Flickr and YouTube Dataset

Objective

- Jointly summarize large sets of online Flickr images and YouTube videos
- The two media types help each other since their properties are complementary

Problem Statement

- (Input) A set of photo streams $P = \{P^1, \ldots, P^m\}$ and videos $V = \{V^1, \ldots, V^n\}$
- (Output1) Image summary as Storyline graph $G = (O, E)$
- (Output2) Keyframe-based video summary

Video Summarization

A simple diversity ranking algorithm + Similarity votes from a collection of images can create semantically-meaningful video summary

- Step1: For each video $V^m$, find the K-nearest photo streams $N(V^m)$
- Step2: Build a similarity graph between video frames and images
- Step3: Solve the following optimization problem of diversity ranking [Kim et al. ICCV 2011]

$$\max_{x \in X} \sum_{i \in X} u(x) s.t. \sum_{i \in X} G(x, \pi(x)) \text{ for } d_i = \sum_{x \in X} G(x, \pi(x))$$

$$u(x) = 0, \text{ for } x \in S^p \subset V^m, |S^p| \leq k$$

Submodular optimization

A greedy solution achieves a constant factor approximation

$$N(V^m)$$

$\pi(x)$

Each image casts $m$ similarity votes

Image Summarization as Storyline Graphs

- A storyline graph $G = (O, E)$
- Vertex set $O$: the set of image clusters
- Edge set $E$: popular transitions recurring in photo streams
- Edges should be Sparse and Time-varying

Evaluation of Image Summarization via AMT

- Quantitative results
- Obtain groundtruths for video summary via Amazon Mechanical Turk (AMT)
- Compare the similarity between GTS and summary be algorithms [2]

Evaluation of Video Summarization via AMT

- Main difficulties of evaluation
  - (i) No groundtruth (ii) Too large graphs from too many images
  - Pairwise preference tests via AMT
  - A crowd of human subjects evaluate a basic unit (i.e. important edges of storyline)
  - An example of application – Predicting likely images between a pair of images

Flickr and YouTube Dataset

- 20 outdoor recreational classes
- # images/photo streams (2,769,504, 35,545) • # videos (15,912)

Advertise for Workshop at Storytelling with Images and Videos (VisStory 2014)

More information at http://vision-storytelling.org/

References
