Outlook: Articulated Tracking

11.06.2014

TECHNISCHE UNIVERSITÄT DARMSTADT

visual inference

Tracking

TECHNISCHE UNIVERSITÄT DARMSTADT

- We typically distinguish 3 cases:
  - Tracking rigid objects
  - Tracking articulated objects, e.g. humans or animals
  - Tracking fully non-rigid objects
- Today:
  - Tracking articulated objects, specifically humans

Bayesian Formulation

TECHNISCHE UNIVERSITÄT DARMSTADT

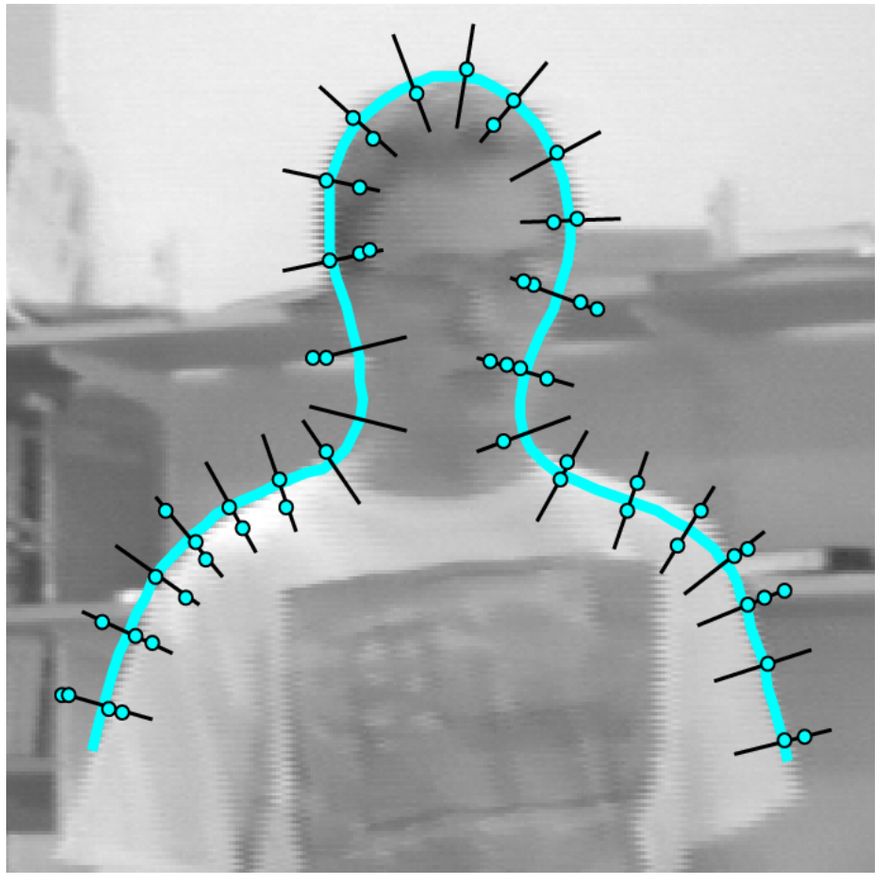
$$p(\boldsymbol{x}_k|\mathbf{Z}_k) = \kappa \cdot p(\boldsymbol{z}_k|\boldsymbol{x}_k) \cdot \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \cdot p(\boldsymbol{x}_{k-1}|\mathbf{Z}_{k-1}) \, \mathrm{d}\boldsymbol{x}_{k-1}$$

$p(\boldsymbol{x}_k\|\mathbf{Z}_k)$	posterior probability at current time step
$p(\boldsymbol{z}_k\|\boldsymbol{x}_k)$	likelihood
$p(\boldsymbol{x}_k\|\boldsymbol{x}_{k-1})$	temporal prior
$p(\boldsymbol{x}_{k-1}\|\mathbf{Z}_{k-1})$	posterior probability at previous time step
κ	normalizing term

Multi-Modal Posteriors

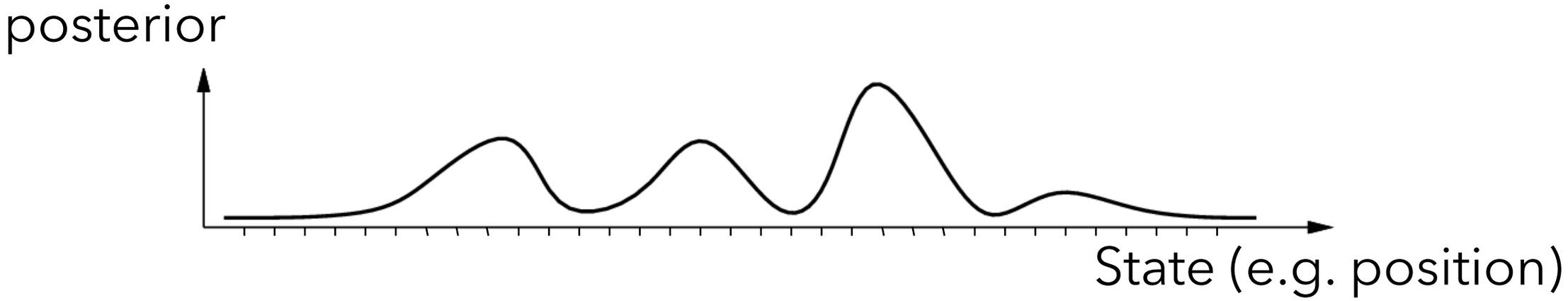
TECHNISCHE UNIVERSITÄT DARMSTADT

- Measurement clutter in natural images causes likelihood functions to have multiple, local maxima.
  - In a particular frame, the observation may be poor so that there are multiple promising looking locations.
  - We cannot resolve these ambiguities until we have seen more data (additional frames).



- To do that, we have to allow for the posterior at each frame to be multi-modal.

Multi-Modal Posteriors

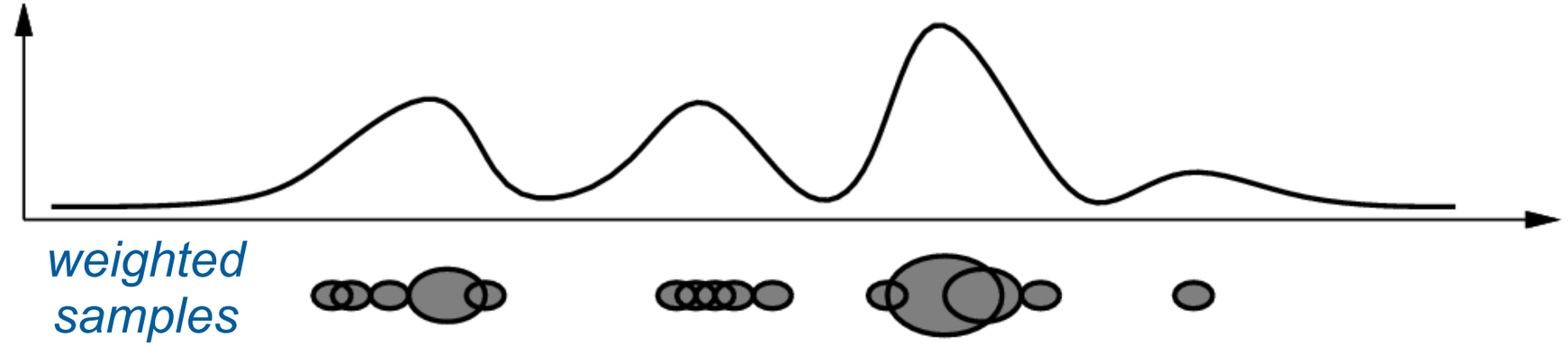


- How can we represent the posterior at each time step in a flexible way that allows for:
  - Multiple modes
    - To encode multiple promising locations.
  - Varying number of modes
    - Modes may appear and disappear again when they are ruled out.

Non-Parametric Approximation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Idea: Sample at irregular intervals and (optionally) weigh samples.

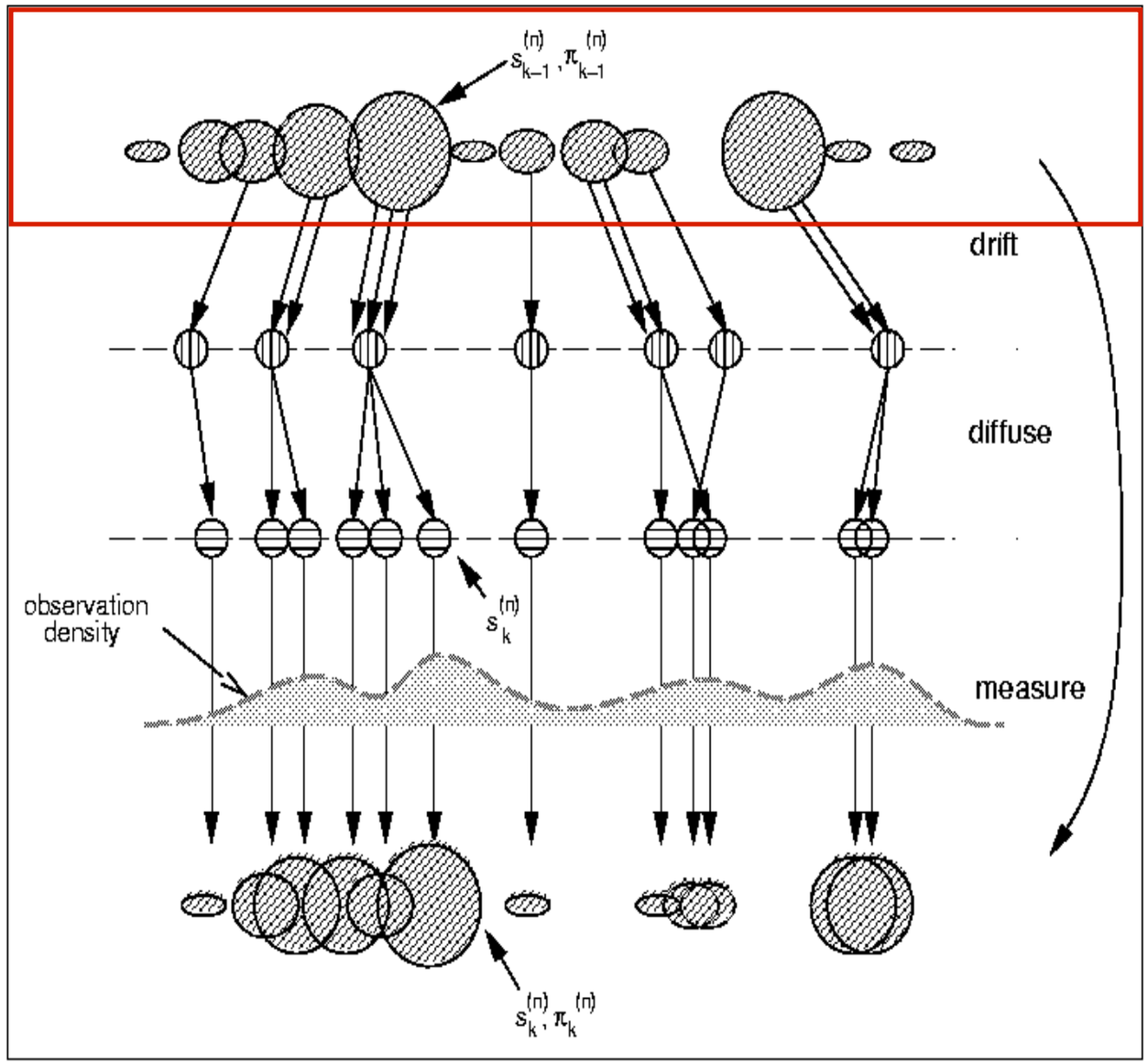


- Weighted samples: $S = \{(\boldsymbol{x}^{(i)}, w^{(i)});\ i = 1, \ldots, N\}$
- Normalized weights $\sum_{i=1}^{N} w^{(i)} = 1$

Particle Filter

TECHNISCHE UNIVERSITÄT DARMSTADT

Posterior $p(\boldsymbol{x}_{k-1}|\mathbf{Z}_{k-1})$



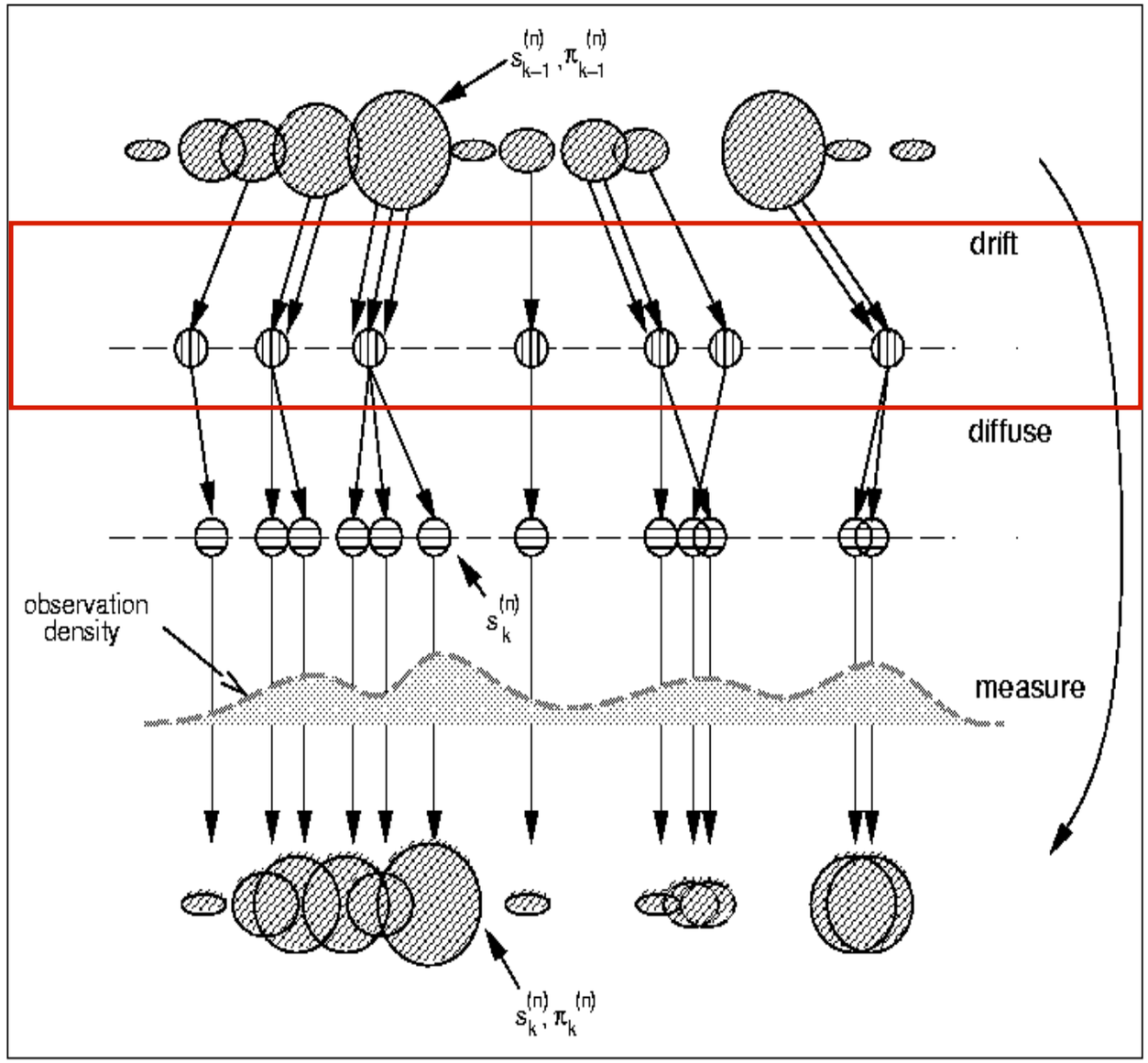
Isard & Blake ‘96

Particle Filter

TECHNISCHE UNIVERSITÄT DARMSTADT

Posterior $p(\boldsymbol{x}_{k-1}|\mathbf{Z}_{k-1})$

↓ **resample**



Isard & Blake ‘96

Particle Filter

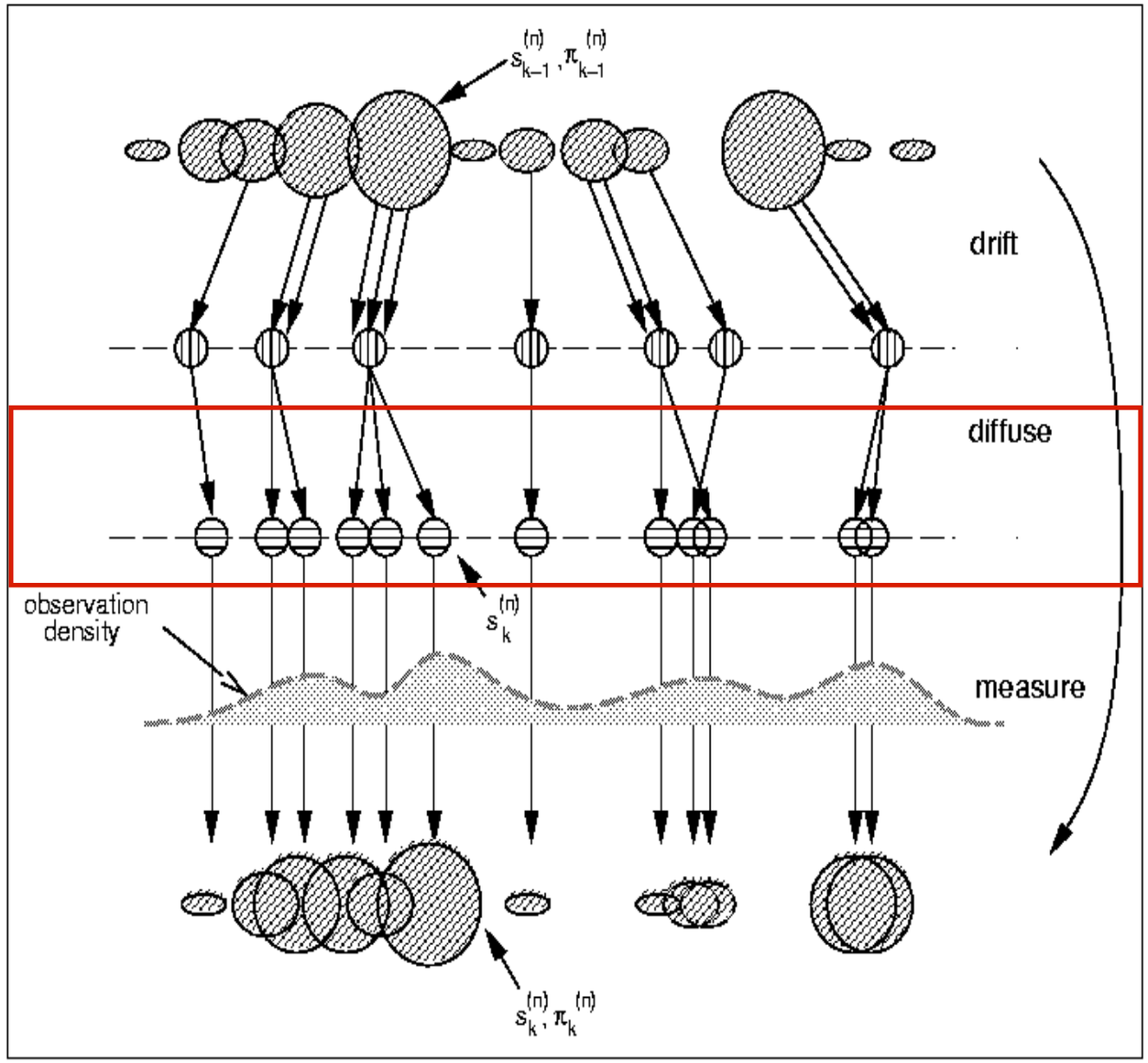
TECHNISCHE UNIVERSITÄT DARMSTADT

Posterior $p(\boldsymbol{x}_{k-1}|\mathbf{Z}_{k-1})$

resample

Apply temporal dynamics

$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$



Isard & Blake ‘96

Particle Filter

TECHNISCHE UNIVERSITÄT DARMSTADT

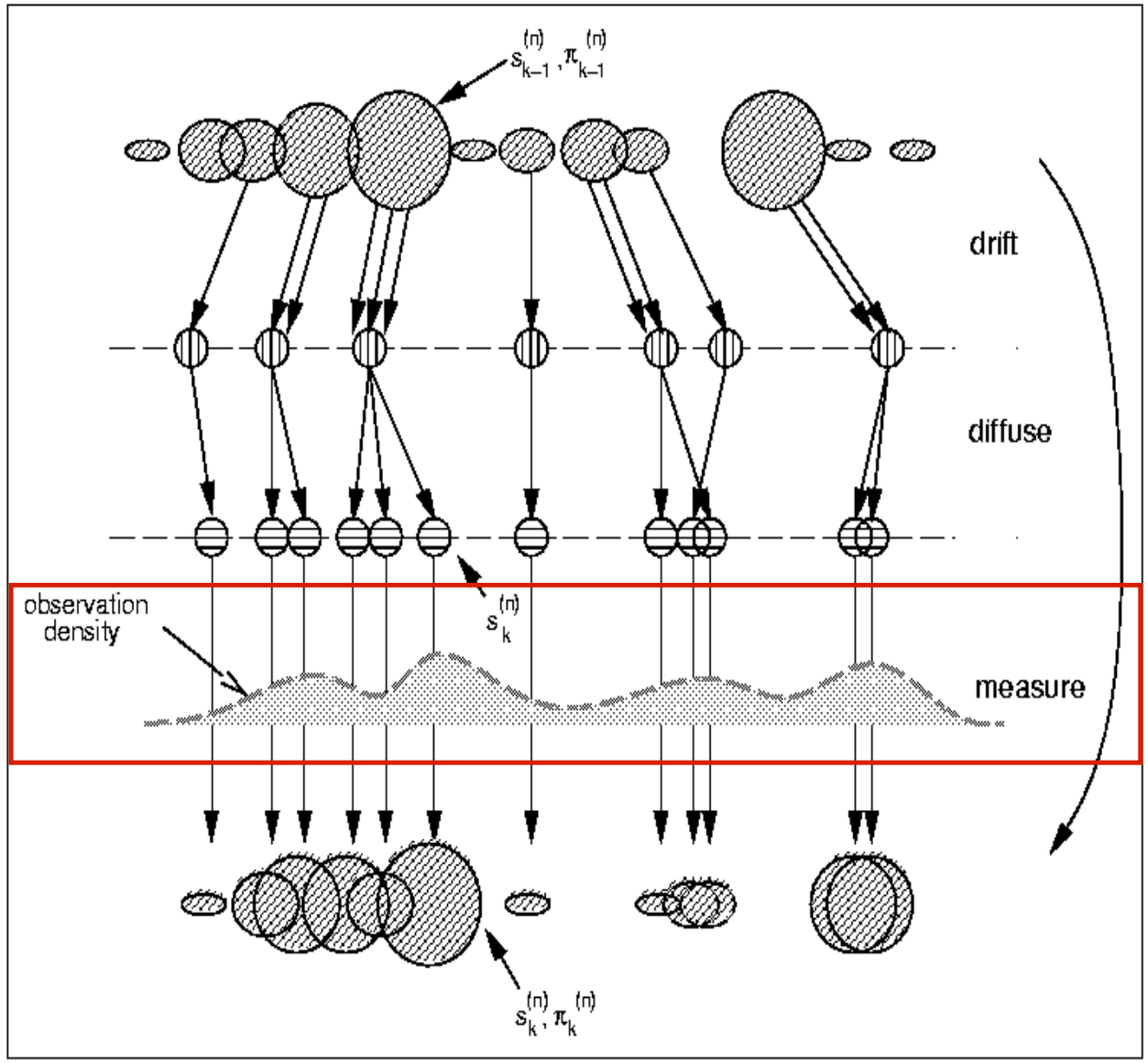
Posterior $p(\boldsymbol{x}_{k-1}|\mathbf{Z}_{k-1})$

↓ **resample**

Apply temporal dynamics $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$

↓ **reweight**

Likelihood $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$



Isard & Blake ‘96

Particle Filter

TECHNISCHE UNIVERSITÄT DARMSTADT

Posterior $p(\boldsymbol{x}_{k-1}|\mathbf{Z}_{k-1})$

↓ **resample**

Apply temporal dynamics

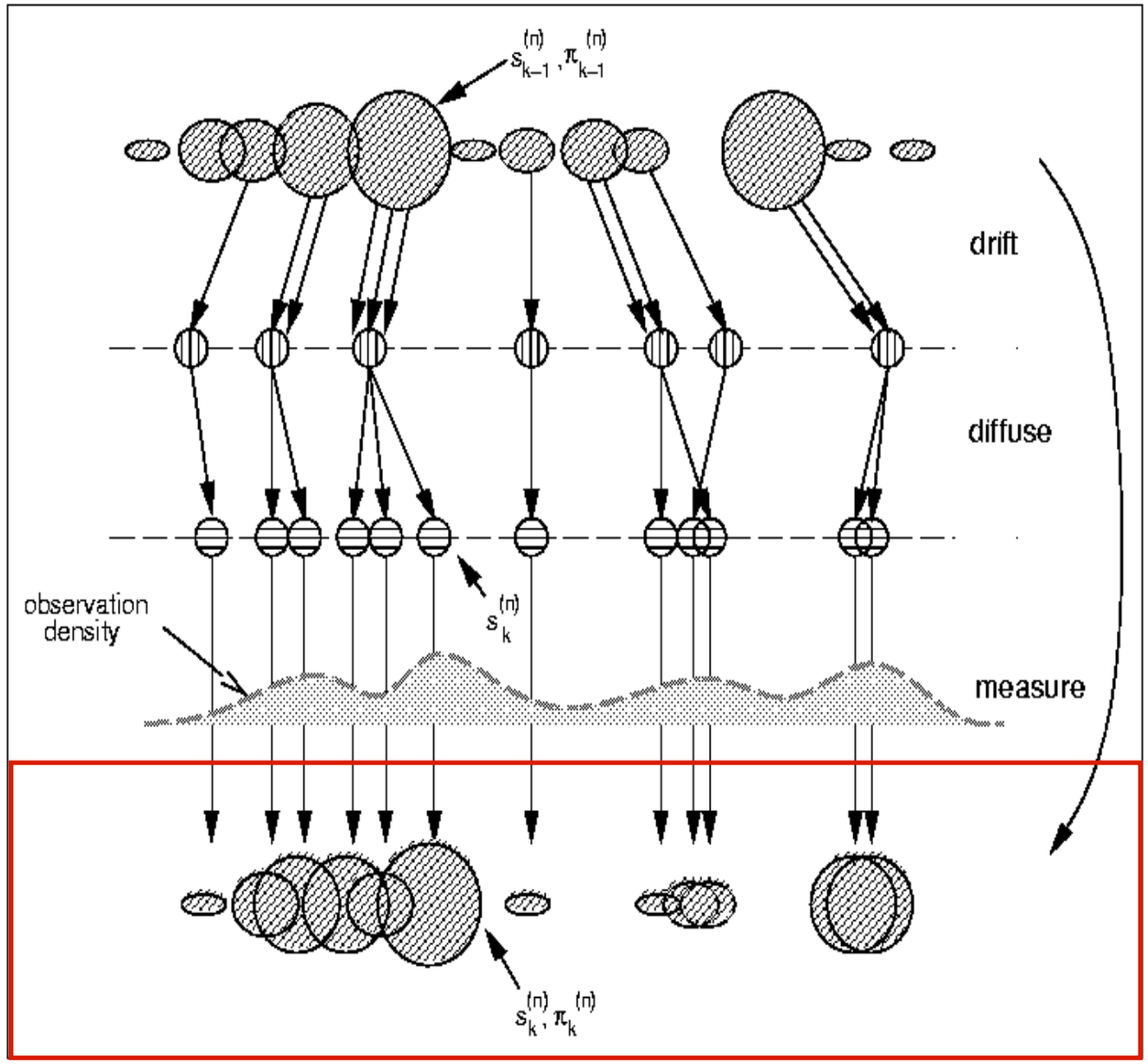
$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$

↓ **reweight**

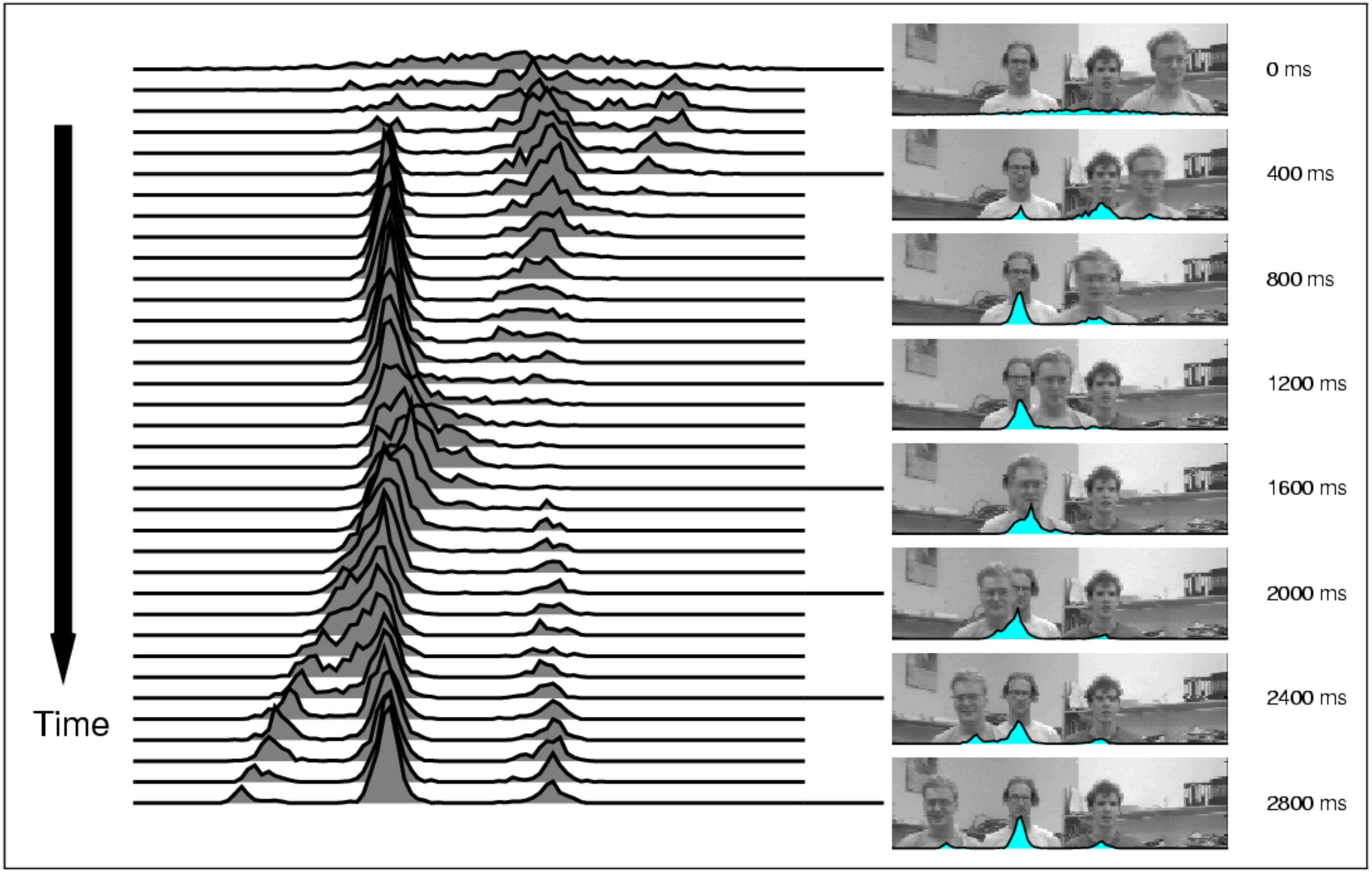
Likelihood $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$

↓ **normalize**

Posterior $p(\boldsymbol{x}_k|\mathbf{Z}_k)$



Isard & Blake '96



[Michael Isard]

Some Properties

TECHNISCHE UNIVERSITÄT DARMSTADT

- It can be shown that in the infinite particle limit this converges to the correct solution [Isard & Blake].
- In practice, we of course want to use a finite number.
  - In low-dimensional spaces we might only need 100s of particles for the procedure to work well.
  - In high-dimensional spaces sometimes 1000s or even 10000s particles are needed.
- There are many variants of this basic procedure, some of which are more efficient (e.g. need fewer particles)
  - See e.g.: Arnaud Doucet, Simon Godsill, Christophe Andrieu: On sequential Monte Carlo sampling methods for Bayesian filtering.

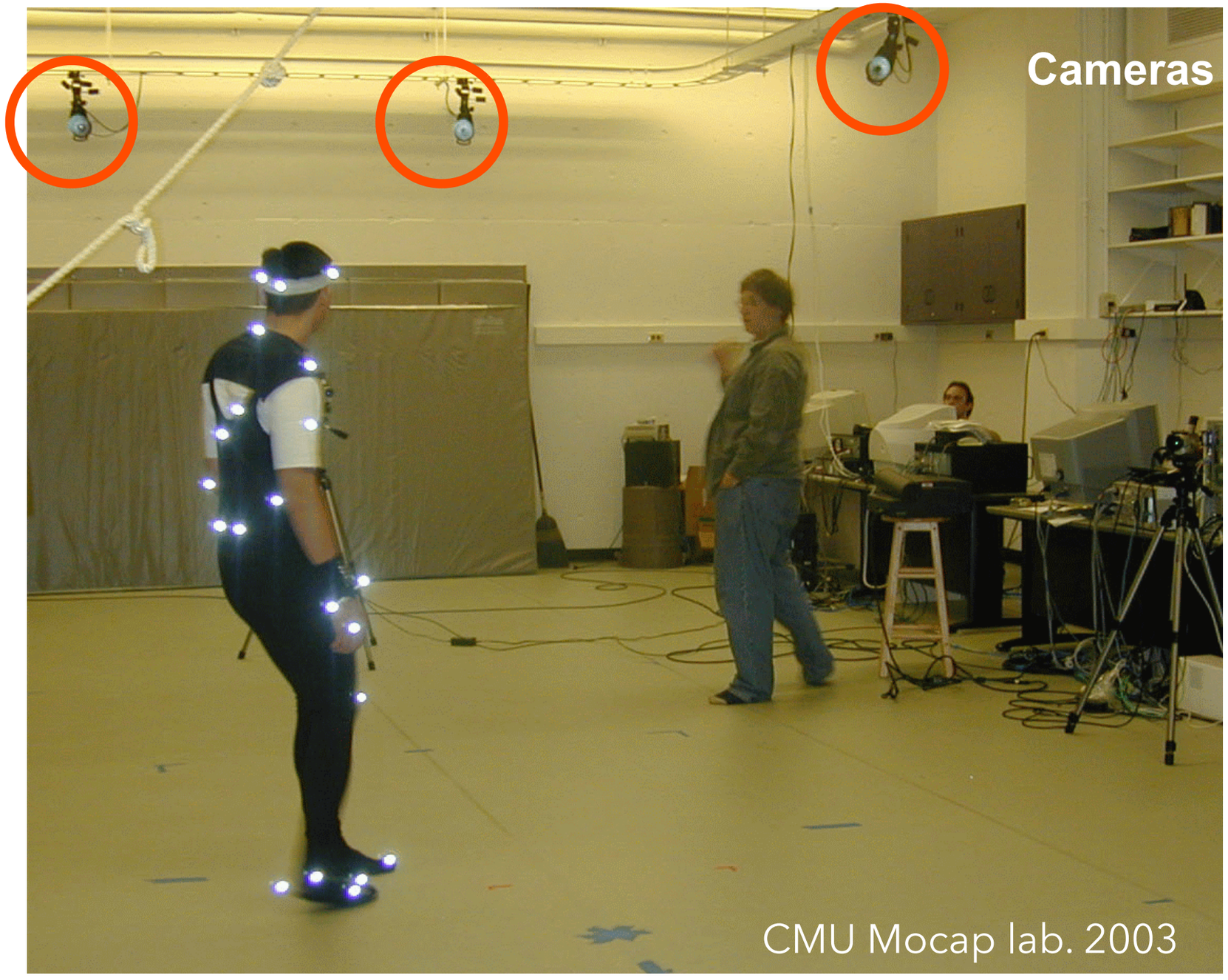
Outlook: Articulated Tracking

TECHNISCHE UNIVERSITÄT DARMSTADT

- So far, we have seen relatively simple tracking applications where the state vector is often just an image position.
  - Can we do more equipped with the tools we have learned?
- Articulated tracking is one such advanced example:
  - Here, we want to track the position and the **configuration** of an articulated object, i.e. an object that consists of several parts that can move (somewhat) independently.
  - Most prominent example: Human tracking.

Motion Capture - “Mocap”

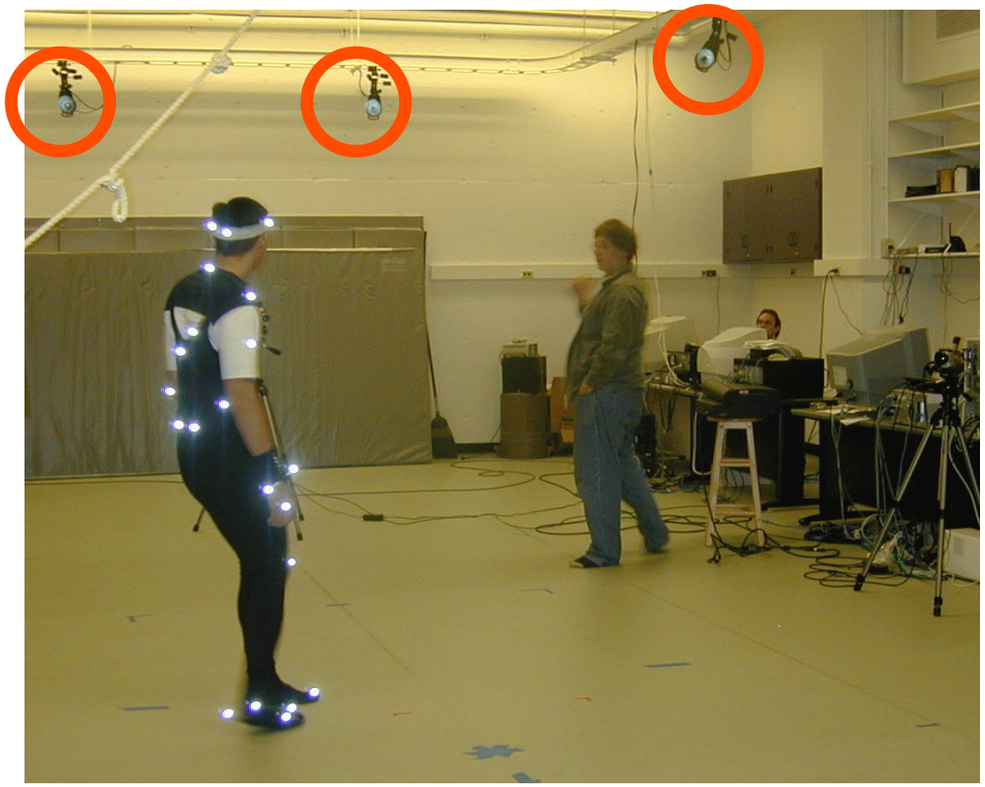
TECHNISCHE UNIVERSITÄT DARMSTADT



Motion Capture - “Mocap”

TECHNISCHE UNIVERSITÄT DARMSTADT

- Motion capture of human subjects has a very wide range of applications.
- Most prominent:
  Entertainment industry
  - Capture humans to animate digital characters.
- But also:
  - Health (e.g. analysis of gait)
  - Sports



Examples

- Tom Hanks in “The Polar Express”

From IGN

Examples

TECHNISCHE UNIVERSITÄT DARMSTADT

- Feature film based on Mocap: Beowolf

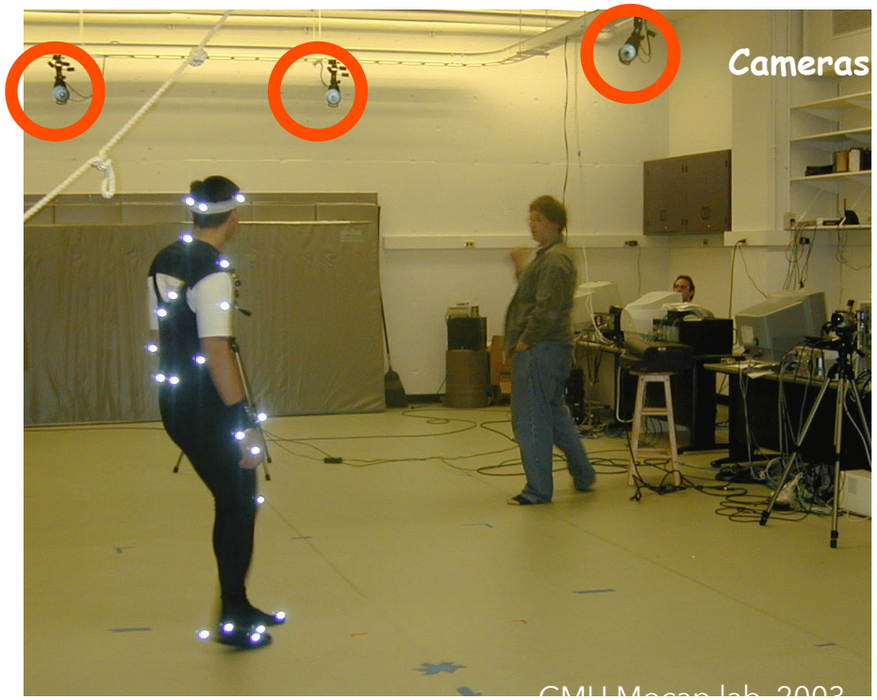
From NY Times

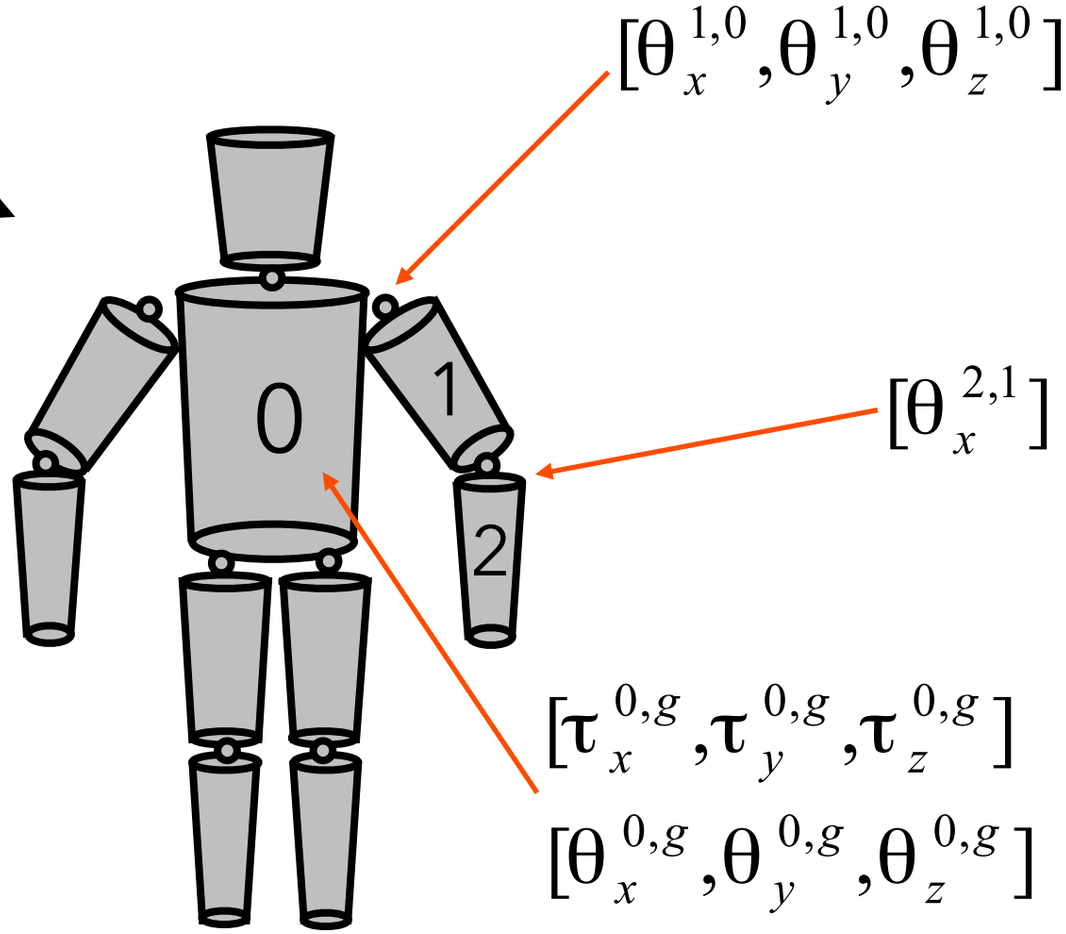
Marker-Based Mocap

TECHNISCHE UNIVERSITÄT DARMSTADT

- Put many reflective markers on the person to be tracked.
  - Record person from different viewpoints simultaneously.
  - Triangulate 3D position of each marker (like in stereo).
  - Convert marker positions into a body model.
- There are robust (but expensive) commercial systems available: e.g. from Vicon.

Body Model: Kinematic Tree





- Tree rooted at torso
  - Global position and orientation
- Other body parts
  - Relative orientation to “parent”

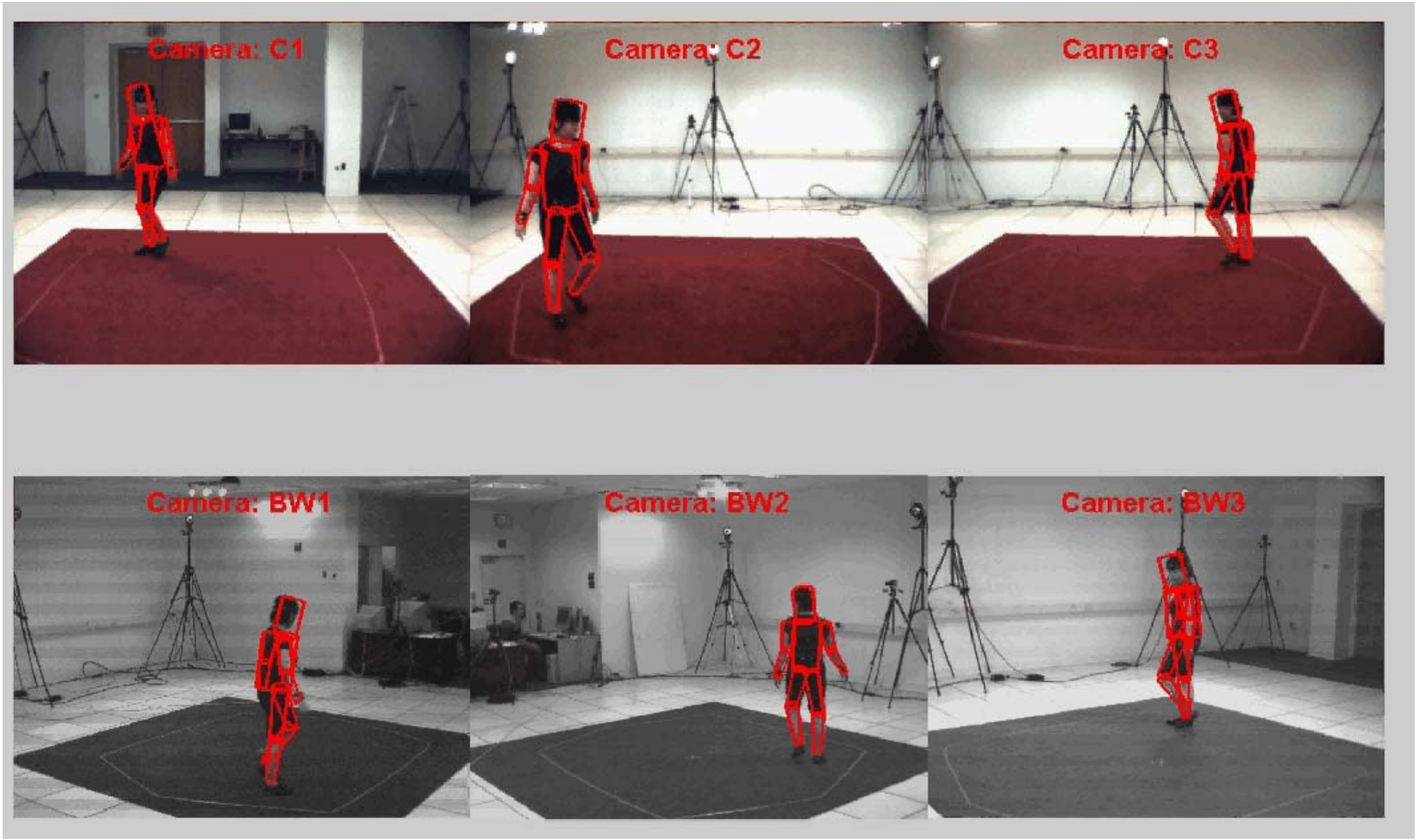
Marker-Less Mocap

TECHNISCHE UNIVERSITÄT DARMSTADT

- Problems with marker-based mocap:
  - Tedious. Many markers have to be placed, they often fall off, etc.
  - Only works in lab setting. Cannot easily be applied in everyday environments.
- Marker-less mocap:
  - Use tracking techniques from computer vision to perform mocap without the need of markers.
  - Not as robust yet as marker-based tracking, but getting there.
- Ultimate goal: only 1 camera
  - This is very difficult.

Where do we want to go (for now)?

TECHNISCHE UNIVERSITÄT DARMSTADT



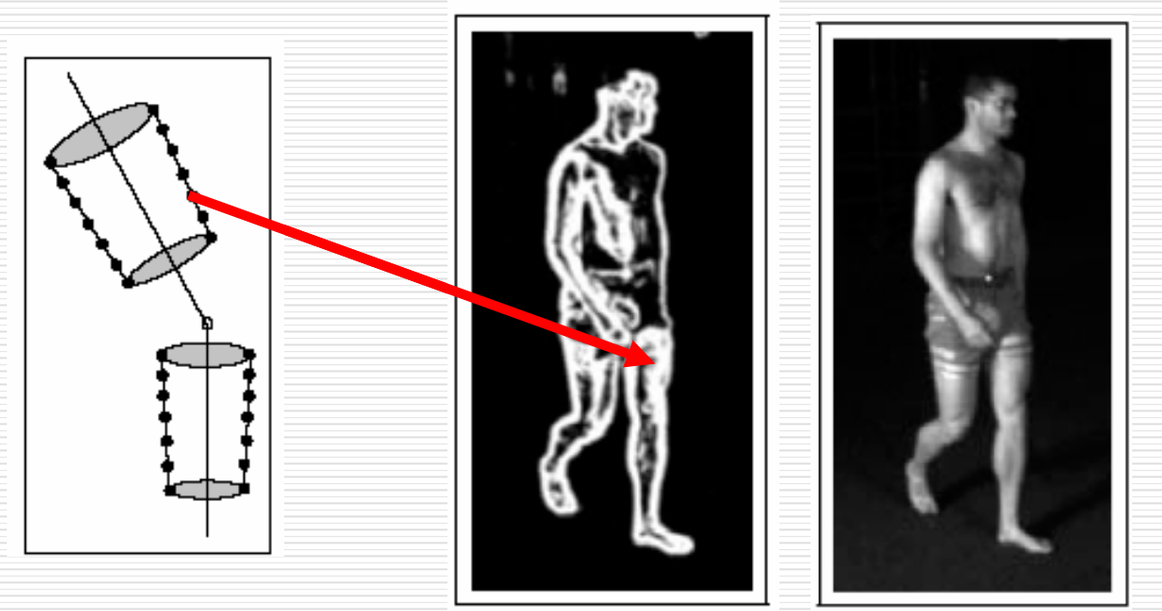
From Leonid Sigal

Particle Filter Approach

TECHNISCHE UNIVERSITÄT DARMSTADT

- In order to apply the particle filter to this problem, we need to define an observation likelihood and a dynamic model.
- Observation likelihood:
  - Historically based on background subtraction

p(bg pixel | limb location and orientation)



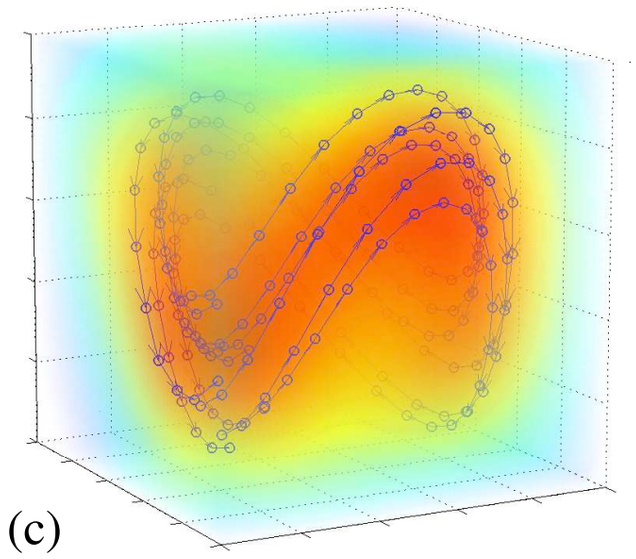
p(edge filter response | limb edge location and orientation)

From Leonid Sigal after [Deutscher et al, '00]

Particle Filter Approach

TECHNISCHE UNIVERSITÄT DARMSTADT

- In order to apply the particle filter to this problem, we need to define an observation likelihood and a dynamic model.
- Dynamical model
  - Wide range from simple Gaussian priors to complex motion models for walking, golf swings, etc.

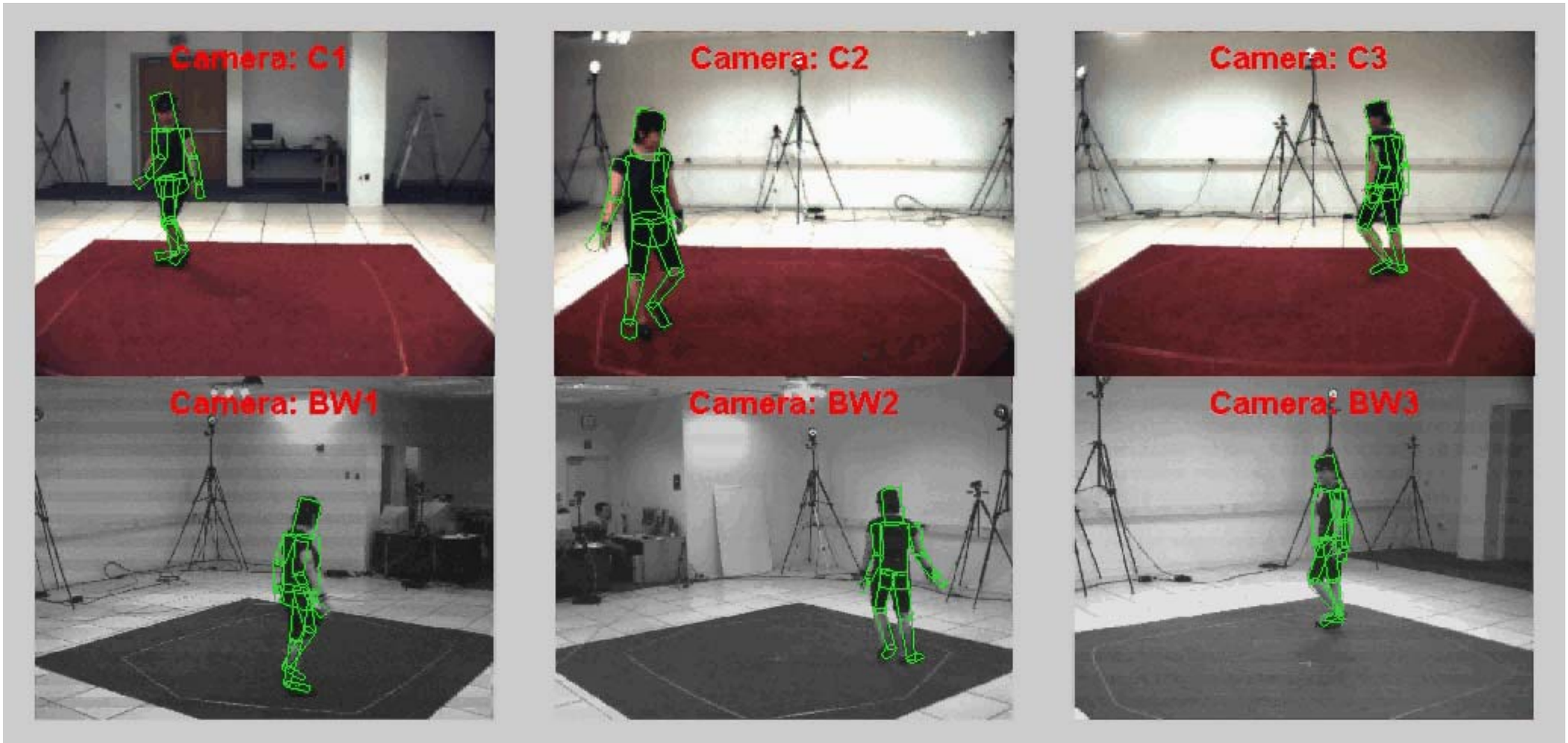


Model of a human walking cycle from [Urtasun et al. ‘06]

Some Results

TECHNISCHE UNIVERSITÄT DARMSTADT

- (Annealed) Particle filtering with a walking prior

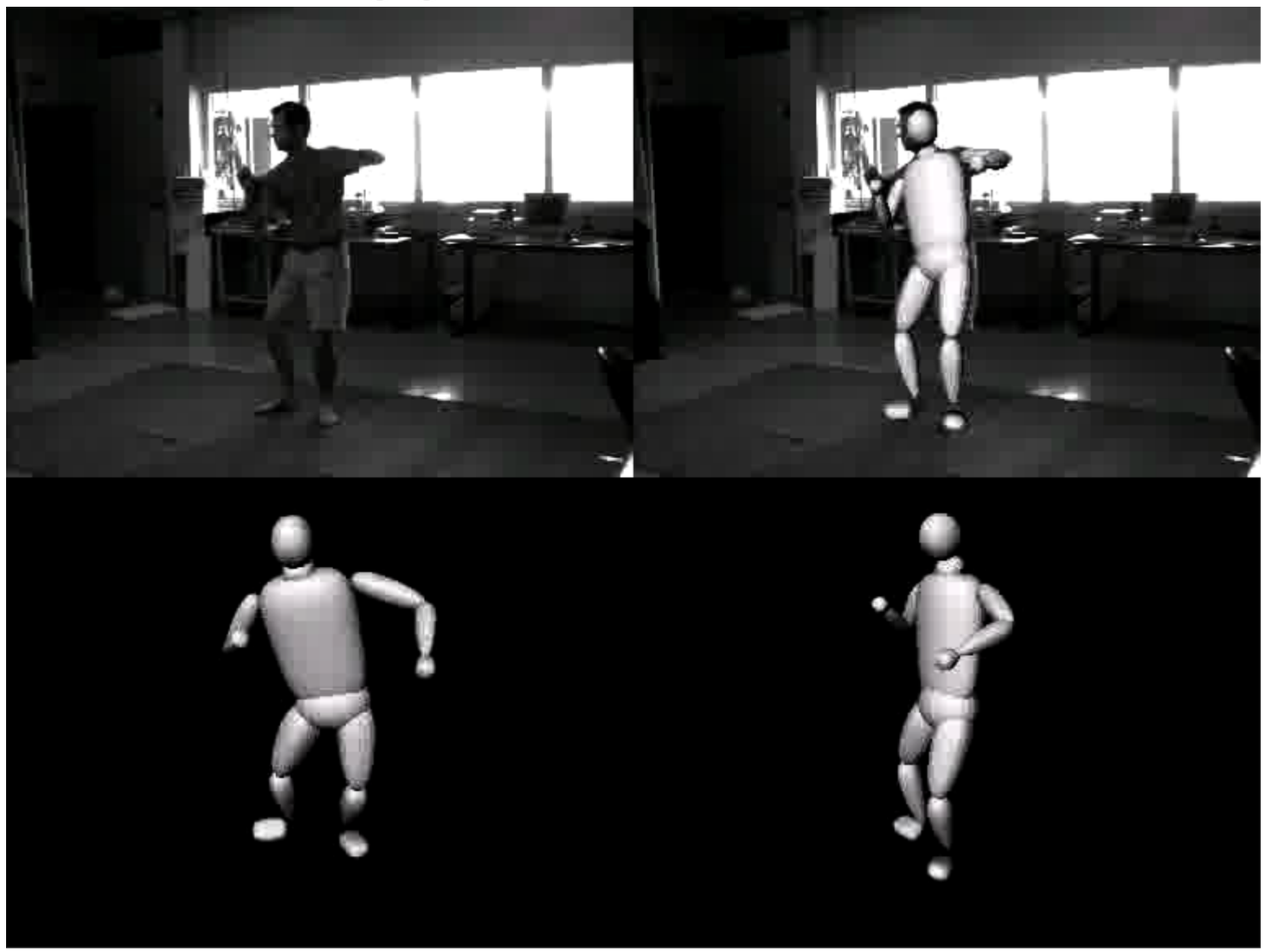


From Leonid Sigal after [Deutscher et al, '00]

Some Results

TECHNISCHE UNIVERSITÄT DARMSTADT

- Stochastic search with kinematic jump sampling [Sminchisescu & Triggs, '03]:



Challenges

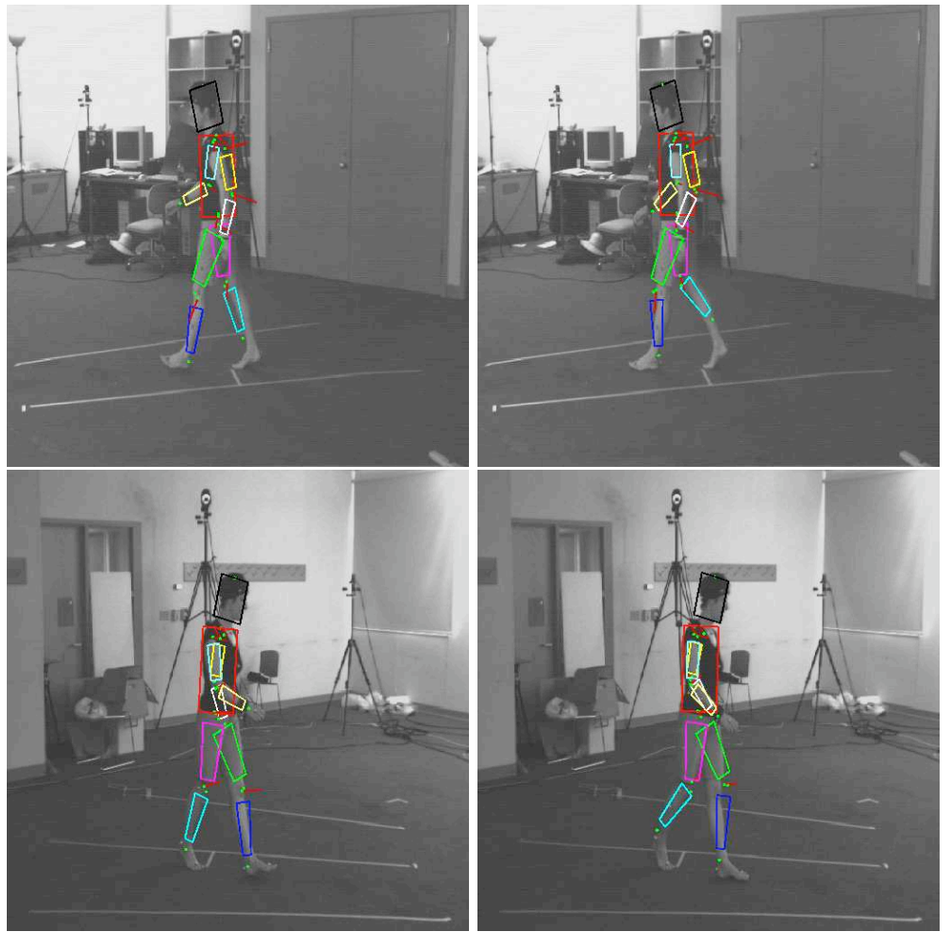
TECHNISCHE UNIVERSITÄT DARMSTADT

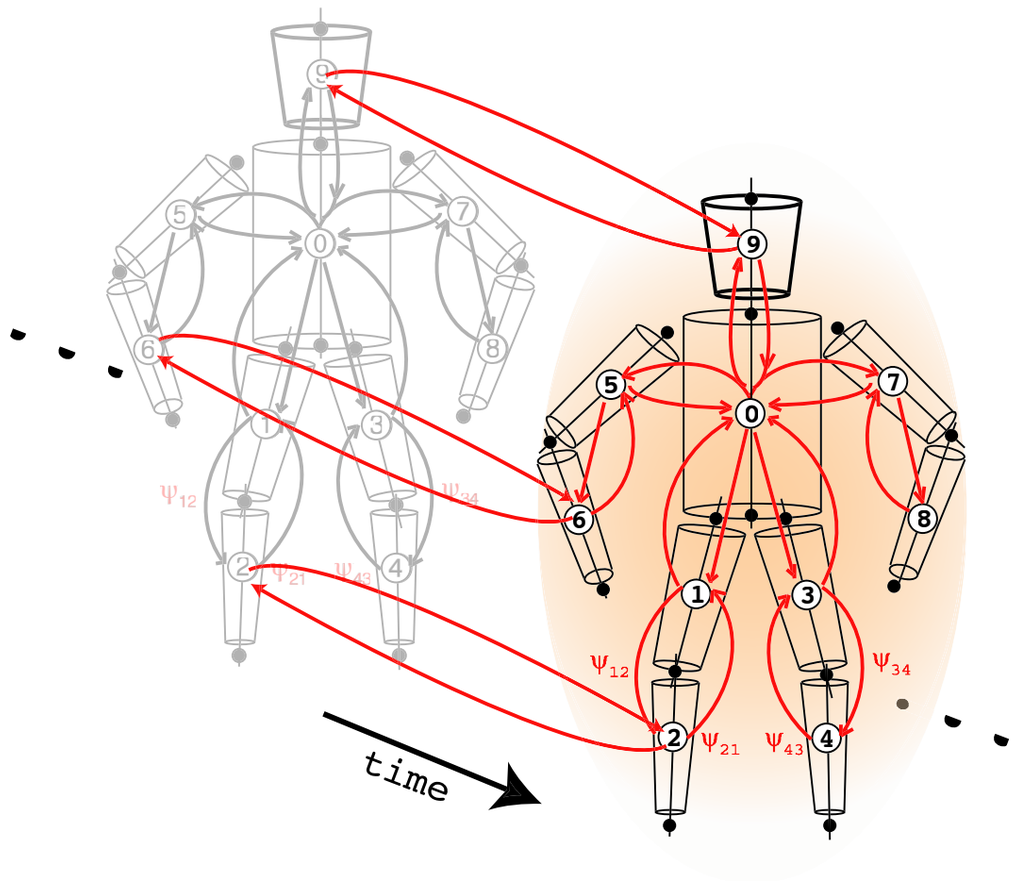
- Yet, the human tracking problem is far from being solved.
- There are many difficult challenges:
  - Self-occlusion
  - Occlusion by other objects or other people
  - Fast motions
  - Complex and cluttered backgrounds
  - Etc.

Current Approaches

TECHNISCHE UNIVERSITÄT DARMSTADT

- Graphical models and belief propagation [Sigal et al.]





- Each part is represented using particles
- Avoids exponentially large state space

Current Approaches

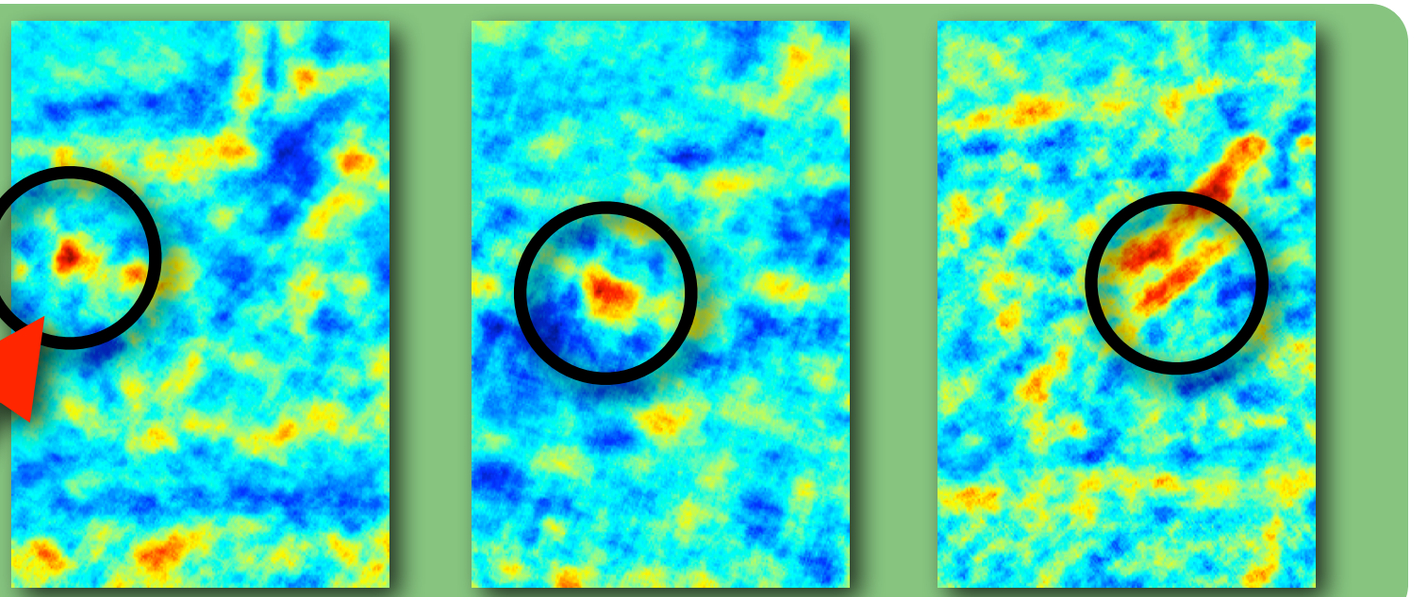
TECHNISCHE UNIVERSITÄT DARMSTADT

- Models of human limbs based on part detectors

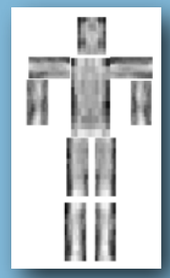
Input image

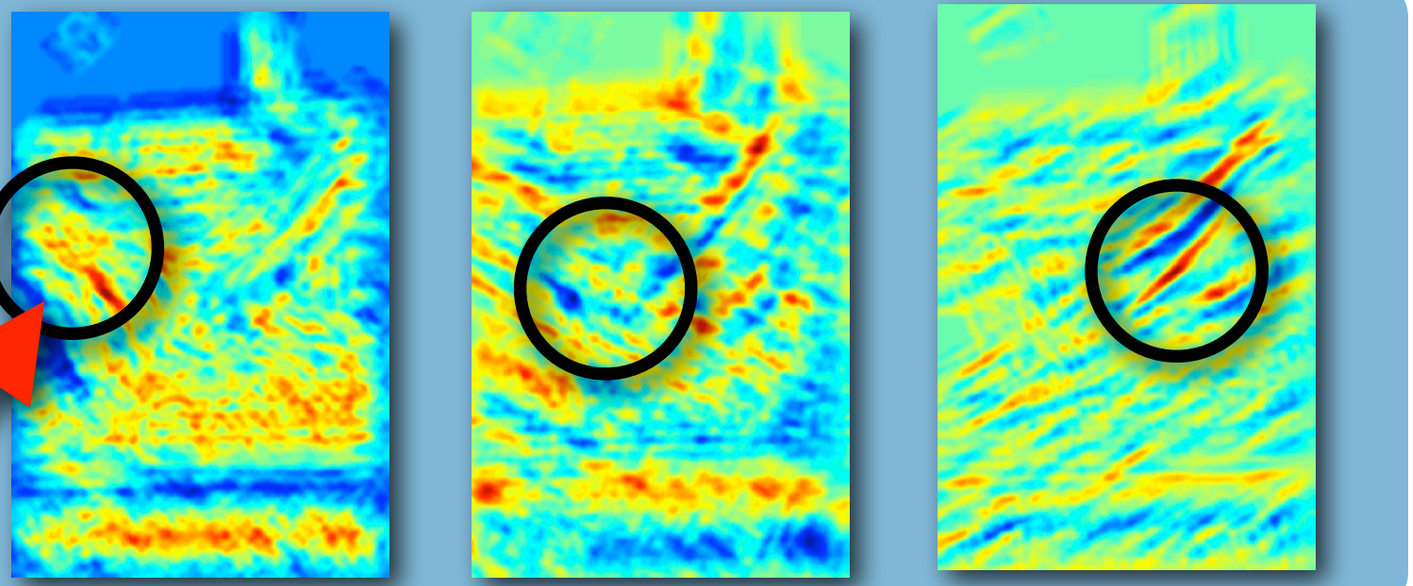
Head Torso Upper leg

part detector



part template





Multi-People Tracking

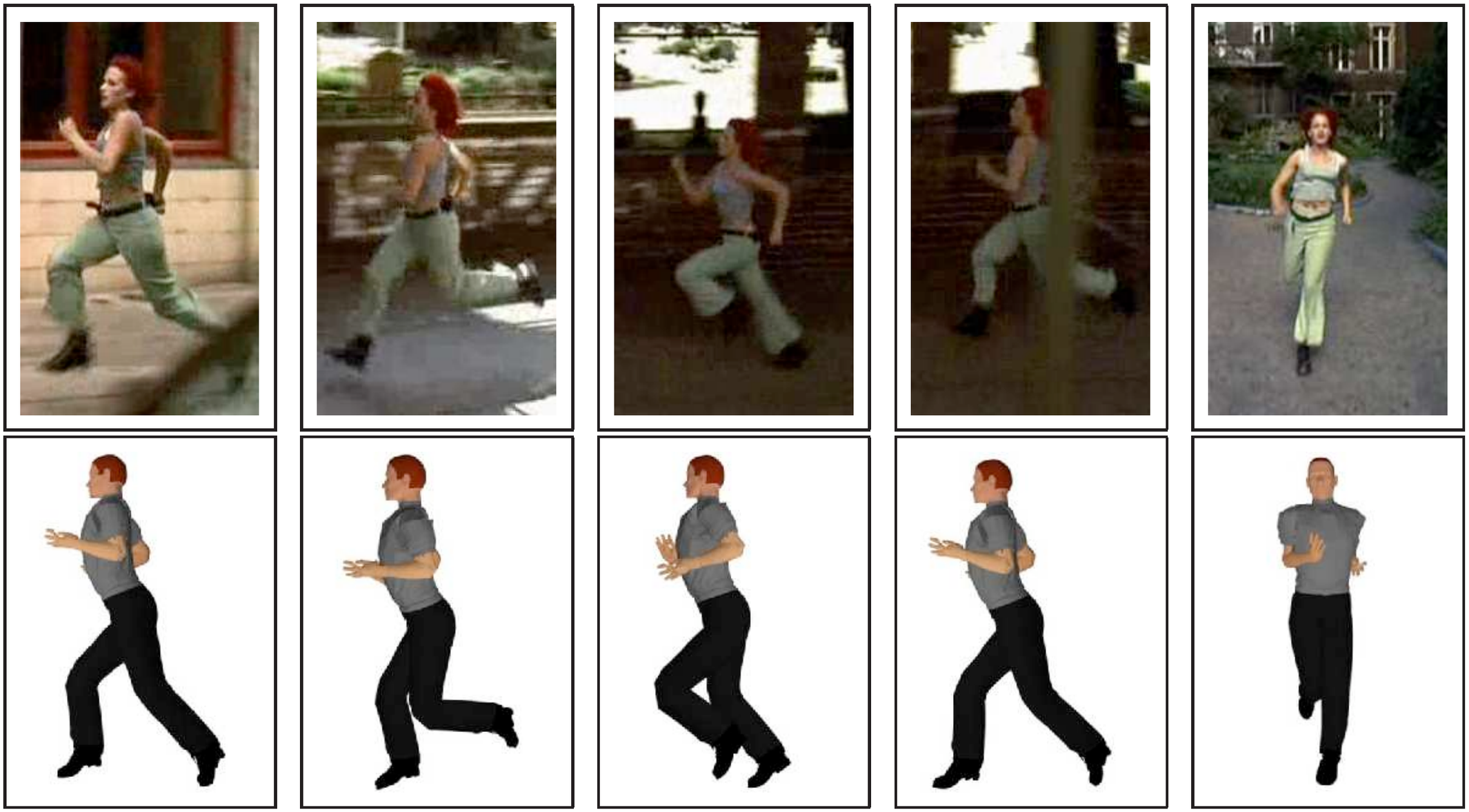
TECHNISCHE UNIVERSITÄT DARMSTADT

[Andriluka et al. 2010]

Pose Prediction from a Single Frame

TECHNISCHE UNIVERSITÄT DARMSTADT

- Even more challenging...



[Sminchisescu et al., '07]

Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

- Particle filtering is a very general tool for temporal inference that we can exploit for tracking.
  - Nonetheless, it applies in a variety of other applications as well.
  - It has problems in high-dimensional spaces, however, but there are a number of variants that alleviate some of these issues.
- Human tracking ("Marker-less mocap") can be performed using particle filtering:
  - Wide range of applications, especially in entertainment.
  - Only a small part of the problems are solved to date.

(Semantic) Segmentation

11.04.2014

TECHNISCHE UNIVERSITÄT DARMSTADT

visual inference

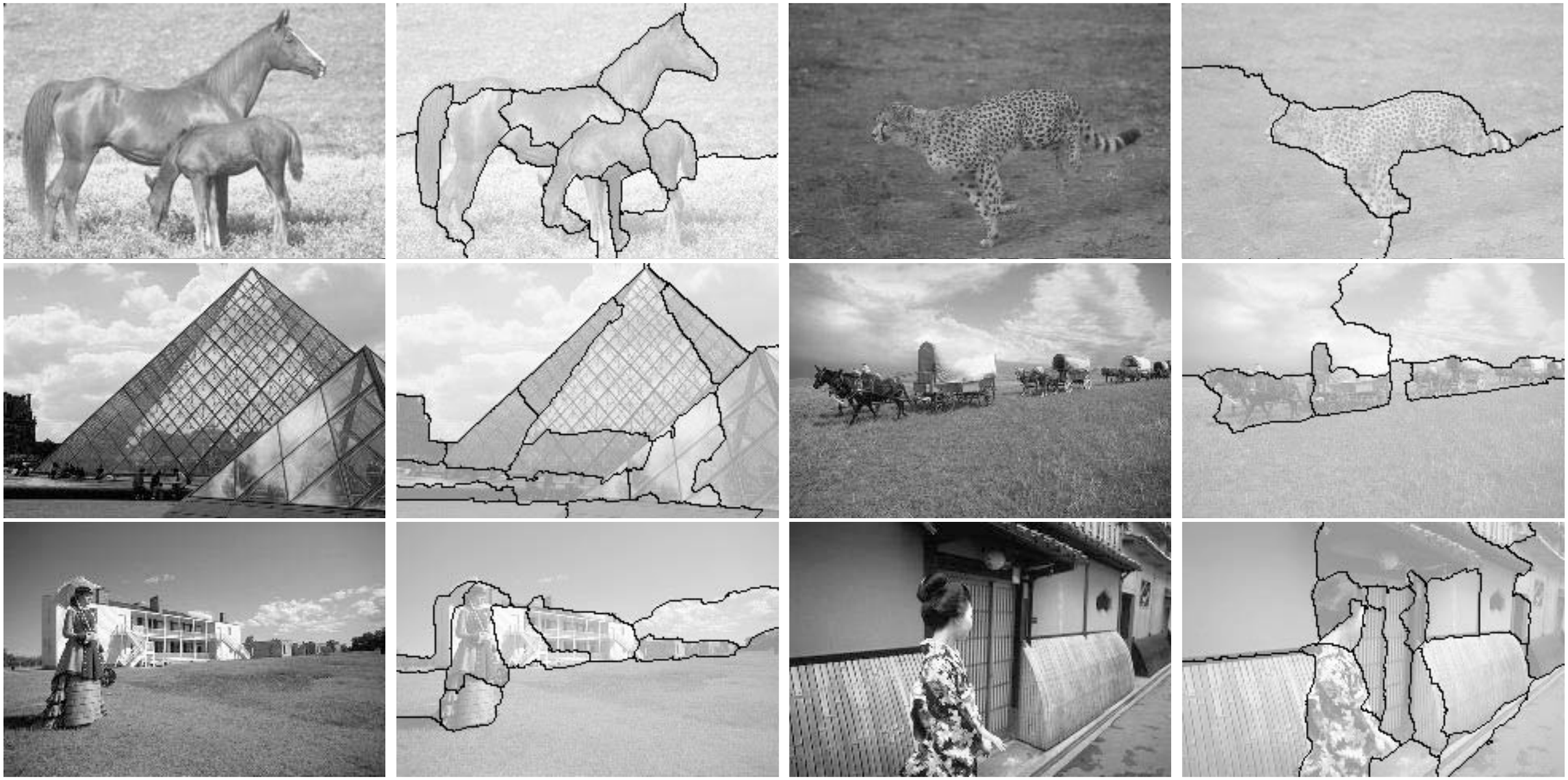
Segmentation Revisited (from CV1)

TECHNISCHE UNIVERSITÄT DARMSTADT

- What do we mean by segmentation and why do we need it?
  - Segmentation can roughly be described as the grouping of similar information in an image.
  - Instead of having to work with all the pixels, a segmentation allows us to work with a much more compact representation.
- This is useful in practice, because this compact representation can make it easier to carry out certain tasks.
  - Scene understanding, object recognition, ...
- Sometimes, we are interested in the segmentation itself.
  - Especially in medical image analysis (e.g., segmenting out a tumor)

Some Examples

TECHNISCHE UNIVERSITÄT DARMSTADT



[Ren & Malik, 03]

Figure-Ground Separation

TECHNISCHE UNIVERSITÄT DARMSTADT

- One very useful way of thinking about segmentation is that it enables the separation of the figure (i.e., foreground) from the background.
- Example:

Full image

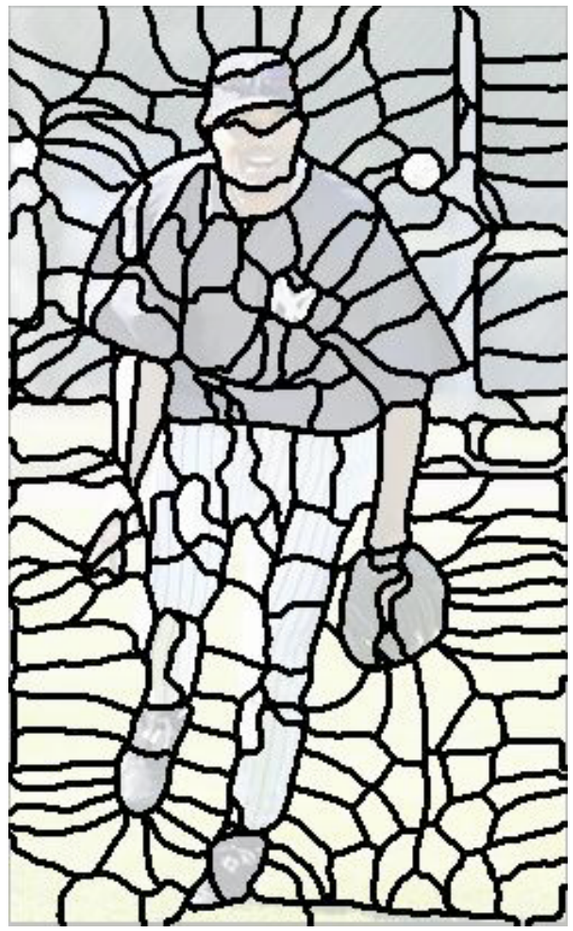
Figure (foregound) portion

[Ren et al., 05]

Superpixels

TECHNISCHE UNIVERSITÄT DARMSTADT

- Superpixels are a form of segmentation, in which the goal is to find many small segments that can substitute using the actual pixels:



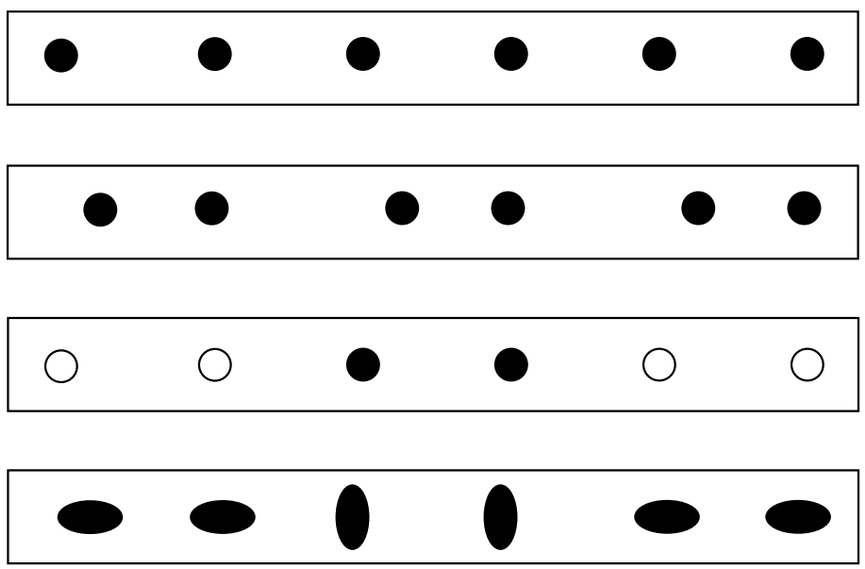
[Mori et al., 04]

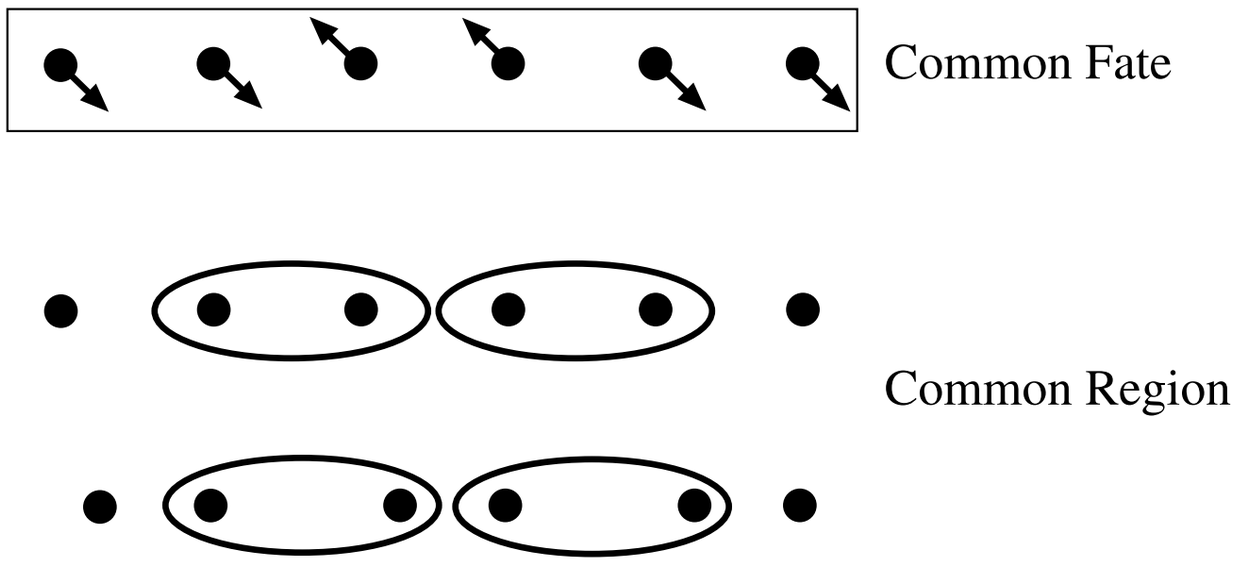
What belongs together?

TECHNISCHE UNIVERSITÄT DARMSTADT

- In order to perform image segmentation, we need to decide which parts of the image belong together.
- We can draw inspiration from various sources.
  - As before, we can try to think about what makes us humans believe parts of an image belong together.
  - Early work: Gestalt psychology in the early 20th century.
    - Max Wertheimer was one of the leading figures.

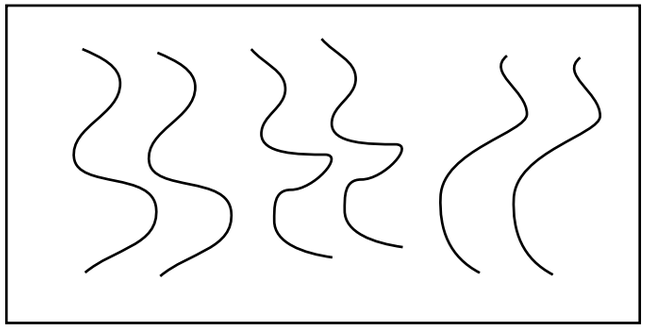
Gestalt Factors



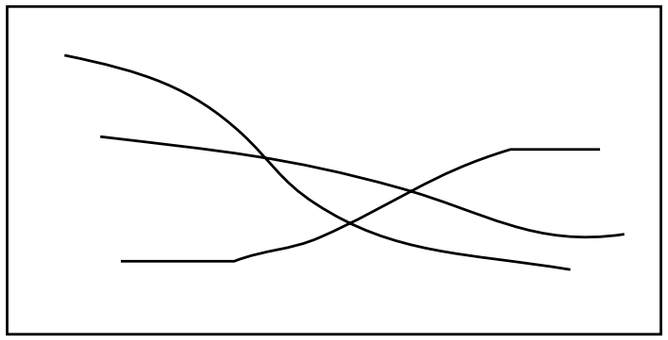


[Gordon]

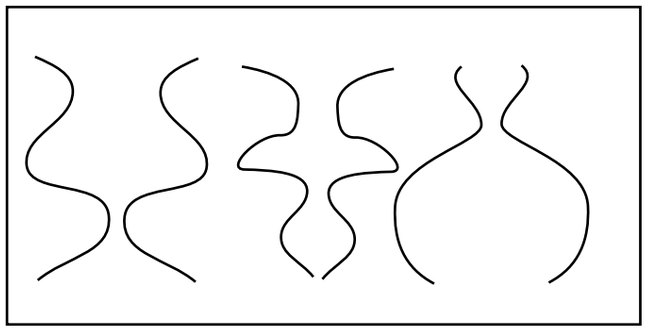
Gestalt Factors



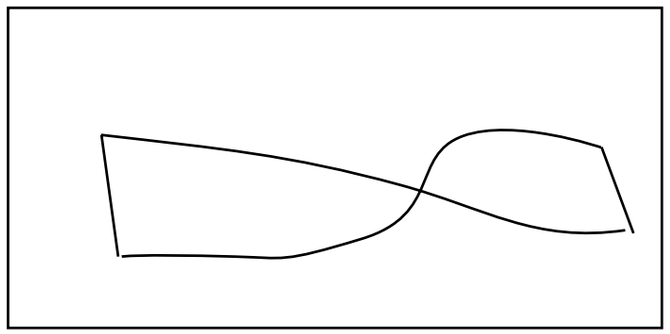
Parallelism



Continuity



Symmetry



Closure

[Gordon]

- These factors offer some insights as to what may be useful from a computer vision point of view.
- Turning them into an algorithm is difficult, however.

Conclusions so far

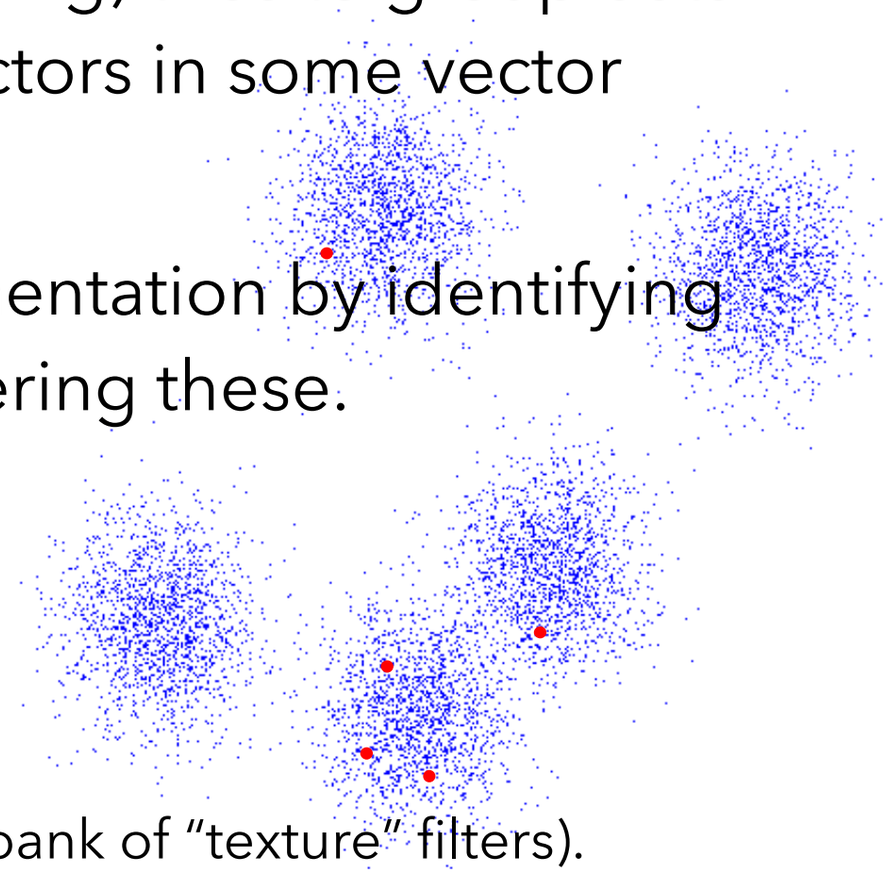
TECHNISCHE UNIVERSITÄT DARMSTADT

- From these examples & rules we can see that:
  - Segmentation is generally a quite difficult problem.
  - It is hard to even characterize what it is.
  - We humans seem to be very good at it, which suggests that it is somehow important for our visual processing.
- Most of these cases are very very challenging to implement on a computer:
  - We will only be able to do something rather simple.
  - In particular, we will not be able to solve these examples.
  - But what we can do is still useful.

Segmentation by Clustering

TECHNISCHE UNIVERSITÄT DARMSTADT

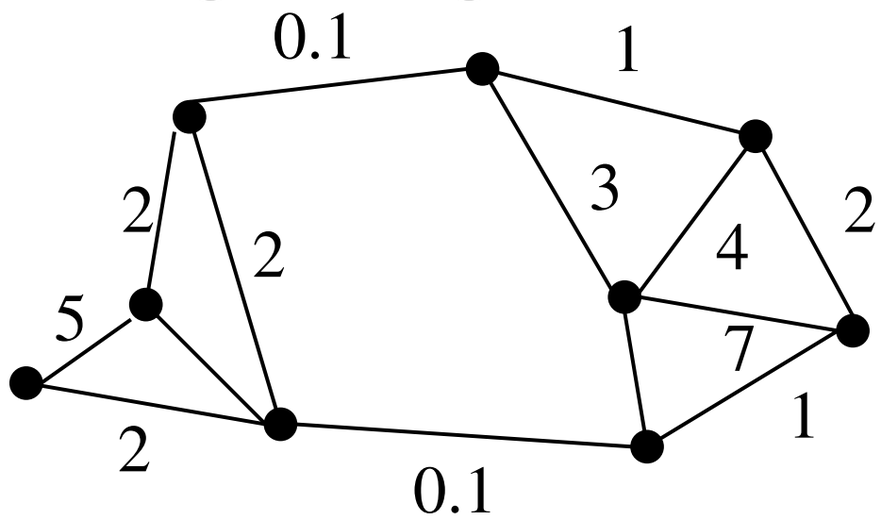
- One simple way of performing segmentation is to use clustering algorithms:
  - Clustering (a problem from machine learning) tries to group data points together. The points are usually vectors in some vector space.
  - We can apply this to the problem of segmentation by identifying each pixel with a **feature vector** and clustering these.
  - This feature vector may include:
    - The pixel's position
    - Pixel intensity or color
    - A description of the local texture (e.g. output of a bank of "texture" filters).



Graph-based Clustering

TECHNISCHE UNIVERSITÄT DARMSTADT

- Clustering can be interpreted as cutting a graph in which each node represents a pixel into pieces.
  - (Note: This graph is not a graphical model!)
- For this we define affinities between the pixels that encode how similar they are.
- These give the edge weights:



Note: Spatial arrangement is arbitrary!

Simple Affinity Criteria

TECHNISCHE UNIVERSITÄT DARMSTADT

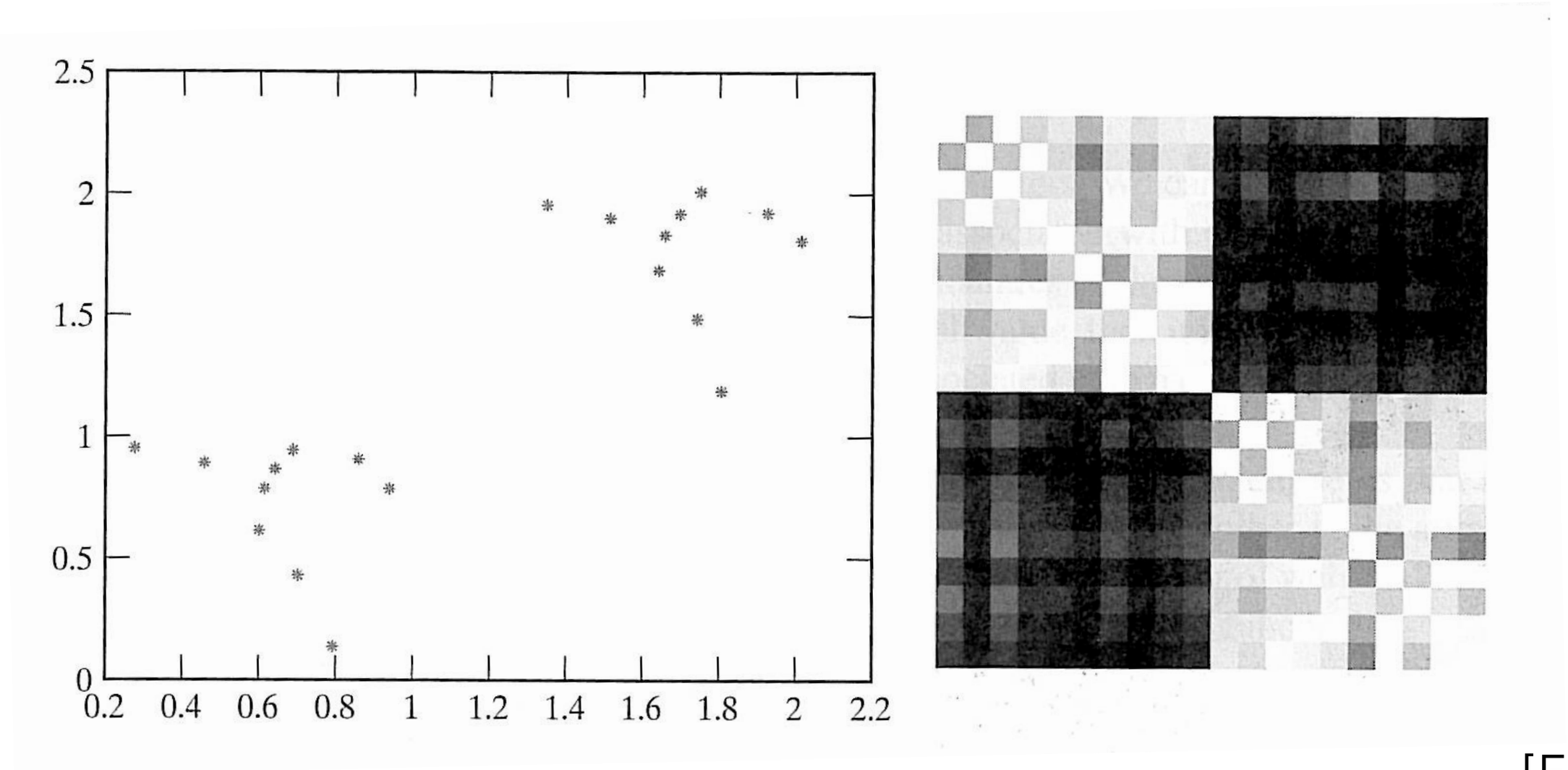
- Define affinities of pixels:
  - Affinity by distance
$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{ -||\boldsymbol{x}-\boldsymbol{y}||^2/(2\sigma_D^2)\right\}$$
  - Affinity by intensity
$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{ -(I(\boldsymbol{x})-I(\boldsymbol{y}))^2/(2\sigma_D^2)\right\}$$
  - Affinity by color
$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{ -\mathrm{dist}(c(\boldsymbol{x}),c(\boldsymbol{y}))^2/(2\sigma_D^2)\right\}$$
  - Affinity by texture
$$\mathrm{aff}(\boldsymbol{x},\boldsymbol{y}) = \exp\left\{ -||\boldsymbol{f}(\boldsymbol{x})-\boldsymbol{f}(\boldsymbol{y})||^2/(2\sigma_D^2)\right\}$$

texture descriptor

Affinity Matrix

TECHNISCHE UNIVERSITÄT DARMSTADT

- From this we can build an affinity matrix with all pairwise affinities:

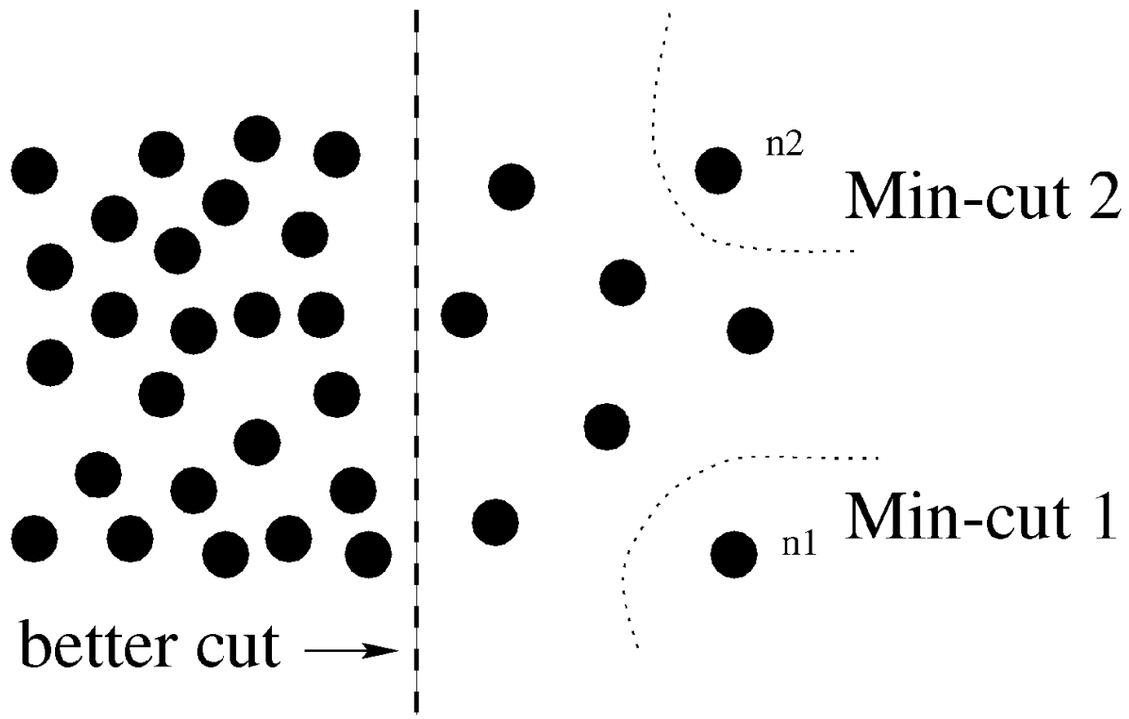


[FP]

Graph-Cut Based Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Approach segmentation by finding the min-cut on the graph:
  - *Remember when we used min-cut to perform inference in MRFs.*
  - The problem with this is that it favors small segments:



[Shi & Malik, 00]

Normalized Cuts

TECHNISCHE UNIVERSITÄT DARMSTADT

- A better approach is to normalize the cut to remove this bias:

$$\frac{\mathrm{cut}(A,B)}{\mathrm{assoc}(A)} + \frac{\mathrm{cut}(A,B)}{\mathrm{assoc}(B)}$$

  - Here $\mathrm{cut}(A,B)$ are the weights that are cut by separating the segments A and B.
  - $\mathrm{assoc}(A)$ is the weight of all edges going into segment A
- Unfortunately, optimizing this objective is NP hard:
  - But there is an efficient approximation as a generalized eigenvalue problem [Shi & Malik, 00].

Some Results

TECHNISCHE UNIVERSITÄT DARMSTADT

[Shi & Malik, 00]

Summary of Methods (so far)

TECHNISCHE UNIVERSITÄT DARMSTADT

- We have seen a whole set of different segmentation techniques:
  - Agglomerative and divisive clustering (CV1)
  - K-Means (CV1)
  - Mean Shift (CV1)
  - Graph-based or spectral clustering methods (CV1)
  - Next up: Energy-based methods & probabilistic methods
- So far, no “golden standard” has been established.

Is there a correct segmentation?

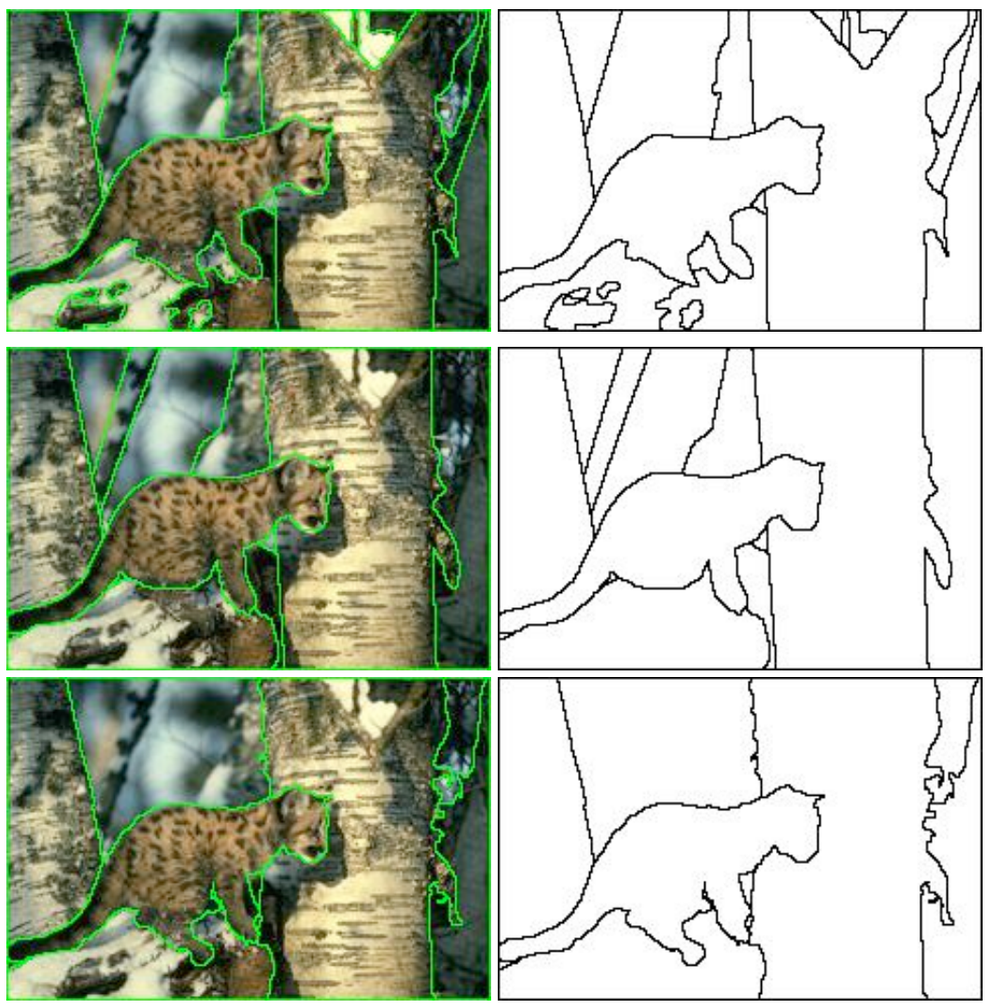
TECHNISCHE UNIVERSITÄT DARMSTADT

- **Unfortunately not!**
- A segmentation can only be right for a certain purpose...
  - Say, if we care about finding a person in an image, we may want the segmentation to separate people from the background.
  - But what if we wanted to know what cash register the person goes to? Then we want to also segment the image into the various cash registers.
  - Segmentation can mean a lot of different things!

Is there a correct segmentation?

TECHNISCHE UNIVERSITÄT DARMSTADT

- If you ask different people to segment an image, you will get many different results:



This makes it also hard to evaluate how good a

What can we hope for?

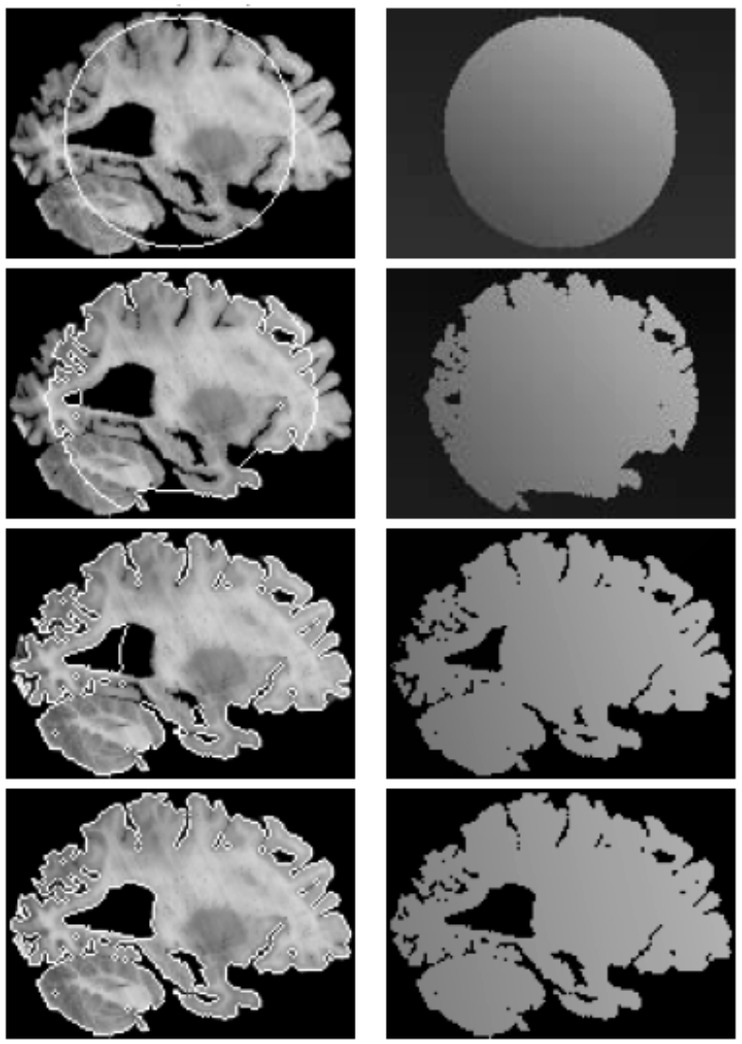
TECHNISCHE UNIVERSITÄT DARMSTADT

- We cannot hope that segmentation does all that we might want it to do.
  - Segmentation is normally “dumb” in the sense that it doesn’t know what we want to do with the result.
  - Chicken-and-egg problem: A good segmentation helps a lot with various vision tasks (e.g. object recognition), but unless we have solved this task already, we can’t hope to get the perfect segmentation.
  - Segmentation should somehow be coupled to the task we want to solve with it. How?
- Of course, all of this doesn’t mean that we should abandon it.

Segmentation with a clear goal

TECHNISCHE UNIVERSITÄT DARMSTADT

- E.g. medical image segmentation



Initialization

Final result

- Clearly defined goal in a specific domain

[Nikos Paragios]

Application: Photomontage





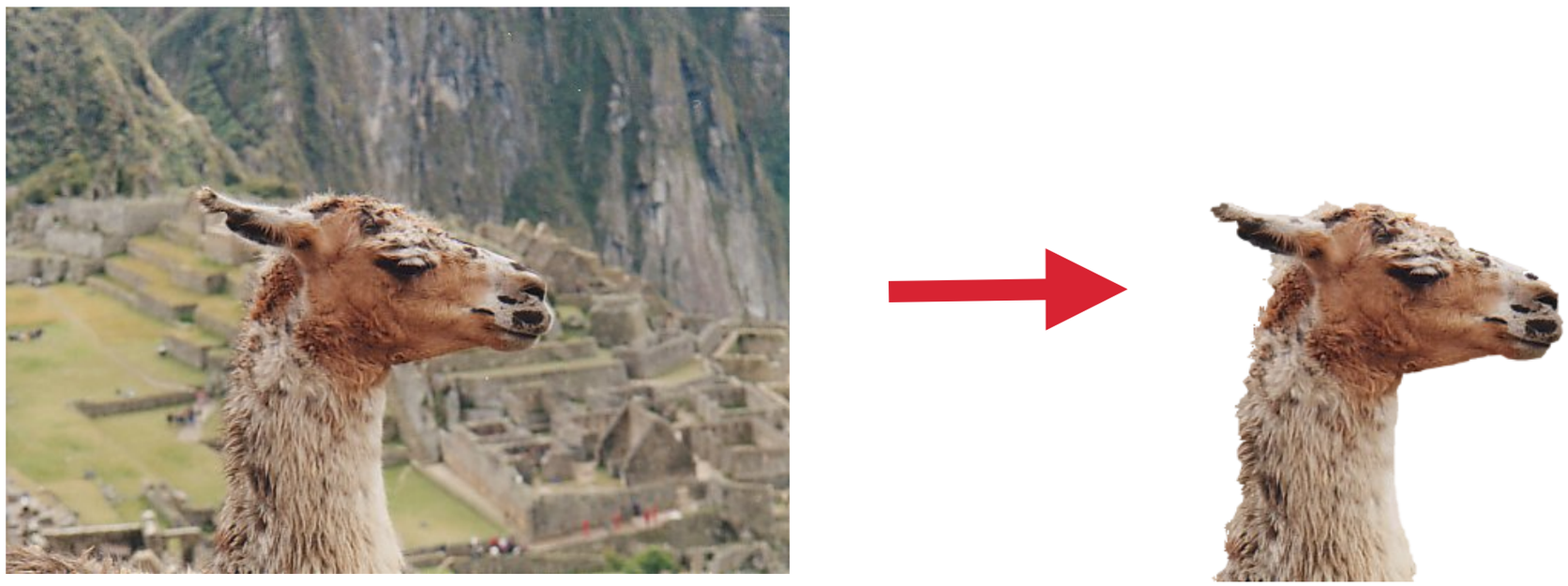
- Exploit user interaction to define a goal

[Carsten Rother]

Photomontage

TECHNISCHE UNIVERSITÄT DARMSTADT

- To assemble several photos into a montage...
  - need to separate the object of interest from the background



- Automatic segmentation?
  - How would the algorithm know what I am interested in?

[Carsten Rother]

Interactive Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Basic idea:
  - Let the user annotate some examples of foreground & background

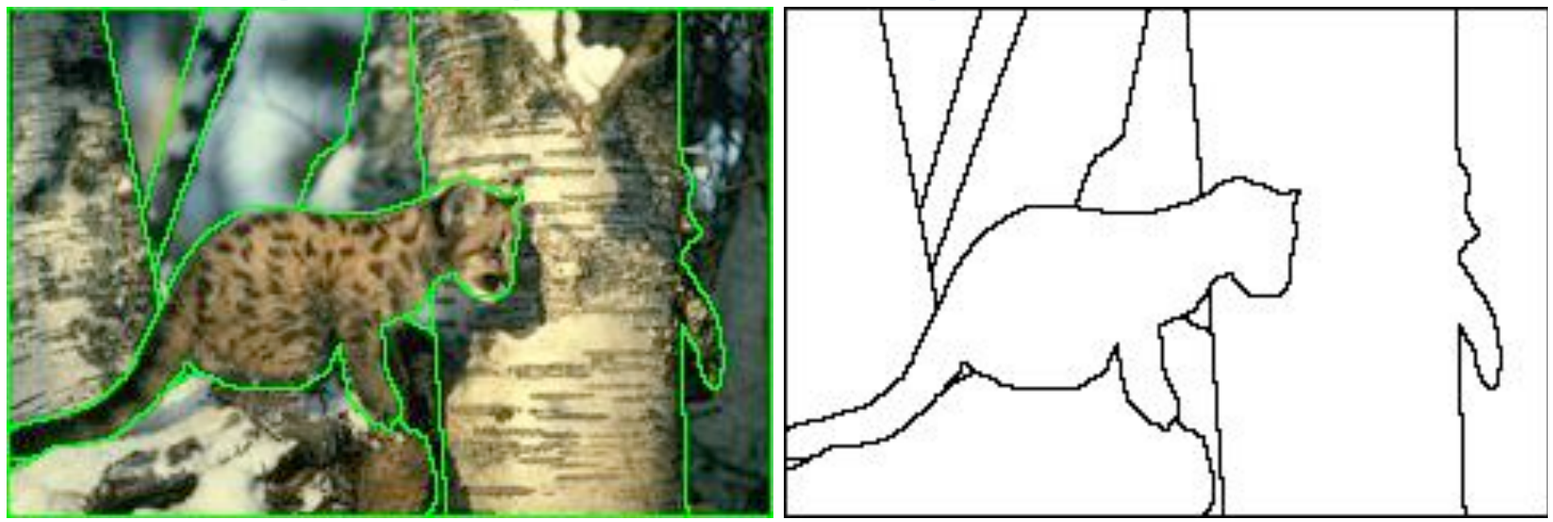
input image

user annotation

[Carsten Rother]

Issue (so far): No notion of spatial coherence

- All segmentation methods discussed so far do not model or encourage spatial regularity of the segments
  - yet, regularity is typical



  - Spatial regularity is only implicitly achieved (through similar appearance)
- CV2: Model and exploit the (dense) spatial regularity of many aspects of vision

Energy-Based Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

- A classical segmentation approach that encourages spatial regularity is based on energy minimization.
- The Mumford-Shah functional approximates an image f with a smooth function u and explicit discontinuities C:

$$E(u, C) = \int_{\Omega} (f - u)^2 \, \mathrm{d}x + \lambda^2 \int_{\Omega - C} |\nabla u|^2 \, \mathrm{d}x + \nu ||C||$$

- $\|C\|$ denotes the boundary length
- tradeoff between describing the image well (term 1 & 2) and having a short (i.e. “regular”) boundary

Energy-Based Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

$$E(u, C) = \int_{\Omega} (f - u)^2 \, \mathrm{d}x + \lambda^2 \int_{\Omega - C} |\nabla u|^2 \, \mathrm{d}x + \nu ||C||$$

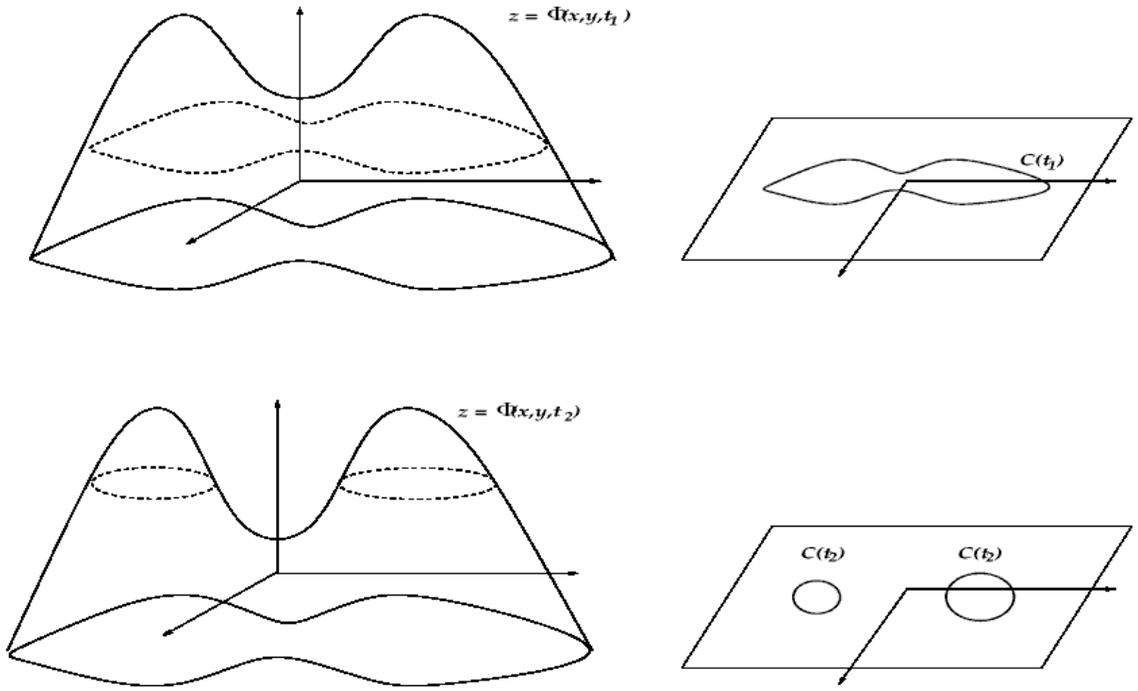
- Some observations:
  - If we disregarded the discontinuities, this would be a denoising approach!
  - The regularizer is spatially continuous; just like in Horn & Schunck
  - To implement it, it has to be spatially discretized

  - The power and “crux” lies in the discontinuities!
  - These make the energy rather hard to optimize.
  - Very often, so called level-set methods are used for approximation

Level-Set Methods

TECHNISCHE UNIVERSITÄT DARMSTADT

- Basic idea: Introduce an additional variable (function) and model the segment boundaries as the zero crossings of that function:

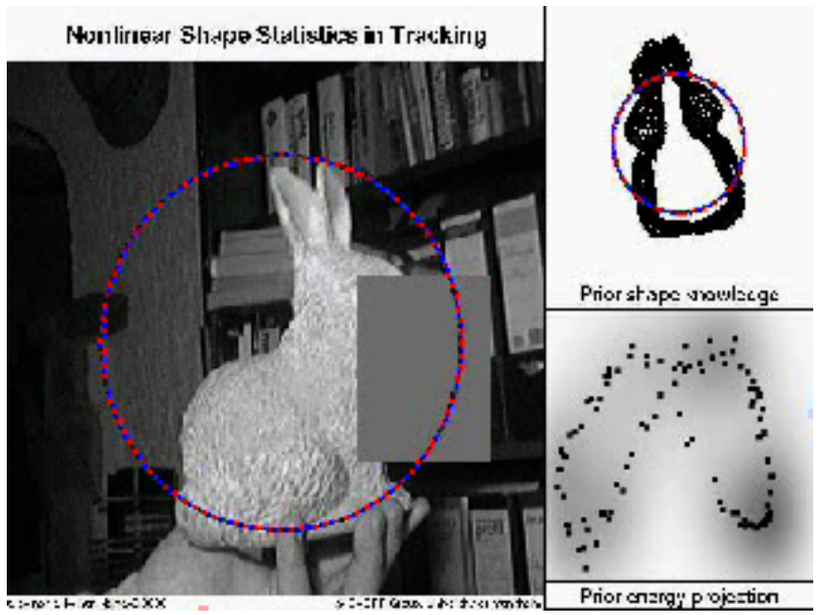


- I will skip all other details as they go beyond what we can reasonably cover here.

Shape Knowledge

TECHNISCHE UNIVERSITÄT DARMSTADT

- One advantage of the energy-based formulation is that we can add terms to modify the energy and make it more appropriate for our task.
- For example, we can add in a term that models **prior knowledge** we may have about the object shape:



[Daniel Cremers]

Probabilistic Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

- This leads us to our final approach: Probabilistic methods.
  - As you might have guessed, energy-based approaches can often be interpreted as probabilistic approaches.
  - We can formulate the segmentation problem using Markov random fields and use our standard inference techniques.

- One advantage:
  - We cannot only do MAP estimation, but we can estimate the marginal probability ("confidence") that a particular pixel will be assigned to a particular segment.
  - There are even techniques that figure out the number of segments automatically [Orbanz & Buhmann, 07].

Basic Approach

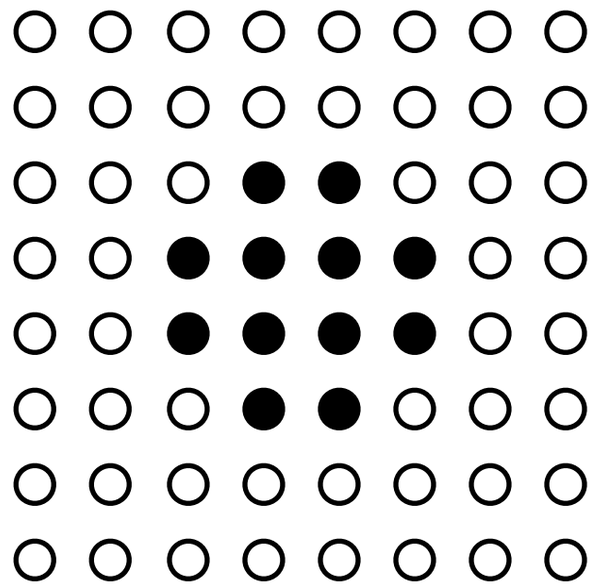
TECHNISCHE UNIVERSITÄT DARMSTADT

- Formulate the (interactive) segmentation problem as a discrete MRF
  - since predominantly MAP inference is used, write it using discrete energies
  - Want: per-pixel segmentation $S \in \{0,1\}^{m \times n}$
  - Energy formulation:

$$E(S) = E_d(S; I) + \lambda \cdot E_s(S)$$

data term (equivalent of likelihood)

smoothness term (equivalent of prior)



MRF-based Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

[Boykov & Jolly, 2001]

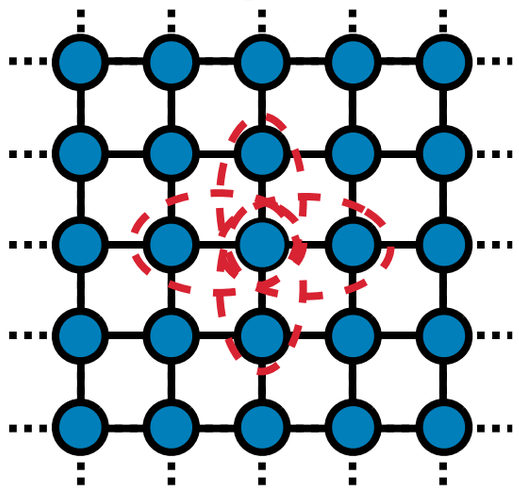
- Smoothness (spatial) term
  - aim for spatially coherent segments
  - expresses our prior belief about what are good segments
- Simplest approach: Potts model

$$E_s(S) = \sum_{(i,j)\in\mathcal{N}} \delta_{S_i \neq S_j}$$

1 if segment indices differ

0 otherwise

neighborhood (e.g. 4 or 8 neighbors)



MRF-based Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

[Boykov & Jolly, 2001]

- Data term
  - prefer pixels that “look like” foreground to be labeled as such & vice versa
  - as usual, assume i.i.d. likelihood model
- Simplest approach: Intensity / color histogram

$$E_d(S; I) = \sum_k -\log h_{S_k}(I_k)$$

- obtain from annotation

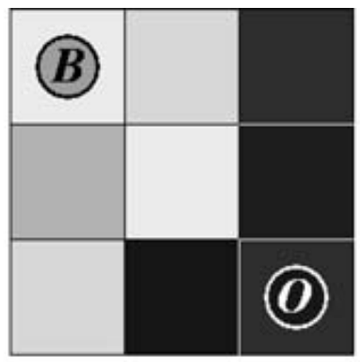
h_0 background histogram
h_1 foreground histogram

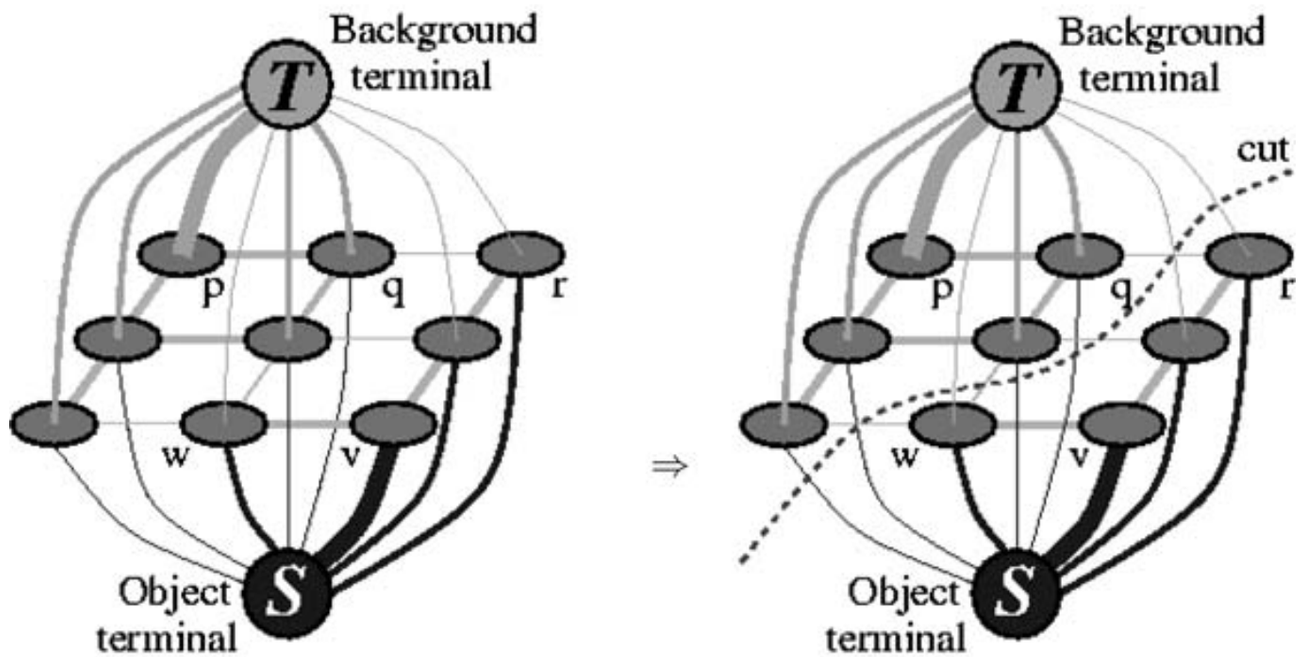
MRF-based Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

[Boykov & Jolly, 2001]

- Inference
  - Use graph cuts to infer the optimal segmentation
  - Globally optimal, energy submodular
  - user annotations form “hard links” with infinite weights
  - cannot be cut





Problem 1: Agnostic smoothness model

TECHNISCHE UNIVERSITÄT DARMSTADT

[Boykov & Jolly, 2001]

- The smoothness model is fully agnostic of the image

$$E_s(S) = \sum_{(i,j)\in\mathcal{N}} \delta_{S_i \neq S_j}$$

  - does not care if a segment boundary aligns well with image boundaries

- Solution: contrast-sensitive Potts model

$$E_s(S;I) = \sum_{(i,j)\in\mathcal{N}} e^{-\beta(I_i - I_j)^2} \cdot \delta_{S_i \neq S_j}$$

  - downweigh the penalty if the intensity of the neighbors differs
  - prefers segment boundaries that are consistent with image

Conditional random fields (CRFs)

TECHNISCHE UNIVERSITÄT DARMSTADT

- This is a particular instance of a conditional random field (CRF) [Lafferty et al., 2001]
  - instead of invoking Bayes' rule $p(S|I) = \frac{p(I|S)p(S)}{p(I)}$
  - and pursuing a generative approach that explains both the segmentation as well as the input image,
  - only regard the output as unknown and always assume the input as a given.
  - In other words: Directly model posterior $p(S|I)$
- In segmentation:

$$E(S) = E_d(S; I) + \lambda \cdot E_s(S; I)$$

contrast sensitive smoothness

  - all terms depend on the image!

Problem 2: Only binary segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

- The approach can be easily generalized to a finite number of segments
  - Generalized Potts model

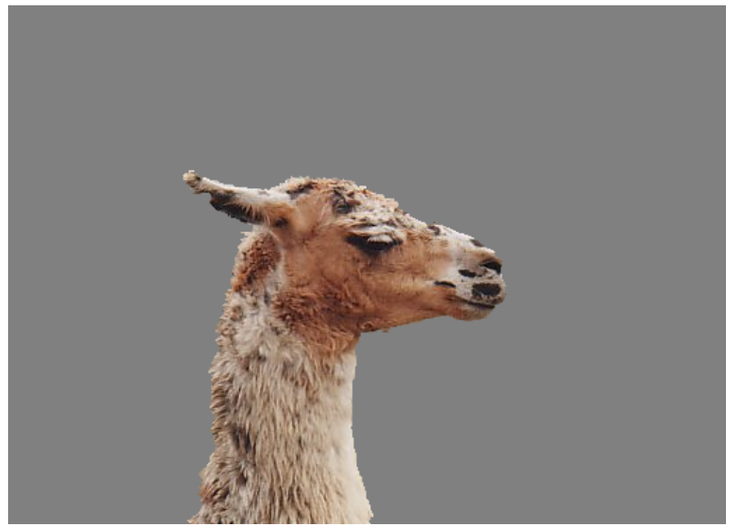
$$E_s(S;I) = \sum_{(i,j)\in\mathcal{N}} e^{-\beta(I_i - I_j)^2} \cdot \delta_{S_i \neq S_j}$$

  - looks just like the regular Potts model…
- But what about the likelihood
  - Either apply as is (requires lots of used input)
  - Or leave it out altogether!
    - Need segment size regularizer -> super pixels

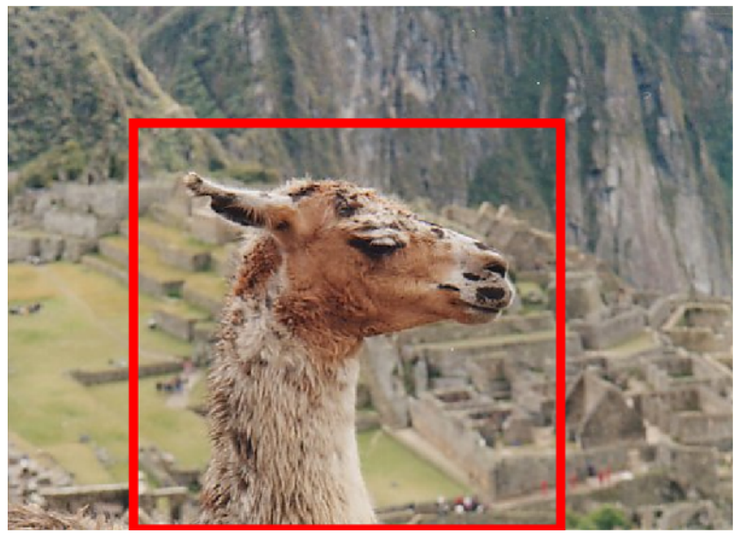
Problem 3: Lots of interaction needed

TECHNISCHE UNIVERSITÄT DARMSTADT

- Method so far [Boykov & Jolly, 2001]:
  - needs lots of user interaction
  - esp. in textured areas

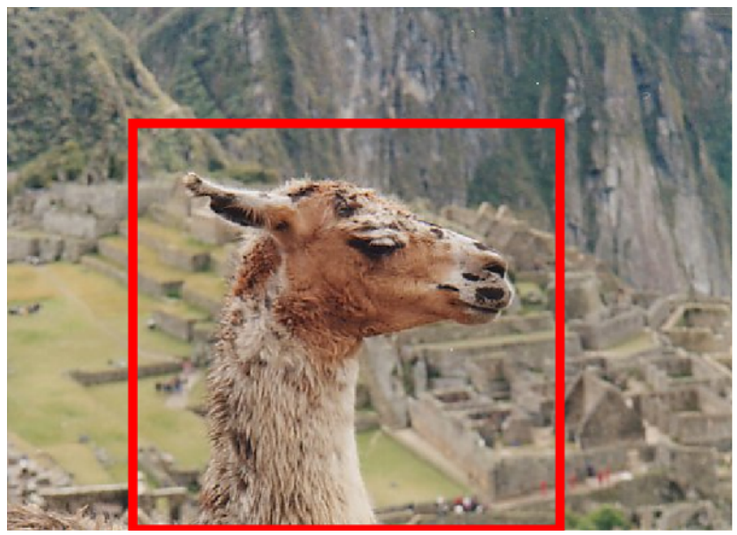


- Ideally:
  - specify only bounding box



Problem 4: Color model too unspecific

- First attempt:
  - Use color histogram of inside and outside of bounding box to define the foreground and background likelihood



Annotation

Result

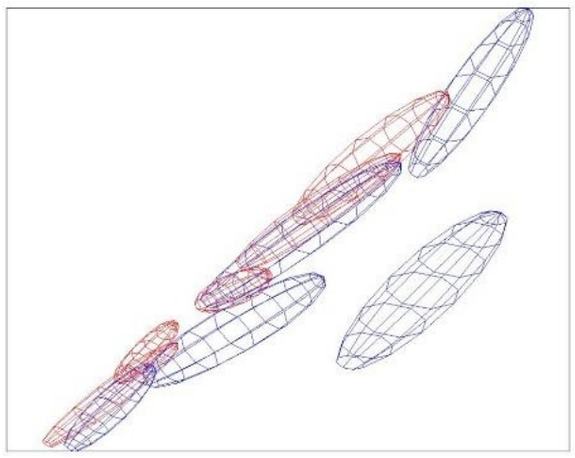
- Too much background is taken as foreground when determining the histograms

Iterated Graph Cuts

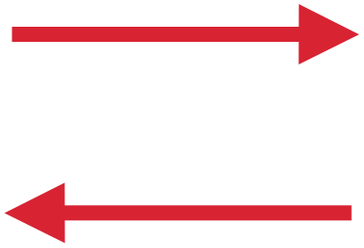
TECHNISCHE UNIVERSITÄT DARMSTADT

Grab Cut [Rother et al., 2004]

- Solution: Iterate between
  - Determining the histograms from current foreground & background
  - Segmenting the image with current likelihood



Gaussian mixture model of FG/BG



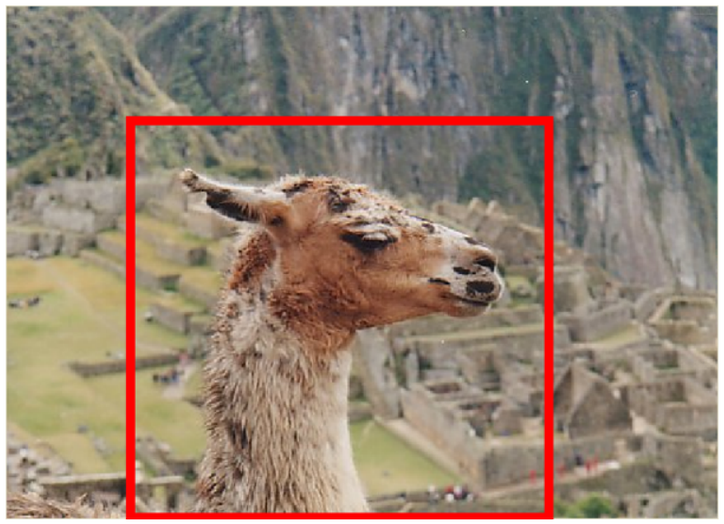
Graph cut segmentation

Iterated Graph Cuts

TECHNISCHE UNIVERSITÄT DARMSTADT

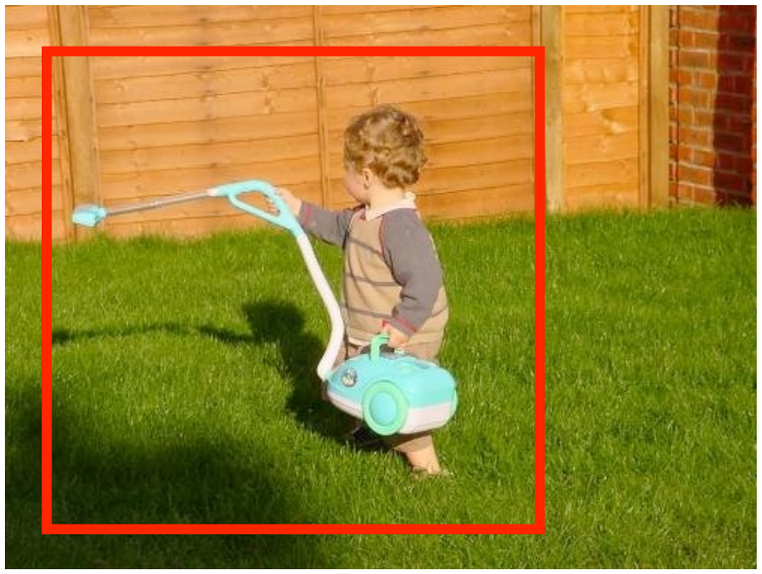
Grab Cut [Rother et al., 2004]

- Solution: Iterate between
  - Determining the histograms from current foreground & background
  - Segmenting the image with current likelihood



Progressing iterations...

Examples

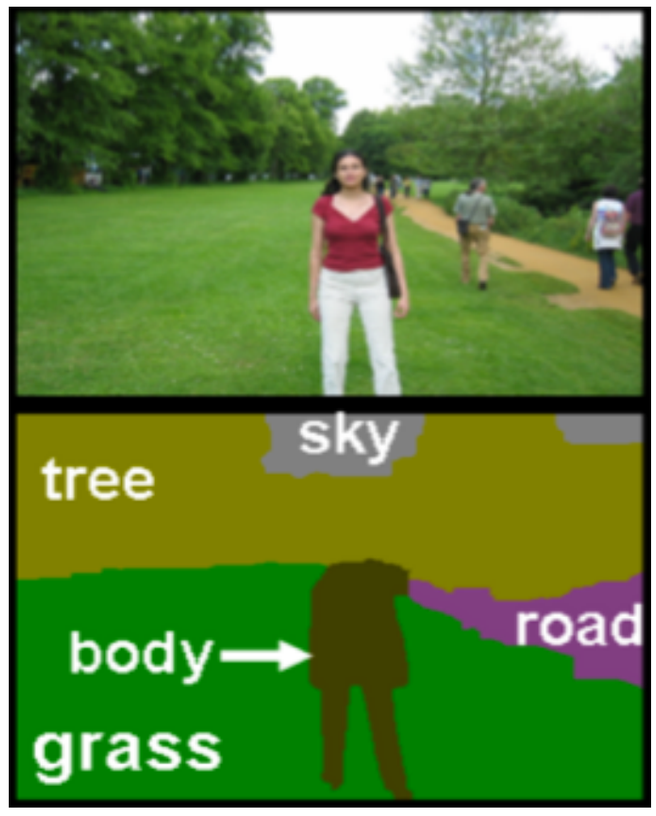


[Rother at al., 2004]

Semantic Segmentation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Segmentation can be made more well-defined by coupling it with high-level reasoning
- Semantic segmentation = Segmentation + Recognition
  - segment the image
  - determine the category at every pixel



[Jamie Shotton]