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
    - ...

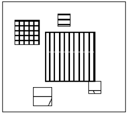
Approaches

- General scheme [MO04]
  - Detection of "interesting" regions in images [MTS+05]
    - 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

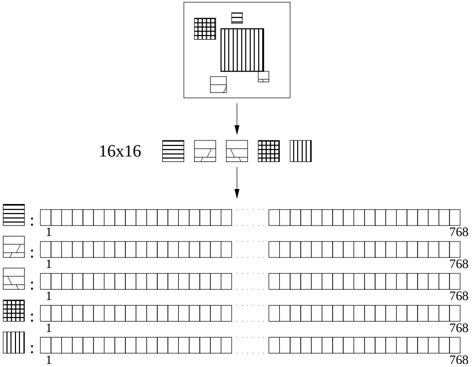
- General scheme [MO04]
  - Detection of "interesting" regions in images [MTS+05]
    - Harris, Hessian, MSER, edge-based, local variance, ...
    - Random extraction of square patches
  - Description by feature vectors [MS05]
    - SIFT, PCA, DCT, moment invariants, ...
    - Pixel-based normalized 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

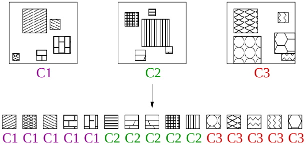
Descriptor: 16x16 Hue-Saturation-Value



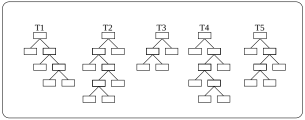
- Resize each subwindow to 16×16
- Describe each subwindow by its 768 pixel values (in HSV)

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

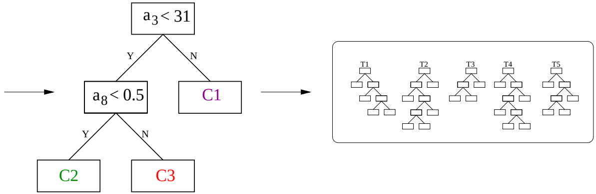


- Build a subwindow classification model by supervised learning



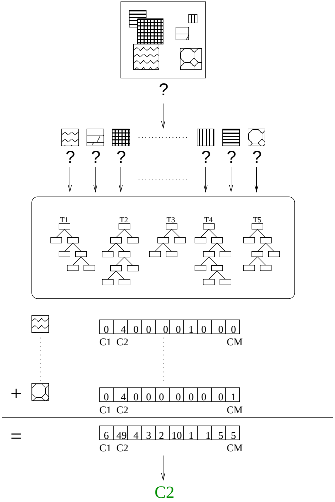
Learning: Extra-Trees [Geu02, GEW05]

a_1	a_2	a_3	a_4	......	a_{766}	a_{767}	a_{768}	Class
60	0.1	99	17	......	1	0.23	164	C1
60	0.37	113	23	......	29	0.07	230	C1
75	0.03	210	1	......	77	0.05	255	C1
2	0.1	97	2	......	0	0.23	88	C2
3	0.2	180	18	......	0	0.12	145	C2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	0.06	55	10	......	10	0.54	100	C3



- 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

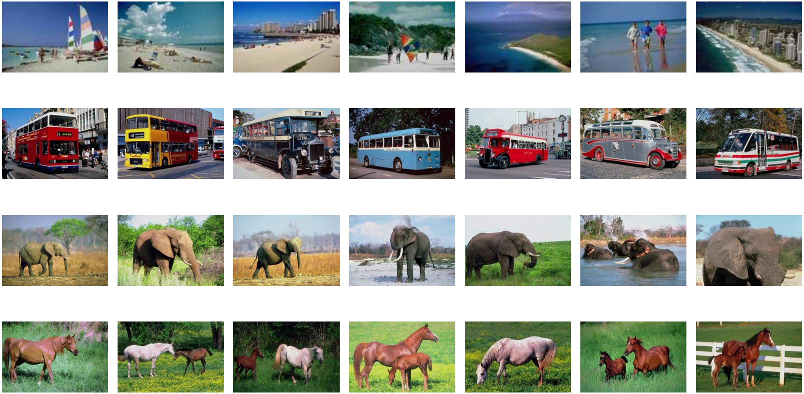
- Standard classification datasets (4 in the paper + 4)
  - Multi-class (up to 201 classes)
  - Illumination/scale/viewpoint changes, partial occlusions, cluttered backgrounds
- Standard protocols
  - Independent test set or leave-one-out validation
  - Directly comparable to other results in the literature
- 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)

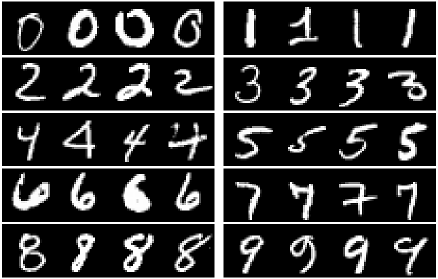
Datasets: ETH-80 [LS03] (8 classes)

Datasets: ZuBuD [SSV03] (201 classes)

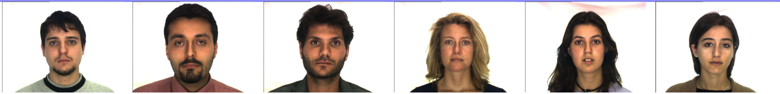
Datasets: WANG [CW04] (10 classes)



Datasets: MNIST [LBBH98] (10 classes)



Datasets: AR Expression Variant Faces [MB98] (100 classes)



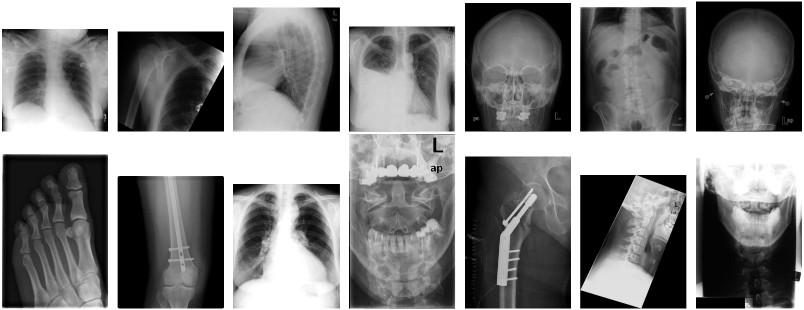
Learning:

Session 1:

Session 2:

Datasets: TSG-20 [FSPB05] (20 classes)

Datasets: IRMA [LGD+05] [iCS05] (57 classes)



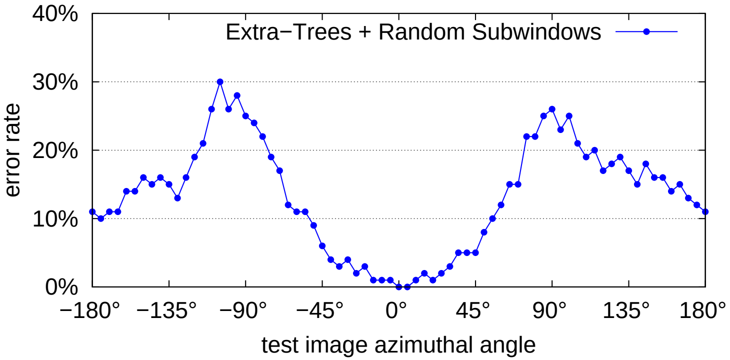
(ImageCLEF 2005 [iCS05])

(courtesy of TM Lehmann, Dept. of Medical Informatics, RWTH Aachen, Germany)

Results: Misclassification error rates

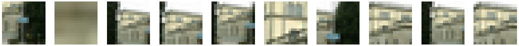
DB	ls/ts	class	us	worst	best
COIL-100	1800/5400	100	0.50%	12.50%	0.10% [MO04]
COIL-100	100/7100	100	13.58%	50%	24% [MO04]
ZuBuD	1005/115	201	4.35%	59%	0% [MO04]
ETH-80	3280/3280	8	25.49%	35.15%	13.60% [LS03]
WANG	1000/1000	10	15.90%	62.5%	15.90% [DKN04a]
MNIST	60000/10000	10	2.13%	12%	0.50% [DKN04b]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

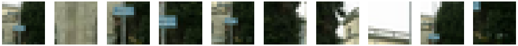
COIL-100: robustness to viewpoint changes



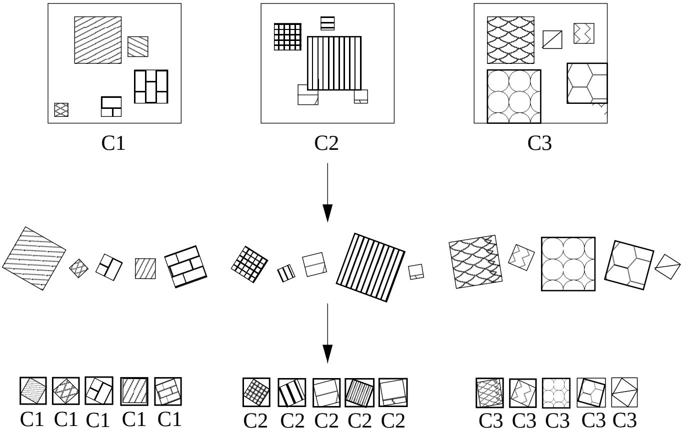
- COIL-100: error rate depending on azimuthal test angle, learning only from the frontal view (0°).

Some observations: subwindow classification

correct: 

misclassified: 

Robustness to orientation changes

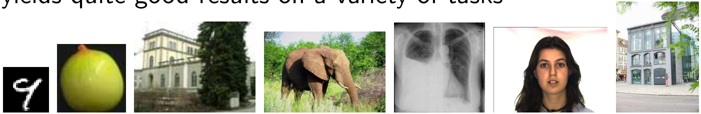


Why does it work?

- 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 ($\pm$ 6m30s on ZuBuD)
    - fast classification (tree depth $\pm$ 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/`

Y. Chen and J. Z. Wang.
Image categorization by learning and reasoning with regions.
Journal of Machine Learning Research, 5:913–939, August 2004.

T. Deselaers, D. Keysers, and H. Ney.
Features for image retrieval: A quantitative comparison.
In *Proc. 26th DAGM Symposium on Pattern Recognition (DAGM 2004)*, volume LNCS 3175, pages 228–236, August/September 2004.

T. Deselaers, D. Keysers, and H. Ney.
Classification error rate for quantitative evaluation of content-based image retrieval systems.
In *Proc. 17th International Conference on Pattern Recognition (ICPR)*, 2004.

G. Fritz, C. Seifert, L. Paletta, and H. Bischof.
Learning informative sift descriptors for attentive object recognition.
In *Proc. 1st Austrian Cognitive Vision Workshop (ACVW 2005)*, 2005.

P. Geurts.
Contributions to decision tree induction: bias/variance tradeoff and time series classification.
PhD thesis, Department of Electrical Engineering and Computer Science, University of Liège, May 2002.

P. Geurts, D. Ernst, and L. Wehenkel.
Extremely randomized trees.
To appear in Machine Learning Journal, 2005.

S. L. N. in Computer Science, editor.
Proc. of Cross Language Evaluation Forum (CLEF), to appear, 2005.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner.
Gradient-based learning applied to document recognition.
Proc. of the IEEE, 86(11):2278–2324, 1998.

T. Lehmann, M. Güld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, H. Ney, and B. Wein.
Automatic categorization of medical images for content-based retrieval and data mining.
Computerized Medical Imaging and Graphics, 29(2):143–155, 2005.

B. Leibe and B. Schiele.
Analyzing appearance and contour based methods for object categorization.
In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Madison, WI, June 2003.

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

R. Marée, P. Geurts, J. Piater, and L. Wehenkel.
Decision trees and random subwindows for object recognition.
In *ICML workshop on Machine Learning Techniques for Processing Multimedia Content (MLMM2005)*, 2005.

H. Murase and S. K. Nayar.
Visual learning and recognition of 3d objects from appearance.
International Journal of Computer Vision, 14(1):5–24, 1995.

J. Matas and S. Obdržálek.
Object recognition methods based on transformation covariant features.
In *Proc. 12th European Signal Processing Conference (EUSIPCO 2004)*, Vienna, Austria, September 2004.

K. Mikolajczyk and C. Schmid.
A performance evaluation of local descriptors.
PAMI, to appear, 2005.

K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool.
A comparison of affine region detectors.
International Journal of Computer Vision, to appear, 2005.

H. Shao, T. Svoboda, and L. Van Gool.
Zubud - Zurich building database for image based recognition.
Technical Report TR-260, Computer Vision Lab, Swiss Federal Institute of Technology, Switzerland, 2003.

X. Tan, S. Chen, Z.-H. Zhou, , and F. Zhang.
Recognizing partially occluded, expression variant faces from single training image per person with som and soft knn ensemble.
IEEE Transactions on Neural Networks, 2005.