



European Ph.D. defense

Segmentation and indexation of complex objects in comic book images

Christophe Rigaud
December 11th, 2014

Jointly supervised by:

Jean-Christophe Burie¹
Dimosthenis Karatzas²
Jean-Marc Ogier¹

Outlines

- Introduction
- Document image analysis
- State of the art of comics analysis
- Contributions
- Experimentations
- Conclusions

Outlines

- Introduction
 - Comic books
 - History of comics art
 - Market place
 - Comics project
 - Objectives of the thesis
- Document image analysis
- State of the art of comics analysis
- Contributions
- Experimentations
- Conclusions

"juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer" Mc Cloud, 1993

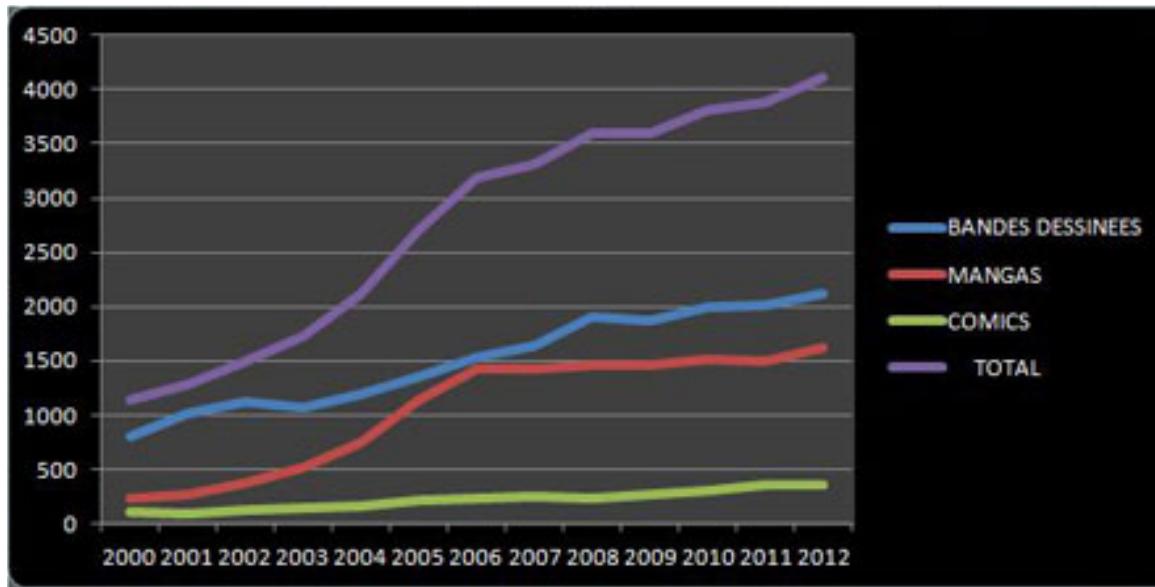
"a visual medium used to express ideas via images, often combined with text or visual information" Wikipédia, 2014

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

- Prehistory: painting of animals and hunters in caves
- 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 ofto social media sites and mobile devices

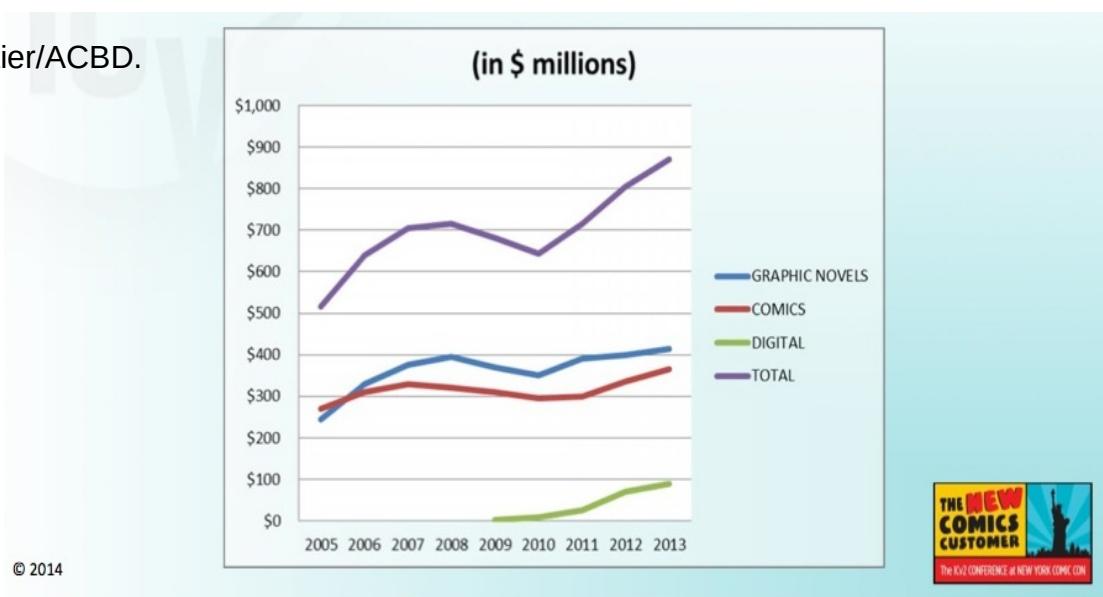
Market place

Introduction



Francophone comics production

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



© 2014

Comics market in the US
Milton Griepp's White Paper, ICv2 Conference 2014

Comics project (eBDtheque)

Introduction

- What?
 - Add value to paper-based comics using the new technologies
- Why?
 - Answer to a real need from librarians, advertisers and readers
 - Allows text/image search, reflowable documents, augmented reading and translation assistance
- How?
 - Extracting content of digitalized comic books (e.g. panels, balloons, text, comic characters)
 - Retrieving the semantic of the elements (e.g. read before, said by, thought by, addressed to)
- Who?
 - Supported by L3i lab
 - 2 Ph.D. students, 6 professors, 1 engineer (one year) and 1 post doc (one year)
 - Public funding
 - CPER 2007-2013 (State-Region Project Contract)
 - PHC-Sakura 2014-2015 (France/Japan Bilateral Joint Research Project)
 - PIA-iiBD 2015-2017 (future investment project with French company/labs)

Objectives of this thesis

Introduction

- Propose generic methods for content extraction of digitalized comic books
- Indexation of content in order to be browsable and exchangeable???
- Duration 36 months
- Challenges:
 - Recent field of research with a largely unknown
 - The documents are semi-unstructured, free-form and with complex background

Outlines

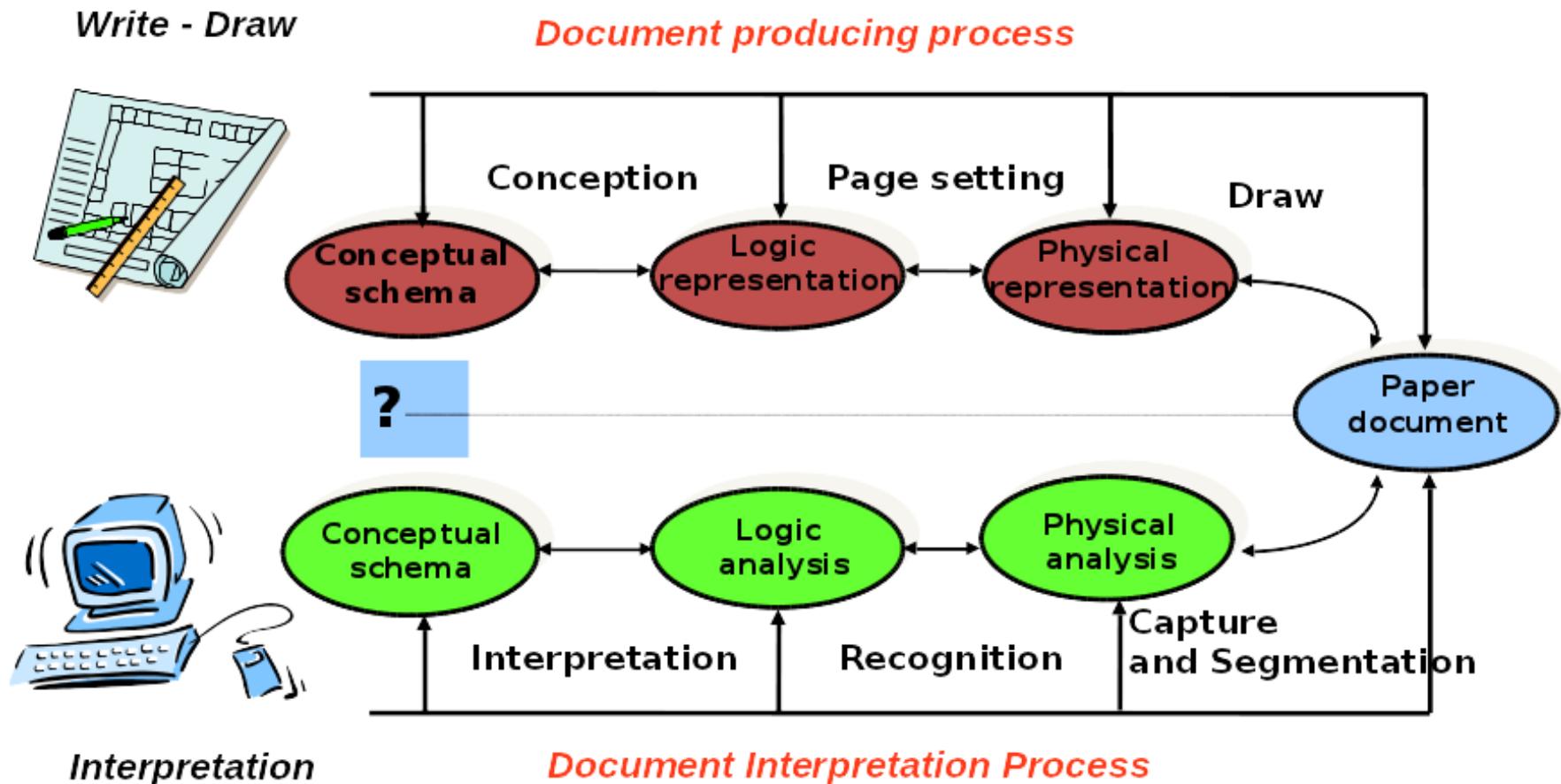
- Introduction
- Document analysis
- State of the art of comics analysis
- Contributions
- Experimentations
- Conclusions

Outlines

- Introduction
- Document analysis
 - Conception to interpretation
 - Comic books production
 - Comic books interpretation
 - (Document type comparison)
- State of the art of comics analysis
- Contributions
- Experimentations
- Conclusions

Conception to interpretation

Document analysis

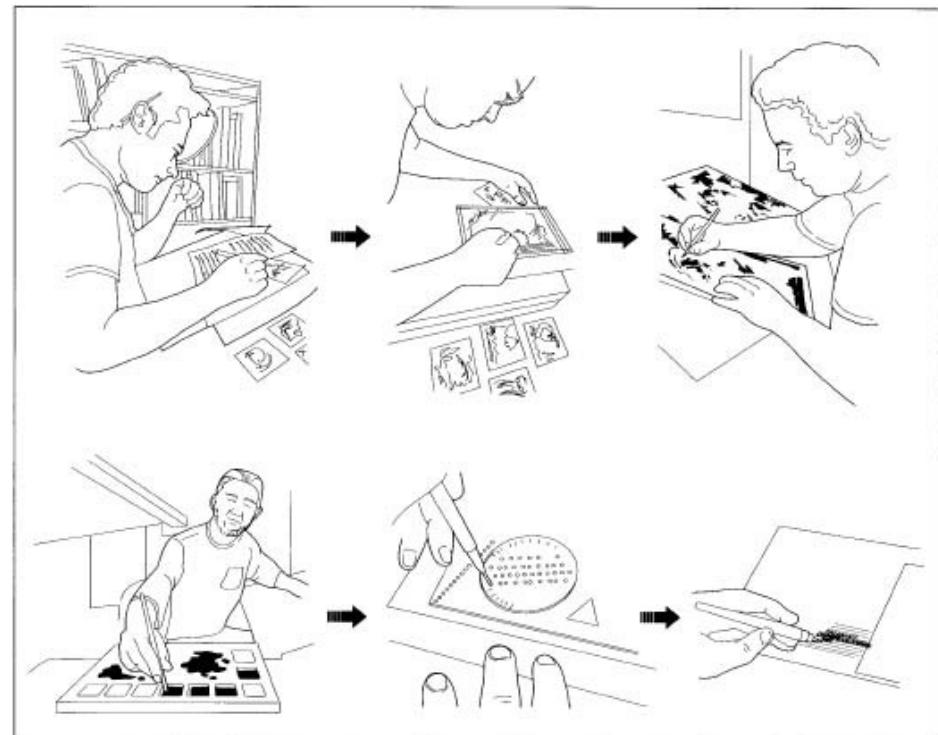


Handbook of Document Image Processing and Recognition. Springer, 2014

Comic books production

Document analysis

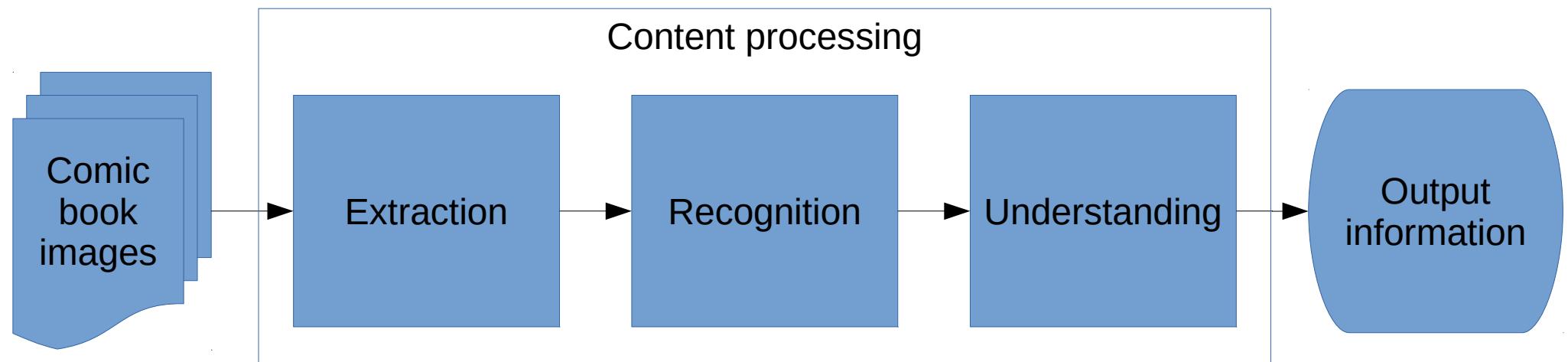
- 1) Synopsis and scenario
- 2) Pencil drawing
- 3) Inking
- 4) Flatting and colouring
- 5) Lettering and sound effects



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

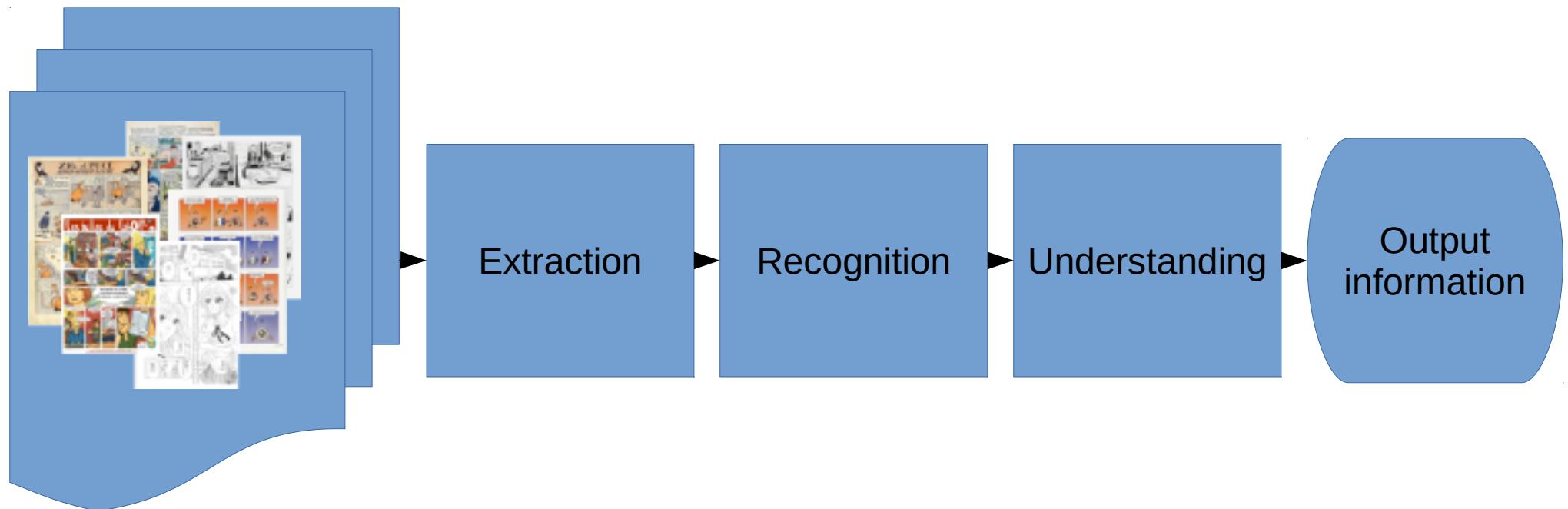
Comic books interpretation

Document analysis



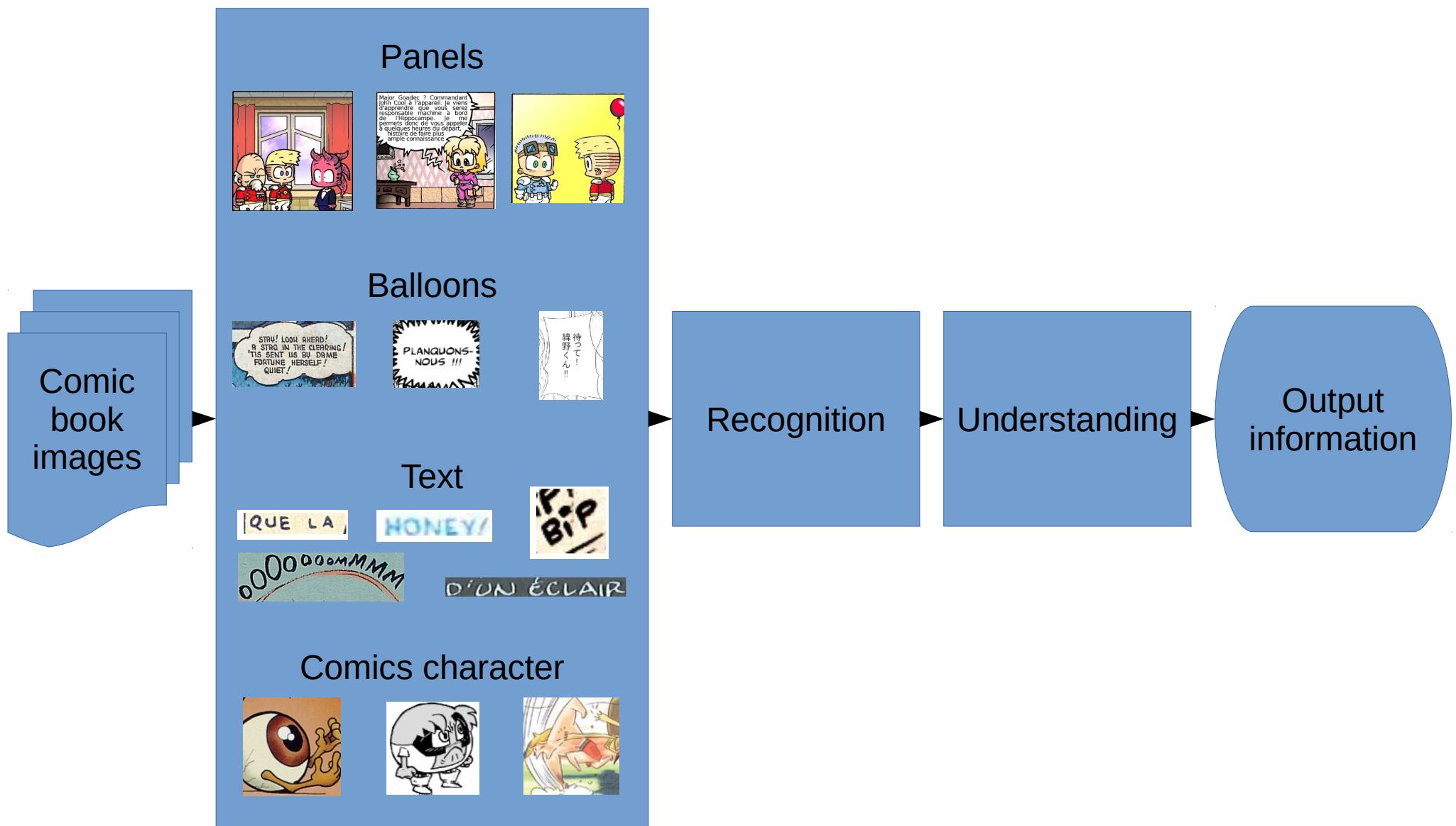
Comic books interpretation

Document analysis



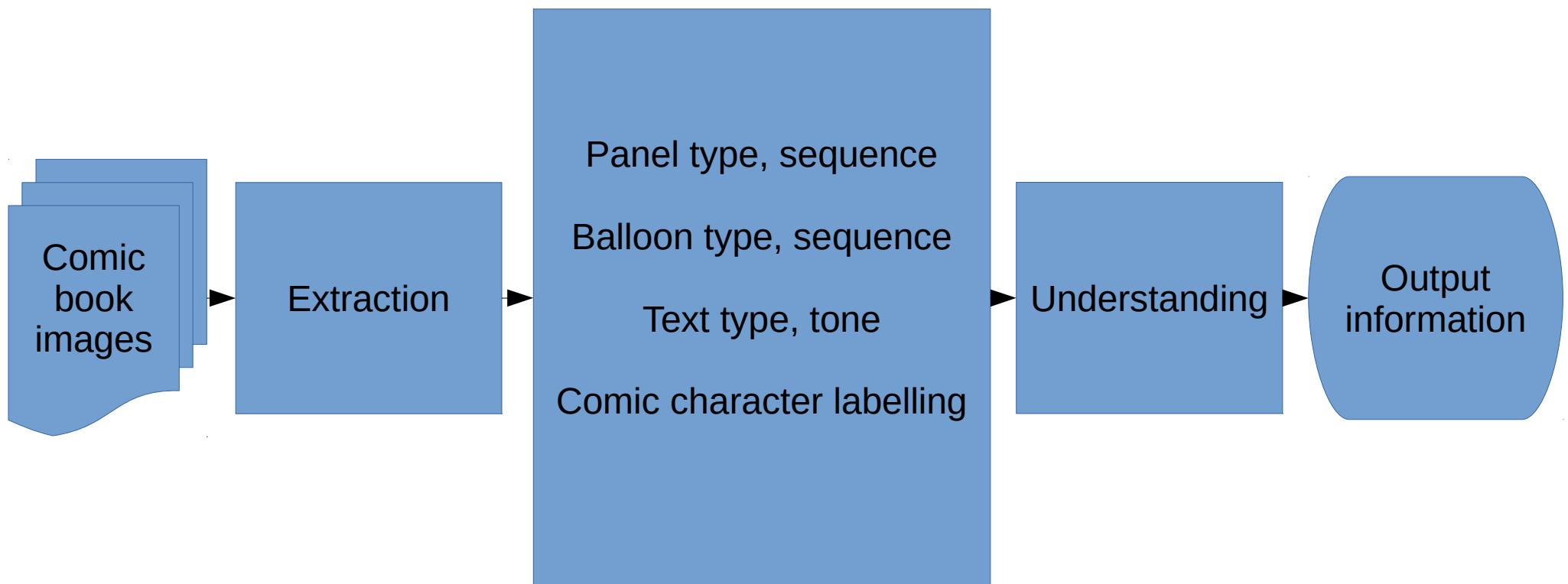
Comic books interpretation

Document analysis



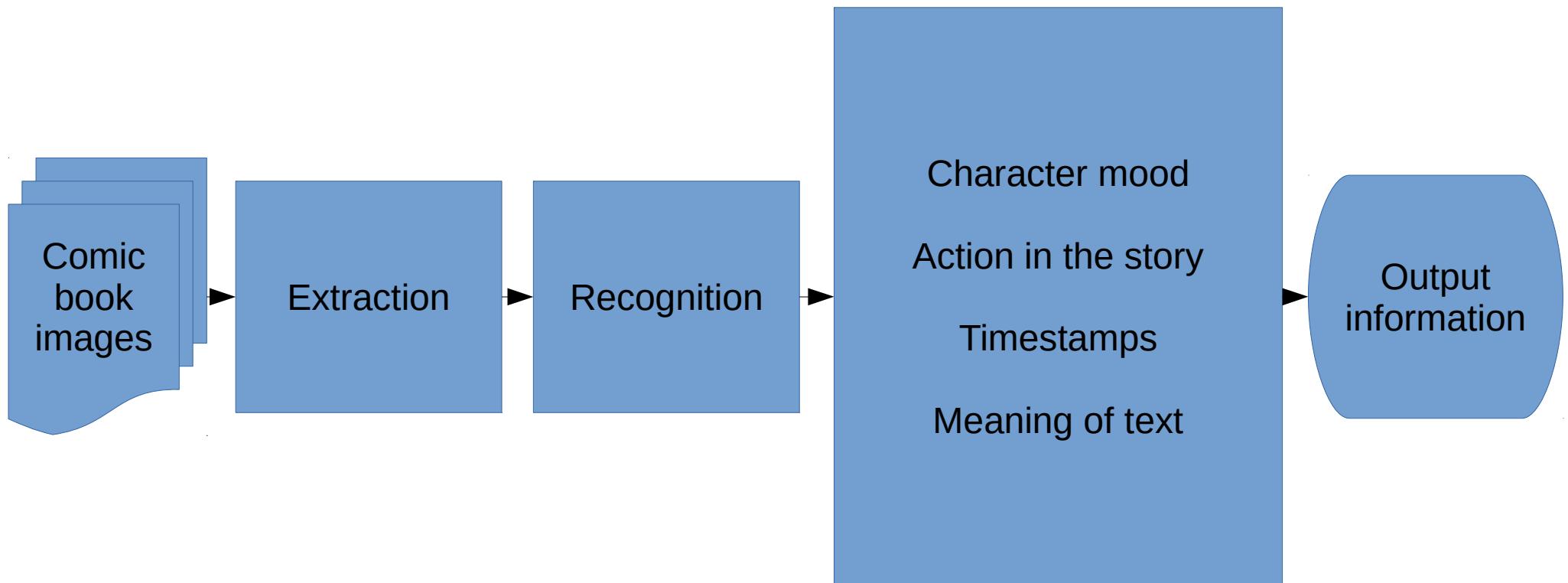
Comic books interpretation

Document analysis

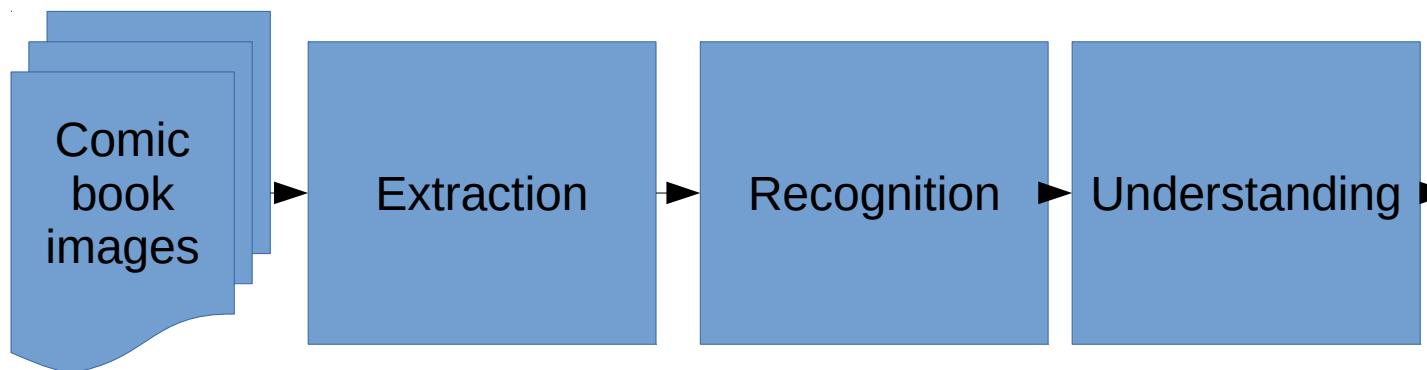


Comic books interpretation

Document analysis



Comic books interpretation



Document analysis

– IMAGE 1 –

Contains 2 panels, 2 **different** characters and 2 balloons

* PANEL 1 *

Contains 2 characters
saying 2 balloons

Character 1's **name** is "Bob"
Bob is **often represented** in blue and dark green

Character 2's **name** is "Tom"
Tom is **always represented** in yellow, purple and grey

Bob is **shouting** to Tom:
"Be quiet!!!"

Tom **answers quietly** "not yet"

* PANEL 2 *

Contains 2 characters,
same as panel 1, **not talking**
Bob and Tom are **fighting**

Outlines

- Introduction
- Document analysis
- State of the art of comics analysis
- Contributions
- Experimentations
- Conclusions

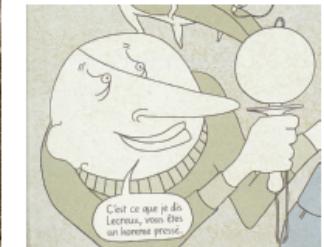
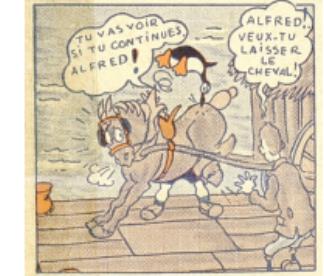
Outlines

- Introduction
- Document analysis
- State of the art of comics analysis
 - Panel extraction and layout analysis
 - Balloons analysis
 - Text extraction and recognition
 - Comic character detection and recognition
 - Holistic understanding
- Contributions
- Experimentations
- Conclusions

Panel and layout analysis

State of the art

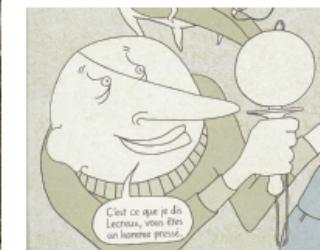
- Challenges
 - Diversity of comic books
 - Semi-structured layout
 - Panel extraction
 - White line cut
 - Recursive X-Y cut algorithm
 - Gradient
 - Connected-components
 - Polygon detection
 - Corners and line segments
 - Layout
 - Reading order (Z-path)
 - Conclusions
 - Problem solved for common ..
 - Remaining difficulties are for ..
 - No approach tested over all ..



Panel and layout analysis

State of the art

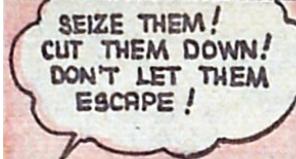
- Challenges
 - Diversity of comic books
 - Semi-structured layout
 - Panel extraction
 - White line cut
 - Recursive X-Y cut algorithm
 - Gradient
 - Connected-components
 - Polygon detection
 - Corners and line segments
 - Layout
 - Reading order (Z-path)
 - Conclusions
 - Problem solved for common
 - Remaining difficulties are for
 - No approach tested over all



Balloon analysis

- Challenges
 - Diversity of balloons intra/inter comics
 - Implicit balloons
 - Interface between text and graphics
- Extraction
 - Shape vs contour
 - Blob detection [Arai 2011, Ho 2012]
- Classification
 - Speech tone information (contour)
- Tail detection
 - Indicate the position of the emitter
- Conclusions
 - Closed balloon solved (sequential)
 - Implicit, classification and tail were not explored

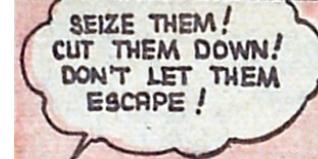
State of the art

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

Balloon analysis

- Challenges
 - Diversity of balloons intra/inter comics
 - Implicit balloons
 - Interface between text and graphics
- Extraction
 - Shape vs contour
 - Blob detection [Arai 2011, Ho 2012]
- Classification
 - Speech tone information (contour)
- Tail detection
 - Indicate the position of the emitter
- Conclusions
 - **Closed** balloon solved (sequential)
 - **Implicit, classification** and **tail** were not explored

State of the art

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

Text extraction and recognition

State of the art

- Challenges
 - Non-standard fonts
 - Multi-script/orientation/scale
 - Complex background (sound effects)
 - Short length, hyphenation
 - Voluntary spelling mistakes
 - Extraction
 - Scene text localization
 - Connected-components approach
 - SVM and Bayesian classifier
 - Sound effects have not been investigated yet
 - Recognition
 - At is early stage
 - OCR trained for a specific comics font



Text extraction and recognition

State of the art



- Conclusions
 - Speech text studied but not solved
 - **Captions and sound effects** unexplored
 - Text recognition not usable yet
 - (Next: automatic font learning?)

Comic character extraction

State of the art

- Challenges

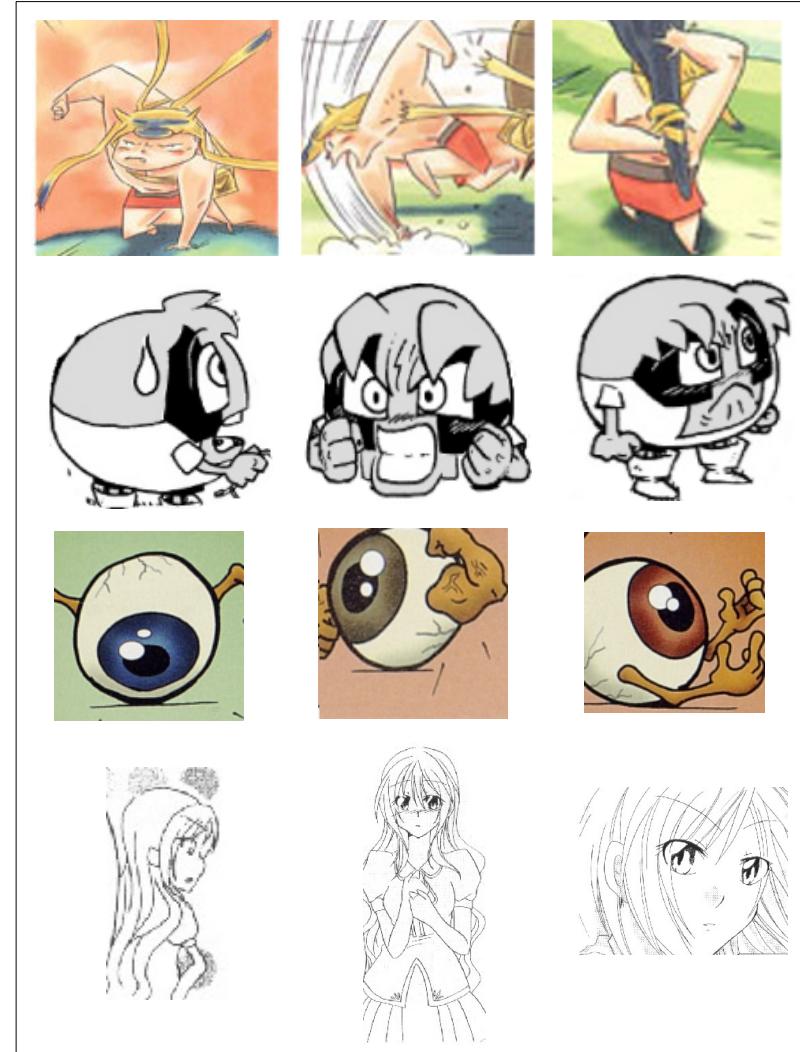
- Hand-drawn, stroke-based
- Intra/inter class variability
- Scale, deformation, posture, occlusion

- Extraction & recognition

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

- Conclusions

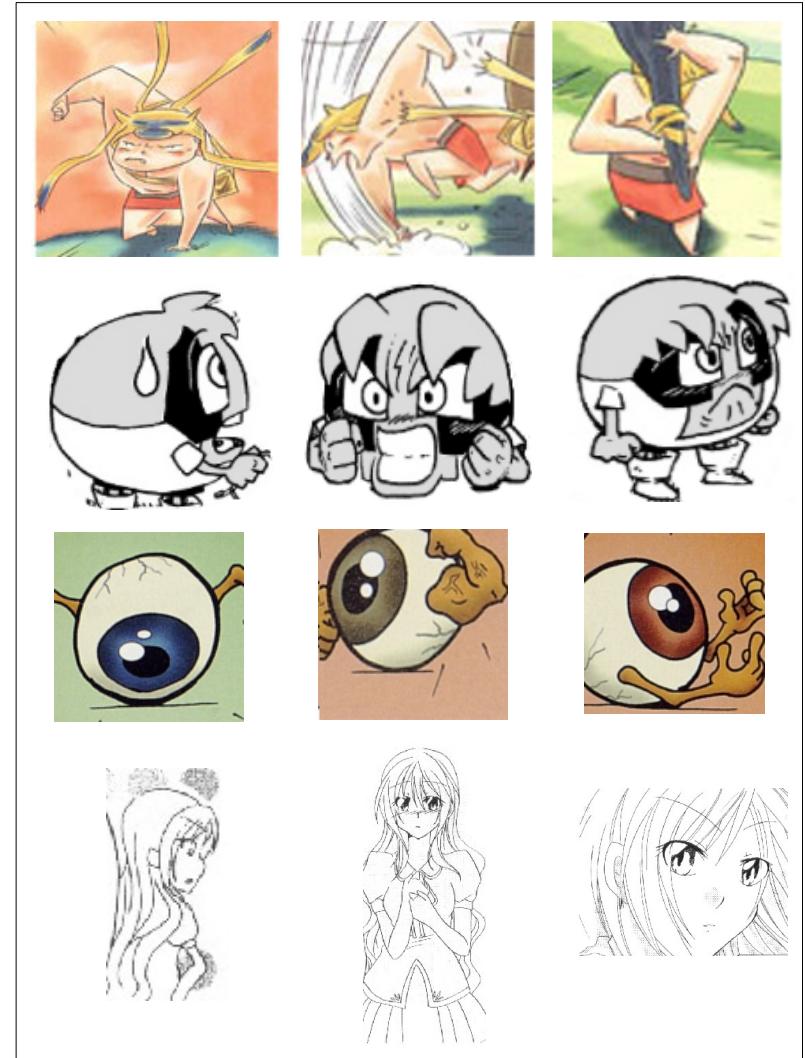
- Preliminary results
- Complex and versatile structure
- Contains most of the interesting information



Comic character extraction

State of the art

- 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



Outlines

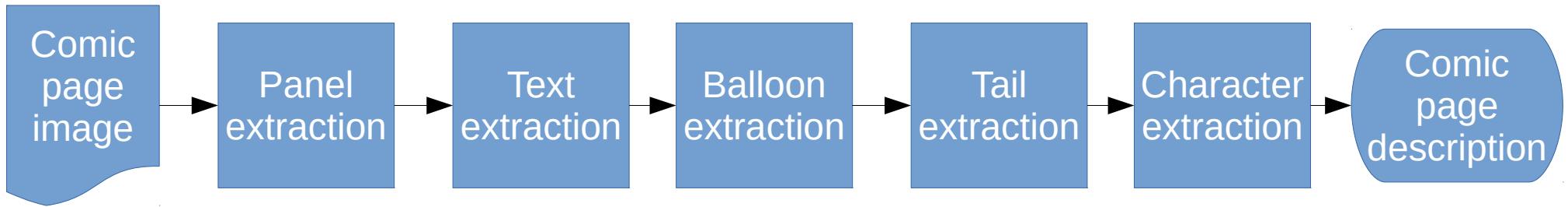
- Introduction
- Document analysis
- State of the art of comics analysis
- Contributions
- Experimentations
- Conclusions

Outlines

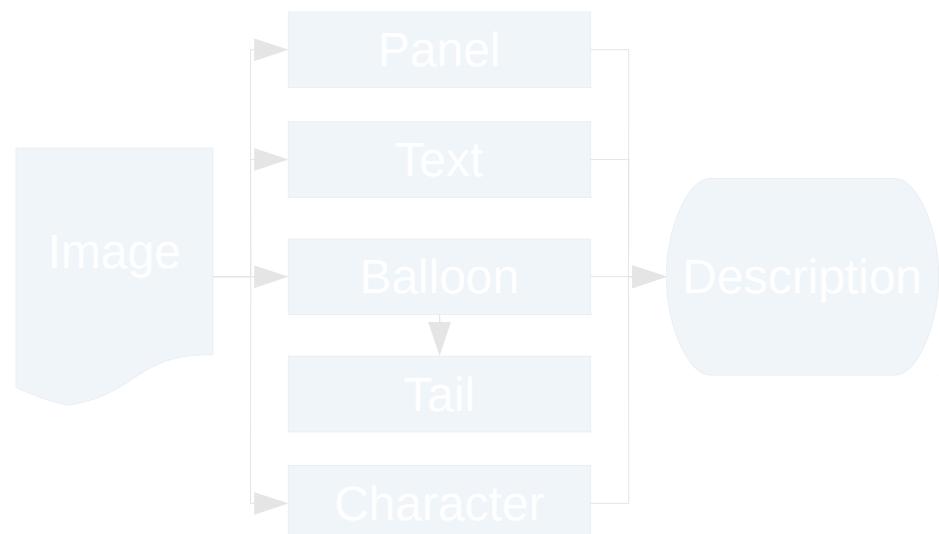
- Introduction
- Document analysis
- State of the art of comics analysis
- Contributions
 - Introduction
 - Content-driven approach
 - Knowledge-driven approach
 - Conclusion
- Experimentations
- Conclusions

Introduction

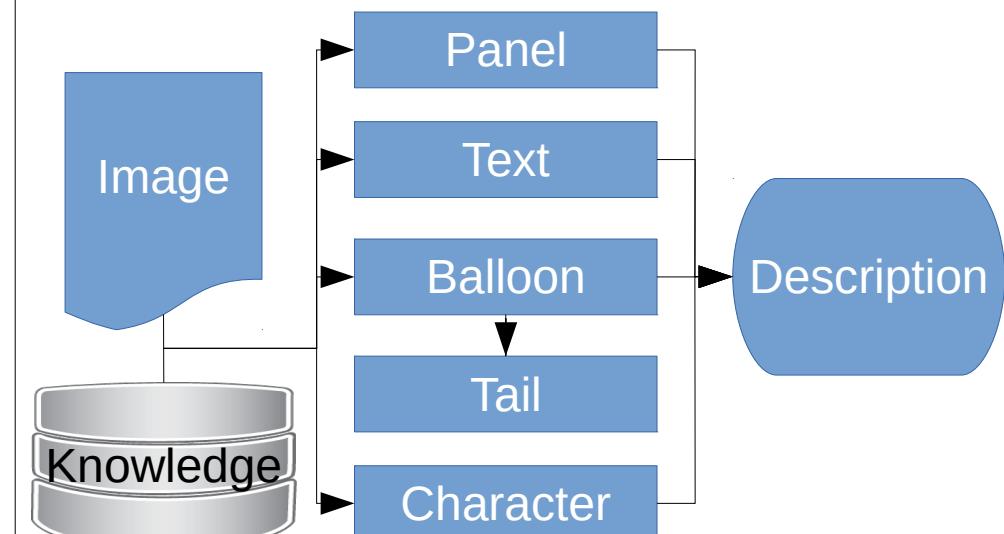
Content-driven (sequential)



Independent



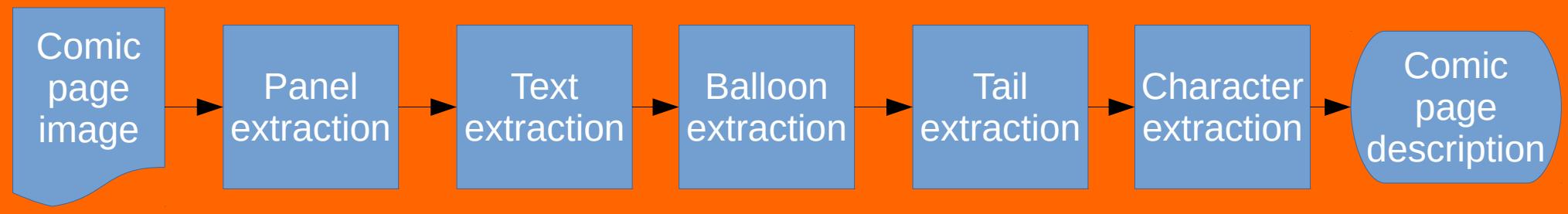
Knowledge-driven



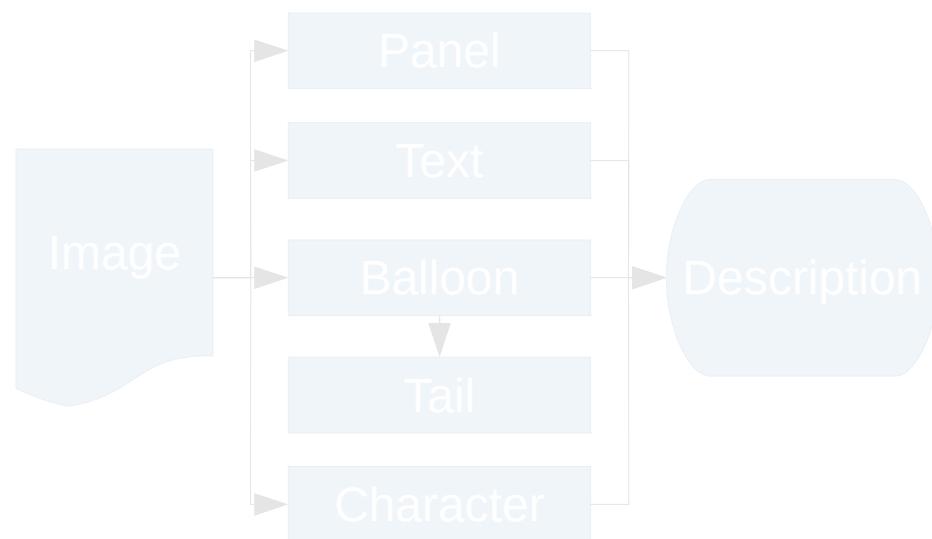
Introduction

Content-driven

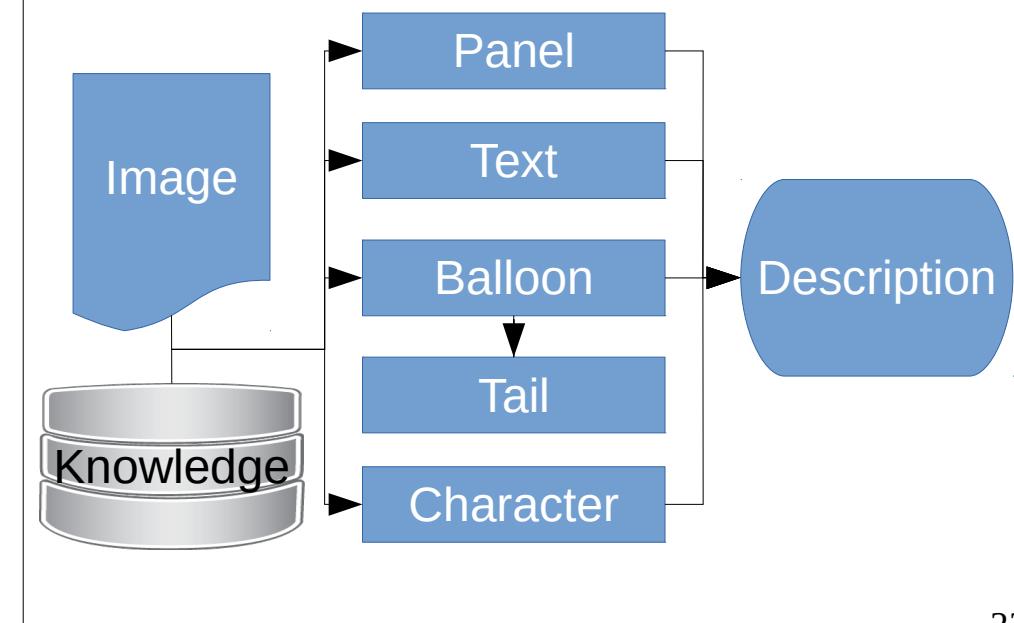
Content-driven (sequential)



Independent

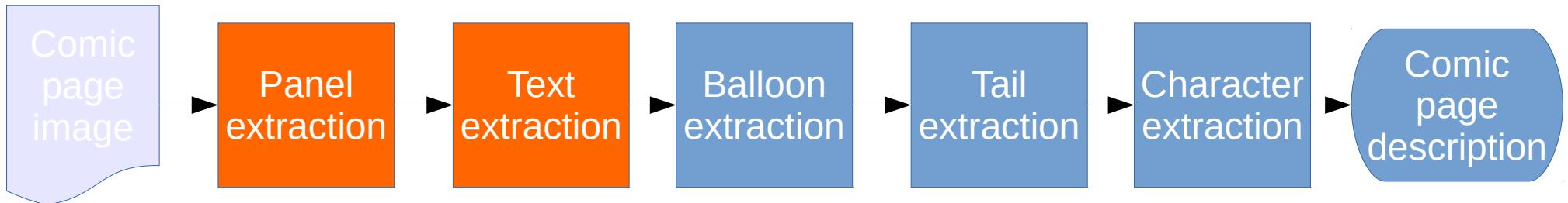


Knowledge-driven

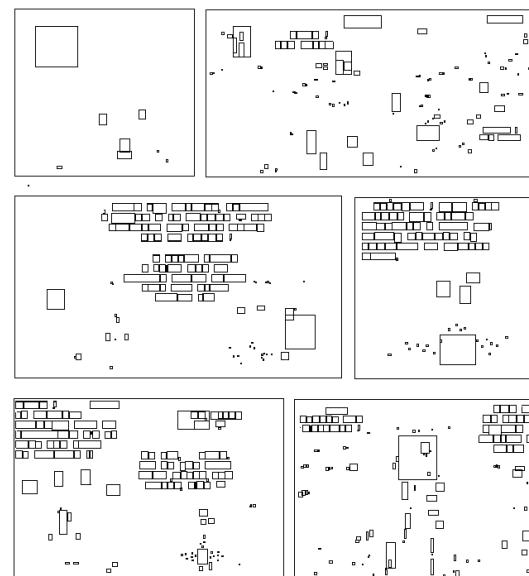


Panel and text extraction

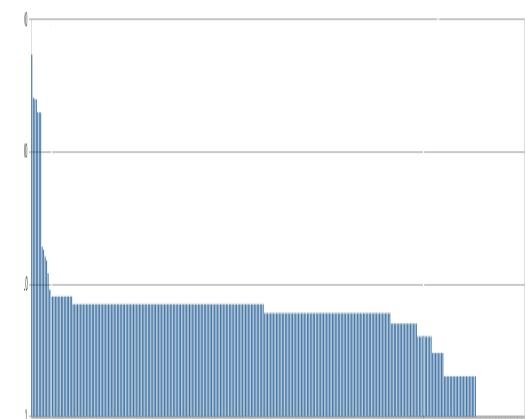
Content-driven



Binary image



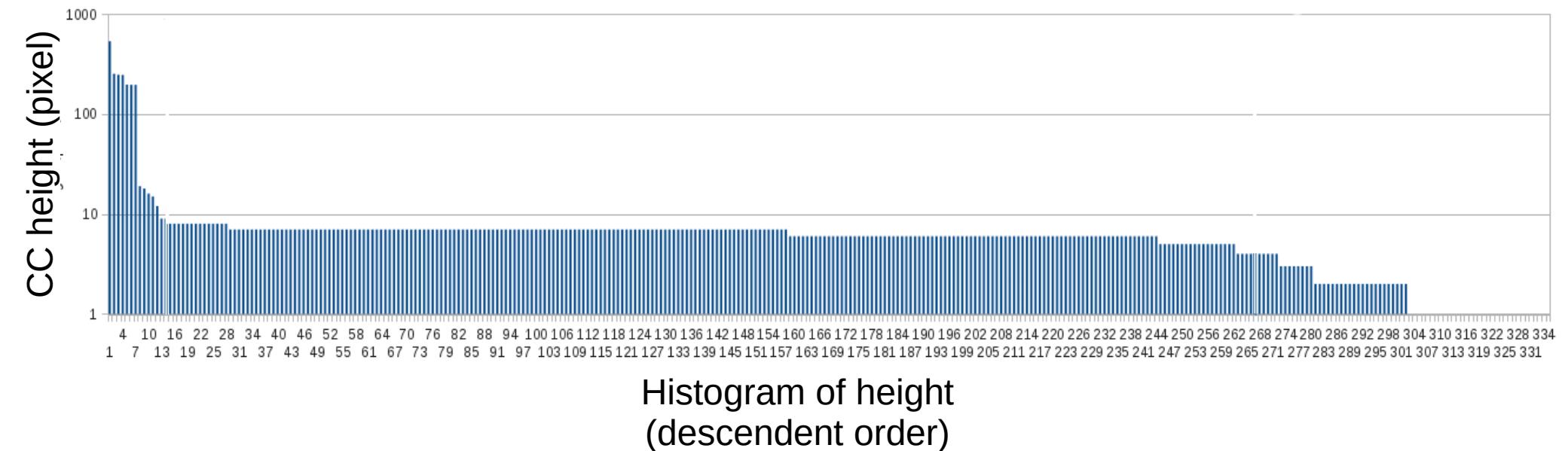
Connected-component (CC) bounding boxes



Histogram of heights of CC

Panel and text extraction

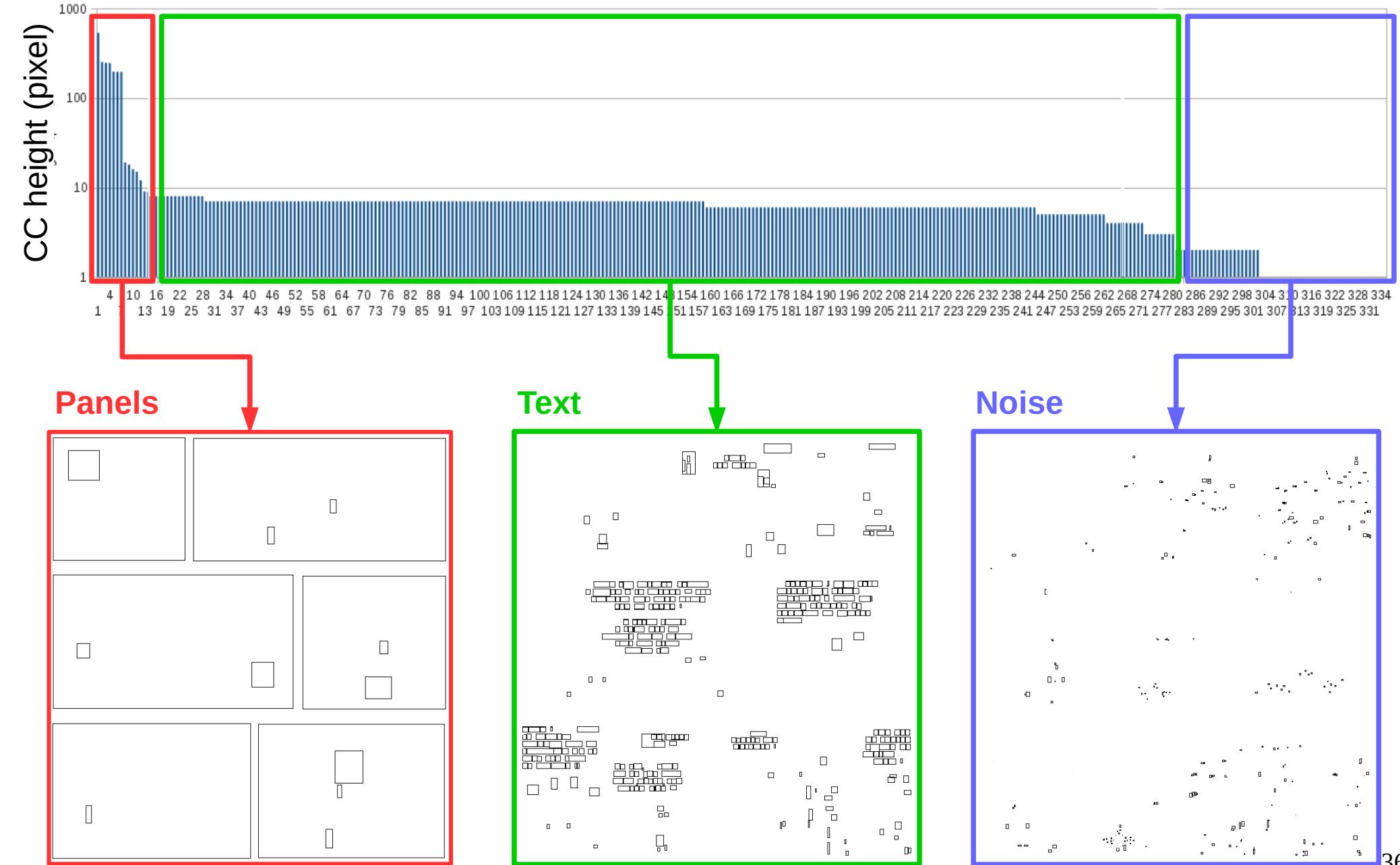
Content-driven



Panel and text extraction

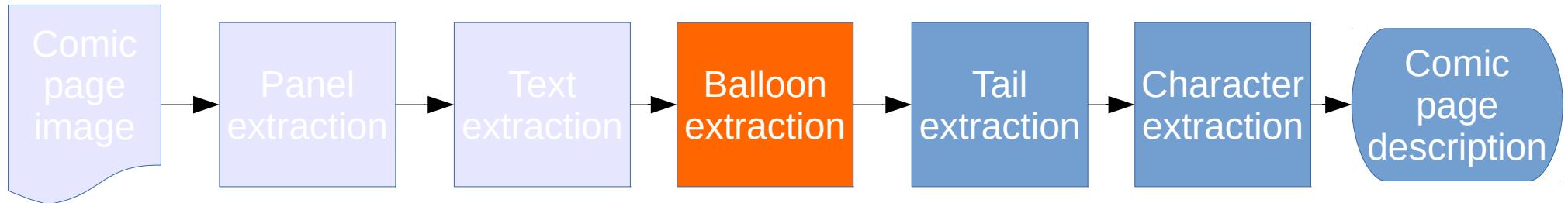
Content-driven

K-means clustering ($k=3$)



Balloon extraction

Content-driven



Regular balloon

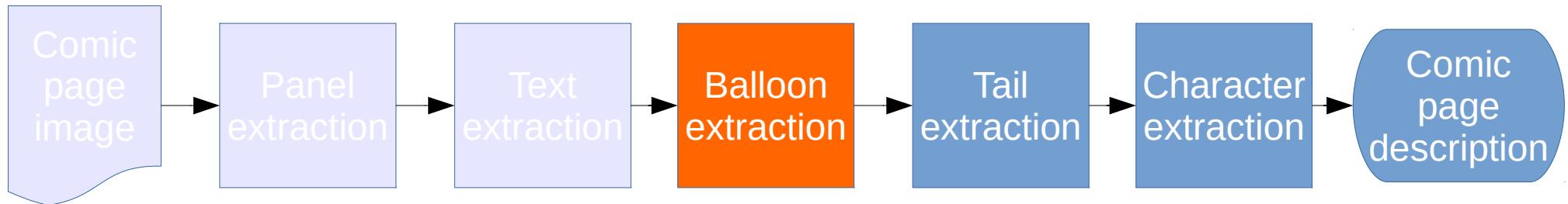


Implicit balloon

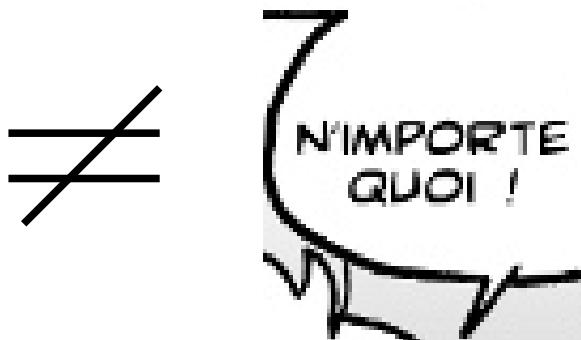
- Literature
 - Top-down approaches: extract white blobs and then text inside
 - Limited to regular balloons
- Contribution
 - Bottom-up approaches: extract text and then surrounding balloons
 - Appropriate for regular and implicit balloons

Balloon extraction

Content-driven



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

Regular balloon extraction

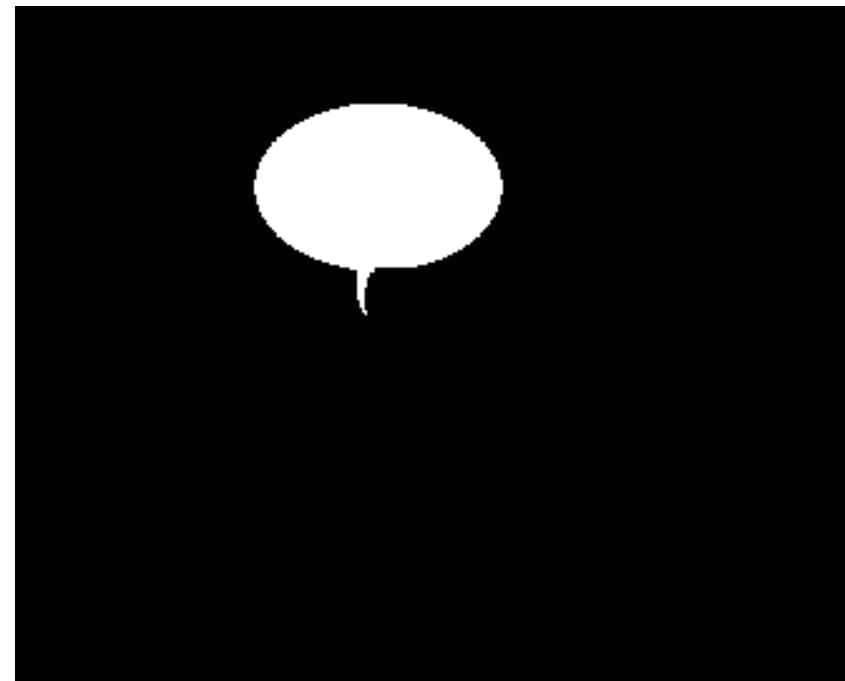
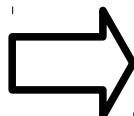
Content-driven

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

Regular balloon extraction

Content-driven

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



Original image

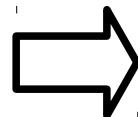
Expected result

Regular balloon extraction

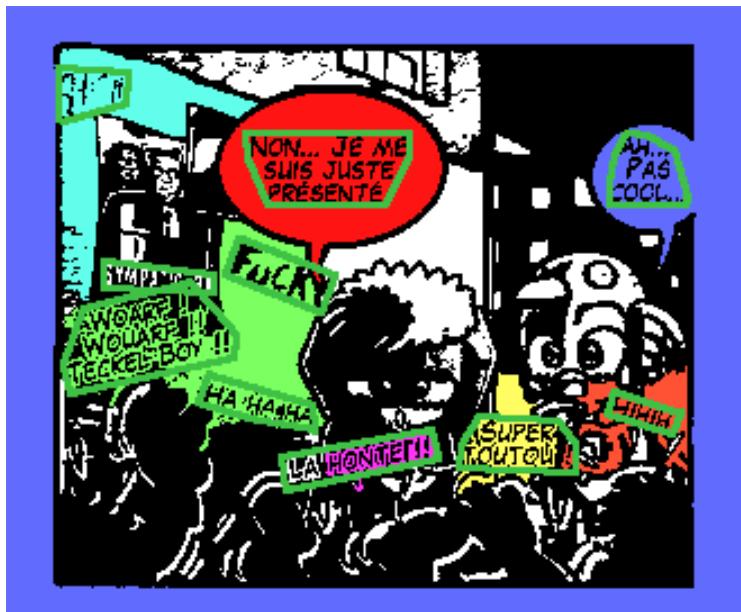
Content-driven



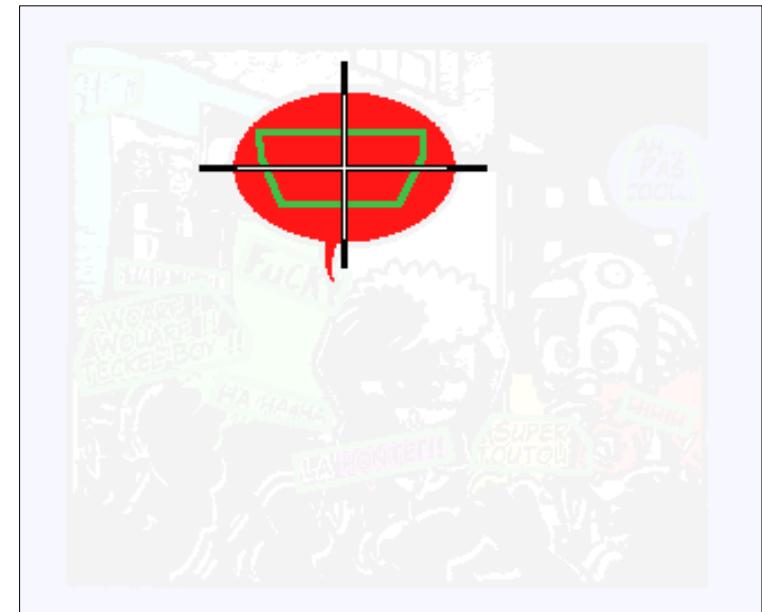
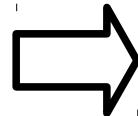
Original image



Text block positions (green)



Regions including text blocks (coloured)



Regions including aligned text block

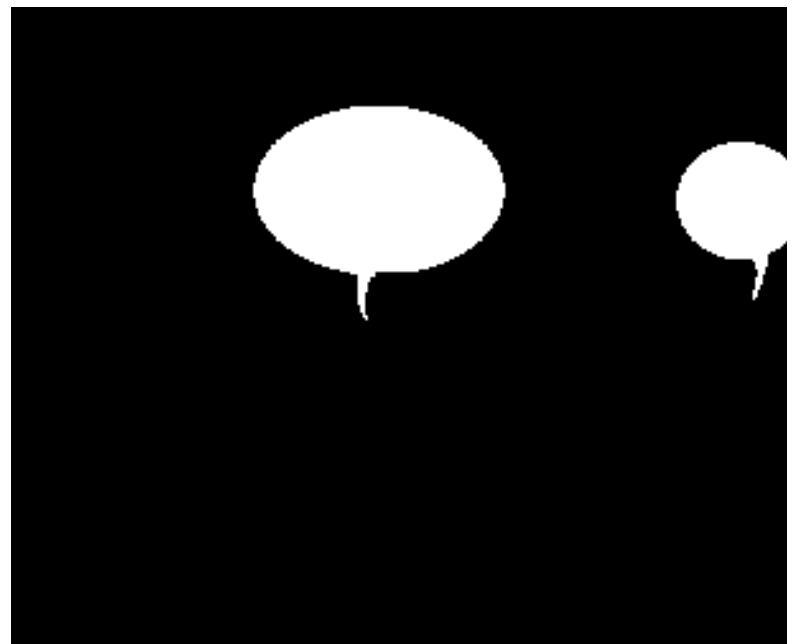
Implicit balloon extraction

Content-driven

- Assumptions
 - Panels and text block positions **are known**
 - Balloons **contain** text
 - Text is **fully contained** and **centred** in balloons
- Proposition → active contour algorithm
 - Extract **implicit contours** from **text regions**

Energy function:

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

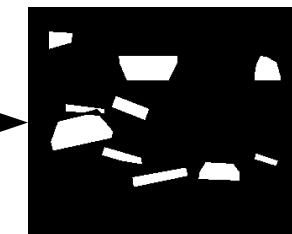
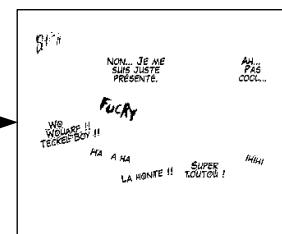
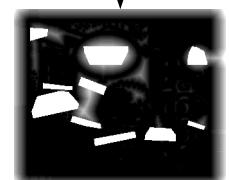
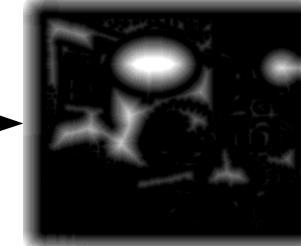


Original image

Expected result

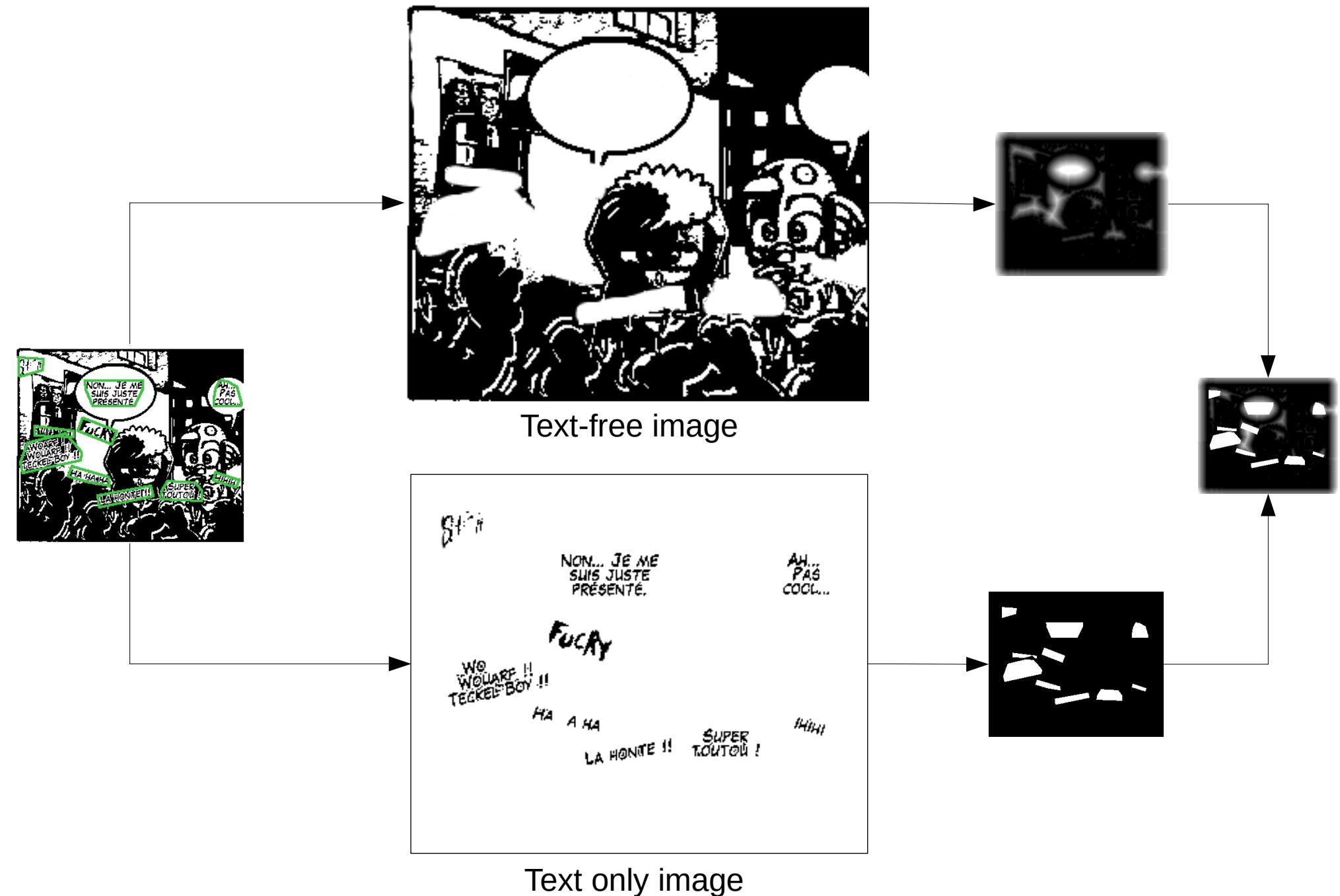
Implicit balloon extraction

Content-driven



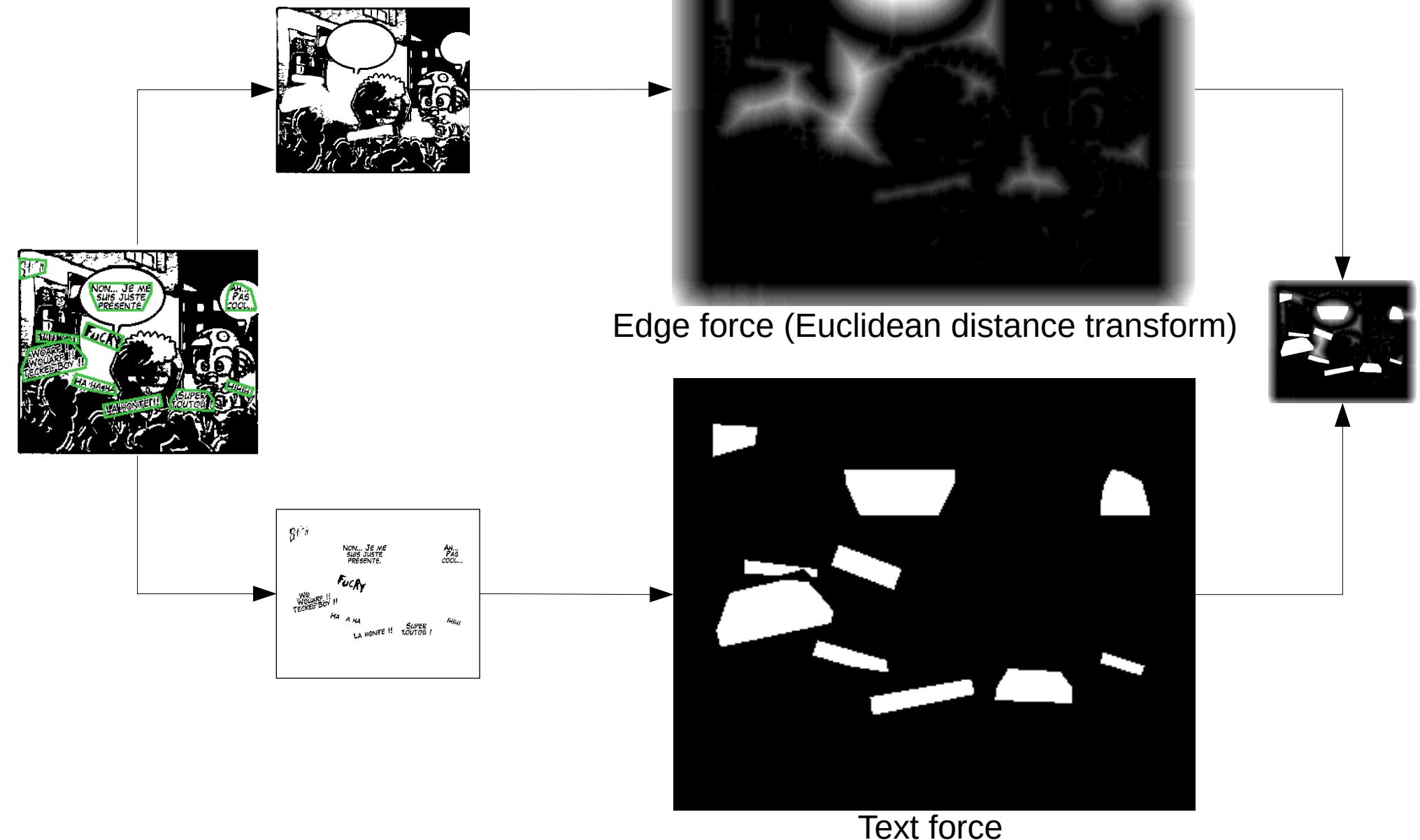
Implicit balloon extraction

Content-driven



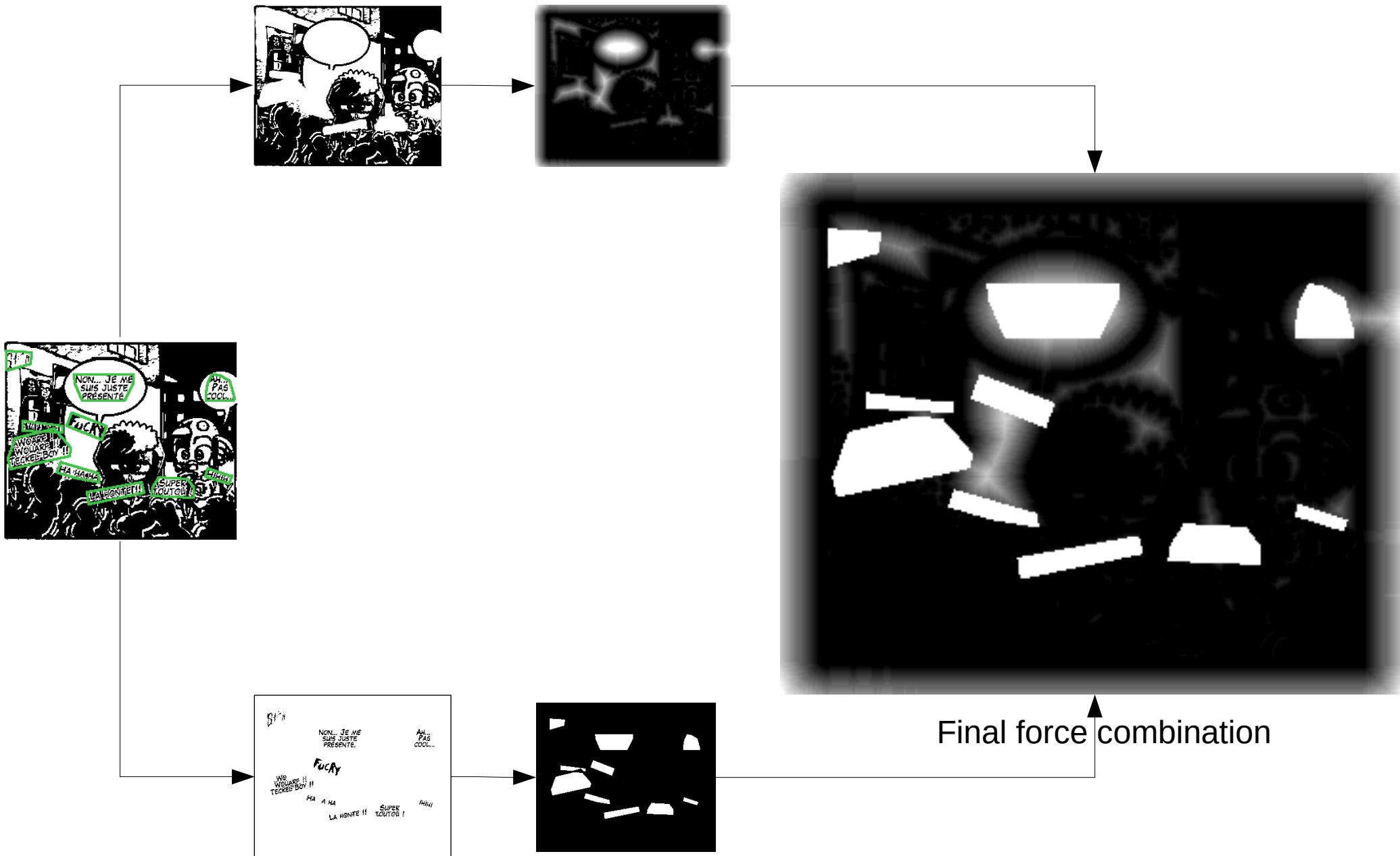
Implicit balloon extraction

Content-driven



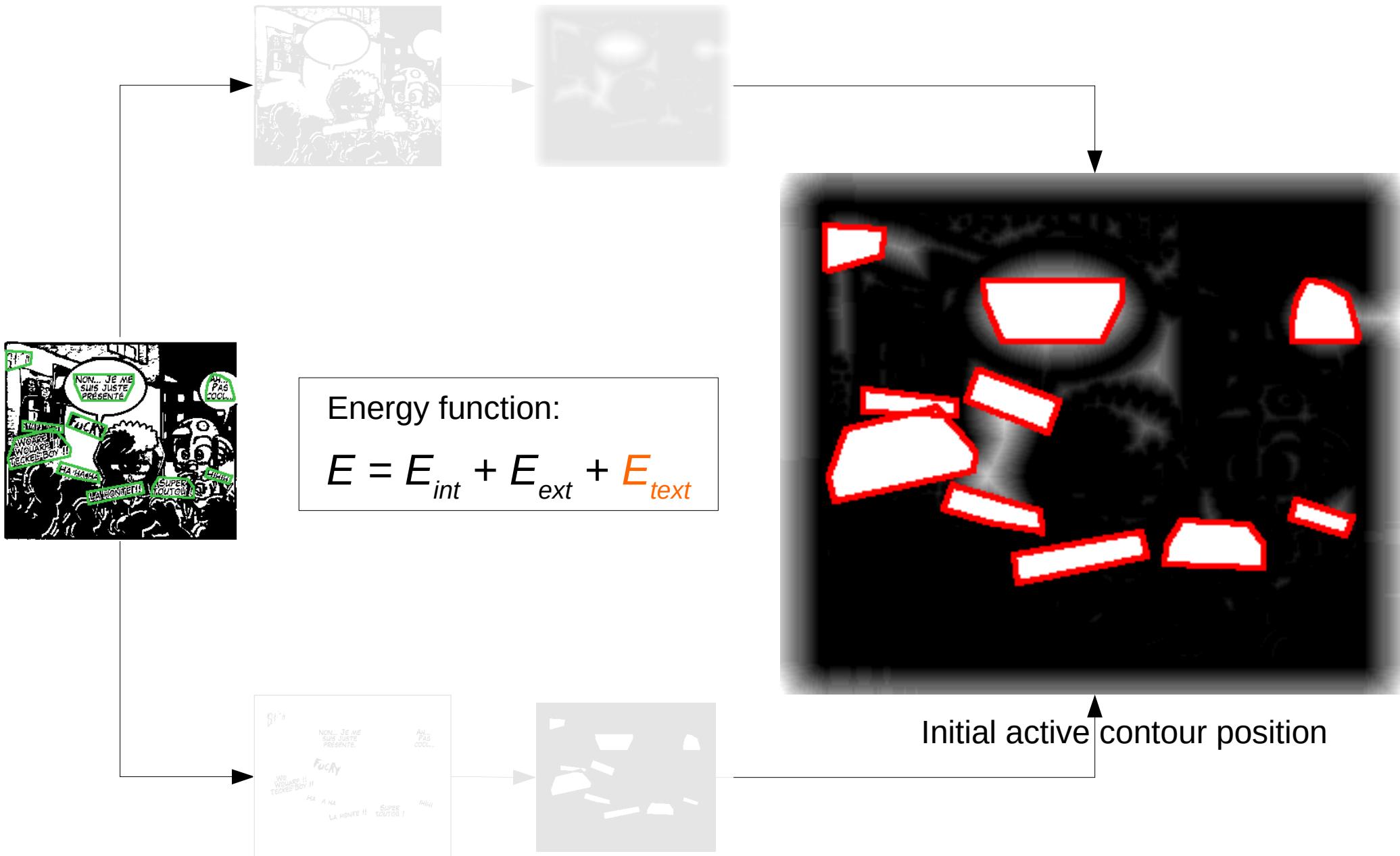
Implicit balloon extraction

Content-driven



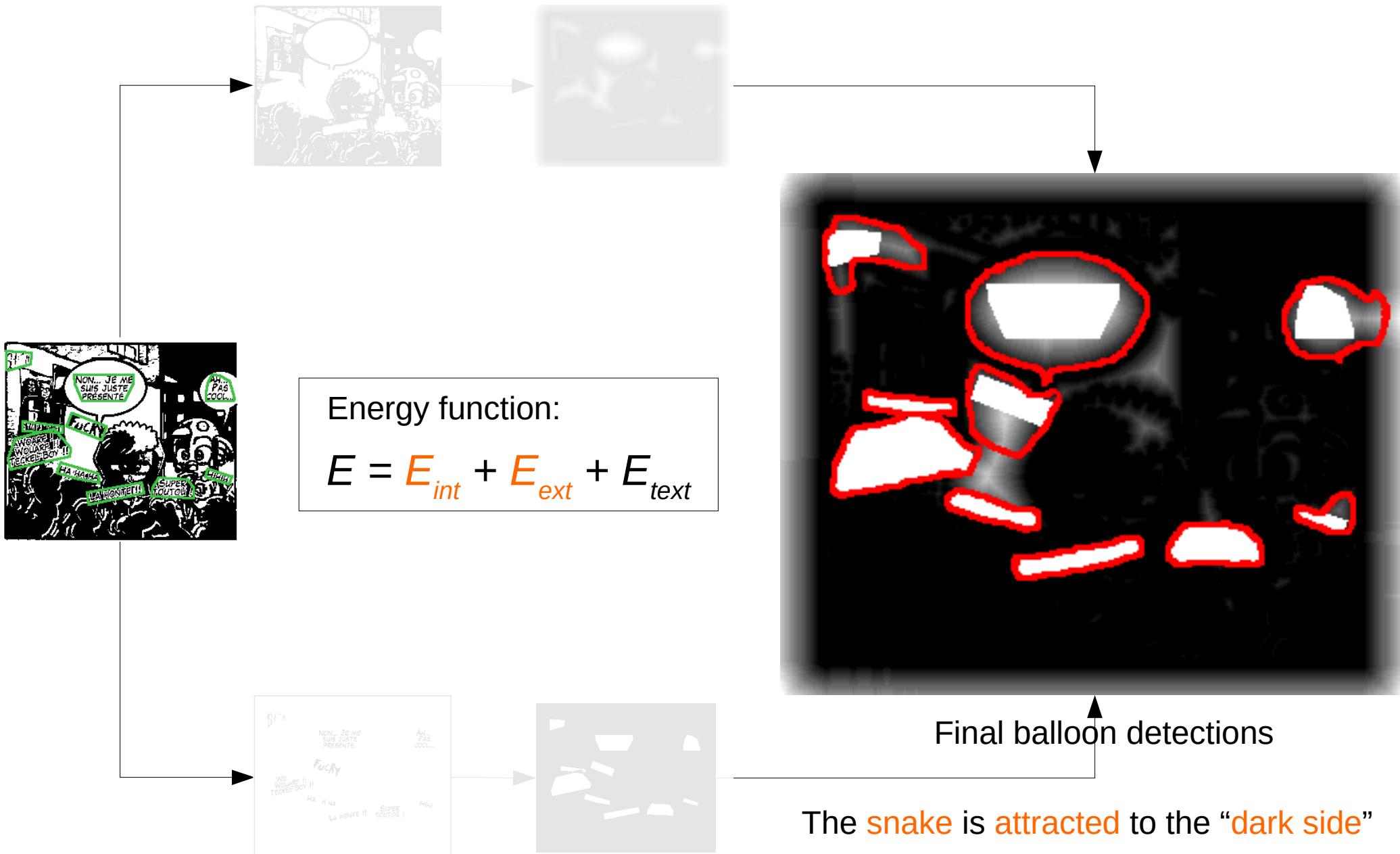
Implicit balloon extraction

Content-driven



Implicit balloon extraction

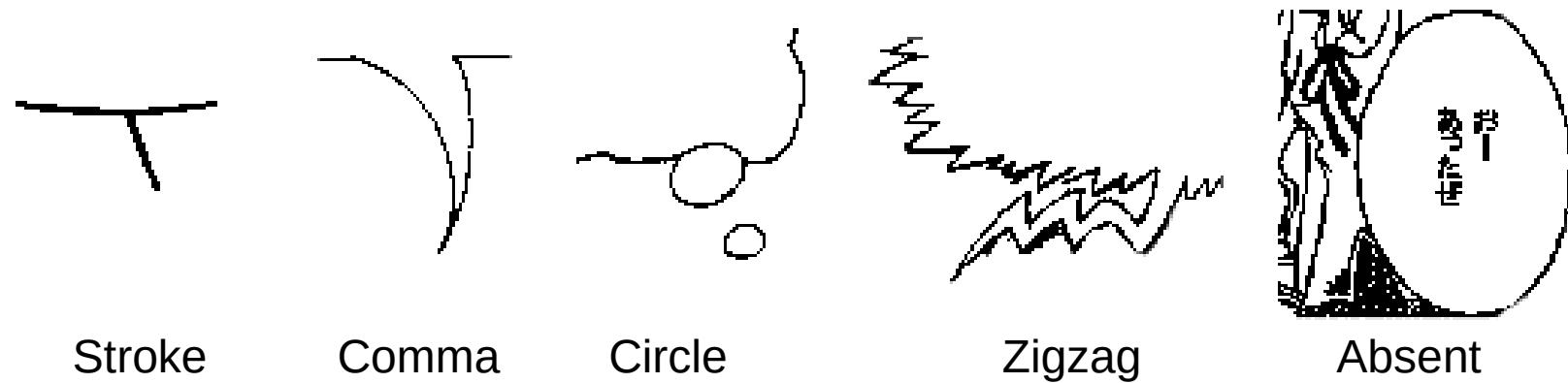
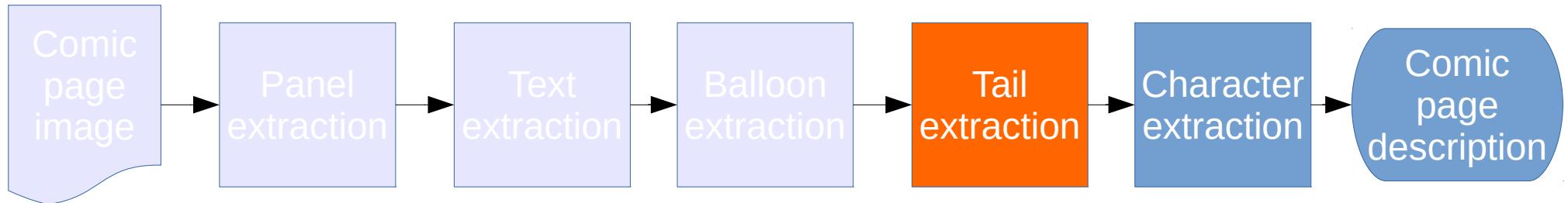
Content-driven



- General conclusion
 - Relies on the performance of text extraction
 - First bottom-up approaches for regular and implicit balloons
- Regular balloons
 - Accurate pixel-level extraction (appropriate for contour analysis)
 - Multi-script according to the text extraction abilities
 - Not suitable for implicit balloons
- Implicit balloons
 - First time studied
 - Appropriate for all type of balloons containing text (regular and implicit)
 - Over-detection if not only speech text blocks are given as input
 - Not able to extract details on the contour (tail, peaks, etc.)
 - Time consuming (active contour)

Tail extraction

Content-driven



- Literature
 - First time studied in image processing
- Objectives
 - Detection of the tail tip position and orientation
 - Method for comma, zigzag and absent types

Tail extraction

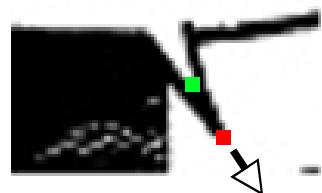
Content-driven

- Challenges

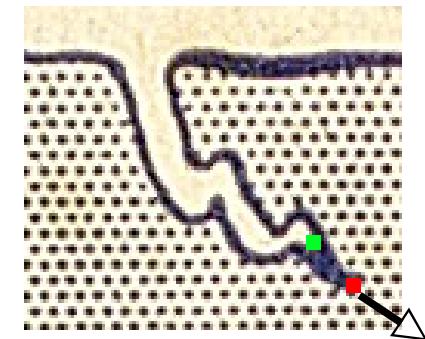
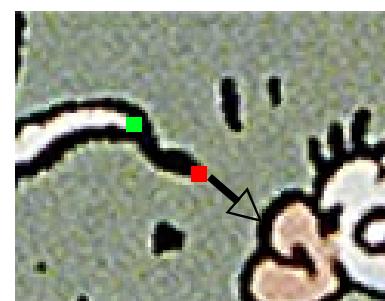
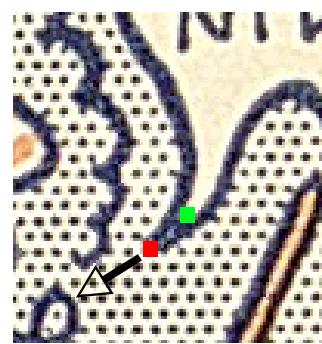
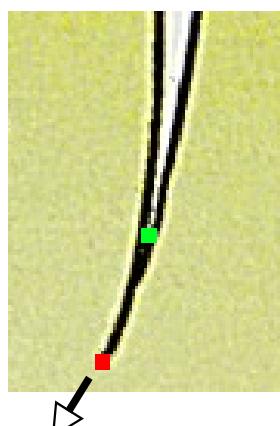
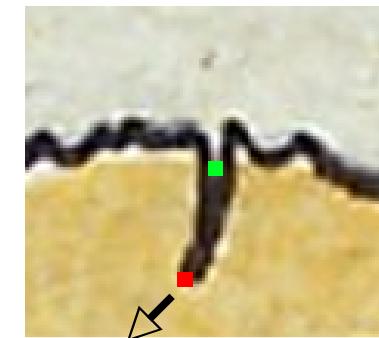
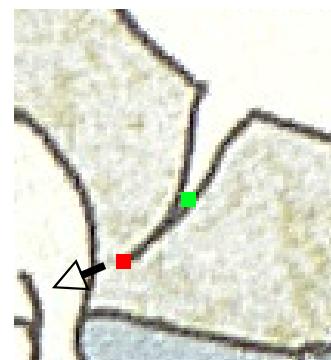
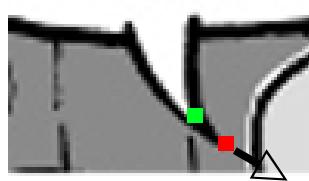
- Tail **tip** position definition
- Tail **direction** definition

■ Tip from background
■ Tip from contour
→ Direction of the tail

MENCE...



...MENCE



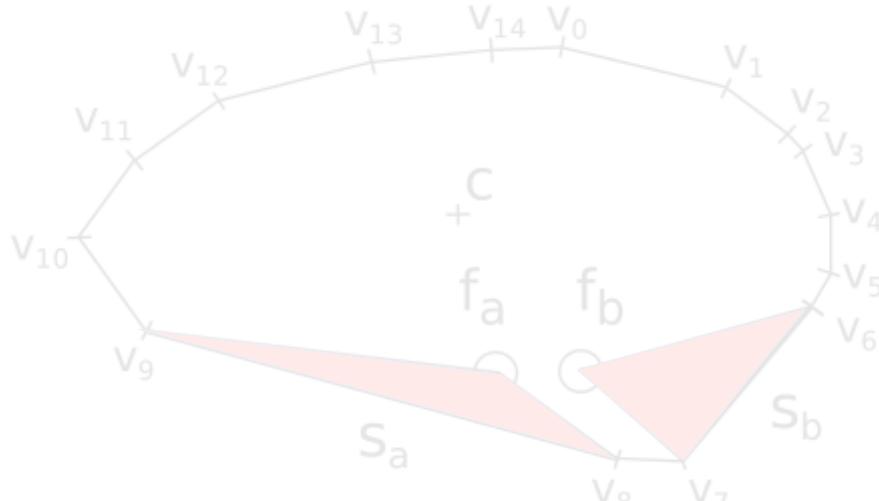
Different interpretation of tail tip positions and directions

Tail extraction: tip position

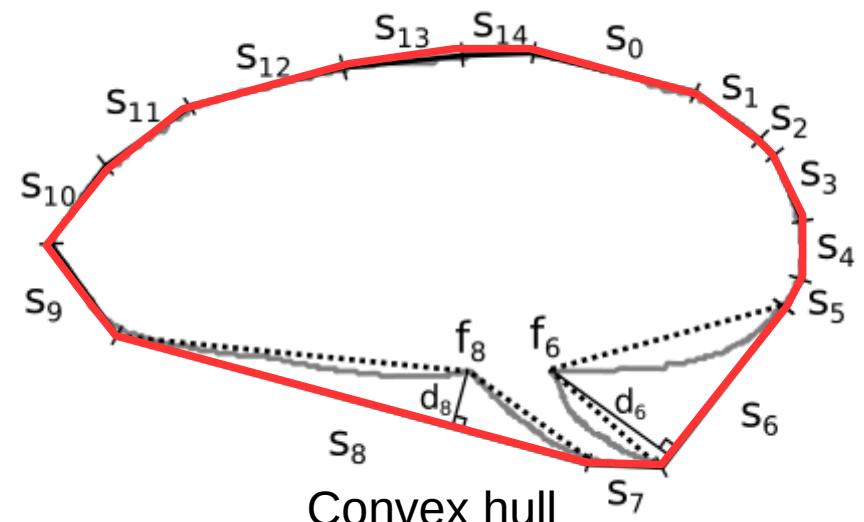
Content-driven



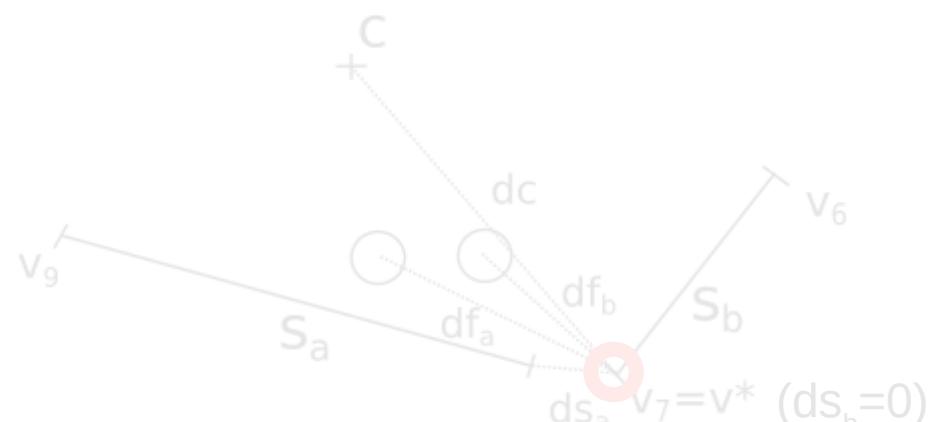
Balloon contour



Two biggest
convexity defects



Convex hull



Tail tip position

Optimal vertex selection:

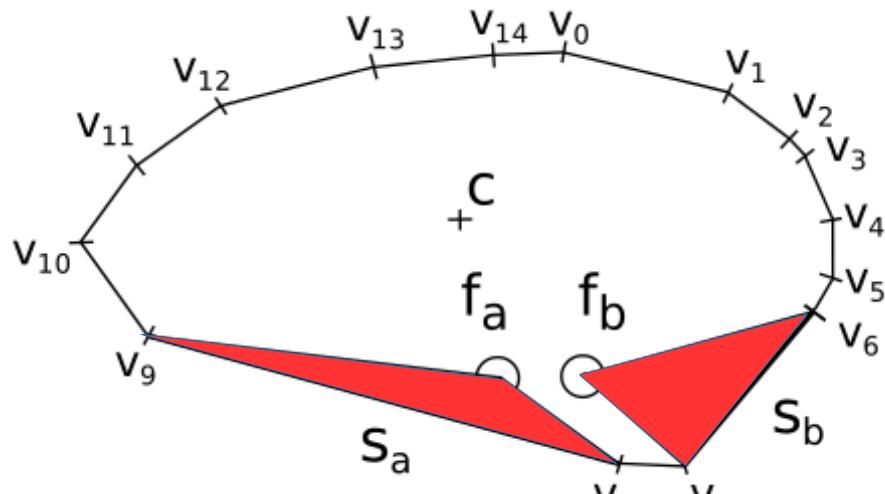
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

Tail extraction: tip position

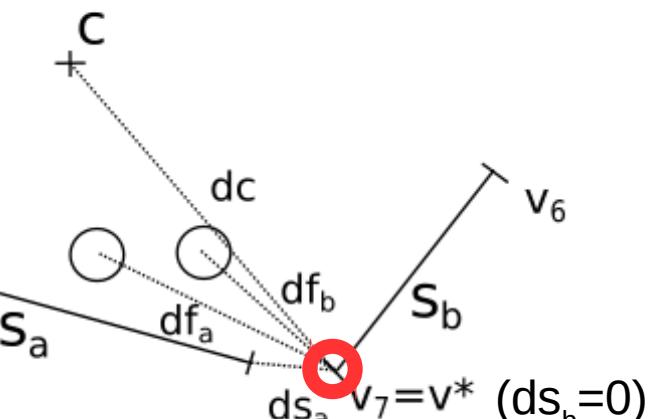
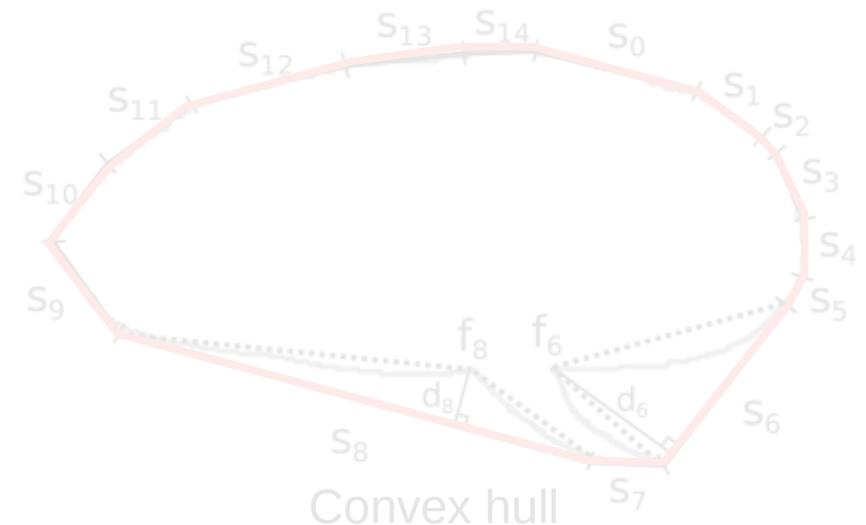
Content-driven



Balloon contour



Two biggest
convexity defects



Tail tip position

Optimal vertex selection:

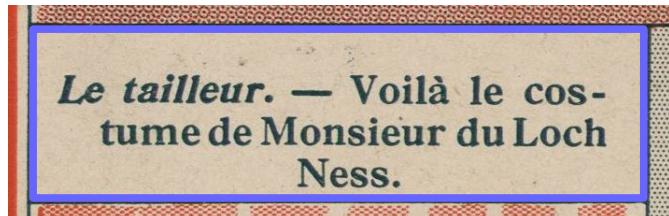
$$v^* = \operatorname{argmax}(\max(dc + df_a + df_b) + \min(ds_a + ds_b))$$

Tail extraction: confidence value

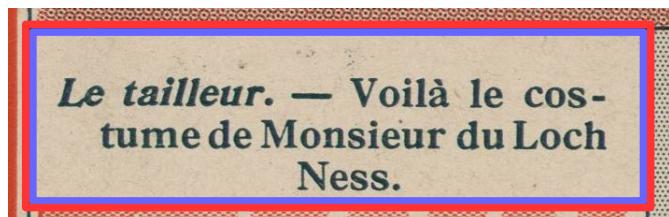
Content-driven

Balloon
contour (blue)

Balloon 1



Convex hull
(red)



Confidence

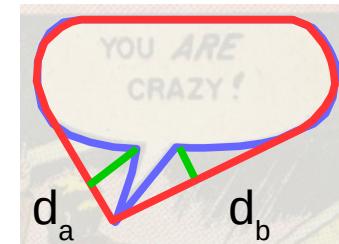
$$C_{tail} = \frac{(d_a + d_b)/2}{meanBalloonSize}$$

$$C_{tail} = 0.0$$

Presence of tail

NO

Balloon 2



$$C_{tail} = 0.73$$

YES (>0)

Tail extraction: tail direction

Content-driven

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



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



Closest point on
external edge
(red)



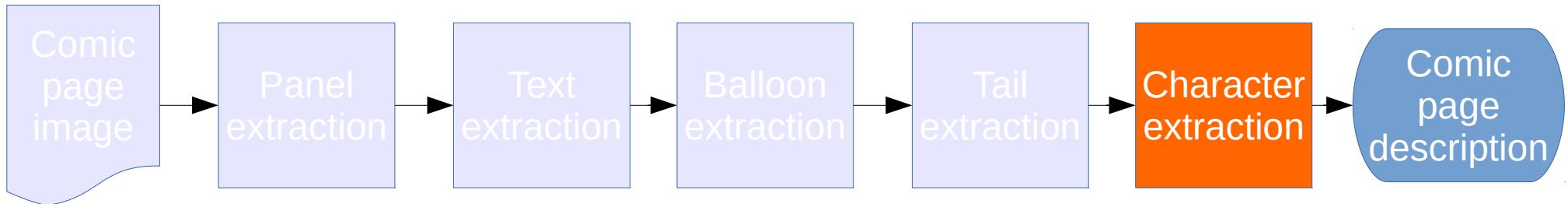
Farthest point
from origin and tip
(red)



Direction from tip
to farthest point
(white arrow)

Comic character extraction

Content-driven



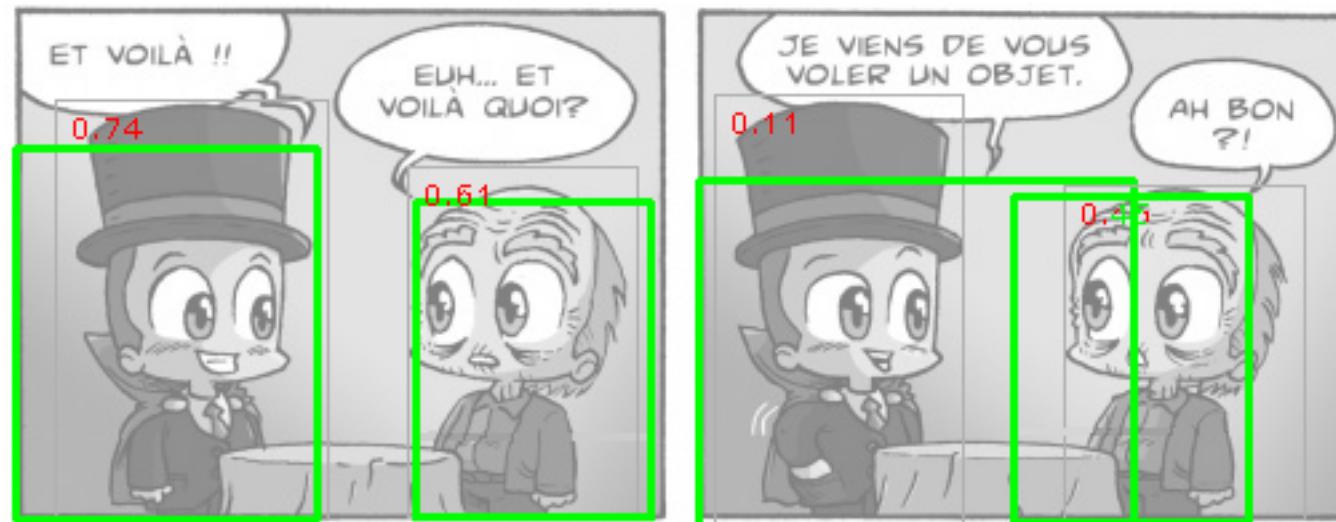
- Literature
 - Supervised approaches for manga and cartoon characters
 - No public dataset (copyright issues)
- Challenges
 - Variety of styles of comic books
 - Intra and extra classes variations of each character instance (e.g. position, scale, pose, occlusion and human-like, invented)
- Objectives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

Comic character extraction

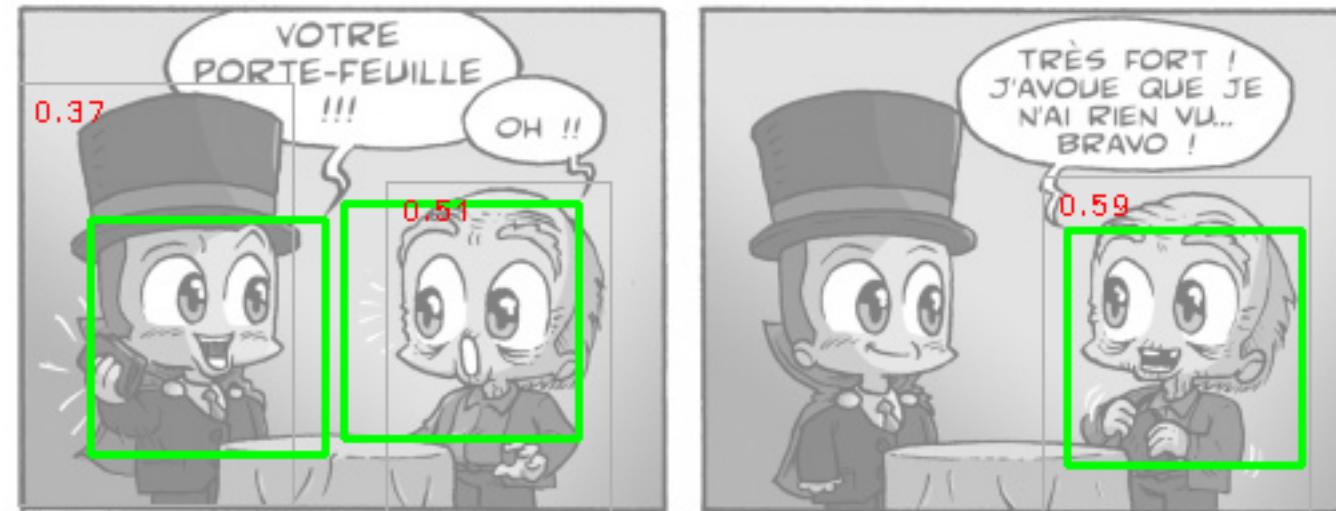
Content-driven

Our proposition: use already extracted information (panels, balloons and tails) to estimate speaking character region of interests

Rectangle ROI



Squared ROI



Comic character extraction

Content-driven

- Preliminary results
- Highly relies on the quality of the tail tip and direction
- Only for “speaking” characters
- Implicitly retrieves the relationship between balloons and characters

Conclusion of the content-driven analysis approach

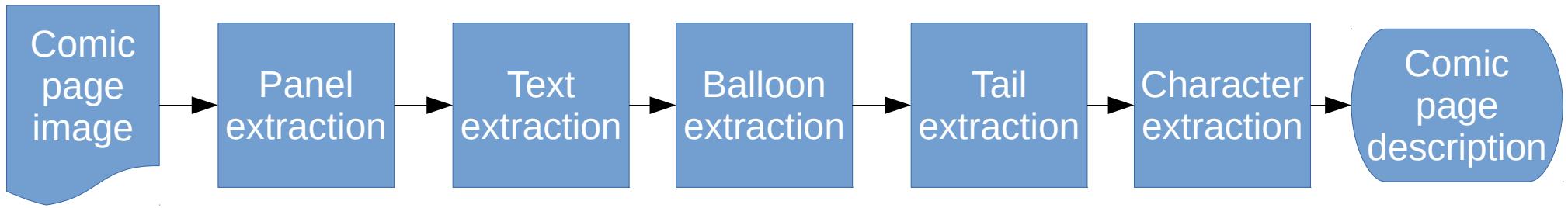
- From simple to complex elements
- Content-driven approach
- Consistence between extracted elements
- Error propagation

Outlines

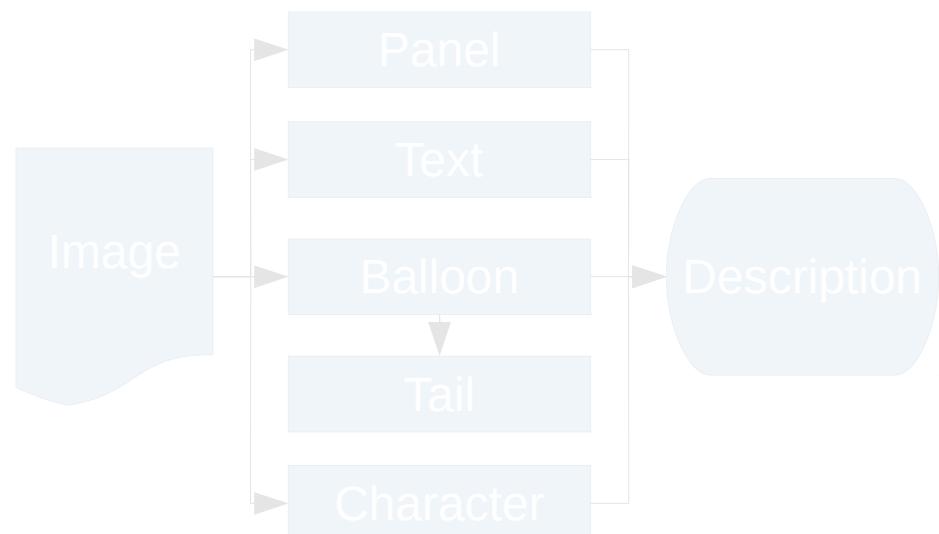
- Introduction
- Document analysis
- State of the art of comics analysis
- Contributions
 - Introduction
 - Content-driven approach
 - Knowledge-driven approach
 - Conclusion
- Experimentations
- Conclusions

Three different approaches for comics analysis

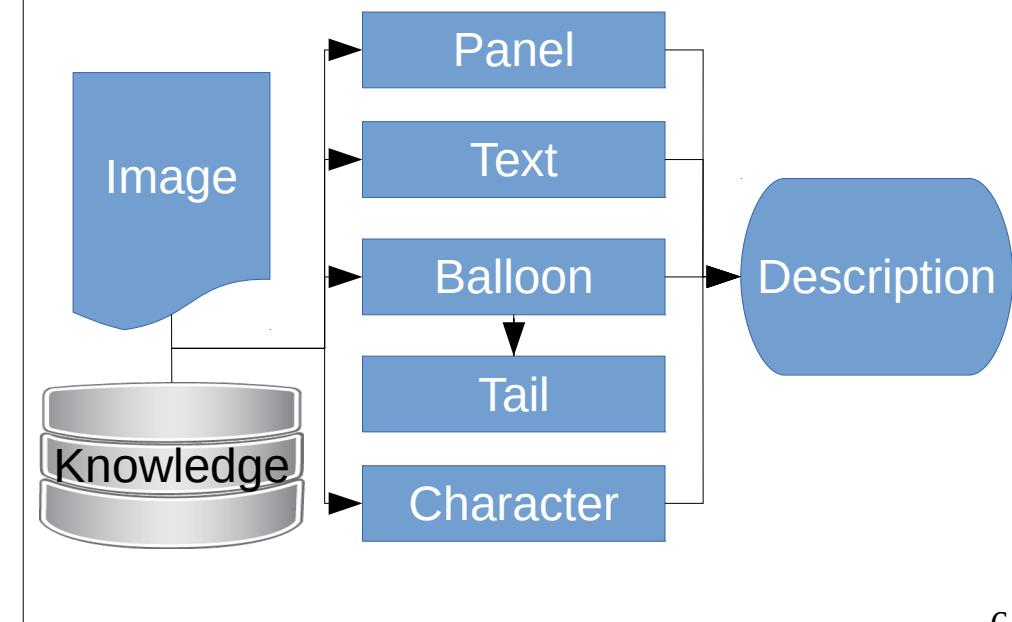
Content-driven (sequential)



Independent

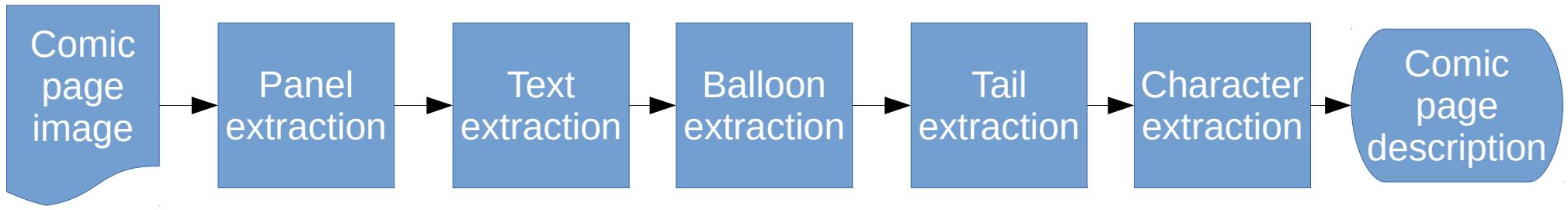


Knowledge-driven

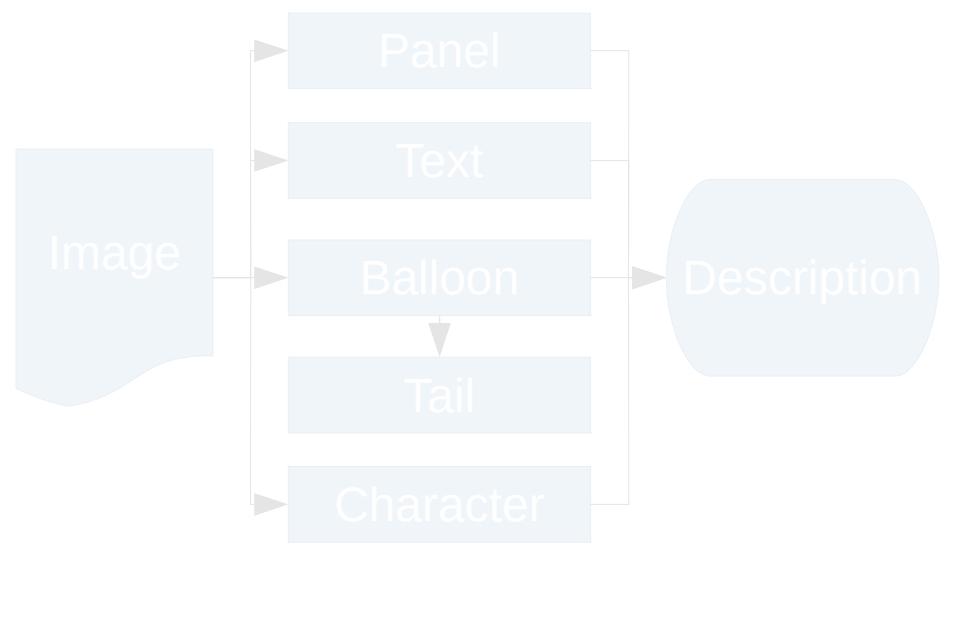


Three different approaches for comics analysis

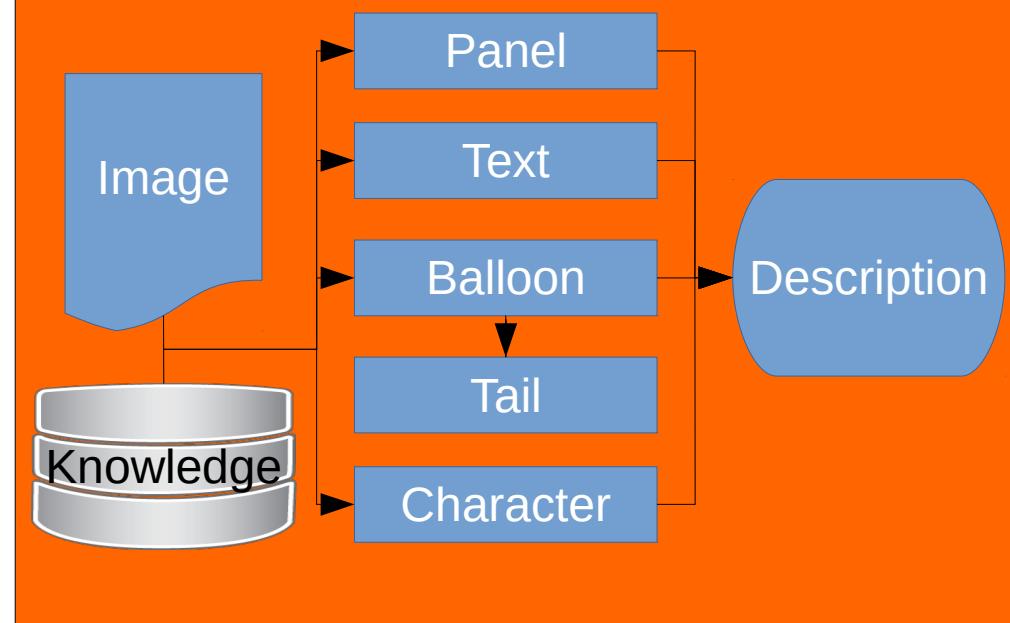
Content-driven (sequential)



Independent



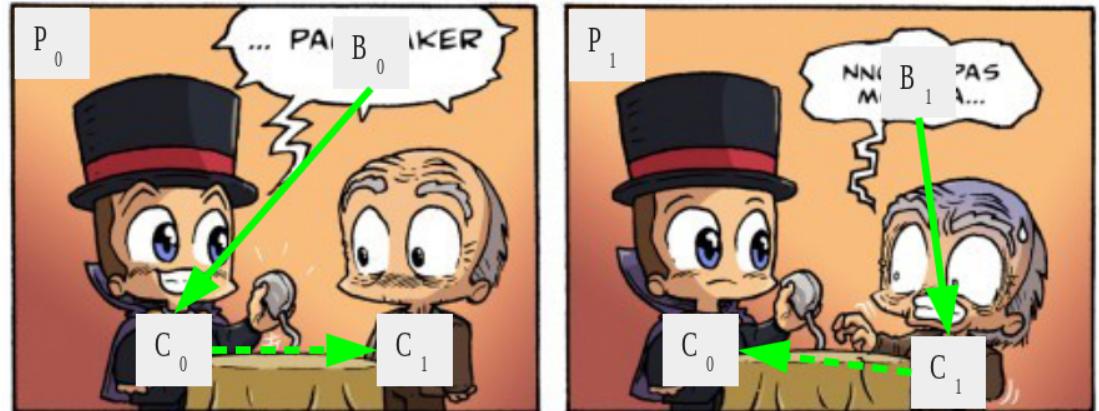
Knowledge-driven



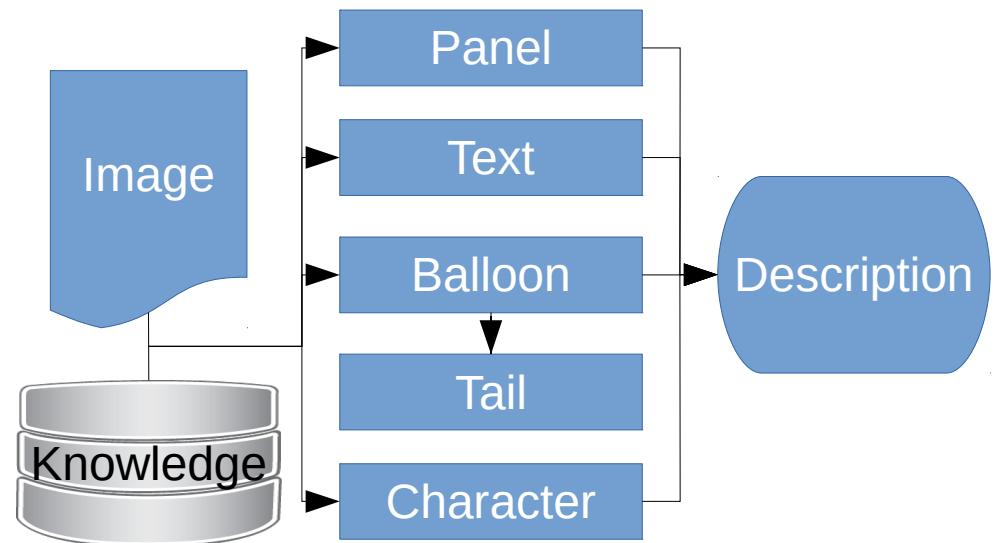
Introduction

Knowledge-driven

- Semantic description and comics understanding
- Independent element extraction
- Image and comics domain knowledge models
- Expert system for contextual analysis
- Processing sequence
- Collaboration with Clément Guérin



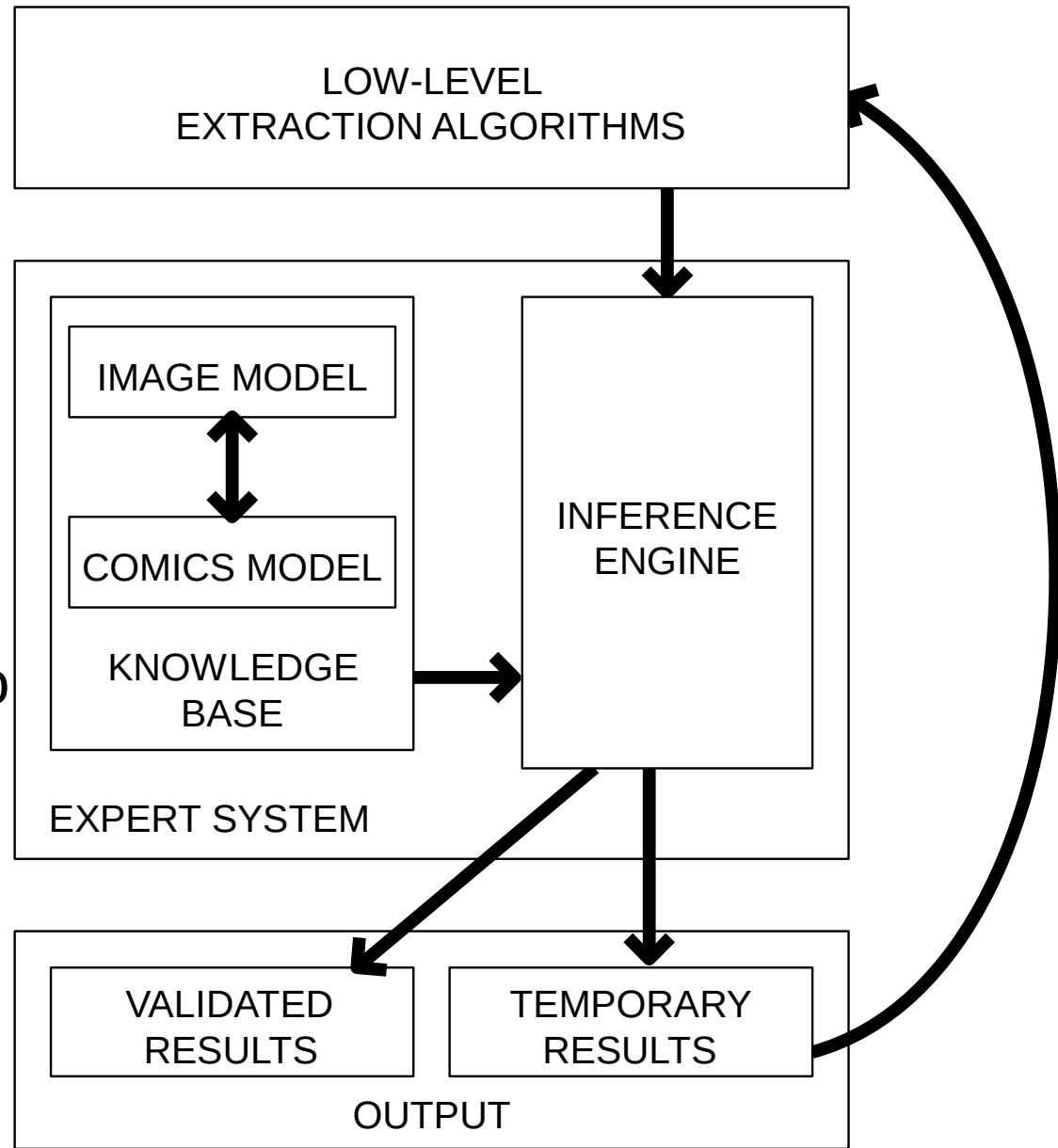
Retrieving and understanding interactions



Knowledge models

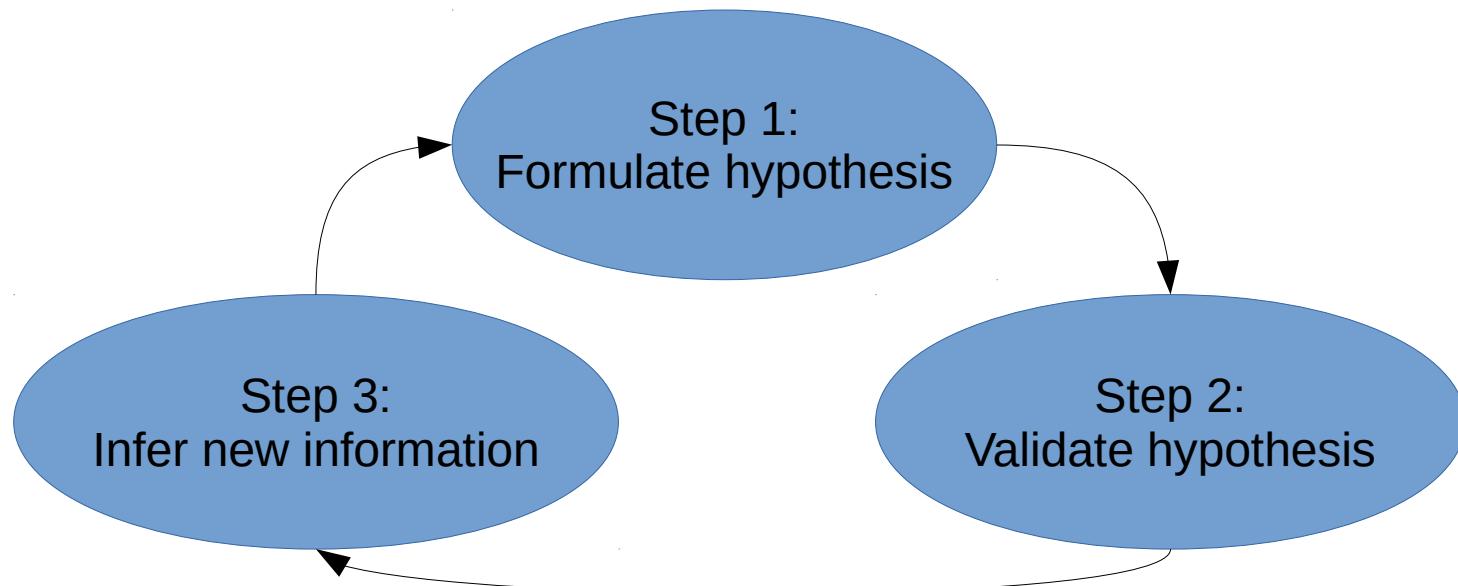
Knowledge-driven

- Image model
 -
- Comics model
 - A panel is related to one page
 - A balloons is related to one panel
 - A characters is related to one panel
 - A text lines is related to one balloon



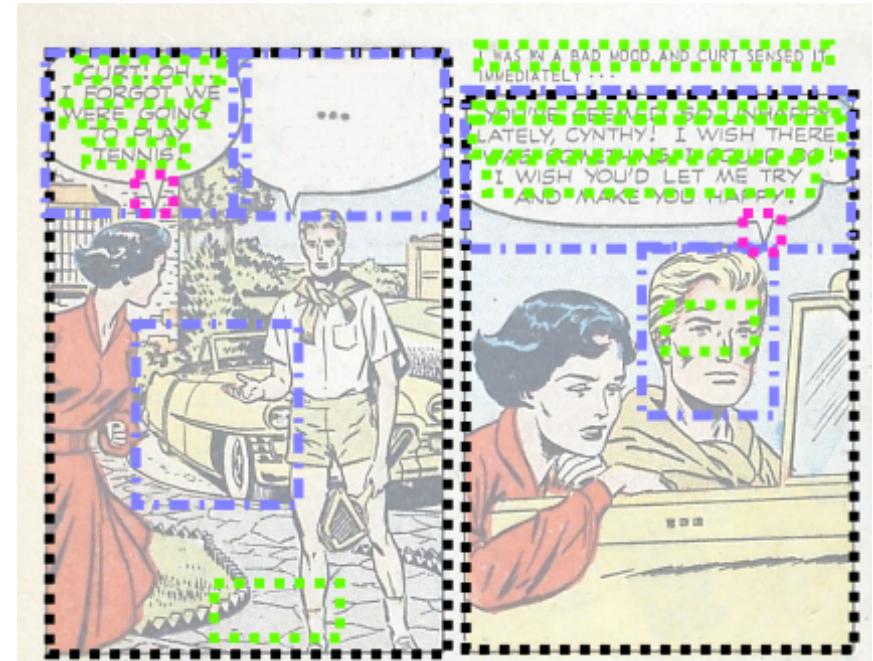
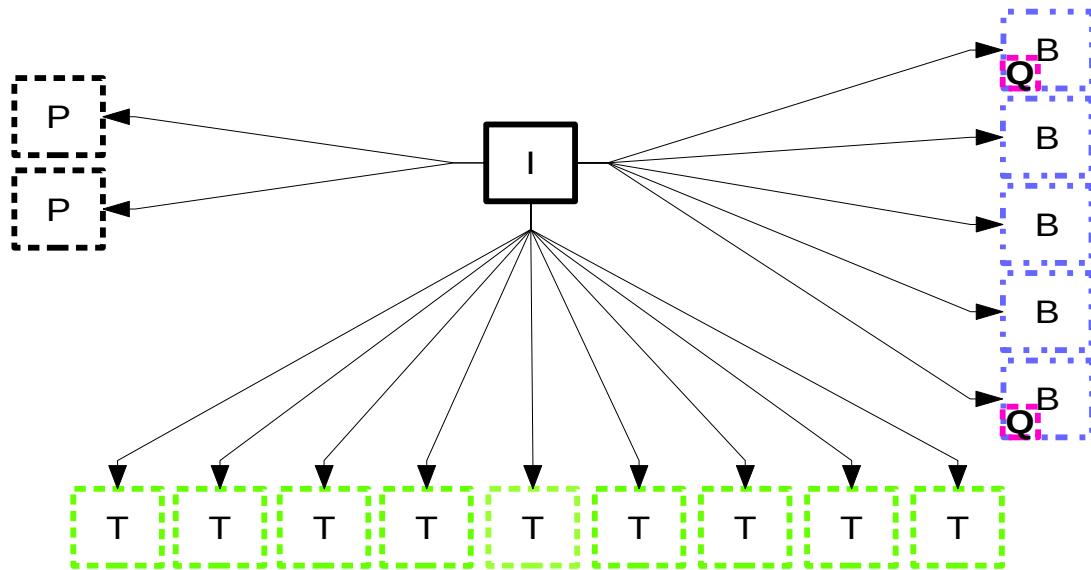
Processing sequence

Knowledge-driven

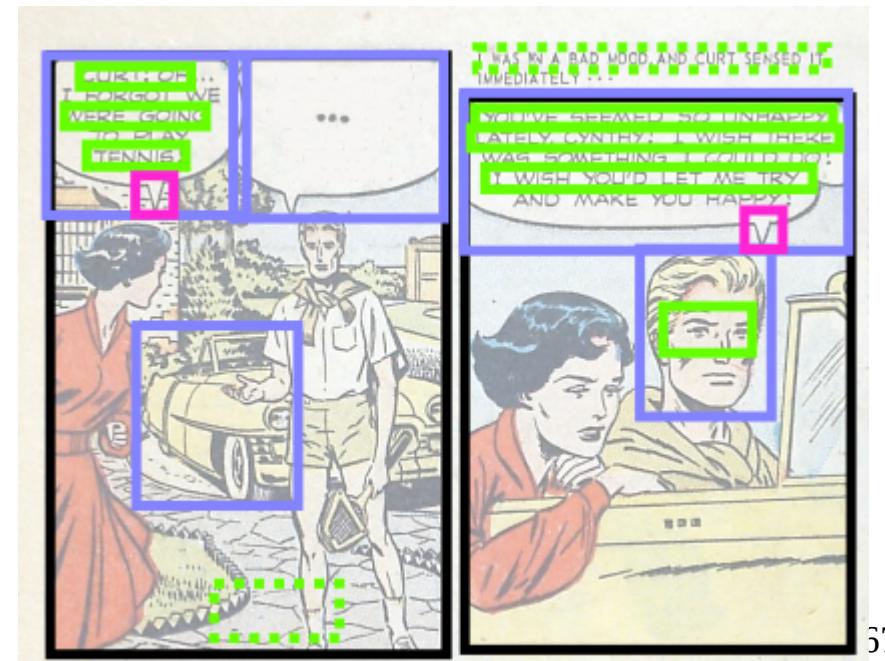
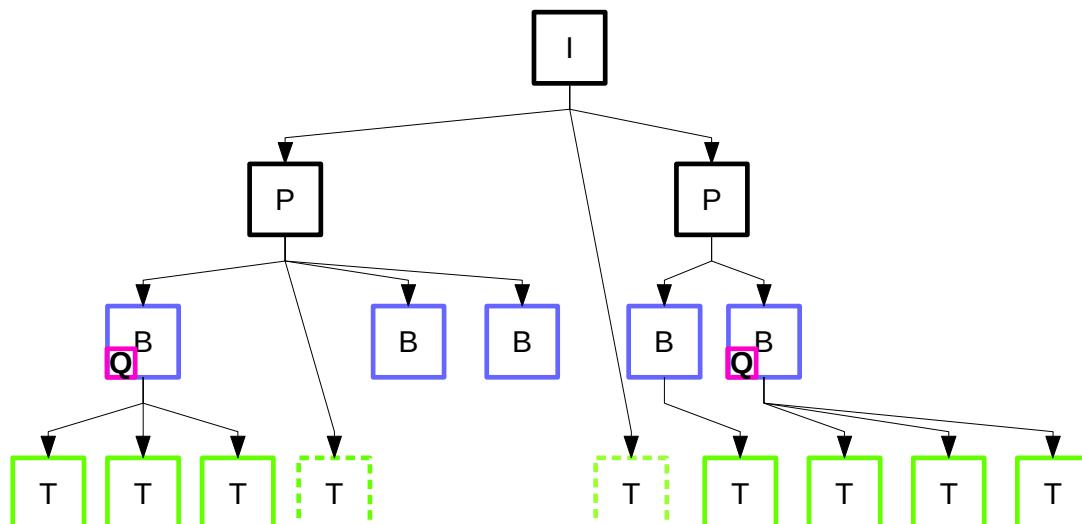


Processing sequence: iteration 1 Knowledge-driven

Step 1: formulation hypothesis

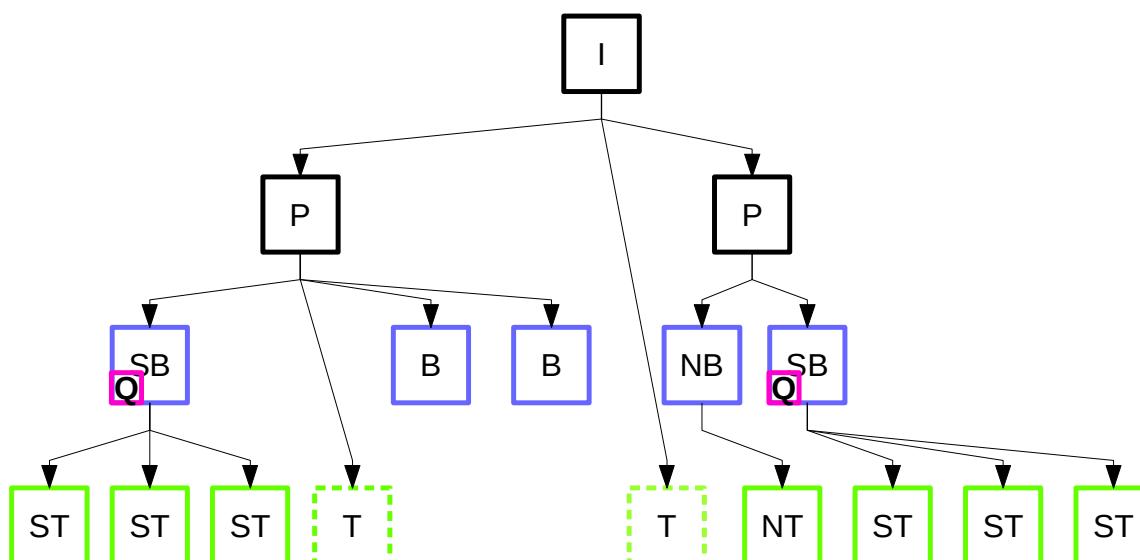
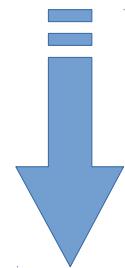
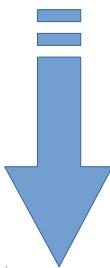
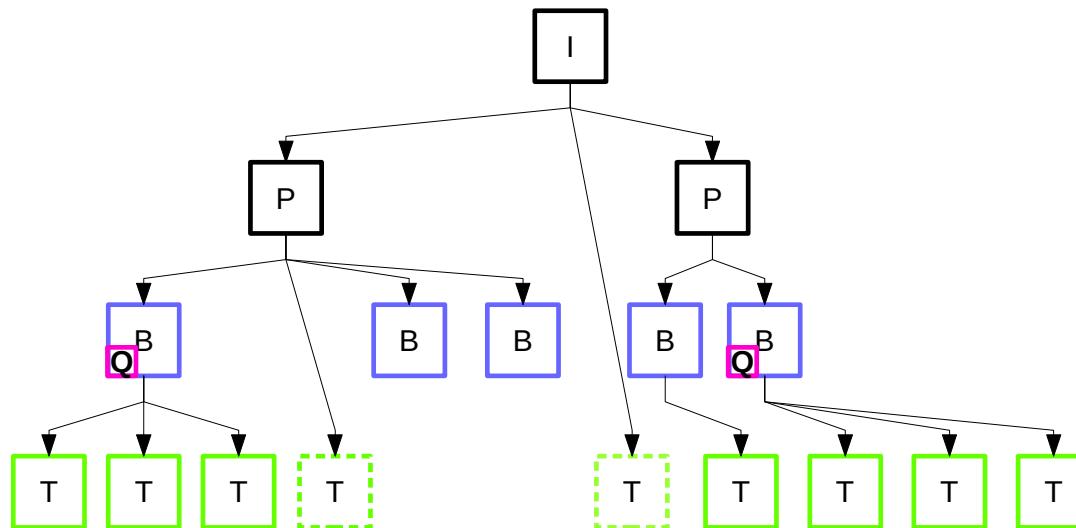


Step 2: validate hypothesis



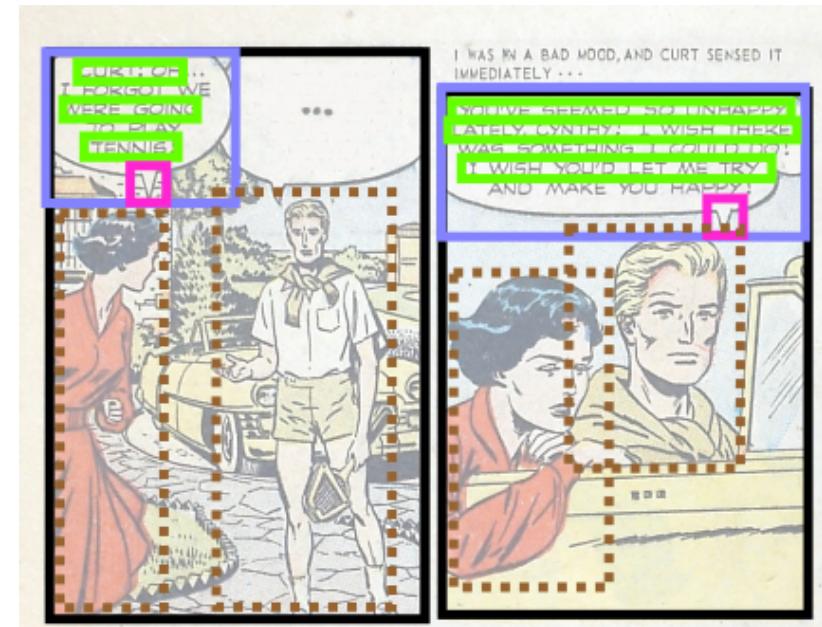
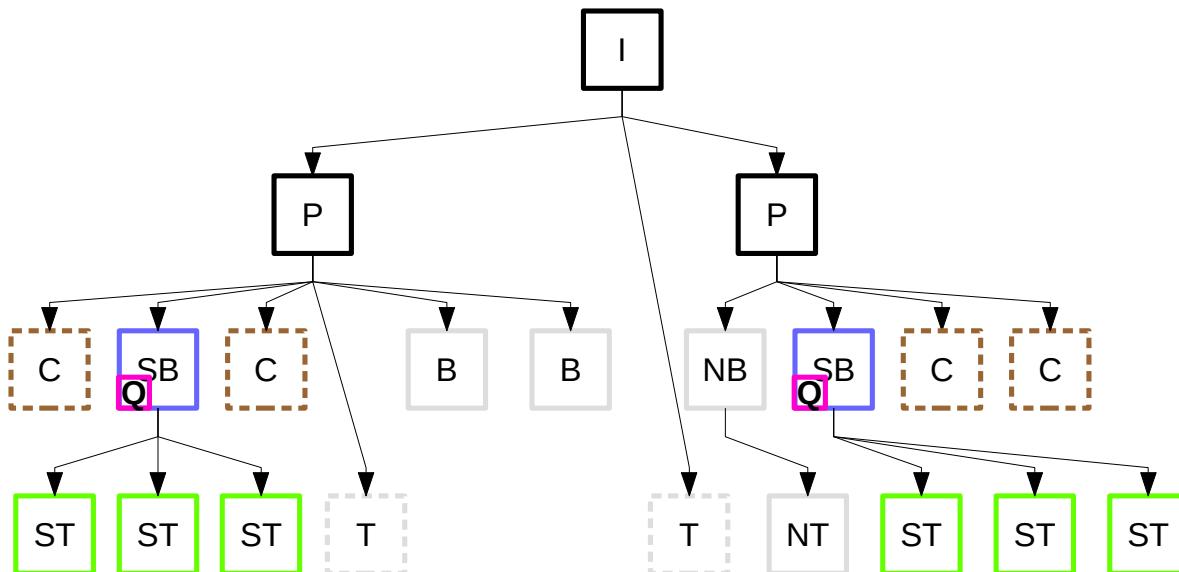
Processing sequence: iteration 1

Step 3: inference of new information

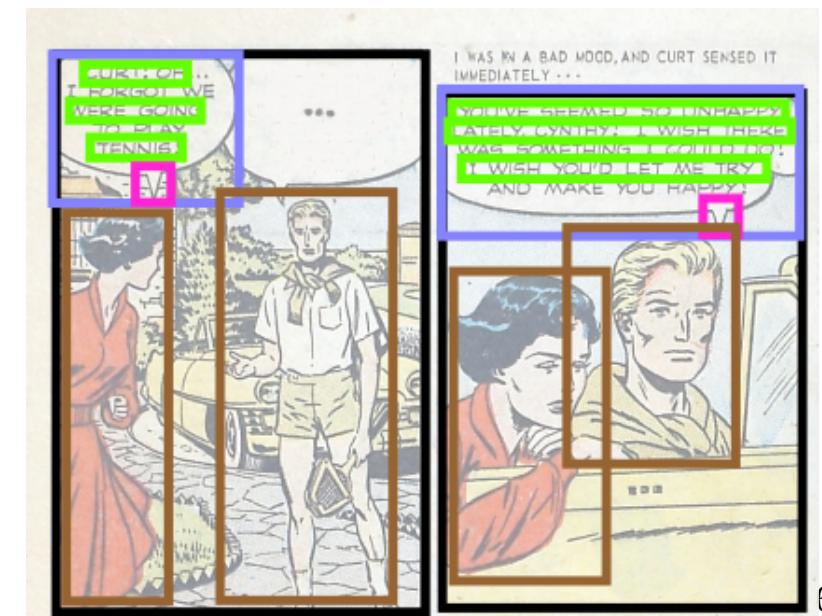
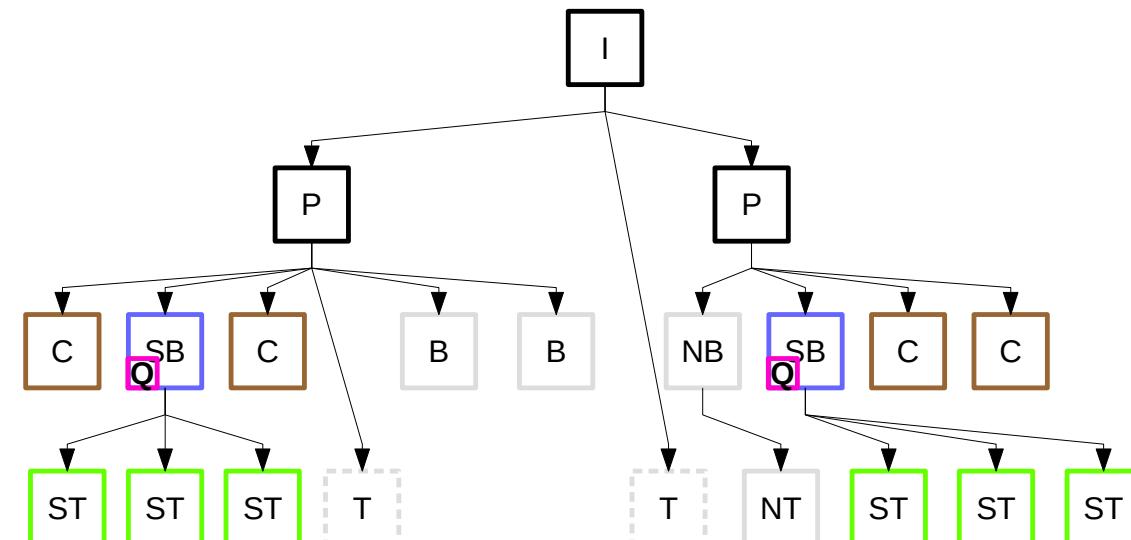


Processing sequence: iteration 2 Knowledge-driven

Step 1: formulation hypothesis

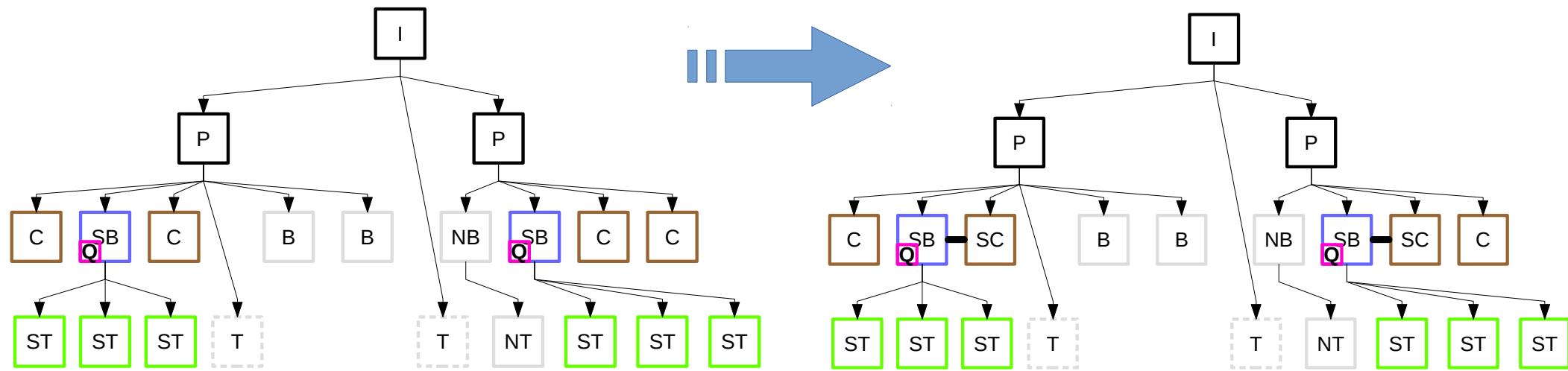


Step 2: validate hypothesis

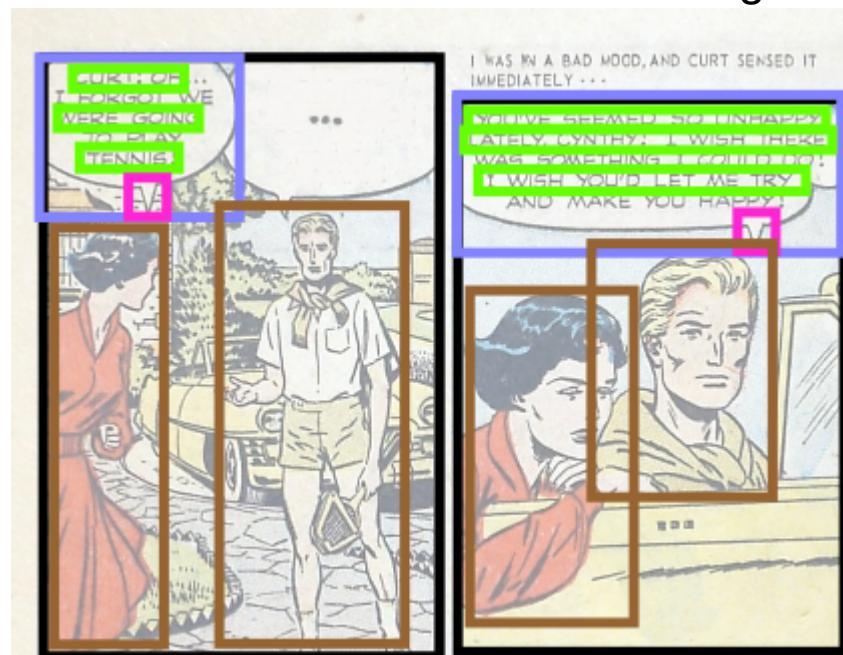


Processing sequence: iteration 1

Step 3: inference of new information



Final information about the image



Conclusion of the knowledge-driven approach

- Contributions
 - Generic, unsupervised and expandable framework for comics understanding
 - Formalization of image and comics concepts
- Limitations
 - Validation process (deletion only)
 -

Outlines

- Introduction
- Document analysis
- State of the art of comics analysis
- Contributions
- Experimentations
- Conclusions

Outlines

- Introduction
- Document analysis
- State of the art of comics analysis
- Contributions
- Experimentations
 - Dataset and ground truth
 - Evaluations
 - Conclusions
- Conclusions

Dataset and ground truth

Experimentations

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



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



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

Evaluations

Experimentations

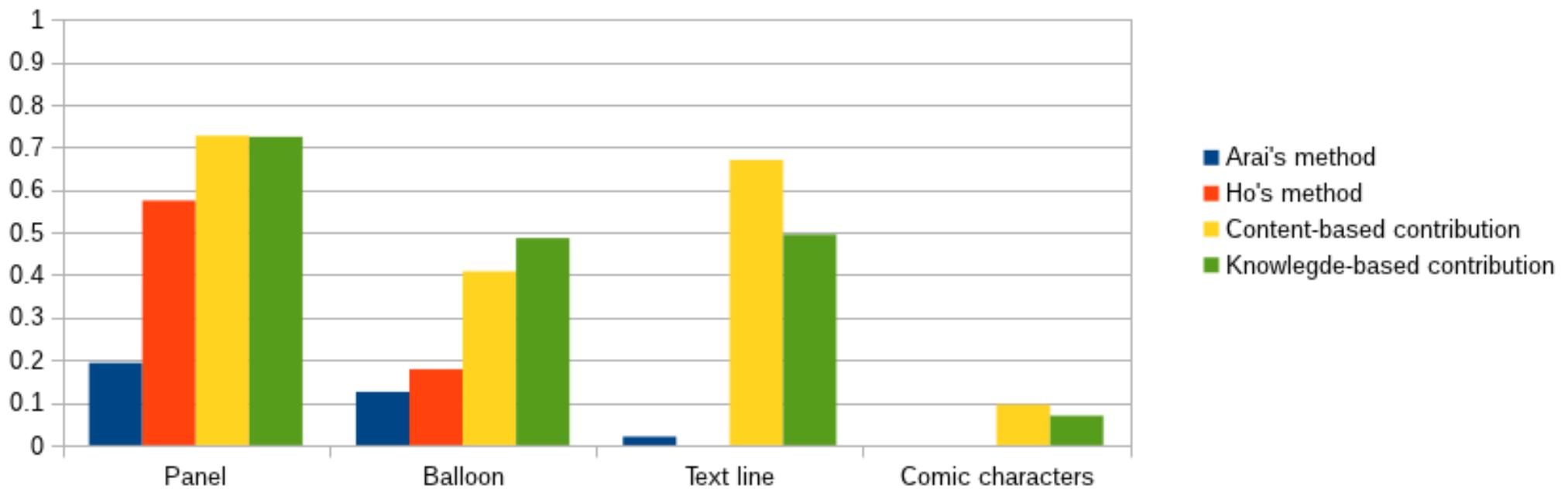
$$a_0 = \frac{\text{area}(B_p \cup B_{gt})}{\text{area}(B_p \cap B_{gt})}$$

B_p = predicted region

B_{gt} = ground truth region

B_p valid if $a_0 > 0.5$

Element localization results (F-score)



Evaluations

Examples of result

Experimentations

W CE JOUR-LÀ, COURANT
SUR LA PLAGE IL NE SE
DOUTAIT PAS QUE SA
VIE ALLAIT BASCULER...

ONE EVENING LEANING OVER THE TAFFRAIL,
I OBSERVED A STRANGE CLOUD THAT STRUCK
A CHILL TO MY HEART.

OK, POUR MOI, LA PRISE EST BONNE !
TU PEUX REMETTRE STEVE DANS NE
LA CHAMBRE DE LA SOEUR DE KID.
ON N'EN A PLUS BESOIN.

... ?!
BATTRE DES
BRAS COMME LES
MOUETTES ?... TU
VEUX DIRE VOLER
COMME DES
MOUETTES, NON
?
SE

Conclusion

Experimentations

Content-driven approach

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

Knowledge-driven approach

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

- Main objective
 - Comics image segmentation and understanding
- Contribution of the thesis
 - Improvement of existing extraction process
 - Several first studies
 - Public dataset and ground truth for researchers
 - Two levels of analysis (content and knowledge driven)
- Research impacts
 - The L3i is now the main actor of comic books analysis in Europe
 - The dataset serves international peers
 - Just starting national project with a French company (Actialuna) and a lab (LIP6) in Paris
 - International project with a group working on Manga analysis at Osaka Prefecture University in Japan

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