Comic Computing: a conceptual framework for decomposition and utilization of comic contents

Ryosuke Yamanishi[†] **Mitsunori Matsushita**[‡]
[†]College of Information Sciences and Engineering, Ritsumeikan University, Japan
ryama@media.ritsumei.ac.jp
[‡]Faculty of Informatics, Kansai University, Japan
mat@res.kutc.kansai-u.ac.jp

Abstract

This paper proposes a conceptual framework of *comic computing*, that intends to manipulate comic contents as a resource for further usages such as comic retrieval and generation. Comic is a medium that conveys a story by using images combined with texts or other visual expressions. With the spread of small devices, a market for digital comics has been growing. Accordingly, the demand for accessing and processing such comics at will has also been increasing. Since comics are cross-modal and mixed-modal contents that utilize both graphical information (e.g., illustrations and comic symbols) and textual information (e.g., dialogues and onomatopoeias) in a coordinated manner. It requires various technologies such as image processing, knowledge processing, and media processing, to handle these contents. To establish the method for conducting these processes beyond the discipline gap among research fields, this paper surveys the existing researches for comic computing and marshals the problems to be tackled in the field of comic computing. The current efforts of research communities that relate to the comic computing and the future ambition of the comic computing are also described.

Keywords: Comic Computing, Image Processing, Langage Processing, Database, Interface

1 Introduction

With the widespread of portable digital tools such as tablets and smart phones, it has become popular for people to read comics on such digital tools. Along with the circumstances, the market for digital comics has grown rapidly in recent years. In Japan, for instance, the sales amount of digital comics in 2015 rose to about 114 billion Yen (approx. one billion US$), up 38.1% from the previous year. With such an increase in the availability of digital comics, the desired comic book contents can be found more efficiently. Digital comics are regarded as having higher applicability than paper-based comics because these digitalized contents are easy to manipulate dynamically with computers and smart devices.

Unlike traditional paper-based comics, digital comics have no physical constraints. This characteristics give comics to have two types of the advantage. The dynamic expression, which is different from the conventional paper-based comics, is realized: e.g., dynamic changing of the contents according to the progress of story, and giving motion to the elements depicted in a comic. Also, that permits users to utilize the comics optimized for their own environment; changing language displayed based on the reader's native tongue, or modifying font shape and font by taking the reader's environment into account.

In the current situation, however, many digital comics remain "static" contents: it simply captures the contents of paper-based comics with a scanner and digitizes it as it is. Therefore, they cannot fully exploit the possibility of digital comics. On the other hand, Matsui et al. have compiled a database consisting of 109 comics for use in academic research [1], and Fujimoto et al. have improved the database with metadata [2] —not only the database but also the sketch-based comics retrieval method is proposed in their other work [3], though. The preparation for computer science research for comics have just been completed. Several types of research for comics have been widely focused on and some research communities for comics have been established in the world. Based on the current situation, this paper proposes a concept of "comic computing" that demonstrates the possibility of innovative digital comic usages.

Figure 1. Example of a frame in comics.

2 Notation of comics

The first step to process comics should be the notation of the comics. Language and image information are collaboratively used to represent all of the stories in a comic; movement, sound, speech, and atmosphere are represented as the mix of the language and image information. The comic readers unconsciously recognize each element in comics. From the combination of the elements, even the movement and the sounds which are sequential information are understood. In order to notate each element in comics, we have to consider the definition of the element of comics at first. And then, how to use the element, that is affording comics to be digitally processed, should be considered.

2.1 Definition of the "element" of comics

In a comic, its content is expressed by utilizing visual expressions such as frame layout and unique symbolic expression. Usually, comic consists of a series of frames. In each frame, various comic elements are depicted that includes personages, items, onomatopoeias, icons, speech balloons, and effect lines (See Figure. 1).

Currently, comics are provided to users as an image format (e.g., JPEG) page by page. In order to make a comic calculable, it is necessary to identify elements composing the comic and extract the elements from the image. Elements to be extracted are represented as binary images composed mainly of lines and dots. Therefore, to encode them, it is necessary to identify the elements by applying image processing prior to the whole process. To meet the requirements, various researches that include frame detection, screentone removal, personage identification, and speech balloon identification are conducted in the field of image processing mainly.

2.2 How to afford comics to be digitally processed

Comic is a cross-modal content in which pictures and texts are used complementarily and cooperatively. Therefore, in order to make these available on computers, the following processes are required;

(1) Making it extractable for each element such as text, personage, serif, item, onomatopoeia from a comic image,

(2) Structuring the relation between the extracted elements and segmenting the consecutive frames,

(3) Identifying the content and meaning conveyed by the set of structured elements, and

(4) Linking the global structure of the story with the local structure from the identified meaning by taking into account the way to use and its context to be used.

These correspond to (1) morphological analysis, (2) syntax analysis, (3) semantic processing, and (4) context processing, respectively, which are processes of applications utilizing natural language processing. Unlike a medium mainly composed of texts such as newspaper articles, texts are placed in a picture, and their positions and font shapes are meaningful. Thus, simply extracting text information from the comic is not sufficient. Information including what kind of expression is applied (e.g, font shape, size information) and where it appears (e.g., position information) must be encoded. Furthermore, since comic uses pictures and texts complementarily and cooperatively, not only text information but also information of personages and items drawn in the picture must be encoded. In order to apply these processes to the existing comics, the deep fusion of pattern processing and symbol processing is required. Especially, such the deep fusion is indispensable executing the steps (1) and (2).

3 Comic contents decomposition

In order to enable processing shown in the preceding sections, the elements drawn in comics (e.g., personages, utterances, effect sounds) should be extracted from the image by using pattern recognition at first. After that, the relationship between the elements should be structured. The following sections show some more researches to decompose the comic elements.

3.1 Frame and speech balloons recognition

In general, a story of comic proceeds by the sequence of frames and the frames are treated as a minimum unit of a comic in terms of semantics. Several methods have been proposed to identify each frame region such as a method to identify dividing lines of a frame by detecting the direction of the density gradient from the target comic image [4; 5], and a method to detect rectangular regions from the target comic image [6]. It is reported that both methods achieved more than 80% accuracy.

Several methods have been studied to identify speech balloons in comic images. Tanaka et al. [7] identified the text area in the page using Ada Boost, and detected balloon candidates based on that area. In addition, balloon shape classification (communication type, curve type, polygonal line type, square type) is performed by SVM (Support Vector Machine). By using this method, 86% of speech balloons have been identified. Rigaud et al. [8; 9] first identifies the location of the text and identifies the speech balloon area based on it. Then focusing on the displacement of the frame line of the balloon, the distance between the balloon region and the frame line is classified by comparing with the typical variation pattern (eg, zigzag, wavy, smooth).

3.2 Screentone removal

Screentone is a sticker affixed to comic. It is used for creating a comic in order to give a background, shades, textures, skin and hair colors, and psychological effect to the comic.

When producing a comic with digital media, effect functions of image drawing tools are often used. To improve recognition accuracy of personages and items, a method to remove the screentones from the comic has been proposed. For example, Ito et al. proposed a method of extracting a line drawing by removing a screentone area [10]. The proposed method separates a screentone area and a line drawing area by using LoG (Laplacian of Gaussian) filter and FDoG (Flow-based Difference-of-Gaussian) filter from black and white comic images. The accuracy of the method is about 55% in average. Though there is still room for improvement, this method will contribute to the improvement of the accuracy of the subsequent processes such as a personage identification.

3.3 Personage recognition

In order to grasp the personages that appear in a comic, a method to identify the personage from face images appeared in the comic has been proposed [11]. The proposed method identifies the face candidates in images based on CNN features and performs matching between the face candidates and the face image stored in the database prepared. Identifying not only the position of the face but also the position of parts (e.g., pupils) has been studied.

Also, identifying a personage by using the determined parts' position has also been conducted. In recent years, a method for identifying face candidates with higher accuracy has been proposed [12]. In this method, deformable part model robust to image deformation is applied to comic images. However, since the variation of accuracy by the personage is large at present, it can not be said that stable face detection is achieved.

The following reasons are assumed for this situation: (1) The face of a personage appearing in a comic is represented by a line drawing in general, and the amount of feature that can be used for identification is limited in comparison with the face recognition of a real image, (2) Because of exaggeration peculiar to comics, variations of facial contours and parts are large. In order to solve this problem, Tani et al. focused on the following method: (1) Differences in hair colors are often used to discriminate personages in comics, (2) The personage appearing on consecutive frames tend to be the same person. They examined to improve personage identification accuracy using these comic peculiar heuristics [13].

3.4 Structuration of comic elements

In order to make use of extracted comic elements, It is necessary to manage the relationship between elements by structuring the segments of the scene and the order in which frames are read. In addition, in the comic, a new expression technique can be created by creator's efforts. This means that the scalability of the management method should be ensured. In order to solve these problems, several kinds of research on modeling metadata to be extracted from comics by using Wikipedia entries and an FRBR (Functional Requirements for Bibliographic Records) [14; 15] and designing a de-

scription framework that takes these metadata into consideration [16; 17] have been conducted.

In these studies, vocabularies that form the basis of metadata are divided into four categories: (1) intellectual content, (2) bibliographic description, (3) structure description, and (4) graphic element. With this modeling, generic knowledge construction that is not limited to a specific use is attempted. On the contrary, Rigaud et al. propose a method for structuring comic in a bottom-up manner [18]. The proposed method acquires structure from image processing results such as frame identification and speech balloon recognition, by referring prior knowledge about domains and comic contents.

Considering these previous studies described above, it can be concluded that basic technologies for extracting comic components. Also, understanding such structures from a comic image are steadily developing toward practical level.

4 Comic contents utilization

Comic contents are not only for the comic itself but also some other domains. Some studies have tried to connect the entertainment with education and communication introducing the advantage of comics. Such applications are some of the typical utilizations that would be realized by the decomposed comic contents.

Shan et al. focused on the fact that many foreigners studied Japanese language through Anime distributed in the world. They analyzed the expressions of dialogues spoken in Anime considering the difficulty in Japanese Language Proficiency Test (JLPT) [19]. They have found that the relationships between the genres of Anime and the difficulty of the dialogues. Through the analysis, it was suggested that the dialogues in especially Anime of daily lives were useful for Japanese learner in the early stage. On the other hand, the dialogues in Anime of fantasy included so many coined words, rude words, and difficult expressions that were rarely used in the real world; those expressions are hard to learn for the beginners. Their future direction is to use the dialogues in Anime as the lecture contents for the Japanese language practice.

The specific and signature drawing style in comics can be also utilized in another domain. Yamanishi et al. proposed a method to recommend speech balloon shapes based on the linguistic features of input text [20]. The speech balloon is typically used in comics to represent the acoustic features of the speech (i.e., speech nuance). Through the experiments, it was confirmed that the difference of the speech balloon shapes told the readers different types of emotion even though the sentence was the same. Then, they modeled the relationships between the shapes of the speech balloons and linguistic features of the serif in its speech balloon. Using the model, the speech balloon shapes are recommended considering how the input text is written. The recommendation system can be used in a chat tool to represent some emotional nuance of the message.

5 Current efforts on comic computing community

The researches targeting at the processing comics were uncooperatively presented in different fields, there were a large number of varied researches handling comics though. Since 2013, we have called for research related to comics in various research field such as computer vision, natural language processing, database, and interface. We started to gather the researchers in the name of “comic computing” in Japan.

We organized the special session named “comic computing” at Human Communication Group Symposium in Japan, on 2013, 2014 and 2015 (HCG 2013, HCG 2014, and HCG 2015). After those sessions were successfully conducted, we started to organize the session named “comic computing and AI” at the Annual Conference of Japanese Society for Artificial Intelligence on 2016 and 2017 (JSAI 2016 and JSAI 2017). Through our organized sessions, 51 papers from unique 18 laboratories in all were presented. The statistic information of those papers are shown in Table 1. As shown in the table, the research fields, which are classified by one of the authors considering the keyword of each paper, were so varied but all of those are for the comics processing and the application of comics.

Also for internationally, the core members of the comic computing community have organized some special sessions and workshops at international conferences. Focusing on the affinity between the cultures because of the geographical proximity, we held the special session “comic Computing” at The Fifth Asian Conference on

Table 1. Statistics of the organized sessions. The papers are classified into each research field by one of the authors considering the keyword of each paper. In the column of the "Laboratories," the unique number of laboratories are shown.

Organized session	Submissions		Research field			
	Papers	Laboratories	Image Processing	Language Processing	Database	HC Interaction
HCG 2013	9	7	1	1	4	3
HCG 2014	12	7	4	2	4	2
HCG 2015	11	7	3	3	4	1
JSAI 2016	7	7	2	3	1	1
JSAI 2017	12	10	4	2	2	4
Total	51	18	14	11	15	11

Information Systems on 2016 [1]. And also in 2016, the international workshop MANPU (coMics ANalysis, Processing and Understanding) [2] was held in conjunction with the 23rd International Conference on Pattern Recognition. At one of the top conferences for image processing, the concepts and future directions of comic computing were shared with the researchers.

Comic is a multimedia content concerning image and language information itself and is popular as media-mix and cross-media content. Based on the characteristics of comics, the community for comic computing should be diverse. We have organized not only research but also communities including publishers and voice actors of Anime. In such organized sessions, we had the invited talks of non-academic comic related specialists. The community has been growing to become more and more international and diverse beyond the media and culture.

6 Conclusion

In this paper, we surveyed the existing researches for comic computing, which handles comic contents in an informatics manner. Traditional paper-based comics consist of only language and image information. Though the comics do not play any kinds of sequential information such as movement and sound, the comic reader can easily grasp the situation in a frame like a movie; in most cases, we know how to read comics without any education. Such characteristics would provide the attraction of comics. Some of the researches have tried to reveal and re-use the characteristics. Recently, the comics have been shifted from papers to digital devices for its distribution field. The digital database of comics for research has also been compiled; it is not just a Japanese culture depended but a kind of the task driven filed of research. We believe that these dynamics have a lot of prospect for comics and the research for comic computing should adapt the dynamics. As manual and digital techniques are collaboratively used in any situations related to comics, the attraction and accessibility of comics would be more and more increased. To tackle this massive task, different types of research have to collaborate with each other beyond the field of the research.

Over several years, we have organized the community for comic computing including the author, publisher and voice actors of Anime in Japan. In a few years later, we would like to expand this community in the whole of the world. May the field of comic computing be with all of the comic lovers.

[1] http://acis2016.reasoning.org/ws-ss/ss2-comic-computing/

[2] http://manpu2016.imlab.jp/

Acknowledgments

This work is supported in part by JSPS Grant-in-Aid for Challenging Exploratory Research #15H02780.

References

[1] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based manga retrieval using manga109 dataset," *Multimedia Tools and Applications*, vol. 64, no. 12, pp. 1–28, 2016.

[2] A. Fujimoto, T. Ogawa, K. Yamamoto, Y. Matsui, T. Yamasaki, and K. Aizawa, "Manga109 dataset and creation of metadata," in *Proc. of the 1st International Workshop on coMics ANalysis, Processing and Understanding*, pp. 2:1–2:5, 2016.

[3] Y. Matsui, K. Aizawa, and Y. Jing, "Sketch2manga: Sketch-based manga retrieval," in *Proc. of Image Processing, 2014 IEEE International Conference on*, 2014.

[4] T. Tanaka, K. Shoji, F. Toyama, and J. Miyamichi, "Layout analysis of tree-structured scene frames in comic images," in *Proc. of the 20th International Joint Conference on Artificial Intelligence*, pp. 2885–2890, 2007.

[5] D. Ishii and H. Watanabe, "A study on frame position detection of digitized comics images," in *Workshop on Picture Coding and Image Processing*, vol. 2010, pp. 124–125, 2010.

[6] S. Nonaka, T. Sawano, and N. Haneda, "Development of gt-scan, the technology for automatic detection of frames in scanned comic," *Fujifilm Research Development*, no. 57, pp. 46–49, 2009.

[7] T. Tanaka, F. Toyama, J. Miyamichi, and K. Shoji, "Detection and classification of speech balloons in comic images," *The Journal of the Institute of Television Engineers of Japan*, vol. 64, no. 12, pp. 1933–1939, 2010. (in Japanese).

[8] C. Rigaud, J.-C. Burie, J.-M. Ogier, and D. Karatzas, "Speech balloon contour classification in comics," in *Proc. of the 10th IAPR International Workshop on Graphics Recognition*, 2013.

[9] C. Rigaud, J.-C. Burie, J.-M. Ogier, D. Karatzas, and J. V. D. Weijer, "An active contour model for speech balloon detection in comics," in *Proc. of the 10th IAPR International Workshop on Graphics Recognition*, 2013.

[10] K. Ito, Y. Matsui, T. Yamasaki, and K. Aizawa, "Separation of manga line drawings and screentones.," in *Eurographics (Short Papers)*, pp. 73–76, 2015.

[11] W.-T. Chu and W.-W. Li, "Manga facenet: Face detection in manga based on deep neural network," in *Proc. of the 2017 ACM on International Conference on Multimedia Retrieval*, pp. 412–415, 2017.

[12] H. Yanagisawa, D. Ishii, and H. Watanabe, "Face detection for comic images with deformable part model," in *Proc. of the IIEEJ Image Electronics and Visual Computing Workshop*, 2014.

[13] Y. Tani, N. Shirozu, and M. Matsushita, "A study for personage extraction from comic contents," in *Proc. of the 75th Annual Conference on Information Processing Society of Japan*, pp. 889–890, 2012. (In Japanese).

[14] S. Nomura, A. Morozumi, M. Nagamori, and S. Sugimoto, "Analysis of an architecture of manga toward an integrated metadata model for manga," in *Digital Libraries*, vol. 36, pp. 3–14, 2009. (in Japanese).

[15] W. He, T. Mihara, M. Nagamori, and S. Sugimoto, "Identification of works of manga using lod resources: An experimental frbrization of bibliographic data of comic books," in *Proc. of the 13th ACM/IEEE-CS joint conference on Digital libraries*, pp. 253–256, ACM, 2013.

[16] A. Morozumi, S. Nomura, M. Nagamori, and S. Sugimoto, "Metadata framework for manga: A multi-paradigm metadata description framework for digital comics," in *International Conference on Dublin Core and Metadata Applications*, pp. 61–70, 2009.

[17] T. Mihara, M. Nagamori, and S. Sugimoto, "A metadata-centric approach to a production and browsing platform of manga," *The Outreach of Digital Libraries: A Globalized Resource Network*, pp. 87–96, 2012.

[18] C. Rigaud, C. Guérin, D. Karatzas, J.-C. Burie, and J.-M. Ogier, "Knowledge-driven understanding of images in comic books," *International Journal on Document Analysis and Recognition*, no. 3, pp. 199–221, 2015.

[19] J. Shan, Y. Nishihara, R. Yamanishi, and J. Fukumoto, "Analysis of japanese animation dialogue level using japanese expressions," in *The Proc. of The Fifth Asian Conference on Information Systems*, pp. 322–325, 2016.

[20] R. Yamanishi, H. Tanaka, Y. Nishihara, and J. Fukumoto, "Speech-balloon shapes estimation for emotional text communication," *Information Engineering Express*, vol. 10, no. 2, pp. 1–10, 2017.