Scene Summarization for Online Photo Collections
Ian Simon, Noah Snavely, Steve Seitz
Scene Summarization for Online Photo Collections
Ian Simon, Noah Snavely, Steve Seitz

- Can you convey the “essence” of Rome in a few images?
We found **278,698 photos matching rome and italy.**
50 images of Rome
20,000 images of Rome
300,000 images of Rome
What about tags?

- User-submitted text labels.
Summarizing Rome

18 largest scenes, computed from 20,000 images on Flickr
Canonical views

from Shimon Edelman

Criteria [Palmer, Rosch, Chase 81]

1. Given a set of photos, which view do you like the best?
2. When taking a photo yourself, which viewpoint do you choose?
3. From which viewpoint is the object easiest to recognize?
4. When imagining the object in your head, which view do you see
Segmenting image collections

- Photos provide a likelihood distribution over viewpoints
- Find small set of views that best describe this distribution
2000 Pantheon images from Flickr
Segmenting image *collections*

- Represent a view $V$ as normalized vector of features

- Similarity defined as $V_i \cdot V_j$

Detect features using SIFT [Lowe, IJCV 2004]
(a) Canonical views selected by the spherical k-means algorithm with $k = 6$.

(b) The output of our greedy k-means canonical views algorithm with $\alpha = 8$.

(c) The output of our greedy k-means algorithm with $\alpha = 5.75$ and orthogonality weight $\beta = 100$.

(d) All six photos from the Wikipedia [2] entry for the Pantheon.
Discovering scenes
Tags Revisited

• Now we can choose better tags.
Tags Revisited

- Now we can choose better tags.

\[
\text{score(cluster, tag)} = \sum_{\text{users}} P(\text{cluster} | \text{tag, user}) P(\text{user})
\]
Vatican Summary (+tags)
Rome Summary (+tags)
Scene Summarization demos

- Rome summary
- 3D browser