Text-independent speech balloon segmentation for comics and manga

Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier

Laboratoire L3i, Université de La Rochelle
Avenue Michel Crépeau 17042 La Rochelle, France
{christophe.rigaud, jcburie, jmogier}@univ-lr.fr

Abstract. Comics and manga are one of the most popular and familiar forms of graphic content over the world and play a major role in spreading country's culture. Nowadays, massive digitization and digital-born materials allow page-per-page mobile reading but we believe that other usages may be released in the near future. In this paper, we focus on speech balloon segmentation which is a key issue for text/graphic association in scanned and digital-born comic book images. Speech balloons are at the interface between text and comic characters, they inform the reader about speech tone and the position of the speakers. We present a generic and text-independent speech balloon segmentation method based on color, shape and topological organization of the connected-components. The method has been evaluated at pixel-level on two public datasets (eBDtheque and Manga109) and the F-measure results are 78.24% and 80.04% respectively.

Keywords: graphic recognition, speech balloon, comics image analysis, manga image analysis.

1 Introduction

The sales of digital comics are now reaching 10% of the comics market and has doubled during the last five years[1]. Such new way of reading allows new capabilities thanks to the richness of the drawings and the recent development of mobile platform reading tools. Apart from layout re-flowing (panel re-arrangement) according to screen size, there are few work exploring other ways of reading.

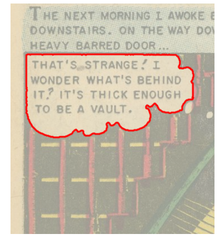
In this paper, we address pixel-level balloon segmentation in order to retrieve position and shape of the speech balloons. Both information are key issues for comics understanding, especially for balloon classification [15] and comic character association [18]. This last information is not explicitly drawn by the cartoonist into the drawing but understood by the reader according to the position of the elements in the images. Speech balloons are placed in a way that helps the reader to associate them with comic characters and follow the story. Panel, balloon and comic character positions are the three information required to associate speech

[1] Milton Griepp's White Paper, ICv2 Conference 2014

(a) Rectangular speech balloon with tail and gray background

(b) Wavy speech balloon with tail and yellowed background

(c) Oval speech balloon with tail and clear background



(d) Unusual thought balloon with tail and complex background

Fig. 1. Pixel-level speech balloon segmentation results (red line) for different balloon shape and background.

balloons and comic characters towards comics understanding. Panel extraction is the easiest task in comics image analysis and several studies exceed 80% recall and precision [20, 10]. Balloon extraction attracted little attention even-though it is an helpful information for text extraction and essential for text/graphics association. Comic character extraction is at its early stage and the information of speech balloon positions together with their tail can be very helpful for such complex graphics extraction [18]. Pixel-level speech balloon extraction appears to be important for further processing such as balloon contour and tail analysis compared to usual bounding boxes (Fig. 1).

We propose a text-independent approach for speech balloon extraction appropriated for extracting any closed balloon (balloons with a fully connected outline, the most common type of balloon). We base our approach on the observation that speech balloons are highly contrasted region which contain aligned elements (text property). In the following, we will use the word *comics* to designate all types of comics including *Manga* (Japanese comic art).

Balloons (or bubbles) are key elements in comics, they contain most of the textual information and go pairwise with comic characters (speakers). Few works about balloon extraction have been done until now and mainly closed speech balloons have been studied. These works are based on connected-component analysis. Arai [1] proposed first a white blob detection method using four filtering rules related to manga image analysis. The rules are based on blob size, white pixel occurrences, presence of vertical spaces and width to length ratio. Later our group uses HSV color space to make a first selection of bright blobs and then consider as balloons the blobs with a ratio between the text area and the blob bounding box higher than 60% [8]. Recently, Liu [12] proposed to learn text and balloon features in order to recognize any closed speech balloons with a white background.

Open balloon extraction (balloons with partial outline) attracted little attention, a first attempt have been proposed by our team by inflating a contour around text regions [16]. Liu *et al.* proposed a systematic blob classification method (SVM) based on both shape and text properties (two for each) [11]. Both approaches require text positions as input.

Section 2 presents the proposed speech balloon segmentation method. Section 3 the experiments we performed. Finally, Sections 4 and 5 discuss and conclude this work respectively.

2 Speech balloon extraction

From the five approaches reviewed in Section 1, the first approach have been developed especially for manga and therefore has several weaknesses for other types of comics [1]. First, the extraction of the connected-components (CC) requires a binary image which is obtained by using a global threshold. This constraint limits its application to images with a clear background color (which tends to white). Second, balloon candidates selection is performed using several heuristics which are not validated and specific to manga. The method proposed by Ho [8] can be very efficient for a particular comics type but the set of parameter makes it not adaptive to all styles of comics and manga (e.g. heuristic of minimum percentage of text inside balloons). Liu *et al.* method achieve very high quality results but it requires speech balloons with a white background (fixed binarization threshold) and it is dependent of the training set (supervised). The two works concerning open balloon extraction requires text positions as input which is a strong constraint because of error propagation issue (from text extraction), but have the advantage to retrieve closed balloons as well [16, 11].

We propose to overcome these limitations by using an unsupervised method based on adaptive thresholding in order to first binarize the gray-level image, extract and then analyze the connected-components. The advantage of using a local and adaptive threshold selection method for comics segmentation has already been demonstrated in our previous work [17]. As for document image analysis in general, it limits original strokes to be broken (Fig. 2). After having extracted all the CC, we select only the ones with a particular color, topology and shape, independently from size and script (written signs) and compute an overall confidence value that is used for the final speech balloon/non-speech balloon classification.

2.1 Adaptive threshold selection

During comics or manga creation process, balloon outlines are first drawn using a black stroke and then filled with text [4]. We propose to rely on these two information which are intrinsic to their design process and thus characteristics of speech balloons.

Speech balloon outlines are intentionally created in a continuous way by the artist whether they are straight or curved (single stroke). Sometimes, they appear to be degraded when reaching the final reader due to image digitization or

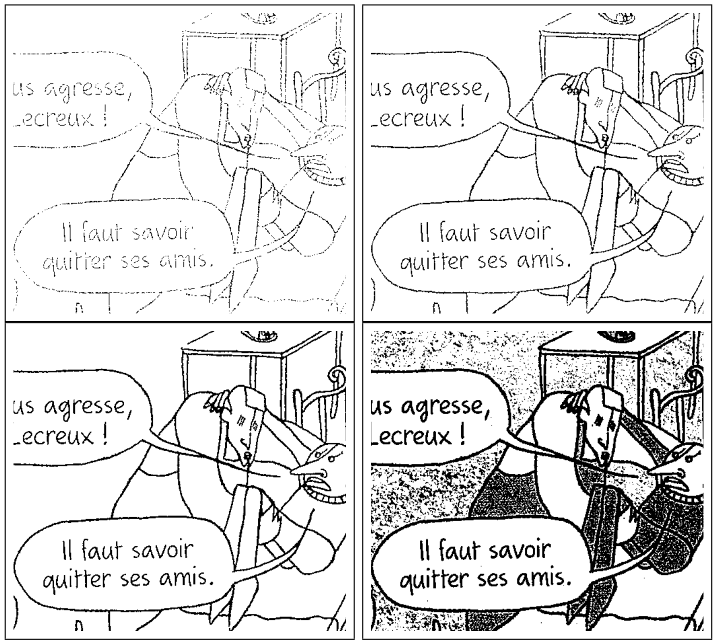
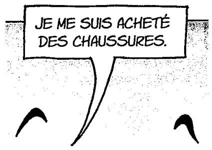


Fig. 2. Binarization results of a 8 bits gray image at different threshold levels from the lower (top-left) to the higher (bottom-right) with threshold 50, 100, 150, 200. We observe that the black strokes are broken at a low threshold level and the background starts to appear as salt and paper noise due to paper texture at high threshold value [17]. The best binarization corresponds to threshold 150 in this example. Images credits: [19].

compression. A perfect outline segmentation and connected-component extraction would result in one single connected-component per balloon outline (Fig. 2).

However, complex background regions complicate this step. There are several adaptive threshold selection method in the literature [9], however, speech balloon being a highly contrasted region it facilitates the separation of background (usually white) and content (black text) in this region. The main difficulties are the shape and the size of the sliding window which is used to determine if a pixel belongs to the background or foreground for each local regions. Having no *a priori* information about speech balloon location, shape and size, we define a squared window of size $blockSize$ relative to the image size. We define the threshold value $T(x, y)$ as the mean of the squared $blockSize * blockSize$ neighborhood of (x, y). The corresponding pixel at position (x, y) is considered as part of the foreground if its gray-value is above T or background else. Fig. 3 shows binarization results of comics and manga from different nature and definition using this approach.

After image thresholding, we obtain a binary image from which we extract and analyze the relationship between white and black connected-components



(a) Recent French comics (A4 page digitized at 300DPI)

(b) Old French comics (A4 page digitized at 300DPI)

(c) Recent manga (digital-born with JPEG compression 150DPI)

(d) Golden age American comics (A4 page digitized at 300DPI)

Fig. 3. Binarization results of comics and manga from different nature and definition. Original images are in Fig. 1.

(CC) using a connected-component labeling algorithm proposed by Suzuki [21]. The CC are split into two sets according to their color in order to ease further processing (speech balloon outline/content separation). White and black sets are called W and B respectively. Note that other region extraction techniques (e.g. MSER [5]) could be used but an extra processing should be added as we also need the topological relations.

2.2 Balloon candidate selection

As introduced earlier, speech balloons can be seen as speech text containers which involves text to be inside speech balloons. From a topological point of view, the first region including speech balloon content (text) is speech balloon background (e.g white region surrounded by a black outline). Note that balloon background regions are assumed to be part of set W because there are usually brighter than their content (part of set B). We propose to combine color and topological feature for selecting from the set W, only the CC (parents) containing other CC (children) which are part of set B. We call this subset of parents the "balloon candidates" CC (Fig. 4). Note that the biggest white CC, which usually corresponds to image background, is ignored as well as very small regions (whose size is lower than 0.5% of image size) usually caused by paper texture or image compression region binarisation.

In the next subsection, shape and spatial organization of each balloon candidates are analyzed in order to determine if it contains text-like information (CC which are aligned).

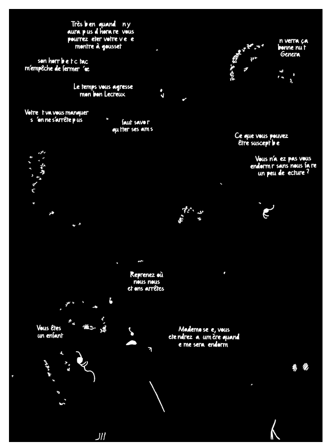
2.3 Balloon candidate analysis

We propose to analyze the content organization of each balloon candidate in order to determine if it contains speech text-like information (assuming speech

(a) Original image

(b) Binary image

(c) Set of parents (W)



(d) Set of children (B)

Fig. 4. Sets of balloon candidate (W) and theirs content (B) for a given image. Corresponding CC are represented in white for both sets. Image credits: [19].

balloons contain speech text). Speech text has several characteristics, some from the text domain and other from comics domain. Text information has some characteristics which are language independent like alignment and equally separated glyphs with noticeable contrast to their background, constant stroke width (thickness), similar color and sizes [2]. When text is used for speech in comics, it is most of the time aligned and centered inside a bubble-like region but glyph space and stroke width are not always stable because text may be handwritten. The difficulty for comics analysis is the important amount of graphics that also consist of aligned elements and confuses balloon/non-balloon separation process (e.g. roofing tile, grass, hairs and face). Such graphics could be disguarded by

(a) 0/18 aligned (b) 7/18 aligned, one line found (c) 14/18 aligned, two lines found (d) 16/18 aligned, three lines found

Fig. 5. Children CC alignment scanning process. The process stops automatically when there is less remaining CC than discovered lines (e.g. in sub-figure 5(d), three lines have been discovered and only two children are remaining).

using an OCR system but currently, commercial OCR systems are not accurate enough for many handwritten comic fonts.

In the following approach, we combine inside balloon CC alignment and balloon shape to compute a confidence value for each balloon candidate. The confidence value is used for the final balloon/non-balloon candidate decision (Section 3).

Content alignment The children are supposed to be horizontally or vertically aligned according to the language (e.g. vertical for Japanese, horizontal for English or French). This is a characteristic of speech text in comics and also for text in general. We propose to "scan" each candidate content and compute the percentage of children which are aligned called inter-child alignment value ($cAlign$). The scanning direction (top-to-bottom or left-to-right) is defined manually here for simplifying but it could be defined automatically by applying both direction separately and then select the one that provides the best results (Section 2.3).

We compute the percentage of aligned children in a specific order, from the longest to the shortest line of aligned children in order to find the longest lines first (the most representative of text-like information). Two children are considered as horizontally aligned if the center of one of them passes through the other. For instance, for two children A and B and their respective center (cx, cy), B is horizontally aligned to A if $A.ymin < B.cy < A.ymax$. Similarly, vertical alignment is computed by replacing y by x. Note that three children minimum are required to compute a relevant alignment.

The process stops automatically when there is less remaining non-aligned children than the number of lines already found. In this manner, we ignore non-aligned children corresponding to punctuations, accents, disconnected-components, etc. (Fig. 5).

Shape analysis Speech balloon shape is similar to a bubble that contains elements (mainly text). It is usually bounded by an outline stroke that has some irregularities throughout its perimeter, the two biggest ones being in the region of the tail [14]. We propose to measure the overall convexity of the outline in order to find how similar to a perfect bubble the balloon candidate is. Several measures of convex polygon exist in the literature [3], some are based on area and others on perimeter analysis. We select a perimeter based approach as it is the most deformable part of the shape. The convexity measure ($cShape$) 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 (Eq. 1).

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

Note that the convexity measure is equal to 100% for perfect rectangles, squares, ovals, circles etc. This could be the case for speech balloon without tail but not for the others because tail region is usually non convex in order to be easily distinguishable. However, tail perimeter is small compared to the overall perimeter of a speech balloon and has a minor impact on the proposed convexity measure. Tail regions could be detected and removed using a tail detection approach [14] but we have preferred to avoid error propagation issue in this process and decided do not use it.

Confidence value The global confidence value C is computed for each balloon candidate, from inter-child alignment $cAlign$ and shape $cShape$ measures according to Formula 2.

$$C = cAlign * \alpha + cShape * \beta \tag{2}$$

where α and β are two weighting parameters defined experimentally (Section 3.2).

3 Experiments

In this section we evaluate the proposed method of speech balloon segmentation using two public datasets and compare our results to other approaches from the literature.

3.1 Datasets

We evaluate the proposed method using the two public datasets eBDtheque [7] and Manga109 [13] in order to show the robustness of the proposed method for French comics, English comics and Japanese manga (script independence evaluation).

The eBDtheque dataset was designed to be as representative as possible of the comics diversity, it includes few pages of diverse albums. It is composed by one hundred images which are composed by 850 panels, 1550 comics characters, 1092 balloons (84.5% are closed) and 4691 text lines. It contains images scanned from French comic books (46%), French webcomics (37%) with various formats and definitions, public domain American comics (11%) and unpublished artwork of manga (6%). In addition to the diversity of styles, formats and definitions, there are also differences in design and printing techniques since 29% of the images were published before 1953 and 71% after 2000. This dataset provides pixel-level object regions as part of the groundtruth.

Manga109 dataset is a selection of 109 manga titles published from the 1970s to the 2010s (21142 images in total). They have been selected from the archive "Manga Library Z" run by J-comi[2]. "*The Manga109 dataset covers various kinds of categories, including humor, battle, romantic comedy, animal, science fiction, sports, historical drama, fantasy, love, romance, suspense, horror, and four-frame cartoons*" [13]. Only text transcription of four titles is publicly available as groundtruth but the authors provide groundtruthing tools. This dataset being quite huge, pixel-level speech balloon groundtruth creation would require a lot of time. We preferred to select few titles and manually count the number of correct/incorrect/missed speech balloons in each image. We randomly selected a subset of three titles that represent 408 images and 3242 balloons in total ("Momoyama Haikagura", "Tetsu San" and "Ultra Eleven").

3.2 Confidence value validation

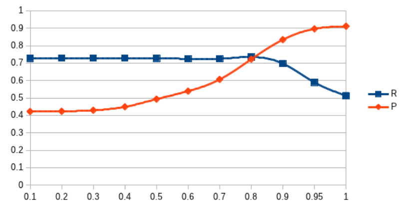
The two measures proposed in Section 2.3 have been evaluated separately and combined in order to define the best weighting parameters α and β from Formula 2 (see Fig. 6). The best performance has been achieved for $\alpha = 0.75$ and $\beta = 0.25$.

3.3 Performance evaluation

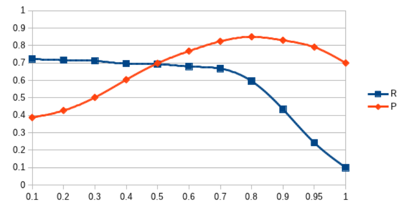
We evaluate the performance of the proposed method at pixel-level and bounding box level for eBDtheque dataset in order to give both precise and comparable results with other methods from the literature providing only bounding box level results. However, the results over the Manga109 subset have only been evaluated visually at object level because of the lack of speech balloon position information in the provided groundtruth. The latter has been performed for speech balloon candidates with a minimum confidence value $C \geq 80\%$ in order to evaluate the performance of the 20% best results provided by the proposed method. Concerning adaptive threshold selection, the $blockSize$ was defined as a square of area 1.3% of the image area according to the validation on the eBDtheque dataset [7].

For pixel-level evaluation, balloon candidates pixels were considered true positives TP if they were corresponding to a balloon pixel in the groundtruth or false positives FP. The number of TP, FP and false negative (missed pixels) FN were used to compute the recall R and the precision P of each of the methods using Formula 3 and 4. We also computed the F-measure F for each result.

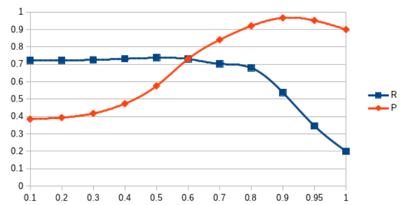
[2] http://www.j-comi.jp/



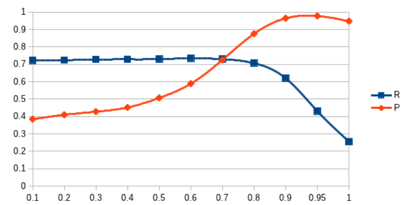
(a) $\alpha = 1.0$ and $\beta = 0.0$



(b) $\alpha = 0.0$ and $\beta = 1.0$



(c) $\alpha = 0.5$ and $\beta = 0.5$



(d) $\alpha = 0.75$ and $\beta = 0.25$

Fig. 6. Confidence value weighting parameter validation (α and β). Vertical axis represents the percentage of recall R and precision P while the horizontal axis is the minimum confidence required for a balloon to be considered as true positive.

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

Concerning bounding box level evaluation, we use the same metric as the PASCAL VOC challenge for visual objects [6]. The detections were assigned to ground truth objects and judged to be true or false positives by measuring bounding box overlaps between detected and groundtruth regions. To be considered as a correct detection, the overlap ratio a_0 between the predicted bounding box B_p and the ground truth bounding box B_{gt} must exceed 0.5 (Formula 5). According to the PASCAL VOC challenge, the predicted objects were considered as true positive TP if $a_0 > 0.5$ or false positive FP otherwise. Missed balloons were counted as false negatives (FN). Note that bounding box level evaluation is less accurate than pixel-level evaluation, especially for speech balloon having a long tail (Fig. 7).

$$a_0 = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \tag{5}$$

(a) Considered as FP at bounding box level ($a_0 < 0.5$)

(b) Considered as TP at pixel level ($>$ 90% pixels match)

Fig. 7. Difference of performance evaluation results for a same balloon extraction using bounding box level groundtruth (left) and pixel level groundtruth (right). Red and blue polygons correspond to detection and groundtruth regions respectively.

3.4 Results analysis

For the first dataset (eBDtheque), we compared the proposed method to other methods from the literature at pixel and bounding box (BdB) levels. Result details are given Table 1. Note that Liu *et al.* [11] reported results at bounding box level only and using a slightly stricter ratio ($a_0 > 0.6$).

Table 1. Average speech balloon segmentation performance in percent for eBDtheque.

	eBDtheque dataset					
	Pixel level			BdB level		
Method	R	P	F_1	R	P	F_1
Arai [1]	18.70	23.14	20.69	13.40	11.76	12.53
Ho [8]	14.78	32.37	20.30	13.96	24.76	17.84
Rigaud [16]	69.81	32.83	44.66	52.68	44.17	48.05
Liu [11]	–	–	–	**90.10**	**86.90**	**88.50**
Proposed	**70.71**	**87.62**	**78.24**	72.21	83.31	77.36

At pixel level, the proposed method provides the best results compared to other method from the literature, including a method proposed by our team which uses a simplified scenario by requiring text positions as input [16]. The recall of 70.71% has been measured over all eBDtheque dataset which contains 15.5% of open balloons (non detectable by the proposed approach). The remaining 13.79% of errors are due to small balloons containing few information or graphics (Fig. 8(a)). The regions that confuse the proposed approach (drop of precision) are often composed by illustrative text or text-like elements (Fig. 8(d) and 8(e)).

Concerning bounding box level evaluation, the proposed approach exceeds our previous approach [16] but not the state of the art methods [11]. Nevertheless, the proposed method has the advantage do not require text position as input.

Table 2. Average speech balloon segmentation performance in percent for Manga109.

	Manga109 dataset		
	Pixel level		
Method	R	P	F_1
Proposed	72.24	89.71	80.04

Concerning the Manga109 dataset subset, we report an average performance of the proposed method for recall, precision and F-measure of 72.24%, 89.71% and 80.04% (see Table 2). The overall performance is similar to eBDtheque dataset which confirms that the proposed method is non sensitive to comics style variations (comics vs manga). The same limits as eBDtheque dataset have been observed: recall and precision drops are mainly due to open balloons (Fig. 8(b)) and illustrative text in balloon like regions (Fig. 8(f)).

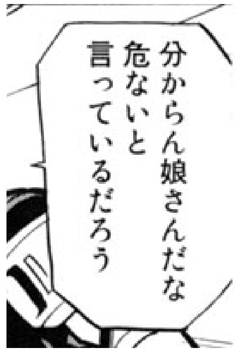
More detailed results and extra material are available on GitHub[3].

4 Discussion

The proposed method uses a simple adaptive thresholding approach which is efficient for the studied scope because speech balloon regions are composed by a background which is easy to binarize when taken apart from the rest of the image (local thresholding). Note that the proposed approach assumes speech balloon background to be brighter than their content, if this is not the case, image color inversion should be applied before binarizing. Speech balloon candidate selection is based on content analysis which sometimes ends up with false positives because other graphics have a similar-to-text organization. Open balloons can be included into this approach by analyzing page background content (contains open balloon text). Note that the presented approach provides balloon background regions (white regions) but some processing require balloon outline (e.g. tail direction retrieval). In such situation, the proposed results have to be post-processed in order to find the external edge of the outline stroke like [12]. This work focus on comics analysis but we believe it can also be applied on other images having strong relations between text and graphic shapes such as engineering flowchart, car plates and road signs.

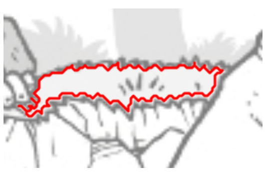
[3] `https://github.com/crigaud/publication/tree/master/2016/LNCS/text-independent_speech_balloon_segmentation_for_comics_and_manga`

(a) Less than three aligned element are inside the balloon (two only)

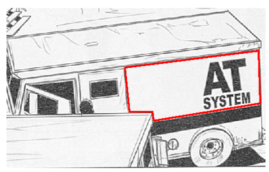


(b) Open balloons not detected by the proposed method

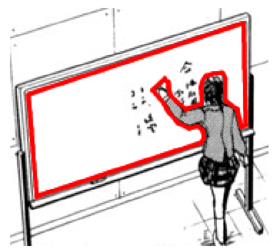
(c) Outline-free balloon not detected by the proposed method



(d) Grass is aligned as text in an balloon-like region



(e) Illustration containing text (truck cover)



(f) Illustration containing text (blackboard)

Fig. 8. Examples of failure cases of the proposed approach that drops down recall on the first row (FN) and precision on the second row (FP). Detected balloons are represented by a red line in the second row.

5 Conclusion

This paper presents an unsupervised speech balloon segmentation approach toward comics and manga text/graphic association. The proposed method combines color, shape and topological connected-component relationship analysis in order to segment speech balloon contours at the level of pixels. We also proposed a segmentation confidence measure mixing speech balloon content and shape which can be useful for several applications. The proposed approach has been tested over most of the comics types with promising performance from two public datasets eBDtheque [7] and Manga109 [13]. In the future, we will include open balloons segmentation in the method.

Acknowledgment

This work was supported by the University of La Rochelle (France), the town of La Rochelle and the PIA-iiBD ("Programme d'Investissements d'Avenir"). We are grateful to all authors and publishers of comics and manga images from eBDtheque and Manga109 datasets for allowing us to use their works.

References

1. Arai, K., Tolle, H.: Method for real time text extraction of digital manga comic. International Journal of Image Processing (IJIP) 4(6), 669–676 (2011)
2. Bigorda, L.G., Karatzas, D.: A fast hierarchical method for multi-script and arbitrary oriented scene text extraction. CoRR abs/1407.7504 (2014), `http://arxiv.org/abs/1407.7504`
3. Chalmeta, R., Hurtado, F., Sacristn, V., Saumell, M.: Measuring regularity of convex polygons. Computer-Aided Design 45(2), 93 – 104 (2013), `http://www.sciencedirect.com/science/article/pii/S0010448512001650`, solid and Physical Modeling 2012
4. Cyb: Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels, pp. 128–153. William Morrow Paperbacks (2006)
5. Donoser, M., Bischof, H.: Efficient maximally stable extremal region (mser) tracking. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. vol. 1, pp. 553–560. IEEE (2006)
6. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision 88(2), 303–338 (2010)
7. Guérin, C., Rigaud, C., Mercier, A., al.: ebdtheque: a representative database of comics. In: Proceedings of International Conference on Document Analysis and Recognition (ICDAR). pp. 1145–1149. Washington DC (2013)
8. Ho, A.K.N., Burie, J.C., Ogier, J.M.: Panel and Speech Balloon Extraction from Comic Books. 10th IAPR International Workshop on Document Analysis Systems pp. 424–428 (Mar 2012)
9. Lamiroy, B., Ogier, J.M.: Analysis and interpretation of graphical documents. In: Doermann, D., Tombre, K. (eds.) Handbook of Document Image Processing and Recognition. Springer (2014)
10. Li, L., Wang, Y., Suen, C.Y., Tang, Z., Liu, D.: A tree conditional random field model for panel detection in comic images. Pattern Recognition 48(7), 2129–2140 (2015), `http://dx.doi.org/10.1016/j.patcog.2015.01.011`
11. Liu, X., Wang, Y., Tang, Z.: A clump splitting based method to localize speech balloons in comics. In: Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR). pp. 901–906. IEEE (2015)
12. Liu, X., Li, C., Zhu, H., Wong, T.T., Xu, X.: Text-aware balloon extraction from manga. The Visual Computer pp. 1–11 (2015), `http://dx.doi.org/10.1007/s00371-015-1084-0`
13. Matsui, Y., Ito, K., Aramaki, Y., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. CoRR abs/1510.04389 (2015), `http://arxiv.org/abs/1510.04389`

14. Rigaud, C., Guérin, C., Karatzas, D., Burie, J.C., Ogier, J.M.: Knowledge-driven understanding of images in comic books. International Journal on Document Analysis and Recognition (IJDAR) 18(3), 199–221 (2015), http://dx.doi.org/10.1007/s10032-015-0243-1
15. Rigaud, C., Karatzas, D., Burie, J.C., Ogier, J.M.: Adaptive contour classification of comics speech balloons. In: Lamiroy, B., Ogier, J.M. (eds.) Graphics Recognition. Current Trends and Challenges, Lecture Notes in Computer Science, vol. 8746, pp. 53–62. Springer Berlin Heidelberg (2014), http://dx.doi.org/10.1007/978-3-662-44854-0_5
16. Rigaud, C., Karatzas, D., Van de Weijer, J., Burie, J.C., Ogier, J.M.: An active contour model for speech balloon detection in comics. In: Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR). pp. 1240–1244. IEEE (2013)
17. Rigaud, C., Karatzas, D., Van de Weijer, J., Burie, J.C., Ogier, J.M.: Automatic text localisation in scanned comic books. In: Proceedings of the 8th International Conference on Computer Vision Theory and Applications (VISAPP). SCITEPRESS Digital Library (2013)
18. Rigaud, C., Le Thanh, N., Burie, J.C., Ogier, J.M., Iwata, M., Imazu, E., Koichi, K.: Speech balloon and speaker association for comics and manga understanding. In: Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR). pp. 351–356. IEEE (2015)
19. Roudier, N.: Les terres creusées, vol. Acte sur BD. Actes Sud (2011)
20. Stommel, M., Merhej, L.I., Müller, M.G.: Segmentation-free detection of comic panels. In: Computer Vision and Graphics, pp. 633–640. Springer (2012)
21. Suzuki, S., et al.: Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing 30(1), 32–46 (1985)