Paper Reading Seminar

September 23, 2013

Overview of the paper review

- We only care about
  - (Are there already standard approaches?)
  - How do they define the problem and the objective?
  - Why is this problem challenging?
  - How do the other researchers solve them?
- $\Rightarrow$ We don't review the experiment sections here

Paper 1

- Fernando de la Torre et. al. Temporal Segmentation of Facial Behavior, ICCV 2007

Problem definition

- Temporal segmentation of facial behavior
- The authors don't give a specific definition. And my understanding is
  - The result segments are expected to be similar within themselves, but different between each other
  - Elements in each segment should be continuous in time
  - There may also be classification involved considering this paper is also about facial expressions

Challenges

- Mainly the unreliability of face tracking
  - Non-frontal pose
  - Moderate out-of-plane head motion
  - Subtle facial actions
- Also some challengings in the algorithm aspect
  - Large variability in the temporal scale

Approach

- Goal
  - Robust to 2D affine transforms
  - (Background introduction: geometric transforms and RANSAC)

Approach

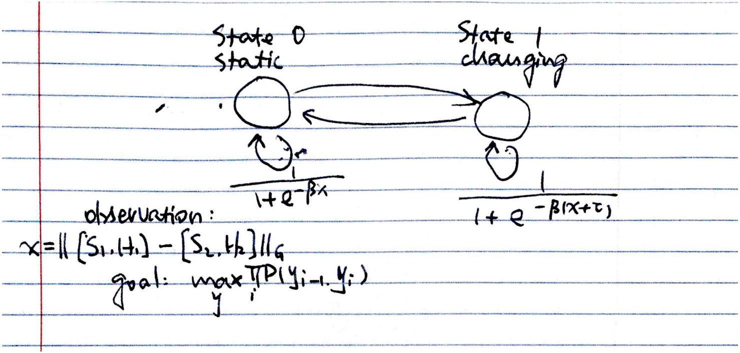
- Framework: feature clustering + temporal clustering (process individually)
- Feature clustering
  - Feed pairwise similarity to *spectral clustering*
  - (Background introduction: graph-cut and normalized-cut)
  - Distance computation: consider 2D transforms
    - Point coordinates $S_1, S_2 \Rightarrow H$
    - Dist $= \mathcal{N}(s_i - Hs_j)\mathcal{N}(h_i - h_j)$

Approach

- Temporal clustering
  - Biased (natural) pose detection and elimination
    - Facial temporal clustering is special because it has a biased/natural pose
    - Frames =HMM=> “static” poses =Spectral Clustering=> natural poses

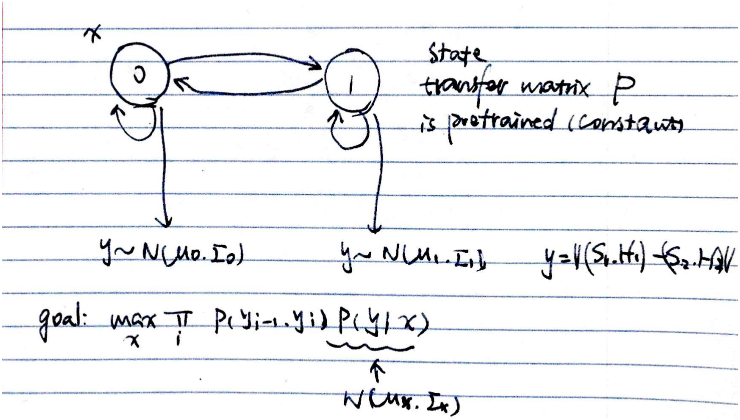
Approach

- Details about HMM



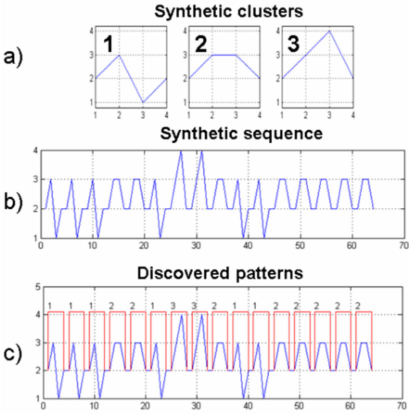
Approach

- Details about HMM



Approach

- Clustering of other poses
  - Temporal greedy scanning



Some thoughts

- HMM looks a slightly better version of feature difference thresholding
- The overall framework is just greedy matching, while largest cluster gets double-confirmed by HMM
  - Our first baseline?
- Feature clustering may act as an intermediate representation
  - May be sensitive. Let's first test whether it has real-world meanings...
- Greedy scanning may act as our first baseline?
  - Can track papers citing this paper to for other baselines
- HMM may be an interesting direction to pursue? (with efficient max-margin inference?)

Paper 2

- Minh Hoai, Fernando de la Torre, Maximum Margin Temporal Clustering, AISTATS 2012

Problem definition

- Factorization of multiple time series into a set of non-overlapping segments that belongs to k temporal clusters
- Previous work
  - Hidden Morkov Model
  - Dynamic Bayesian Network
  - Support Vector Machine
- The related work section is very useful

Multi-class Max Margin Clustering

- Background introduction: SVM)

$$
\begin{aligned}
&\min \frac{1}{2m} \sum_j \|w_j\|^2 + C \sum_i \xi_i \\
&\text{s.t. } \forall i : w_{y'}^T x_i - w_y^T x_i \geq 1 - \xi_i, \forall y \neq y' \\
&\qquad \forall j, j' : -\lambda \leq (w_j - w_{j'})^T \sum_i x_i \leq \lambda
\end{aligned}
$$

- Cluster balance?

Membership requirement MMC

- Formulation

$$
\begin{aligned}
&\min \frac{1}{2m} \sum_j \|w_j\|^2 + C \sum_i \xi_i + C_2 \sum_j \beta_j \\
&\text{s.t. } \forall i : w_{y'}^T x_i - w_y^T x_i \geq 1 - \xi_i, \forall y \neq y' \\
&\quad \forall j : \exists l \text{ different indexers } i\text{s} : \\
&\quad w_j^T x_i - w_{j'}^T x_i \geq 1 - \beta_j, \forall j' \neq j
\end{aligned}
$$

- Soft constraint requiring each cluster to have at least l members
- Optimize with coordinate descent

Joint segmentation and clustering

- Adding changing points $\{s_i\}$ (heuristic)

$$\underset{\substack{\mathbf{w}_j, k_i, s_t^i, y_t^i \\ \xi_t^i \geq 0, \beta_j \geq 0}}{\text{minimize}} \quad \frac{1}{2m} \sum_{j=1}^{m} ||\mathbf{w}_j||^2 + C \sum_{i=1}^{n} \sum_{t=1}^{k_i} \xi_t^i + C_2 \sum_{j=1}^{m} \beta_j, \quad (7)$$

$$\text{s.t. } \forall i, t : s_{t+1}^i - s_t^i \leq l_{max}, s_1^i = 0, s_{k_i+1}^i = n_i, \quad (8)$$

$$\forall i, t : (\mathbf{w}_{y_t^i} - \mathbf{w}_y)^T \varphi(\mathbf{X}^i_{(s_t^i, s_{t+1}^i]}) \geq 1 - \xi_t^i \ \forall y \neq y_t^i, \quad (9)$$

$$\forall j : \exists\ l \text{ segments, i.e., index pairs } (i, t) :$$

$$(\mathbf{w}_j^T - \mathbf{w}_{j'}^T)\varphi(\mathbf{X}^i_{(s_t^i, s_{t+1}^i]}) \geq 1 - \beta_j \ \forall j' \neq j. \quad (10)$$

- Optimize with block coordinate descent and alternating optimization

Some thoughts

- Large-margin approaches is appealing in that it also selects the important features (especially with l_1 norms)
- The training process may be troublesome, requiring upper bound approximation and cutting-plane optimization. But may also bring more technical contributions