Towards Automated Large Scale Discovery of Image Families

Mohamed Aly\textsuperscript{1}  
Mario Munich\textsuperscript{2} 

Peter Welinder\textsuperscript{1}  
Pietro Perona\textsuperscript{1}

\textsuperscript{1} Computational Vision Lab, Caltech, Pasadena, CA  
\textsuperscript{2} Evolution Robotics, Pasadena, CA
Image Families

Eiffel Tower
Image Families

Aeon Flux
Automated Family Discovery

 Importance

 - Content-based Image Search Engines
 - Organization of Personal Image Collections
 - Computer Vision Research Datasets
Automated Family Discovery

 Importance
  - Content-based Image Search Engines
  - Organization of Personal Image Collections
  - Computer Vision Research Datasets

 Challenges
  - Ease of Gathering Huge Image Collections
  - Large Scale: Number of Images & Families
  - Need for Unsupervised Methods
Goal

- Explore automated family discovery
Goal

🔹 Explore automated family discovery
🔹 Compare different methods
  - Image representation
  - Similarity measures
  - Clustering scenarios
Goal

✿ Explore automated family discovery

✿ Compare different methods
  ✿ Image representation
  ✿ Similarity measures
  ✿ Clustering scenarios

✿ Scale up to ~ 11,400 images and ~ 6,300 families
Goal

Explore automated family discovery

Compare different methods
- Image representation
- Similarity measures
- Clustering scenarios

Scale up to ~ 11,400 images and ~ 6,300 families

Introduce two new annotated datasets
Approach

Compute Image Features
Approach

- Compute Image Features
- Local Vs Global Features
Approach

- Compute Image Features
- Compute Similarity Matrix
- Local Vs Global Features
Approach

1. Compute Image Features
2. Compute Similarity Matrix
3. Local Vs Global Features
4. Euclidean Vs Matching
Approach

Compute Image Features

Compute Similarity Matrix

Cluster Images

Local Vs Global Features

Euclidean Vs Matching
Approach

- Compute Image Features
- Compute Similarity Matrix
- Cluster Images

Local Vs Global Features
Euclidean Vs Matching
Semi- Vs Unsupervised
Approach

1. Compute Image Features
2. Compute Similarity Matrix
3. Cluster Images
4. Image Families
   - Local Vs Global Features
   - Euclidean Vs Matching
   - Semi- Vs Unsupervised
Image Features: Global

- SIFT [Lowe '04]
- Gist [Oliva & Torralba '01]
- HOG [Dalal & Triggs '05]
- Bag-of-Words (BoW)
  - 1K, 5K, 10K, 25K, 50K Dictionary
  - Raw and Tf-Idf Weighted Histograms
Image Features: Local
Image Features: Local

- Harris Affine Covariant Detector
  - SIFT Descriptor

[Mikolajczyk & Schmid '04]
Image Similarity: Global

★ Euclidean Distance
Image Similarity: Global

★ Euclidean Distance

$\text{Image } i$
Image Similarity: Global

★ Euclidean Distance

Image $i$ × $i$ = Similarity Matrix
Image Similarity: Local

(1) Simple
(1) Simple

Image Similarity: Local

Dataset

Kd-Forest
(1) Simple
Image Similarity: Local

(1) Simple

Image $i$

Dataset

Kd-Forest

Nearest Neighbors
Image Similarity: Local

(1) Simple

Dataset

Kd-Forest

Image $i$

Nearest Neighbors

Similarity Matrix $S$

$\text{# neigh. } i$
Image Similarity: Local

(2) Image-aff

Image $i$

Dataset

Kd-Forest

Nearest Neighbors
(2) Image-aff

Image $i$ → Dataset → Kd-Forest → Nearest Neighbors ≥ $t_c$ neighb. → Exhaustive Search
(2) Image-aff

Image $i$

Dataset

Kd-Forest

$\geq t_c$ neighbors

Nearest Neighbors

Exhaustive Search

RANSAC Affine
Image Similarity: Local

(2) Image-aff

Image $i$ → Kd-Forest → Nearest Neighbors $\geq t_c$ neigh. → Exhaustive Search → RANSAC Affine

Dataset

Similarity Matrix S

$\#$ inliers
Image Similarity: Local

(3) Region-aff

Image $i$ -> Kd-Forest

Dataset

Nearest Neighbors

$\geq t_c$ neigh. -> Exhaustive Search
Image Similarity: Local

(3) Region-aff

Image $i$ → Kd-forest → Nearest Neighbors $\geq t_c$ neighb. → Exhaustive Search → RANSAC Affine
Image Similarity: Local

(3) Region-aff

Dataset

Kd-Forest

Image $i$

Similarity Matrix $S$

Nearest Neighbors

Exhaustive Search

$\geq t_c$ neighb.

RANSAC Affine

# inliers
Clustering: Semi-supervised
Clustering: Semi-supervised

Normalized Cuts (NC) [Shi & Malik 2000]

\[
\max \frac{1}{K} \sum_{i=1}^{K} \frac{\text{links}(V_i, V_i^C)}{\text{degree}(V_i)}
\]

\[
\text{links}(A, B) = \sum_{i \in A, j \in B} s_{ij} \quad \text{degree}(A) = \text{links}(A, S)
\]
Clustering: Semi-supervised

🌟 Normalized Cuts (NC)  

\[
\max \frac{1}{K} \sum_{i=1}^{K} \frac{\text{links}(V_i, V_i^C)}{\text{degree}(V_i)}
\]

\[
\text{links}(A, B) = \sum_{i \in A, j \in B} s_{ij} \quad \text{degree}(A) = \text{links}(A, S)
\]

🌟 Agglomerative Clustering (Ag)  

\[
\text{Average Linkage} = \frac{\sum_{i \in A, j \in B} s_{ij}}{|A||B|}
\]

[Shi & Malik 2000]  

[Jain et al. '99]
Clustering: Unsupervised

⭐ Clustering with Ranked Connected Components Labeling (Cranclе)

Similarity Matrix $S$

```
  5 2 0 7 8 1 2
```
Clustering: Unsupervised

🌟 Clustering with Ranked Connected Components Labeling (Cranclle)

Similarity Matrix $S$

Top $r$ values

\[
\begin{array}{ccccccc}
& 5 & 2 & 0 & 7 & 8 & 1 & 2 \\
\end{array}
\]
Clustering: Unsupervised

Cluster with Ranked Connected Components Labeling (Crancl)

Similarity Matrix $S$

Connectivity Matrix $C$

Top $r$ values
Clustering: Unsupervised

Clustering with Ranked Connected Components Labeling (Crancle)

Similarity Matrix $S$

$$
\begin{bmatrix}
5 & 2 & 0 & 7 & 8 & 1 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

Connectivity Matrix $C$

$$
\begin{bmatrix}
1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

Top $r$ values
Clustering with Ranked Connected Components Labeling (Crancle)

Similarity Matrix $S$

Connectivity Matrix $C$

Image Families
Performance Measures

Mean Confusion Matrix Performance (MCMP)

\[ \frac{1}{K} \sum_{f=1}^{K} \left( \frac{u_{ff}}{\sum_{k=1}^{K} u_{fk}} \right) \times 100\% \]
Performance Measures

★ Mean Confusion Matrix Performance (MCMP)

$$ \frac{1}{K} \sum_{f=1}^{K} \left( \frac{u_{ff}}{\sum_{k=1}^{K} u_{fk}} \right) \times 100\% $$

★ F-Measure (FM)  

[Hammouda & Kamel '06]

$$ prec(f, k) = \frac{L_{fk}}{|k|} \quad rec(f, k) = \frac{L_{fk}}{|f|} $$
Performance Measures

Mean Confusion Matrix Performance (MCMP)

\[
\frac{1}{K} \sum_{f=1}^{K} \left( \frac{u_{f f}}{\sum_{k=1}^{K} u_{f k}} \right) \times 100\%
\]

F-Measure (FM)  [Hammouda & Kamel '06]

\[
p_{red}(f, k) = \frac{L_{f k}}{|k|} \quad r_{ed}(f, k) = \frac{L_{f k}}{|f|}
\]

\[
\frac{1}{N} \sum_{f} \left( \max_{k} \frac{2 \times p_{red}(f, k) \times r_{ed}(f, k)}{p_{red}(f, k) + r_{ed}(f, k)} \right) \times |f| \times 100\%
\]
Datasets: Caltech Games

- CD/DVD Game Covers
- 11,431 images
- 6,361 families
- Manually labelled

[http://vision.caltech.edu/malaa/datasets.php]
Datasets: Caltech Games

- CD/DVD Game Covers
- 11,431 images
- 6,361 families
- Manually labelled

<table>
<thead>
<tr>
<th>Subset</th>
<th>#images</th>
<th>#families</th>
</tr>
</thead>
<tbody>
<tr>
<td>Games 01</td>
<td>11,431</td>
<td>6,361</td>
</tr>
<tr>
<td>Games 02</td>
<td>7,054</td>
<td>1,984</td>
</tr>
<tr>
<td>Games 03</td>
<td>5,212</td>
<td>1,063</td>
</tr>
<tr>
<td>Games 04</td>
<td>3,961</td>
<td>646</td>
</tr>
<tr>
<td>Games 06</td>
<td>2,312</td>
<td>273</td>
</tr>
<tr>
<td>Games 08</td>
<td>1,380</td>
<td>127</td>
</tr>
<tr>
<td>Games 12</td>
<td>645</td>
<td>43</td>
</tr>
<tr>
<td>Games 16</td>
<td>210</td>
<td>10</td>
</tr>
</tbody>
</table>

[http://vision.caltech.edu/malaa/datasets.php]
Datasets: Caltech Buildings

50 Buildings, 5 images each = 250 images

[http://vision.caltech.edu/malaa/datasets.php]
Datasets: Oxford Landmarks

"Good" Subset
11 landmarks with 272 images

[http://www.robots.ox.ac.uk/~vgg/data/oxbuildings] [Philbin et al. '07]
Experiments: Semi-Supervised

BoW Dictionary Size

Games 16
- bag1k-ag: 89
- bag5k-ag: 100
- bag10k-ag: 100
- bag25k-ag: 100
- bag50k-ag: 100

Games 08
- bag1k-ag: 56
- bag5k-ag: 80
- bag10k-ag: 87
- bag25k-ag: 91
- bag50k-ag: 94

Games 03
- bag1k-ag: 55
- bag5k-ag: 73
- bag10k-ag: 78
- bag25k-ag: 81
- bag50k-ag: 84

Games 01
- bag1k-ag: 79
- bag5k-ag: 82
- bag10k-ag: 82
- bag25k-ag: 83
- bag50k-ag: 84

Caltech
- bag1k-ag: 49
- bag5k-ag: 53
- bag10k-ag: 64
- bag25k-ag: 75
- bag50k-ag: 82

Oxford
- bag1k-ag: 30
- bag5k-ag: 38
- bag10k-ag: 42
- bag25k-ag: 51
- bag50k-ag: 68
Experiments: Semi-Supervised

BoW Tf-Idf Weighting

Games 16
- bag10k-ag: 100
- bag10k-ag-tf-idf: 100
- bag25k-ag: 100
- bag25k-ag-tf-idf: 5
- bag50k-ag: 100
- bag50k-ag-tf-idf: 5

Games 08
- bag10k-ag: 87
- bag10k-ag-tf-idf: 87
- bag25k-ag: 91
- bag25k-ag-tf-idf: 94
- bag50k-ag: 94
- bag50k-ag-tf-idf: 96

Games 03
- bag10k-ag: 78
- bag10k-ag-tf-idf: 76
- bag25k-ag: 81
- bag25k-ag-tf-idf: 81
- bag50k-ag: 84
- bag50k-ag-tf-idf: 85

Games 01
- bag10k-ag: 82
- bag10k-ag-tf-idf: 82
- bag25k-ag: 83
- bag25k-ag-tf-idf: 83
- bag50k-ag: 84
- bag50k-ag-tf-idf: 85

Caltech
- bag10k-ag: 64
- bag10k-ag-tf-idf: 68
- bag25k-ag: 75
- bag25k-ag-tf-idf: 80
- bag50k-ag: 82
- bag50k-ag-tf-idf: 20

Oxford
- bag10k-ag: 42
- bag10k-ag-tf-idf: 52
- bag25k-ag: 51
- bag25k-ag-tf-idf: 55
- bag50k-ag: 68
- bag50k-ag-tf-idf: 11
Experiments: Semi-Supervised Clustering Method

Cluster:

- Games 16:
  - bag25k-nc: 100
  - bag25k-ag: 100
  - bag50k-nc: 63
  - bag50k-ag: 100
  - simple-nc: 100
  - simple-ag: 100

- Games 08:
  - bag25k-nc: 82
  - bag25k-ag: 91
  - bag50k-nc: 82
  - bag50k-ag: 94
  - simple-nc: 93
  - simple-ag: 96

- Games 03:
  - bag25k-nc: 30
  - bag25k-ag: 81
  - bag50k-nc: 33
  - bag50k-ag: 84
  - simple-nc: 75
  - simple-ag: 92

- Games 01:
  - bag25k-nc: 13
  - bag25k-ag: 83
  - bag50k-nc: 13
  - bag50k-ag: 84
  - simple-nc: 29
  - simple-ag: 91

- Caltech:
  - bag25k-nc: 86
  - bag25k-ag: 75
  - bag50k-nc: 72
  - bag50k-ag: 82
  - simple-nc: 74
  - simple-ag: 72

- Oxford:
  - bag25k-nc: 38
  - bag25k-ag: 51
  - bag50k-nc: 28
  - bag50k-ag: 68
  - simple-nc: 43
  - simple-ag: 50
Experiments: Semi-Supervised

Local Vs Global

Games 16
sift-ag: 18
hog-ag: 19
gist-ag: 29
bag50k-ag: 100
simple-ag: 100
image-aff-ag: 100
region-aff-ag: 100

Games 08
sift-ag: 22
hog-ag: 25
gist-ag: 33
bag50k-ag: 94
simple-ag: 96
image-aff-ag: 94
region-aff-ag: 94

Games 03
sift-ag: 36
hog-ag: 36
gist-ag: 39
bag50k-ag: 84
simple-ag: 92
image-aff-ag: 90
region-aff-ag: 91

Games 01
sift-ag: 76
hog-ag: 75
gist-ag: 76
bag50k-ag: 84
simple-ag: 91
image-aff-ag: 90
region-aff-ag: 91

Caltech
sift-ag: 29
hog-ag: 34
gist-ag: 31
bag50k-ag: 82
simple-ag: 72
image-aff-ag: 57
region-aff-ag: 62

Oxford
sift-ag: 17
hog-ag: 18
gist-ag: 19
bag50k-ag: 68
simple-ag: 50
image-aff-ag: 52
region-aff-ag: 45
Experiments: Semi-Supervised
Experiments: Unsupervised

Local Vs Global

Games 16 (10 fam.)
- bag25k-cc: 69
- bag50k-cc: 46
- simple-cc: 19
- image-aff-cc: 34
- region-aff-cc: 35

Games 08 (127 fam.)
- bag25k-cc: 83
- bag50k-cc: 90
- simple-cc: 85
- image-aff-cc: 87
- region-aff-cc: 94

Games 03 (1063 fam.)
- bag25k-cc: 82
- bag50k-cc: 82
- simple-cc: 84
- image-aff-cc: 83
- region-aff-cc: 84

Games 01 (6361 fam.)
- bag25k-cc: 80
- bag50k-cc: 80
- simple-cc: 78
- image-aff-cc: 77
- region-aff-cc: 78

Caltech (50 fam.)
- bag25k-cc: 79
- bag50k-cc: 87
- simple-cc: 89
- image-aff-cc: 61
- region-aff-cc: 64

Oxford (11 fam.)
- bag25k-cc: 7
- bag50k-cc: 7
- simple-cc: 7
- image-aff-cc: 7
- region-aff-cc: 7
Experiments: Unsupervised
Summary

👀 Need for diverse datasets for comparisons
Summary

- Need for diverse datasets for comparisons
- SIFT, HOG, Gist not suitable
Summary

- Need for diverse datasets for comparisons
- SIFT, HOG, Gist not suitable
- BoW comparable and promising
Summary

- Need for diverse datasets for comparisons
- SIFT, HOG, Gist not suitable
- BoW comparable and promising
- Need for more research on the problem
  - Scaling up to millions of images
  - Scaling up to hundreds of thousands of families
  - Unsupervised algorithms
Questions?

Thank You