



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

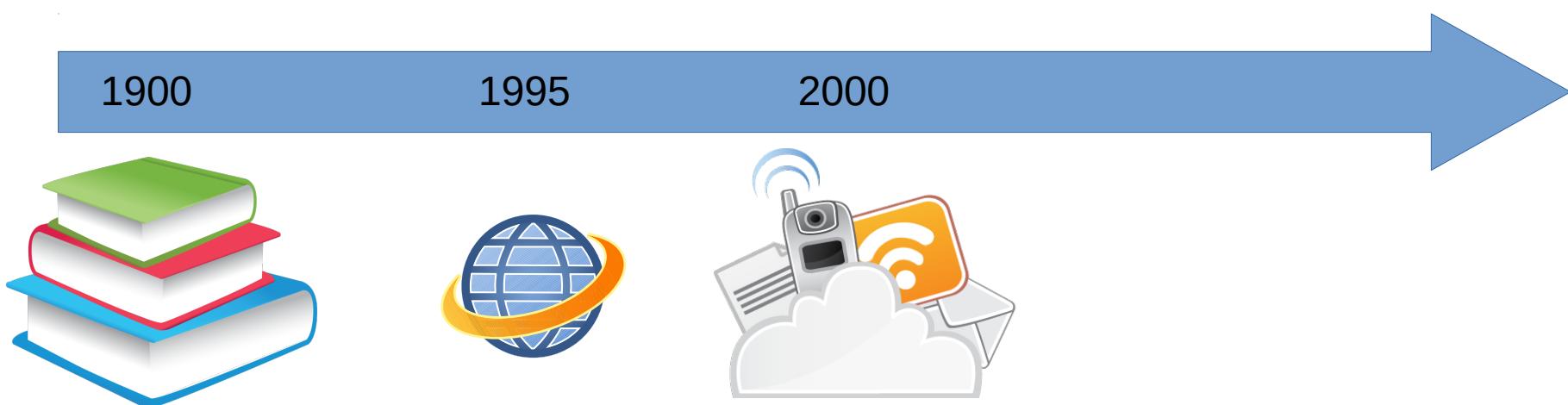
Introduction

“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

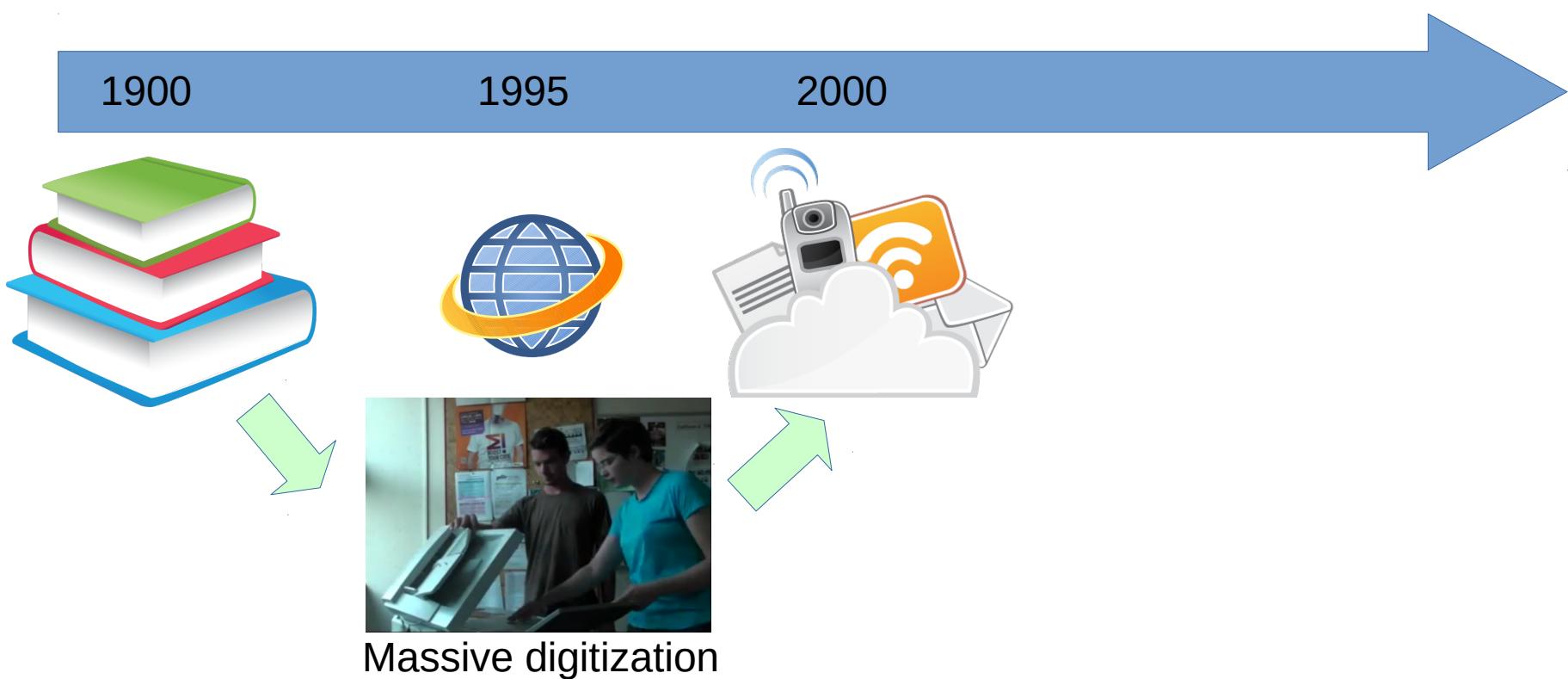
Introduction

“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

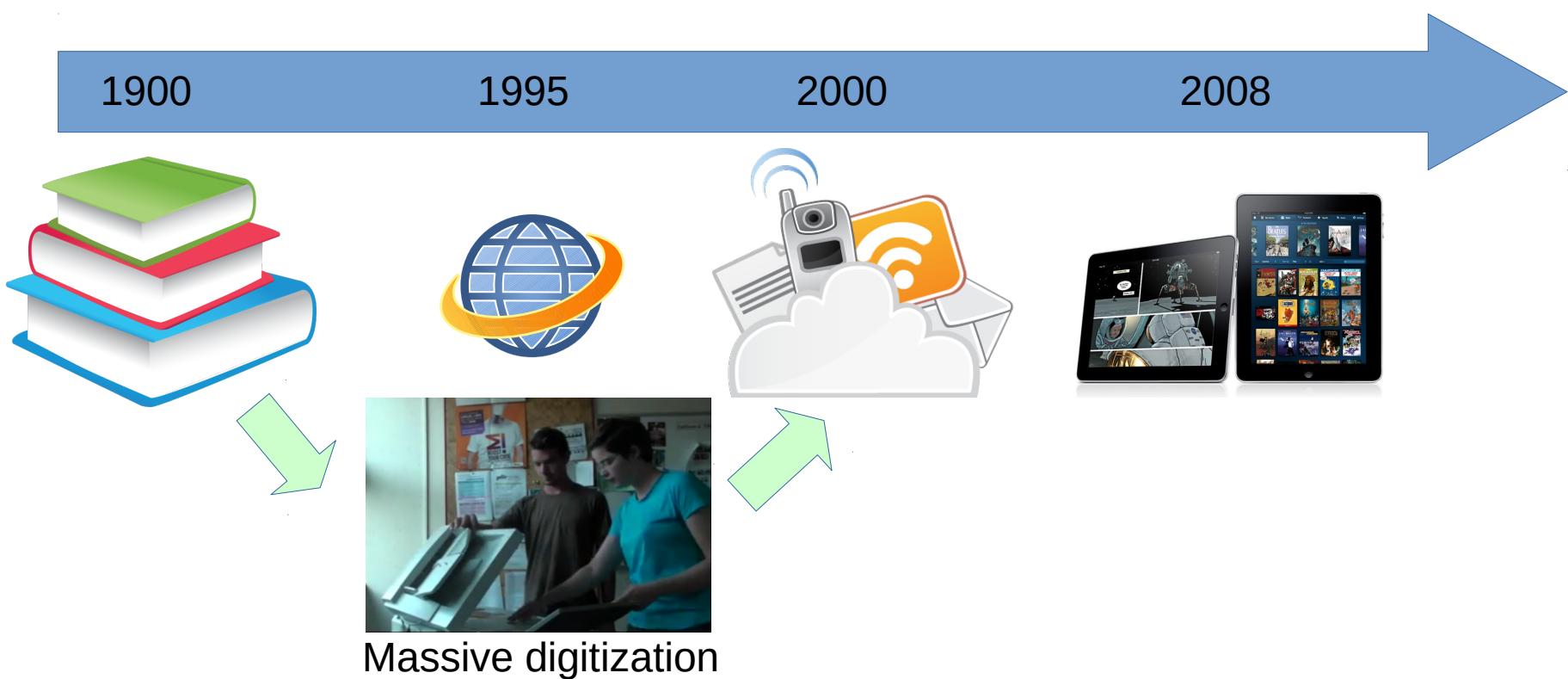
Introduction

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

Wikipédia, 2014

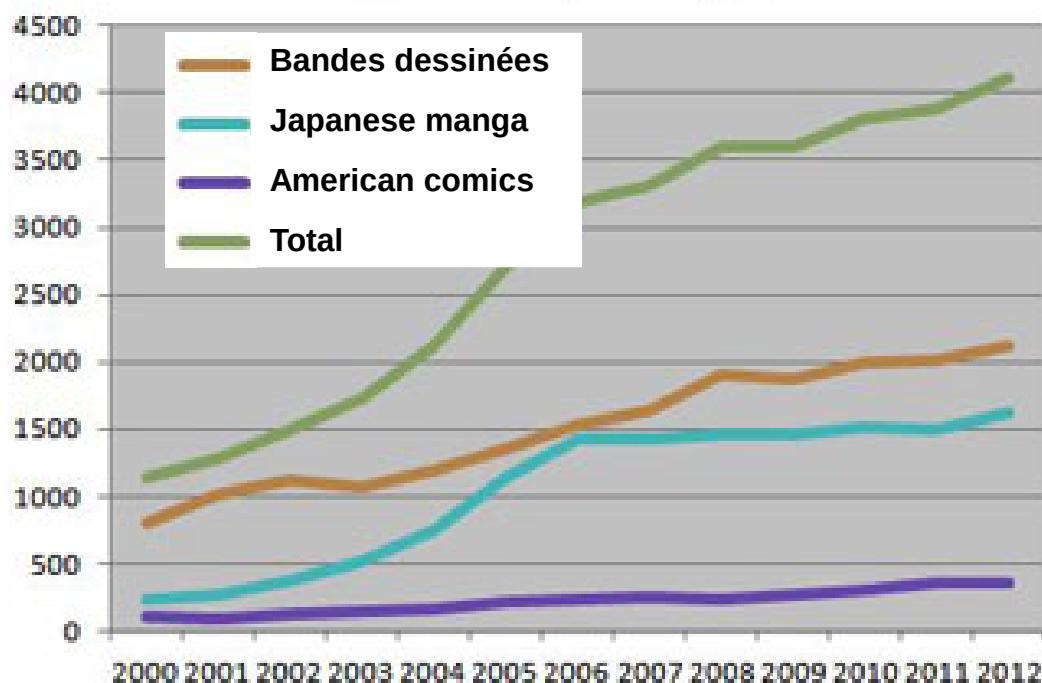
“one of the most popular and familiar forms of graphic content”

Hiroaki Tobita, Sony CSL Interaction Laboratory, 2014



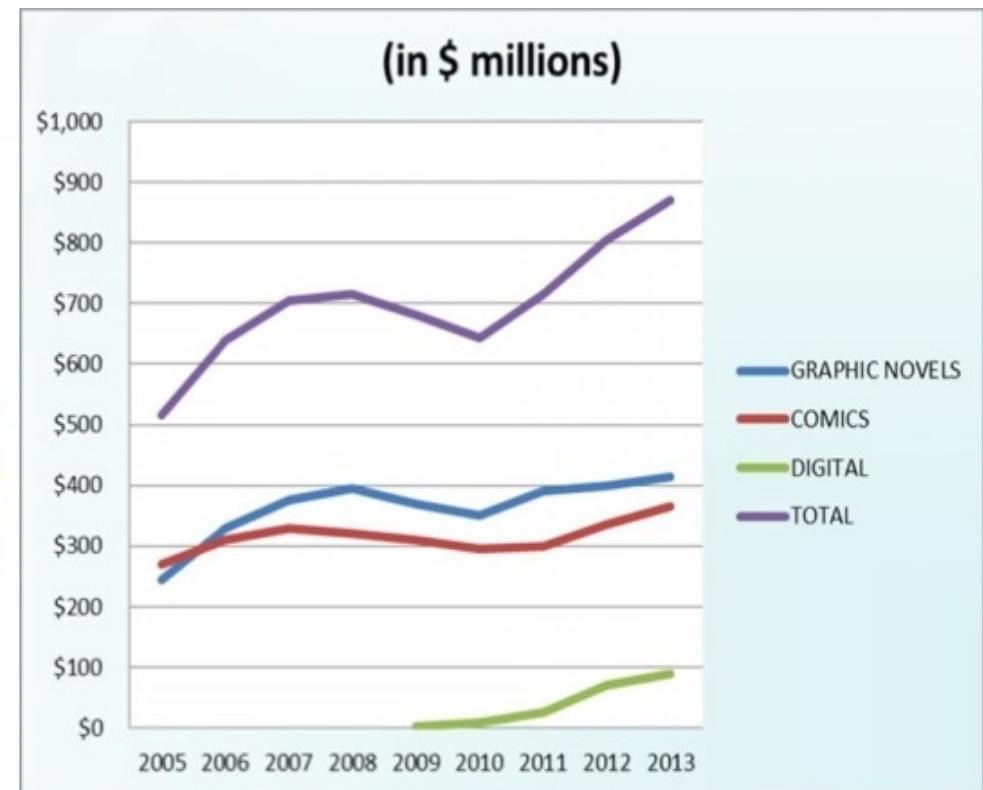
Comic books

Introduction



1) Francophone comics production

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



2) 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**
 - Content extraction (thesis of Christophe Rigaud)
 - Knowledge representation (thesis of Clément Guérin)
 - Public founding CPER 2007-2013
 - 2 Ph.D. students, 1 engineer, 1 post doc, 6 professors (L3i)
- Scientific challenges
 - Mixed contents of a **graphical** and **textual** nature
 - Combination of the difficulties of **free-form** and **complex background** documents
 - Recent and **unexplored** field of research
- Objectives
 - Propose **generic approaches** able to **retrieve** as many **elements** as possible from **any comic book image**
 - Provide a **first dataset** and **ground truth**

- eBDthèque project (since 2011)
 - Add value to **digitized comics** using the **new technologies**
 - Content extraction (thesis of Christophe Rigaud)
 - Knowledge representation (thesis of Clément Guérin)
 - Public founding CPER 2007-2013
 - 2 Ph.D. students, 1 engineer, 1 post doc, 6 professors (L3i)
- Scientific challenges
 - Mixed contents of a **graphical** and **textual** nature
 - Combination of the difficulties of **free-form** and **complex background** documents
 - Recent and **unexplored** field of research
- Objectives
 - Propose **generic approaches** able to **retrieve** as many **elements** as possible from **any comic book image**
 - Provide a **first dataset** and **ground truth**

- eBDthèque project (since 2011)
 - Add value to **digitized comics** using the **new technologies**
 - Content extraction (thesis of Christophe Rigaud)
 - Knowledge representation (thesis of Clément Guérin)
 - Public founding CPER 2007-2013
 - 2 Ph.D. students, 1 engineer, 1 post doc, 6 professors (L3i)
- Scientific challenges
 - Mixed contents of a **graphical** and **textual** nature
 - Combination of the difficulties of **free-form** and **complex background** documents
 - Recent and **unexplored** field of research
- Objectives
 - Propose **generic approaches** able to **retrieve** as many **elements** as possible from **any comic book image**
 - Provide a **first dataset** and **ground truth**

Introduction

Background

Contributions

Experiments

Conclusion



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

- Panel extraction
- Balloon extraction
- Text extraction & recognition
- Comic character extraction
- Overall progress



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

Panel extraction

Background

- Challenges

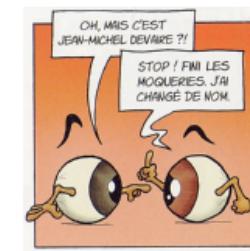
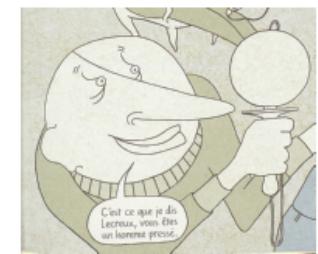
- Diversity of styles (gutter, implicit)
 - Semi-structured layout

- Panel extraction

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

- Conclusions

- Specific approaches only
 - Remaining difficulties for non-rectangle and implicit panels
 - Copyrighted images (not shareable)

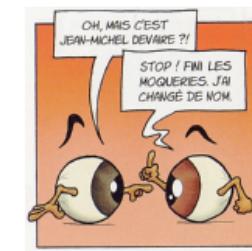


Panel extraction

Background

- Challenges

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



- Panel extraction

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

- Conclusions

- Specific approaches only
 - Remaining difficulties for non-rectangle and implicit panels
 - Copyrighted images (not shareable)

Balloon extraction

- Challenges

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

- Extraction

- **Connected-components** [Arai11, Ho12]

- Conclusions

- **Closed** balloon with text inside
- Several **unexplored** fields (e.g. **implicit** balloon positions, balloon **classification**, **tail detection**)

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

Balloon extraction

- Challenges

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

- Extraction

- **Connected-components** [Arai11, Ho12]

- Conclusions

- **Closed** balloon with text inside
- Several **unexplored** fields (e.g. **implicit** balloon positions, balloon classification, tail detection)

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

Text extraction & recognition

Background

- Challenges

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

- Extraction

- Sliding Concentric Windows + SVM [Su11]
- Connected-components [Ho12, Pang14]
- SVM and Bayesian classifier [Li14b]

- Recognition

- OCR trained for a specific comics font [Ponsard12]



- Conclusions

- Speech text (from balloons)
- Captions and sound effects unexplored
- Text recognition very poor

Text extraction & recognition

Background

- Challenges
 - Non-standard fonts
 - Multi-script/orientation/scale
 - Complex background (sound effects)
 - Hyphenation, voluntary spelling mistakes
- Extraction
 - Sliding Concentric Windows + SVM [Su11]
 - Connected-components [Ho12, Pang14]
 - SVM and Bayesian classifier [Li14b]
- Recognition
 - OCR trained for a specific comics font [Ponsard12]



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

Comic character extraction

Background

- Challenges

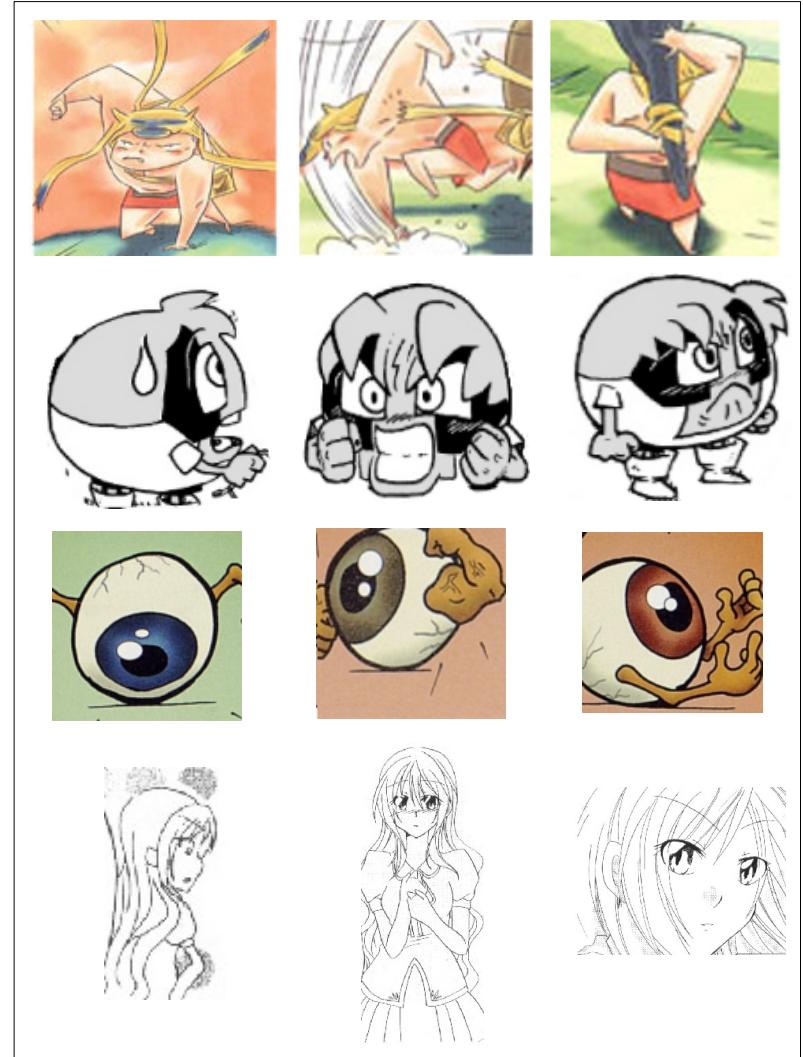
- Hand-drawn, stroke-based
- Intra/inter class variability
- Scale, deformation, posture, occlusion

- Extraction & recognition

- Manga faces [Cheung08, Sun10, Kohei12]
- Cartoons [Khan12]

- Conclusions

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



Comic character extraction

Background

- Challenges

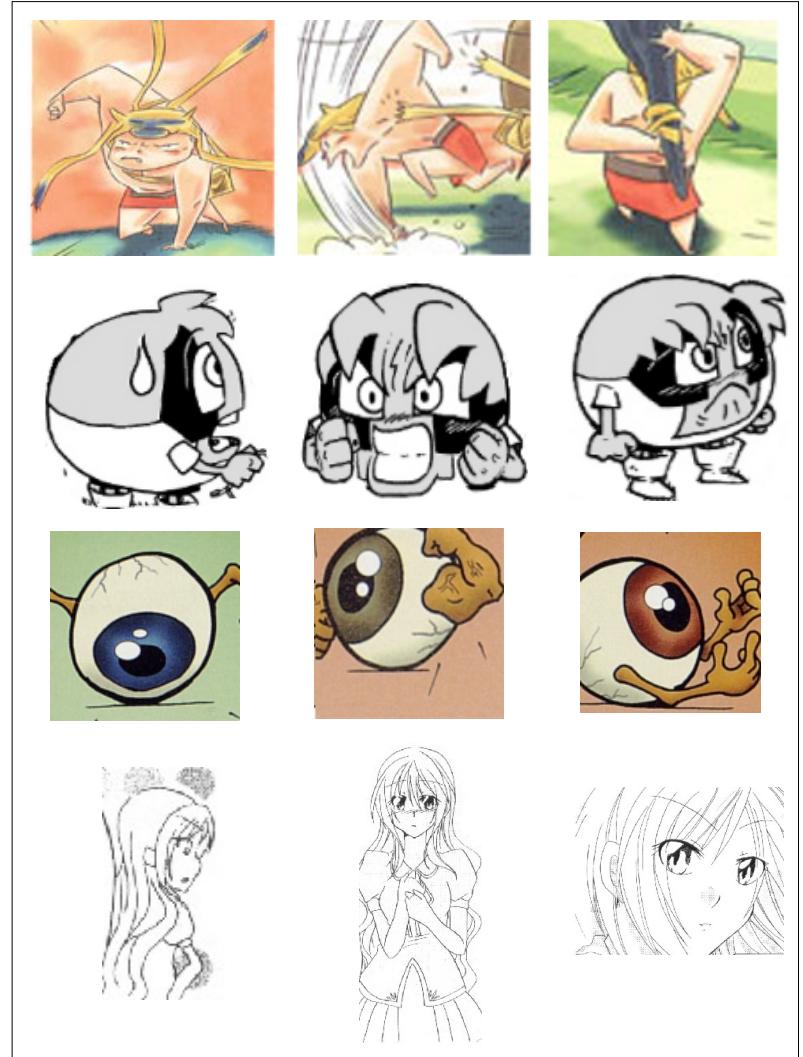
- Hand-drawn, stroke-based
- Intra/inter class variability
- Scale, deformation, posture, occlusion

- Extraction & recognition

- Manga faces [Cheung08, Sun10, Kohei12]
- Cartoons [Khan12]

- Conclusions

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



Overall progress

Background

Element	Process type	Status
Panel	Localisation	
	Classification	
Balloon	Localisation	
	Classification	
	Tail detection	
Text	Localisation	
	Recognition	
Comic character	Localisation	
	Identification	
	Face/pose	
Context	Inter-element link	
	Situation retrieval	
	Timestamps	
Dataset	Localisation	
	Semantic	

-  Solved
-  Advanced
-  Medium
-  Early stage
-  Unexplored

- Introduction
- Sequential approach
- (Independent approach)
- Knowledge-driven approach



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

- **Introduction**
- Sequential approach
- (Independent approach)
- Knowledge-driven approach



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

Introduction

- Objective: cover the widest possible scope of study

Contributions

Introduction

- Objective: cover the widest possible scope of study

1) Creation of heterogeneous dataset

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

Contributions

Bibliographic annotations



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

Visual and semantic annotations

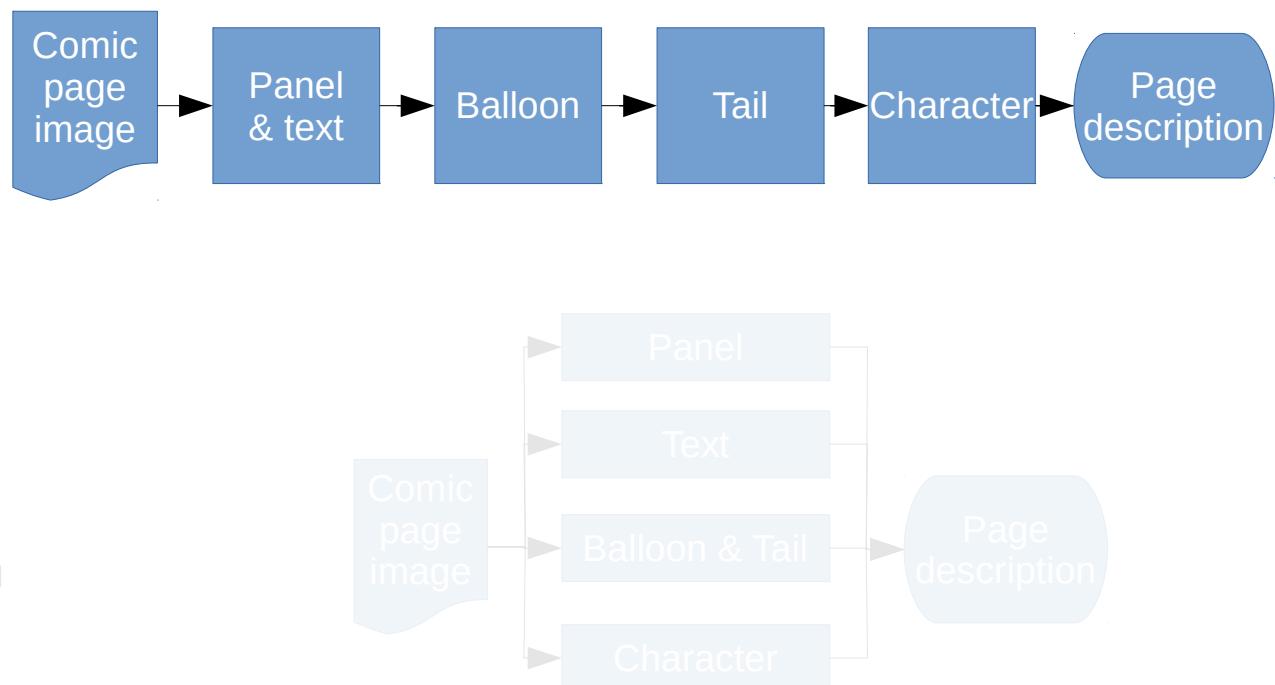


- PANEL**
Rank: 1
- BALLOON**
Rank: 2
Shape: Oval
Tail direction: South-West
- TEXT LINE**
Text: « STARK RAVING »
- CHARACTER**
LinkedToBalloon: 2

Introduction

- Objective: cover the widest possible scope of study
 - 1) Creation of heterogeneous dataset
 - 2) Three approaches
- Content-driven
 - Sequential approach
 - Similar to literature
 - Intuitive
 - Sensible to error propagation
 - Independent approach
 - Avoid error propagation
- Knowledge-driven
 - Knowledge-driven approach
 - Based on domain knowledge
 - Retrieve context

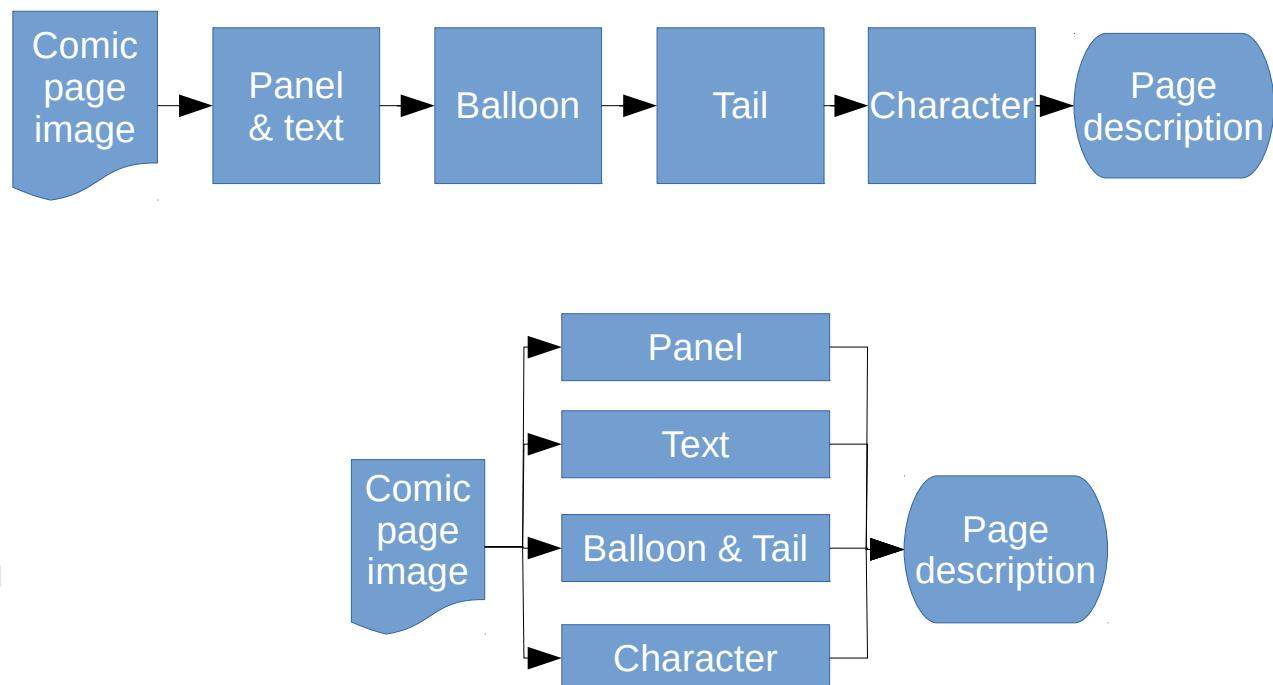
Contributions



Introduction

- Objective: cover the widest possible scope of study
 - 1) Creation of heterogeneous dataset
 - 2) Three approaches
- Content-driven
 - Sequential approach
 - Similar to literature
 - Intuitive
 - Sensible to error propagation
 - Independent approach
 - Avoid error propagation
- Knowledge-driven
 - Knowledge-driven approach
 - Based on domain knowledge
 - Retrieve context

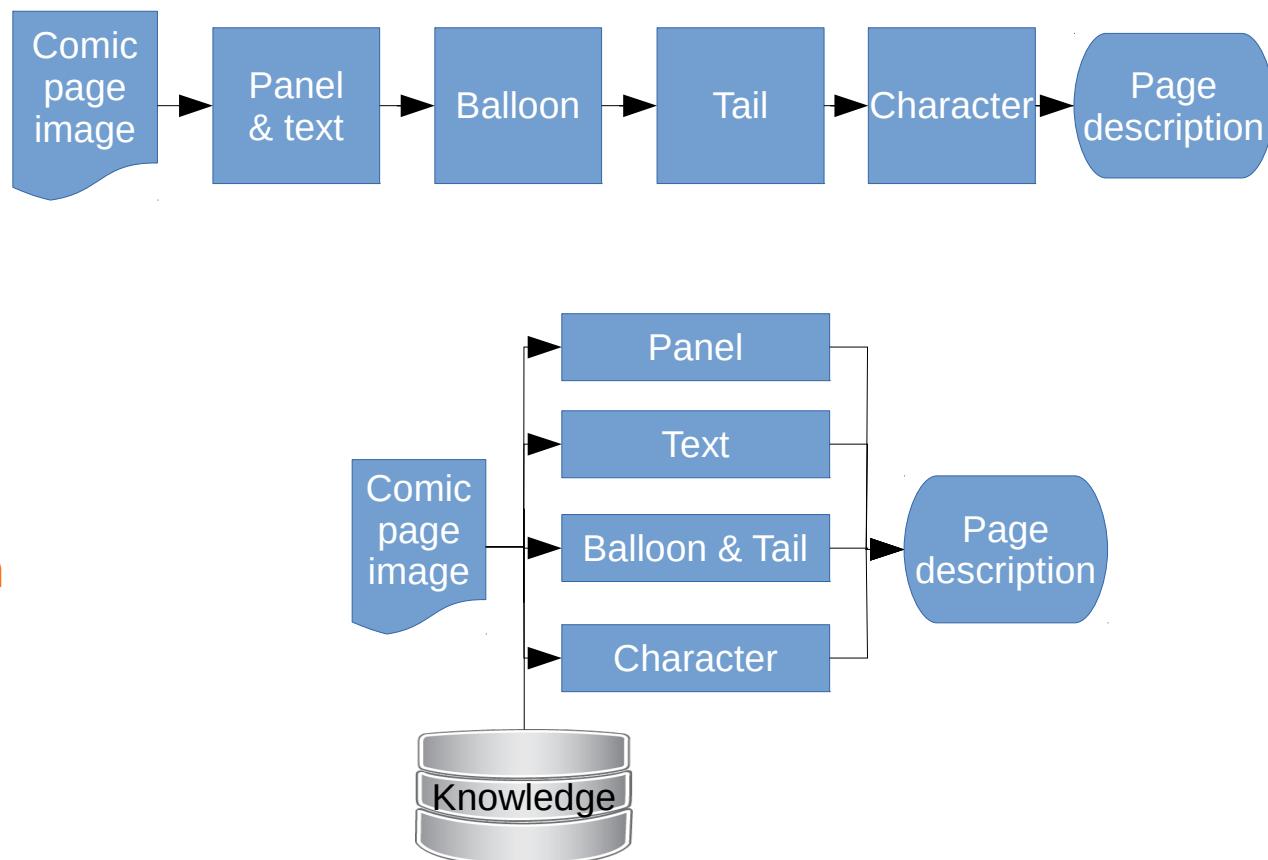
Contributions



Introduction

- Objective: cover the widest possible scope of study
 - 1) Creation of heterogeneous dataset
 - 2) Three approaches
- Content-driven
 - Sequential approach
 - Similar to literature
 - Intuitive
 - Sensible to error propagation
 - Independent approach
 - Avoid error propagation
- Knowledge-driven
 - Knowledge-driven approach
 - Based on domain knowledge
 - Retrieve context

Contributions



- Introduction
- Sequential approach
- (Independent approach)
- Knowledge-driven approach



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

- Introduction
- Sequential approach
 - Panel & text extraction
 - Balloon extraction
 - Tail extraction
 - Comic character extraction
- (Independent approach)
- Knowledge-driven approach



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

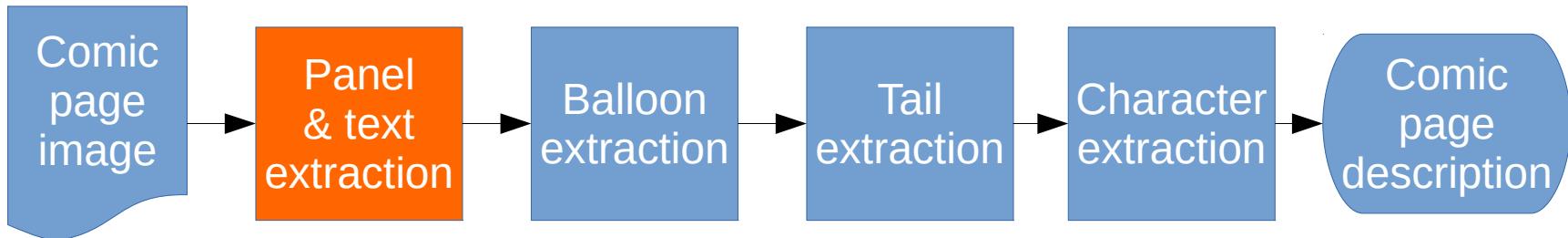
- Introduction
- Sequential approach
 - Panel & text extraction
 - Balloon extraction
 - Tail extraction
 - Comic character extraction
- (Independent approach)
- Knowledge-driven approach



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

Panel & text extraction

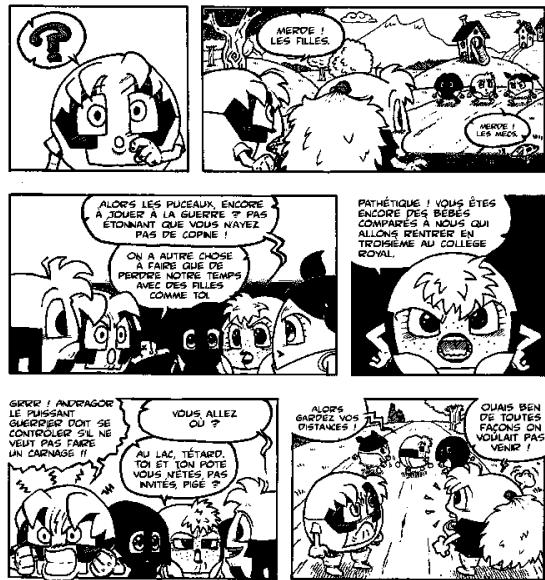
Contributions
Sequential approach



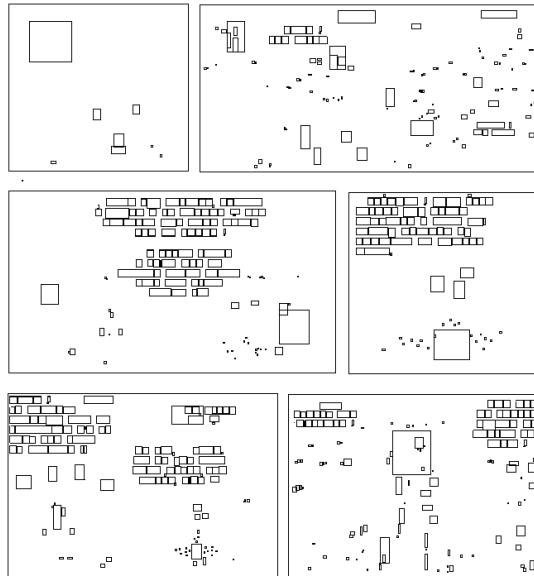
- Literature
 - Panel with frame, separated by **gutters** or black **line**
 - Text located **inside** balloons
- Contribution
 - **Simultaneous panel and text** extraction from a binary image
 - Consider **implicit** and **non-rectangle** panels
 - **Location-independent** text extraction

Panel & text extraction

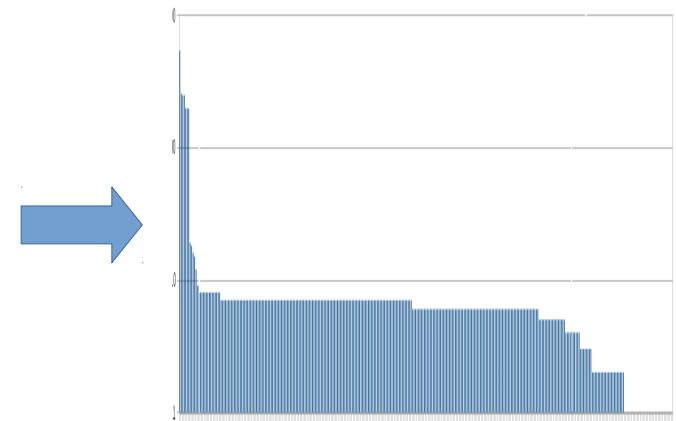
Contributions Sequential approach



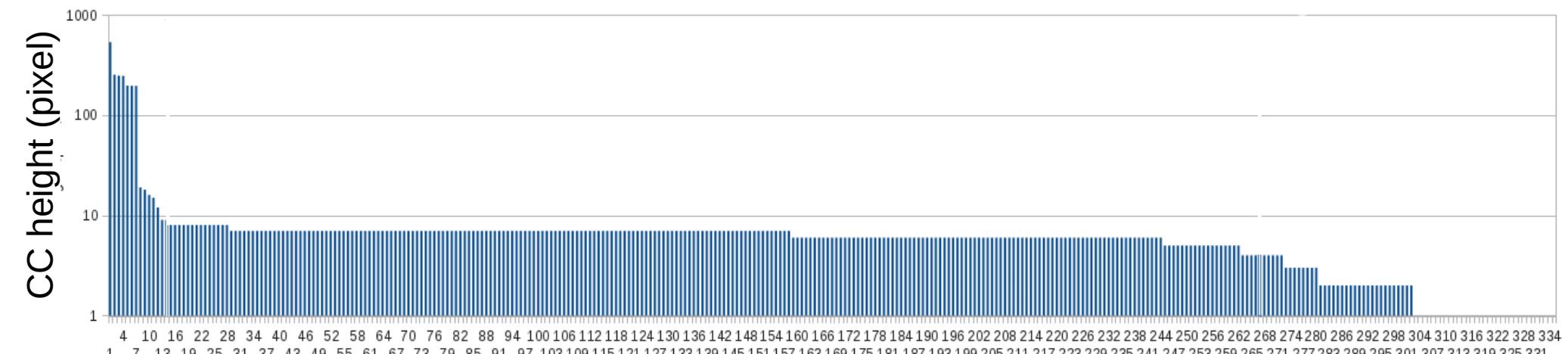
1) Binary image



2) Black connected-component (CC) bounding boxes



3) Histogram of heights of CC

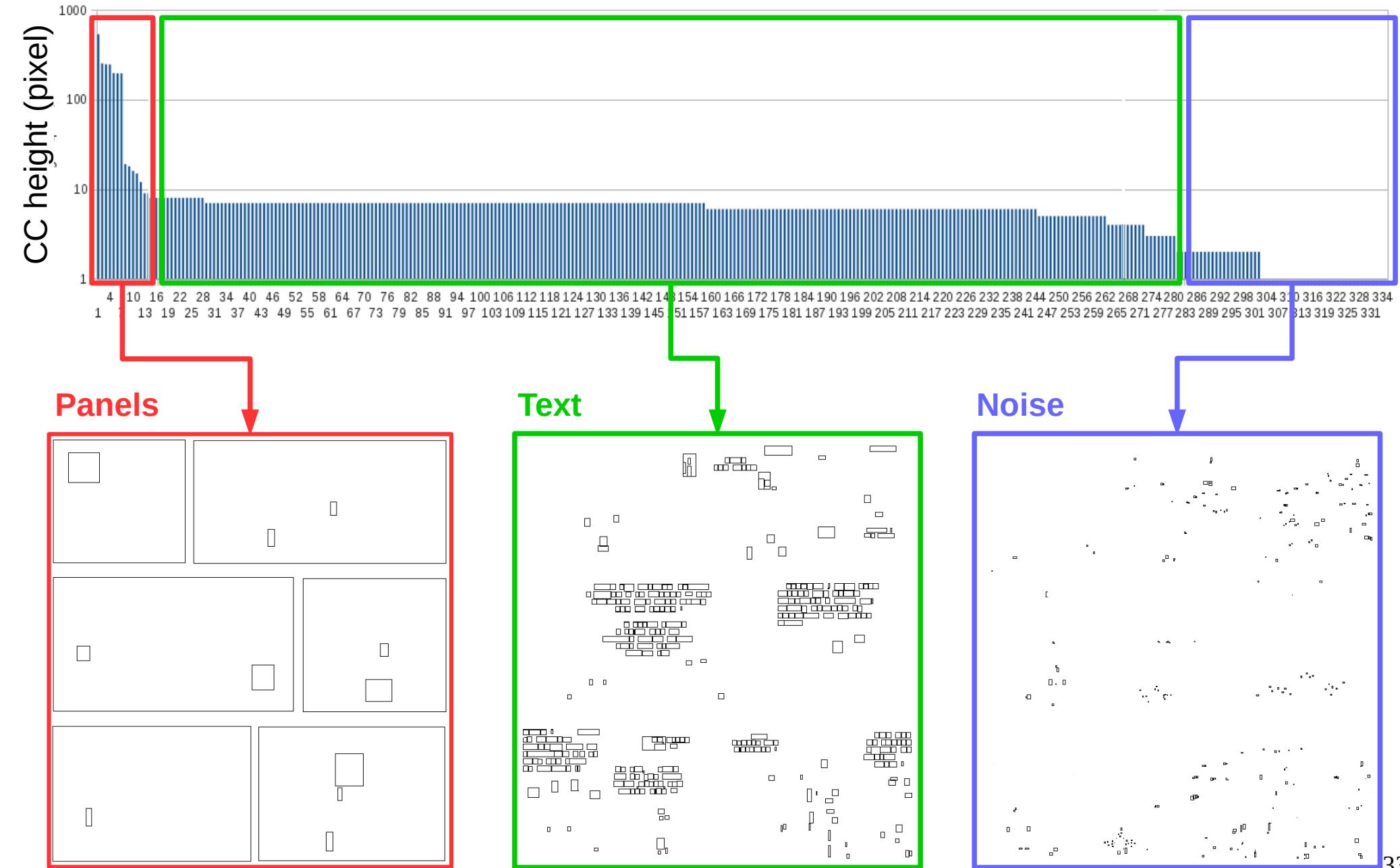


4) Histogram of heights (descendent order)

Panel & text extraction

Contributions Sequential approach

K-means clustering ($k=3$)

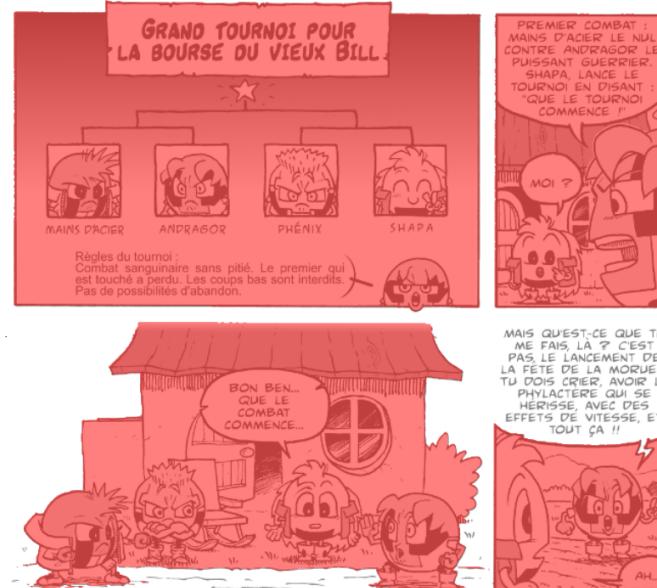
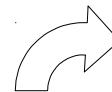


Panel & text extraction: results

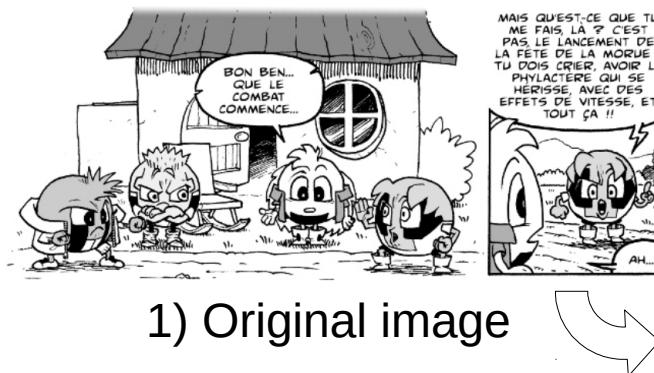
Contributions Sequential approach

Panel

Text



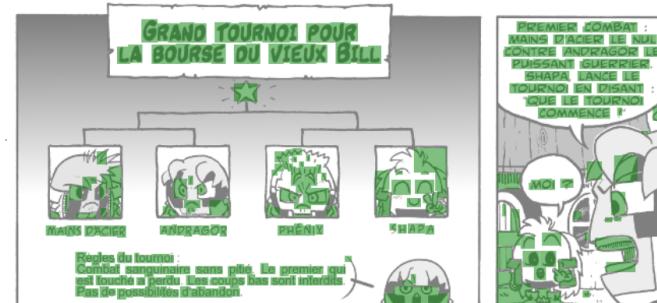
2) CC regions



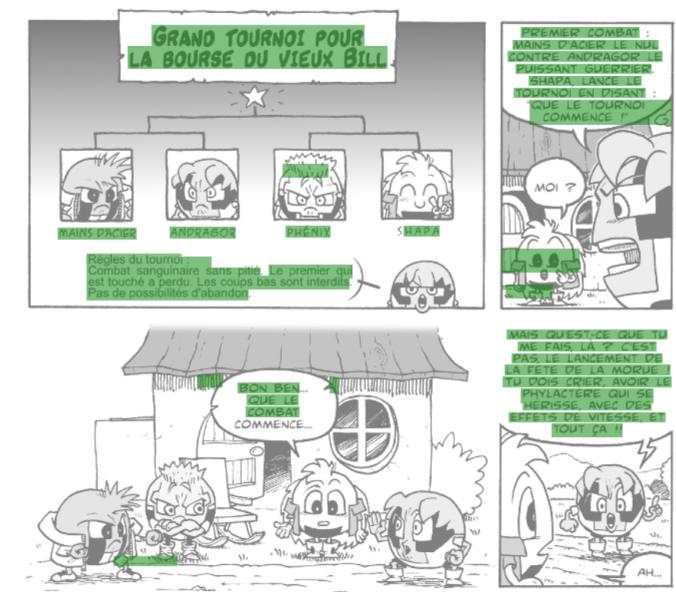
3) Panel boxes

eBDtheque dataset

	Recall	Precision
Panel	64.24 %	83.81 %
Text	54.91 %	57.15 %



4) CC region boxes



5) Text line boxes

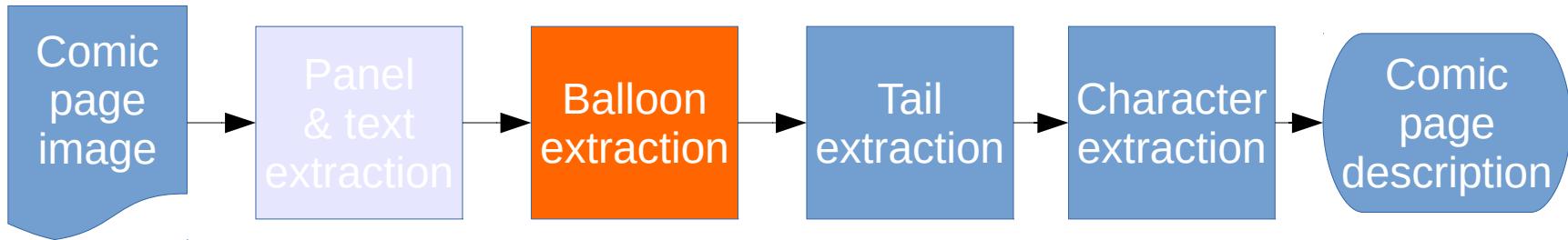
- Introduction
- Sequential approach
 - Panel & text extraction
 - **Balloon extraction**
 - Tail extraction
 - Comic character extraction
- (Independent approach)
- Knowledge-driven approach



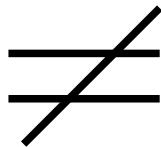
Inking. Image credits: Le cycle des bulles,
Christophe Rigaud, 2012

Balloon extraction

Contributions
Sequential approach



Regular balloon

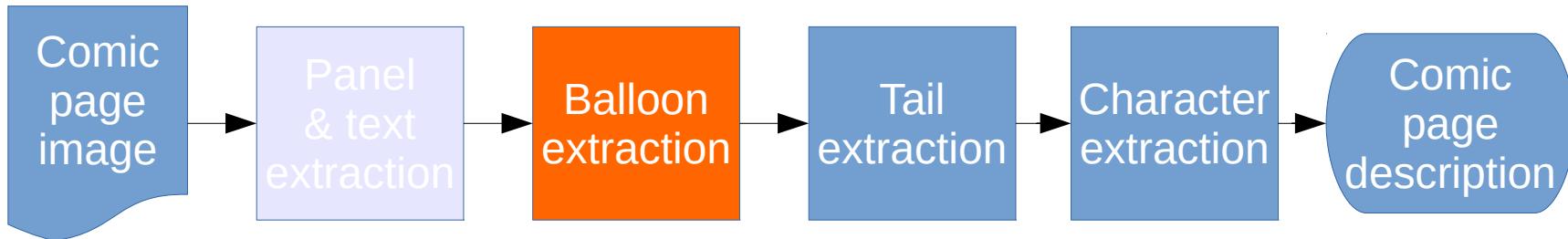


Implicit balloon

- 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
Sequential 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

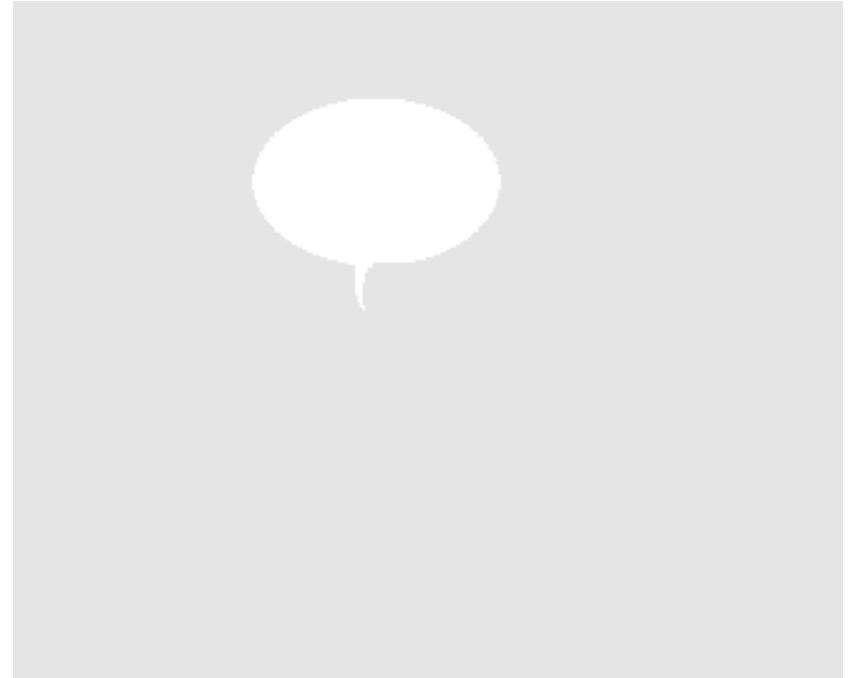
Balloon extraction: regular

Contributions
Sequential approach

- Assumptions
 - Panels and text block positions are known
 - Regular balloons contain centred text
- Proposition → structural analysis
 - Extract closed contours that fully include centred text



Original image



Expected result

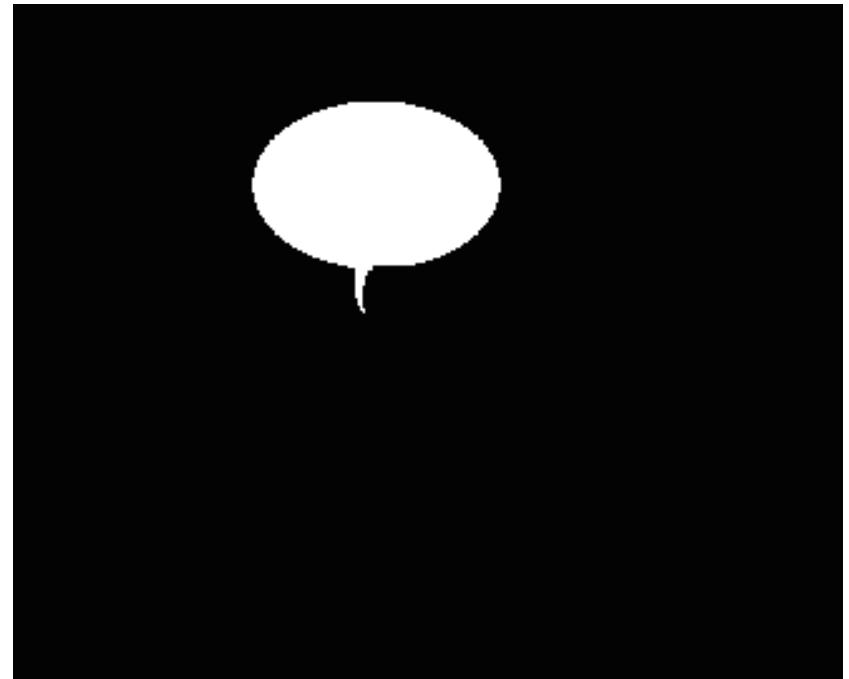
Balloon extraction: regular

Contributions
Sequential approach

- Assumptions
 - Panels and text block positions are known
 - Regular balloons contain centred text
- Proposition → structural analysis
 - Extract closed contours that fully include centred text



1) Original image



2) Expected result

Balloon extraction: regular

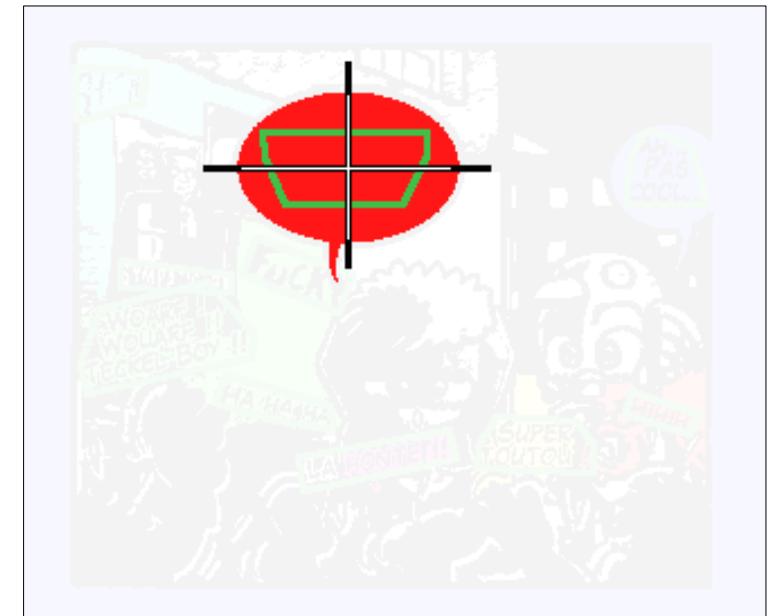
Contributions
Sequential approach



1) Original image



2) Text block positions (green)



3) Regions including text blocks (coloured)
4) Regions including aligned text blocks 40

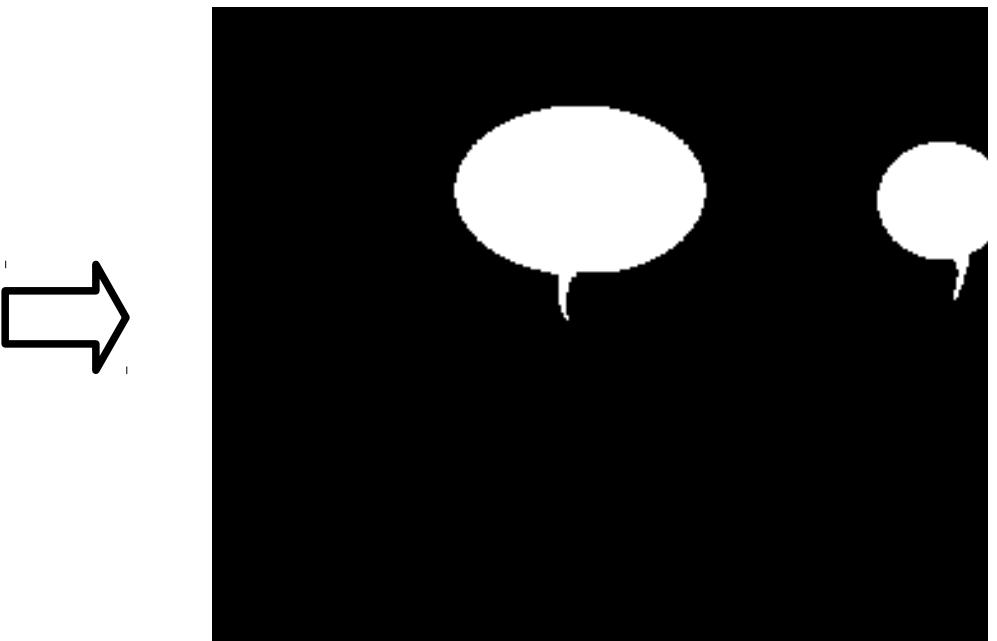
Balloon extraction: implicit

Contributions
Sequential 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)



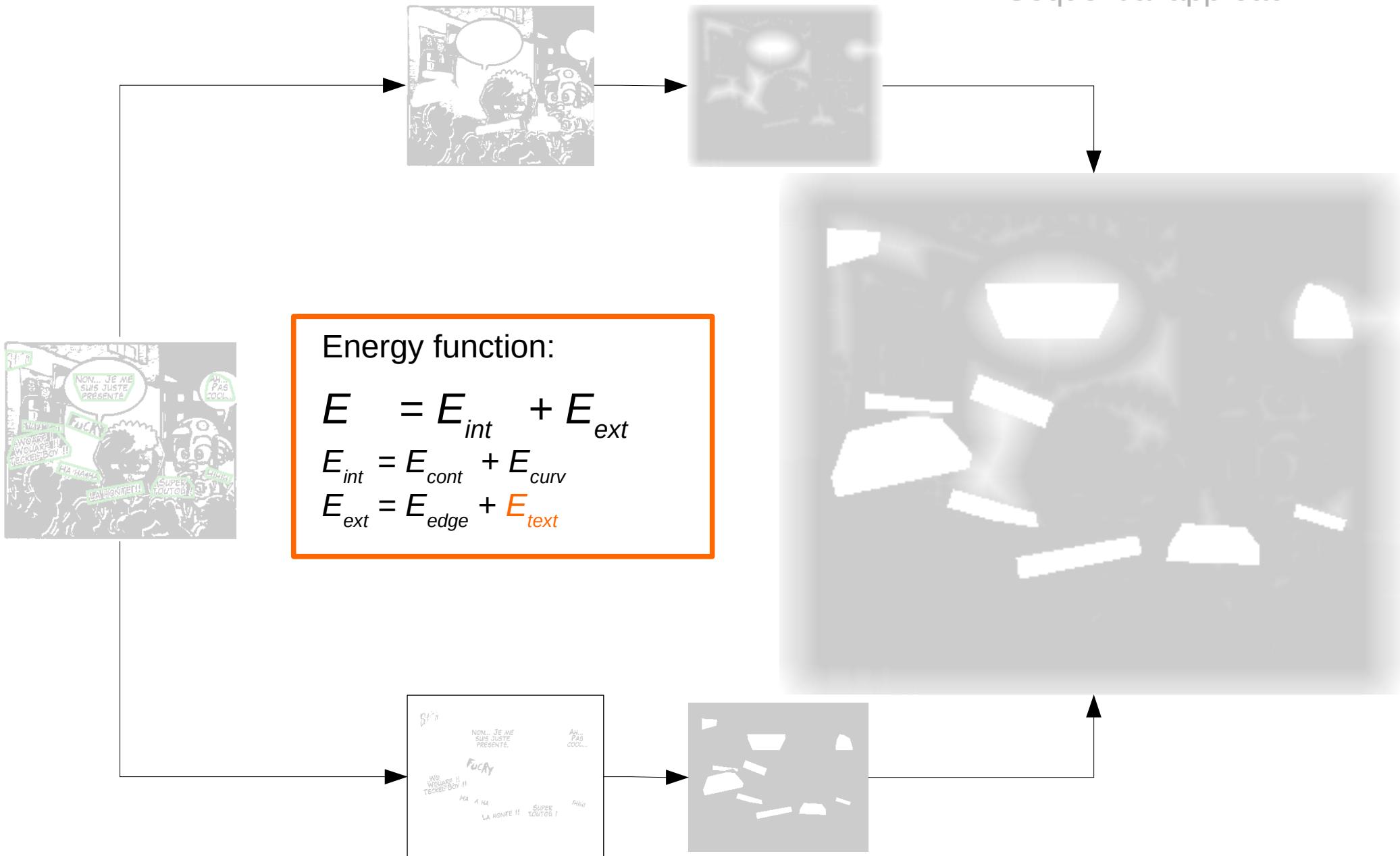
1) Original image and text locations



2) Expected result

Balloon extraction: implicit

Contributions
Sequential approach



Balloon extraction: implicit

Contributions
Sequential approach

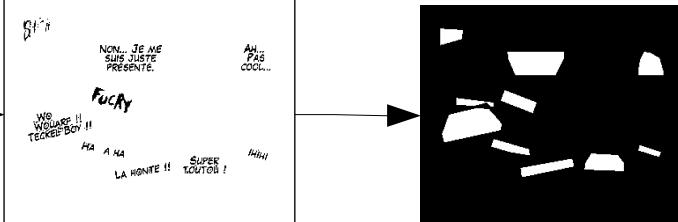
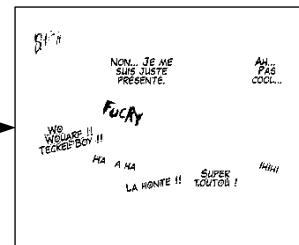


Energy function:

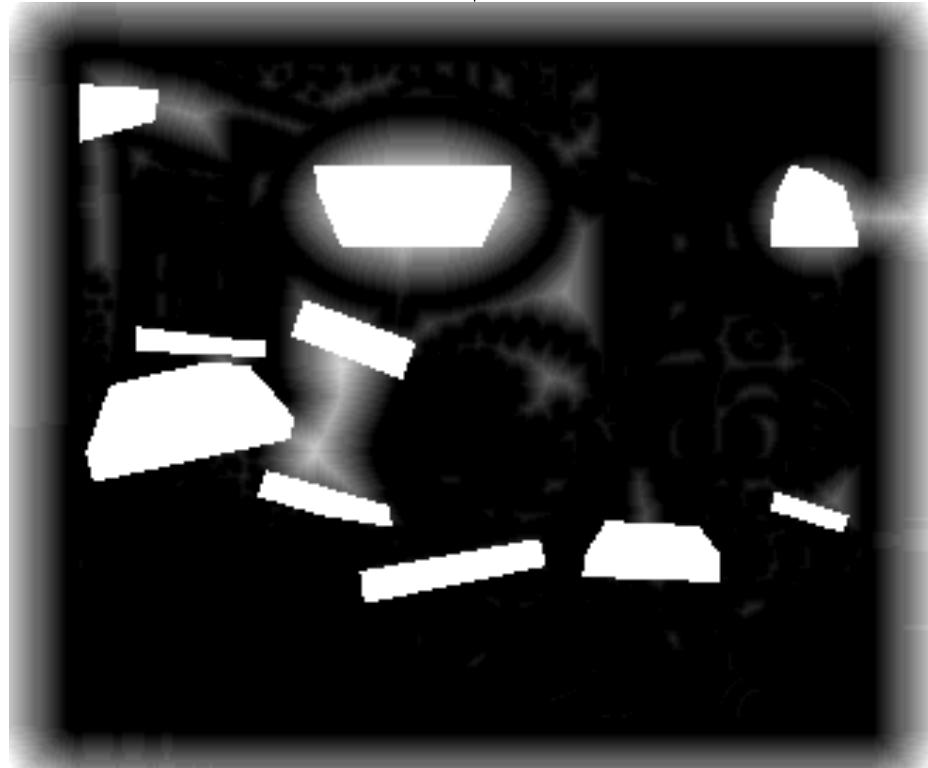
$$E = E_{int} + E_{ext}$$

$$E_{int} = E_{cont} + E_{curv}$$

$$E_{ext} = E_{edge} + E_{text}$$

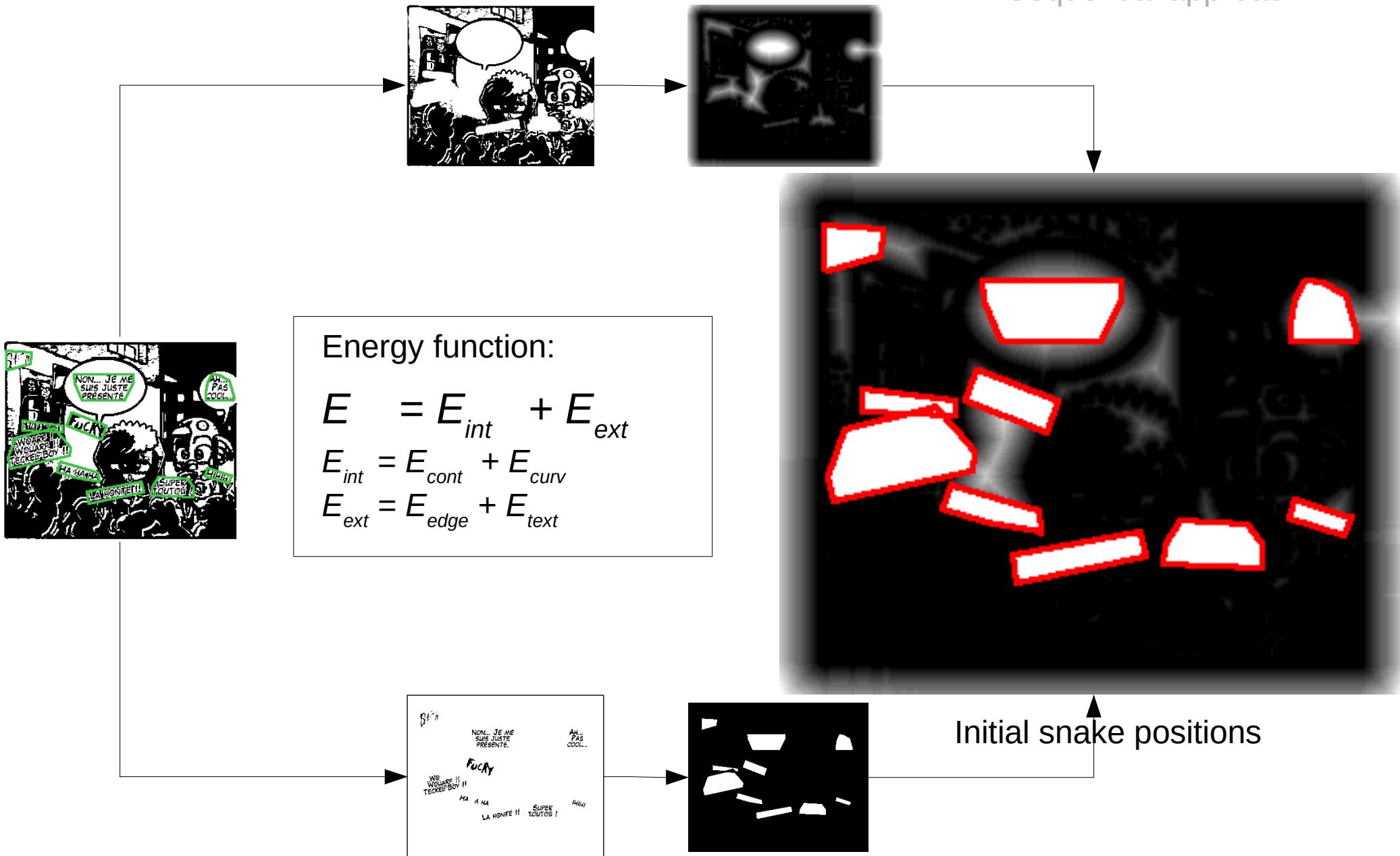


External energy E_{ext}



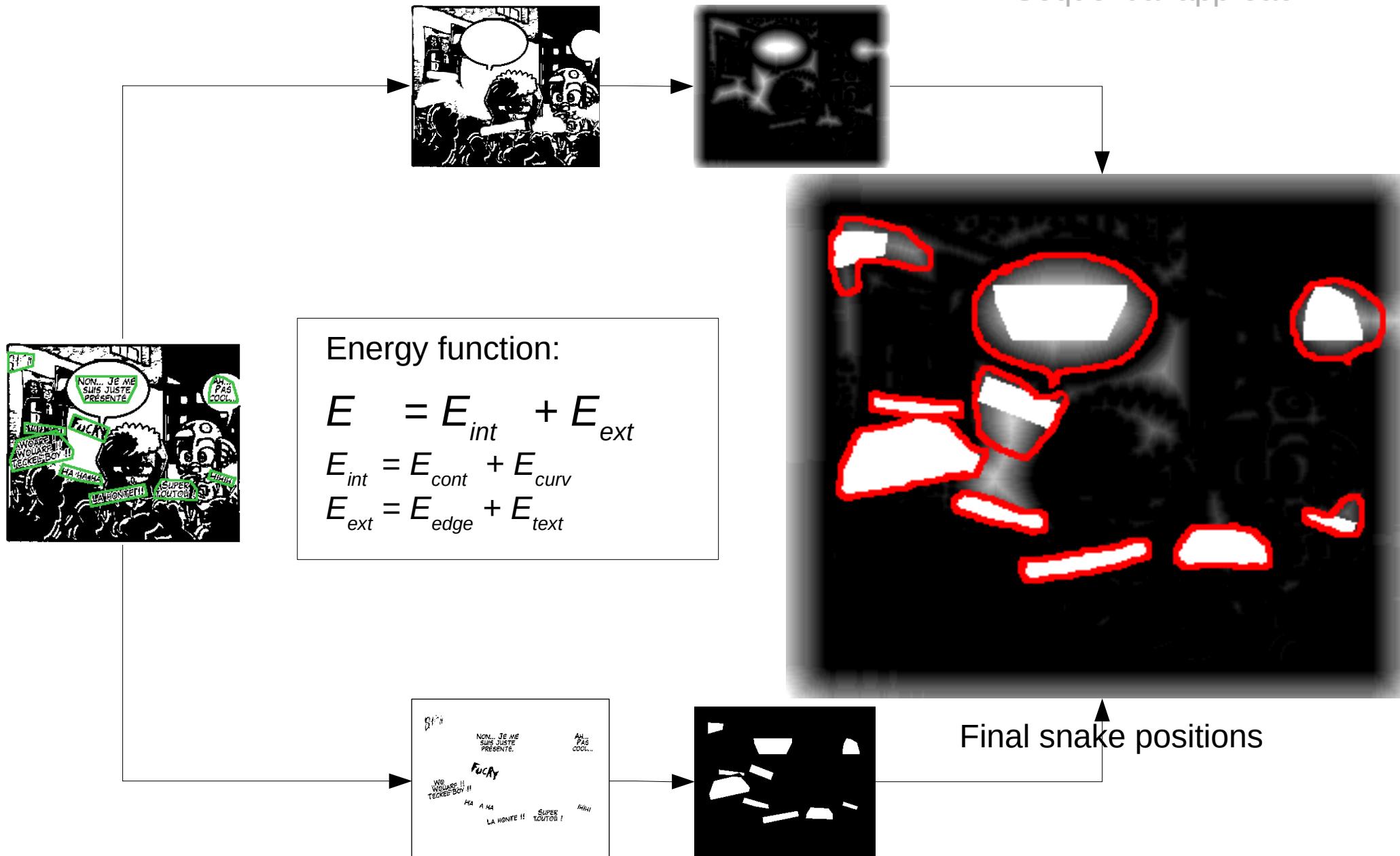
Balloon extraction: implicit

Contributions
Sequential approach



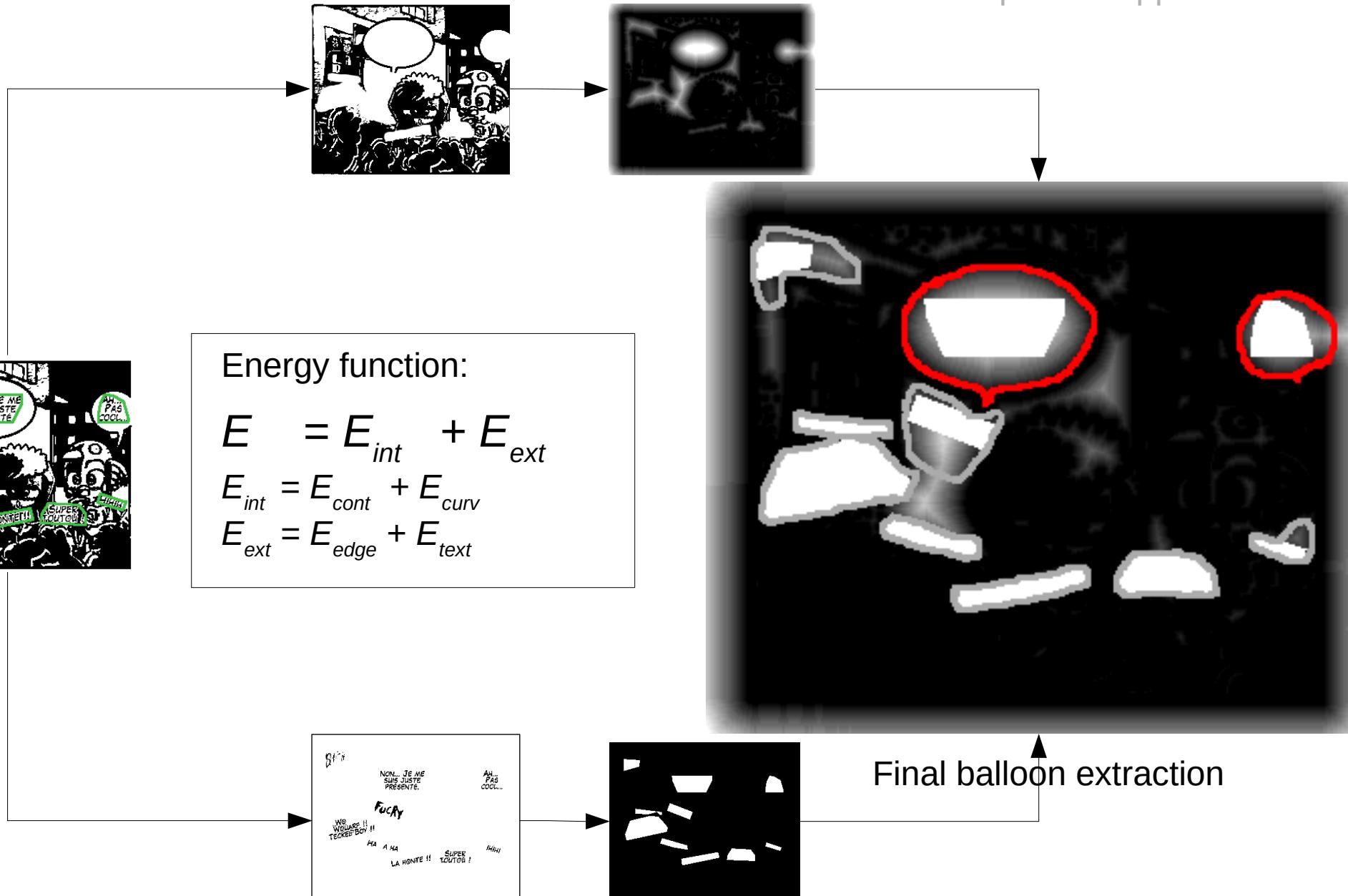
Balloon extraction: implicit

Contributions
Sequential approach



Balloon extraction: implicit

Contributions
Sequential approach



Balloon extraction: results

Contributions Sequential approach

— Text block
— Balloon extraction



Regular balloons



Implicit balloons



eBDtheque dataset

	Recall	Precision
Regular	37.25 %	45.19 %
Implicit	57.76 %	41.62 %



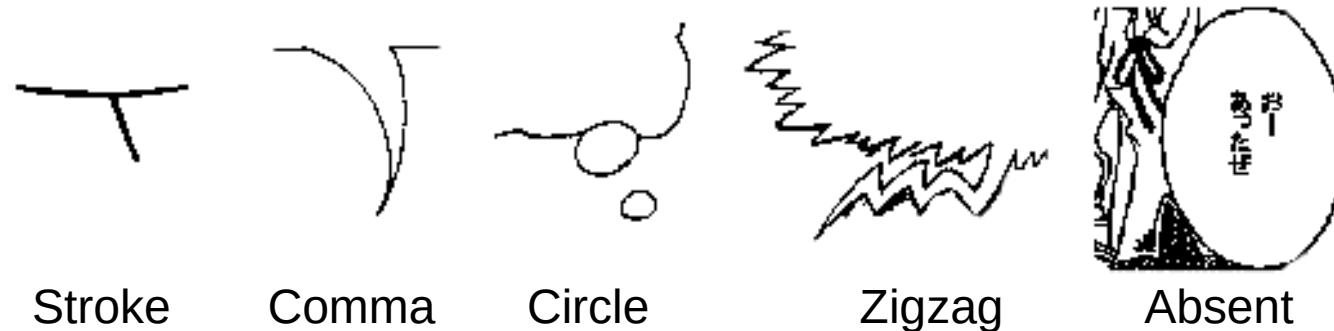
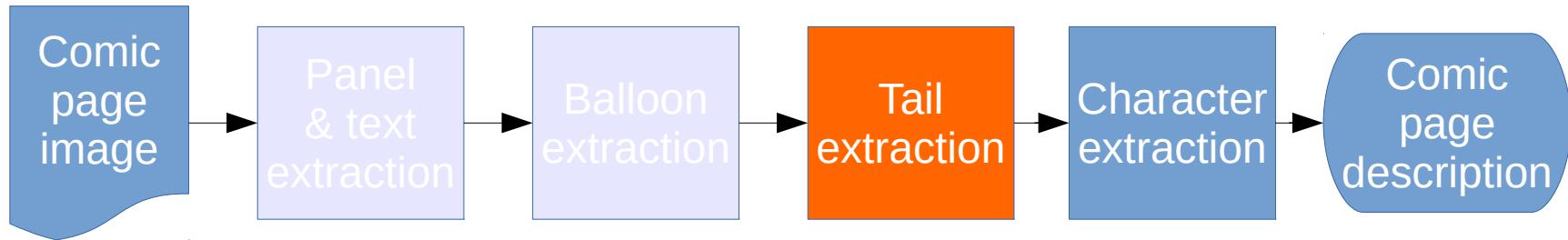
- Introduction
- Sequential approach
 - Panel & text extraction
 - Balloon extraction
 - Tail extraction
 - Comic character extraction
- (Independent approach)
- Knowledge-driven approach



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

Tail extraction

Contributions
Sequential approach



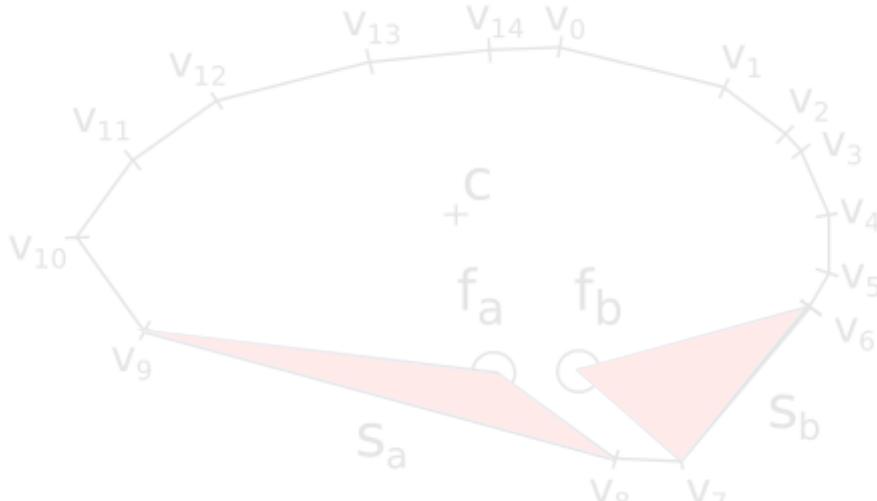
- Literature
 - First time studied in document image analysis
- Objectives
 - Extraction of tail tip position and direction
 - Focus on comma, zigzag and absent types

Tail extraction: tip position

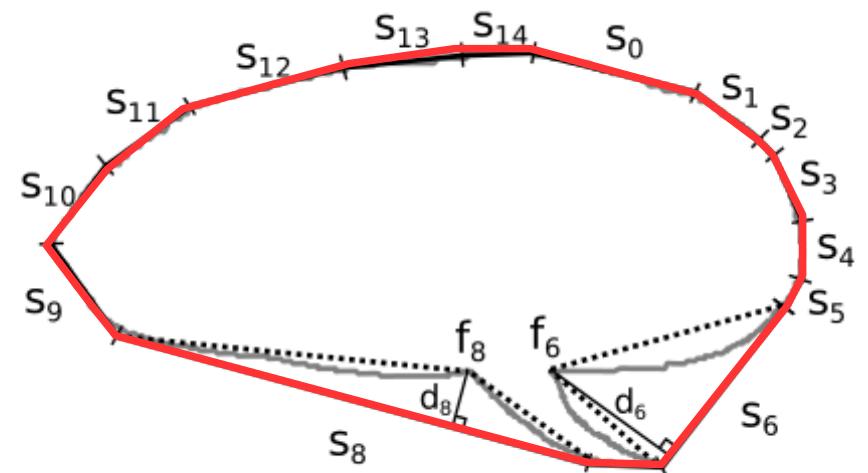
Contributions Sequential approach



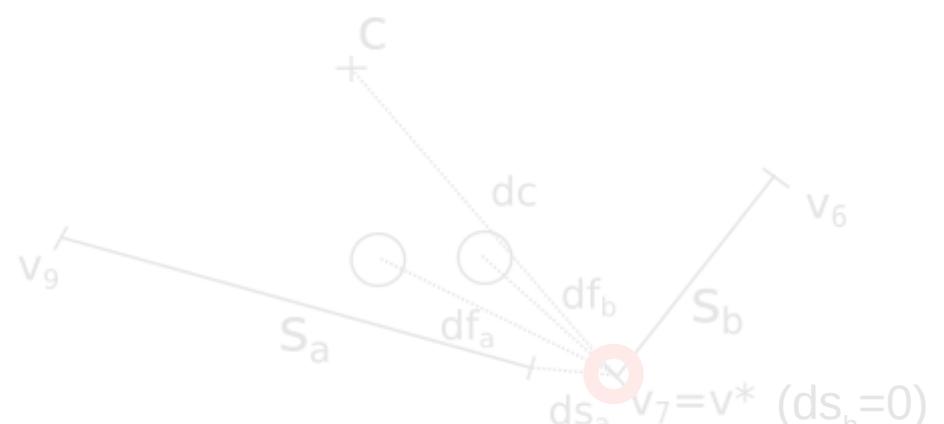
1) Balloon contour



3) Two biggest
convexity defects



2) Convex hull



4) Tail tip position

Optimal vertex selection:

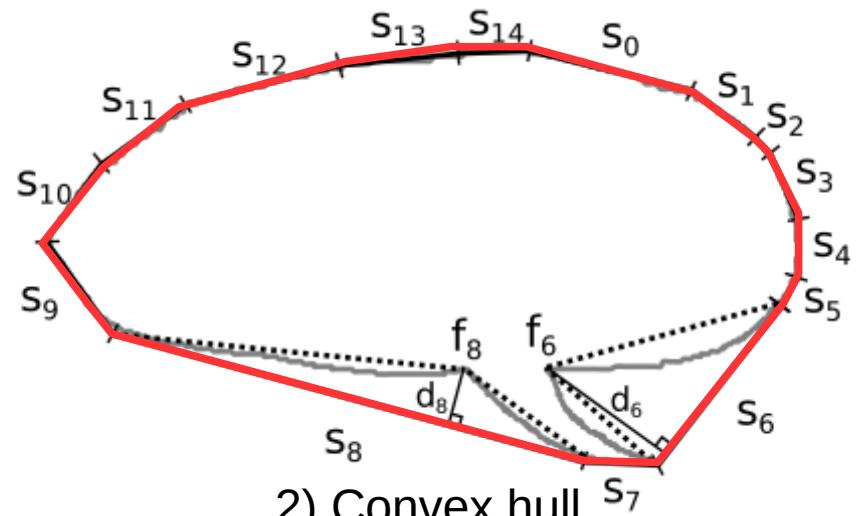
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

Tail extraction: tip position

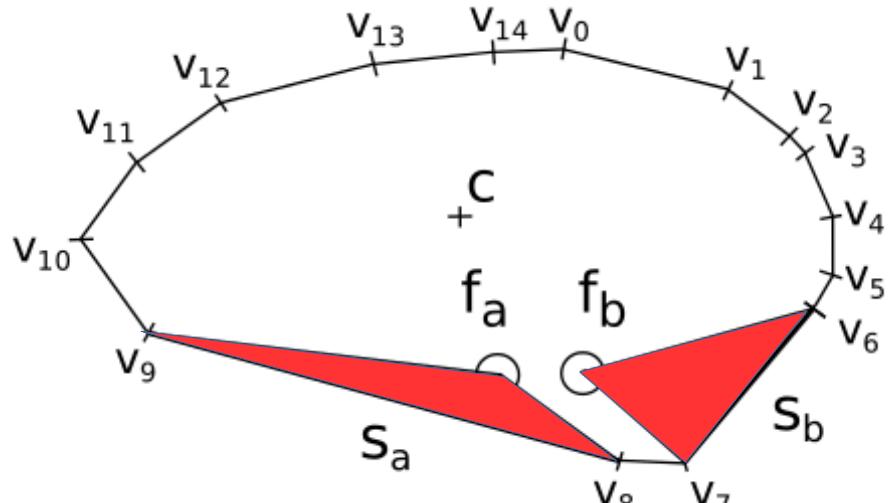
Contributions
Sequential approach



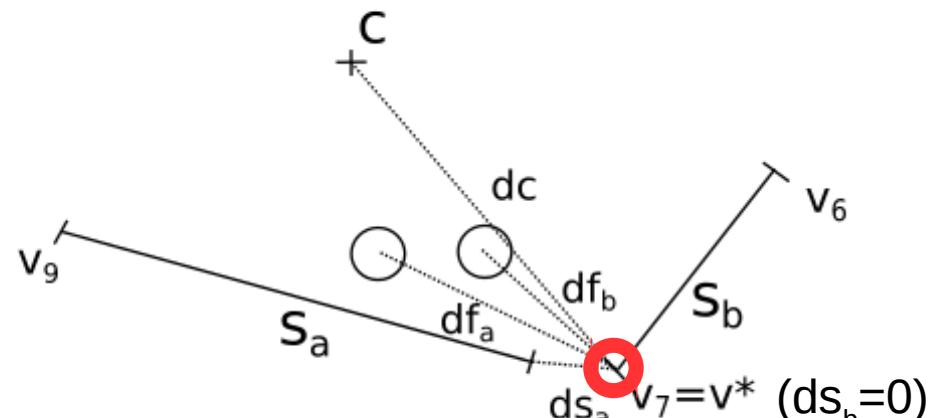
1) Balloon contour



2) Convex hull



3) Two biggest
convexity defects



4) Tail tip position

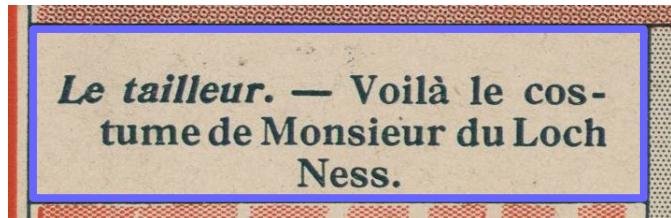
Optimal vertex selection:

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

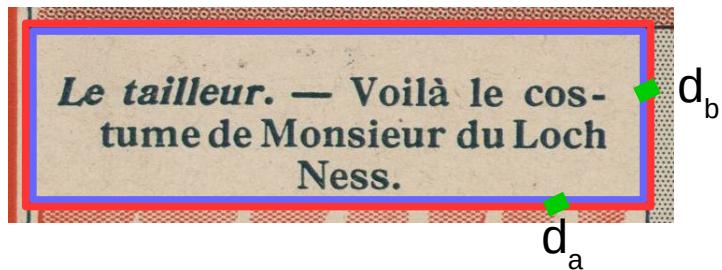
Tail extraction: confidence value

Balloon
contour (blue)

Balloon 1



Convex hull
(red)



Confidence

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

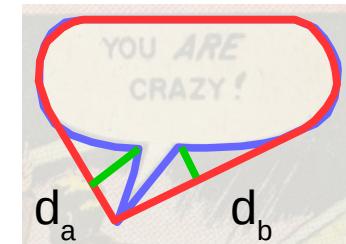
$$C_{tail} = 0.0$$

Presence of tail

NO

Contributions
Sequential approach

Balloon 2



$$C_{tail} = 0.73$$

YES (>0)

Tail extraction: tail direction

Contributions
Sequential 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)



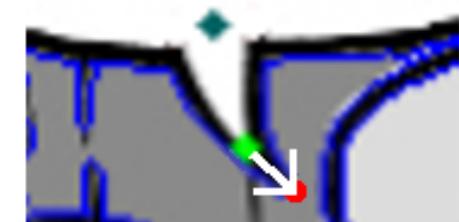
1) Background tail tip (green) and external edge (blue)



2) Closest point on external edge (red)



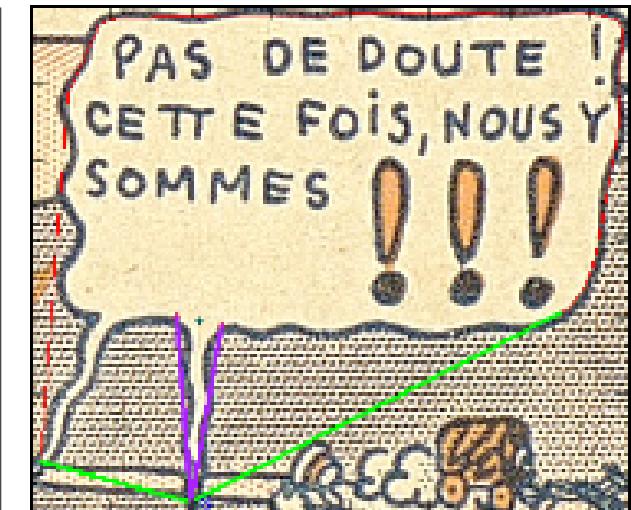
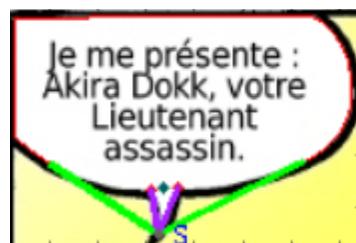
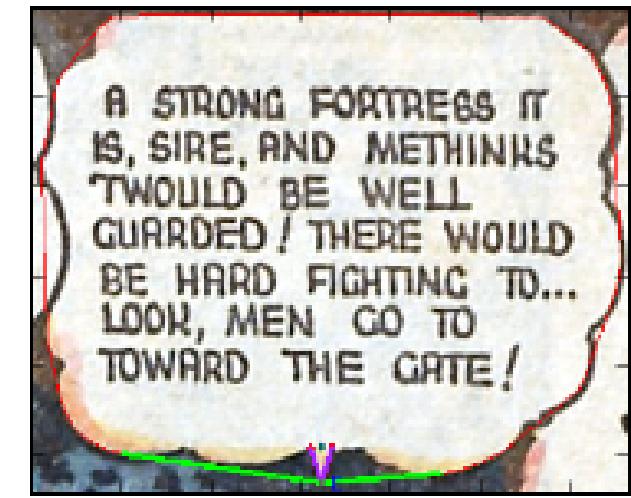
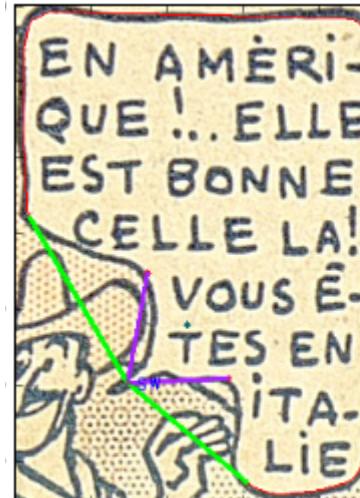
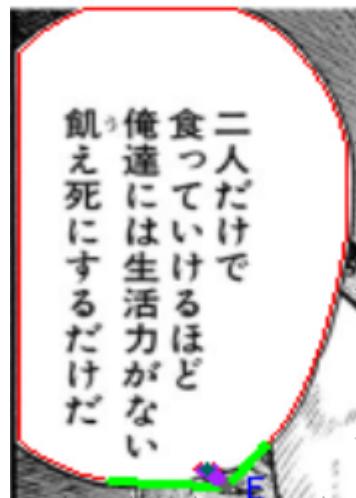
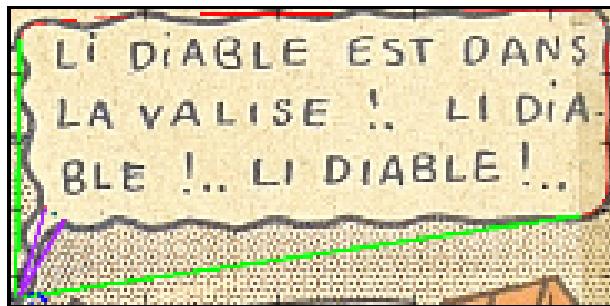
3) Farthest point from origin and tip (red)



4) Direction from tip to farthest point (white arrow)

Tail extraction: results

Experiments



eBDtheque dataset

	Accuracy of tail tip	Accuracy of tail direction
GT balloons	96.61 %	80.40 %
Auto balloons	55.77 %	27.59 %

- Biggest convexity defects
- Tail region
- Balloon extraction

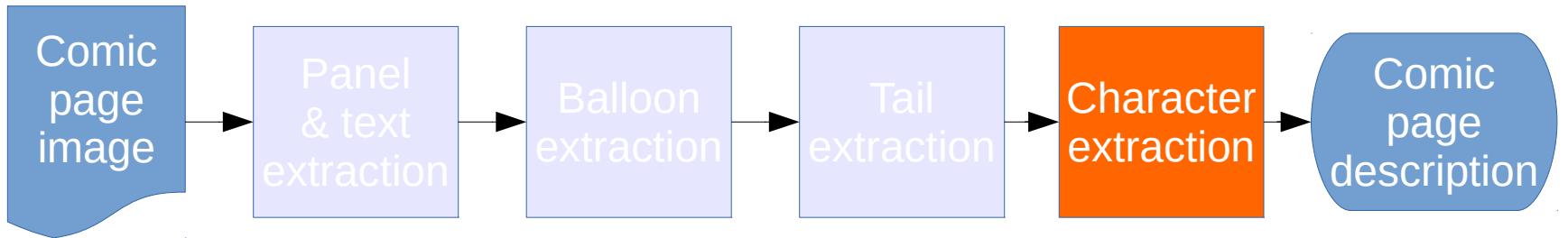
- Introduction
- Sequential approach
 - Panel & text extraction
 - Balloon extraction
 - Tail extraction
 - Comic character extraction
- (Independent approach)
- Knowledge-driven approach



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

Comic character extraction

Contributions
Sequential approach

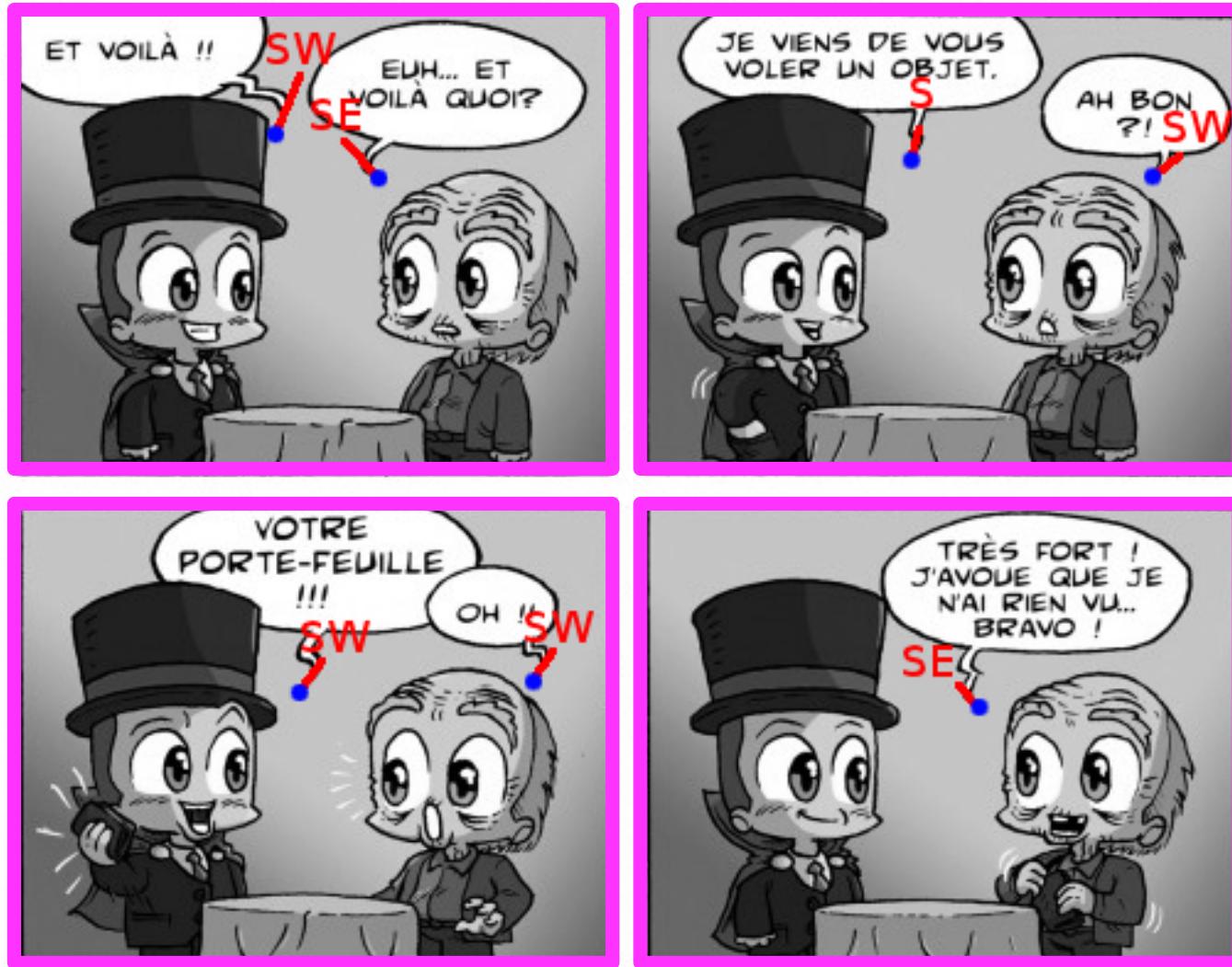


- Challenges
 - Variety of styles of comic books
 - Intra and extra class variations of each character instance (e.g. pose, occlusion, human-like or invented)
- Literature
 - Supervised approaches for manga and cartoon characters
- Contribution
 - Unsupervised and generic approach for all styles of comic books

Comic character extraction

Contributions
Sequential approach

Panels + Tails = ?



Comic character extraction

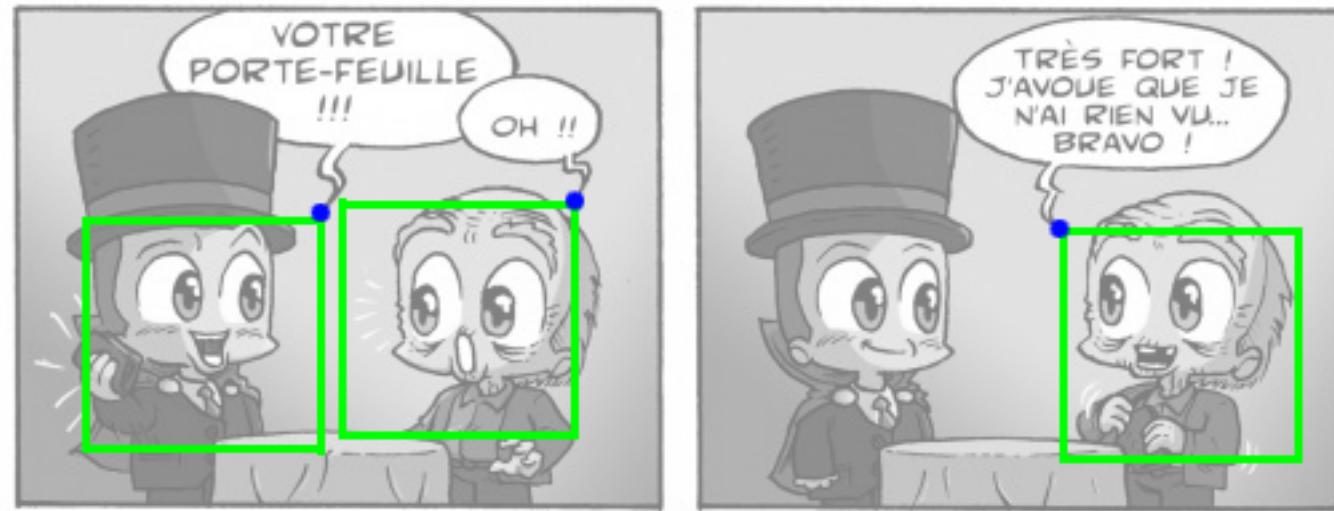
Contributions
Sequential approach

Panels + Tails = Comic character ROIs

Large ROI



Small ROI

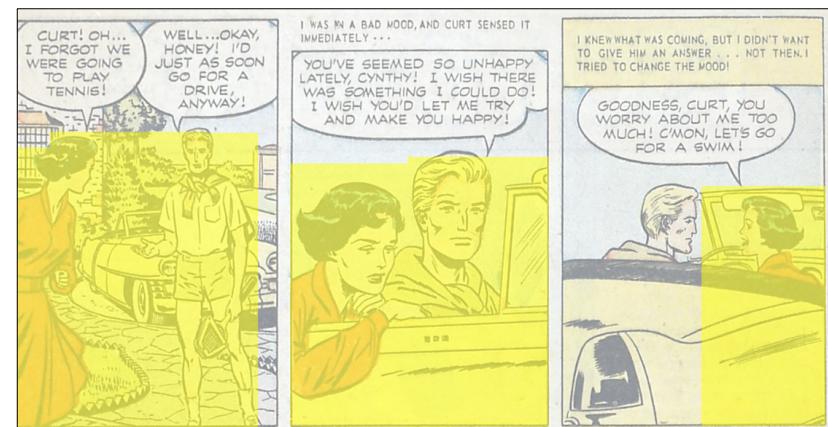
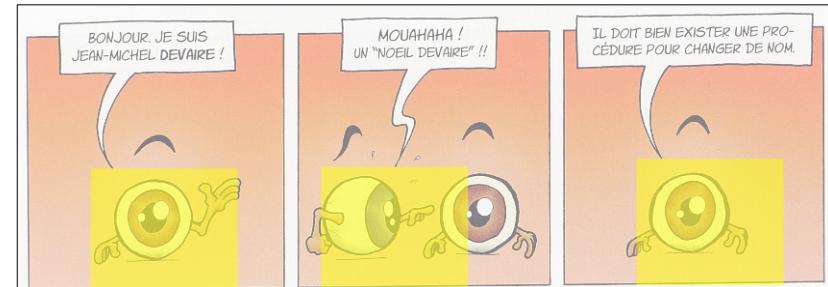
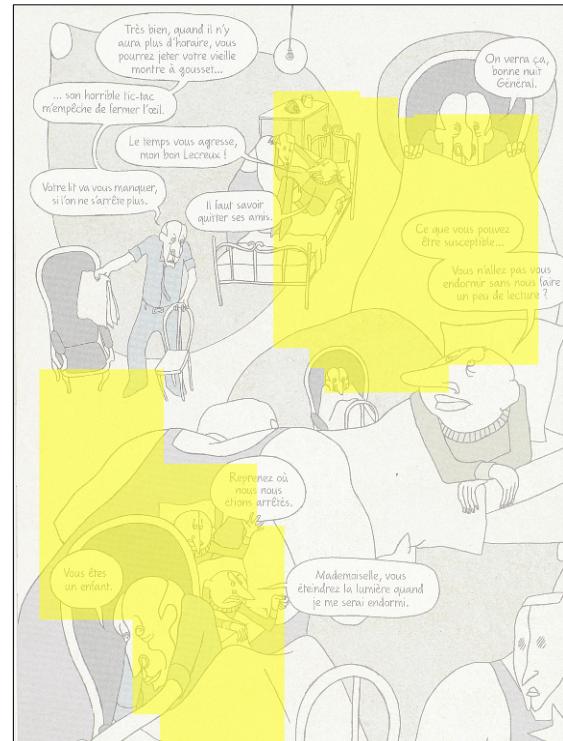
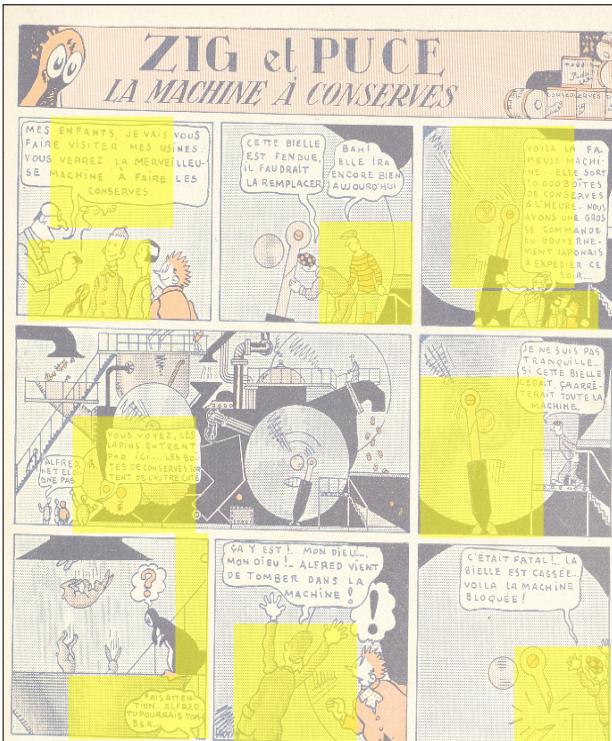


Comic character extraction: results

Contributions

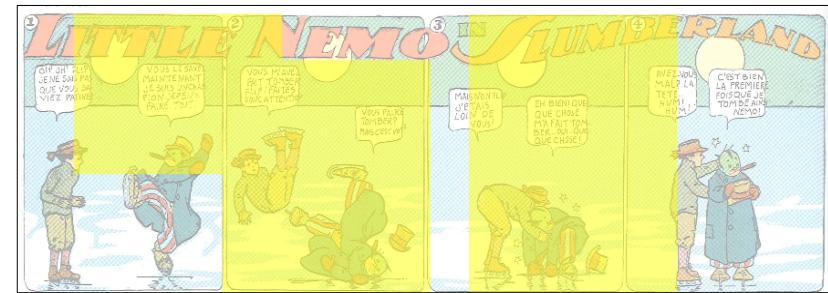
Sequential approach

Comic character ROIs (Regions Of Interests)



eBDtheque dataset

	Recall	Precision
GT (panels + balloons)	15.59 %	23.18 %
Auto	6.84 %	12.13 %



- Introduction
- Sequential approach
- (Independent approach)
- Knowledge-driven approach
 - Introduction
 - Knowledge representation
 - Processing sequence



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

Introduction

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

Contributions

Knowledge-driven approach

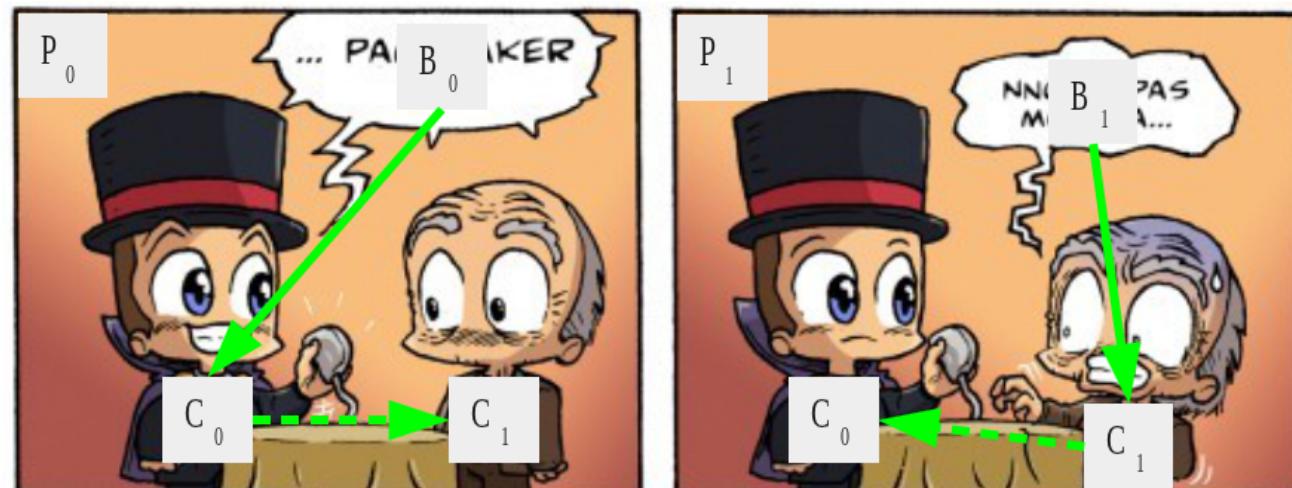
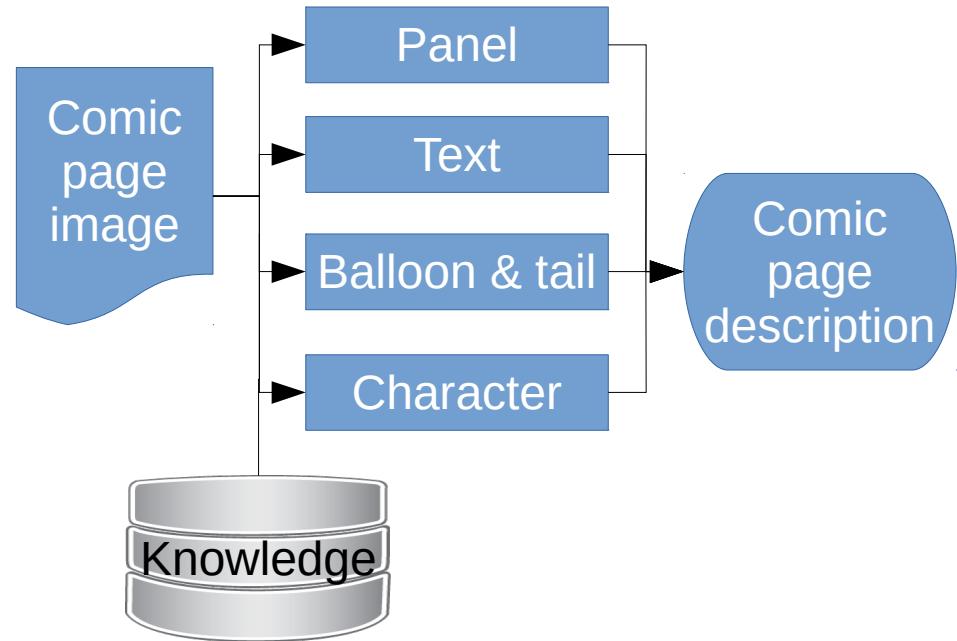
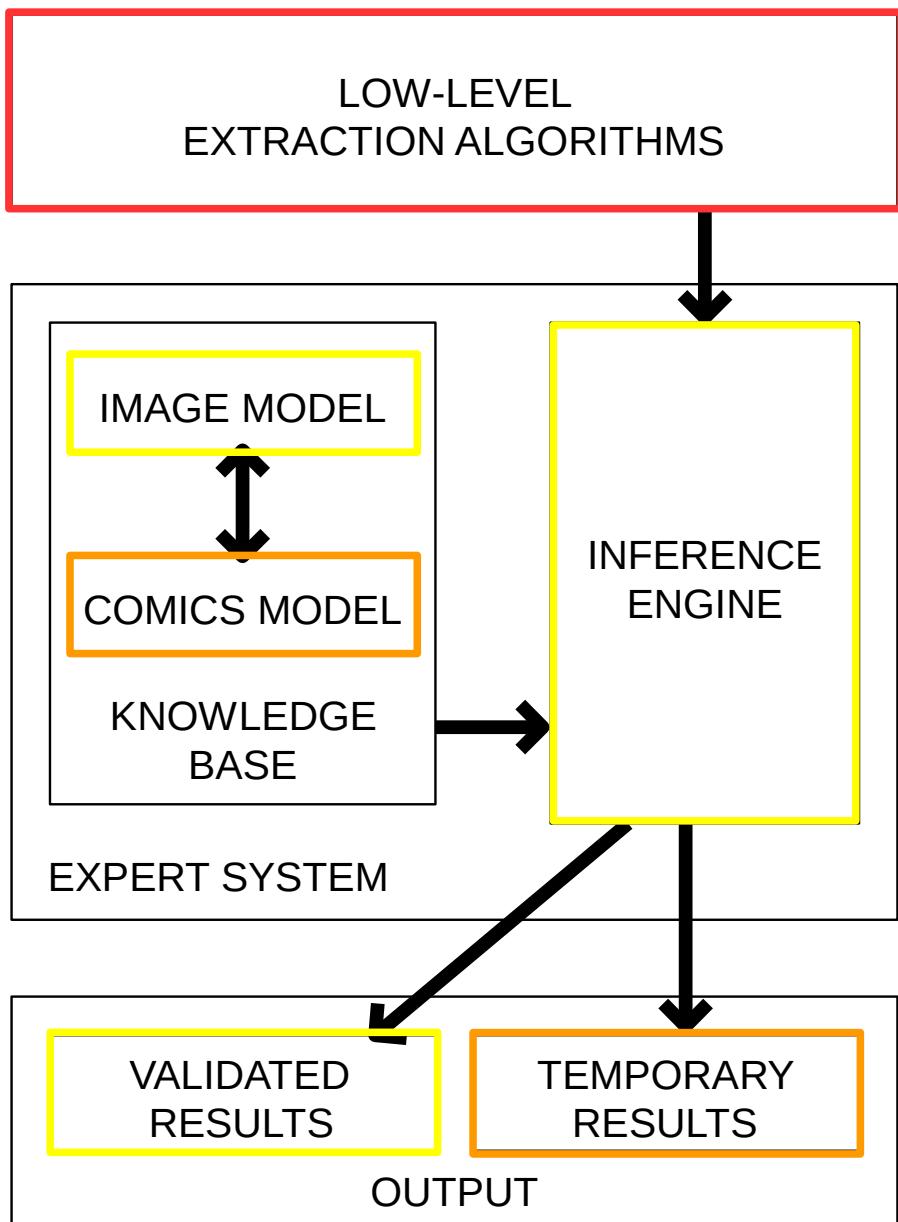


Illustration of high level image description

Knowledge representation



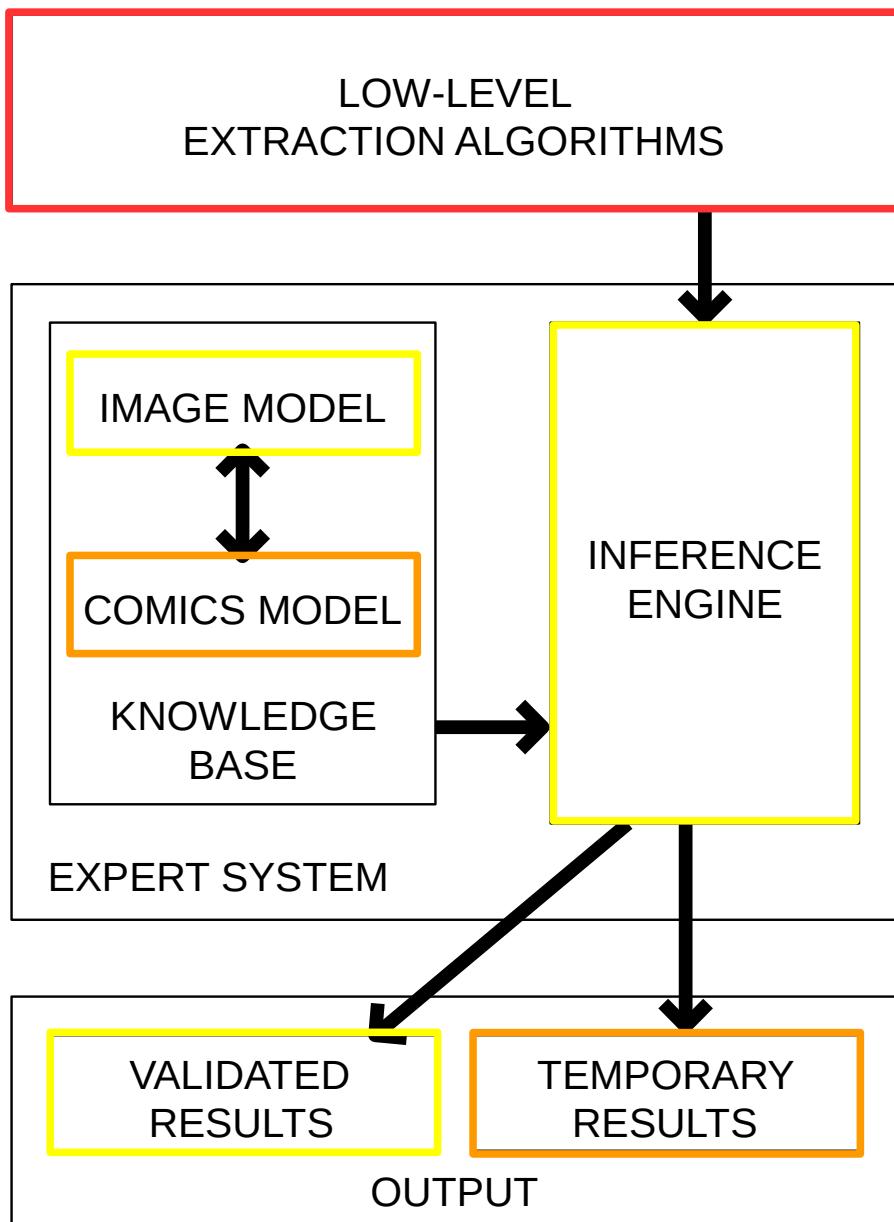
Contributions

Knowledge-driven approach

- Rigaud's thesis
- Collaboration
- Guérin's thesis

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

Knowledge representation



Contributions

Knowledge-driven approach

- Rigaud's thesis
- Collaboration
- Guérin's thesis

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

Processing sequence

Contributions

Knowledge-driven approach

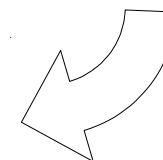


Formulate hypotheses



Validate hypotheses

Infer new information



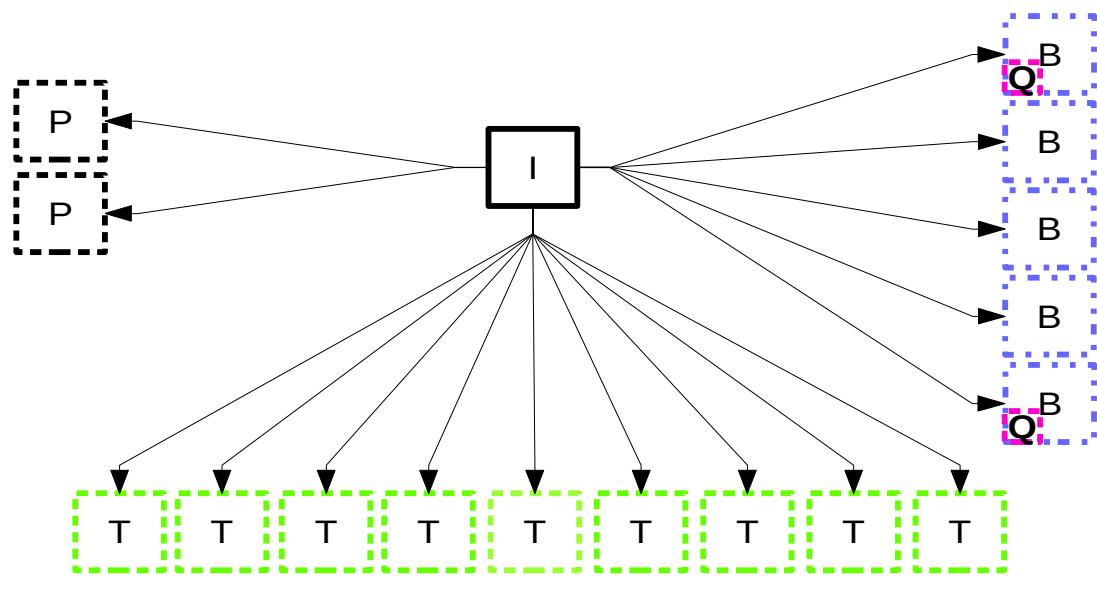
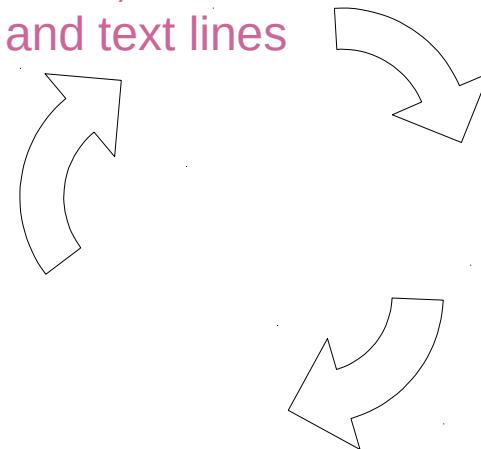
- Iteration 1
 - Step 1: hypotheses of **simple element** positions
 - Step 2: validation of the positions
 - Step 3: inference a new information
- Iteration 2
 - Step 1: hypotheses of more **complex elements**
 - Step 2: validation of the positions
 - Step 3: inference a new information
 - ...

Processing sequence

Contributions
Knowledge-driven approach



Hypotheses of
panels, balloons
and text lines



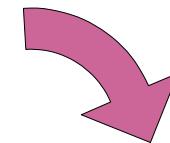
		Valid
P	P	Panel
B	B	Balloon
Q	Q	Tail
T	T	Text

Processing sequence

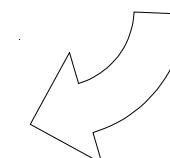
Contributions
Knowledge-driven approach



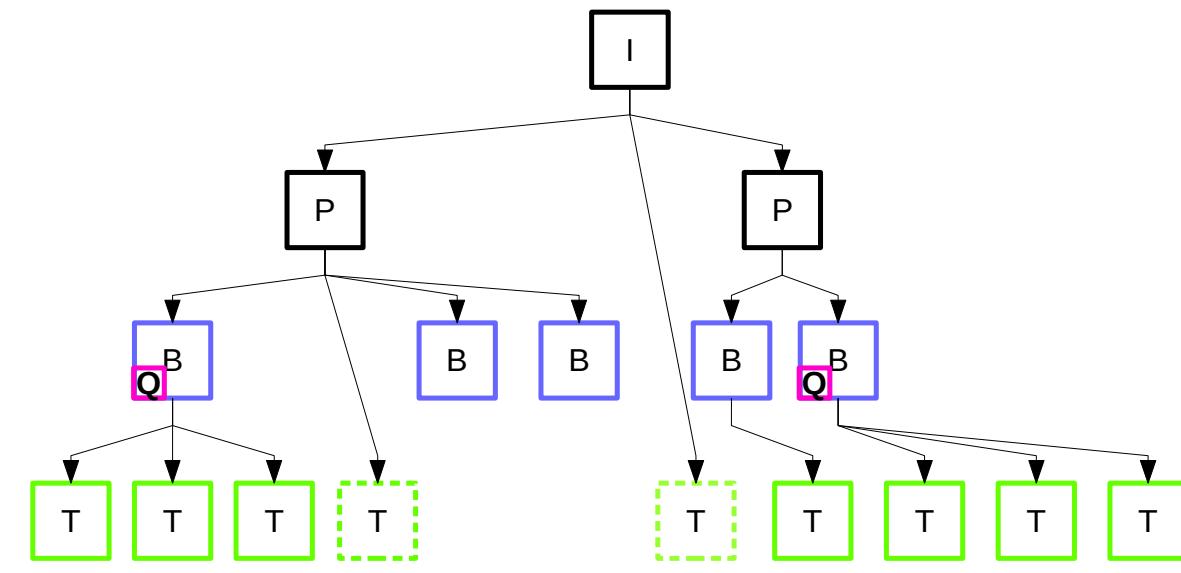
Hypotheses of
panels, balloons
and text lines



Validation of the
hypotheses



	Non valid	Valid	
P			Panel
B			Balloon
Q			Tail
T			Text

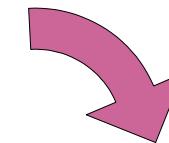


Processing sequence

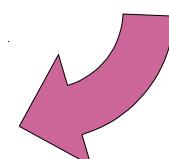


Contributions
Knowledge-driven approach

Hypotheses of
panels, balloons
and text lines

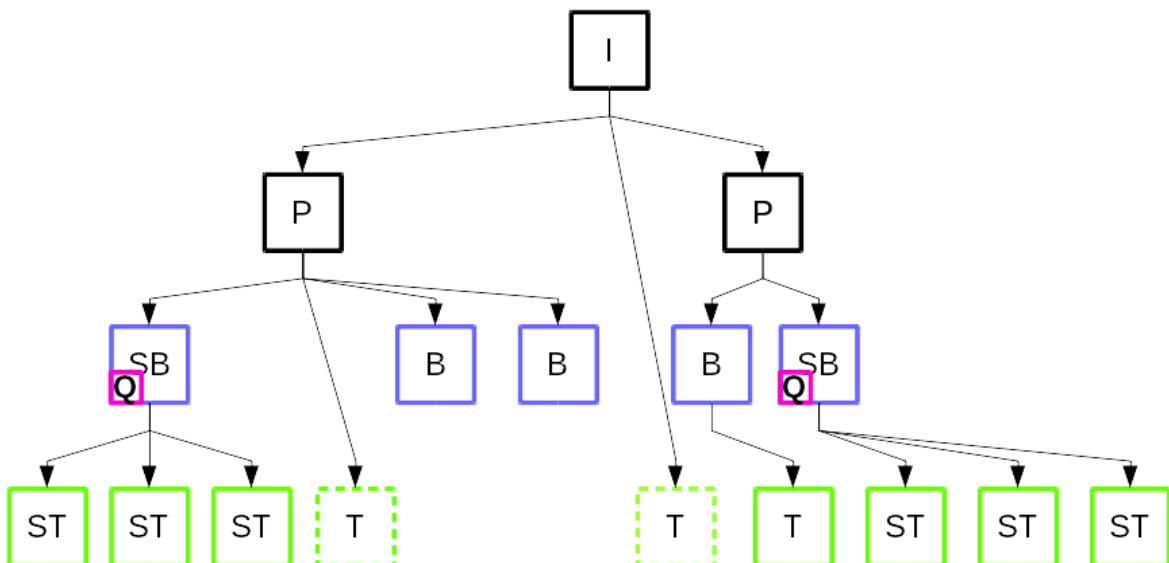


Inferences of
specific types

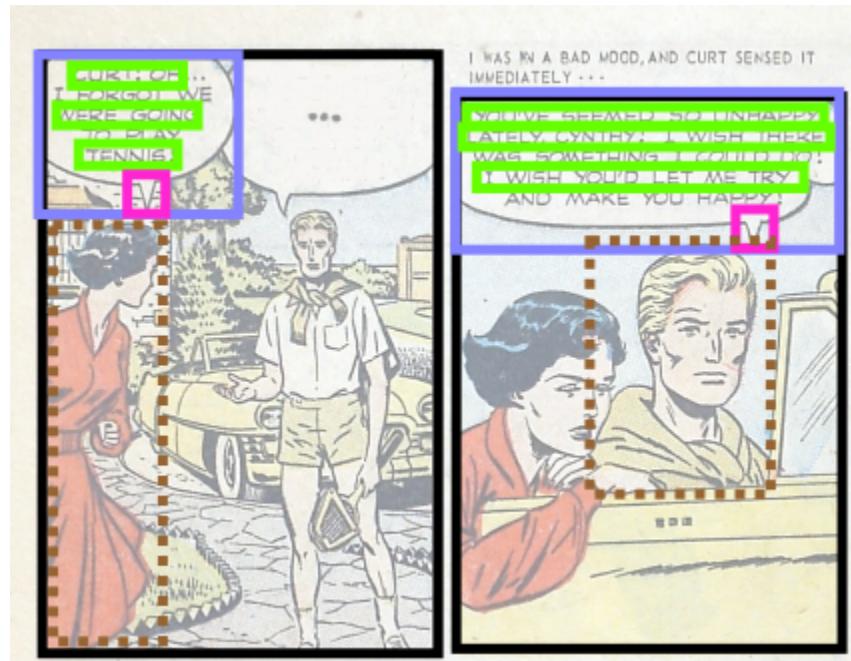


	Non valid	Valid
P		
SB		
Q		
ST		

Panel
Speech balloon
Tail
Speech text

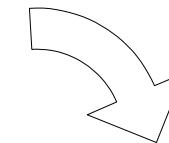


Processing sequence

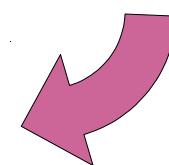


Contributions
Knowledge-driven approach

Hypotheses of
comic characters



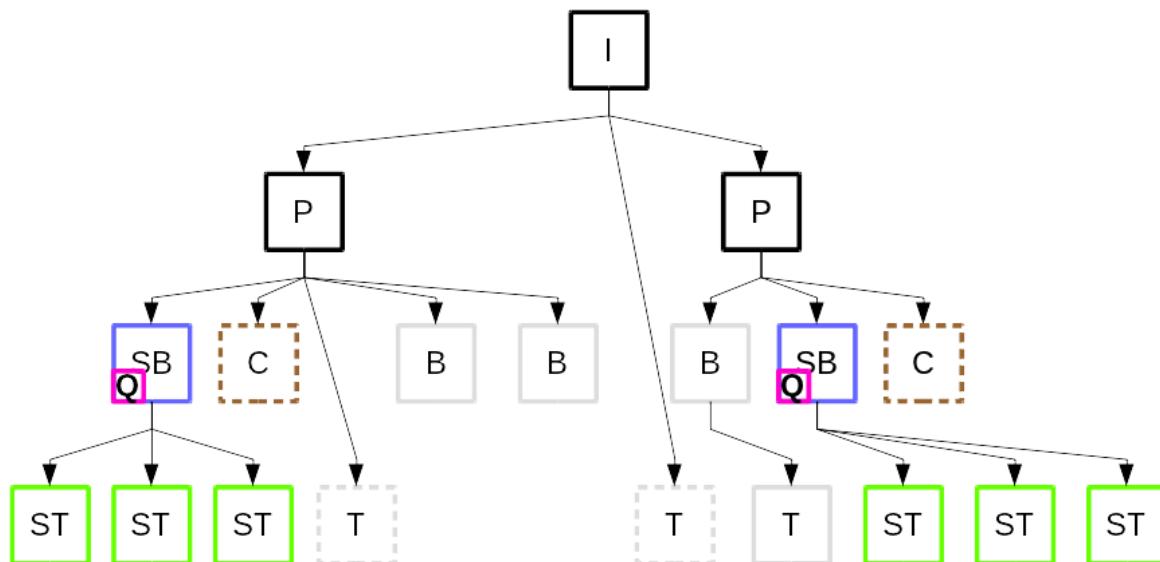
Inferences of
specific types



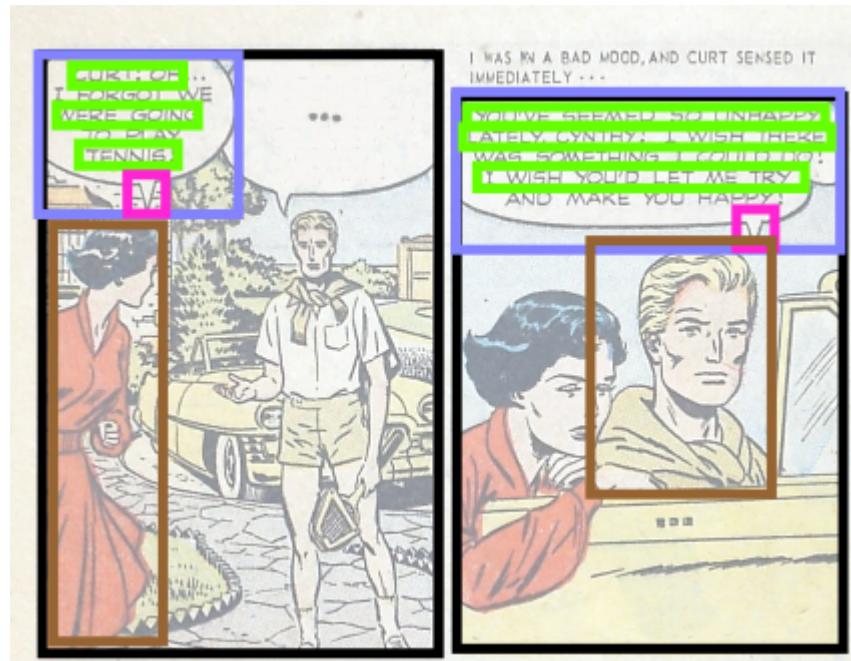
Non
valid Valid



Comic
character



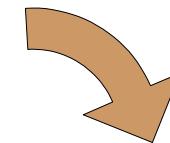
Processing sequence



Contributions

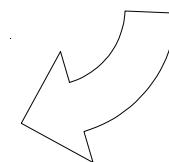
Knowledge-driven approach

Hypotheses of comic characters



Validation of the hypotheses

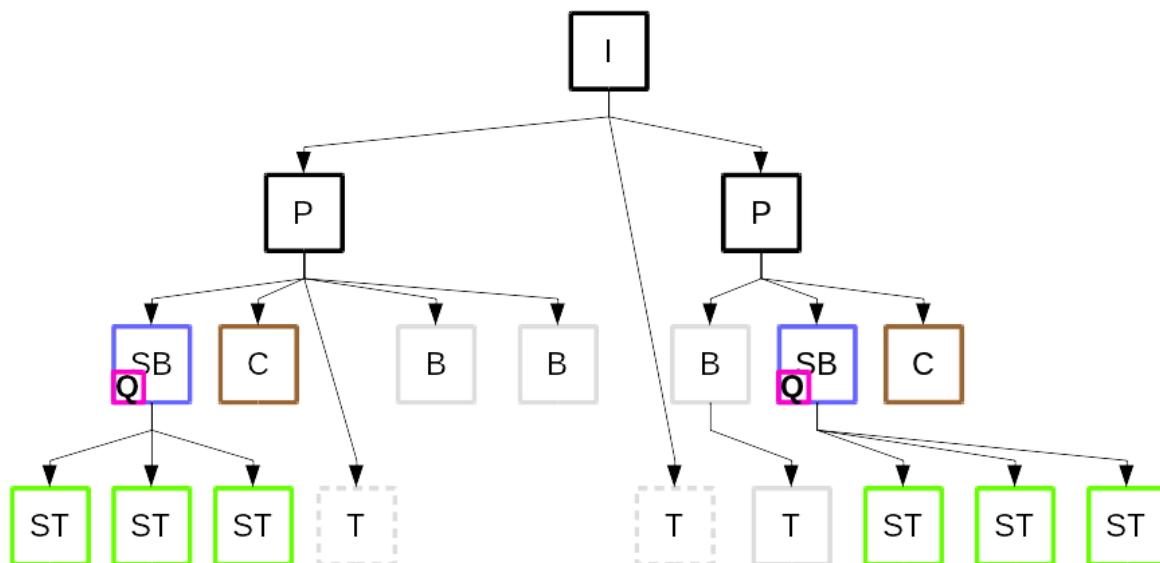
Inferences of specific types



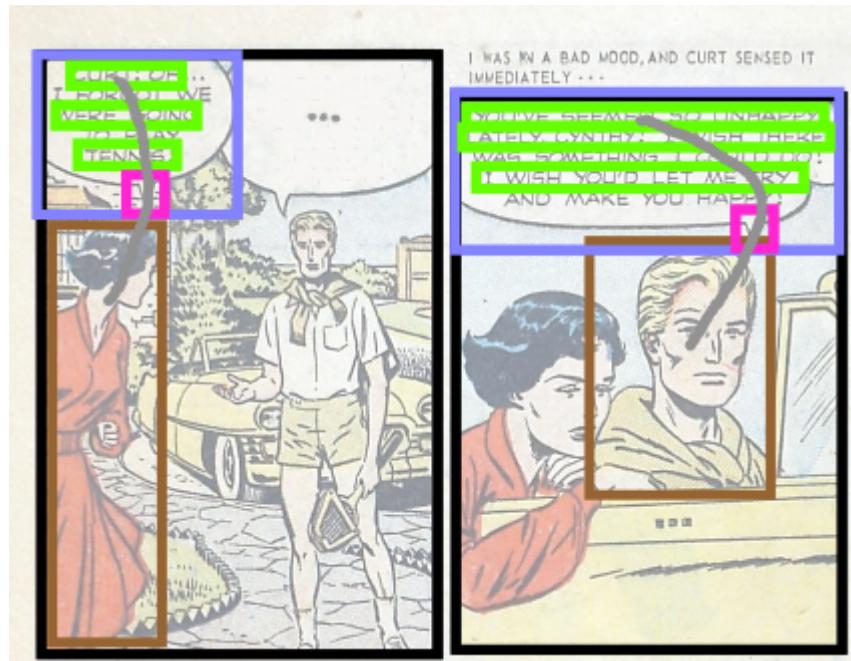
Non valid Valid



Comic character



Processing sequence

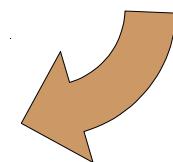


Contributions
Knowledge-driven approach

Hypotheses of
comic characters



Inferences of
specific types
+ semantic links



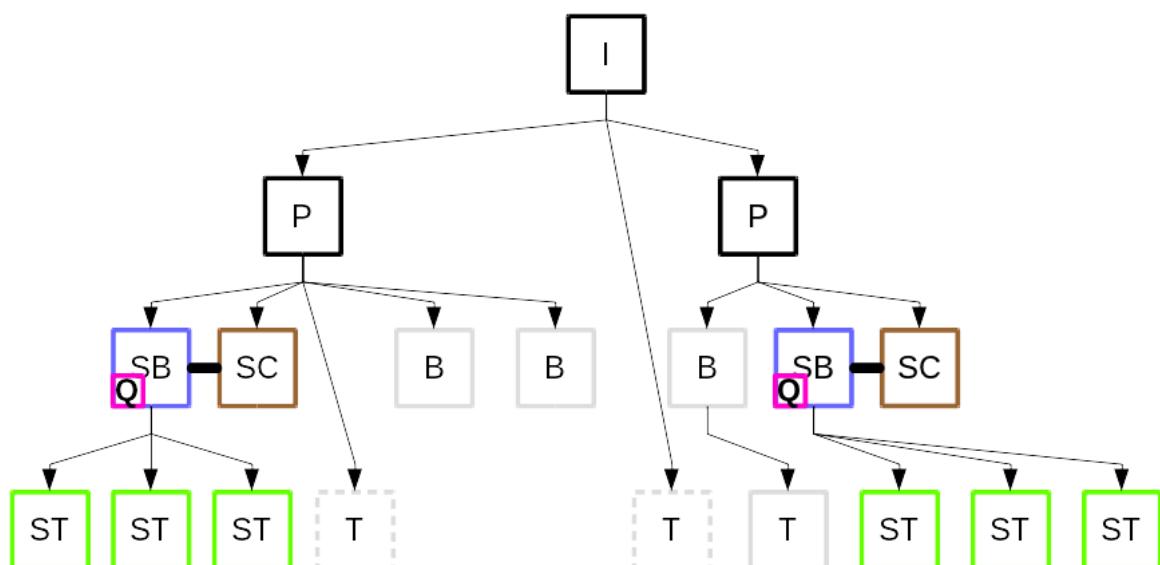
Non
valid



Valid



Speaking
comic
character



- Evaluations
- Overall contribution



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

Evaluations

B_p = predicted region

B_{gt} = ground truth region

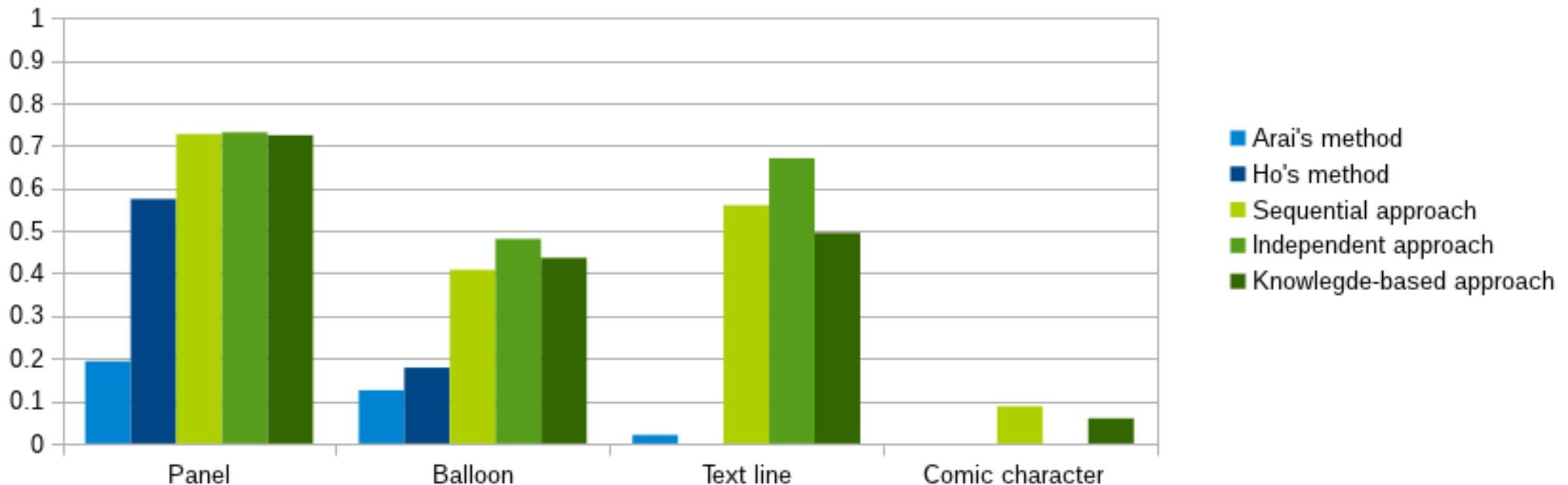
B_p is valid if $a_0 > 0.5$

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

Experiments

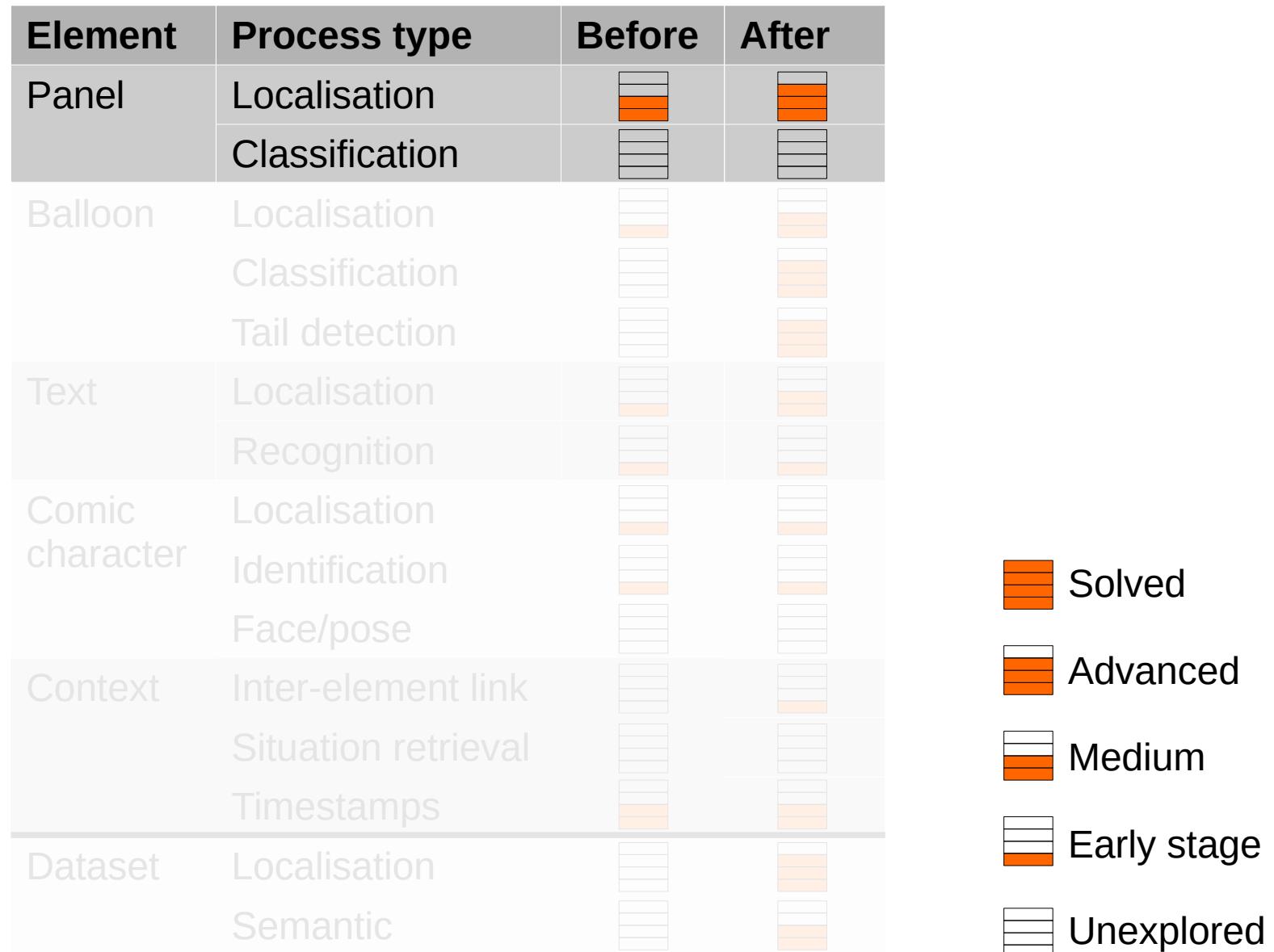
Number of element in the dataset	
Panel	850
Balloon	1092
Text line	4691
Comic character	1550

Extraction results on the eBDtheque dataset (F-score)



Overall contribution

Experiments



Overall contribution

Experiments



Overall contribution

Experiments



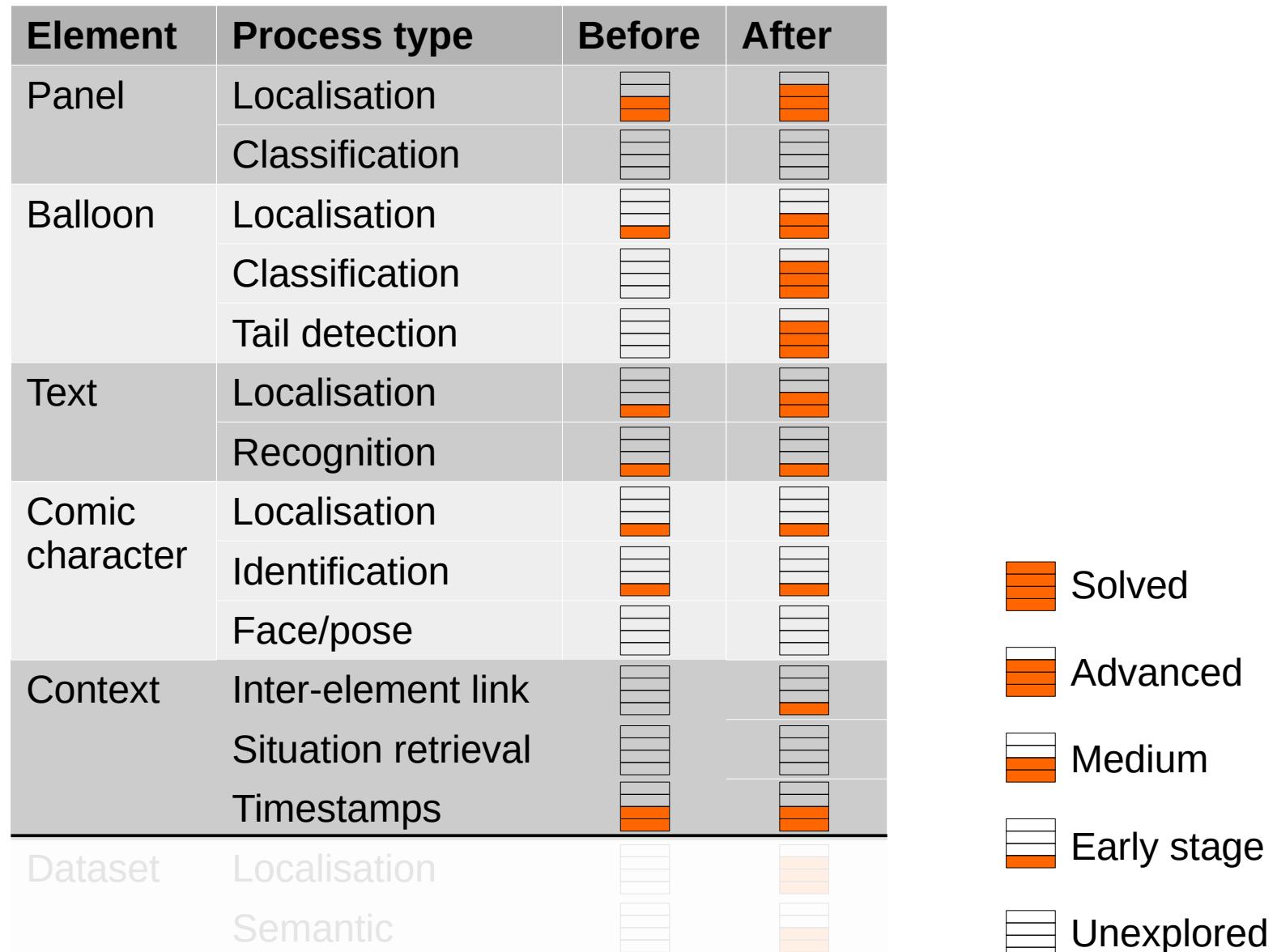
Overall contribution

Experiments



Overall contribution

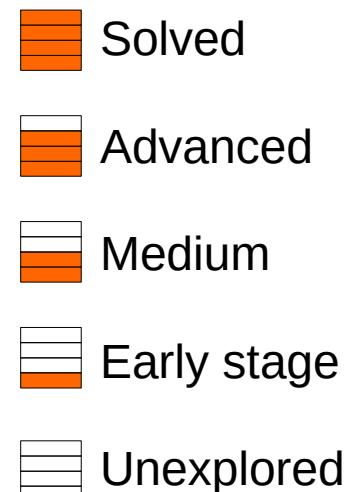
Experiments



Overall contribution

Experiments

Element	Process type	Before	After
Panel	Localisation		
	Classification		
Balloon	Localisation		
	Classification		
	Tail detection		
Text	Localisation		
	Recognition		
Comic character	Localisation		
	Identification		
	Face/pose		
Context	Inter-element link		
	Situation retrieval		
	Timestamps		
Dataset	Localisation		
	Semantic		



- Global conclusions
- Global perspectives
- Publications



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

- Reached objectives
 - Efficient **panel**, **balloon**, **text** and **tail** extraction methods
 - First approaches for **comic character** extraction and **context retrieval**
 - Public **dataset** and **ground truth** (<http://ebdtheque.univ-lr.fr>)
- Publications
 - 1 journal, 3 book chapters, 4 conferences, 5 workshops (3 national)
 - 6 local **seminars**
- 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 (Germany, India, China, Japan)
 - **National projects** (PIA BigData Actialuna/LIP6, ANR EXPION 2015)
 - **International project** on manga analysis (PHC-SAKURA with Japan)

- Content extraction
 - Consider overlapping panel extraction
 - Investigate text recognition
 - Improve implicit balloon extraction and evaluation
 - Extract and identify non-speaking comic characters
- Content understanding
 - Situation retrieval (e.g. landscape, outdoor, night)
 - Action recognition (e.g. running, driving, dreaming)
 - Interaction retrieval (e.g. balloon said by/to)
 - Labelling from text analysis (e.g. auto tagging)
- Dataset
 - Increase and balance the number of images of each category
 - Add more annotation (e.g. panel situation, comic character body parts and roles)

Publications

Conclusion

JOURNAL

Christophe Rigaud, Clément Guérin, Dimosthenis Karatzas, Jean-Christophe Burie and Jean-Marc Ogier. “**Knowledge-driven understanding of images in comic books**”. International Journal on Document Analysis and Recognition (**IJDAR**), 2015 (accepted with minor reviews).

BOOK CHAPTERS

Christophe Rigaud, Dimosthenis Karatzas, Jean-Christophe Burie and Jean-Marc Ogier. “**Adaptive contour classification of comics speech balloons**”. In Graphic Recognition. New Trends and Challenges. Lecture Notes in Computer Science (**LNCS**), Vol. 8746, pp. 53-62, 2014.

Nam Ho Hoang, Christophe Rigaud, Jean-Christophe Burie and Jean-Marc Ogier. “**Detecting recurring deformable objects: an approximate graph matching method for detecting characters in comics books**”. In Graphics Recognition. Current Trends and Challenges. Lecture Notes in Computer Science (**LNCS**), Vol. 8746, pp. 122-134, 2014.

Christophe Rigaud, Norbert Tsopze, Jean-Christophe Burie and Jean-Marc Ogier. “**Robust frame and text extraction from comic books**”. In Graphic Recognition. New Trends and Challenges. Lecture Notes in Computer Science (**LNCS**), Vol. 7423, pp. 129-138, 2013.

Publications

Conclusion

CONFERENCES

Christophe Rigaud, Dimosthenis Karatzas, Jean-Christophe Burie and Jean-Marc Ogier. “**Color descriptor for content-based drawing retrieval**”. In the Proceedings of the 11th IAPR International Workshop on Document Analysis Systems (**DAS**), pp. 267-271 , Tours, France, April, 2014.

Christophe Rigaud, Dimosthenis Karatzas, Joost Van de Weijer, Jean-Christophe Burie and Jean-Marc Ogier. “**An active contour model for speech balloon detection in comics**”. In the Proceedings of the 12th International Conference on Document Analysis and Recognition (**ICDAR**), pp. 1240-1244, Washington DC, USA, August, 2013.

Clément Guérin, Christophe Rigaud, Antoine Mercier, Farid Ammar-Boudjelal, Karelle Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier and Arnaud Revel. “**eBDtheque: a representative database of comics**”. In the Proceedings of the 12th International Conference on Document Analysis and Recognition (**ICDAR**), pp. 1145-1149, Washington DC, USA, August, 2013.

Christophe Rigaud, Dimosthenis Karatzas, Joost Van de Weijer, Jean-Christophe Burie and Jean-Marc Ogier. “**Automatic Text Localisation in Scanned Comic Books**”. In the Proceedings of the 8th International Conference on Computer Vision Theory and Applications (**VISAPP**), pp. 814-819, Barcelona, Spain, February, 2013.

Publications

Conclusion

WORKSHOPS

Clément Guérin, Christophe Rigaud, Karell Bertet, Jean-Christophe Burie, Arnaud Revel and Jean-Marc Ogier. “**Réduction de l'espace de recherche pour les personnages de bandes dessinées**”. In the Proceedings of the 19ème congrès national sur la Reconnaissance de Formes et l'Intelligence Artificielle (**RFIA**), Rouen, France, July, 2014.

Christophe Rigaud, and Clément Guérin. “**Localisation contextuelle des personnages de bandes dessinées**”. In the Proceedings of the 13ème Colloque International Francophone sur l'Ecrit et le Document (**CIFED**), pp. 367–370, Nancy, France, March 2014.

Christophe Rigaud, Dimosthenis Karatzas, Jean-Christophe Burie and Jean-Marc Ogier. “**Speech balloon contour classification in comics**”. Proceedings of the 10th International Workshop on Graphics RECognition (**GREC**), pp. 23-25, Bethlehem, USA, August, 2013.

Hoang Nam Ho, Christophe Rigaud, Jean-Christophe Burie and Jean-Marc Ogier. “**Redundant structure detection in attributed adjacency graphs for character detection in comics books**”. In the Proceedings of the 10th IAPR International Workshop on Graphics RECognition (**GREC**), pp. 109-113, Bethlehem, PA, USA, August, 2013.

Christophe Rigaud, Norbert Tsopze, Jean-Christophe Burie and Jean-Marc Ogier. “**Extraction robuste des cases et du texte de bandes dessinées**”. In the Proceedings of the 10ème Colloque International Francophone sur l'Ecrit et le Document (**CIFED**), pp. 349-360, Bordeaux, France, March 2012.

References

Conclusion

- [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.
- [Cheung08] S.C.S. Cheung, City University of Hong Kong. Run Run Shaw Library, and City University of Hong Kong. **Face Detection and Face Recognition of Human-like Characters in Comics.** Outstanding academic papers by students. Run Run Shaw Library, City University of Hong Kong, 2008.
- [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 Information Processing - PCM 2007, volume 4810 of Lecture Notes in Computer Science, pages 775–784. Springer Berlin Heidelberg, 2007.
- [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.
- [Ho12] Anh Khoi Ngo Ho, Jean-Christophe Burie, and Jean-Marc Ogier. **Panel and speech balloon extraction from comic books.** In 2012 10th IAPR International Workshop on Document Analysis Systems, pages 424–428. IEEE, 2012.

References

Conclusion

- [Khan12] Fahad Shahbaz Khan, Muhammad Anwer Rao, Joost van de Weijer, Andrew D. Bagdanov, Maria Vanrell, and Antonio Lopez. **Color attributes for object detection**. In 25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012), 2012.
- [Li14a] Luyuan Li, Yongtao Wang, Zhi Tang, and Liangcai Gao. **Automatic comic page segmentation based on polygon detection**. Multimedia Tools Applications, 171–197, 2014, Kluwer Academic Publishers.
- [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
- [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.
- [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.
- [Stommel12] Martin Stommel, Lena I Merhej, and Marion G Müller. **Segmentation-free detection of comic panels**. In Computer Vision and Graphics, pages 633–640, 2012.

References

Conclusion

- [Su11] Chung-Yuan Su, Ray-I Chang, and Jen-Chang Liu. **Recognizing text elements for svg comic compression and its novel applications**. In Proceedings of International Conference on Document Analysis and Recognition (ICDAR), pages 1329–1333, Washington, DC, USA, 2011.
- [Sun10] Weihan Sun and Koichi Kise. **Similar partial copy detection of line drawings using a cascade classifier and feature matching**. In Hiroshi Sako, Katrin Franke, and Shuji Saitoh, editors, ICWF, volume 6540 of Lecture Notes in Computer Science, pages 126–137. Springer, 2010.
- [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.



GRACIAS ARIGATO SHUKURIA JUSPAXAR TASHAKKUR ATU YAQHANYELAY MAAKE SUKSAMA EKKHMET MEHRBANI PALDIES GRAZIE GOZAIMASHITA EFCHARISTO LAH BOLZIN TINGKI BİYAN SHUKRIA THANK YOU MERCI

SPASSIBO NUHUN CHALTU WABEEJA MAITEKA HUI GUI EKOJU SIKOMO MANETAI
SNAZHAIHUYA DHANYABAAD ANNA UNALCHEESH
SANCO ATTO DENKAUJA MERESTI
MERASTAWHY GAEJTHO SPASIBO
FAKAAUE DENKAUJA
BAINKA EFCHARISTO AGUYJE
TAVTAPUCH MEDAWAGSE
KOMAPSUMNIDA
LAH
MINMONCHAR



<https://github.com/crigaud/thesis>
<http://www.christophe-rigaud.com>

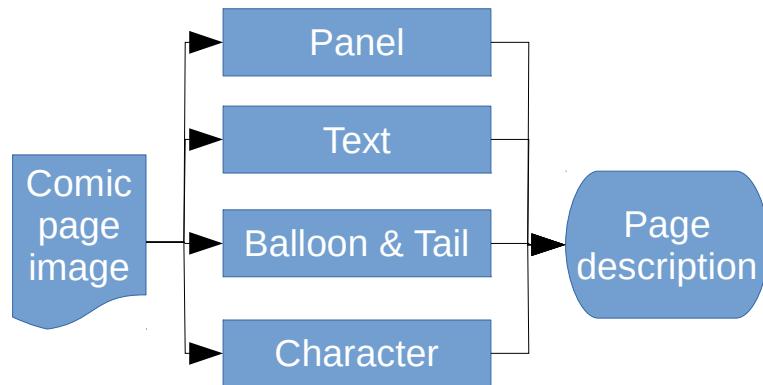
Applications examples

Experiments

Independent approach details

Contributions

Objective: **avoid error propagation**



Independent panel extraction

Contributions



(a) Original image



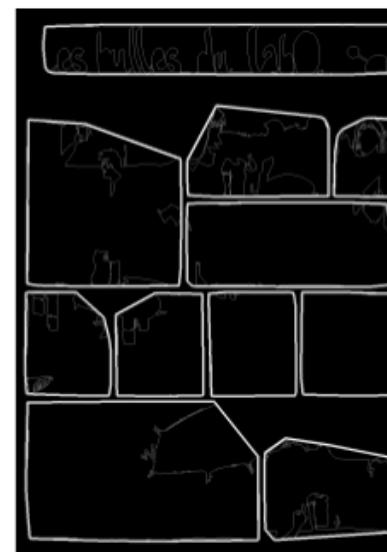
(b) Binary segmentation



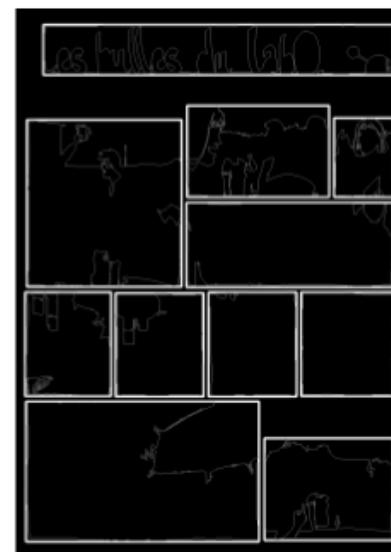
(c) Outermost contours



(d) Panel contours



(e) Panel contour hulls



(f) Panel contour bounding boxes

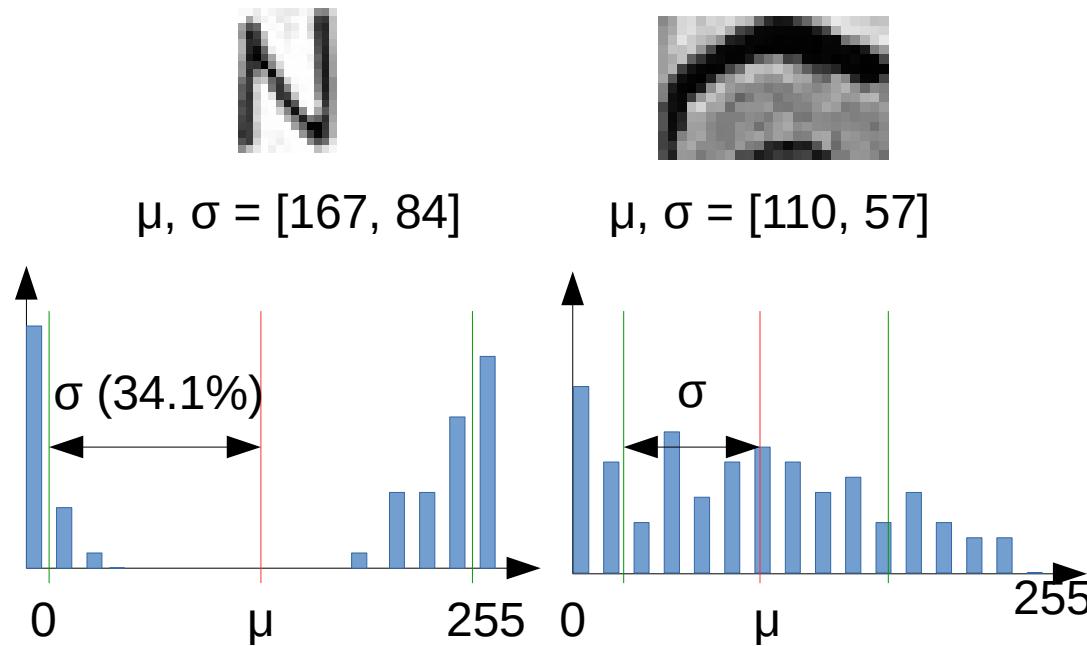
Independent text extraction

Contributions

Key idea: the speech text regions are the more contrasted region (black over white).

"We compare the standard deviation of the pixel grey level distribution of each CC bounding box with the contrast ratio of the page to make a low/high local contrast CC separation, similarly to [101] using the Mahalanobis distance"

Contrast ratio of the page $r = \max \text{ gray value} - \min \text{ gray value}$
Optimal Mahalanobis distance $dm = 2$ (according to validation)



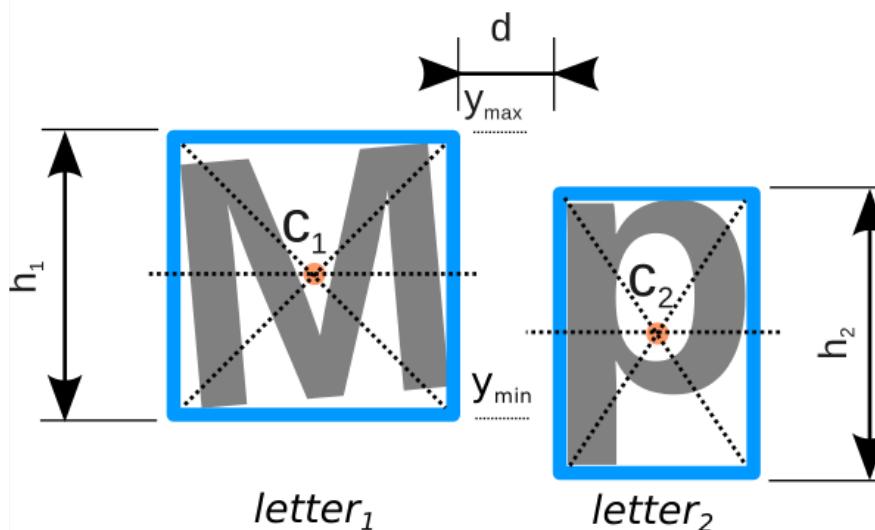
If $\sigma > r / dm$ then the CC is a letter candidate
Else the CC is not considered as candidate

Letter to text line

Contributions

“Similarly to [29, 100], we first search for the first letter (connected-component) of each text line and then attach the close and similar connected-components to the same text line.”

“A letter is considered first if it is positioned on the left, on the same horizontal line and if there is no intersection found with any other letters at a distance equal to the letter height. Then the other letters on the right are added by checking their relative horizontal and vertical positions.”

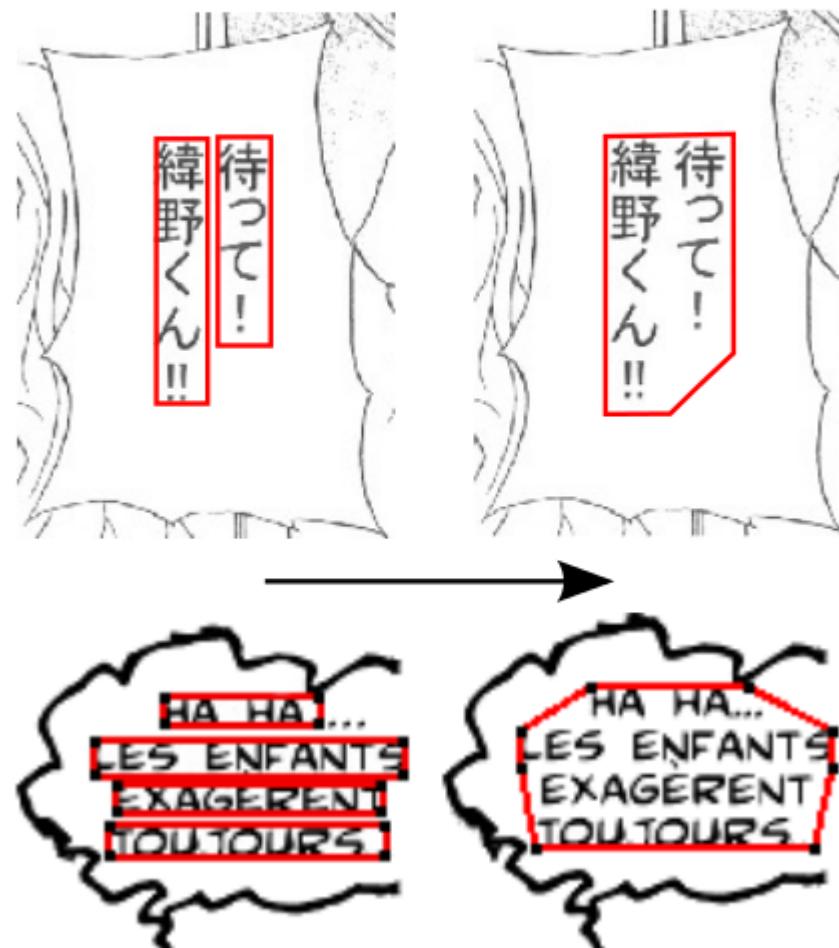


- The horizontal inter-letter distance d should be smaller than the maximal height of the letters ($d < \max(h_1, h_2)$);
- The vertical alignment is considered as correct if the vertical coordinate of the centre of the next letter c_2 passes through the first letter ($y_{min}(\text{letter}_1) > c_{2y}$ and $y_{max}(\text{letter}_1) > c_{2y}$);

Text line to text block/paragraph

Contributions

“First, we require that the candidate text lines to group have similar heights (or width in case of vertical text) and second that the inter-line distance is smaller than the average text line height of the potential paragraph region.”



Independent balloon extraction

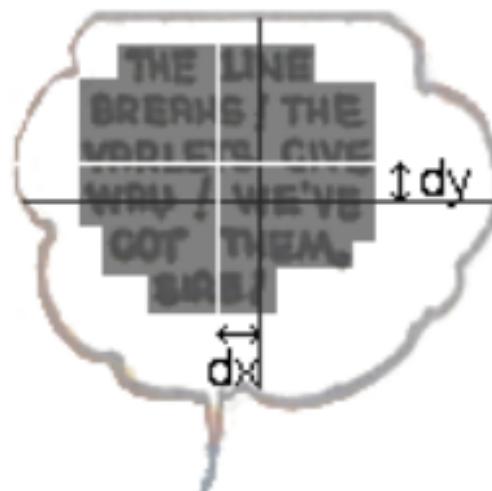
Contributions

- Aligned children (horizontal or vertical)

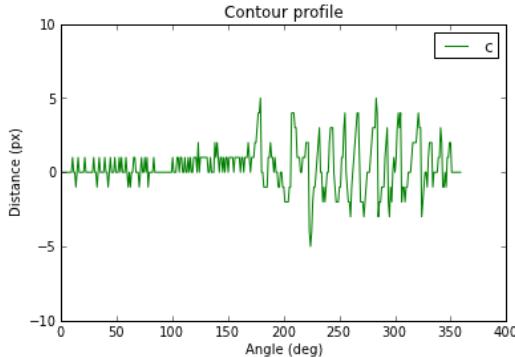
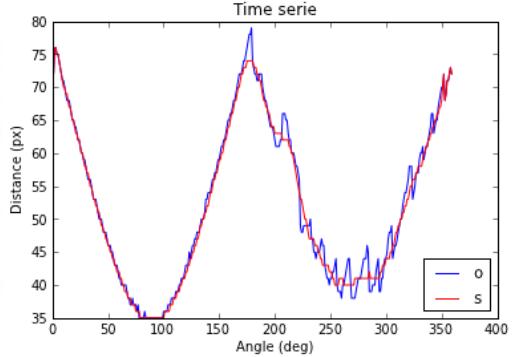
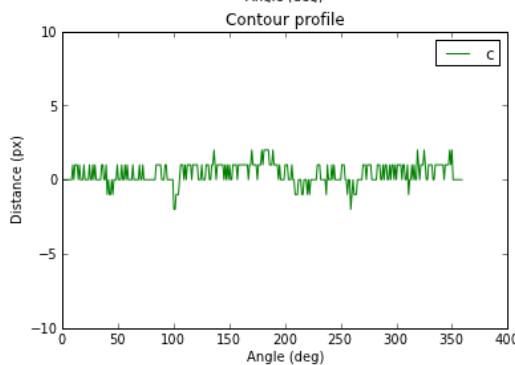
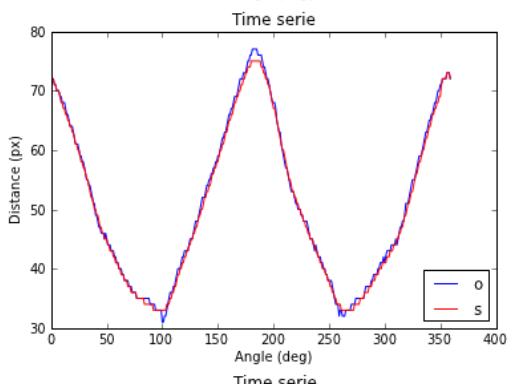
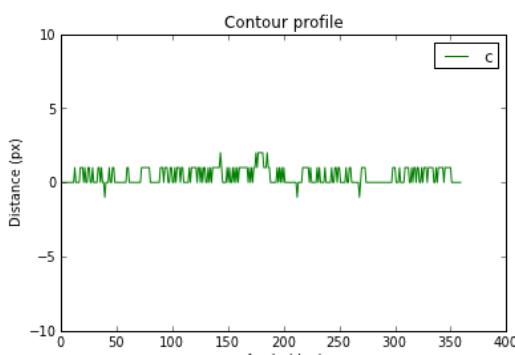
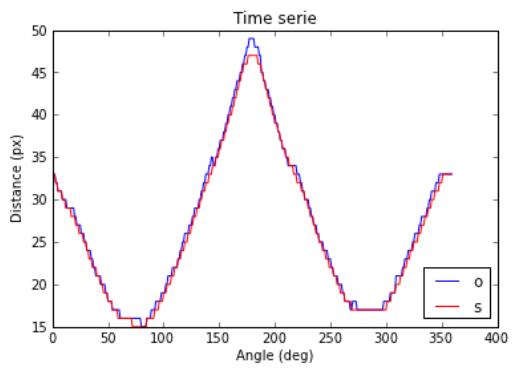
$$align(ch_i) = \frac{|CHA(ch_i)|}{n}$$

- Confidence based on inter-children alignment and children block alignment to candidate CC

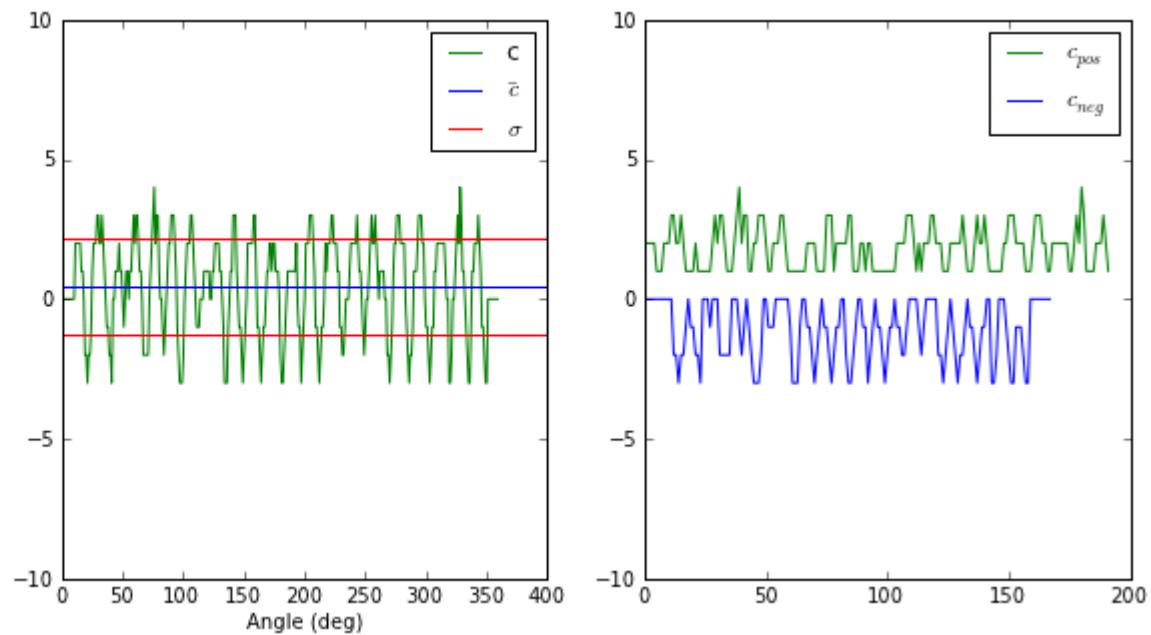
$$C_{balloon} = \frac{1}{3} * (\overline{align} + d_x + d_y)$$



Independent balloon classification 1 Contributions



Independent balloon classification 2 Contributions



Feature 1:

$$f_1 = \frac{\sigma_c}{\bar{o}}$$

σc = standard deviation of signal c

Feature 2:

$$f_2 = \frac{\sigma_{neg} - \sigma_{pos}}{\bar{o}}$$

Independent comic character 1/6

Contributions

1) Colour quantization

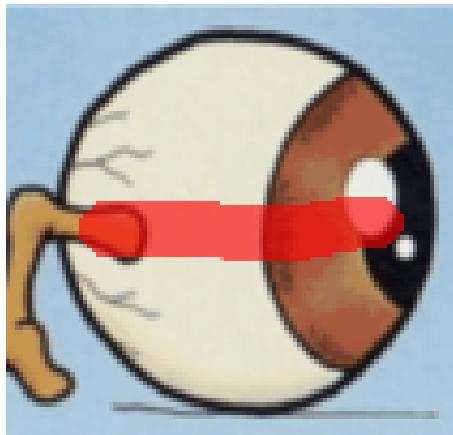
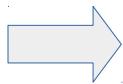


Colour quantization

Independent comic character 2/6

Contributions

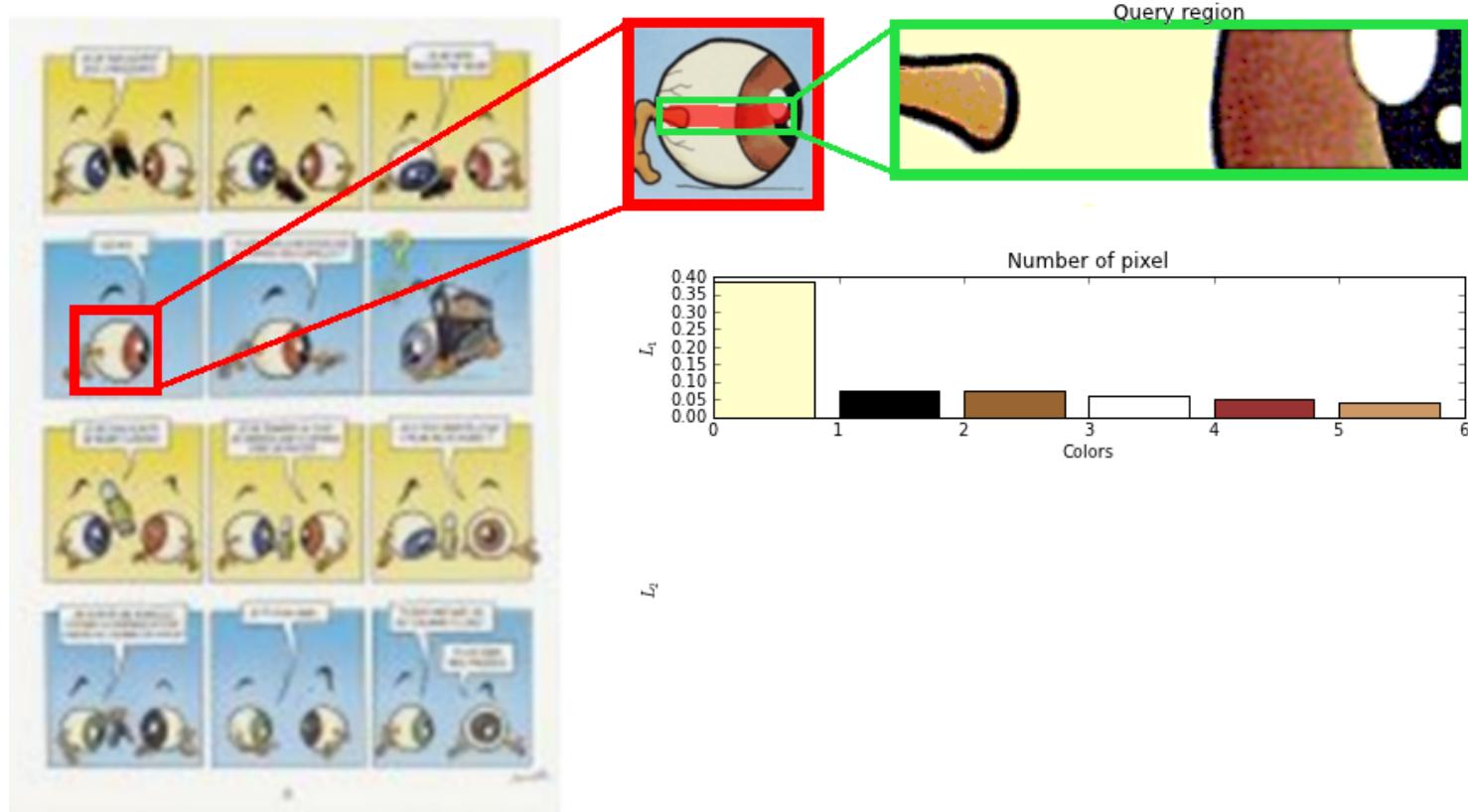
2) User query



Independent comic character 3/6

Contributions

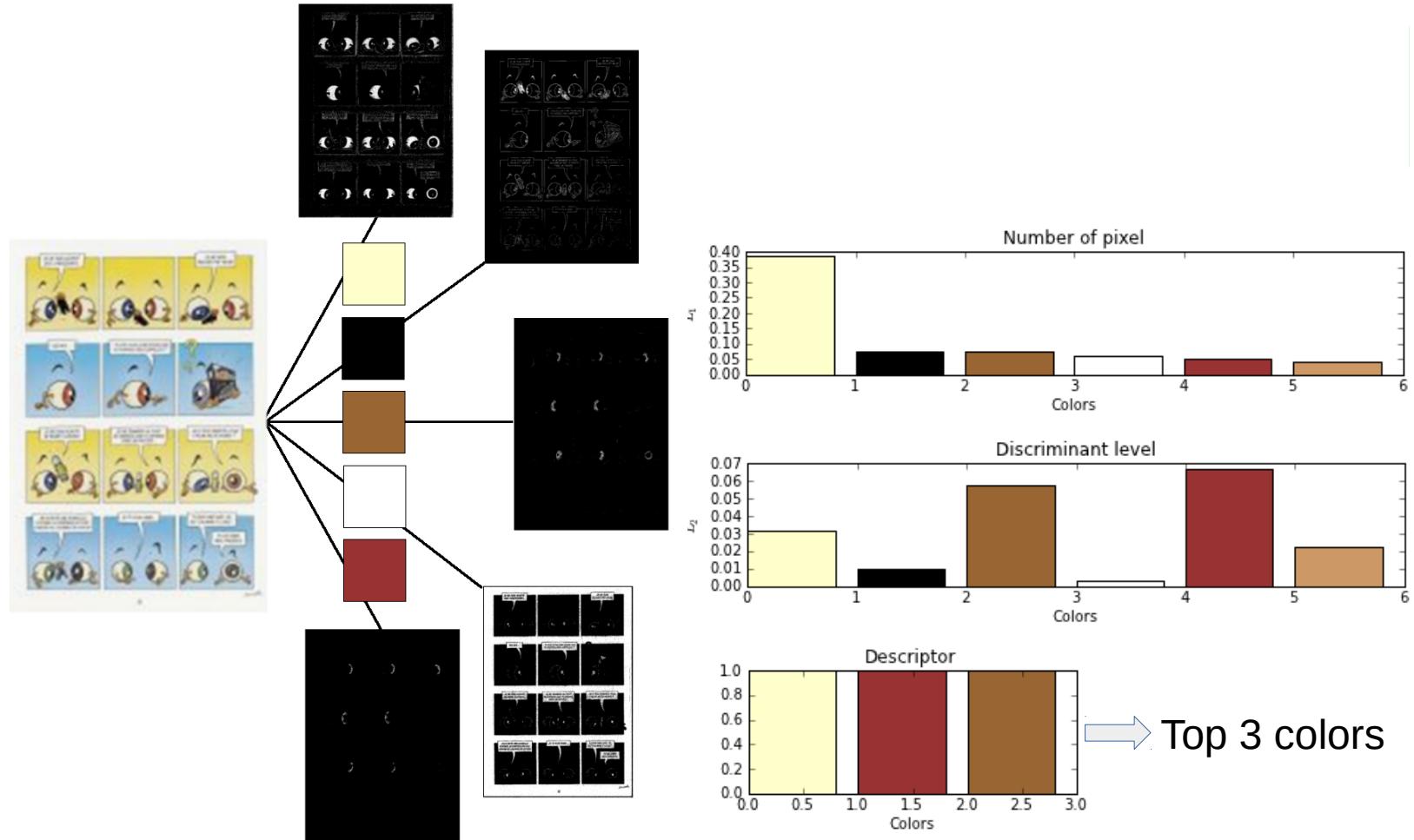
3) Color selection: occurrence level



Independent comic character 4/6

Contributions

3) Color selection: occurrence + discriminability levels



Independent comic character 5/6

Contributions

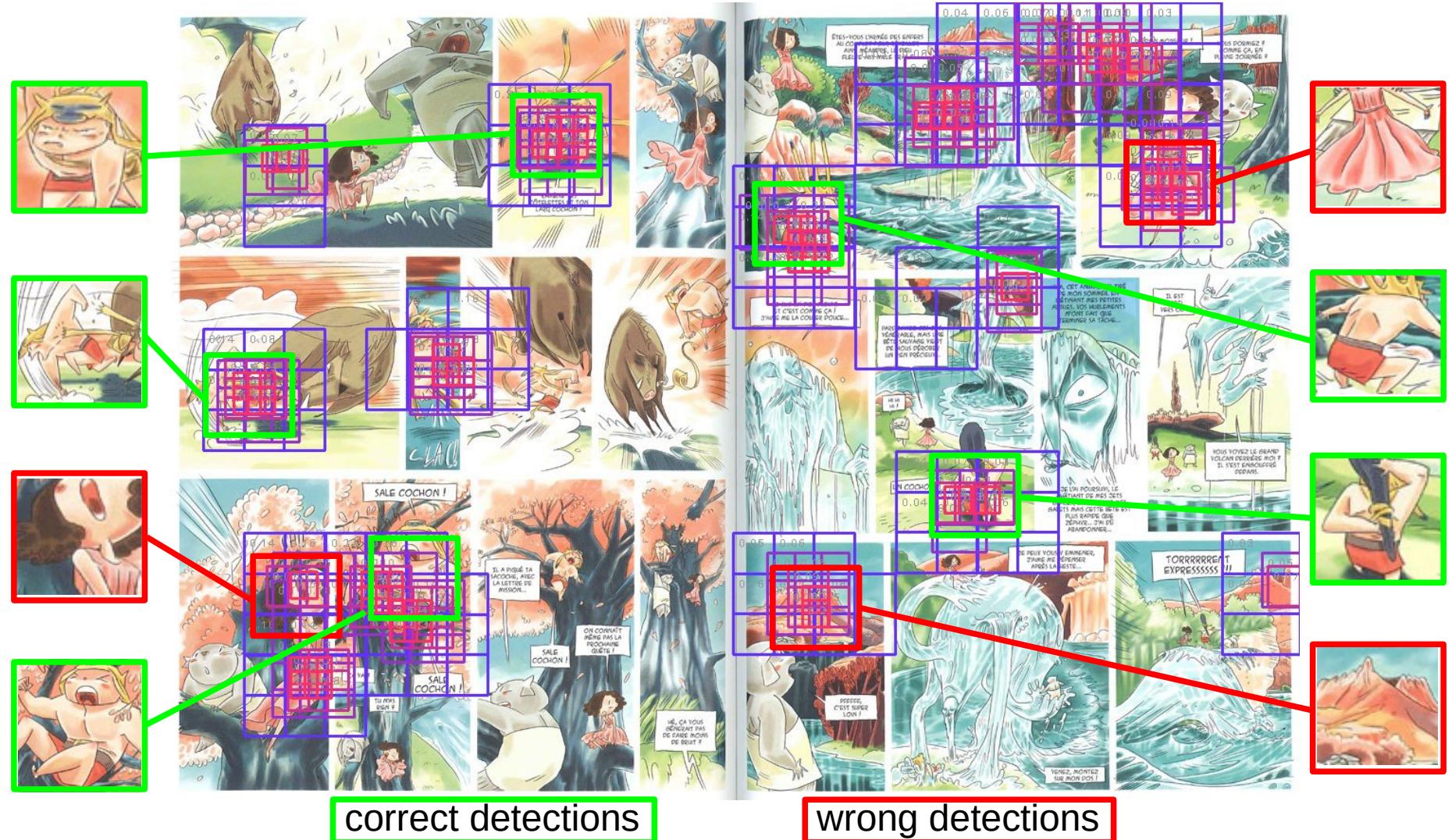
4) Character spotting: example



Image credits: (Prunelle, la fille du cyclope, Ankama, 2010)

Independent comic character 6/6

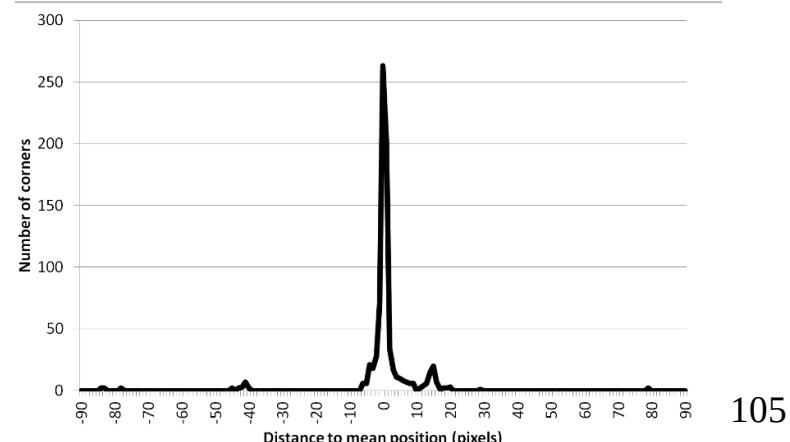
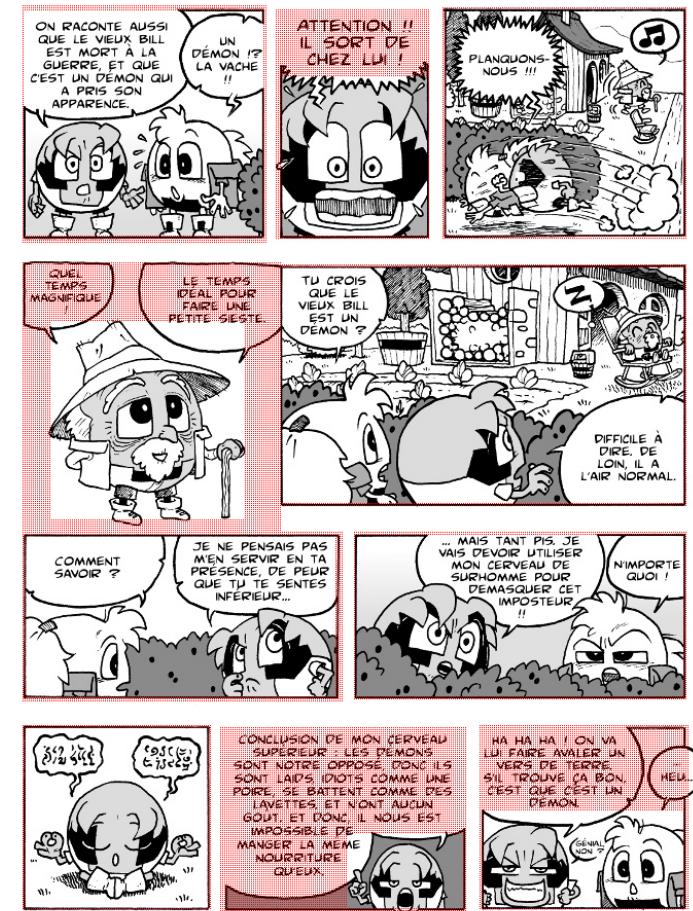
Contributions



Dataset & quality assessment

IDs	Date	Category	Colour	Nb
CYB_BUBBLEGOM	2009	French webcomics	Grey	10
CYB_COSMOZONE	2008	French webcomics	Colour	5
CYB_MAGICIENLOOSE	2011	French webcomics	Colour	1
CYB_MOUETTEMAN	2011	French webcomics	Colour	4
CYB_PSYCOCHON	2010	French webcomics	Colour	1
CYB_SUPERCOMPTOIR	2011	French webcomics	Colour	1
CYB_TRAFFIC	2009	French webcomics	B&W	3
FOX_CHILLINTALES	1953	American comics	Colour	4
FRED_PHILEMON	2004	French comics	Colour	5
INOUE_KYOUMEN	2008	Japanese manga	B&W	6
JOLIVET_BOSTONPOLICE	2010	French comics	Colour	2
JOLIVET_BOSTONPOLICE	2010	French comics	B&W	3
LAMISSEB_ETPISTAF	2012	French webcomics	B&W	5
LAMISSEB_LESNOEILS1	2011	French comics	Colour	5
LUBBIN_LESBULLESDULABO	2012	French webcomics	Colour	2
MCCALL_ROBINHOOD	1946	American comics	Colour	4
MCCAY_LITTLENEMO	1905	French comics	B&W	1
MCCAY_LITTLENEMO	1905	French comics	Colour	2
MIDAM_GAMEOVER	2009	French comics	Colour	2
MIDAM_KIDPADDLE	2001	French comics	Colour	2
MONTAIGNE_				
TUMOURRASMOINS	2011	French webcomics	Colour	5
PIKE_BOYLOVEGIRLS	1953	American comics	Colour	3
ROUDIER_				
LESTERRESCREUSEES	2011	French comics	Colour	5
SAINTOGAN_PROSPER	1934	French comics	Colour	5
SAINTOGAN_ZIGPUCE_				
ET_ALFRED	1952	French comics	Colour	5
SAINTOGAN_ZIGPUCE_				
MILLIONAIRES	1928	French comics	Colour	5
TRONDHEIM_				
LES_TROIS_CHEMINS	2000	French comics	Colour	2
WARE_ACME	2005	American comics	Colour	2

Contributions



Metrics

Experiments

B_p = predicted region

B_{gt} = ground truth region

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

B_p is a true positive **TP** if $a_0 > 0.5$, else false positive **FP**

A false negative **FN** corresponds to a missed detection.

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$F-score = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}}$$

New binarisation approach

Contributions



Original

Luminance

th=250 [Arai11]

th=Otsu

Production to interpretation

Background

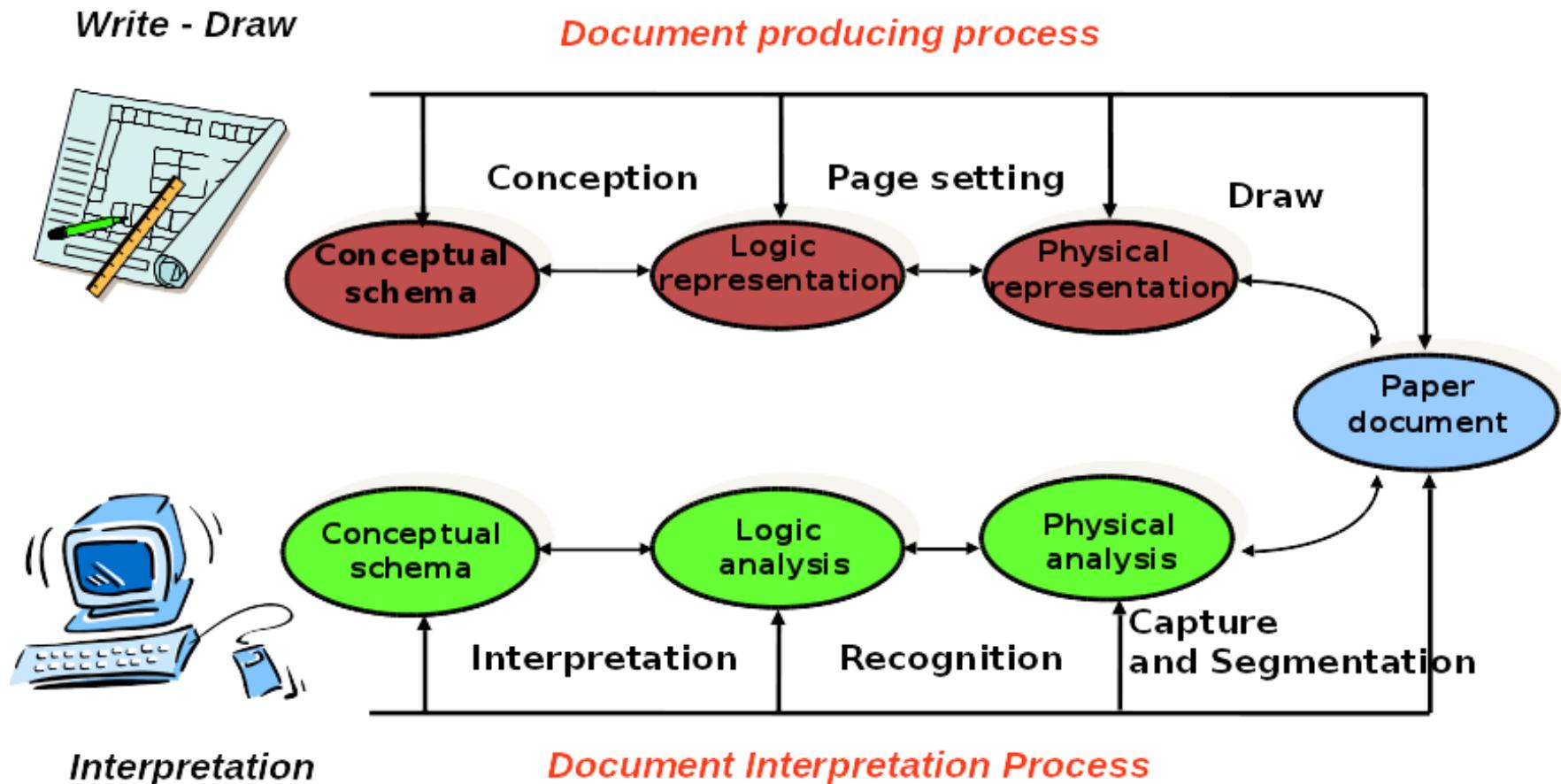


Image source: Handbook of Document Image Processing and Recognition. Springer, 2014

Element	Process type	Analysis method	Method	Publication	Before	After	Δ
Panel	Localisation	Histogram	S	CIFED'12 LNCS'13	60%	70%	+10%
		Topology	I, K	IJDAR'15			
Balloon	Localisation	Topology	S	-	20%	50%	+30%
		Region	I, K	-			
		Edge	S	ICDAR'13			
	Classification	Contour	S, I, K	GREC'13 LNCS'14	0%	80%	+80%
Text	Localisation	Topology	S		5%	60%	+55%
		Structure	I, K	VISAPP'13			
	Recognition	-	-	-	10%	10%	0%
Comic characters	Localisation	Graph	-	GREC'13 (2nd)	5%	10%	+5%
		Context	S, K	CIFED'14 RFIA'14 (2nd)			
		Spotting	Colour	I			
	Identification	-	-	-	10%	10%	0%
	Face/pose	-	-	-	0%	0%	0%
Context	Inter-element	Proximity	S, K	CIFED'14 RFIA'14 (2nd) IJDAR'15	0%	10%	+10%

History of comics art

Background

- 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)

Contributions



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