WISE: Large Scale Content-Based
Web Image Search

Michael Isard
Joint with: Qifa Ke, Jian Sun, Zhong Wu
Microsoft Research Silicon Valley
Query by Images

“A picture is worth a thousand words.”

What leaf?

Artist? Higher resolution?

©: Who else using this?

Bank website?

Email?

Ad. in e-bay?

...
Partial-Duplicate Image Search

• Given a query image, find its partial duplicates from a database of web-images
Two Major Challenges

• How to represent images
  – No text annotations or labels
  – Noise and modification

• How to efficiently index and query images
  – Large number of images (millions)
Image Representation: Bag-of-Words
[CVPR’09, ICCV’09]

1: Feature extraction: Bundle Features
Detection

Normalization

Description
[Lowe 2004, Winder et al 2007]

2: Quantization
Code-book

3: Representation
Bag-of-words
[Sivic&Zisserman’2003]
Matching query to database

• Use an index
  – Each visual word has a ‘posting list’
  – Lists every image containing the word

• At query time
  – Look up the posting list for each query word
  – Merge lists to find candidate images
    • Partial match: don’t need every word to be present
How much work to query?

• Disk-based index, bottleneck is random reads
  – One seek per posting list
• Also one seek per matching image
  – To fetch thumbnail etc
• Keep as little information as possible in posting lists, to keep index size small
Index Pipeline

• Implemented in a large computer cluster
  – 256-nodes, using Dryad/DryadLINQ
Query Pipeline
Bag-of-Words: Limitations

- Quantization
  - Lost discriminative power
  - Sensitive to image variations and noises
  - Soft quantization [Philbin et al, CVPR 2008]
  - Hamming embedding [Jegou et al, ECCV 2008]
Geometric verification

• In practice, bag of words is too weak
• Does not exploit any geometry
• Post-process to check spatial layout of matching features
• Requires a disk seek per image
  – Only used as a re-ranking step to shortlist of matched images
Geometric Re-ranking

Re-rank top 300 images

![Diagram showing mAP vs. Number of images for baseline (bag-of-words) and baseline + reranking methods.](image-url)
Geometry in the index

• Previous works:
  – Jegou et al ECCV 2008
    • Try to match similar orientations and scales
  – Perdoch et al CVPR 2009
    • Match oriented features more effectively

• Still feature-by-feature
  – Global geometric consitency applied at the end
Single Feature is Weak
Neighboring Features ?
Define Neighboring Features

• Previous works
  – kNN voting [Sivic&Zisserman 2003]
  – Higher-order spatial features
    [Liu et al][Yuan et al][Tirilley et al][Quack et al]
  – Post geometric spatial verification
  – Geometric Min-Hash [Chum et al 2009]

• Challenges
  – Repeatable
  – Partial matching
  – Scalable: simple enough to build into index
Define Neighboring Features

DoG Features [Lowe 2004]
- point features
- repeatable

MSER Features [Matas et al 2002]
- region features
- repeatable
Define Neighboring Features

DoG Features [Lowe 2004]
- point features
- repeatable

MSER Features [Matas et al 2002]
- region features
- repeatable

*region groups points?*
Bundled Feature: Definition

- Bundled Feature =
  A set of DOG features bundled by a MSER region
Bundled Feature: Definition

Bundled Features
Matching Bundles: Membership

Query bundle \( q = \{q_j\} = \{\text{diamond}, \text{triangle}, \times, \circ\} \)

Matched bundle \( p = \{p_i\} \)

Membership score:

\[
M_m(q; p) = |q \cap p| = 4
\]

Voting weight:

\[
v(q_j) = M_m(q; p) = 4
\]

\[
Sim(I_1, I_2) = \sum_{\{q_j\}} v(q_j) = \sum_{\{q_j\}} 4 = 16
\]
Matching Bundles: Membership

Query bundle $q = \{q_j\} = \{\Diamond \Delta \times \circ\}$

Matched bundles $p_1, p_2, p_3$

Membership score:

$M_m(q; p_1) = |q \cap p_1| = 2$

$M_m(q; p_2) = |q \cap p_2| = 1$

$M_m(q; p_3) = |q \cap p_3| = 2$

$v(q_j) = \max_{p_k} \left\{ M_m(q; p_k) \mid q_j \in q \right\}$

$v(q_2) = 2 \quad v(q_1) = 2 \quad v(q_3) = \max(1, 2) = 2 \quad v(q_4) = 2$

$Sim(I_1, I_2) = \sum_{\{q_j\}} v(q_j) = 8$
Matching Bundles: Geometric Constraint

Penalize inconsistent relative orders:

\[ M_g(q; p) = -\sum \delta(O_q[p_i] > O_q[p_{i+1}]) \]
Matching Bundles: Formulation

- **Bundle matching score:**

\[
M(q; p) = M_m(q; p) + \lambda M_g(q; p)
\]

- **Image matching score:**

\[
v(q_j) = \max_{p_k} \left\{ M(q; p_k) \mid q_j \in q \right\}
\]

\[
Sim(I_1, I_2) = \sum_{\{q_j\}} v(q_j)
\]

- **Repeatable**
- **Partial matching**
- **Scalable?**
Inverted Index (without Bundles)

Image ID = 27
Inverted Index with Bundles

Image ID = 27
Retrieval

Query Image $I_q$

Inverted index with bundle bits

Top candidate images
Experimental Settings

• Image database:
  – 1M web images from query-click log

• Ground truth partial duplicates
  – 780 known partial duplicate images in 19 groups

• Baseline bag-of-words
  – Visual word vocabulary size = 1 M
  – Soft quantization factor = 4
  – 500 features per image
Partial Duplicate Example
Partial Duplicate Example
Example Query Results

Query

Challenging cases
Evaluation: Precision-Recall

• A query returns $N$ images
  – $T$: correct matches
  – $A$: expected matches

Precision = \frac{T}{N} \quad \text{Recall} = \frac{T}{A}
Comparison: Precision-Recall

Query image:

Baseline bag-of-words (started from 13th)

Bundled features (started from 13th)
More Precision-Recall Comparisons
Evaluation: mAP

• Average Precision (AP) for one query:
  – Area under Precision-Recall curve

• mAP: mean of AP’s from all testing queries

\[
\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i
\]
mAP: Baseline Bag-of-Words

Number of images

mAP
mAP: Hamming Embedding (HE)
mAP: Bundle (Membership)
mAP: Bundle (both terms)

![Graph showing mAP against Number of images]
mAP: Bundle + HE

![Graph showing mAP performance over number of images]

- **Baseline**
- **HE**
- **Bundled (membership)**
- **Bundled**
- **Bundled + HE**

Number of images:
- 50,000
- 200,000
- 500,000
- 1,000,000

mAP values:
- 0.7
- 0.6
- 0.5
- 0.4
- 0.35

Baseline HE bundled bundled(membership) bundled + HE

49% improvement
Bundle VS. Geometric Re-ranking

Re-rank top 300 images

mAP

baseline (bag-of-words)
bundle
baseline + reranking

Number of images

50000 200000 500000 1000000
Bundle + Geometric Re-ranking

Re-rank top 300 images

24% 77%

mAP vs. Number of images
More Results
Failure Case
Demo Setup

Client

Web Server

Index Servers

Document Server

6 million images
Demo

Query image

Results
Conclusion

• Bundle feature
  – More discriminative
  – Enforce spatial constraints while traversing index
  – Partial match
  – Scalable: built into index
Thanks!