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

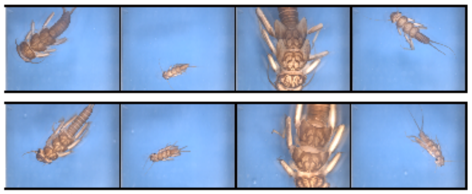
Gonzalo Martínez-Muñoz[1] Wei Zhang[1] Nadia Payet[1] Sinisa Todorovic[1]
Natalia Larios[2] Asako Yamamuro[1] David Lytle[1] Andrew Moldenke[1]
Eric Mortensen[3] Robert Paasch[1] Linda Shapiro[2] Thomas G. Dietterich[1]

[1]Oregon State University, Corvallis, OR 97331 USA
[2]University of Washington, Seattle, WA 98195 USA
[3]Lucidyne Technologies, Inc., Corvallis, OR 97333 USA

Project Home Page: http://web.engr.oregonstate.edu/~tgd/bugid/

MOTIVATION

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



Top row Calineuria californica; Bottom row Doroneuria baumanni;

Even trained human experts cannot readily categorize these images, but have to examine the insects themselves!

PROBLEM

How to categorize images showing very similar object categories?

OUR 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

BIOMONITORING 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

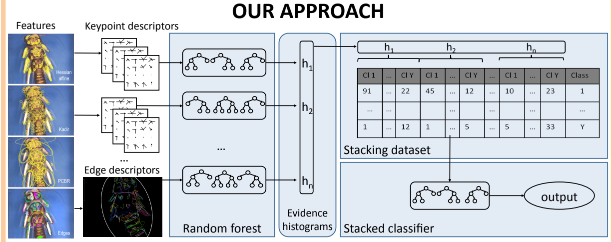
STONEFLY9 DATASET

•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

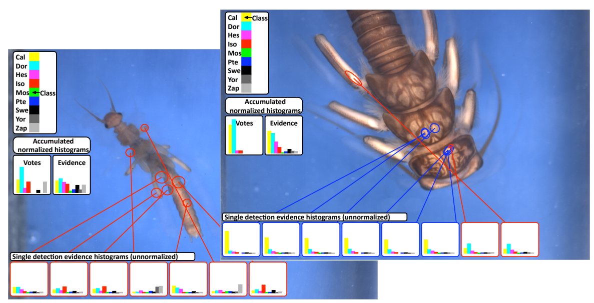
•Dictionaries constructed using purely unsupervised methods
•Information lost in quantizing keypoints to dictionary entries
•Requires manual tuning of: number of clusters, quantization, etc.

OUR APPROACH



FIRST STAGE (Random forest)

1. Random forest is trained directly on descriptors
   - Training images are sampled from the training set with replacement
   - Descriptors of features extracted from a training image are labeled with the class of that image
   - Descriptors are "dropped" through each tree in the random forest
   - In each leaf, a class histogram is stored



Evidence/class histograms

SECOND STAGE (Stacking)

2. Stacking dataset is created:
   - Leaf histograms are summed over all trees and descriptors
   - The histograms of each descriptor are concatenated
3. Boosting ensemble of decision trees classifies the concatenated vector

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

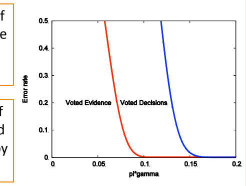
MATHEMATICAL MODEL

Proposition 1. The error rate ε_{vd} (for "voted decisions") of classifying each SIFT vector separately and then taking the majority vote is bounded by

$$\varepsilon_{vd} \le exp[-2d\,(\pi\,\gamma\,(1-2\,\varepsilon\,))^2].$$

Proposition 2. The error rate ε_{vd} (for "voted evidence") of accumulating the leaf histograms for each SIFT vector and then taking the class with the highest count is bounded by

$$\varepsilon_{ve} \le exp[\,-8d\,C(\,\gamma\,\pi)^4].$$

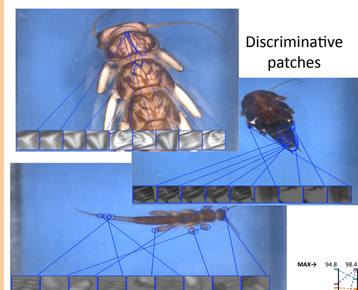


RESULTS

STONEFLY9

- Edge + {Kadir+Hessian Affine+PCBR} x {SIFT} ➔ 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

Discriminative patches



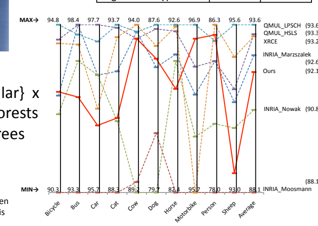
Classification error

Detector	Stacked R. Forest	Visual Dict.
Hessian	15.5±2.8	34.0±3.1
Kadir	11.1±1.1	23.9±1.3
PCBR	11.2±1.8	28.1±1.0
Edges	36.3±1.1	-
Hessian + Edges	11.4±1.9	-
Kadir + Edges	9.8±0.4	-
PCBR + Edges	9.2±1.8	-
Hessian + Kadir	8.1±1.5	19.0±2.5
Kadir + PCBR	7.8±1.6	19.2±1.7
PCBR + Hesaff	7.8±2.3	20.2±2.6
All keypoints	6.4±1.8	16.1±1.8
Edges + all keypoints	5.6±2.1	-

PASCAL 2006

- {Harris Aff.+Hessian Aff.+PCBR+Regular} x {SIFT, color SIFT, filt bank} ➔ 12 rand forests
- Stacking: Boosting of 200 decision trees

AUC:



Published, top 6 methods and ours. Max/min AUC values have been rescaled separately for each task. Average AUC is shown in parenthesis

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.

IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami Beach, FL, U.S.A. June, 2009