MCA-P2.6 International Conference on Multimedia and Expo ICME2009

NTT 国立沖縄高専 Okinawa National College of Technology

Saliency-based video segmentation with graph cuts and sequentially-updated priors

Ken Fukuchi (2), Kouji Miyazato (2), Akisato Kimura (1), Shigeru Takagi (2), Junji Yamato (1)

(1) NTT Communication Science Laboratories, NTT Corporation (2) Okinawa National College of Technology

Notice!! None of the authors cannot attend ICME2009 due to the policy of the affiliations related to the swine flu. If you have any questions and/or comments for this paper, please feel free to contact the corresponding author, ***Akisato Kimura <akisato@ieee.org>***. Some demonstration movies can be seen in a web site ***http://www.brl.ntt.co.jp/people/akisato/saliency3.html***. Sorry for the inconvenience.

Background :

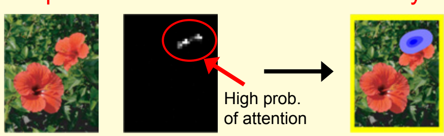
- Extracting "important" regions from videos is a challenging and crucial task
  - Especially for video compression, object recognition, video annotation and retrieval, etc.
- It can be formulated as a problem of binary segmentation.
  - Important regions = "objects", the remaining regions = "backgrounds"
- A promising way for precise segmentation: graph-cuts based methods
  - Interactive Graph Cuts [Boykov 2006], extension to videos [Kohli 2007], etc.

Problem 1: Need to provide cues for segmentation manually and carefully.

[White] "obj" labels
[Black] "bkg" labels

Input frame with labels

Contributions 1: Segmentation priors are provided based on visual saliency.



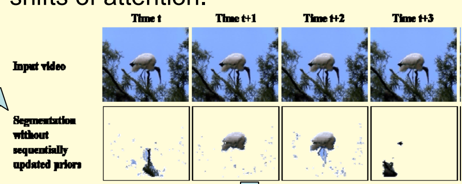
Input frame Attention Priors

[Blue] High prior, Ex. $p(A_x = \text{"obj"}) = 0.8$

[Transparent] Low prior, Ex. $p(A_x = \text{"obj"}) = 0.2$

[Yellow] Zero prior, e.g. $p(A_x = \text{"obj"}) = 0$

Problem 2: Segmented regions may be randomly switched as a result of the shifts of attention.

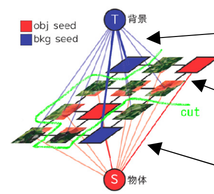


Contributions 2: Sequential update of priors with previous results

Original prior Segmentation Updated prior

Segmentation with Graph Cuts

Minimizing the energy function is equivalent to deriving the minimum cut of the following (directed) graph.



Cost of edge from each pixel to sink: $\psi_1(D|A_x = \text{"obj"}) + \xi_1(A_x = \text{"obj"})$

Cost of edge between neighboring pixels: $\psi_2(D|A_x, A_y) + \xi_2(A_x, A_y)$

Cost of edge from source to each pixel: $\psi_1(D|A_x = \text{"bkg"}) + \xi_1(A_x = \text{"bkg"})$

There exists several ways to derive the minimum cut in polynomial time. [Ford-Fulkerson 1956] [Boykov 2001]

↓

The MAP solution of the MRF for segmentation can be exactly solved in polynomial time.

Attention-based priors (Contribution 1)

Prior term

$\xi_1(A_x) = -\log p(A_x)$

Prior density for "obj" $p(A_x = \text{"obj"})$

- GMM fitting of the density of visual attention
- For the edge of each frame, $p(A_x = \text{"obj"}) = 0$

Prior density for "bkg"

$p(A_x = \text{"bkg"}) = 1 - p(A_x = \text{"obj"})$

Likelihood term

$\psi_1(D|A_x) = -\log p(D|A_x)$

Likelihood for "obj" $p(D|A_x = \text{"obj"})$

- GMM of RGB color features estimated with EM
- Samples are weighted by the prior density

Likelihood for "bkg" $p(D|A_x = \text{"bkg"})$

- GMM of RGB color features estimated with EM
- Samples are taken from the edge of each frame

Sequential update of priors (Contribution 2)

Updating prior density for "obj"

Assume the following two relationships:

$$\Pr\{p(A_{x,t} = \text{"obj"}; t)\} = \mathcal{N}(f(\widehat{a}_{x,t-1}), \sigma_1)$$

- A position whose label was "obj" at time t-1 is likely to have a label "obj" at time t.

$$\Pr\{q(A_{x,t} = \text{"obj"}; t)\} = \mathcal{N}(p(A_{x,t} = \text{"obj"}; t), \sigma_2),$$

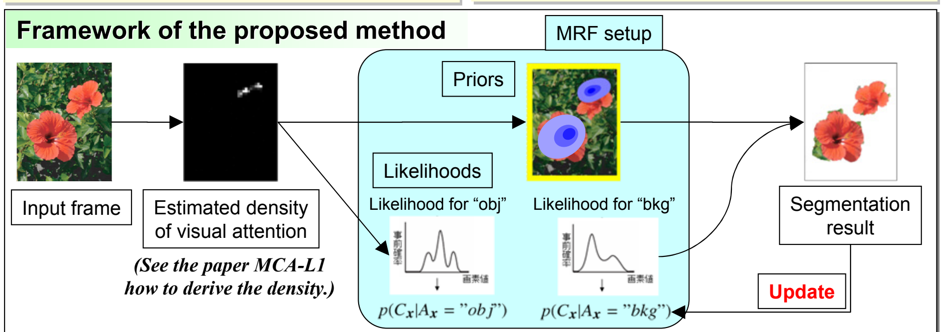
- A prior density $p(A_{x,t} = \text{"obj"}; t)$ at time t is quite similar to the prior density $q(A_{x,t} = \text{"obj"}; t)$ derived only from the density of visual attention at time t.

- The prior density $p(A_{x,t} = \text{"obj"}; t)$ at time t can be derived through Kalman filter, where the observation is $q(A_{x,t} = \text{"obj"}; t)$

$$\widehat{p}(A_{x,t} = 1; t) = \frac{\sigma^2_{\xi_1}(t)}{\sigma^2_2 + \sigma^2_{\xi_1}(t-1)} f(\widehat{a}_{x,t-1}) + \frac{\sigma^2_{\xi_1}(t)}{\sigma^2_1} q(A_{x,t} = 1; t),$$

$$\sigma^2_{\xi_1}(t) = \frac{\sigma^2_1 \cdot (\sigma^2_2 + \sigma^2_{\xi_1}(t-1))}{\sigma^2_1 + \sigma^2_2 + \sigma^2_{\xi_1}(t-1)},$$

Framework of the proposed method



(See the paper MCA-L1 how to derive the density.)

Energy function of MRF for segmentation

$$E(A|D) = \sum_{x \in I} \left\{ \underline{\psi_1(D|A_x)} + \underline{\xi_1(A_x)} + \sum_{y \in N_x} \left(\underline{\psi_2(D|A_x, A_y)} + \underline{\xi_2(A_x, A_y)} \right) \right\},$$

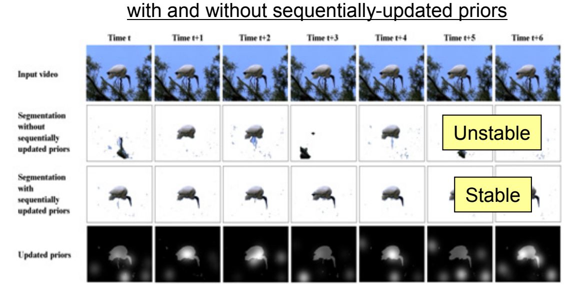
(ξ_i: prior terms, ψ_i: likelihood terms)

Based on GMM of prior-weighted color features | attention-based | Similarity of intensities among neighboring pixels | Generalized Potts model

Snapshots

Examples of segmentation results

Comparisons of the proposed method with and without sequentially-updated priors



Special thanks to Mr. Derek Pang of Simon Fraser Univ., Canada for supporting this presentation.