

# Panel and speech balloon extraction from comic books

Anh Khoi Ngo ho, Jean-Christophe Burie, Jean-Marc Ogier  
*Laboratoire L3i, University of La Rochelle,*  
*Avenue Michel Crepeau, 17042 La Rochelle Cedex 1, France*  
*anh.ngo\_ho@etudiant.univ-lr.fr; jean-christophe.burie@univ-lr.fr; jean-marc.ogier@univ-lr.fr*

**Abstract**—Comic books represent an important cultural heritage in many countries. However, few researches have been done in order to analyse the content of comics such as panels, speech balloons or characters. At first glance, the structure of a comic page may appear easy to determine. In practice, the configuration of the page, the size and the shape of the panels can be different from one page to the next. Moreover, authors often draw extended contents (speech balloon or comic art) that overlap two panels or more. In some situations, the panel extraction can become a real challenge. Speech balloons are other important elements of comics. Full text indexing is only possible if the text can be extracted. However the text is usually embedded among graphic elements. Moreover, unlike newspapers, the text layout in speech balloons can be irregular. Classic text extraction method can fail. We propose, in this paper, a method based on region growing and mathematical morphology to extract automatically the panels of a comic page and a method to detect speech balloons. Our approach is compared with other methods find in the literature. Results are presented and discussed.

**Keywords**-comic book; comic panel extraction; speech balloon extraction; comic page segmentation; region growing; mathematical morphology

## I. INTRODUCTION

Born in the 19th century, comics spread worldwide and became an important industry. Nowadays, the development of new devices such as PDA or tablet PC gave birth to e-comics, specially designed for this kind of devices. However, *traditional* comic books represent an important cultural heritage in many countries. For example, the CIBDI [1] (French acronym for International City of Comics and Images) has already digitized tens of thousands comic albums. Till now, few researches have been carried out in order to exploit the content of digitized comics.

The characteristics of comics are quite different compared with newspapers or magazines. At first glance, the structure of a comic page may appear easy to extract. In practice, the configuration of the page, the size and the shape of the panels can be different from one page to the next. Even if some conventions have been established for the design of comics, authors have a certain liberty when drawing comic pages. With a complex and relatively free layout, the application of methods used for other media (newspapers, magazines) is inefficient on comic images.

Different approaches have been published recently to detect and extract panels in comics. Han [2] used a X-Y re-

cursive cut algorithm associated to a multi-layer perceptron to extract, respectively, candidate positions and verification. Ponsard and Fries [3] used watershed segmentation to detect panels. Chan and al [4] identify uniform colour stripes and used them as separators to segment colour comic page images. Then, he adds an elimination process to identify false stripes and to improve the accuracy of the comic cuts. Tanaka [5] has developed a recursive method using the density gradient to divide comic pages in panels. Ishii [6] has extended this iterative separation method utilizing consistency of density gradient, and has proposed a panel corner detection method. Most of these methods are effective on comics where :

- the content is drawn inside the panel, but some methods can deal with some extended contents (speech balloons, characters,...).
- the shape of the panels is geometrical (rectangle, square)
- the layout is regular

From our point of view, panel extraction is essential because

- each panel contain a key moment of the story
- the reading order can be deduced from the layout.
- panels can then be further analysed to extract features used for content based indexing and retrieval.

But the extraction method has to be robust enough to take into account the diversity of comics in term of size, shape and extended contents.

Speech balloons are also important elements where the text of the story is printed. These balloons are inserted in panels among the drawings. Sometimes, they can also overlap two panels. An approach to extract these balloons consists in detecting the text inside. Many methods have been proposed to extract text from document image or to separate text and graphic. One of the best known algorithms, based on the analysis of the connected components, was proposed by Fletcher and Kasturi [7]. Some improvements have been proposed as in [8]. Another well-known method, called RLSA, was proposed by Wong and al [9]. To our knowledge, few approaches have been published to extract the speech balloons from comic books. Arai [10] considers that a balloon has a white background and suggests to extract speech balloon as a white pixel area. Some rules are then applied to select the balloons among the candidate areas.

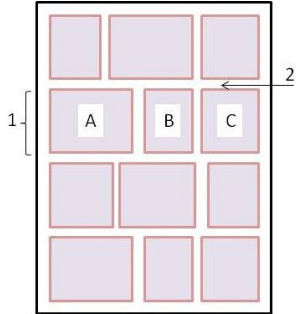


Figure 1. Example of plate : (1) strip ; (A) , (B) , (C) panels ; (2) gutter

These methods often fail because comics are very complex.

Here, we propose a method based on region growing and mathematical morphology to extract automatically the panels of a comic page. The method to extract speech balloon is also based on mathematical morphology and connected component analysis. The paper is organised as follow. Section II presents an overview of comics. A review of panel extraction methods is given in section III and a special attention is given to the case of extended contents. Sections IV and V described respectively the proposed methods for panel and speech balloon extraction. Experimental results are presented in section VI. Finally, the section VII concludes this paper.

## II. OVERVIEW OF COMICS

Comics can be found throughout the world. The most famous comics are published in Europe, America or in Japan. If the style is quite different, the comics structure is somewhat similar.

### A. Comics vocabulary

A comic book (or *album*) consists of pages also called *plates* by cartoonists. In traditional comics, a plate (*cf.* figure 1) is split up into *strips*. A strip is sequence of *panels*. Each panel consists of a single drawing depicting a frozen moment of the story. It contains drawings (characters, scenery, ...) and text (speech balloon, narrative box, ...). Panels are usually separated one another by gutters whose the colour is the one of the plate. However, sometimes, some drawings go out the panel and overlap the adjacent panel. This drawing can be a character, a speech balloon or a comic art (for example an arrow). This kind of drawing is called, in this paper, extended object or overlap object. In some comic books, it is the panel itself which can overlap another panel.

### B. Type of comics

There is a variety of comics with different layouts. Automatic analysis of all these comics is a real challenge. In this work, we consider only traditional comics with uniform colour plates. Among them, we have defined 3 types of comics. The type *simple* consists of pages with regular panels totally separated by gutters (*cf.* figure 1). The

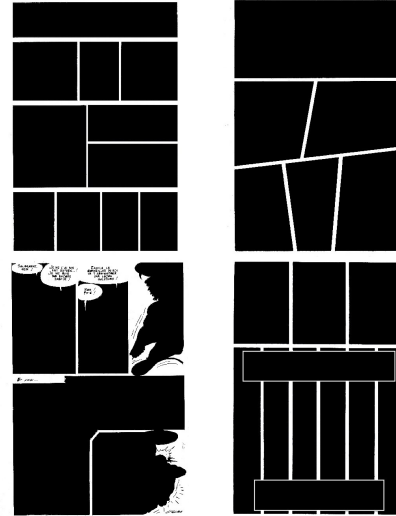


Figure 2. Examples of comic layout

type *complex* consists of regular panels with some extended objects. The last type *hard* contains free-style panels with complex overlaps of some panels. Our approach will be tested on the 3 types of comics we have defined. The aim of our method will be to recognise area corresponding to panels, whatever their size, their shape or the presence of extended contents between panels.

## III. REVIEW OF PANEL SEGMENTATION METHODS

In traditional comics, the layout of the panels is often in a grid. However, this grid is usually not regular. Figure 2 shows 4 examples of layout (upper left : regular, upper right and bottom left : irregular, bottom right : overlap of panels)

### A. Case of regular layout

Tanaka [5] analysed layout of comic images using density gradient. The page is divided recursively into panels to extract the structure of the plate and to determine the reading order. However the method assumes that quadrangle regions (the panels) are filled before applying the density gradient but this pre-processing step isn't detailed by the authors. Chan and al. [4] propose to find uniform colour lines (called gutter above *cf.* II-A). When a line is found, the page is divided into sub-regions. Each sub-region is processed recursively until no more uniform colour lines are found. This is achieved by checking the intensity histogram of the pixels along each line. But false lines can be found if the panels contain regions with uniform colour.

### B. Case of irregular layout

Usually, the layout is not regular as shown on figure 3. In this case, methods based on gradient density fail because the gutters between panels are difficult to identify as one can see on the horizontal and vertical histograms of gradient

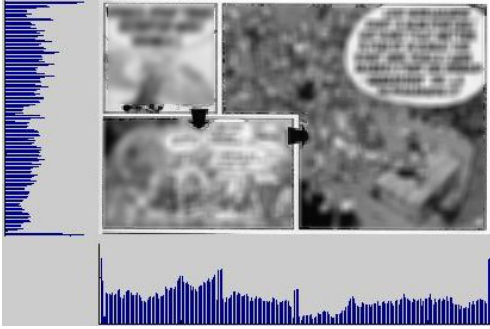


Figure 3. Irregular grid with overlap objects [11]

density. The gutter isn't long enough to be detected by the histogram. Arai [10] has proposed a method that can be applied on regular and irregular layout. His approach considers that each panel can be detected as a single blob object. He considers also that the rest of the page : the background is a single blob object of connected white pixels. First a binarisation with fixed threshold is used to extract the background of the plate. The author assumes that the colour of the background is always white. Then, he applied connected component labelling algorithm to detect single blob objects which, in other words, correspond to panels.

### C. Case of extended contents

Extended contents are often drawn by artists whatever the comic layout : regular or not. Figure 3 shows two examples of comic arts (arrows) which overlap two adjacent panels. Arai's method [10] fails in this case because his method links the 3 panels together. To take into account overlap objects, Arai proposes a checking process for each panel. It consists in detecting the beginning and the end of the horizontal gutters located respectively on the left and the right edge of the page. The same reasoning can be used for the vertical gutters. A white pixel line is added in order to break the blob objects. The connected component labelling method is then applied to detect new panels. However, this approach doesn't work on horizontal gutter of the figure 3, since the end of the gutter cannot be detected from the right border.

## IV. PANEL EXTRACTION

Our approach consists in extracting the panels of comic pages and solving the problem of extended contents in order to determine the page structure.

### A. Step 1 : binarisation phase

In this work, we consider only traditional comics with uniform colour plates. The background is usually white. However, for old comics, the colour can be different due to colour degradation, yellowing of the paper. Like Arai [10], the first step of our method is to extract the background of the page. But, instead of using a simple thresholding,

region growing algorithm is applied to accomplish this task. Region growing [12] is a simple region-based image segmentation method. This approach examines neighbouring pixels of initial "seed points" and determines whether the pixel neighbours should be added to the region. The process is iterated on, in the same manner as general data clustering algorithms. A five pixel frame is defined on the perimeter of the page and the average colour is computed in this area. Four seed points are defined respectively in the top, right, bottom and left area of the frame to initiate the process. Each seed point corresponds to a pixel with a colour close to the average colour. When the region growing algorithm stops, the background is filled in white and the rest of the page in black. All the panels are, thus, black solid blocks. Figure 4 (a) shows the original image and figure 4 (b) the result of the segmentation with region growing method.

### B. Step 2 : panel checking

Panels are now represented as black solid blocks. However, when extended contents are presents adjacent panels can be linked. Mathematical morphology [13] is used to transform the blocks and to break the links. The main idea in binary morphology is to compare an image with a pre-defined shape (structuring element) drawing conclusions on how this shape fits or misses the shapes in the image. In this work, a 3x3 square is used. The basic morphological transformation called dilatation allows to shrink black objects of a binary image and to expand white ones. To break the link, it might be necessary to apply  $N$  dilatation operations. A stopping criterion based on the size of the black objects is used to stop the dilatation process. Experimentally, a width (resp. a height) of 1/6 of the width (resp. height) of the page has been fixed. To give each black object their initial size again, the same number  $N$  of erosion is applied. Erosion is the opposite of dilatation. It allows to expand black objects and to shrink white ones. Figure 4 (c) shows the result after applying  $N$  dilatations : the links have been broken. Figure 4 (d) presents the result after  $N$  erosions : the original size of the black object has been restored.

### C. Step 3 : Connected component labelling

The final step consists in applying connected component labelling algorithm. A filtering based on the size of the objects is applied to remove noise. Each black object, i.e. each panel, is then characterized by its position and its size in order to determine the structure of the page.

## V. SPEECH BALLOON EXTRACTION

### A. Brief review

The first idea to extract the speech balloons is to detect the text inside. The Run Length Smoothing Algorithm (RLSA) is a method that can be used for block segmentation and text discrimination [9]. However, to be efficient, the width of the text area need to be large enough and the number of

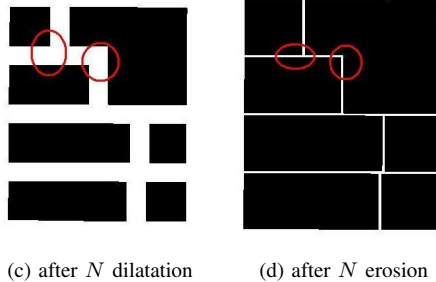
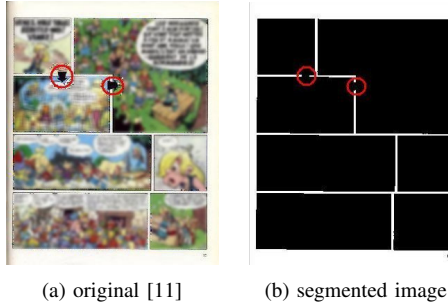


Figure 4. Panel extraction

lines sufficient to locate text blocks. The RLSA is inefficient with comic books because the text in speech balloon is often short, so the number of misdetections is important. With RLSA, we obtain a recognition rate of about 50%.

Arai assumes in [10] that speech balloons are white pixel areas. He is right, whatever the type of comics (manga, european or american comics) the background of speech balloons is almost always white or painted with light colours. First, Arai uses a fixed threshold on the 3 components (Red, Green, Blue) of the image to obtain a binary image and extract candidate areas. Then, balloon selection is based on four rules for classification : (1) minimal size of the candidate areas, (2) minimal ratio of white pixels in the area, (3) presence of two white vertical lines with a minimal length, (4) fixed width to length ratio. This method gives good results. However it has been designed for japanese manga and its sensitivity to noise is important.

### B. Our approach

Speech balloons are usually inside the panels. Even, if a speech balloon can sometimes overlap two panels, this case isn't considered in this work. The balloon detection is applied, here, on each panel detected with the above method.

1) *Candidate area selection*: Candidate areas are selected according to their shape and their colour. RGB image is converted in HSV (Hue, Saturation, Value) colour space. In this colour space, white and light colour areas have a high value in terms of *Value* (V) and low value in terms of *Saturation* (S). So, the first selection is done according to these criteria. To reduce the number of candidate areas, a second selection is applied according the size and the shape of selected areas. First, small regions are removed. Then,

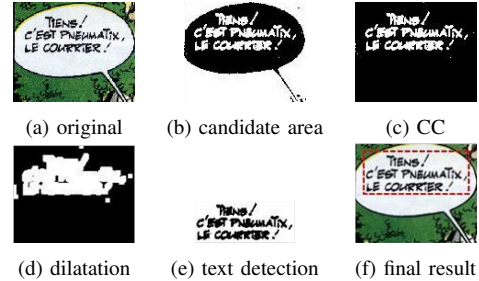


Figure 5. Speech balloon extraction

only regions whose the ratio between the number of pixels of the selected area and the ones of the bounding box is higher than 60% are kept. The result of this step is a binary image where white (or light colour) pixels of the selected areas are represented in black and the rest of the pixels in white (*cf.* figure 5 (b)).

2) *Text detection*: The speech balloons contain text. So, the next step consists in determining if the candidate areas contain text or not. Only connected components (CC) inside the candidate areas are kept (figure 5 (c)). The structure of text is supposed to be regular when the one of the other connected components, considered as noise, is supposed to be irregular. The idea is to link the text elements (characters, words, lines of text), which are close from each other, in order to obtain a text block. A dilatation is applied to expand white pixels and link connected components (figure 5 (d)). After this operation, small connected components are considered as noise and the biggest one is extracted as a text block 5 (e)). To avoid false detections, a filter based on the size of the block is applied.

3) *Balloon extraction*: Each candidate area, where a block has been detected, is labelled as a speech balloon (*cf.* figure 5 (f)).

## VI. EXPERIMENTS AND RESULTS

### A. Panels extraction

A data set with 42 pages extracted from 7 comic books has been created. Each type (*simple*, *complex* and *hard*) is represented by 14 pages. This data set contains 355 panels with various size and shape. This data set is rather small but it was the only one available at the time of the experiment. To evaluate the results, two segmentation rates have been computed. The first one is the success rate for pages. A page is considered to be well segmented if **ALL** the panels of the page have been correctly extracted. This rate allows to estimate the quality of the extracted layout. The second one is the success rate for panels. This rate gives the percentage of well extracted panels among the 355 panels of the data set. Note that in this experiment, the evaluation is based on visual inspection.

In the first experiment, our method has been applied separately on each type of pages. The results are given in

Table I  
SUCCESS RATES BY TYPE

Type	Simple	Complex	Hard
Page (%)	100	85.7	7.1
Panel (%)	100	91.8	55.6

Table II  
COMPARISON WITH OTHER METHODS

Method	Ishi	Tanaka	Arai	Our Method
Page (%)	30.9	42.8	47.6	<b>64.3</b>
Panel (%)	56.3	63.9	75.6	<b>87.3</b>

Table III  
SPEECH BALLOON EXTRACTION

Method	Recall(%)	Precision(%)
Arai	56.80	94.69
Our approach	75.53	94.67

table I. The first line represents the success rate for pages and, the second one, the success rate for panels. A success rate of 100% is reached for pages of type *simple* both for pages and panels. For pages and panels with extended contents (type *complex*), the success rate is also very good with a rate of 91.8% for panels and 85.7% for pages. These results show the pertinence of our approach. For the type *hard*, the success rate drops to 7.1%. Indeed, the segmentation of this type of pages is very difficult. The shape of panels is totally irregular and many objects overlap the panels. Even for a reader, to define the structure of such pages is sometimes not so easy. Due to limited space, we can't show cases where our method fails. Anyway, if we consider only panels, the success rate reaches 55.6% what is promising. In a second experiment, a comparison with other methods (Ishii [6], Arai [10], Tanaka [5]) has been carried out. The segmentation rates for the whole data set are given in table II. Our method obtains the best success rates both for pages and panels.

### B. Speech balloon extraction

The data set used to test our method consists in 150 speech balloons present in 116 panels extracted from 7 comic books. We have compared our approach to the Arai's one. The table III shows the results in terms of recall/precision computed as follow :

$$precision(\%) = \frac{\# \text{ of correctly detected balloons}}{\# \text{ of detected balloons}}$$

$$recall(\%) = \frac{\# \text{ of correctly detected balloons}}{\# \text{ of desired balloons}}$$

With our data set, the proposed approach gives approximately the same precision rate than Arai's method. However, it increases the recall rate significantly. These results show the pertinence of our approach.

## VII. CONCLUSION

In this paper, we proposed two methods to extract panels and speech balloons of comic pages. The first one based on

region merging and mathematical morphology obtains good results in the experimentations both for page segmentation and panel extraction. An effort has to be done to improve the results for pages of type *hard* which are very difficult to analyse. The second method increases the results of the only method find in the literature. We are working on a bigger data set and the corresponding ground truth in order to make them publicly available. Future works will concern the study of the panels in order to analyse their content.

## REFERENCES

- [1] Cibdi : Cité internationale de la bande dessinée et de l'image. [Online]. Available: <http://www.citebd.org/>
- [2] E. Han, K. Kim, H. Yang, and K. Jung, "Frame segmentation used mlp-based x-y recursive for mobile cartoon content," in *LNCS: Human-Computer Interaction. Intelligent Multimodal Interaction Environments*, vol. 4552, 2007, pp. 872–881.
- [3] C. Ponsard and V. Fries, "An accessible viewer for digital comic books," in *LNCS : Computers Helping People with Special Needs*, vol. 5105, 2008, pp. 569–577.
- [4] C. H.Chan, H. Leung, and T. Komura, "Automatic panel extraction of color comic images," *LNCS, Advances in Multimedia Information Processing, PCM 2007*, vol. 4810, pp. 775–784, 2007.
- [5] T. Tanaka, K. Shoji, F. Toyama, and J. Miyamichi, "Layout analysis of tree-structured scene frames in comic images," in *Proc. International Joint Conference on Artificial Intelligence, IJCAI-07*, Hyderabad, India, Jan. 2007, pp. 2885–2890.
- [6] D. Ishii and H. Watanabe, "A study on frame position detection of digitized comics images," in *Proc. Workshop on Picture Coding and Image Processing, PCSJ2010/IMPS2010*, Nagoya, Japan, Dec. 2010, pp. 124–125.
- [7] L. A. Fletcher and R. Kasturi, "A robust algorithm for text string separation from mixed text/graphics images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, pp. 910–918, November 1988.
- [8] K. Tombre, S. Tabbone, L. Plissier, B. Lamiroy, and P. Dosch, "Text/graphics separation revisited," in *Workshop on Document Analysis Systems, DAS2002*, 2002, pp. 200–211.
- [9] K. Y. Wong, R. G. Casey, and F. M. Wahl, "Document analysis system," *IBM Journal of Research and Development*, vol. 26, pp. 647–656, November 1982.
- [10] K. Arai and H. Tolle, "Automatic e-comic content adaptation," *International Journal of Ubiquitous Computing (IJUC)*, vol. 1, no. 1, pp. 1–11, Nov. 2010.
- [11] R. Goscinny and A. Uderzo, "Les aventures d'astérix le gaulois - astérix et les normands," *Editor : Hachette*, 1967.
- [12] S. W. Zucker, "Region growing: Childhood and adolescence," *Computer Graphics and Image Processing*, vol. 5, no. 3, pp. 382–399, 1976.
- [13] J. Serra, *Image Analysis and Mathematical Morphology*. Academic Press, Inc., 1983.