



## European Ph.D. defense

Communauté  
d'Agglomération de  
**La Rochelle**

# Segmentation and indexation of complex objects in comic book images

Christophe Rigaud  
December 11<sup>th</sup>, 2014

Co-supervised by:

Jean-Christophe Burie<sup>1</sup>  
Dimosthenis Karatzas<sup>2</sup>  
Jean-Marc Ogier<sup>1</sup>

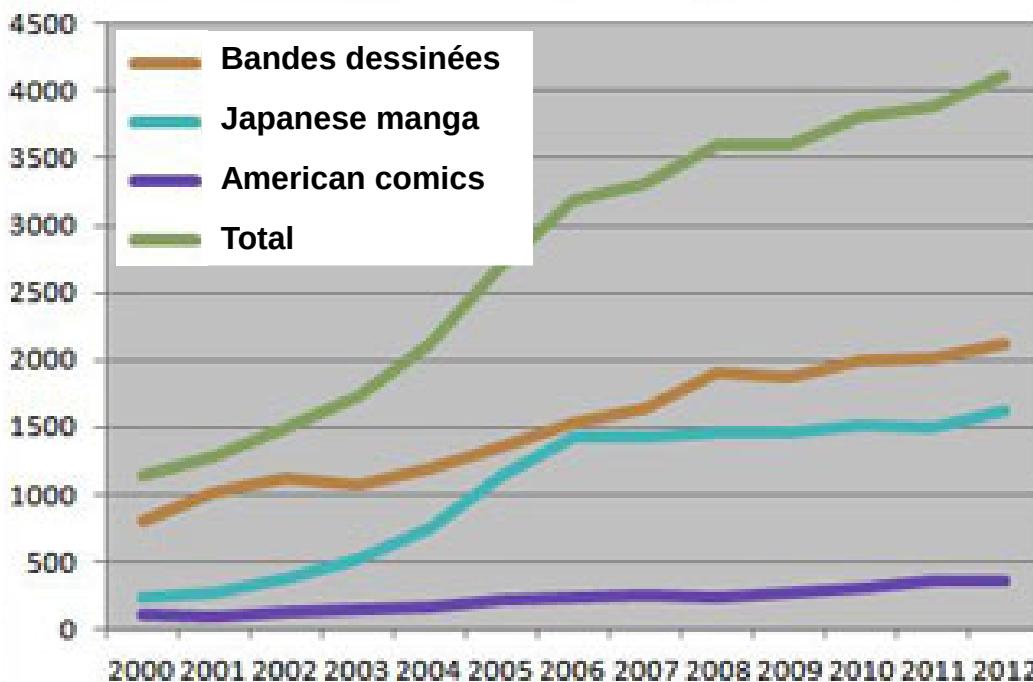
# Comic books

## Introduction

*“a visual medium used to express ideas via images, often combined with text or visual information”*

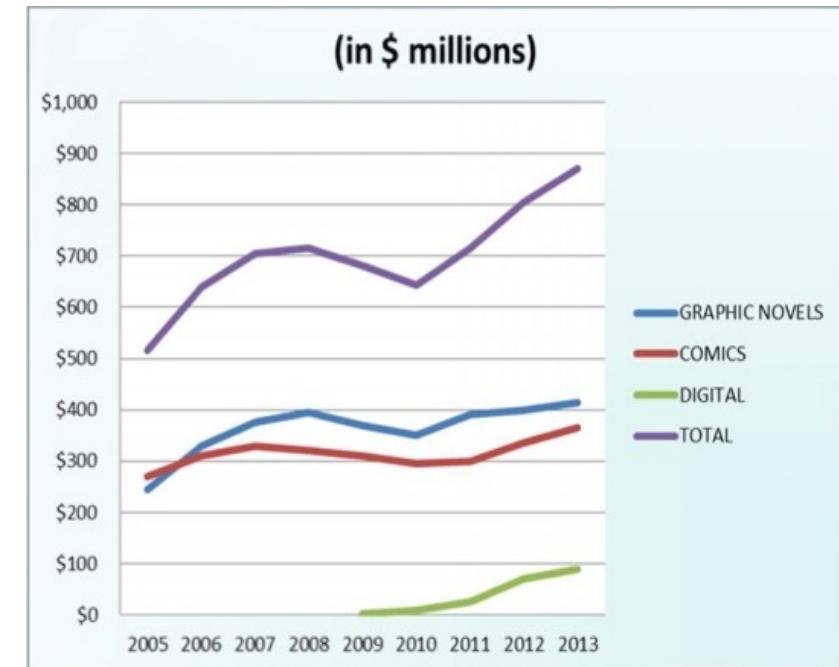
Wikipédia, 2014

One of the most **popular** and **familiar** forms of **graphic content**



Francophone comics production

Infographie (c) L'Agence BD d'après les chiffres de Gilles Ratier/ACBD.



Comics market in the US

Milton Griepp's White Paper, ICv2 Conference 2014

- **TODO**

Problematic

COMICS

- Fast evolution
- Massive digitization
- Digital images contain poor information (pixel position + colour)
- Need human-readable description to be searched and retrieved by humans
- but...
- Cost is too high if done by hand

SOLUTION: COMPUTER SCIENTISTS

- Provide methods for generating automatic description of digital images

SCIENTIFIC ISSUES RELATED TO COMICS

- Locate and identify image content → low level description
- Understand image content to produce → high level description

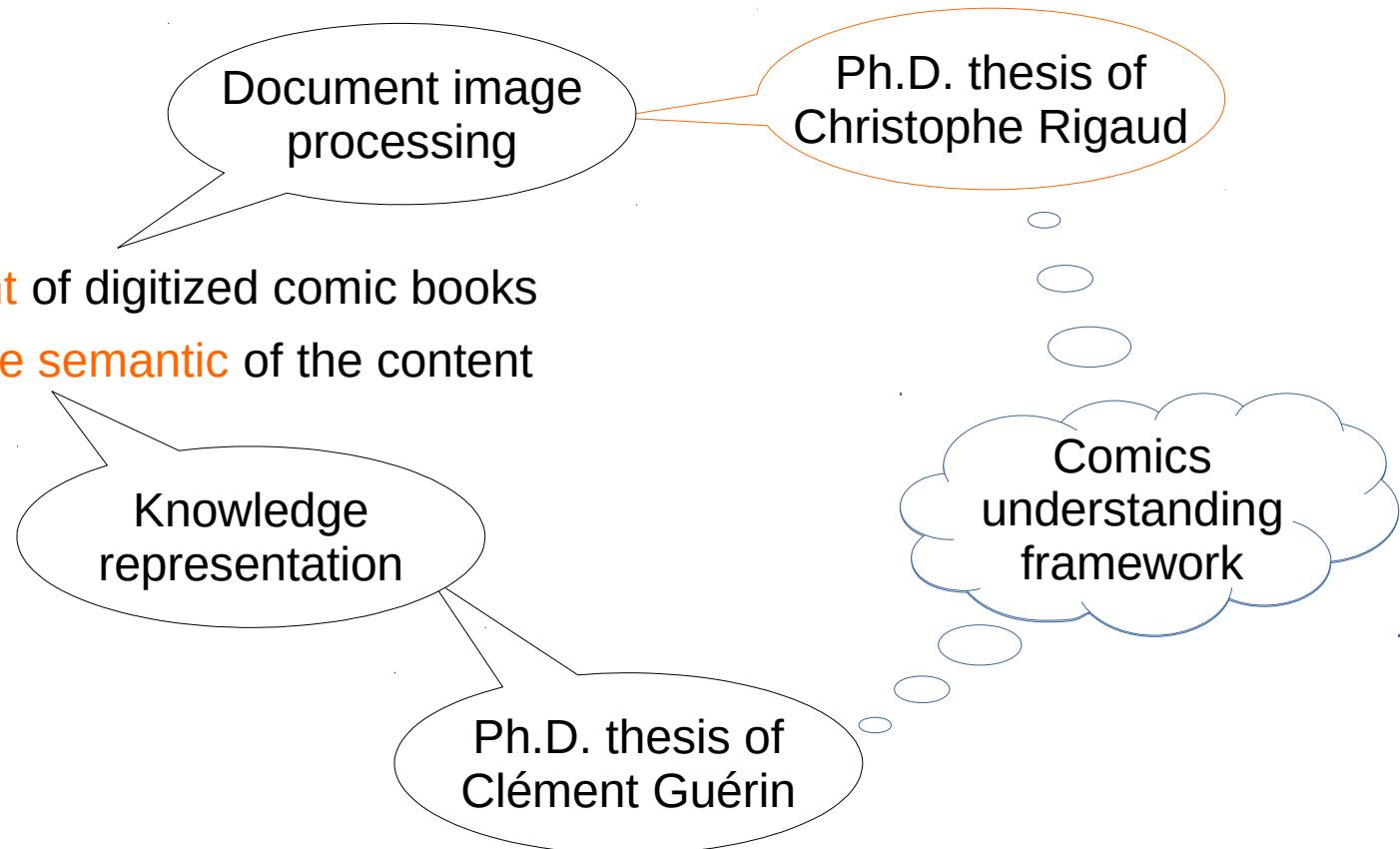
# Solution from L3i lab

## Introduction

- eBDthèque project (since 2011)
  - Add value to paper-based comics using the new technologies
  - Public founding CPER 2007-2013
  - 2 Ph.D. students, 1 engineer, 1 post doc
  - 6 professors

- Research axes

- Extracting content of digitized comic books
  - Understanding the semantic of the content



- Document production to interpretation
- Application to comic books



Pencil drawing. Image credits: Le cycle des bulles, Christophe Rigaud, 2012

# Production to interpretation

# Document analysis

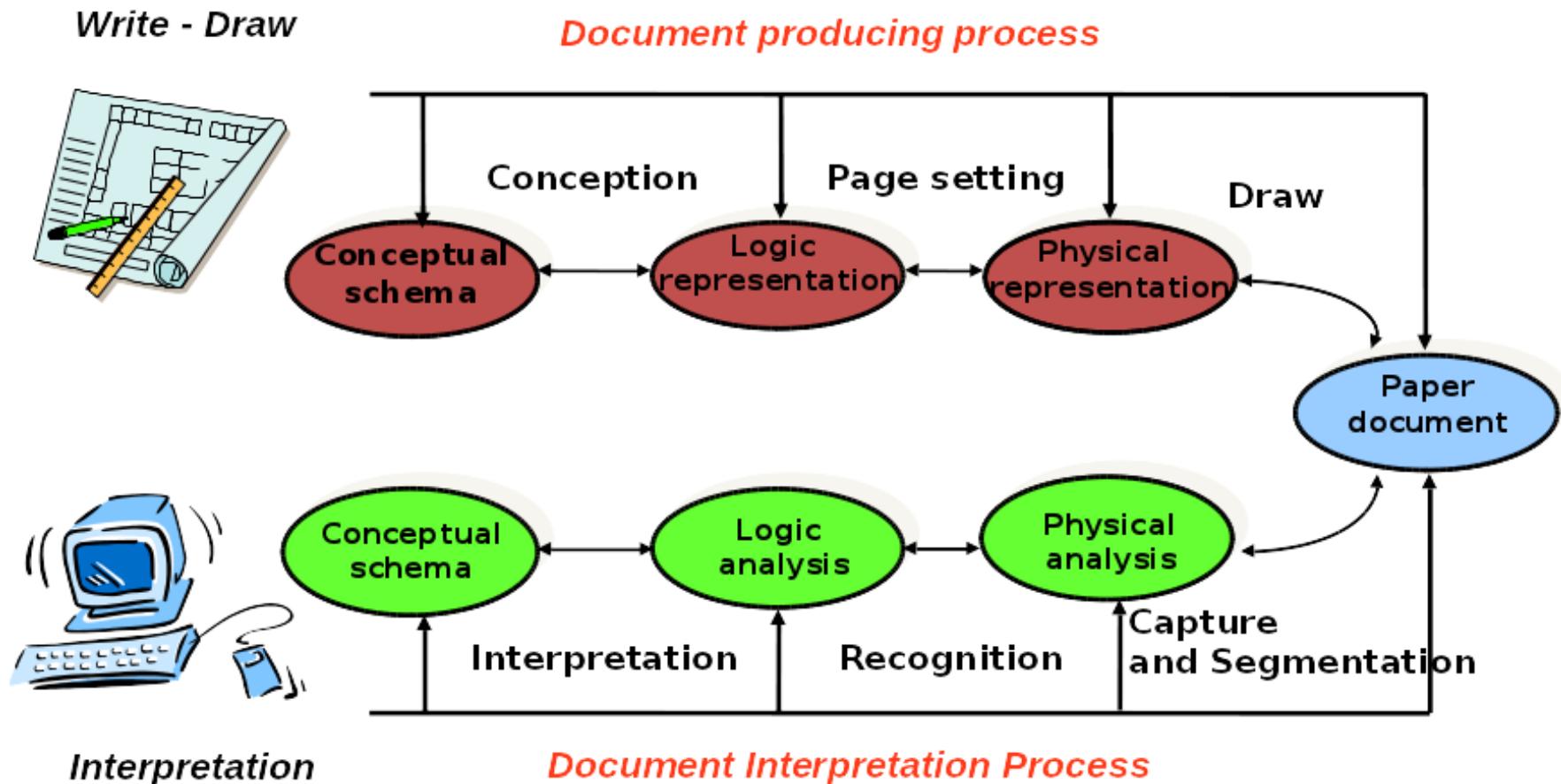


Image source: Handbook of Document Image Processing and Recognition. Springer, 2014

# Application to comic books

## Document analysis

- Challenges:
  - Recent field of research with a **largely unknown**
  - **Semi-structured** and **free-form** document mixing text and graphics

- Production process
  - 1) Synopsis and storyboarding
  - 2) Pencil drawing
  - 3) Inking
  - 4) Flattening and coloring
  - 5) Lettering and sound effects

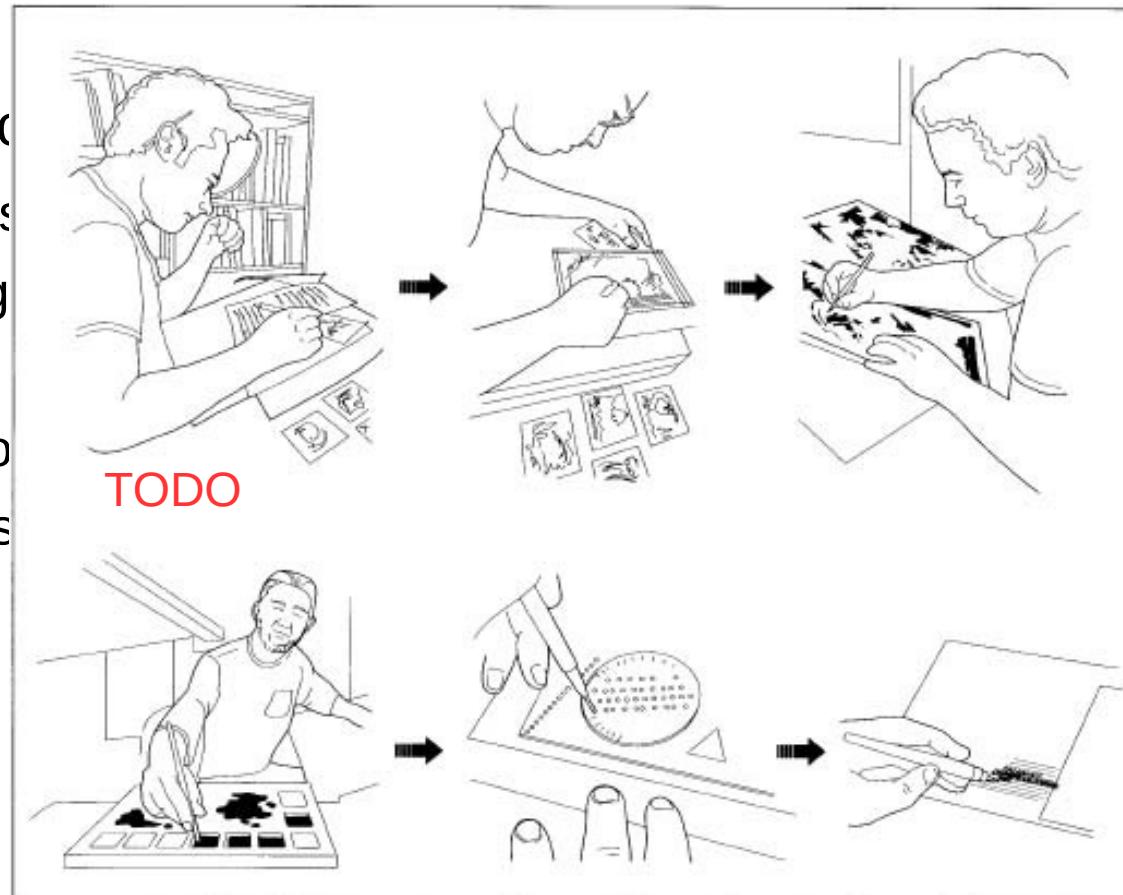
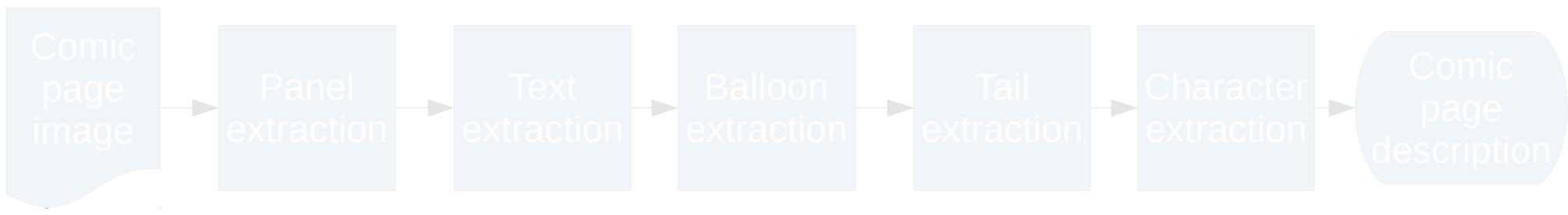
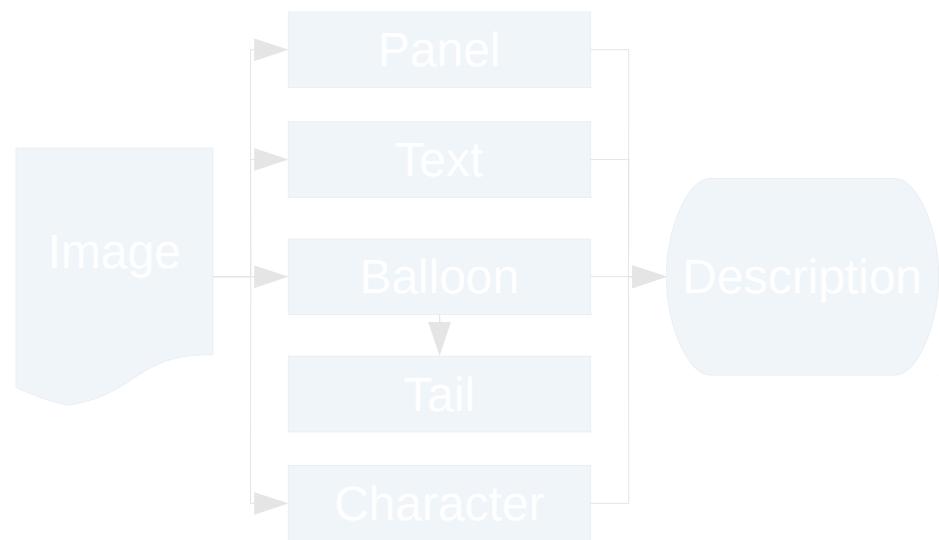


Image source:  
<http://www.madehow.com/Volume-6/Comic-Book.html>

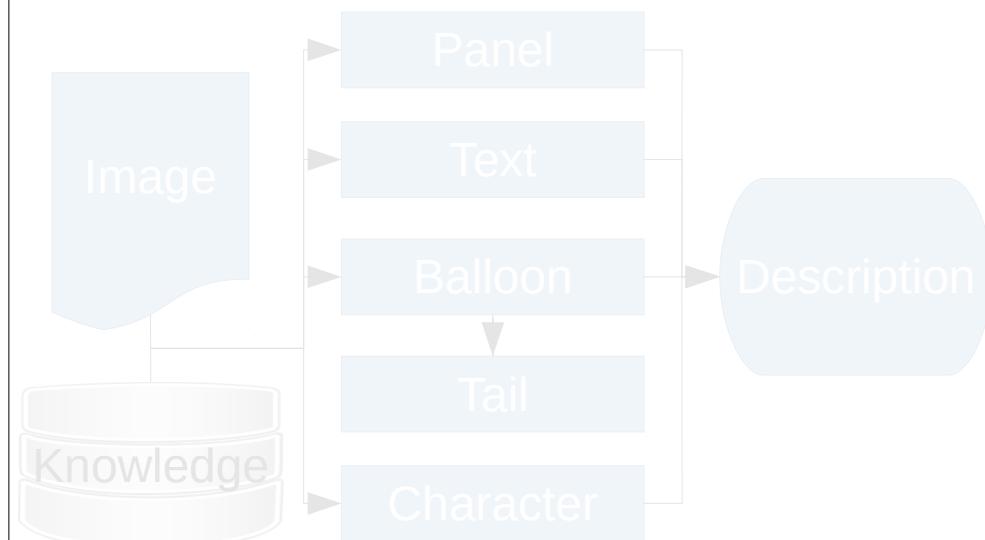
## Content-driven (sequential)



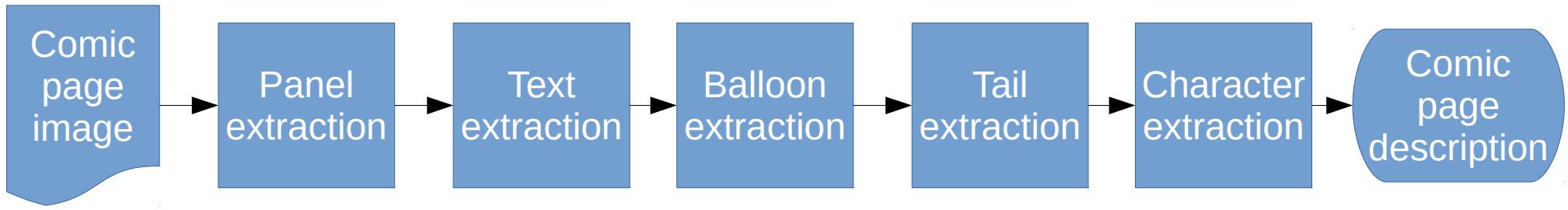
## Content-driven (independent)



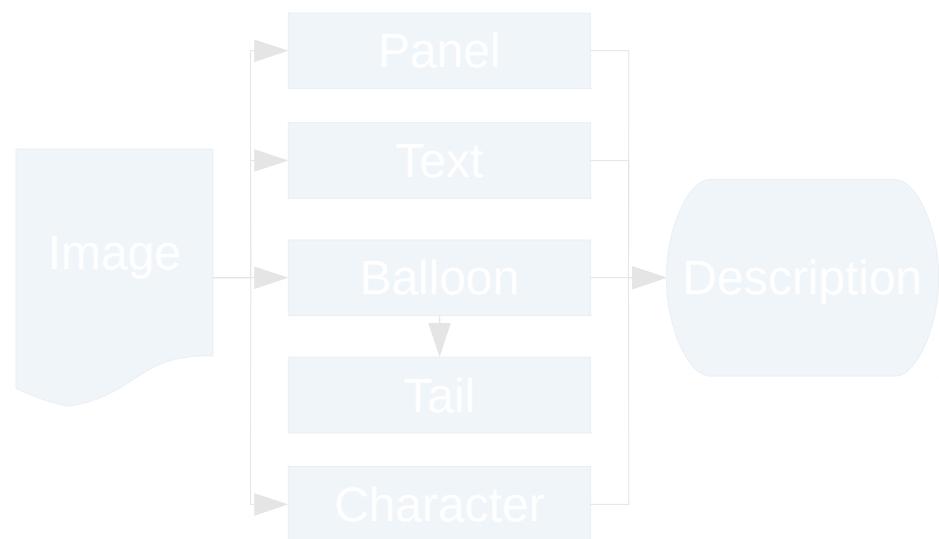
## Knowledge-driven (independent)



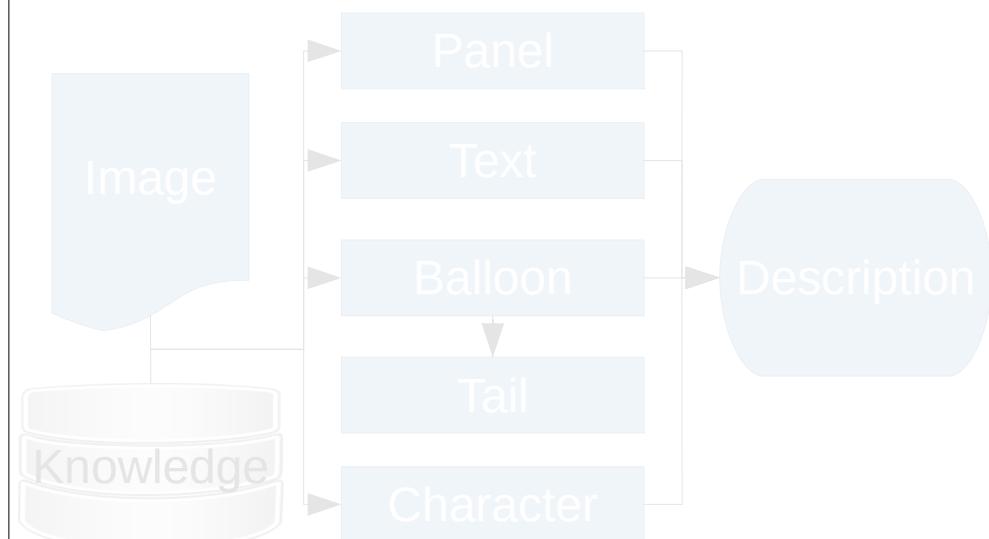
## Content-driven (sequential)



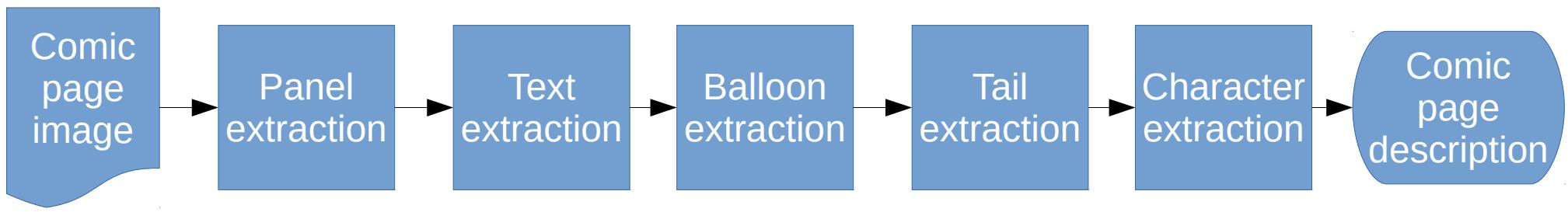
## Content-driven (independent)



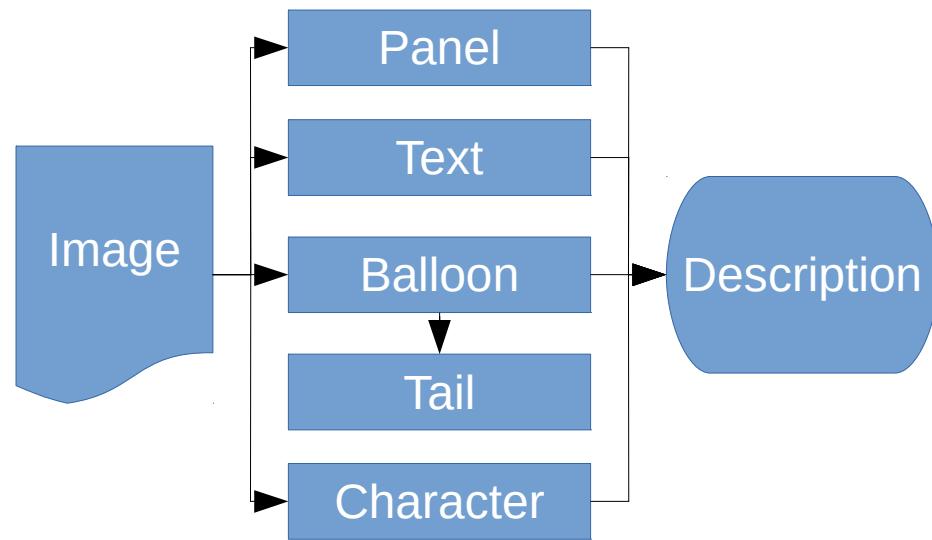
## Knowledge-driven (independent)



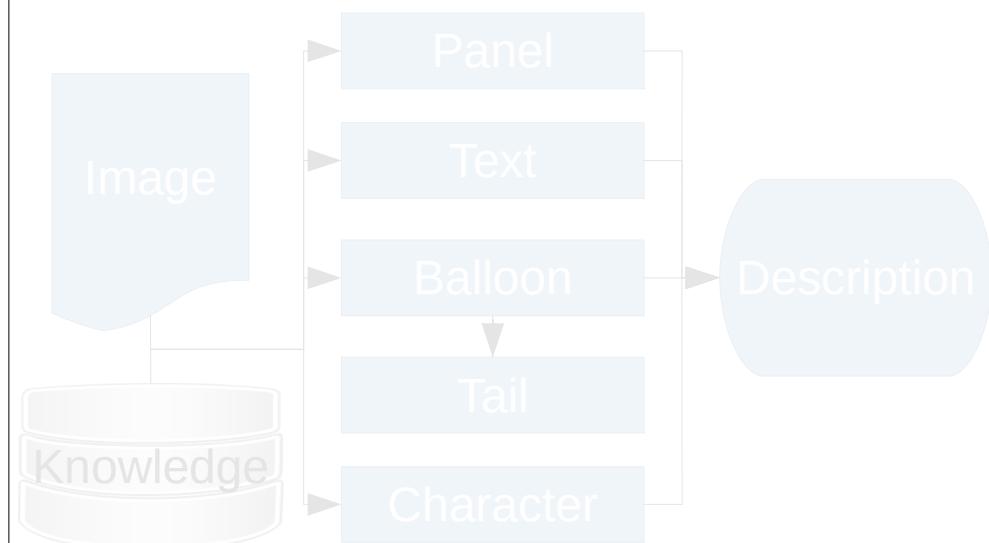
## Content-driven (sequential)



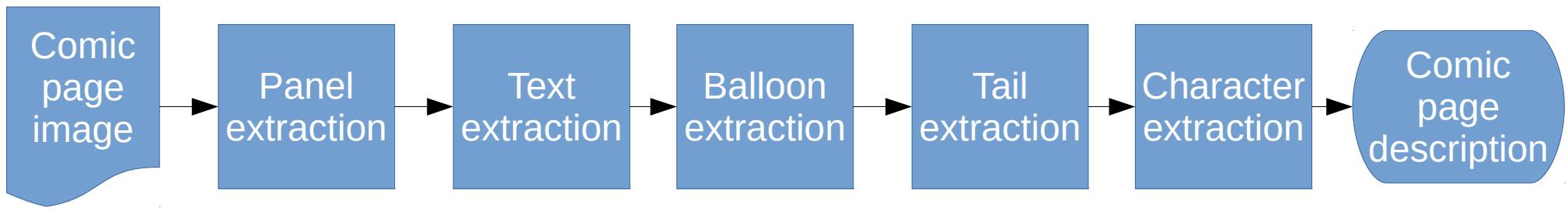
## Content-driven (independent)



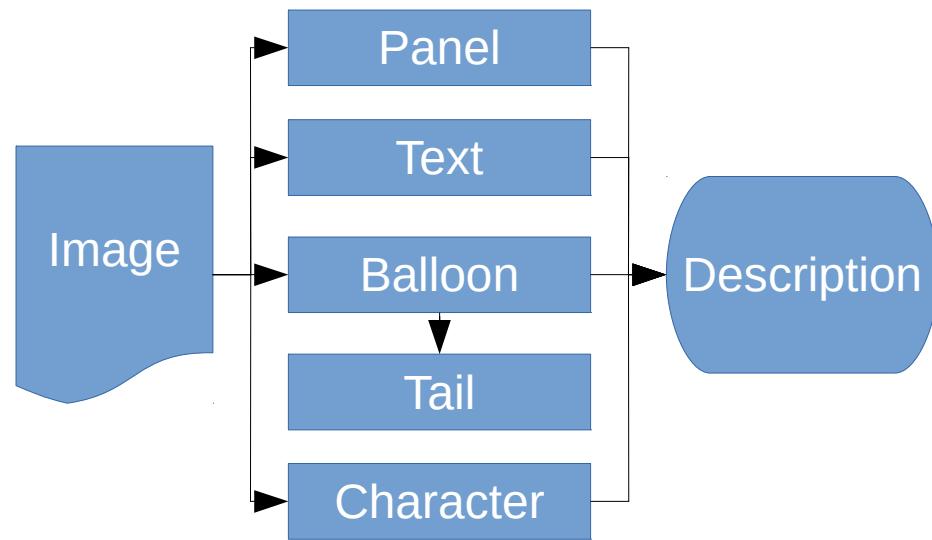
## Knowledge-driven (independent)



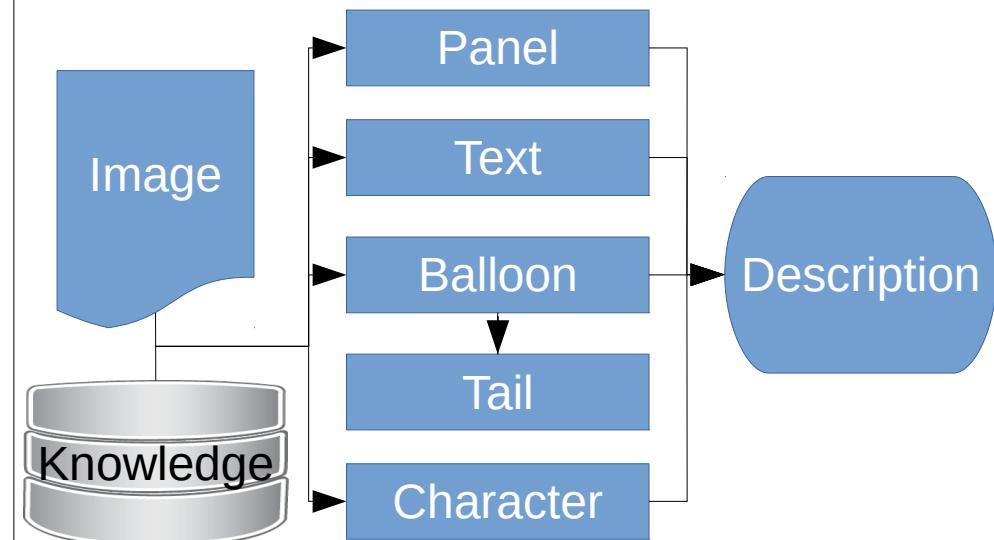
## Content-driven (sequential)



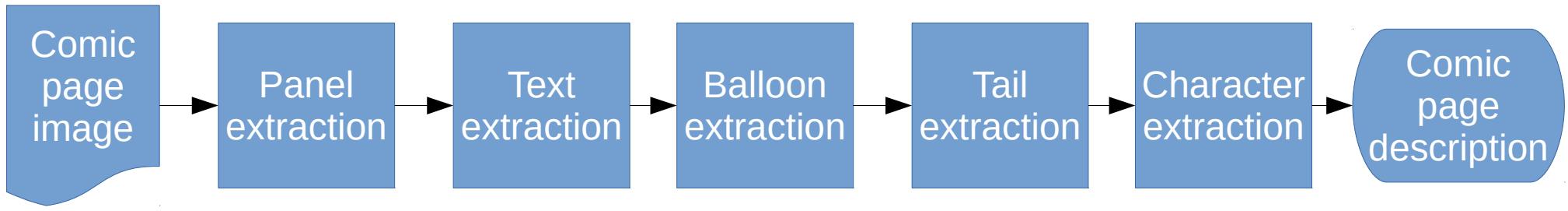
## Content-driven (independent)



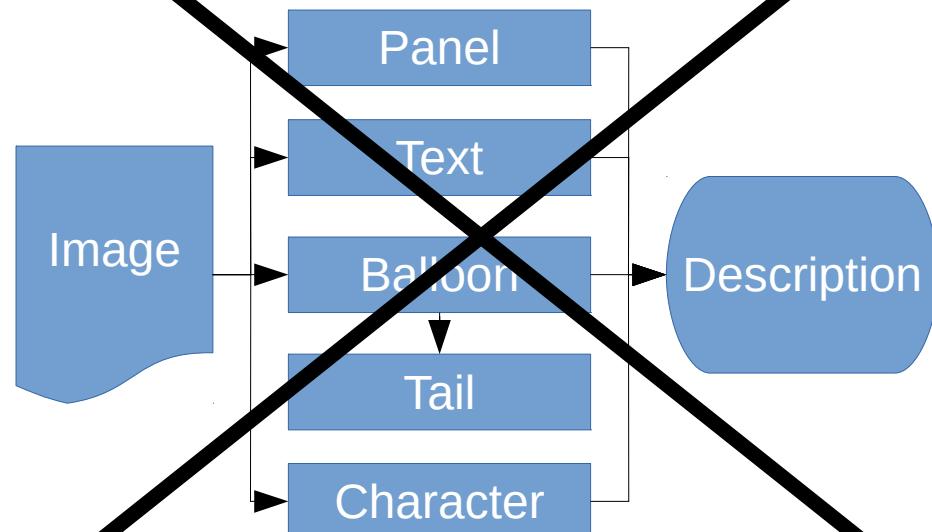
## Knowledge-driven (independent)



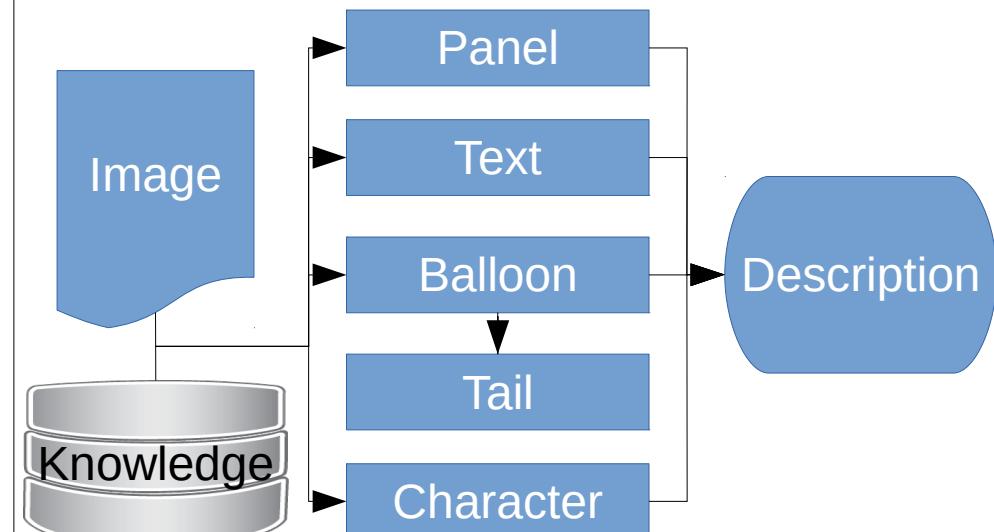
## Content-driven (sequential)



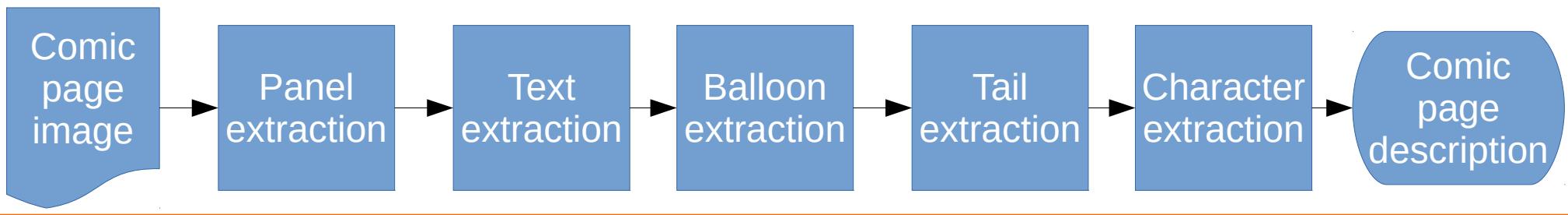
## Content-driven (independent)



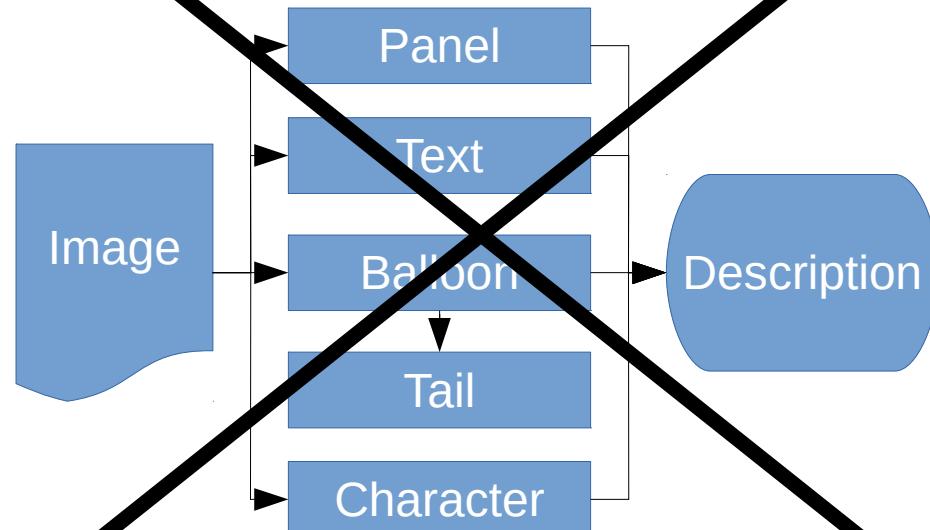
## Knowledge-driven (independent)



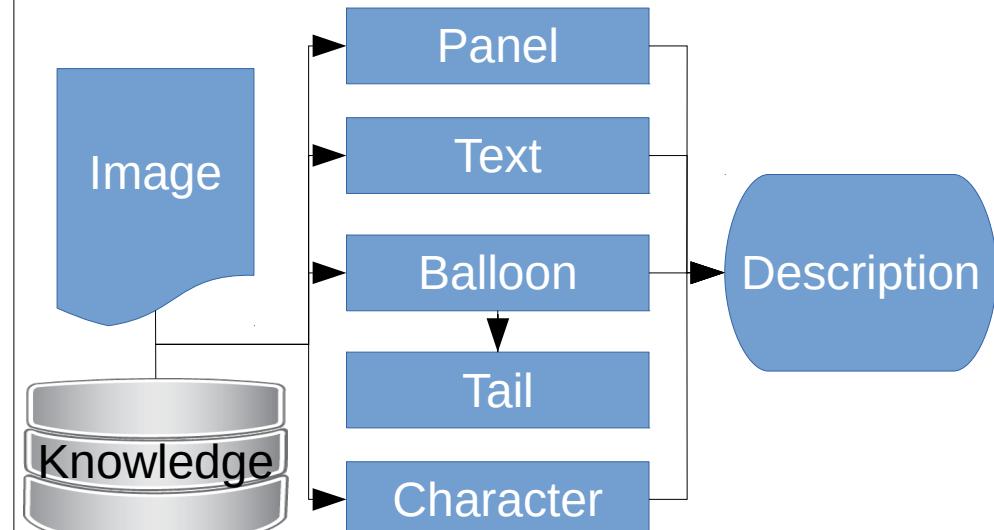
## Content-driven (sequential)



## Content-driven (independent)



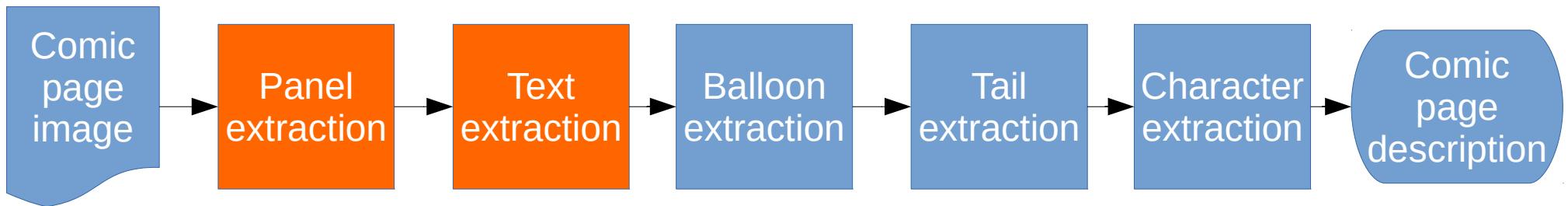
## Knowledge-driven (independent)



# Panel and text extraction

Contributions  
Content-driven approach

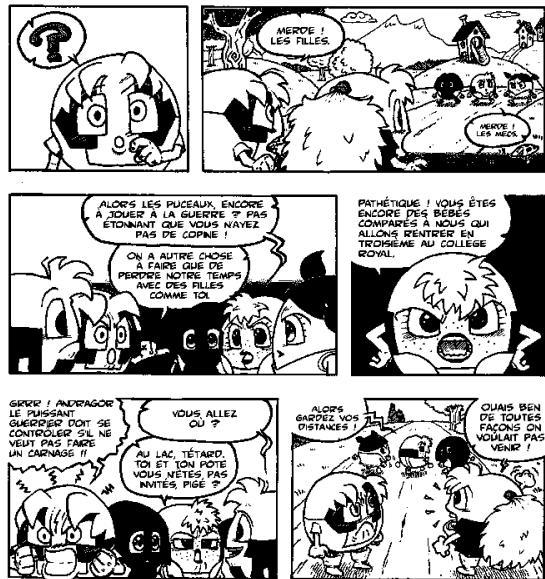
- Processing sequence



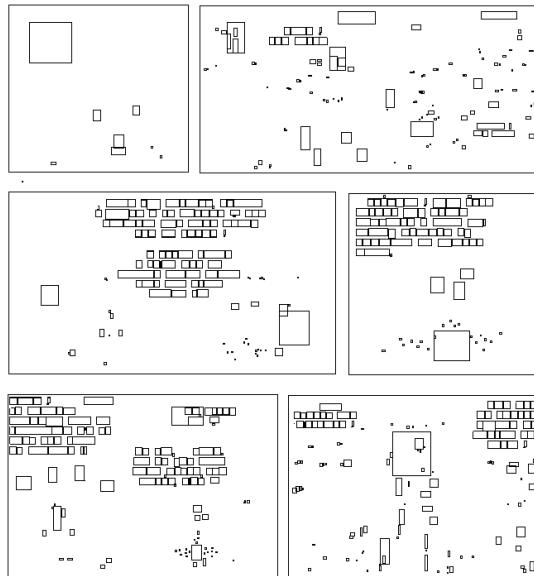
- Literature
  - Panel, balloon, text extraction [Ho2012, Arai2010]
  - TODO M. Stommel, L. Merhej, and M. Müller Segmentation-free detection of comic panels. In LNCS 2012.
  - Text only extraction [Li2013]
- Contribution
  - Simultaneous panel and text extraction

# Panel and text extraction

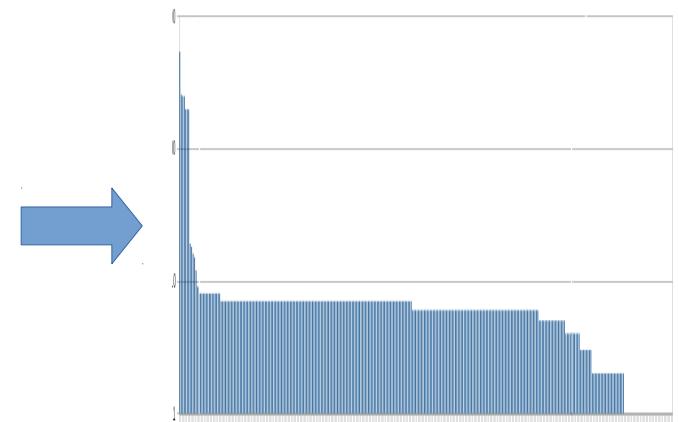
Contributions  
Content-driven approach



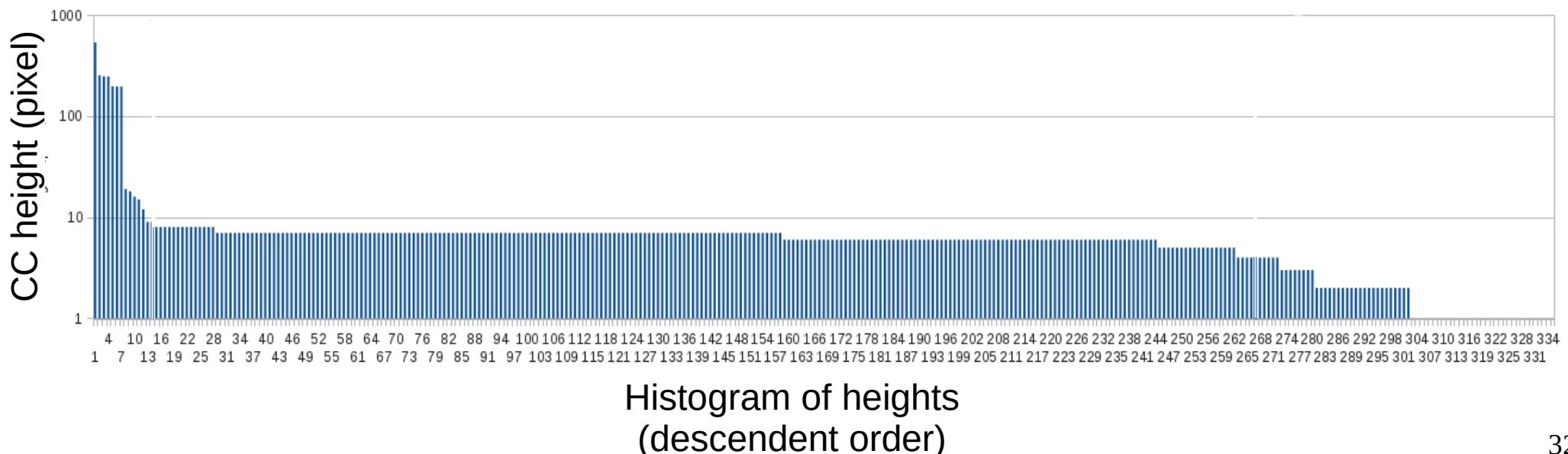
Binary image



Connected-component (CC) bounding boxes



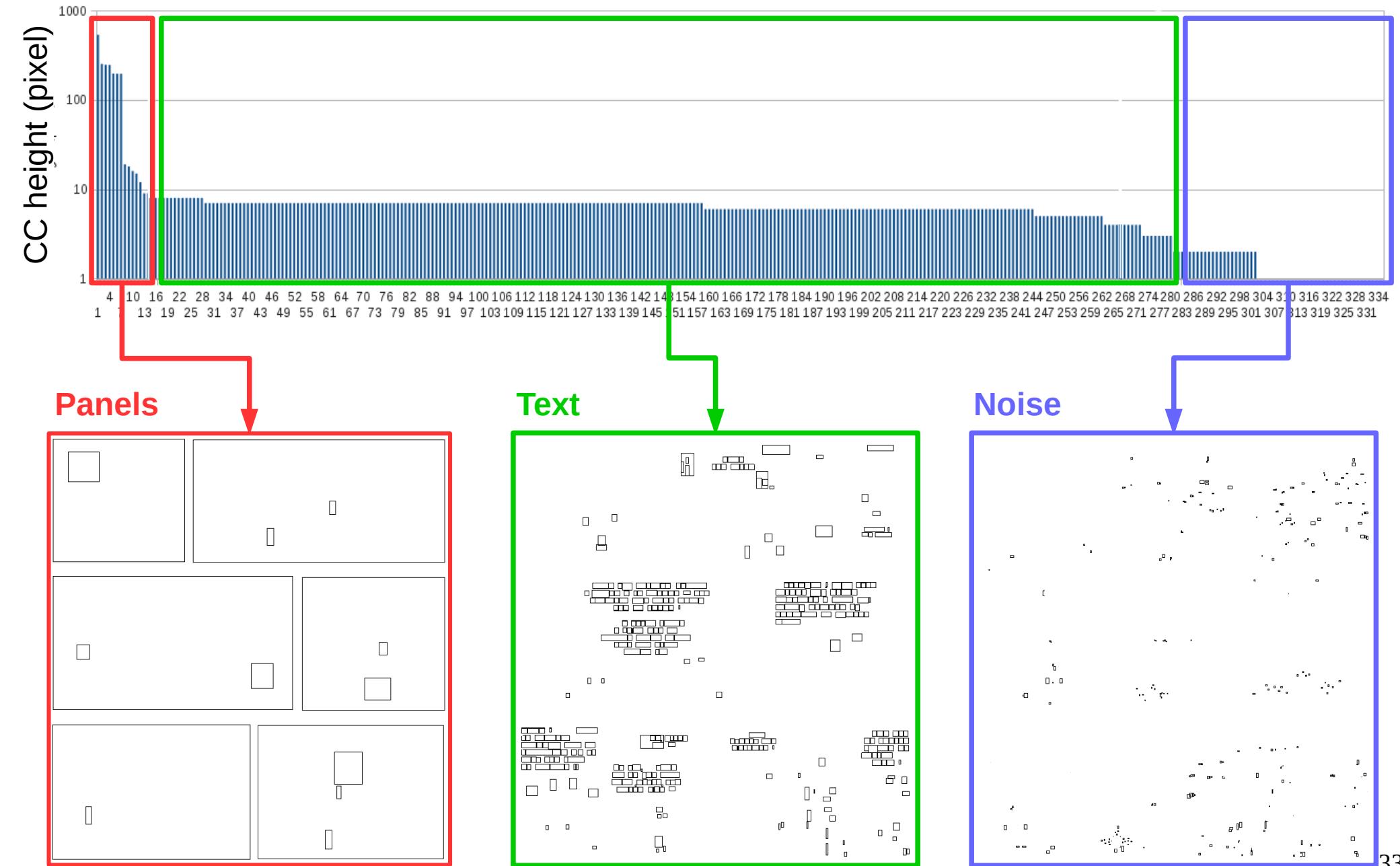
Histogram of heights of CC



# Panel and text extraction

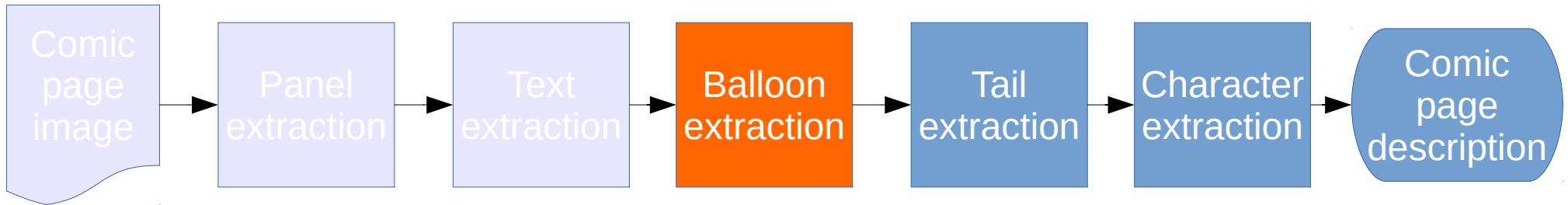
Contributions  
Content-driven approach

K-means clustering ( $k=3$ )

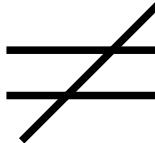


# Balloon extraction

Contributions  
Content-driven approach



Regular balloon

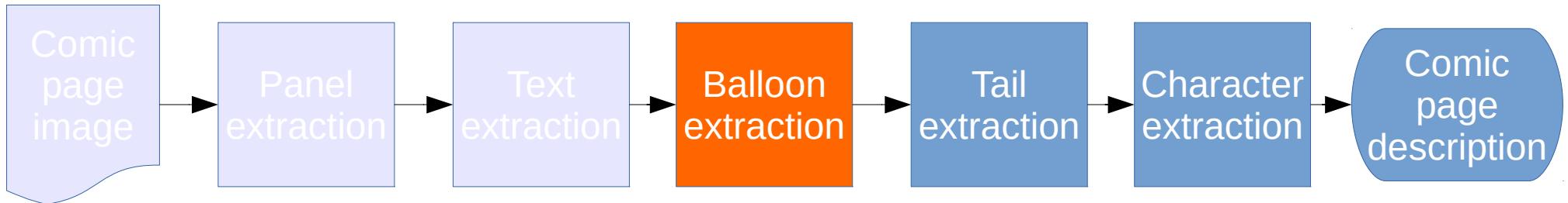


Implicit balloon

- Literature
  - Top-down approaches: extract white blobs and then text inside
  - Limited to regular balloons
- Contribution
  - Bottom-up approaches: extract text and then surrounding balloons
  - Appropriate for regular and implicit balloons

# Balloon extraction

Contributions  
Content-driven approach



Regular balloon



Implicit balloon

- Literature
  - Top-down approaches: extract white blobs and then text inside [Arai2010]
  - Limited to regular balloons
- Contribution
  - Bottom-up approaches: extract text and then surrounding balloons
  - Improvement of regular and a first approach for implicit balloon extractions

# Balloon extraction: regular

Contributions  
Content-driven approach

- Assumptions
  - Panels and text block positions are known
  - Balloons contain text
  - Text is fully contained and centred in balloons
- Proposition → structural analysis
  - Extract closed contours that includes centred text

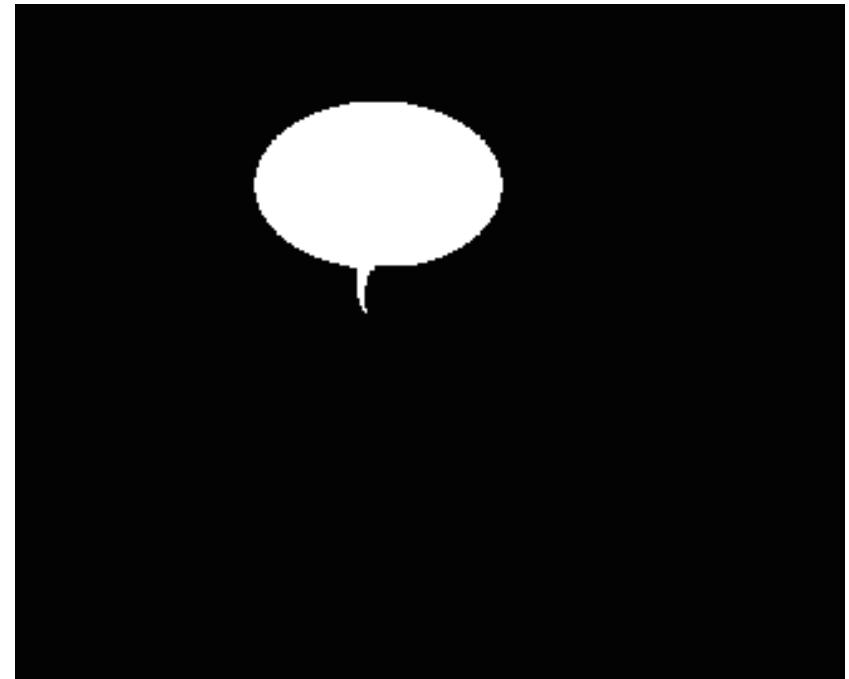
# Balloon extraction: regular

Contributions  
Content-driven approach

- Assumptions
  - Panels and text block positions are known
  - Balloons contain text
  - Text is fully contained and centred in balloons
- Proposition → structural analysis
  - Extract closed contours that includes centred text



Original image



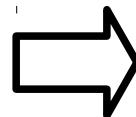
Expected result

# Balloon extraction: regular

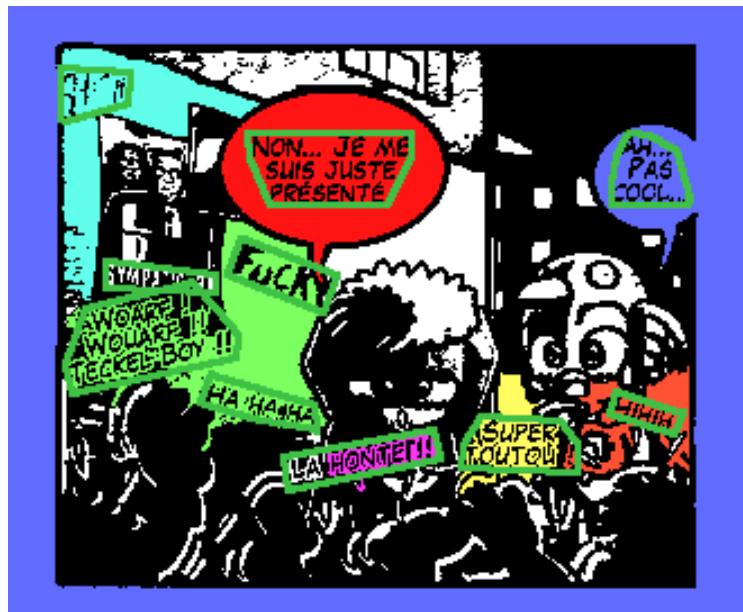
Contributions  
Content driven approach



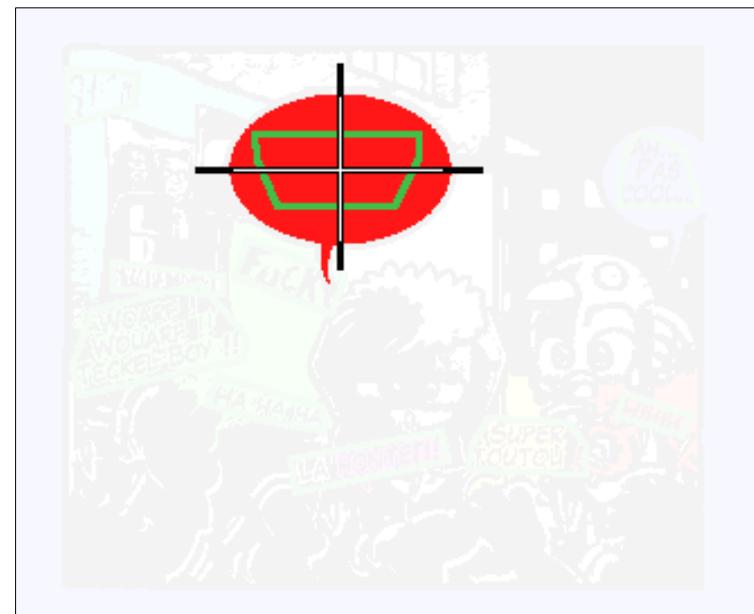
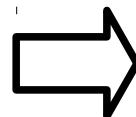
Original image



Text block positions (green)



Regions including text blocks (coloured)



Regions including aligned text blocks

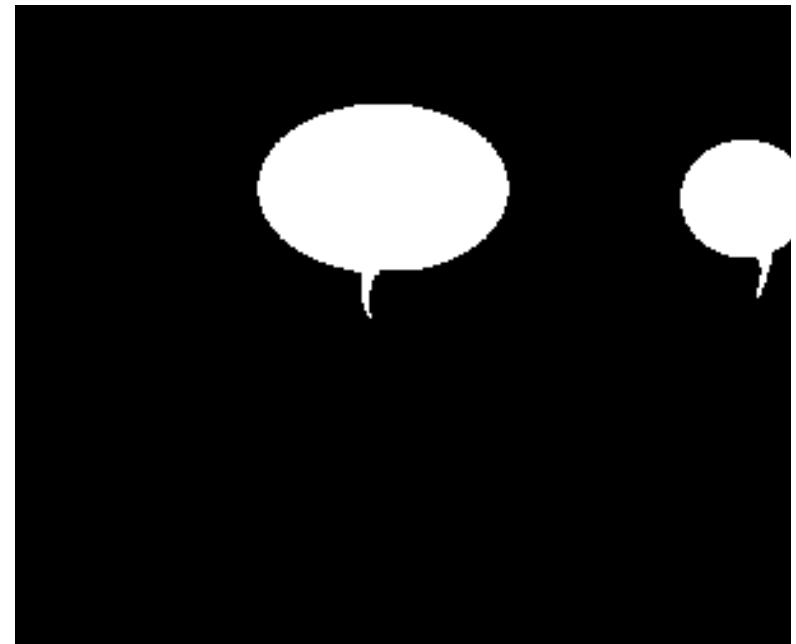
# Balloon extraction: implicit

Contributions  
Content-driven approach

- Prerequisites
  - Text block positions are known
  - Implicit balloons contain text
- Observation
  - Text is fully included and centred in balloons
- Proposition
  - Extract implicit balloons from text regions by inflating a deformable contour
  - Adaptation of active contour model (snake)



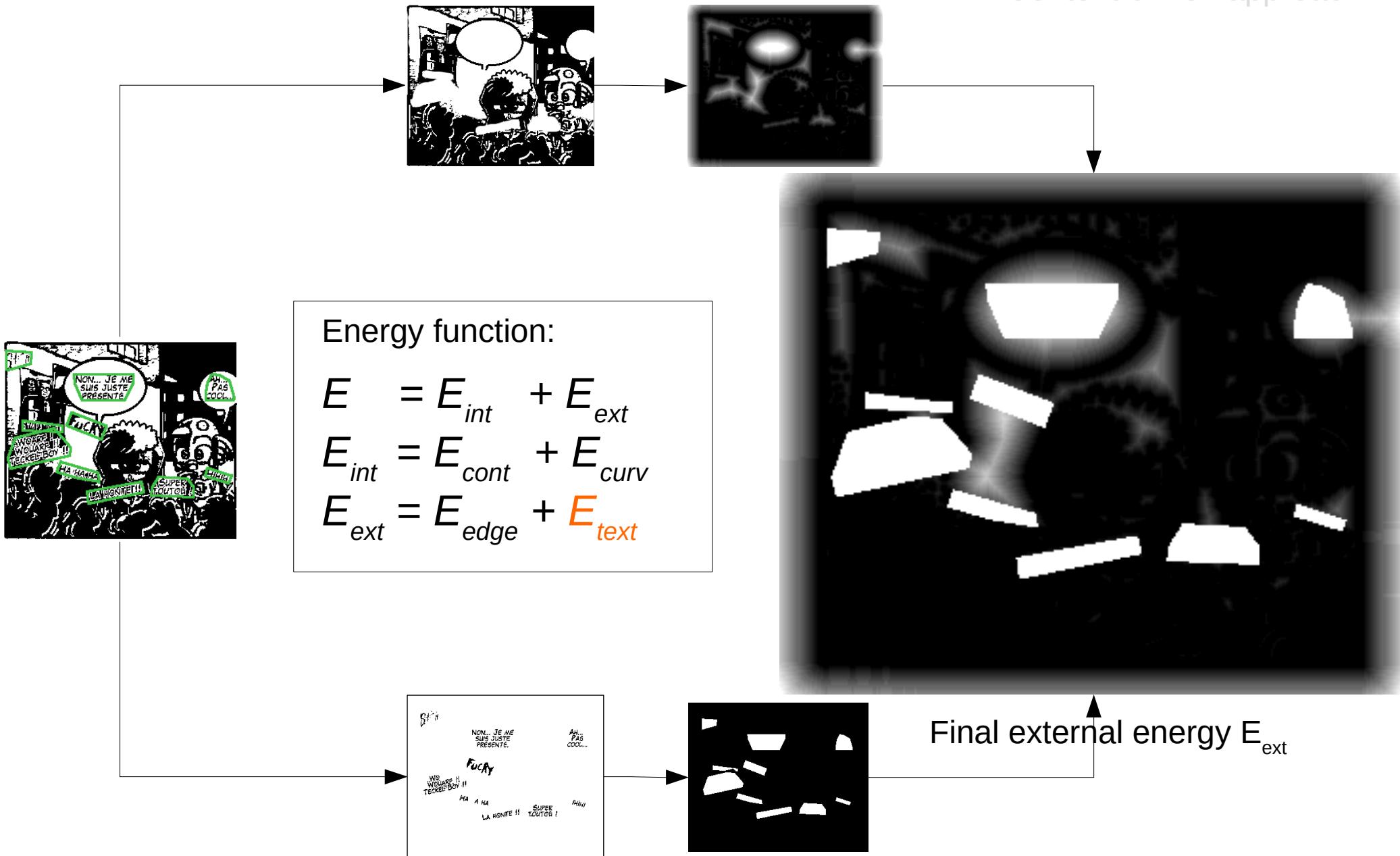
Original image and text locations



Expected result

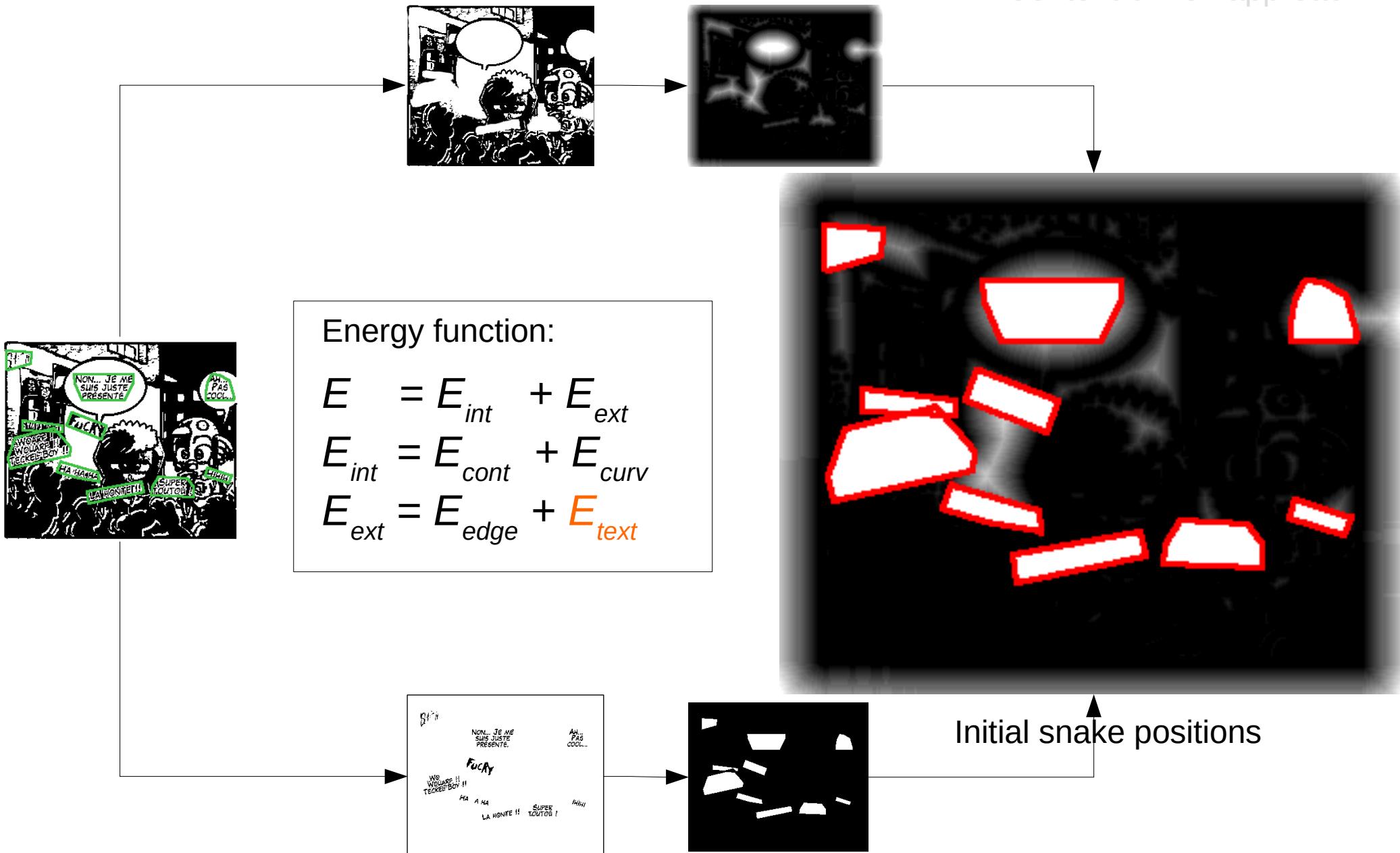
# Balloon extraction: implicit

Contributions  
Content-driven approach



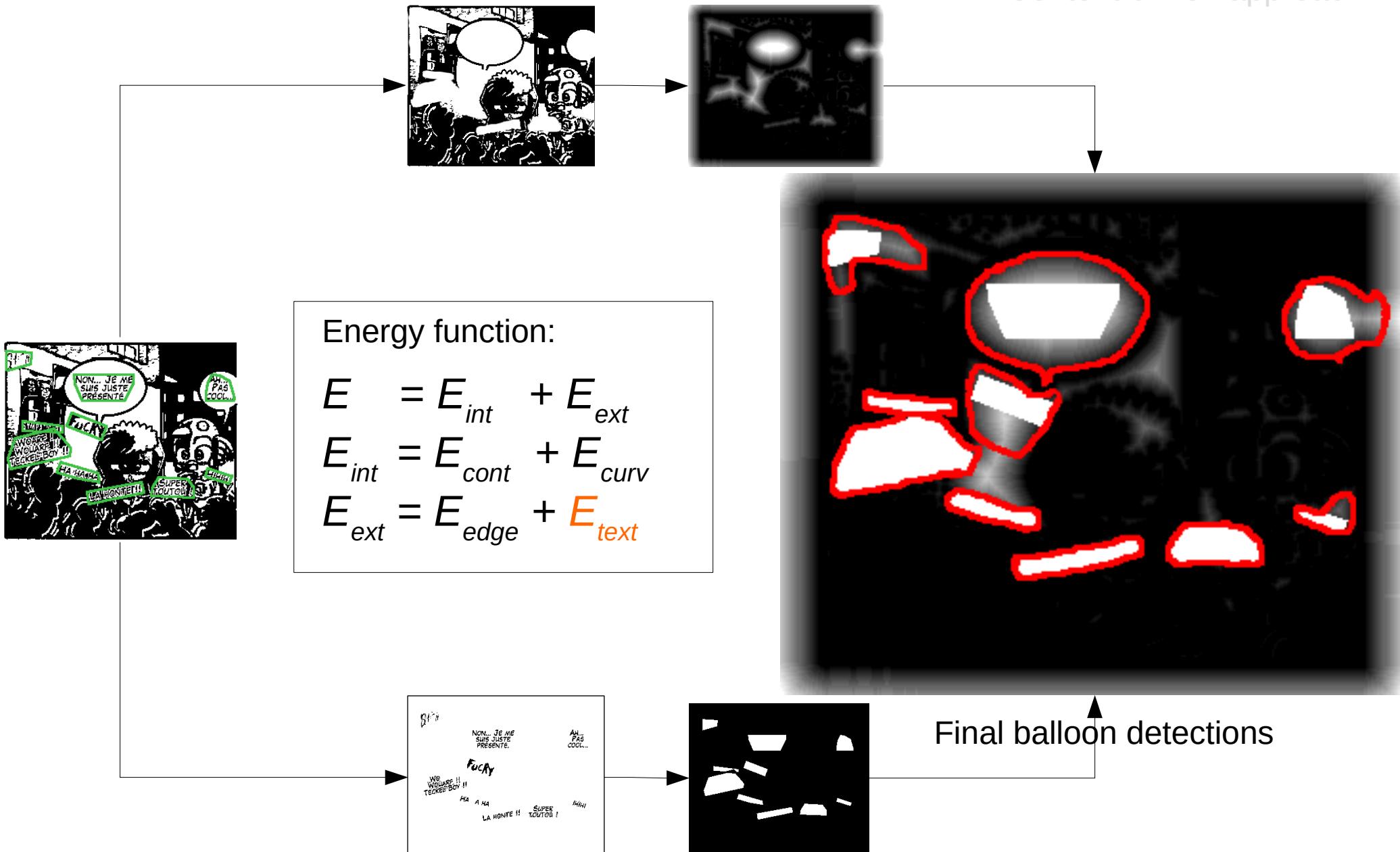
# Balloon extraction: implicit

Contributions  
Content-driven approach



# Balloon extraction: implicit

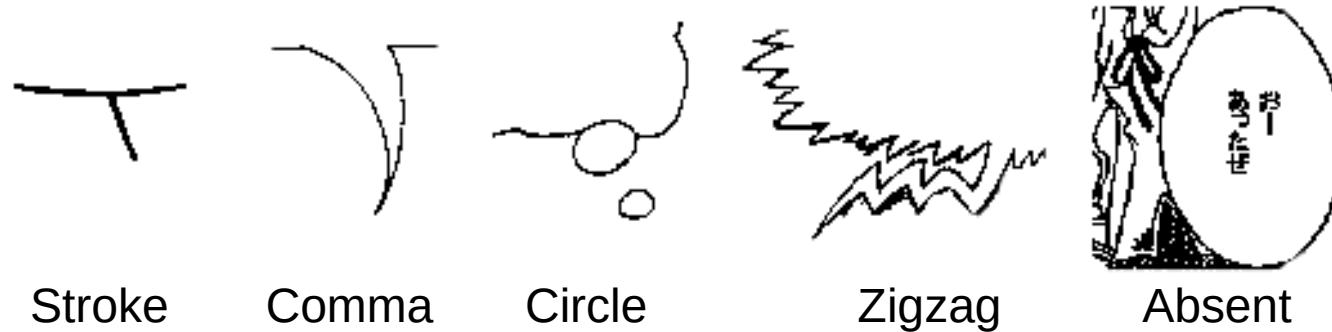
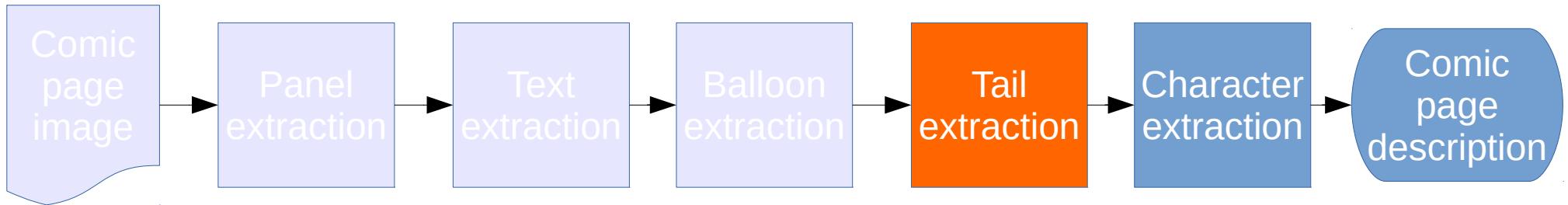
Contributions  
Content-driven approach



The snake is attracted to the “dark side”

# Tail extraction

Contributions  
Content-driven approach

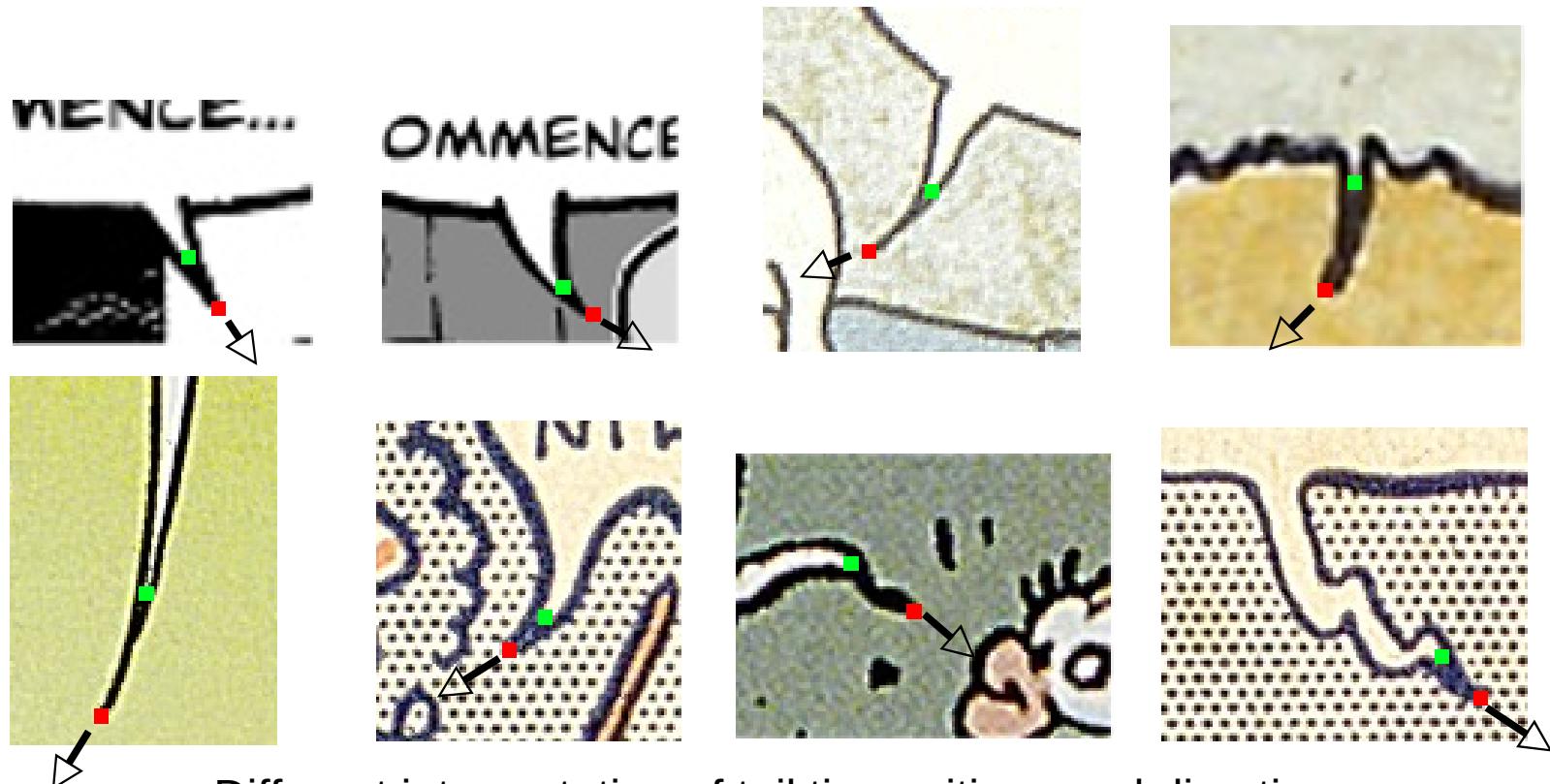


- Literature
  - First time studied in document image analysis
- Objectives
  - Detection of the tail tip position and orientation
  - Focus on comma, zigzag and absent types

# Tail extraction

Contributions  
Content-driven approach

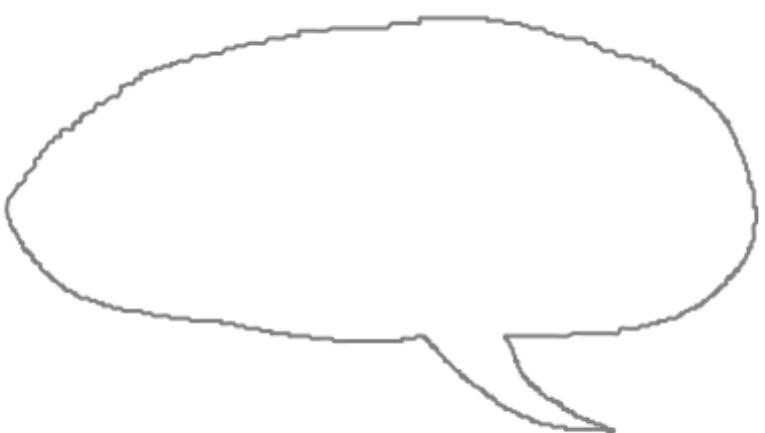
- Tip from background
- Tip from contour
- Direction of the tail



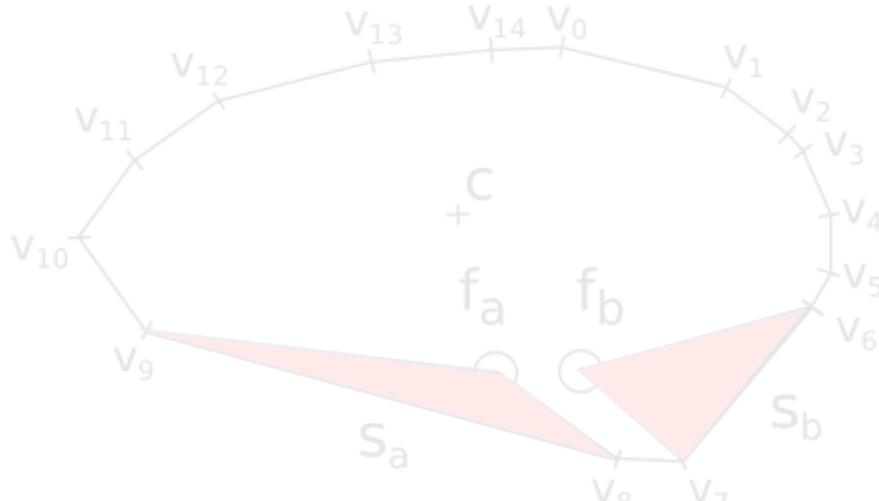
Different interpretation of tail tip positions and directions

# Tail extraction: tip position

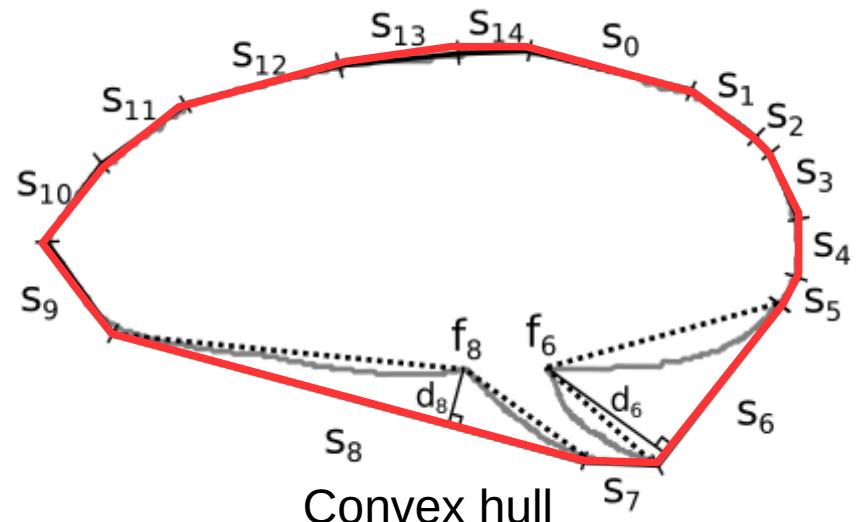
Contributions  
Content-driven approach



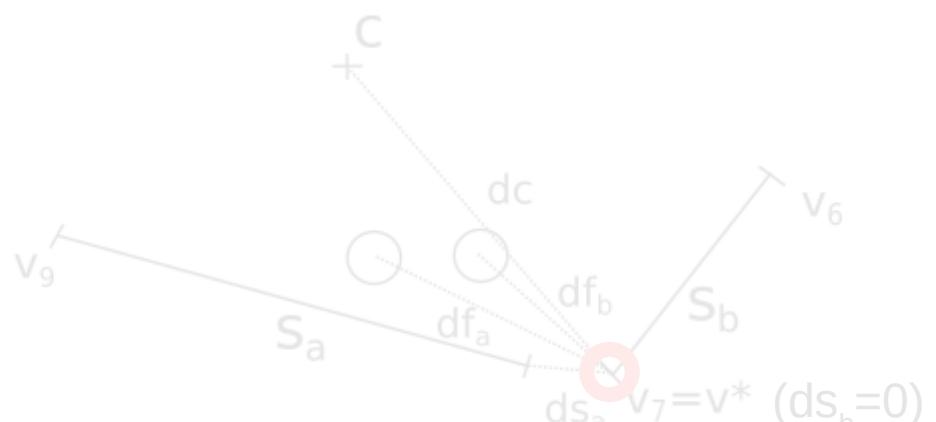
Balloon contour



Two biggest  
convexity defects



Convex hull



Tail tip position

Optimal vertex selection:

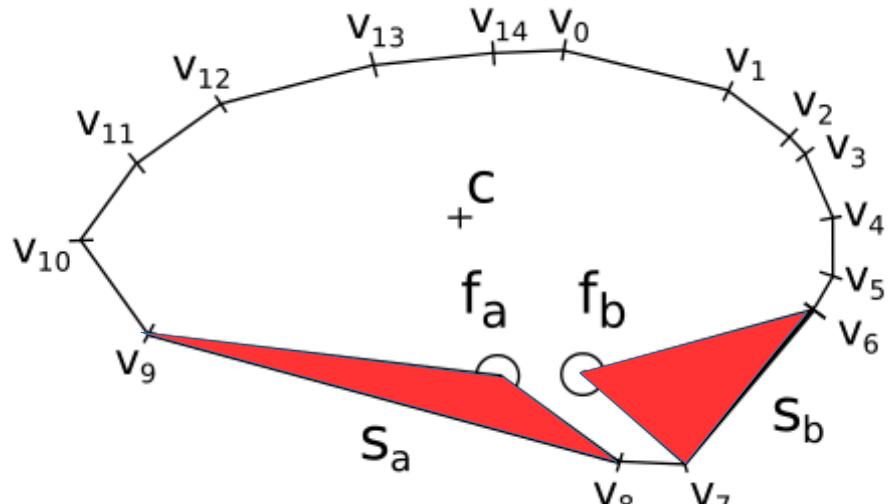
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

# Tail extraction: tip position

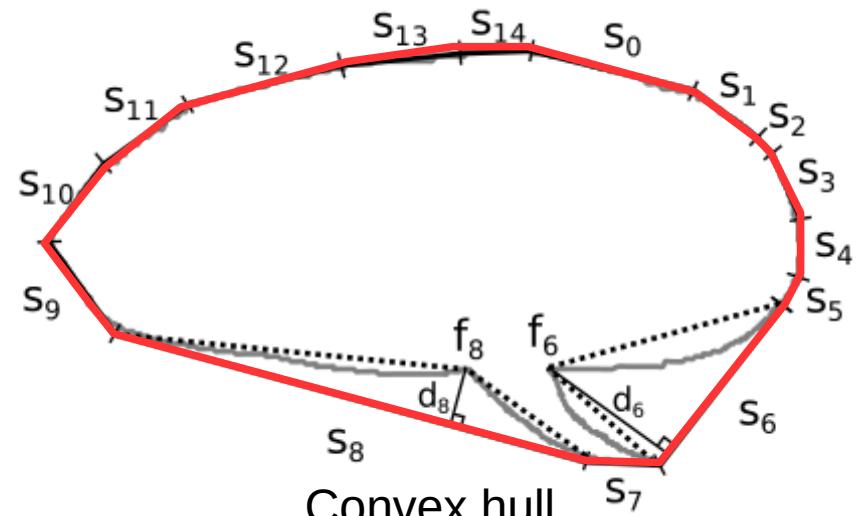
Contributions  
Content-driven approach



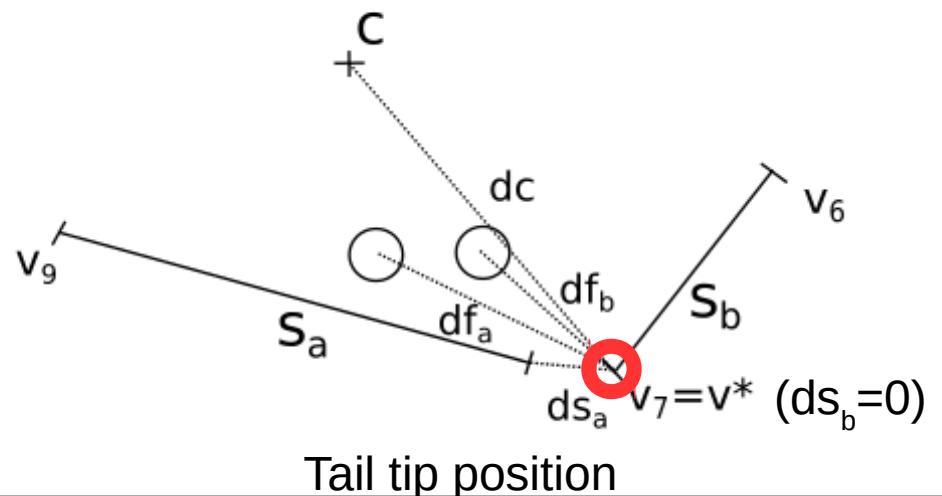
Balloon contour



Two biggest  
convexity defects



Convex hull



Tail tip position

Optimal vertex selection:

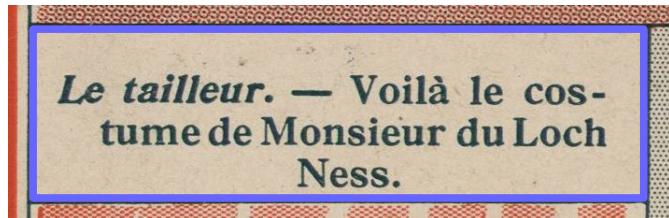
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

# Tail extraction: confidence value

Contributions  
Content-driven approach

Balloon  
contour (blue)

Balloon 1

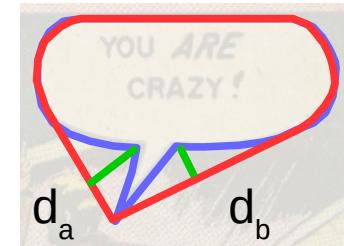
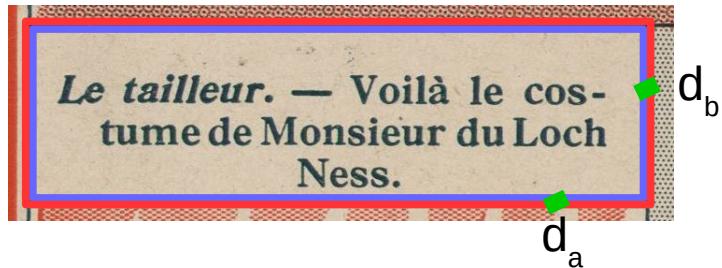


Convex hull  
(red)

Balloon 2



Confidence



$$C_{tail} = \frac{(d_a + d_b)/2}{meanBalloonSize}$$

$$C_{tail} = 0.0$$

$$C_{tail} = 0.73$$

Presence of tail

NO

YES (>0)

# Tail extraction: tail direction

Contributions  
Content-driven approach

- Definition
  - Vector starting from “background” to “external edge” tail tip positions
- Approach
  - Extract **external edge**
  - Find **external edge tail tip coordinates**
  - Define the **tail direction** (N, NE, E, SE, S, SW, W, NW)



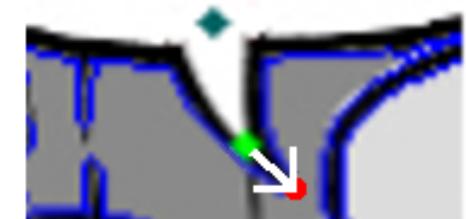
Background tail tip  
(green) and  
external edge (blue)



Closest point on  
external edge  
(red)



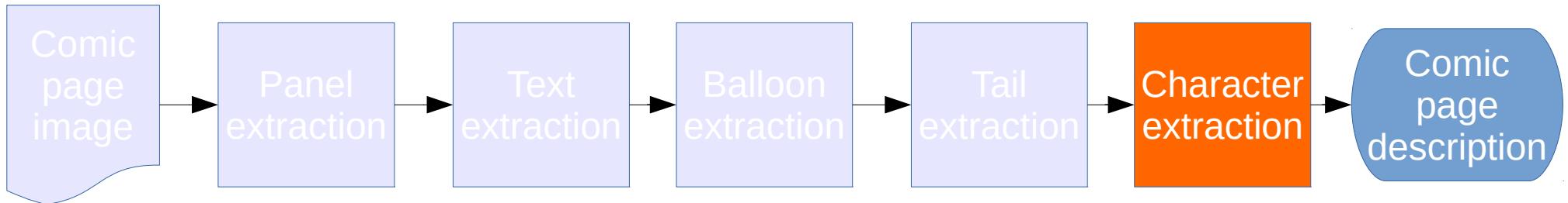
Farthest point  
from origin and tip  
(red)



Direction from tip  
to farthest point  
(white arrow)

# Comic character extraction

Contributions  
Content-driven approach

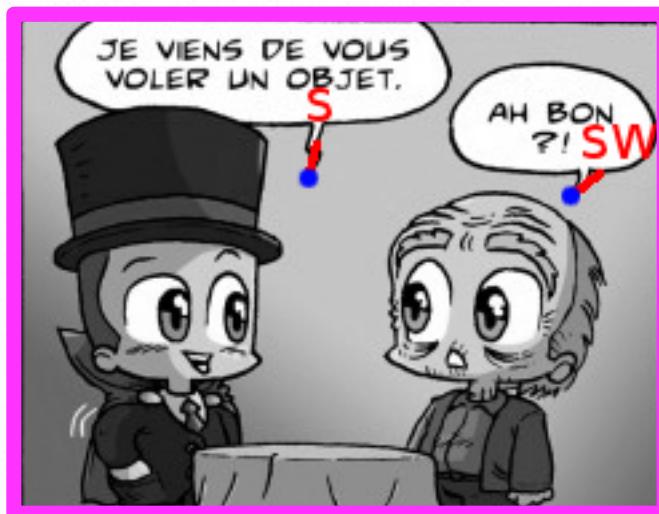


- Literature
  - Supervised approaches for manga and cartoon characters [TODO]
  - No public dataset (copyright issues)
- Challenges
  - Variety of styles of comic books
  - Intra and extra class variations of each character instance (e.g. position, scale, pose, occlusion and human-like, invented)
- Objective
  - Unsupervised and generic approach for all styles of comic books

# Comic character extraction

Contributions  
Content-driven approach

Panels + Tails = ?



# Comic character extraction

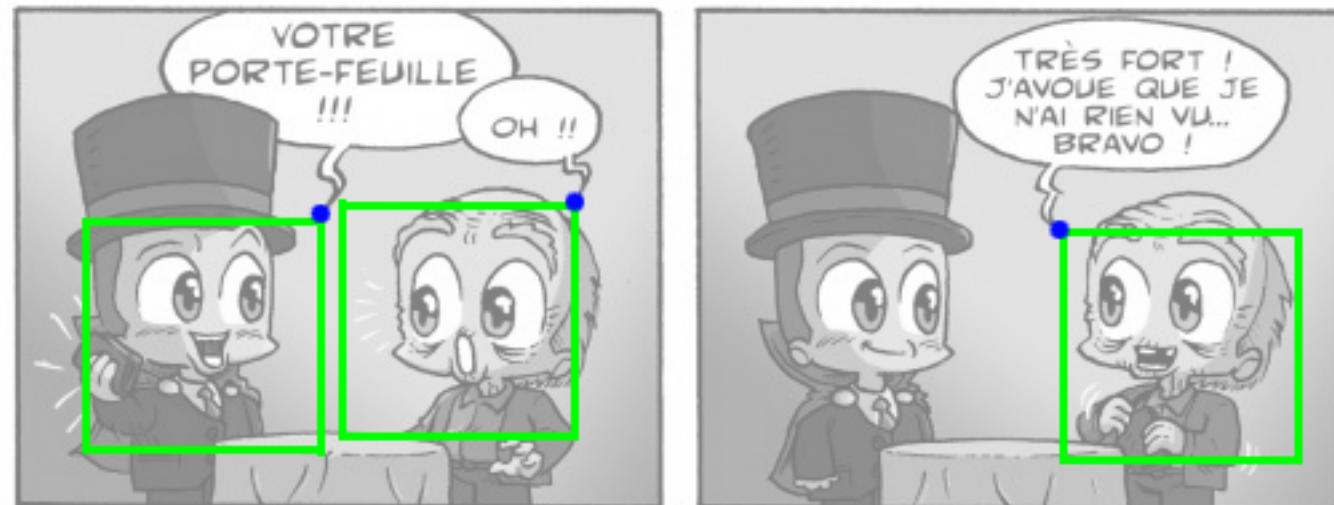
Contributions  
Content-driven approach

Panels + Tails = Comic character ROIs

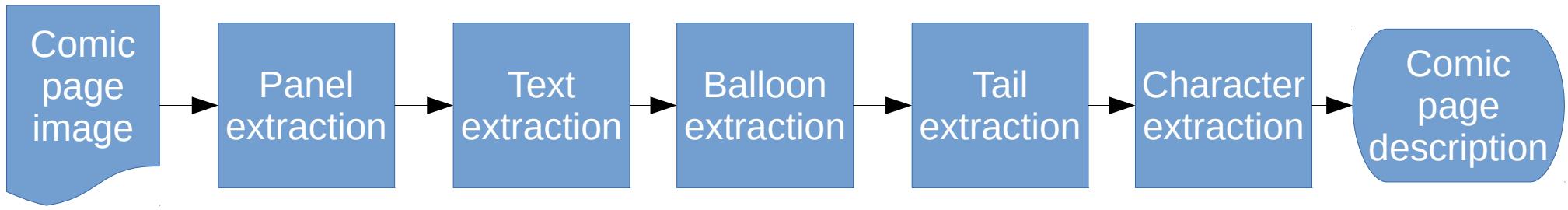
Large ROI



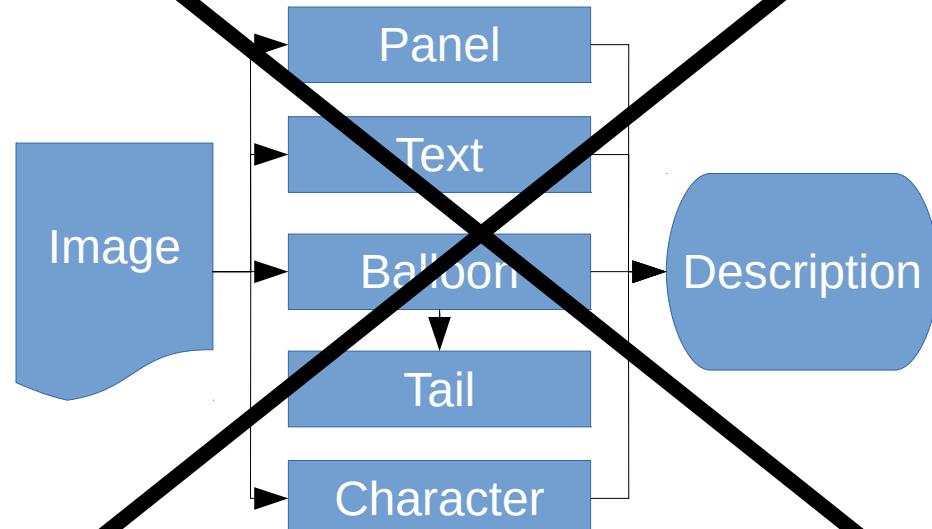
Small ROI



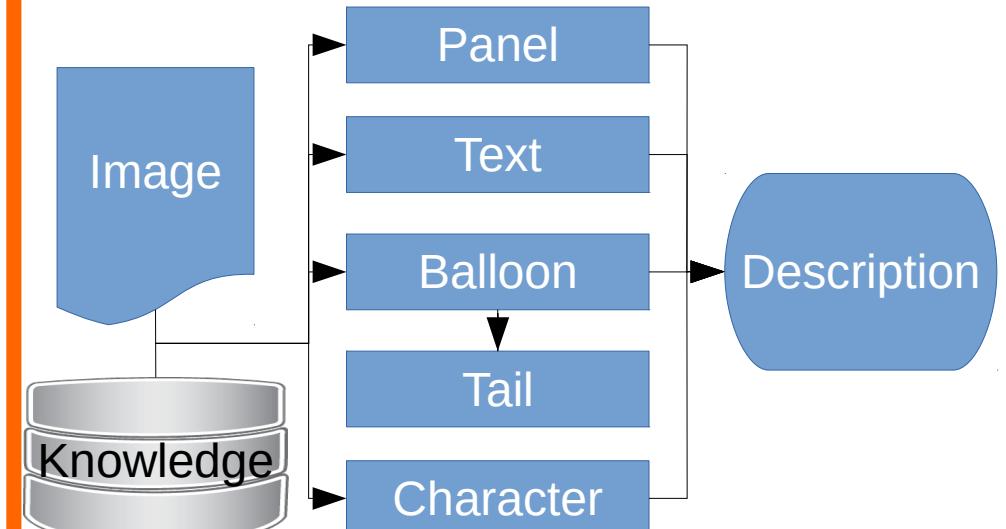
## Content-driven (sequential)



## Content-driven (independent)



## Knowledge-driven (independent)

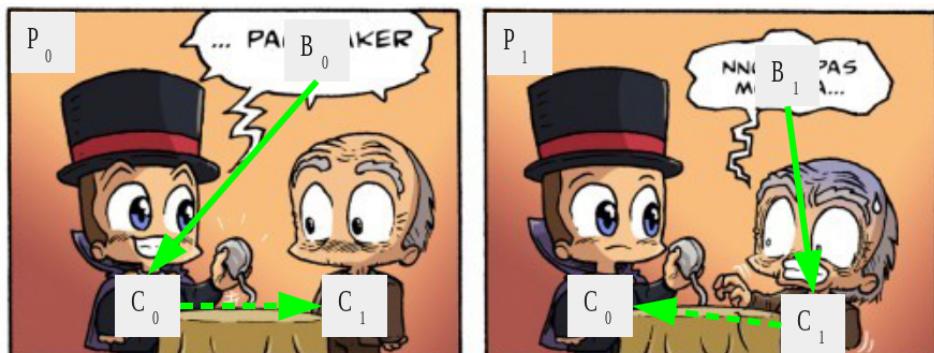


# Introduction

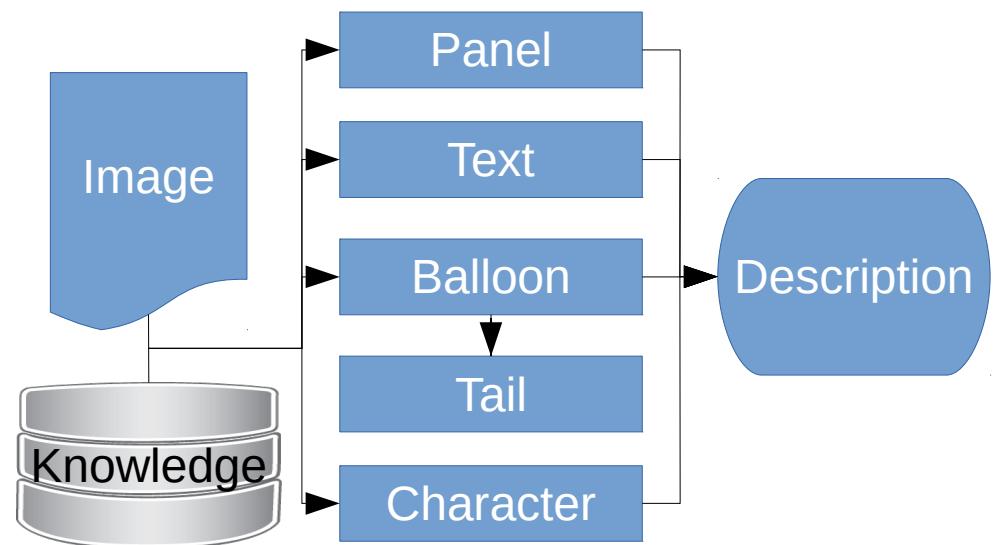
# Contributions

Knowledge-driven approach

- High level image description
- Independent element extraction
- Framework for comics understanding
- Collaboration with Clément Guérin

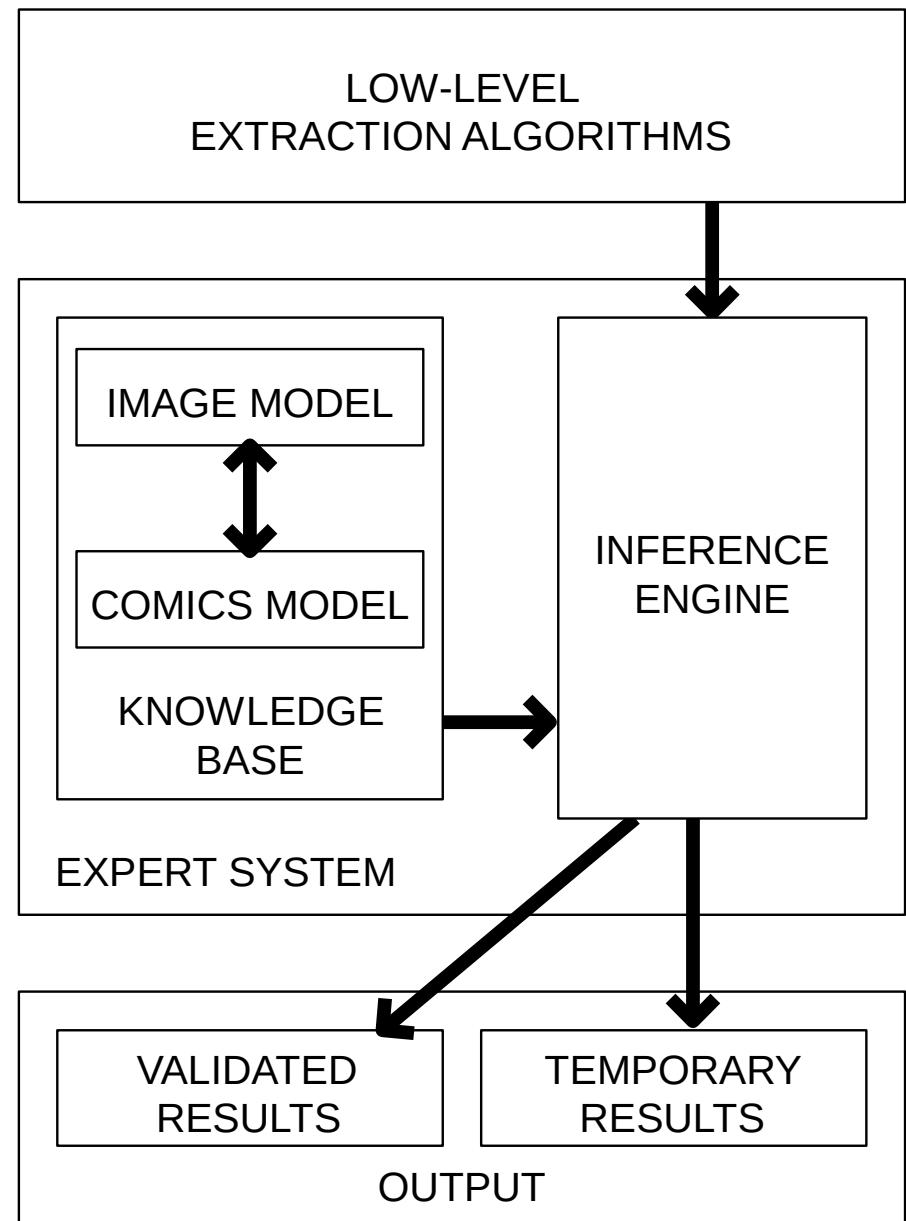


Example of situation understanding



# Knowledge representation

- Image model
  - Physical support
  - Region of interests
- 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**



# Processing sequence

Contributions  
Knowledge-driven approach

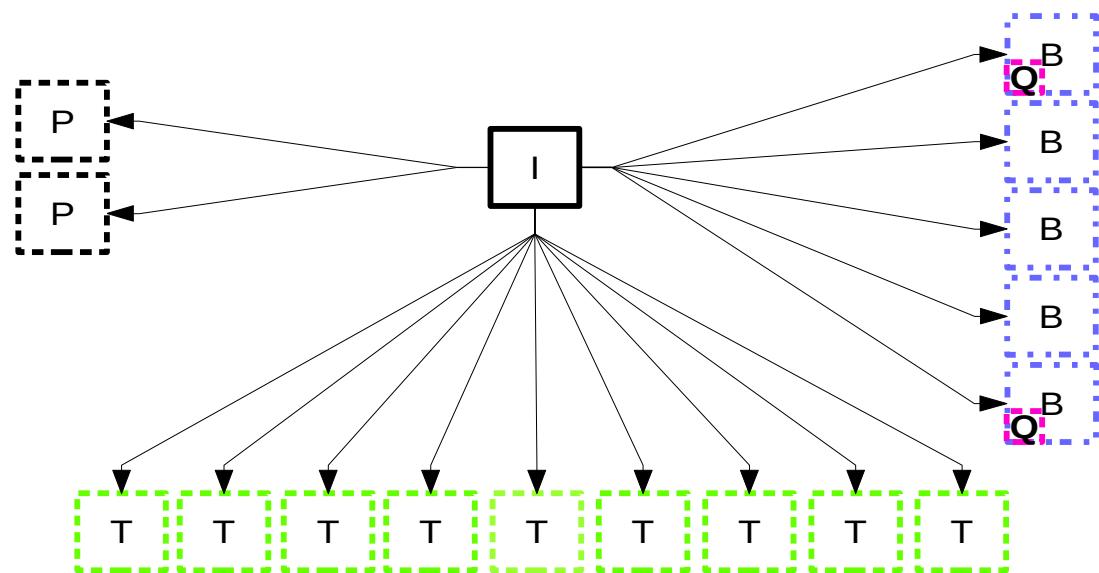


# Processing sequence

Contributions  
Knowledge-driven approach

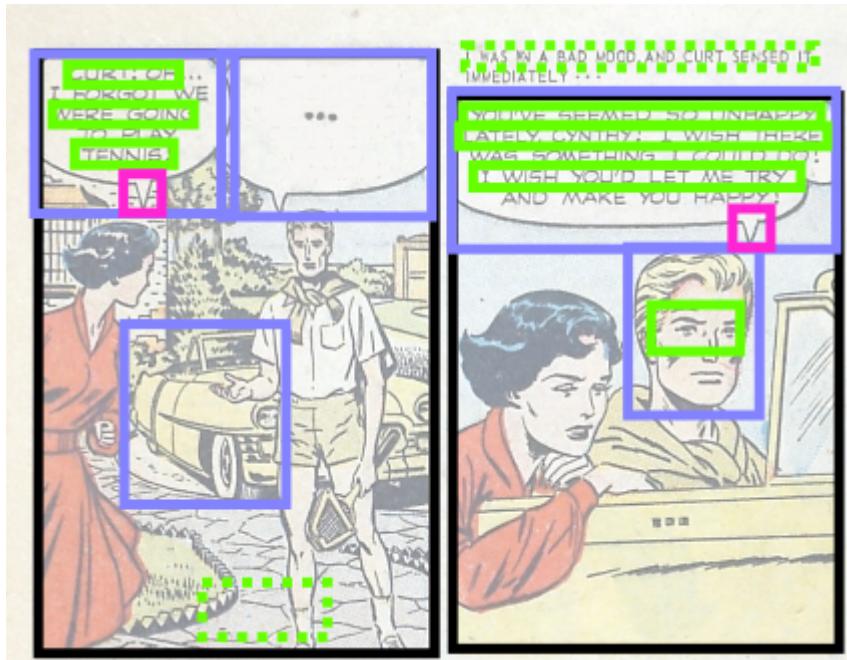


Hypotheses of  
panels, balloons  
and text lines



# Processing sequence

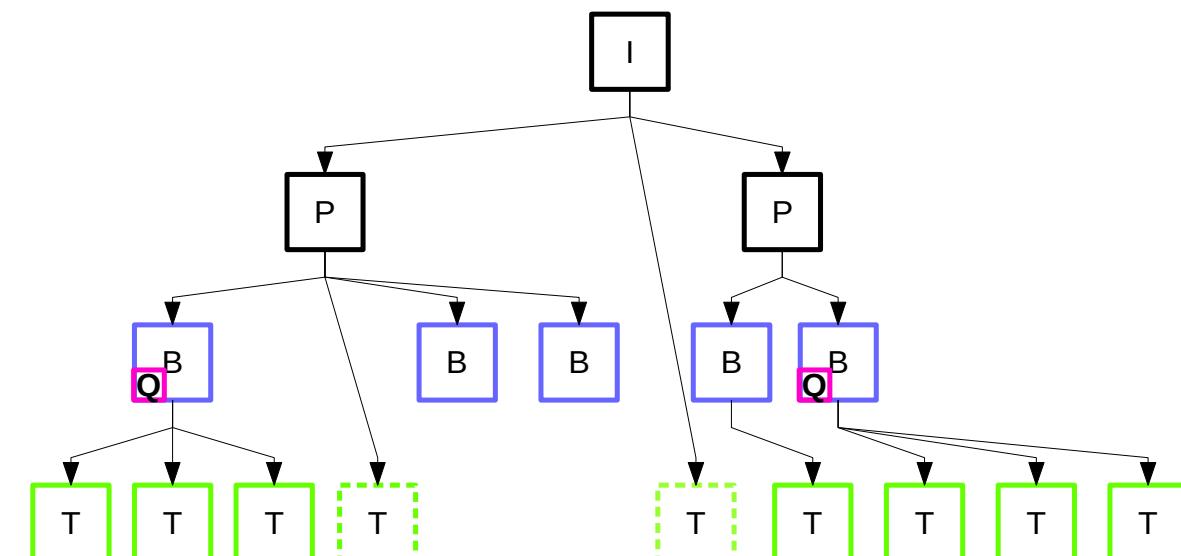
Contributions  
Knowledge-driven approach



Hypotheses of  
panels, balloons  
and text lines



Validation of the  
hypotheses

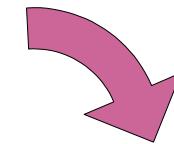


# Processing sequence

Contributions  
Knowledge-driven approach

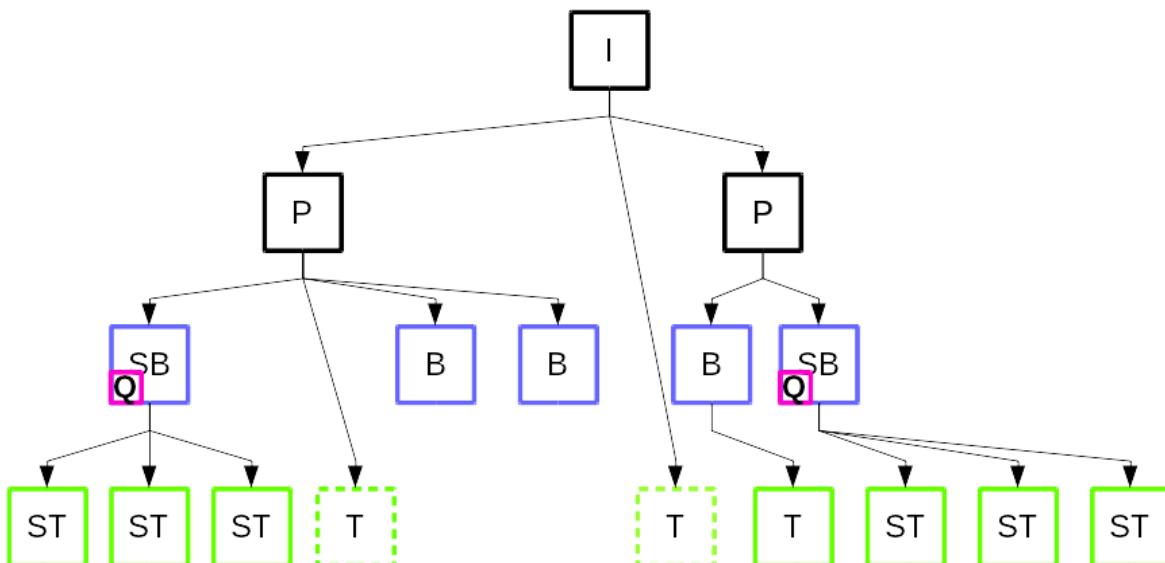


Hypotheses of  
panels, balloons  
and text lines



Validation of the  
hypotheses

Inferences of  
specific types



# Processing sequence

Contributions  
Knowledge-driven approach



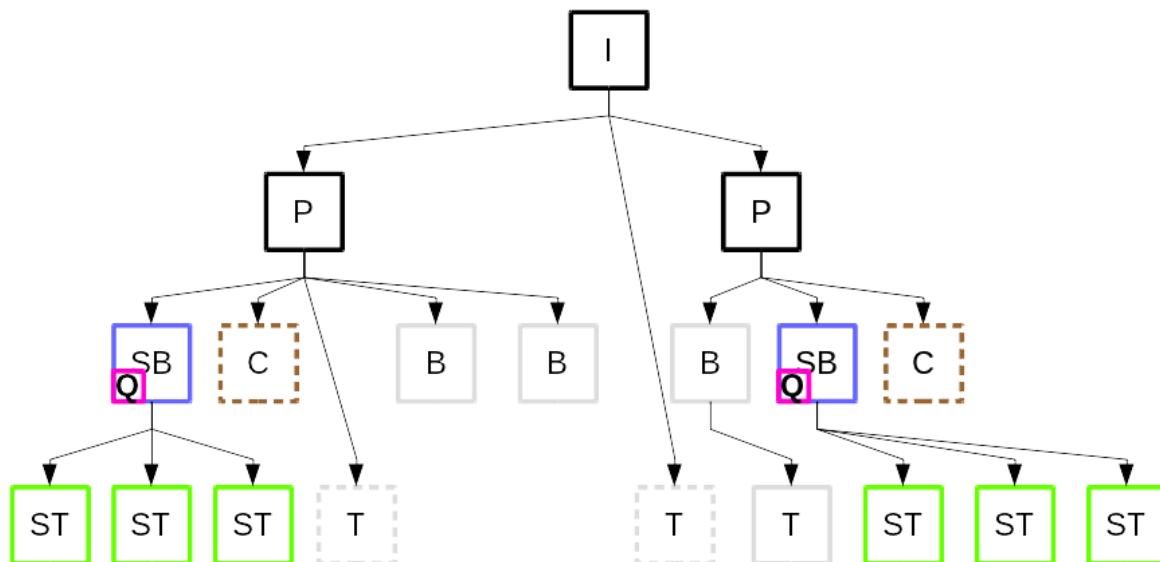
Hypotheses of  
comic characters



Validation of the  
hypotheses



Inferences of  
specific types

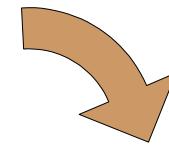


# Processing sequence

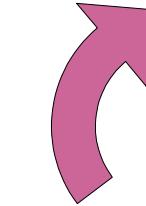
Contributions  
Knowledge-driven approach



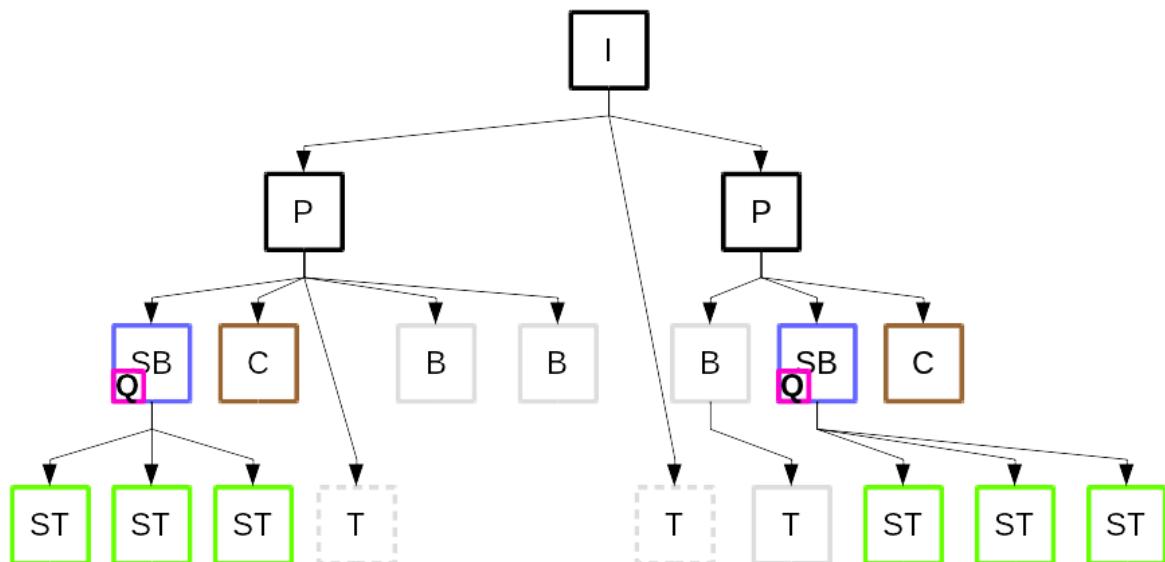
Hypotheses of  
comic characters



Validation of the  
hypotheses

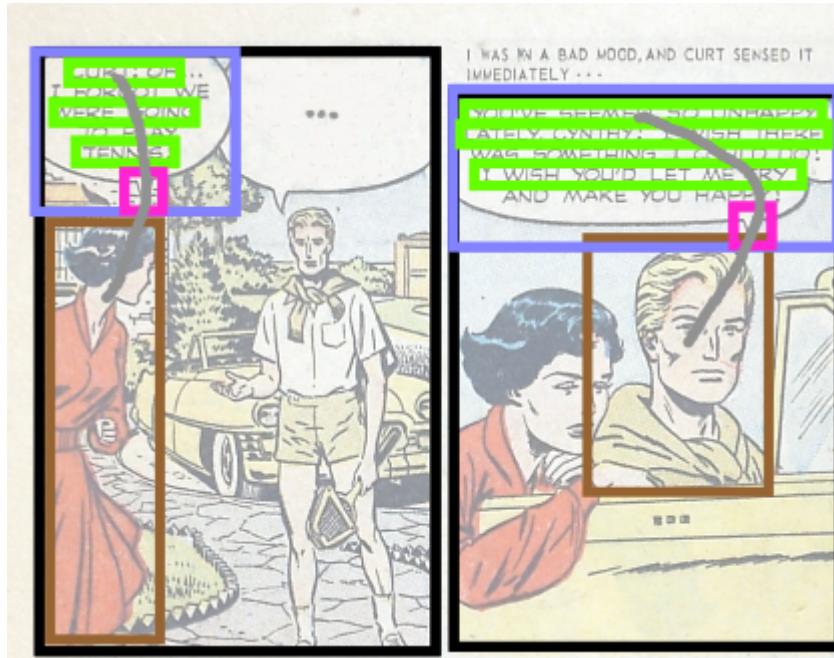


Inferences of  
specific types

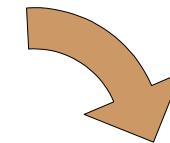


# Processing sequence

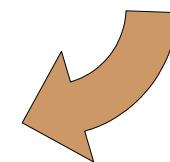
Contributions  
Knowledge-driven approach



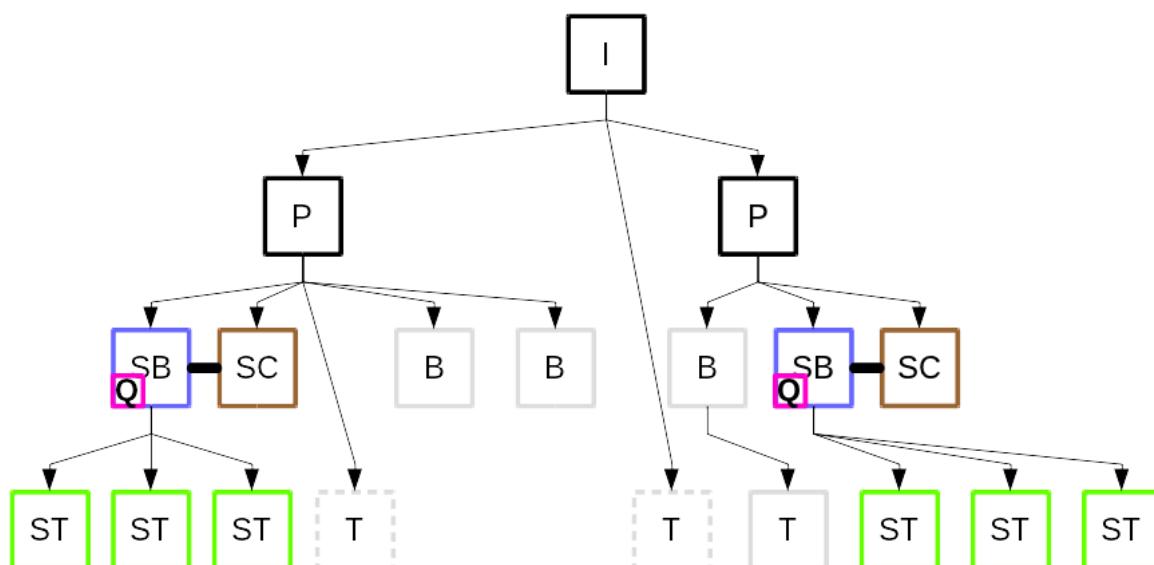
Hypotheses of  
comic characters



Validation of the  
hypotheses



Inferences of  
specific types  
+ semantic links



- Dataset and ground truth
- Evaluations
- Conclusions



Lettering. Image credits: Le cycle des bulles,  
Christophe Rigaud, 2012

# Dataset and ground truth

# Experiments

- Absence of public dataset
- Creation of heterogeneous dataset
  - 100 mixed pages from 20 albums
  - Franco-Belgium “bandes dessinées”, American comics and Japanese manga
  - From 1905 to 2012, paper and webcomics
  - Rights holder permissions agreement
- Online: <http://ebdtheque.univ-lr.fr>

## Bibliographic annotations



**PAGE** (100)  
**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** (850)  
**Rank:** 1  
**BALLOON** (1092)  
**Rank:** 2  
**Shape:** Oval  
**Tail direction:** South-West  
**TEXT LINE** (4691)  
**Text:** « STARK RAVING »  
**CHARACTER** (1550)  
**LinkedToBalloon:** 2

# Evaluations

# Experiments

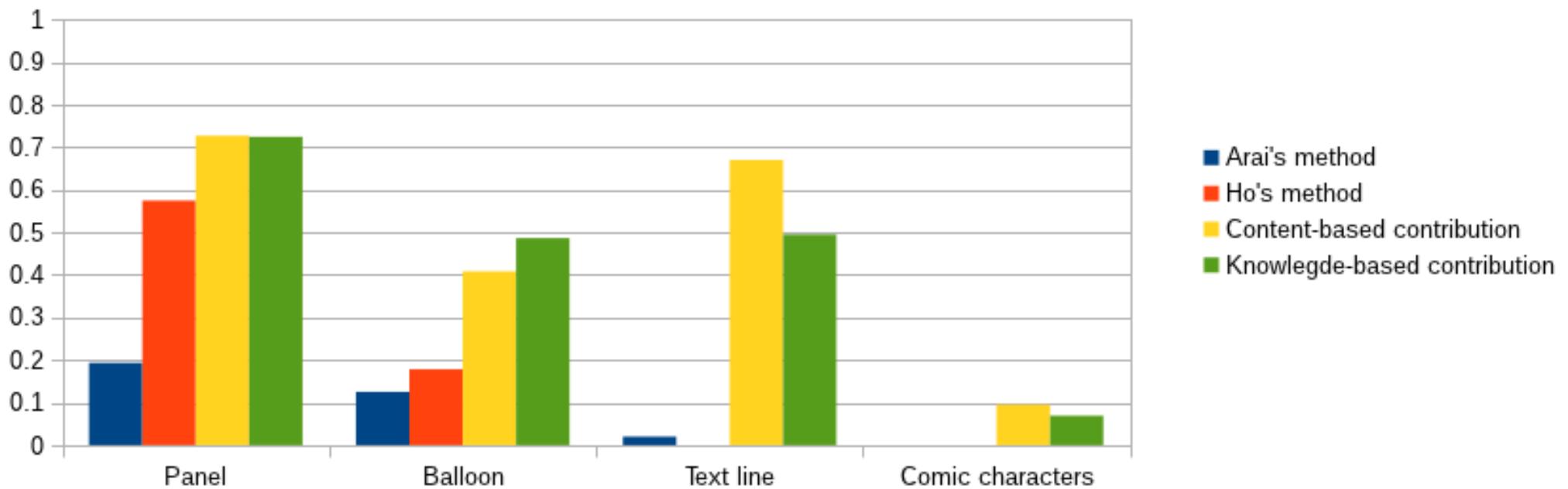
$$a_0 = \frac{\text{area}(B_p \cup B_{gt})}{\text{area}(B_p \cap B_{gt})}$$

$B_p$  = predicted region

$B_{gt}$  = ground truth region

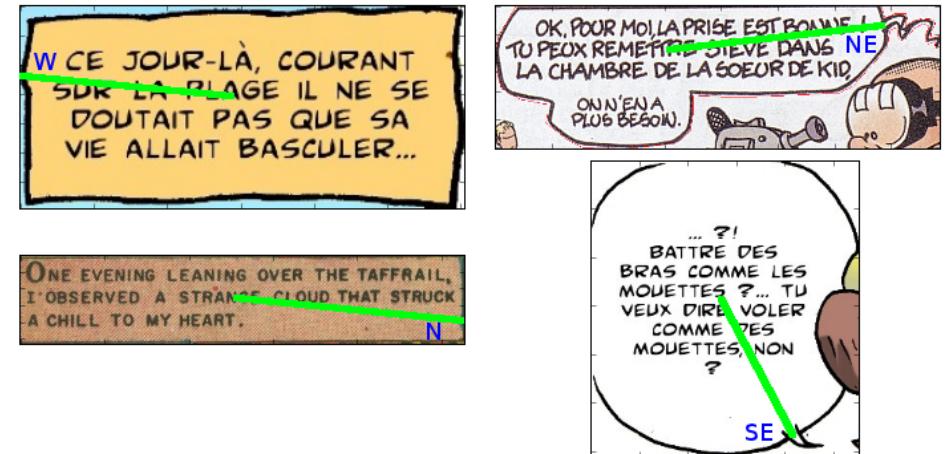
$B_p$  valid if  $a_0 > 0.5$

Element localisation results (F-score)



# Evaluations

## Examples of result



TODO

- Global conclusions
- Global perspectives
- Publications



Lettering. Image credits: Le cycle des bulles,  
Christophe Rigaud, 2012

# Global conclusions

# Conclusion

- Reached objectives
  - Comics **image segmentation** and understanding
- Contributions of the thesis
  - **Comic book** content **extraction** methods **improvement**
  - **First approaches** for tail detection, balloon classification and speaker extraction
  - Public **dataset** and ground truth
- Research impacts
  - **L3i** is now a **main actor** of comic book analysis in Europe
  - New Ph.D. **thesis** started in 2013 (Nam Le Thanh)
  - **Dataset** used by international peers (**TODO: list them all?**)
  - **National projects** (PIA BigData Actialuna + LIP6, ANR EXPION 2015)
  - **International project** on manga analysis (Campus France SAKURA with Japan)

# Global perspectives

# Conclusion

- Content extraction
  - Panels: implicit, overlapping and connected
  - Text: recognition
  - Balloon: implicit extraction and evaluation
  - Comic characters: non-speaking and identification
- Content understanding
  - Situation retrieval
  - Object interaction retrieval
  - Label elements from text analysis
- Dataset
  - Increase the number of pages
  - Panel view angle and situation
  - Multi-part comic character segmentation
  - Comic character names and roles

# Publications

# Conclusion

TODO

Thank you + github + c-r

Conclusion

- Authors, publishers, CIBDI
- Founding: Communauté d'Agglomération de La Rochelle
- Time for demo on easy case with comic character ROI computation?



# Complementary information

# History of comics art

- Pre-history: **painting of animals** and hunters in **caves** [Marx, 2007]
- 1846: **Rodolphe Töpffer**, the inventor of the “**bandes dessinées**”
- 1930s: magazine-style **comic books** production in the **US**
- 1950s: massive production of **manga** in **Japan** (Osamu Tezuka)
- 1971: the term of **ninth art** is attributed to comics art (Francis Lacassin)
- 1996: explosion of the **Internet bubble** and **webcomics**
- 2007: adaptation to **social media sites** and **mobile devices**



Pre-history and ancient Egypt



Rodolphe Töpffer, Histoire de Monsieur Cryptogame (1830)



Ted McCall, Robin Hood And Company (1946)



Tezuka Osamu, Manga Classroom (1953)



Marion Montaigne, Tu mourras moins bête (2013)

# GT validation

# Knowledge-driven analysis detail

- Comic character region refinement