Home Photo Content Modeling for Personalized Event-Based Retrieval

Joo-Hwee Lim and Qi Tian
Institute for Infocomm Research, Singapore
Philippe Mulhem
Multimedia, IEEE
Digital Object Identifier 10.1109/MMUL.2003.1237548

指導教授：林信志博士
報告者：林宸宇
報告日期：96.2.9

Computer Graphics & Interactive Techniques Lab.
Outline

- Abstract
- Related Works
- Home photo event taxonomy
- Modeling visual events
- Experimental evaluation
- Future work
Abstract

• Digital camera – more digital photos
• Need effective and efficient tools
• User
  – Event
  – People
  – Time
  – Place
Abstract

- Semantic gap
- Home photo event taxonomy
Related Works

• 人工
  – Inconsistent
  – Erroneous
  – Time

• An efficient annotation system
  – Limit the possible keyword choice
Related Works

- **MiAlbum system**
  - Query catch Keyword find picture
  - None a complete taxonomy for home photos

- **Unlabeled photos**
  - Event-based retrieval
Home photo event taxonomy

• Home photos
  – Family and friends (900 万)
  – Others (500 万)
    • Scenery and nature
    • Travel

• Typical event taxonomy for home photos
Home photo event taxonomy

- The notion of event usually encompasses four aspects:
  - Who
    - takes part in the event
  - What
    - occasion or activity is involved
  - Where
    - the event takes place
  - When
    - the event takes place
Modeling visual events

• R (Mi, x)
  – Photo x and Mi (event model)
  – 0~1

• Minimize manual annotation
• Personalization
Modeling visual events
Modeling visual events

- Visual keyword
  - Automatically extract relevant semantic token
- Mi (event model)
  - Mvi
  - Mgi
Modeling visual events

- Visual keywords indexing
  - Visual keywords
    - Visual prototypes extracted
    - Learned from a visual content domain
    - Relevant semantics label
  - Visual vocabulary
    - 8 classes
    - Each class into two to five
    - Total 26 labels
Modeling visual events

• Labels
  – using
    • A three-layer feed-forward neural network
    • Dynamic node creation capabilities
  – Learning
    • Color feature
    • Texture feature
  – 375 labels images
Modeling visual events

- People: Face, figure, crowd, skin
- Sky: Clear, cloudy, blue
- Ground: Floor, sand, grass
- Water: Pool, pond, river
- Foliage: Green, floral, branch
- Mountain: Far, rocky
- Building: Old, city, far
- Interior: Wall, wooden, China, fabric, light

Computer Graphics & Interactive Techniques Lab.
Modeling visual events

Figure 4. Swimming pool image (a) to be indexed and (b) its index represented as a tessellation of visual keyword histograms.
Modeling visual events

\[ S_r(V, x_{vk}) = \max_j(S(v_j, x_{vk})) \quad (1) \]

• Compare a single image histogram Xvk and a group of images histogram Vj, find the similarity

• Mohan
  – Can recognize some combination objects
  – But not use here, not have a well-defined part structure
Modeling visual events

- Visual event graph

![Diagram]

*Figure 5. Excerpt of a concept type hierarchy.*
Modeling visual events
Modeling visual events

[Diagram showing relationships and probabilities between visual events such as "On top 0.5", "Water pool 1.0", "Touches 1.0", "Object 1.0", "Nature 0.5", "Man made 0.5", "Foliage 0.5", "Building 0.5".]
Modeling visual events

- Equations related to graph matching.

\[ S_g (Mg_i, x_{cg}) = \max_{g_p \in \pi(g_i, x_{cg})} \text{Match concepts}(g_p, x_{cg}) + \text{Match arches}(g_p, x_{cg}) \]
Modeling visual events

- Visual event learning
- Learning of a model $M_i$
  - local visual keyword histograms
  - graph-based abstraction
- Compare $M_{vi}$ and $M_{gi}$

$$R(M_i, x) = \lambda \cdot S_v(M_{vi}, x_{vi}) + (1 - \lambda) \cdot S_g(M_{gi}, x_{gi})$$
Experimental evaluation

- 2,400 heterogeneous home photos collected over a five-year period with both indoor and outdoor settings
- Smallest 256×384, noisy marginal pixels, the images become 240×360
Experimental evaluation

- Normal

- Noisy
Experimental evaluation
Experimental evaluation

• Event-based learning and query
• Four event
  – Park 306, swimming pools 52, waterside 114, wedding 241
  – Choose each 10 photos for learning
Experimental evaluation

• Comparison and analysis

• Three method
  – Hue Saturation Value (HSV) – global
  – Hue Saturation Value (HSV) – grid
  – Visual Event Retrieval (VER)
    • VER- $\lambda = 0.5$
Experimental evaluation

Table 1. Average precision ratios for event-based retrieval.

<table>
<thead>
<tr>
<th>Event</th>
<th>HSV-global</th>
<th>HSV-grid</th>
<th>VER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parks</td>
<td>0.30</td>
<td>0.42</td>
<td>0.56</td>
</tr>
<tr>
<td>Swimming pools</td>
<td>0.12</td>
<td>0.22</td>
<td>0.29</td>
</tr>
<tr>
<td>Waterside</td>
<td>0.10</td>
<td>0.13</td>
<td>0.20</td>
</tr>
<tr>
<td>Wedding</td>
<td>0.38</td>
<td>0.32</td>
<td>0.41</td>
</tr>
<tr>
<td>Overall</td>
<td>0.23</td>
<td>0.27</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Experimental evaluation

Table 2. Average precision ratios at top 20 and 30 documents.

<table>
<thead>
<tr>
<th>Event</th>
<th>HSV-global</th>
<th>HSV-grid</th>
<th>VER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(20/30 docs)</td>
<td>(20/30 docs)</td>
<td>(20/30 docs)</td>
</tr>
<tr>
<td>Parks</td>
<td>0.62/0.59</td>
<td>0.78/0.79</td>
<td>0.94/0.92</td>
</tr>
<tr>
<td>Swimming pools</td>
<td>0.26/0.21</td>
<td>0.32/0.27</td>
<td>0.52/0.43</td>
</tr>
<tr>
<td>Waterside</td>
<td>0.24/0.22</td>
<td>0.24/0.23</td>
<td>0.31/0.29</td>
</tr>
<tr>
<td>Wedding</td>
<td>0.62/0.65</td>
<td>0.60/0.53</td>
<td>0.90/0.85</td>
</tr>
<tr>
<td>Overall</td>
<td>0.44/0.42</td>
<td>0.49/0.46</td>
<td>0.67/0.62</td>
</tr>
</tbody>
</table>
Experimental evaluation

![Graph showing precision vs recall for different methods]

- HSV-global
- HSV-grid
- Visual event retrieval

Computer Graphics & Interactive Techniques Lab.
Experimental evaluation

![Graph showing precision and recall for different categories: Park, Pool, Waterside, and Wedding.](image-url)
Future work

- Event model
- More model
Thanks!

報告結束～