Introduction

• Motion-based image segmentation
  – Fundamental to object based video coding
  – Motion Estimation ⇔ Motion Segmentation

• Common approach:
  – MAP formulation on dense fields using MRFs
  – Parametric fitting within segments

• Our Approach:
  – Direct ML formulation using parametric models and support maps
  – Solution using Mean Field Theory in EM procedure
Parametric Motion Field Modeling

- $I(t)$ can be related to $I(t - 1)$ by
  
  $I(x_i, t) \approx I(x_i - u_{\theta}(x_i), t - 1) \quad \forall i \in S$

- For $k$th model ($k = 1, 2, \ldots K$) form
  
  - **Predicted:** $\tilde{I}^k(x_i, t) = I(x_i - u_{\theta^k}(x_i), t - 1)$
  
  - **Residual:** $r^k(x_i, t) = I(x_i, t) - \tilde{I}^k(x_i, t)$
  
  - Assume $r^k(t) \sim \mathcal{N}(0, \sigma^k)$

- **Parameter vector:** $\Phi(t) = [\Theta(t) \quad \Sigma(t)]^T$
  
  where $\Theta(t) = [\theta^1 \theta^2 \ldots \theta^K]^T$ and $\Sigma(t) = [\sigma^1 \sigma^2 \ldots \sigma^K]^T$
Segmentation Modeling

- Segmentation, $s(t)$, assigns each pixel $x_i$ to a unique model

$$s_i(t) = [s_1^i(x_i, t) \ s_2^i(x_i, t) \ \ldots \ s^K_i(x_i, t)]^T$$

where

$$s^K_i(x_i, t) = \begin{cases} 1 & : I(x_i, t) \in \tilde{I}^k(t) \\ 0 & : I(x_i, t) \notin \tilde{I}^k(t) \end{cases}$$

- 

$$p(s(t)) = \frac{1}{Z} \exp \left[ -\beta \sum_{i=1}^{N} \left( \lambda_1 \sum_{j \in N_i} [1 - 2\delta(\mathcal{H}(s_i(t) - s_j(t))) \right) \right]$$

$$\delta(z_i(t) - z_j(t)) \right) \right]$$

$$= \frac{1}{Z} \exp[-\beta \ U(s(t))]$$

$\delta$: Kronecker delta function

$\mathcal{H}$: Hamming distance function

$z(t)$: Static (gray level) segmentation field of $I(t)$
Maximum Likelihood Formulation

\[
\hat{\Phi}_{ML}(t) = \arg \max_{\Phi(t)} \left\{ p(I(t)|I(t-1), \Phi(t)) \right\}
\]

\[
= \arg \max_{\Phi(t)} \left\{ \log[p(I(t)|I(t-1), \Phi(t))] \right\}
\]

- Part of data \( (s_i(t)) \) is \textit{unobservable} or \textit{hidden}
  \( \implies \textbf{Incomplete} \) data problem

- Solve using Expectation Maximization procedure
**Expectation Maximization Procedure**

\[ \mathbf{0} = \{ o_i, i \in \mathcal{S} \} : \textit{observations} with likelihood } p(\mathbf{0}|\Phi_o, \mathbf{h}) \]

\[ \mathbf{h} = \{ h_i, i \in \mathcal{S} \} : \textit{hidden variables} with prior } p(\mathbf{h}|\Phi_h) \]

To solve

\[ \hat{\Phi}_{ML} = \arg \max_{\Phi} \left\{ \log p(\mathbf{0}|\Phi) \right\} \]

Alternate between

**E step:** \[ Q\left(\Phi|\Phi^{(p)}\right) = E\{ [\log p(\mathbf{0}|\mathbf{h}, \Phi_o) + \log p(\mathbf{h}|\Phi_h)] | \mathbf{0}, \Phi^{(p)} \} \]

**M step:** \[ \Phi^{(p+1)} = \arg \max_{\Phi} Q\left(\Phi|\Phi^{(p)}\right) \]
Expectation Step

- Compensate for $s(t)$ with conditional expectation
  \[ g_i(t) = E^* [s_i(t)] = E^* \left\{ [s^1(x_i, t) \quad s^2(x_i, t) \quad \ldots \quad s^K(x_i, t)]^T \right\} \]
  where
  \[ g^k(x_i, t) = E[s^k(x_i, t)|I(t), I(t-1), \Phi^{(p)}(t)] \]
  \[ = \text{Prob} \quad [I(x_i, t) \in \tilde{I}_k(t)|I(t), I(t-1), \Phi^{(p)}(t)] \]
  \[ \implies \text{In effect, compute “soft segmentation”!} \]
- $s$ is an MRF $\implies s|I(t), I(t-1), \Phi^{(p)}$ is an MRF

  \[ E[s_i|I(t), I(t-1), \Phi^{(p)}] = \sum_s s_i \quad p(s|I(t), I(t-1), \Phi^{(p)}) \]

  \[ = Z^{-1}(\Phi^{(p)}) \sum_s s_i \exp[-\beta U(s|o, \Phi^{(p)})] \]

Note: Precise calculation is exponentially complex!

- Use Mean Field Approximation:

  \[ E[s_i|o, \Phi^{(p)}] \simeq \frac{1}{Z_i^{mf}} \sum s_i \exp(-\beta U_i^{mf}) \]

  where

  \[ U_i^{mf} = s_i^T \left[ \frac{-1}{\beta} W_1(o_i, \Phi^{(p)}) \right] + \sum_{j \in \mathcal{N}_i} s_i^T V_2(\Phi^{(p)}) E[s_j|o, \Phi^{(p)}] \]
\[ Z_{i}^{mf} = \text{local partition function} \]
\[ W_{1}(o_{i}, \Phi_{o}) : K \text{ dim. vector, } W_{1}(k) = \log p(o_{i}|h_{i} = e_{k}, \Phi_{o}) \]
\[ V_{2}(\Phi_{h}) : K \times K \text{ matrix, } V_{2}(k, l) = V_{c}(h_{i} = e_{k}, h_{j} = e_{l}) \]
\[ e_{k} : K \text{ dim. vector with 1 at } k\text{th component and 0 elsewhere} \]

- Parallel implementation possible using coding method!
Maximization Step

1. Recover intermediate binary labeling
   \[
   l_i^j = 1 \quad \text{iff} \quad j = \arg \max_{k \in 1, 2, \ldots, K} E[s_i^k|o, \Phi^{(p)}]
   \]

2. Update \( \sigma^k \) using
   \[
   \sigma^{k,(p+1)} = 1.4826 \quad \text{median}_{i: l_i^k = 1} |r^k(x_i, t)|
   \]

3. Update \( \Theta^k \) using
   \[
   \Theta^{k,(p+1)}(t) = \arg \min_{\Theta(t)} \sum_{i=1}^{N} E[s_i^k|r, \Phi^{(p)}(t)] - \log p^k(r^k(x_i, t)|\Phi(t))
   \]
Results
Extensions for Content-based Video Coding

- Temporally coherent labeling of video objects
- Per-object coding of shape, texture and motion