Scene text localization based on the ultimate opening. Application on imagEval database campaign.

ISMM 2007:Session Image Processing

Thomas Retornaz
Advisor:Beatriz Marcotegui Iturmendi

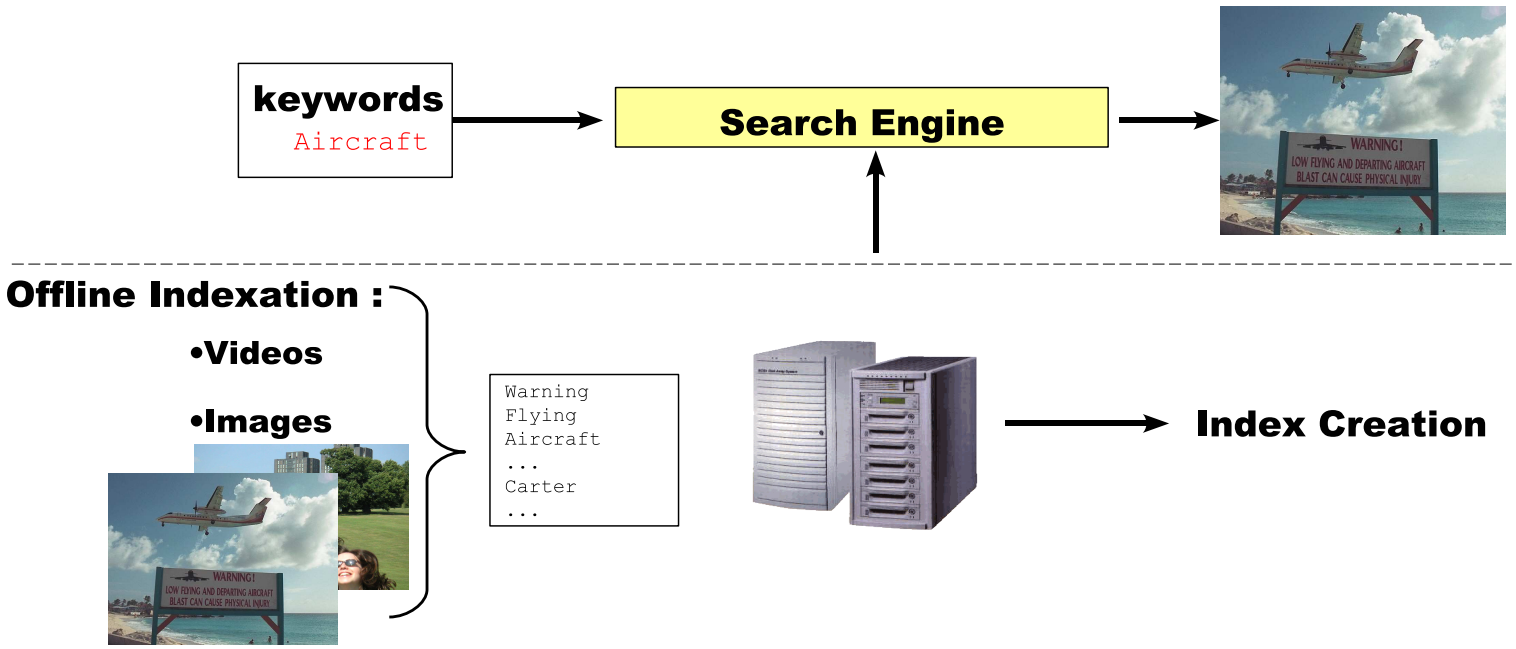
Centre de Morphologie Mathématique
École des Mines de Paris

October 12, 2007

Outline of the presentation

Outline of the presentation

CBIR Interest on text localization



Interest in text localization/recognition

It is a high level descriptor which gives semantic information

- Direct extraction of key words
- These key words are (usually) linked with the semantic of the image

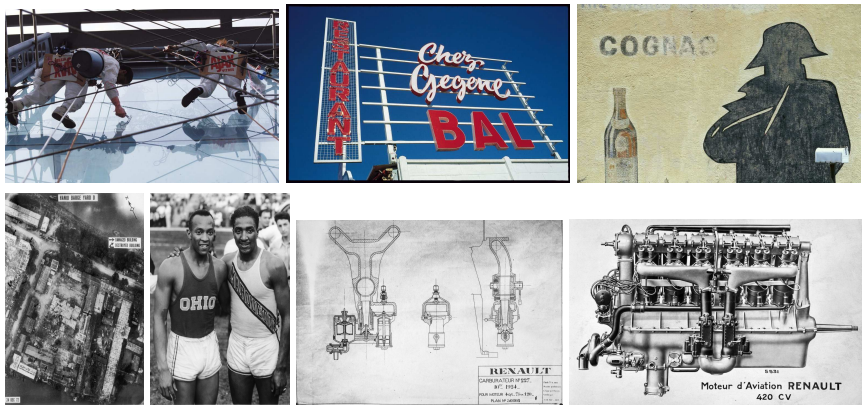
Our Job

We have to localize them

Application :ImagEval Evaluation Campaign

ImagEVAL objective is double :

- constitute data-bases representing real needs (Dataset : Hachette, . . .)
- evaluate technologies held by national and foreign research laboratories



Task 3 : Text localization contest, 2x500 images With Or Without Text

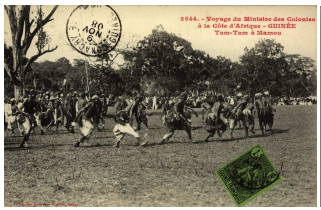
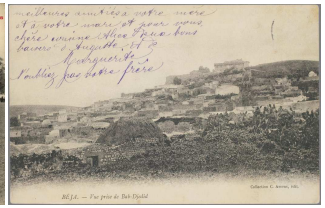
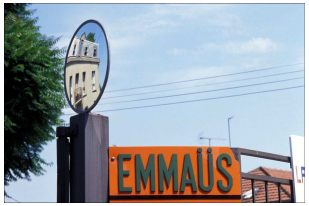
Difficult database : Annotated by ImagEval committee

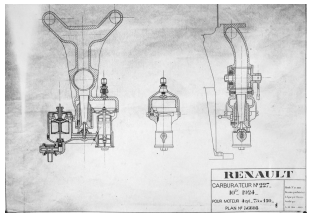
- Old postcards : overlay and scene text, various fonts, manuscript, stamps
- Natural images : geometric distortions, word art, manuscript, scales
- Natural images : occulted text, stuck character, textured background
- Archived images : overlay and scene text, various font and typo

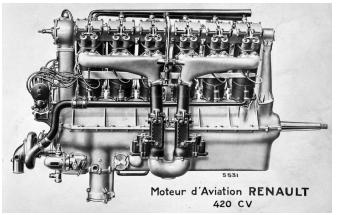
Outline of the presentation

We have to build an application dealing with :

- Multiple scales of text zones
- Font/color variability
- Relaxed alignment constraints
- Compliant/or not compliant background








Hypothesis

- Contrast : Characters are contrasted (VS local background)
- Global geometry : Common geometrical characteristics on same text zones
- Relaxed alignment constraint
- VERY common hypothesis : At least three characters

Flowchart of the algorithm

Bottom-up approach

Note : Each polarity of text is being processed separately

1. Extract connected components
2. Filter extracted connected components

   discard most of false positives (no letters)
3. Merge the remaining components

   Create bounding boxes around text zones, which are required for the evaluation.
4. Merge the result of both polarities

Starting point

First we have to extract connected components using local contrast information

$\Longrightarrow$ For this step, we use ultimate opening operator

Step1 :use ultimate opening to expose letters

A residual operator for gray level images : its goal ?

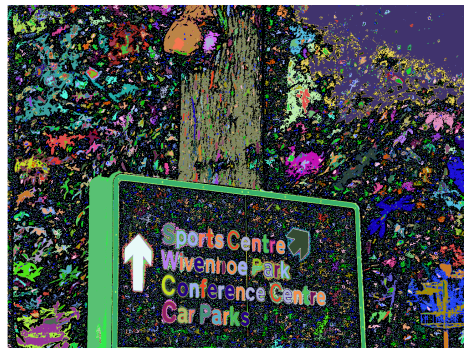
To extract the most "significant" structures from an image based on "local" information about contrast.

Definition

- Compute for each point of an image I, residue
  The difference between two successive openings of increasing size
  $r_\lambda(I) = \gamma_\lambda - \gamma_{\lambda+1} \forall \lambda \geq 1$ with γ_λ a family of opening
- Keep two pieces of information for each pixel x :

Example Image
Attribute opening
Vertical ferret diameter

Residual image : R_θ
Registers the value
of the maximal residue
Conveys local contrast



Associated image : q_θ
registers the size of the opening
that produced the maximal residue
Conveys size information

Some results

Original Image | R_θ ($\gamma = 3$) | q_θ (Labeled)

Example using attribute opening : criteria used vertical ferret diameter

Step1 :extract connected components

Thresholding of R_θ

- R_θ contains contrast information
- Based on the assumption that letters have a minimal contrast, we could basically threshold them

Contrast Image : R_θ

Thresholding of R_θ

Each Polarity
Residual Operator
Coarse Filtering
Fine Filtering
Iterative AMA
Final Merge

Next Step : we have to discard false positive (ie no letters)

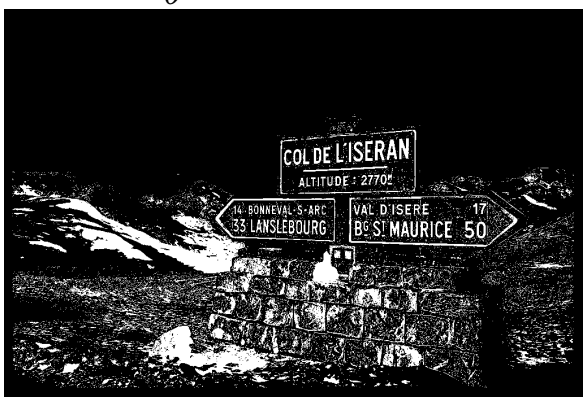
Step2 :coarse filtering strategy

Design two simple filters to discard a lot of false positives

- First using spatial coherence of text zone
- Second based on stroke thickness estimation

Remove all cc whose height is smaller than twice its thickness

R_θ thresholded



Result of two step coarse filtering

Each Polarity
Residual Operator
Coarse Filtering
Fine Filtering
Iterative AMA
Final Merge

Conclusion

- Reduce quickly a great number of false positives, and accelerate the following steps
- We have to design a new filter based on learning strategy

Step3 :Fine Filtering Strategy

Goal : Discriminate characters/no charaters using machine learning

Training : a CC subset from the blind campaign dataset, annotated as letters VS other

Feature stability ? ?

- Digitalization
- Slant
- Scale

Overall : 27 basic features

1. Stroke thickness estimation and coherence : RLE, distance
2. Geometric features : height, width, area of BB -> provide normalization
3. Shape regularity features : Euler, compacity, complexity, ...
4. Contrast : Maximum Inter Class Variance ($M.V.I$)

Step3 :Result of fine filtering strategy

Learning Strategy

LDA quadratic
Classical cross-validation scheme

Detected / *GT*	*Characters*	*No Characters*
Characters	89.1	10.9
No Characters	9.7	90.3

90% of text/non text is correctly classified

Previous Image

After learning Strategy

- Each Polarity
- Residual Operator
- Coarse Filtering
- **Fine Filtering**
- Iterative AMA
- Final Merge

Conclusion

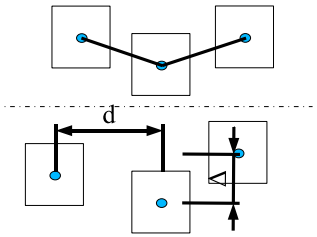
Satisfactory filtering

Aligning and merging analysis

Provide bounding box required for evaluation

White:
CC validated by learning

Gray :
CC Invalidate



Predicate:
- **Intercharacter distance**
- **distance between centroids**
- **Height difference**
- **....**

1 Previous Image

2 Seeds of at least 3 CCs

3 Expand seed and (re) validate cc

4 Recursive Merging

- Each Polarity
- Residual Operator
- Coarse Filtering
- Fine Filtering
- **Iterative AMA**
- Final Merge

Conclusion

- Relaxation strategy : some misclassifications are recovered.
- May produce new false positives

Outline of the presentation

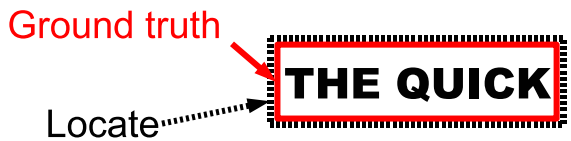
Metrics for text localization

Evaluation based on rectangle overlap

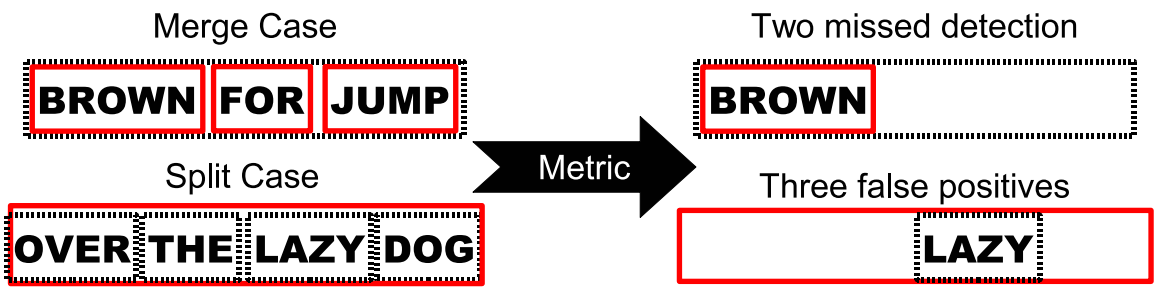
S.M. Lucas etal. IJDAR, 7 2-3 105–122 2005.

Provide just one-to-one match

1. Perfect match : Precision/Recall=1



2. Split or Merge : keep just the best match



Recall that this metric depends on annotation granularity

Results on ImagEval contest

Experiment

ImagEval Task 3 : Text localization contest, 500 images With Or Without Text

Precision	Recall	F-Mean
0.490	0.650	0.559
Ground Truth Boxes		580
Located		745

In spite of the difficulty of the base, half of the text has been correctly detected

Results

- Same parameters for the whole dataset
- Taking into account : scale, slant, font variability

Results on ICDAR Dataset

International Reference Dataset (downloadable for free)

Experiment on ICDAR Dataset

IDCAR Robust Reading and Text Locating contest : 500 images With Or Without Text

Precision	Recall	F-Mean
0.41	0.57	0.48

Half of the text has been (also) correctly detected

Results

- Same parameters for the whole dataset ICDAR and ImagEval
- "Good Result" on ICDAR dataset compared to other systems
  Have a look at [Lucas etal(2005)]

Typical example of missed detections :

Not reachable text by our approach

Text embedded in texture, vertical text, connected characters, manuscript

Text embedded in texture

Vertical text

Connected characters

Outline of the presentation

Conclusion

- Criteria-based ultimate opening introduced for CC extraction
- Machine learning strategy for characterizing letters (90% correctly classified)
- Satisfactory results in generic ImagEval database
  - First position in ImagEval text detection campaign
  - half of the text has been correctly detected in spite of the variability of text zones
- Satisfactory results in the international ICDAR database (with the same parameter set used for ImagEval)

Prospect

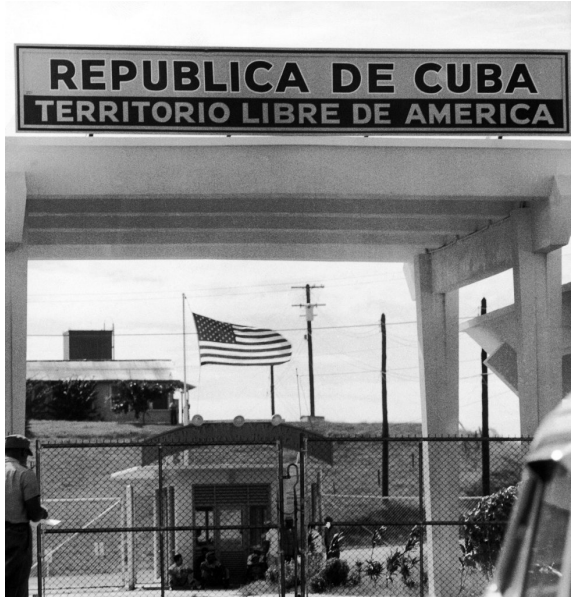
- Localization step
  - Use the image "size" q_θ (already done, www.imageval.org)
  - Evolution of algorithm performance/stability with various sets of parameters (already done in my Phd work)
  - Tackle Heuristics -> Statistical Framework -> Global Optimization
- Achieve recognition step
  - Less biased metric for evaluation purpose
  - Real feature for CBIR system

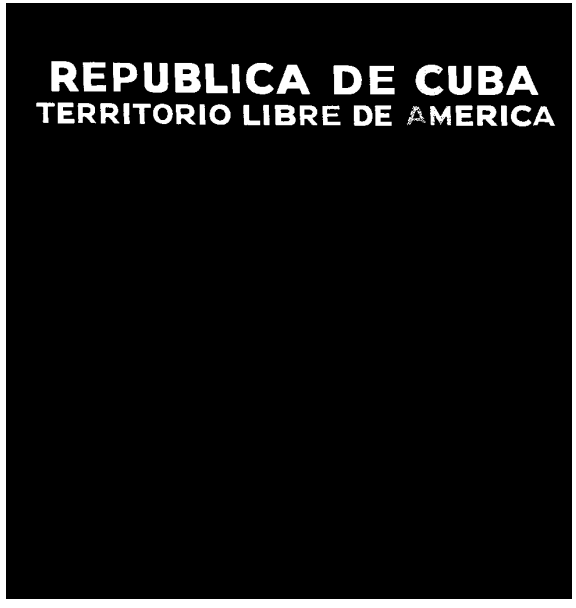
Thanks you Any questions ?

Outline of the presentation

we use a connected component approach

so, we have some CCs tagged as text. . .



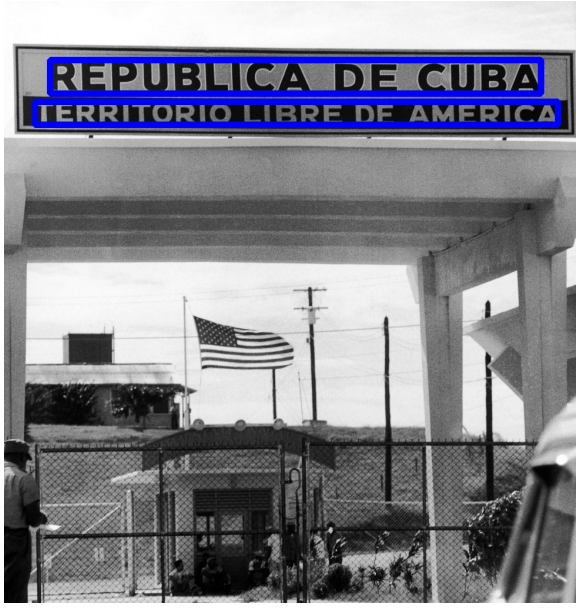


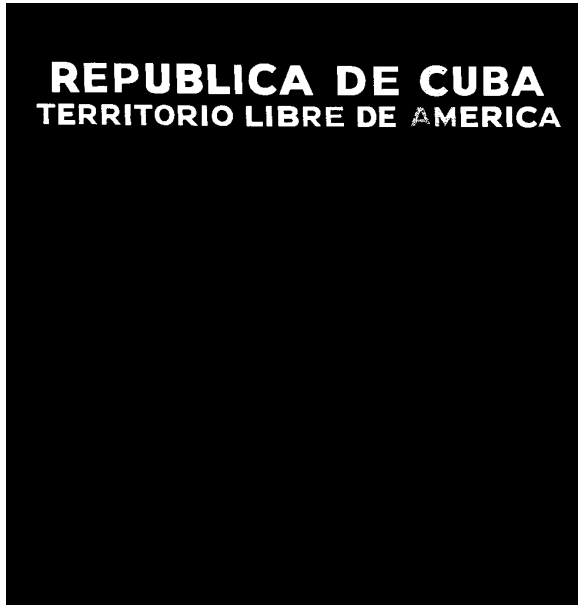
Open Source OCR Tesseract (http://code.google.com/p/tesseract-ocr/)

- free of course
- **open learning module**

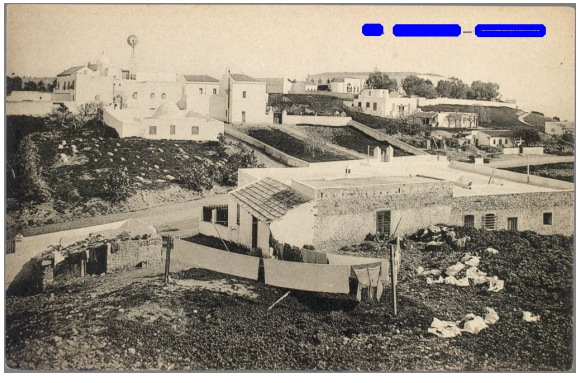
No peculiar attention has been payed to improve OCR ouput

Goal 1 :achieve recognition, if possible





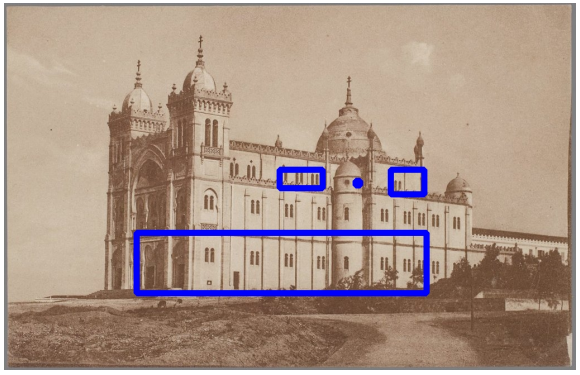
- Black :
  TERRITORIO LIBRE DE AMERICA
- White :
  REPUBLICA DE CUBA



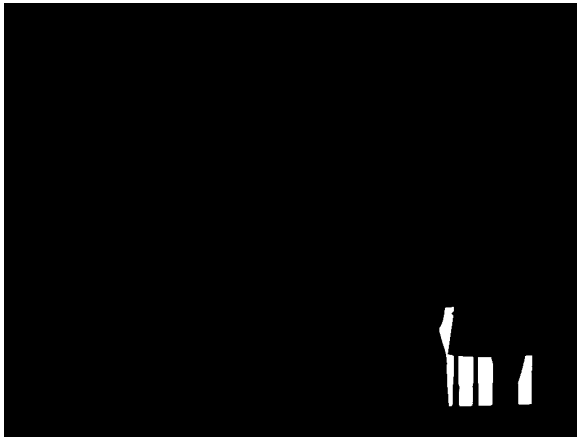
117 CARTHAGE Vue Generale

- Black :
  MY CARTHAGE Vue Generals
- White :

Goal 2 :discard false positives

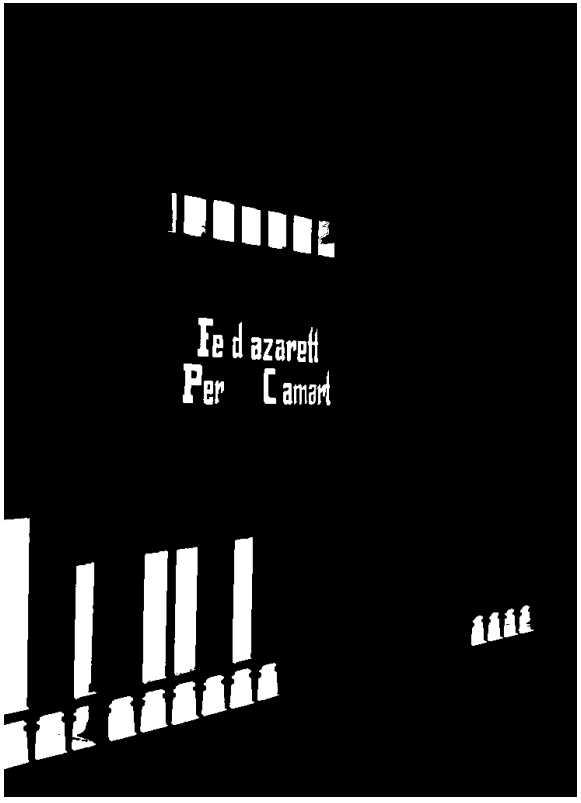


- Black :
  Uu L va \TT\ I
- White :
  WH IM we



- Black :
- White :
  tll I

BUT TRUE positives may be discarded



- Black :
  IIÍ¨ ?. I‘ ! llllll ma
- White :
  In cl BZBFEH PB ?’ [
  Emit‘ !

We have to

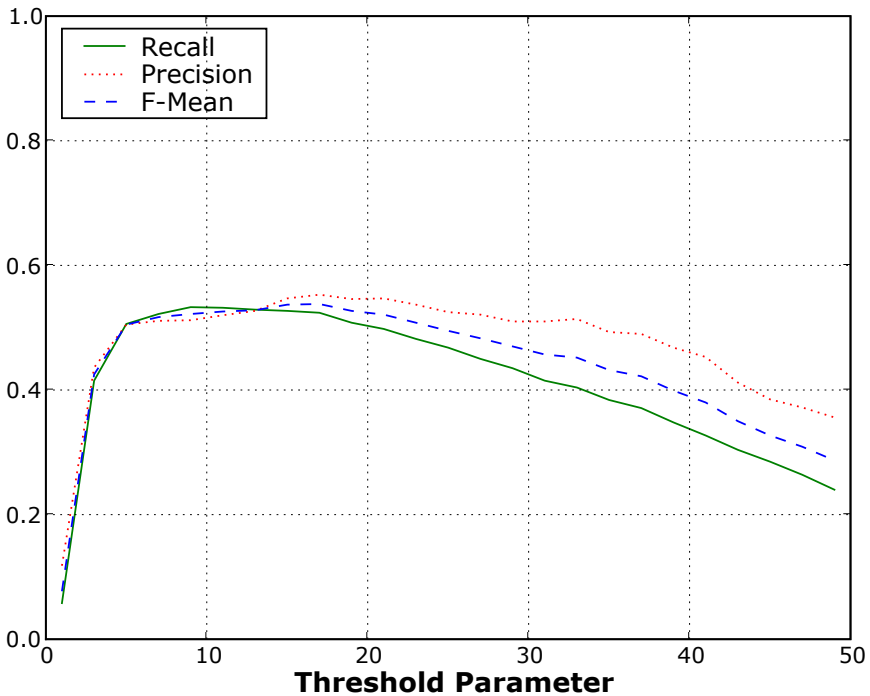
- Correct scale and distortion
- Character restoration ? MRF, . . .
- Achieve learning on OUR DATA

Outline of the presentation

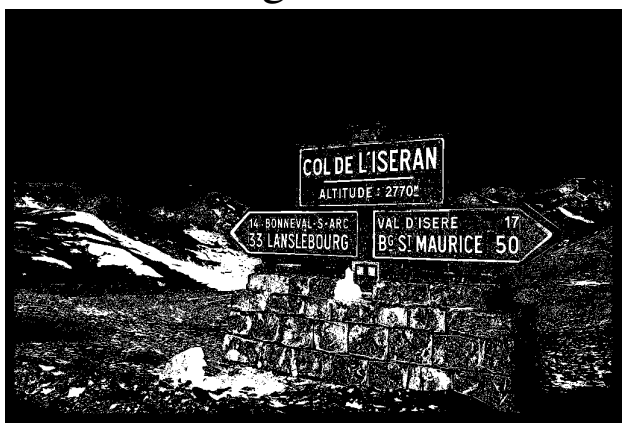
Stability respect to threshold parameter

Evolution of precision/recall/f-mean on ImagEval dataset

$\Longrightarrow$ Various threshold parameters



Contrast Image : R_θ



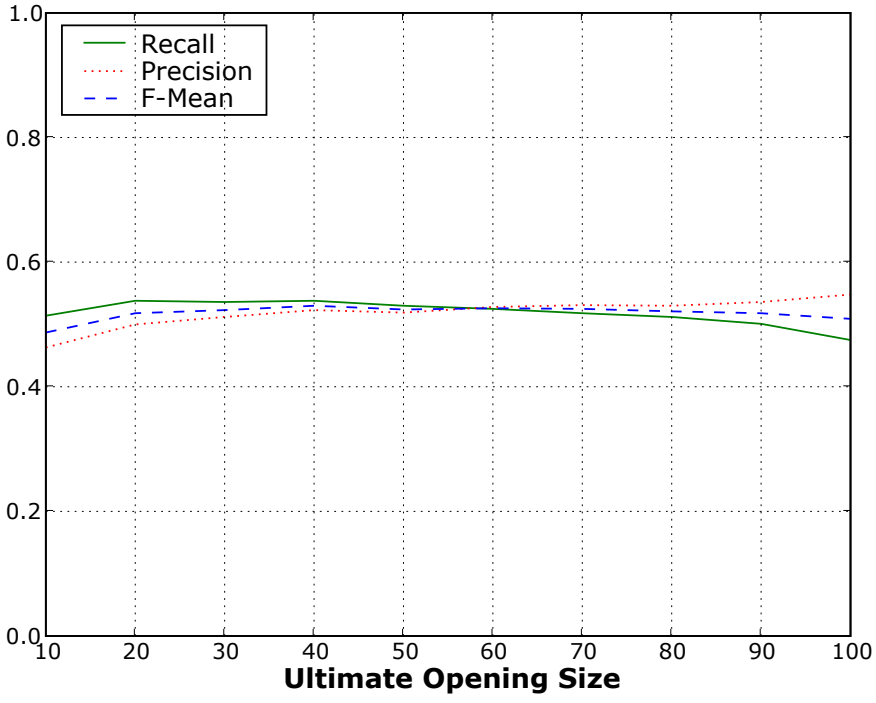
Thresholding of R_θ

Use alternative metric C. Wolf and J.M. Jolion IJDAR, 8 4, p.280 296, 2006

Stability respect to maximal opening size parameter

Evolution of precision/recall/f-mean on ImagEval dataset

$\Longrightarrow$ Various maximal opening size, with fixed threshold parameter



Initial Image

R_θ : Stop Half Image Size

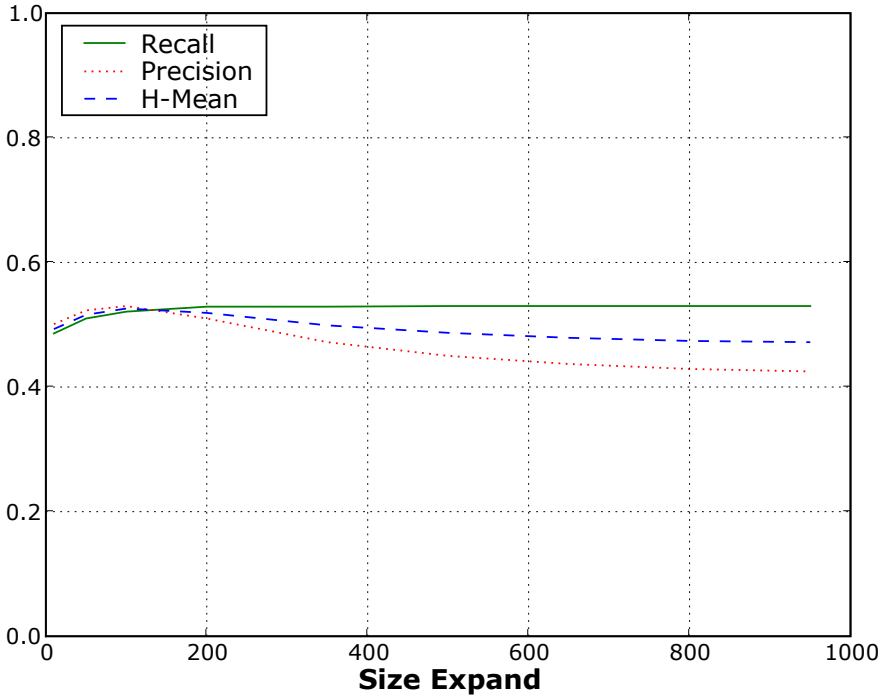
R_θ : Stop third Image Size

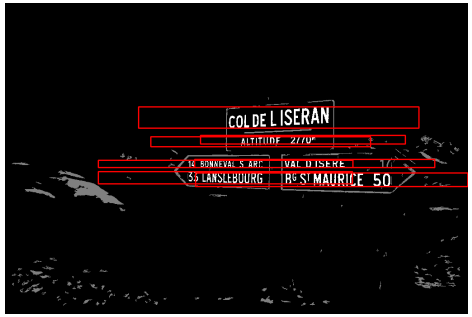
Use alternative metric C. Wolf and J.M. Jolion IJDAR, 8 4, p.280 296, 2006

Stability respect to "expansion" parameter

Evolution of precision/recall/f-mean on ImagEval dataset

$\Longrightarrow$ Various expansion parameter, with fixed threshold and stop parameter





Use alternative metric C. Wolf and J.M. Jolion IJDAR, 8 4, p.280 296, 2006