Random Subwindows for Robust Image Classification

Raphaël Marée, Pierre Geurts, Justus Piater, Louis Wehenkel

Institut Montefiore, University of Liège, Belgium

CVPR05, 22th June 2005
Image classification

Given a training set of $N$ labelled images (i.e. each image is associated with a class), build a model to predict the class of new images.

Challenges

- To avoid manual adaptation to specific task
- To be able to discriminate between a lot of classes
- To be robust to uncontrolled conditions
  - Illumination/scale/viewpoint/orientation changes
  - Partial occlusions, cluttered backgrounds
  - ...
Approaches

- General scheme [MO04]
  - Detection of “interesting” regions in images [MTS05]
    - Harris, Hessian, MSER, edge-based, local variance, …

- Description by feature vectors [MS05]
  - SIFT, PCA, DCT, moment invariants, …

- Matching of feature vectors
  - Nearest neighbor with Euclidian, Mahalanobis distance, …
Approaches

- General scheme [MO04]
  - Detection of “interesting” regions in images [MTS+05]
    - Harris, Hessian, MSER, edge-based, local variance, ...
    - Random extraction of square patches
  - Description by feature vectors [MS05]
    - SIFT, PCA, DCT, moment invariants, ...
    - Pixel-based normalized representation
- Matching of feature vectors
  - Nearest neighbor with Euclidian, Mahalanobis distance, ...
  - Recent machine learning algorithms able to handle high-dimensional data, e.g.: Ensemble of Decision Trees, SVMs
Detector: Random Subwindows

- Extract Subwindows of random sizes, at random locations
Descriptor: 16x16 Hue-Saturation-Value

- Resize each subwindow to $16 \times 16$
- Describe each subwindow by its 768 pixel values (in HSV)
Learning: subwindow classification model

- Extract $N_w (>> N)$ subwindows from training images
  - Random detector, 16x16 HSV descriptor
  - Label each subwindow with the class of its parent image

- Build a subwindow classification model by supervised learning
Learning: Extra-Trees [Geu02, GEW05]

- Ensemble of $T$ decision trees, generated independently
- Top-down growing by recursive partitioning
  - Internal test nodes compare a pixel-location-channel to a threshold ($a_i < v_i$), terminal nodes output class probability estimates
  - Choice of internal tests at random
  - Fully developed (perfect fit on LS)
Recognition: aggregation of subwindows and tree votes
Experiments

- Standard classification datasets (4 in the paper + 4)
  - Multi-class (up to 201 classes)
  - Illumination/scale/viewpoint changes, partial occlusions, cluttered backgrounds

- Standard protocols
  - Independent test set or leave-one-out validation
  - Directly comparable to other results in the literature

- Parameters
  - Number of learning subwindows: $N_w = 120000$ (total)
  - Number of trees built: $T = 10$
  - Number of test subwindows: $N_{w,test} = 100$ (per image)
Datasets: COIL-100 [MN95] (100 classes)
Datasets: ETH-80 [LS03] (8 classes)
Datasets: ZuBuD [SSV03] (201 classes)
Datasets: WANG [CW04] (10 classes)
Datasets: MNIST [LBBH98] (10 classes)
Datasets: AR Expression Variant Faces [MB98] (100 classes)
Datasets: TSG-20 [FSPB05] (20 classes)
Datasets: IRMA [LGD$^+_{05}$] [iCS05] (57 classes)

(ImageCLEF 2005 [iCS05])

(courtesy of TM Lehmann, Dept. of Medical Informatics, RWTH Aachen, Germany)
## Results: Misclassification error rates

<table>
<thead>
<tr>
<th>DB</th>
<th>ls/ts</th>
<th>class</th>
<th>us</th>
<th>worst</th>
<th>best</th>
</tr>
</thead>
<tbody>
<tr>
<td>COIL-100</td>
<td>1800/5400</td>
<td>100</td>
<td>0.50%</td>
<td>12.50%</td>
<td>0.10% [MO04]</td>
</tr>
<tr>
<td>COIL-100</td>
<td>100/7100</td>
<td>100</td>
<td>13.58%</td>
<td>50%</td>
<td>24% [MO04]</td>
</tr>
<tr>
<td>ZuBuD</td>
<td>1005/115</td>
<td>201</td>
<td>4.35%</td>
<td>59%</td>
<td>0% [MO04]</td>
</tr>
<tr>
<td>ETH-80</td>
<td>3280/3280</td>
<td>8</td>
<td>25.49%</td>
<td>35.15%</td>
<td>13.60% [LS03]</td>
</tr>
<tr>
<td>WANG</td>
<td>1000/1000</td>
<td>10</td>
<td>15.90%</td>
<td>62.5%</td>
<td>15.90% [DKN04a]</td>
</tr>
<tr>
<td>MNIST</td>
<td>60000/10000</td>
<td>10</td>
<td>2.13%</td>
<td>12%</td>
<td>0.50% [DKN04b]</td>
</tr>
<tr>
<td>AR EVF</td>
<td>100/600</td>
<td>100</td>
<td>15.83%</td>
<td>29.83%</td>
<td>12% [TCZ⁺ 05]</td>
</tr>
<tr>
<td>TSG-20</td>
<td>40/40</td>
<td>20</td>
<td>5.0%</td>
<td>2.5%</td>
<td>0% [FSPB05]</td>
</tr>
<tr>
<td>IRMA</td>
<td>9000/1000</td>
<td>57</td>
<td>14.7%</td>
<td>73.3%</td>
<td>12.6% [iCS05]</td>
</tr>
</tbody>
</table>
COIL-100: robustness to viewpoint changes

- COIL-100: error rate depending on azimuthal test angle, learning only from the frontal view ($0^\circ$).
Some observations: subwindow classification

correct:

misclassified:
Robustness to orientation changes

C1

C2

C3

C1 C1 C1 C1 C1 C2 C2 C2 C2 C2 C3 C3 C3 C3 C3

Marée et al.

Random Subwindows + Extra-Trees
Why does it work?

- Random Subwindows
  - Aggregation of a large amount of information
    - Use both local, global, (un)homogeneous regions, ...
  - Pixel-based normalized representation
    - Normalization to a fixed size
    - HSV limits the effect of illumination changes
  - Tolerance to partial occlusions and cluttered backgrounds

- Extra-trees
  - Accurate even with high-dimensional data (variance reduction)
Summary

- Novel image classification method that...
  - combines Random Subwindows and Extra-Trees
  - yields quite good results on a variety of tasks

- could be quickly evaluated on new classification problems
  - few parameters (the more trees/subwindows, the better)
  - fast learning (± 6m30s on ZuBuD)
  - fast classification (tree depth ± 18.26 on ZuBuD)

- is now implemented in Java:
  http://www.montefiore.ulg.ac.be/~maree/
Extensions and Future Work

- **Method**
  - Comparison with other detectors and other descriptors
  - Comparison with other machine learning algorithms
    - CART, Bagging, Boosting, Random Forests: [MGPW05]
    - KNN, SVM
  - Filtering Subwindows for heavily cluttered backgrounds?

- **Evaluation**
  - ALOI, Butterflies, Birds, Caltech 101, NORB, . . . , ?
  - Ongoing real-world applications: metal powders, marbles, flowers, license plates, . . .
Acknowledgments

- Raphaël Marée is supported by GIGA-Interdisciplinary Cluster for Applied Genoproteomics, hosted by the University of Liège
- Pierre Geurts is a Postdoctoral Researcher at the National Fund for Scientific Research (FNRS, Belgium)
- IRMA database courtesy of TM Lehmann, Dept. of Medical Informatics, RWTH Aachen, Germany
- PEPITe for the release of PiXiT, a Java implementation of the method, available for evaluation purpose at: http://www.montefiore.ulg.ac.be/~maree/
Image categorization by learning and reasoning with regions.

T. Deselaers, D. Keysers, and H. Ney.
Features for image retrieval: A quantitative comparison.

T. Deselaers, D. Keysers, and H. Ney.
Classification error rate for quantitative evaluation of content-based image retrieval systems.

Learning informative sift descriptors for attentive object recognition.

P. Geurts.
*Contributions to decision tree induction: bias/variance tradeoff and time series classification.*

P. Geurts, D. Ernst, and L. Wehenkel.
Extremely randomized trees.

S. L. N. in Computer Science, editor.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner.
Gradient-based learning applied to document recognition.
Automatic categorization of medical images for content-based retrieval and data mining.

B. Leibe and B. Schiele.
Analyzing appearance and contour based methods for object categorization.

A. Martinez and R. Benavente.
The ar face database.
Technical report, School of Electrical & Computer Engineering Purdue University, West Lafayette, Indiana, 1998.

Decision trees and random subwindows for object recognition.

Visual learning and recognition of 3d objects from appearance.

J. Matas and S. Obdržálek.
Object recognition methods based on transformation covariant features.

A performance evaluation of local descriptors.
A comparison of affine region detectors.

Zubud - Zurich building database for image based recognition.

Recognizing partially occluded, expression variant faces from single training image per person with som and soft knn ensemble.
*IEEE Transactions on Neural Networks, 2005.*