



European Ph.D. defense

Communauté
d'Agglomération de
La Rochelle

Segmentation and indexation of complex objects in comic book images

Christophe Rigaud
December 11th, 2014

Co-supervised by:

Jean-Christophe Burie¹
Dimosthenis Karatzas²
Jean-Marc Ogier¹

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”

Hiroaki Tobita, Sony CSL Interaction Laboratory, 2014

Comic books

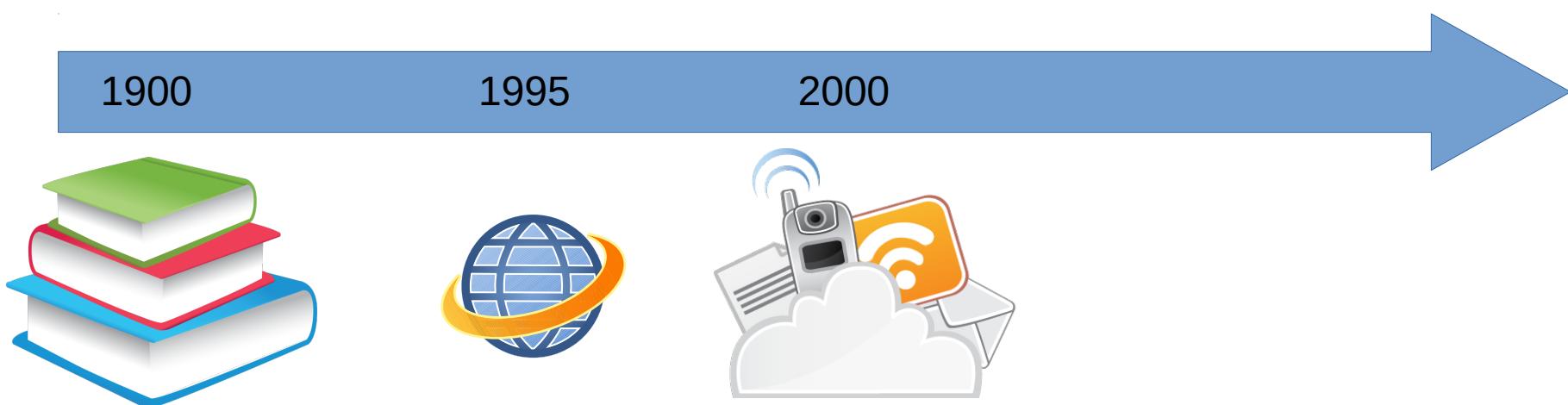
Introduction

“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”

Hiroaki Tobita, Sony CSL Interaction Laboratory, 2014



Comic books

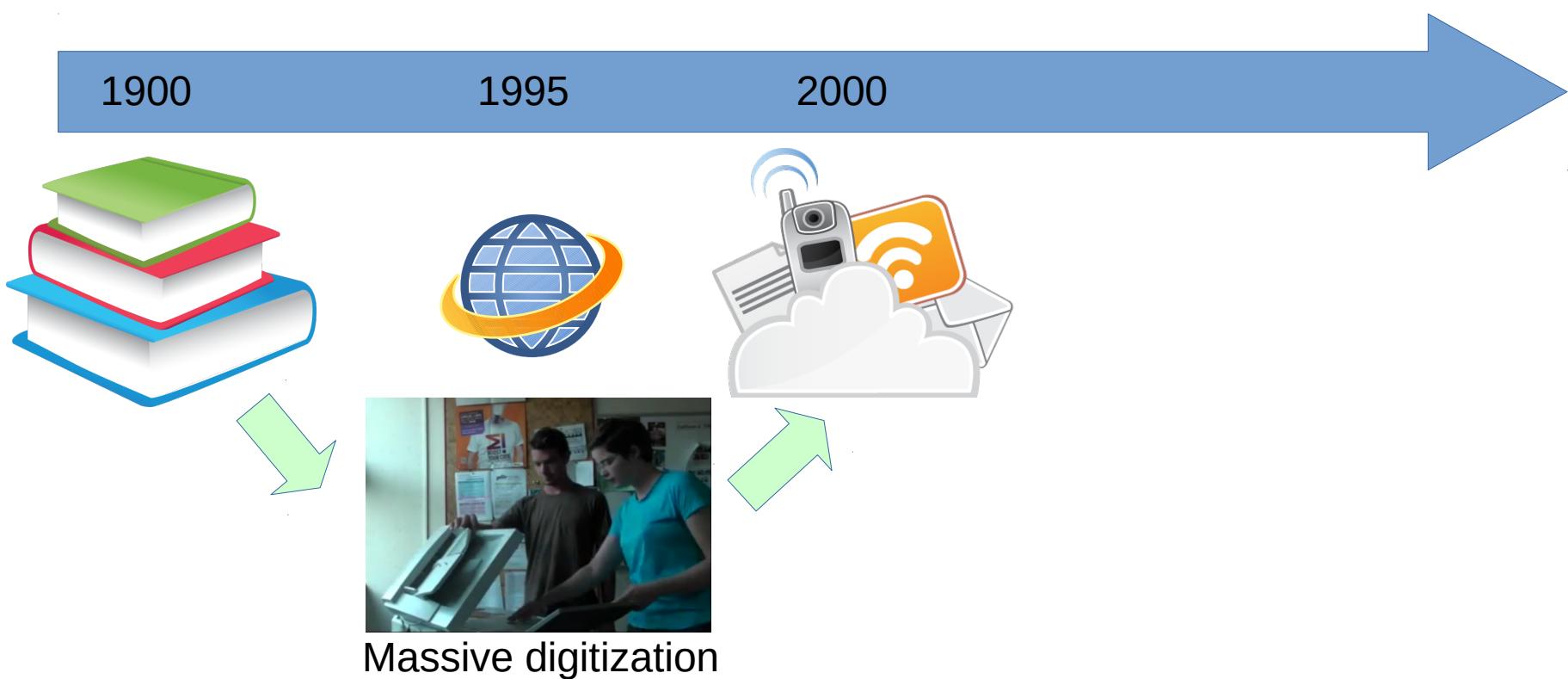
Introduction

“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”

Hiroaki Tobita, Sony CSL Interaction Laboratory, 2014



Comic books

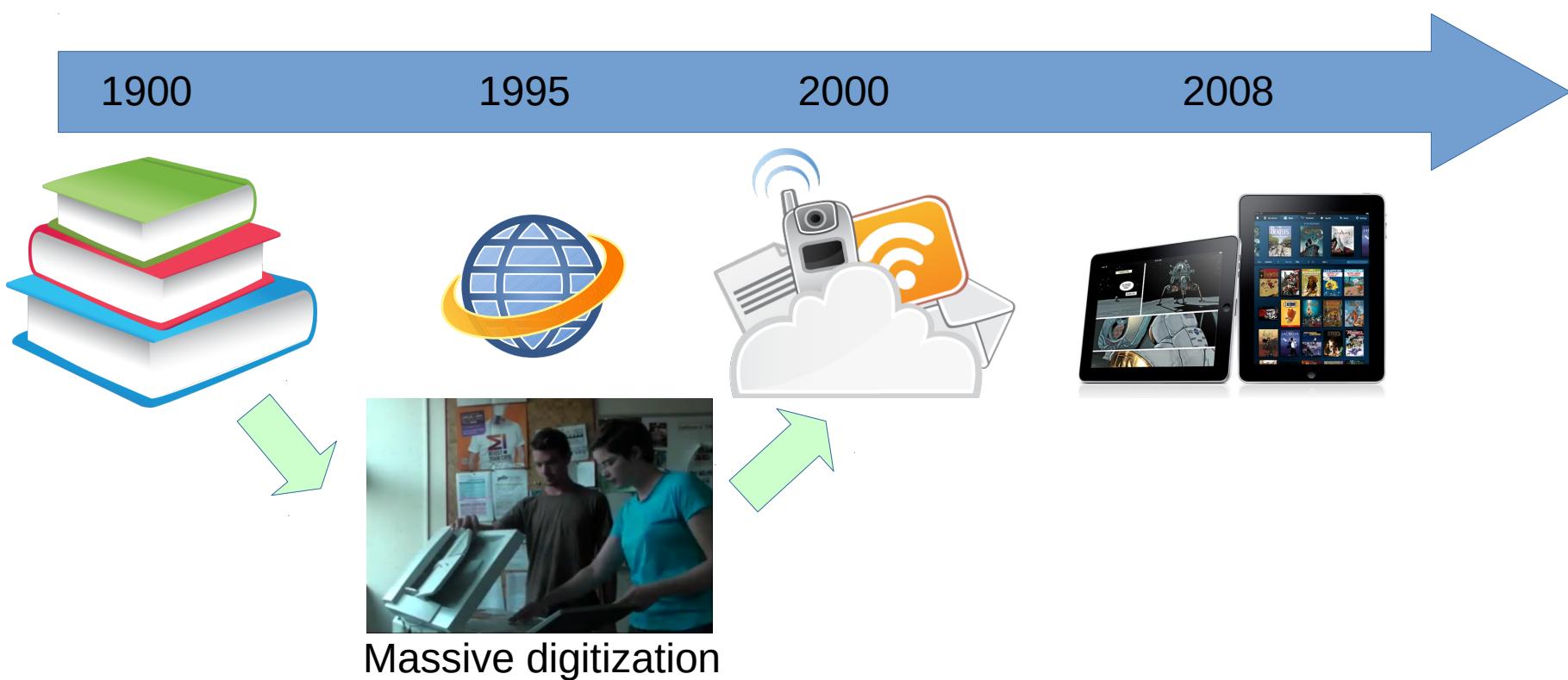
Introduction

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

Wikipédia, 2014

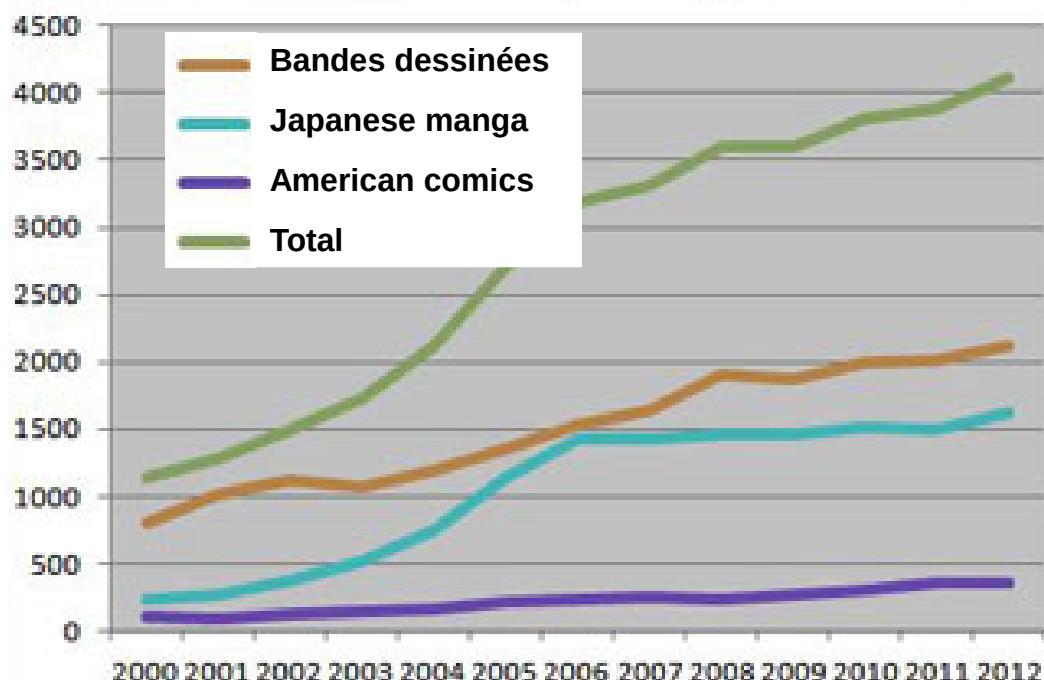
“One of the most popular and familiar forms of graphic content”

Hiroaki Tobita, Sony CSL Interaction Laboratory, 2014



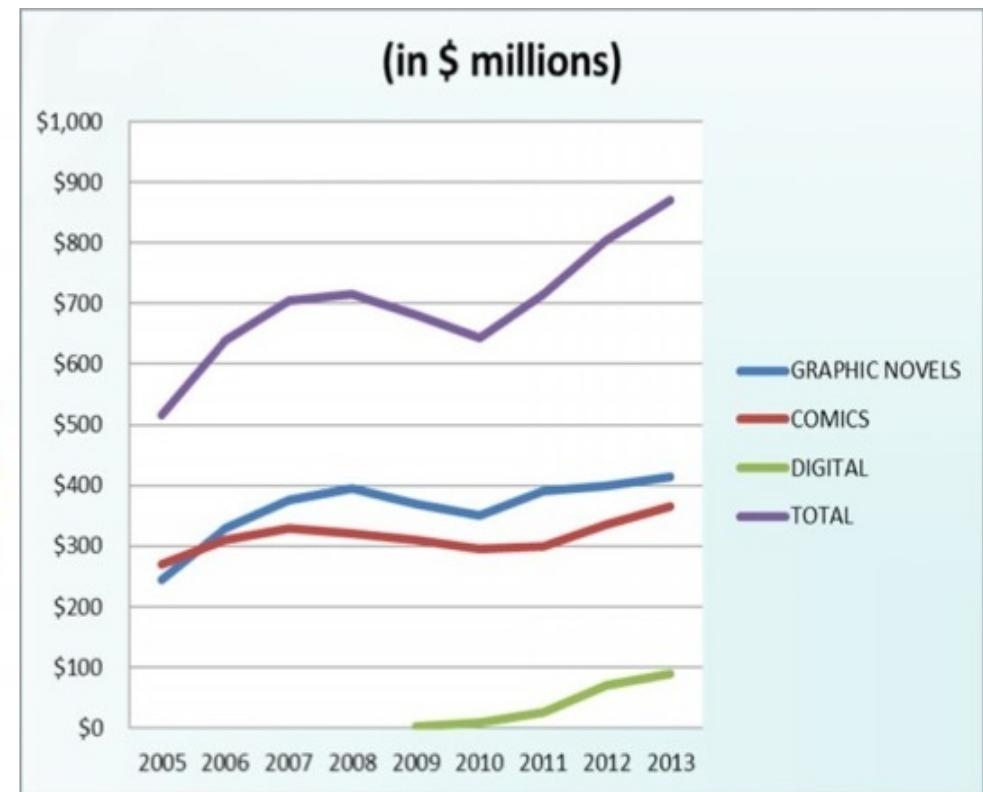
Comic books

Introduction



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

- eBDthèque project (since 2011)
 - Add value to digitized comics using the new technologies
 - Public founding CPER 2007-2013
 - 2 Ph.D. students, 1 engineer, 1 post doc, 6 professors
- Research axes
 - Content extraction (Christophe Rigaud)
 - Knowledge representation (Clément Guérin)
- Scientific challenges
 - TODO
 - TODO
 - TODO (make challenges appear only on a second slide)

- Comic book images
- Content extraction
 - Panels
 - Balloons
 - Text
 - Comic characters



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

Comic book images

Background

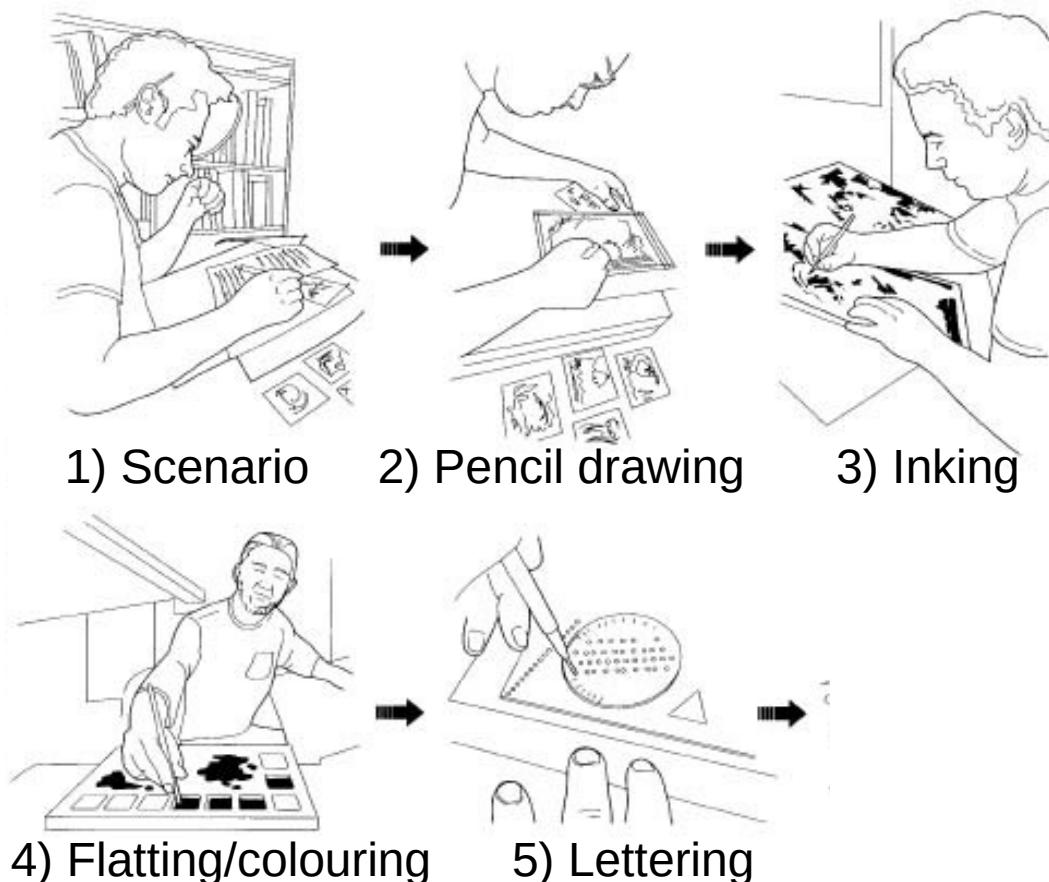
- Challenges:
 - Recent field of research with a **largely unknown**
 - **Semi-structured** and **free-form** document mixing text and graphics
- Design process:



Comic book images

Background

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



Panel extraction

Background

- Challenges

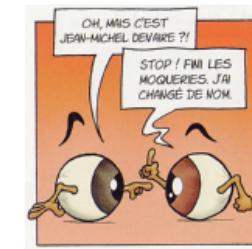
- Diversity of styles (gutter, implicit)
- Semi-structured layout

- Panel extraction

- White line cut [Chung07]
- Recursive X-Y cut [Eunjung07]
- Gradient [Tanaka07]
- Connected-components [Arai10, Pang14]
- Polygon detection [Li14a]
- Corners and line segments [Stommel12]

- Conclusions

- Problem solved for common manga and European comics if treated separately
- Remaining difficulties are for connected, nested and implicit panels
- No approach tested over all comics styles, no dataset, no reproductive results



Panel extraction

Background

- Challenges

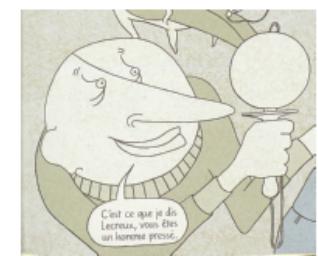
- Diversity of styles (gutter, implicit)
- Semi-structured layout

- Panel extraction

- White line cut [Chung07]
- Recursive X-Y cut [Eunjung07]
- Gradient [Tanaka07]
- Connected-components [Arai10, Pang14]
- Polygon detection [Li14a]
- Corners and line segments [Stommel12]

- Conclusions

- Problem solved for manga and other comics if treated separately
- Remaining difficulties are for connected, nested and implicit panels
- No public dataset, no reproductive results



Balloon extraction

- Challenges
 - Shape and contour variations
 - Implicit balloons

- Extraction
 - Connected-components [Arai 2011, Ho 2012]
- Conclusions
 - Closed balloon solved (with text)
 - Implicit balloons, classification and tail to explore

Image	Shape	Contour
	Oval	Smooth
	Rectangle	Smooth
	Oval	Wavy
	Oval	Spiky
	Oval / implicit	Smooth / Implicit

Balloon extraction

- Challenges
 - Shape and contour variations
 - Implicit balloons
- Extraction
 - Connected-components [Arai 2011, Ho 2012]
- Conclusions
 - Closed balloon solved (from text)
 - Implicit balloons, classification and tail are to explore

Image	Shape	Contour
	Oval	Smooth
	Rectangle	Smooth
	Oval	Wavy
	Oval	Spiky
	Oval / implicit	Smooth / Implicit

Text extraction and recognition

Background

- Challenges
 - Non-standard fonts
 - Multi-script/orientation/scale
 - Complex background (sound effects)
 - Hyphenation, voluntary spelling mistakes
 - Extraction
 - Connected-components [Ho12,??]
 - SVM and Bayesian classifier [Li14b]
 - Recognition
 - OCR trained for a specific comic font [Ponsard12]



Text extraction and recognition

Background

- Challenges
 - Non-standard fonts
 - Multi-script/orientation/scale
 - Complex background (sound effects)
 - Hyphenation, voluntary spelling mistakes
 - Extraction
 - Connected-components [Ho12, ???]
 - SVM and Bayesian classifier [Li14b]
 - Recognition
 - OCR trained for a specific comic font [Ponsard12]



- Conclusions
 - Speech text (from speech balloon)
 - Captions and sound effects unexplored
 - Text recognition very poor

Comic character extraction

Background

- Challenges

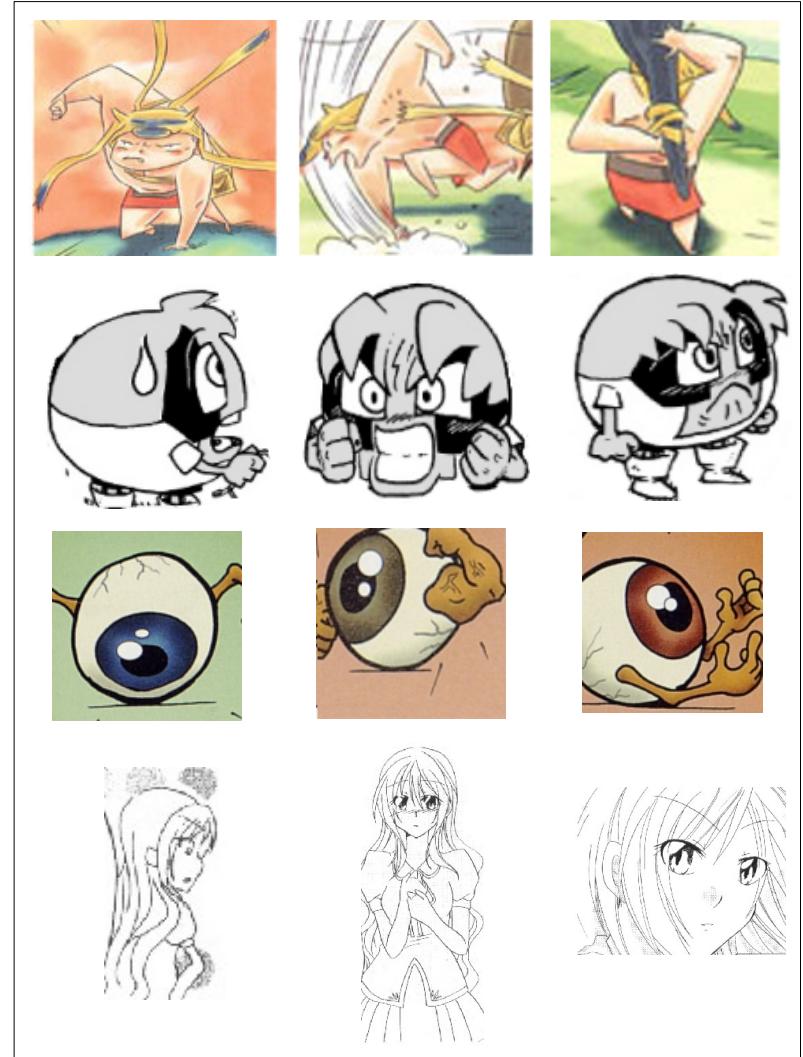
- Hand-drawn, stroke-based
- Intra/inter class variability
- Scale, deformation, posture, occlusion

- Extraction & recognition

- Manga faces [Cheung2008, Sun2010, Kohei2012]
- Cartoons [Khan2012]

- Conclusions

- Preliminary results
- Complex and versatile structure
- Contains most of the interesting information



Comic character extraction

Background

- Challenges

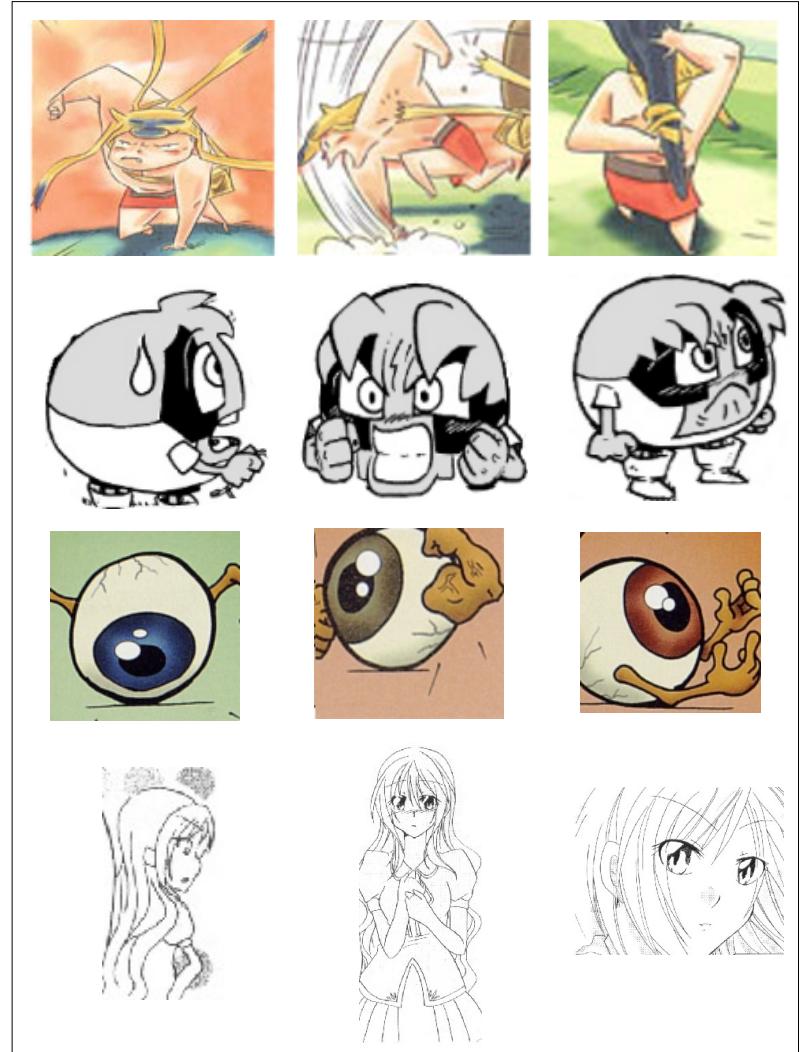
- Hand-drawn, stroke-based
- Intra/inter class variability
- Scale, deformation, posture, occlusion

- Extraction & recognition

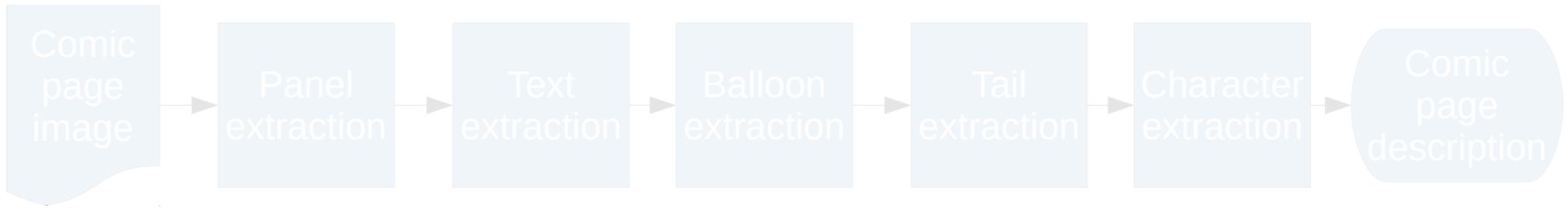
- Manga faces [Cheung2008, Sun2010, Kohei2012]
- Cartoons [Khan2012]

- Conclusions

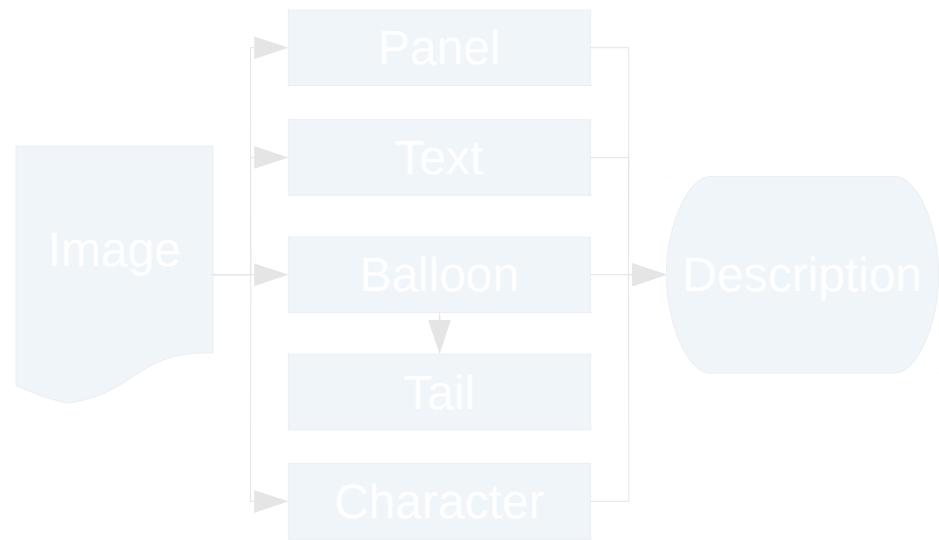
- Preliminary results
- Complex and versatile structure
- Contains most of the interesting information



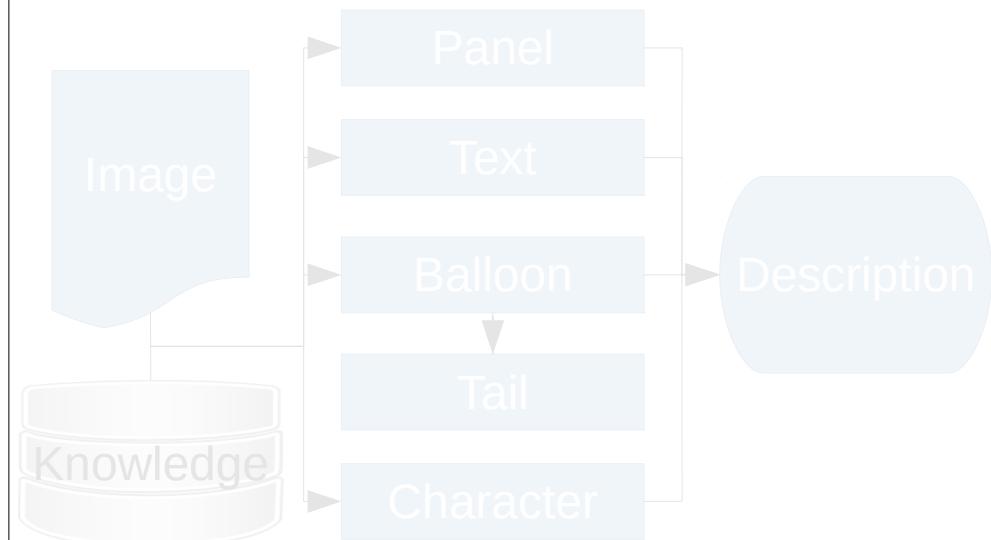
Content-driven (sequential)



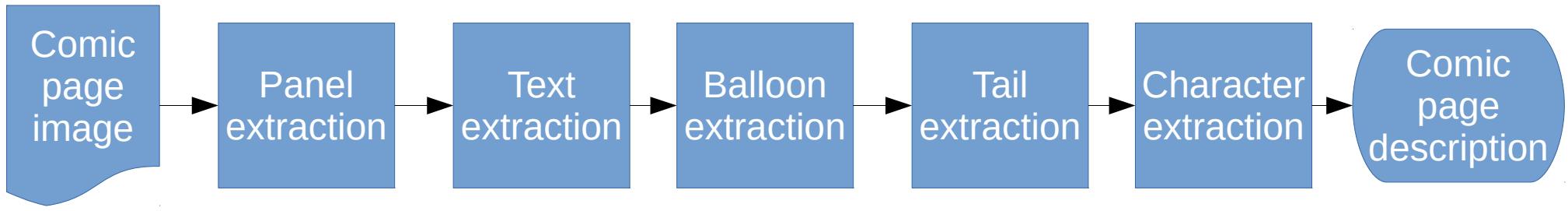
Content-driven (independent)



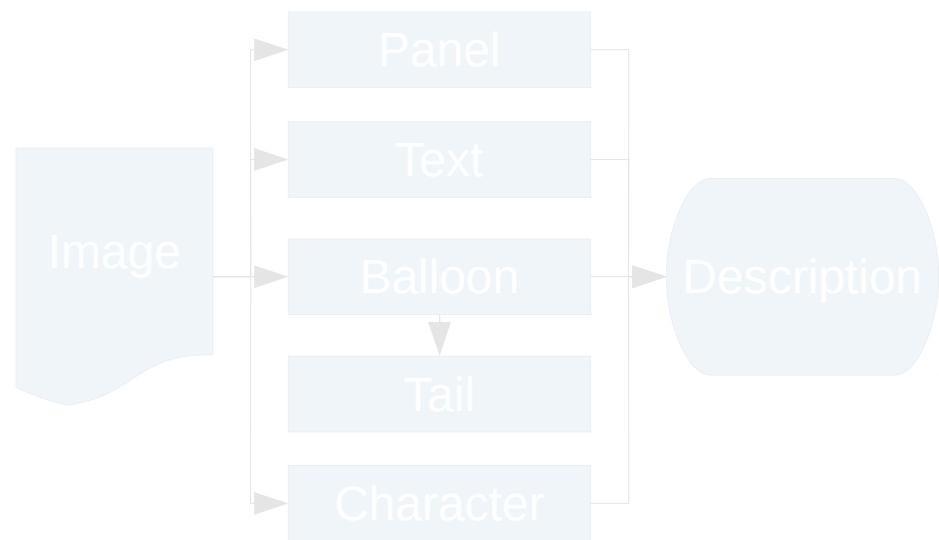
Knowledge-driven (independent)



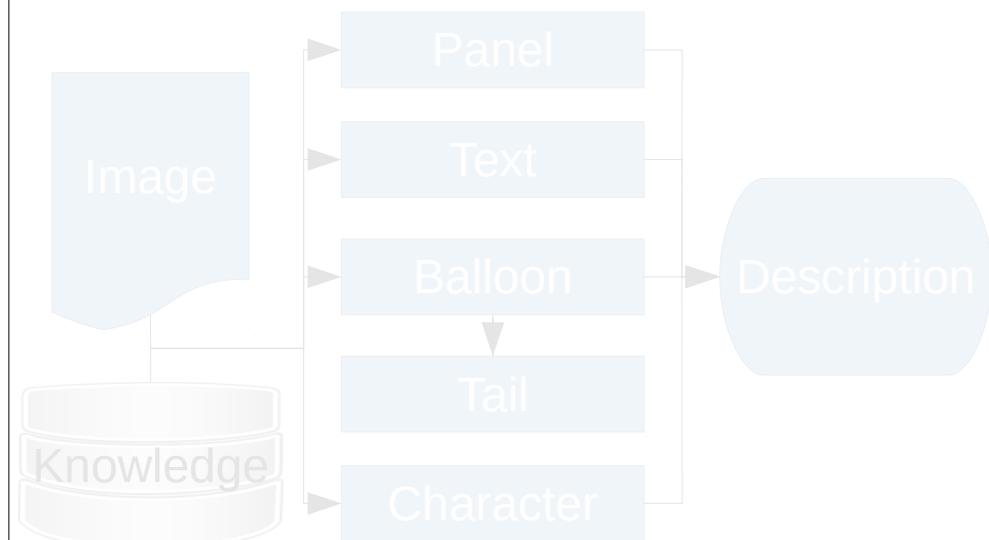
Content-driven (sequential)



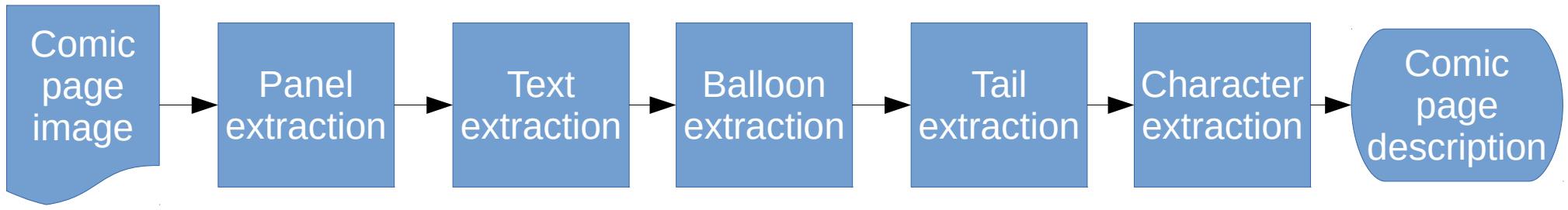
Content-driven (independent)



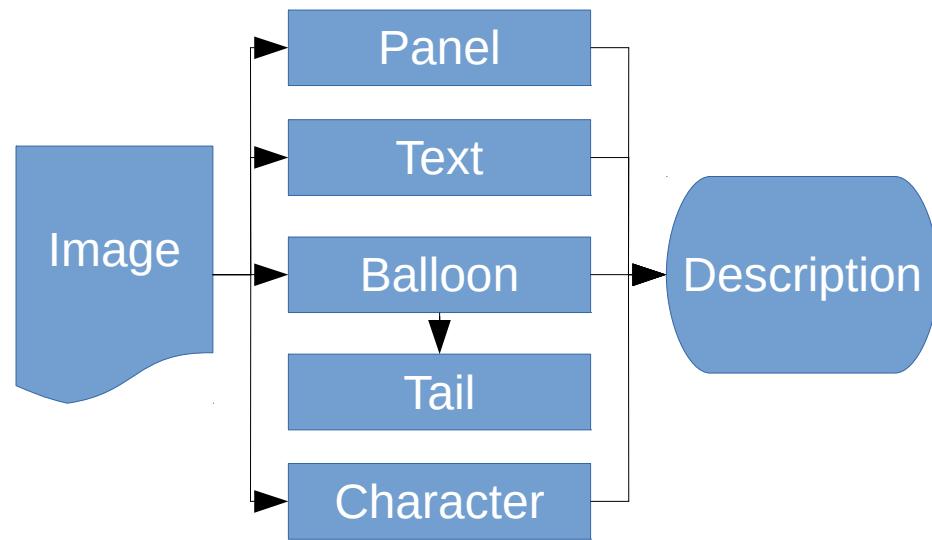
Knowledge-driven (independent)



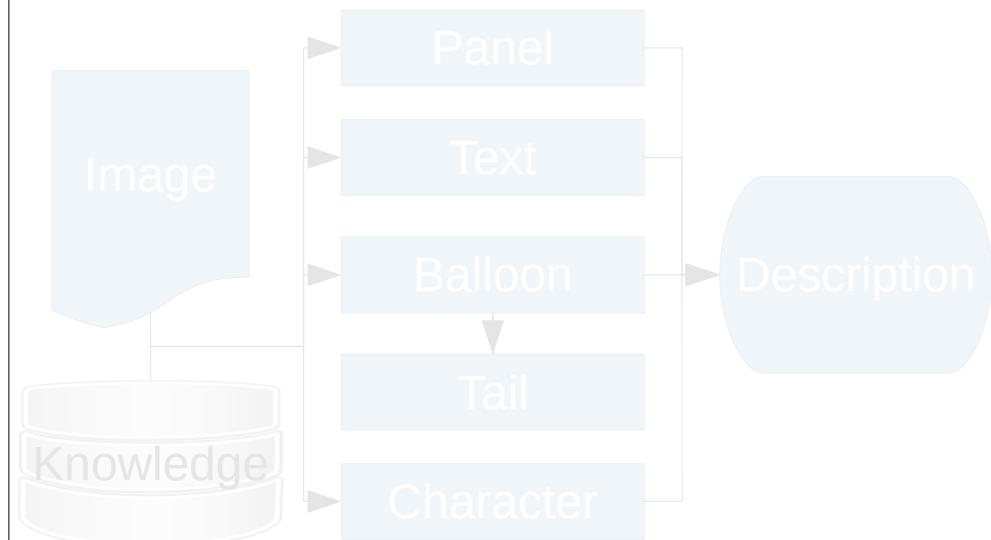
Content-driven (sequential)



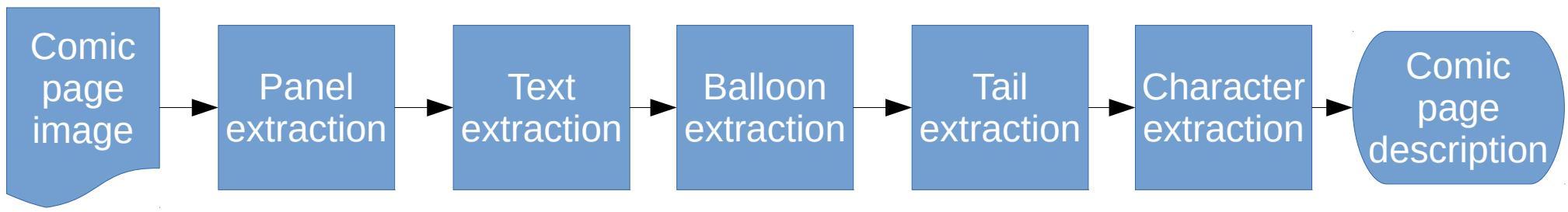
Content-driven (independent)



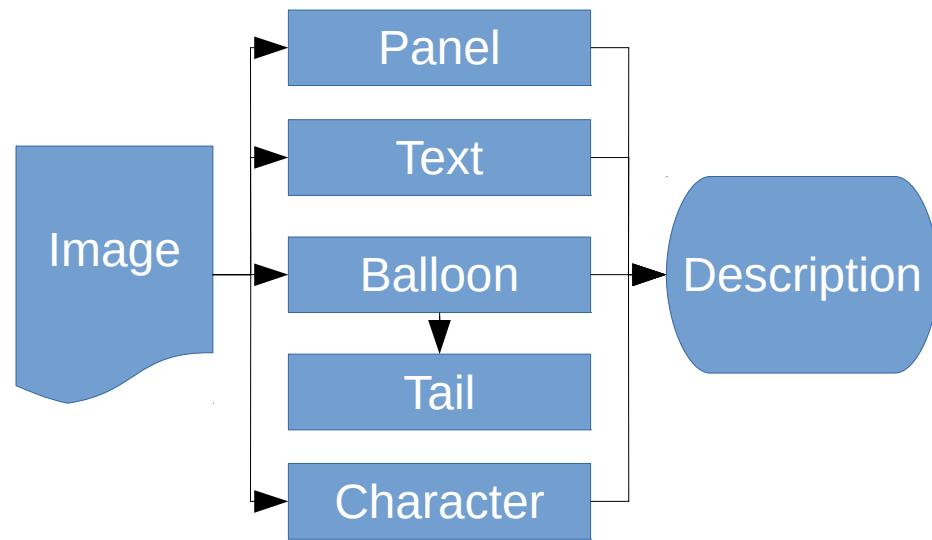
Knowledge-driven (independent)



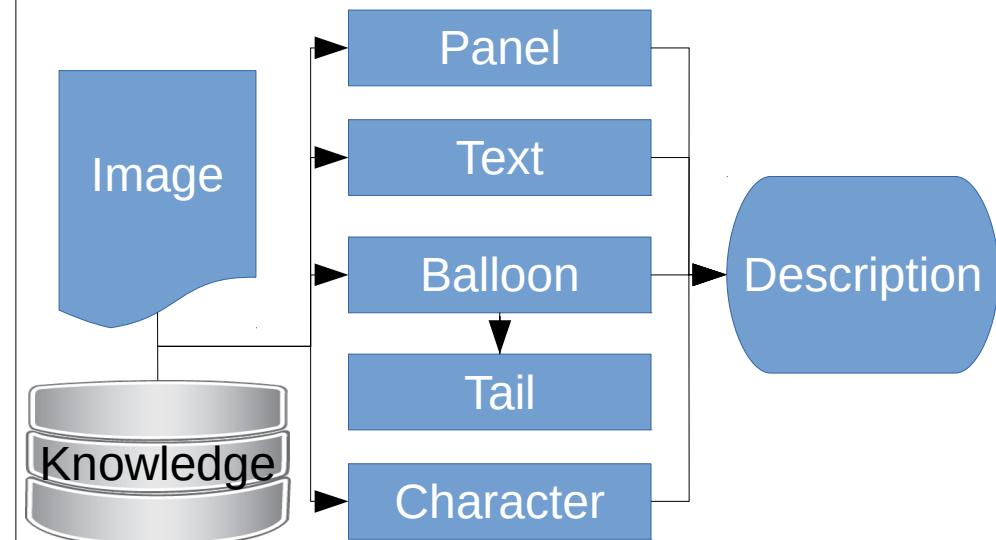
Content-driven (sequential)



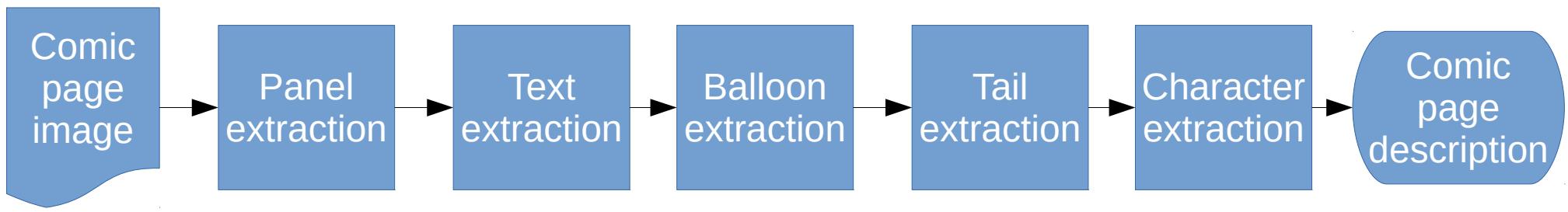
Content-driven (independent)



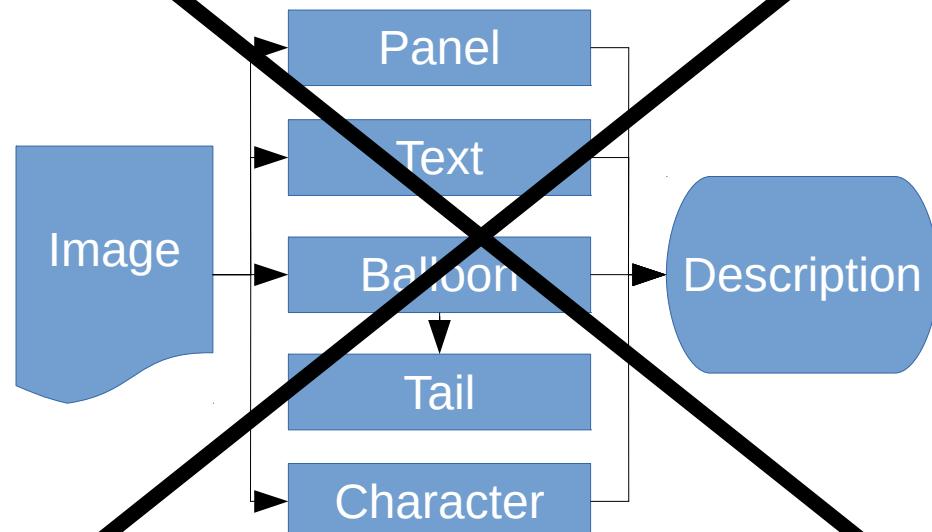
Knowledge-driven (independent)



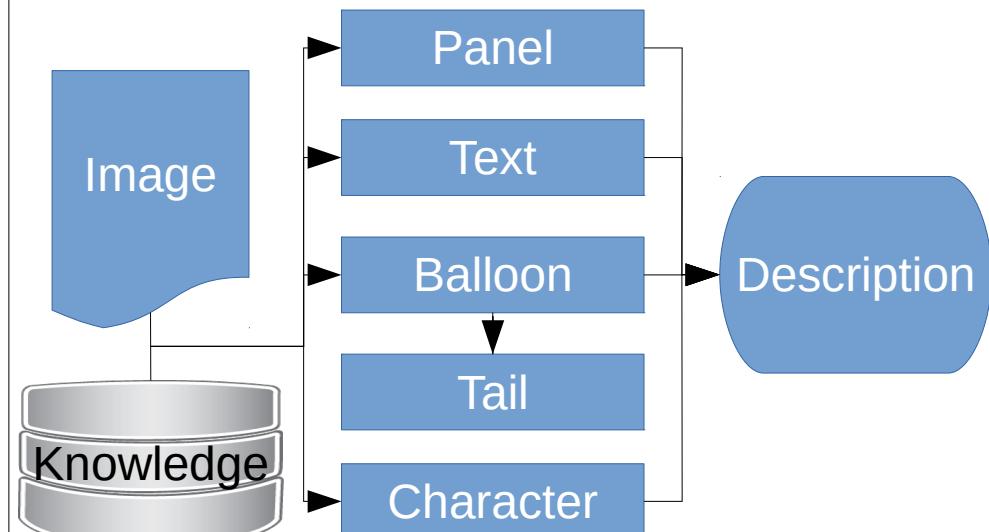
Content-driven (sequential)



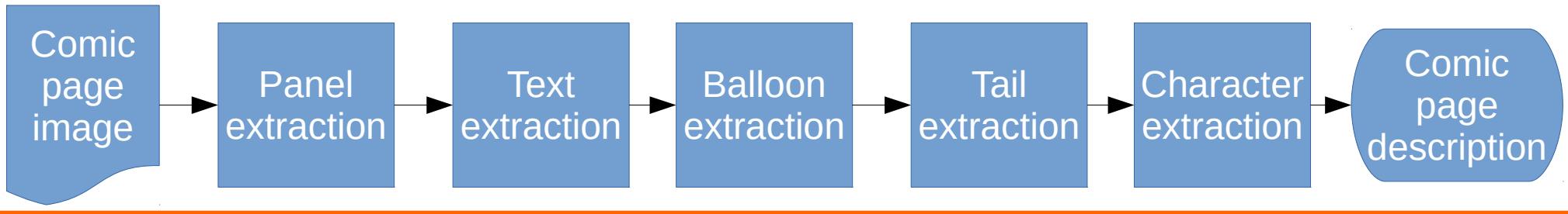
Content-driven (independent)



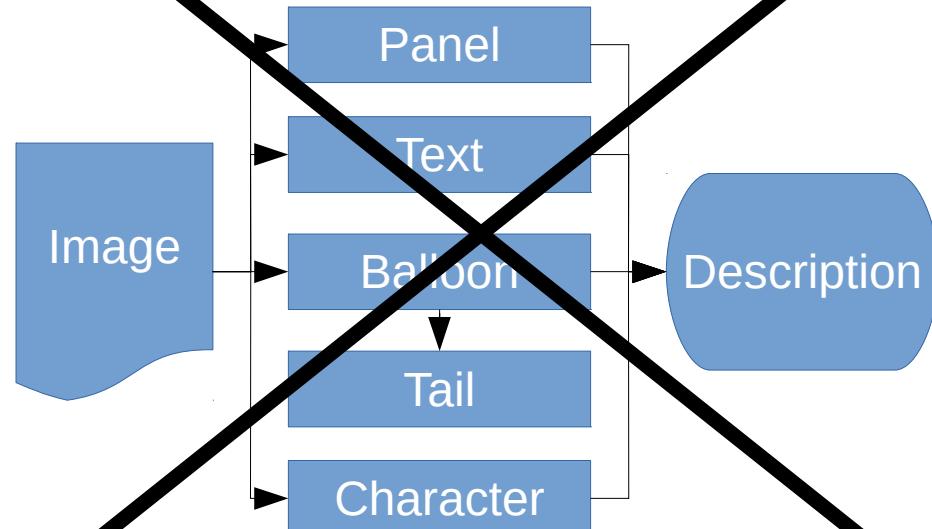
Knowledge-driven (independent)



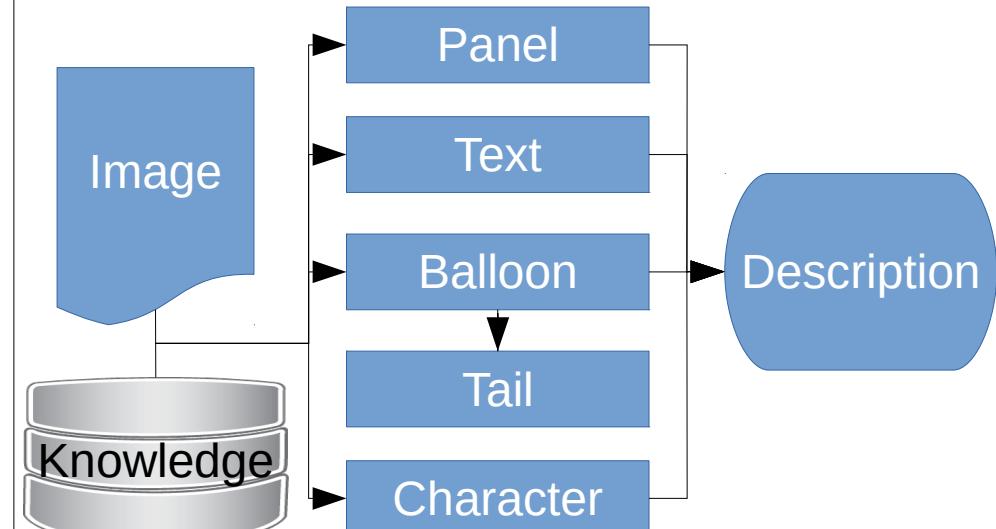
Content-driven (sequential)



Content-driven (independent)



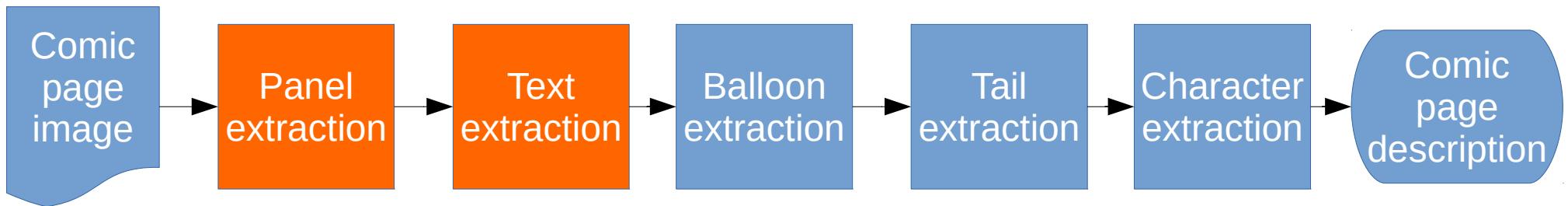
Knowledge-driven (independent)



Panel and text extraction

Contributions
Content-driven approach

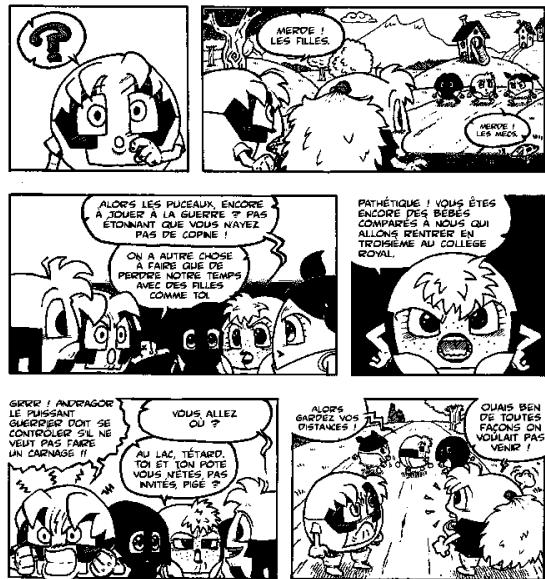
- Processing sequence



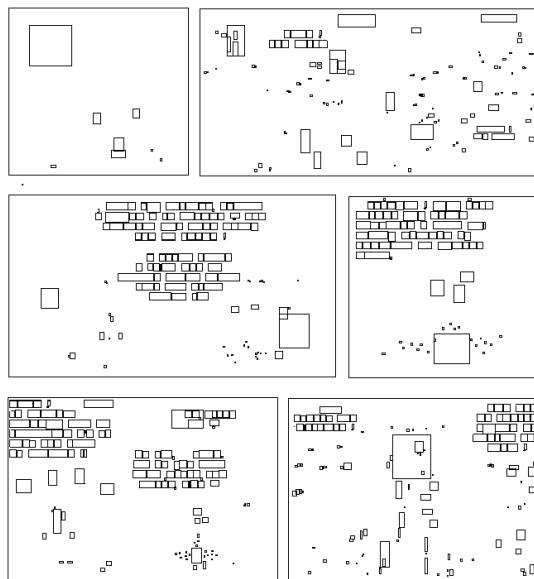
- Literature
 - Sequential or separated panel and text extraction
 - With or without gutters
- Contribution
 - Simultaneous panel and text extraction
 - Implicit and with gutters

Panel and text extraction

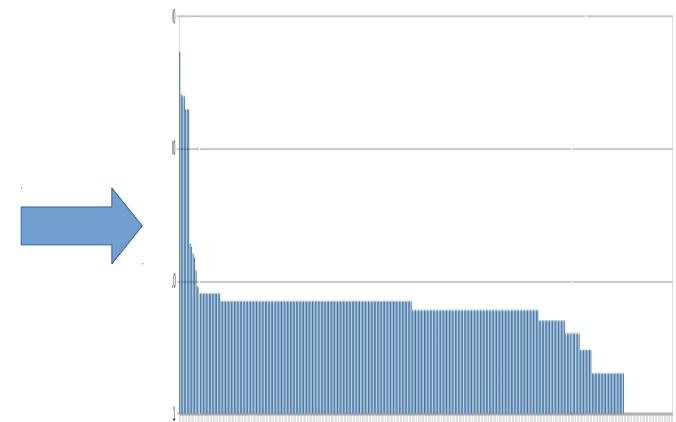
Contributions
Content-driven approach



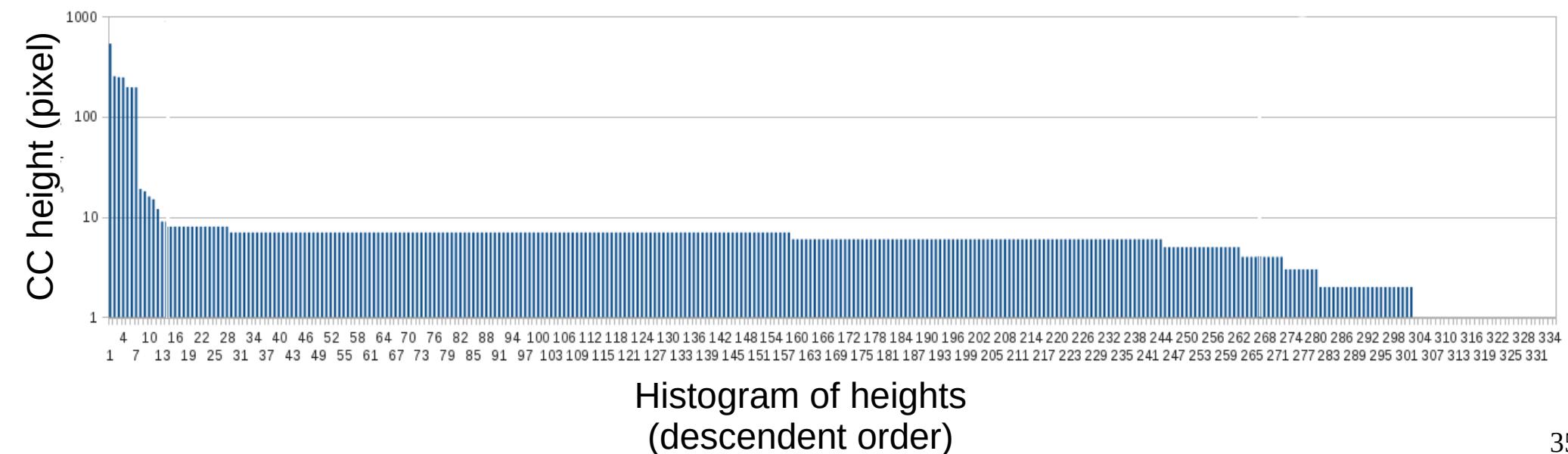
Binary image



Connected-component (CC) bounding boxes



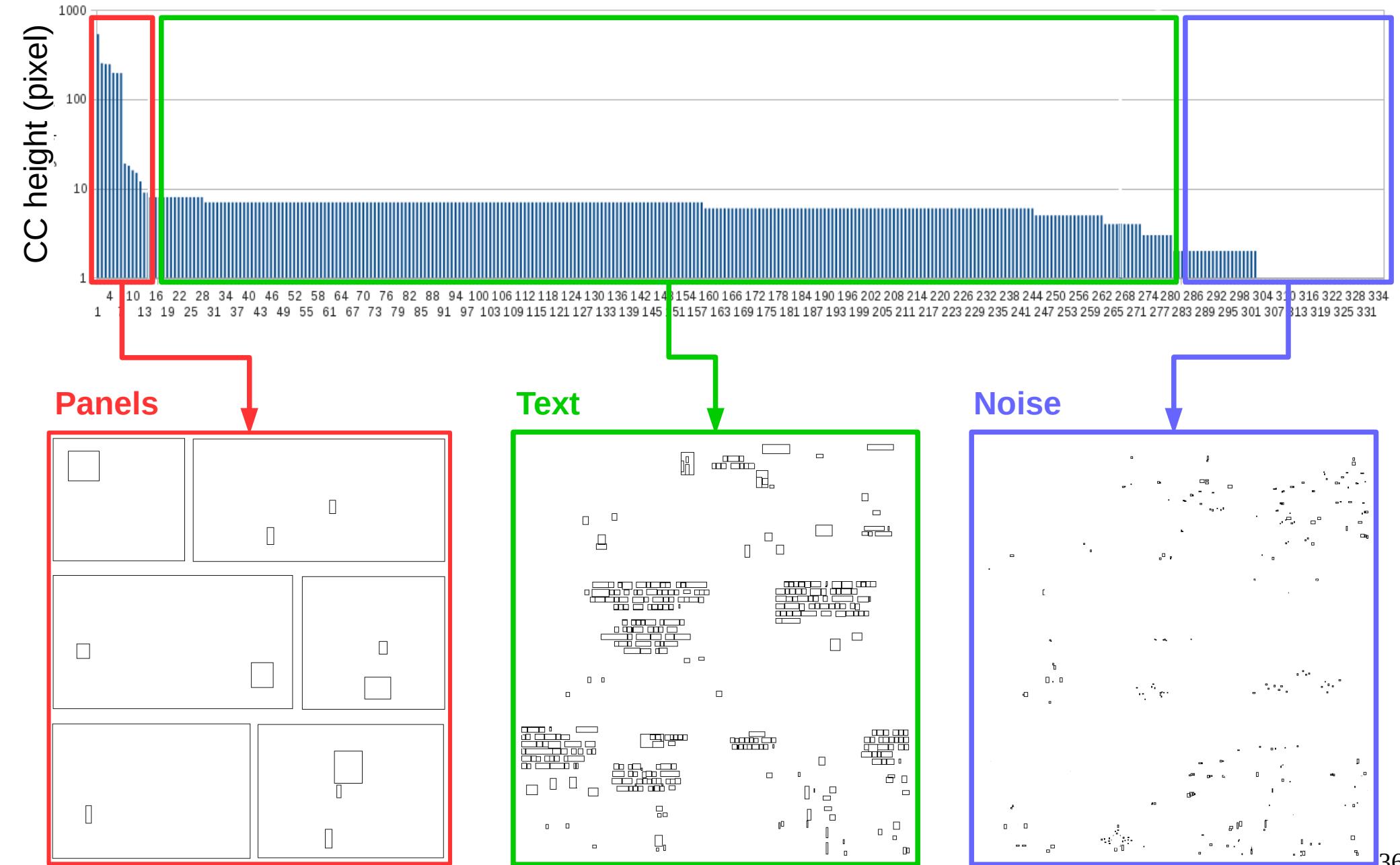
Histogram of heights of CC



Panel and text extraction

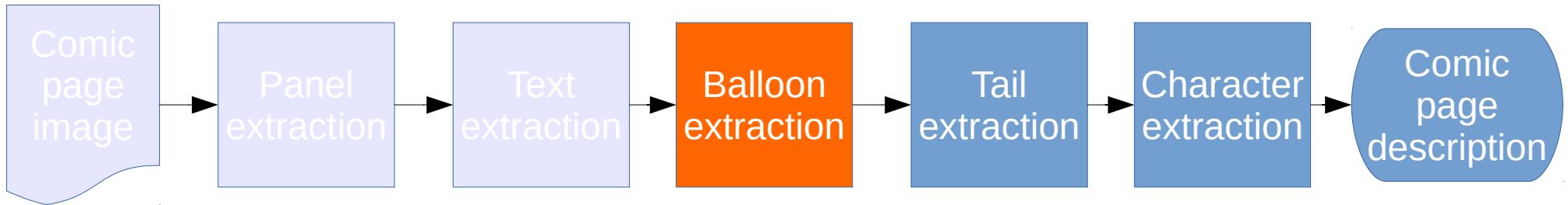
Contributions
Content-driven approach

K-means clustering ($k=3$)



Balloon extraction

Contributions
Content-driven approach



Regular balloon

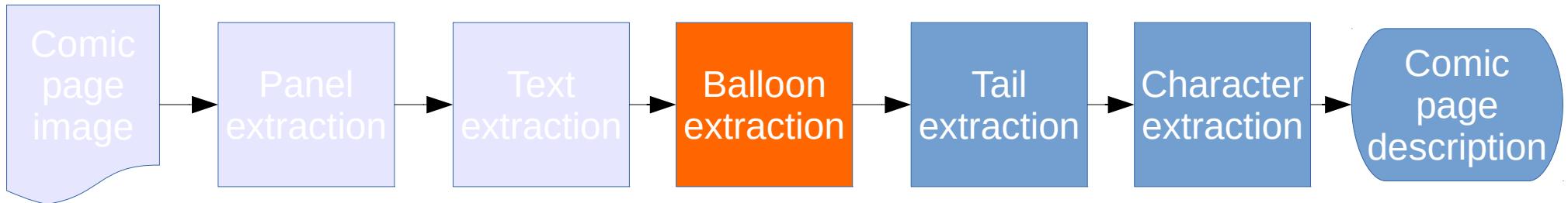


Implicit balloon

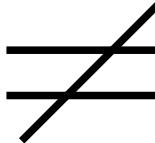
- Literature
 - Top-down approaches: extract white blobs and then text inside
 - Limited to regular balloons
- Contribution
 - Bottom-up approaches: extract text and then surrounding balloons
 - Appropriate for regular and implicit balloons

Balloon extraction

Contributions
Content-driven approach



Regular balloon



Implicit balloon

- Literature
 - Top-down approaches: extract white blobs and then text inside
 - 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

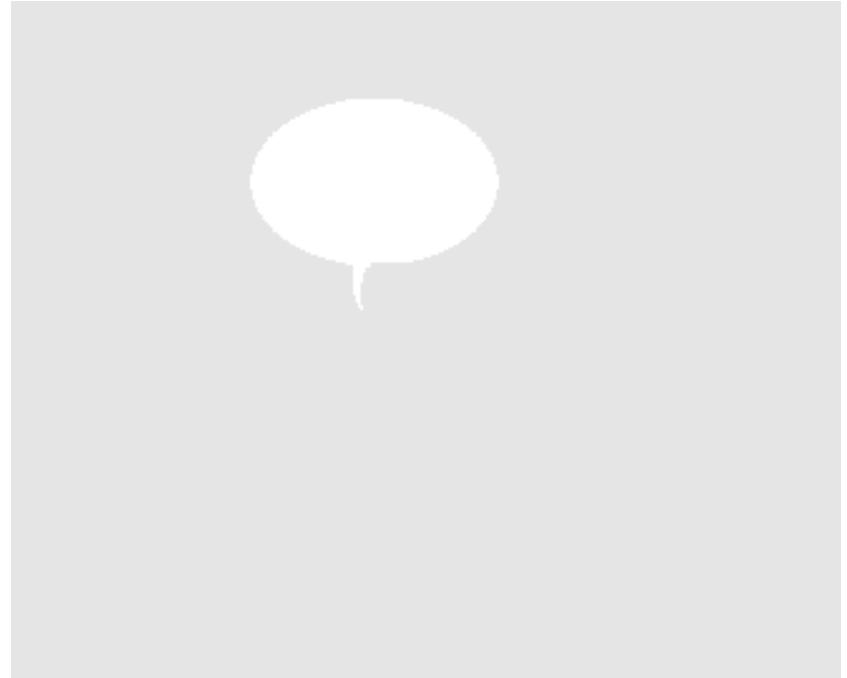
Balloon extraction: regular

Contributions
Content-driven approach

- Assumptions
 - Panels and text block positions are known
 - Balloons contain centred text
- Proposition → structural analysis
 - Extract closed contours that includes centred text



Original image



Expected result

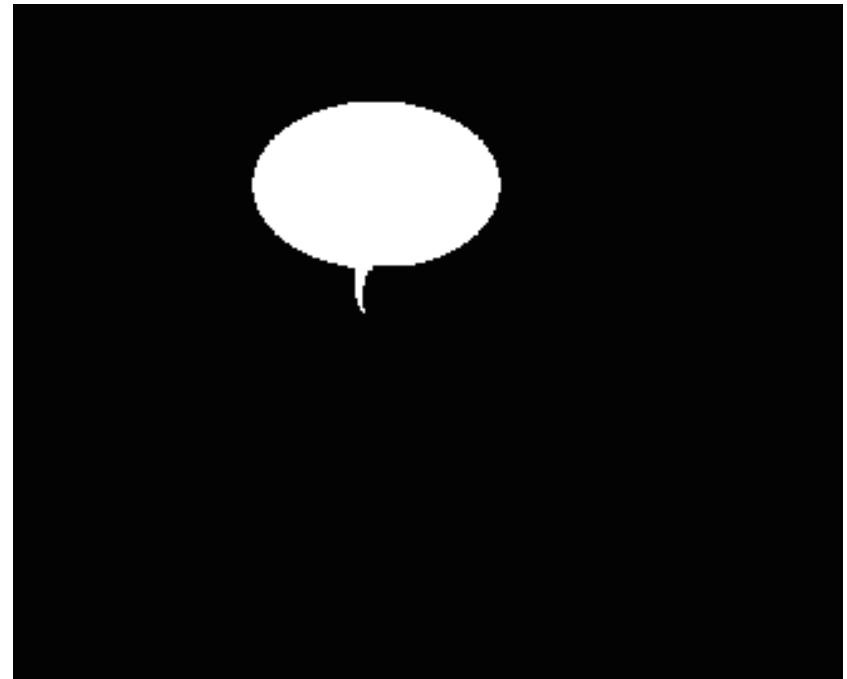
Balloon extraction: regular

Contributions
Content-driven approach

- Assumptions
 - Panels and text block positions are known
 - Balloons contain centred text
- Proposition → structural analysis
 - Extract closed contours that includes centred text



Original image



Expected result

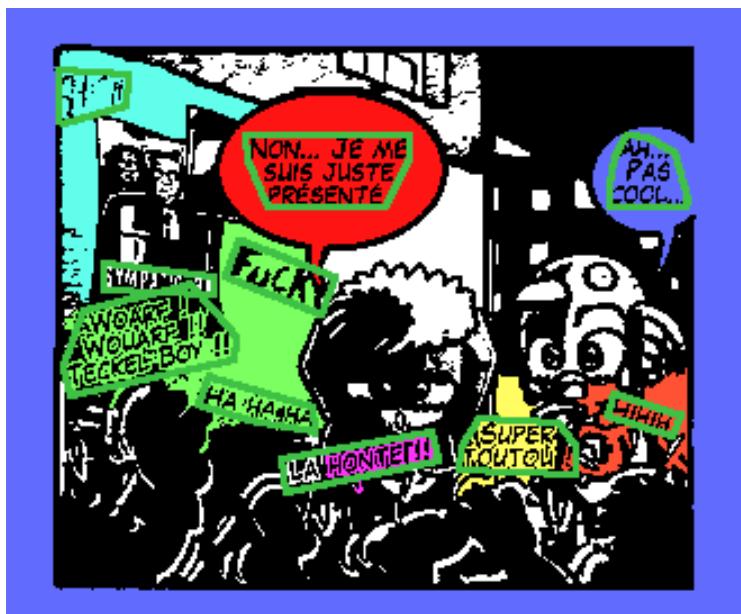
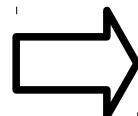
Balloon extraction: regular



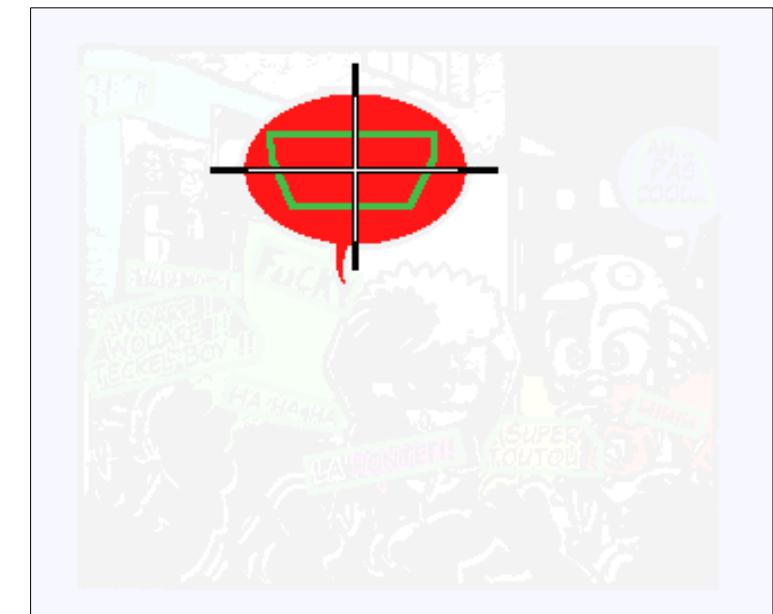
Original image



Text block positions (green)



Regions including text blocks (coloured)



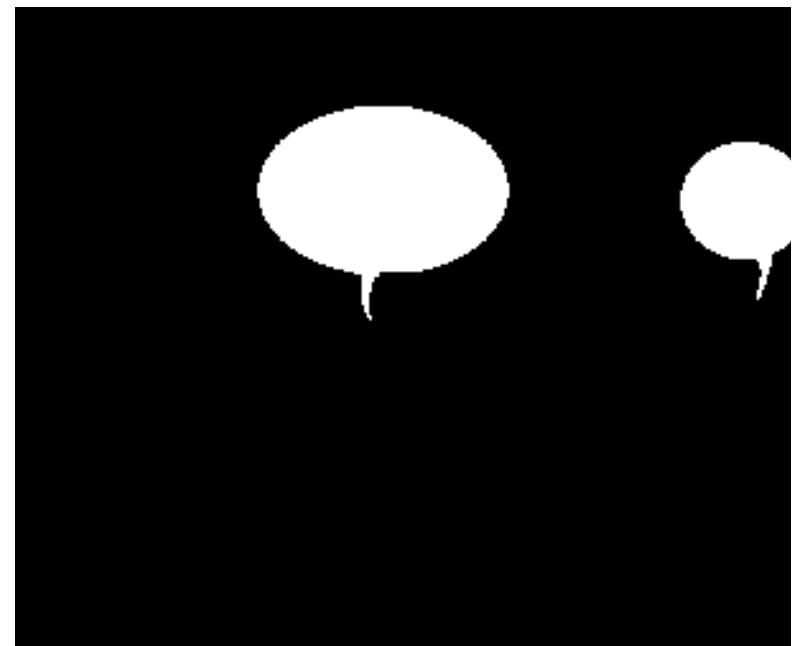
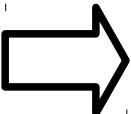
Regions including aligned text blocks

Contributions
Content driven approach

Balloon extraction: implicit

Contributions
Content-driven approach

- Assumptions
 - Panel and text blocks positions are known
 - Implicit balloons contain centred text
- Proposition
 - Extract implicit balloons from text regions by inflating a deformable contour
 - Adaptation of active contour model (snake)

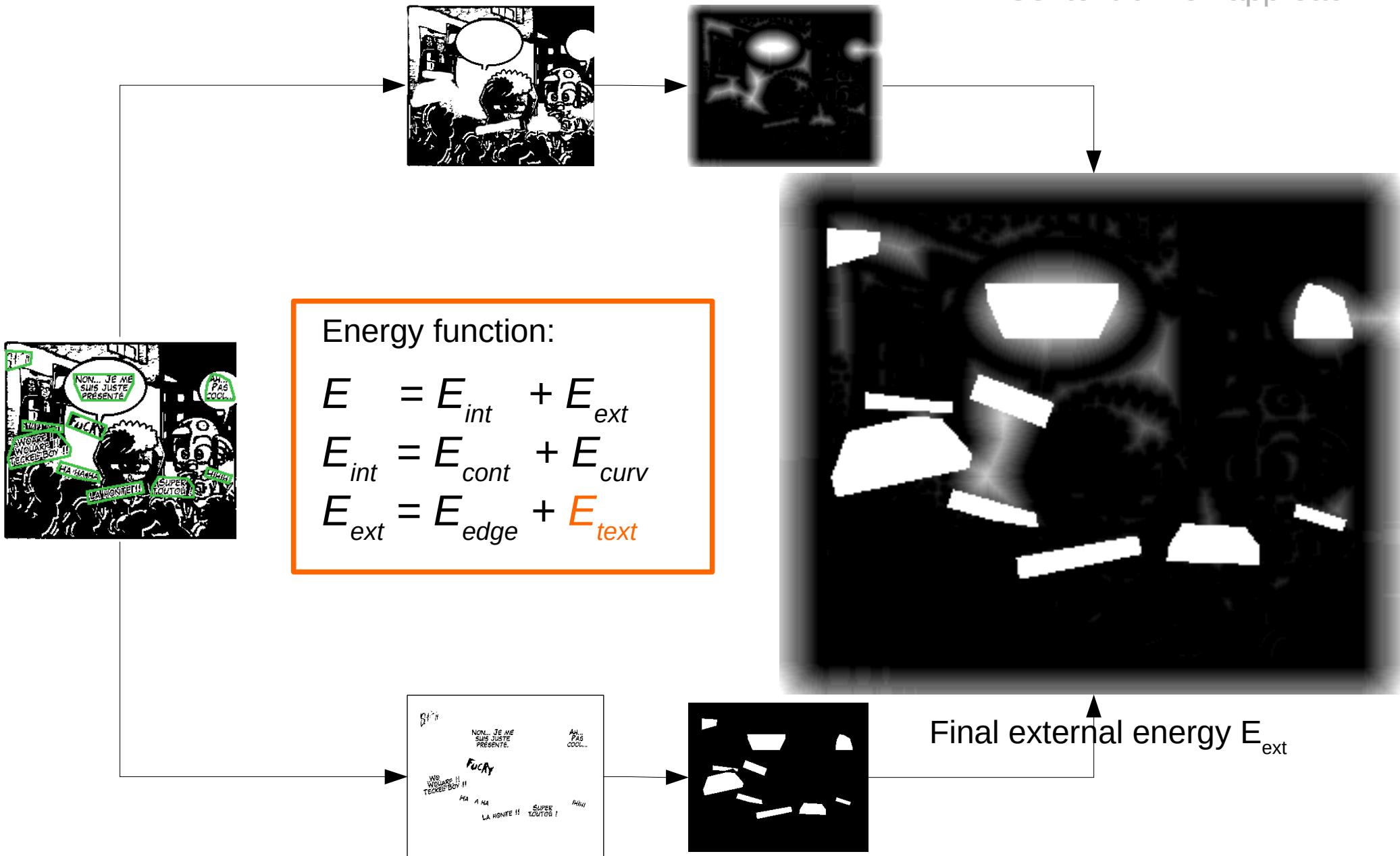


Original image and text locations

Expected result

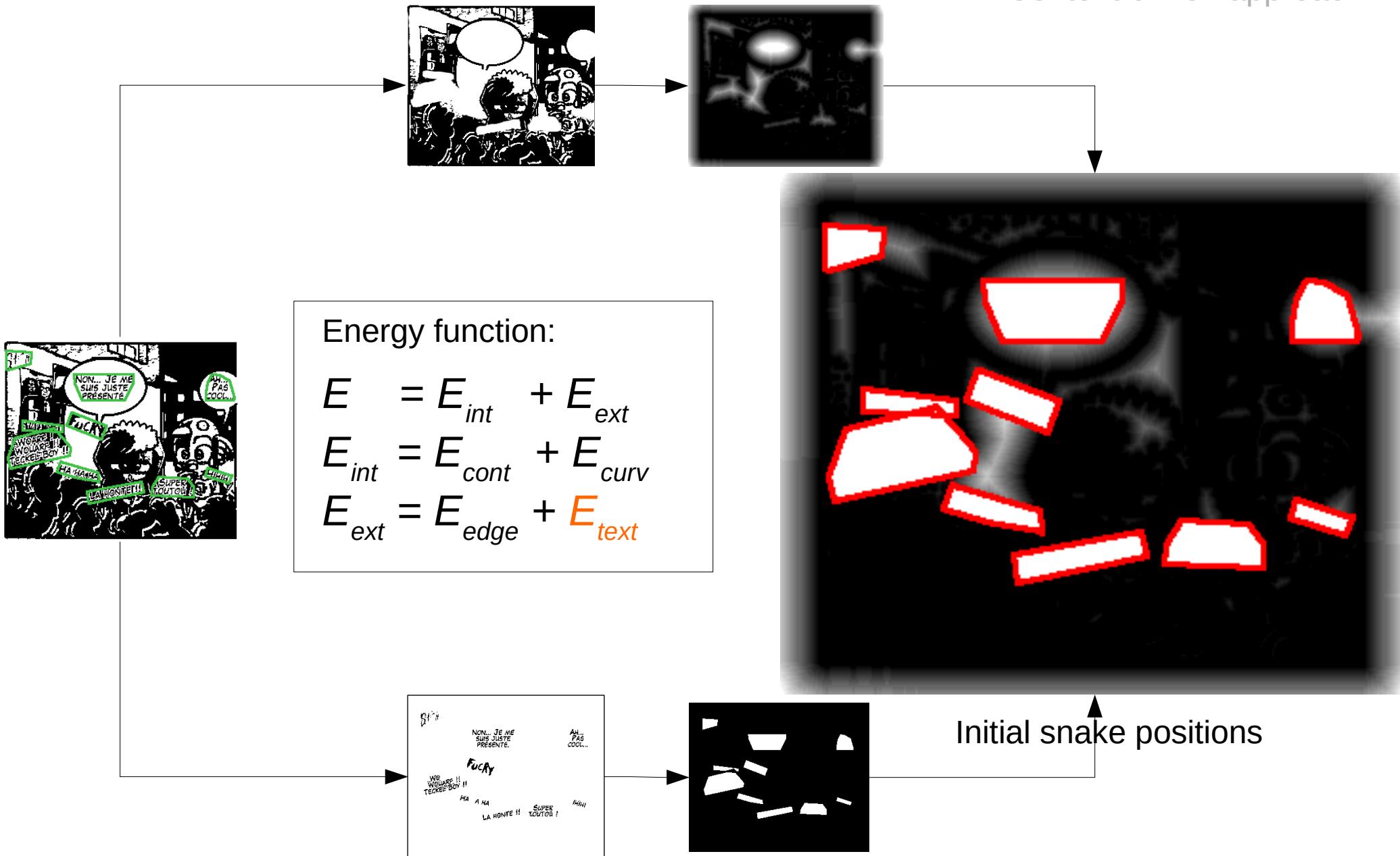
Balloon extraction: implicit

Contributions
Content-driven approach



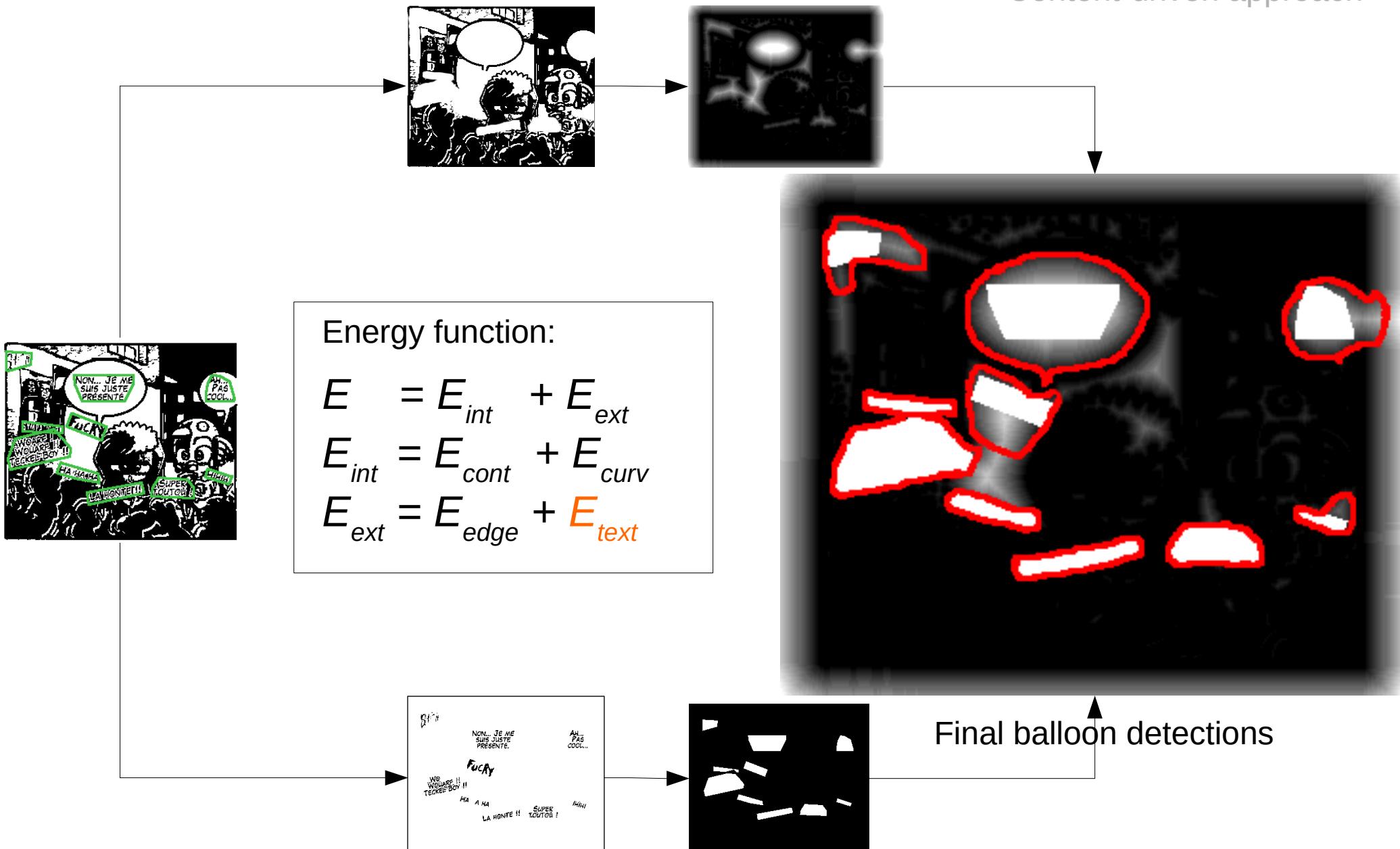
Balloon extraction: implicit

Contributions
Content-driven approach



Balloon extraction: implicit

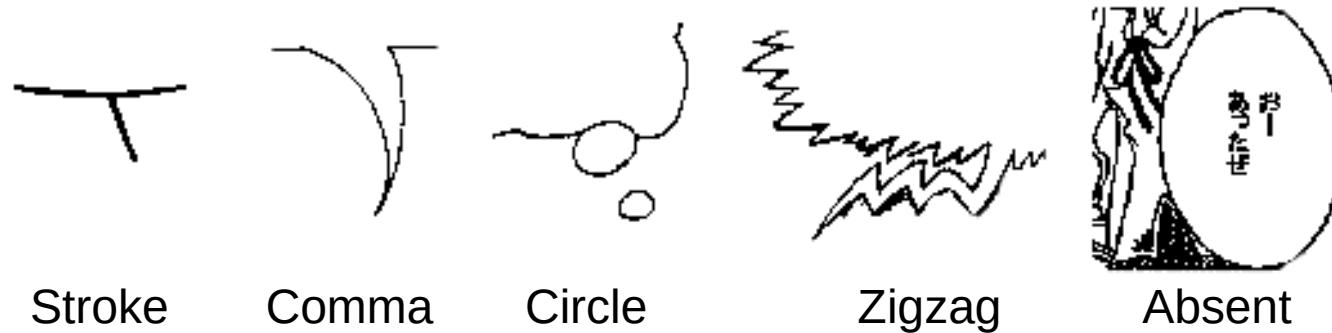
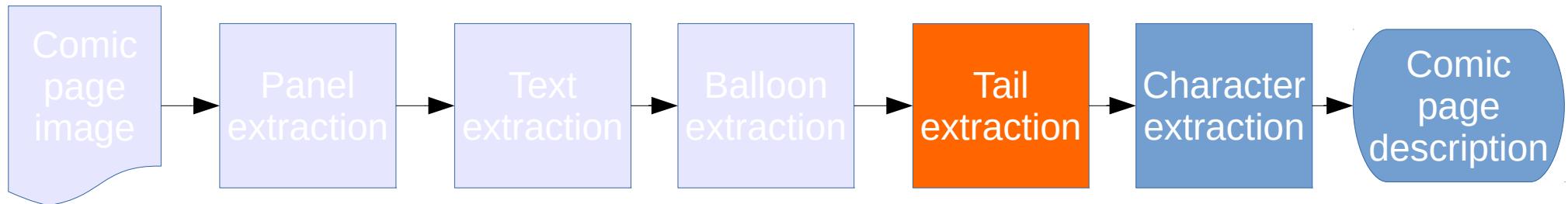
Contributions
Content-driven approach



The snake is attracted to the “dark side”

Tail extraction

Contributions
Content-driven approach



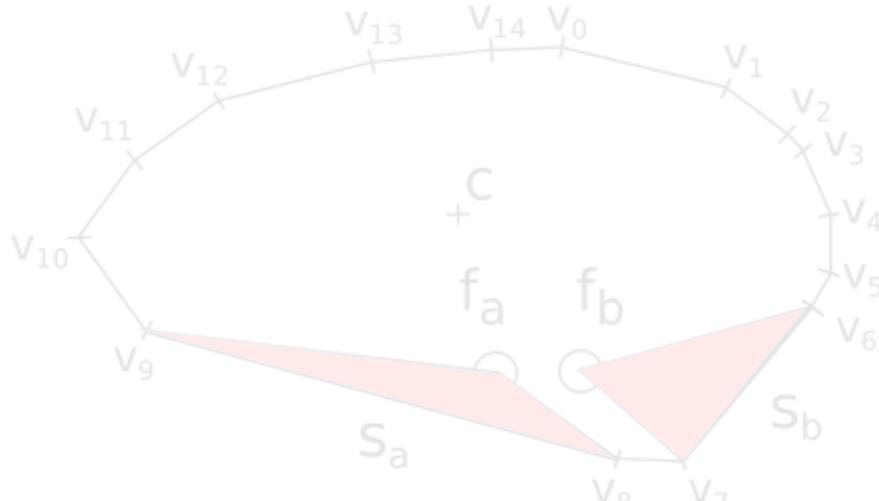
- Literature
 - First time studied in document image analysis
- Objectives
 - Detection of the tail tip position and orientation
 - Focus on comma, zigzag and absent types

Tail extraction: tip position

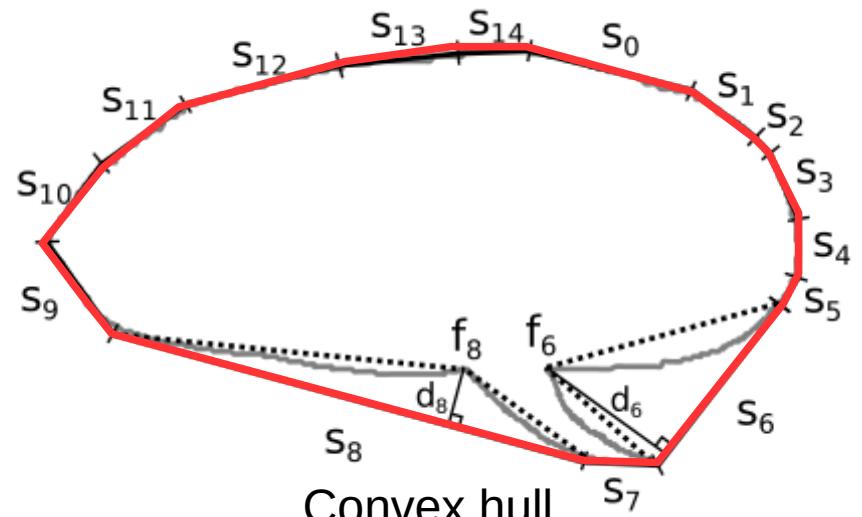
Contributions
Content-driven approach



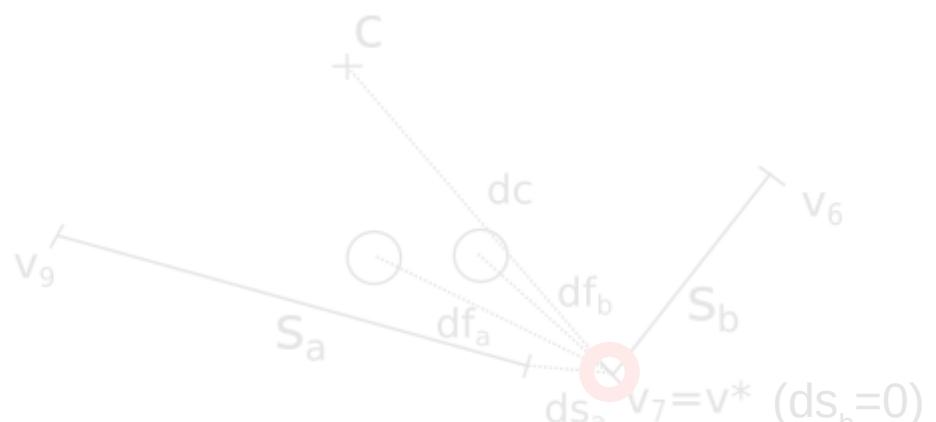
Balloon contour



Two biggest
convexity defects



Convex hull



Tail tip position

Optimal vertex selection:

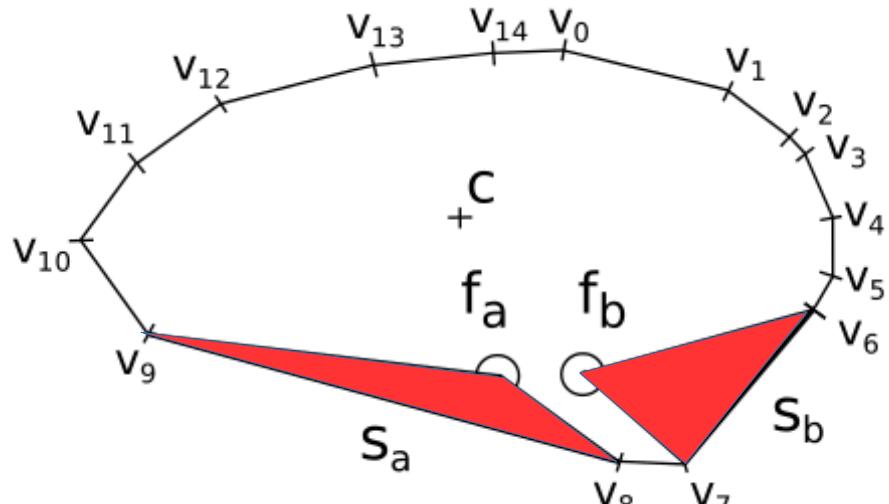
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

Tail extraction: tip position

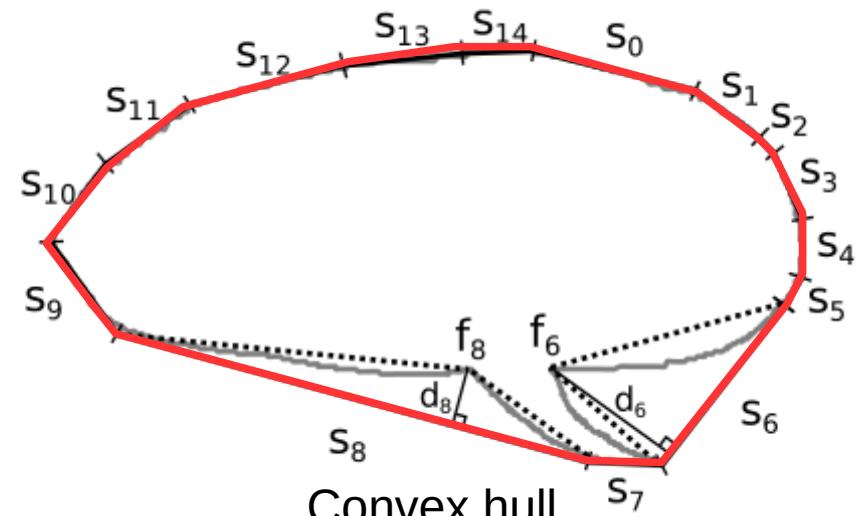
Contributions
Content-driven approach



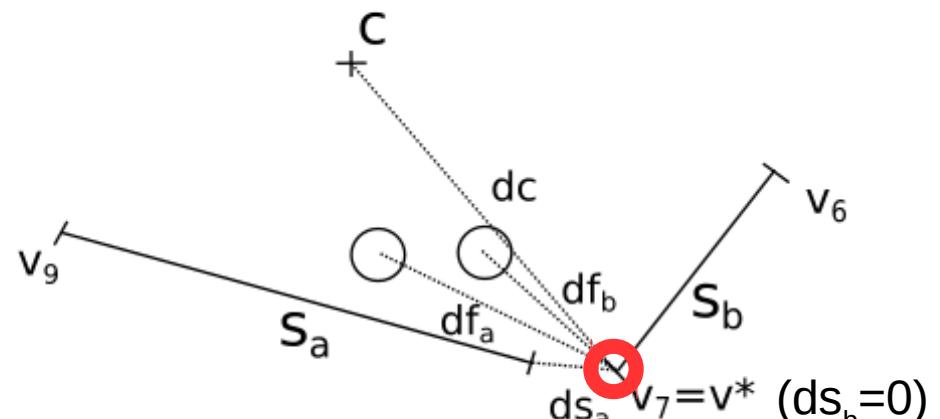
Balloon contour



Two biggest
convexity defects



Convex hull



Tail tip position

Optimal vertex selection:

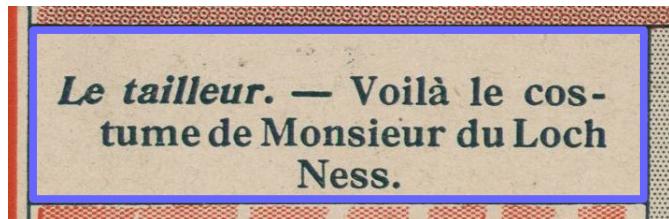
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

Tail extraction: confidence value

Contributions
Content-driven approach

Balloon
contour (blue)

Balloon 1

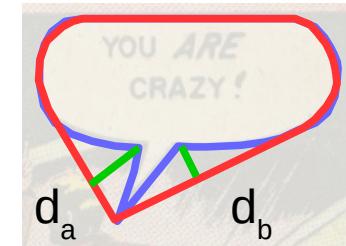
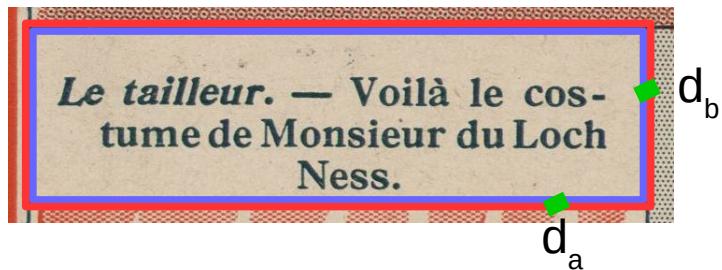


Convex hull
(red)

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)



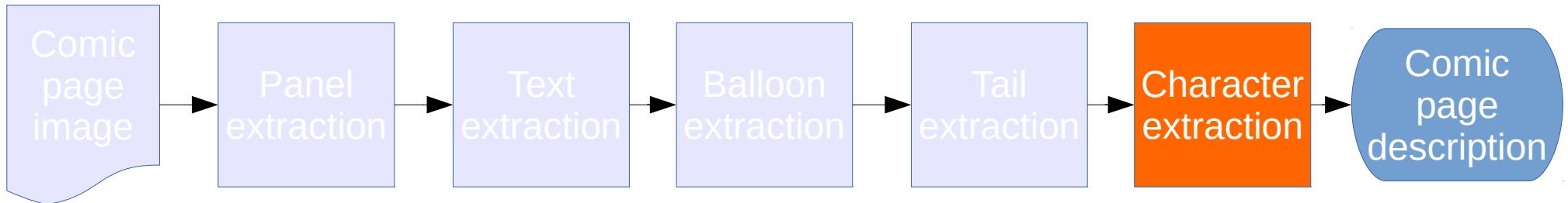
Farthest point
from origin and tip
(red)



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

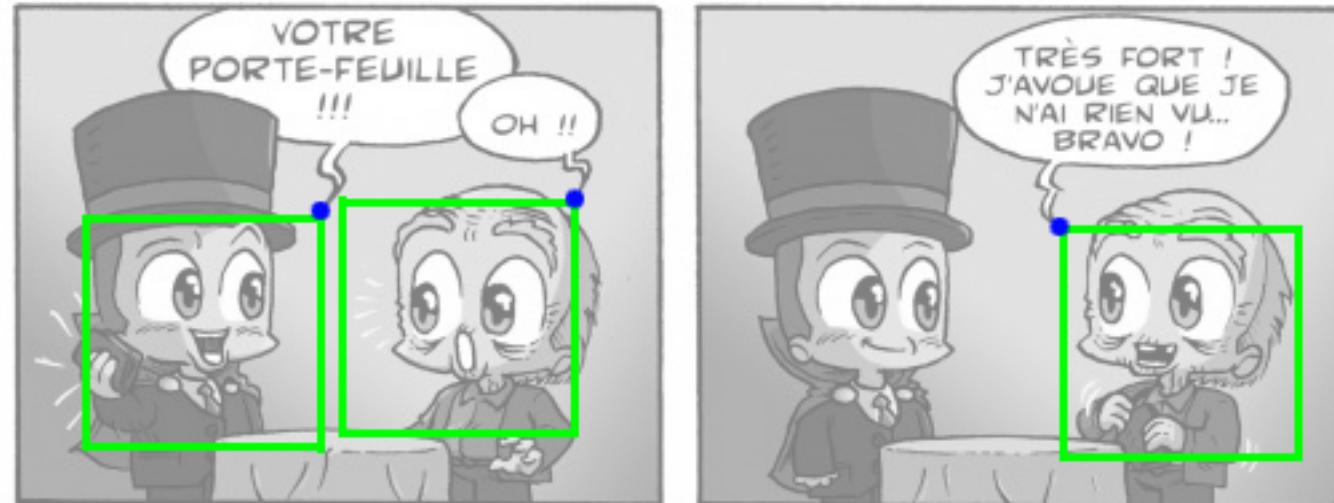
Contributions
Content-driven approach

Panels + Tails = Comic character ROIs

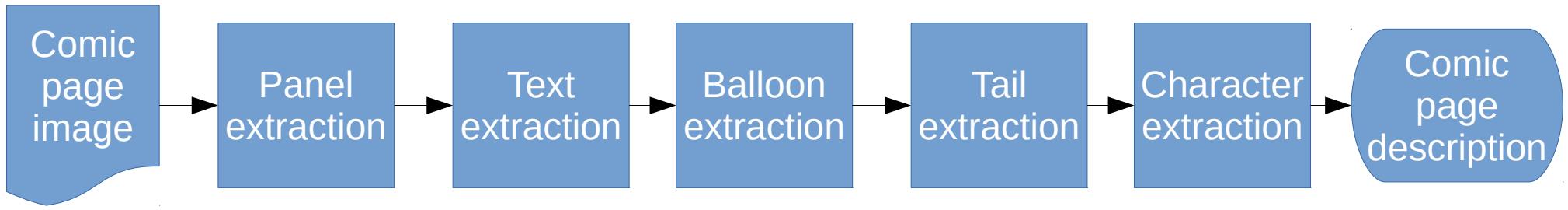
Large ROI



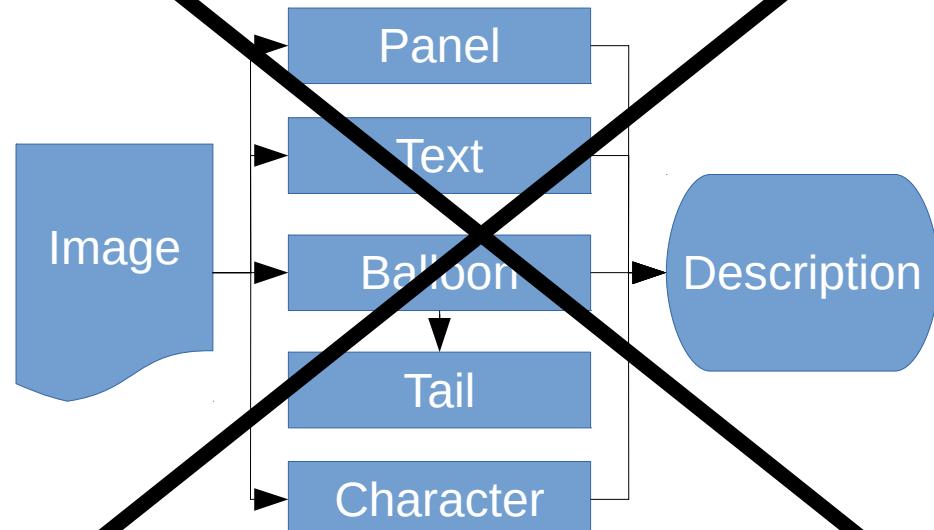
Small ROI



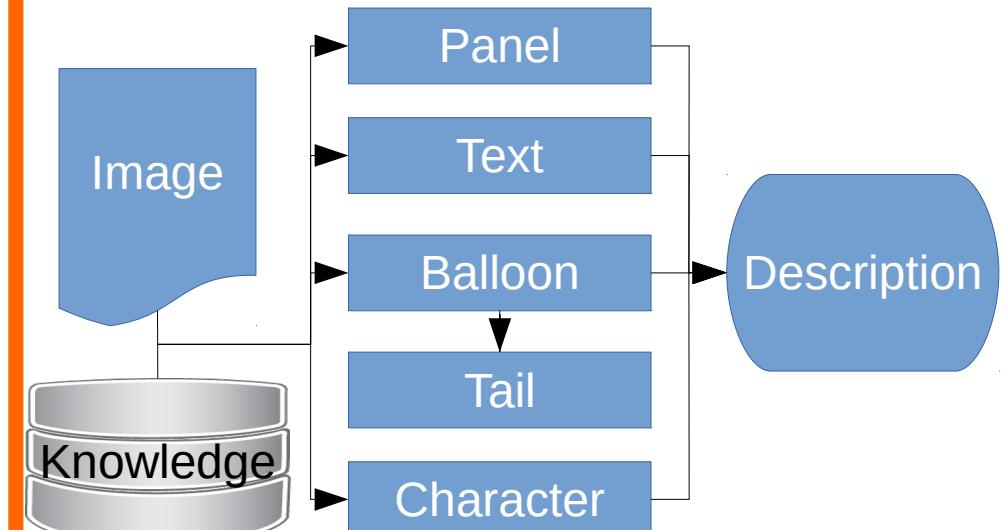
Content-driven (sequential)



Content-driven (independent)

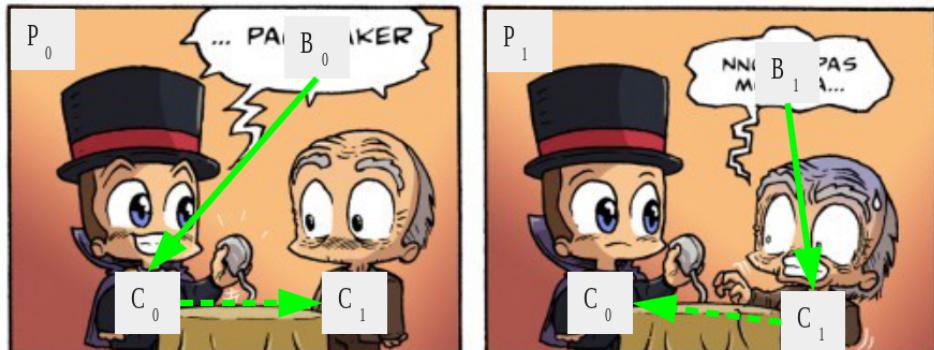


Knowledge-driven (independent)



Introduction

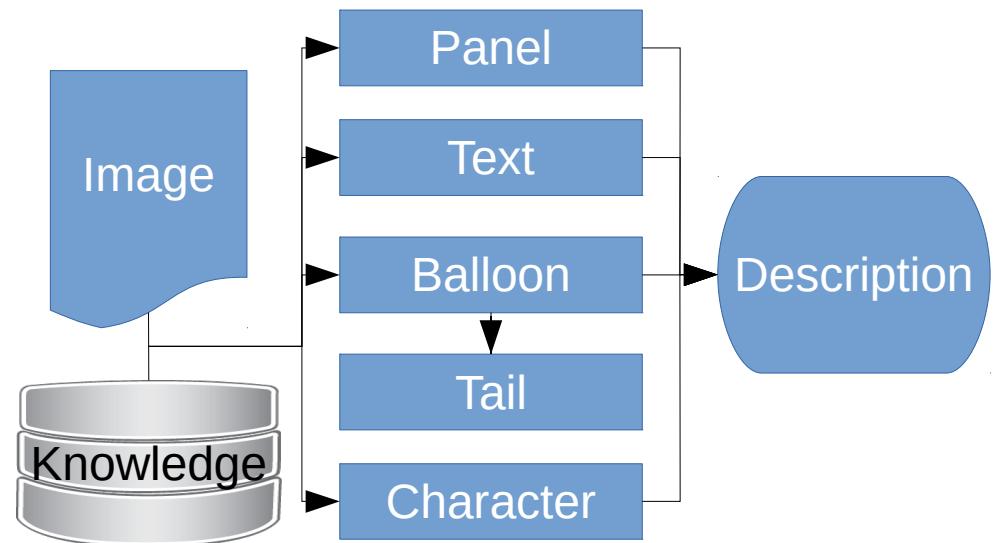
- High level image description
- Independent element extraction
- Framework for comics understanding
- Collaboration with Clément Guérin



Example of situation understanding

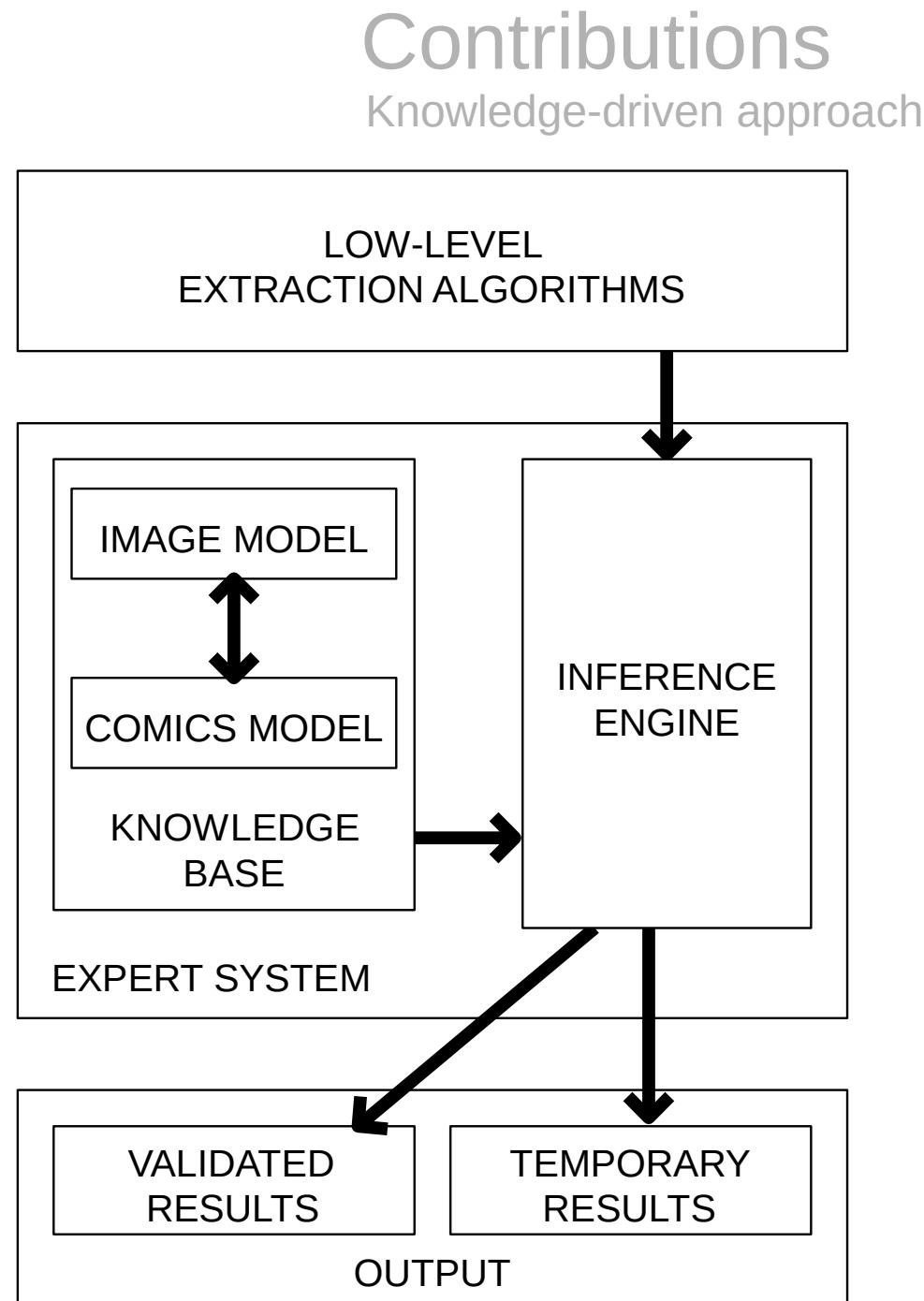
Contributions

Knowledge-driven approach



Knowledge representation

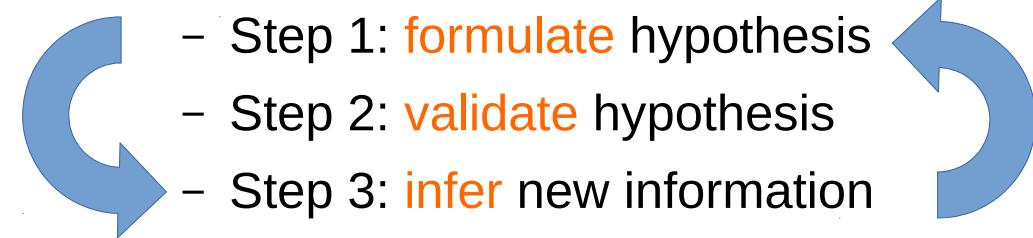
- Image model
 - Physical support
 - Region of interests
- Comics model
 - Validations
 - A **panel P** is related to one page
 - A **balloon B** is related to one panel and may have a **tail Q**
 - A **character C** is related to one panel
 - A **text line T** is related to one balloon
 - Inferences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



Processing sequence

- 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
 - ...

Contributions

Knowledge-driven approach



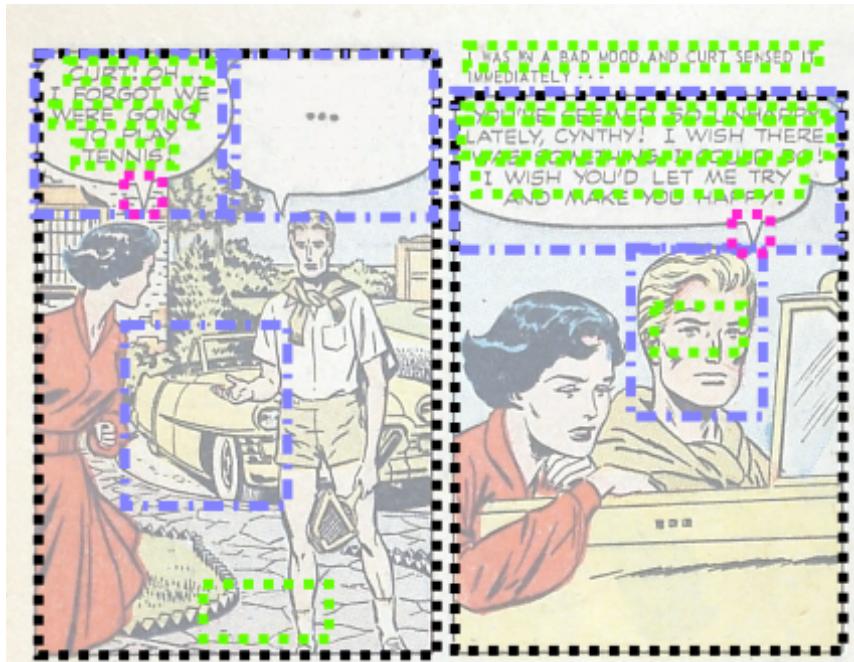
Processing sequence

Contributions
Knowledge-driven approach

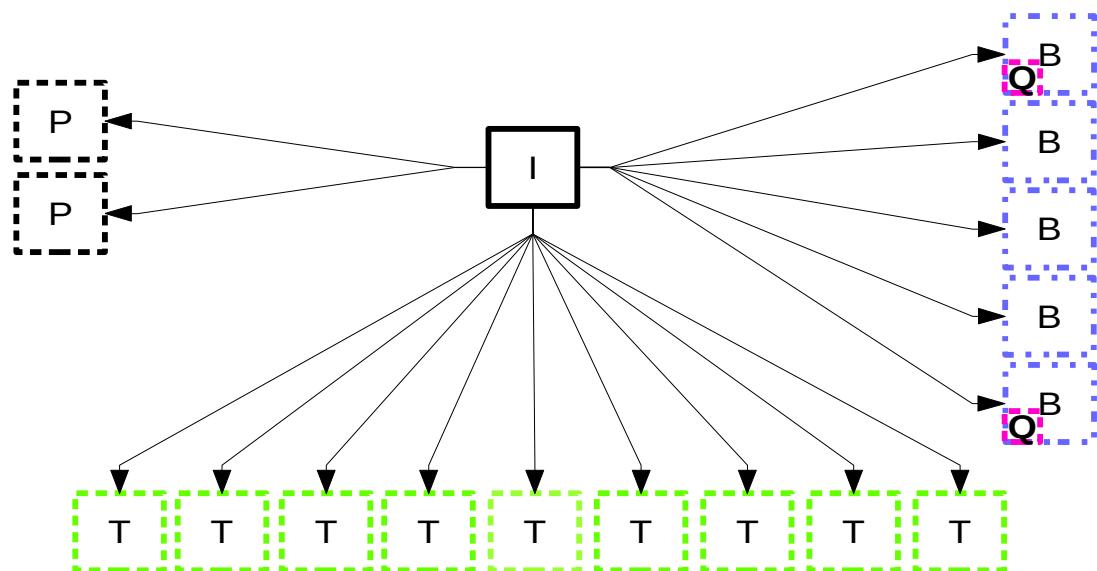


Processing sequence

Contributions
Knowledge-driven approach



Hypotheses of
panels, balloons
and text lines

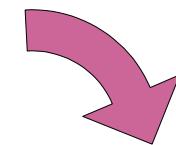


Processing sequence

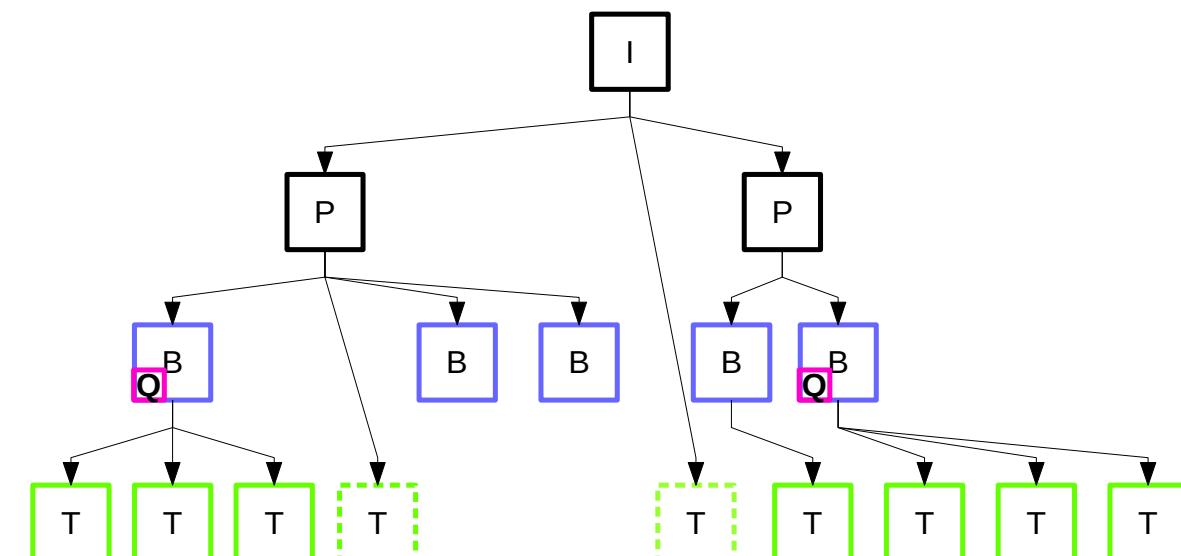
Contributions
Knowledge-driven approach



Hypotheses of
panels, balloons
and text lines

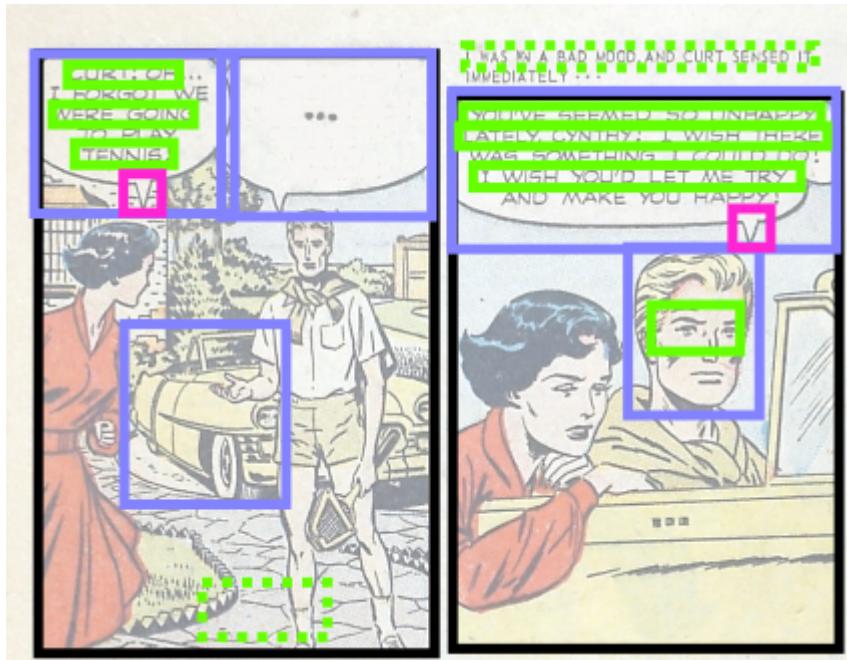


Validation of the
hypotheses

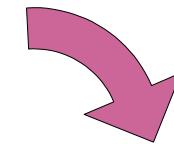


Processing sequence

Contributions
Knowledge-driven approach



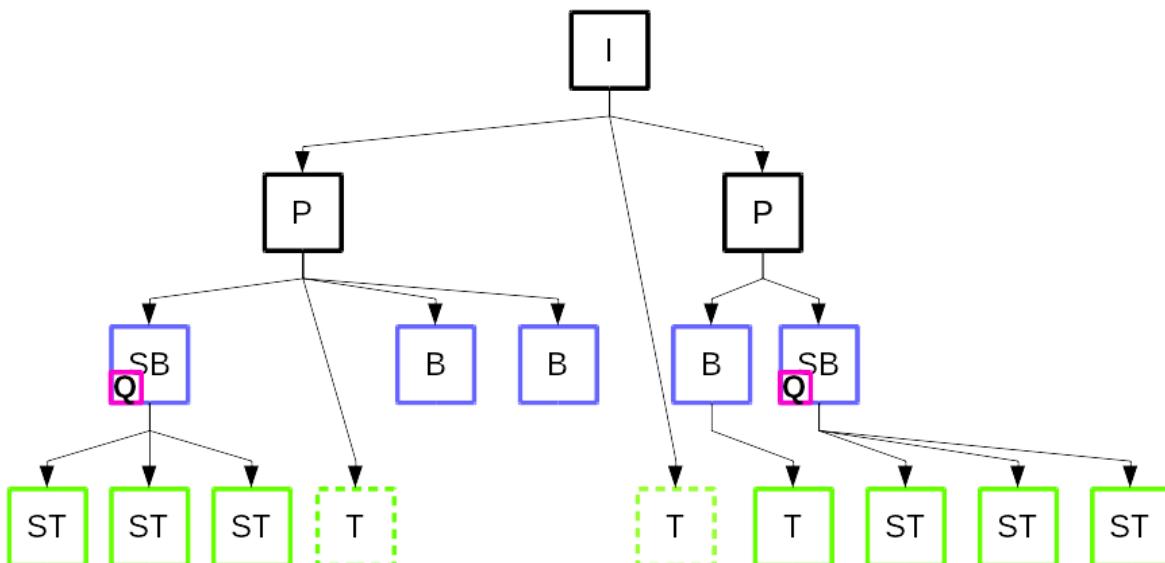
Hypotheses of
panels, balloons
and text lines



Validation of the
hypotheses



Inferences of
specific types



Processing sequence

Contributions
Knowledge-driven approach

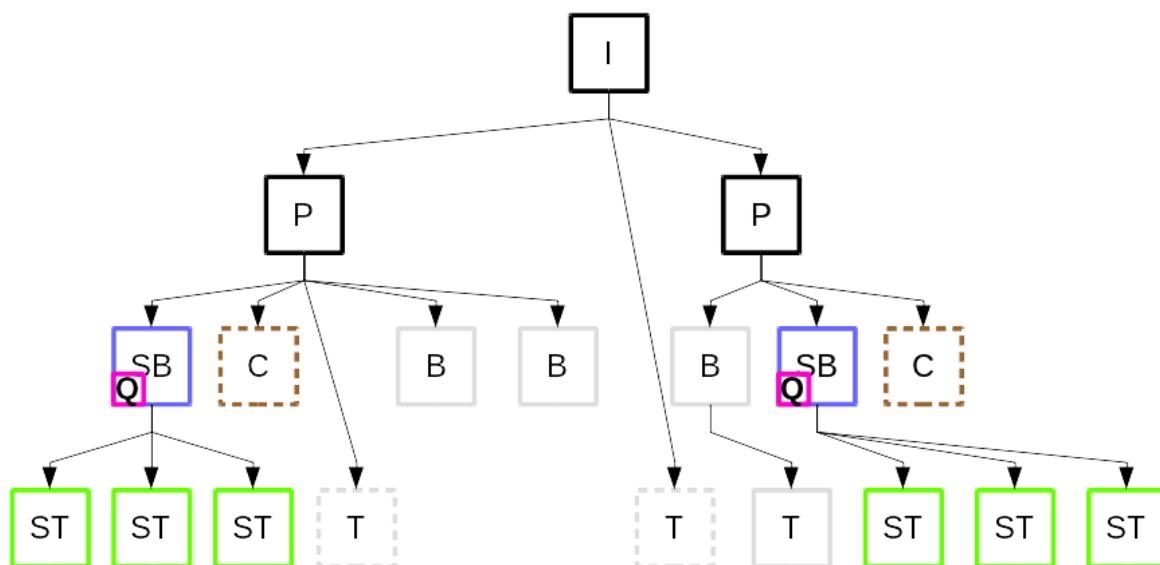


Hypotheses of
comic characters



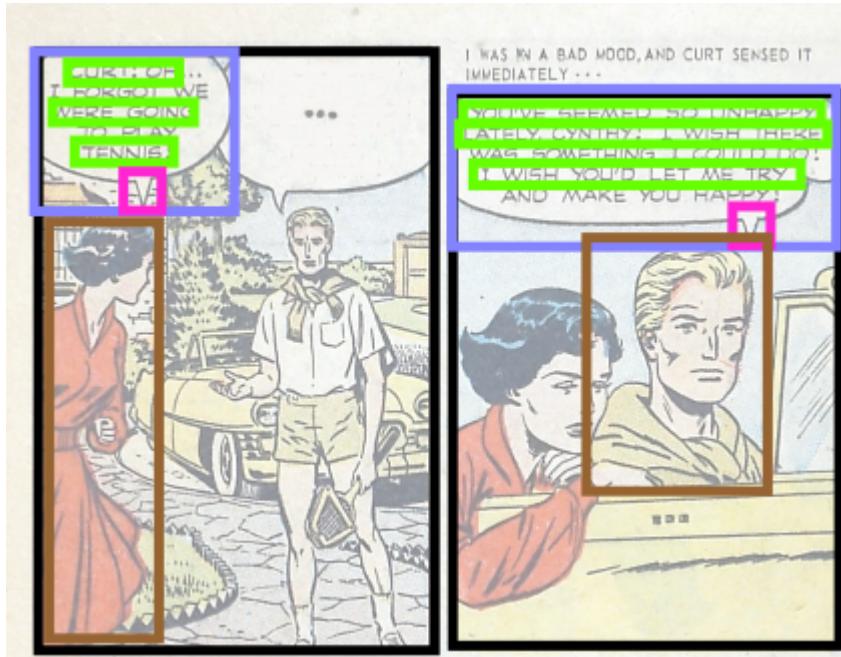
Validation of the
hypotheses

Inferences of
specific types

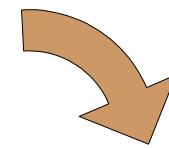


Processing sequence

Contributions
Knowledge-driven approach



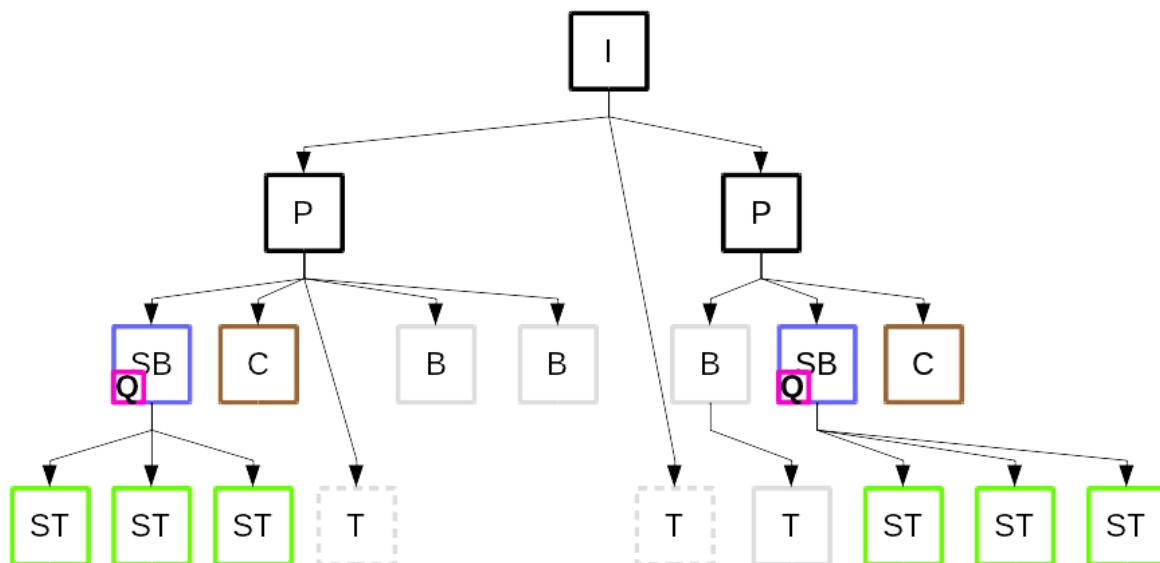
Hypotheses of
comic characters



Validation of the
hypotheses

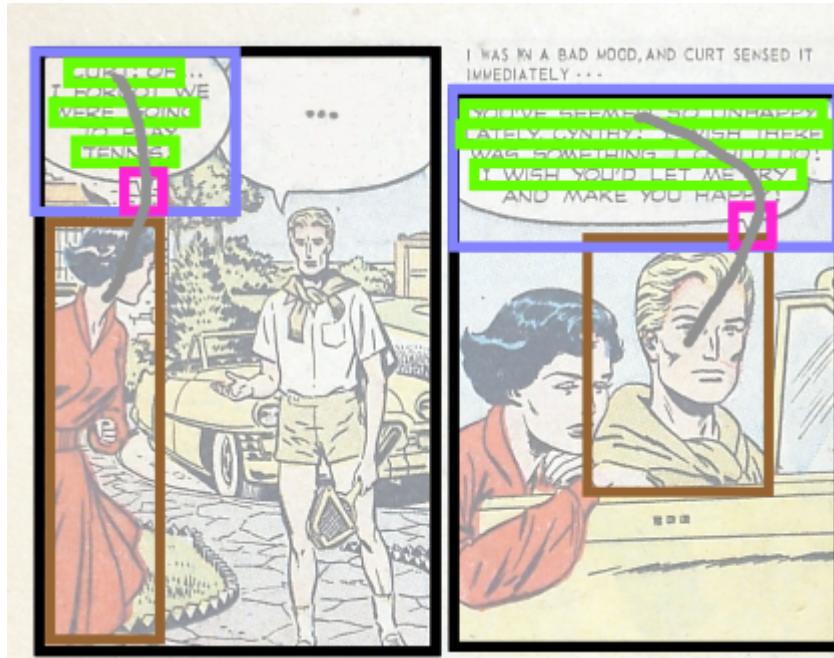


Inferences of
specific types

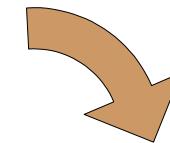


Processing sequence

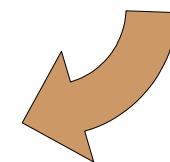
Contributions
Knowledge-driven approach



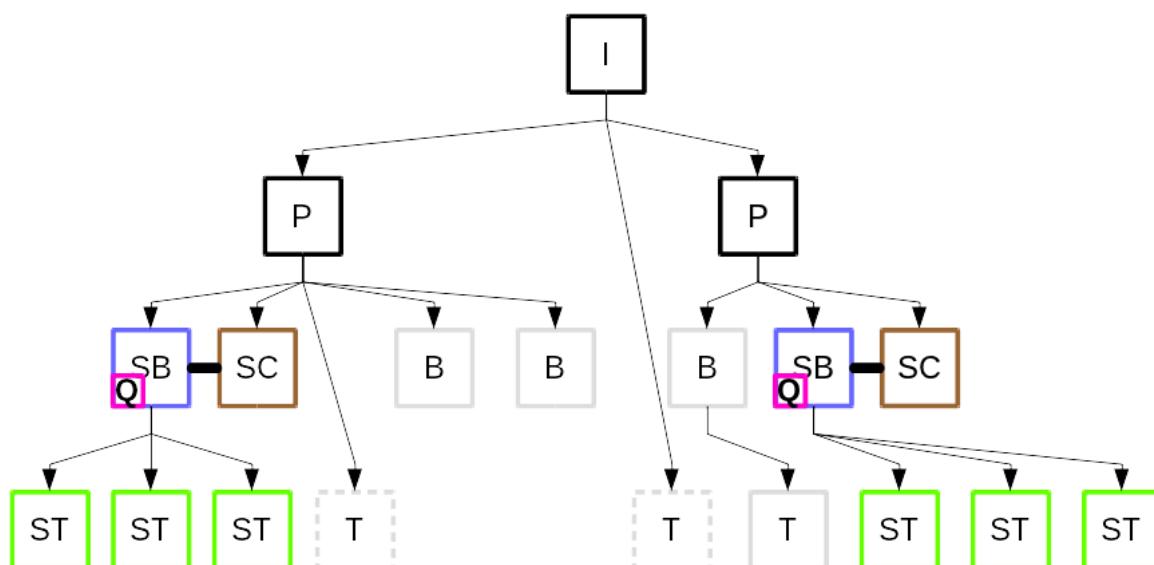
Hypotheses of
comic characters



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

- 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>

Experiments

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

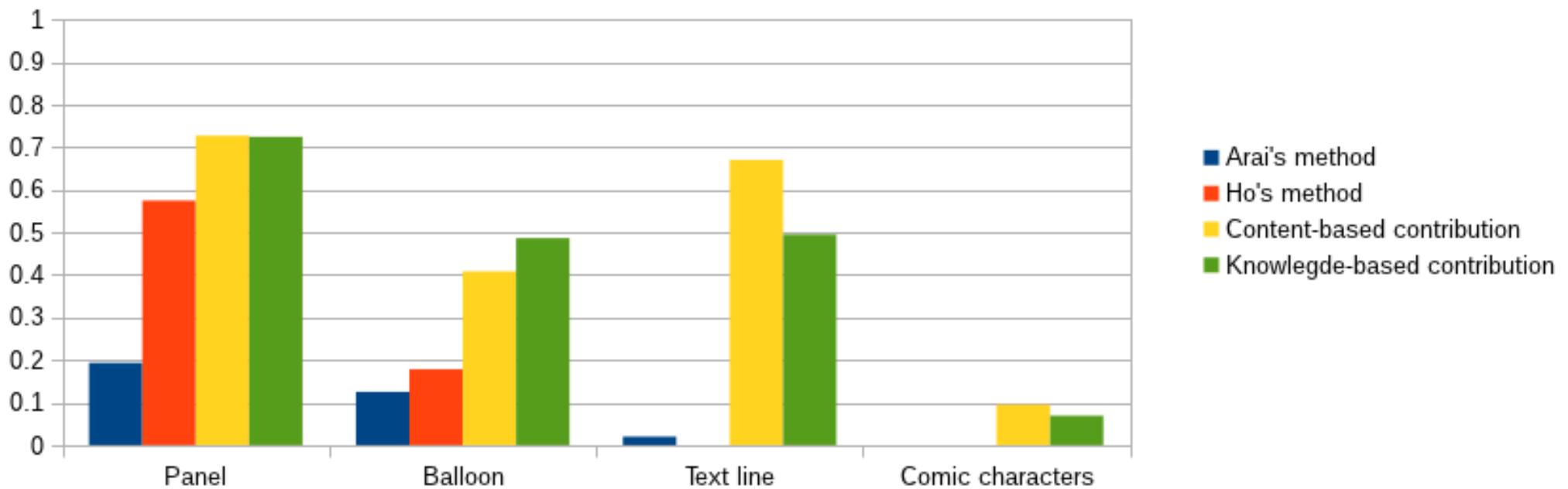
$$a_0 = \frac{\text{area}(B_p \cup B_{gt})}{\text{area}(B_p \cap B_{gt})}$$

B_p = predicted region

B_{gt} = ground truth region

B_p valid if $a_0 > 0.5$

Element localisation results (F-score)



- 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
 - **TODO: revisit objectives/challenges from intro**
- 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 (**TODO: replace by numbers and move to next page?**)
 - **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 (PHC-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

- TODO

Conclusion

References

Conclusion

- [Chung07] ChungHo Chan, Howard Leung, and Taku Komura. Automatic panel extraction of color comic images. In HoraceH.-S. Ip, OscarC. Au, Howard Leung, Ming-Ting Sun, Wei-Ying Ma, and Shi-Min Hu, editors, *Advances in Multimedia*
- [Eunjung07] Eunjung Han, Kirak Kim, HwangKyu Yang, and Keechul Jung. Frame segmentation used mlp-based x-y recursive for mobile cartoon content. In *Proceedings of the 12th international conference on Human-computer interaction: intelligent multimodal interaction environments, HCI'07*, pages 872–881, Berlin, Heidelberg, 2007. Springer-Verlag
- [Tanaka07] Takamasa Tanaka, Kenji Shoji, Fubito Toyama, and Juichi Miyamichi. Layout analysis of tree-structured scene frames in comic images. In *IJCAI'07*, pages 2885–2890, 2007
- [Arai10] Kohei Arai and Herman Tolle. Method for automatic e-comic scene frame extraction for reading comic on mobile devices. In *Seventh International Conference on Information Technology: New Generations, ITNG '10*, pages 370–375, Washington, DC, USA, 2010. IEEE Computer Society
- [Pang14] Xufang Pang, Ying Cao, Rynson W.H. Lau, and Antoni B. Chan. A robust panel extraction method for manga. In *Proceedings of the ACM International Conference on Multimedia, MM '14*, pages 1125–1128, New York, NY, USA, 2014

References

Conclusion

- [Stommel12] Martin Stommel, Lena I Merhej, and Marion G Müller. Segmentation-free detection of comic panels. In Computer Vision and Graphics, pages 633–640. Springer, 2012
- [Li14a] Luyuan Li, Yongtao Wang, Zhi Tang, and Liangcai Gao. Automatic comic page segmentation based on polygon detection. *Multimedia Tools Applications*, 69(1):171–197, 2014
- [Li14b] Luyuan Li, Yongtao Wang, Zhi Tang, Xiaoqing Lu, and Liangcai Gao. Unsupervised speech text localization in comic images. In Proceedings of International Conference on Document Analysis and Recognition (ICDAR), pages 1190–1194, Aug 2013
- [Ponsard12] Christophe Ponsard, Ravi Ramdoyal, and Daniel Dziamski. An ocr-enabled digital comic books viewer. In Computers Helping People with Special Needs, pages 471–478. Springer, 2012

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

Production to interpretation

Background

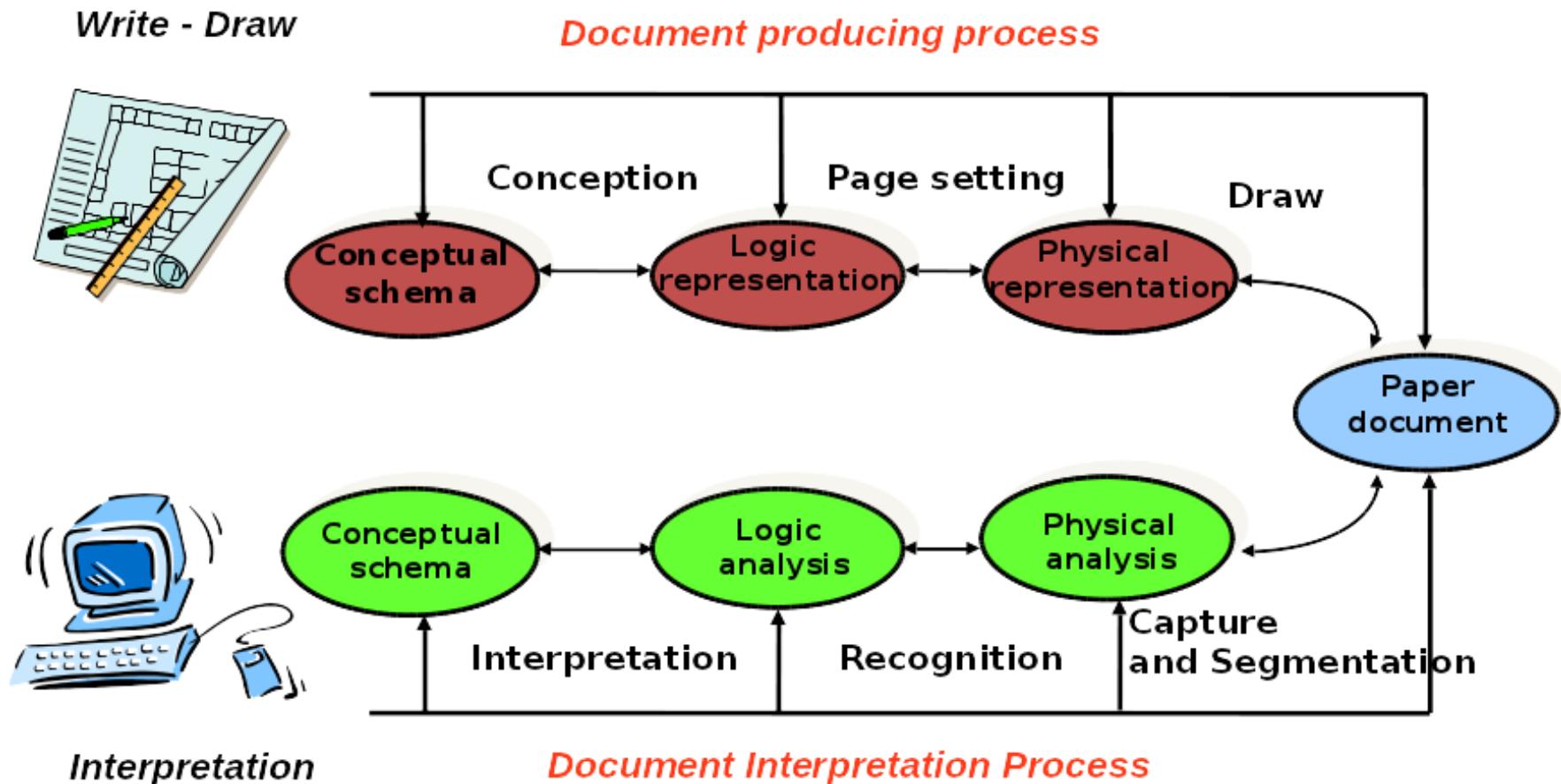


Image source: Handbook of Document Image Processing and Recognition. Springer, 2014

Background



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