

Computer Vision Applied to Comic Book Image

Christophe RIGAUD, Jean-Christophe BURIE



Summary

↗ L3i lab

- > Presentation
- > Research strategy

↗ e-BDthèque project

- > Presentation
- > Dataset
- > Markup Languages

↗ Image analysis

- > Content-driven approach
- > Knowledge-driven approach

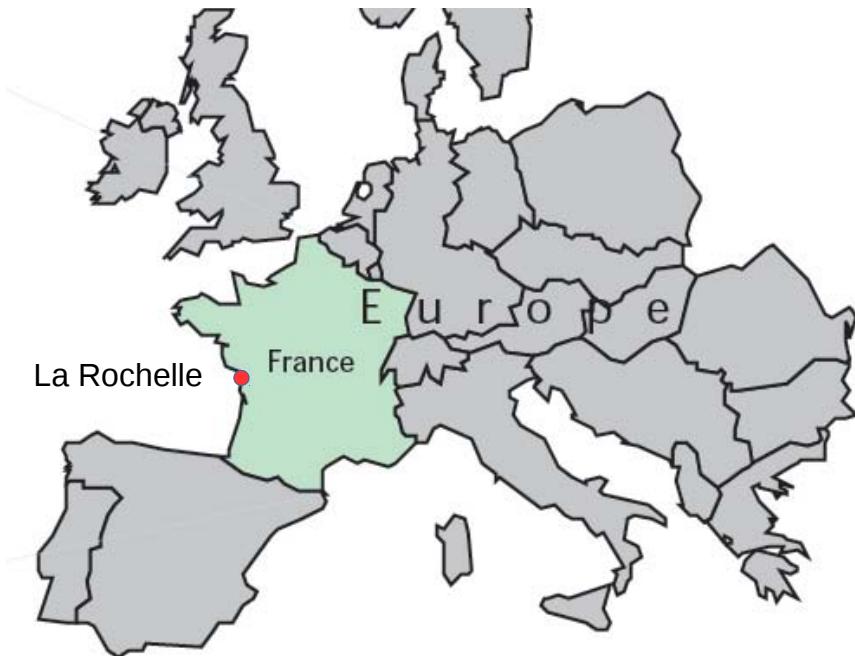
↗ Conclusion



L3i lab

L3i lab Presentation

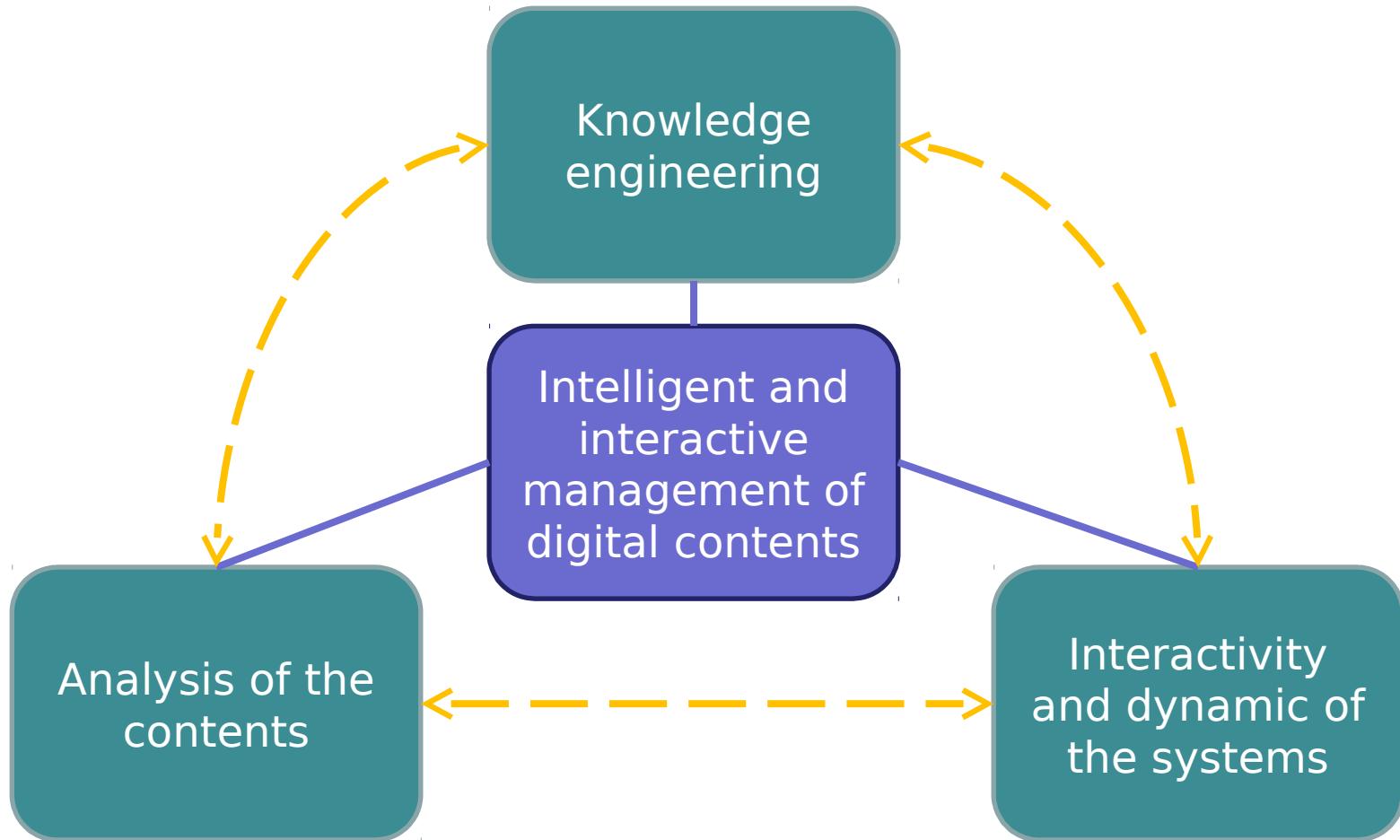
- L3i = Computer science,
Image and Interactions
- About 100 people

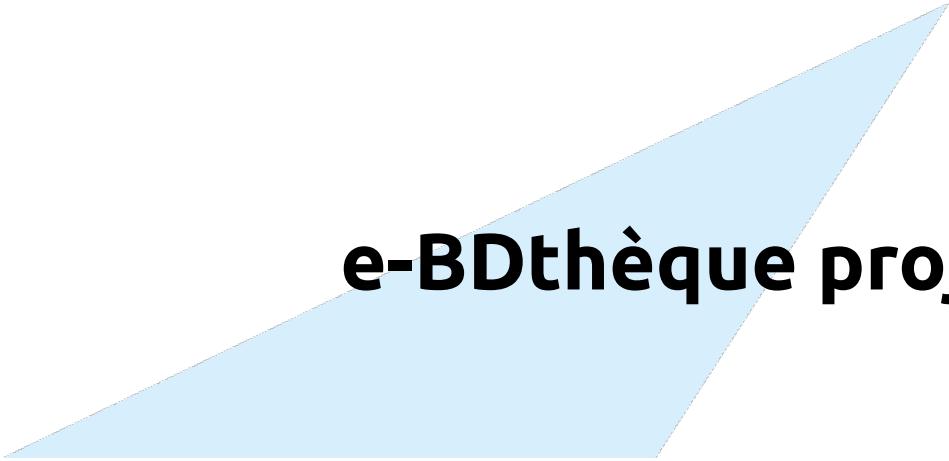


➤ <http://l3i.univ-larochelle.fr/>



Research strategy





e-BDthèque project

e-BDthèque project Presentation

↗ Objectives

- > Valorisation of cultural heritage
- > Content-based image retrieval

↗ 2011-2014

- > 10 participants (2 PhD candidates)

↗ 2013 -2015

- > Collaboration with OPU (Japan)

↗ 2015-2017

- > French startup partnership
- > Online comics library
- > <https://www.sequencity.com>

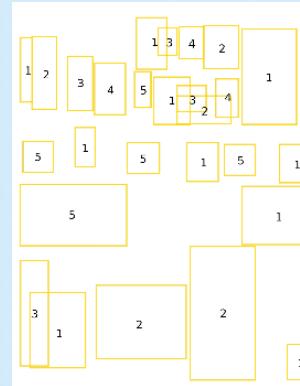
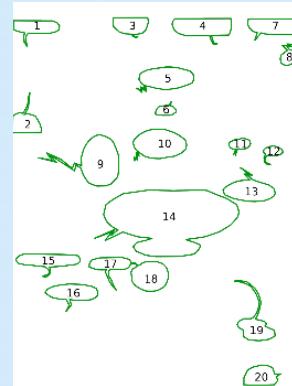
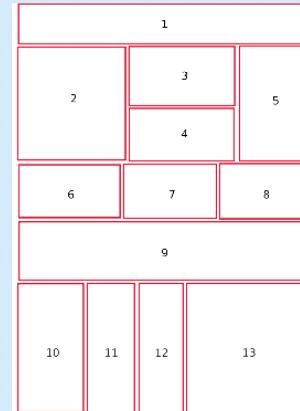


Image credits:
les bulles du labo, Gérald Lubbin, Doc en Stock, 2012

e-BDthèque project

Dataset (<http://ebdtheque.univ-lr.fr/>)

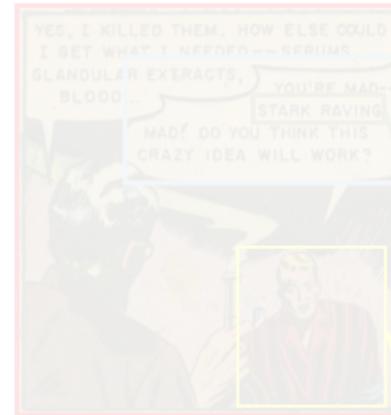
- ↗ 100 mixed pages from 20 albums
- ↗ Franco-Belgium BD, American comics and Japanese manga
- ↗ From 1905 to 2012, digitized and webcomics
- ↗ Rights holder permissions agreement

Bibliographic annotations



PAGE
Collection: Chilling Tales
Album: 17 Geo
Editor: Youthful Magazines
Drawer: Matt Fox
Writer: Matt Fox
Language: English
Page number: 16
Release date: 1953

Visual and semantic annotations



PANEL
Rank: 1

BALLOON
Rank: 2
Shape: Oval
Tail direction: South-West

TEXT LINE
Text: « STARK RAVING »

CHARACTER
LinkedToBalloon: 2

e-BDthèque project

Dataset (<http://ebdtheque.univ-lr.fr/>)

- ↗ 100 mixed pages from 20 albums
- ↗ Franco-Belgium BD, American comics and Japanese manga
- ↗ From 1905 to 2012, digitized and webcomics
- ↗ Rights holder permissions agreement

Bibliographic annotations



PAGE

Collection: Chilling Tales
Album: 17 Geo
Editor: Youthful Magazines
Drawer: Matt Fox
Writer: Matt Fox
Language: English
Page number: 16
Release date: 1953

Visual and semantic annotations



PANEL

Rank: 1

BALLOON

Rank: 2

Shape: Oval

Tail direction: South-West

TEXT LINE

Text: « STARK RAVING »

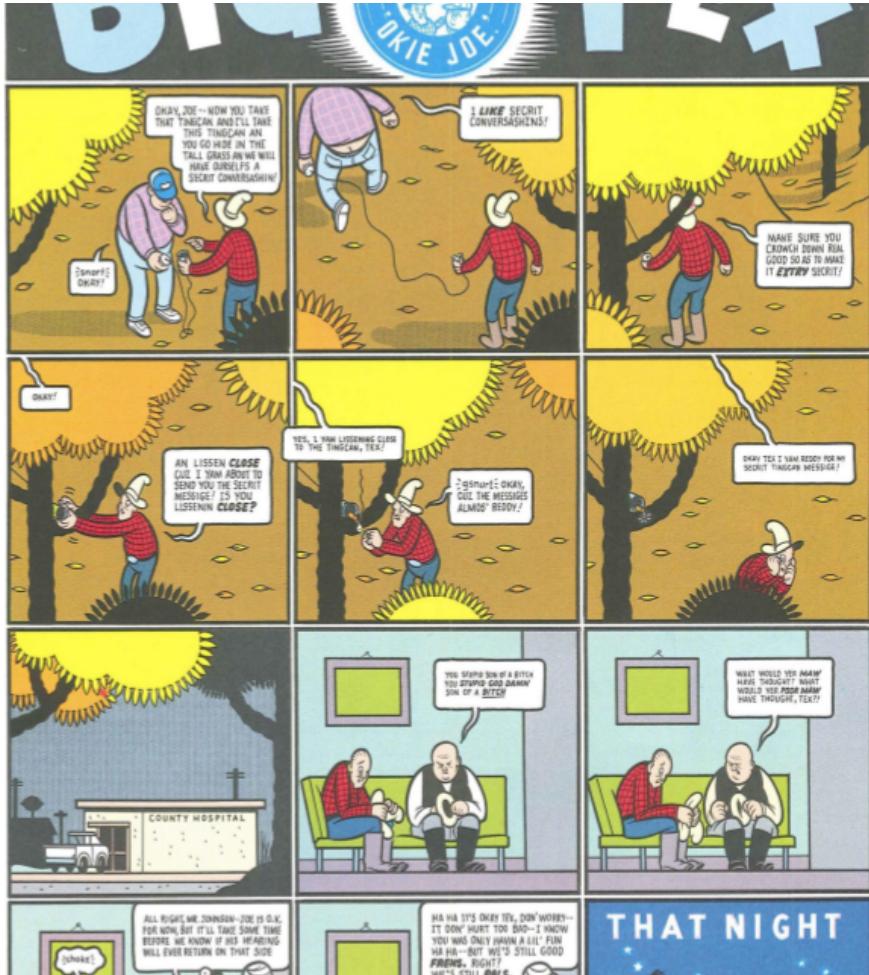
CHARACTER

LinkedToBalloon: 2

e-BDthèque project

Scalable Vector Graphics Markup Language

(<https://www.w3.org/Graphics/SVG/>)



```
<svg class="Panel">
  <polygon points="1023,629 1807,629 1807,1367
              1023,1367 1023,629">
    <metadata rank="3"/>
  </polygon>
</svg>
```



```
<svg class="Balloon">
  <polygon points="2238,1620 2570,1620 2570,1728
              2238,1728 2238,1620">
    <metadata shape="rectangle" rank="9"
              tailDirection="NW"/>
  </polygon>
</svg>
```

e-BDthèque project Comic Book Markup Language

(<http://dcl.slis.indiana.edu/cbml/>)



```
<cbml:panel  
n="5"  
characters="#cap #anon_man"  
ana="#action-to-action"  
xml:id="eg_000"  
  
xmlns:cbml="http://www.cbml.org/ns/1.0">  
<cbml:caption>  
Cap acts quickly to tranquilize the  
gun-happy pedestrian...  
</cbml:caption>  
<cbml:balloon xml:id="eg_007"  
type="speech" who="#cap">  
A little <emph  
rendition="#b">sleep</emph>  
will do wonders for you!  
</cbml:balloon>  
<sound>SPLAT!</sound>  
<cbml:balloon type="speech"  
who="#anon_man">  
Ugh!  
</cbml:balloon>  
</cbml:panel>
```



Image analysis

Image analysis Presentation

↗ Content-driven

- Similar to the literature
- Intuitive
- Sensible to error propagation

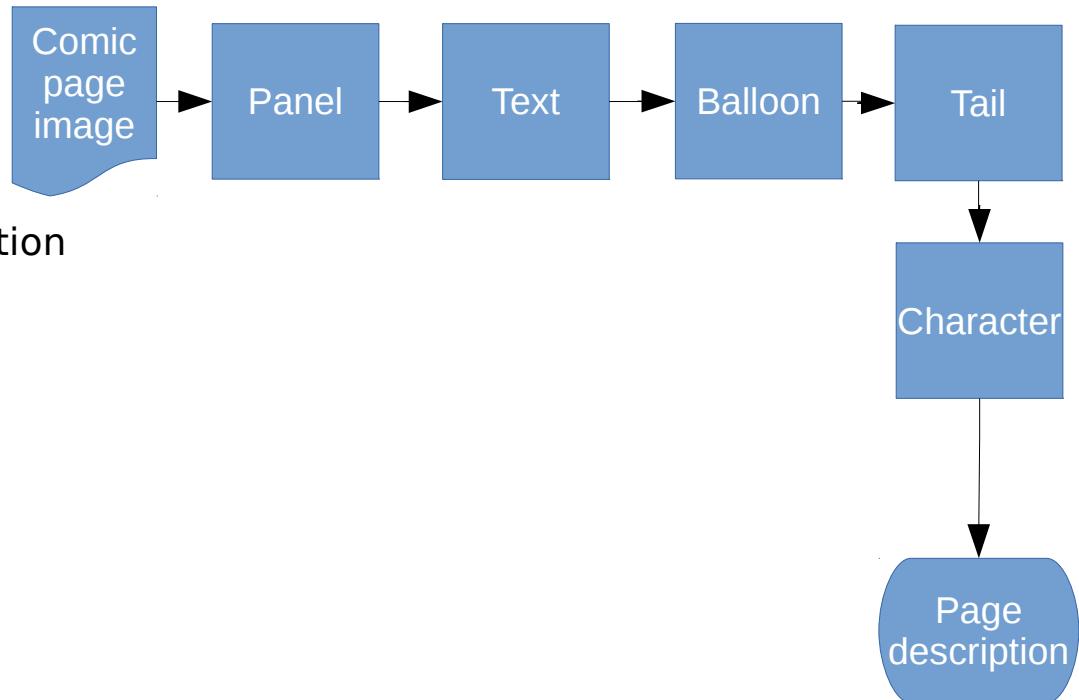
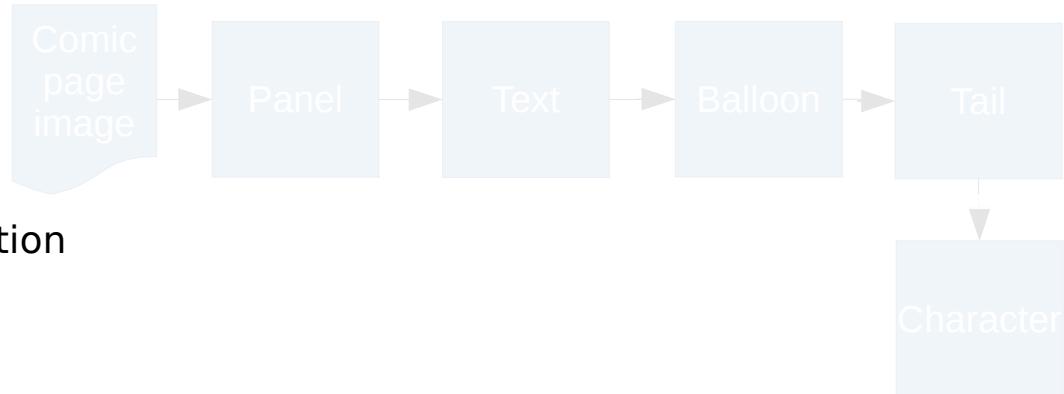


Image analysis Presentation

↗ Content-driven

- Similar to the literature
- Intuitive
- Sensible to error propagation



↗ Knowledge-driven

- > Based on domain knowledge
- > Retrieve context

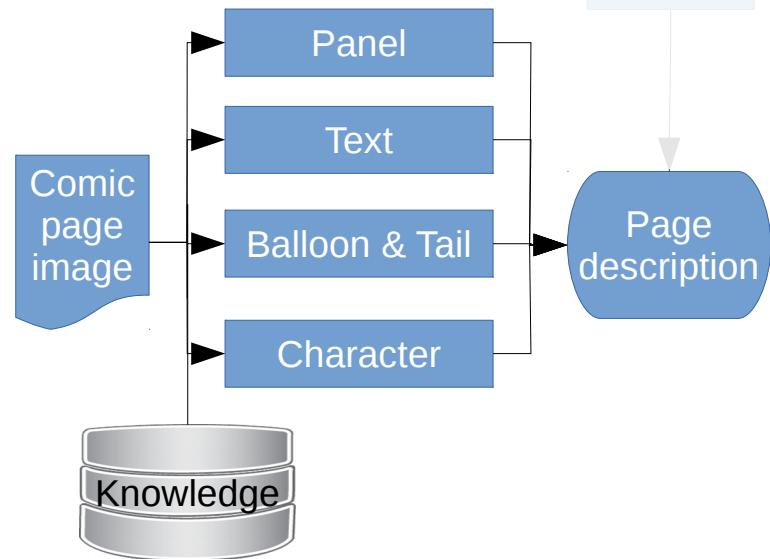
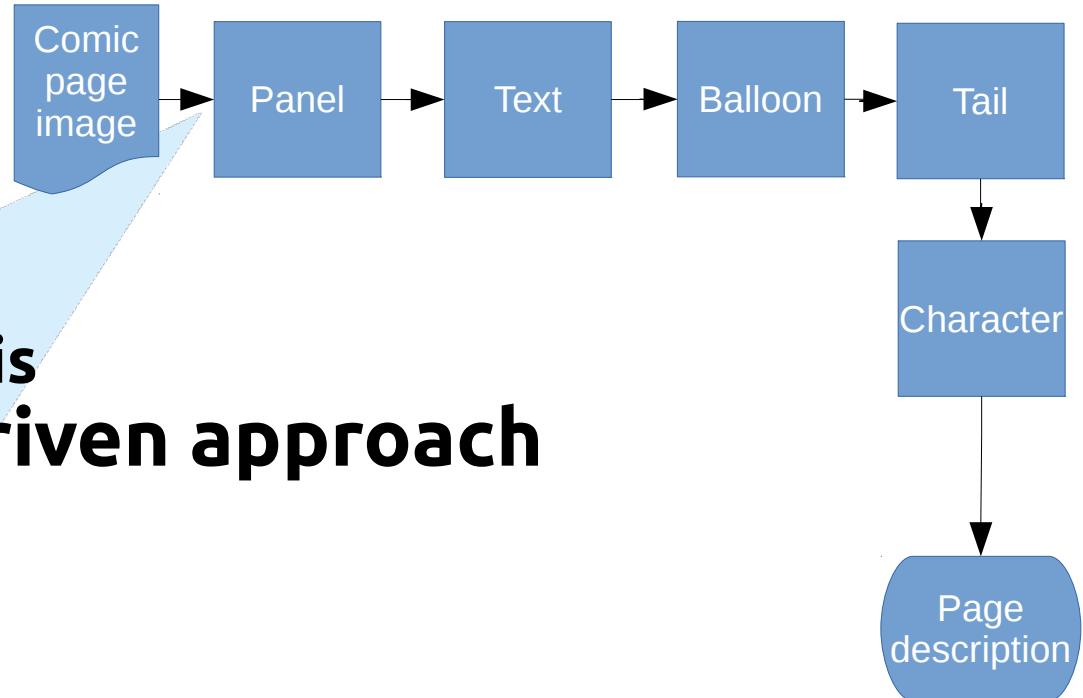
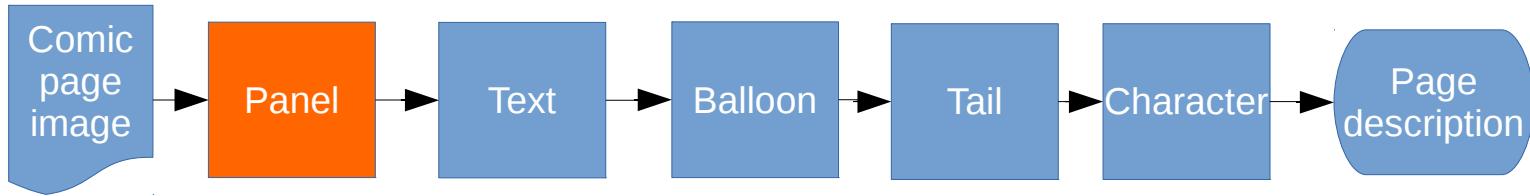


Image analysis Content-driven approach



Content-driven approach

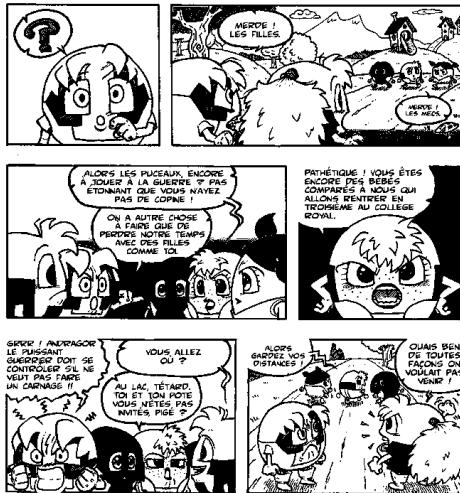
Panel extraction



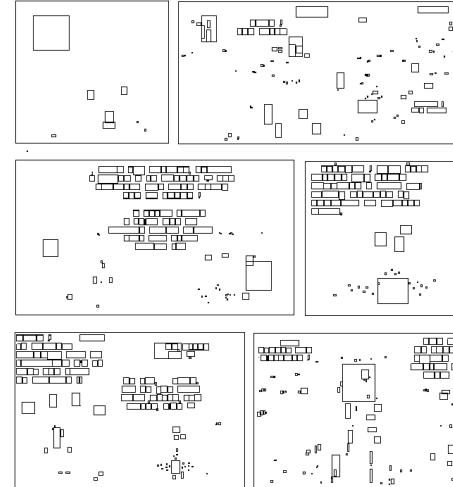
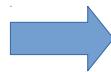
↗ Challenges

- > **Diversity** of styles (gutter, implicit, non-rectangle)
- > Semi-structured layout

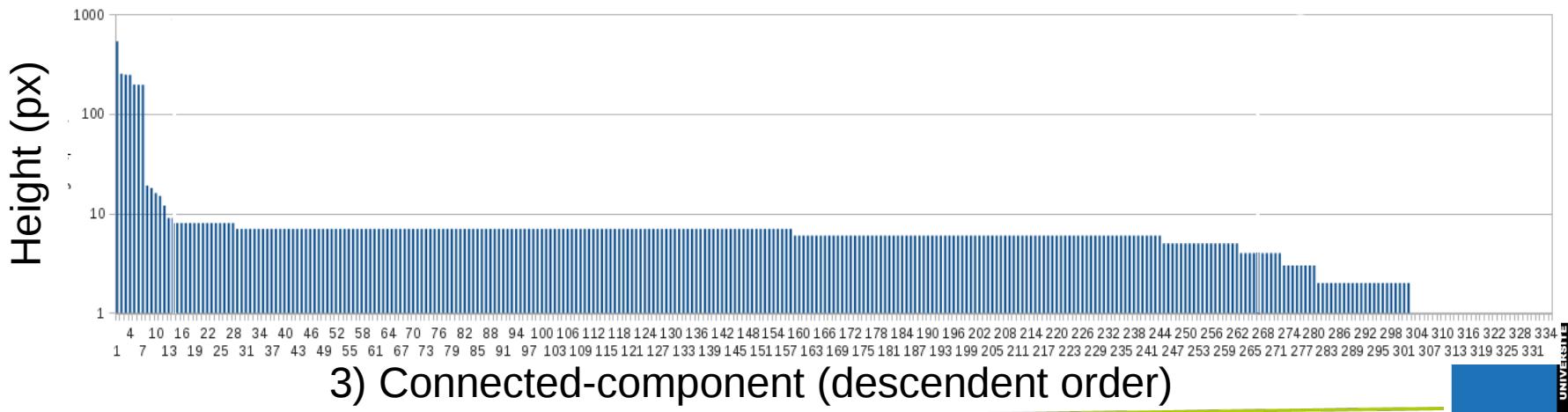
Content-driven approach Panel extraction



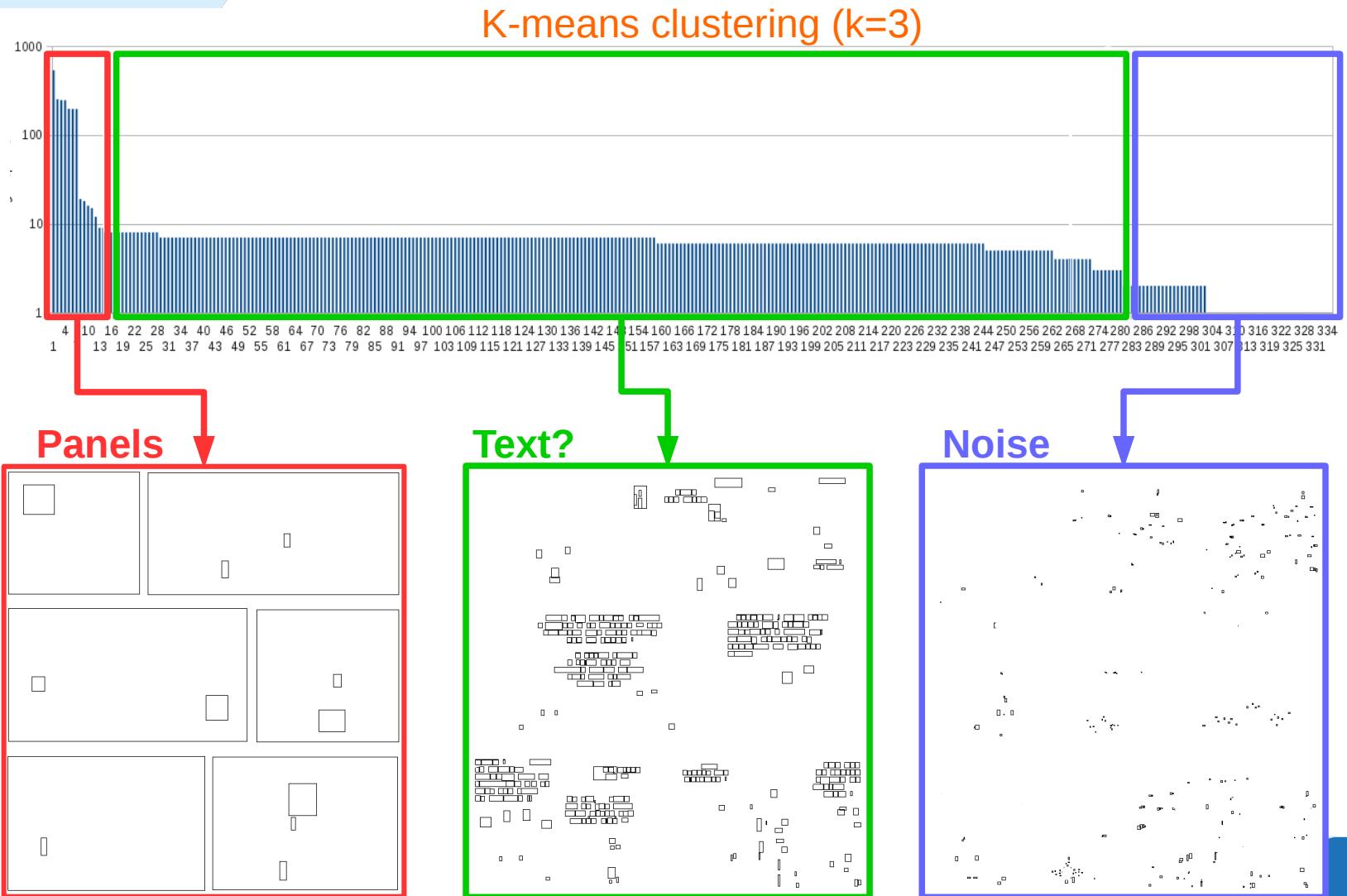
1) Binary image



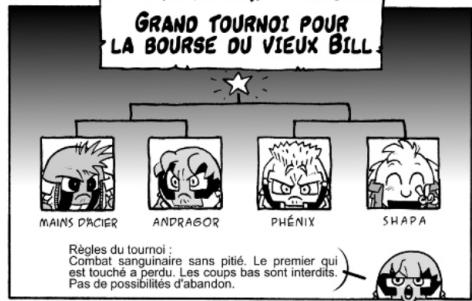
2) Connected-component



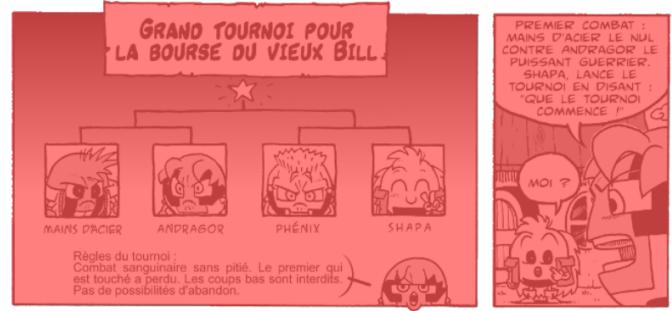
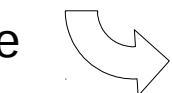
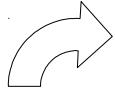
Content-driven approach Panel extraction



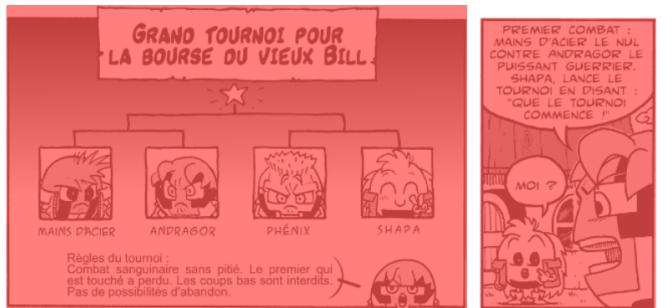
Content-driven approach Panel extraction



1) Original image

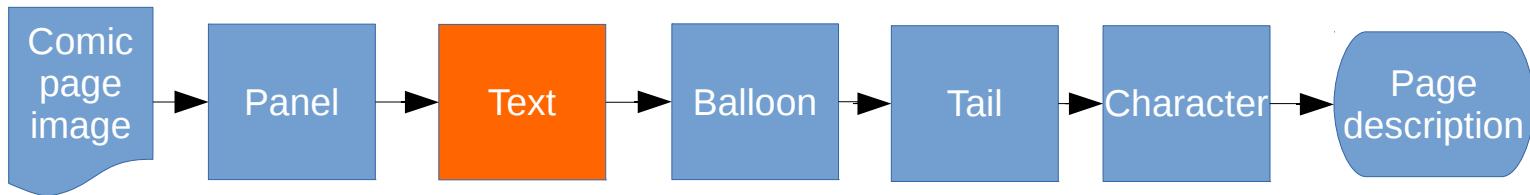


2) Pixel regions



3) Panel boxes

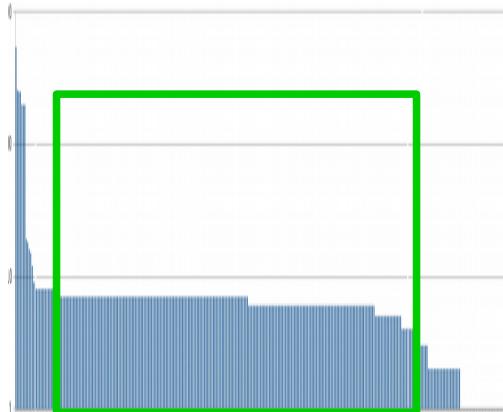
Content-driven approach Text extraction



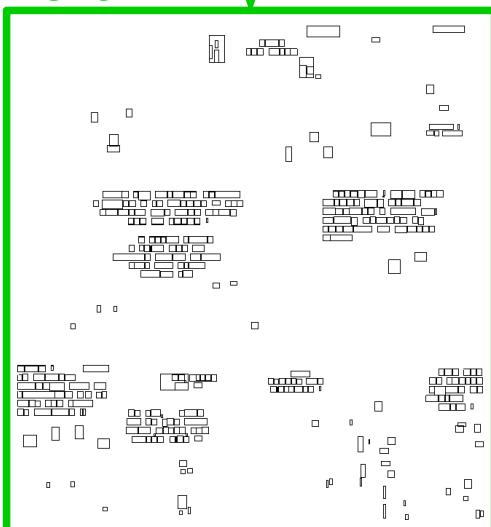
↗ Challenges

- > Non-standard fonts
- > Multi-script/orientation/scale
- > Complex background (sound effects)
- > Hyphenation, voluntary spelling mistakes

Content-driven approach Text extraction



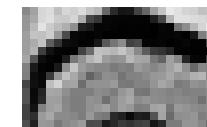
Text?



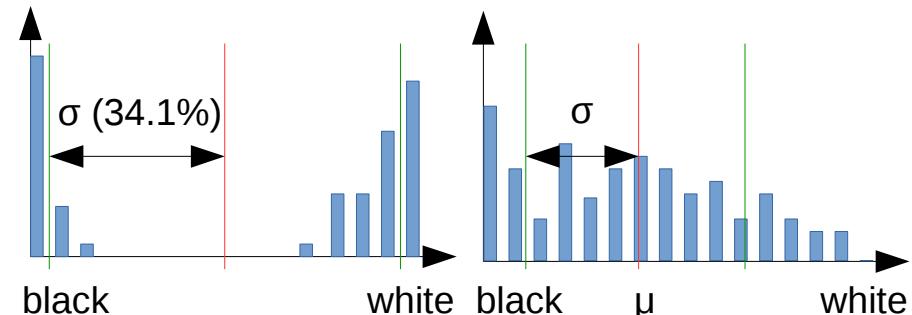
Contrast filtering



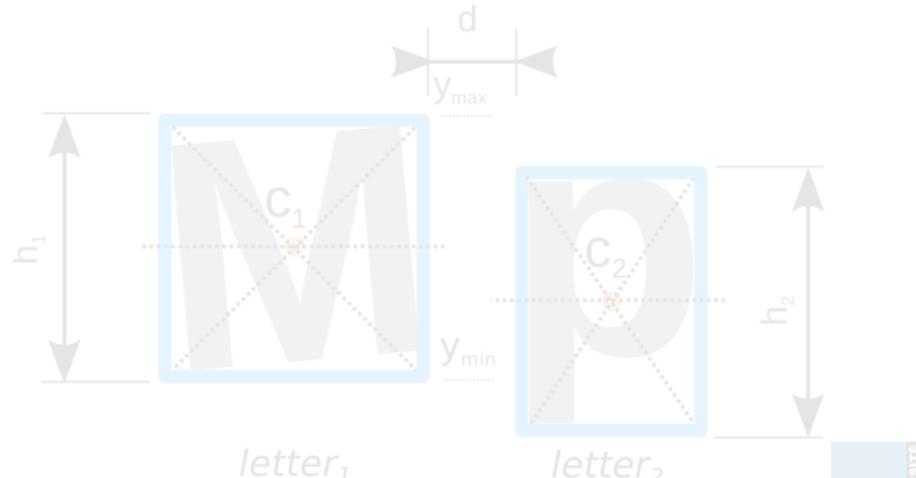
$\mu, \sigma = [167, 84]$



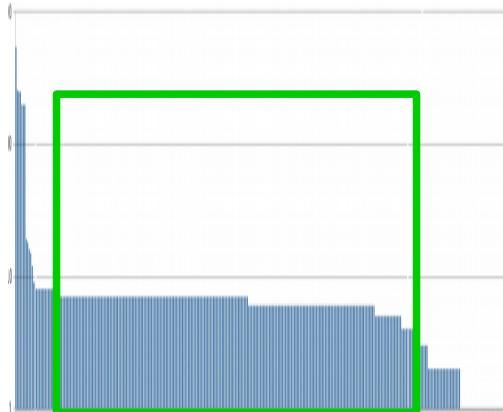
$\mu, \sigma = [110, 57]$



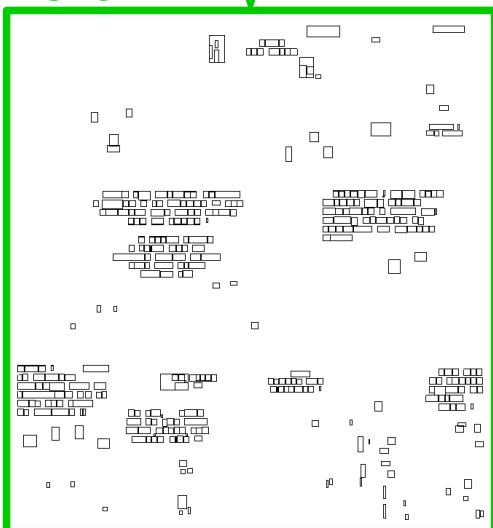
Alignment filtering



Content-driven approach Text extraction



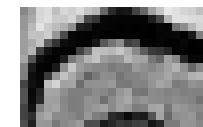
Text?



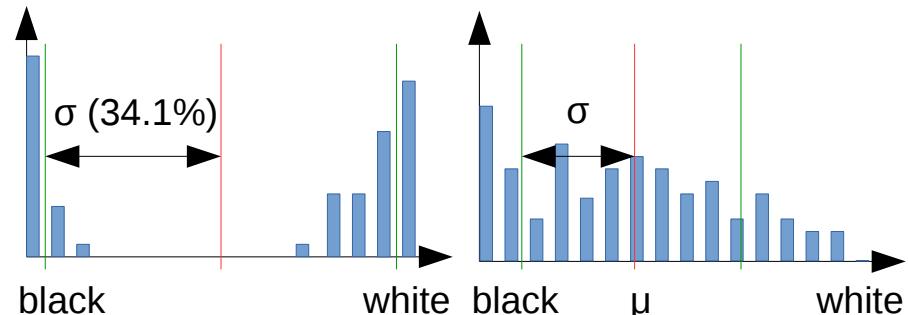
Contrast filtering



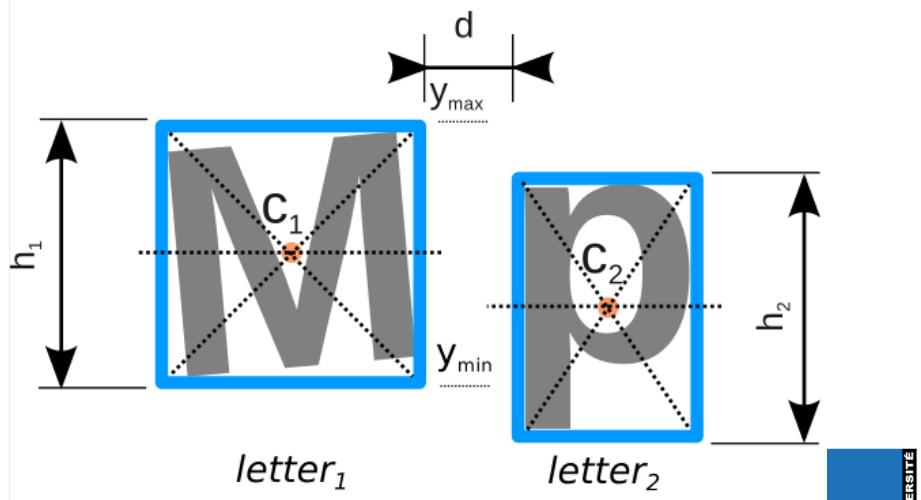
$\mu, \sigma = [167, 84]$



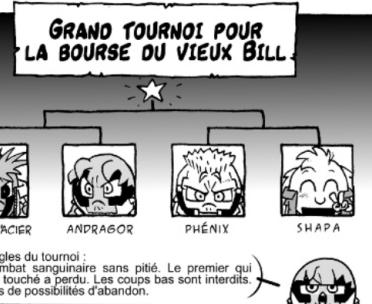
$\mu, \sigma = [110, 57]$



Alignment filtering



Content-driven approach Text extraction



1) Original image

2) Text positions

3) Transcription

GRANO TOURNOI POUR LA BOURSE DU VIEUX BILL

MAINSDACIER ANDRAGOR
PHENIY SHAPA

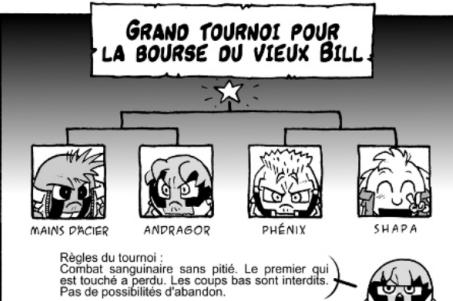
Regles du tournoi
Combat sanguinaire sans pitié. Le premier qui est touché a perdu. Les coups bas sont interdits Pas de possibilités d'abandon

PREMIER COMBAT MAINS Dacierr LE NUL CONTRE ANDRAGOR LE PUSSANT GUERRIER SHAPA, LANCE LE TOURNOI EN DISANT QUE LE TOURNOI COMMENCE

BON BEN QUE LE COMBAT COMMENCE

MAIS QUEST CE QUE TU ME FAIS LA ? C'EST PAS LE LANCEMENT DE LA FETE DE LA MORUE ! TU DOIS CRIER, AVOIR LE PHYLACTERE QUI SE HERISSE, AVEC DES EFFETS DE VITESSE, ET TOUT CA !!

Content-driven approach Text extraction



1) Original image



2) Text positions

GRANO TOURNOI POUR LA BOURSE DU VIEUX BILL

MAINS DACIER ANDRAGOR PHENIY SHAPA

Regles du tournoi

Combat sanguinaire sans pitié. Le premier qui est touché a perdu. Les coups bas sont interdits Pas de possibilités d'abandon

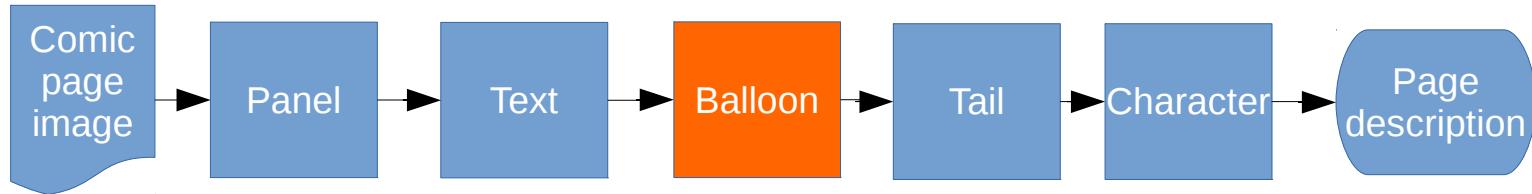
PREMIER COMBAT MAINS DACIER R LE NUL CONTRE ANDRAGOR LE PUISSANT GUERRIER SHAPA, LANCE LE TOURNOI EN DISANT QUE LE TOURNOI COMMENCE

BON BEN QUE LE COMBAT COMMENCE

MAIS QUEST CE QUE TU ME FAIS, LA ? C'EST PAS LE LANCEMENT DE LA FETE DE LA MORUE TU DOIS CRIER, AVOIR LE PHYLACTERE QUI SE HERISSE, AVEC DES EFFETS DE VITESSE, ET TOUT CA !

3) Transcription

Content-driven approach Balloon extraction



↗ Challenges

- > Difference between **shape** and **contour**
- > Implicit **balloon** positions
- > **Semantics** related to text

↗ Observation

- > Balloons contain **centered text** and are mainly **convex**



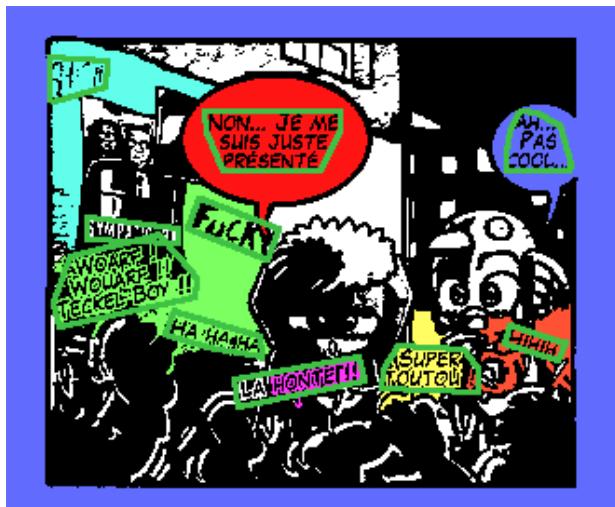
Content-driven approach Balloon extraction



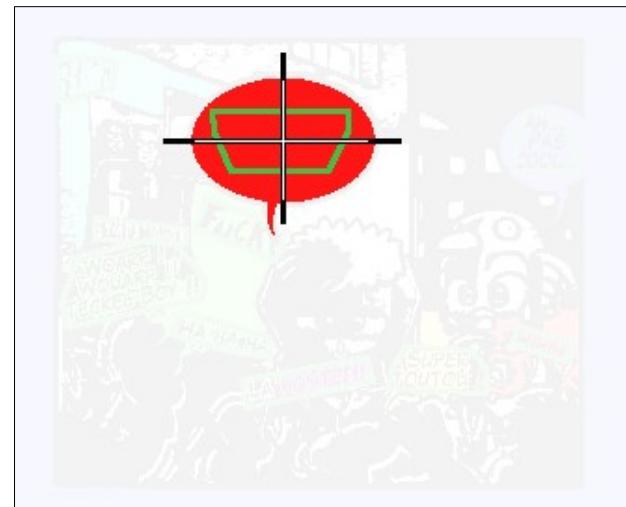
1) Original image



2) Text block positions (green)

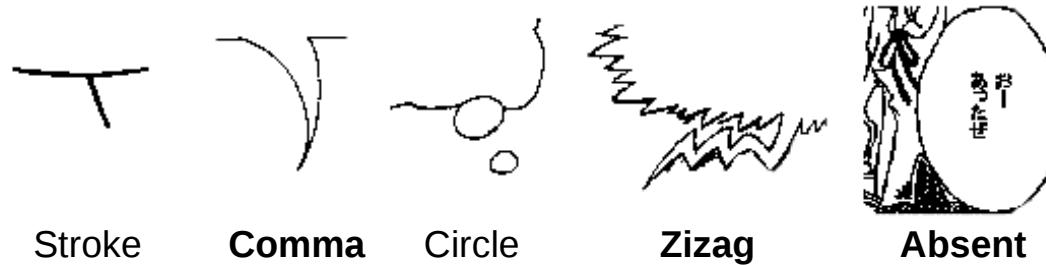
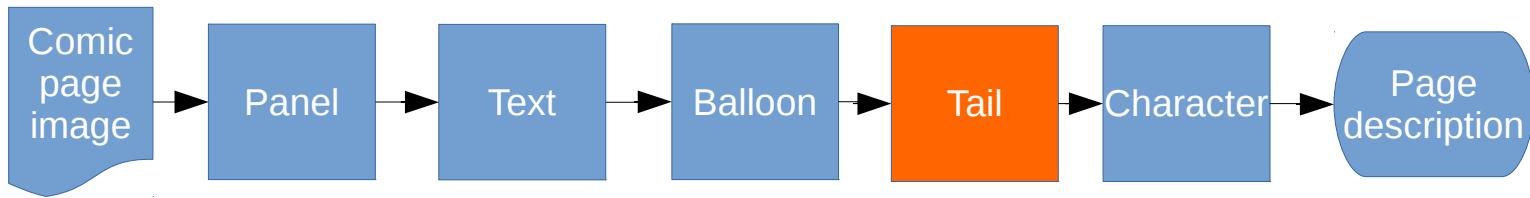


3) Regions including text blocks (coloured)



4) Convex regions with aligned text

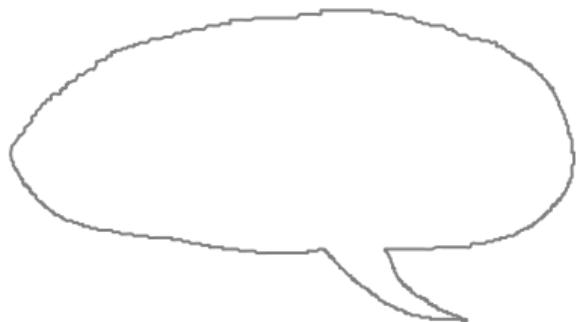
Content-driven approach **Tail tip localisation**



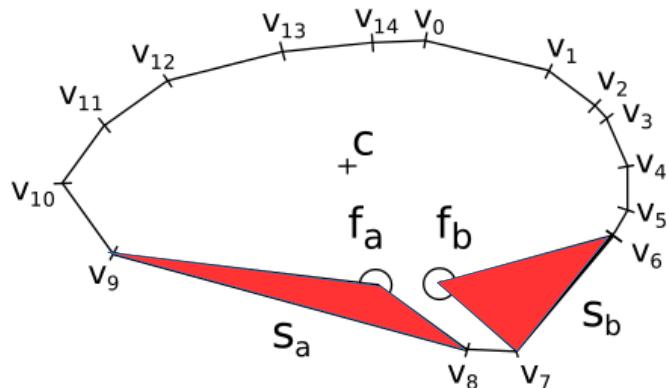
↗ Challenges

- > Shape variation
- > Tiny details

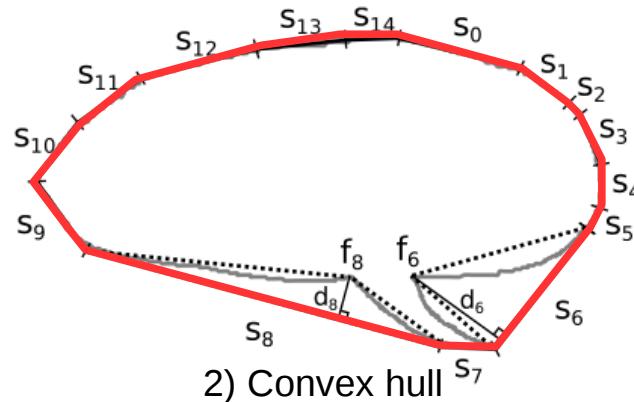
Content-driven approach Tail tip localisation



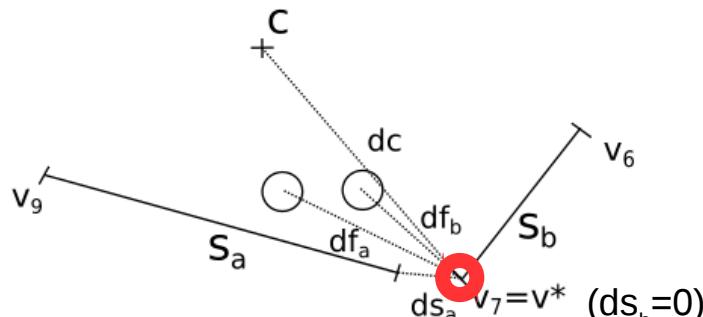
1) Balloon contour



3) Two biggest
convexity defects



2) Convex hull

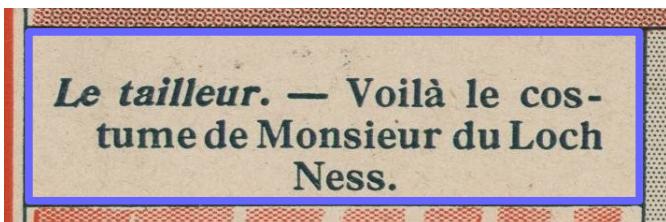


4) Tail tip position

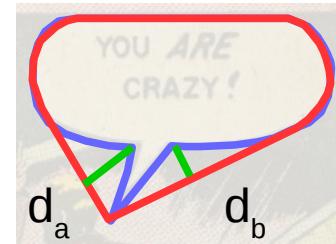
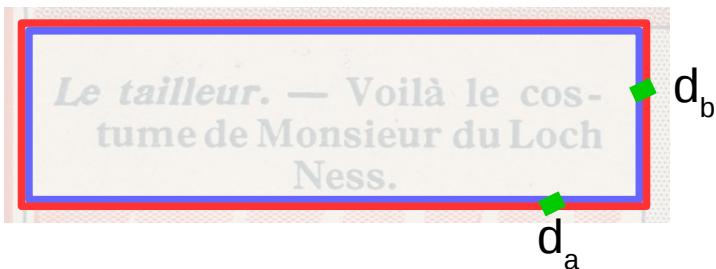
Optimal vertex selection:
 $v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$

Content-driven approach **Tail tip localisation**

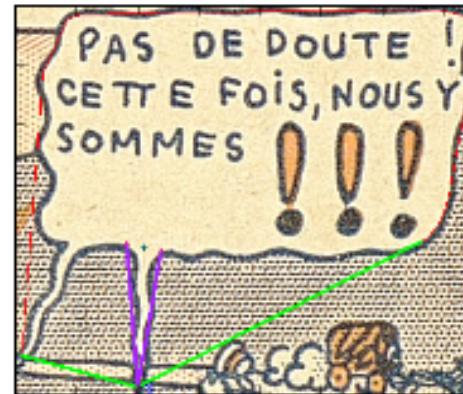
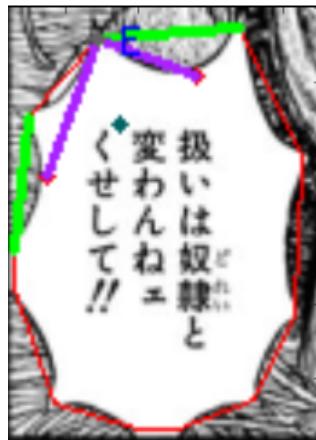
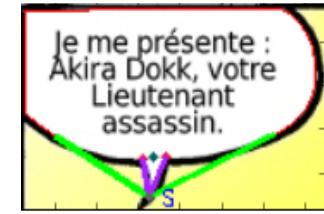
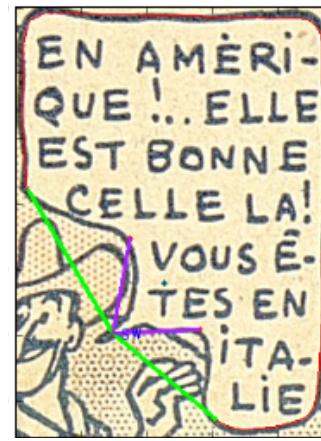
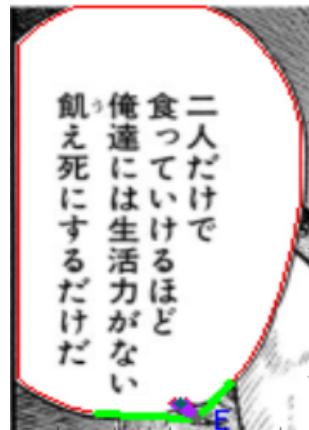
Balloon 1



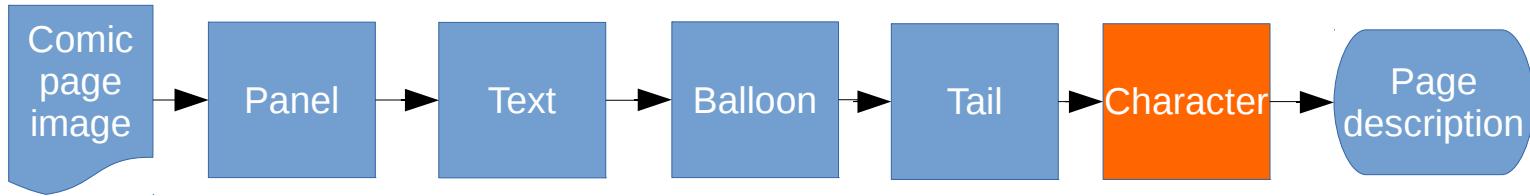
Balloon 2



Content-driven approach Tail tip localisation



Content-driven approach Comic character extraction



↗ Challenges

- > Hand-drawn and stroke-based
- > Human/invented like
- > Scale, deformation, posture, occlusion variations

Content-driven approach **Comic character extraction**

Approach 1: ROI estimation

Panels + Tails = Comic character ROIs



Content-driven approach Comic character extraction

Approach 2: subgraph matching

1) Panel to graph



2) Similar subgraph retrieval



Content-driven approach Comic character extraction

Approach 2: subgraph matching

1) Panel to graph

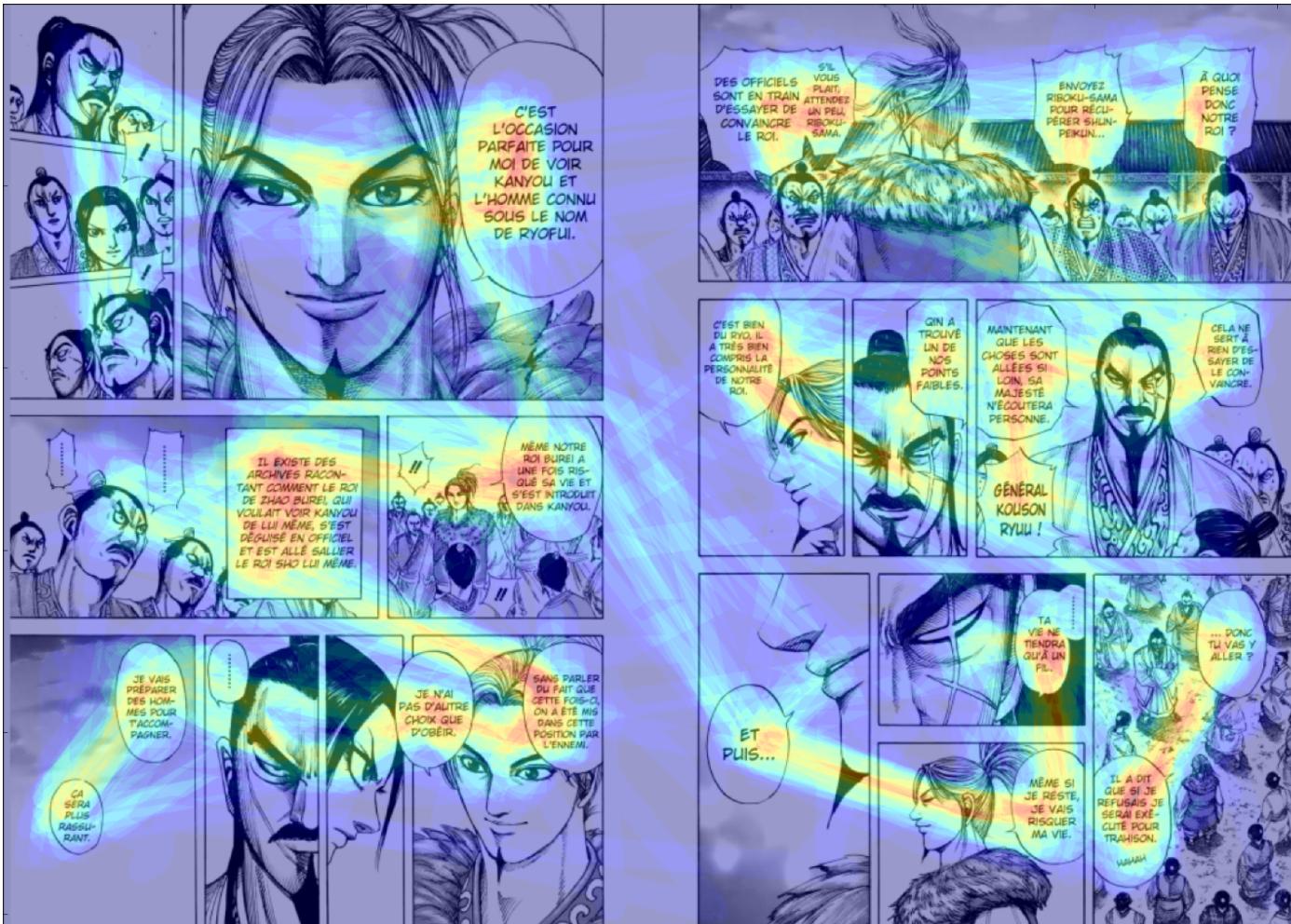
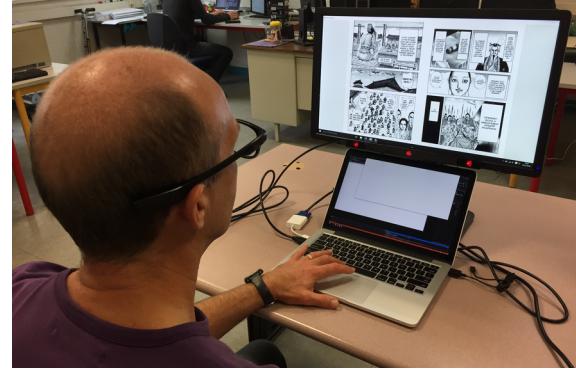


2) Similar subgraph retrieval



Content-driven approach Comic character extraction

Approach 3: eye tracking



Content-driven approach **Comic character extraction**

Approach 4: deep learning

↗ Data preparation

- > 12206 images of “human, cat, cow, dog, horse” in Pascal dataset
- > eBDthèque dataset => 537 panels with characters
- > + 4 private albums (4AS, Yoko, Asterix, Midam) => 426 panels with characters
- > Categories: human, human-like, near-humain-like, far-humain-like, humain-like-sketch, object-like

Content-driven approach Comic character extraction

Approach 4: deep learning

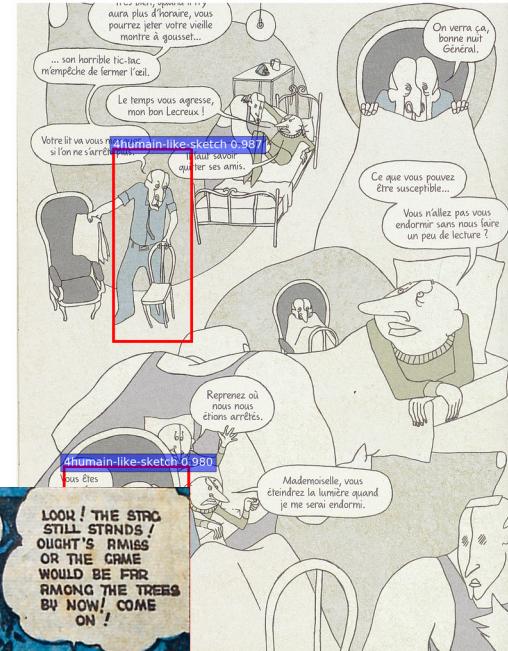
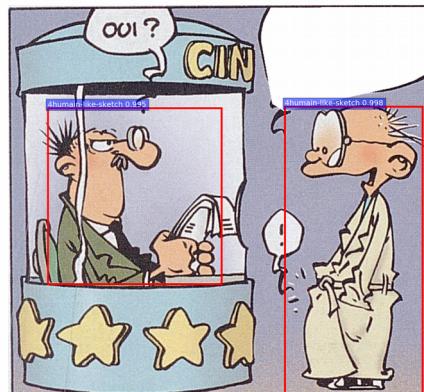
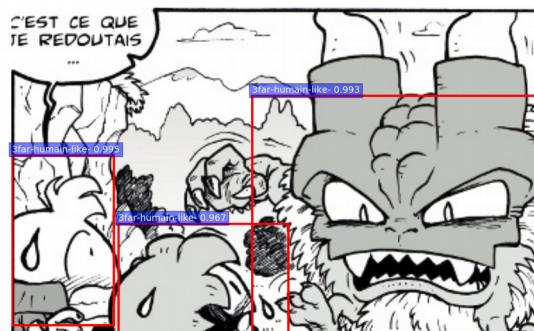
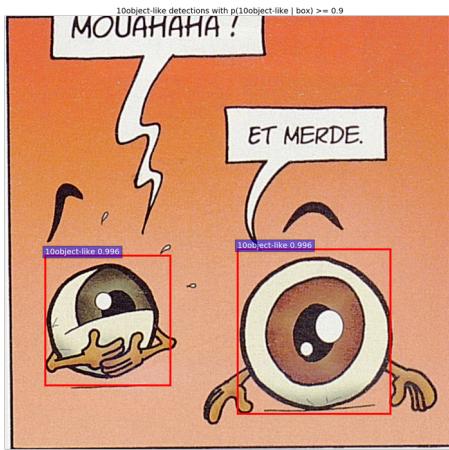
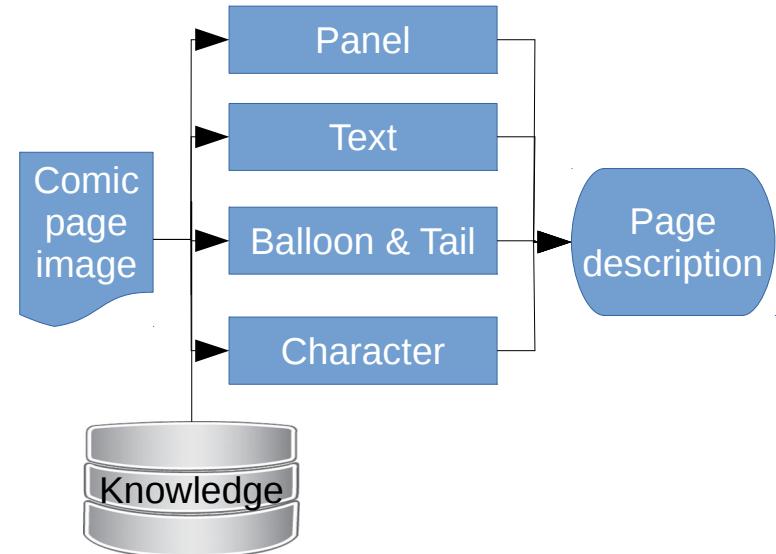
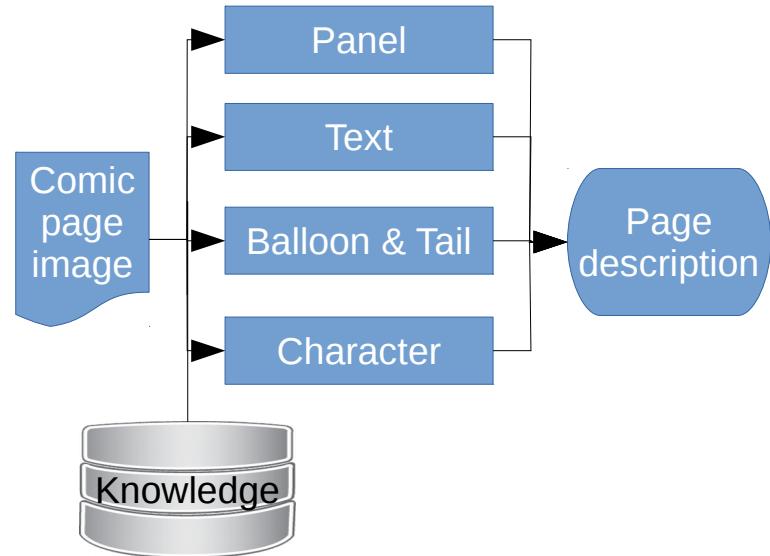


Image analysis Knowledge-driven approach

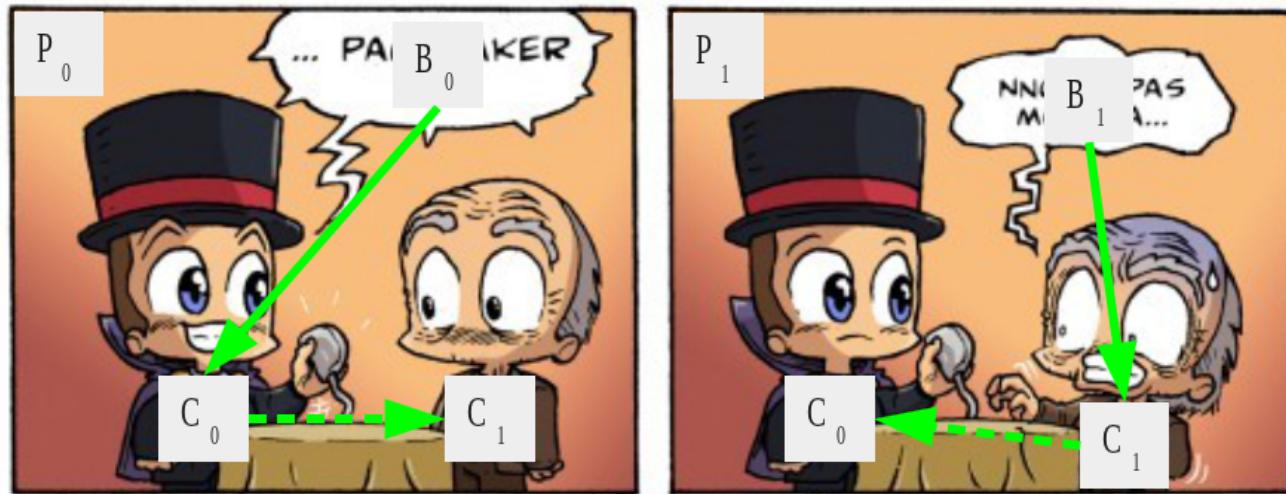


Knowledge-driven approach Introduction



➤ Challenges

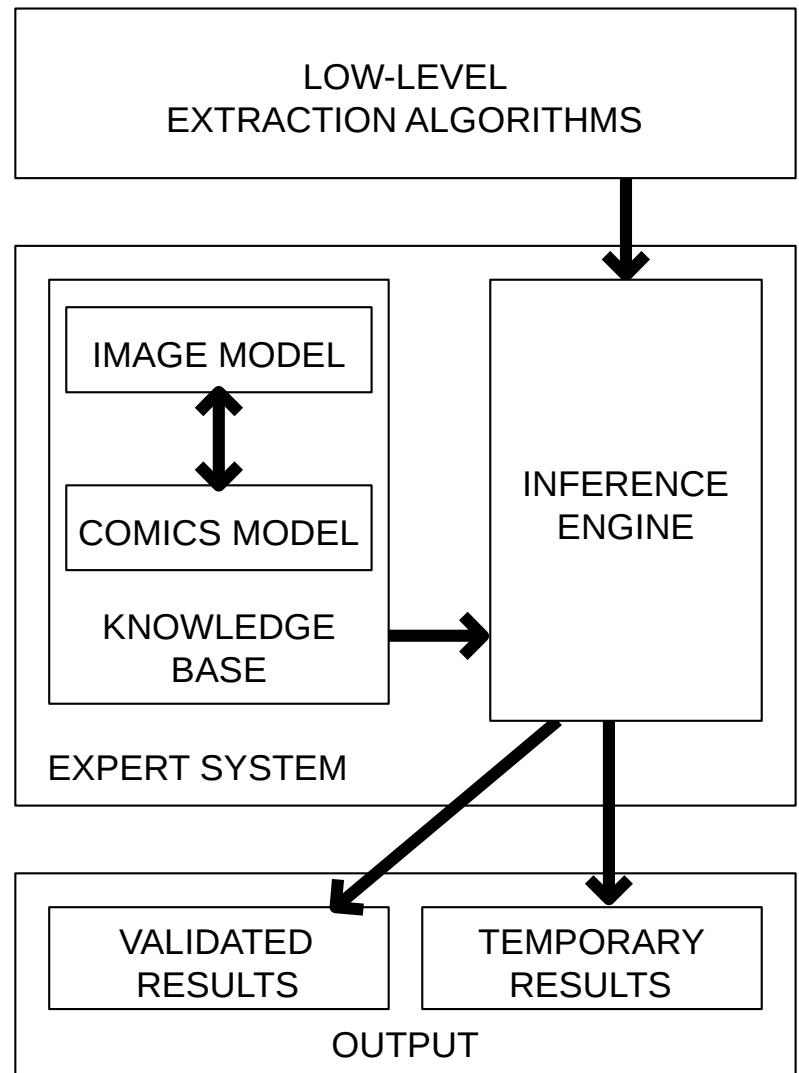
- > Increase overall precision
- > Framework for automatic comics understanding



Knowledge-driven approach

Knowledge representation

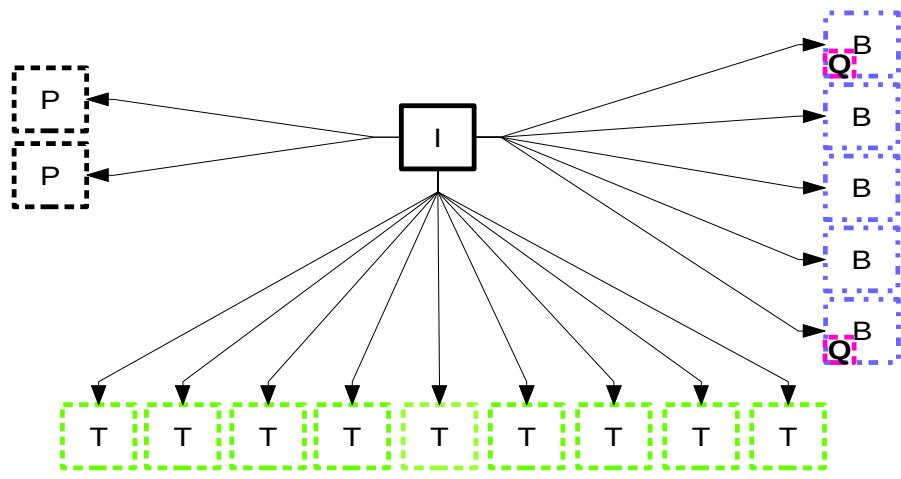
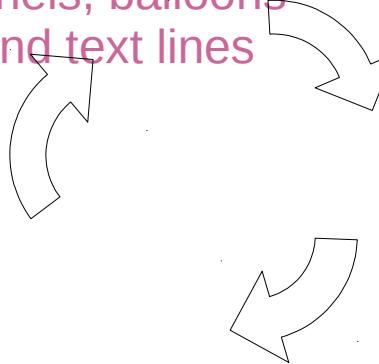
- Image model
 - > Physical support
 - > Regions of interest
- Comics model
 - > Validations
 - A **panel P** is related to one page
 - A **balloon B** is related to one panel and may have a **tail Q**
 - A **character C** is related to one panel
 - A **text line T** is related to one balloon
 - > Inferences
 - **B + Q + T => speech balloon SB**
 - **SB + T => speech text ST**
 - **SB + C => speaking character SC**



Knowledge-driven approach Processing sequence



Hypotheses of
panels, balloons
and text lines



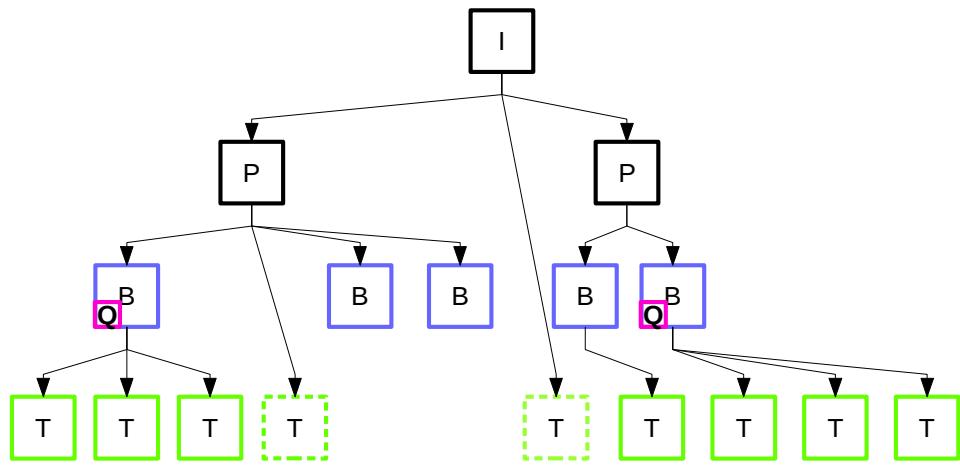
	Non valid	Valid
P	[P]	Panel
B	[B]	Balloon
Q	[Q]	Tail
T	[T]	Text

Knowledge-driven approach Processing sequence



Hypotheses of
panels, balloons
and text lines

Validation of
the
hypotheses



Non valid
valid



Panel



Balloon



Tail



Text

Knowledge-driven approach Processing sequence

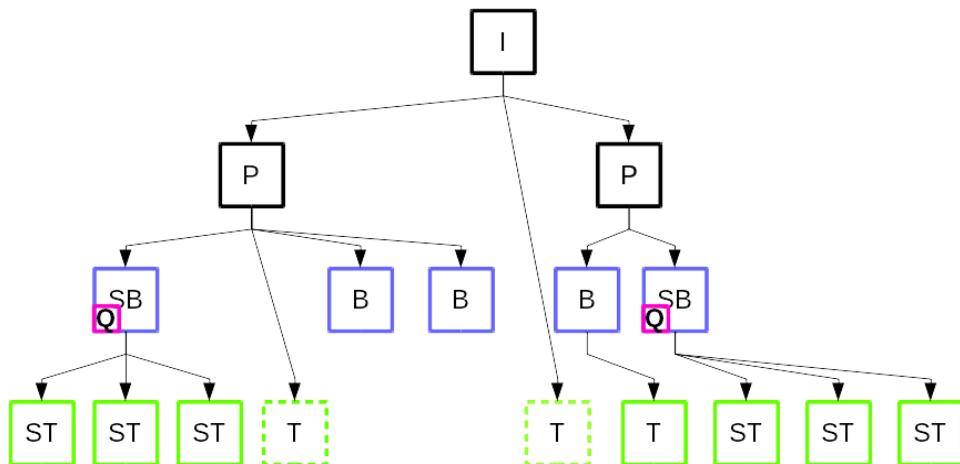


Hypotheses of panels, balloons and text lines



Validation of the hypotheses

Inferences of specific types



Non valid
valid



Panel



Speech balloon



Tail



Speech text

Knowledge-driven approach Processing sequence

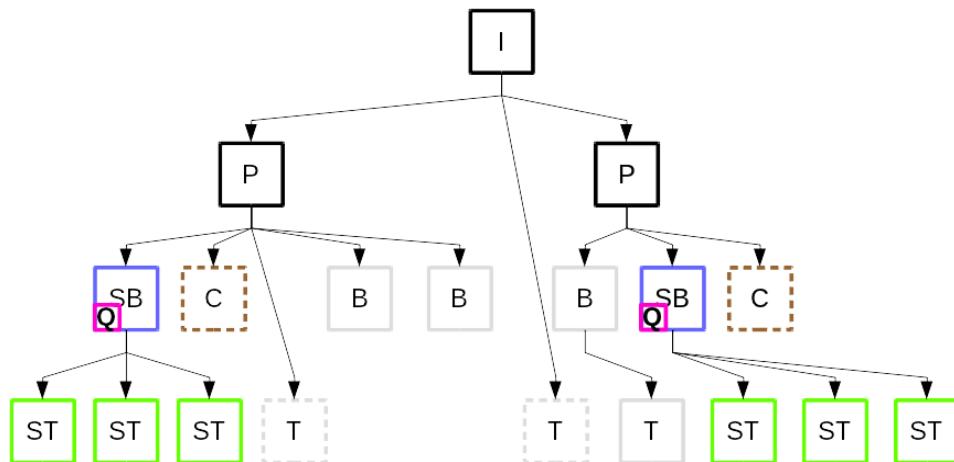


Hypotheses of
comic
characters

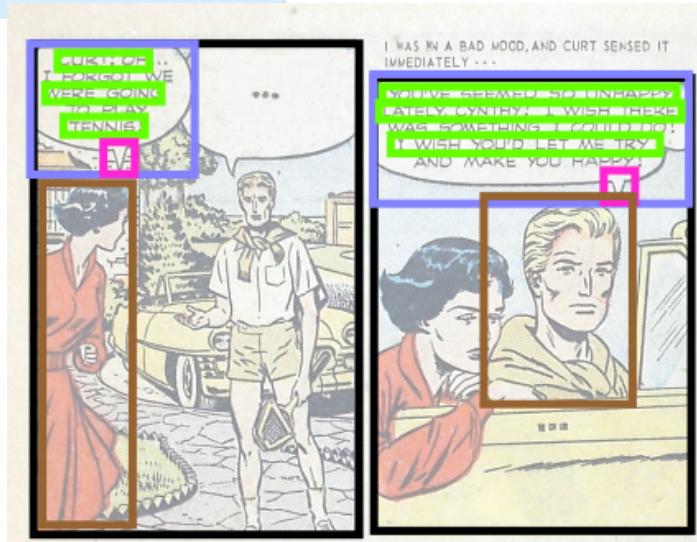
Inferences of
specific types

Validation of
the
hypotheses

Non valid
valid
 C C Comic character



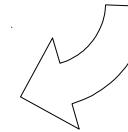
Knowledge-driven approach Processing sequence



Hypotheses of
comic
characters



Validation of
the
hypotheses

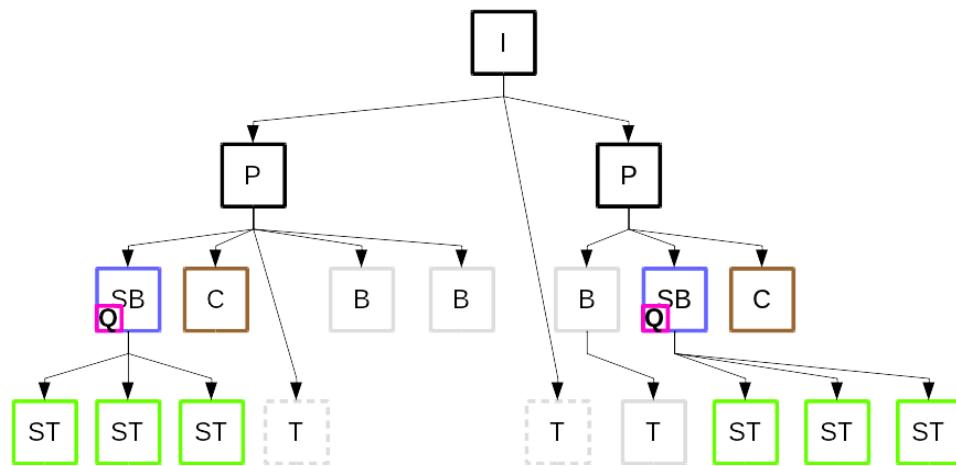


Inferences of
specific types

Non valid
valid



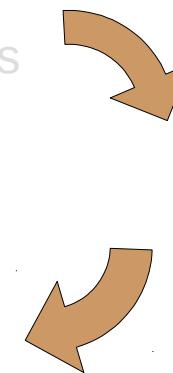
Comic character



Knowledge-driven approach Processing sequence



Hypotheses of
comic
characters



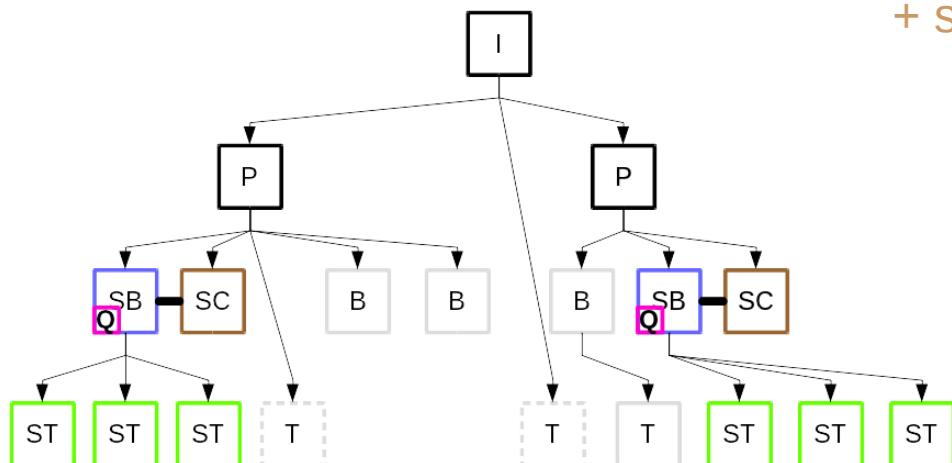
Validation of
the
hypotheses

Inferences of
specific types
+ semantic links

Non valid
valid



Speaking
comic character



Conclusions

- > Efficient **panel**, **balloon**, **text** and **tail** extraction methods
 - > Base of a **framework** for automatic comics understanding
 - > Public **dataset** (<http://ebdtheque.univ-lr.fr>)
-
- > Improvement of automatic text recognition
 - > Extraction of more non-human-like comic characters
 - > **Scene** and **action/interaction** recognition

Participants

↗ Professors

> Karell Bertet, Jean-Christophe Burie, Arnaud Revel, Jean-Marc Ogier

↗ PhD candidates

> Nam Le Thanh

↗ Post-doc

> Clément Guérin, Christophe Rigaud, Srikanta Pal, Van Nhu Nguyen



Christophe Rigaud, Clément Guérin, Dimosthenis Karatzas, Jean-Christophe Burie and Jean-Marc Ogier. “**Knowledge-driven understanding of images in comic books**”. International Journal on Document Analysis and Recognition (IJDAR), Volume 18, 2015, Pages 199-221 ([link](#))

Clément Guérin, Christophe Rigaud, Karelle Bertet, Arnaud Revel, **An ontology-based framework for the automated analysis and interpretation of comic books’ images**, Information Sciences, Volume 378, 1 February 2017, Pages 109-130 ([link](#))