An MPEG-7 Based Content-aware Album System for Consumer Photographs

Chen-Hsiu Huang, Chih-Hao Shen, Chun-Hsiang Huang and Ja-Ling Wu
Communication and Multimedia Laboratory,
National Taiwan University,
E-mail: {chenhsiu,shen,bh,wjl}@cmlab.csie.ntu.edu.tw
Introduction

- Devices used to create or digitize photographs, such as digital cameras, have become increasingly affordable.
- It’s ease for consumers to shoot pictures but not trivial when it comes to deal with many of them.
  - Contents that we can not handle or manage are of no values.
- Many album system are designed to solve this by using EXIF information or textual metadata, but we think that’s not quite straight forward.
- In our opinion, an ideal album system should be able to identify the difference between photographs and realize some semantic information about the content; that is, it should be a content-aware album system.
Core Functionalities

- **Locating:** Query images by face
  - Face detection & recognition
- **Adaptation:** Smart Thumbnail
  - Photo Focus identification
- **Browsing:** Photo Similarity
  - Find relevant photos with similarity calculation
Query Images by Face

Steps for querying photos by face:

- Face Detection
- Training
- Face database
- Add photos to album
- Recognize faces in photographs
- Query photos by faces

PS: We use Intel OpenCV Library as face detection & recognition module
Before thumbnailing for image adaptation, we should first identify what’s the focus in photos.

For photos with people, human faces are surely our focus when viewing.

The user attention model has applied to find some saliency points:

- **Red**: Intensity based
- **Green**: Color based
- **Blue**: Skin color based
For each face & saliency points, a weighting function was applied to calculate its importance. User can select the cropping ratio, the cropping region is adaptive decided according to the weighting value.

Cropping the focus region first, then scaling. Better then direct scaling, but not so good.
Adaptive Selection

- For all the visual objects (faces, saliency points), calculate its importance by:
  $$W_i = \left( FR_w \times FR_h \right)^2 / d_c$$

- When adaptive selection, sort those visual objects by importance, dropping the least import object to achieve the goal cropping ratio.
Photo Similarity

- Two descriptors borrowed from MPEG-7 standard:
  - **Color Layout Descriptor**
    - It is designed to efficiently represent spatial distribution of colors
  - **Dominant Color Descriptor**
    - The representative colors in an image or image region

- Face descriptors used in our system
  - **Face Number Descriptor**
    - The number of faces detected in image

- By using the faces information and MPEG-7 descriptors, we can calculate the similarities between images.
Similarity Modeling

- Distance of face number descriptor between photos is defined as:
  \[ \text{dist}_{FND} = \frac{|FN_i - FN_j|}{\max(FN_i, FN_j)} \]

- Similarity modeling with descriptor distance combination
  \[ \text{Sim}_{ij} = (\text{dist}_{CLD} + \text{dist}_{DCD} + \text{dist}_{FND}) / 3 \]

Images of photos:
- IB02-155.jpg
- IB02-160.jpg
- IB02-023.jpg
- IB02-162.jpg
- IB02-163.jpg
- IB02-001.jpg
- IB02-159.jpg
- IB02-003.jpg
- IB02-144.jpg
- IB02-031.jpg
We can get more semantic meanings from low level features by combining those kernel modules.
In the Future

- The album system can be improved both systematically and component-wise:

  - **System aspect:**
    - The album syntax should be fully conform to the MPEG-7 standard.
    - The album should be able to process other media type such as audio and video.

  - **Component aspect:**
    - More low level features or descriptors in MPEG-7 standard will be used and combined for further semantic meaning extraction.
    - The face detection & recognition library could be fine tuned to meet the needs of album system.
Discussion

- Any comments are welcomed.
- Thank you.