La Rochelle Université

Informatique Image Interaction
Univ. La Rochelle

Intelligent Media Processing Group
Osaka Prefecture University

PhD era

CVC Centre de Visió per Computador

SEQUENCITY

PostDoc era

la cité internationale de la bande dessinée et de l'image

SAiL
Sequential Art Image Laboratory

2011 2014 2019

Element	Process type	2011	2014	2019
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			

Christophe Rigaud
et al.

Sequential Art image Laboratory

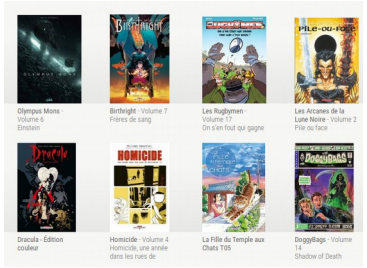
Confidence criterion for speech balloon segmentation

Christophe Rigaud – Van Nguyen and Jean-Christophe Burie
Laboratoire Informatique Image Interaction

GREC 2019 – September 20th – Sydney, Australia

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

How to measure detection without GT?



Online comic bookstore

Balloon segmentation

Method 1

Method 2

Method 3

Method N

Is this result good enough?

Yes

No

Show interaction to the reader

Highlighting,
Zoom,
Translation,
...

RELATED WORKS

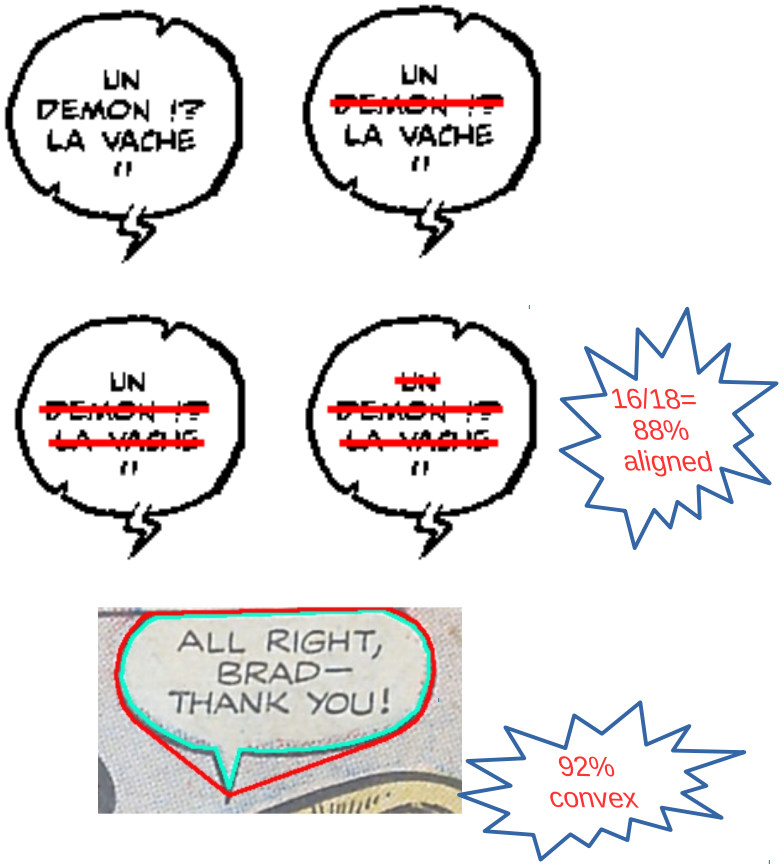
Speech balloon segmentation

- Xicheng Liu et al. 2015 [1]
  - Closed+**unclosed** balloons, clump splitting
  - Recall, precision, F1 score
- Xueting Liu et al. 2016 [2]
  - Closed balloon (Manga), **text-aware** detection [3]
  - **Outside contour** detection
  - Recall, precision
- C. Rigaud et al. 2017 [4]
  - Closed balloon (mixed styles)
  - Recall, precision, F1 + **confidence value**

[1] X. Liu, Y. Wang, and Z. Tang, "**A clump splitting based method to localize speech balloons in comics**" in 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 901–905, Aug 2015
[2] X. Liu, C. Li, H. Zhu, T.-T. Wong, and X. Xu, "**Text-aware balloon extraction from manga**" in The Visual Computer, vol. 32, no. 4, pp. 501–511, Apr 2016
[3] Epshtein, B., Ofek, E., Wexler, Y., "**Detecting text in natural scenes with stroke width transform**" in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2963–2970, 2010

Speech balloon segmentation

- Xicheng Liu et al. 2015 [1]
  - Closed+**unclosed** balloons, clump splitting
  - Recall, precision, F1 score
- Xueting Liu et al. 2016 [2]
  - Closed balloon (Manga), **text-aware** detection [3]
  - **Outside contour** detection
  - Recall, precision
- C. Rigaud et al. 2017 [4]
  - Closed balloon (mixed styles)
  - Recall, precision, F1 + **confidence values**



[4] C. Rigaud, J.-C. Burie, and J.-M. Ogier, "**Text-independent speech balloon segmentation for comics and manga**" in Graphic Recognition. Current Trends and Challenges, Springer, pp. 133–147, 2017.

Speech balloon segmentation

- D. Dubray et al. 2019 [5]
  - Closed/**unclosed** balloons (illusory)
  - Based on **U-Net** architecture + **VGG-16** encoder
  - Recall, precision, F1 score

- N.-V. Nguyen et al. 2019 [6]
  - Closed/**unclosed** balloons (+ character/face/panel/narrative text/link)
  - Region Proposal Network (**RPN**) + **Mask R-CNN**
  - Recall, precision, F1 score

[5] D. Dubray and J. Laubrock (2019). « **Deep CNN-based Speech Balloon Detection and Segmentation for Comic Books »**. CoRR, abs/1902.08137. 2019.
[6] N-V. Nguyen, C. Rigaud, and J.-C. Burie, « **Comic MTL: optimized multi-task learning for comic book image analysis »** International Journal on Document Analysis and Recognition (IJDAR), July 2019

PROPOSED APPROACH

Generic confidence value

S = balloon candidate contour

$$cShape = \frac{arcLength(hull(S))}{arcLength(S)}$$

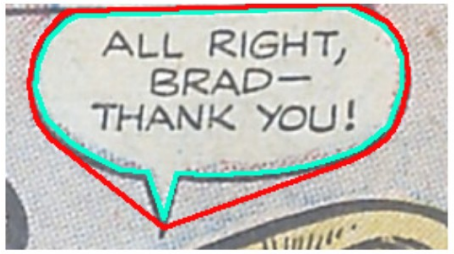
[4]: $C = \alpha \times cAlign + \beta \times cShape$

[7]: *cLex = (1 - mean Levenshtein distance per character ratio)*

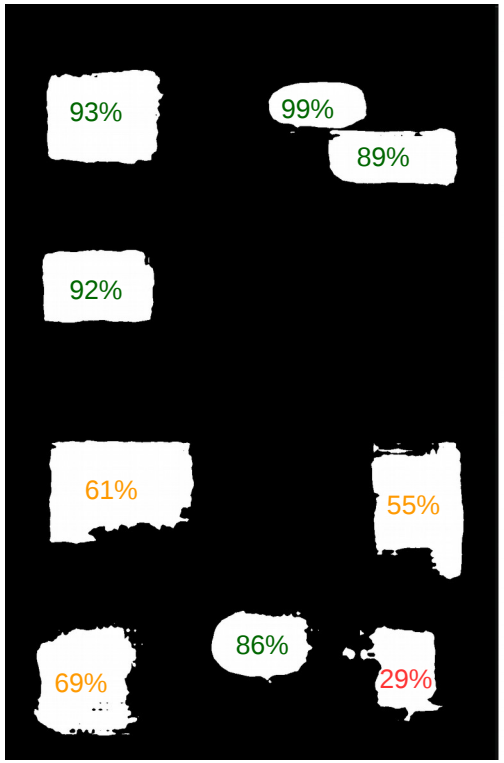
Proposed: $C = \alpha \times cLex + \beta \times cShape$

[4] C. Rigaud, J.-C. Burie, and J.-M. Ogier, "**Text-independent speech balloon segmentation for comics and manga**" in Graphic Recognition. Current Trends and Challenges, Springer, pp. 133–147, 2017.
[7] U. Springmann, F. Fink, and K. U. Schulz. "**Automatic quality evaluation and (semi-) automatic improvement of mixed models for OCR on historical documents**". CoRR, abs/1606.05157, 2016.

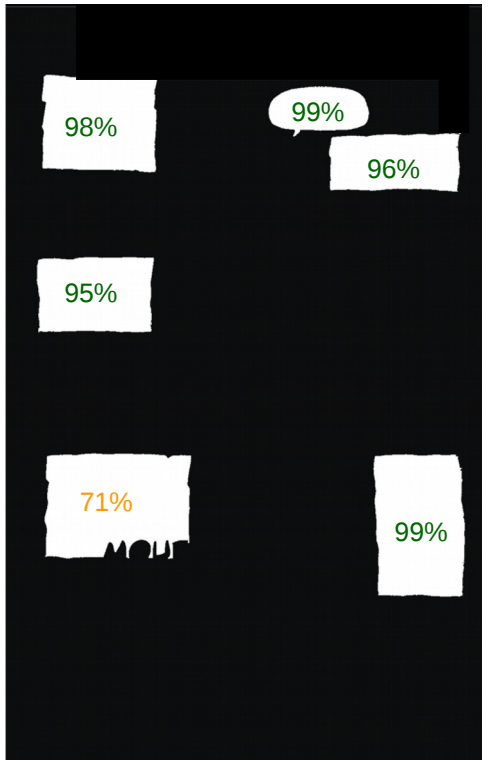
Example 1

		
OCR output	ALL RIGHT, BRAD !HANK you!	(i - . . ,
Confidence	$cShape$ = 92% $cLex$ = 95% (=19/20)	$cShape$ = 22% $cLex$ = 0%

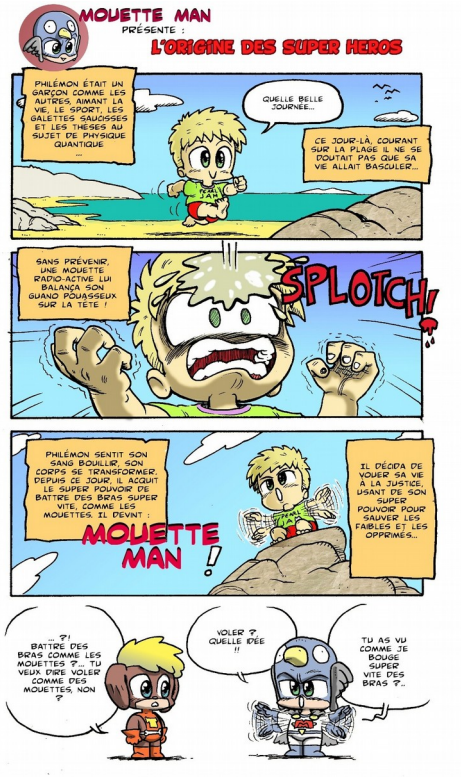
Example 2



NN-based approach



CC-based approach

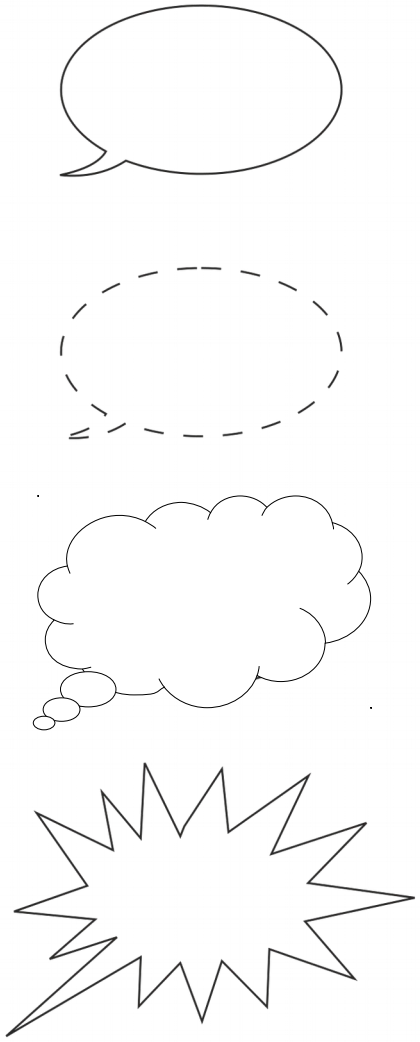


Original image

CONCLUSION

Conclusion

- Simple, fast and generic confidence criterion
- Applications
  - Continious quality control
  - Automatic GT for machine learning
- Limitations
  - OCR/lexicon dependent
  - Whisper, thought, scream balloons
- Future
  - Compare with previous/next balloons



THANK YOU!

https://github.com/crigaud/publication/tree/master/2019/GREC/