eBDtheque: a representative database of comics

Clément Guérin, Christophe Rigaud, Antoine Mercier,
Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju,
Jean-Christophe Burie, Georges Louis,
Jean-Marc Ogier, Arnaud Revel

Laboratoire L3i
Université de La Rochelle
Avenue Michel Crépeau, 17000 La Rochelle, France
{firstname.lastname, l3i-ebdtheque}@univ-lr.fr

***Abstract*—We present eBDtheque, a database of various comic book images and their ground truth for panels, balloons and text lines plus semantic annotations. The database consists of a hundred pages of various comic book albums, Franco-Belgian, American comics and mangas. Additionally, we present the piece of software used to establish the ground truth and a tool to validate results against this ground truth. Everything is publicly available for scientific use on http://ebdtheque.univ-lr.fr.**

***Keywords*—scientific comics collection, ground truth, semantic annotation, image, database**

I. Introduction

Over the past century, a gigantic amount of comic books has been produced, mostly by United-States, Japan and western Europe. Just like music, movies and books a few years ago, comics books are now on their way to the digital world pulling a noticeable amount of research works regarding enhancement concerns. Perspectives are various, starting from interactive reading, automatic translation and text-to-speech to information retrieval into big comic books databases. A special interest is taken into reading-on-small-devices, such as smartphones and tablets, which seems to be the current background target of most researches. The two main bottlenecks to these goals achievements are the image segmentation and the semantic retrieval. The first problem is addressed in numerous papers with a special focus on panel extraction [1], [2], speech balloons [3], [4] and text recognition [5], [6]. But we are also witnessing a growing interest for the latter. A few works have recently been dedicated to screenplay panels ordering [7]–[9] or automatic content conversion of comics images using a knowledge ontology assisted approach [10].

With the analysis and processing of data comes the need of the output results evaluation. Traditionally, this evaluation is made by validating the results of an algorithm with a ground truth that represents what an ideal output should be [11]–[13]. Ideally, such a ground truth is made publicly available so anyone can challenge his own algorithm to the community [14]. This can be applied to any kind of results from image segmentation to classification or information retrieval.

Being in need of comic books material and an associated ground truth for the ongoing eBDtheque project[1], we noticed that there is not such dataset publicly available for scientific purpose. Therefore, we decided to gather the first comic books database in association with several renowned authors and to build up the corresponding ground truth according to our current concerns which are image segmentation and semantic analysis.

This paper introduces the eBDtheque database. The second section presents the comic books corpus, how it has been selected and what it is made of. The ground truth itself, its construction's protocol, structure and content are detailed in the third section. The fourth part is dedicated to the presentation of the construction and evaluation tools made available to the community Section V concludes this paper.

II. Corpus

Scott McCloud defined a comic as being "juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer" [15]. His definition is voluntary large enough to cover all the different style of comic books that have been produced so far. However, this heterogeneity in style has to be taken into account during the image analysis process for anyone who wants to claim for a sound and robust comic books' elements extraction algorithm. Indeed, the efficiency of an algorithm will depend on its ability to deal with position, shape and style of the elements, as well as their graphical relationship, it is aiming to recognize. For example, for panel detection there are several methods either based on corners and edges [16], connected components [6] or straight lines [2] but they all fail for certain comics' styles. Considering that point, several comic books authors with their own specific style have been asked to provide material for the construction of this ground truth database. We collected a corpus of a hundred pages with the following properties:

- Published between 1905 and 2012. 29 pages of the corpus have been released before 1953 and 71 between 2000 and 2012. The result comes with different degrees of quality, paper degradation and printing process (e.g. 3 color ink dots, ink jet, laser).
- 46 pages, from 14 different albums, have been digitized with a resolution of 600 dots-per-inch into lossless PNG files by the A3DNum[2] company. The

[1] http://ebdtheque.univ-lr.fr/context/

[2] http://www.a3dnum.fr

others 54 have been downloaded straight from 11 web comics websites with JPEG loss compression. The resolution varies from 72 to 300 dots-per-inch, resulting in heterogeneous image quality properties.

- 72 of these pages are colorful (tint areas, watercolors, manually or computer-assisted), shades of gray have been used for 16 of them and 12 were printed in black and white. One album provided both colored and black and white pages so image analysis algorithms can be tested and tuned with or without colors out of the equation. Additionally, 5 images are double pages. Moreover, each page has different structure and content.
- Panels' frame shapes are various as well. While most of the panels are fully fenced with a black line, a noticeable amount of them are only half-closed, i.e. a part of the panel is indistinguishable from the page background. On two pages, panels are not even out-framed with a black line, their boundaries being materialized by the difference in color between the panel and the background. Nine pages contain overlapping panels and a lot more of them contain couples of panels sharing one or more pictorial objects. Finally, 12 pages contain only full frame-less panels.
- Balloons and text shapes are also quite heterogeneous. 33 pages contain text out of any balloon, 8 of them don't contain any balloon at all. Balloons can be closed or half-closed, oval, rectangular, peaky, wavy, with or without tail and with a white or yellow background.
- Text is handwritten on 39 pages and computer-written on the other 61. 18 pages contain out of balloon onomatopoeias. 13 pages are written in English, 6 in Japanese, the rest of them are in French.

Due to copyright concerns, we do not have the rights of use on recent American comics yet. That is the reason why the presented corpus is mainly made of French and royalty-free material. We are looking forward to extend it both in size and completeness for the next version of the dataset.

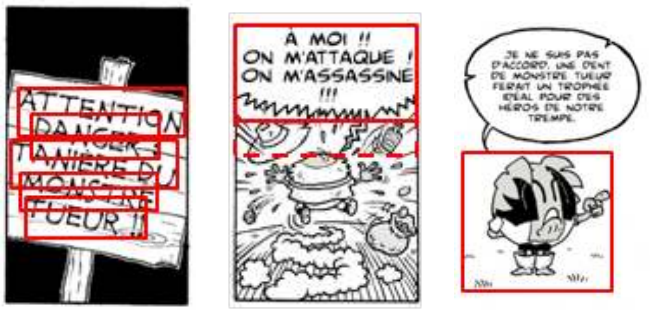
This first version is composed of 100 pages, from which 848 panels, 1091 balloons and 4667 text lines have been annotated.

III. Ground Truth Construction

Building an annotated ground truth always requires a lot of time and concentration from the experts. We kindly had the contribution of 20 people during one day to build this first ground truth. They are mostly all related to research in computer science and some are from the administration department. Specific instructions have been given before and throughout the session to all the participants regarding the use of the groundtruthing software (see section IV) and the protocol to follow for clipping and annotating the pages.

A. Visual segmentation protocol

In order to cover a wide range of possible research matters, it has been decided to extract three different type of objects from the corpus: text lines, balloons and panels. We decided to do this first ground truth by drawing horizontal bounding boxes as close as possible from the feature and including all its pixels in order to proceed a maximum of pages in the allotted time. This precision level is used in several, widely used datasets [11], [17]. Each element of the page follows specific guidelines according to the rules below:



a) b) c)

Fig. 1. Segmentation protocol. a) Regular bounding boxes are not always fitted to crooked text. b) Balloon's are segmented as close as possible from the text, without the tail. c) Panels are segmented as close as possible from their graphic content. Credits: [18]

1) Text lines: Text lines are defined as a group of written characters aligned towards the same direction. As emotions and expressions are often materialized with a single character (for instance "!" is used to express surprise and "?" a lack of understanding), it is also considered as being a text line. We labelled all the type of text (e.g. speech text, illustrative, graphic sound, narrative) at line level. Note, we kept using the horizontal bounding box level for homogeneous purpose but it is not appropriate for non horizontal text line, see Fig. 1a.

2) Balloons: Balloons are defined as an area surrounding a block of text lines, graphically represented by a boundary and/or a tail[3]. It is very uncommon that a balloon does not contain anything at all but, even if so and that it matches the above criteria, it is still considered as being a balloon. The bounding box passes through the tail. The reason is the shape and size of the tail, often very different from the balloon's, that may alter the bounding box (see Fig. 1b). In case of suggested balloon contour, the bounding box is defined around the contained feature (e.g. text, drawing), see Fig. 1b.

3) Panels: Panels are defined as an image area picturing a single scene. They can be framed and the bounding box will be defined as close as possible to the frame. They can be frame-less, in that case the bounding box will be set according to the contained drawings, see Fig. 1c. In both cases, text and overlapping features are ignored. There is necessarily at least one panel in a page.

These three kinds of element are independent from each others. A balloon does not necessarily have to be contained in a panel, as well as a text line in a balloon (nor in a panel by extension either).

[3]We call a tail the arrow-shaped area of a balloon, pointing to the speaking (or thinking) character.

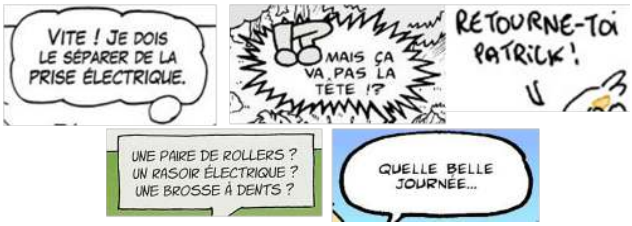


Fig. 2. Different speech balloon shapes. Top-down, from left to right: cloud, peak, suggested, rectangular and oval.

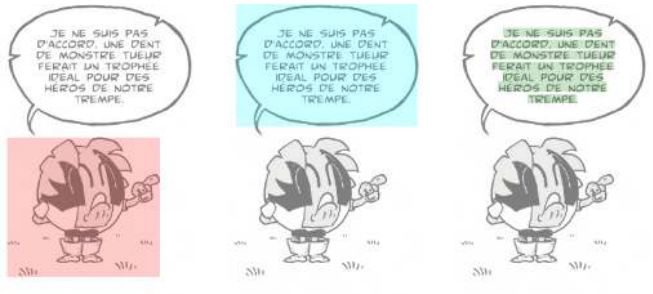


Fig. 3. Left-to-right: segmentation of a panel, a speech balloon and text lines. Credits: [18]

B. Semantic annotation

Once segmented, each object is annotated with a set of predefined metadata as follows.

1) Panels: Each panel is annotated with a `rank` metadata which stand for its position in the reading sequence. The first panel to be read on a given page has its rank property set to 1, while the last one's is set to n, with n the number of panels in the page.

2) Balloons: Balloons are also annotated with a `rank` property. However, as balloons are not explicitly bounded to any panel, their rank is set according to the page as a whole. For a page containing m balloons, the first balloon's rank will be 1 and the last will be m. Moreover, two additional metadata are given. First, the `shape` is indicated, picking a value from the enumeration {`cloud, oval, peak, rectangle, suggested`} as pictured in Fig. 2. Finally, the pointing direction of a balloon's tail is given through the `queueDirection` property. The possible values are the eight cardinal directions plus a ninth additional value for the lack of a tail: {`N, NE, E, SE, S, SW, W, NW, none`}.

3) Text lines: Each text line is associated with its transcription in capital letters.

4) Pages: Each page has been annotated with bibliographical information, so that anyone using this database can find the appropriate comic books. Among the first ones comes the page number (`pageNumber`) the comic book title, from which the page has been picked up, and its release date (`albumTitle, releaseDate`), the series it belongs to (`collectionTitle`), the authors and editor names (`writerName, drawerName, editorName`) and, finally, the website and/or ISBN (`website, ISBN`). The album title is not mandatory for webcomics. Structural information about the page content has been added as well, such as resolution (`resolution`), reading direction (`readingDirection`), main language of the written text (`language`) and single or double page image (`doublePage`).

The combination of visual segmentation and semantic annotation provides the advantage to make this ground truth database relevant for image analysis evaluation and semantic retrieval.

C. Structure

As we wanted to keep the database file system simple and easy to share, semantic and visual annotations on a given page are gathered in a single SVG (Scalable Vector Graphics) file. Besides being royalty-free and actively maintained by a W3C working group, the SVG format fulfils two essential needs for this database.

First, in association with a recent Internet browser or your favorite image viewer, it provides a simple, fast and elegant way to display the visual segmentation of any desired object over a comic book page. Different kind of objects (e.g. panels, balloons) can be displayed in different ways according to CSS (Cascading Style Sheets) properties defined for each of them, see Fig. 3. Each layer can be displayed or not in order to enhance the clearness of the annotations when browsing the database. Secondly, SVG being a XLM-based language, it makes the integration of semantic annotation very easy via the use of the predefined `metadata` element.

One ground truth file contains the complete description of one comics image. There is no hierarchical link between pages from a same comic book. Following the basic XML and encoding information, a SVG file starts with a root `<svg>` element containing the title of the document, `<title>`, and four `<svg>` children with different class attributes. The annotations are the same for all the `<svg>` nodes, each of them describing one kind of element (e.g. panel, balloon) according to its `class` attribute. The first `<svg>` element, `<svg class="Page">`, has two children. The first one is `<image>` and contains a link to the corresponding image file and the size it has to be displayed. The next child is a `<metadata>` element containing the bibliographical information described in III-B4. The three following `<svg>` siblings, `<svg class="Panel">`, `<svg class="Balloon">` and `<svg class="Line">` respectively contain the annotations on panels, balloons and text lines. They all contain SVG `<polygon>` elements with a list of five points in a `point` attribute that define the position of the bounding box's corners. Note that the fifth point equal the first one to "close" the polygon according to the SVG format. Those points are used by the viewer to draw polygons over the page. Each `<polygon>` has a `<metadata>` child to store information on the corresponding polygon, according to the attributes list described in III-B.

D. Error evaluation

When so many different persons are involved in the creation of a graphical ground truth, it is very difficult, if not impossible, to have a perfectly homogeneous segmentation. Therefore, in addition to the package of pages he was in charge

of, each participant has been asked to annotate the panels of a final page. This page was the same for everybody and was chosen for its graphical components heterogeneity. It contains ten panels from which, four are full-fenced, five half-fenced and one is completely frameless, see Fig. 4. This heterogeneity is somehow representative of the whole corpus.

We defined an acceptable error for the position of a corner given by several persons. Image being of different definitions, using a percentage of the page size makes more sense than using a specific number of pixels. We set this percentage p at 0.5% of the page height and width in x and y. Given the test image's definition of 750x1060 pixels, this makes a delta of +/- 5 pixels in y axis and +/- 4 pixels in x axis.

We asked to each one of the twenty involved persons to draw the four points bounding box of the panels ignoring text area. A mean position from the twenty different values has been calculated for each of them. Then, the distance of each point to its mean value is computed. Fig. 5 shows the amount of corners for a distance, centered on zero. Given the threshold $p = 0.5$, 87.5% of pointed corners can be considered as being homogeneous over the group of labeling people. The overall mean standard deviation on this page reaches 1.13 pixels for the width, and 1.28 pixels for the height. The two bumps, at -40 and 15, are related to the missegmentation of 13 of the 80 panels. Indeed, instructions have been misunderstood by some people who included text area outside of the panels or missed some panel's part. Fig. 4 shows the difference between areas labeled as a panel by at least one person and areas labeled as a panel by every participant. However, mistakes of this kind have been manually corrected by a post-production pass.

Even though the error criterion has only been estimated on panels, it is reasonable to extend it to balloons and text lines as well. Indeed, the segmentation protocol being quite similar for all features (bounding box as close as possible to the object), the observed standard deviation of panel corner positions has no reason to be different from balloons and text lines.

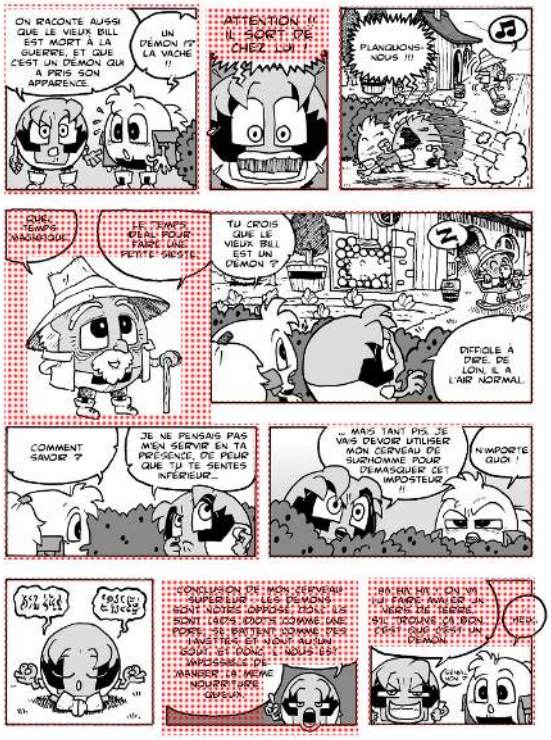


Fig. 4. Error measurement page. Hatched areas are the difference between areas labeled as panels by at least one person, and areas labeled by everybody. Credits: [18]

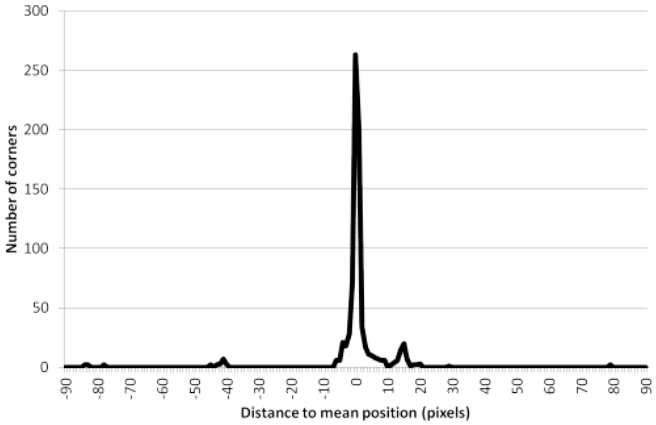


Fig. 5. Number of corners for a given standard deviation value. This has been calculated on y axis, x axis produces similar plot.

IV. Associated Tools

A. ToonShop

We developed a ground truth software called ToonShop, that assisted users during the construction of this dataset. It has been developed using Java programming language version 6 and uses public libraries, like Java Advanced Imaging (JAI) and Apache Batik, to handle images and SVG files. This tool allows users to create, modify or delete polygons on separate layers. Each layer refers to a kind of feature, namely panels, balloons, lines of text and with a possible extension to characters, objects and so on. Polygons are automatically colorized with a unique color, accordingly to their attached layer (e.g. panel, balloon, text). Semantic annotations on pages and polygons have been added through ToonShop as well.

B. Validation

In addition to the database, we provide a tool to evaluate the recall and precision of our panels and balloons extraction's algorithm regarding the ground truth. The validation process is based on the work of Wolf [19] who addressed the issue of region segmentation evaluation and its specificities.

Let $S_a = \{a_1, ..., a_n\}$ and $S_b = \{b_1, ..., b_m\}$ be respectively the sets of n extracted and m ground truthed same kind of objects (panels or balloons). The number of correctly segmented objects in S_a is obtained by comparing each of its elements to the matching element, or elements, in the ground truth. Region splitting and region merging are not unfamiliar to image segmentation. Indeed, it happens that a visual feature is well recognized except for a mismatch in the amount of segmented features. So, rather than putting those results to a null score, a penalty function f is introduced to

lower the output value of the evaluation. An element a_i is considered validated if the recall and precision of the shared pixels with its matching element(s) are respectively higher than two thresholds t_r and t_p. Then, this element is given a score s depending on which of the following cases it belongs to:

- One-To-One validation, a_i is matching one element b_j of S_b. $s = 1$.
- One-To-Many validation, a_i is matching a subset S'_b of S_b, $|S'_b| > 1$. $s = 1 - f(|S'_b|)$.
- Many-To-One validation, a_i is part of a subset S'_a of S_a, $|S'_a| > 1$, which is matching one element b_j of S_b. $s = 1 - f(|S'_a|)$

The penalty function f does not necessarily have to be the same for merges and splits, so they can be sanctioned differently. We chose $f(x) = ln(x)$.

Thresholds t_r and t_p are not set to any particular values. They can vary from 0 (which will validate any intersection) to 1 (which will only validate perfect matchings) by 0.1 steps. This produces a 10-by-10 matrix that can be summed to obtain a final performance value. Results for a given couple of values can be exported to a set of comparison SVG files.

V. Conclusion and Future Work

We presented the eBDtheque database, a ground truth on comic books containing spatial and semantic annotations. The corpus has been introduced as well as the construction protocol. The database is available to the community on http://ebdtheque.univ-lr.fr, with the evaluation tools.

New semantic annotations, such as the view angle or the shot type of a panel, will soon be added by domain experts. Type of balloons (e.g. speech, thought, narration, illustrative) and the affiliation between objects are two points that will be expressed as well.

Acknowledgment

This work was supported by the European Regional Development Fund, the region Poitou-Charentes (France), the General Council of Charente Maritime (France) and the town of La Rochelle (France). The authors would like to thank the L3i members and every people that help to the construction of this database.

Sincere thanks to all the comics writers who kindly agreed to the use of their pieces of art in this database. In alphabetical order: Pascal Boisgibault, Cyb, Fred, Sergio Garcia, Olivier Jolivet, Lamisseb, Grald Lubbin, Winsor McCay, Midam, Marion Montaigne, Nicolas Roudier, Alain Saint Ogan, Trébla and Lewis Trondheim. Thanks also to their editors: Actes Sud, Ankama, Bac@BD, Clair de Lune, Dargaud, Delcourt, Doc En Stock, Dupuis, Hachette and Studio Cyborga. Finally, a special thank to the CIBDI[4], Free Public Domain Golden Age Comics and the Department of Computer Science and Intelligent Systems of Osaka Prefecture University who kindly provided material from their personal collection.

[4]Cité Internationale de la Bande Dessinée et de l'Image

References

[1] E. Han, K. Kim, H. Yang, and K. Jung, "Frame segmentation used mlp-based x-y recursive for mobile cartoon content," in *Proceedings of the 12th international conference on Human-computer interaction: intelligent multimodal interaction environments*, ser. HCI'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 872–881.

[2] Y. In, T. Oie, M. Higuchi, S. Kawasaki, A. Koike, and H. Murakami, "Fast frame decomposition and sorting by contour tracing for mobile phone comic images," *Internatinal journal of systems applications, engineering and development*, vol. 5, no. 2, pp. 216–223, 2011.

[3] K. Arai and H. Tolle, "Method for automatic e-comic scene frame extraction for reading comic on mobile devices," in *Seventh International Conference on Information Technology: New Generations*, ser. ITNG. Washington, DC, USA: IEEE Computer Society, 2010, pp. 370–375.

[4] 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.

[5] M. Yamada, R. Budiarto, M. Endo, and S. Miyazaki, "Comic image decomposition for reading comics on cellular phones." *IEICE Transactions*, vol. 87-D, no. 6, pp. 1370–1376, 2004.

[6] C. Rigaud, N. Tsopze, J.-C. Burie, and J.-M. Ogier, "Robust frame and text extraction from comic books," in *Graphics Recognition. New Trends and Challenges*, ser. Lecture Notes in Computer Science, Y.-B. Kwon and J.-M. Ogier, Eds. Springer Berlin Heidelberg, 2013, vol. 7423, pp. 129–138.

[7] C. Guérin, "Ontologies and spatial relations applied to comic books reading," in *PhD Symposium of Knowledge Engineering and Knowledge Management (EKAW)*, Galway, Ireland, 2012.

[8] C. Ponsard and V. Fries, "Enhancing the Accessibility for All of Digital Comic Books," vol. I, no. 5, 2009.

[9] L. Li, Y. Wang, Z. Tang, and D. Liu, "Comic image understanding based on polygon detection," in *IS&T/SPIE Electronic Imaging*, vol. 8658. International Society for Optics and Photonics, 2013.

[10] E. Han, J. Yang, H. Yang, and K. Jung, "Automatic mobile content conversion using semantic image analysis," in *Proc. of the HCI'07*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 298–307.

[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[12] A. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and trecvid," in *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*. ACM, 2006, pp. 321–330.

[13] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Institute of Technology, Tech. Rep. 7694, 2007.

[14] B. Lamiroy, D. Lopresti, H. Korth, and J. Heflin, "How Carefully Designed Open Resource Sharing Can Help and Expand Document Analysis Research," in *Document Recognition and Retrieval XVIII - DRR 2011*, C. V.-G. Gady Agam, Ed., vol. 7874, SPIE. San Francisco, United States: SPIE, Jan. 2011.

[15] S. McCloud, *Understanding comics*. William Morrow Paperbacks, 1994.

[16] M. Stommel, L. Merhej, and M. Mller, "Segmentation-free detection of comic panels," in *Computer Vision and Graphics*, ser. Lecture Notes in Computer Science, L. Bolc, R. Tadeusiewicz, L. Chmielewski, and K. Wojciechowski, Eds. Springer Berlin Heidelberg, 2012, vol. 7594, pp. 633–640.

[17] B. Yao, X. Yang, and S. Zhu, "Introduction to a large-scale general purpose ground truth database: methodology, annotation tool and benchmarks," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer, 2007, pp. 169–183.

[18] Cyb, *Bubblegôm Gôm vol. 1, pp. 3*. Goven, France: Studio Cyborga, 2009, vol. 1.

[19] C. Wolf and J. Jolion, "Object count/area graphs for the evaluation of object detection and segmentation algorithms," *International Journal on Document Analysis and Recognition*, vol. 8, no. 4, pp. 280–296, 2006.