SKETCH2MANGA: SKETCH-BASED MANGA RETRIEVAL

Yusuke Matsui *Kiyoharu Aizawa* *Yushi Jing*

The University of Tokyo

ABSTRACT

We propose a sketch-based method for manga image retrieval, in which users draw sketches via a Web browser that enables the automatic retrieval of similar images from a database of manga titles. The characteristics of manga images are different from those of naturalistic images. Despite the widespread attention given to content-based image retrieval systems, the question of how to retrieve manga images effectively has been little studied. We propose a fine multi-scale edge orientation histogram (FMEOH) whereby a number of differently sized squares on a page can be indexed efficiently. Our experimental results show that FMEOH can achieve greater accuracy than a state-of-the-art sketch-based retrieval method [1].

Index Terms— Image Retrieval, Sketch, Manga

1. INTRODUCTION

Manga (Japanese comics) are popular in many parts of the world. Nowadays, manga are being distributed not only in print but also electronically via online stores such as Amazon. They are read on devices such as Kindles, iPads, and PCs. Some of the e-manga archives are very large, (e.g., the Amazon Kindle store sells more than 30,000 e-manga titles[1]), but offer very limited search support (keyword-based search by title or author). For this reason, applying content-based techniques would have the potential to make the manga-search experience more intuitive, efficient, and enjoyable.

There are two unique challenges associated with content-based manga retrieval. First, the visual characteristics of manga images are very different from those of naturalistic images. For example, as shown in Figure 1(c), manga images are usually line drawings made up of black lines on a flat white ground. Local-feature characterization, such as SIFT [2], is not suited to describing manga images, which do not have varying gradient intensities. Furthermore, a manga image (a page) comprises several frames (rectangular areas). It is therefore necessary to retrieve not only an image but also part of the image, as shown in Figure 1(c), where the red square region within the image is detected. A traditional bag-of-features (BoF) image representation framework [3] is not suited to retrieving an area of an image because it discards spatial information. Therefore, even if two BoF vectors are similar, it is hard to identify the areas that are similar within the images.

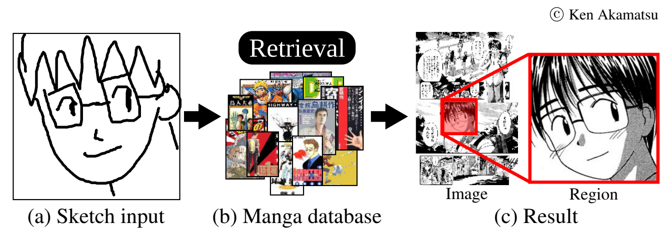


Fig. 1. (a) A user draws a rough sketch. (b) Similar images are retrieved from a manga database. (c) An image containing a region that best matches the sketch is retrieved.

The second challenge is to avoid limitations in the query modality. Because manga images are mostly black-and-white drawings, users should have access to some form of drawing, such as examples of manga images or sketches of their own. For this reason, it is important to provide a natural sketch-based interface for manga retrieval.

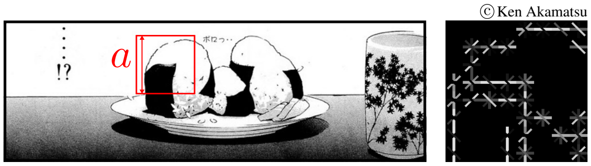
In this paper, we propose a content-based manga retrieval system that addresses the above two challenges. First, we propose a manga-specific image feature, namely the fine multi-scale edge orientation histogram (FMEOH), and a framework for extracting it. The system is composed of three steps: the labeling of margin areas, FMEOH feature description, and binary embedding. From a comparative study, we confirm that the proposed method can obtain better accuracy than a previous method, namely the edgel index [1]. Second, we study the feasibility of a sketch-based interface as a more natural way for people to interact with manga content. Sketch-based interfaces for retrieving images have been explored previously, and we conjecture that they are particularly suited to manga. An outline of the method is shown in Figure 1.

2. RELATED WORK

In the image processing literature, there is some research dealing with manga; colorization [4, 5], vectorization [6], layout recognition [7], layout generation [8, 9], element composition [10], manga-like rendering [11], speech balloon detection [12], segmentation [13], and retargeting [14]. Although

[1] Amazon.co.jp Kindle Comic. Retrieved from January 28, 2014, from `http://amzn.to/WWXSuw`

©Ken Akamatsu



(a) Input image and selected area (b) Visualized FMEOH

Fig. 2. (a) An image and a selected square area. a measures the side of the feature. (b) The visualized FMEOH feature extracted from the red selected area in (a).

some commercial products such as Google image search can retrieve cropped manga frames using text queries via the Web, they rely heavily on textual information related to the frame. The scope of the retrieval is therefore limited. Several methods have tackled the sketch-based retrieval of naturalistic images, where queries are sketches drawn by users [15, 16, 1, 17, 18, 19, 20, 21].

3. MANGA-SPECIFIC IMAGE REPRESENTATION

In content-based image retrieval, the basic form of image representation is the BoF representation, whereby local features such as SIFT [2] are extracted from the image and quantized into a sparse vector [3]. When a manga image is represented as a BoF, there are two problems. First, local features are not suitable because the intensity gradients are low. We show that BoF matching achieves poor results in Section 5. Second, because BoF discards spatial information, it is difficult to locate the part of the image containing the similarity.

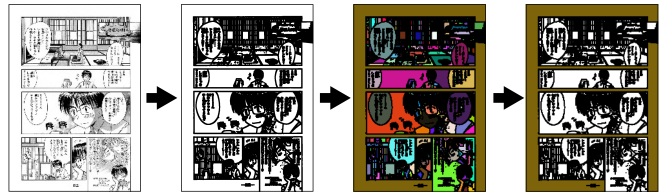
To deal with these problems, we propose the FMEOH, which defines spatial areas in terms of a multi-scale sliding window, and comprises a histogram of edge orientation for each area. It is based on the EOH [22], and we generate the FMEOH by using a fine multi-scale moving window. The method can locate features in images because a feature denotes a region in the image.

We show an example in Figure 2. If the red square area in Figure 2(a) is selected, its FMEOH is visualized in Figure 2(b). Such FMEOH descriptors are extracted at many different scales and for many positions on the page. Each FMEOH feature represents a square region, and the retrieval is performed by matching of an FMEOH feature of the query against the FMEOH features in the database.

3.1. FMEOH features

A square area is divided into 8×8 cells. The edges of each cell are quantized into 4 orientation bins and normalized, and whole vector is then renormalized. For manga, the features need to be robust against scale changes because we want to find areas of any size that are similar to the input sketch. For example, if a certain character's face is the target, it can be either small or large. To achieve matching across scales, we use

©Ken Akamatsu



(a) Input page (b) Erosion of white areas (c) Connected component labeling (d) Label of the margin areas

Fig. 3. (a) An input image. (b) Erosion is applied to white regions to thicken the lines. (c) White-connected areas are labeled with the same value. (d) The margin areas are selected.

square blocks with finely changing multiple sizes. FMEOH features are extracted at different scales and from different locations in the manga page. The page is then represented by a set of FMEOH features: $P = \{p_1, \ldots, p_n\}$ where P means the page and p_i denotes an FMEOH feature.

We extract FMEOH features in a sliding window manner. We first set a small constant value $a = a_0$ and start to extract FMEOH features via a sliding window, left top to right bottom, with a sliding step size of $a/8$. The scale is then changed, with a being slightly enlarged via $a \leftarrow 2^{1/3}a$. This procedure iterates until a exceeds the size of the side of the manga page.

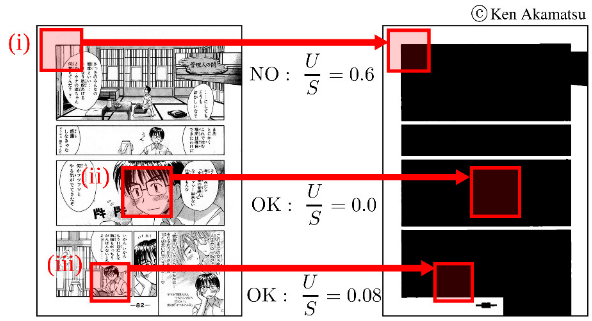
This sliding-window-based representation is simple and efficient at finding the matching area in the image because the feature implicitly describes the position in the image. This property is particularly useful for manga page. Its disadvantage is that it can require much memory and computation time to compare features, and we solve them in Sections 3.2.

3.2. Binary embedding

Given a set of FMEOH features, we convert them into binary strings for efficient computing and for saving memory. After the features are converted to binary strings, the Hamming distance between two binary strings can be computed quickly using low-level machine operations. The search can therefore be performed efficiently even for large numbers of database vectors. We leverage shift-invariant kernel LSH (SKLSH) [23] for converting features into binary strings.

By converting each feature (p_i) to a binary string ($B(p_i)$), an image is represented as $P_B = \left\{B(p_i) \in \{1,0\}^b \middle| i = 1, \ldots, n\right\}$ where P_B is a set of binary strings for a page. b is the length of the string, and $B(\cdot)$ is the binarize operator in SKLSH.

If a binary string from a query sketch is converted into $B(q_j)$, the search engine computes the nearest neighbors using $\langle p_i^*, q_j^* \rangle = \arg\min d_H(B(p_i), B(q_j))$, where $d_H(\cdot,\cdot)$ denotes a Hamming distance.



(a) Input page (b) The unnecessary label

Fig. 4. (a) Input image. (b) Its margin labels. In case (i), the red area (i) in (a) is skipped because $U/S = 0.6 > 0.1$. In case (ii), in contrast, the corresponding area is all black, and the feature is therefore extracted. In case (iii), $U/S = 0.08 < 0.1$, and the FMEOH is extracted.

3.3. Skipping margins

Manga comprises a series of frames, with the inter-frame space (margin) being unimportant. For efficient searching, margin exclusion from retrieval candidates should be performed. We label the margins as shown in Figure 3. First, the lines of the manga page image (Figure 3(a)) are thickened by applying an erosion [24] to the white areas (Figure 3(b)), thereby filling small gaps between black areas. Next, the white-connected areas are labeled by connected-component labeling [25] as shown in Figure 3(c), where different colors for the labels have been used for visualization. Finally, areas are selected as margins by finding the most frequent label appearing in the outermost peripheral regions (Figure 3(d)). Because inter-frame spaces tend to connect to the outer areas, the method succeeds in most cases.

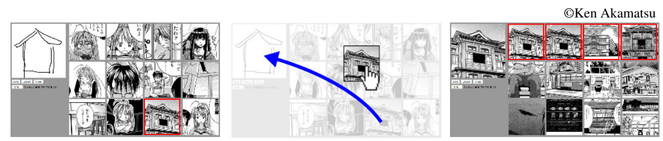
We define the square area for which the FMEOH feature p is extracted as $S(p)$, and the margins as $U(p)$ (e.g., the colored areas shown in Figure 3(d)). We extract a feature only if its area ratio U/S is less than a threshold, which we set to 0.1. P_B is therefore rewritten as $P_B = \{B(p_i) | U(p_i)/S(p_i) < 0.1, \ i = 1, \ldots, n\}$. Intuitively, this means that if an area belongs to the margin, it is skipped in the feature extraction step. An example of the process is shown in Figure 4.

4. QUERYING BY SKETCH

Querying is a difficult issue for manga retrieval. We cannot employ textual or tag data because FMEOH features do not involve such information. For natural and intuitive interface, we would prefer a sketch-based query. Because manga itself comprises sketches drawn by authors, sketching is compatible with manga. Figure 5 shows an example of the proposed method. The method is fully implemented on the Amazon EC2 parallel platform and the interface is provided by a Web

(a) Querying by "two guys" (b) The result of the manga page corresponding to the 1st ranked in (a)

Fig. 5. (a) Interface for the retrieval. The left area is the canvas for the user's sketch. In this case, two guys are drawn. The right area shows the retrieved results. (b) When the user clicks on one of the results, the page containing it is shown.



(a) First sketch (b) Dragging a result to the canvas (c) Retrieval by reusing the result

Fig. 6. Relevance feedback. (a) A user draws strokes and finds a target building on 11^{th} result (red colored). (b) By dragging it to the canvas, another retrieval is performed automatically with this as the query. (c) A wide range of images of the target building is obtained.

browser. The retrieval is performed automatically after each stroke is completed.

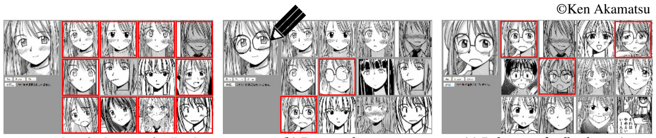
In addition, we can make use of sketching not only for the initial query but also for additional interaction with the retrieved results. The queries that can be performed using our framework are summarized as follows: (1) sketch querying, the proposed sketch-based retrieval described above. (2) *relevance feedback*, that reuses the retrieved results. (3) *query retouch*, where the result of relevance feedback is modified and reused. (4) combinations of the above.

Relevance feedback extends the existing method. Users can reuse a retrieved result simply by dragging the result to the canvas as shown in Figure 6. With relevance feedback, even novice users can use professional manga images as queries. Query retouch is a new methodology we propose in this paper. In query retouch, we can modify either the initial sketch or a query taken from the results of a retrieval (Figure 7). As the query is changed by adding lines or partial erasure, the results will change immediately. As a result, we can intuitively change the results of the retrieval in the direction that we want[2].

5. EVALUATION1: COMPARISON OF FEATURES

We compare the proposed FMEOH to BoF using a SIFT method [3] (as a baseline) and to edgel index [1] (represent-

[2]Because of a copyright issue, we show in Figure 6, 7 results using a small database, and it does not include binary embedding.



(a) Results of relevance feedback (b) Draw a glass (c) Relevance feedback, again

Fig. 7. Query retouch. (a) Results of relevance feedback where target characters are red colored. (b) The user adds strokes and draws a glass with the target character. Another target characters with a glass are then retrieved (colored red). Note users can erase lines using an eraser tool. (c) Both relevance feedback and query retouch can be conducted multiple times.

ing the state of the art), to assess retrieval ability. In the comparison, we conducted a character retrieval task using the manga *Dragon Ball*. In addition, we examined how drawing ability might affect the results.

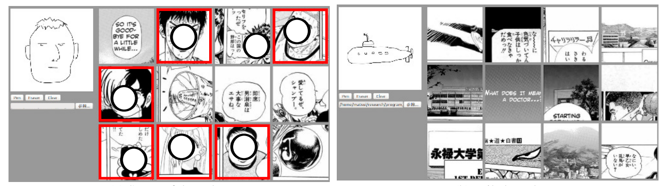
We gathered 18 subjects (9 novices and 9 semiprofessional drawers). First, we showed them a paper containing object characters for 10 seconds, and removed it. Next, they were instructed to draw the character on a computer screen. We followed Rodden's report [26] for this setup. After obtaining the sketches, a retrieval was performed, for which we calculated the mean average precision[3] (mAP) and the average of the top 10 precision values (precision@10) of each image. The character *Frieza* from *Dragon Ball* was used as the object. We used frames (manually cropped from the manga) as the database because the purpose here was to compare the features. The initial value for the square's side in FMEOH, a_0, was set to 160, and the bit length was set to 256 byte. For the evaluation, we manually cropped 1,503 frames from *Dragon Ball* Vols. 26 and 27. Among these, we identified 208 images with *Frieza* as positive samples, and the remainder as negative.

Table 1 gives the results of the comparative study, with mAP measuring retrieval precision and recall. Here, the proposed FMEOH method has the best performance for both novice and semiprofessional users. Precision@10 measures how correct the returned top 10 results are. If precision@10 $= 0.1$, a user would obtain one correct *Frieza* in 10 returned results. FMEOH is also the best method in terms of precision@10. One reason for the worse results with edgel index is that the method is not robust against position changes, e.g., it cannot match a query with an "apple" at the center of the canvas to an image where the "apple" is in the top leftmost corner. For a Web-image retrieval task, this would not be a problem because the quantity of Web images is large and there would be many "apple" images. For manga images, however, we would have to pay more attention to the position.

[3]We considered the top 50 results.

Table 1. Results of the comparative study.

	mAP		mPrecision@10	
	Novice	Semi-professional	Novice	Semi-professional
BoF [3]	0.179	0.228	0.111	0.148
Edgel index [1]	0.242	0.195	0.210	0.086
Proposed	**0.284**	**0.292**	**0.247**	**0.210**



(a) Successful resulsts (b) Failed result

Fig. 8. (a) Successful results for the larger-scale evaluation. Results of a query "head" are shown (red results are successful). White holes are drawn because of their copyright. (b) The query was "submarine", and no results matched.

6. EVALUATION2: LARGER-SCALE EVALUATION

Next, we evaluated the method using a larger dataset. We used *representative sketches* [27] as queries. These comprised 347 sketches, with each having a category name, e.g., "panda". For the evaluation, we used 30 manga titles (5,719 pages) as the database, which covered many manga genres. Note that the average size of a page was 1326×2036 pixels. We set a_0=360, and the number of bits as 256 byte. The average number of features per page was 1,512.

Figure 8(a) shows successful examples. We could retrieve objects from a manga database. Because faces are the most frequent type of object in manga images, such queries are well matched. In contrast, some queries such as "submarine" failed (Figure 8(b)). One reason is that the range of queries is very wide and there may be no manga that contains the category of the query, e.g., the query "submarine" did not appear in our manga database.

7. CONCLUSION

We have proposed a sketch-based manga retrieval method. Users draw sketches, with similar images being retrieved automatically from a database of manga titles. The method comprises the labeling of margin areas, FMEOH feature description, binary embedding, and a sketch interface. From our comparative study, we showed that the proposed method performed the best.

Acknowledgements: We would like to thank Stephen Holiday's help with the system integration.

8. REFERENCES

[1] Yang Cao, Changhu Wang, Liqing Zhang, and Lei Zhang, "Edgel index for large-scale sketch-based image search," in *Proc. CVPR*. IEEE, 2011.

[2] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[3] Josef Sivic and Andrew Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. ICCV*. IEEE, 2003.

[4] Yingge Qu, Tien-Tsin Wong, and Pheng-Ann Heng, "Manga colorization," in *Proc. SIGGRAPH*. ACM, 2006.

[5] Daniel Sýkora, John Dingliana, and Steven Collins, "Lazybrush: Flexible painting tool for hand-drawn cartoons," in *Proc. Eurographics*, 2009.

[6] Johannes Kopf and Dani Lischinski, "Digital reconstruction of halftoned color comics," in *Proc. SIGGRAPH Asia*. ACM, 2012.

[7] Takamasa Tanaka, Kenji Shoji, Fubito Toyama, and Juichi Miyamichi, "Layout analysis of tree-structured scene frames in comic images," in *Proc. IJCAI*, 2007.

[8] Keiichiro Hoashi, Chihiro Ono, Daisuke Ishii, and Hiroshi Watanabe, "Automatic preview generation of comic episodes for digitized comic search," in *Proc. MM*. ACM, 2011.

[9] Ying Cao, Antoni B. Chan, and Rynson W. H. Lau, "Automatic stylistic manga layout," in *Proc. SIGGRAPH Asia*. ACM, 2012.

[10] Ying Cao, Rynson W. H. Lau, and Antoni B. Chan, "Look over here: Attention-directing composition of manga elements," in *Proc. SIGGRAPH*. ACM, 2014.

[11] Yingge Qu, Wai-Man Pang, Tien-Tsin Wong, and Pheng-Ann Heng, "Richness-preserving manga screening," in *Proc. SIGGRAPH Asia*. ACM, 2008.

[12] Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier, Dimosthenis Karatzas, and Joost Van de Weijer, "An active contour model for speech balloon detection in comics," in *Proc. ICDAR*. IEEE, 2013.

[13] Yuji Aramaki, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa, "Interactive segmentation for manga," in *Proc. SIGGRAPH Poster*. ACM, 2014.

[14] Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa, "Interactive manga retargeting," in *Proc. SIGGRAPH Poster*. ACM, 2011.

[15] Rui Hu, Mark Barnard, and John Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *Proc. ICIP*. IEEE, 2010.

[16] Rui Hu, Tinghuai Wang, and John Collomosse, "A bag-of-regions approach to sketch-based image retrieval," in *Proc. ICIP*. IEEE, 2011.

[17] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *IEEE TVCG*, vol. 17, no. 11, pp. 1624–1636, 2011.

[18] Kai-Yu Tseng, Yen-Liang Lin, Yu-Hsiu Chen, and Winston H. Hsu, "Sketch-based image retrieval on mobile devices using compact hash bits," in *Proc. MM*. ACM, 2012.

[19] Rong Zhou, Liuli Chen, and Liqing Zhang, "Sketch-based image retrieval on a large scake database," in *Proc. MM*. ACM, 2012.

[20] Xinghai Sun, Changhu Wang, Chao Xu, and Lei Zhang, "Indexing billions of images for sketch-based retrieval," in *Proc. MM*. ACM, 2013.

[21] Shuai Ren, Cheng Jin, Chang Sun, and Yuejie Zhang, "Sketch-based image retrieval via adaptive weightinge," in *Proc. ICMR*. ACM, 2014.

[22] Kobi Levi and Yair Weiss, "Learning object detection from a small number of examples: the importance of good features," in *Proc. CVPR*. IEEE, 2004.

[23] Maxim Raginsky and Svetlana Lazebnik, "Locality-sensitive binary codes from shift-invariant kernels," in *Proc. NIPS*, 2009.

[24] Jean Serra, *Image Analysis and Mathematical Morphology*, Academic Press, 1983.

[25] Stéphane Marchand-Mailet and Yazid M. Sharaiha, *Binary Digital Image Processing: A Discrete Approach*, Academic Press, 1999.

[26] Kerry Rodden, "Evaluating similarity-based visualizations as interfaces for image browsing," Technical Report YCAN-CL-TR-543, University of Cambridge, 2002.

[27] Mathias Eitz, James Hays, and Marc Alexa, "How do humans sketch objects?," in *Proc. SIGGRAPH*. ACM, 2012.