How to Learn and Evaluate an Interactive Computer Vision

Carsten Rother

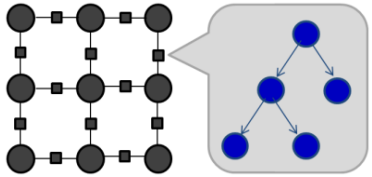
TU Dresden & Microsoft Research Cambridge

Computer Vision Lab Dresden (CVLD)

COMPUTER VISION LAB DRESDEN

Interactive Image and Data manipulation

Applied Optimization, Models, and Learning



Inverse rendering from moving images

3D Scene Understanding

Benchmarking and Label collection

BioImaging

… looking for PostDocs and PhD students

Inverse Rendering & Scene Understanding

[extension of Barron, Malik, ECCV '12]

[Vibhav, Torr et al. ECC '12]

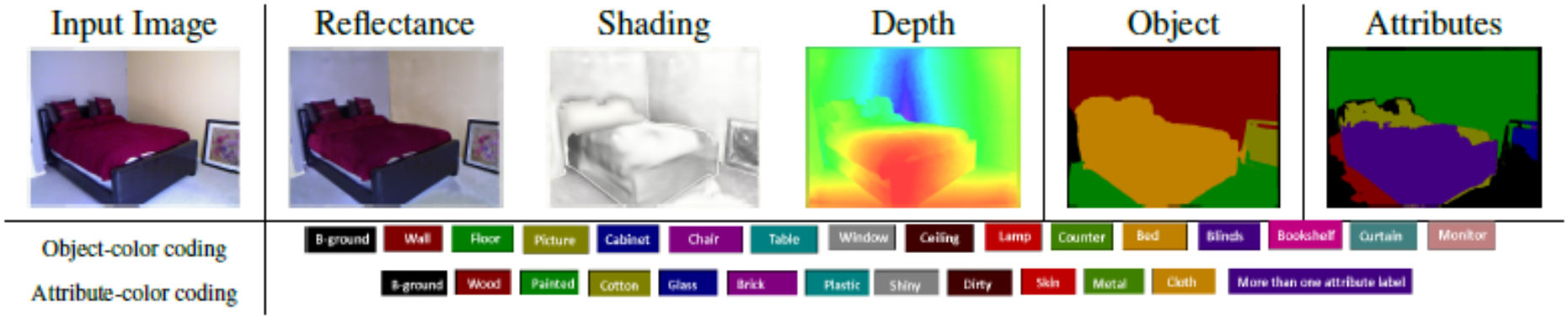


Figure 1: *Given an image, our algorithm jointly estimates the intrinsic properties such as reflectance, shading and depth maps, along with the estimation of the per-pixel object and attribute labels.*

[under submission]

How to Learn and Evaluate an Interactive Computer Vision

Joint work with:
Pushmeet Kohli, Hannes Nickisch, Christoph Rhemann

Learning and Evaluation of Interactive Segmentation Systems

Interactive Image Segmentation

Image (X)

[Boykov and Jolly ' 01] [Blake et al. '04] [Rother, Kolmogorov and Blake `04]

Demo

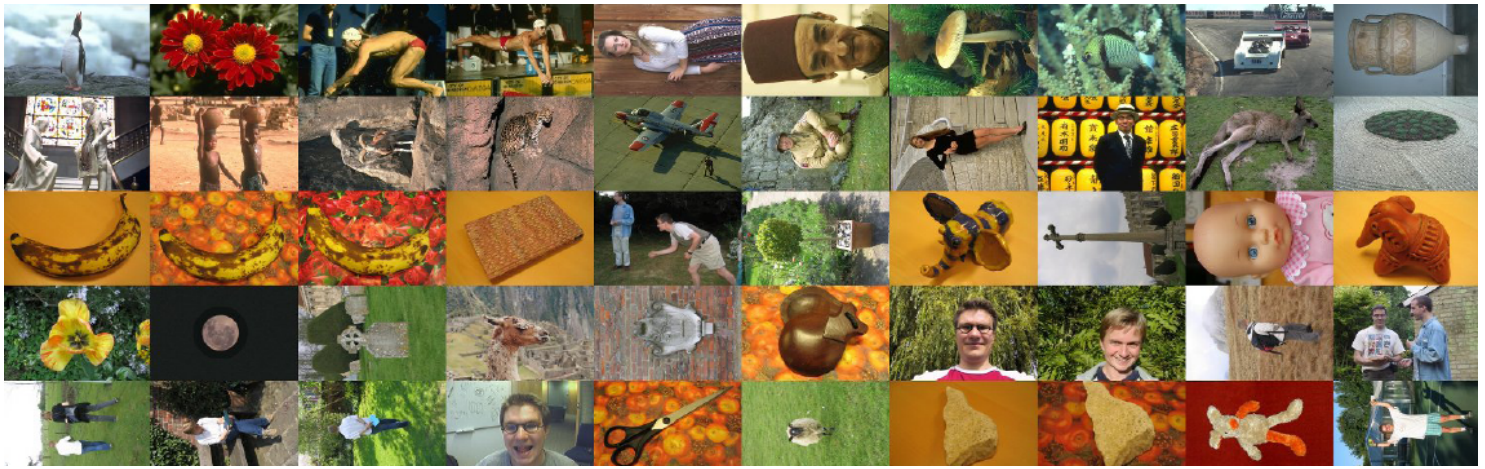
Key Questions

1. How to evaluate different segmentation systems?

2. How to learn the parameters of a given segmentation system?

… How GrabCut got into Office 2010+

[Markov Random Fields for Computer Vision, Blake, Kohli and Rother]



Database of 200 Images

Image

Ground Truth

Rectangle Labelling

Lasso Labelling

… How GrabCut got into Office 2010+

[Markov Random Fields for Computer Vision, Blake, Kohli and Rother]

Product team got big spread sheet with error rates

Product team: “What do these numbers mean?”

Me: “Explain Error rate”

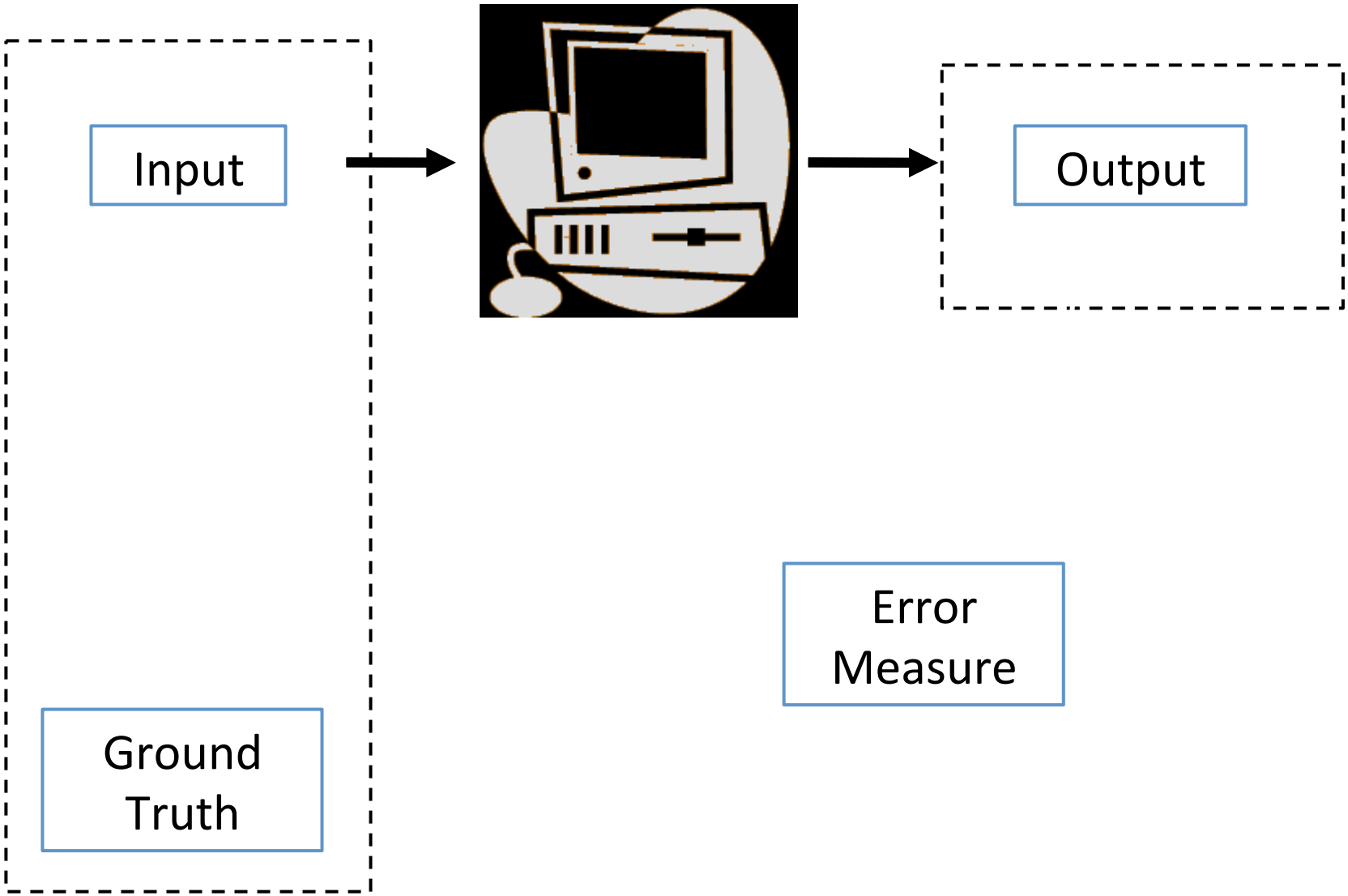
Product team: “Hm”

Me: “Ok we do something better”

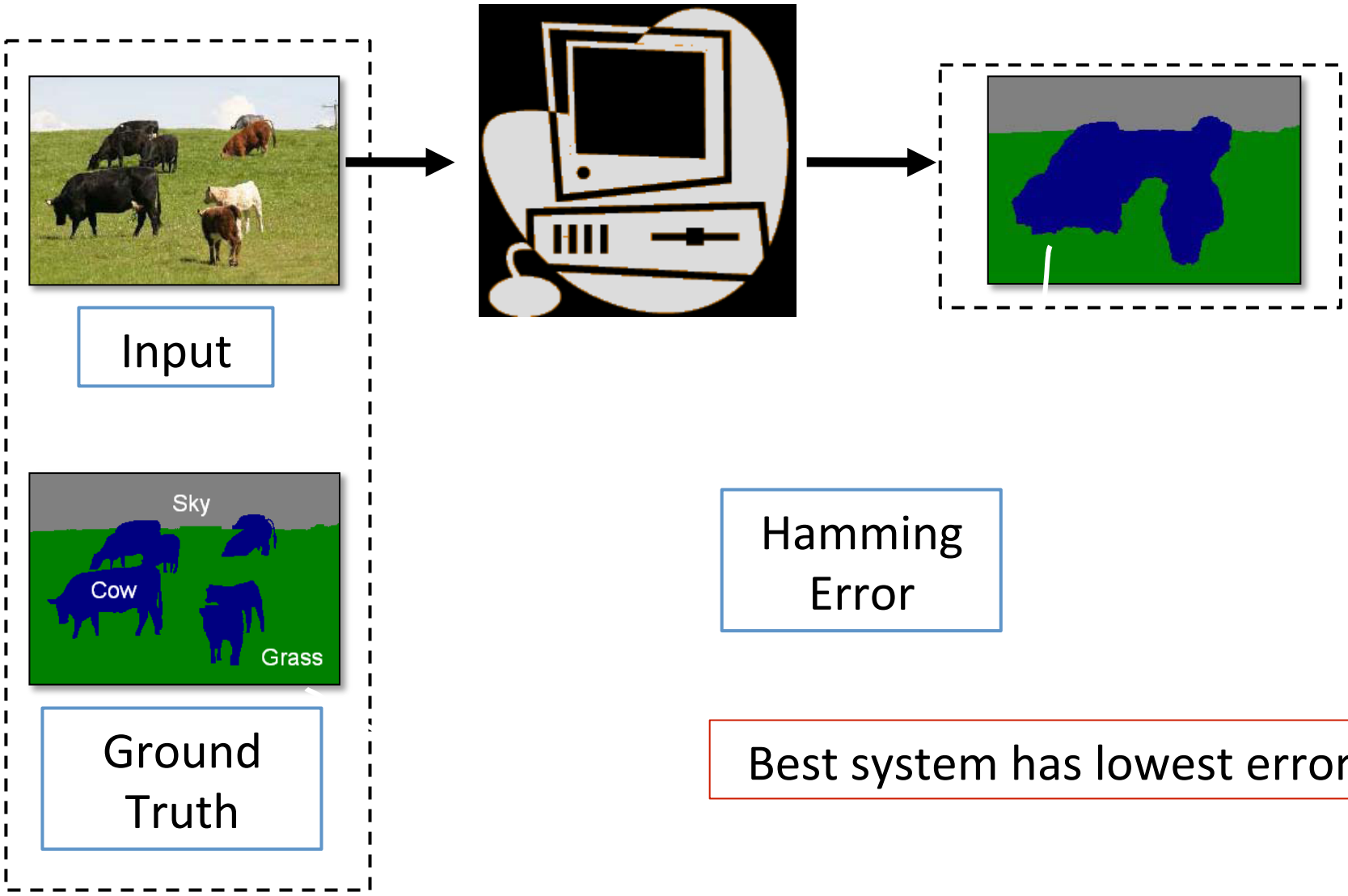
5% error

5% error

The Evaluation Task

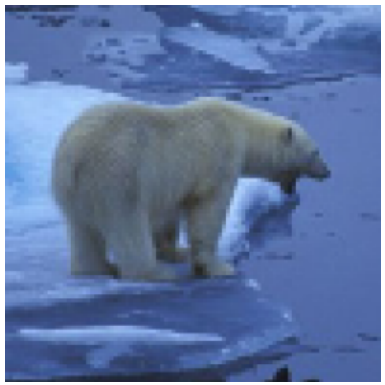


The Evaluation Task



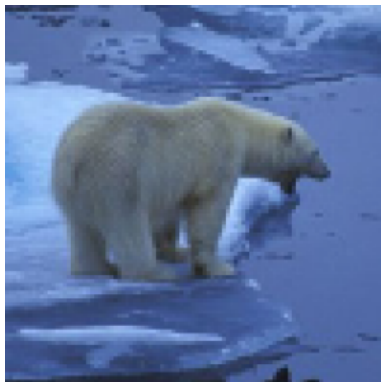
The Evaluation Task

User Scribbles
(Effort)



The Evaluation Task

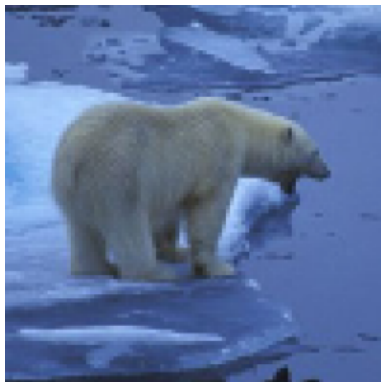
User Scribbles
(Effort)



Segmentation
(Quality)

The Evaluation Task

User Scribbles
(Effort)



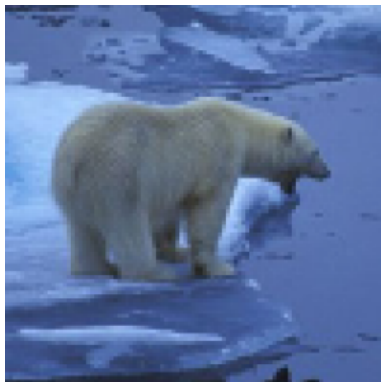
Segmentation
(Quality)

Interactive Algorithms

Best system requires lowest amount of effort to reach a reasonable answer

The Evaluation Task

User Scribbles
(Effort)

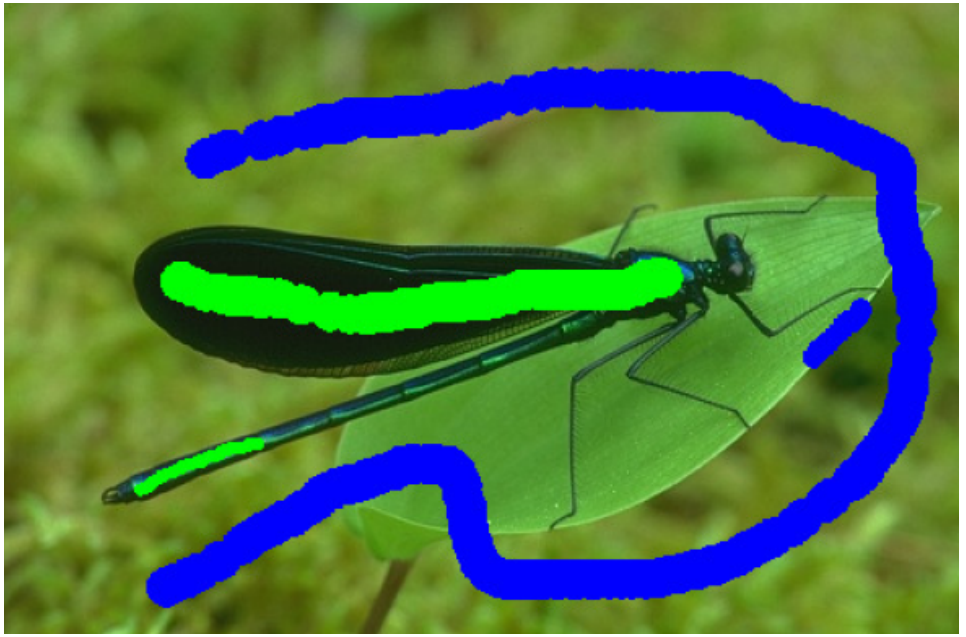


Segmentation
(Quality)

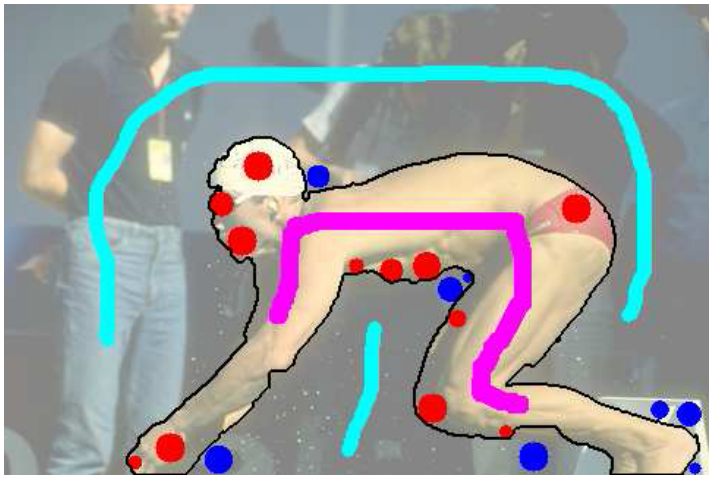
Interactive Algorithms
Best system requires lowest
amount of effort to reach a
reasonable answer

How to measure “effort” and “reasonable”?

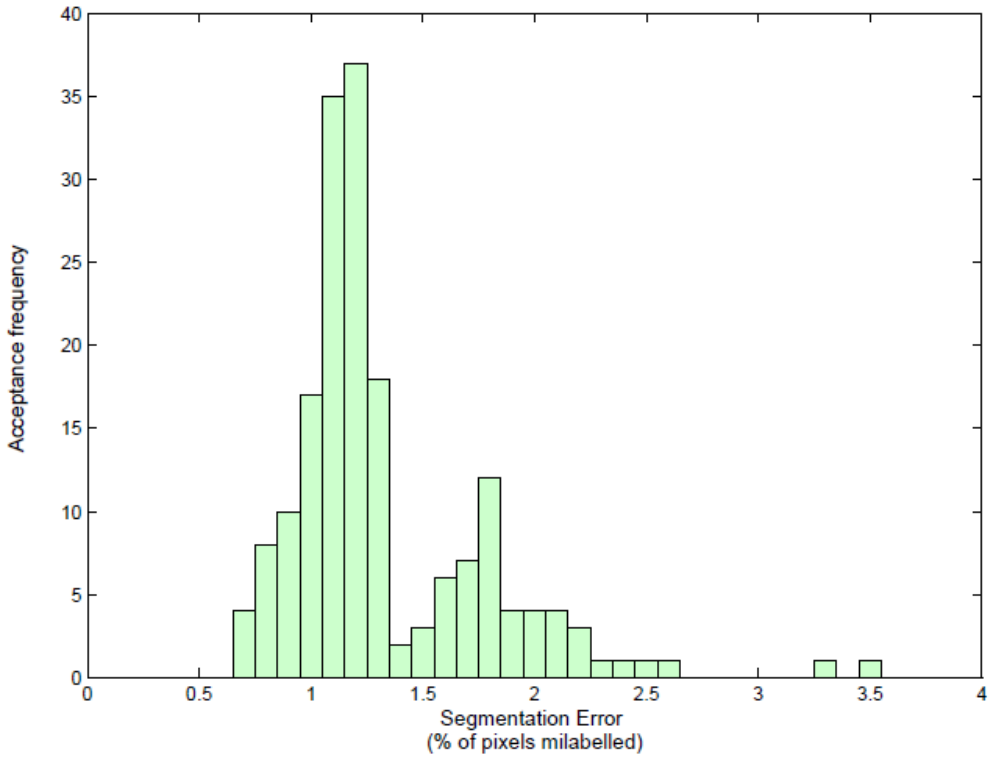
- Number of brush strokes
- Stroke complexity
- Time taken



Circular brush



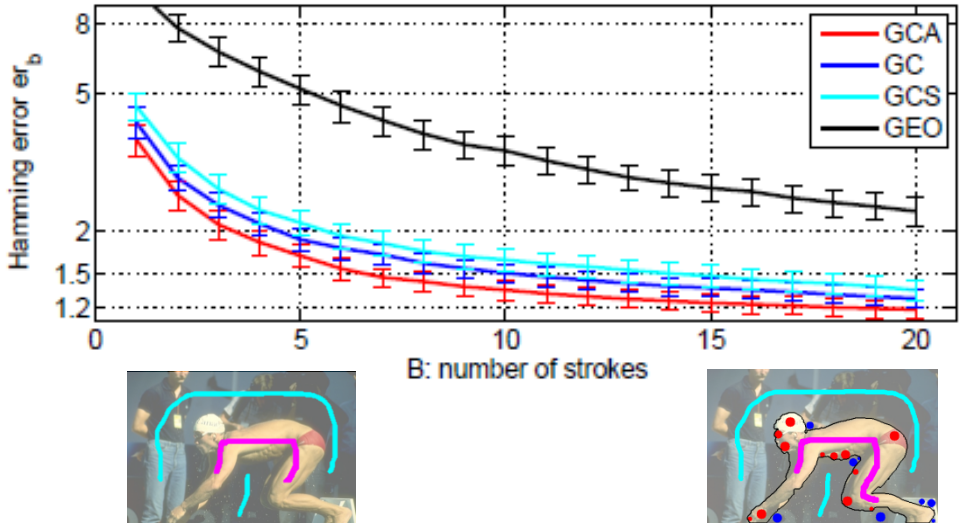
How to quantify reasonable accuracy?

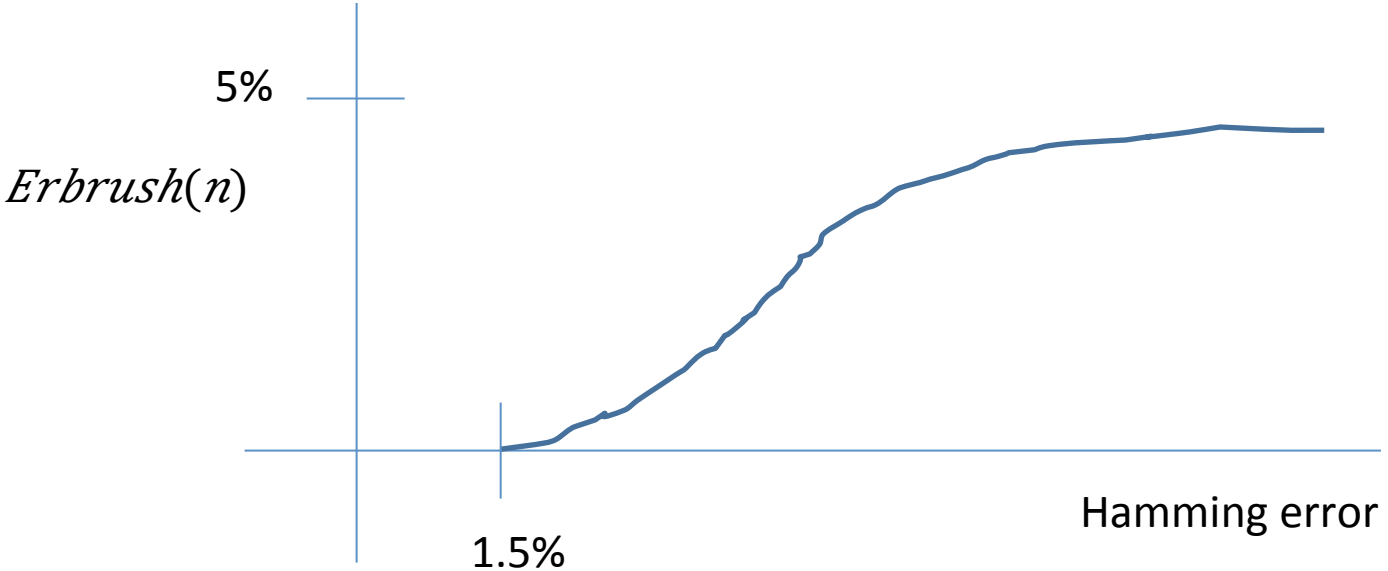


We truncate the hamming error with 1.5%

Error measure

$Er = \sum_{n=1}^{B} Erbrush(n)$





User Study or Crowdsourcing

- Measure performance using actual users
  - Few users

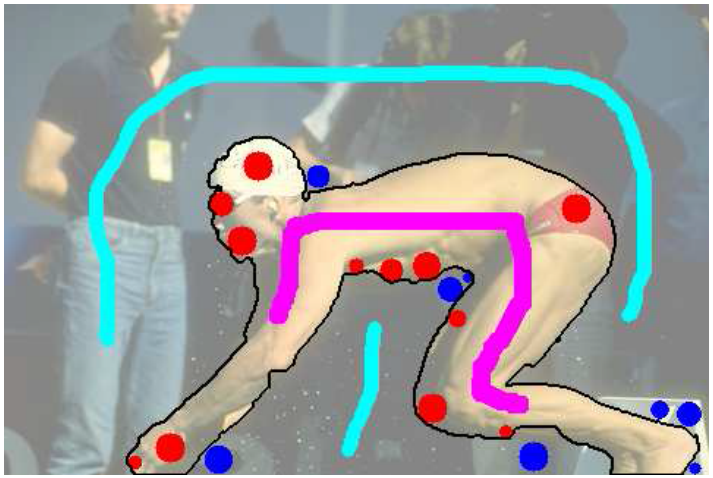
    [Li et al. 2004, Bai & Sapiro 2007]
  - Crowd-sourcing with different incentives

    [MTurk, Community-based (Labelme), ESP game]

- Problems:
  - Expensive
  - Slow (participants should be allowed learning time)
  - Unsuitable for parameter learning

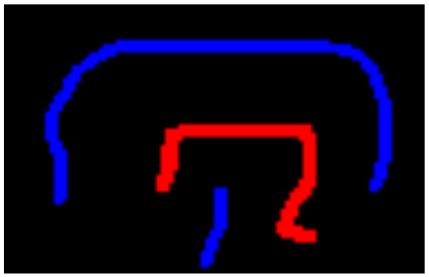
    (where you might need to evaluate many systems)

Robot User



Where to place the brush strokes?

Image

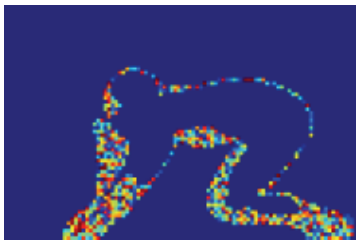


Interactions

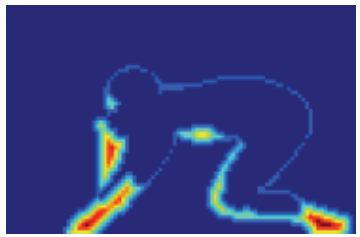
Ground Truth

Current Solution

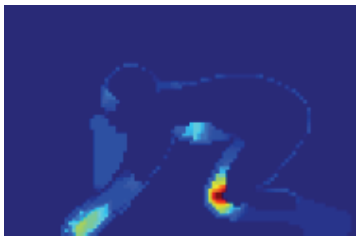
Interaction chosen by taking the maximum in the preference map



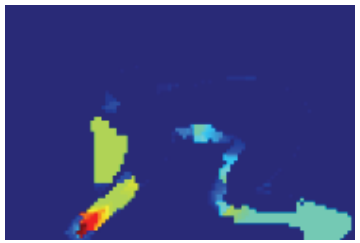
Random



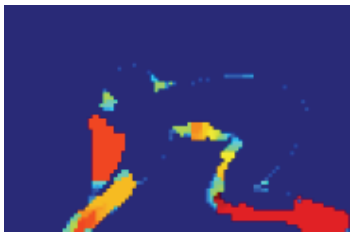
Center



Confidence
Max-marginal Difference



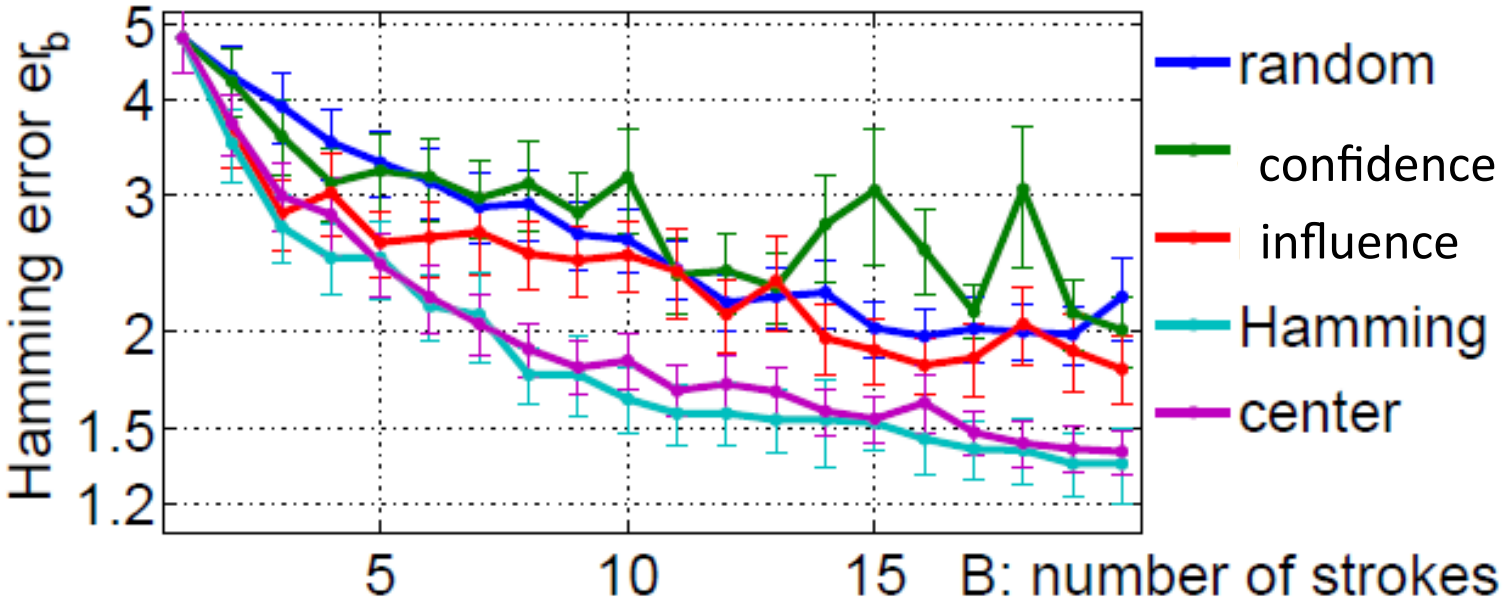
Amount of Influence*



Decrease in Hamming error

Red – preferred
Light blue – not preferred

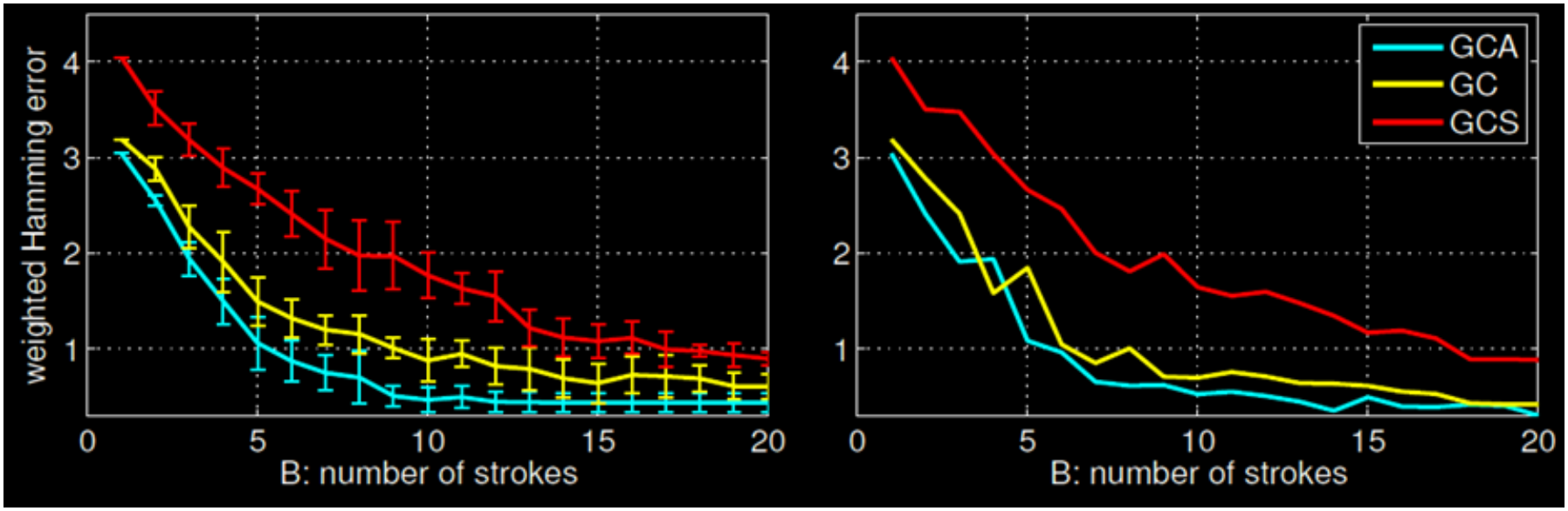
Where to place brush strokes



Number of Interactions

Side comment: using the uncertainty of the system is worse than random!

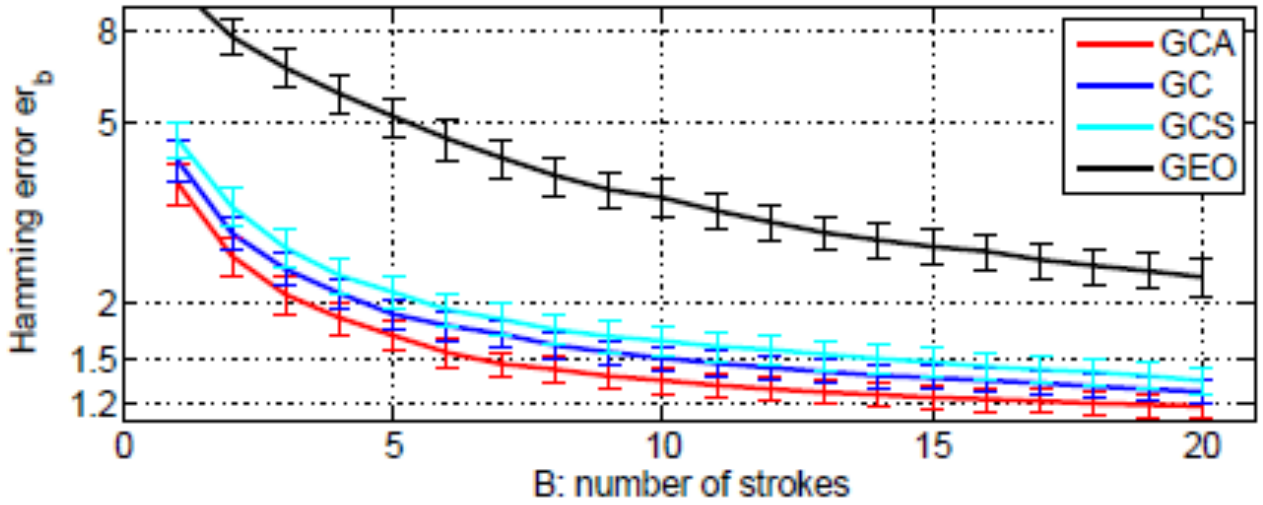
Is it a good model for Users?



Human

Computer

Comparison



- Grabcut

  [Boykov Jolly 2001] [Rother et al SIGGRAPH 2004]

– GCS

  (Simplified – fixed colour models)

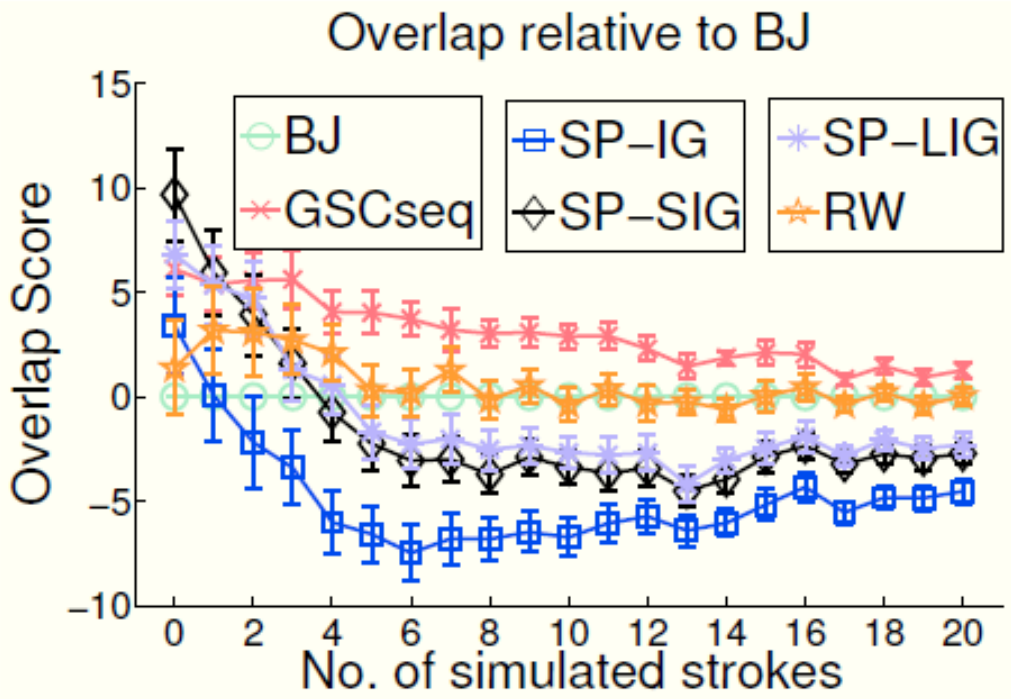
– GCA

  (Advanced – enforces connectivity of segmentation)

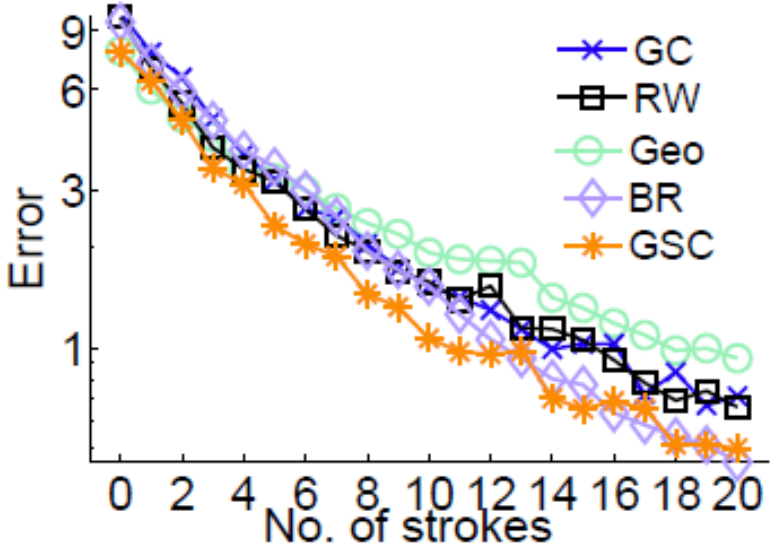
– GEO

  (Geodesic distance based methods) [Bai Sapiro, ICCV 07]

Used to compare systems



[Gulshan et al. CVPR '10]



[Rother, MRF Book]

What is ignored ?

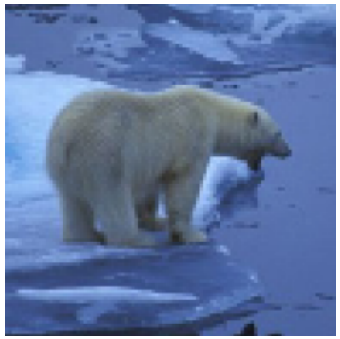
- Speed of system? (faster better)
- User adjust to the system

Key Questions

1. How to evaluate different segmentation systems?

2. How to learn the parameters of a given segmentation system?

The Learning Task



Segmentation System

System Parameters

(w_s)

[Generative Learning]
Maximum Likelihood
[Kumar et al 2005]
[Pletscher, Nowozin, Kohli, Rother, DAGM 2011]
....

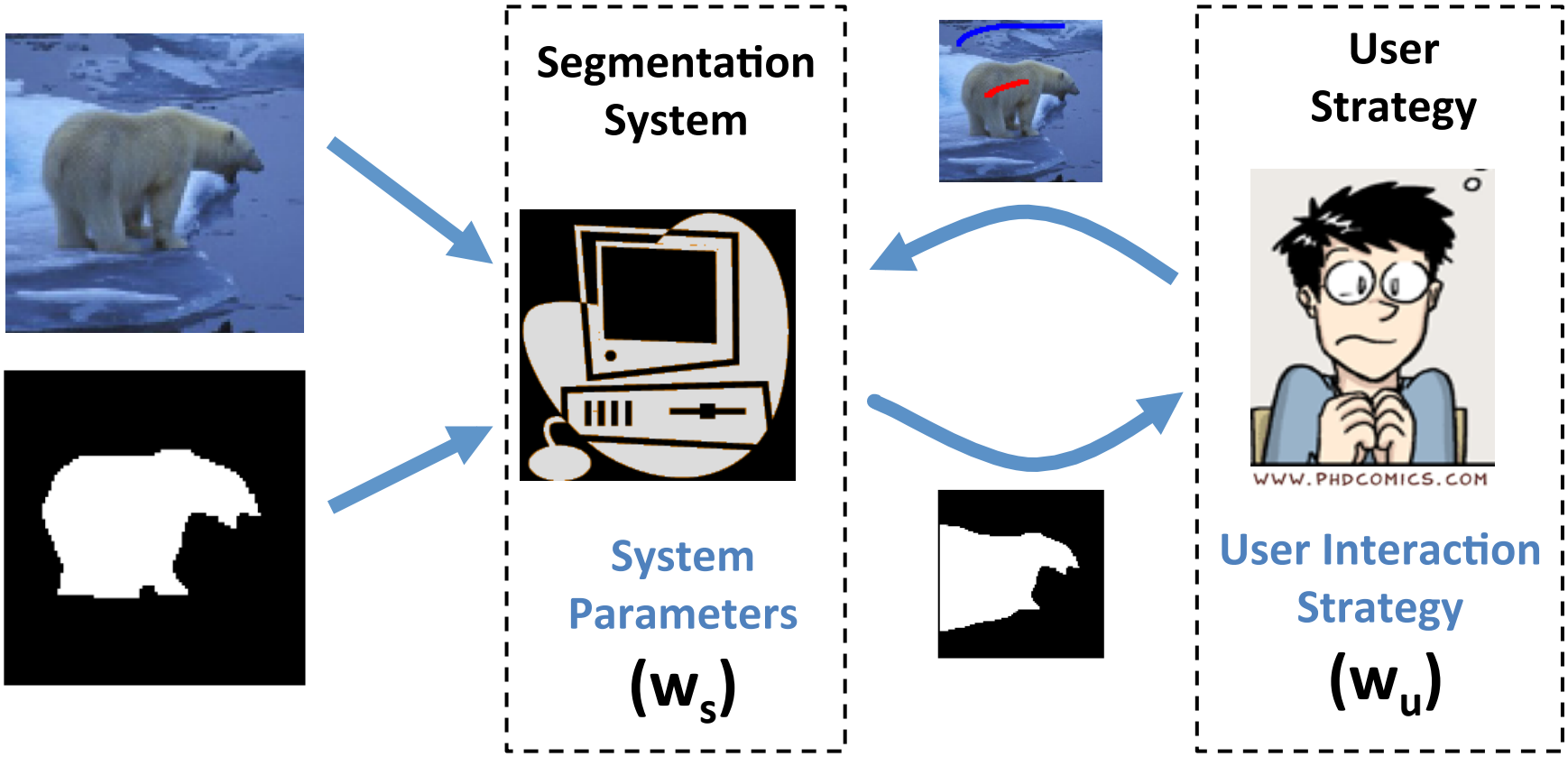
$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \left[\prod_{k=1}^{K} \mathbb{P}_{\mathbf{w}}(\mathbf{x}^k, \mathbf{y}^k)\right]$$

$$\mathbb{P}_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \exp\left(-E_{\mathbf{w}}(\mathbf{x}, \mathbf{y})\right)$$

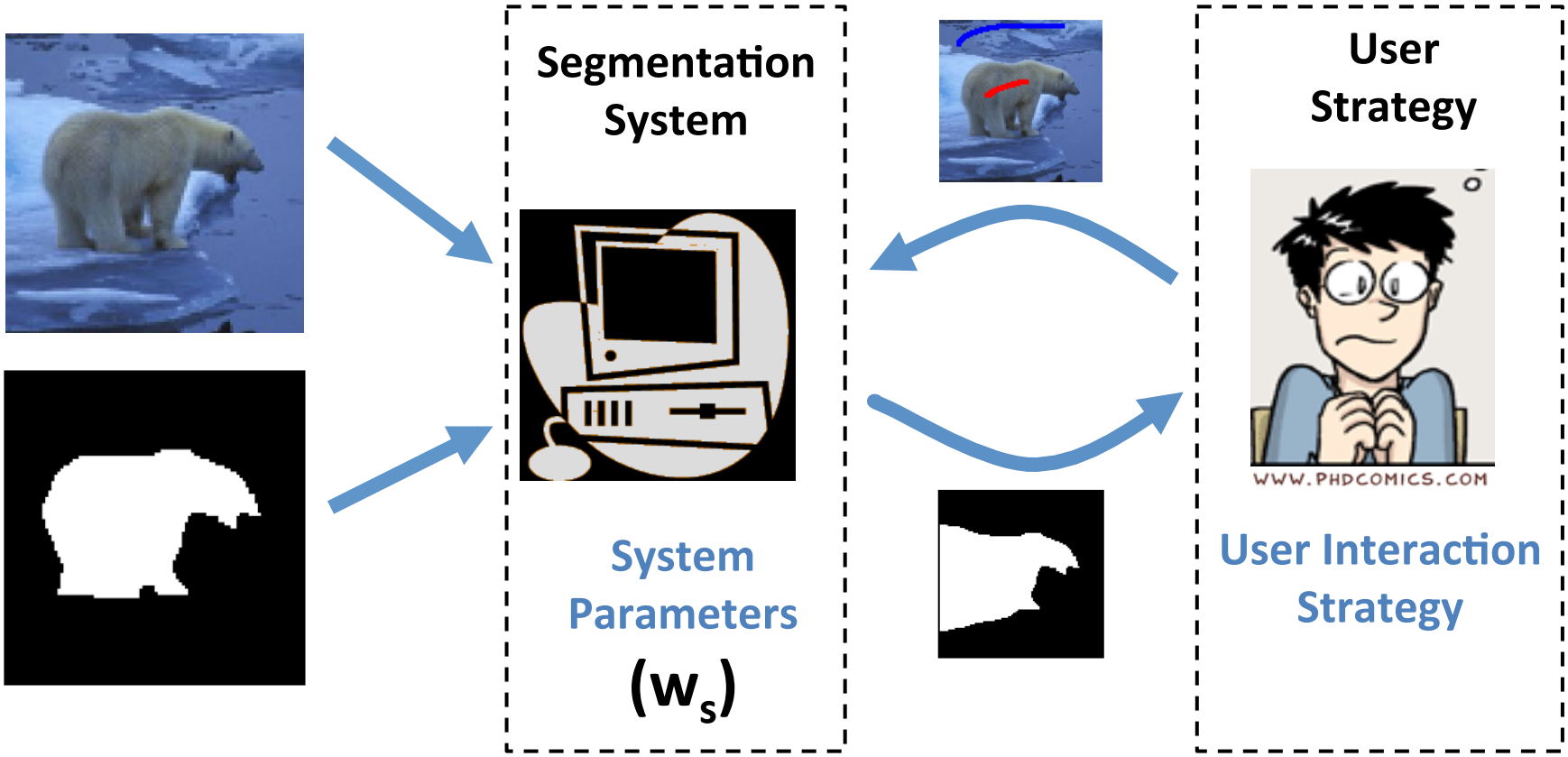
[Discriminative Learning]
Max-margin based methods
[Tsochantaridis et al 2001]
[Taskar et al 04,05]
[Szummer et al 08]

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \left[\min_{\mathbf{y} \in \mathcal{Y} \setminus \mathbf{y}^k} \left\{E_{\mathbf{w}}(\mathbf{x}^k, \mathbf{y}) - E_{\mathbf{w}}(\mathbf{x}^k, \mathbf{y}^k)\right\}\right]$$

The Learning Task



The Learning Task



Few parameters → Learn by grid-search over the evaluation score on test data

System

$$E(\mathbf{y}) = \sum_{p \in \mathcal{V}} E_p(y_p) + \sum_{(p,q) \in \mathcal{E}} E_{pq}(y_p, y_q)$$

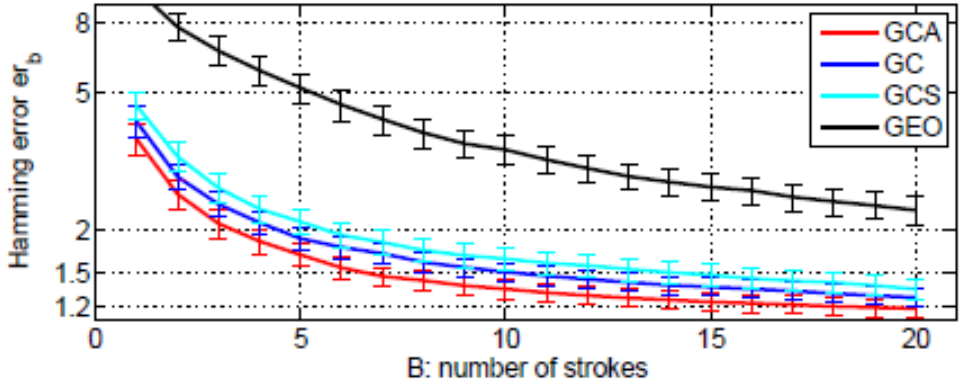
$$E_{pq}(y_p, y_q) = \frac{|y_q - y_p|}{\text{dist}\,(p,q)} \left(w_i + w_c \exp\left[-\beta \|x_p - x_q\|^2 \right] \right)$$

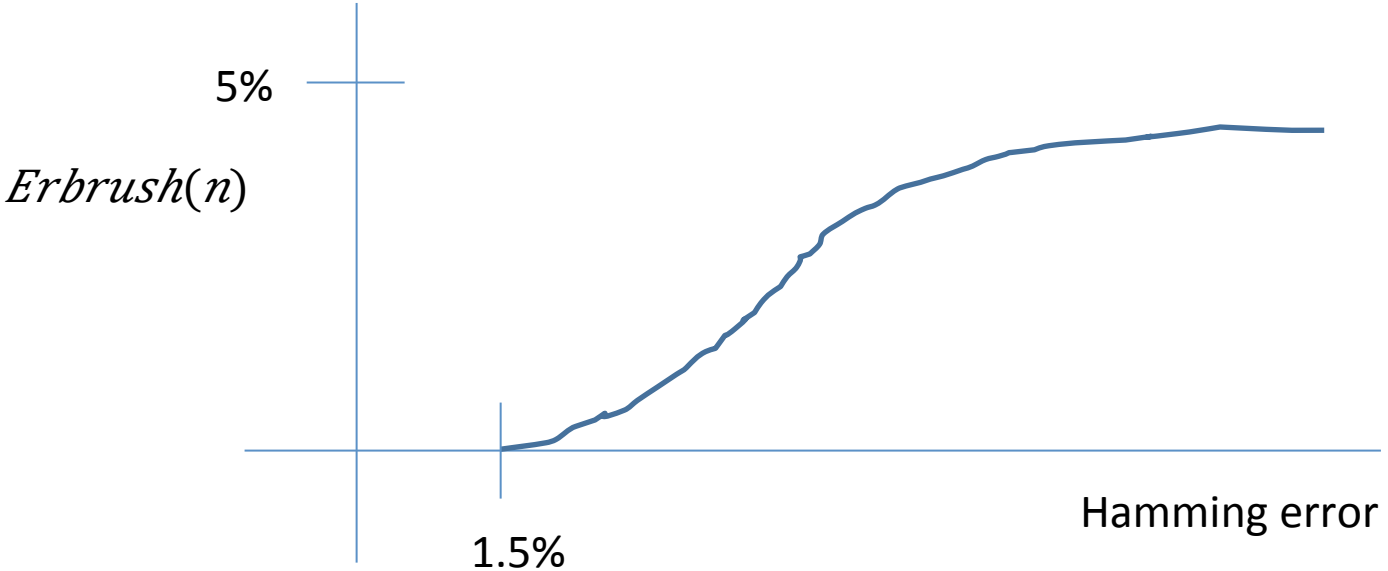
$$\beta = 0.5 \cdot w_\beta / \left\langle \|x_p - x_q\|^2 \right\rangle$$

3 free parameters

Training objective

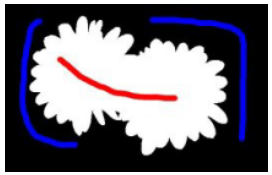
$$Er = \sum_{n=1}^{B} Erbrush(n)$$



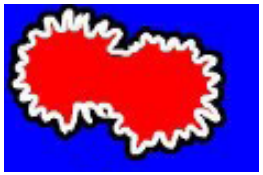


Extreme scenarios

Training Time



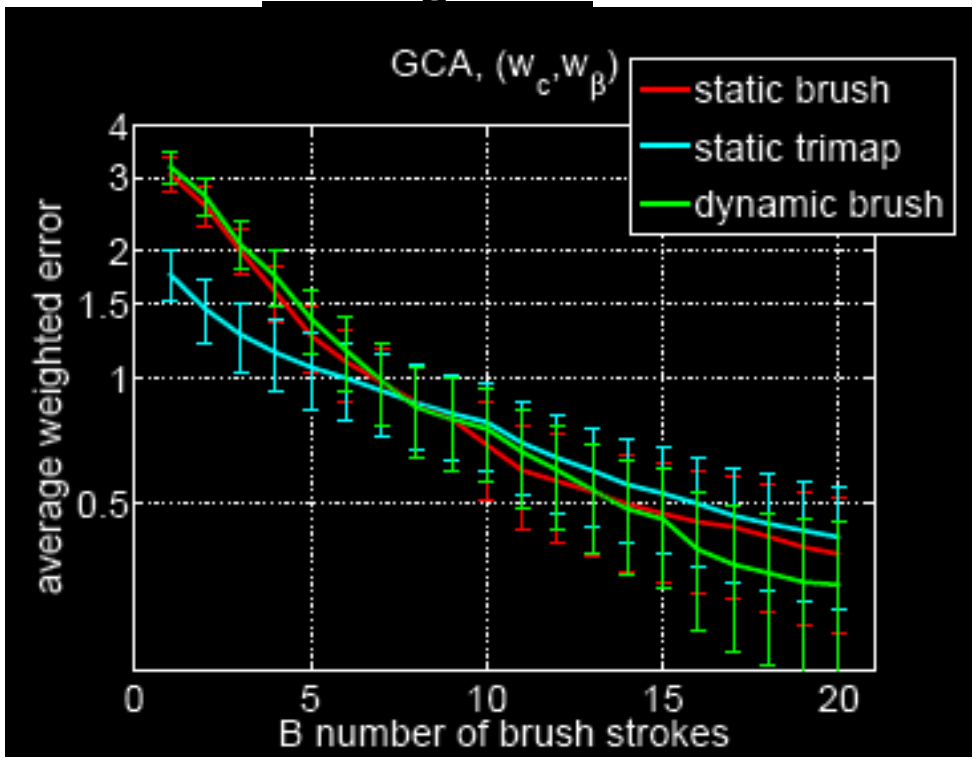
Static brush



Static trimap

Trimap	w_c
Static trimap	0.07 +/- 0.09
Static brush	0.22 +/- 0.52

Testing Time



Other Application Scenarios

Interactive systems is more than just a work-around to not yet automatic systems

- Computer vision:
  - segmentation, optical flow, shape, etc.
  - Mechantical turk for gathering ground truth

- Biology
- Physical simulations
- Human-Computer-Interaction
- Interactive visualization

Open Questions

- Speed of system? (faster better)
- User adjust to the system
- Other solutions …