Streaming News Image Summarization

Hao Li, Shangfu Peng, Hanan Samet

{haoli, shangfu, hjs}@cs.umd.edu
NewsStand: A News Aggregation System

Indexed 10K+ RSS news feeds, extracted ~50K+ news articles/day

Samet et al, “Reading news with maps by exploiting spatial synonyms”. CACM, 2014
http://newsstand.umiacs.umd.edu/
PhotoStand: A News Image Aggregation System

A picture is worth a thousand words

Samet et al, “PhotoStand: A Map Query Interface for a Database of News Photos”. VLDB, 2013
http://newsstand.umiacs.umd.edu/photostand/
Temporal-based News Browsing
Topic-based News Summarization
Needs for Automatic News Image Summarization

• The “Boston Bombing” news topic has 835 new articles and 1540 images during April 15 - May 12, 2013

• News photos from different medias can be highly similar.
Problem Formulation

**Input:** Given a news photo stream \( \mathcal{I}_e = \{I_1, \ldots, I_N\} \) on topic \( e \), each image is associated with an article \( a \), a timestamp \( t \) and a caption \( d \)

**Output:** a subset of images \( V \) to summarize the topic \( e \) in a time interval \( T \)

**Goals**

- The selected images should be *important* and *representative*
- The visual timeline should maximize
  - Information Value
  - Coverage
  - Diversity
Redundancy in News Photos

- News photos taken by different media for the same event
  - Yahoo!
  - NBC News
  - Business Insider

- The same photo is cropped or modified by different media
  - CNN
  - The Guardian
  - BBC
How to measure the importance of a news photo?

**Observations**: the greater the importance of the news photo, the higher the frequency and diversity of use.

- The importance can be evaluated by the number of its near-duplicates.
Pipeline of News Image Summarization

News topic clustering → Photo stream → Global feature clustering → Local feature clustering → Representative image selection → Timeline generation

\[ I_c \rightarrow I_1 \rightarrow I_2 \rightarrow \cdots \rightarrow I_N \]
Overview

- Image in News
- News Image Summarization Pipeline
  - Near-Duplicate Image Clustering
  - Representative Image Selection
  - Timeline Generation
- Evaluation
Near-Duplicate Image Clustering

• Global feature
  • Hierarchical Color Histogram
  • **Pro**: efficient to compute, store and compare (512-bytes)
  • **Con**: not robust to significant cropping and geometric distortions.

• Local feature
  • SIFT + RANSAC verification
  • **Pro**: more robust to various transformations
  • **Con**: more time consuming for matching.

encode and compare global and local features of the images, preserving spatial information

Samet et al, “PhotoStand: A Map Query Interface for a Database of News Photos” . VLDB, 2013
Streaming Two-Stage Image Clustering

Algorithm 1: Online near-duplicate image clustering

Input: Image stream \( \{I_1, \ldots, I_N\} \) to be processed
Output: Clusters \( C \)

while \( i \leq N \) do
  \( c_g \leftarrow \text{match}\_\text{global}\_\text{feature}(C, f_g(I_i)) \)
  if \( c_g \neq \emptyset \) then
    \( C[c_g] \leftarrow C[c_g] \cup \{I_i\} \)
  else
    \( c_l \leftarrow \text{match}\_\text{local}\_\text{feature}(C, f_l(I_i)) \)
    if \( c_l \neq \emptyset \) then
      \( C[c_l] \leftarrow C[c_l] \cup \{I_i\} \)
    else
      \( C \leftarrow C \cup \{I_i\} \)

strict threshold for global feature matching

local feature matching merges near-duplicate images

The number of images in cluster \( C \) is used as the importance score \( s \) for images in that cluster.

The criteria for SIFT matching:
1) at least 15 local descriptors are matched
2) the percentage of matched descriptors compared to the average number of descriptors of two images is greater than 5%.
Overview

• Image in News

• News Image Summarization
  • Duplicate Detection

• Representative Image Selection
  • Timeline Generation

• Evaluation
VisualRank for Image Search Re-ranking

Identifying the “authority” images on a visual-similarity graph

search result  visual hyperlink  similarity graph
Representative Image Selection

\[ p(I_u, I_v) = \frac{m(I_u, I_v)}{n(I_u)} \]

\[ PR(I_v) = \frac{\alpha}{|c|} + (1 - \alpha) \sum_{I_u \in c} PR(I_u) \frac{p(I_u, I_v)}{\sum_{I_x \in c} p(I_u, I_x)} \]
Overview

• Image in News

• News Image Summarization
  • Duplicate Detection
  • Representative Image Selection
  • **Timeline Generation**

• Evaluation
Timeline Generation

**Basic approach**

- select the top $K$ images for each day while the importance score $s_i$ of each image is above a threshold $\epsilon$.

$$\max \sum_{i \in V_t} s_i \quad \text{st} \quad |V_t| \leq K \quad \text{and} \quad \forall s_i > \epsilon$$

**Cons**

- fixed $K$ may fail to display important images.
- thresholding the important score may result in too many (few) images.

---

Xu et al. A Cross-media Evolutionary Timeline Generation Framework Based on Iterative Recommendation. ICMR 2013
Yan et al. Visualizing Timelines: Evolutionary Summarization via Iterative Reinforcement between Text and Image Streams. CIKM 2012
Dynamic Timeline

- We treat this task as a window query.
- Users can zoom in or zoom out on the timeline to see the detailed or the general summarization result.

Visual Timeline with Constraints

- **Constraint 1**: total $M$ images are visible in the window
  
- **Constraint 2**: screen width $W$
  
<table>
<thead>
<tr>
<th>Constraint</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$s_A &gt; s_B &gt; s_C$</td>
</tr>
<tr>
<td></td>
<td>$w_A + w_B &gt; W$</td>
</tr>
<tr>
<td></td>
<td>$w_A + w_C &gt; W$</td>
</tr>
<tr>
<td>2</td>
<td>$s_B + s_C &gt; s_A$</td>
</tr>
<tr>
<td></td>
<td>$w_B + w_C &lt; W$</td>
</tr>
</tbody>
</table>

sometimes two narrow images are better than one wide image
0-1 Knapsack Problem

**Objective:** select at most images $M$ to maximize the accumulated scores in a query window (time period $T = [t_s, t_e]$ and width $W$).

\[
\hat{V} = \arg \max_V \sum_{i \in V} s_i
\]

subject to:
- $|V| \leq M$ (visible image constraint)
- $\sum_{i \in V_k} w_i \leq W$ (screen width constraint)
- $t_s \leq t_i \leq t_e$ (time window constraint)

We solve this 0 – 1 knapsack problem by dynamic programming:

- $f(i, n, m, w) + s_i \rightarrow \begin{cases} f(i + 1, n, m + 1, w + w_i) \\ f(i + 1, n + 1, m + 1, 0) \end{cases}$
- $f(i, n, m, w) \rightarrow \begin{cases} f(i + 1, n, m, w) \\ f(i + 1, n + 1, m, 0) \end{cases}$
“Prince William and Kate seen after the birth of their first child”
More Results

Timeline summarization with $M = 30$, $W = 600$ and $|T| = 8$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Disabled Cruise</td>
<td>Disabled Cruise</td>
<td>Disabled Cruise</td>
<td>Disabled Cruise</td>
<td>Disabled Cruise</td>
<td>Disabled Cruise</td>
<td>Disabled Cruise</td>
<td>Disabled Cruise</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Bombing</td>
<td>Boston Bombing</td>
<td>Boston Bombing</td>
<td>Boston Bombing</td>
<td>Boston Bombing</td>
<td>Boston Bombing</td>
<td>Boston Bombing</td>
<td>Boston Bombing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SFO Crash</td>
<td>SFO Crash</td>
<td>SFO Crash</td>
<td>SFO Crash</td>
<td>SFO Crash</td>
<td>SFO Crash</td>
<td>SFO Crash</td>
<td>SFO Crash</td>
</tr>
</tbody>
</table>
More Results

Timeline summarization of “Boston Bombing” with $M = 30$, $W = 600$ and $|T| = 8$
Summary

• Utilize the crowd-wisdom to determine the importance of news images.

• Dynamic timeline display as a window query and solve it as a 0-1 knapsack problem.
Thank you!
Performance

| Event                  | #imgs | $|C|_{GT}$ | Dup%  | $\max(|c|)$ | $|C|$ | False Match | Miss Match |
|------------------------|-------|---------|-------|-------------|------|-------------|------------|
| Disabled Cruise        | 325   | 175     | 46.2  | 44          | 219  | 160         | 170        |
| Boston Bomning         | 1540  | 713     | 53.7  | 65          | 963  | 675         | 698        |
| George Zimmerman      | 1374  | 706     | 48.6  | 27          | 954  | 608         | 695        |
| SFO Plane crash        | 788   | 305     | 61.3  | 25          | 427  | 266         | 301        |
| William and Kate       | 1273  | 766     | 39.8  | 58          | 913  | 726         | 736        |
| California Kidnapping  | 396   | 97      | 75.5  | 128         | 176  | 91          | 97         |

<table>
<thead>
<tr>
<th>Color</th>
<th>SIFT</th>
<th>Hybrid</th>
<th>Color</th>
<th>SIFT</th>
<th>Hybrid</th>
<th>Color</th>
<th>SIFT</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>645</td>
<td>100</td>
<td>170</td>
<td>48</td>
<td>160</td>
<td>170</td>
<td>33.98</td>
<td>11.81</td>
<td>0.040</td>
</tr>
<tr>
<td>121.99</td>
<td>848.55</td>
<td>58.61</td>
<td>432.21</td>
<td>7.00</td>
<td>0.167</td>
<td>105.89</td>
<td>7.69</td>
<td>0.106</td>
</tr>
<tr>
<td>91.04</td>
<td>432.21</td>
<td>7.00</td>
<td>1149.02</td>
<td>11.97</td>
<td>0.139</td>
<td>1313.00</td>
<td>1.03</td>
<td>3.47</td>
</tr>
<tr>
<td>4.74</td>
<td>1149.02</td>
<td>11.97</td>
<td>11.47</td>
<td>20.57</td>
<td>0.108</td>
<td>76.36</td>
<td>0.19</td>
<td>0.22</td>
</tr>
</tbody>
</table>

TABLE II
Performance of Near-Duplicate Clustering

<table>
<thead>
<tr>
<th>Feature Extraction</th>
<th>Duplicate Clustering</th>
<th>Image Selection</th>
<th>Total</th>
<th>Average</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>SIFT</td>
<td>Stage 1</td>
<td>Stage 2</td>
<td>Graph Building</td>
<td>PageRank</td>
</tr>
<tr>
<td>1.75</td>
<td>22.58</td>
<td>0.15</td>
<td>33.98</td>
<td>11.81</td>
<td>0.040</td>
</tr>
<tr>
<td>12.44</td>
<td>121.99</td>
<td>8.33</td>
<td>848.55</td>
<td>58.61</td>
<td>0.210</td>
</tr>
<tr>
<td>10.59</td>
<td>91.04</td>
<td>4.74</td>
<td>432.21</td>
<td>7.00</td>
<td>0.167</td>
</tr>
<tr>
<td>7.62</td>
<td>64.57</td>
<td>0.90</td>
<td>105.89</td>
<td>7.69</td>
<td>0.106</td>
</tr>
<tr>
<td>9.03</td>
<td>137.74</td>
<td>5.09</td>
<td>1149.02</td>
<td>11.97</td>
<td>0.139</td>
</tr>
<tr>
<td>3.81</td>
<td>40.27</td>
<td>0.13</td>
<td>11.47</td>
<td>20.57</td>
<td>0.108</td>
</tr>
</tbody>
</table>
Failure Case

2013-08-24 16:39:47
President Barack Obama

2013-08-28 03:45:25
White House Press Secretary Jay Carney speaks about Syria during a press briefing at the White House

2012-12-27 15:46:10
Syrian President Bashar al-Assad meets with peace envoy Lakhdar Brahimi in the capital Damascus

2013-02-03 09:06:32
A picture released by the official Syrian Arab News Agency shows Syrian president Bashar al-Assad talking with Iran's Saeed Jalili.