Automatic Object Extraction in Single-Concept Videos
Kuo-Chin Lien and Yu-Chiang Frank Wang
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

Goal
Extracting the main concept (i.e. foreground object) of a video without user interaction or training data

Our Approach
- Learning of sparse shape representation of objects
- Ability to observe both foreground and background color models via GMM
- Integration of motion, shape, and color features by Conditional Random Field (CRF)

Contribution
- Ability to extract single concept with multiple instances
- Allows marginal but complex background motion
- No prior knowledge or assumptions on objects required (e.g. scale, pose, object category, etc.)
- No need to collect training data of any particular objects

System Flow

Collecting motion-driven salient parts
- Determine moving pixels \(q_t\) at frame \(t\) by forward-backward optical flow

\[q_t = \hat{q}_{t,t-1} \cup \hat{q}_{t,t+1}\]
- Each moving part of a foreground object is assumed to form a complete sampling of the entire object of interest.

Learning of sparse shape representation
- Extraction of HOG feature from moving parts
- Learning of shape dictionary \(D\) under sparse constraint

\[\min_{\mathbf{h}_k} \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{h}_n - \mathbf{D} q_n \| + \lambda \|\mathbf{\alpha}_n\|_1\]

(a) Example codewords for sparse shape representation

(b) Corresponding image patches (only top 5 matches)

Learning of color models
- Modeling of foreground and background color via GMM

Estimated color distributions of foreground and background

Foreground object extraction by minimizing total energy

Integration of multiple cues by CRF

Experimental Results & Comparison

<table>
<thead>
<tr>
<th>Input video</th>
<th>Felzenszwalb et al. PAMI’10</th>
<th>Babenko et al. CVPR’09</th>
<th>Liu and Gleicher CVPR’09</th>
</tr>
</thead>
</table>

Our method