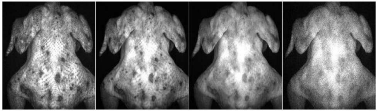
A Machine Learning Approach for Material Detection in Hyperspectral Images

Raphaël Marée[1], Benjamin Stévens[2], Pierre Geurts[1], Yves Guern[3], Philippe Mack[2]

[1]: GIGA, Dept. EE & CS, University of Liège, Belgium
[2]: PEPITe S.A., Liège, Belgium
[3]: ATIS S.A., Aix-en-Provence, France

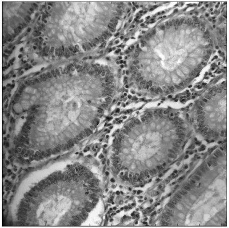
6th IEEE Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum , 20^{th} June 2009, Miami

Hyperspectral image analysis: applications



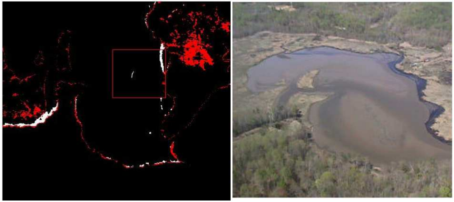
Poultry skin tumor detection

(Myongji University, KR)



Colon cancer classification

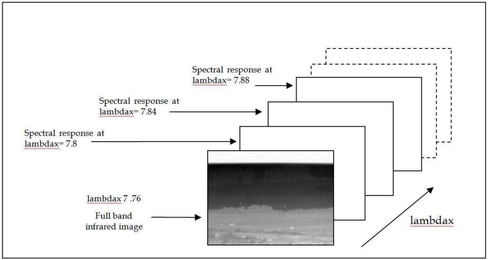
(University of Warwick, UK)

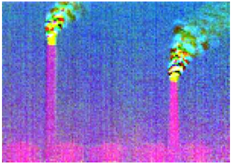


Monitoring of oil spills

(George Mason University, USA)

Hawkeye project (EU-FP6)



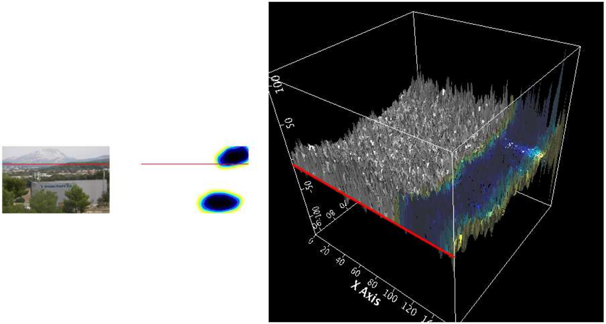


Goal: Detection of gaseous traces in still images of real-world scenes

Applications: chemical accident detection, surveillance of industries, long-term pollution follow-up, . . .

Sensor: thermal infrared domain (8 to 12 microns)

Data analysis challenges



- High-dimensional
- Various acquisition conditions, noisy
- Multiple types of materials to detect

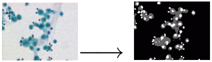
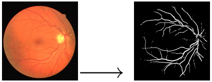
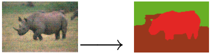
General approach

The proposed method is generic and relies on

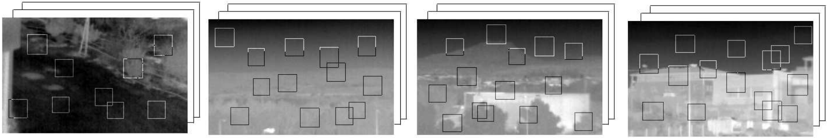
- raw hyperspectral data (no explicit pre-processing)
- both spatial and spectral information (dense sampling of local subcubes)
- supervised learning (ensembles of extremely randomized trees with multiple outputs)

Related work

- *Extremely Randomized Trees*
  Geurts, P., Ernst, D., Wehenkel, L., Machine Learning, 2006
- *Random Subwindows for Robust Image Classification*
  Marée, R., Geurts, P., Piater, J., Wehenkel, L., CVPR, 2005
- *Fast Multi-Class Image Annotation with Random Subwindows and Multiple Output Randomized Trees*
  Dumont, M., Marée, R., Wehenkel, L., Geurts, P., VISAPP, 2009

Generating a training set of random subcubes

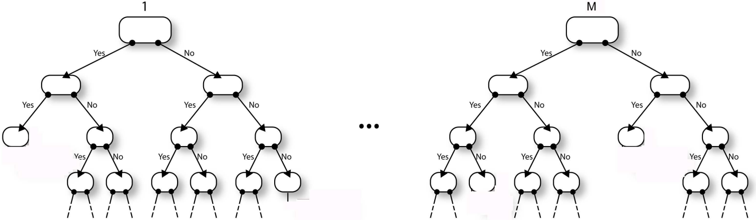


- From all training hypercubes, extract a total of N subcubes of size $w \times h$ at random locations in the x, y position plane
- Each subcube is
  - described by spectral information for each position *ie.* $w \times h \times l$ raw hyperspectral values (input features)
  - annotated with the classes of each of its $w \times h$ pixels (outputs). A class is one of the gas type or background

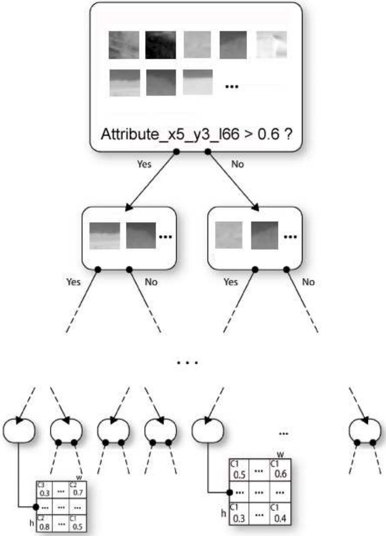
Training a subcube annotation model

Using the training set of subcubes, we build a subcube annotation model

- It is based on an extension of extremely randomized trees to deal with multiple outputs
- Each extra-tree is learnt to map subcube numerical features (raw hyperspectral values) to discrete outputs (a class for each pixel)



Learning an extra-tree with multiple outputs



Node splitting algorithm for multiple outputs

Split_a_node(S)

Input: the local learning subset S corresponding to the node we want to split
Output: a split $[a < a_c]$ or leaf
- If **Stop_split**(S) is TRUE then attach $w \times h$ output predictions.
- Otherwise select randomly K attributes $\{a_1, \ldots, a_K\}$ among all non constant (in S) input attributes;
- Draw K splits $\{T_1, \ldots, T_K\}$, where T_t = **Pick_a_random_split**(S, a_t), $\forall t = 1, \ldots, K$;
- Return a split T_t such that MeanScore$(S, T_t) = \max_{j=1,\ldots,K}$ MeanScore(S, T_j)

Pick_a_random_split(S,a)

Inputs: a subset S and an attribute a
Output: a split
- Let $a^S_{\max}$ and $a^S_{\min}$ denote the maximal and minimal value of a in S;
- Draw a random cut-point a_{th} uniformly in $]a^S_{\min}, a^S_{\max}]$;
- Return the split $[a < a_{th}]$.

Stop_split(S)

Input: a subset S
Output: a boolean
- If $|S| < n_{\min}$, then return TRUE;
- If if all attributes are constant in S, then return TRUE;
- If all outputs are constant in S, then return TRUE;
- Otherwise, return FALSE.

Université de Liège

Scoring the K random splits at each node

For each test $T = [a < a_{th}]$ and given the subset S of subcubes, we compute the average of the Gini entropy based score for each output pixel. We select the test that maximizes the score.

$$
\begin{aligned}
MeanScore(S, T) &= \frac{1}{w.h} \sum_{o=1}^{w*h} Score^{o}(S, T) && (1)\\
Score^{o}(S, T) &= G_C^{o}(S) - G_{C|T}^{o}(S) && (2)\\
G_C^{o}(S) &= \sum_{i=1}^{m} \frac{n_{i.o}}{n_{...}}(1 - \frac{n_{i.o}}{n_{...}}) && (3)\\
G_{C|T}^{o} &= \sum_{j=1}^{2} \sum_{i=1}^{m} \frac{n_{ijo}}{n_{...}}(1 - \frac{n_{ijo}}{n_{.j.}}) && (4)
\end{aligned}
$$

T is the split to evaluate, S the subsample of size $n_{...}$ associated to the node to split, m is the number of pixel classes, $n_{i.o}(i = 1, \ldots, m, o = 1, \ldots, wh)$ is the number of subcubes in S of class i at pixel o, n_{i1o} (resp. n_{i2o}) the number of subcubes in S of class i at pixel o and which satisfy (resp. do not satisfy) test T, $n_{.1.}$ (resp. $n_{.2.}$) the total number of subcubes in S which satisfy (resp. do not satisfy) test T.

Embedded feature selection

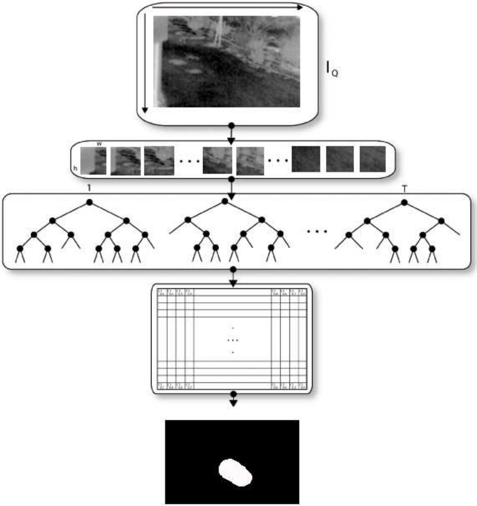
It is possible to compute from the ensemble of trees a ranking of the input features (*i.e.* spectral bands) according to their relevance for predicting the output classes (*i.e.* types of gas).

The overall importance of a feature for class i is computed by summing the $V_i(T)$ values for all test nodes T of the ensemble of trees where this feature is used to split:

$$V_i(T) = \frac{1}{w.h} \sum_{o=1}^{w*h} \{ n_{i.o}(1 - \frac{n_{i.o}}{n_{...}}) - \sum_{j=1}^{2} n_{ijo}(1 - \frac{n_{ijo}}{n_{.j.}}) \}, \quad (5)$$

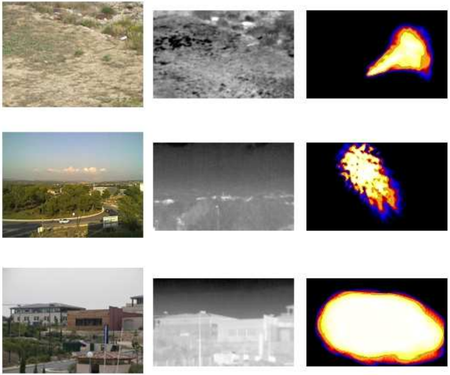
$V_i(T)$ is such that $\sum_{i=1}^{m} V_i(T) = n_{...} Score(S, T)$ with $Score(S, T)$ as defined in (2).

Hypercube annotation by aggregating sliding subcubes



Experimental setup (1/2)

Gaseous traces simulated and incrusted in real-world scenes under different conditions



Experimental setup (2/2)

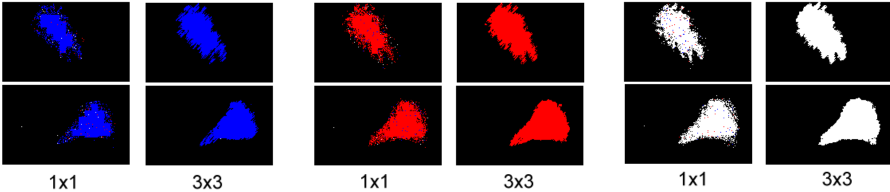
Three gases (SF6, Acetone, Methanol) incrusted in three scenes with various concentrations and thermal contrasts

21 hybercubes for each gas (10, 10, 1) $\longrightarrow$ 63 hypercubes

Hypercubes of dimensions $160 \times 100 \times 94$ (spectral range: $7.76 \mu m$ to $12.98 \mu m$)

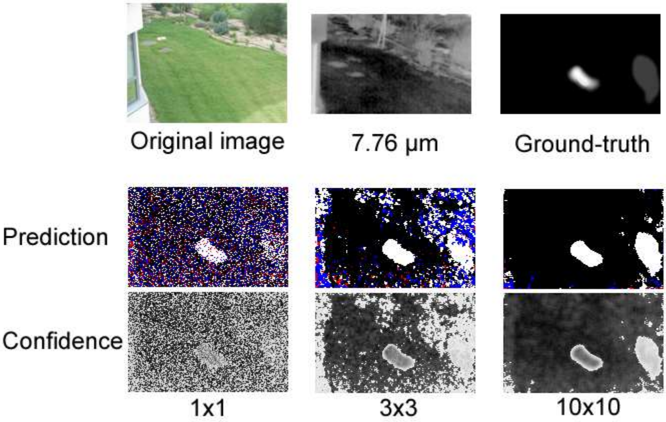
Results: Known Scene, unknown conditions

Multi-class prediction results on two known scenes with unknown (slightly changing) conditions for the three gases:



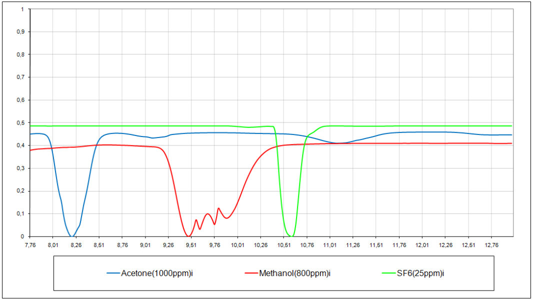
Results: Unknown Scene, unknown conditions

Multi-class prediction results on an unknown scene:



Ranking of spectral bands: SF6

Relevant spectral bands extracted from the set of 63 hypercubes

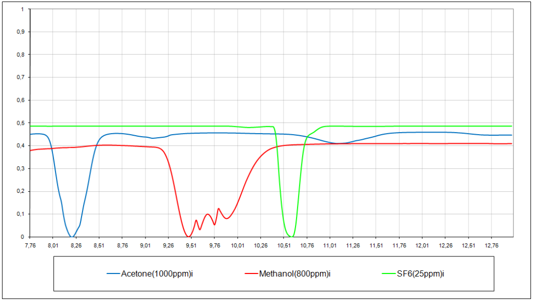


Wavelength	Computed Importance
SF6	
70 (10.56 μm)	16.45%
69 (10.52 μm)	9.82%
71 (10.60 μm)	9.44%
72 (10.64 μm)	8.31%
0 (7.76 μm)	5.21%
68 (10.48 μm)	2.55%
73 (10.68 μm)	1.58%
59 (9.72 μm)	1.12%

Theoretical spectral responses between 7.76 μm and 12.98 μm.

Ranking of spectral bands: Acetone

Relevant spectral bands extracted from the set of 63 hypercubes

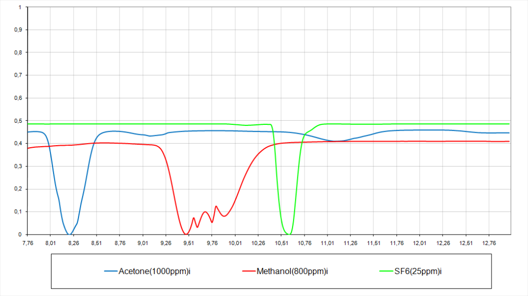


Wavelength	Computed Importance
Acetone	
11 (8.20 μm)	9.08%
12 (8.24 μm)	8.10%
13 (8.28 μm)	5.21%
10 (8.16 μm)	4.16%
0 (7.76 μm)	3.37%
14 (8.32 μm)	2.17%
9 (8.12 μm)	1.90%
44 (9.52 μm)	1.07%
46 (9.60 μm)	1.05%
43 (9.48 μm)	1.05%
71 (10.60 μm)	1.04%
70 (10.56 μm)	1.02%

Theoretical spectral responses between 7.76 μm and 12.98 μm.

Ranking of spectral bands: Methanol

Relevant spectral bands extracted from the set of 63 hypercubes



Theoretical spectral responses between 7.76 μm and 12.98 μm.

Wavelength	Computed Importance
Methanol	
0 (7.76 μm)	6.87%
44 (9.52 μm)	6.35%
42 (9.44 μm)	5.51%
43 (9.48 μm)	5.38%
45 (9.56 μm)	3.98%
46 (9.60 μm)	3.03%
48 (9.68 μm)	2.48%
41 (9.40 μm)	2.34%
47 (9.64 μm)	1.74%
49 (9.72 μm)	1.36%
51 (9.80 μm)	1.17%
50 (9.76 μm)	1.11%
52 (9.84 μm)	1.11%
54 (9.92 μm)	1.02%
53 (9.88 μm)	1.01%

Université de Liège

Influence of parameters

We observed similar trends than in our previous works in 2D image classification and segmentation

- More subcubes yields better detection results
- More trees yields better detection results, but 10 trees is already good
- Higher values of K allow to better filter out irrelevant attributes, and reduce the size of models (number of nodes)

Computational requirements

- Training
  - $\mathcal{O}(MwhlKN\log_2(N))$
  - Training time was less than 9 minutes with a database of $63 \times 5000 \times 3 \times 3 \times 94 = 266,490,000$ floating point values
  - The model had about 500000 nodes.
- Prediction
  - $\mathcal{O}(M\log_2 N)$ tests for each of the $(w_i - w + 1)(h_i - h + 1)$ subwindows
  - Less than 5 seconds to load and predict a new hypercube
- Implementation
  - Implemented in JAVA and runned on a single 2.4Ghz CPU
  - The method is highly parallelizable

Summary and future work

- A generic machine-learning approach for annotation of hyperspectral images
  - Uses dense spatial and spectral information
  - Without explicit pre-processing/dimensionality reduction
  - Provides embedded spectral band selection
  - It is conceptually rather simple and fast

- Preliminary results on thermal infrared images are promising but need further validation and comparison with other works
- The approach might also be useful in other applications with different sensors/imaging spectrometers
- Extension to multiple numerical outputs (regression) for prediction of material "density"

Université de Liège

Acknowledgments

- European Sixth Framework Programme and Regional Development Fund
- Walloon Region
- Belgian National Fund for Scientific Research
- Pepite