

Extraction du texte enfoui dans les images de scènes naturelles

Projet ANR iTowns
CMM – LIP6

J.Fabrizio, B. Marcotegui, M. Cord

R. Minetto, J. Guyomard, F. Preciosó







I - CONTEXTE a - Texte enfoui







I - CONTEXTE b – Pourquoi détecter le texte?

Quelles applications de la détection automatique du texte :

- Annotation automatique de grandes bases de données d'images, vidéo, multi-média
- Aide aux personnes malvoyantes
- Navigation enrichie en ville

Contexte itowns:

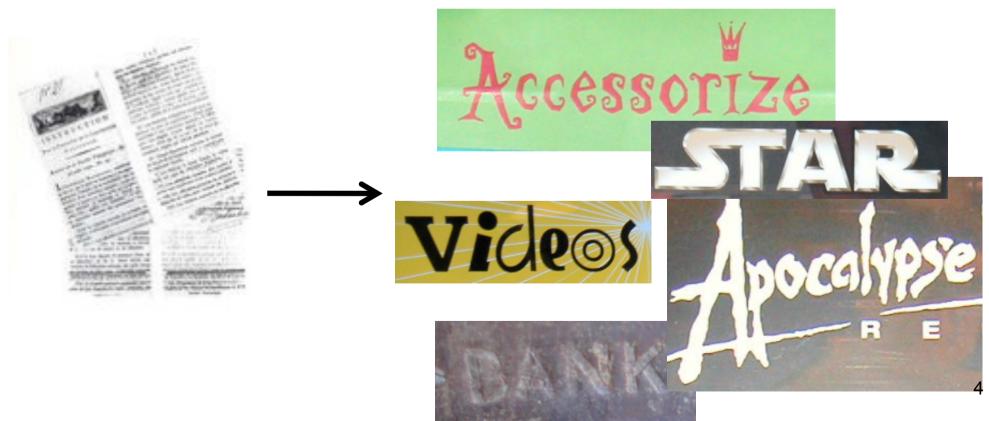
1. Enrichissement de la navigation dans le viewer



2. Enrichissement des index textuel pour la recherche de lieux par mots-clés

Problèmes rencontrés par rapport à un OCR (Optical Character Recognition) classique :

Grande variété des polices de caractères



Problèmes rencontrés par rapport à un OCR :

- Grande variété des polices de caractères
- Variations de luminosité



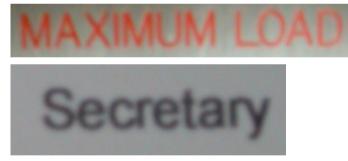






Problèmes rencontrés par rapport à un OCR :

- Grande variété des polices de caractères
- Variations de luminosité
- Images bruitées
- Images floues







Problèmes rencontrés par rapport à un OCR :

- Grande variété des polices de caractères
- Variations de luminosité
- Images bruitées
- Images floues
- Pas de mise en page définie
- Fond non uniforme

. . .

II - ÉTAT DE L'ART 2 grandes familles de méthodes

APPROCHES BASÉES "TEXTURE":

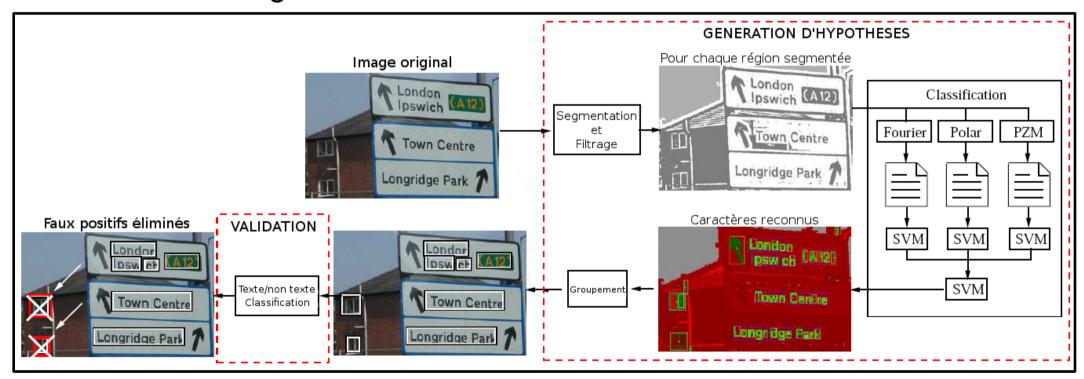
- 1/ Apprentissage d'1 fonction de classification Texte/ Non_Texte (à partir de vignettes exemples)
- 2/ Détection par Fenêtres glissantes à plusieurs échelles
- 3/ Fusion des boîtes et validation

APPROCHES BASÉES RÉGIONS

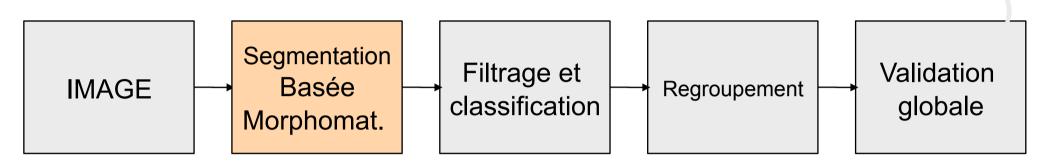
- 1/ Segmentation de l'image (au niveau des caractères)
- 2/ Détection des régions de caractères
- 3/ Groupement des caractères pour former le mot/ligne/texte

J. Fabrizio (post-doc commun LIP6 et CMM)

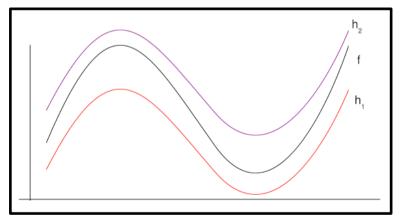
Chaîne globale:



J. Fabrizio (post-doc commun LIP6 et CMM)



Inspiré du Toogle Mapping [Serra]



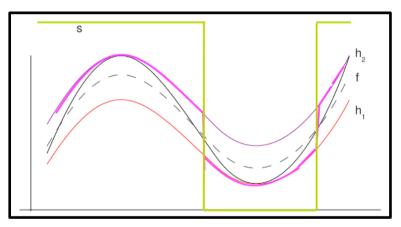
f(x): image

h1(x): érosion morphologique de

l'image

h2(x): dilatation morphologique de

l'image

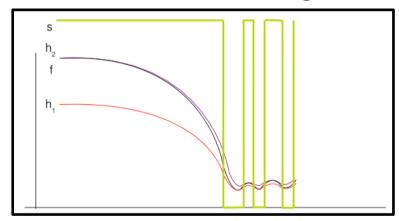


s(x): binarisation de l'image

J. Fabrizio (post-doc commun LIP6 et CMM)



Cas des zones homogènes:



s(x): binarisation de l'image Oscillations dues à l'homogénéité de la zone.

 \rightarrow Ajout d'une contrainte.



J. Fabrizio (post-doc commun LIP6 et CMM)



Image segmentée →



Filtrage rapide des zones segmentées, basé sur des heuristiques géométriques →

Classification des régions segmentées à l'aide de SVM (Support Machine Vector) →



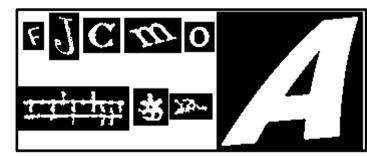


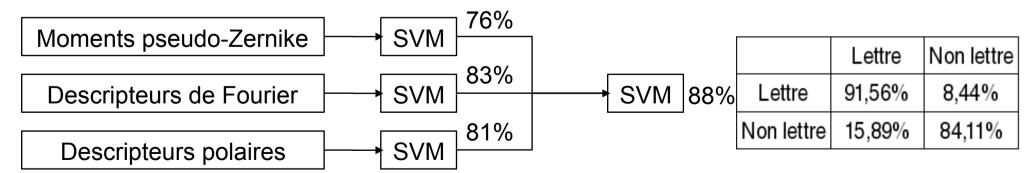
J. Fabrizio (post-doc commun LIP6 et CMM)



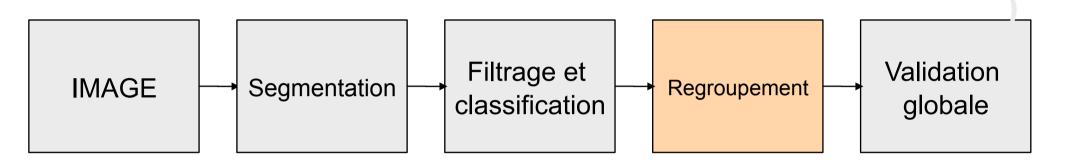
Entraînement de classifieur de type SVM (Support Vector Machine) sur:

- une base de données de 32400 exemples
- une base de test de 3600 exemples

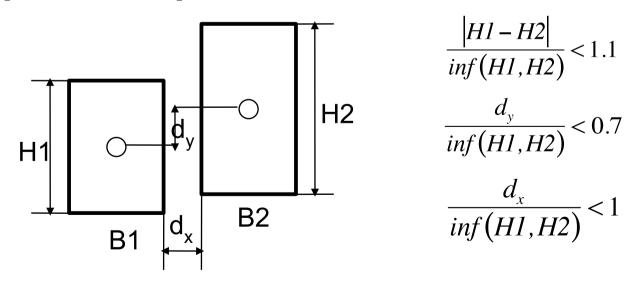




J. Fabrizio (post-doc commun LIP6 et CMM)



Le regroupement s'effectue selon plusieurs heuristiques simples [Retornaz et al] :



J. Fabrizio (post-doc commun LIP6 et CMM)

IMAGE Segmentation Filtrage et classification Regroupement globale



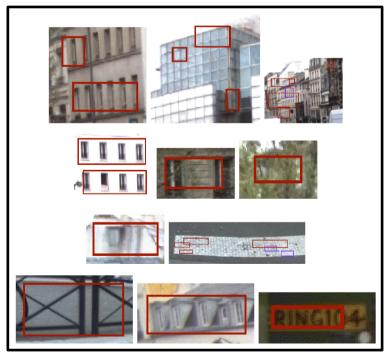
J. Fabrizio (post-doc commun LIP6 et CMM)

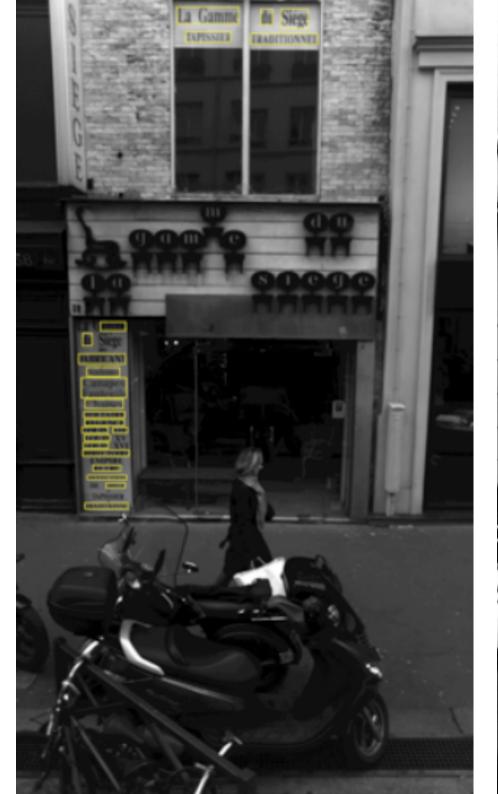


J. Fabrizio (post-doc commun LIP6 et CMM)



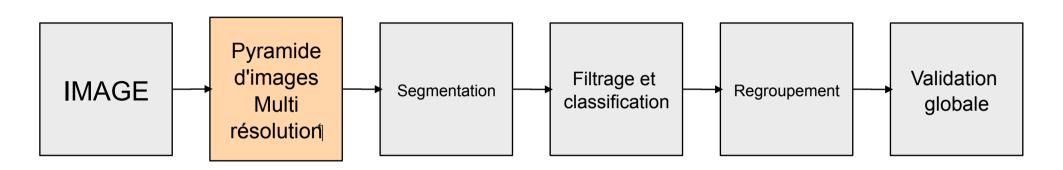
Exemples de fausses détections rejetées par le processus de validation basé sur l'apprentissage d'histogrammes de gradients orientés







IV – SnooperText Thèse R. Minetto (LIP6-Brésil)









Détection multi-résolution texte détecté sur 3 niveaux de la pyramide

V – RECONNAISSANCE J. Guyomard & F. Precioso (LIP6)

Coupler la détection avec un OCR en sortie Tesseract:

- Développé par Hewlett Packard de 1985 à 1995
- Open source depuis 2006
- Actuellement développé par Google
- http://code.google.com/p/tesseract-ocr/

Abbyy FineReader:

- Logiciel fermé
- Meilleur OCR du marché actuellement

Développés initialement pour reconnaître du texte dans des cas simples, en particulier de numérisation de livres et de documents papiers

V – RECONNAISSANCE OCR

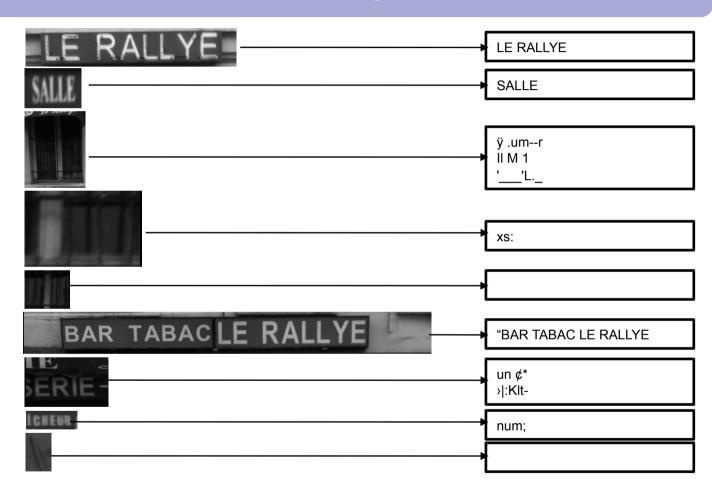
Intérêt de la détection de texte avant traitement par l'OCR



```
Sortie Tesseract:
`«¿\` HH > -, .V- Y
lx Âhs.
- .N «~*>~ -.--« *- .A vw-«L-x . A
mum . au
Cz«vI'-fr
'Mng'IQ
| || >/ _ _
BAR TABAC E RAU YE
ΙVÎ
A. Ö/*e aÿåe Tvi3- P
--A'-
9 "' A
QA\rightarrow A
* '/<f****
*~v'\;¿.¿_
. A- \_; \rightarrow \_ \rightarrow /I ,
..*>/*
_*',>'f;~;¿_`\Y,Y *¿_v.¿,_>_,;* _ 1 /r
'2':
** a ~¢>..=*"" 1
.J 1>_fî` @L
` >' fl
F"'. ain ¢.5ä§?`
flflfliflfl
```

V – RECONNAISSANCE SnooperText + OCR



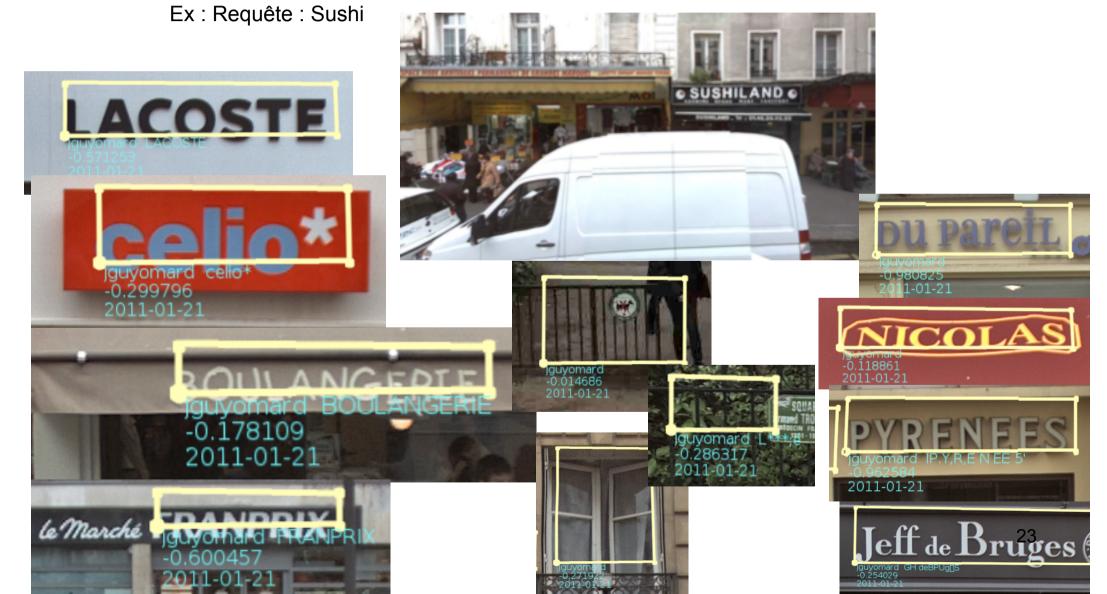


	Sans filtrage	Tesseract	Abbyy FineReader
Precision	0,380	0,468	0,496
Recall	0,598	0,551	0,527
F	0,440	0,481	0,478

CONCLUSION

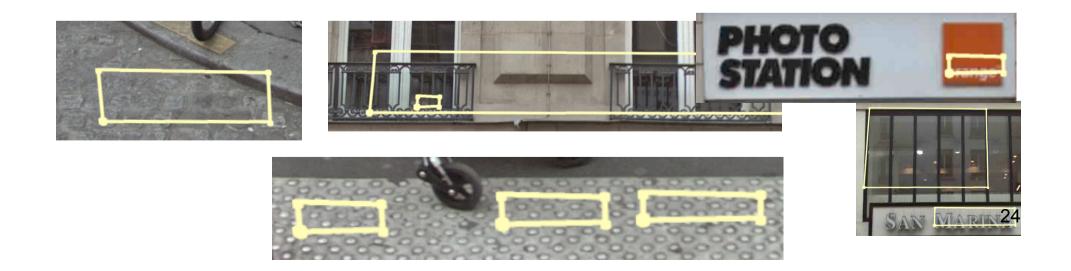
Méthode actuellement intégrée dans le viewer d'itowns

Moteur de recherche par mots clés enrichis par la reconnaissance automatique



PERSPECTIVES

- · Passage à l'échelle : actuellement 2min/image
- . Il reste des faux positifs et faux négatifs !
- → Améliorer le filtrage en sortie
- → Améliorer l'étape de classification des régions



PERSPECTIVES Classification par Boosting

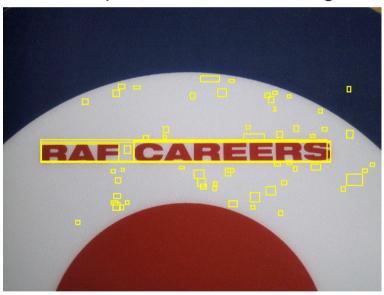
Le facteur limitant en temps actuellement : classif SVM régions

Snoopertext

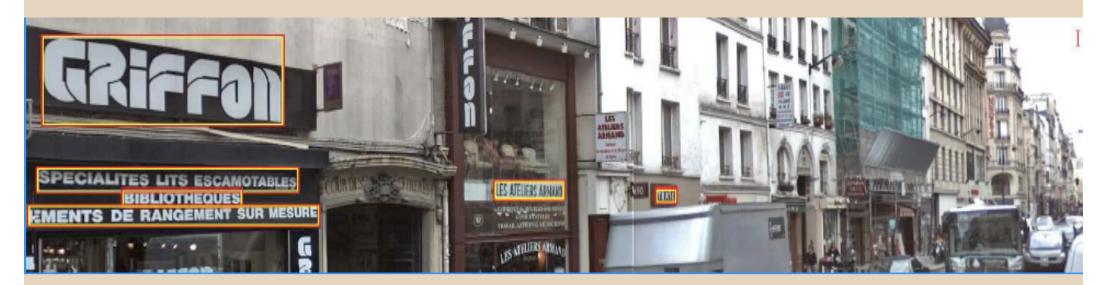


- Temps de calcul: ~2min

Snoopertext avec Boosting



- Temps de calcul: ~10sec



Extraction du texte enfoui dans les images de scènes naturelles

Projet ANR iTowns
CMM – LIP6

J.Fabrizio, B. Marcotegui, M. Cord

R. Minetto, J. Guyomard, F. Preciosó





