Tiling Slideshow

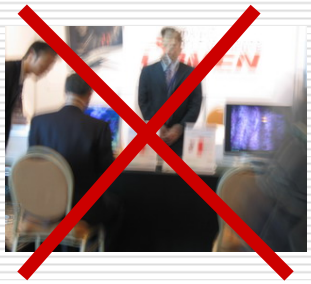
Jun-Cheng Chen, Wei-Ta Chu, Jin-Hau Kuo, Chung-Yi Weng, and Ja-Ling Wu

wtchu@cmlab.csie.ntu.edu.tw

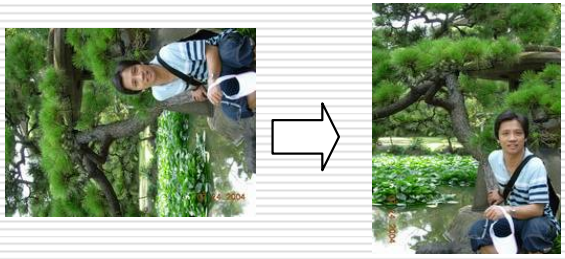
Communications & Multimedia Laboratory
National Taiwan University

Motivation

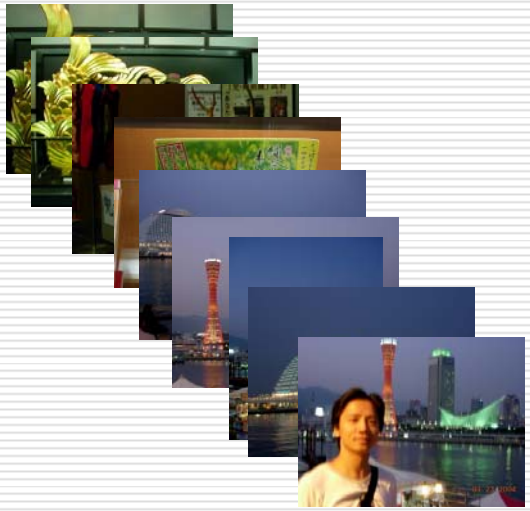
- Large amounts of **consumer photos** derive the following problems:
  - Filtering or correcting are annoying.
  - Browsing photos takes much time.
  - Sequential presentation makes users boring.



blurred photo

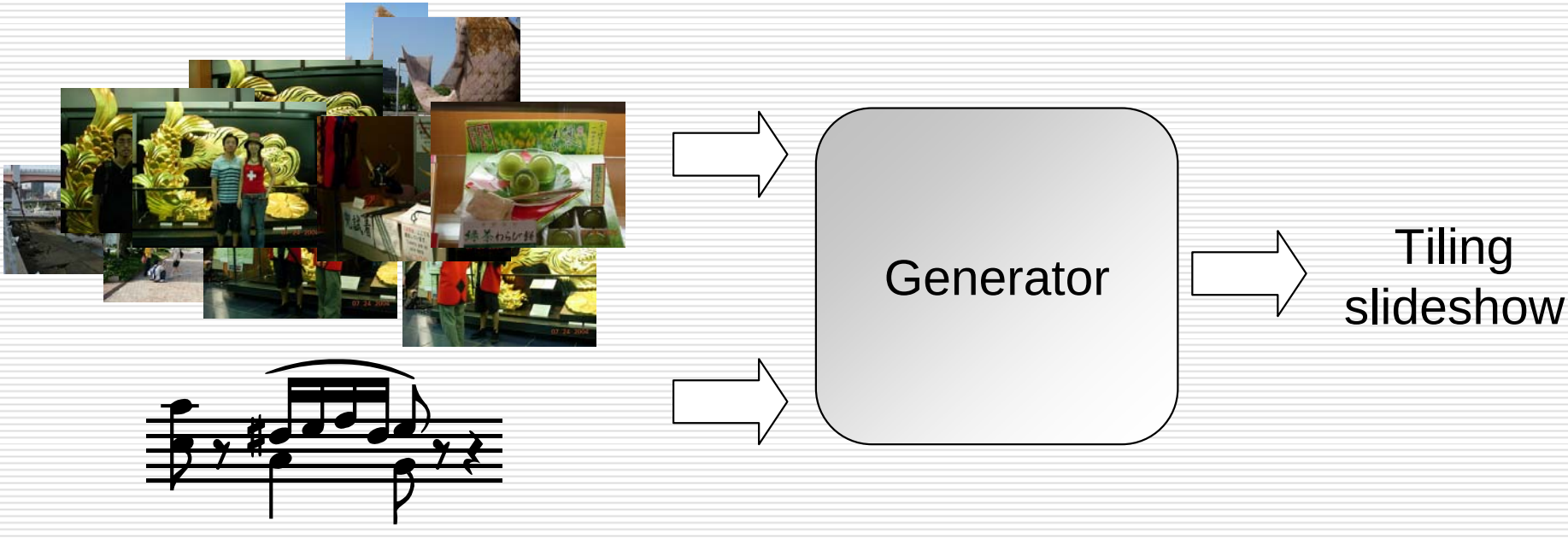


orientation correction



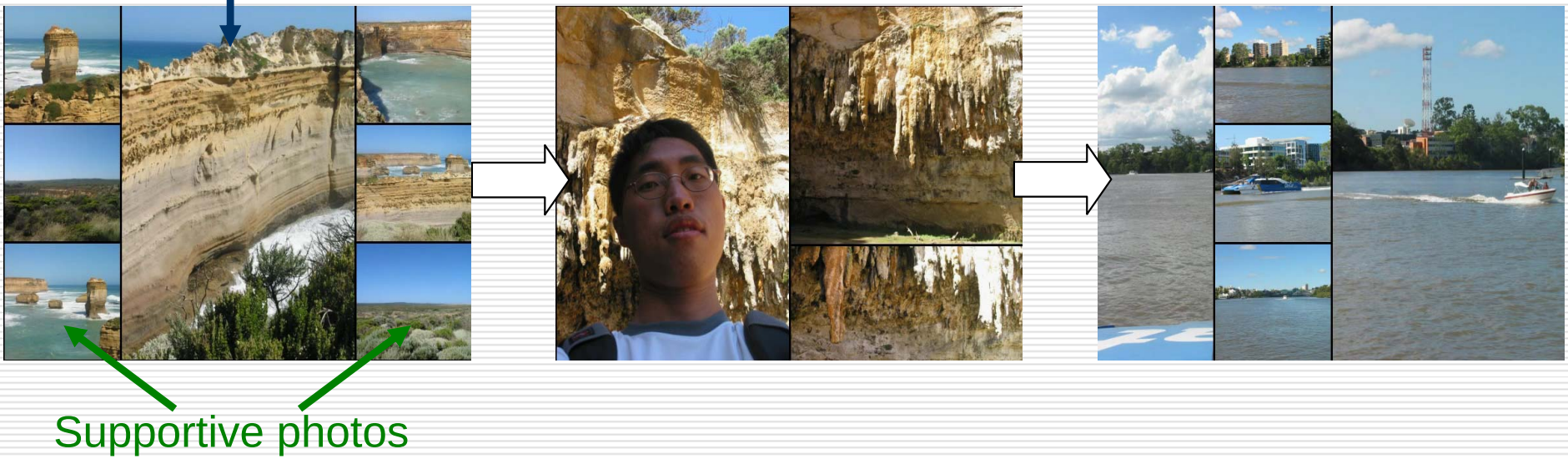
Goal

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

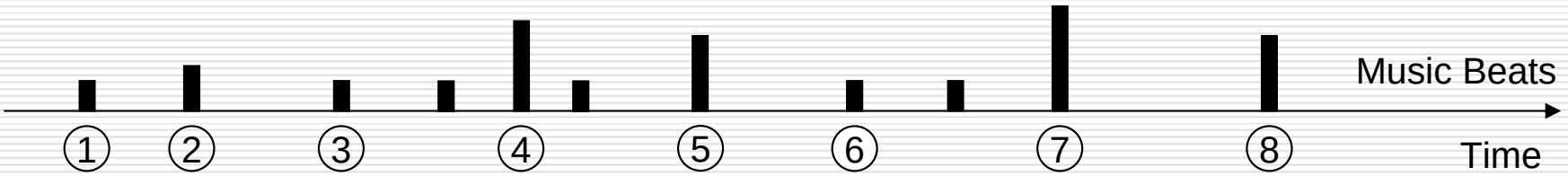


Photographic Story

- 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



Outline

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

System Overview

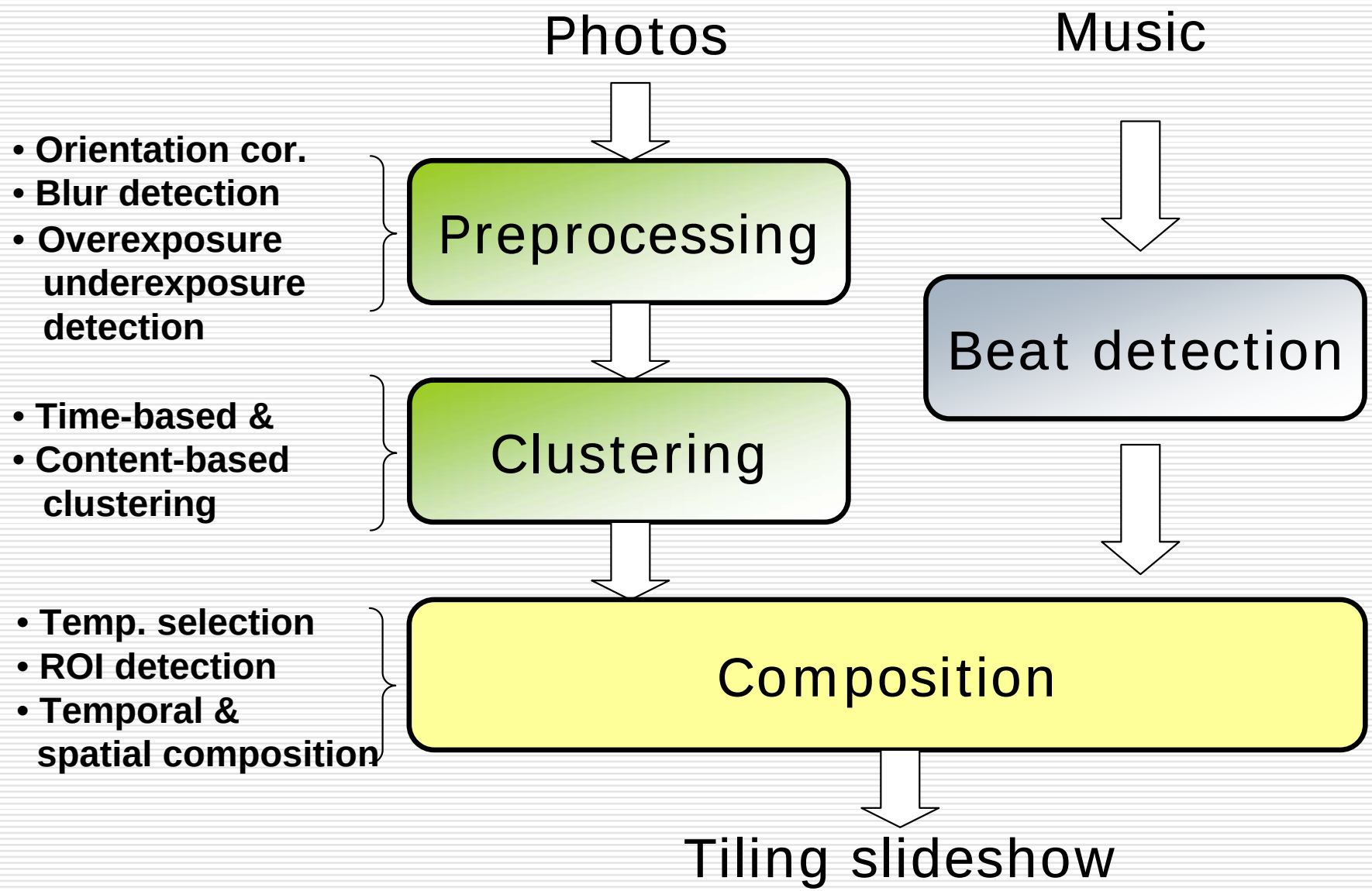
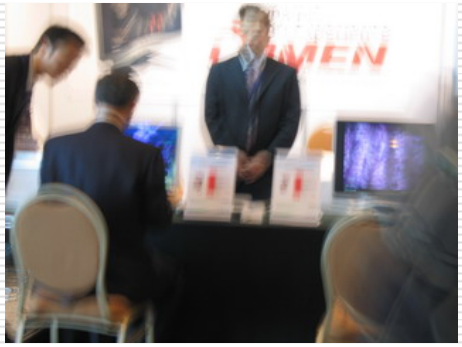


Photo Processing

- Orientation correction
  - EXIF (Exchangeable Image File Format) metadata
- Photo Filtering
  - Blur detection
    - Check edge information in diff. resolutions
  - Overexposure/Underexposure detection
    - Check intensity information of each photo

Blurred photo



Underexposure photo

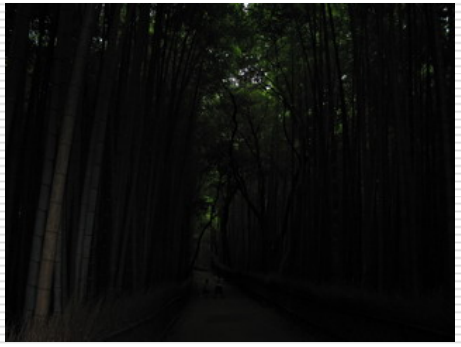


Photo Clustering

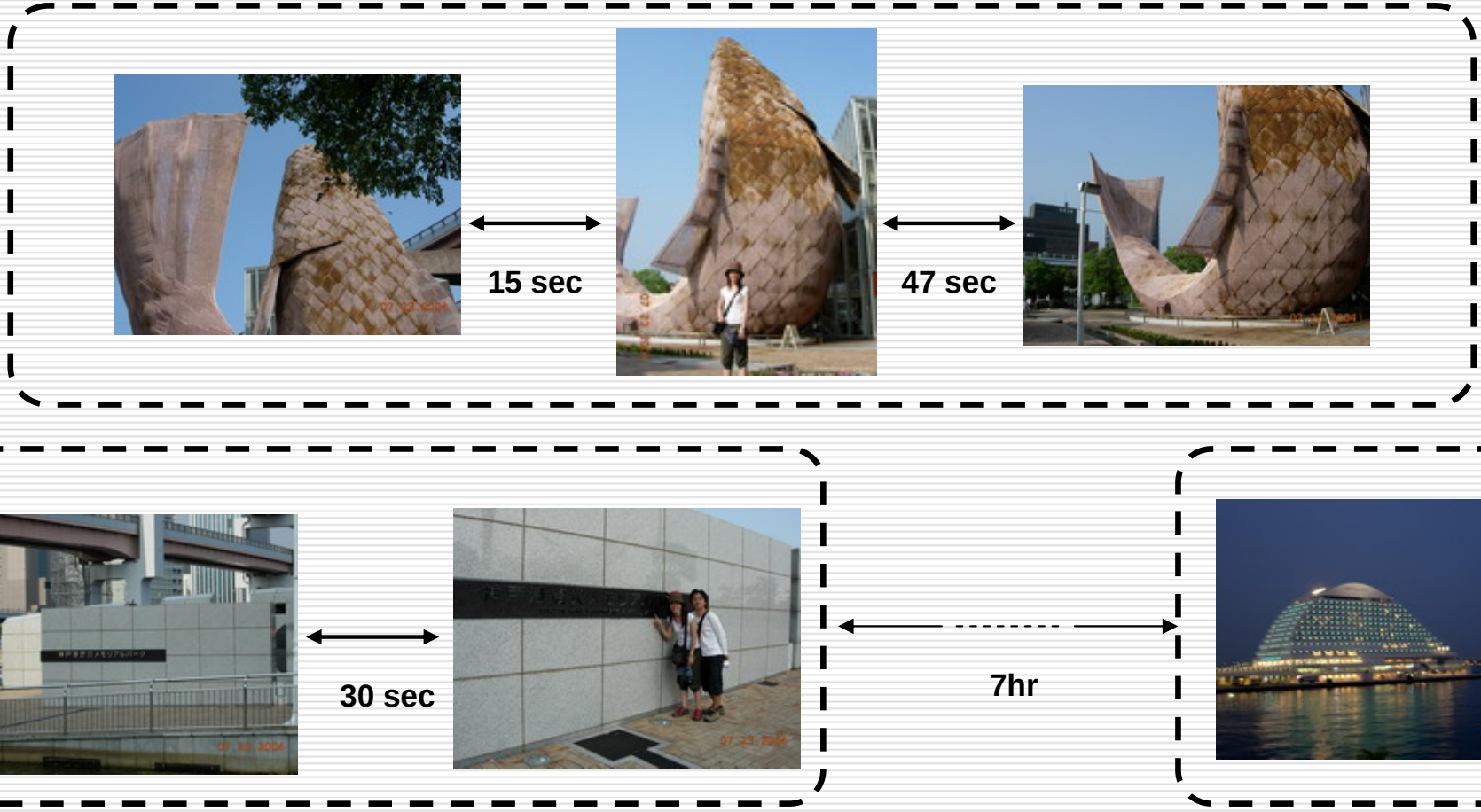
- Displaying photos that are in the same scenic spot or the same event would strengthen audiovisual perception.
- Clustering
  - Time characteristics – event
  - Content characteristics – visually homogenous

(O)

(X)

Time-based Clustering

- Check the time gap between adjacent photos



Content-based Clustering (1/3)

- Given a time-based photo cluster, finer clustering is performed based on content-based features. (dominant color and color layout)

Within-cluster distance:

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

$d(.)$ is the average of normalized dominant color and color layout distances.

Between-cluster distance:

$$S_b = \min_{\substack{g \neq h \\ g=1,\ldots,m \\ h=1,\ldots,m}} \frac{1}{n_g n_h} \sum_{i=1}^{n_g} \sum_{j=1}^{n_h} d(P_i, P_j)$$

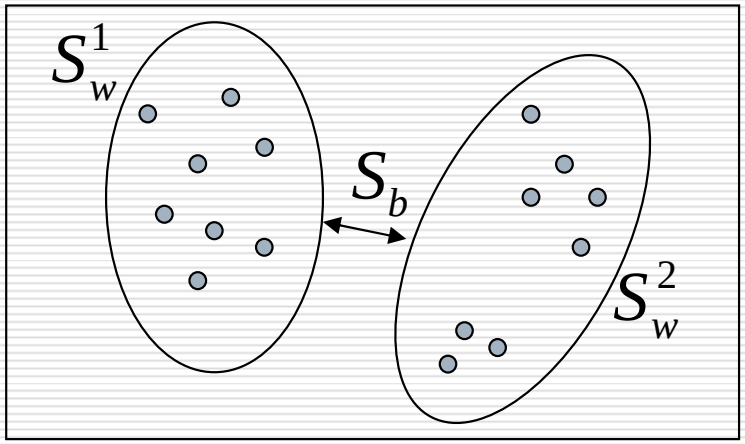
Goodness of a clustering case:

$$R_i = \frac{S_b}{S_w}$$

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)

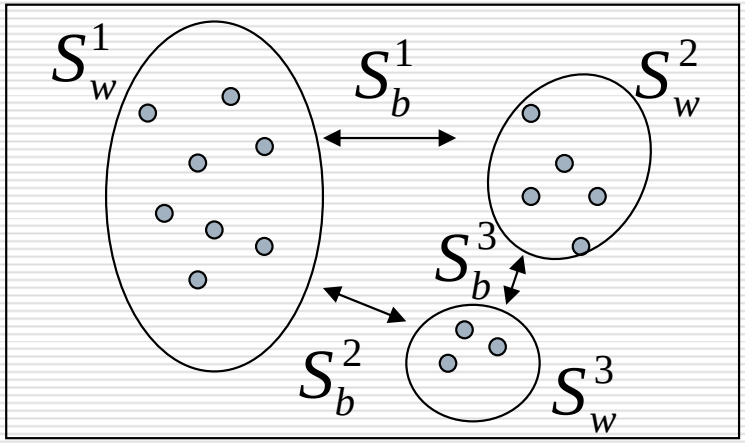
Clustering case 1



$$R_1 = \frac{S_b}{S_w^2}$$

$$\left(S_w^2 > S_w^1\right)$$

Clustering case 2



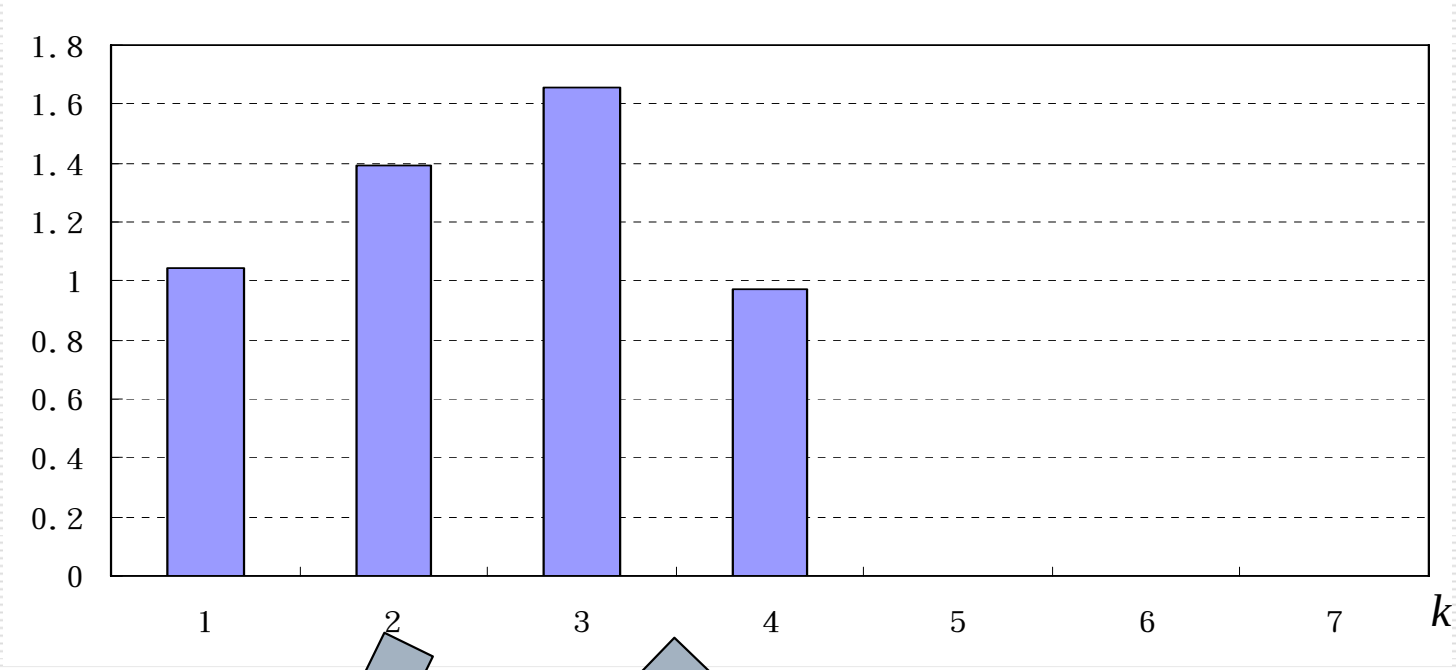
$$R_2 = \frac{S_b^3}{S_w^1}$$

$$\left(S_w^1 > S_w^2 > S_w^3\right)$$

$$\left(S_b^1 > S_b^2 > S_b^3\right)$$

Content-based Clustering (3/3)

$$R = S_b / S_w$$



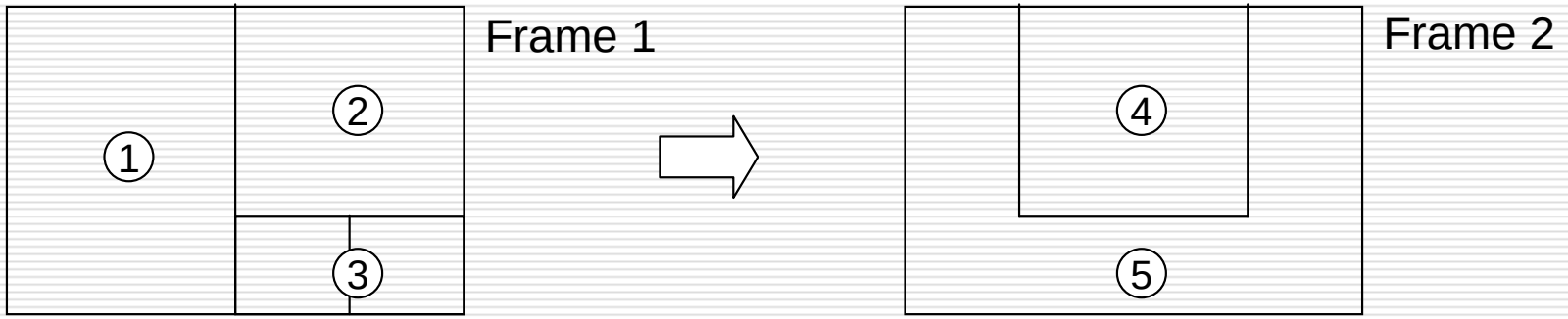
Clustering Results

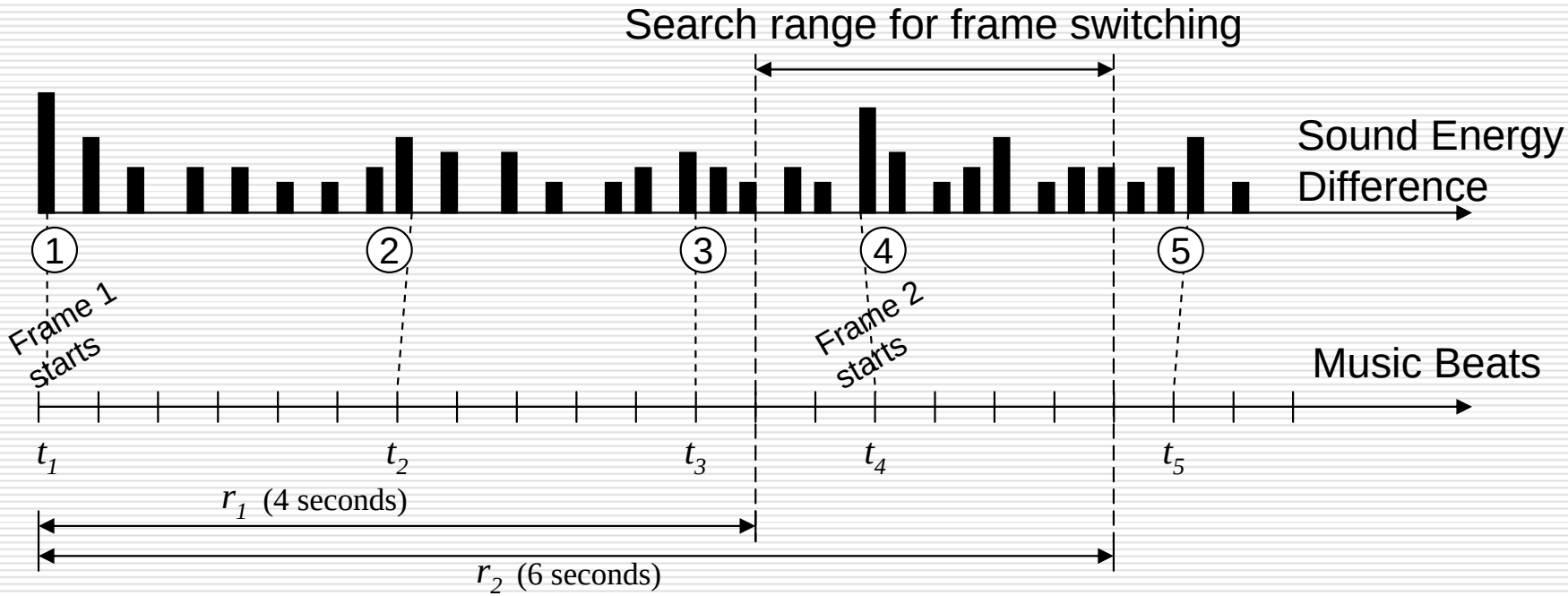
Clustering Results

Music Analysis

- Beat detection
- Music segmentation

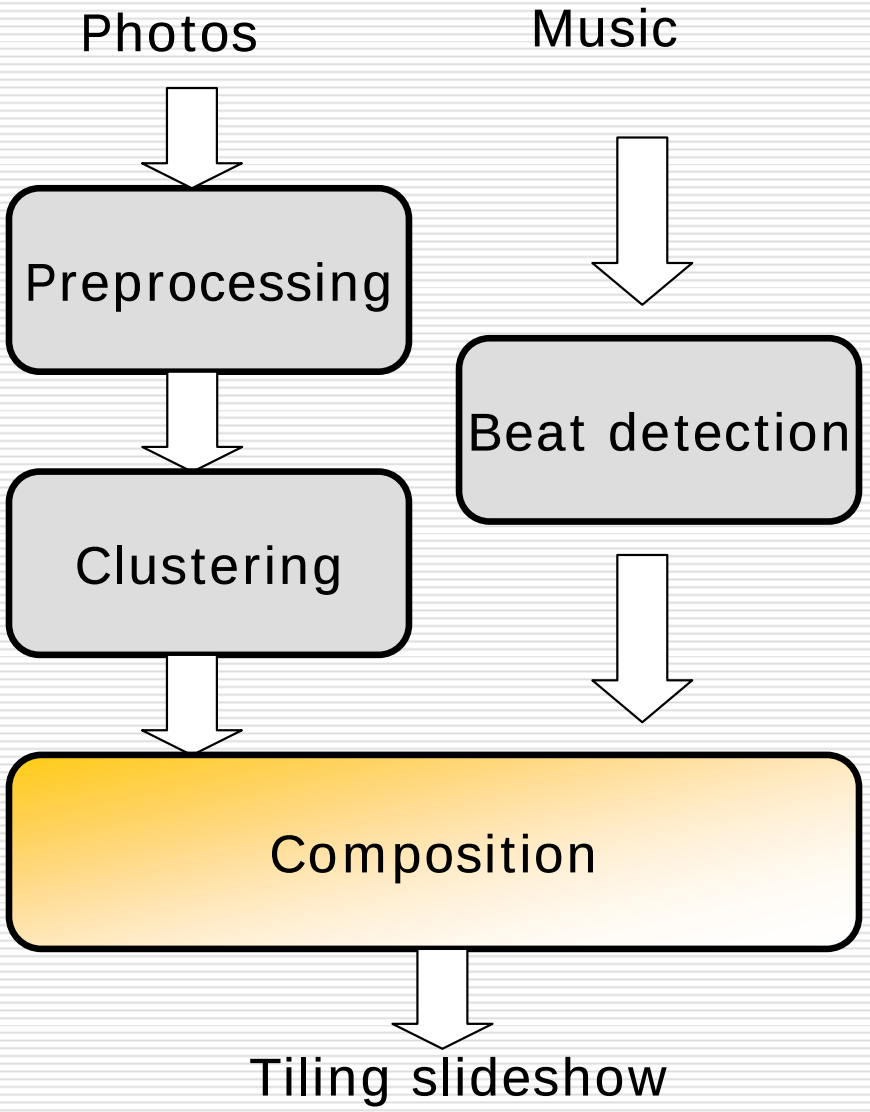
} For frame switching and photo displaying





Short Summary

- Photo
  - Filter out defective photos
  - Organize photos in terms of time and content characteristics
- Music
  - Segment into smaller pieces



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

$$PPM(C_g) = N(\Psi)/Time\_Duration(\Psi)$$ — Shooting frequency

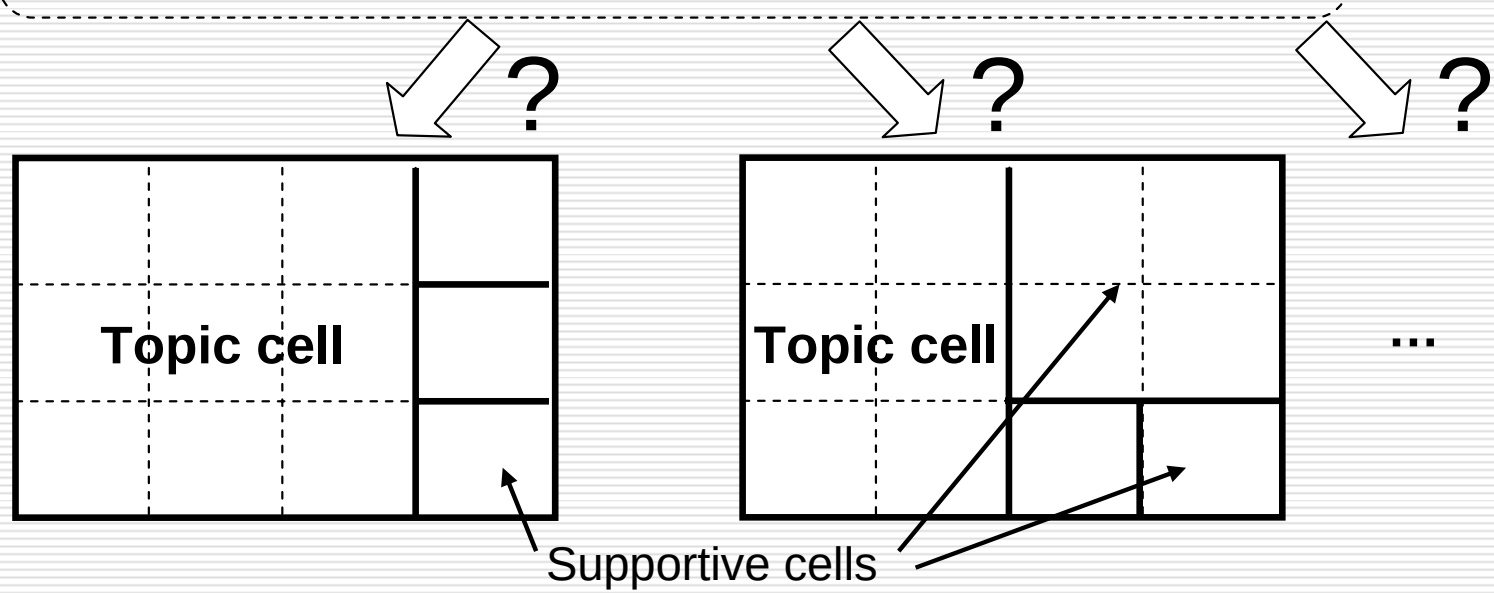
$$PC(C_g) = 1 - \frac{1}{n_g(n_g - 1)} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} d(P_i, P_j)$$ — Opposite to within-cluster distance

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



Template Determination (for Challenge 2)

- Templates importance

$$Ic_i = Area(Tc_i) / Area(T) \qquad \left(Ic_1 \geq Ic_2 \geq \ldots \geq Ic_k\right)$$

$TV = \left(Ic_1, Ic_2, \ldots, Ic_k\right)$ — Template importance vector

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

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

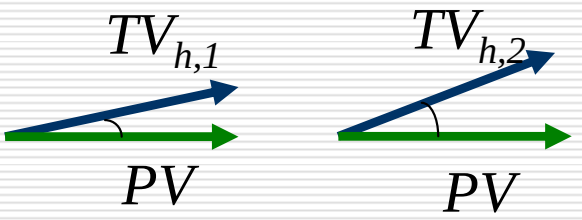
$PV = \left(PI_1, PI_2, \ldots, PI_k\right)$ — Photo importance vector

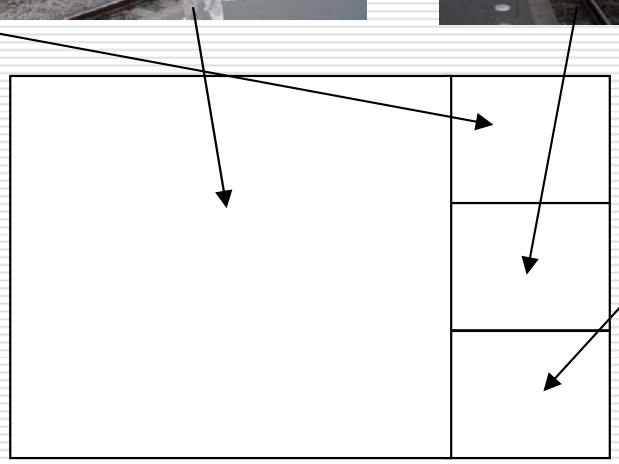
$\left(PI_1 \geq PI_2 \geq \ldots \geq PI_k\right)$

Template Determination (for Challenge 2)

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

$$T_{h,i} = \arg \min_{i=1,2,\ldots,s} \operatorname{acos}\left(\frac{PV \cdot TV_{h,i}}{\| PV \| \| TV_{h,i} \|} \right)$$

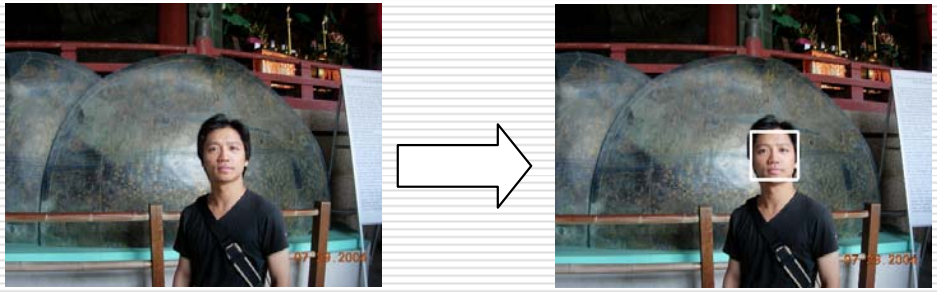




Composition (for Challenge 3)

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

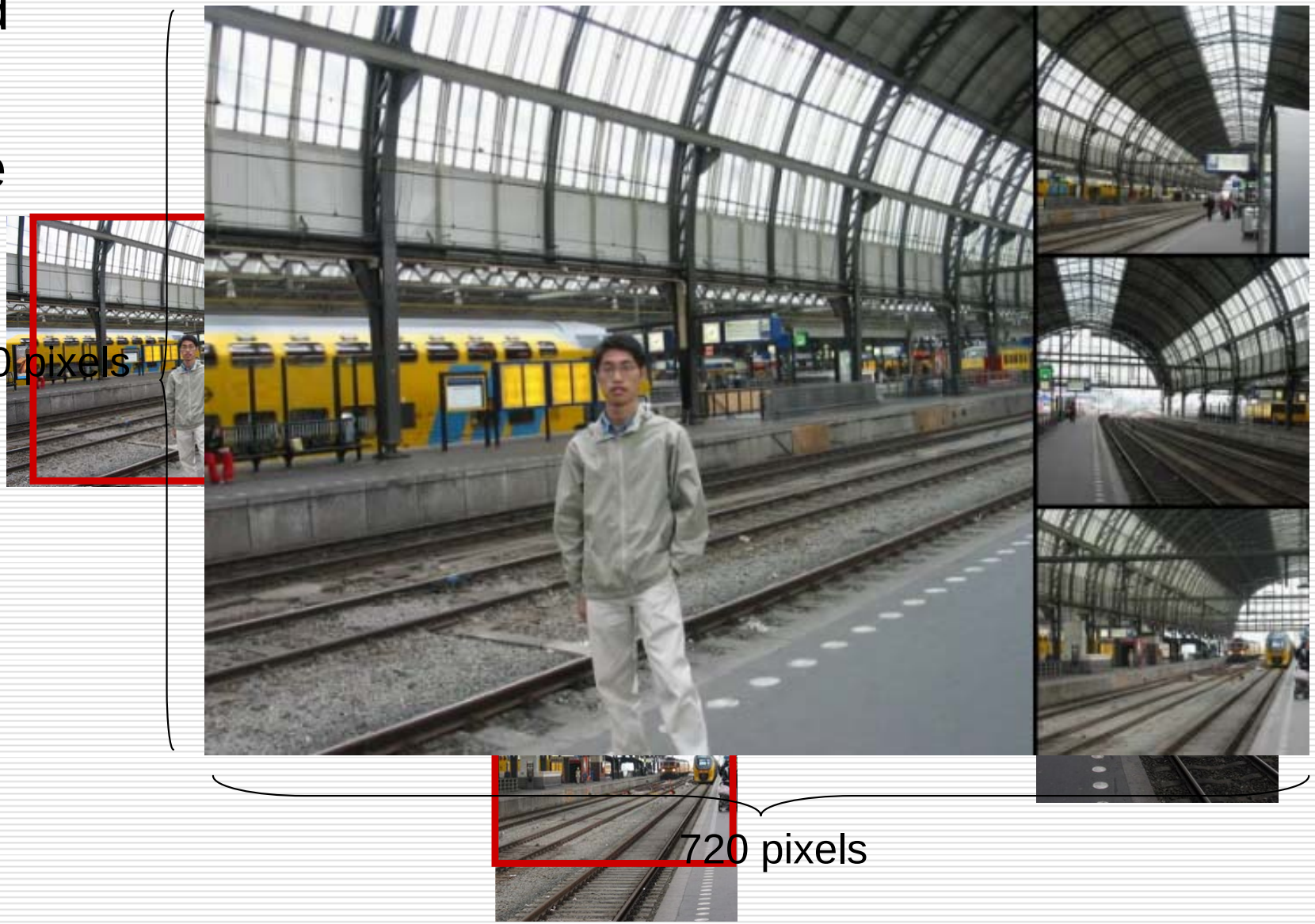
Top-down case:
(photo with face)



Bottom-up case:
(photo without face)

Composition (for Challenge 3)

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



Demo

Evaluation

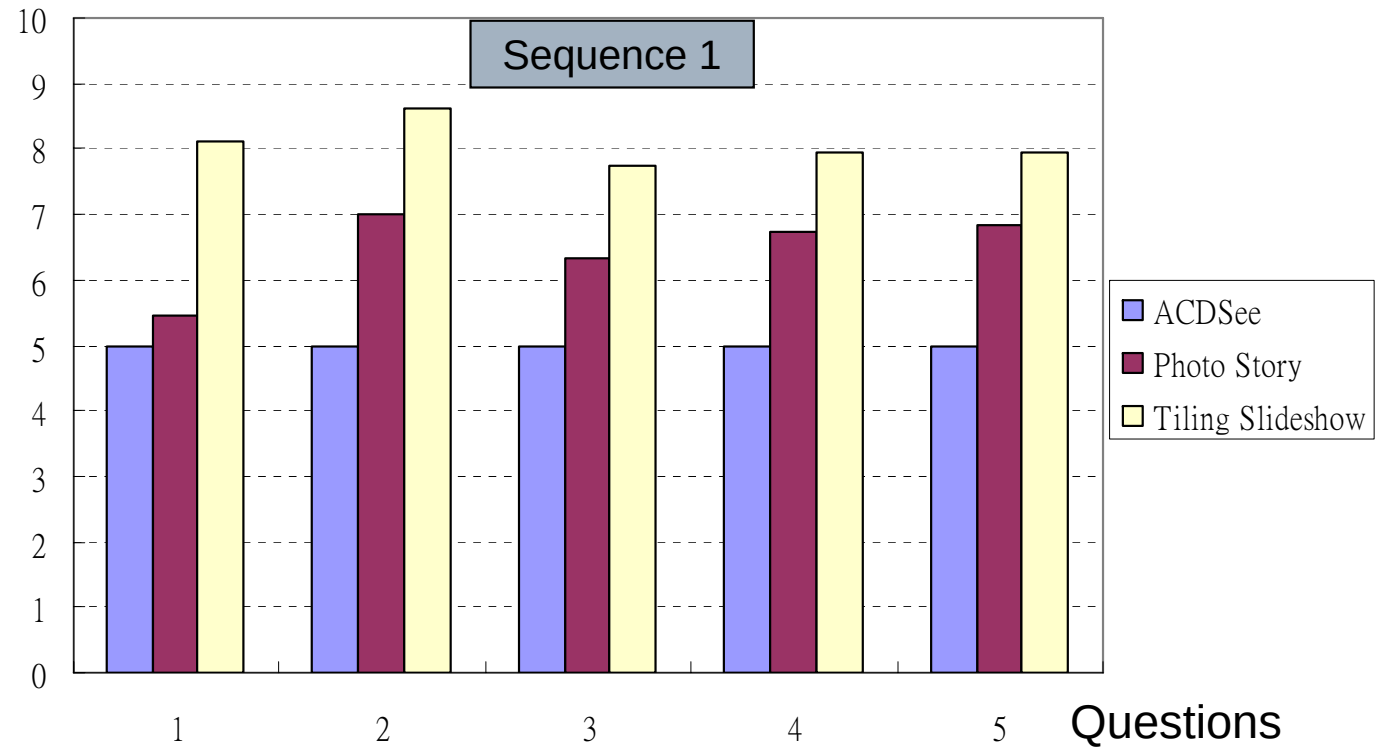
- Photos taken by amateurs in three different trips.

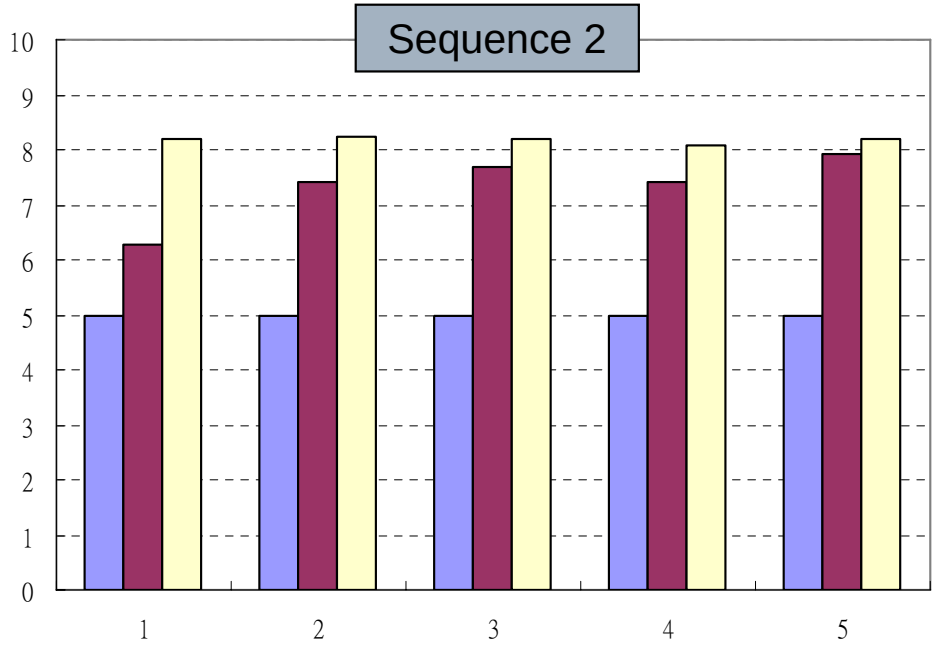
Data Set 1:	Data Set 2:	Data Set 3:
		
780 photos Music: 3m31s	522 photos Music: 4m38s	1257 photos Music: 4m06s
Osaka, Kyoto, Kobe, Nagoya, Tokyo (Japan)	Melbourne, Brisbane (Australia) Amsterdam (Netherlands)	Osaka, Kyoto, Kobe (Japan)

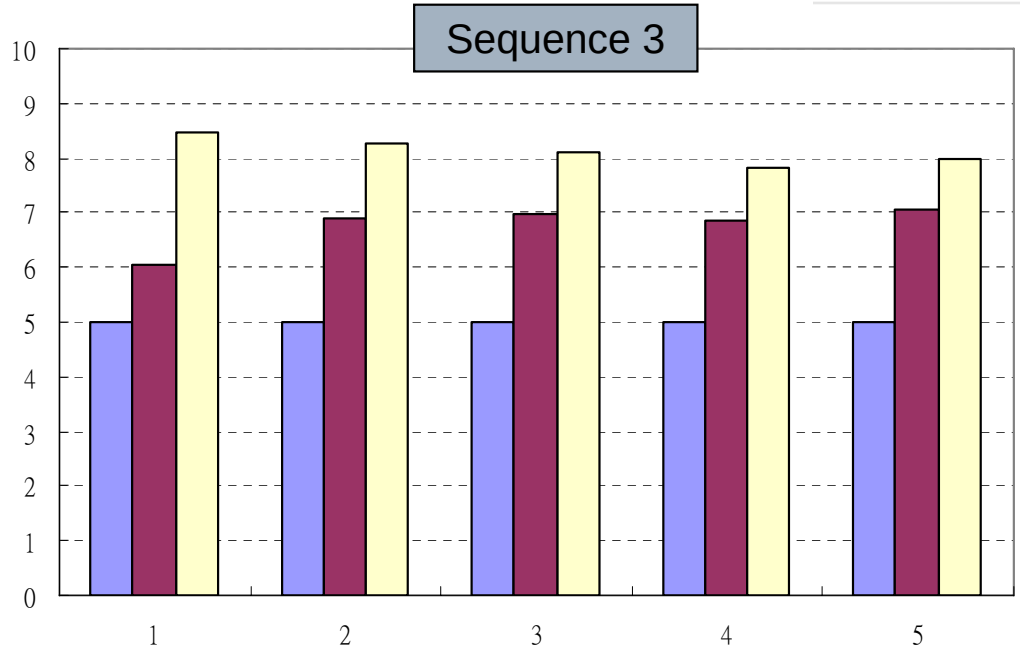
User Study

- 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?

Subjective Scores





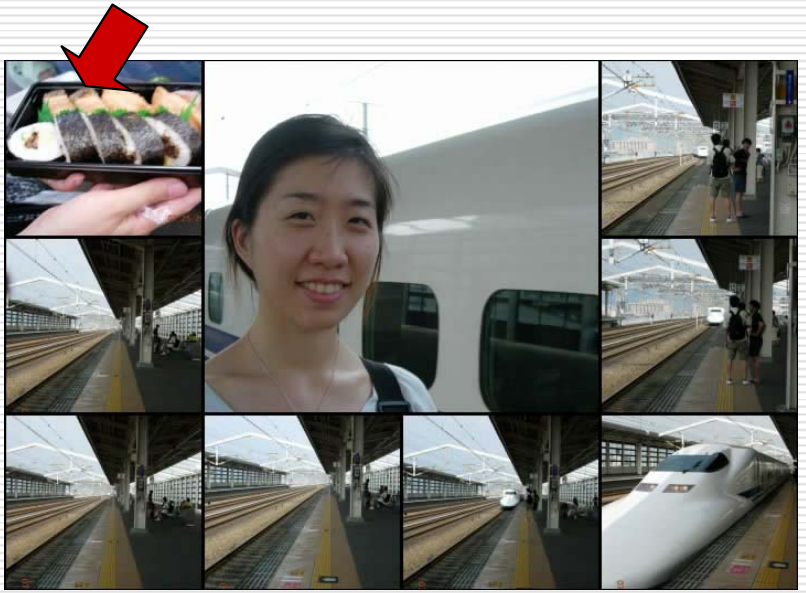


Objective Tests (1/2)

- Clustering performance evaluation

	#frames	# photos	# frame with clustering error	Avg. number of photos in a frame
Slideshow 1	37	127	1	3.43
Slideshow 2	48	172	1	3.58
Slideshow 3	43	184	2	4.28

Ill-clustered photo



Ill-clustered photo

Objective Tests (2/2)

- Cropping performance evaluation

	# photos	# ill-cropped photos	# ill-cropped photos in topic cell
Slideshow 1	127	5	1
Slideshow 2	172	5	0
Slideshow 3	184	6	3

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.