Confidence criterion for speech balloon segmentation

Christophe Rigaud, Van Nguyen, Jean-Christophe Burie

▶ **To cite this version:**

Christophe Rigaud, Van Nguyen, Jean-Christophe Burie. Confidence criterion for speech balloon segmentation. 13th IAPR International Workshop on Graphics Recognition, Sep 2019, Sydney, Australia. hal-02305078

HAL Id: hal-02305078
https://hal.science/hal-02305078

Submitted on 3 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Confidence criterion for speech balloon segmentation

Christophe Rigaud, Van Nguyen and Jean-Christophe Burie
Laboratoire L3i,
University of La Rochelle
17042 La Rochelle CEDEX 1, France
{*christophe.rigaud, nhu-van.nguyen, jean-christophe.burie*}*@univ-lr.fr*

***Abstract*—This short paper investigates how to improve the confidence of speech balloon segmentation algorithms from comic book images. It comes from the need of precise indications about the quality of automatic processing in order to accept or not each segmented regions as a valid result, according to the application and without requiring any ground truth. We discuss several applications like result quality assessment for companies and automatic ground truth creation from high confidence results to train machine learning based systems.We present some ideas to combine several domain knowledge information (e.g. shape, text, etc.) and produce an improved confidence criterion.**

I. Introduction

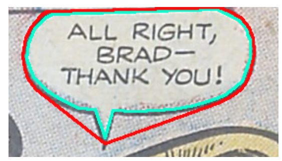
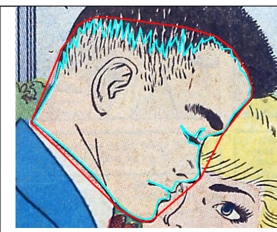
Digital comic content is produced to facilitate transport, reduce publishing cost and allow reading on screens from television to smartphones like newspapers and other documents. To get a user-friendly experience of digital comics on all mediums, it is necessary to extract, identify and adapt comic book content originally designed for paper printing [1]. Comic book image content is composed of different textual and graphical elements such as panel, balloon, text, comic character and background.

The speech balloon is a major link between graphic and textual elements. They can have various shapes (e.g., oval, rectangular) and contours styles (e.g., smooth, wavy, spiky, partial, absent). Speech balloons entirely surrounded by a black line (closed) have attracted most of the researches so far. They were initially based on region detection, segmentation and filtering rules [2], [3] and evaluated on small privates datasets. Liu *et al.* [4] proposed a clump splitting based localization method which can detect both closed and open speech balloons. They performed the evaluation on the eBDtheque [5] dataset using recall and precision metrics.

In addition, Liu *et al.* [6] and Rigaud *et al.* [7] proposed approaches making use of text contained in speech balloons for segmenting speech balloon contours at pixel level. They both used the same evaluation metric (recall and precision) and dataset (eBDtheque) with a little difference that Rigaud *et al.* computed a confidence value within their contour candidate filtering operation.

All these approaches have been tested on datasets which give indication about how well they perform and make comparison possible with new researches. However, they can not give an indicator on a single detection out of the tested dataset, except [7]. Private companies may not be satisfied by such evaluation because published datasets may not exactly reflect their private dataset characteristics. Moreover, a confidence value associated to each detected element would be helpful for deciding if it fits their requirements or requires extra processing.

Table I
Examples of confidence criterion for correct and wrong balloon segmentation.

	
ALL RIGHT, BRAD WHANK you!	(i - . . ,
$cShape$ = 0.92% $cLex$ = 0.95%	$cShape$ = 0.22% $cLex$ = 0.00%

In this paper we propose an improvement of an existing confidence criterion for speech balloon segmentation quality.

II. Proposed method

We define the speech balloon segmentation confidence criterion as a score between zero and one encoding the confidence of each segmented region similarly to the confidence value introduced in [7]. In [7], the confidence value is based on two weighted variables such as the contained text (all connected components) alignment $cAlign$ and segmented region shape $cShape$. We propose to strengthen the first parameter initially based on alignment features (distance and position analysis of neighbouring connected components) by replacing it by the confidence value computed from at least one Optical Character Recognition (OCR) system. OCR systems take into account a lot of features to recognise characters, words and text lines from images (e.g. alignment, shape, size, spaces, contrast) which increases their results reliability. Such systems usually provide a confidence value associated to each results (classification likelihood) but instead of using it , we preferred to use a readily observable quantity that correlates well with true accuracy of the recognized text as describe in [8]. This metric was originally proposed to compare OCR system output performances but in our case, we use it as an OCR independent indicator about the OCR output quality. We compute, for each OCR token, the minimum edit distance (Levenshtein distance) to its most probable lexical equivalent from a lexicon of the corresponding language (e.g. Grammalecte, WordNet). The sum of these distances d over all tokens is, therefore, a statistical measure for the OCR output, and the lexicality defined as *cLex = (1 - mean Levenshtein distance per character ratio)* is a measure for accuracy.

The second term $cShape$, as described in the original publication [7], encodes the overall convexity of the balloon outline in order to find how similar to a perfect bubble (or rectangle) the balloon candidate is. It is defined as the ratio between the Euclidean perimeter of the convex hull of the measured shape S and the Euclidean perimeter of the measured shape S as follows:

$$cShape = \frac{arcLength(hull(S))}{arcLength(S)} \quad (1)$$

The original Equation (2) from [7] becomes as follows when we replace $cAlign$ by $cLex$:

$$C = \alpha \times cLex + \beta \times cShape \quad (2)$$

The best weighting parameter values were validated as $\alpha = 0.75$ and $\beta = 0.25$ in [7] ($\alpha + \beta = 1$). However, because $cLex$ is based on really different features compared to the original $cAlign$ parameter, they both need to be re-validated according to the desired application. The main advantage of replacing the alignment-based measure by a OCR-based measure is that it is much more reliable to detect the presence of text with a high confidence inside segmented regions, thanks to the growing progress of OCR systems [9]. However, if the OCR system is not able to recognize a part of text because it is written with an "unseen" typewritten font or handwritten style, it will result in a poor confidence score even if the segmentation region is a true positive. Also, words may be recognized as others, or with minor errors and still get a good confidence score in some cases.

An example of the proposed confidence criterion is given Table I. In this table, segmented contours are represented in cyan and convex hulls in red in the first row. Wrong transcriptions are highlighted in red in OCR output in the second row. Corresponding confidences are given in the last row.

III. Conclusion

This short paper investigates how to compute a confidence criterion that can indicate speech balloon segmentation quality without requiring any ground truth. It relies on comics domain knowledge i.e. bubble-like shape and text content. It may be suitable for continuous quality control in large digitization and indexation processes or automatic ground truth generation for machine learning techniques.

In the future, we would like to investigate other features that can improve further the quality of such confidence criterion. The combination with external information like the position in the panel and the overlap with other elements could be other sources of information to aggregate.

Acknowledgment

This work is supported by Research National Agency (ANR) in the framework of 2017 LabCom program (ANR 17-LCV2-0006-01), CPER NUMERIC program funded by the Region Nouvelle Aquitaine, CDA, Charente-Maritime French Department, La Rochelle conurbation authority (CDA) and the European Union through the FEDER funding.

References

[1] O. Augereau, M. Iwata, and K. Kise, "A survey of comics research in computer science," *CoRR*, vol. abs/1804.05490, 2018. [Online]. Available: http://arxiv.org/abs/1804.05490

[2] K. Arai and H. Tolle, "Method for real time text extraction of digital manga comic," *International Journal of Image Processing (IJIP)*, vol. 4, no. 6, pp. 669–676, 2011.

[3] A. K. N. Ho, J.-C. Burie, and J.-M. Ogier, "Panel and Speech Balloon Extraction from Comic Books," *2012 10th IAPR International Workshop on Document Analysis Systems*, pp. 424–428, mar 2012.

[4] X. Liu, Y. Wang, and Z. Tang, "A clump splitting based method to localize speech balloons in comics," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, Aug 2015, pp. 901–905.

[5] C. Guérin, C. Rigaud, A. Mercier, F. Ammar-Boudjelal, K. Bertet, A. Bouju, J. C. Burie, G. Louis, J. M. Ogier, and A. Revel, "eBDtheque: A representative database of comics," in *2013 12th International Conference on Document Analysis and Recognition*, Aug 2013, pp. 1145–1149.

[6] X. Liu, C. Li, H. Zhu, T.-T. Wong, and X. Xu, "Text-aware balloon extraction from manga," *The Visual Computer*, vol. 32, no. 4, pp. 501–511, Apr 2016. [Online]. Available: https://doi.org/10.1007/s00371-015-1084-0

[7] 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*, B. Lamiroy and R. Dueire Lins, Eds. Cham: Springer International Publishing, 2017, pp. 133–147.

[8] C. Rigaud, J. Burie, and J. Ogier, "Segmentation-free speech text recognition for comic books," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 03, Nov 2017, pp. 29–34.

[9] T. M. Breuel, "High performance text recognition using a hybrid convolutional-lstm implementation," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, Nov 2017, pp. 11–16.