MOTIVATION
Can you distinguish between insects in the top and bottom rows?

Problems
How to categorize images showing very similar object categories?

Solution
Train a classifier directly on descriptors of image features, instead of building a visual dictionary and training on the dictionary words
Use class evidence accumulated from all descriptors, instead of voting class decisions made on individual descriptors

Challenge
How to handle volumes of unquantized data? => Evidence trees

APPLICATION: BIOMONITORING
Biomonitoning by categorizing stoneflies
- Sensitive and robust indicator of water-stream health and quality
- Easy to collect specimens
- Limitation: High degree of expertise required to classify specimens

STONEFLY9 DATABASE
- Small inter-class differences and large intra-class variations
- No guarantee of fully frontal, dorsal views of insects
- Insects may be only partially visible
- Size, color, and texture change significantly with the insect’s age
- Insects appear in a wide range of poses

VISUAL DICTIONARIES GIVE MEDIOCRE RESULTS ON STONEFLY9
- Dictionaries constructed using purely unsupervised methods
- Information lost in quantizing keypoints to dictionary entries
- Requires manual tuning of: number of clusters, quantization, etc.

Dictionary-Free Categorization of Very Similar Objects via Stacked Evidence Trees

Gonzalo Martinez-Muñoz1, Wei Zhang1, Nadia Payet1, Sinisa Todorovic1, Natalia Larios1, Asako Yamamuro1, David Lytle1, Andrew Moldenke1
Eric Mortensen1, Robert Paasch1, Linda Shapiro2, Thomas G. Dietterich1
1Oregon State University, Corvallis, OR 97333 USA 2University of Washington, Seattle, WA 98195 USA

PROJECT HOME PAGE: http://web.engr.oregonstate.edu/~tgd/bugid/


 leans separately for each task. Average AUC is shown in parenthesis

Table:

<table>
<thead>
<tr>
<th>Detector</th>
<th>Dataset</th>
<th>Discriminative Patches</th>
<th>Visual dict.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harris + Hessian Affine + PCBR</td>
<td>INRIA_Moosmann</td>
<td>16.1±1.8</td>
<td>34.0±3.1</td>
</tr>
<tr>
<td>Harris + Hessian Affine + PCBR</td>
<td>INRIA_Marzszalek</td>
<td>23.9±1.3</td>
<td>34.0±3.1</td>
</tr>
<tr>
<td>Harris + Hessian Affine + PCBR</td>
<td>XRCE</td>
<td>28.1±1.0</td>
<td>23.9±1.3</td>
</tr>
<tr>
<td>Harris + Hessian Affine + PCBR</td>
<td>QMUL_HSLS</td>
<td>22.8±2.5</td>
<td>23.9±1.3</td>
</tr>
<tr>
<td>Harris + Hessian Affine + PCBR</td>
<td>INRIA_Nowak</td>
<td>19.0±2.5</td>
<td>23.9±1.3</td>
</tr>
</tbody>
</table>

RESULTS

- Edge + (Kadir + Hessian Affine + PCBR) x (SIFT, color SIFT, Hessian) => 4 random forests
- Stacking: Boosting of 200 decision trees
- Visual dictionary:
  - K-means 100 clusters per detector/descriptor and class
  - Mapping: nearest cluster center and accumulated into a histogram
  - Final classifier: Boosted decision-tree classifier containing 200 trees

Advantages over visual-dictionary methods

1. No information loss, because no quantization
2. Evidence trees are grown discriminatively => no unsupervised steps
3. No manual parameter tuning
4. Low sensitivity to a wide range of values of input parameters

CONCLUSIONS

- We categorize highly articulated objects with large intra-category variations and small inter-category differences by using evidence random forests trained directly on descriptors
- We have provided a mathematical model of our approach
- Experiments on STONEFLY9 and PASCAL06 datasets demonstrate validity and generality of our approach.