Efficient Matching for Retrieval and Recognition

Kristen Grauman
Trevor Darrell
MIT

DOE CSGF
Annual Fellows’ Conference
June 22, 2006
Recognition

Specific

Wild card
Tower Bridge
Steve Carell

Categories

butterfly
butterfly
building
building
Key challenges: robustness

Illumination
Object pose
Clutter

Occlusions
Intra-class appearance
Viewpoint
Key challenges: efficiency

- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search
- 18 billion+ prints produced from digital camera images in 2004
- 295.5 million camera phones sold in 2005
Global image representations

Map image to a single vector based on overall characteristics

- vector of pixel intensities
- grayscale / color histogram
- bank of filter responses …
Limitations of global representations

- Success may rely on alignment
- All parts of image impact description
Local image representations

Describe component regions or patches separately

- SIFT [Lowe]
- Shape context [Belongie et al.]
- Superpixels [Ren et al.]
- Maximally Stable Extremal Regions [Matas et al.]
- Salient regions [Kadir et al.]
- Harris-Affine [Schmid et al.]
- Spin images [Johnson and Hebert]
- Geometric Blur [Berg et al.]
Comparing images

A meaningful way of comparing two image descriptions is key for various tasks:

- Nearest neighbor
- Classifying
- Regression
- Clustering

- Image retrieval
- Object recognition
- Pose inference
- Category discovery
How to handle sets of features?

- Each instance is unordered set of vectors
- Varying number of vectors per instance
Previous approaches

- Voting
  [Schmid, Lowe, Tuytelaars et al.,…]

- Fit (parametric) distributions to sets
  [Kondor & Jebara, Moreno et al., Cuturi & Vert, Wolf & Shashua]

- Compare histograms over prototype features
  [Csurka et al., Sivic & Zisserman, Lazebnik & Ponce, Agarwal & Triggs]

- Explicit search for correspondences
  [Wallraven et al., Lyu, Boughorbel et al., Belongie et al., Rubner et al., Berg et al., Gold & Rangarajan, Shashua & Hazan,…]

- Multi-resolution matching approximations
  [Indyk & Thaper, Charikar, Agarwal & Varadarajan…]

Limitations

- Costly comparisons
- Costly training
- Sensitivity to clutter
- Restrictive assumptions
- Assumes feature independence
- Restricted to sets of equal sizes
- Inapplicable for discriminative learning
Partial matching

Compare sets by computing a *partial matching* between their features.

\[
\max_{\pi: X \rightarrow Y} \sum_{x_i \in X} S(x_i, \pi(x_i))
\]
Computing the partial matching

- Optimal matching \( O(m^3) \)
- Greedy matching \( O(m^2 \log m) \)
- Pyramid match \( O(mL) \) \[Grauman and Darrell, ICCV 2005]\n
for sets with \( O(m) \) features
Pyramid match overview

\[
X = \{\bar{x}_1, \ldots, \bar{x}_m\} \quad Y = \{\bar{y}_1, \ldots, \bar{y}_n\}
\]

\[
\max_{\pi : X \rightarrow Y} \sum_{x_i \in X} S(x_i, \pi(x_i))
\]

optimal partial matching
Pyramid match overview

Pyramid match measures similarity of a partial matching between two sets:

- Place multi-dimensional, multi-resolution grid over point sets
- Consider points matched at finest resolution where they fall into same grid cell
- Approximate optimal similarity with worst case similarity within pyramid cell

*No explicit search for matches!*
Pyramid extraction

\[ X = \{ \vec{x}_1, \ldots, \vec{x}_m \}, \quad \vec{x}_i \in \mathbb{R}^d \]

Histogram pyramid: level \( i \) has bins of size \( 2^i \)

\[ \Psi(X) = [H_0(X), \ldots, H_L(X)] \]
Counting matches

Histogram intersection

\[ \mathcal{I}(H(X), H(Y)) = \sum_{j=1}^{r} \min(H(X)_j, H(Y)_j) \]

\[ \mathcal{I}(H(X), H(Y)) = 3 \]
Counting new matches

Histogram intersection

\[ I(H(\mathbf{X}), H(\mathbf{Y})) = \sum_{j=1}^{r} \min(H(\mathbf{X})_j, H(\mathbf{Y})_j) \]

matches at this level

\[ N_i = I(H_i(\mathbf{X}), H_i(\mathbf{Y})) - I(H_{i-1}(\mathbf{X}), H_{i-1}(\mathbf{Y})) \]

matches at previous level

Difference in histogram intersections across levels counts *number of new pairs* matched
Pyramid match

\[ K_\Delta \left( \Psi(X), \Psi(Y) \right) = \sum_{i=0}^{L} \frac{1}{2^i} \left( I(H_i(X), H_i(Y)) - I(H_{i-1}(X), H_{i-1}(Y)) \right) \]

- For similarity, weights inversely proportional to bin size
- Normalize kernel values to avoid favoring large sets

measure of difficulty of a match at level \( i \)

number of newly matched pairs at level \( i \)
Efficiency

Pyramid match complexity \( O(dmL) \)

\( d \) feature dimension
\( m \) set size
\( L = \log(D) \) number of pyramid levels
\( D \) range of feature values
Example pyramid match

\[ \mathcal{I}(H_0(X), H_0(Y)) = 2 \quad \rightarrow \quad N_0 = 2, \quad w_0 = 1 \]
Example pyramid match

\[ \mathcal{I}(H_1(X), H_1(Y)) = 4 \quad \rightarrow \quad N_1 = 4 - 2 = 2 \]

\[ w_1 = \frac{1}{2} \]
Example pyramid match

\[ I(\mathcal{H}_2(X), \mathcal{H}_2(Y)) = 5 \quad \rightarrow \quad N_2 = 5 - 4 = 1 \]
\[ w_2 = \frac{1}{4} \]
Example pyramid match

Pyramid match

\[ K_\Delta = \sum_{i=0}^{L} w_i N_i \]

\[ = 1(2) + \frac{1}{2}(2) + \frac{1}{4}(1) = 3.25 \]

Optimal match

\[ K = \max_{\pi: X \rightarrow Y} \sum_{x_i \in X} S(x_i, \pi(x_i)) \]

\[ = 1(2) + \frac{1}{2}(3) = 3.5 \]
Learning with the pyramid match

• Kernel-based methods
  – Embed data into a Euclidean space via a similarity function (kernel), then seek linear relationships among embedded data
  – Include classification, regression, clustering, dimensionality reduction,…
  – Efficient and good generalization

• Pyramid match forms a Mercer kernel
### Category recognition results

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match [Wallraven et al.]</td>
<td>$O(dm^2)$</td>
</tr>
<tr>
<td>Pyramid match</td>
<td>$O(dmL)$</td>
</tr>
</tbody>
</table>

**ETH-80 data set**

![ETH-80 data set](image)

---

**Graphs**

- **Accuracy**
- **Time (s)**
- **Mean number of features**
Category recognition results

- 101 categories
  40-800 images per class
  8677 total images
- One-vs-all SVM with PMK
- Features:
  - Densely sampled
  - SIFT descriptor + spatial
  - Average $m=1140$ per set

Caltech 101 data set

Data provided by Fei-Fei Li, Rob Fergus and Pietro Perona
Category recognition results

Caltech 101 Categories data set

mean recognition rate per class

0.002 s / match

5 s / match

Averaged over 10 runs with randomly selected training examples

number of training examples per class

PMK
Holub, Welling, & Perona
Berg, Berg, & Malik
Serre, Wolf, & Poggio
Fei-Fei, Fergus, & Perona
SSD baseline
Our 5 best categories

1. 100%
2. 100%
3. 99.7%
4. 99.1%
5. 98.2%
Our 5 worst categories

1. Ants - 7.7%
2. Beavers - 11.2%
3. Crabs - 11.5%
4. Birds - 11.8%
5. Anchors - 12.3%
Most confused category pairs

**schooner**  *n.*
A fore-and-aft rigged sailing vessel having at least two masts, with a foremast that is usually smaller than the other masts.

**ketch**  *n. Nautical*
A two-masted fore-and-aft-rigged sailing vessel with a mizzenmast stepped aft of a taller mainmast but forward of the rudder.
Most confused category pairs

lotus

water lily
Most confused category pairs

- Gerenuk
- Kangaroo
Most confused category pairs

nautilus

brain
Pyramid match applications

A meaningful way of comparing two image descriptions is key for various tasks:

- Nearest neighbor
- Classifying: Object recognition
- Regression: Pose inference
- Clustering: Category discovery

Image retrieval
Contributions

**Pyramid match**: a new similarity measure over sets of vectors that efficiently forms an implicit partial matching

- linear time complexity
- no independence assumption
- model-free
- insensitive to clutter
- positive-definite function (a **kernel**)

Demonstrated effectiveness for retrieval, recognition, and regression tasks with local image features
Sets of features elsewhere

diseases as sets of gene expressions
methods as sets of instructions
documents as bags of words
Acknowledgements

Trevor Darrell

MIT Vision Interface Group

DOE Computational Science Graduate Fellowship