Joint Photo Stream and Blog Post Summarization and Exploration

Gunhee Kim       Seungwhan Moon       Leonid Sigal

June 29, 2015
Outline

- Problem Statement
- Approach
- Experiments
- Conclusion
Background

Online sharing of our experiences becomes popular

Photos

Videos

Blogs

Experiences are recorded in different forms (ex. Visiting Disneyland)
Our Objective

Jointly summarize online sets of **photo streams** and **blog posts**

- Two two social media modalities are *mutual-rewarding*

What can **blog posts** do for **photo streams**?

---

**Mater’s Junkyard Jamboree** features a square dance in a wagon pulled by a baby Tractor.

Blogs are written in a way of *storytelling*

- Blog pictures are selected as the most canonical ones out of photo albums
- Informative info about location, sentiments, time, actors, ...
Our Objective

Jointly summarize online sets of **photo streams** and **blog posts**

- Two two social media modalities are **mutual-rewarding**

What can **blog posts** do for **photo streams**?

Blogs transfer semantic knowledge to photo streams

Ex1) Automatic Image Titling

![Image](disneyland-california-adventure)

**radiator springs racers**

![Image](img_3467)

**storybook land canal boats stay fairly crowded**

Ex2) Image Localization

![Image](Jumpin' Jellyfish)

**Jumpin' Jellyfish (0.2028)**

![Image](Jedi Training Academy)

**Jedi Training Academy (0.4786)**
Our Objective

Jointly summarize online sets of **photo streams** and **blog posts**

- Two social media modalities are **mutual-rewarding**

What can **photo streams** do for **blog posts**?

Photo streams help interpolation between blog images
Problem Statement

**(Input1)** A set of photo streams and blog posts

**(Output1)** Alignment from blog pictures to photo streams

**(Output2)** Photo Stream Summarization

**(Application1: Blogs ➔ Photos)**
Semantic Knowledge Transfer

Automatic Image Titling  Localization

dsc00117.jpg  Autopia

**(Application2: Blogs ← Photos)**
Interpolation between Blog Pictures

original submarine voyage
Disneyland Dataset for Photos and Blogs

Photo Streams from **flickr**

- 542,217 unique images of 6,026 valid photo streams

Blog posts

- 53,091 unique blog posts and 128,563 associated pictures

<table>
<thead>
<tr>
<th>Labels</th>
<th>Blogger</th>
<th>WordPress</th>
<th>TypePad</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ALL</strong></td>
<td>15,257 (74,218)</td>
<td>37,240 (53,467)</td>
<td>594 (878)</td>
</tr>
<tr>
<td><strong>Travelogues</strong></td>
<td>5,152 (71,934)</td>
<td>4,815 (48,554)</td>
<td>108 (763)</td>
</tr>
<tr>
<td><strong>Disney</strong></td>
<td>3,270 (58,311)</td>
<td>2,350 (33,831)</td>
<td>28 (378)</td>
</tr>
<tr>
<td><strong>Junks</strong></td>
<td>1,638 (12,268)</td>
<td>1,682 (9,123)</td>
<td>56 (299)</td>
</tr>
</tbody>
</table>

# posts (# images)
Outline

• Problem Statement
• Approach
• Experiments
• Conclusion
Blog Processing

For each blog post from $B = \{B^1, \ldots, B^N\}$

$$B^n \rightarrow \{(I^n_1, m^n_1, v^n_1), \ldots, (I^n_{Nn}, m^n_{Nn}, v^n_{Nn})\}$$

where $m^n_i = \{l^n_i, k^n_i\}$

Locations Key phrases

<table>
<thead>
<tr>
<th>Location</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lightening and Martar</td>
<td>0.4</td>
</tr>
<tr>
<td>Cozy Cone Motel</td>
<td>0.32</td>
</tr>
<tr>
<td>Lightening and Martar to sit in the driveway</td>
<td>0.7</td>
</tr>
<tr>
<td>Cozy Cone Motel for photo shoots with the kids</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Blog Processing

For each blog post from $B = \{B^1, \cdots, B^N\}$

$$B^n \rightarrow \{ (I^n_1, m^n_1, v^n_1), \cdots, (I^n_N, m^n_N, v^n_N) \}$$

- **Image Meta-data Confidences**
- **Locations Key phrases**

where $m^n_i = \{l^n_i, k^n_i\}$

- Location extraction
  CRF-based named entity recognizer (NER)

- Key phrases extraction
  Unsupervised automatic keyword extraction

- Confidence of Photo Association
  Weighted by the distance btw text blocks and images
Joint Exploration btw Blogs and Photo Streams

Input: A set of blogs $\mathcal{B} = \{B^1, \ldots, B^N\}$ and photo streams $\mathcal{P} = \{P^1, \ldots, P^L\}$

Two key subproblems

Alignment from blog images to photo streams

- Goal: Build similarity links from blog pictures to photo streams ($i.e.$ A bipartite graph $\mathbf{W} \in \mathbb{R}^{\mathcal{I} \times \mathcal{P}}$)

(A1) $\mathbf{W}$ should be sparse

(A2) Blog photos are encouraged to align to the images in $\mathcal{S}^l$

(A3) Consecutive blog images are to matched to the same photo stream

Photo stream Summarization

- Goal: Predict a best subset $\mathcal{S}^l \subseteq \mathcal{P}^l$
**Joint Exploration btw Blogs and Photo Streams**

Input: A set of blogs $\mathcal{B} = \{B^1, \ldots, B^N\}$ and photo streams $\mathcal{P} = \{P^1, \ldots, P^L\}$

Two key subproblems

- Alignment from blog images to photo streams
  - Goal: Build similarity links from blog pictures to photo streams (i.e. A bipartite graph $\mathbf{W} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{P}|}$)

- Photo stream Summarization
  - Goal: Predict a best subset $S^l \subset P^l$

(S1) $S^l$ should have more inlinks as possible (i.e., $\mathbf{W}_{S^l}$ have more non-zeros)

(S2) $S^l$ should have high coverage and diversity
Joint Exploration btw Blogs and Photo Streams

Input: A set of blogs $\mathcal{B} = \{B^1, \ldots, B^N\}$ and photo streams $\mathcal{P} = \{P^1, \ldots, P^L\}$

Two key subproblems

Alignment from blog images to photo streams

- Goal: Build similarity links from blog pictures to photo streams (i.e. A bipartite graph $\mathbf{W} \in \mathbb{R}^{\mathcal{I} \times \mathcal{P}}$)

  (A1) $\mathbf{W}$ should be sparse
  
  (A2) Blog photos are encouraged to align to the images in $\mathcal{S}^l$
  
  (A3) Consecutive blog images are to matched to the same photo stream

Photo stream Summarization

- Goal: Predict a best subset $\mathcal{S}^l \subset \mathcal{P}^l$

  (S1) $\mathcal{S}^l$ should have more inlinks as possible (i.e., $\mathbf{W}_{\mathcal{S}^l}$ have more non-zeros)

  (S2) $\mathcal{S}^l$ should have high coverage and diversity
Joint Exploration btw Blogs and Photo Streams

Input: A set of blogs $\mathcal{B} = \{B^1, \cdots, B^N\}$ and photo streams $\mathcal{P} = \{P^1, \cdots, P^L\}$

Two key subproblems

1. Use K-means to obtain $S^l \subset P^l$

2. Compute $\mathbf{W} \in \mathbb{R}^{I \times |\mathcal{P}|}$ by solving the following ranking SVM

$$\min_{\mathbf{W}, \xi} \frac{1}{2} \|\mathbf{W}\|_1 + \frac{\lambda_A}{M} \sum_{i=1}^{M} \xi_i$$

s.t. $\forall (i, j, k) \in C_d \cup C_s$: $W^n_{ij} - W^n_{ik} \geq \Delta(\sigma_{ij}, \sigma_{ik}) - \xi_i$

$\forall (i, j, k) \in C_c$: $W^n_{ij} - W^n_{ik} \geq \Delta(#_{ij}, #_{ik}) - \xi_i$

(A1) $\mathbf{W}$ should be sparse

(A2) Blog photos are encouraged to align to the images in $S^l$

(A3) Consecutive blog images are to matched to the same photo stream
Joint Exploration btw Blogs and Photo Streams

Input: A set of blogs $B = \{B^1, \ldots, B^N\}$ and photo streams $\mathcal{P} = \{P^1, \ldots, P^L\}$

Two key subproblems

Alignment from blog images to photo streams

Photo stream Summarization

(1) Use K-means to obtain $S^l \subset P^l$

(2) Compute $W \in \mathbb{R}^{I \times |\mathcal{P}|}$ by solving the following ranking SVM

(3) Update $S^l \subset P^l$ by solving the following ranking SVM

(S1) $S^l$ should have more inlinks as possible (i.e., $W_{S^l}$ have more non-zeros)

(S2) $S^l$ should have high coverage and diversity
Outline

• Problem Statement
• Approach
• Experiments
• Conclusion
Results of Image Localization

Goal: Find out at which attraction a given image is taken

• Two CA parks: Disney California Adventure Park and Disneyland
• 108 selected attractions and restaurants from 18 districts
• Groundtruth: Human experts labels 3,000 images of locations
• Ten experiments with randomly selected 2,000 images.

Quantitative results

<table>
<thead>
<tr>
<th>Method</th>
<th>Top-1 Attr.</th>
<th>Top-5 Attr.</th>
<th>Top-1 Dist.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(JointRSVM)</td>
<td>9.12%</td>
<td>22.83%</td>
<td>28.81%</td>
</tr>
<tr>
<td>(KNN+KM)</td>
<td>7.34%</td>
<td>18.62%</td>
<td>24.27%</td>
</tr>
<tr>
<td>(DTW+KM)</td>
<td>4.05%</td>
<td>15.31%</td>
<td>21.03%</td>
</tr>
<tr>
<td>(VKNN)</td>
<td>5.16%</td>
<td>16.85%</td>
<td>22.12%</td>
</tr>
<tr>
<td>(VSVM)</td>
<td>4.63%</td>
<td>15.80%</td>
<td>20.63%</td>
</tr>
<tr>
<td>(Rand)</td>
<td>0.93%</td>
<td>4.63%</td>
<td>5.56%</td>
</tr>
</tbody>
</table>

(JointRSVM): our method.
(KNN/DTW+KM): KNN/DTW alignment + K-means summarization
(VKNN)/VSVM: Vision-based method using KNN/SVM classifiers
Examples of Image Localization

Success

(JointRSVM) Minnie's House (0.3001)
(KNN+KM) Mickey's Soundsational Parade (0.2026)

(JointRSVM) Storybook Land Canal Boats (0.4694)
(DTW+KM) Mr. Toad's Wild Ride (0.0173)

(JointRSVM) Autopia (0.5986)
(VKNN) Mad Tea Party (0.4988)

(JointRSVM) Luigi's Flying Tires (0.1376)
(KNN+KM) The Disney Gallery (0.01163)

(JointRSVM) Astro Orbiter (0.0188)
(KNN+KM) Pirate's Lair on Tom Sawyer Island (0.0183)

(JointRSVM) Mickey's Fun Wheel (0.0183)
(VSVM) Ariel's Undersea Adventure (0.1984)

Near-miss

1. Muppet Vision 3D (0.019786)
2. Sleeping Beauty Castle Walkthrough (0.019637)

1. Goofy's Sky School (0.019259)
2. Twilight Zone Tower of Terror (0.019088)

1. Pixar Play Parade (0.017526)
2. Big Thunder Ranch (0.017462)
Results of Automatic Image Titling

Goal: Automatically generate an informative title of an image

- Titles of online images are (i) frequently missing or (ii) meaningless (e.g., IMG1136.jpg).
- Randomly sample 500 images out of photo streams
- Original title vs. predicted title by blog posts

Quantitative results

<table>
<thead>
<tr>
<th></th>
<th># Votes</th>
<th># Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original title</td>
<td>5 4 3 2 1 0</td>
<td>156 92 94 60 55 43</td>
</tr>
</tbody>
</table>

64.2% (1605/2500)
Examples of Automatic Image Titling

**Success**
- after exiting original submarine voyage attraction opened
- 2010111 disneyland family vacation 0510 simply fountains blasting water high
- first photoshop hdr attempt newly refurbished big thunder mountain
- california adventure radiator springs racers
- disneyland feb 2011 (120) paradise garden grille opens
- IMG_0160 sailing ship columbia

**Near-miss**
- bessie at big thunder ranch service animals
- nightmare dia de los muertos haunted mansion holiday
- celebrate! a street party bippidi boppidi boutique
Results of Interpolation btw Blog Pictures

Goal: Given a pair of blog images, predict a likely sequence btw them using photo streams

• No groundtruth! Perform user studies via Amazon Mechanical Turk

Quantitative results

<table>
<thead>
<tr>
<th>Algorithm #1 vs. (KNN+KM)</th>
<th>Algorithm #2 vs. (DTW+KM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.9% (929/1500)</td>
<td>66.5% (997/1500)</td>
</tr>
</tbody>
</table>
Examples of Interpolation btw Blog Pictures

Summarization results according to iterations
Outline

• Problem Statement
• Approach
• Experiments
• Conclusion
Conclusion

Jointly summarize online sets of **photo streams** and **blog posts**

- The characteristics of two media are **complementary**
  - **Photo streams**: Dense visual samples on the events
  - **Blog posts**: transferring semantic knowledge

Automatic image titling  
Image localization  
Interpolation btw blog pictures

Structured ranking SVM framework for the alignment and summarization

- Large-scale dataset of 10K blogs (120K associated images) and 6K photo streams (540K images)
Thank you!