Speech balloon contour classification in comics

Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier
Laboratory L3i
Université de La Rochelle
Avenue Michel Crépeau 17042 La Rochelle, France
{christophe.rigaud, jean-marc.ogier, jean-christophe.burie}@univ-lr.fr

Dimosthenis Karatzas
Computer Vision Center
Universitat Autònoma de Barcelona
E-08193 Bellaterra (Barcelona), Spain
dimos@cvc.uab.es

***Abstract*—Comic books digitization combined with subsequent comic book understanding create a variety of new applications, including mobile reading and data mining. Document understanding in this domain is challenging as comics are semi-structured documents, combining semantically important graphical and textual parts. In this work we detail a novel approach for classifying speech balloon in scanned comics book pages based on their contour time series.**

***Keywords*—contour classification, comics understanding, time series, speech balloon**

I. Introduction

Comic books are a widespread cultural expression and are commonly accepted as the “ninth art”. Comics are a hybrid medium, combining textual and visual information in order to convey their narrative. Digitization combined with subsequent document understanding of comic books is therefore of interest, both in order to add value to existing paper-based comic heritage, but also to bridge the gap between the paper and electronic comic media.

In comics content understanding, speech balloons present a lot of interest since they offer the links between the textual content and comic characters providing two major informations, one about the location of the characters with the tail, and another one about the speech tone according to the contour shape. Apart for document understanding, shape classification is also beneficial for image retrieval such as marine creature [1], leaves [2], illicit tablet [3] and visual shape descriptor [4]. Also shape representation overviews are given by [5] and [6].

Contour-based classification method, especially in a template matching context, is a well developed field. Three well know methods are Curvature Scale Space (CSS) [7], Fourier Descriptor [8] and Invariant Moments [9]. A recent work [3] explains that CSS is more appropriate for shapes containing a high number of inflections as marine animals [7].

Image moment have been used extensively for Optical Character Recognition [10], plant leaf recognition [11], vehicle detection in satellite images [12], and detecting and localizing duplicated areas in digital images [13]. Nevertheless, they are not discriminant enough for our purpose because speech balloon are most of the time compact, symmetric and the gravity centre right in the middle. The discriminant information is in the boundary, not in the shape of the balloon.

Speech balloon shapes are different from the already studied domains (e.g. marine animals, leaves, characters) because their are part of the hand drawing art of the comics and therefore unique. Depending on the style of the comics and the structure of the panel, the drawer make them similar to rectangle, square, circle, oval or unknown (see figure 1). Independently from the shape, the type of line gives an information about the way the text is spoken, for instance sharp for exclamations and explosions or wavy for dreams and thoughts (see figure 2).

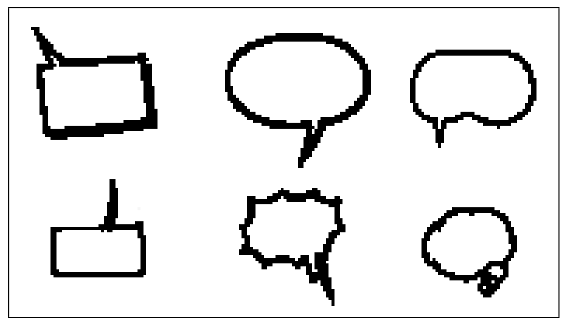


Fig. 1. Examples of speech balloon shapes. Rectangle on the left, oval in the middle and unknown on the right side.

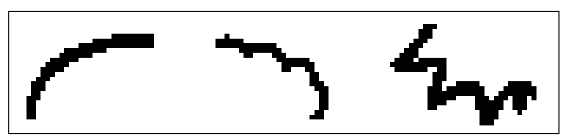


Fig. 2. Examples of contour line types. Smooth on the left, wavy in the middle and zigzag on the right side.

Even for human being, it is often difficult to say whether the balloon shape is oval, rectangular or unknown. We can not determine an exhaustive list of balloon shapes, nevertheless we can understand different meanings according to the features that compose its contour (e.g. straight line, corners, inflections, oscillations). Those features guide our well trained eyes (and brain) through the story and give us the tone of the story without having to explicitly write it.

We aim is to determine the semantic of the balloon (e.g. thought, speech, exclamation, explosion) according to the low level features that compose its boundary. The classification at this level, has to do with the type of line used to draw the balloon (e.g. zigzag, continuous, cloud-like, wavy) and not with the overall shape (e.g. circular, oval, semi-rectangular, rectangular).

If we were able to automatically determine which balloons are for general purpose and the more disruptive ones then we will make a progress in comics content understanding.

II. PROPOSED METHOD

Balloon line type classification is complementary to speech balloon detection. For balloon detection, we use a recent work [14] which is able to detect both closed and suggested (not explicitly draw) speech balloon based on active contour model initialized on text regions. The algorithm assume that text line bounding boxes have already been detected and group them into text blocks by using spacial information. Then, an active contour model (curve) is initialized on the convex hull of each text blocks and iteratively fit the balloon contour by minimizing the internal and external energy terms of its energy function. This approach gives good results for balloon localization at bounding box level but suffer from a lack of precision for contour fitting. In our case, we consider only closed balloon because we based the classification on the balloon boundary. Because suggested balloons have only partial or totally missing boundary, the proposed method is not appropriate. In this particular case, different features have to be investigated (e.g. tail type recognition, speech analysis).

We propose to adapt the balloon detection, measure the difference to the balloon contour and then convert it into a one dimension signal (time series) that we analyse in spacial and frequency domains for the classification.

A. Balloon shape detection

There are several ways to find the global shape of a speech balloon (the seed in our case), for instance we can approximate it by computing the enclosed or surrounded ellipse or rectangle (the most frequent shape in comics) or using a deformable shape as active contour (see figure 3). We chose the last one (active contour) because they already proved their evidence for balloon detection [14] and it is invariant to rotation, scale and translation. Moreover, we change the energy function weighting parameters in order to give more importance to internal energies, maintain the snake smooth and keep a certain distance with the boundary features (see original equation 1).

$$E = E_{int} + E_{ext} + E_{text} \tag{1}$$

The weighting parameters γ for the external energy term (E_{ext}), α and β for the internal energy terms (E_{int}) and κ for the text energy term (E_{text}) where defined as follows: $\gamma = 0.1$, $\alpha = 0.25$, $\beta = 0.25$, $\kappa = 0.4$.

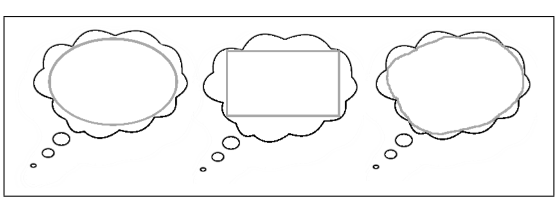
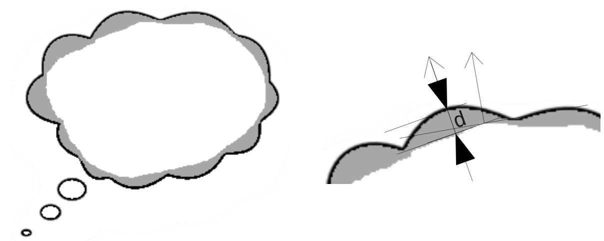


Fig. 3. Example of balloon shape approximation by the maximal included ellipse (left), rectangle (center) and active contour (right).

B. Contour detection and analysis

As introduced in section I, the discriminant information for balloon classification are in the contour variations, not in the global shape of the balloon. The idea here is to focus on the region between the shape detection and the balloon contour and measure those variations by computing the Euclidean distance all around the contour (see figure 4a). Note, the average distance will be close to zero if the balloon has a smooth round contour or high in case of wavy and zigzag contour. For each point of the shape detection we compute the distance to the first perpendicularly edge to the tangent and from the outside of the balloon (see figure 4b). The inter-point distance has to be smaller than the feature to detect as explained in [14]. The distances are then interpreted as time series to be classified.



a) Region of interest (grey area) b) Euclidean distance d

Fig. 4. Difference between the balloon shape detection (left) and Euclidean distance measure (right).

C. Contour classification

In this section we propose a method to classify speech balloon contours by converting the distance between the global balloon shape and the contour details into time series. Another approach could be based on the centre of mass [15]. We consider three classes as “smooth”, “wavy” and “zigzag” that can be respectively interpreted as “dialogue”, “thought” and “exclamation” from a comics content understanding point of view. Each class has a particular profile, “smooth” has a low average distance value, “wavy” and “zigzag” have both a strong response in the frequency domain but their point repartition around the maximums is discriminant (see figure 5).

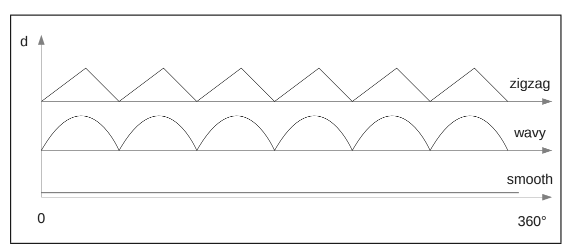


Fig. 5. Typical time series for “smooth” (bottom), “wavy” (middle) and “zigzag” (top) classes.

An alternative would be to use a corner detector to get an idea about the contour distortions but that ignore the spatial information.

III. DISCUSSION

The proposed solution covers the comics contour classification problem in general but comics contains a lot of creativities.

Two cases are not taken into account in the proposed method, first the case of non closed contour that are actually handled by the balloon shape detector we use [14] but not in the contour detection and classification process (see section II-B and II-C). Second, the case of multiple contour types for the same balloon (e.g. if only the bottom of a balloon has zigzag style and all the rest is smooth). Also, other distance measures have to be investigated.

IV. Conclusion

In this paper we propose a contour classification method for comics speech balloons based on balloon shape approximation and contour time series. This is a preliminary work that will be continued to handle more comics specificities until having a complete comics understanding framework.

Acknowledgement

This work was supported by the European Doctorate founds of the University of La Rochelle, European Regional Development Fund, the region Poitou-Charentes (France), the General Council of Charente Maritime (France), the town of La Rochelle (France) and the Spanish research projects TIN2011-24631, RYC-2009-05031.

References

[1] F. Mokhtarian and S. Abbasi, "Shape similarity retrieval under affine transforms," vol. 35, pp. 31–41, 2002.

[2] Z. Wang, Z. Chi, and D. Feng, "Shape based leaf image retrieval," *Vision, Image and Signal Processing, IEEE Proceedings*, vol. 150, no. 1, pp. 34–43, 2003.

[3] M. Lopatka and W. V. Houten, "Science and Justice Automated shape annotation for illicit tablet preparations: a contour angle based classification from digital images," *Science & Justice*, vol. 53, no. 1, pp. 60–66, 2013.

[4] M. Bober, "Mpeg-7 visual shape descriptors," *IEEE Trans. Circuits Systems*, 2001.

[5] R. C. Veltkamp and M. Tanase, "Content-based image retrieval systems: A survey," 2000.

[6] D. Zhang and G. Lu, "Review of shape representation and description techniques," *PR*, vol. 37, no. 1, 2004.

[7] S. Abbasi, F. Mokhtarian, and J. Kittler, "Curvature scale space image in shape similarity retrieval," *Multimedia Syst.*, vol. 7, no. 6, pp. 467–476, Nov. 1999.

[8] C. T. Zahn and R. Z. Roskies, "Fourier Descriptors for Plane Closed Curves," *Transactions on Computers, IEEE*, vol. c-21, no. 3, pp. 269–281, Mar. 1972.

[9] M.-K. Hu, "Visual pattern recognition by moment invariants," *Information Theory, IRE Transactions on*, vol. 8, no. 2, pp. 179–187, 1962.

[10] J. Flusser, B. Zitova, and T. Suk, *Moments and Moment Invariants in Pattern Recognition*. Wiley Publishing, 2009.

[11] X. Wang, D.-S. Huang, J.-X. Du, H. Xu, and L. Heutte, "Classification of plant leaf images with complicated background." *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 916–926, 2008.

[12] L. Eikvil, L. Aurdal, and H. Koren, "Classification-based vehicle detection in high-resolution satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 65–72, Jan. 2009.

[13] B. Mahdian and S. Saic, "Detection of copymove forgery using a method based on blur moment invariants," *Forensic Science International*, vol. 171, no. 2-3, pp. 180–189, Sep. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.forsciint.2006.11.002

[14] C. Rigaud, D. Karatzas, J. Van de Weijer, J.-C. Burie, and J.-M. Ogier, "An active contour model for speech balloon detection in comics," in *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2013.

[15] E. Keogh, L. Wei, X. Xi, S. hee Lee, and M. Vlachos, "Lb keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures," in *IN VLDB, 2006*, 2006, pp. 882–893.