Paper Reading Seminar

Yan Wang

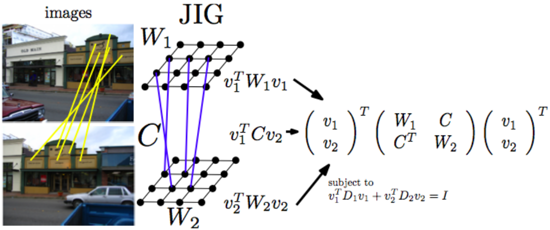
October 3, 2012

Image Matching via Saliency Region Correspondences

- Intuition
  - Segmentation is not perfect. Weak connections $\Rightarrow$ graph-based segmentation (NCut)
  - Matching is not perfect. Weak connections $\Rightarrow$ spatial consistency check
  - Combine the two tasks together $\Rightarrow$ use a single graph to get a joint optimal
  - Encoding context in the matching process $\Rightarrow$ no need for spatial consistency check
- Outline
  - Formulation
  - Optimization
  - Features

Formulation

- Graph formulation
  - Vertex: pixel
  - Layer: two-layer, one for each image
  - Intra-layer edges: how strongly the pixels are similar (connected) in this image
  - Inter-layer edges: how strongly the pixels are similar (connected) across images



Formulation

- Intra-layer optimization goal

$$\max_{v} \frac{v_1^T W_1 v_1 + v_2^T W_2 v_2}{v^T D v}$$

  - Binary segmentation problem (extend to multi-class with 1-vs-all binary mask)
  - Each segment should be internally consistent $\Rightarrow v_1^T W_1 v_1$. Only if v_{1j}, v_{1k} both positive, will $W_{1(i,j)}$ be counted
  - $v_i = \{0, 1\}^{n_i}, \sum_j v_{ij} = 1$, indicator vector for image i, pixel j, and cluster k, $v = [v_1^T, v_2^T]^T$
  - Normalize on the size: $\sum_j v_{1j}^T W_1 = \mathbf{1}^T W_1 = D_1$

Formumation

- Inter-layer optimization goal

$$\max_{v} \frac{v_1^T C v_2}{v^T D v}$$

  - "Co-saliency" regions (where $v_{ij} = v_{ik}$) should be similar
  - Also normalize on size
  - Use "context" to refine segmentation as well as matching

Formulation and optimization

- Final optimization goal

$$F(v^{(c)}, C) = \mathrm{IntraIS}(v^{(c)}, C) + \mathrm{InterIS}(v^{(c)}, C)$$

$v^{(c)}$: binary segmentation result for the cth segment.

- Relaxation for optimizaiton
  - v to real vector
  - Joint "soft" segmentation result V (set of eigenvectors) lies in the subspace spanned by the individual "soft" segmentation results S_1 and S_2.
  - Relax the synchronization with orthogonal constraints.
  - EM iterative optimization, with no explicit proof of convergence (in some speicial cases, EM algorithms can arrive in stationary points)

Features

- MSER detector + SIFT descriptor
- Intra-image W
  - x, y are considered in the same segment $\Leftrightarrow$ no edges with large magnitude spatially separate them
  - Edge detection with large magnitude $\Rightarrow$ get W
- Inter-image C
  - 1. Feature detection: also consider the ellipse (orientation/scale) from the detector
  - 2. Feature matching
  - 2.0 Simplest pixel-wise matching: Gaussian kernel + descriptor + ellipse matrix

$$m_{x,y}(p,q) = e^{-\|d_p - d_q\|^2/\sigma_i^2} e^{-\|H_p(x) - H_q(y)\|^2/\sigma_p^2}$$

Features

- Inter-image C
  - 2.1 Adopt patch-wise feature matching score as pixel-wise score for C

    $$M_{x,y} = \max\{m_{x,y}(p,q) \mid p \in P, q \in Q, x \in R_p, y \in R_q\}$$

    One point can be involved in multiple MSER interest regions. For every point pair x, y, check all the MSER pairs (with different H settings), and pick up the highest matching score
  - 2.2 Compute patch-wise similarity matrix M, and then normalize to C

    $$D_1^{-1/2} C D_2^{-1/2} = P \circ M$$