Objective

Jointly summarize large sets of online blog posts and photo streams in a mutually-rewarding way.

Input: large sets of blog posts and photo streams.

What can blog posts do for photo streams?

1. Automatic Image titling
   - Disneyland California adventure
   - Radiator Springs racers

2. Image Localization
   - Storybook Land canal boats
   - Stay fairly crowded

What can photo streams do for blog posts?

- Photo streams help interpolation between blog images.
- Blogs transfer semantic knowledge to photo streams.

Problem Statement

Input: A set of photo streams \{P_1, \ldots, P_l\} and a set of Blogs \{B_1, \ldots, B_l\}.

- Each blog is decomposed by NLP tools.

\[ B_l = \{(i_l^j, m_l^j, v_l^j), \ldots, (i_l^n, m_l^n, v_l^n)\} \]

Two key problems:

1. Alignment from blog pictures to photo streams.
2. Photo stream summarization.

Goal: Build similarity links from blog pictures to photo streams (i.e., a bipartite graph \( W \in \mathbb{R}^{l \times p} \)).

- (A1) Sparsity: \( W \) should be sparse.
- (A2) Similarity: If a blog image \( i \) is more similar to a photo \( j \) than \( k \), \( W_{ij} > W_{ik} \).
- (A3) Summary: Blog images are encouraged to align to the image in \( S \).
- (A4) Continuity: Consecutive blog images are matched to the same photo stream.

Goal: Predict a best subset \( S' \subset P \).

- (S1) Alignment: \( S' \) should have more inlinks as possible (i.e., \( W_{S'} \) have more non-zeros).
- (S2) Coverage and Diversity: Summary \( S' \) should contain fewer redundant images, and not miss any important images of photo streams.

Approach

- A formulation using Latent Ranking SVM.
- Alignment \( W \) requires the output of summarization \( S' \), which requires the output of alignment \( W \).

1. Use K-means to obtain initial \( S' \subset P \).
2. Compute \( W \in \mathbb{R}^{l \times p} \) by solving
   \[ \min_{w_{ij}} \left\| \frac{1}{l} \sum_{i \in S'} w_{ij} \right\|_1 \quad \text{s.t.} \quad W_{ij} - W_{ik} \geq \Delta(s_i, s_k) \Rightarrow \zeta \]
   - (A1)-(A4) Implementation constraints.
3. Update \( S' \subset P \) by solving
   \[ \min_{s_i} \left\{ \frac{1}{l} \sum_{i \in S'} \left( W_{ij} - W_{ik} \right) \right\} \quad \text{s.t.} \quad \Delta(S_i, S_k) \leq \zeta \]
   - Alternately solve step 2-3.

Results of Image Localization

Goal: Find at which attraction an image is likely to be taken

- Qualitative localization examples:
  - Minnie’s House (0.3001)
  - Storybook Land Canal Boats (0.4694)
  - Mickey’s Soundsational Parade (0.2028)
  - Mr. Toad’s Wild Ride (0.0173)

- Quantitative results:
  - Top-1 Attr: 9.12% vs. Original title 66.5% (997/1500)
  - Top-5 Attr: 22.83% vs. Original title 66.5% (997/1500)
  - Top-1 Dist: 24.27% vs. Original title 66.5% (997/1500)

Results of Automatic Image Titling

Goal: Automatically generate a descriptive title for an image

- Qualitative titling examples:
  - Cutting-Plane Training of Structural SVMs.
  - Vision-based method

- AMT evaluation: Estimated title by blogs vs original one
  - vs. Original title

Results of Interpolation between Blog Pictures

Goal: Interpolate consecutive blog images with photo stream images

- Qualitative results:
  - AMT evaluation: Comparison btw our approach vs. other baselines?
    - vs. Original title
    - vs. VSVM

Results of Interpolation between Blog Pictures

Goal: Interpolate consecutive blog images with photo stream images

- Qualitative results:
  - AMT evaluation: Comparison btw our approach vs. other baselines?
    - vs. Original title
    - vs. VSVM

Reference