



## European Ph.D. pre-defense

Communauté  
d'Agglomération de  
**La Rochelle**

# Segmentation and indexation of complex objects in comic book images

Christophe Rigaud  
November 10<sup>th</sup>, 2014

Jointly supervised by:

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

Introduction

Documents

Contributions

Experiments

Conclusion

UNIVERSITÉ

La Rochelle

- Comic books
- Comics project at L3i



Thinking about the scenario. Image credits:  
Wissam Shekhani, 2010

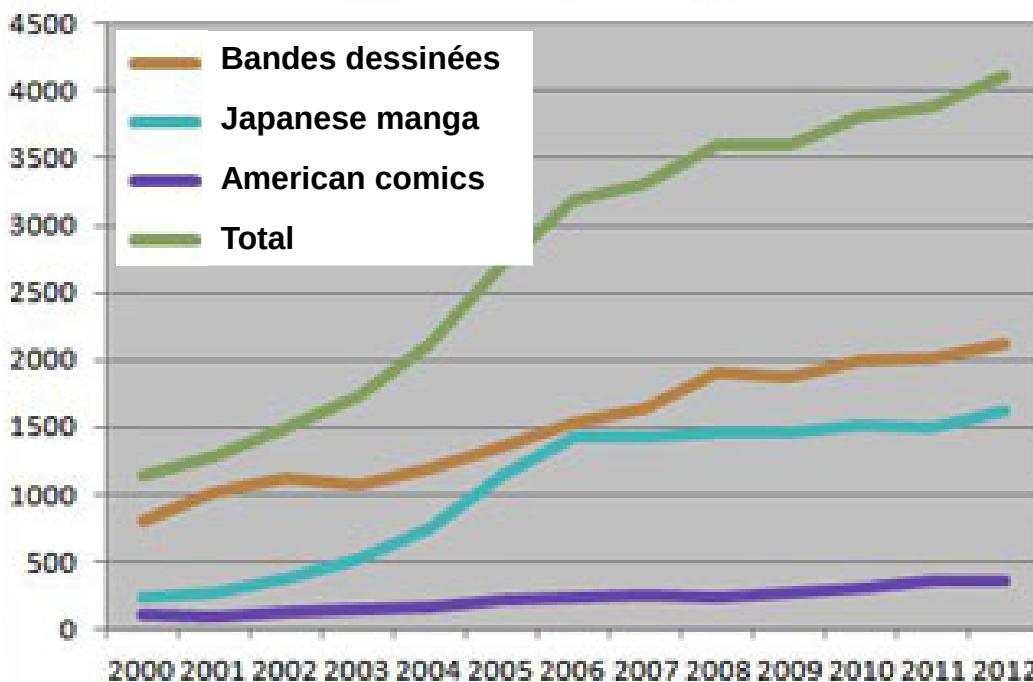
# Comic books

## Introduction

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

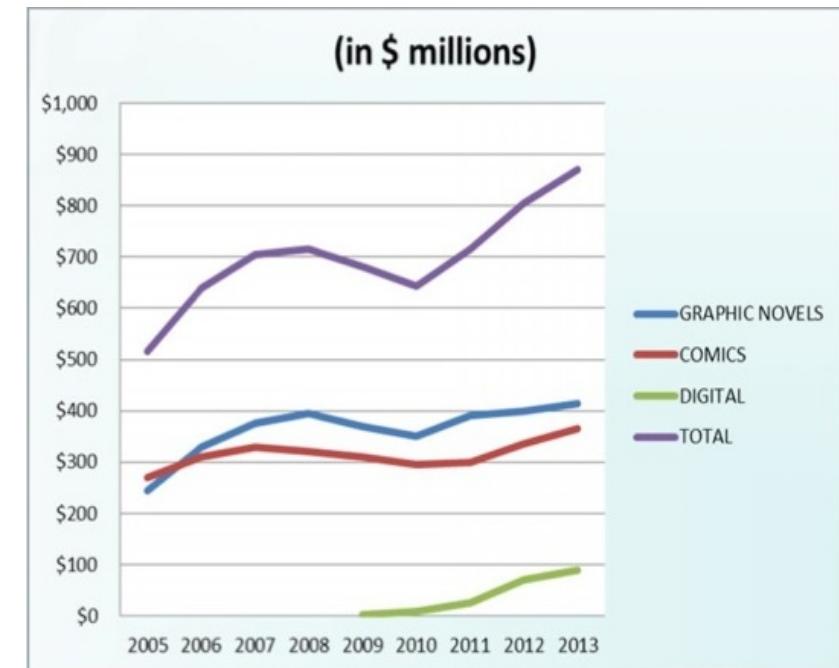
Wikipédia, 2014

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

# Comics project (eBDthèque)

## Introduction

- Objectives
  - Add value to paper-based comics using the new technologies
  - Answer to a real need from librarians, advertisers and readers
- Applications
  - Text/image search, reflowable documents, augmented reading and translation assistance
- Research axes
  - Extracting content of digitalized comic books (e.g. panels, balloons, text, comic characters)
  - Retrieving the semantic of the content (e.g. said/thought by, walking/running with, happened after/before)
- Participants
  - 2 Ph.D. students, 6 professors, 1 engineer (one year) and 1 post doc (one year)
- Initial public founding
  - CPER 2007-2013 (State-Region Project Contract)

# Comics project (eBDthèque)

## Introduction

- Objectives
  - Add value to paper-based comics using the new technologies
  - Answer to a real need from librarians, advertisers and readers
- Applications
  - Text/image search, reflowable documents, augmented reading and translation assistance
- Research axes
  - Extracting content of digitalized comic books (e.g. panels, balloons, text, comic characters)
  - Retrieving the semantic of the content (e.g. said/thought by, walking/running with, happened after/before)
- Participants
  - 2 Ph.D. students, 6 professors, 1 engineer (one year) and 1 post doc (one year)
- Initial public founding
  - CPER 2007-2013 (State-Region Project Contract)

- Document production to interpretation
- Application to comic books



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

# Production to interpretation

# Document analysis

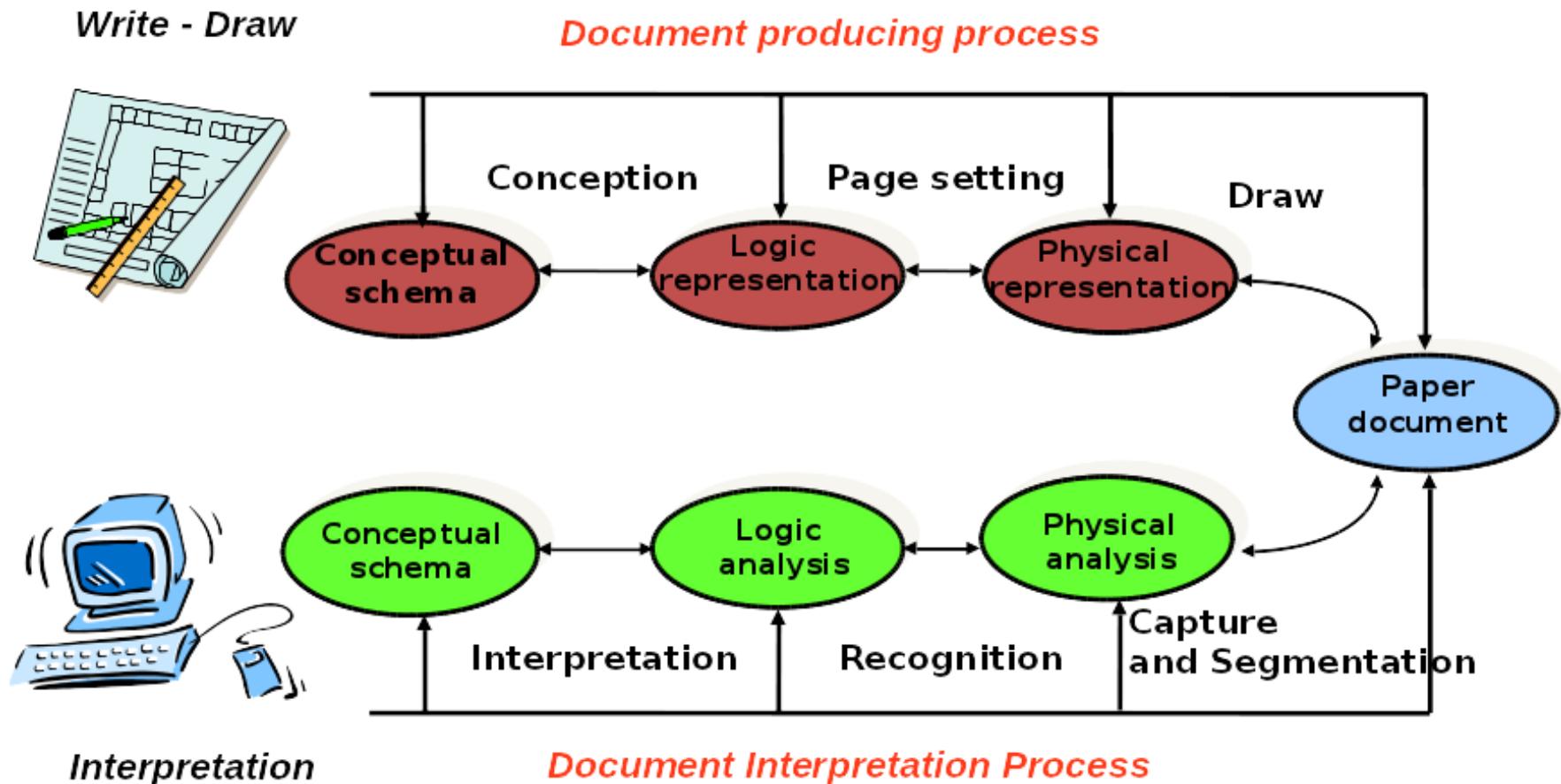


Image source: Handbook of Document Image Processing and Recognition. Springer, 2014

# 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 scenario
  - 2) Pencil drawing
  - 3) Inking
  - 4) Flattening and colouring
  - 5) Lettering and sound effects

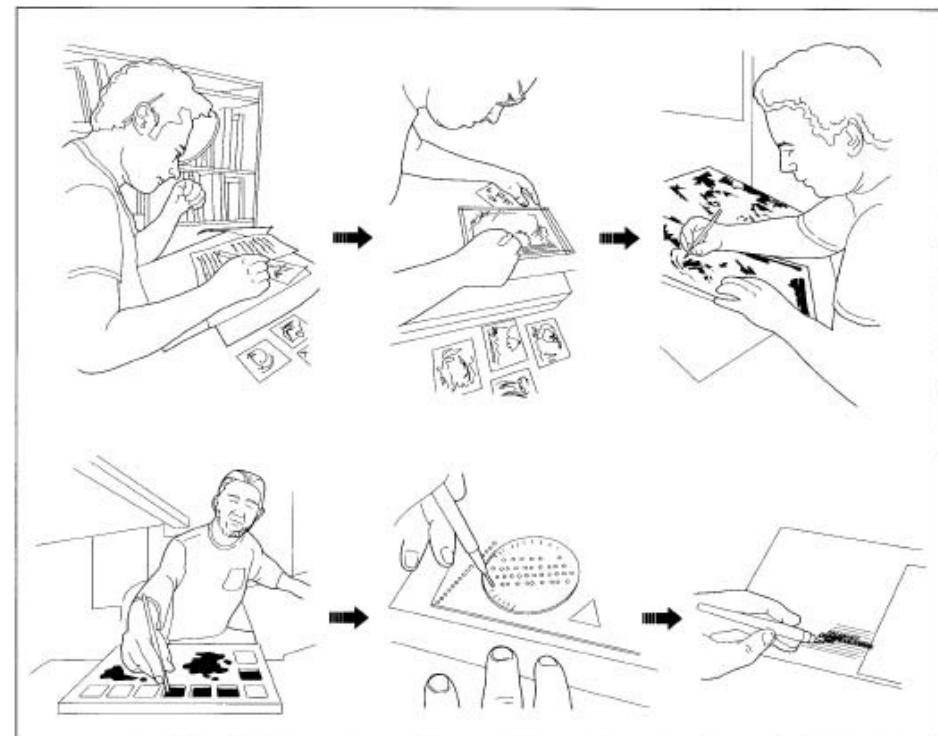


Image source:  
<http://www.madehow.com/Volume-6/Comic-Book.html>

- Introduction
- Content-driven approach
- Knowledge-driven approach

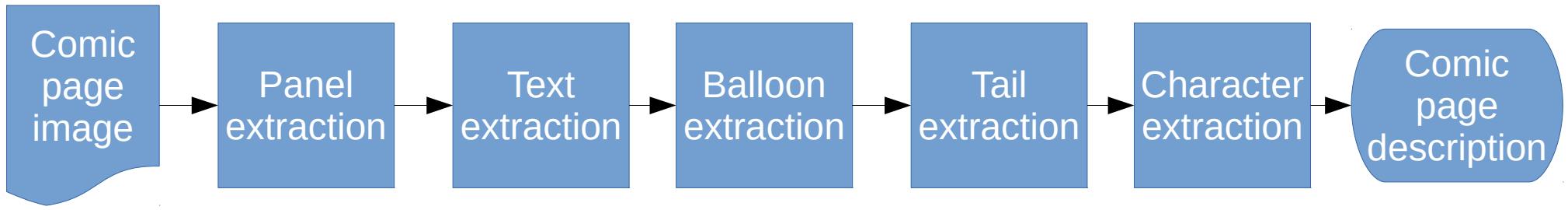


Inking. Image credits: Le cycle des bulles,  
Christophe Rigaud, 2012

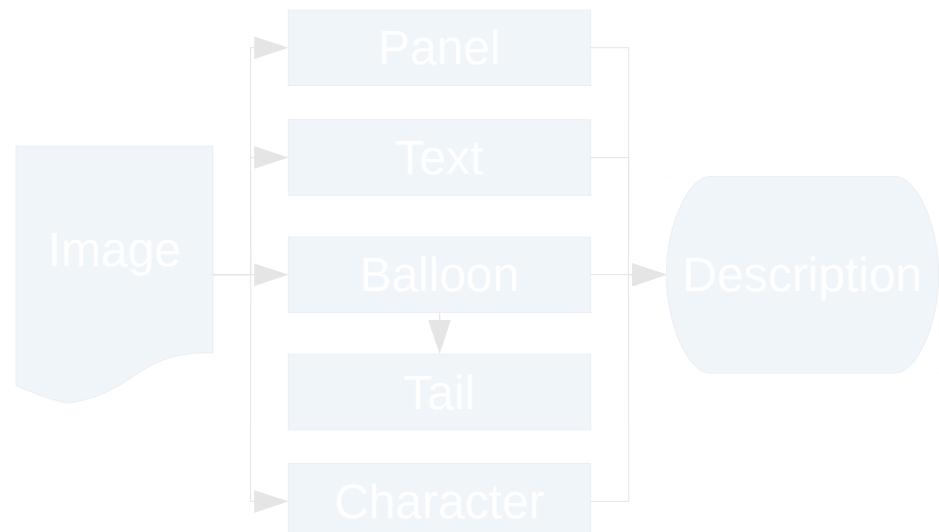
# Introduction

# Contributions

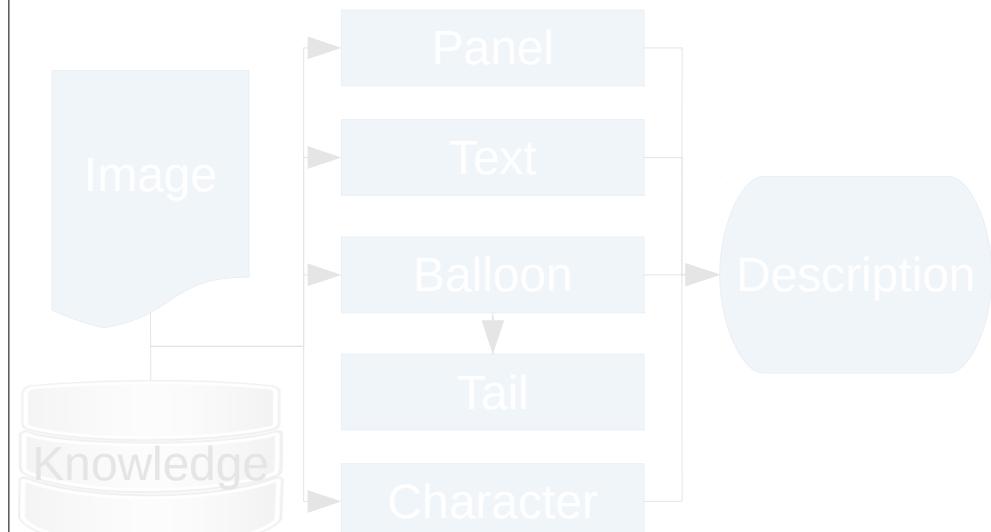
## Content-driven (sequential)



## Content-driven (independent)



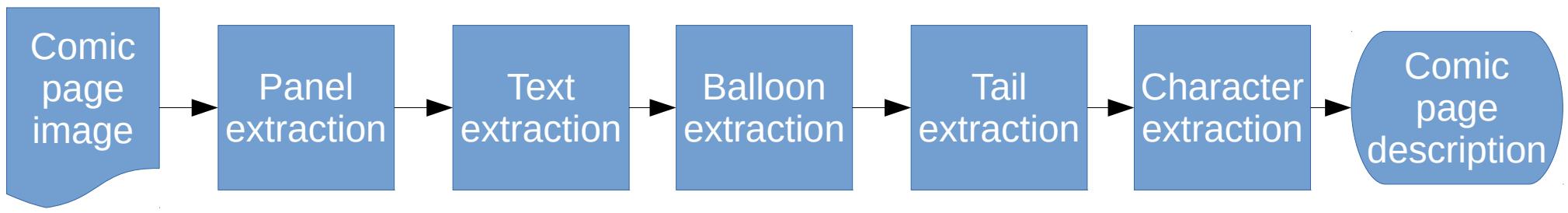
## Knowledge-driven (independent)



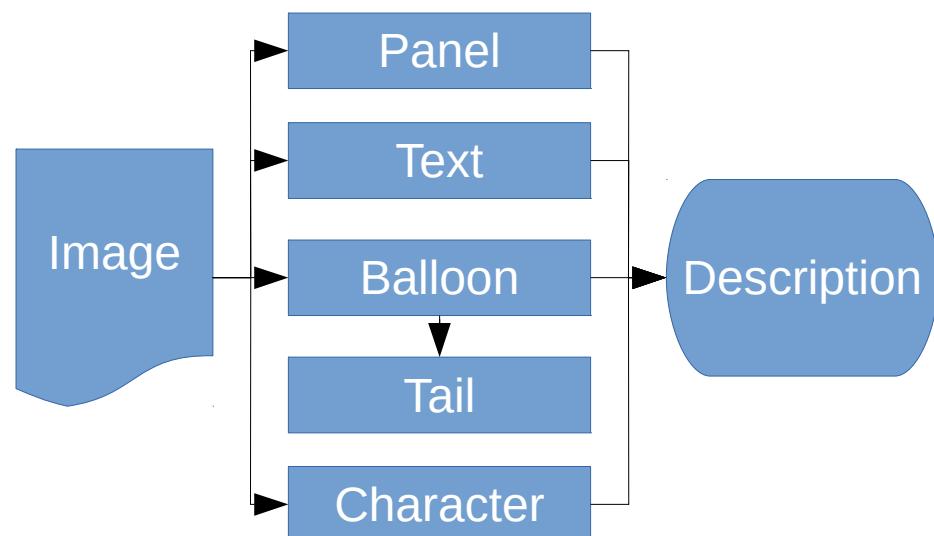
# Introduction

# Contributions

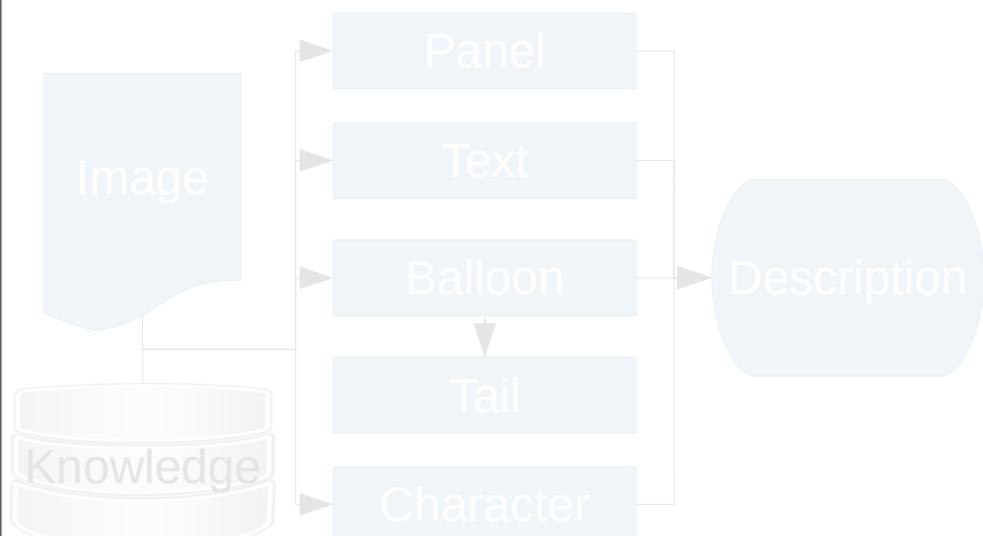
## Content-driven (sequential)



## Content-driven (independent)



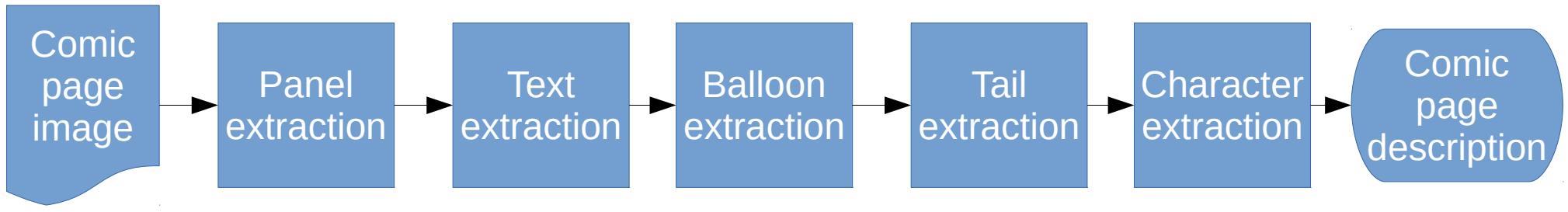
## Knowledge-driven (independent)



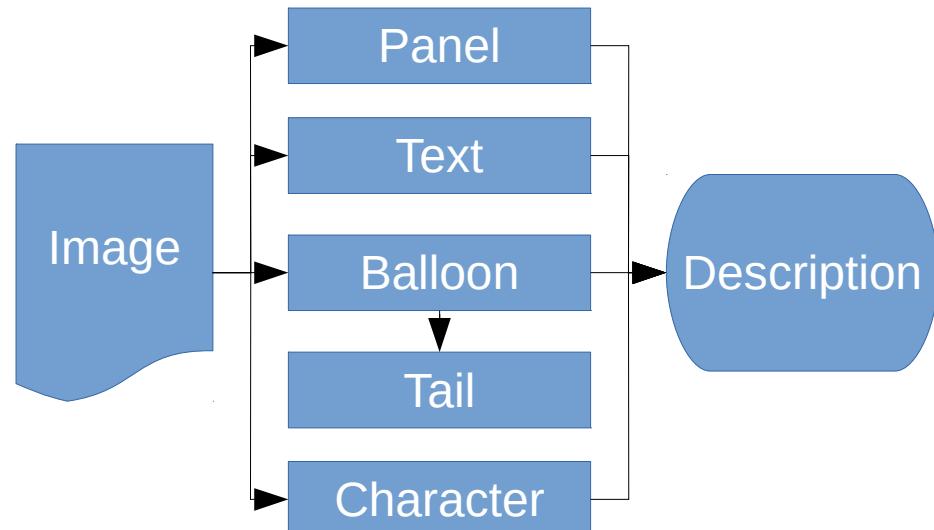
# Introduction

# Contributions

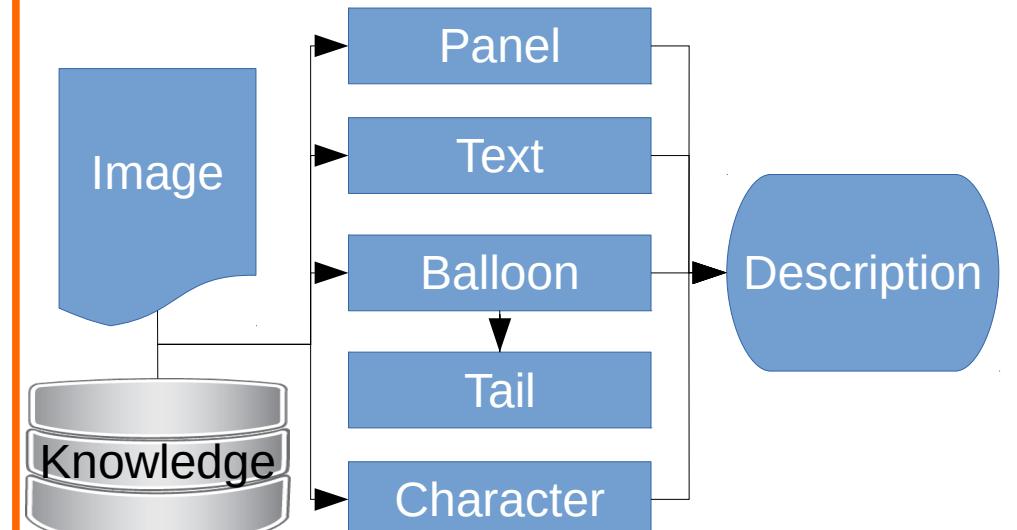
## Content-driven (sequential)



## Content-driven (independent)



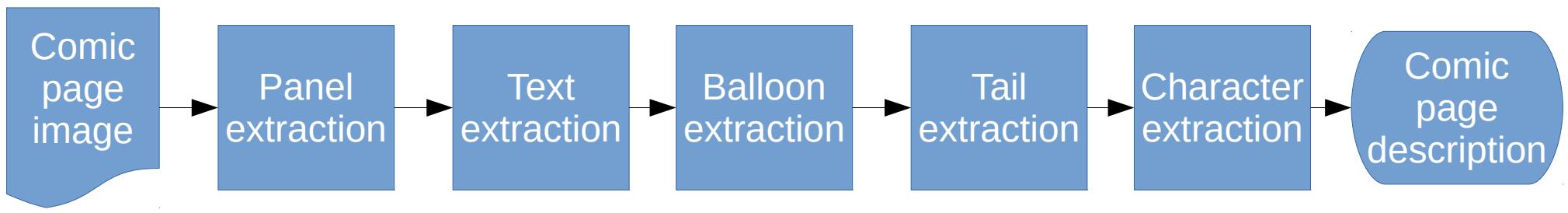
## Knowledge-driven (independent)



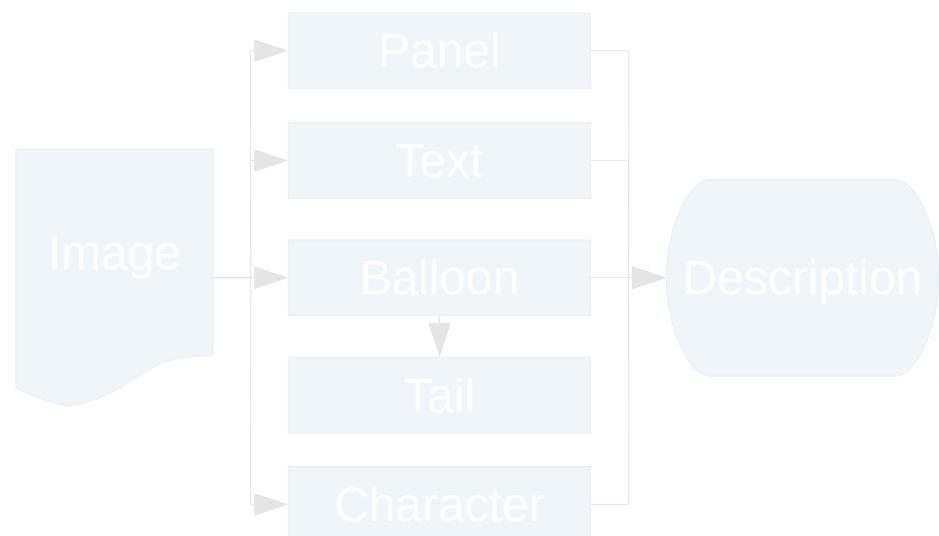
# Introduction

# Contributions

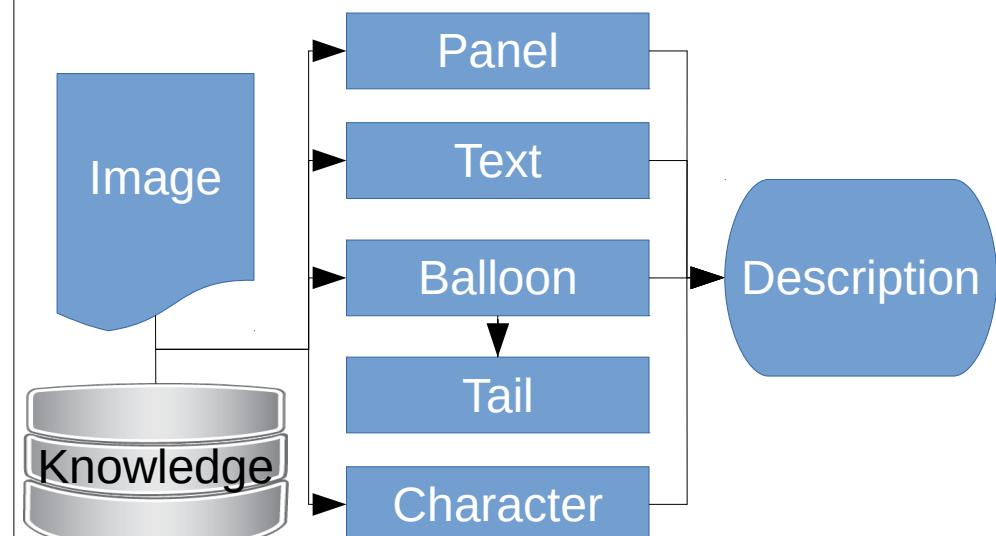
## Content-driven (sequential)



## Content-driven (independent)



## Knowledge-driven (independent)



- Introduction
- Content-driven approach
  - Panel & text extraction
  - Balloon extraction
  - Tail extraction
  - Comic character extraction
  - Conclusion
- Knowledge-driven approach

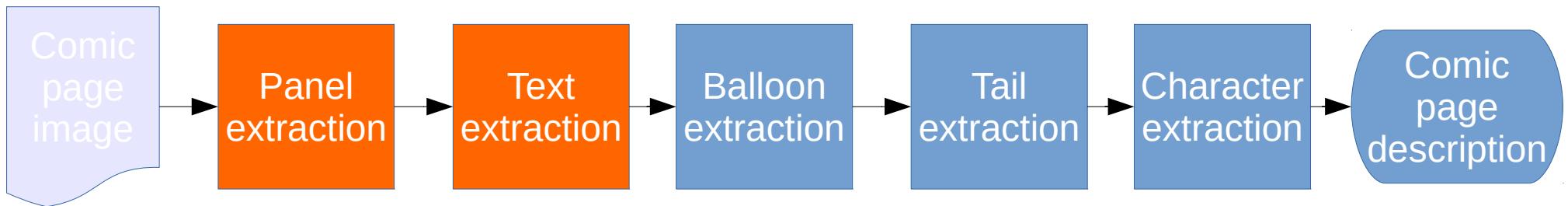


Gray-scale Image credits: Le cycle des bulles,  
Christophe Rigaud, 2012

# Panel and text extraction

Contributions

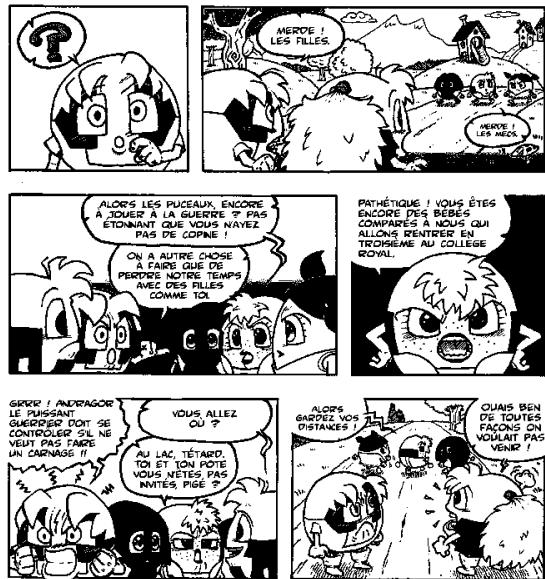
- Processing sequence



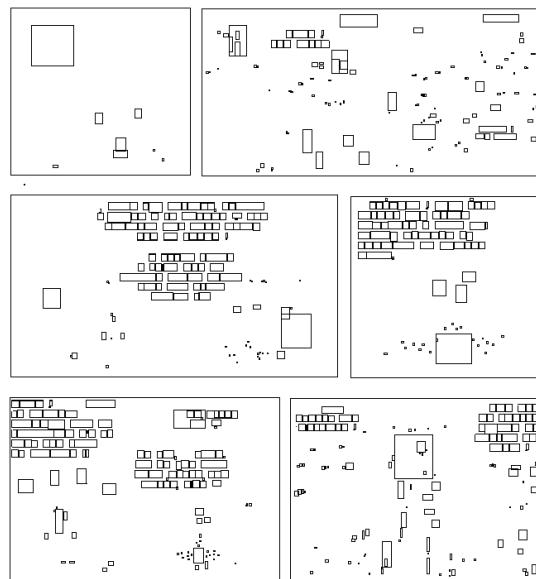
- Literature
  - Panel, balloon, text extraction [Ho2012, Arai2010]
  - Text only extraction [Li2013]
- Contribution
  - Simultaneous panel and text extraction

# Panel and text extraction

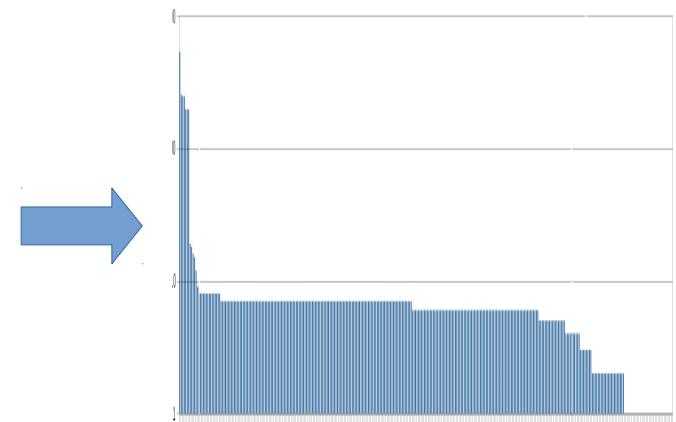
# Contributions



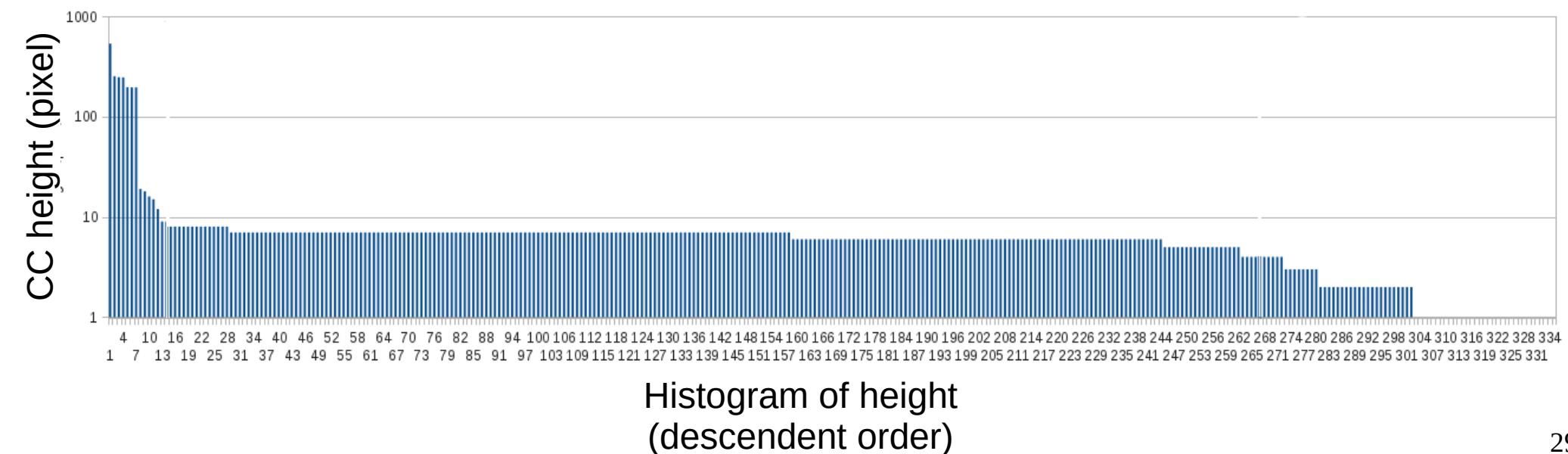
Binary image



Connected-component (CC) bounding boxes



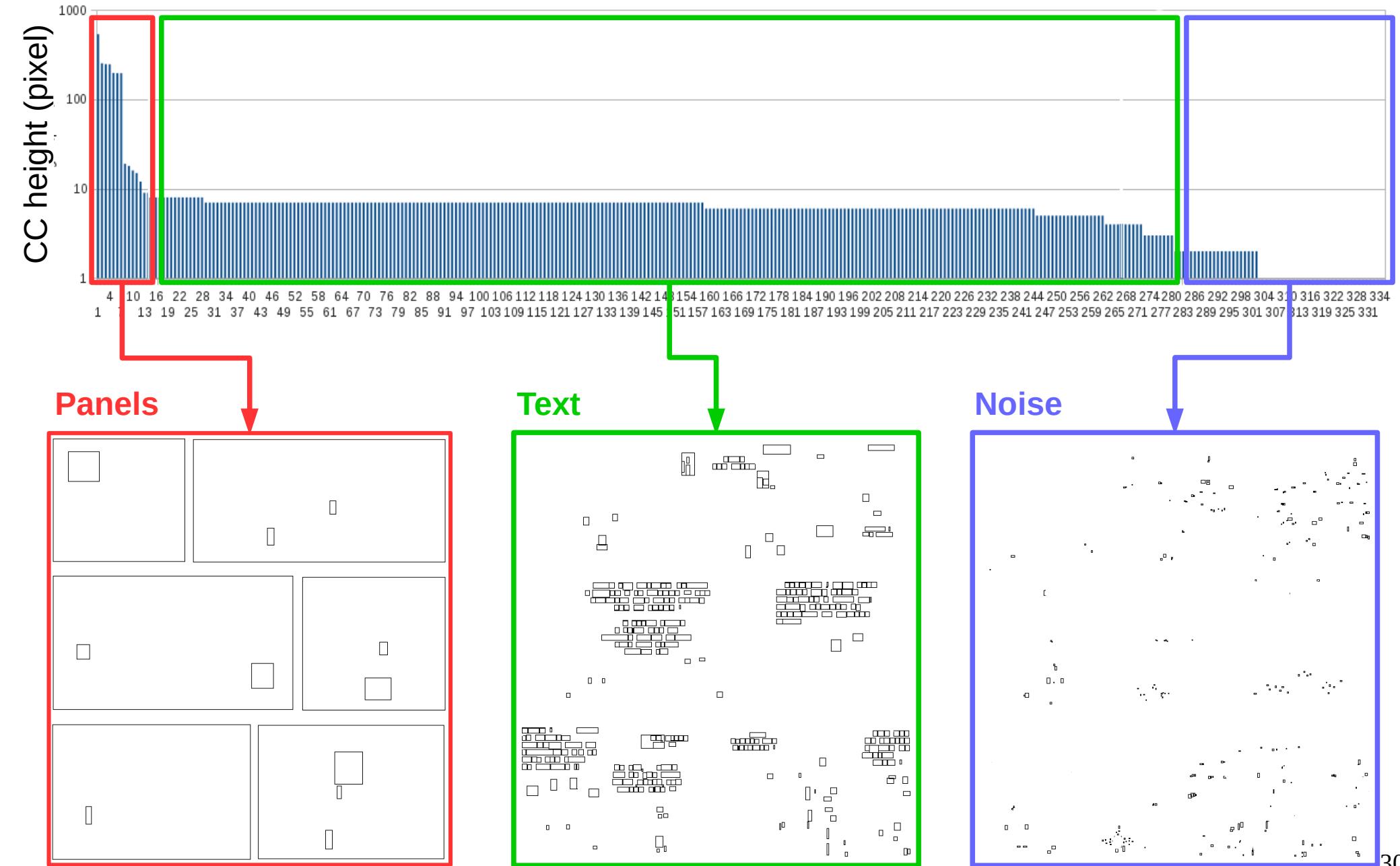
Histogram of heights of CC



# Panel and text extraction

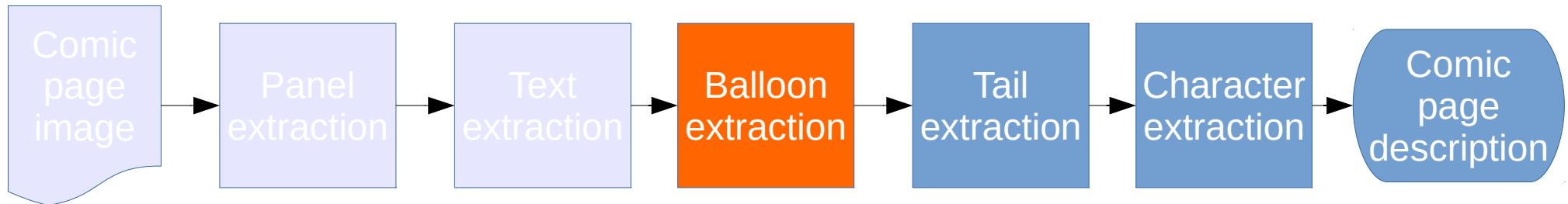
# Contributions

K-means clustering ( $k=3$ )

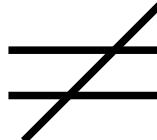


# Balloon extraction

# Contributions



Regular balloon

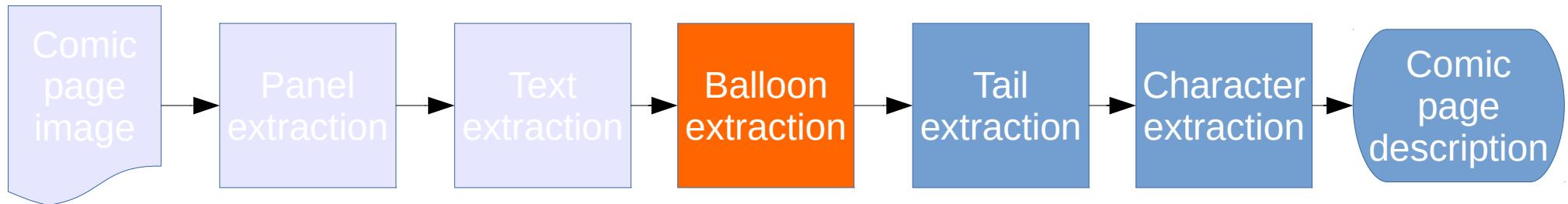


Implicit balloon

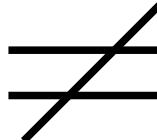
- Literature
  - Top-down approaches: extract white blobs and then text inside
  - Limited to regular balloons
- Contribution
  - Bottom-up approaches: extract text and then surrounding balloons
  - Appropriate for regular and implicit balloons

# Balloon extraction

# Contributions



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

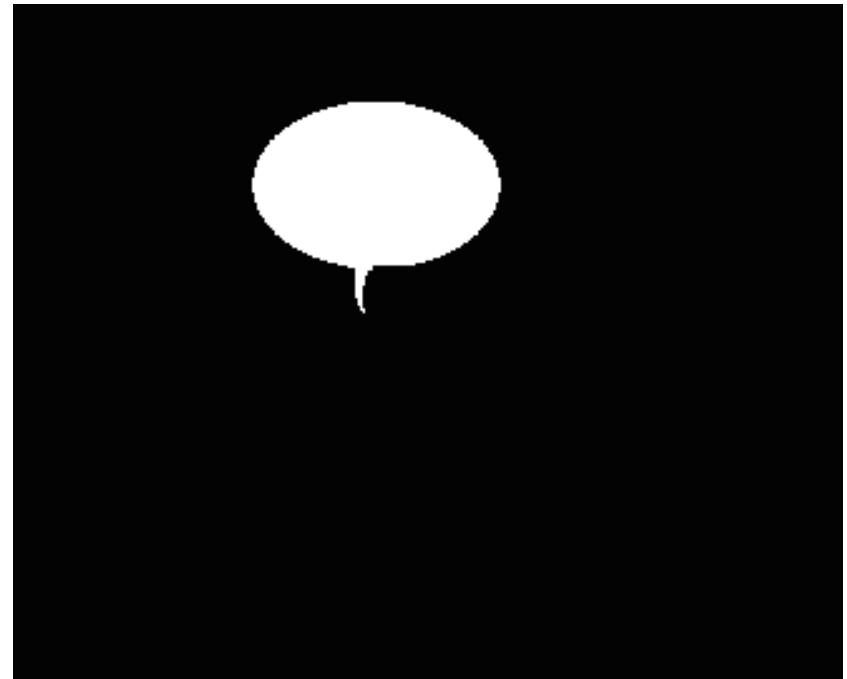
## Contributions

- 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

- 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

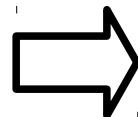
Expected result

# Balloon extraction: regular

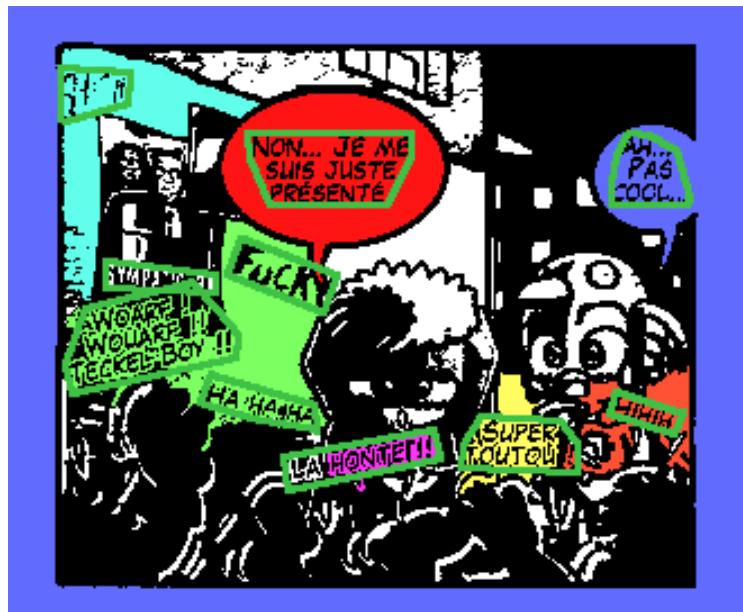
Content-driven



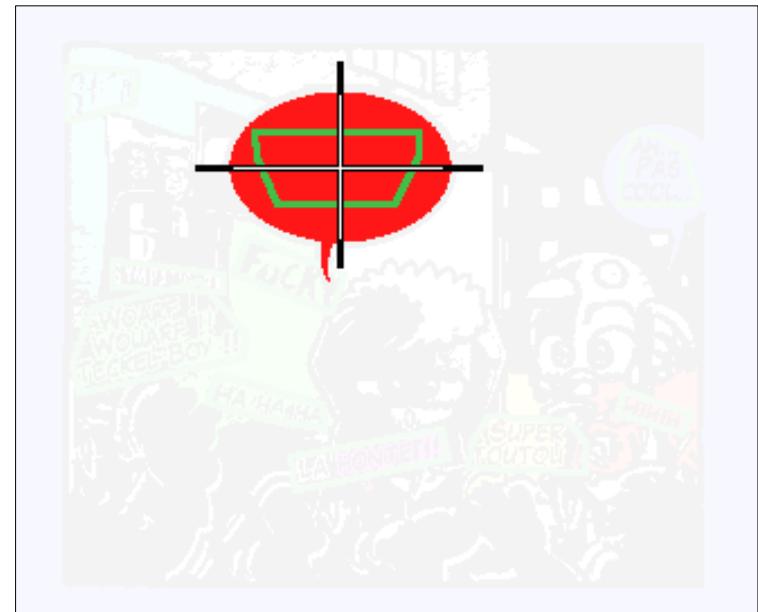
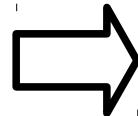
Original image



Text block positions (green)



Regions including text blocks (coloured)



Regions including aligned text block

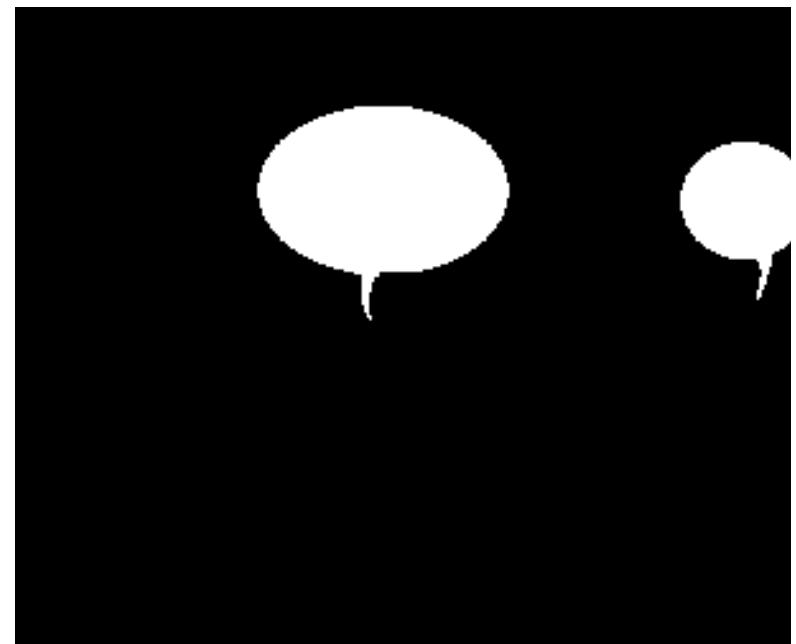
# Balloon extraction: implicit

Content-driven

- 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)



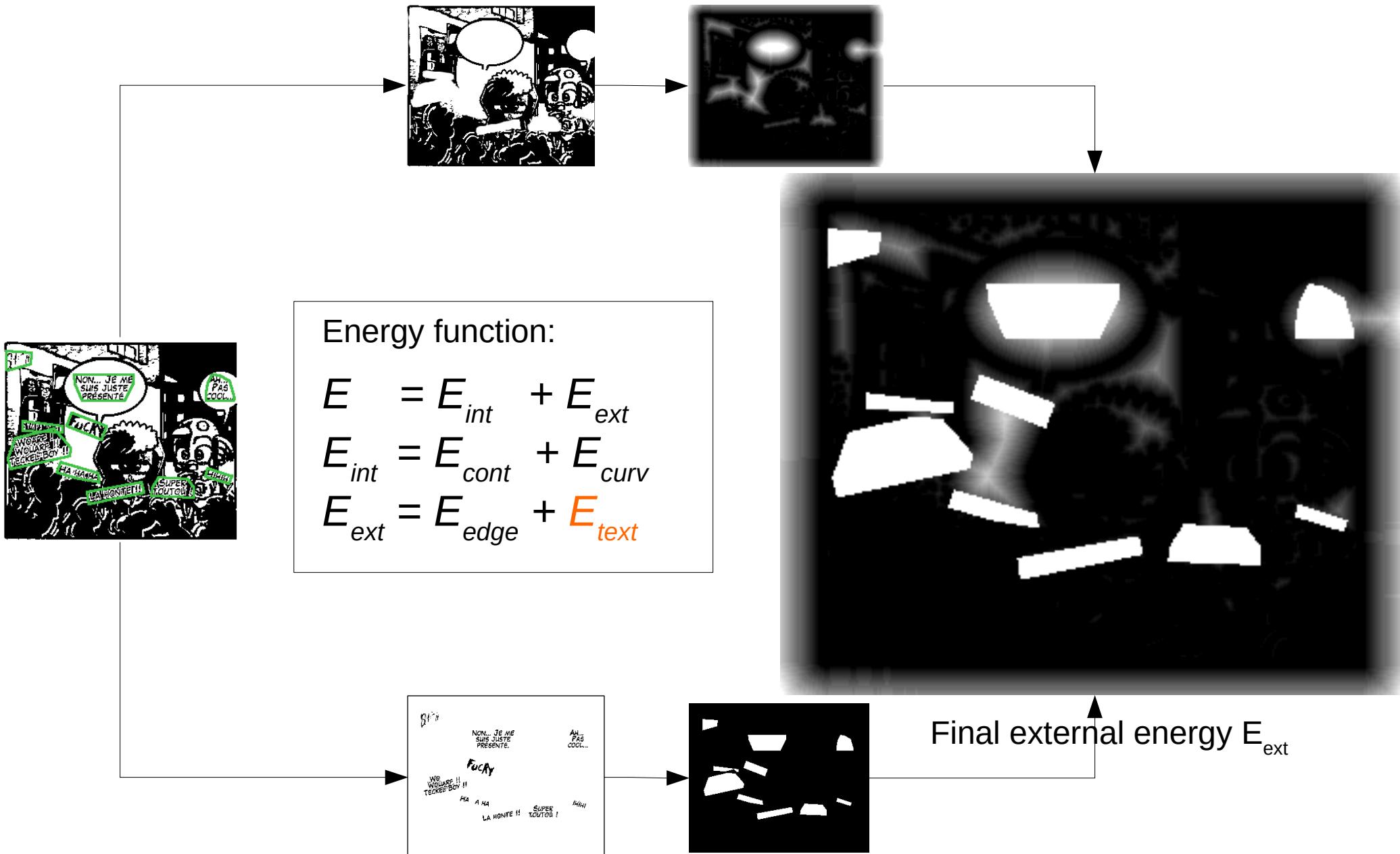
Original image and text locations



Expected result

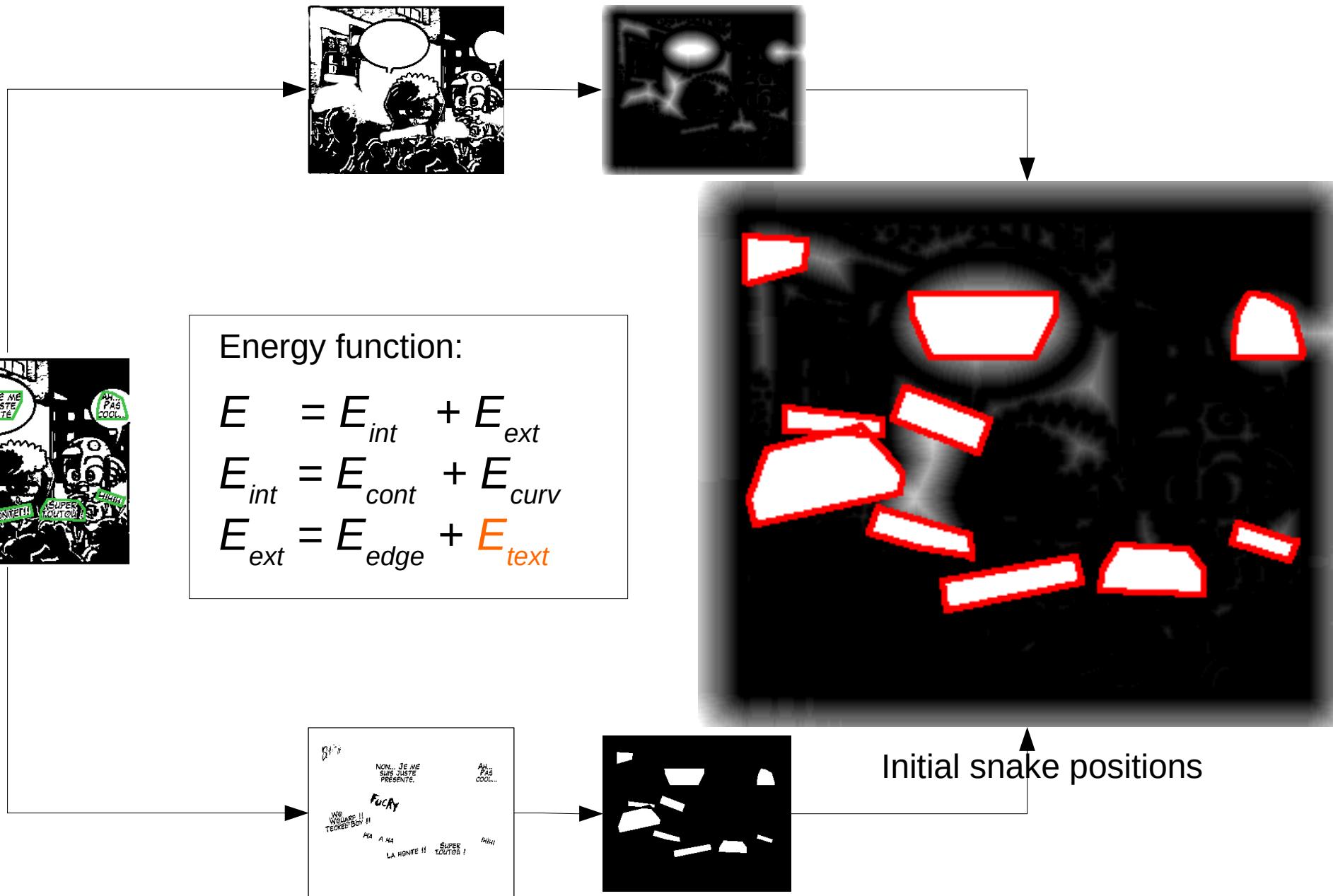
# Balloon extraction: implicit

Content-driven



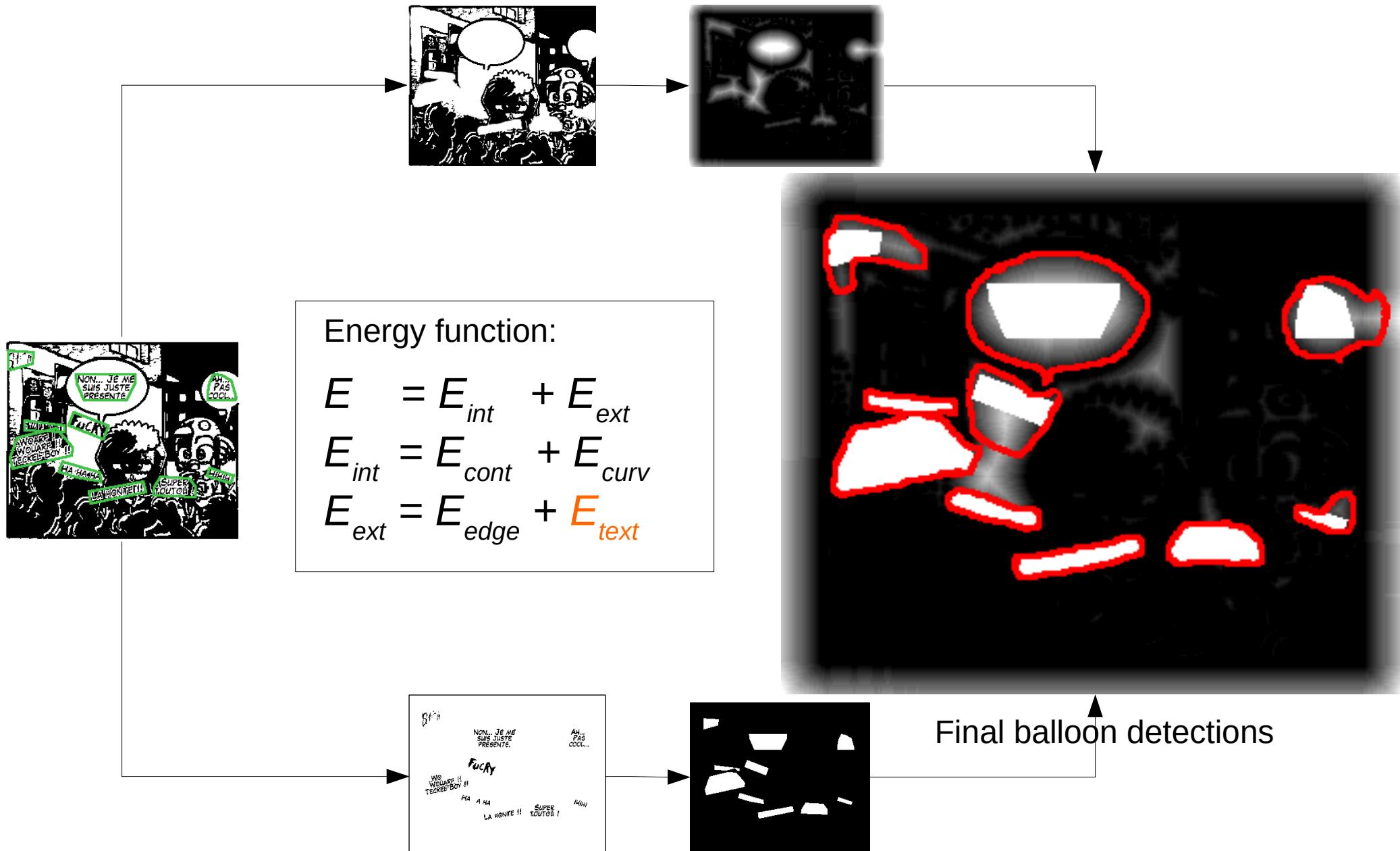
# Balloon extraction: implicit

Content-driven



# Balloon extraction: implicit

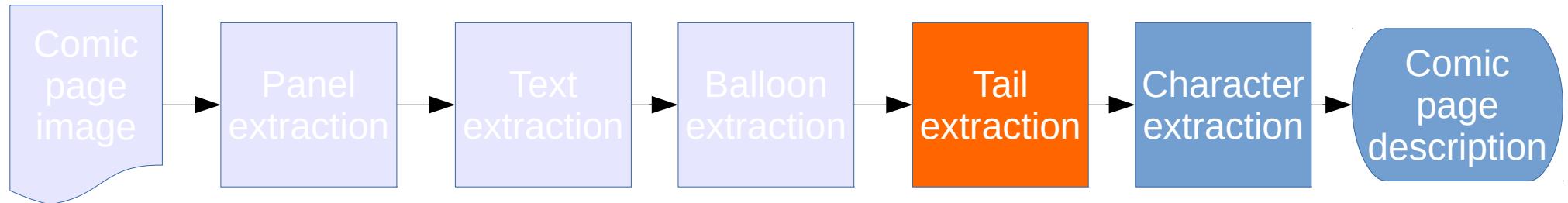
Content-driven



The snake is attracted to the “dark side”

# Tail extraction

Content-driven



- Definitions

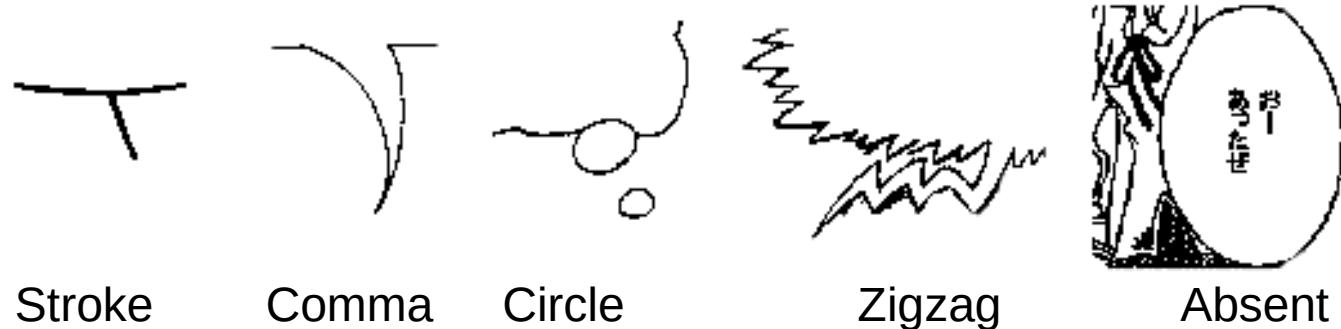
- Tail: a pointer that indicates where is the emitter (speaker) of the balloon
- Tail tip: the extremity of the tail, the closest part to the emitter
- Tail direction: direction pointed by the tail

- Literature

- First time studied in image processing

- Objectives

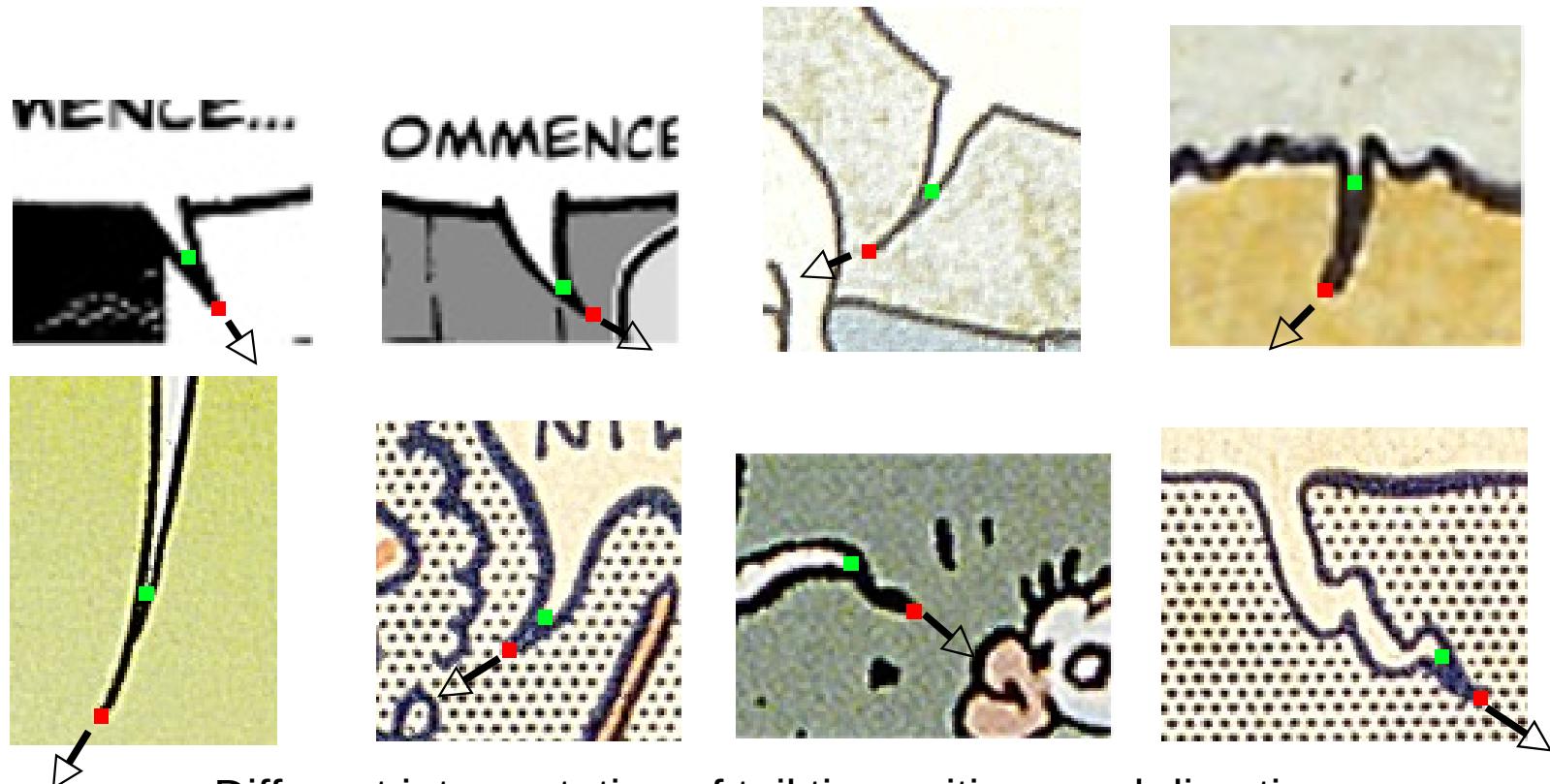
- Detection of the tail tip position and orientation
- Focus on comma, zigzag and absent types



# Tail extraction

Content-driven

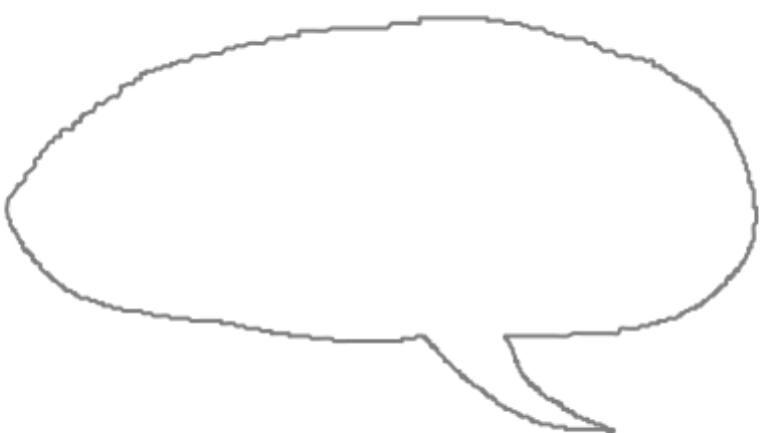
- Tip from background
- Tip from contour
- Direction of the tail



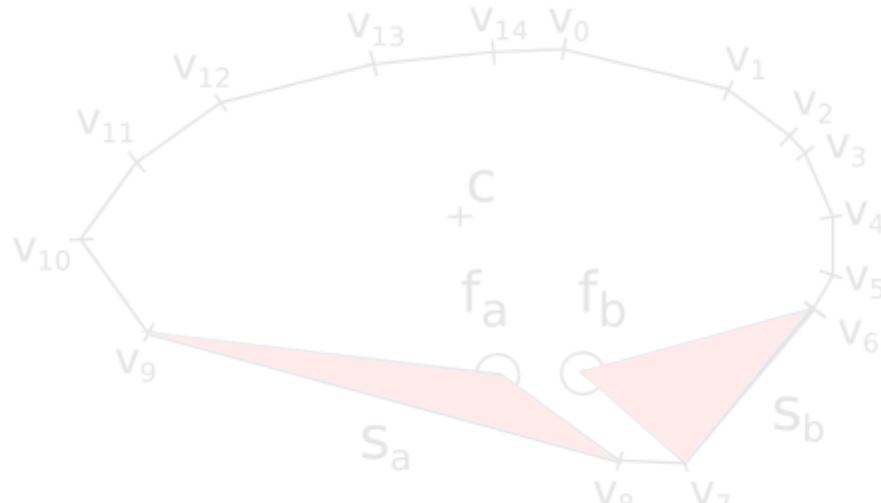
Different interpretation of tail tip positions and directions

# Tail extraction: tip position

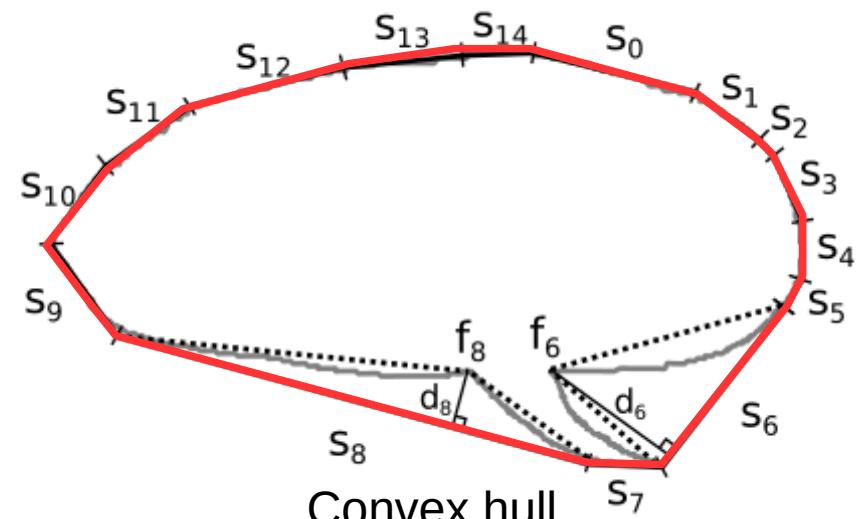
Content-driven



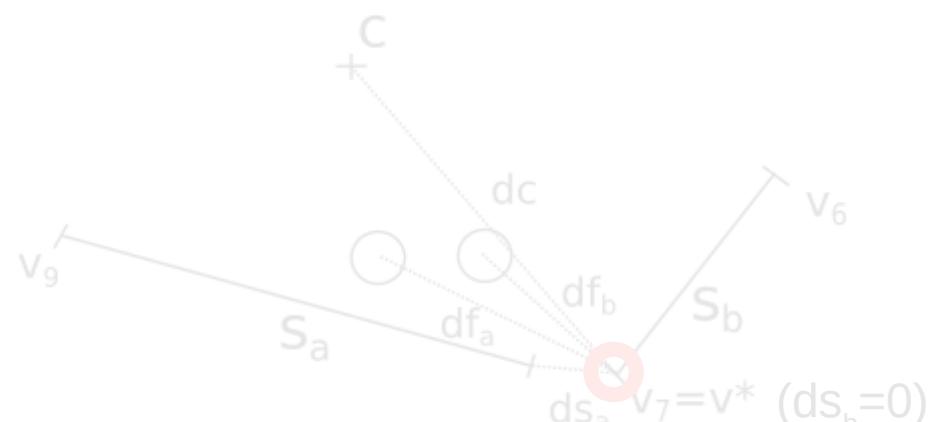
Balloon contour



Two biggest  
convexity defects



Convex hull



Tail tip position

Optimal vertex selection:

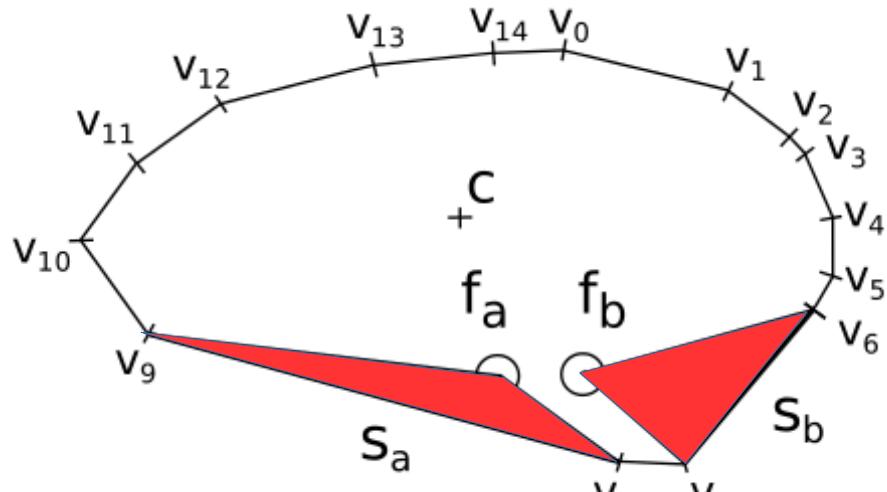
$$v* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

# Tail extraction: tip position

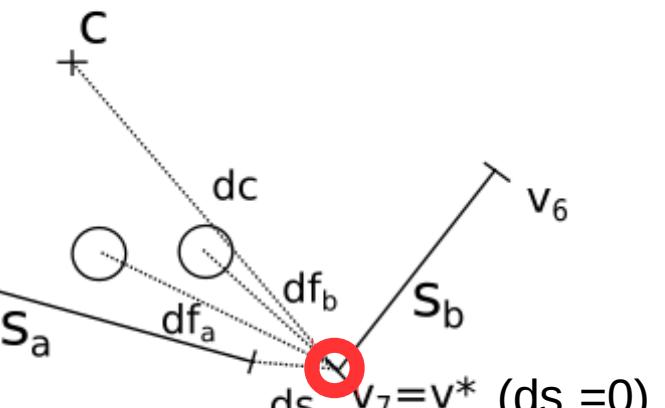
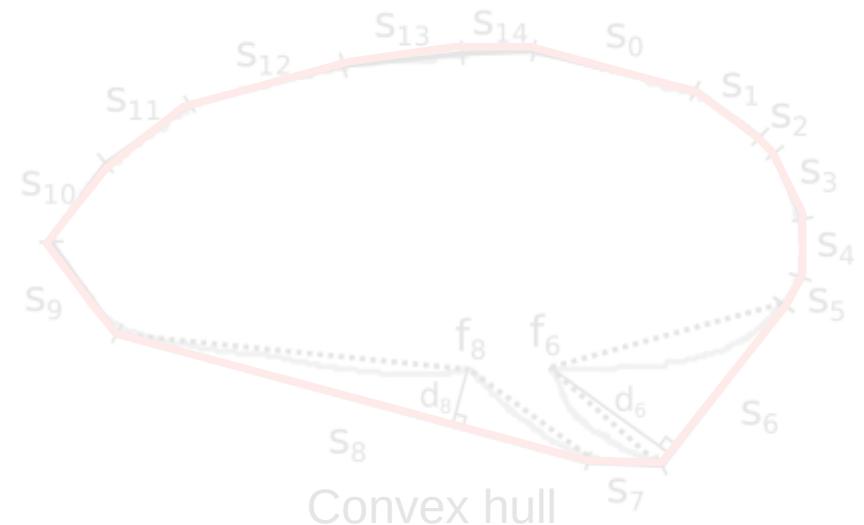
Content-driven



Balloon contour



Two biggest  
convexity defects



Tail tip position

Optimal vertex selection:

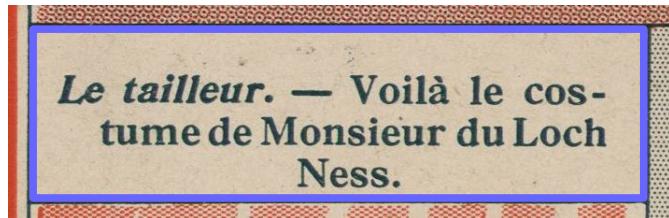
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

# Tail extraction: confidence value

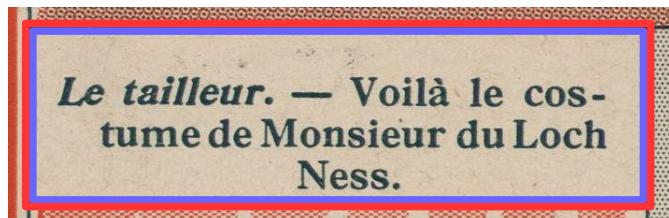
Content-driven

Balloon  
contour (blue)

Balloon 1



Convex hull  
(red)



Confidence

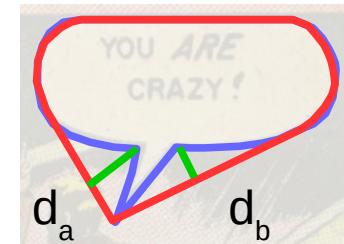
$$C_{tail} = \frac{(d_a + d_b)/2}{meanBalloonSize}$$

$$C_{tail} = 0.0$$

Presence of tail

NO

Balloon 2



$$C_{tail} = 0.73$$

YES (>0)

# Tail extraction: tail direction

Content-driven

- 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)



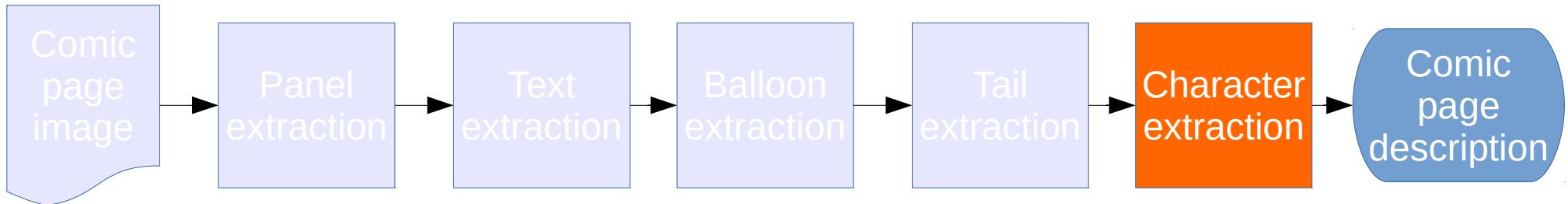
Farthest point  
from origin and tip  
(red)



Direction from tip  
to farthest point  
(white arrow)

# Comic character extraction

Content-driven



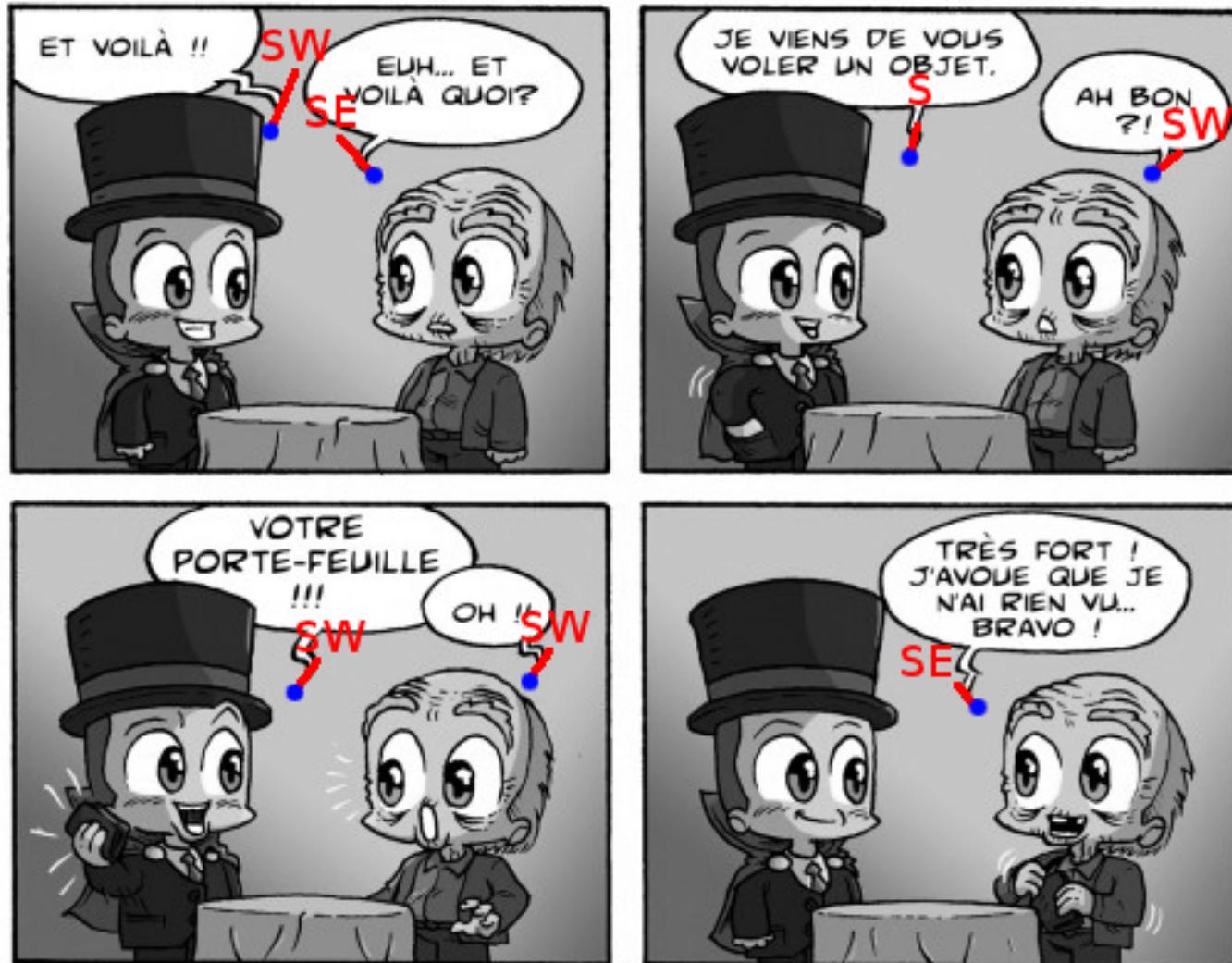
- Literature
  - Supervised approaches for manga and cartoon characters
  - No public dataset (copyright issues)
- Challenges
  - Variety of styles of comic books
  - Intra and extra class variations of each character instance (e.g. position, scale, pose, occlusion and human-like, invented)
- Objectives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

# Comic character extraction

Content-driven

## Prerequisites

- Position and direction of the tails in the panels
- Our proposition
  - use already extracted information (panels, balloons and tails) to estimate speaking character region of interests



# Comic character extraction

Content-driven

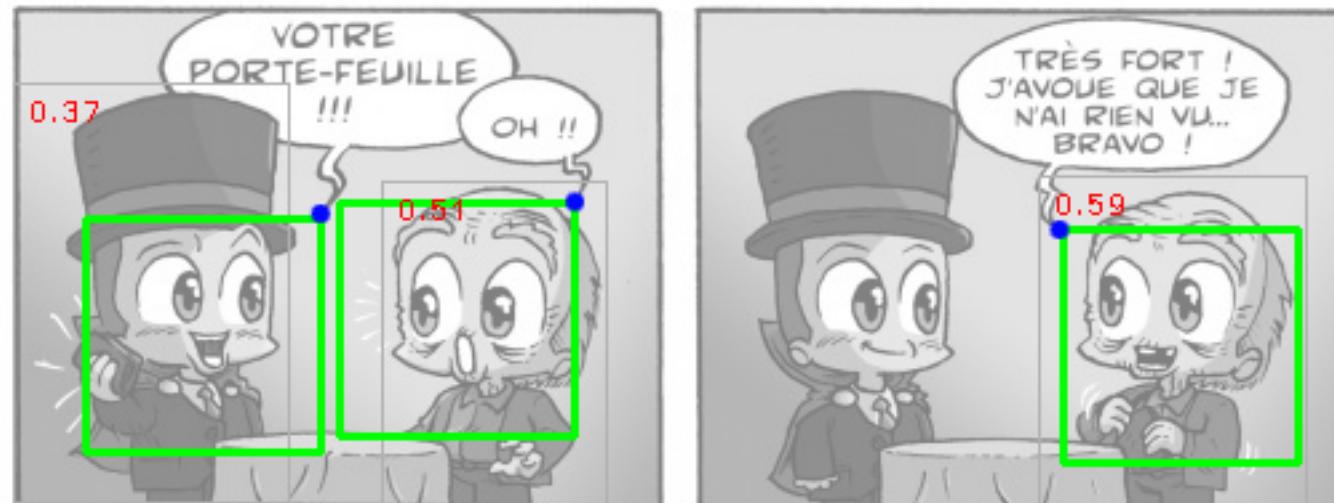
## Prerequisites

- Position and direction of the tails in the panels
- Our proposition
  - use already extracted information (panels, balloons and tails) to estimate speaking character region of interests

## Rectangle ROI



## Squared ROI



# Conclusion

Content-driven

- Panel & text extraction
  - Simple and **fast**
  - **Disconnected panels only**
- Balloon extraction
  - Accurate for closed balloons
  - Implicit balloons would require more **contextual** information
- Tail extraction
  - Relies on the **balloon contour** extraction
  - Limited to **certain types** of tail
- Comic character extraction
  - **Highly relies** on the quality of the tail extraction
  - **Implicitly retrieves the relationship** between balloons and characters
  - **Only for “speaking” characters**
- Global
  - **Content-driven** approach
  - Sequential element processing
  - From **simple to complex** elements
- Pros
  - **Consistence** between extracted elements
- Cons
  - **Error propagation**

- Introduction
- Content-driven approach
- Knowledge-driven approach
  - Introduction
  - Knowledge models
  - Processing sequence
  - Conclusion

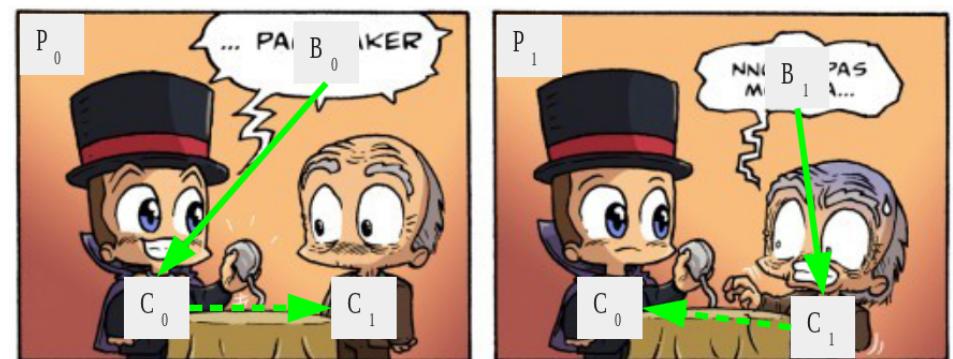
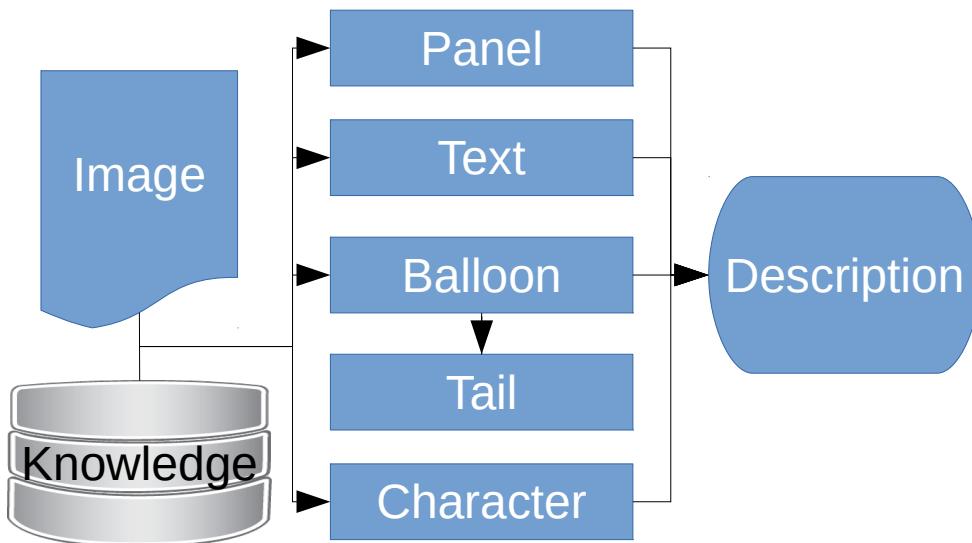


Colouring. Image credits: Le cycle des bulles, Christophe Rigaud, 2012

# Introduction

Knowledge-driven

- Objective
  - High level image description for comics understanding (retrieve semantic information)
  - Independent element extraction (prevent error propagation)
  - Collaboration with Clément Guérin (Ph.D student from the same project)
- Process flow

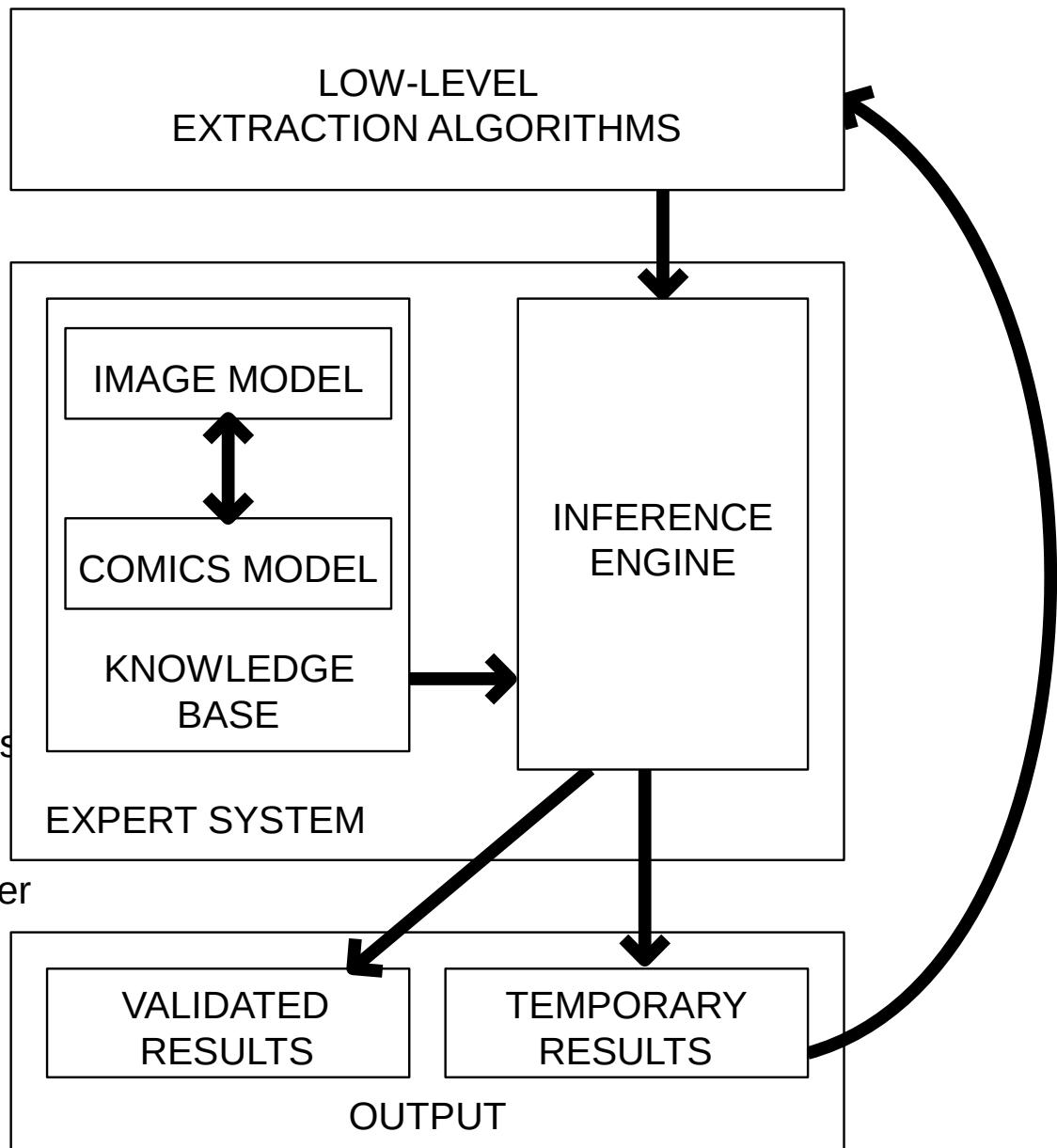


Example of situation understanding

# Knowledge models

Knowledge-driven

- Image domain model
  - Comics image analysis
- Comics domain model
  - Validation
    - A panel **P** is related to one page
    - A balloon **B** is related to one panel
    - A character **C** is related to one panel
    - A text line **T** is related to one balloon
  - Inference
    - A balloon with a tail **Q** is a speech balloon **SB**
    - A text contained in a speech balloon is speech text **ST**
    - A comic character associated to a speech balloon is a speaking character **SC**



# Processing sequence

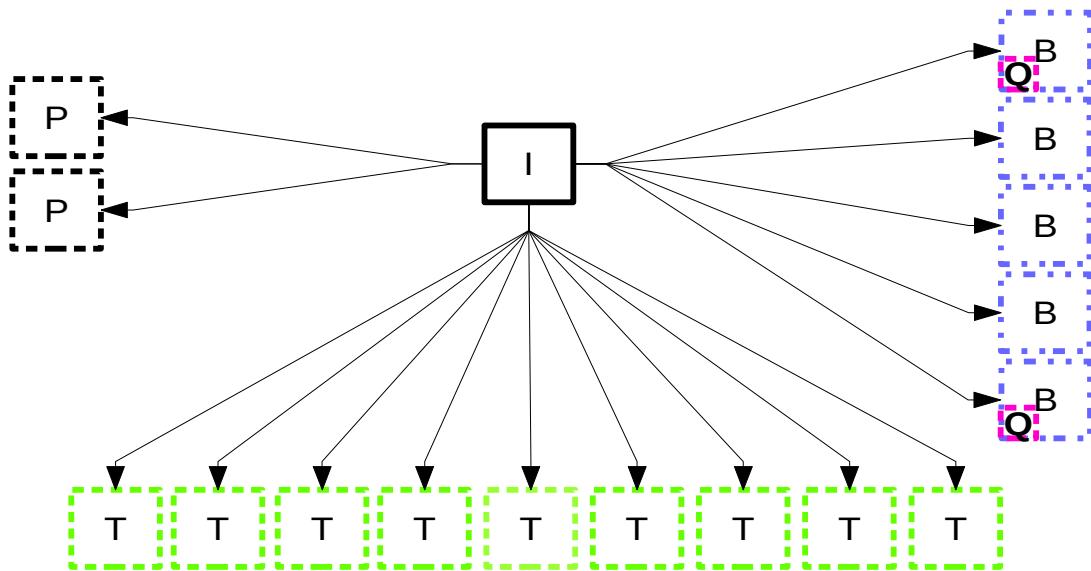
Knowledge-driven

- Process loop
  - Step 1: **formulate** hypothesis
  - Step 2: **validate** hypothesis
  - Step 3: **infer** new information
- Study case
  - Iteration 1
    - Step 1: hypotheses of **simple element** positions
    - Step 2: validation of the positions
    - Step 3: inference a new information
  - Iteration 2
    - Step 1: hypotheses of **more complex** elements
    - Step 2: validation of the positions
    - Step 3: inference a new information
  - ...



# Processing sequence 1/2

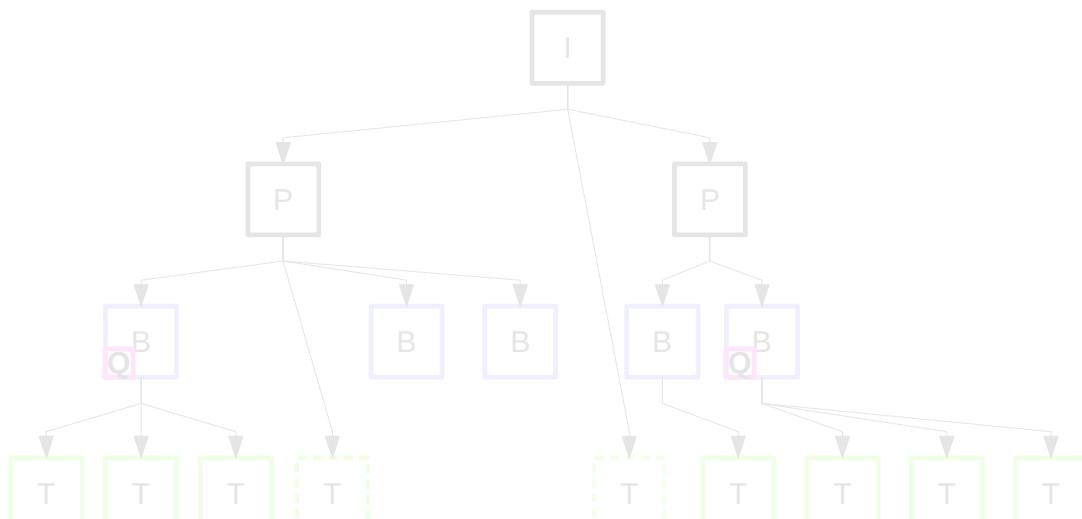
## Step 1: formulation of hypotheses



Knowledge-driven

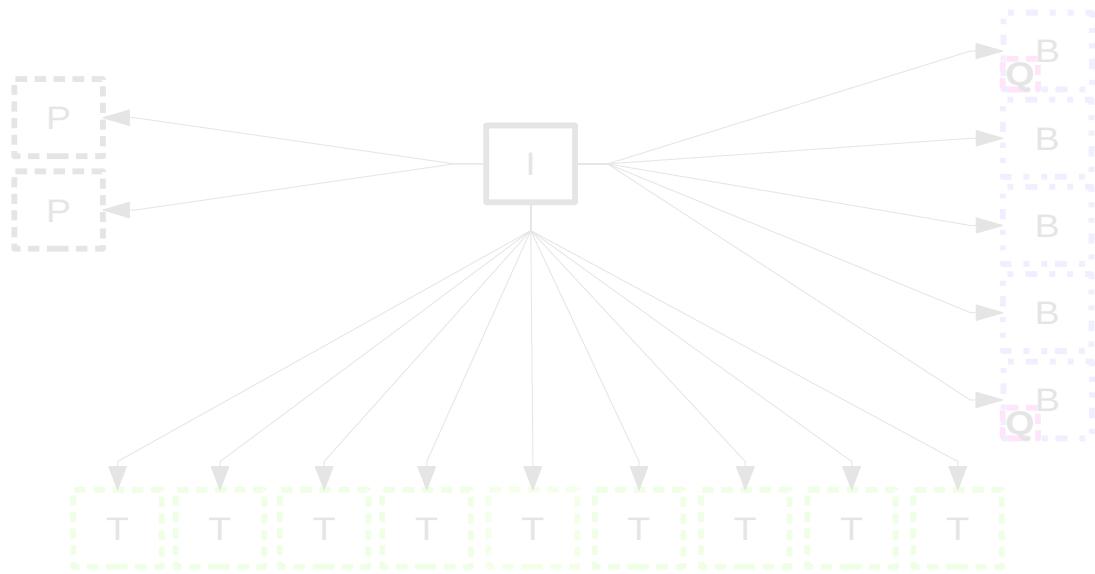


## Step 2: validation of the hypotheses

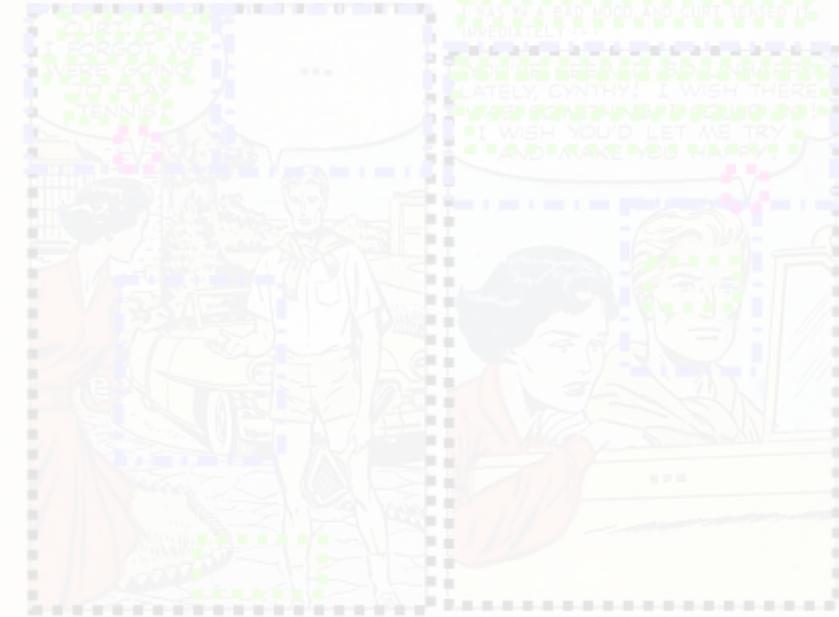


# Processing sequence 1/2

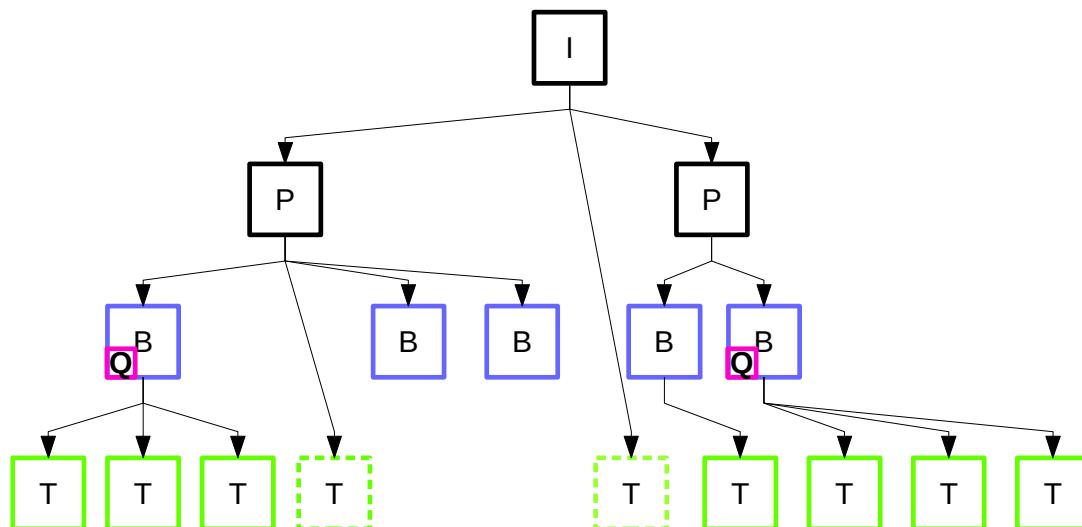
Step 1: formulation of hypotheses



Knowledge-driven



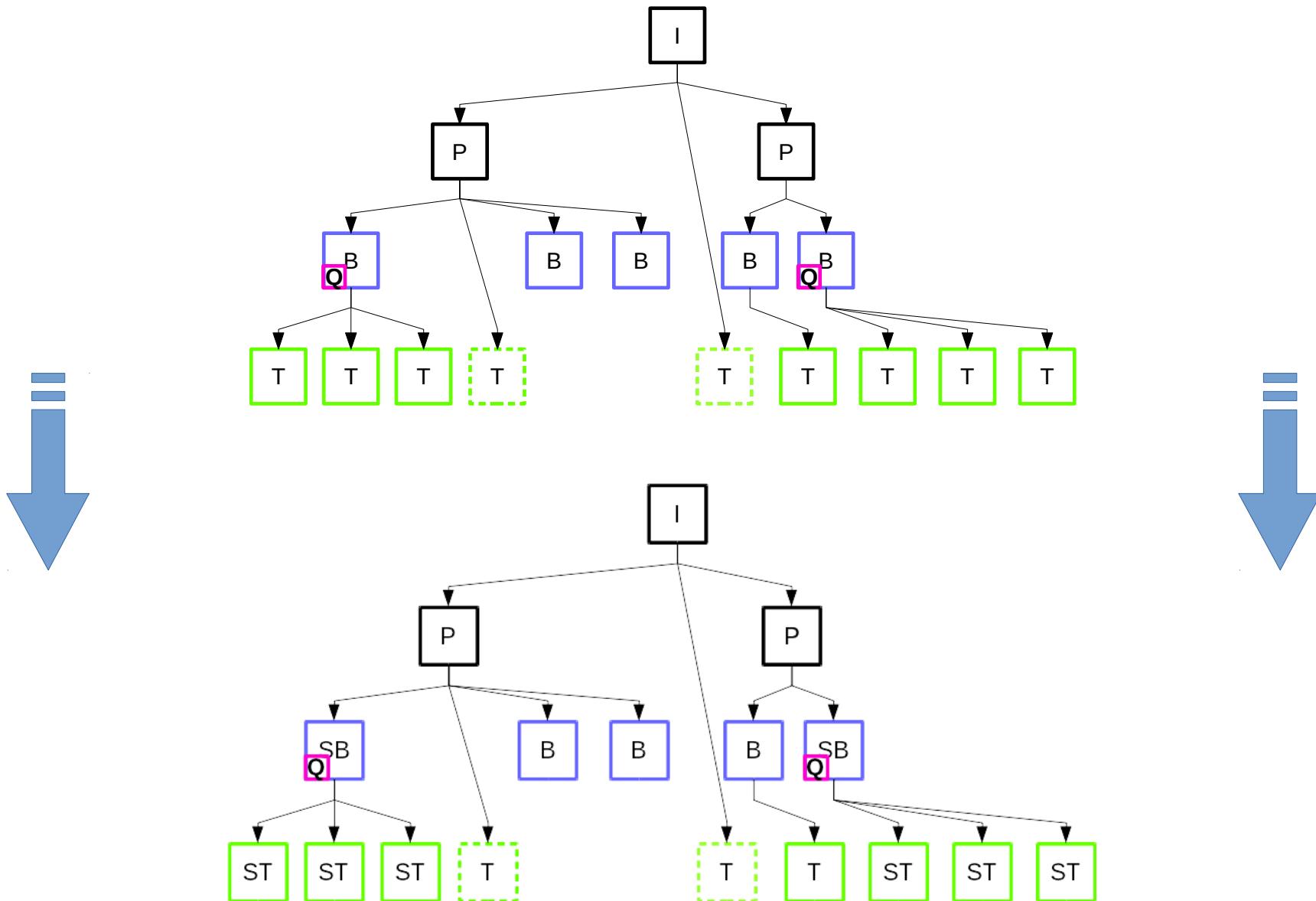
Step 2: validation of the hypotheses



# Processing sequence 1/2

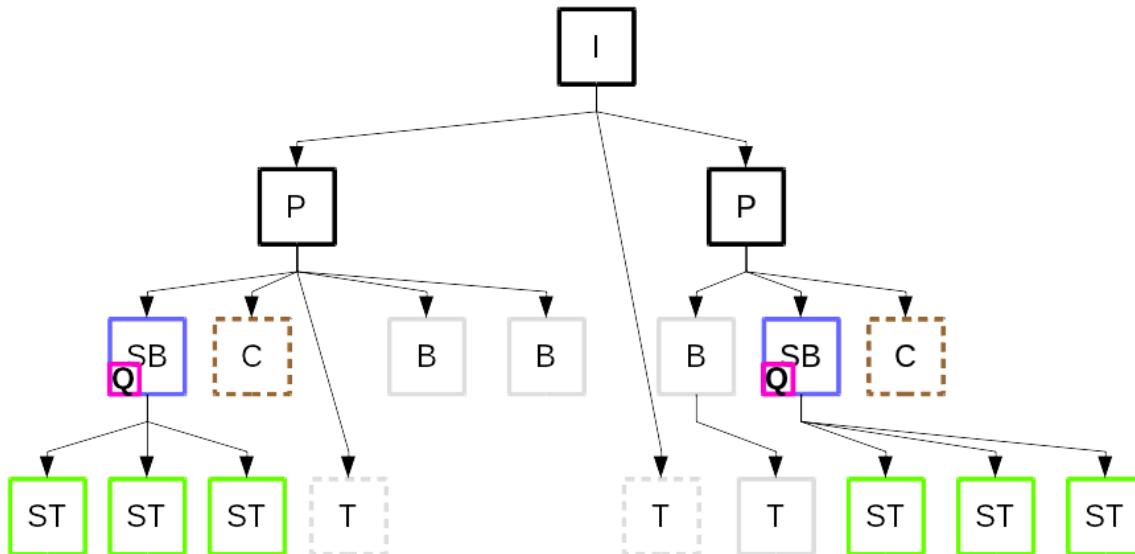
Knowledge-driven

Step 3: inference of new information

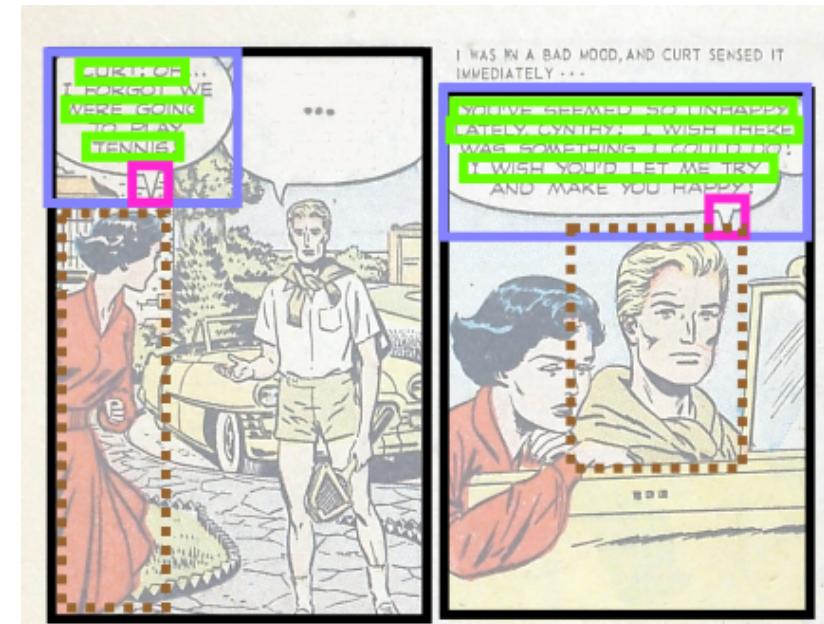


# Processing sequence 2/2

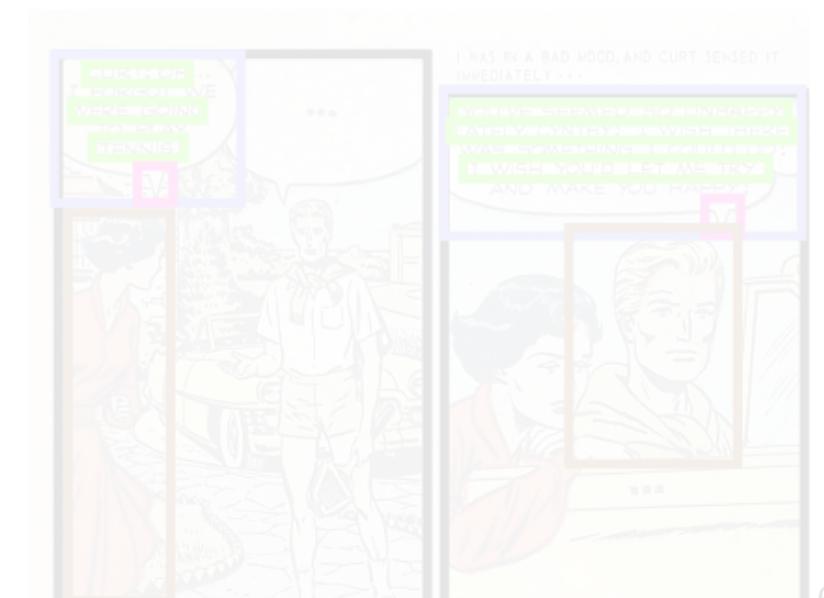
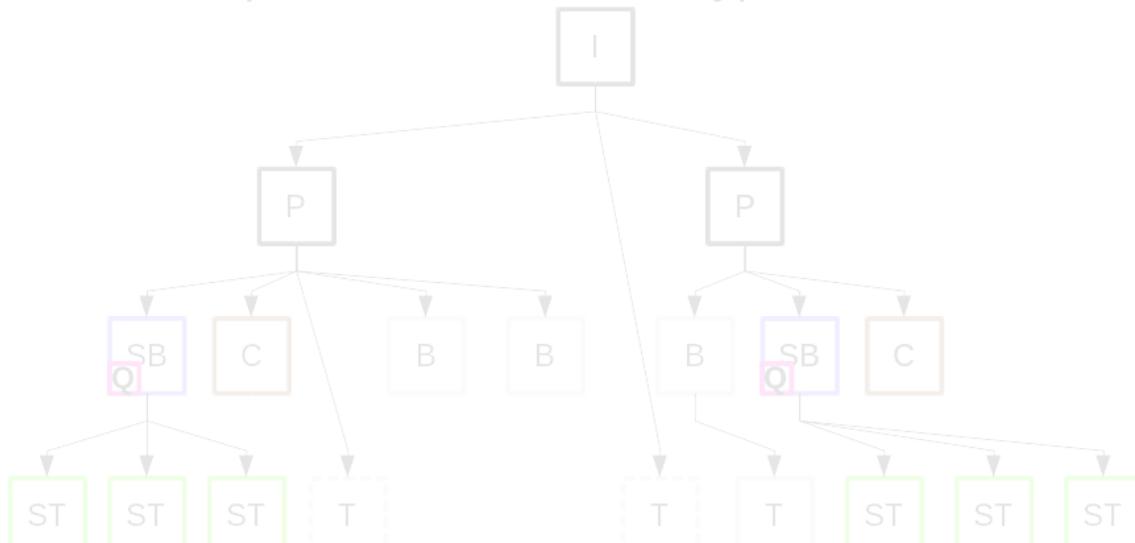
## Step 1: formulation of hypotheses



Knowledge-driven

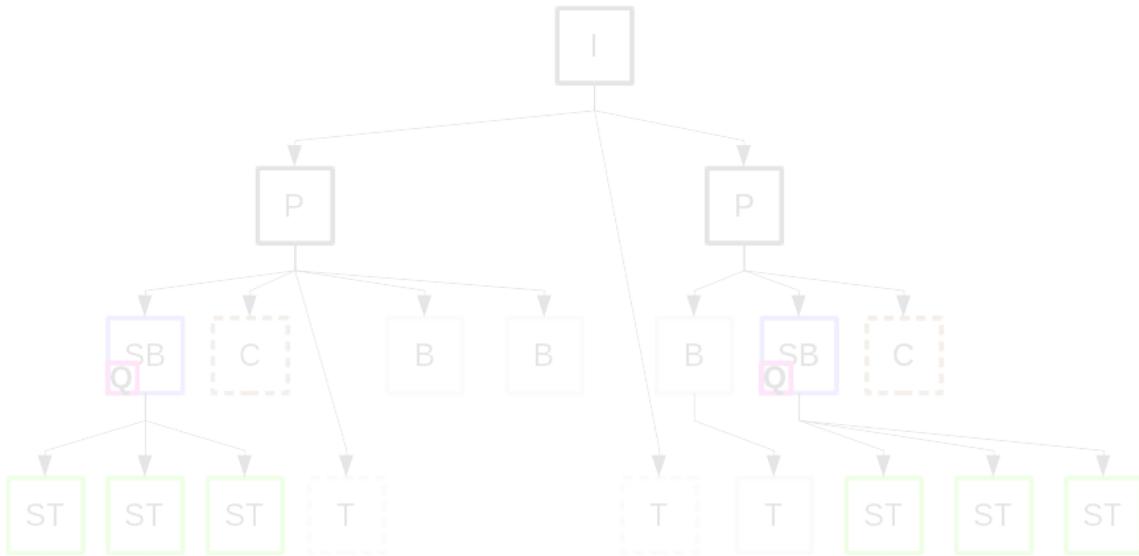


## Step 2: validation of the hypotheses

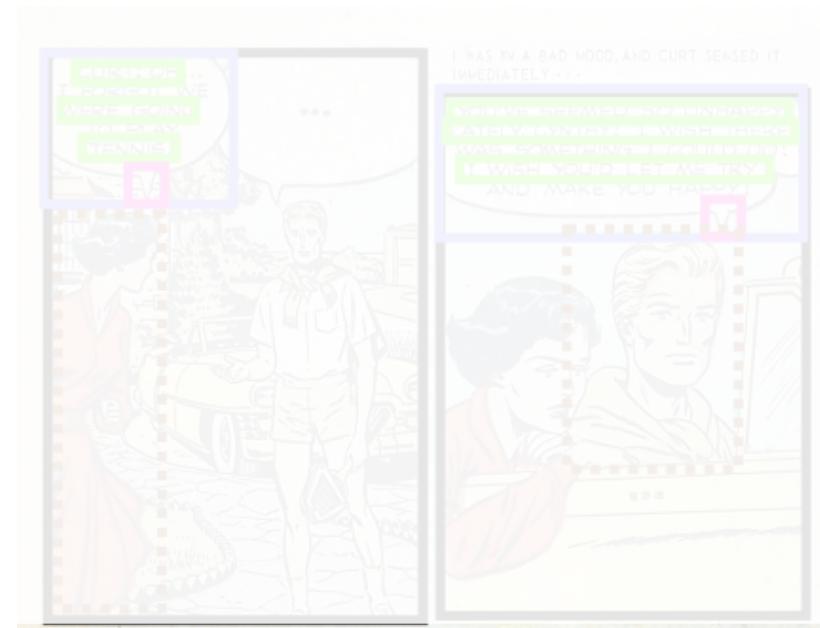


# Processing sequence 2/2

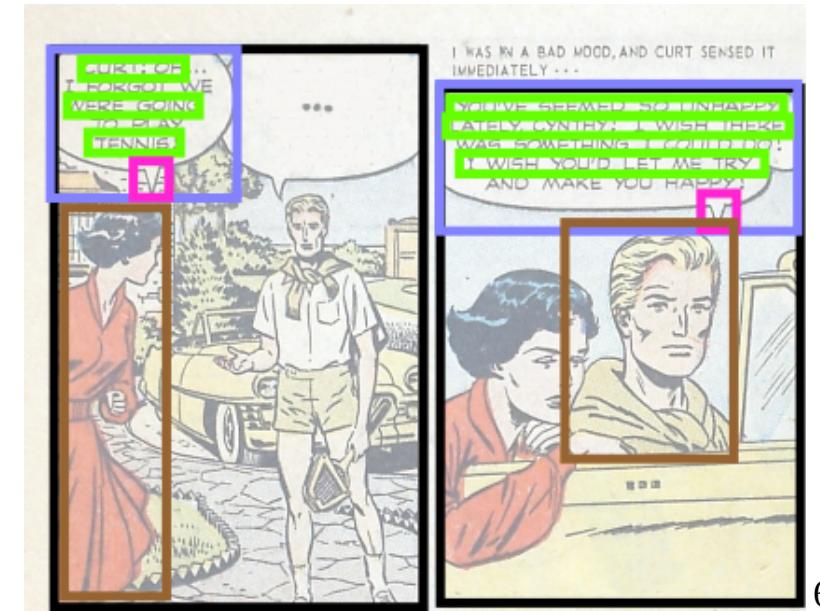
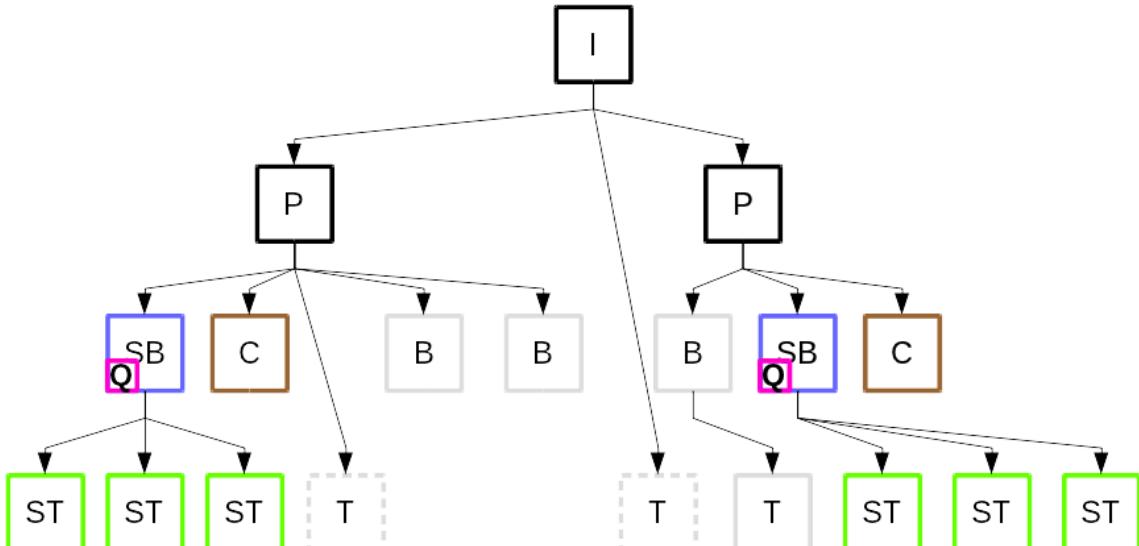
Step 1: formulation of hypotheses



Knowledge-driven



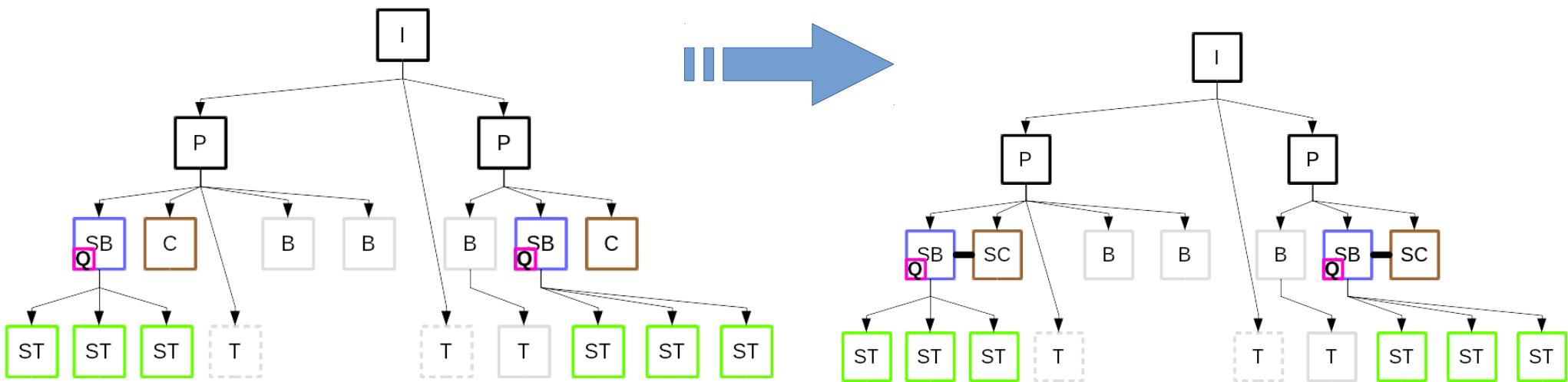
Step 2: validation of the hypotheses



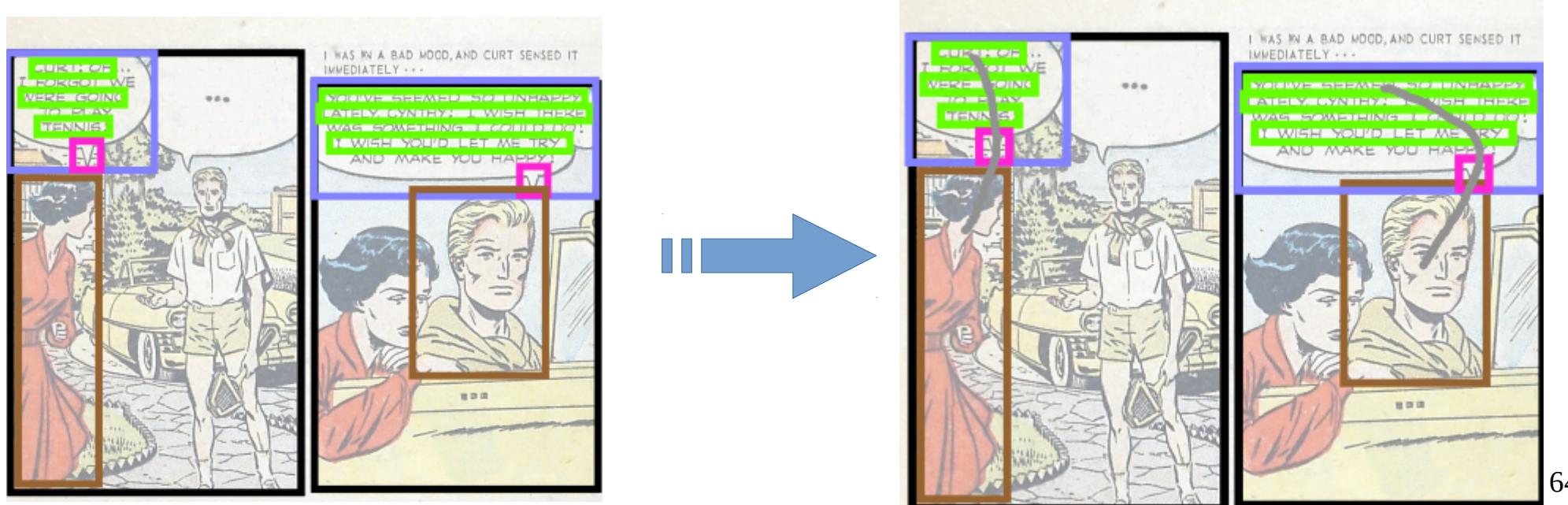
# Processing sequence 2/2

Knowledge-driven

Step 3: inference of new information



Final information about the image



- Contributions
  - Generic, unsupervised and expandable framework for comics understanding (pending acceptance IJDAR)
  - Formalization of image and comic book domains
- Limitations
  - Validation process (deletion only)

- Dataset and ground truth
- Evaluations
- Conclusions



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

# Dataset and ground truth

- Composition
  - 100 **mixed** pages
  - From 1905 to 2012
  - Franco-Belgium, American, Japanese
- **Visual** annotations
  - 850 panels
  - 1092 balloons
  - 1550 comic characters
  - 4691 text lines
- **Semantic** annotations
  - **Text type** (e.g. speech, thought, narrative)
  - **Character** ↔ **balloon** relationships
- Meta-data annotations
  - ISBN, author/publisher names, year
  - Page number, language, reading order



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



PANEL
<b>Rank:</b> 1
BALLOON
<b>Rank:</b> 2
<b>Shape:</b> Oval
<b>Tail direction:</b> South-West
TEXT LINE
<b>Text:</b> « STARK RAVING »
CHARACTER
<b>LinkedToBalloon:</b> 2

# Evaluations

# Experimentations

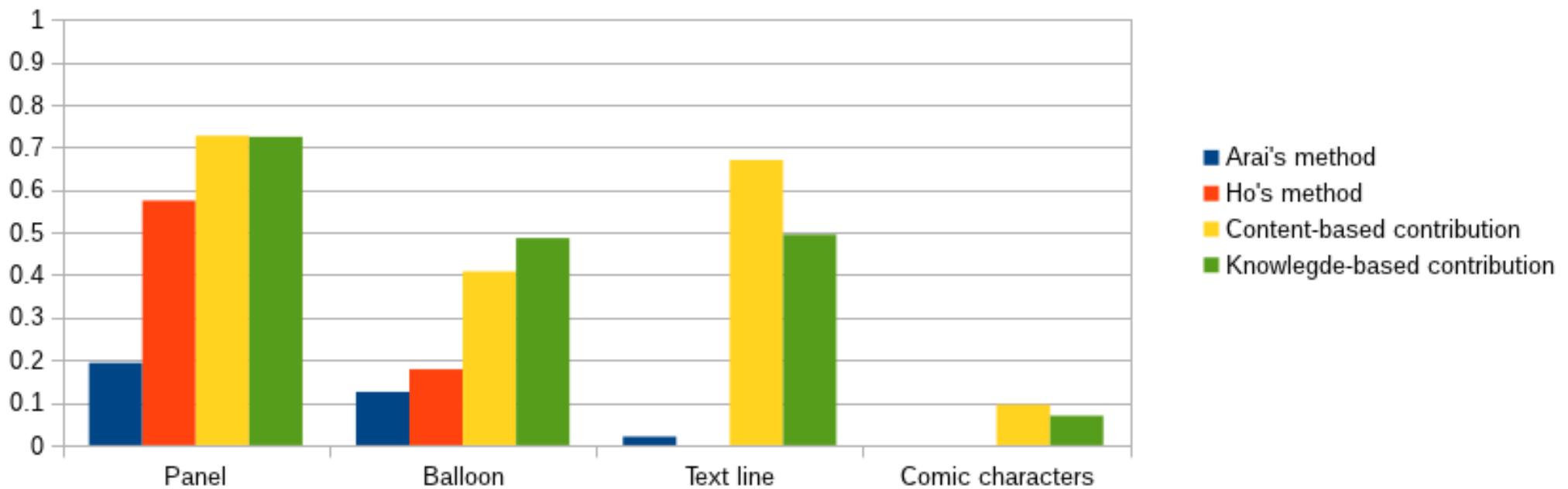
$$a_0 = \frac{\text{area}(B_p \cup B_{gt})}{\text{area}(B_p \cap B_{gt})}$$

$B_p$  = predicted region

$B_{gt}$  = ground truth region

$B_p$  valid if  $a_0 > 0.5$

Element localization results (F-score)



# Conclusion

## Experimentations

### Content-driven approach

- Pros
  - Efficient for “common” comics
  - Easily reusable for specific needs
  - Short processing time
- Cons
  - Rigid when used as a processing sequence
  - Error propagation

### Knowledge-driven approach

- Pros
  - Extensive, standard and exchangeable content description
  - Allows information recovery
  - No error propagation
  - Consistent information
- Cons
  - Validation process by deletion only
  - Processing time (reasoning)

- 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
  - In 36 months
- Contributions of the thesis
  - **Improvement** of existing extraction methods for panel, balloon and text extraction
  - Proposition of **first approaches** for , tail detection, balloon classification
  - Public **dataset** and ground truth for researchers
- Research impacts
  - **L3i** is now a **main actor** of comic book analysis in Europe
  - **Dataset** used by international peers
  - Just starting **national project** with a French company (Actialuna) and a lab (LIP6) in Paris
  - **International project** with a group working on Manga analysis at Osaka Prefecture University in Japan

- Retrieve implicit, overlapping and connected panels
- Deeply investigate text extraction and recognition
- Improve implicit and external border balloon extraction and evaluation
- Non-speaking comic character extraction and identification
- Increase the number of pages of the eBDtheque dataset
- Continue to enrich the ground truth (e.g. panel view angle and situation, text tone, multi-part character segmentation, character names and roles)

# Publications

# Conclusion

Thank you + github + c-r

Conclusion