” state of the world?
  - Using probabilistic inference, e.g. we maximize w.r.t. state x

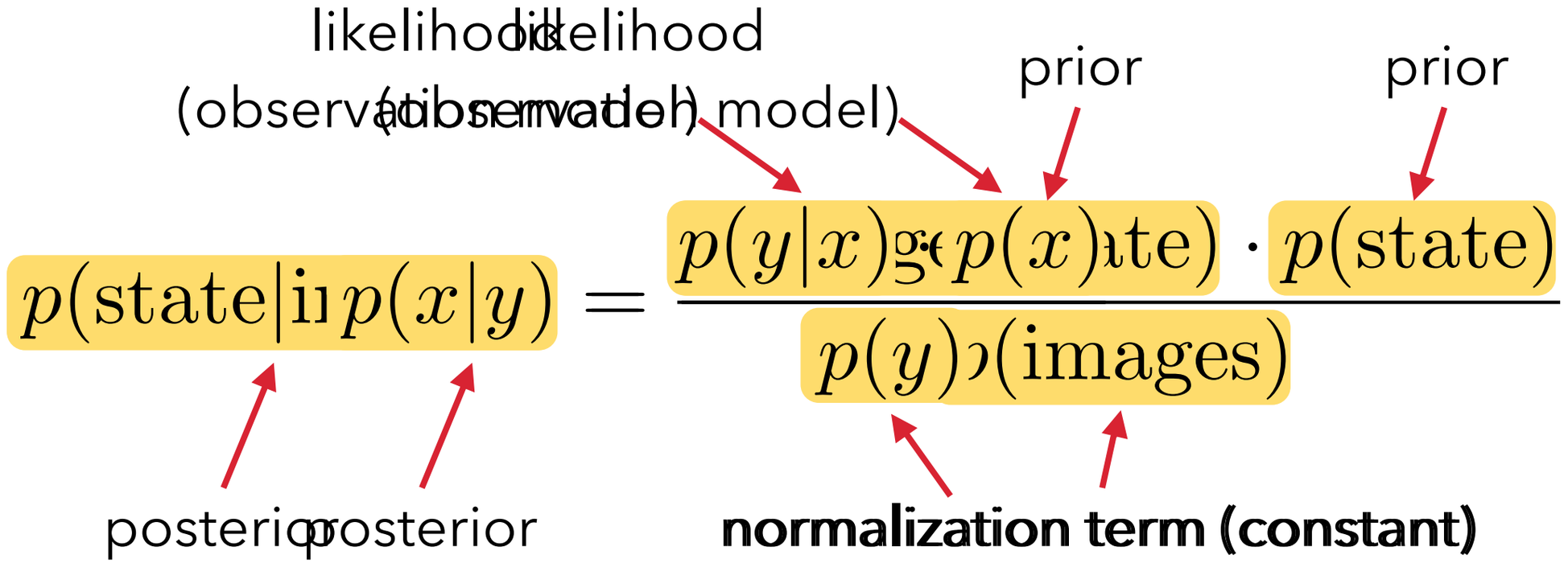
Modeling the Posterior

TECHNISCHE UNIVERSITÄT DARMSTADT

- How do we model the posterior?
  - This can be done directly (discriminative approaches), but we will not do this now as it is more difficult.
- Instead, we simplify the modeling problem by applying Bayes' rule (generative approach):



Modeling the Posterior



- The likelihood $p(y|x)$is an observation model that describes how we obtained the image measurements, given a particular state of the world.
- The prior $p(x)$models our a-priori assumptions about the world, or the state of the world.
- The normalization term $p(y)$can often be ignored, because it only depends on the image measurements, which are given to us.

Three problems

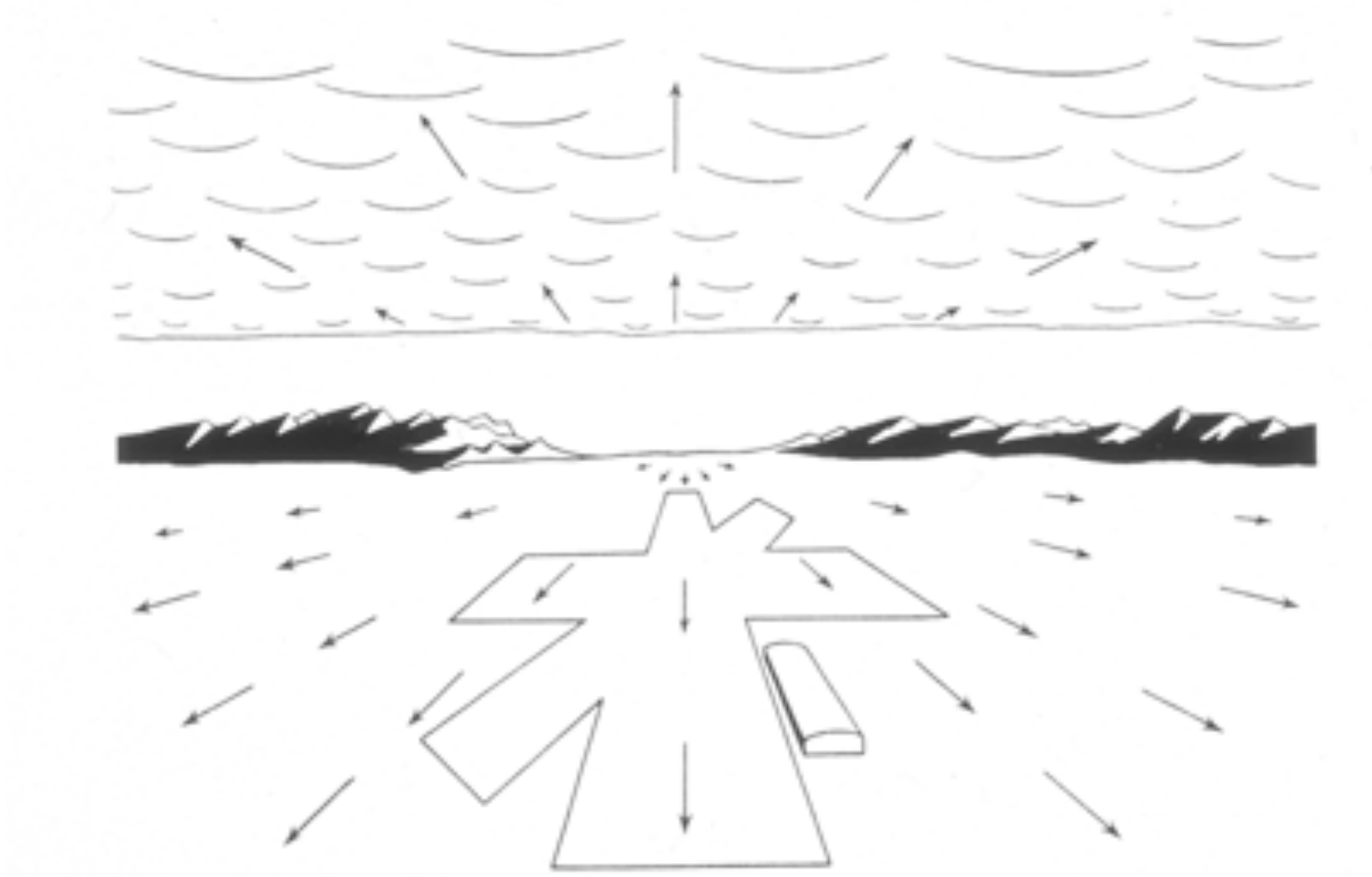
- How do we model the likelihood $p(y|x)$?
  - Stereo, flow, image restoration, etc.

- How do we model the prior $p(x)$?
  - Again, many applications
  - disparity, motion, natural images, image segments, etc.

- Given the posterior, how do we compute a solution?
  - What is the right objective?
  - What are good algorithms?

Optical Flow

TECHNISCHE UNIVERSITÄT DARMSTADT

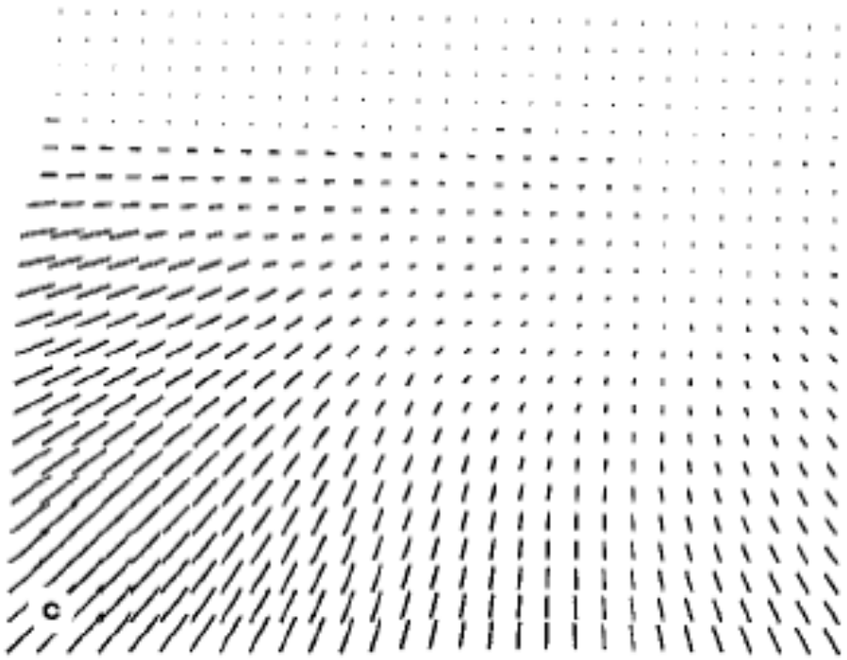


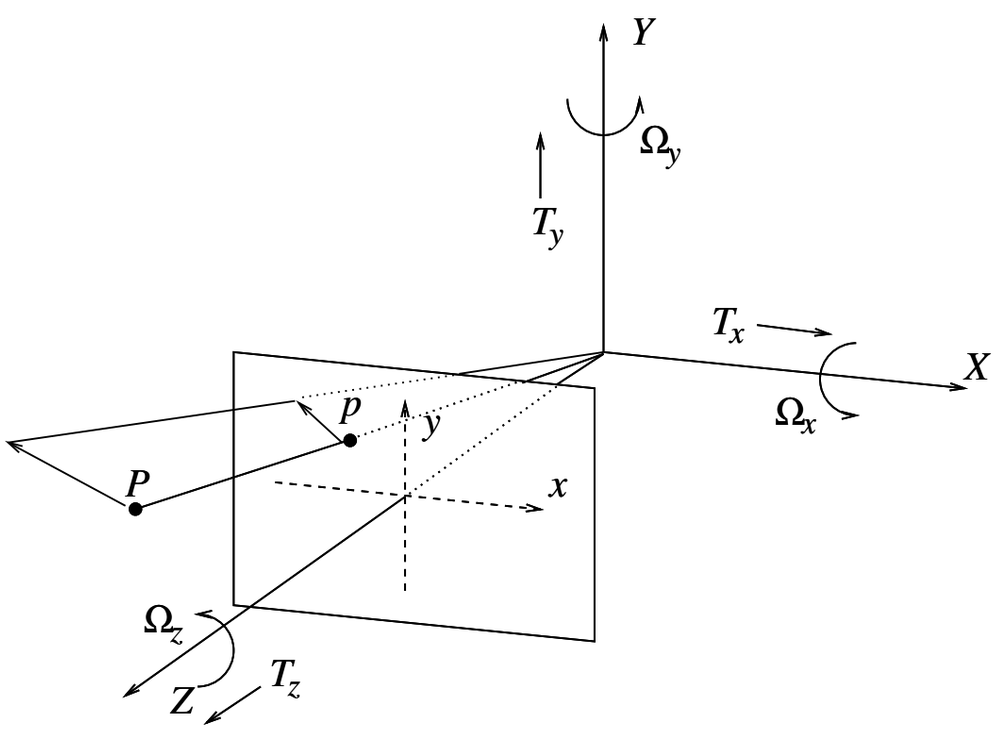
J. J. Gibson, The Ecological Approach to Visual Perception

Optical Flow Field

Image irradiance at time t and location $\mathbf{x} = (x, y)$:

$$I(x, y, t)$$



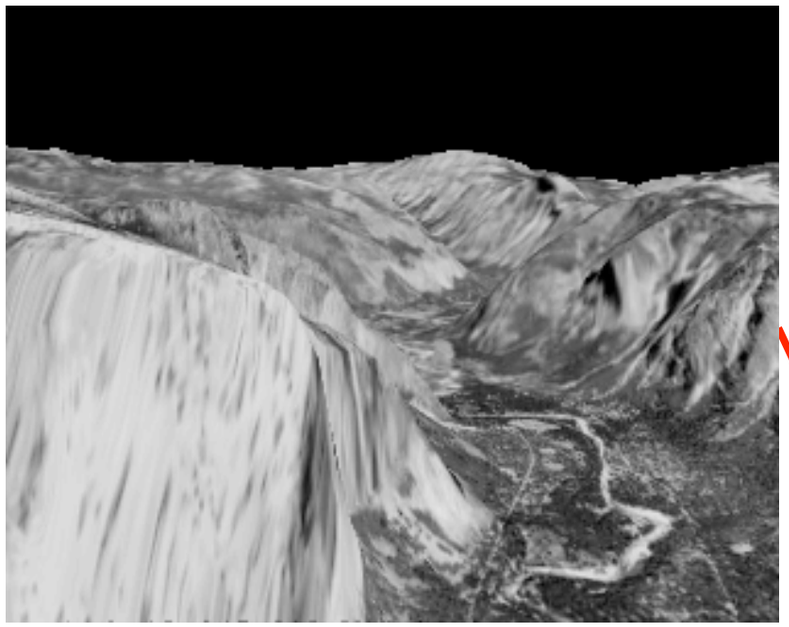


$u(x, y)$ Horizontal component

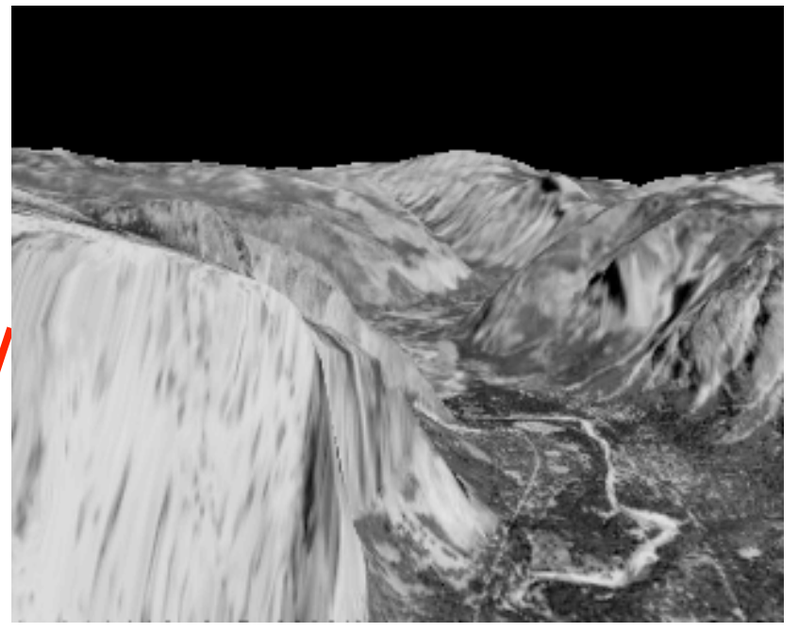
$v(x, y)$ Vertical component

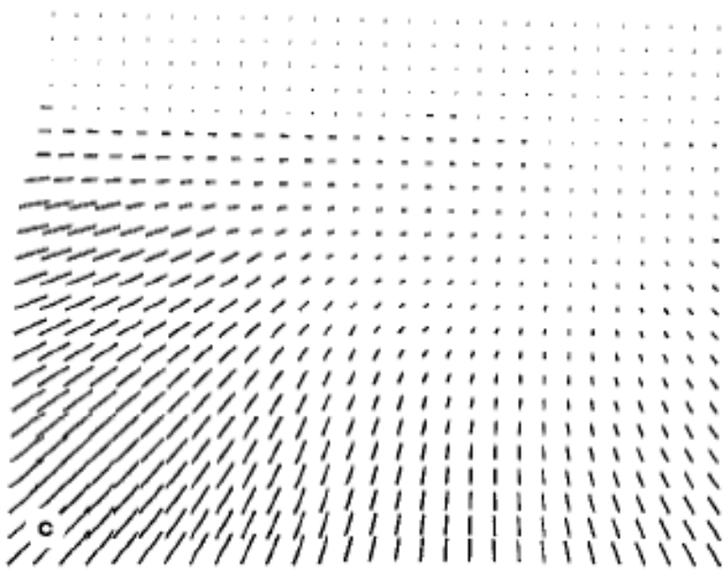
Optical Flow Estimation

time t

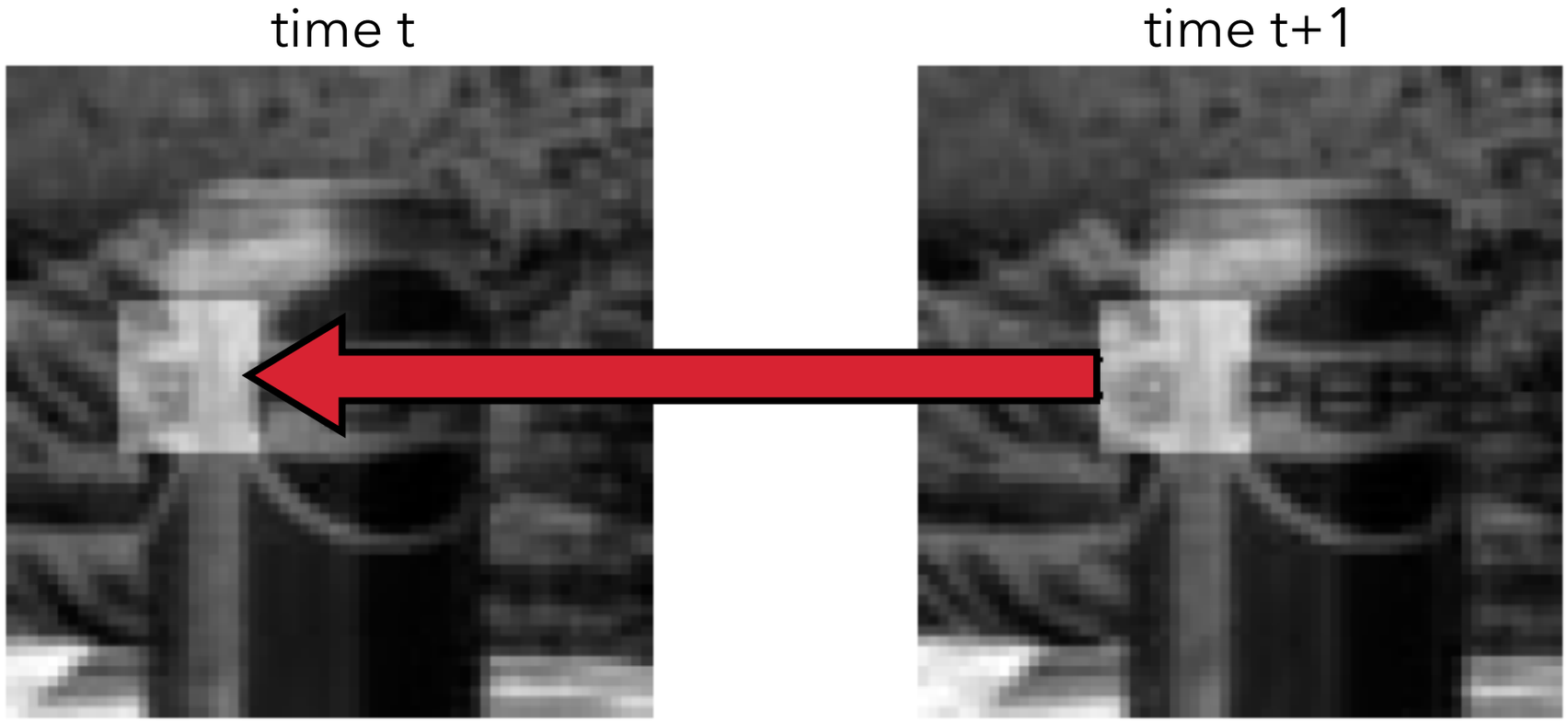


time t+1





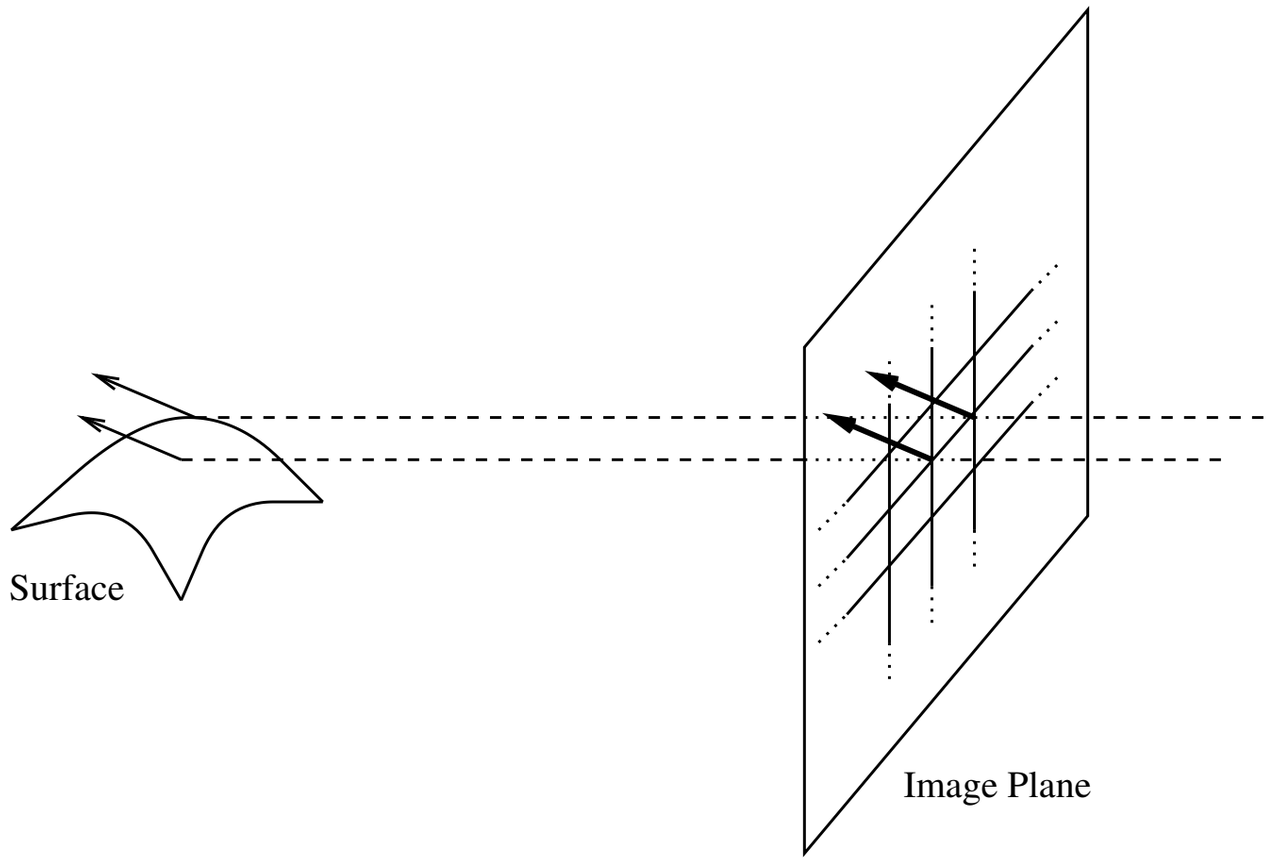
Brightness Constancy



Assumption 1:

Image measurements (e.g. brightness) in a small region remain the same although their location may change.

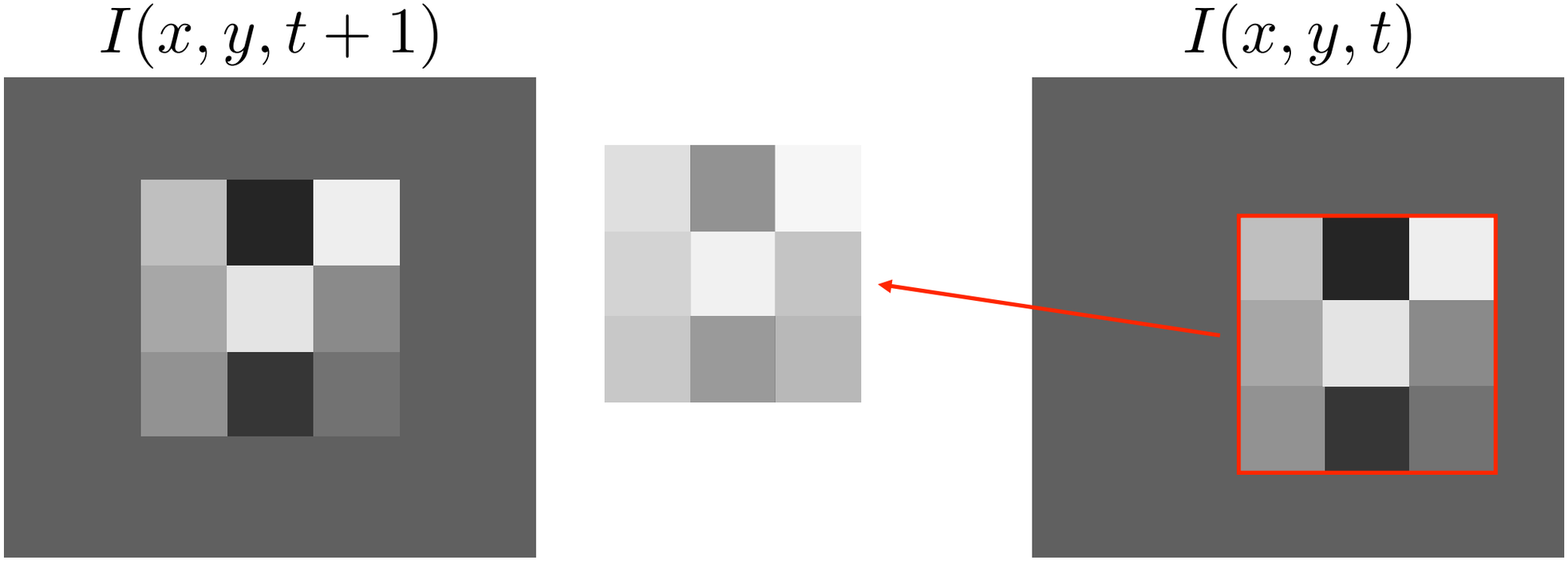
Spatial Coherence



Assumption 2:

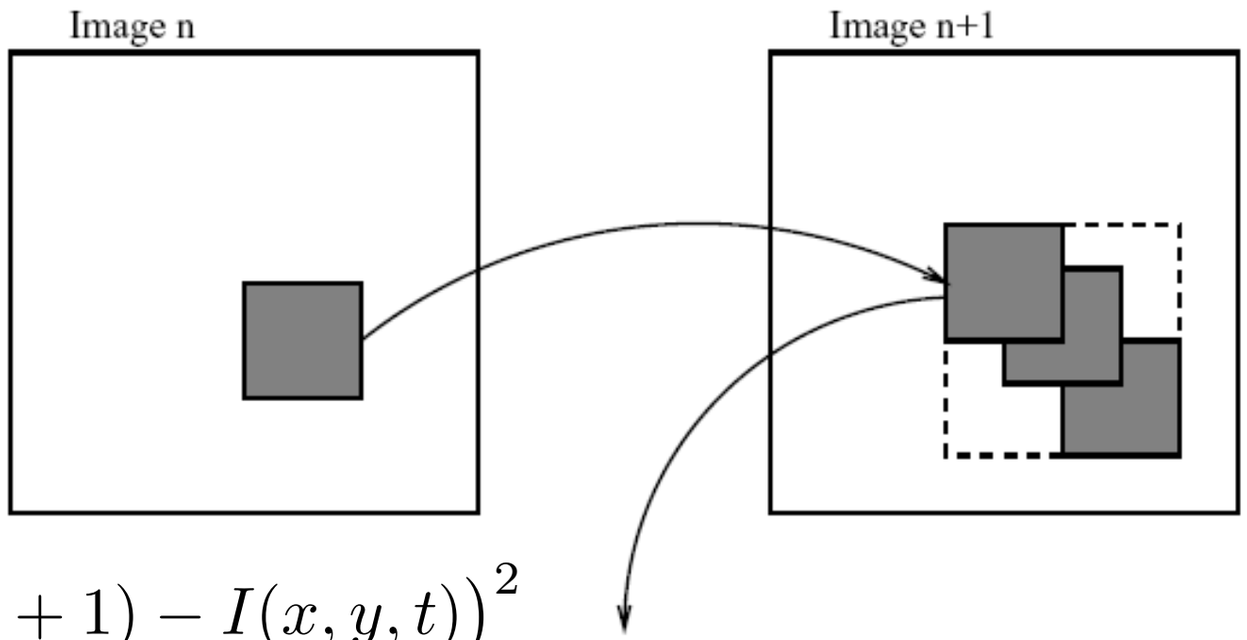
- Neighboring points in the scene typically belong to the same surface and hence typically have similar 3D motions.
- Since they also project to nearby points in the image, we expect spatial coherence in the image flow.

Minimize Brightness Difference

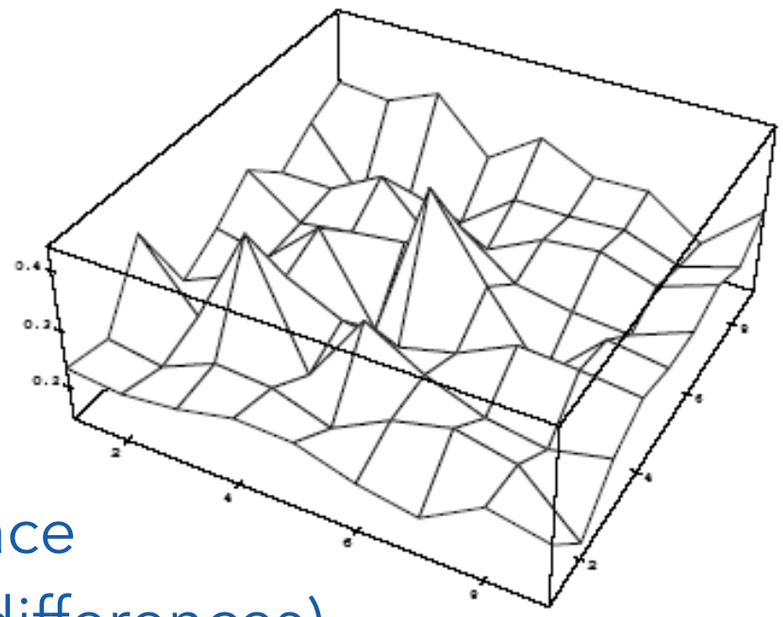


$$E_{SSD}(u, v) = \sum_{(x,y) \in R} \left(I(x + u, y + v, t + 1) - I(x, y, t) \right)^2$$

Sum of Squared Differences



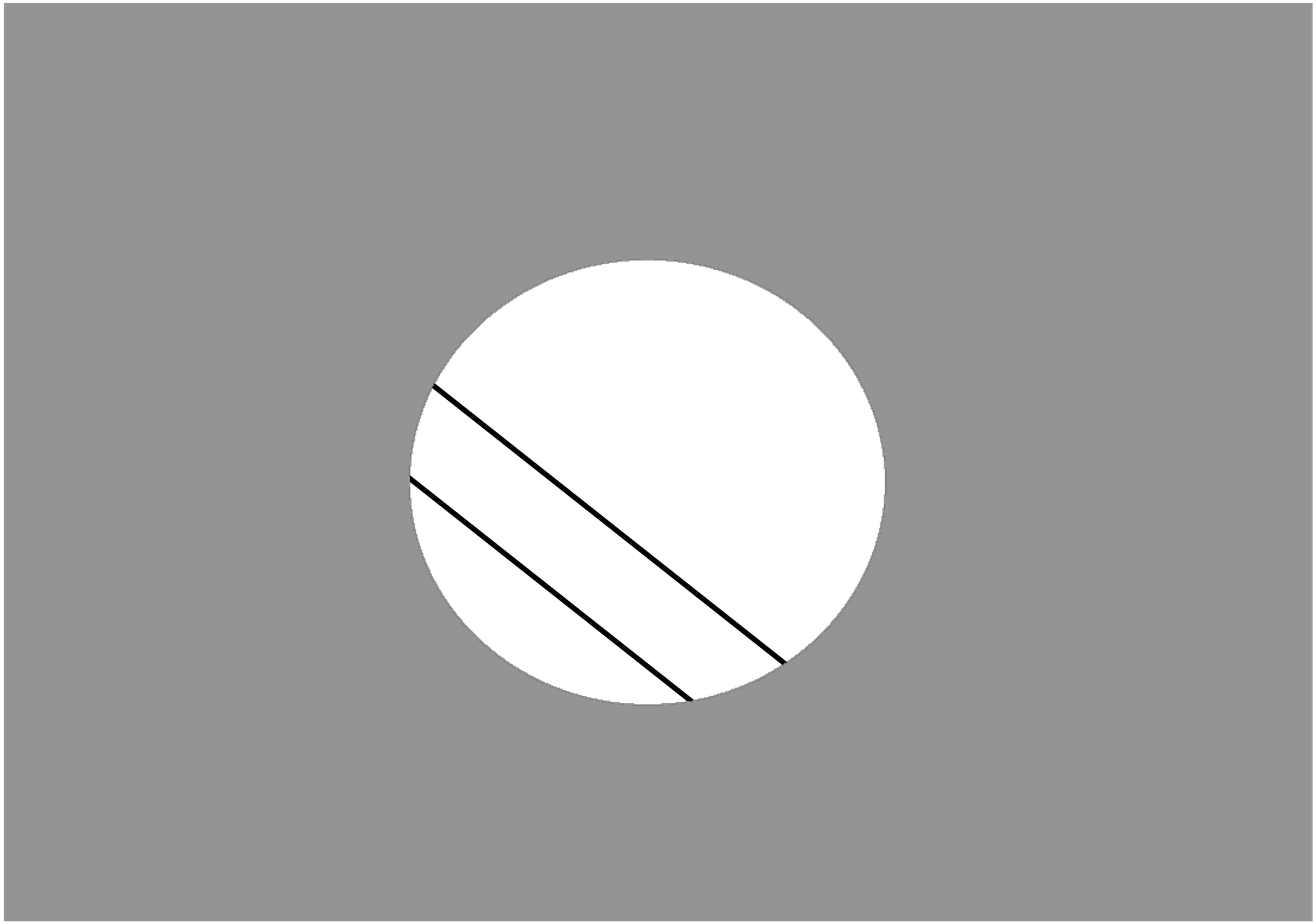
$$E_{SSD}(u,v) = \sum_{(x,y)\in R} \big(I(x+u,y+v,t+1) - I(x,y,t)\big)^2$$



SSD surface
(sum of squared differences)

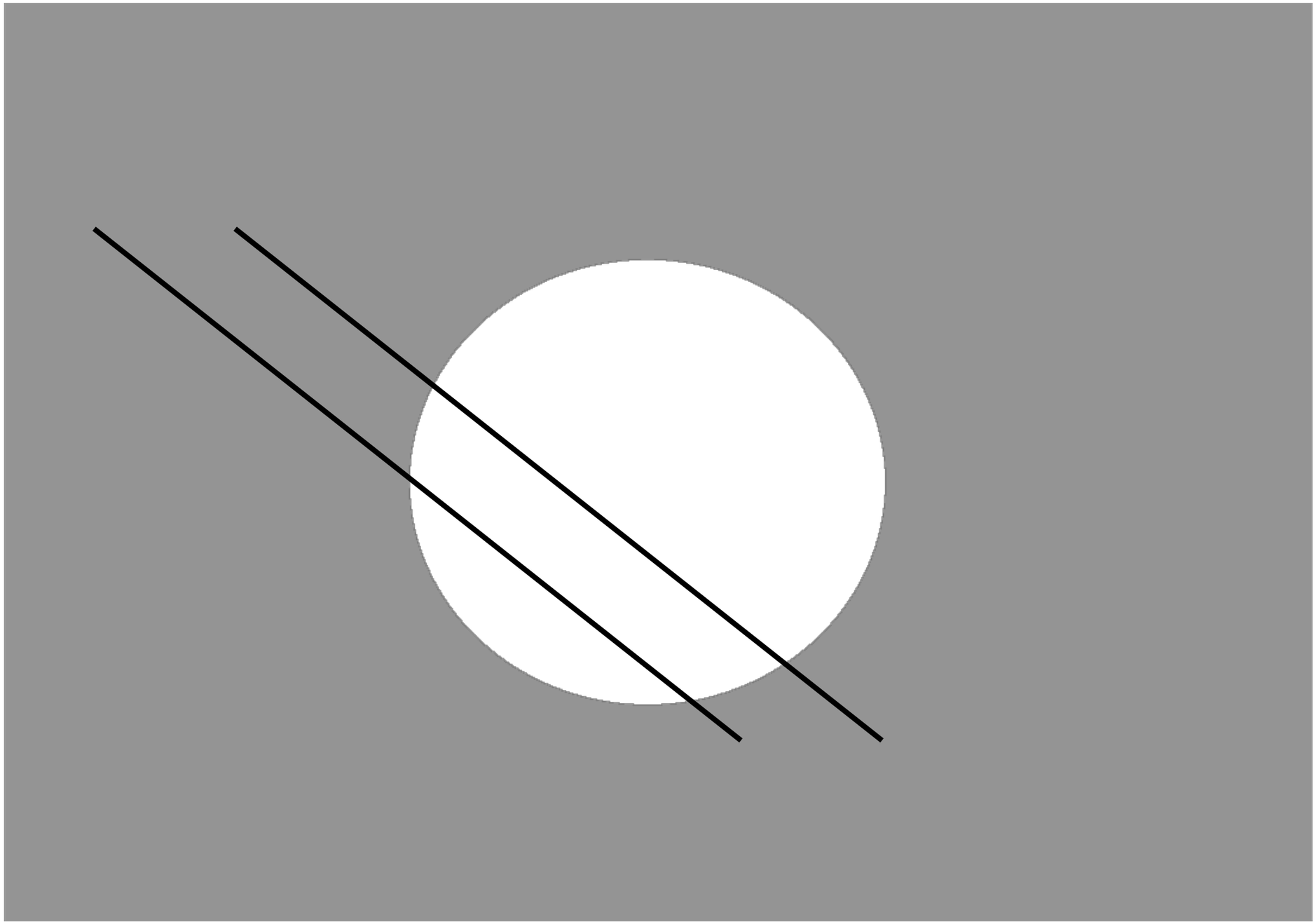
Aperture Problem

TECHNISCHE UNIVERSITÄT DARMSTADT



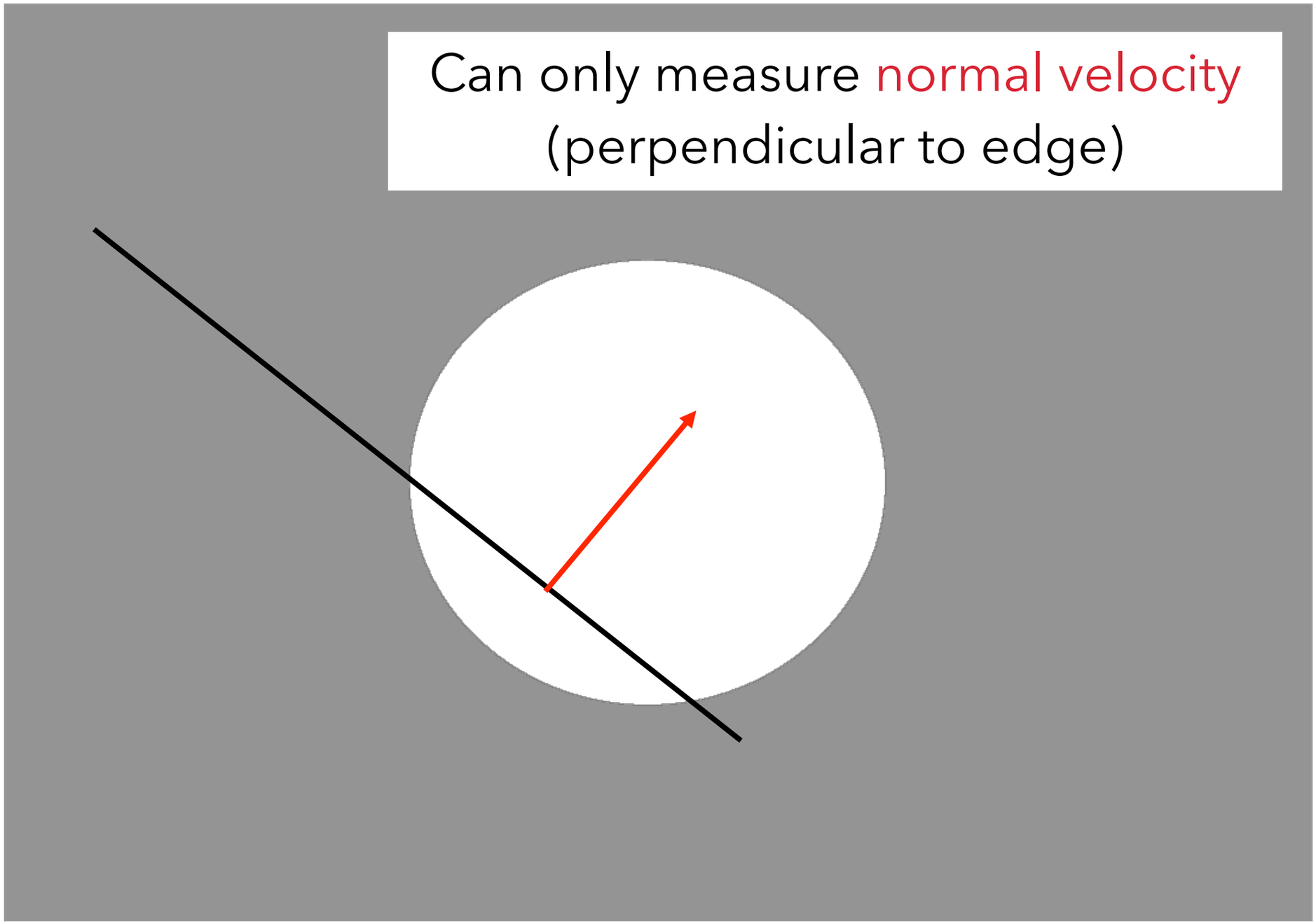
Aperture Problem

TECHNISCHE UNIVERSITÄT DARMSTADT

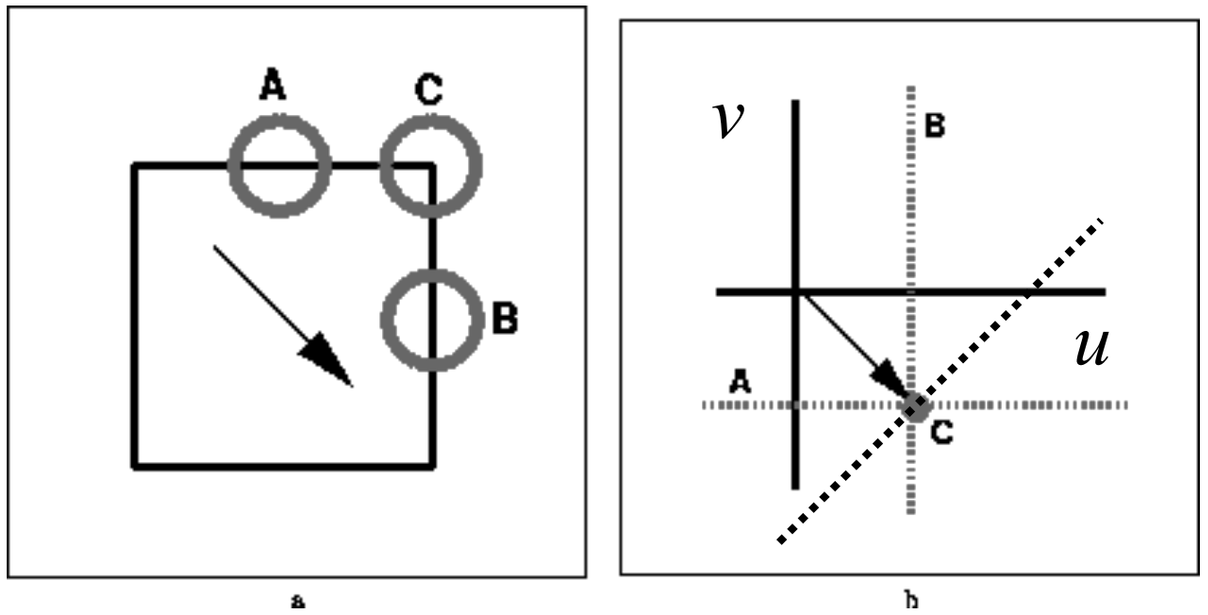


Aperture Problem

TECHNISCHE UNIVERSITÄT DARMSTADT



Multiple Constraints



[Yair Weiss]

Combine multiple constraints to get an estimate of the velocity (not only the normal component).

Area-Based Flow

TECHNISCHE UNIVERSITÄT DARMSTADT

- How do we combine multiple constraints?
  - Remember our assumptions.
  - We will assume spatial smoothness of the flow.
- More specifically:
  - Assume that the flow is constant in a region.

$$E_{SSD}(u,v) \approx \sum_{(x,y)\in R} \left(u \cdot I_x(x,y,t) + v \cdot I_y(x,y,t) + I_t(x,y,t)\right)^2$$

  - This is what we have been doing (sliding window).
- But how do we solve for the motion?

Solving for u

$$\mathbf{u} = -\left(\sum_R \nabla I \nabla I^{\mathrm{T}}\right)^{-1} \left(\sum_R I_t \nabla I\right)$$

- This is a classical flow technique:

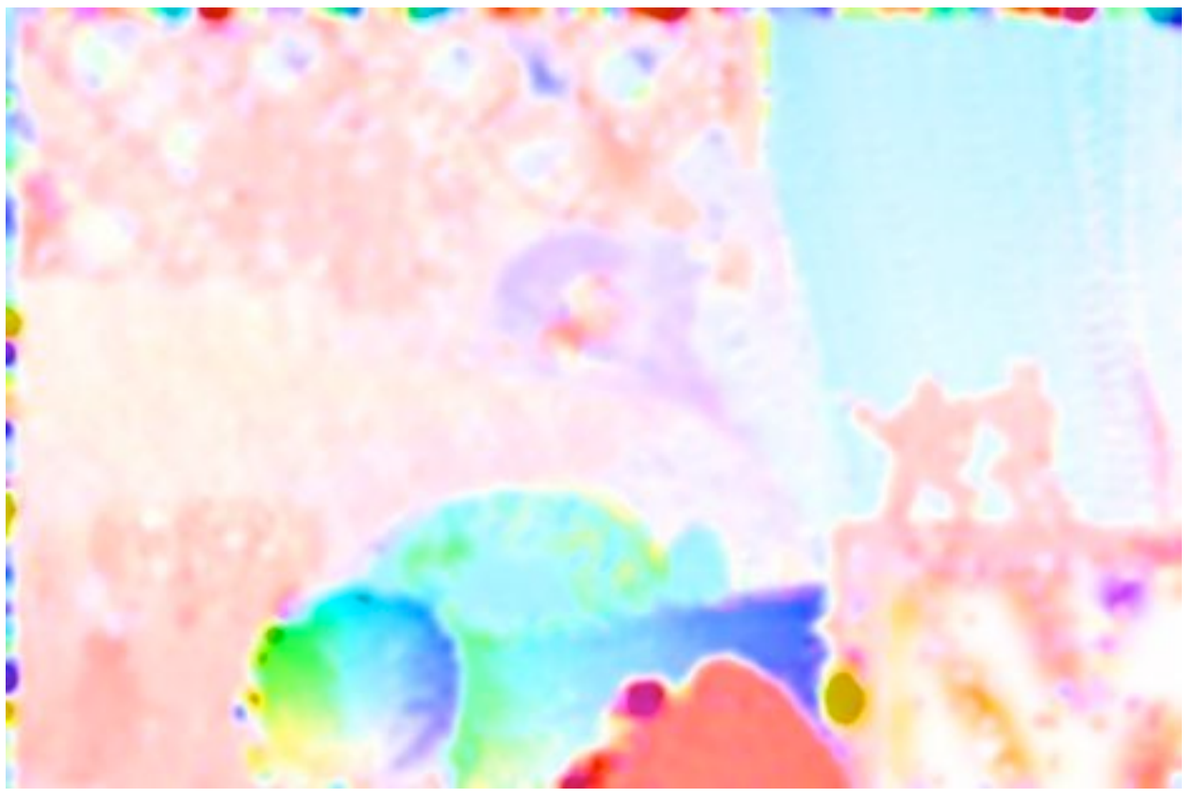
  B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. IJCAI, pp. 674–679, 1981.

- Use this to compute dense optical flow
  - Iterative estimation.
  - Coarse-to-fine estimation.

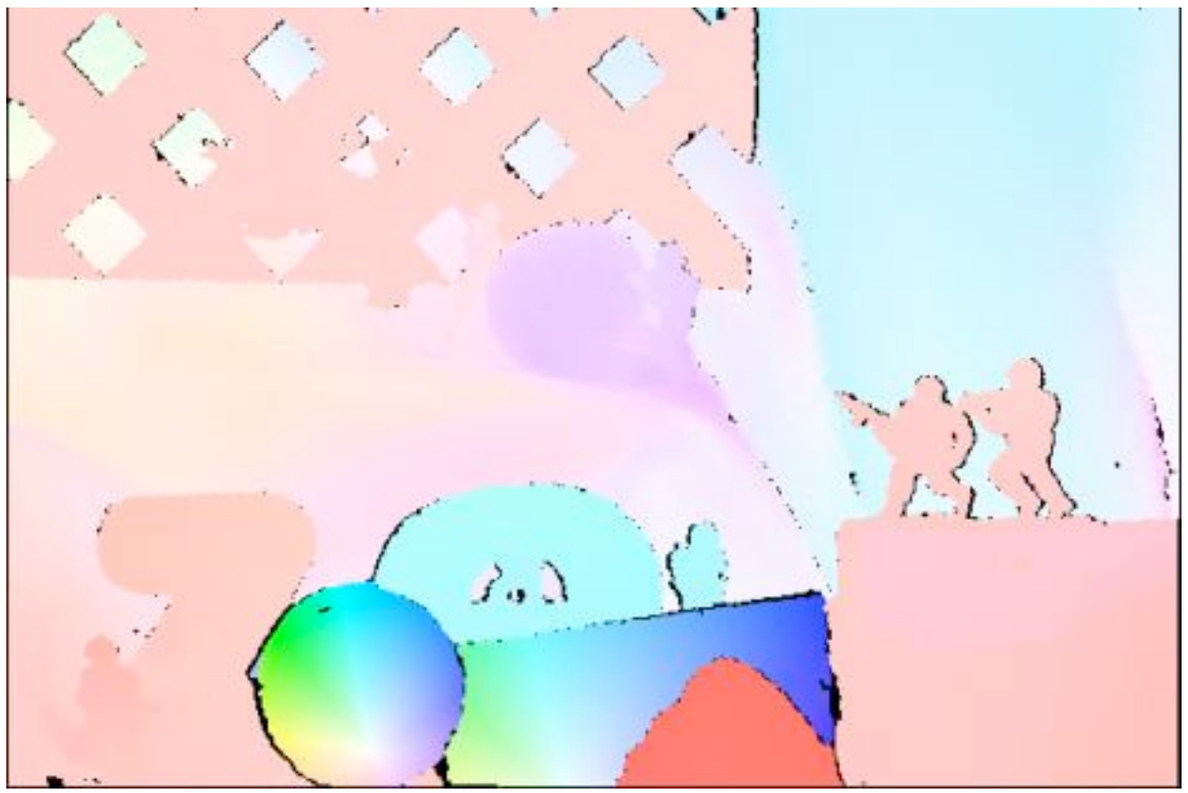
Do we get good results?

TECHNISCHE UNIVERSITÄT DARMSTADT

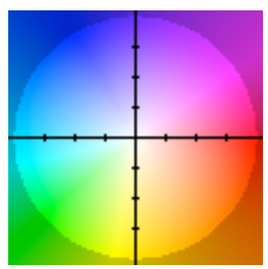
- Not really...



Pyramid LK method



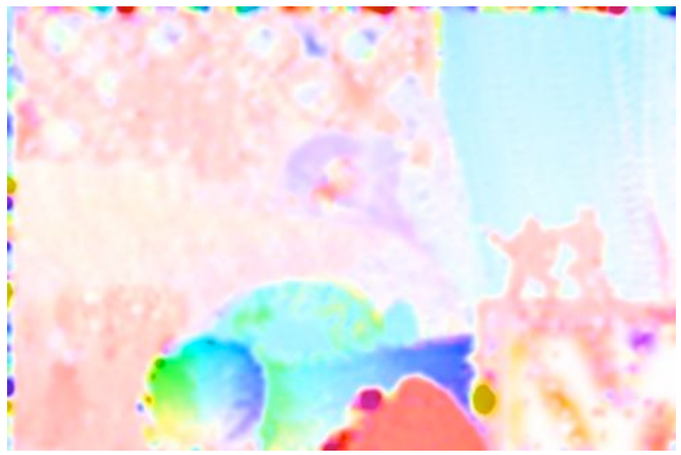
Ground truth



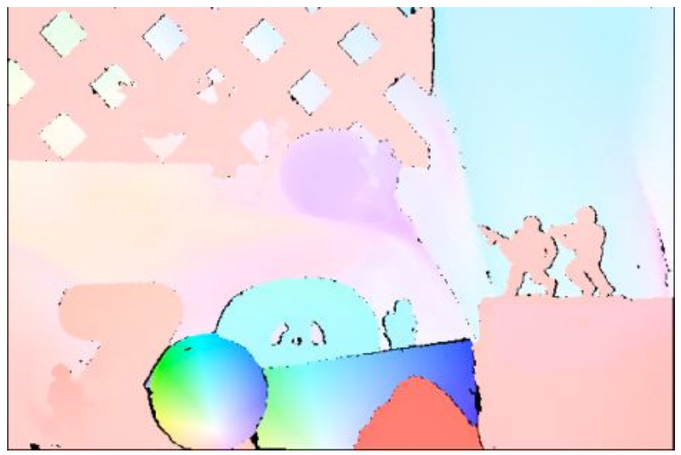
What is the problem?

TECHNISCHE UNIVERSITÄT DARMSTADT

- The window is **too big**!
  - All the discontinuities in the flow are smoothed over.
  - But discontinuities do exist, e.g., at motion boundaries.



- But: The window is also **too small**!
  - In some areas, the flow is really bad, because there is not enough image information in the window.
  - This happens in areas with little texture.



- LK is a local optical flow method.
  - Global flow methods to the rescue