A Real-time MRF Based Approach for Binary Segmentation

Dmitrij Schlesinger

TECHNISCHE UNIVERSITÄT DRESDEN

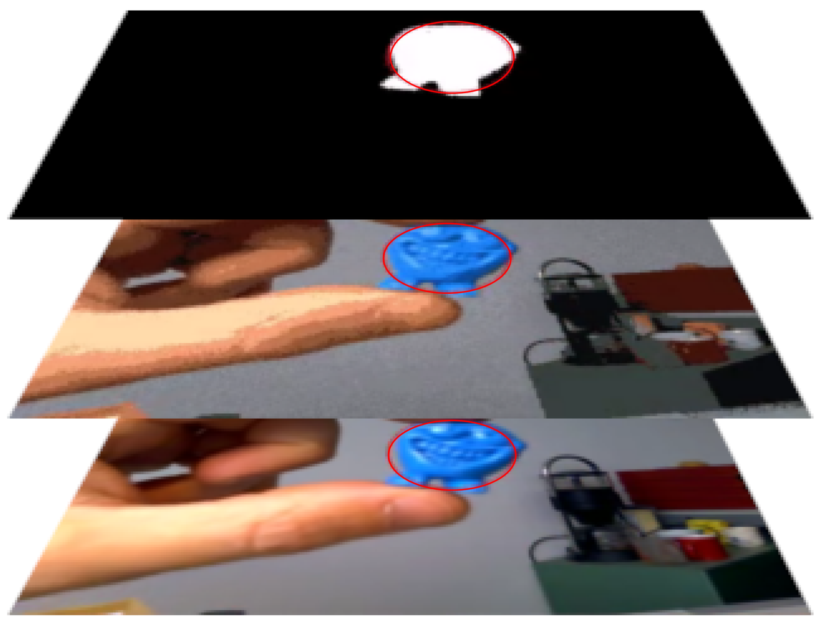
Application: real-time segmentation of live video streams

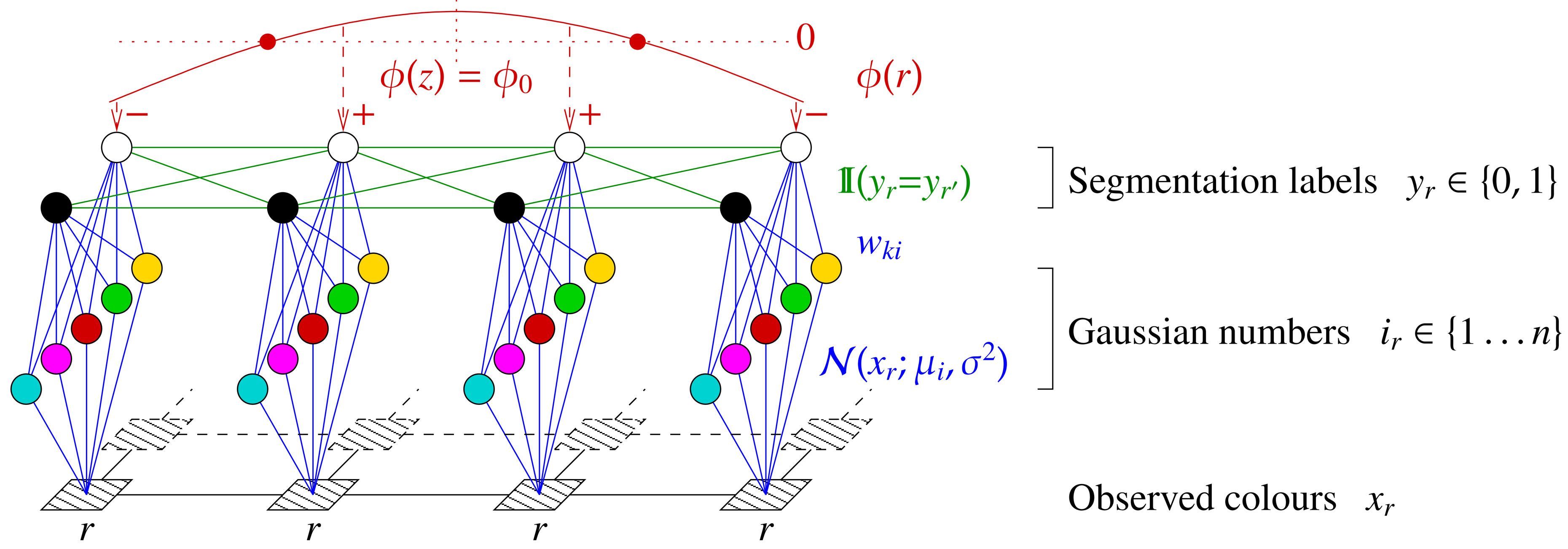
- Both inference and learning should be very fast.
- No user interactions $\rightarrow$ fully unsupervised learning $\rightarrow$ a *generative* model is necessary.

Make it right before you make it fast – model

An **MRF** (see [5]) with: Potts prior, shape prior, appearance model

$$p(x,y;\phi,\theta)=\frac{1}{Z(\phi)}\exp\left[\alpha\cdot\sum_{rr'}\mathbb{I}(y_r{=}y_{r'})+\lambda\cdot\sum_r y_r\cdot\left[\phi_0-\frac{(r_x-z_x)^2}{2\sigma_x^2}-\frac{(r_y-z_y)^2}{2\sigma_y^2}\right]\right]\cdot\prod_r\sum_i w_{y_r i}\cdot\frac{1}{\sigma^3}\exp\left[-\frac{\|x_r-\mu_i\|^2}{2\sigma^2}\right]$$

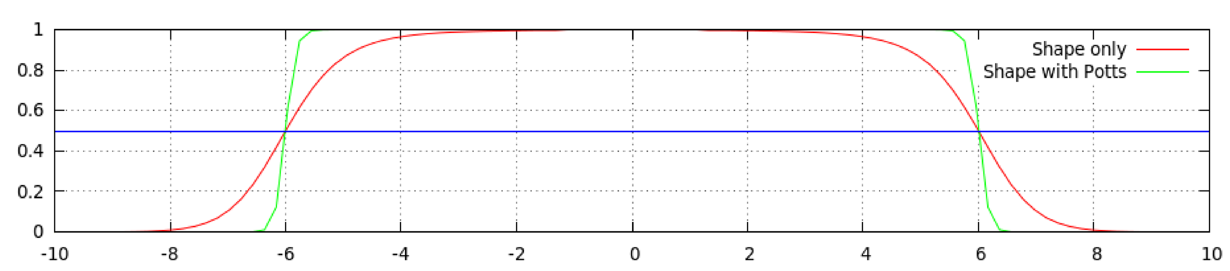




- Inference: Max-Marginal decision $y_r^*=\arg\max_k p(y_r{=}k|x;\phi,\theta)$ $\rightarrow$ posterior sampling
- Learning: Maximum Likelihood
  - appearance model $\theta=(\mu,\sigma)$ $\rightarrow$ Expectation-Maximization + posterior sampling
  - shape $\phi=(\phi_0,z_{xy},\sigma_{xy})$ $\rightarrow$ Expectation-Maximization + posterior sampling + prior sampling

Keep it right when you make it fast – implementation

Prior label probabilities

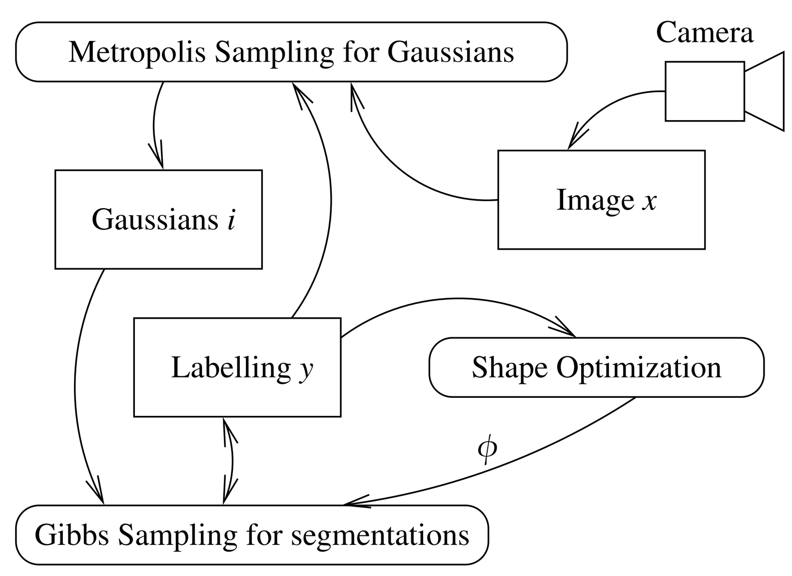


Prior label probabilities are very distinctive
$\rightarrow$ can be approximated by 0/1
$\rightarrow$ prior statistics can be computed explicitly
$\rightarrow$ prior sampling is not necessary

Fast generation

- Gibbs Sampling for segmentations
  $\rightarrow$ the necessary energy *difference*
  $q(x_r,1)-q(x_r,0)+\phi(r)+\alpha\sum_{r'\in N(r)}(2y_{r'}-1)$
  can be computed very fast
- Metropolis Sampling for Gaussian numbers
  $\rightarrow$ energy difference is a scalar product
- Look-up tables are easy to use

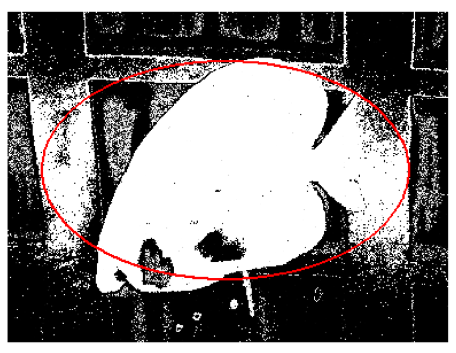
Multithreading



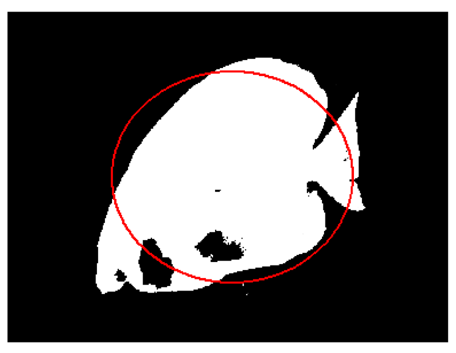
$\rightarrow$ working in parallel

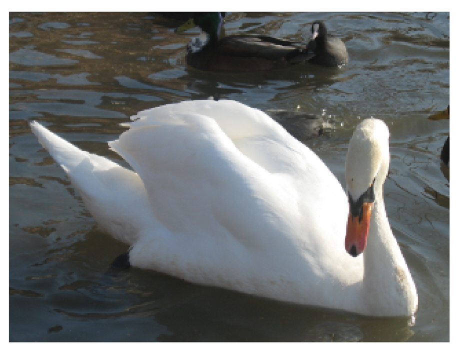
Some results

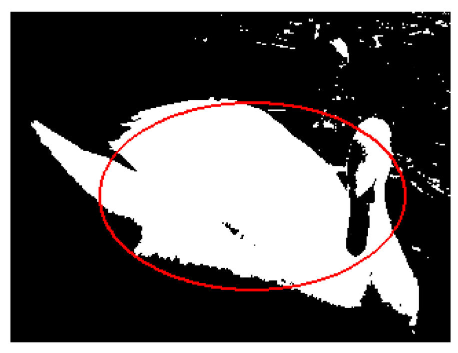

Input image


Shape prior only


Potts prior only

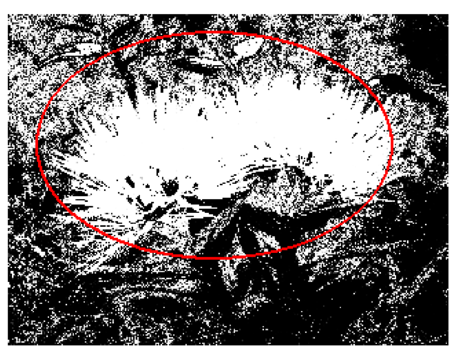

Full prior


Input image

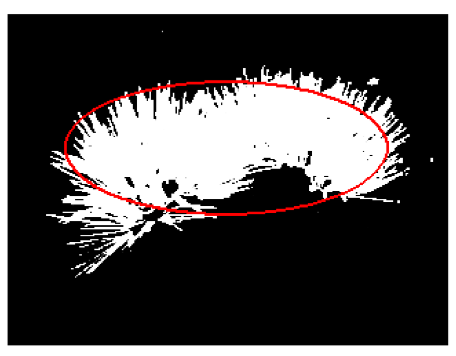

Weak prior

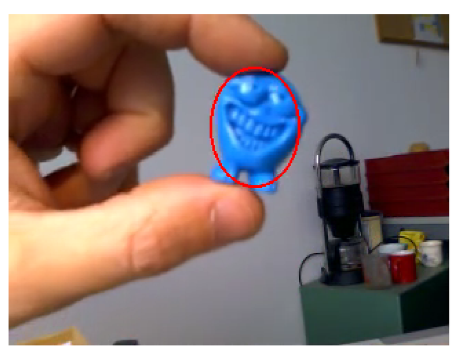

Strong prior

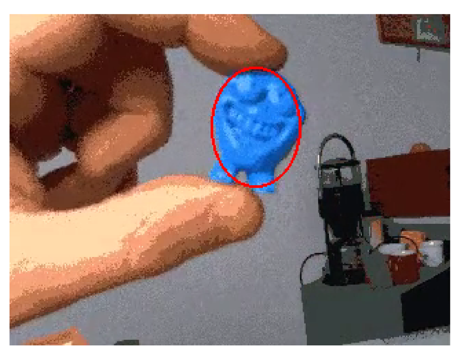

Input image

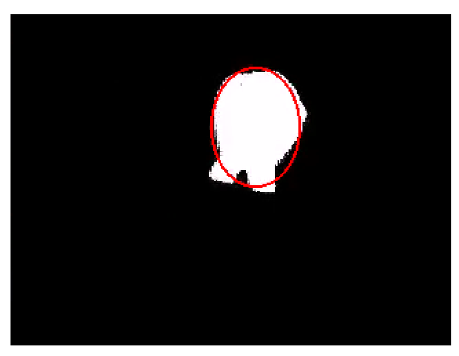

Shape prior only


Potts prior only


Full prior


Input image


Colouring


Segmentation

Keep it right and fast when you make it better – future work

- The assumption for the prior label probabilities seems to hold for more complex shape models as well
- Many foreground labels $\rightarrow$ more complex objects/scenes, object part segmentation, etc.
- User interactions $\rightarrow$ non real-time semi-interactive video segmentation

...