CrossMark

Sketch-based manga retrieval using manga109 dataset

Yusuke Matsui[1] · Kota Ito[1] · Yuji Aramaki[1] ·
Azuma Fujimoto[1] · Toru Ogawa[1] ·
Toshihiko Yamasaki[1] · Kiyoharu Aizawa[1]

Abstract Manga (Japanese comics) are popular worldwide. However, current e-manga
archives offer very limited search support, i.e., keyword-based search by title or author.
To make the manga search experience more intuitive, efficient, and enjoyable, we propose
a manga-specific image retrieval system. The proposed system consists of efficient mar-
gin labeling, edge orientation histogram feature description with screen tone removal, and
approximate nearest-neighbor search using product quantization. For querying, the system
provides a sketch-based interface. Based on the interface, two interactive reranking schemes
are presented: relevance feedback and query retouch. For evaluation, we built a novel dataset
of manga images, Manga109, which consists of 109 comic books of 21,142 pages drawn
by professional manga artists. To the best of our knowledge, Manga109 is currently the
biggest dataset of manga images available for research. Experimental results showed that

✉ Yusuke Matsui
matsui@hal.t.u-tokyo.ac.jp

Kota Ito
ito@hal.t.u-tokyo.ac.jp

Yuji Aramaki
aramaki@hal.t.u-tokyo.ac.jp

Azuma Fujimoto
fujimoto@hal.t.u-tokyo.ac.jp

Toru Ogawa
t_ogawa@hal.t.u-tokyo.ac.jp

Toshihiko Yamasaki
yamasaki@hal.t.u-tokyo.ac.jp

Kiyoharu Aizawa
aizawa@hal.t.u-tokyo.ac.jp

[1]   The University of Tokyo, Tokyo, Japan

the proposed framework is efficient and scalable (70 ms from 21,142 pages using a single computer with 204 MB RAM).

Keywords Manga · Sketch · Retrieval · Dataset

1 Introduction

Manga are Japanese black-and-white comics (Fig. 1). They are popular worldwide. Nowadays, manga are distributed not only in print, but also electronically via online stores. Some e-manga archives are very large; e.g., the Amazon Kindle store sells more than 130,000 e-manga titles.[1] Therefore, users of online stores must retrieve titles from large collections when they purchase new manga.

However, current e-manga archives offer very limited search support, i.e., keyword-based search by title or author. They are not suitable for large-scale search and searches cannot take the images (contents) of manga into consideration. For this reason, applying content-based multimedia retrieval techniques to manga search would have the potential to make the manga-search experience more intuitive, efficient, and enjoyable.

In this paper, we propose a content-based manga retrieval framework. The framework is composed of three steps: labeling of margin areas, an objectness-based edge orientation histogram (EOH) feature description [35] with screentone removal [23], and approximate nearest-neighbor (ANN) search using product quantization (PQ) [26]. For querying, the framework provides a sketch-based interface. Based on the interface, two interactive reranking schemes are proposed for an intuitive search: **relevance feedback** and **query retouch**. The main contribution of this paper is to build an efficient large-scale framework for manga search.

An overview of the GUI implementation of the proposed system is shown in Fig. 2. Given a query sketch (Fig. 2a), the system retrieves similar areas from a manga dataset in real time, and shows the top results in thumbnail windows (Fig. 2c). The retrieval is performed automatically after each stroke is completed. If a user clicks on one of the retrieved results in the thumbnail windows, a page containing the result is shown in a preview window (Fig. 2d). In this case, the page containing the first-ranked result is shown. Please see the supplemental video for more details.

There are two technical challenges associated with content-based manga retrieval. First, image description for naturalistic images may not be suited to describing manga images because the visual characteristics of manga images are different from those of naturalistic images. Second, it is necessary to retrieve not only an image but also a part of the image because a manga page comprises several frames (rectangular areas). Both problems are tested using the proposed framework.

For evaluation, we built a novel dataset of manga images drawn by professional manga artists, Manga109, which consists of 109 comic books with a total of 21,142 pages. Manga109 will be very beneficial for the multimedia research community because it has been difficult for researchers to use "real" manga in experiments due to the copyright problem. Research so far has been limited to the small scale. We made this dataset with permission from 94 professional creators. The dataset covers a wide range of manga genres, and is publicly available for academic use.

---

[1]Amazon.co.jp Kindle Comic. Retrieved from September 27, 2015, from http://amzn.to/1KD5ZBK
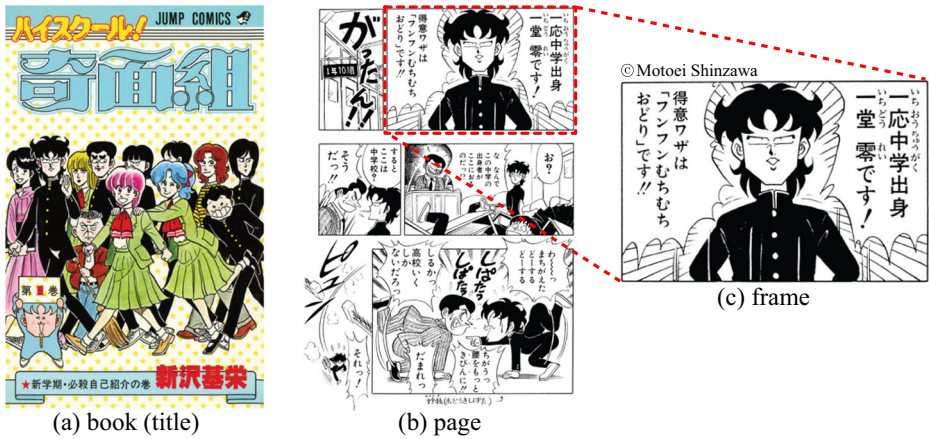
Fig. 1 Manga examples. **a** A manga book (title) consists of 200 pages on average. **b** A page (an image) is composed of several frames (rectangular areas). **c** Characters, text balloons, and background are drawn in each frame. Typically, manga are drawn and printed in black and white

Our contributions are summarized as follows:

– We built an efficient large-scale framework for manga search. The framework enables us to retrieve a manga image in 70 ms from 21,142 pages with 204 MB using a single PC.

– We provided a publicly available manga dataset, **Manga109**, which includes 109 manga titles containing 21,142 pages that will be useful for manga image-processing research.

– We verified that EOH with screentone removal is effective to manga image retrieval compared to those used in the state of the art sketch based retrieval system [54, 62].

– As far as we know, this is the first work to make use of sketch to retrieve manga or comics.
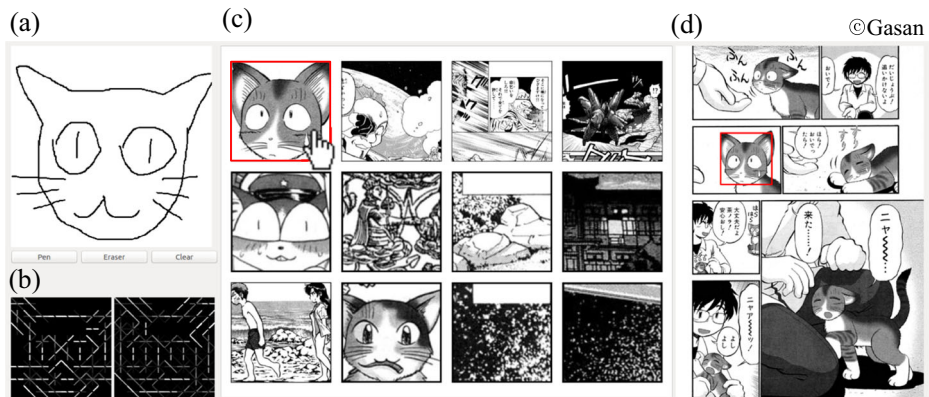

Fig. 2 An interface for the retrieval system. **a** A canvas for the user's sketch. **b** Visualized EOH features, where the left figure shows a query and the right shows the retrieved result. **c** Thumbnail windows for the retrieved results. **d** A preview window of a manga page

A preliminary version of this paper appeared in our recent conference paper [39]. This paper contains significant differences: (1) we constructed and released the publicly available manga dataset; (2) we significantly improved the proposed algorithm by leveraging an objectness measure, PQ, and screentone removal; and (3) we performed massive experimental evaluations using the new dataset, including comparisons with state-of-the-art methods.

The rest of the paper is organized as follows: Section 2 introduces related work. Section 3 presents our retrieval system and Section 4 shows the proposed query interaction schemes. A novel dataset is introduced in Section 5 and evaluations are given in Section 6.

2 Related work

We discuss related studies in content-based manga retrieval by categorizing them into manga description, retrieval–localization, ANN search, and querying. Additionally, we introduce research related to manga.

2.1 Manga image description

In the literature of content-based image retrieval, many methods have been proposed to describe images. A query and dataset images are converted to some form of vector representation to evaluate their similarity. A typical form is the bag-of-features (BoF) representation [57], in which local features such as SIFT [37] are extracted from an image and quantized into a sparse histogram. The distance between a query histogram and a dataset histogram can be computed efficiently using an inverted index structure. The BoF form has been studied widely because it achieved successful image description with fast computation.

Various extensions of BoF have been proposed so far, including query expansion [12], soft voting [45], and Hamming embedding [24]. The state-of-the-art in this line are the vector of locally aggregated descriptors (VLAD) and Fisher vector (FV) approaches [27, 59], which are interpreted as generalized versions of BoF [52].

However, all of the above methods were designed for naturalistic images for both query and dataset. For example, texture-based features such as Local Binary Pattern (LBP) [42] is not effective for manga because manga images do not contain texture information (Fig. 3). Thus, special attention has been paid to cases in which a query is a line drawing, i.e., a sketch.

If a query is presented as a form of sketch, several special features tailored for sketch queries have been proposed. BoF using a histogram of oriented gradients [13] was proposed [16, 54, 75]. Because sketch images usually have a flat background of uniform white pixels, several methods tried to fill such blank areas with meaningful values and interpreted them as a feature vector; e.g., distance-transformed values from edge pixels [65] and Poisson image-editing-based interpolation values [21, 22]. Chamfer matching, which is a successful method for computing the similarity between two contours of objects, was incorporated in the sketch-based retrieval framework [8, 48, 62].

In the manga feature description task, we compared the proposed method with (1) BoF, (2) large-window FV as the state-of-the-art of BoF-based methods [54], and (3) compact oriented chamfer matching (Compact OCM) as the state-of-the-art of chamfer-based methods [62].

Note that deep-learning techniques are becoming quite popular in image recognition because of their superior performance. These technologies are being applied to the sketch
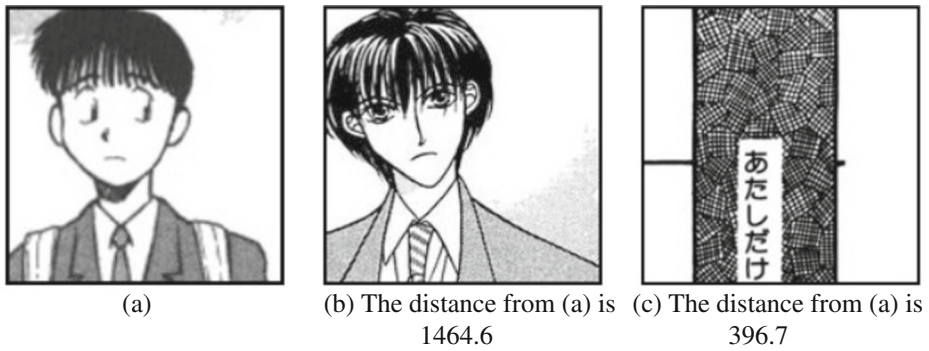
(a)   (b) The distance from (a) is 1464.6   (c) The distance from (a) is 396.7

Fig. 3 Poor performance of LBP. For each image, streentone is removed [23] , and the LBP feature is extracted. The differences between LBP features are then computed. Clearly, the visual characteristics of (a) and (b) are similar, whereas (a) and (c) are dissimilar. However, the distance between the LBP features for (a) and (b) is 1464.6, and between (a) and (c) is 396.7. This illustrates that the LBP feature cannot express visual similarities of line drawings. Note that the results become even worse if we do not remove screentone

processing. However, the effectiveness of such methods has not yet been discussed sufficiently. For example, Yu and colleagues [72] reported that traditional methods such as FV are still competitive for a sketch recognition task, and therefore such deep learning-based features are beyond the scope of this paper.

2.2 Retrieval and localization

As shown in Fig. 2d, our objective is to find not only a similar page, but also identify a similar region in the retrieved image. This kind of task has been tackled in spatial verification-based reranking methods [7, 25, 33, 44, 71, 74]. These methods first find candidate images by BoF, and rerank the top results by spatial information such as relative positions among extracted SIFT features. However, such methods cannot be applied to manga for two reasons. First, the manga image's structure is more complex than that of a naturalistic image. A single manga page usually consists of several frames, each of which is a totally different visual scene (see Fig. 1b). Therefore, a manga page can be interpreted as a set of images if we treat each frame as a single image. This is a much harder condition than those usual in image processing tasks such as retrieval or recognition, which assume that at least an image is a single image rather than a set of images. Second, BoF does not work well for manga as shown in the experimental section.

Another related area is object-detection methods, in which the position of a target object is identified in an input image. A number of methods have been proposed so far. Among them, contour-based object localization methods [56, 70] are closely related to our manga problem because an object is represented by its contour. In such methods, an instance of a target object such as "giraffe" is localized from an image by matching a query contour of giraffe, which is learned from training data. If we apply such methods to all dataset pages and select the best result, we might achieve both retrieval and localization at once. However, this is not realistic because: (1) these methods usually require learning to achieve reasonable performance, i.e., many giraffe images are required to localize an instance of giraffe, but a query is usually a single sketch in our case; and (2) localization usually takes time and cannot be scaled to a large number of images.

The proposed method makes use of a window-based matching approach. It is simple but usually slow. We formulate the search and localization as a single nearest-neighbor problem, and solve it using ANN methods.

2.3 Approximate nearest neighbor search

To determine similar vectors efficiently, several approximate nearest neighbor search methods have been proposed. The two main branches of ANN methods include product-quantization-based methods [18, 26] and hashing-based methods [36, 68, 69, 73]. Hashing-based methods mainly focus on supervised searches, i.e., label information is considered, such as image categories. Product-quantization-based methods approximate a distance measure more directly, which is more closely related to our considerations. Therefore, we employ product quantization in our pipeline for an efficient search.

2.4 Querying

How to query multimedia data has been an important problem in content-based multimedia retrieval. Although many approaches have been proposed, including keywords, images, and spatial/image/group predicates [58], we focus on sketches drawn by users. Sketch-based interaction is the most natural way for humans to describe visual objects, and is widely adapted not only for image retrieval, but also in many applications such as shape modeling [43], image montage [9], texture design [29], and as a querying tool for recognizing objects [15, 54].

2.5 Manga

From an image-processing perspective, manga has a distinctive visual nature compared with naturalistic images. Several applications for manga images have been proposed: colorization [47, 53, 63], vectorization [31], layout recognition [64], layout generation [5, 20], element composition [6], manga-like rendering [46], speech balloon detection [49], segmentation [2], face tailored features [11], screen tone separation [23], and retargeting [40]. The research group at Université de La Rochelle constructed a comic database [19], which we mention in Section 5, and analyzed the structure of visual elements in a page to understand the comics content semantically [50].

Several studies have explored manga retrieval [14, 34, 60, 61]. In these methods, inputs are usually not pages but cropped frames or characters, and evaluations were conducted using a small number of examples. In addition, the runtimes of these methods have not been discussed. Therefore, it is not known whether these methods can scale to a large dataset. On the other hand, our proposed method can perform a retrieval from 21,142 pages in 70 ms on a single computer.

3 Manga retrieval system

In this section, we present the proposed manga retrieval system. Figure 4 shows the framework of the system. First, input pages are preprocessed offline (Fig. 4d, as discussed in Section 3.4). Next, region of interest (ROI) areas are detected and EOH features are extracted with screentone removal (Fig. 4a, as described in Section 3.1). These features are
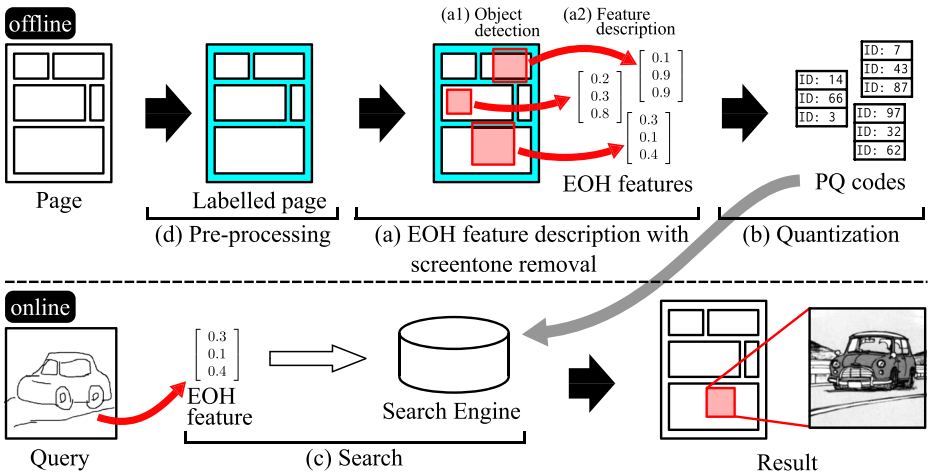
Fig. 4 Framework of the proposed system

then compressed into PQ codes (Fig. 4b, as discussed in Section 3.2). At the query phase, given a query sketch, similar EOF regions are retrieved (Fig. 4c, as discussed in Section 3.3).

The proposed framework is based on the window search paradigm; i.e., several features are extracted from dataset images using a generic object detector, and a query feature is compared with all features. Compared with the traditional image retrieval problem, such a window system is usually too slow to handle a large number of features. We formulate the problem as a single ANN search, and solve it using PQ, with effective preprocessing including the labeling of margin areas.

3.1 EOH feature description of object windows with screentone removal

We first show a core part of the system, feature description (Fig. 4a). Given a page, candidate areas of objects are detected automatically (Fig. 4(a1)), then EOH features [35] are described from selected areas (Fig. 4(a2)). The retrieval is performed by matching an EOH feature of the query against the EOH features in the dataset. Note that we delete screentones in advance by applying our previously proposed method [23]. This process significantly improves both the ROI detection and the feature description. Figure 5 shows an example of the feature representation. The EOF feature of the red square area in Fig. 5a is visualized as shown in Fig. 5b. The result of the screen tone removal is shown in Fig. 5c, and the feature is visualized in Fig. 5d. Comparing these two figures, we can observe that screentone removal effectively extracts line structures of the target, and suppresses noisy features caused by screentones. EOH is not new, but we found that it is effective to sketch based manga retrieval when screentone removed is applied in advance. Hereafter EOH in this paper means EOH with screentone removal.

The candidate areas of objects are detected by generic object detectors [1, 10, 66], which are successful ways to replace a sliding window paradigm. (Fig. 4(a1)). Intuitively, these detectors compute bounding boxes, which are more likely to contain any kinds of objects. These bounding boxes are considered as candidates for object instances, and a processing
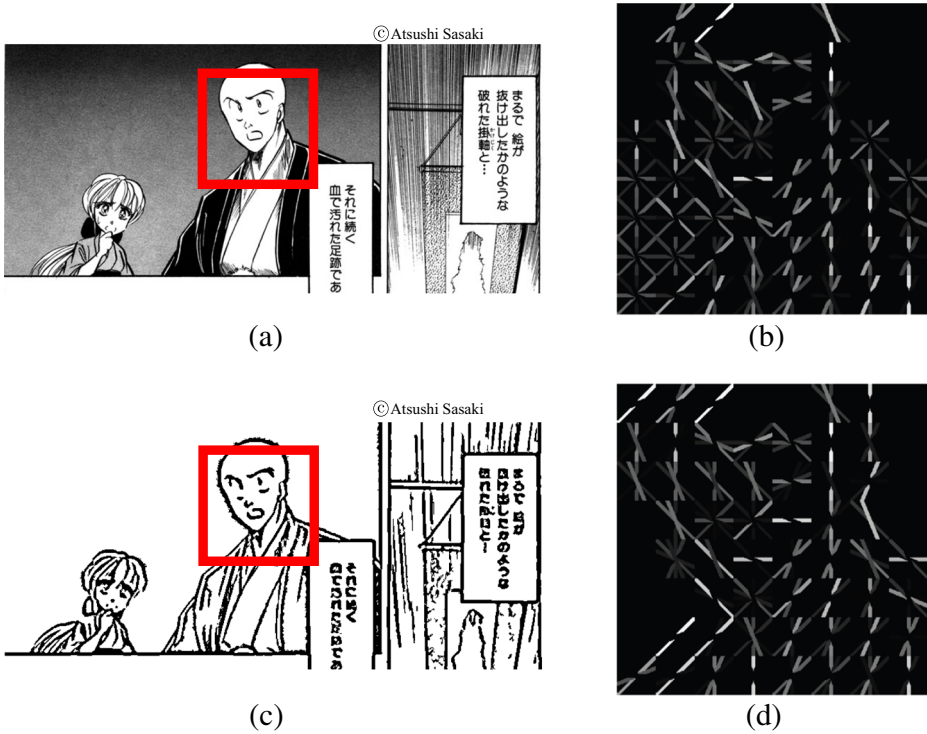
Fig. 5 An example of an EOH feature. **a** An image and a selected area. **b** The visualized EOH feature extracted from the area in (**a**). As can be seen, the visual characteristics of a manga page can be captured using such a simple histogram. **c** The image with screen tone removal [23]. **d** The visualized EOH feature extracted from the area in (**c**). By removing the screen tone, the quality of the feature is improved, e.g., the left bottom area of the feature in (**b**) includes random directions because of the background, where the feature in (**d**) does not

task such as image recognition is applied only to the bounding box areas, which greatly reduces computational costs compared with a simple sliding window approach.

We found that objectness measures can detect candidate manga image objectives more effectively than the sliding window approach. We conducted an experiment using a small dataset in which ground-truth objects (a head of a man) are annotated. The detection rate of a few methods were evaluated as shown in Fig. 6. BING [10], selective search [66], and sliding window (baseline) [39] are compared. From the results, we concluded that selective search is the best approach for this dataset. In the rest of this paper, we use selective search for our object detector. Note that a horizontally or vertically long bounding box is divided into several square boxes with overlaps because we restrict the shape of object areas to a square.

After object areas have been detected (typically, 600 to 1000 areas are selected from a page), their EOF features are described (Fig. 4(a2)). The selected square area is divided into $c \times c$ cells. The edges in each cell are quantized into four orientation bins and normalized, and the whole vector is then renormalized. Therefore, the dimension of the EOH vector is $4c^2$. Note that we discard features in which all elements are zero. For manga, the features should be robust against scale changes because we want to find areas of any size that are
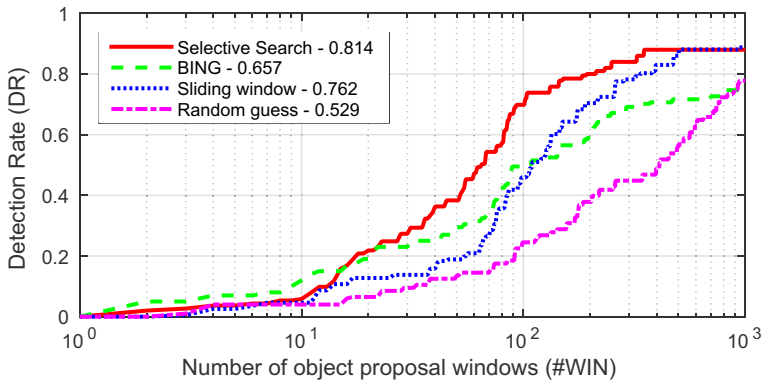
Fig. 6 Detection rate (DR) given #WIN proposals, which is a standard evaluation metric for evaluating objectness [1, 10, 66]. DR shows the percentage of ground-truth objects covered by proposals, which are generated by each method. An object is considered covered only when the PASCAL overlap criteria is satisfied (intersection over union is more than 0.5 [17]). The area under the curve (AUC), which is a single-value summary of performance for each curve, is shown on the legends. We evaluated methods using the dataset used for the localization task, see details on Section 6.2

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 different scales, we simply accept any sizes of patches, but restrict the shape of an area to square. Whatever the patch sizes, the same-dimensional EOH features are extracted and used for matching.

By describing EOF features of candidate areas, a page is represented by a set of EOH features:

$$\mathcal{X} = \{\boldsymbol{x}_n\}_{n=1}^{N}, \tag{1}$$

where $\mathcal{X}$ means the page, $\boldsymbol{x}_n$ denotes an EOH feature for the nth window, and N is the number of features extracted from the page. Note that integral images are utilized to enable the fast computing of features in the same manner as in [35]. In the comparative study, we show this simple representation achieves better accuracy than previous sketch-based retrieval methods, and we confirm that EOH-based description gives a good solution to the manga image representation problem.

In summary, given a page, screenton removal is applied, candidate areas are detected by selective search, and EOH features are extracted, then the page is represented as a set of EOH features. This window-based representation is simple and efficient at finding a matching area in the page because a feature implicitly contains its position in the page, which is particularly useful for manga pages.

3.2 Feature compression by product quantization

Given a set of EOH features, we compress them into binary codes using PQ [26], which is a successful ANN method, as shown in Fig. 4b. Applying PQ greatly reduces the computational cost of matching and reduces memory usage. After the features are compressed to binary codes, the approximate distances between a query vector and dataset codes can be computed efficiently by lookup operations (asymmetric distance computation (ADC) [26]). Therefore, the search can be performed efficiently even for large number of dataset vectors.

Each feature $\boldsymbol{x}_n$ is compressed into a binary code, $\boldsymbol{q}(\boldsymbol{x}_n)$, and the page is represented as

$$\mathcal{X}_{PQ} = \{\boldsymbol{q}(\boldsymbol{x}_n)\}_{n=1}^{N}, \tag{2}$$

where $\mathcal{X}_{PQ}$ is a set of quantized EOH features for a page. Because each $\boldsymbol{q}(\boldsymbol{x}_n)$ is recorded as a tuple of M subcentroid indices (8 bits `uint8_t`), which take $8M$ bits, $\mathcal{X}_{PQ}$ is stored by $8MN$ bits.

3.3 Search

Let us describe how to retrieve the nearest PQ-encoded feature from many manga pages (Fig. 4c). Suppose there are P manga pages, and the quantized EOH features of the pth page are represented as $\mathcal{X}_{PQ}^{p} = \left\{\mathbf{q}(\mathbf{x}_n^p)\right\}_{n=1}^{N^p}$, where N^p means the number of EOH features from the pth page, and $p \in \{1, \ldots, P\}$. Given an EOH feature $\mathbf{y}$ from a query sketch, the search engine computes the nearest neighbors using:

$$\langle p^*, n^* \rangle = \underset{p \in \{1,\ldots,P\},\ n \in \{1,\ldots,N^p\}}{\arg\min} d_{AD}\left(\boldsymbol{y}, \boldsymbol{q}(\boldsymbol{x}_n^p)\right), \tag{3}$$

where $d_{AD}(\cdot, \cdot)$ measures an approximate Euclidean distance between an uncompressed vector and a compressed PQ code [26]. This can be computed efficiently by lookup operations (ADC [26]).

By computing d_{AD} from $\mathbf{y}$ to each $\boldsymbol{q}(\boldsymbol{x}_n^p)$ and selecting the smallest one, p^* and n^* are computed. This is a linear search problem for PQ codes, which can be solved efficiently; e.g., 16 ms for searching one million 64-bit codes. Then we can say that the n^*th feature from the p^*th page is the nearest to the query sketch. The point here is that the problem is separated into a feature description and an ANN search. Therefore, even if the number of features is very large, we can employ the ANN techniques. In our system, searching from 21,142 pages takes 793 ms (without parallel thread implementation) or 70 ms (with parallel thread implementation) on a single computer using a PQ linear scan, which is fast enough. The search could be faster if we employ ANN data structures such as [3, 28].

3.4 Skipping margins

Although PQ computing is fast, effective preprocessing of the manga pages of the dataset is essential (Fig. 4d). A manga page comprises a series of frames and the interframe spaces (margins) are not important. For efficient searching, margin exclusion from retrieval candidates should be performed before extracting features. We label the margins as shown in Fig. 7. First, the lines of the manga page image (Fig. 7a) are thickened by applying an erosion [55] to the white areas (Fig. 7b), thereby filling small gaps between black areas. Next, the white-connected areas are labeled by connected-component labeling [38] as shown in Fig. 7c, 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 (Fig. 7d). Because interframe spaces tend to connect to the outer areas, the method succeeds in most cases.

Let us define the patch area of $\boldsymbol{x}$ as $S(\boldsymbol{x})$, and the margins as $U(\boldsymbol{x})$ (e.g., the colored areas shown in Fig. 7d). We extract a feature only if its area ratio U/S is less than a threshold, which we set to 0.1. Equation (2) is therefore rewritten as

$$\mathcal{X}_{PQ} = \left\{\boldsymbol{q}(\boldsymbol{x}_n) \left| \frac{U(\boldsymbol{x}_n)}{S(\boldsymbol{x}_n)} < 0.1, \ n \in \{1, \ldots, N\} \right. \right\}. \tag{4}$$

©Ken Akamatsu
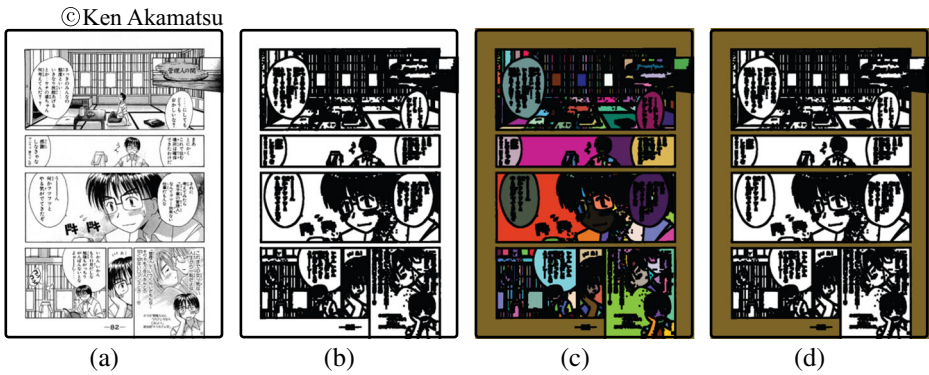

| (a) | (b) | (c) | (d) |

Fig. 7 **a** An input page. **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

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 Fig. 8.

4 Query interaction

Querying is a difficult issue for manga retrieval. For a natural and intuitive interface, we prefer sketch-based queries. Because manga itself comprise sketches drawn by authors, sketching is compatible with manga.

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**, which reuses the retrieved results; and (3) **query retouch**, in which the results of relevance feedback are modified and reused.
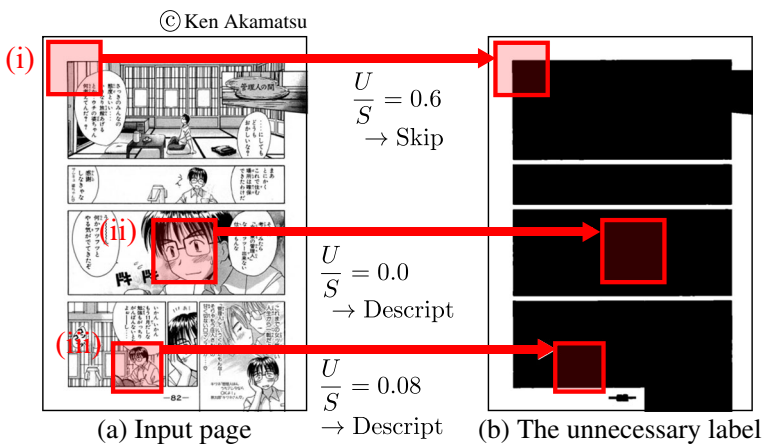
©Ken Akamatsu

(i) $\dfrac{U}{S} = 0.6$ → Skip

(ii) $\dfrac{U}{S} = 0.0$ → Descript

(iii) $\dfrac{U}{S} = 0.08$ → Descript

(a) Input page (b) The unnecessary label

Fig. 8 **a** Input image. **b** Its margin labels. In case of (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 an EOH feature is extracted

Relevance feedback We propose an interaction to reuse retrieved results, which is inspired by the relevance feedback techniques proposed in the information retrieval literature [41]. Users can reuse a retrieved result simply by selecting a region in the retrieved

©Chikae Ide

(a)

(b)

(c)

Fig. 9 Relevance feedback. **a** A user draws strokes and finds a target (Japanese-style clothes) in the third and fifth results (red borders). **b** By selecting the region in the retrieved page, another retrieval is performed automatically with this as the query. **c** A wide range of images of Japanese-style clothes is obtained

manga page, as shown in Fig. 9. With relevance feedback, even novice users can use professional manga images as queries. Note that a user can use any region in the page as a query.
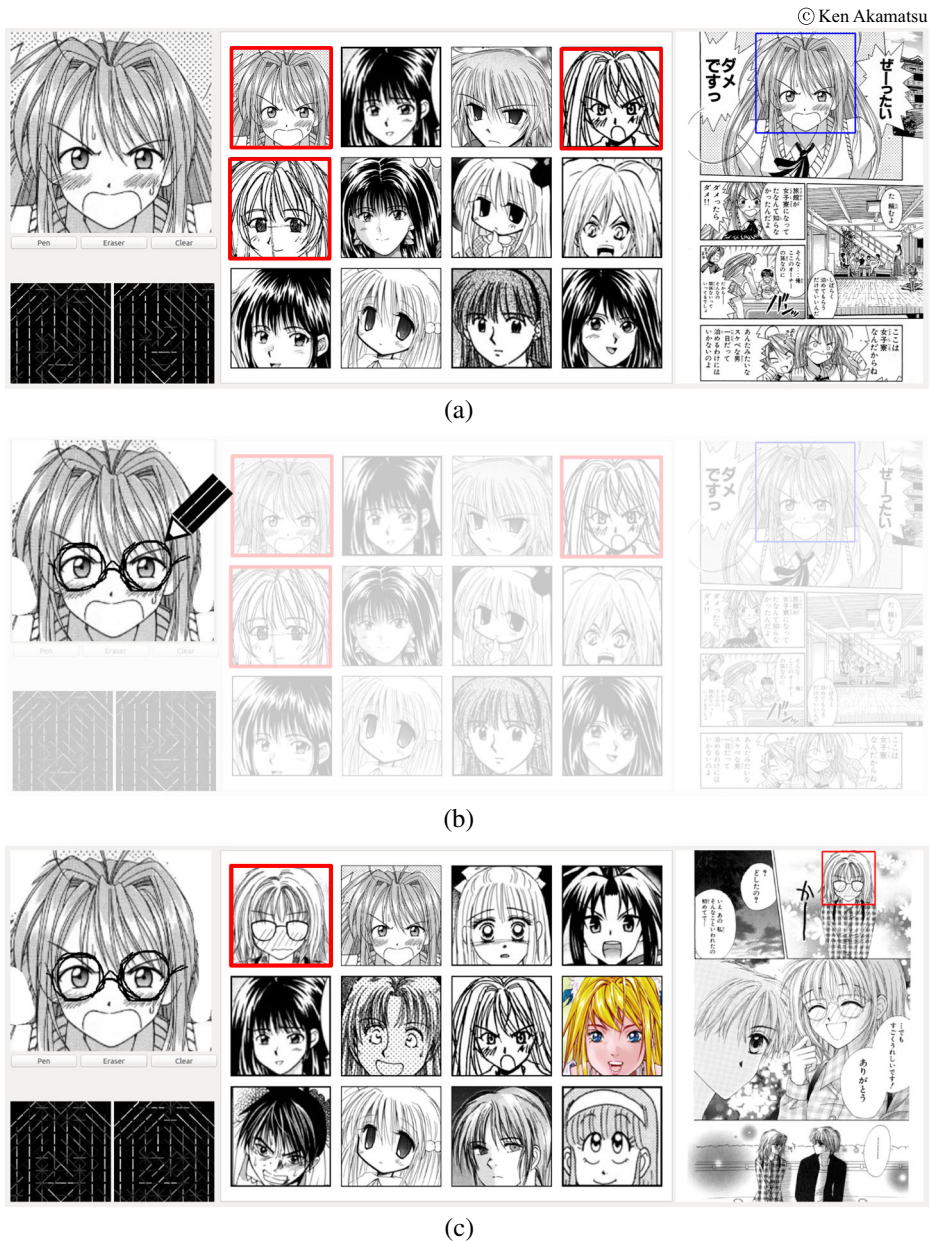

Fig. 10 Query retouch. **a** Results of relevance feedback. Target characters have red borders. **b** A user adds strokes and draws glasses on the target character. Other target characters with glasses are then successfully retrieved (red borders). Note that users can erase lines using an eraser tool. **c** Both relevance feedback and query retouch can be conducted multiple times

Query retouch We propose a new methodology for modifying and reusing the results. In query retouch, we can modify either the initial sketch or a query taken from the results of a retrieval (Fig. 10). As the query is changed by adding lines or partial erasure, the results will change immediately. As a result, we can intuitively change the retrieval results in the direction that we want.

According to Mei and colleagues [41], these interactions are classified as "interactive reranking," and are especially useful for the visual search problem because "*visual data is particularly suited for interaction, as a user can quickly grasp the vivid visual information and thus judge the relevance at a quick glance.*" [41].

Note that we have designed the proposed pipeline to be scale-**invariant**, rotation-**variant**, and flip-**variant**. Scale-invariance is required because the query should match with image regions of any size. This is achieved by extracting features of the same dimension from a given region (Section 3.1). Regarding rotations, the system does not accept results that are similar to the query if rotated. We decided on this design in order to inhibit unusual results, such as upside-down faces. Regarding flips, we decided to reject flipped results in order to make the sketch interaction more intuitive. Note that a flip-invariant search is technically easy to implement, simply by flipping the query feature and performing the retrieval again.

5 Manga dataset

In this section, we introduce a new dataset of manga images, namely the **Manga109** dataset for evaluation. The Manga109 dataset consists of 109 manga titles, and is made publicly available for academic research purposes with proper copyright notation. Figure 11 shows example pages from the Manga109 dataset.

A large-scale dataset of manga images is important and useful for manga research. Although several papers related to manga image processing have been published, fair comparisons among the proposed methods have not been conducted because of the lack of a dataset with which the methods can be evaluated. In preparing a dataset, the most serious and intractable problem is copyright. Because manga is artwork and protected by copyright, it is hard to construct a publicly available dataset. Therefore, researchers have collected their own manga images or used manga images drawn by amateurs. These are not

(a) Gakuen Noise, vol.1, p.117. ©Daisuke Inohara

(b) PLANET7, vol.2, p.36. ©Shuji Takeya

(c) Aosugiru Haru. p.26. ©Momoko Okuda

(d) Hanzai Kousyounin Minegishi Eitarou, vol.1, p.17. ©Takashi Ki

(e) OL Lunch, p.18. ©Youko Sanri

Fig. 11 Example images from the Manga109 dataset. Each caption shows the bibliographic information of an image (title, volume, and page)

appropriate for manga research because: (1) the number of manga images is small; (2) the artistic quality of the manga images is not always high; and (3) they are not publicly available. The Manga109 dataset, on the other hand, clears these three issues. It contains 21,142 pages in total, and we can say that it is currently the largest manga image dataset. The quality of the images is high because all titles were drawn by professional manga authors.

As is widely known, well-crafted image datasets have played critical roles in evolving image processing technologies, e.g., the PASCAL VOC datasets [17] for image recognition in the 2000s, and ImageNet [51] for recent rapid progress in deep architecture. We hope the Manga109 dataset will contribute to the further development of the manga research domain.

Image collection All manga titles in the Manga109 dataset have been previously published in manga magazines,[2] i.e., they were drawn by professional manga authors. The manga titles are in an archive "Manga Library Z" run by J-Comic Terrace. It has more than 3,000 titles, most of which are currently out of print. With the help of J-Comic Terrace, we chose 109 titles from the archive that cover a wide range of genres and publication years. We obtained permission from the creators to use them for academic research purposes. Thus researchers can use them freely with appropriate citation for their research challenges, including not only for retrieval and localization, but also for colorization, data mining from manga, and so on. The manga titles were originally published from the 1970s to the 2010s. The Manga109 dataset covers various kinds of categories, including humor, battle, romantic comedy, animal, science fiction, sports, historical drama, fantasy, love romance, suspense, horror, and four-frame cartoons. Please see the details (titles, author names, and so on) in our project page [32].

Dataset statistics The dataset consists of 109 manga titles. Each title includes 194 pages on average, with a total of 21,142 pages. The average size of images is 833×1179, which is bigger than the image sizes usually used for object recognition and retrieval tasks, e.g., 482×415 pixels on average for ILSVRC 2013.[3] This is because manga pages contain multiple frames and thus require higher resolutions.

Relation to other datasets *eBDtheque* [19] is a publicly available comic dataset. It consists of 100 comic pages with detailed meta-information such as text annotations. Compared with eBDtheque, our Manga109 dataset contains a much larger number of pages. Manga109 is useful for large-scale experiments, whereas eBDtheque is beneficial for small but detailed evaluations such as object boundary detection.

6 Experimental results

In this section, we present three evaluations: (1) a comparative study with previous methods for manga description (Section 6.1); (2) a localization evaluation (Section 6.2); and (3) a large-scale qualitative study (Section 6.3). In the comparative study and the localization evaluation, we used a single-thread implementation for a fair comparison, and employed a parallel implementation for the large-scale study.

---

[2]Japanese manga is usually published in a manga magazine first, which contains many manga titles. Then it comes out in an independent book form for each manga title.
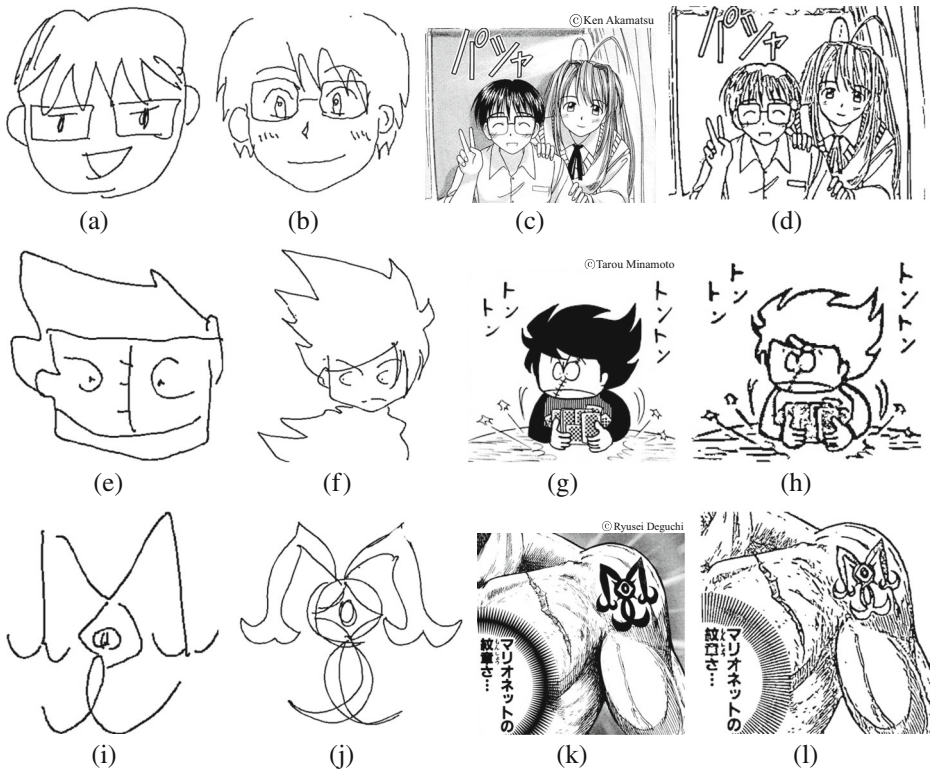
[3]http://www.image-net.org/challenges/LSVRC/2014/index#data

Fig. 12 Targets for the comparative study. Left to right: query sketches by novice artists, skilled artists, ground-truth images, and ground-truth images with screentone removal. Top to bottom: targets, *Boy-with-glasses*, *Chombo*, and *Tatoo*

6.1 Comparative study

A comparative study was performed to evaluate how well the proposed framework could represent manga images compared with previous methods such as those introduced in Section 2. We compared our proposal with a baseline (BoF with large window SIFT [54]), the state-of-the-art of BoF-based methods (FV [54]), and the state-of-the-art of chamfer-based methods (Compact OCM [62]). All experiments were conducted on a PC with a 2.8 GHz Intel Core i7 CPU and 32 GB RAM, using C++ implementations.

Frame image dataset For evaluation, we cropped frames from 10 representative manga titles from the Manga109 dataset.[4] There were 8,889 cropped frames, and the average size was 372 × 341. We used these frames for retrieval. Note that we used frames instead of pages for this comparison because the features of BoF, FV, and Compact OCM are only

---

[4]*Lovehina vol.1*, *Gakuen Noise*, *DollGun*, *Hanzai Kousyounin Minegishi Eitarou*, *Bakuretsu KungFu Girl*, *Aosugiru Haru*, *OL Lunch*, *Mukoukizu no Chombo*, *Highschool Kimengumi vol.1*, and *PLANET7*

comparable within a frame. Although the frame is less complex than the page, retrieval is not easy because the size of frames varies greatly, and the localization problem still exists as shown in Fig. 12c, where the target (a head of a boy) is small and the frame includes other objects and backgrounds.

Target images To make the comparisons, we chose three kinds of targets: *Boy-with-glasses* (Fig. 12c), *Chombo* (Fig. 12g), and *Tatoo* (Fig. 12k), which are from manga titles *Lovehina vol. 1*, *Mukoukizu no Chombo*, and *DollGun*, respectively. *Boy-with-glasses* is an easy example, *Chombo* is much harder because the face might be either to the right or left, and *Tatoo* is the most difficult because it is so small compared with the size of the frame. These target images are treated as ground truths.

Query images To prepare query sketches, we invited 10 participants (seven novices and three skilled artists, such as a member of an art club), and asked them to draw sketches. First, we showed them query images for 20 s. Next, they were asked to draw a sketch for the query. Each participant drew all target objects; therefore, we collected $10 \times 3 = 30$ queries. Examples of queries are shown in Fig. 12. We used a 21-inch pen display (WACOM DTZ-2100) for drawing.

Results of comparative study Using the queries, we evaluated each method using standard evaluation protocols in image retrieval: *recall@k* and *mean average precision* [4]. Note that the ground-truth target images were labelled manually. All frame images were used for the evaluation, and the images of the different targets were regarded as distractors. The statistics of the frame images are shown in Table 1.

Figure 13 shows the results for each target. Note that this task is challenging, and all scores tend to be small. The queries from users are not always similar to the target. On the contrary, some novices even drew queries that were dissimilar to the target, as shown in Fig. 12e. Still, in all cases, the proposed method achieved the best scores. In particular, the proposed method outperformed the other methods for *Boy-with-glasses*. Note that BoF and FV received almost zero scores for the *Tatoo* case because they are not good at finding a relatively small instance from an image. From these experiments, we can say that BoF-based methods do not satisfy the purpose in manga retrieval.

Note that we did not apply any approximation steps for a fair comparison. Compact OCM measures an original chamfer distance without an approximation (a sparse projection). We did not compress a feature using PQ in this experiment.

About quantization When we applied PQ to the features, the score decreased according to the quantization level, as shown in Fig. 14. There is a clear trade-off between the rate of compression and accuracy. From the result, we accepted $M = 16$ as a reasonable option; i.e., a feature is divided into 16 parts for PQ compression and encoded to a 16-byte code.

Table 1 Image statistics for comparative study. All images are cropped frames from the Manga109 dataset

target type	difficulty	#ground truth	#dataset	#query
Boy-with-glasses	easy	67	8563	30
Chombo	normal	178	8563	30
Tatoo	hard	81	8563	30

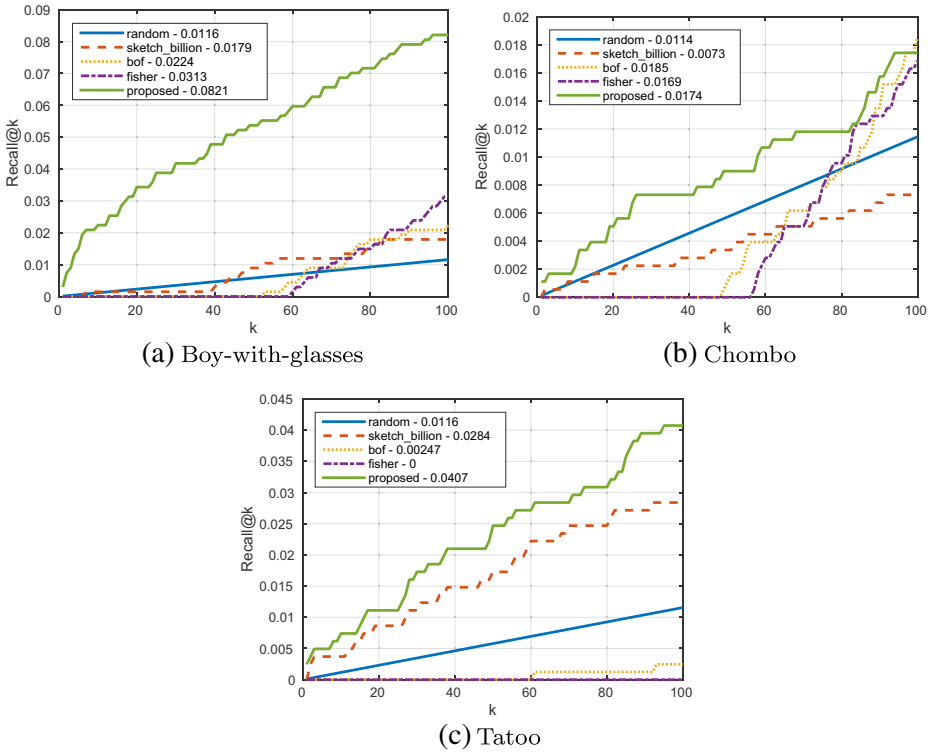(a) Boy-with-glasses

(b) Chombo

(c) Tatoo

Fig. 13 Results of the comparative study. Values in the legend show Recall@100

Note that, interestingly, there is not a clear relation between accuracy and the number of cells (c^2). The feature description becomes finer with a large number of cell divisions, but it does not always achieve higher recall. This indicates that some level of abstraction is required for the sketch retrieval task. We employed $c = 8$ for all experiments, i.e., a selected area is divided into 8×8 areas for feature description.
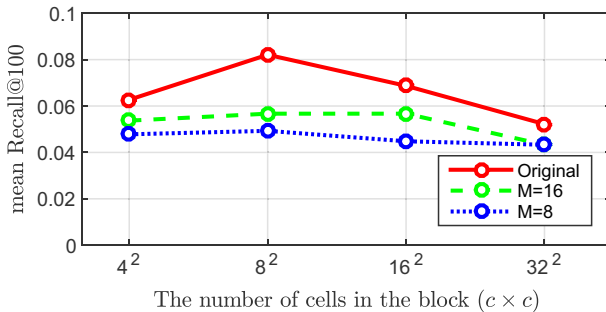

Fig. 14 The effect of feature compression by PQ. The Y-axis represents the retrieval performance. The X-axis shows the number of cells. Each line corresponds to a compression level. As the features are compressed by PQ (i.e., the feature is represented by a smaller number of subvectors (M)), the score decreases compared with the original uncompressed feature

Parameter settings and implementation details We show parameter settings and implementation details used for the evaluation. For BoF, SIFT features are densely extracted where the size of a SIFT patch is 64×64 pixels, with a 2×2 spatial pyramid, and the number of vectors in the dictionary is set to 1024. The final number of dimension is 4096. For FV, a Gaussian Mixture Model with 256 Gaussians was used, with a 2×2 spatial pyramid. The dimensions of SIFT are reduced from 128 to 80 by PCA. For BoF and FV, we leveraged the same parameter settings as in Schneider and colleagues [54], with the `vlfeat` implementation [67]. For selective search, we employed the `dlib` implementation [30], with little Gaussian blurring before applying selective search. To train code words for BoF, FV, and PQ, we randomly selected 500 images from the dataset. Note that they were excluded and not used for testing. To eliminate small patches, we set a minimum length of a patch as 100 pixels and discarded patches that were smaller than that.

6.2 Localization evaluation

Next, we evaluated how well the proposed method can localize a target object in manga pages. The setup is similar to that for image detection evaluation [17].

Images As query sketches, we used the Boy-with-glasses sketches collected in Section 6.1. We prepared two datasets for evaluation. (i) A Lovehina dataset, which is a title of *Lovehina vol.1*, and consists of 192 pages, including the pages containing ground-truth windows. (ii) The Manga109 dataset, which is the sum of all manga data, consists of 109 titles with a total of 21,142 pages. Note that the Lovehina data is included in the Manga109 dataset, so (i) is a subset of (ii). The ground-truth area (69 windows) were manually annotated from a Lovehina dataset.

In contrast to the previous comparative study, this is an object localization task, i.e., given a query image, find a target instance in an image. In our case, we must find the target from many manga pages (21,142 pages, for Manga109).

Evaluation criteria For evaluation, we employed a standard PASCAL overlap criterion [17]. Given a bounding box (retrieved result) from the method, it is judged to be true or false by measuring the overlap of the bounding box and ground-truth windows. Denote the predicted bounding box as B_p, and the ground-truth bounding box as B_{gt}, the overlap is measured by:

$$r = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}. \tag{5}$$

We judged the retrieved area is true if $r > 0.5$. If multiple bounding boxes are produced, at most one among them is counted as correct.

For each query, we find the top 100 areas using the proposed retrieval method from the dataset (Lovehina or Manga109), then the retrieved areas are judged using (5). Then we can compute the standard mAP@100 from the result (true/positive sequence).

Table 2 Results for localization evaluation (single thread implementation)

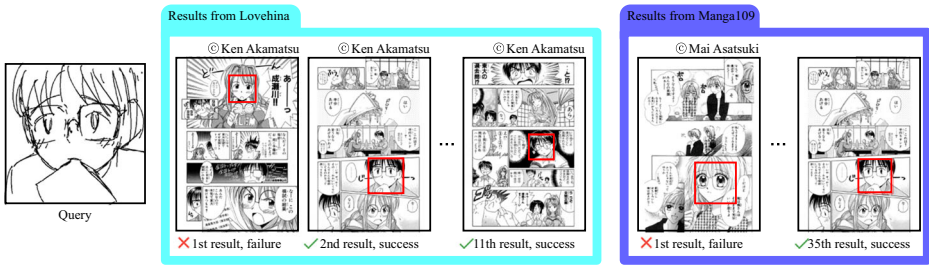|	#image	#patch	mAP@100	memory	runtime
Lovehina	192	138K	1.12×10^{-2}	2.21 MB	11.6 ms
Manga109	21,142	14M	1.43×10^{-4}	204 MB	331 ms

Fig. 15 Examples of localization experiments for the Lovehina dataset and Manga109 dataset

Result of localization evaluation We show the results in Table 2. With our single implementation, searching the Manga109 dataset (14M patches) took 331 ms. This is fast enough for interaction, and the computation cen be further improved (70 ms) using a parallel implementation as discussed in Section 6.3. We also show theoretical values of memory consumption for EOH features (#patch$\times 8M$). The whole Manga109 dataset consumes only 204 MB. As can be seen from mAP, this task is difficult because there are possibly hundreds of thousands of candidate areas in the windows (138K for Lovehina, and 14M for Manga109) for only 69 ground-truth areas. Examples of retrieved results are shown in Fig. 15. For the Lovehina data, the first result is a failure but the second is correct. For the Manga109 dataset, the first success can be found at the 35th result. The first result shares similar characteristics to the query (wearing glasses) even though the result is incorrect.

6.3 Large-scale qualitative study

In this section, we show a qualitative study of retrieval from the Manga109 dataset. The whole system was implemented using a GUI as shown in Fig. 2.

    We employed a parallel implementation using the Intel Thread Building Library, and the average computation time was 70 ms for the Manga109 dataset (21,142 images). The parallelization was straightforward. Neighbors were computed for each manga title in parallel. In the implementation, we selected the most similar feature from a page (not keeping all features per page), then merged the results.

Qualitative study using a sketch dataset We qualitatively evaluated the proposed method using a public sketch dataset as queries. We used representative sketches [15] as queries. The 347 sketches each had a category name, e.g., "panda."

    Figure 16a and b show successful examples. We could retrieve objects from the Manga109 dataset successfully. In particular, the retrieval works well if the target consists of simple geometric shapes such as squares, as shown in Fig. 16b. This tendency is the same as that for previous sketch-based image retrieval systems [16, 62]. Figure 16c shows a failure example, although the retrieved glass is similar to the query.

    As can be seen in Fig. 16c, text regions are sometimes retrieved and placed at the top of the ranking. Because users usually do not require such results, detecting and eliminating text areas would improve the results, which remains as a future work.
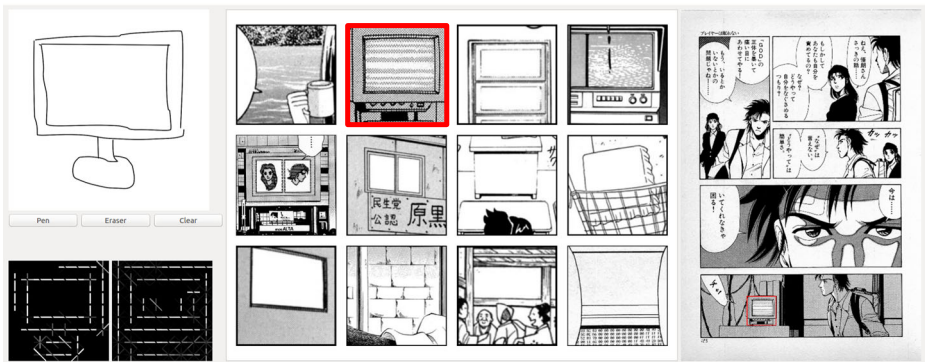
More results by relevance feedback We show more results from queries of character faces using the proposed relevance feedback in Fig. 17. We see that the top retrieved results were the same as (or similar to) the characters in the query. In this case, all the results were

© Hidehisa Masaki

(a) Trousers

© Tetsuya Kurosawa / Hidehisa Masaki

(b) Computer monitor

© Masako Yoshi

(c) Hourglass

Fig. 16 Results of the subjective study using representative sketches [15] as queries

drawn by the same author. Interestingly, our edge histogram feature captured the characteristic of authors. Figure 17b shows the results of "blush face." In Japanese manga, such blush faces are represented by hatching. By the relevance feedback, blushed characters

(a)

(b)

(c)

Fig. 17 More results for relevance feedback from Manga109 dataset

were retrieved from various kinds of manga titles. These character-based retrievals are made possible by content-based search. This suggests that the proposed query interactions are beneficial for manga search.

7 Conclusions

We proposed a sketch-based manga retrieval system and novel query schemes. The retrieval system consists of three steps: labeling of margin areas, EOH feature description with screen tone removal, and approximate nearest-neighbor search using product quantization. The query schemes include relevance feedback and query retouch, both of which are new schemes that interactively rerank the retrieved results. We built a new dataset of manga images, Manga109, which consists of 21,142 manga images drawn by 94 professional manga artists. To the best of our knowledge, Manga109 is currently the biggest manga image dataset. It is available to the research community. Experimental results showed that the proposed framework is efficient and scalable (70 ms from 21,142 pages using a single computer with 204 MB).

Combining the sketch- and keyword-based searches is a promising direction for future work.

References

1. Alexe B, Deselaers T, Ferrari V (2012) Measuring the objectness of image windows. IEEE TPAMI 34(11):2189–2202
2. Aramaki Y, Matsui Y, Yamasaki T, Aizawa K (2014) Interactive segmentation for manga. In: Proceedings SIGGRAPH Poster. ACM
3. Babenko A, Lempitsky V (2014) The inverted multi-index. IEEE TPAMI
4. Buttcher S, Clarke CLA, Cormack GV (2010) Information Retrieval: Implementing and Evaluating Search Engines The MIT Press
5. Cao Y, Chan AB, Lau RWH (2012) Automatic stylistic manga layout. In: Proceedings SIGGRAPH Asia. ACM
6. Cao Y, Lau RWH, Chan AB (2014) Look over here: Attention-directing composition of manga elements. In: Proceedings SIGGRAPH. ACM
7. Cao Y, Wang C, Li Z, Zhang L, Zhang L (2010) Spatial-bag-of-features. In: Proceedings CVPR. IEEE
8. Cao Y, Wang C, Zhang L, Zhang L (2011) Edgel index for large-scale sketch-based image search. In: Proceedings CVPR. IEEE
9. Chen T, Cheng MM, Tan P, Shamir A, Hu SM (2009) Sketch2photo: Internet image montage. In: Proceedings SIGGRAPH Asia. ACM
10. Cheng MM, Zhang Z, Lin WY, Torr P (2014) Bing: Binarized normed gradients for objectness estimation at 300fps. In: Proceedings CVPR. IEEE
11. Chu WT, Chao YC (2014) Line-based drawing style description for manga classification. In: Proceedings MM. ACM
12. Chum O, Philbin J, Sivic J, Isard M, Zisserman A (2007) Total recall: Automatic query expansion with a generative feature model for object retrieval. In: Proceedings ICCV. IEEE
13. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings CVPR. IEEE
14. Delalandre M, Iwata M, Kise K (2014) Fast and optimal binary template matching application to manga copyright protection. In: Proceedings DAS
15. Eitz M, Hays J, Alexa M (2012) How do humans sketch objects? In: Proceedings SIGGRAPH. ACM
16. Eitz M, Hildebrand K, Boubekeur T, Alexa M (2011) Sketch-based image retrieval: Benchmark and bag-of-features descriptors. IEEE TVCG 17(11):1624–1636

17. Everingham M, Gool LV, Williams CKI, Winn J, Zisserman A (2010) The pascal visual object classes (voc) challenge. IJCV 88(2):303–338
18. Ge T, He K, Ke Q, Sun J (2014) Optimized product quantization. IEEE TPAMI 36(4):744–755
19. Guérin C, Rigaud C, Mercier A, Ammar-Boudjelal F, Betet K, Bouju A, Burie JC, Louis G, Ogier JM, Revel A (2013) eBDtheque: A Representative Database of Comics. In: Proceedings ICDAR. IEEE
20. Hoashi K, Ono C, Ishii D, Watanabe H (2011) Automatic preview generation of comic episodes for digitized comic search. In: Proceedings MM. ACM
21. Hu R, Barnard M, Collomosse J (2010) Gradient field descriptor for sketch based retrieval and localization. In: Proceedings ICIP. IEEE
22. Hu R, Wang T, Collomosse J (2011) A bag-of-regions approach to sketch-based image retrieval. In: Proceedings ICIP. IEEE
23. Ito K, Matsui Y, Yamasaki T, Aizawa K (2015) Separation of line drawings and screentones of manga. In: Proceedings Eurographics
24. Jégou H, Douze M, Schmid C (2008) Hamming embedding and weak geometric consistency for large scale image search. In: Proceedings ECCV
25. Jégou H, Douze M, Schmid C (2009) Packing bag-of-features. In: Proceedings ICCV. IEEE
26. Jégou H, Douze M, Schmid C (2011) Product quantization for nearest neighbor search. IEEE TPAMI 33(1):117–128
27. Jégou H, Perronnin F, Douze M, Sánchez J, Pérez P, Schmid C (2012) Aggregating local image descriptors into compact codes. IEEE TPAMI 34(9):1704–1716
28. Kalantidis Y, Avrithis Y (2014) Locally optimized product quantization for approximate nearest neighbor search. In: Proceedings CVPR. IEEE
29. Kazi RH, Igarashi T, Zhao S, Davis RC (2012) Vignette: Interactive texture design and manipulation with freeform gestures for pen-and-ink illustration. In: Proceedings CHI. ACM
30. King DE (2009) Dlib-ml: A machine learning toolkit. J Mach Learn Res 10:1755–1758
31. Kopf J, Lischinski D (2012) Digital reconstruction of halftoned color comics. In: Proceedings SIG-GRAPH Asia. ACM
32. Lab AY (2015) Manga109 dataset. http://www.manga109.org
33. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings CVPR. IEEE
34. Le TN, Luqman MM, Burie JC, Ogier JM (2015) A comic retrieval system based on multilayer graph representation and graph mining. In: Graph-Based Representations in Pattern Recognition, LNCS, vol 9069, pp 355–364
35. Levi K, Weiss Y (2004) Learning object detection from a small number of examples: the importance of good features. In: Proceedings CVPR. IEEE
36. Lin G, Shen C, Shi Q, van den Hengel A, Suter D (2014) Fast supervised hashing with decision trees for high-dimensional data. In: Proceedings CVPR. IEEE
37. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. IJCV 60(2):91–110
38. Marchand-Mailet S, Sharaiha YM (1999) Binary Digital Image Processing: A Discrete Approach Academic Press
39. Matsui Y., Aizawa K., Jing Y. (2014) Sketch2manga: Sketch-based manga retrieval. In: Proceedings ICIP. IEEE
40. Matsui Y, Yamasaki T, Aizawa K (2011) Interactive manga retargeting. In: Proceedings SIGGRAPH Poster. ACM
41. Mei T, Rui Y, Li S, Tian Q (2014) Multimedia search reranking: A literature survey. ACM Comp Surv 46(3):38:1–38:38
42. Ojala T, inen MP, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE TPAMI 24(7):971–987
43. Olsen L, Samavati FF, Sousa MC, Jorge JA (2009) Sketch-based modeling: A survey. Comput Graph 33(1):85–103
44. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In: Proceedings CVPR. IEEE
45. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2008) Lost in quantization: Improving particular object retrieval in large scale image databases. In: Proceedings CVPR. IEEE
46. Qu Y, Pang WM, Wong TT, Heng PA (2008) Richness-preserving manga screening. In: Proceedings SIGGRAPH Asia. ACM
47. Qu Y, Wong TT, Heng PA (2006) Manga colorization. In: Proceedings SIGGRAPH. ACM
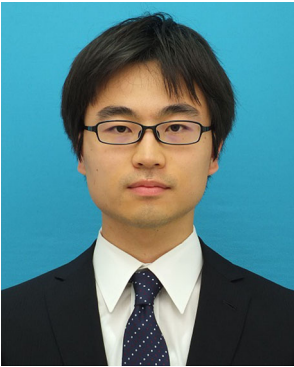
48. Ren S, Jin C, Sun C, Zhang Y (2014) Sketch-based image retrieval via adaptive weightinge. In: Proceedings ICMR. ACM

49. Rigaud C, Burie JC, Ogier JM, Karatzas D, de Weijer JV (2013) An active contour model for speech balloon detection in comics. In: Proceedings ICDAR. IEEE

50. Rigaud C, Guérin C, Karatzas D, Burie JC, Ogier JM (2015) Knowledge-driven understanding of images in comic books. IJDAR 18(3):199–221

51. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2014) Imagenet large scale visual recognition challenge. CoRR abs/1409.0575

52. Sànchez J, Perronnin F, Mensink T, Verbeek JJ (2013) Image classification with the fisher vector: Theory and practice. IJCV 105(3):222–245

53. Sato K, Matsui Y, Yamasaki T, Aizawa K (2014) Reference-based manga colorization by graph correspondence using quadratic programming. In: Proceedings SIGGRAPH Asia Technical Brief. ACM

54. Schneider RG, Tuytelaars T (2014) Sketch classification and classification-driven analysis using fisher vectors. In: Proceedings SIGGRAPH Asia. ACM

55. Serra J (1983) Image Analysis and Mathematical Morphology Academic Press

56. Shotton J, Blake A, Cipolla R (2008) Multi-scale categorical object recognition using contour fragments. IEEE TPAMI 30(7):1270–1281

57. Sivic J, Zisserman A (2003) Video google: A text retrieval approach to object matching in videos. In: Proceedings ICCV. IEEE

58. Smeulders AW, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE TPAMI 22(12):1349–1380

59. Spyromitros-Xioufis E, Papadopoulos S, Kompatsiaris IY, Tsoumakas G, Vlahavas I (2014) A comprehensive study over vlad and product quantization in large-scale image retrieval. IEEE TMM 16(6):1713–1728

60. Sun W, Burie JC, Ogier JM, Kise K (2013) Specific comic character detection using local feature matching. In: Proceedings ICDAR. IEEE

61. Sun W, Kise K (2013) Detection of exact and similar partial copies for copyright protection of manga. IJDAR 16(4):331–349

62. Sun X, Wang C, Xu C, Zhang L (2013) Indexing billions of images for sketch-based retrieval. In: Proceedings MM. ACM

63. Sýkora D, Dingliana J, Collins S (2009) Lazybrush: Flexible painting tool for hand-drawn cartoons. In: Proceedings Eurographics

64. Tanaka T, Shoji K, Toyama F, Miyamichi J (2007) Layout analysis of tree-structured scene frames in comic images. In: Proceedings IJCAI

65. Tseng KY, Lin YL, Chen YH, Hsu WH (2012) Sketch-based image retrieval on mobile devices using compact hash bits. In: Proceedings MM. ACM

66. Uijlings JRR, van de Sande KEA, Gevers T, Smeulders AWM (2013) Selective search for object recognition. IJCV 104(2):154–171

67. Vedaldi A, Fulkerson B (2010) Vlfeat: An open and portable library of computer vision algorithms. In: Proceedings MM

68. Wang J, Liu W, Kumar S, Chang SF (2016) Learning to hash for indexing big data - a survey. Proc IEEE 104(1):34–57

69. Wang J, Zhang T, Song J, Sebe N, Shen HT (2016) A survey on learning to hash. CoRR abs/1606.00185

70. Wang X, Bai X, Ma T, Liu W, Latecki LJ (2012) Fan shape model for object detection. In: proceedings CVPR. IEEE

71. Wu Z, Ke Q, Isard M, Sun J (2009) Bundling features for large scale partial-duplicate web image search. In: proceedings CVPR. IEEE

72. Yu Q, Yang Y, Song YZ, Xiang T, Hospedales TM (2015) Sketch-a-net that beats humans. In: proceedings BMVC

73. Zhang X, Liu W, Dundar M, Badve S, Zhang S (2015) Towards large-scale histopathological image analysis: Hashing-based image retrieval. IEEE TMI 34(2):496–506

74. Zhang Y, Jia Z, Chen T (2011) Image retrieval with geometry-preserving visual phrases. In: proceedings CVPR. IEEE

75. Zhou R, Chen L, Zhang L (2012) Sketch-based image retrieval on a large scake database. In: proceedings MM. ACM

Yusuke Matsui received the B.E. degree in information and communication engineering in 2011, and the M.A. degree in interdisciplinary information studies in 2013, from the University of Tokyo, Japan. He is currently working towards the Ph.D. degree at the University of Tokyo. He worked as a research intern at Microsoft Research Asia in 2013, and at Microsoft Research Redmond in 2014. His research interests include computer vision and computer graphics, particularly comic image processing and approximate nearest neighbor search.

Kota Ito received the B.E. degree in information and communication engineering in 2014 from the University of Tokyo. He is currently working towards M.E. degree at the University of Tokyo. His research interests include computer vision and applications.

Yuji Aramaki received the B.S. degree in information and communication engineering in 2014, and the M.S. degree in interdisciplinary information studies in 2016, from the University of Tokyo. His research topics were in computer vision and multimedia processing, particularly manga processing.
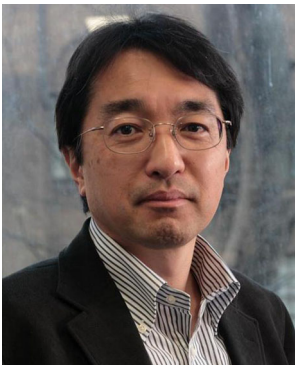

Azuma Fujimoto received the B.E. degree in information and communication engineering in 2015 from the University of Tokyo. He is currently working towards M.A. degree at the University of Tokyo. His research interests include computer vision and image processing.

Toru Ogawa received the B.E. degree in information and communication engineering in 2016 from the University of Tokyo. He is currently working towards M.E. degree at the University of Tokyo. His research interests include computer vision and applications.

Toshihiko Yamasaki received the B.S. degree in electronic engineering, the M.S. degree in information and communication engineering, and the Ph.D. degree from The University of Tokyo in 1999, 2001, and 2004, respectively. From April 2004 to Oct. 2006, he was an Assistant Professor at Department of Frontier Informatics, Graduate School of Frontier Sciences, The University of Tokyo. He is currently an Associate Professor at Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo. He was a JSPS Fellow for Research Abroad and a visiting scientist at Cornell University from Feb. 2011 to Feb. 2013. His current research interests include multimedia big data analysis, pattern recognition, machine learning, and so on. His publication includes three book chapters, more than 50 journal papers, more than 150 international conference papers, more than 420 domestic conference papers. He has received around 30 awards including Geomm 2013 Best Paper Award, ICMLA 2012 Best Special Session Paper Award, IEICE Young Researcher Award, so on. Dr. Yamasaki is a member of IEEE, ACM, IEICE, ITE, IPSJ.

Kiyoharu Aizawa received the B.E., the M.E., and the Dr.Eng. degrees in Electrical Engineering all from the University of Tokyo, in 1983, 1985, 1988, respectively. He is currently a Professor at the Department of Information and Communication Engineering of the University of Tokyo. He was a Visiting Assistant Professor at University of Illinois from 1990 to 1992. His current research interest is in image processing and multimedia applications, and he is currently engaged in multimedia life log and 3D video etc. He received the 1987 Young Engineer Award and the 1990, 1998 Best Paper Awards, the 1991 Achievement Award, 1999 Electronics Society Award from IEICE Japan, and the 1998 Fujio Frontier Award, the 2002 and 2009 Best Paper Award from ITE Japan. He received the IBM Japan Science Prize in 2002. He is currently a Senior Associate Editor of IEEE Trans. Image Processing and on Editorial Board of ACM TOMCCAP, Journal of Visual Communications and Image Processing and APSIPA Transactions on Signal and Information Processing. He served as the Editor in Chief of Journal of ITE Japan, an Associate Editor of IEEE Trans. Image Processing, IEEE Trans. CSVT and IEEE Trans. Multimedia. He has served a number of international and domestic conferences; he was a General co-Chair of ACM Multimedia 2012 and IEEE VCIP2012.