“Clustering by Composition”: Unsupervised Discovery of Image Categories

Alon Faktor & Michal Irani
The Weizmann Institute of Science
**Goal:** Separate into 2 clusters (Yoga & Ballet)

**Previous Work:**
Unsupervised Category Discovery

**Simple pairwise affinities:**
- Grauman & Darrell 2006 (Pyramid match kernel)

**Discover common “cluster model”:**
- Sivic et al 2005 (PLSA)
- Russell et al 2006 (segments)
- Kim et al 2008 (link analysis)
- Lee & Grauman 2009 (foreground)
- Payet & Todorovic 2010 (shape)
- ...
**Goal:** Separate into 2 clusters (Yoga & Ballet)

**Overview of Our Approach:**
- “Affinity by composition”

“Good region” ≠ “Good segment”
1) “Rare” (low chance to occur at random).
2) Shared by 2 images.
Overview of Our Approach:

- “Affinity by composition”
- Random Search process + Image collaboration → Linear complexity

Goal: Separate into 2 clusters (Yoga & Ballet)

“Good region” ≠ “Good segment”
1) “Rare” (low chance to occur at random).
2) Shared by 2 images.
Goal: Separate into 2 clusters (Yoga & Ballet)

Cluster#1  (mostly Ballet images)  Cluster#2  (mostly Yoga images)

Purity = 18/19 = 95%  Purity = 19/21 = 90%
“Affinity by Composition”

extending [Boiman & Irani, NIPS’06]

“Good region” $R$:

1) Shared by 2 images.

2) “Rare” (low chance to occur at random).
“Affinity by Composition”

extending [Boiman & Irani, NIPS’06]

"Good region" $R$:

1) Shared by 2 images.
2) "Rare" (low chance to occur at random).

$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)} \xrightarrow{\text{contribution of } R \text{ to affinity}(I_1,I_2)}$$
“Affinity by Composition”

extending [Boiman & Irani, NIPS’06]

\[
\log \frac{p(R | I_1, I_2)}{p(R | H_0)} = \sum_{d_i \in R} \left[ \log p(d_i | I_1, I_2) - \log p(d_i | H_0) \right]
\]

Quality of descriptor match
“Affinity by Composition”

extending [Boiman & Irani, NIPS’06]

\[
\log \frac{p(R | I_1, I_2)}{p(R | H_0)} = \sum_{d_i \in R} \left[ \log p(d_i | I_1, I_2) - \log p(d_i | H_0) \right]
\]

Quality of descriptor match
“Affinity by Composition”

extending [Boiman & Irani, NIPS’06]

\[
\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)} = \sum_{d_i \in R} \left[ \log p(d_i \mid I_1, I_2) - \log p(d_i \mid H_0) \right] - \text{Err}(d_i \mid I_1, I_2)
\]

How rare is the descriptor?
Generate a “codebook” (quantized descriptors):

- Frequent descriptors $\rightarrow$ Low error
- Rare descriptors $\rightarrow$ High error

Statistically Significant Descriptors

How rare is each descriptor:

$$- \log p(d_i \mid H_0) = ?$$
Statistically Significant Descriptors

How rare is each descriptor:

$$- \log p(d_i \mid H_0) = ?$$

Codebook → Compute error to Nearest Codeword
How rare is each descriptor:

\[- \log p(d_i \mid H_0) = ?\]

\[- \log p(d_i \mid H_0) \approx \text{Err}(d_i \mid \text{Codebook})\]
Statistically Significant Descriptors

How rare is each descriptor:

\[- \log p(d_i \mid H_0) = ?\]

\[- \log p(d_i \mid H_0) \approx \text{Err}(d_i \mid \text{Codebook})\]

The most informative descriptors are the rare ones!
Contribution of $R$ to affinity($I_1, I_2$)

\[
\log \frac{p(R | I_1, I_2)}{p(R | H_0)}
\]

How good is the match

How rare is $R$
“Affinity by Composition”

Contribution of $R$ to affinity($I_1,I_2$)

$$\log \frac{p(R | I_1, I_2)}{p(R | H_0)}$$

$$= \sum_{d_i \in R} [\text{Err}(d_i | \text{codebook}) - \text{Err}(d_i | I_1, I_2)]$$
Contribution of $R$ to affinity($I_1, I_2$)

$$\log \frac{p(R | I_1, I_2)}{p(R | H_0)}$$

$$= \sum_{d_i \in R} \left[ Err(d_i | codebook) - Err(d_i | I_1, I_2) \right]$$
Contribution of $R$ to affinity($I_1, I_2$)

$$\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)}$$

$$= \sum_{d_i \in R} \left[ \text{Err}(d_i \mid \text{codebook}) - \text{Err}(d_i \mid I_1, I_2) \right]$$
“Affinity by Composition”

Contribution of $R$ to affinity $(I_1, I_2)$

\[
\log \frac{p(R \mid I_1, I_2)}{p(R \mid H_0)} = \sum_{d_i \in R} [\text{Err}(d_i \mid \text{codebook}) - \text{Err}(d_i \mid I_1, I_2)]
\]
“Affinity by Composition”

\[ \sum_{R} \log \frac{p(R | I_1, I_2)}{p(R | H_0)} \]

\[ = \sum_{R} \sum_{d_i \in R} \left[ \text{Err}(d_i | \text{codebook}) - \text{Err}(d_i | I_1, I_2) \right] \]
Affinity$(I_1, I_2) = \sum_R \sum_{d_i \in R} \left[ Err(d_i \mid codebook) - Err(d_i \mid I_1, I_2) \right]$
“Region Match”

Using Randomized Search

Extending “Patch Match”

[Barnes et al 2009]
“Region Match”
Using Randomized Search

Extending “Patch Match”
[Barnes et al 2009]
“Region Match”
Using Randomized Search

Extending “Patch Match”
[Barnes et al 2009]
“Region Match”
Using Randomized Search

Extending “Patch Match”
[Barnes et al 2009]
“Region Match”

Using Randomized Search

Extending “Patch Match”

[Barnes et al 2009]
“Region Match”
*Using Randomized Search*

Coherently mapped descriptors ➔ Large shared regions
“Region Match”
Using Randomized Search

CLAIM: 40 samples per descriptor

Find shared regions $|R| \geq 10\%$ with very high probability $\geq 98\%$

$\Rightarrow$ linear complexity $O(n)$

Explicit segmentation $\Rightarrow$ NO NEED!
Going back to our Clustering problem...
Going back to our Clustering problem...

\[ M \text{ images} \]

Uniform Sampling across all images

**CLAIM:** 40 samples per descriptor regardless of \# images

\(\Rightarrow\) At least one strong connection per image
Going back to our Clustering problem...

$M$ images

Uniform Sampling across all images

CLAIM: 40 samples per descriptor regardless of # images

$\Rightarrow$ At least one strong connection per image
Going back to our Clustering problem...

\[ M \text{ images} \]

Sampling Distribution

Guided sampling

\[ \text{Images} \]

Image collaboration

\[ \rightarrow \text{Sparse set of meaningful affinities} \]

\[ \rightarrow \text{Linear complexity (in size of dataset)} \]
Going back to our Clustering problem...

\[ M \text{ images} \]

Sampling Distribution

Guided sampling

“Ideas of crowds of images”

Sparse set of meaningful affinities

Linear complexity (in size of dataset)
Our Full Clustering Algorithm

Sparse set of meaningful affinities

Sampling Distribution

Guided sampling

Images

Update sampling distribution

“Wisdom of crowds of images”

Normalized Cuts

Update Affinity Matrix
Experiments

- **Comparisons on Benchmark Datasets** (Caltech, ETHZ)
  - Significant improvement over *state-of-the-art*
    - (up to 30%)

- **Experiments on more challenging datasets**
  - Tiny datasets
  - PASCAL-VOC
Experiments on Tiny Dataset
20 images (4 classes)
Experiments on Tiny Dataset

Purity = 100%
Experiments on Tiny Dataset
Statistically Significant Descriptors

\[- \log p(d_i \mid H_0) \approx Err(d_i \mid Codebook)\]
Experiments on Tiny Dataset

Statistically Significant Descriptors

\[-\log p(d_i \mid H_0) \approx Err(d_i \mid \text{Codebook})\]

The statistically significant descriptors ➔ on the animals!
PASCAL subset (4 classes)
CARS, BICYCLES, CHAIRS, HORESES
PASCAL subset (4 classes)

CARS, BICYCLES, HORSES, CHAIRS
PASCAL subset (4 classes)

CARS, BICYCLES, HORSES, CHAIRS

Mean purity = 67%

20% **better** than a baseline of: SPM + Ncuts
1. “Affinity by composition”
   - Look for RARE shared regions

2. Codebook Quantization Error
   - Estimate how rare a region is.

3. Randomized search using the “Wisdom of Crowds of images”
   - Find shared regions
   - Linear complexity

4. State of the art results
   - Benchmark datasets
   - New challenging datasets

Thank you!