Automatic Preview Generation of Comic Episodes for Digitized Comic Search *

Keiichiro Hoashi, Chihiro Ono
KDDI R&D Laboratories Inc.
2-1-15 Ohara Kamifukuoka
Saitama 356-8502 Japan
{hoashi,ono}@kddilabs.jp

Daisuke Ishii, Hiroshi Watanabe
GITI, Waseda University
1011 Nishitomida Honjo
Saitama 367-0035 Japan
dai.ishii@fuji.waseda.jp
hiroshi.watanabe@waseda.jp

ABSTRACT

This research proposes a novel method to present "thumbnails" of episodes of digitized comics, in order to improve the efficiency of comic search. Comic episode thumbnails are generated based on image analysis technologies developed especially for comic images. Namely, the following procedures are developed for our system: automatic comic frame segmentation, text balloon extraction, and a linear regression based model to calculate the importance score of each extracted frame. The system then selects frames from each episode with high importance score, and aligns the selected frames to create the episode thumbnail, which is presented to the system user as a compact preview of the episode. User experiments conducted with actual Japanese comic images prove that the proposed method significantly decreases the time necessary to search for specific episodes from a large scaled comic data collection.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval—*retrieval models, search process*; H.2.4 [**Database Management**]: Systems—*multimedia databases*

General Terms

Algorithms, Experimentation.

1. INTRODUCTION

Recently, digital distribution of comics (especially for mobile phones) has developed into a huge business in Japan. Furthermore, the spread of content over the Web has increased the worldwide popularity of Japanese comics, which indicates the potential growth of the global comic distribution market. As the amount of digitized comics increases, technology to support comic search is expected to become more and more important. However, the search function of existing online comic distribution sites is basically restricted to metadata search, *e.g.*, keyword search of comic title/artist.

The objective of this research is to improve the usability of digitized comic search systems. In order to achieve this goal, we propose a novel method to generate "thumbnails" of comic episodes, by selecting and aligning significant comic frames from a given comic episode, *i.e.*, a series of comic page images. Component technologies based on image analysis are developed to extract features from comic images, such as automatic segmentation of frames and text balloon extraction. Features extracted from the frames and text balloons are input to a linear regression based model, which calculates the importance score of each frame. Finally, frames with high importance score are presented to the user as the preview of the comic episode. User experiments are conducted to evaluate this method on a virtual comic distribution site. Results of this experiment indicate that our proposal significantly improves the efficiency of comic episode search.

2. BACKGROUND

By far, the most popular format of Japanese comics is the "story comic," which is typically serialized on weekly/monthly issued comic magazines. The basic unit of a story comic, *i.e.*, an *episode*, typically consists of 20-30 pages, depending on the publication which the comic is serialized on. In most digital comic distribution services, comic titles are sold in units of episodes. However, especially in comic distribution services for mobile devices, users can only search for comics based on their metadata, *e.g.*, comic title, author name, genre, etc. This means that, while users can easily search for specific titles of comics, it is extremely difficult for them to search for specific episodes from the title, which is a major user need. In fact, most existing services merely present a list of numbered episodes for each title, making it virtually impossible for users to find episodes without (mistakenly) downloading undesired ones.

Presenting a preview of comic episodes is assumed to be a helpful method for users who want to find specific episodes from a large scaled comic data collection. A typical way to generate such previews is to provide access to the first few pages of a comic episode, similar to the approach implemented for digital book distribution sites, such as Amazon.com. However, the pages presented in the preview may not be sufficient for users to grasp the story of the episode.

*Area chair: Bernard Merialdo

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

This is especially true for comics, since the amount of (textual) information on a page is sparse compared to usual books. Furthermore, browsing through numerous comic pages is expected to be a time-consuming process, especially when the users' demand is to find an episode from a large data collection, and/or when users are using digital comic distribution services on mobile devices with limited screen size.

3. RELATED WORK

Existing research efforts in the area of content-based analysis of comic images have focused on automatic frame segmentation. The mainstream approach is to iteratively segment frames from the input comic image based on density gradient, as proposed in [1, 2]. These methods enable the extraction of individual frames from a given comic page image. Furthermore, methods which utilize comic frame segmentation results to transform comics for viewing on mobile devices have also been developed [3, 4, 5].

Extraction of text information from comic page images is also an essential technology. Most existing text extraction research [6, 7] utilize edges as features to localize regions which include text. This approach is effective for general photographs, but is difficult to be applied for comics, since all elements (including the non-textual areas) of a typical comic image consist of strong edges. Thus, text extraction methods which are specialized for comics images becomes necessary. Guo *et al.* have developed a method to extract text balloons from comic strips, by extracting areas which are not connected with the frame boundary lines [8]. The efficacy of this method has only been evaluated on comic strips in relatively fixed formats, thus may not be able to handle the wide variety of expressions in story comics.

4. COMIC EPISODE SUMMARIZATION

4.1 General framework

The main proposal of this research is the automatic generation of comic episode "thumbnails," and its presentation to comic search system users for previewing comic episodes. The general framework of our proposal is illustrated in Figure 1[1]. The proposed system consists of four major steps. First, for each page (image) of the input comic episode, frame segmentation (*ref.* (1) of Figure 1) is conducted. A set of features is extracted from each segmented frame (*ref.* (2)), which is then used to calculate the frame importance score (*ref.* (3)). Frames with high importance scores are selected (*ref.* (4)) and presented to the user as the thumbnail of the input episode.

A conceptual image which illustrates the implementation of comic episode summarization results for an online comic distribution service, is shown in Figure 2. In existing comic distribution sites, users access to their favorite comic titles by text-based search (as illustrated in the "Top page" and "Title search result" screens of Figure 2), and select the episode which they want to purchase from the list of episodes in the search result. By utilizing our proposal, a summary of each episode can be generated beforehand, so that users can confirm the content of an episode by clicking on the preview link, as illustrated in the "Preview" screen in Figure 2. This user interface is expected to reduce erroneous downloading/purchasing of unwanted comic episodes, and improve the overall usability of comic distribution services. Details of each step are provided in the following sections.

[1]Comic images in this paper are provided from the website "Manga no Shimbun" (http://manganews.com/)

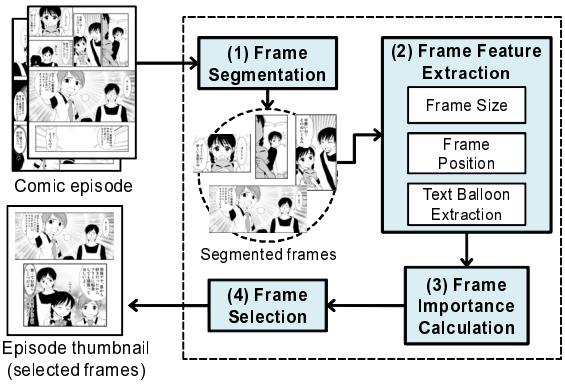

Figure 1: Outline of proposed system, which selects frames from comic images to generate a thumbnail summarization for each episode.

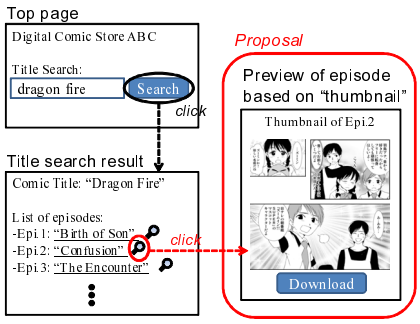

Figure 2: Conceptual illustration of comic search interface using summarization results

4.2 Frame segmentation

The frame segmentation method developed for our system utilizes the results of two existing approaches: iterative frame separation (IFS), and frame corner detection (FCD). The combination of these two methods has been proved to improve frame segmentation accuracy in [2].

First, IFS is conducted to obtain a rough estimation of segmented frames, following the method proposed in [1]. The basic approach of this method is to recursively split the input comic image to two sub-regions until no separation line can be detected. Next, FCD is implemented by extracting corners from the comic image based on two methods: selecting intersection points of lines extracted by Hough transformation, and Harris's corner detection method. Based on the IFS results, a multiple number of frame boundary candidates are generated by drawing lines which intersect with the corners detected by FCD. The average density of gradient values for all frame candidates are calculated, based on the result of applying the Sobel filter to the page image. Finally, the candidate with the highest score is selected as the optimal segmentation result for each of the initially segmented frames.

4.3 Text balloon extraction

Japanese text consist of a wide variety of characters, especially compared to English. Furthermore, as mentioned in

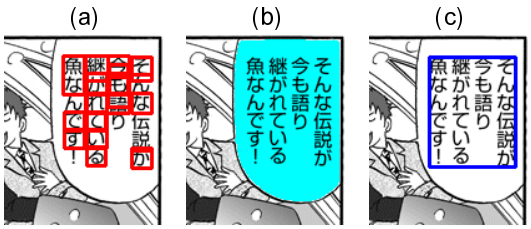

Figure 3: Example images of text balloon extraction steps: (a) text character extraction, (b) paint bucket filling, (c) gradual expansion.

Table 1: Extracted frame features

Type	Feature
Frame	width (pixels)
	height (pixels)
	area (pixels)
	coordinates of upper left corner (x, y)
Text balloon	area (pixels)
	number of characters
	mean character size (pixels)

Section 3, text character extraction from comic images is not a trivial problem. Therefore, development of a technology to extract text balloons from Japanese story comics for feature extraction of comic images is necessary. The method we have developed to achive this goal is based on the following two steps. First, initial text character regions are extracted from the input comic image, based on text character detection. Next, the area of the text balloon is determined by expanding the extracted candidate regioins.

The first step is to extract text character candidates from the comic page image. This procedure is implemented by utilizing the face detection method of Viola *et al* [9]. A text detector is built based on training data, which consists of images of Japanese text characters as the positive data set, and randomly extracted non-textual sub-regions from the comic images, as the negative data set. Haar-like features are extracted from all images in the training data, and the text detector is constructed by Adaboost learning.

Next, the area of the text balloon is determined by expanding the region of the extracted text characters. Area expansion is conducted by a hybrid of two methods. One is the paint bucket filling approach, which simply fills the white space surrounding each extracted character. However, there are many cases which result in over-filling errors, due to small gaps in the balloon boundaries. In order to compensate with such errors, we also implement the gradual expansion method, where the candidate area is gradually expanded until all edges of the surrounding area become white. This approach is applied when the result of the paint bucket method verges to multiple edges of the comic page image. All overlapping results are merged to obtain the final text balloon area. Example images of each step in the text balloon extraction procedure are shown in Figure 3.

4.4 Frame feature extraction

Based on the results of the frame segmentation and text balloon extraction steps, a feature vector is generated for each segmented frame. The frame features extracted in this process are listed in Table 1.

As written in Table 1, frame features are roughly categorized to two types: features of the frame itself, and features of the text balloon associated with each frame. The number of characters per text balloon is derived by applying a commercial Japanese optical character recognition (OCR) module. The mean character size is estimated by dividing the text balloon area by the number of characters in the OCR results. As a result, the features of frame F are expressed by an 8-dimensional feature vector $\vec{F} = (f_1, \cdots, f_8)$, where each element of $\vec{F}$ expresses the value of the features.

4.5 Frame importance score calculation and episode summarization

Features extracted from each frame are then used to calculate the importance score of the frame in question. A linear regression model is constructed to calculate the importance score of each frame F, denoted as follows: $Score(F) = \sum_{i=1}^{8} \beta_i \cdot f_i + \epsilon$, where β_i denotes the regression coefficient of the i-th feature.

In order to construct the importance score calculation model, we prepared a training data set by collecting subjective evaluation data of significant comic frames. In this process, 5 subjects were asked to read 8 episodes of Japanese comics, and select up to a maximum of 5 frames which they judged to be important for the episode. The ratio of subjects who selected each frame as important, and the feature vector of each frame are used as the regressand and regressors to estimate the coefficients β_i, and ϵ.

The thumbnail of a comic episode is generated by calculating the importance score of all frames of the episode, and selecting the frames with high importance. Namely, the top N frames, according to their importance scores, are selected to represent the episode. Since the chronological order of the frames is essential to express the episode content, the selected N frames are sorted according to their order of appearance, when the thumbnail is presented to the user.

5. EXPERIMENTS

A subjective user experiment is conducted to evaluate the efficacy of our proposal for comic search systems. For this experiment, we simulate a situation where users of a virtual comic distribution site are searching for a specific comic episode, from comic titles which consist of a large number of episodes. The subjects are assigned to use one of the following two methods to preview the content of comic episodes during the search task: our proposed thumbnail generation method, and the conventional (Amazon-like) previewing method, in which the subjects browse the first few pages of an episode. The overall efficiency required to successfully search the target episode is measured for evaluation. Details of the experiment are as follows.

5.1 Framework of experiment

In this experiment, the subjects are asked to search for specific episodes of popular Japanese comic titles. Namely, a short text abstract of the target episode is presented to the subjects as the query. The list of all episodes from the specified comic title is also presented to the subjects on a Web-based interface. By clicking on an episode link, subjects are able to see the "preview" of the episode. The episode preview is generated on one of the two previously mentioned methods. Based on the content of the episode preview, the subject judges whether or not the selected episode matches with the query. If the subject decides that the current episode

Table 2: Summary of comic titles used for evaluation experiments

ID	# of episodes	Avg page #	Avg # frames/page
T_1	64	26.3	4.65
T_2	276	19.5	4.47
T_3	108	21.4	4.61
T_4	233	18.8	5.61

matches the query, he/she clicks the button to "download" the episode. The search task concludes if the downloaded episode is a correct match. If not, the subject can browse the entire content of the downloaded episode, in order to grasp the story of the comic title.

For comparison of the proposed and conventional methods, the following measures are recorded for each experiment: time required to search target episode, and the number of episodes browsed per task. These measures are used to evaluate the efficiency of the episode search task for both methods. In addition to these objective measures, all subjects were asked to answer a questionnaire after the experiment. The main objective of the questionnaire is to collect information about each subjects' familiarity to the comic titles used for this experiment. Namely, the subjects are asked if they have read each comic title prior to the experiment, and also their general acquaintance with (both digital and paper) comics.

5.2 Experimental data

Four popular Japanese comic titles $T_1, \cdots, T_4$ are used for our experiments. The number of episodes per title, average number of pages per episode, and average number of frames per page are presented in Table 2.

84 subjects have cooperated for the experiment (41 male, 43 female, age range from late teens to 50s). All subjects have general experience in reading Japanese story comics. Two queries are given to the subject for each title $T_1, \cdots, T_4$, thus, each subject is asked to conduct the search task eight times. The preview method for each title in the experiment is decided so that each subject previews two of the four titles by the proposed method, and the other two by the conventional method. The combination of the title and preview method is determined randomly for each subject.

5.3 Results

Table 3 lists the average time (secs), and the number of previewed episodes that were necessary for each user to search for a single query. As clear from this result, the proposed method has enabled users to search for comic episodes in significantly less time, compared to the conventional method. It is also notable that the number of episode previews for the proposed method is higher than that of the conventional method. These results indicate that the conventional method to present the first few pages of a comic episode is more time-consuming than the proposal, which is assumed to be the main cause of the search time difference between the two methods.

Furthermore, we focus on the time required for users to search for episodes of comic titles which they are familiar with. Table 4 lists the average search time per title, for the subjects who have read more than half of all episodes of each comic title. The number of such familiar subjects per title is also written in this Table.

Table 3: Average time and number of browsed episodes in search experiment

Method	Avg time (secs)	# of browsed episodes
Preview	324.7	34.1
Conventional	373.5	22.7

Table 4: Average search time (secs) for subjects familiar to each comic title

	Title ID				
Avg time	T_1	T_2	T_3	T_4	Avg
Prev	306.4	200.2	221.9	328.2	264.2
Conv	365.9	405.8	295.4	341.4	352.1
# of subjects	36	51	44	33	41.0

As clear from Table 4, the average time difference between the proposed and conventional methods is 87.9 secs, which is significantly higher than that of all experiments (48.8 secs, see Table 3). This result indicates that the proposed method is especially effective for users who understand the general outline of the comic title.

6. CONCLUSIONS

This paper proposes a novel method to generate and present thumbnail-like summaries of episodes for digitized comics. This method is expected to improve the efficiency to search comic episodes on online comic distribution services. Subjective experimental results conducted with actual Japanese comic data have proved that our proposal can reduce the time necessary to search for specific comic episodes from a large-scaled data collection. We believe our research will contribute to improve the usability of digitized comic distribution services, and support the future prosperity of the comic distribution industry.

7. REFERENCES

[1] T. Tanaka, K. Shoji, F. Toyama, J. Miyamichi: "Layout analysis of tree-structured scene frames in comic images," Proceedings of IJCAI 2007, pp. 2885-2890, 2007.

[2] D. Ishii, H. Watanabe: "A study on frame position detection of digitized comics images," 2010 Workshop on Picture Coding and Image Processing (WPCIP), pp. 124-125, 2010.

[3] M. Yamada, R. Budiarto, M. Endo, S. Miyazaki: "Comic image decomposition for reading comics on cellular phones," IEICE Trans. Info. Sys., Vol. E87-D, No. 6, pp. 1370-1376, 2004.

[4] K. Arai, H. Tolle: "Method for automatic e-comic scene frame extraction for reading comic on mobile devices," Proceedings of ITNG 2010, pp. 370-375, 2010.

[5] K. Arai, H. Tolle: "Automatic e-comic content adaptation," Int'l Journal of Ubiquitous Computing, Vol. 1, Issue 1, pp. 1-11, 2010.

[6] K. Junga, K.I. Kimb, A.K. Jainc: "Text information extraction in images and video: a survey," Pattern Recognition, Vol. 37, Issue 5, pp. 977-997, 2004.

[7] V. Wu, R. Manmata, E.M. Riseman: "TextFinder: an automatic system to detect and recognize text in images," IEEE Trans PAMI, Vol. 21, No. 11, pp. 1224-1229, 1999.

[8] Q. Guo, K. Kato, N. Sato, Y. Hoshino: "An algorithm for extracting text strings from comic strips," Proc. of ACM SIGGRAPH 2006, 2006.

[9] P. Viola, M. J. Jones: "Robust real-time face detection," Int. J. Comput. Vis. Vol. 57, No. 2, pp. 137-154, 2004.