Interactive Retrieval and Mining of Visual Content

Mihail Crucianu

http://cedric.cnam.fr/vertigo/

Outline

- Why interactive retrieval and mining?
- Active semi-supervised clustering
- Relevance feedback with global or local features
- Scalability issues for relevance feedback
- Scalable video mining by copy detection
- Interactive retrieval after prior mining
- Information retrieval beyond ranking

Why interactive retrieval / mining

- Define/find task-dependent or user-dependent complex visual concepts/patterns
- Available information
  1. Data-issued similarities (visual, spatial relations…)
     - Inherent to the data!
  2. User-provided valuable information
     - Class labels, pairwise constraints, …
     - Typically approximate and/or uncertain
     - Typically scarce and expensive

  → Combine these information sources!

23/09/2010 M. Crucianu 3

Active semi-supervised clustering

- Supervision (information regarding the target: class labels, constraints, …) is only available for (a small) part of the data, while data-issued similarities are available for all the data

→ semi-supervised learning

- High cost of
  - Acquiring supervision (requires interaction with the user)
  - Using the data (algorithmic complexity)

→ active learning: the algorithm selects the data for which supervision information is required from the user

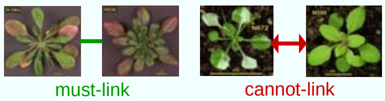
→ Maximal performance improvement at minimal cost

23/09/2010 M. Crucianu 4

Case: similarities + constraints

- Application context (image database categorization)
  - Large unknown (or little known) database
  - Direct clustering has poor performance
    - → Supervision is needed
  - Image classes are unknown *a priori*
    - → Users cannot provide class label but can say whether 2 images should be in a same class (*must-link* constraint) or in different classes (*cannot-link* constraint)
    - → Given the size of the database, the amount of supervision should be minimal

© http://arabidopsis.info

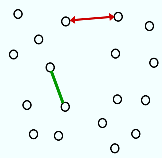


23/09/2010 M. Crucianu 5

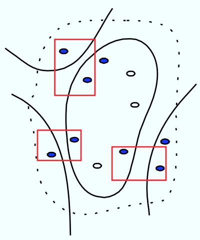
Cédric

Active clustering with constraints

- Semi-supervized aspect
  - Combine two sources of information:
    1. Similarities between image descriptors
    2. Pairwise constraints
  - → New objective function (based on CA)



- Active aspect
  - Minimize number of needed constraints ⇐ maximize information transfer user → system
  - → 2 complementary selection criteria:
    1. Informative constraints: ambiguous images from the least well defined clusters
    2. Low redundancy between the constraints



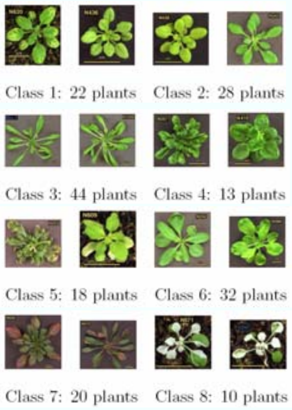
[GCB08]

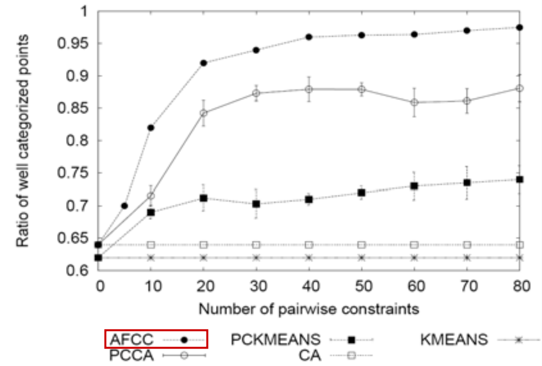
23/09/2010 M. Crucianu 6

Cédric

Illustrative results

- *Arabidopsis thaliana* database



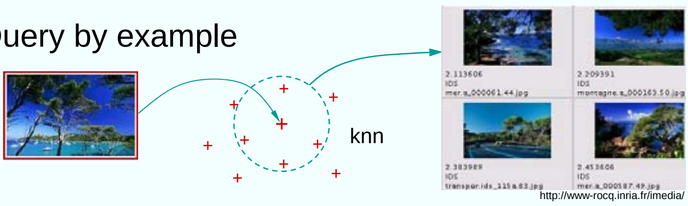


[GCB08]

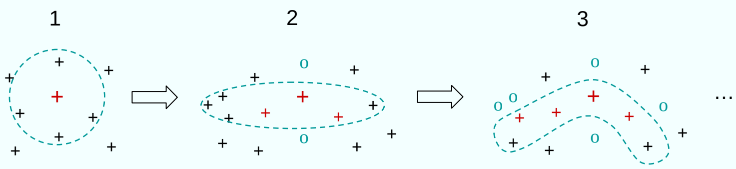
Images provided by NASC (http://arabidopsis.info), ground truth by INRA (http://www.inra.fr)

Retrieval with relevance feedback

1. Query by example



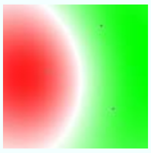
2. Iterative interactive retrieval with relevance feedback



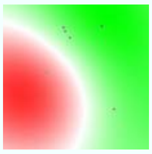
Active learning for RF

- System must select unlabeled sample so as to maximize the transfer of information from the user to the system

→ Ambiguousness

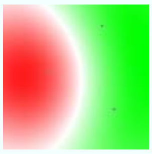


Before selection

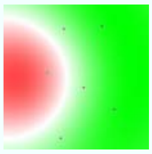


After selection, feedback, estimation

→ Low redundancy



Before selection



After selection, feedback, estimation

23/09/2010 M. Crucianu 9 INRIA

Illustrative example 1

Goal: find portraits

Base of 7500 images, including 110 portraits

Available description: global (colour, texture, shape)

First page after 4 iterations →

[FCB04/1]

23/09/2010 M. Crucianu 10 INRIA

Illustrative example 2

Goal: find regions representing villages

Base of 24000 regions, 87 belong to the class

Available description: region features (colour, texture, shapes inside)

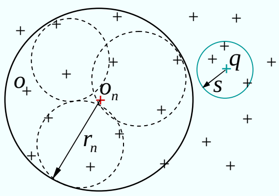
First page after 6 positive examples and 28 negative

[FCB04/2]

23/09/2010 M. Crucianu 11

Does relevance feedback scale?

- Two stages
  1. Learning: typically very few examples, can be fast if a fast learner is employed (e.g. SVM); still, is expensive with transductive learning!
  2. Selection of unlabeled sample the user should label: if all the unlabeled items in the database are evaluated, complexity O(N)!
- Many proposals for scaling retrieval by similarity (query by example), most of them fail when distance distribution is narrow
- An example: M-tree (metric search tree)

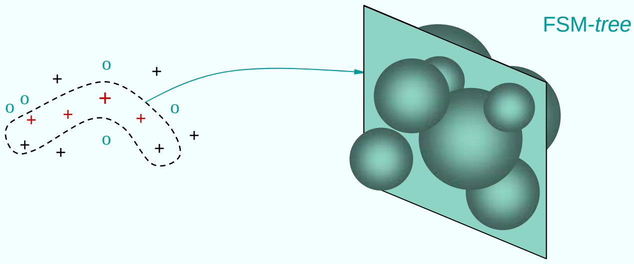


Subtree $N(o_n)$: objects o such that $d(o,o_n) \le r_n$
Query q of range s
If $d(q,o_n) > r_n + s$ then the entire subtree $N(o_n)$ can be rejected (from triangular inequality)

23/09/2010 M. Crucianu 12

FSM-tree and hyperplane queries

- What about relevance feedback?
- Principle of the method in [CEOT08]:
  - Classification with a 2-class SVM
  - Build an M-tree in the feature space of the kernel (FSM-*tree*)
  - Return the images that are closest to the boundary, found by a knn query in the FSM-tree with a hyperplane

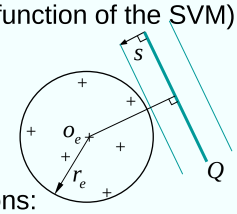


23/09/2010 M. Crucianu 13

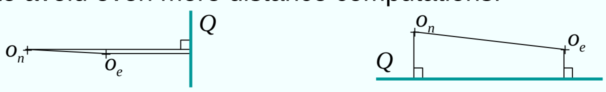
FSM-tree and hyperplane queries

- Distance to the hyperplane: $d(Q,o_e)=\dfrac{\left|\sum_i \alpha_i\, y_i\, K(o_e,x_i)+b\right|}{\sqrt{\sum_{i,j}\alpha_i\,\alpha_j\, y_i\, y_j\, K(x_i,x_j)}}$

  ($f(o_e)=\sum_i \alpha_i\, y_i\, K(o_e,x_i)+b$ being the decision function of the SVM)
- Pruning principle (test to reject a subtree):
  - If $d(Q,o_e)>r_e+s$ then the node is not retained for subsequent exploration



- How to avoid even more distance computations:



$d(Q,o_e)+d(o_e,o_n)\geq d(Q,o_n)$ but $d(Q,o_e)+d(Q,o_n)\ngeq d(o_e,o_n)$

$\Rightarrow d(Q,o_e)\geq d(Q,o_n)-d(o_e,o_n)$ ($d(Q,o_e)\ngeq |d(Q,o_n)-d(o_e,o_n)|$), so

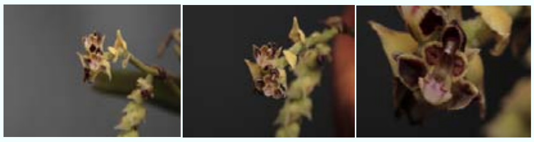
if $d(Q,o_n)-d(o_e,o_n)>r_e+s$ then $d(Q,o_e)>r_e+s$

23/09/2010 M. Crucianu 14

RF with local features

- Application context: assisted plant species identification
  - Relevant information only concerns a part of the image
  - Additional difficulties:
    - Strong background variations

    - Scale and viewpoint variation



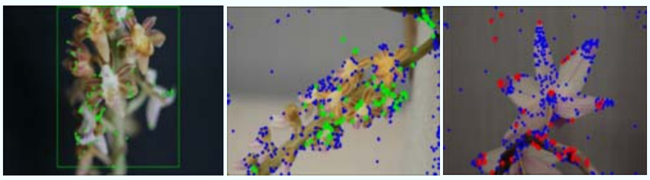
All plant images are provided by INRA

RF with local features

→ Use of local features (LF) with appropriate invariance and robustness characteristics

- User selects an image region to label as “relevant” or “irrelevant”
- Which part should the system consider in unlabeled images?
- Which unlabeled images should be considered? (all: too slow!)

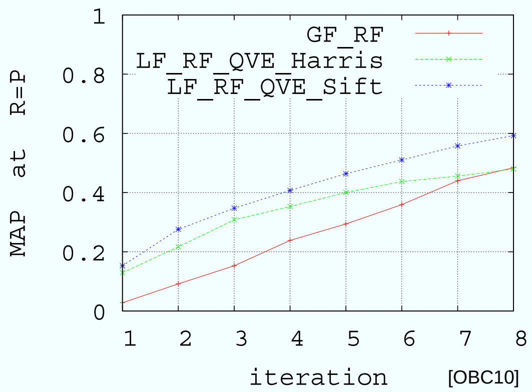
→ Add implicit feature selection to the user-provided selection



User target (left) and two candidate images with LF belonging to the target (green, middle) or not (red, right); the other LF (blue) are ignored

Retrieval performance

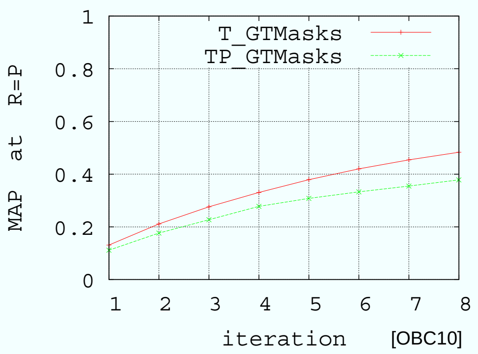
- Comparison between two types of local features and global features (context can play a significant role for some databases…)

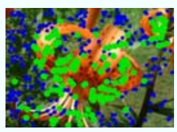


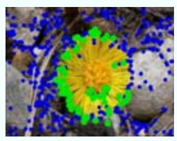
23/09/2010 M. Crucianu 17

Localization performance

- Implicit object localization (by LF similarity) is close to explicit localization (prior segmentation)







Images from Oxford Flower 17 database

23/09/2010 M. Crucianu 18

Scalable RF with local features

- Kernel approximation by hashing $k_h(\mathbf{x},\mathbf{y}) \approx k(\mathbf{x},\mathbf{y})$

$$k_h(\mathbf{x},\mathbf{y}) = \langle \phi_h(\mathbf{x}), \phi_h(\mathbf{y}) \rangle \qquad \phi_h(\mathbf{x}) = [\underbrace{\underbrace{\underbrace{0\ldots0}_{s-1}\ o_1}_{h_1(\mathbf{x})=s}\ 0\ldots0}_{\|\Omega\|}\ \ldots\ \underbrace{\underbrace{\underbrace{0\ldots0}_{t-1}\ o_n}_{h_n(\mathbf{x})=t}\ 0\ldots0}_{\|\Omega\|}]^T$$

- Hashing and active learning

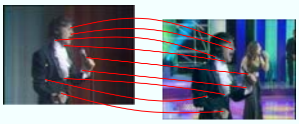
$$\begin{aligned} f_h(\mathbf{x}) &= \sum_{j=1}^{s} \alpha_j y_j k_h(\mathbf{x}, \mathbf{v}_j) - \beta \\ &= \sum_{j=1}^{s} \alpha_j y_j \langle \phi_h(\mathbf{x}), \phi_h(\mathbf{v}_j) \rangle - \beta \\ &= \left\langle \phi_h(\mathbf{x}),\ \sum_{j=1}^{s} \alpha_j y_j \phi_h(\mathbf{v}_j) \right\rangle - \beta \end{aligned}$$

23/09/2010 M. Crucianu 19

Cédric

Video mining by copy detection

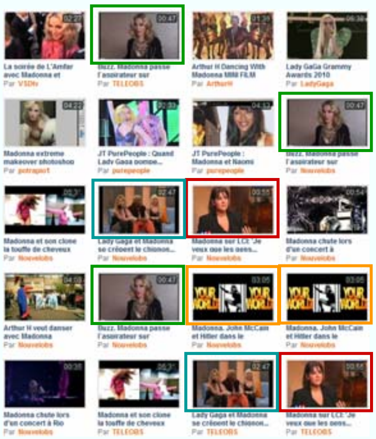
- Goal: scalable content-based video copy detection for video stream surveillance and mining of large video databases
- "Copy" = transformed version of original content (photometric, geometric, temporal changes, post-production…)



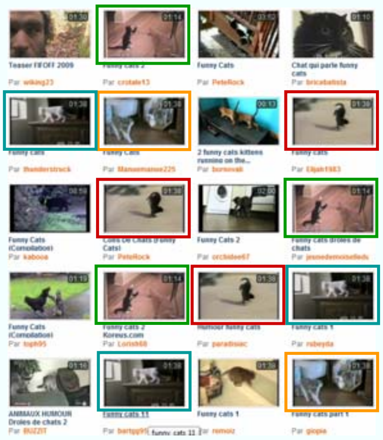
23/09/2010 M. Crucianu 20

Cédric

Video mining by copy detection



Source: DailyMotion



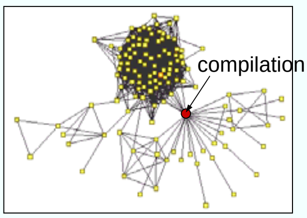
Source: DailyMotion

23/09/2010 M. Crucianu 21

Cédric

Video mining by copy detection

- Large content archive (e.g. Institut National de l'Audiovisuel) context
  - Segmentation, labelling
  - Aided annotation
  - Media impact analysis
  - Media offer analysis
- Video sharing web site context
  - Cleanup (remove/reduce redundancy)
  - Organize: select representatives, identify characteristics
  - Mutualisation/filtering of keywords
  - New tools for navigation/visualisation



23/09/2010 M. Crucianu 22

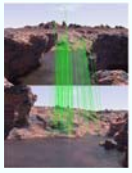
Cédric

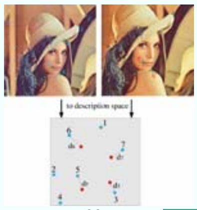
Description for copy detection

- Global description?
  - Not robust to expected transformations

→ Local description
  - SIFT, PCA-SIFT, GLOH, SURF?
    - Too robust to changes in scale, viewpoint…
    - Expensive (extraction, retrieval)

  → Improved Harris detector with differential descriptor
    - Robust to changes in contract, limited robustness to changes in scale
    - Quite light (dimension = 20)





Scalable mining by copy detection

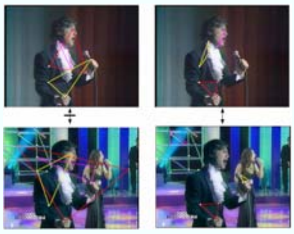
- Preliminary work: stream surveillance
  - Z-grid, models of distortions and of the local density of signatures

  → Deferred real-time surveillance of one video stream against a database of 280 000 hours with 1 PC, detection rate ~95% for video fragments of more than 5 seconds
- Mining by copy detection = similarity self-join on the video database → complexity O(N^2) (without an index)
- Alternative solutions:
  1. Direct use of the stream surveillance method (sequential reading the database, search of each keyframe in the database)
  2. Direct mining:
     - Similarity-based segmentation (redundant segments!) of the database
     - Similarity self-join within each segment

Scalable mining by copy detection

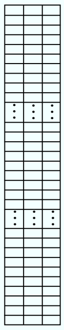
- Quantization of the description space in cells, each hash bucket is defined by a triple of cells [PCS09]
- Triples of points defined by knn in the image (→ local)
- Basic geometric information

$$\rho = \frac{d\left(q, 1^{st}\ \mathrm{nn}\right)}{d\left(q, 2^{nd}\ \mathrm{nn}\right)}$$



[PCS09]

Bucket sorted by ρ



Examples of detections

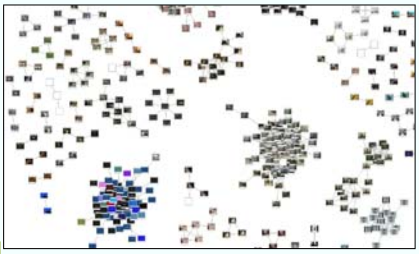
Transformed copy | Original

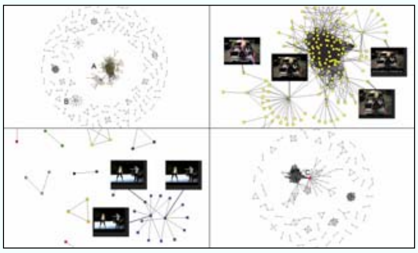
[PCS09]

Illustrative results

- Web2.0 data (from YouTube)

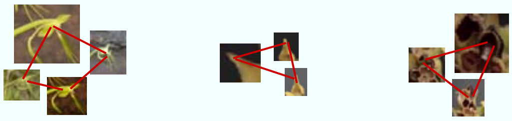
- ◆ Performance
- → Quality: recall 0.8-0.85 at précision 0.95-0.96 (base CIVR 2007)
- → Cost: 10 000 hours of video in ~3.3 days; 63 hours in 42 seconds





Interactive retrieval after mining

- Complex concepts → very large search space
  - Learning: requires too much training data
  - Difficult to strongly reduce computational complexity of search
- Strong assumption: meaningful retrieval results rely on elementary regularities that can be found *a priori*
- Example for plant identification:
  - Identify complex patterns that are redundant in the database

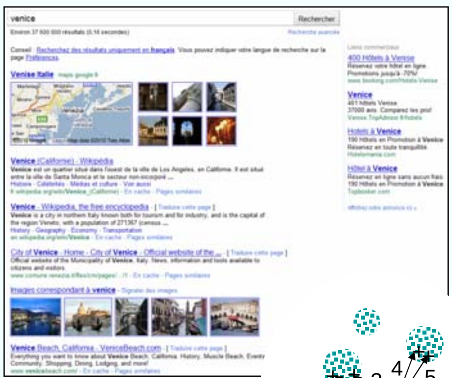


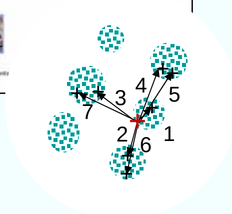
  - Can be done with methods related to copy detection!

IR beyond ranking

- What is informative in the results of search?
  - Potentially several meaningful views, each with a specific grouping
  - What about ranked answers?
    - ⇒ Mixture of many different dimensions of similarity
    - ⇒ Individual answers rather than relevant clusters

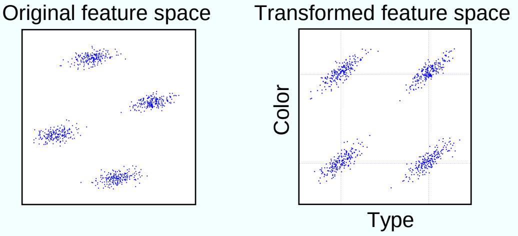
⇒ Ranking leads to a loss of information





Complementary clusterings [PC10]

- Complementary clusterings for a set of vectors, each clustering in a specific, arbitrarily oriented subspace
- Complementarity: one clustering provides little or no information regarding the other(s)



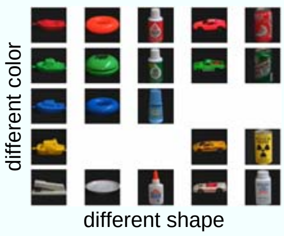
- Method: inspired by Tree-Component Analysis, but mutual information is computed between clusters in different subspaces

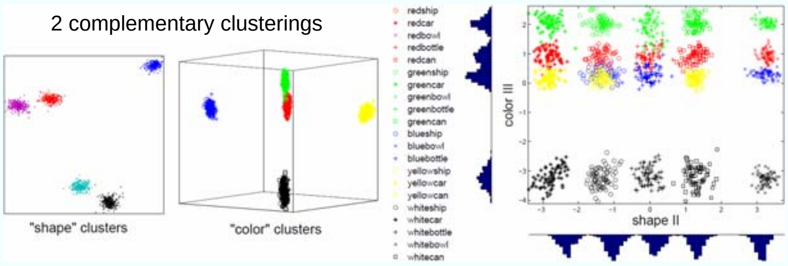
23/09/2010 M. Crucianu 29

Cédric

Illustrative results

Data:
21 object classes,
72 viewpoints/class
Global descriptions





23/09/2010 [PC10] M. Crucianu 30

Cédric

Conclusion

- Interactive retrieval/mining can allow to define/find task-dependent or user-dependent complex visual concepts/patterns
- Difficult but possible to scale such methods to large databases
- Search should provide more informative answers to user queries
- Prior mining allows to improve subsequent interactive retrieval (quality, speed, information content…)

23/09/2010 M. Crucianu 31

References

[CEOT08] Crucianu, M., Estevez, D., Oria, V., Tarel, J.-Ph. Speeding Up Active Relevance Feedback with Approximate kNN Retrieval for Hyperplane Queries, *International Journal of Imaging Systems and Technology*, *Special issue on Multimedia Information Retrieval*, Vol. 18 (2-3): 150-159.

[FBC04/1] Ferecatu, M., Crucianu, M., Boujemaa, N. Retrieval of Difficult Image Classes Using SVM-Based Relevance Feedback, *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, October 2004, New York, pp. 23-30.

[FBC04/2] Ferecatu, M., Crucianu, M., Boujemaa, N. Sample selection strategies for relevance feedback in region-based image retrieval. In K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): *Pacific-Rim Conference on Multimedia 2004*, Lecture Notes in Computer Science 3332, pp. 497-504, Springer-Verlag.

[GCB08] N. Grira, M. Crucianu, N. Boujemaa. Active Semi-Supervised Fuzzy Clustering, *Pattern Recognition*, Vol. 41, No 5, pp. 1851-1861.

[OBC10] W. Ouertani, P. Bonnet, M. Crucianu, N. Boujemaa, D. Barthélémy. Iterative Search with Local Visual Features for Computer Assisted Plant Identification. In *BioIdentify 2010*. Paris, September 20-22, 2010.

[PCS09] S. Poullot, M. Crucianu, and S. Satoh. Indexing local configurations of features for scalable content-based video copy detection. In LS-MMRM: *1st Workshop on Large-Scale Multimedia Retrieval and Mining*, with *ACM Multimedia 2009*, pages 43–50, New York, NY, USA, 2009. ACM.

[PC10] Pröscholdt, T., Crucianu, M. On Finding Complementary Clusterings, In ESANN 2010: *European Symposium on Artificial Neural Networks*, Brügge, Belgium, 2010.

23/09/2010 M. Crucianu 32