Tiling Slideshow

Jun-Cheng Chen, Wei-Ta Chu, Jin-Hau Kuo, Chung-Yi Weng, and Ja-Ling Wu

wtchu@cmlab.csie.ntu.edu.tw

Communications & Multimedia Laboratory
National Taiwan University
Motivation

☐ Large amounts of **consumer photos** derive the following problems:

- Filtering or correcting are annoying.
- Browsing photos takes much time.
- Sequential presentation makes users boring.

![blurred photo](image1)

![orientation correction](image2)
Goal

- Generate a kind of new media that provides user elaborate photo browsing experience.
  - Photo filtering & organization
  - Vivid audiovisual presentation
  - Value-added results
Photographic Story

- Paragraph: describe by text
  - Contains a topic sentence and several supportive sentences.

- Photographic paragraph: describe by photos
  - Contains a topic photo and several supportive photos
The Proposed Slideshow

Music Beats

Time

1 2 3 4 5 6 7 8

Music Waveform
Outline

- System Overview
- Visual Processing
- Music Analysis
- Tiling Slideshow Composition
- Evaluation
- Conclusion and Future Work
System Overview

Photos

- Orientation correction
- Blur detection
- Overexposure/underexposure detection

Preprocessing

Clustering

- Time-based & content-based clustering
- Temporal & spatial composition
- Temporal selection
- ROI detection

Composition

Music

Beat detection

Tiling slideshow
Photo Processing

- Orientation correction
  - EXIF (Exchangeable Image File Format) metadata

- Photo Filtering
  - Blur detection
    - Check edge information in different resolutions
  - Overexposure/Underexposure detection
    - Check intensity information of each photo

Blurred photo

Underexposure photo
Photo Clustering

- Displaying photos that are in the same scenic spot or the same event would strengthen audiovisual perception.

- Clustering
  - Time characteristics – event
  - Content characteristics – visually homogenous

(O) (X)
Time-based Clustering

- Check the time gap between adjacent photos

- 15 sec
- 47 sec
- 30 sec
- 7 hr
Given a time-based photo cluster, finer clustering is performed based on content-based features. (dominant color and color layout)

Within-cluster distance:

\[ S_w = \max_{g=1,\ldots,m} \frac{1}{n_g(n_g-1)} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} d(P_i, P_j) \]

Between-cluster distance:

\[ S_b = \min_{g \neq h} \frac{1}{n_g n_h} \sum_{i=1}^{n_g} \sum_{j=1}^{n_h} d(P_i, P_j) \]

Goodness of a clustering case:

\[ R_i = \frac{S_b}{S_w} \]

Prefer that photos in the same cluster are similar, and photos in different clusters are distinct as much as possible.
Content-based Clustering (2/3)

Clustering case 1

\[ R_1 = \frac{S_b}{S_w^2} \]
\[ (S_w^2 > S_w^1) \]

Clustering case 2

\[ R_2 = \frac{S_b^3}{S_w^1} \]
\[ (S_w^1 > S_w^2 > S_w^3) \]
\[ (S_b^1 > S_b^2 > S_b^3) \]
Content-based Clustering (3/3)

\[ R = \frac{S_h}{S_w} \]

Clustering Results

Clustering Results
Music Analysis

- Beat detection
- Music segmentation

For frame switching and photo displaying

For frame switching and photo displaying

Search range for frame switching

Sound Energy Difference

Music Beats

Frame 1

1

2

3

Frame 2

4

5

Frame 1 starts

t_1

t_2

t_3

Frame 2 starts

t_4

t_5

r_1 (4 seconds)

r_2 (6 seconds)
Short Summary

- **Photo**
  - Filter out defective photos
  - Organize photos in terms of time and content characteristics

- **Music**
  - Segment into smaller pieces

Diagram:
- Photos: Preprocessing \rightarrow Clustering \rightarrow Composition \rightarrow Tiling slideshow
- Music: Beat detection \rightarrow Composition
Tiling Slideshow Composition

- Challenge 1
  - Given a time-limited music clip, only a subset of photo clusters can be displayed.

- Challenge 2
  - For a cluster of photos to be displayed, more important photos should occupy larger space.

- Challenge 3
  - Photos should be smartly manipulated to fit in with the limited displaying space.
Cluster Selection (for Challenge 1)

- Cluster-based importance
  - Defined based on “photo per minute (PPM)” and “photo conformance (PC)”
  - Higher shooting frequency (PPM), more important
  - Larger conformance (PC), more important

For each content-based cluster $C_g$ in a time-based cluster

\[
PPM(C_g) = \frac{N(\psi)}{\text{Time}_\text{Duration}(\psi)} \quad \text{— Shooting frequency}
\]

\[
PC(C_g) = 1 - \frac{1}{n_g(n_g - 1)} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} d(P_i, P_j) \quad \text{— Opposite to within-cluster distance}
\]
Templates (for Challenge 2)

- Design various templates that contain a topic cell and several supportive cells – to form photographic paragraphs.

A cluster with 4 photos

4-cell Templates

Topic cell

Supportive cells
## Template Determination (for Challenge 2)

### Templates importance

$$Ic_i = \frac{\text{Area}(Tc_i)}{\text{Area}(T)} \quad (Ic_1 \geq Ic_2 \geq ... \geq Ic_k)$$

$$TV = (Ic_1, Ic_2, ..., Ic_k) \quad \text{— Template importance vector}$$

### Photo-based importance

- Defined based on “face region (FR)” and “attention value (AV)”

$$PI_i = W_{\text{face}} \times FR(P_i) + W_{\text{attention}} \times AV(P_i)$$

$$PV = (PI_1, PI_2, ..., PI_k) \quad \text{— Photo importance vector} \quad (PI_1 \geq PI_2 \geq ... \geq PI_k)$$
Template Determination (for Challenge 2)

- Find the most matching between template importance and photo importance
- Find the minimum included angle between them

\[
T_{h,i} = \arg \min_{i=1,2,...,s} \cos \left( \frac{PV \cdot TV_{h,i}}{\| PV \| \| TV_{h,i} \|} \right)
\]
Composition (for Challenge 3)

- Find the region that conveys most “content value” and conforms to the aspect ratio of the targeted cell – constrained optimization problem.

Top-down case: (photo with face)

Bottom-up case: (photo without face)
Composition (for Challenge 3)

1. Find ROI
2. Extend
3. Crop
4. Resize

480 pixels

720 pixels
Demo
## Evaluation

- Photos taken by amateurs in three different trips.

<table>
<thead>
<tr>
<th>Data Set 1:</th>
<th>Data Set 2:</th>
<th>Data Set 3:</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Images" /></td>
<td><img src="image2.png" alt="Images" /></td>
<td><img src="image3.png" alt="Images" /></td>
</tr>
<tr>
<td>780 photos</td>
<td>522 photos</td>
<td>1257 photos</td>
</tr>
<tr>
<td>Music: 3m31s</td>
<td>Music: 4m38s</td>
<td>Music: 4m06s</td>
</tr>
<tr>
<td>Osaka, Kyoto, Kobe, Nagoya, Tokyo (Japan)</td>
<td>Melbourne, Brisbane (Australia) Amsterdam (Netherlands)</td>
<td>Osaka, Kyoto, Kobe (Japan)</td>
</tr>
</tbody>
</table>
User Study

- Compare the satisfaction of ACDSee, PhotoStory, and Tiling slideshow

- Questionnaire (to 27 evaluators)
  - Q1: How do you feel the photo variety in a time unit?
  - Q2: Do you think it's a funny presentation?
  - Q3: Do you think the sequence helps you experience this trip?
  - Q4: Are you willing to use it to generate your own slideshow?
  - Q5: How do you feel the audiovisual effects of this slideshow?
## Objective Tests (1/2)

### Clustering performance evaluation

<table>
<thead>
<tr>
<th>Slideshow</th>
<th>#frames</th>
<th># photos</th>
<th># frame with clustering error</th>
<th>Avg. number of photos in a frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37</td>
<td>127</td>
<td>1</td>
<td>3.43</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>172</td>
<td>1</td>
<td>3.58</td>
</tr>
<tr>
<td>3</td>
<td>43</td>
<td>184</td>
<td>2</td>
<td>4.28</td>
</tr>
</tbody>
</table>

[Image of ill-clustered photo]

[Image of ill-clustered photo]
Objective Tests (2/2)

- Cropping performance evaluation

<table>
<thead>
<tr>
<th>Slideshow</th>
<th># photos</th>
<th># ill-cropped photos</th>
<th># ill-cropped photos in topic cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slideshow 1</td>
<td>127</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Slideshow 2</td>
<td>172</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Slideshow 3</td>
<td>184</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

ill-cropped photo

ill-cropped photos
Summary

- We propose a new type of audiovisual presentation for **consumer photos**.
- Perform both visual and music analysis for **organized presentation**.
- We deal with issues on content selection and smart manipulation to display qualified content in limited time and limited space.
Future Work

- Semantic features or user intervention can be added to facilitate more elaborate filtering, clustering, cropping, and interactive browsing.

- Possible applications
  - Apply to different types of photos, e.g. wedding and party.
  - Include videos in cells.
  - Extend it to be an on-line version and provide an on-line tour.
  - Extract keyframes from videos. Slideshow of keyframes could be a kind of video summary.