Journal of Communication and Computer 9 (2012) 414-421

Comic Image Category Classification Using SIFT Features

Yusuke In, Nakamura Kentaro, Masakazu Higuchi, Jonah Gamba, Atushi Koike and Hitomi Murakami
Seikei University, 3-3-1 Kitijyoujikitamati, musashino, Tokyo, Japan

Received: October 25, 2011 / Accepted: December 14, 2011 / Published: April 30, 2012.

Abstract: The recent development of the information society leads to many multimedia data on the Web. Especially in Japan, there are a lot of comic images on the Web. However, such a phenomenon causes copyright problems. In order to develop useful information society, the need for comic managing system identifying the author of comic images and ignoring the copied images is increasing. Such a system is indispensable for our future information society. In this paper, we propose a system for identifying the image of comics, and evaluate its performance against conventional methods. Furthermore, we examine the performance with subjective evaluation result by specialists of cartoonist.

Key words: Comic image, similarity-based image retrieval, image content, image local features.

1. Introduction

Due to the recent development of the information society, there are many multimedia data on the Web. Especially in Japan, there are various kinds of comic images on the Web, but effective methods of searching for category classification are not available. The need for a system that stores and manages image content will clearly increase in the future. As one approach to address this issue, there has not been any effective method presented for searching comic images when category classification is performed.

With comic images, even for the same work, various designs exist and are most the time different from the actual scene. For example, objects represented by monochrome pictures, and also for the frontal and lateral views for the face of the same character, the size of the eyes and the position and shape of the mouth changes significantly. In fact, even for the same work, there is no guarantee that the design would be the same. A typical example of a comic image is shown in Fig. 1. One character is represented in a various ways.

Fig. 1 Image examples of comic.

Nakamura Kentaro, undergraduate student, research field: image signal processing.
Masakazu Higuchi, Ph.D., post doctoral, researcher, research field: signal processing.
Jonah Gamba, Ph.D., post doctoral, researcher, research field: signal processing
Atsushi Koike, Ph.D., professor, research field: signal processing
Hitomi Murakami, Ph.D., professor, research field: signal processing.
Corresponding author: Yusuke In, graduate student, research fields: image signal processing, smart phone contents. E-mail: yuhsuke.in@gmail.com.

Comic image detection and classification are basically achieved by processing on a page by page basis. However, for page based processing, each frame is displayed in small scale which leads to an expected decrease in the correct image detection and classification characteristic. We therefore, considered a high speed frame-based method to detect and classify human images and evaluated its characteristics. In this paper, for the purpose of performing frame-based detection and classification human images, we first describe the frame decomposition method in Section 2.

Next, we performed experiments to determine the degree to which the proposed method improved the detection rate and processing speed when compared to the page-based method. Thesis described in Section 3. Additionally, in Section 4 with the results, the degree to which the detection and classification of human images was successful when compared to the manual results by semi-professional cartoonist was confirmed by subjective evaluation.

2. Classification of Comic Image Categories

Generally, the following two methods are used for searching a desired image from a large collection of images.

- Text-Based Image Retrieval (TBIR);
- Context-Based Image Retrieval (CBIR).

In the case of TBIR, it is necessary to attach keyword information to every image. However, because it is difficult to represent the respective images by keyword information only, most of the time good search results cannot be obtained. On the other hand, the CBIR uses image features to search the image, thus, it is possible to quantify the level of similarity. In this paper, we investigate the classification of comic image categories by the CBIR.

The bag-of-keypoints method for distinguishing images that was proposed by G. Csurka et al. is employed [1]. We investigated and compared the page-based bag-of-keypoints with the frame-based image decomposition method [2-3]. The flowcharts for the classification of comic image categories are shown in Figs. 2 and 3.

2.1 Frame Decomposition Method

With frame decomposition, after detecting the region that contains information, line detection is performed and image is recursively bisected. The details of the method are given below.

The region in which the comic image information exists is detected by contour scanning. In order to clarify the outline of the closed region obtained, weighting is applied to the pixels using Eq. (1).

$$C(x,y)=\begin{cases} a_0 : \text{Contour} \quad \text{pixel} \\ a_1 : \text{Anyother} \quad \text{pixel} \end{cases} \tag{1}$$

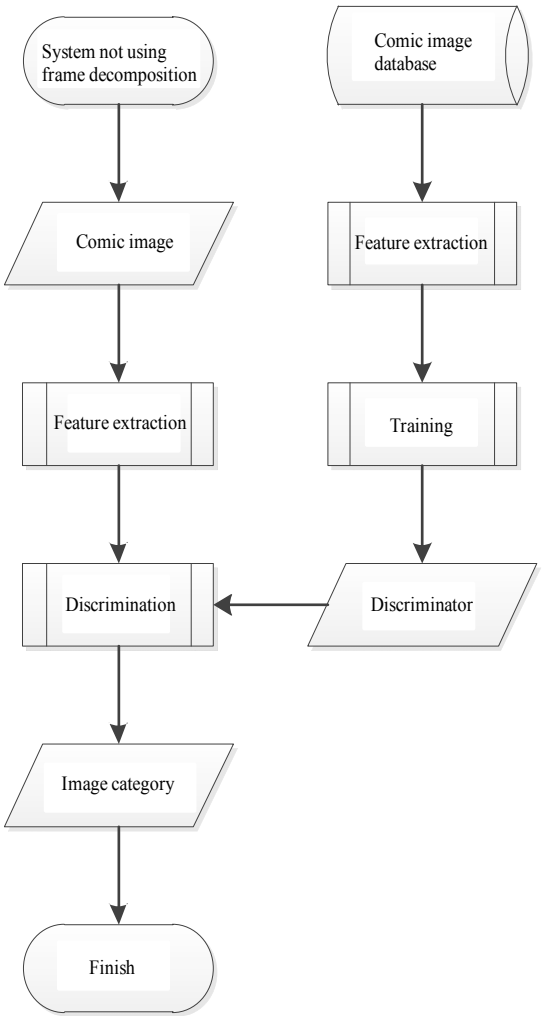


Fig. 2 Page-based classification of comic image categories.

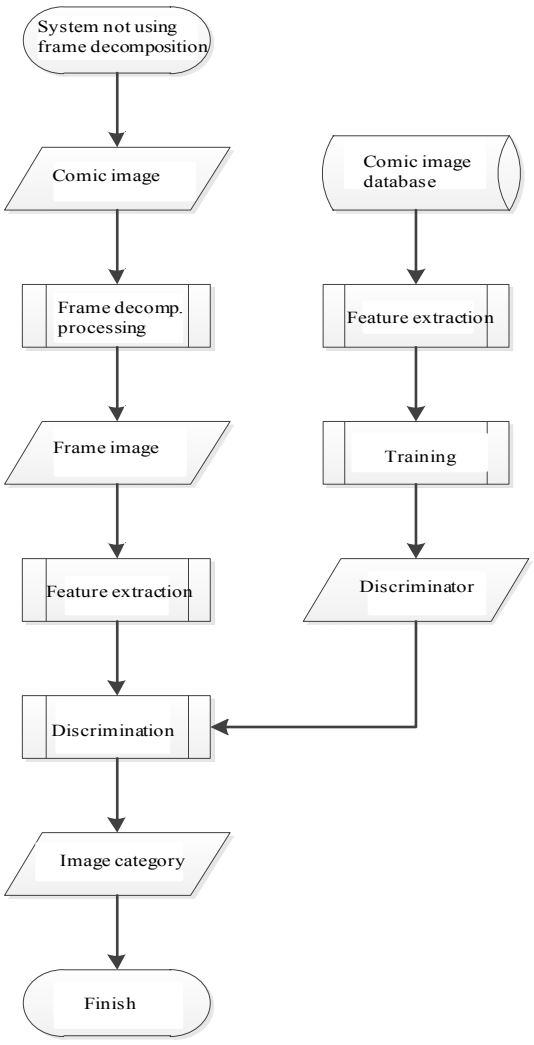


Fig. 3 Frame-based classification of comic image categories.

The parameters (x, y) of above equation represent the coordinates of the pixel on the image. In addition, empirically determined values, $a_0 = 1$ and $a_1 = 0.75$ are used. By this approach, it would be easy to detect regions where image information exists and regions which make up the margin when the line detection for frame decomposition is performed. In this paper, contour scan refers to the operation of raster scanning followed by contour extraction.

We explain the process of using division lines to perform region decomposition. As a property of comics, every frame is bounded by lines. For this reason, in order to detect lines that are almost straight, we use a morphological operator for line thinning and then use the Hough transform [4] for line detection. A point in the Hough space, (ρ, θ), is found and Eq. (2) is used to express the equation of the straight line.

$$\rho = x \cos \theta + y \sin \theta \quad (2)$$

In the above equation, ρ is the length of the straight line from the origin and θ is the angle of the line with the x-axis.

For the detected lines, all lines with θ less than or equal to K° and the difference in ρ within M are treated as one line. Here, we use the settings K° = 1 and M = 10. In this way, extremely narrow blank margins regions between the frames are ignored. The resulting straight line image pixels are denoted by L(ρ, θ). Additionally, by using pixels L(ρ, θ) and the neighboring pixels denoted by M(ρ,θ), the intensity gradient at pixel (x, y) is calculated by Eq. (3) as follows:

$$g_\theta(x,y) = g_x(x,y)\cos\theta + g_y(x,y)\sin\theta \quad (3)$$

$g_x(x,y)$: Horizontal intensity gradient

$g_y(x,y)$: Vertical intensity gradient

There are cases in which part of the picture and the Serif extends outside the frame into the blank margin and cases where the shape of the frame is not polygonal. In these cases, frame division lines cannot be detected well. To deal with these cases, we consider L(ρ, θ)'s neighboring pixels, M(ρ, θ).

Based on our much experience, when frame decomposition is performed, the center of the image is likely to have straight lines. For this reason, the Gauss function given by Eq. (4) and shown in Fig. 4 is used as weighting for images with straight lines.

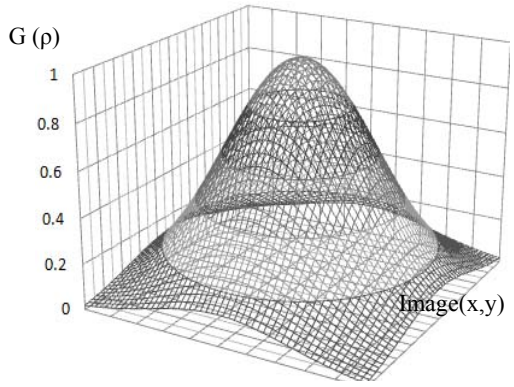


Fig. 4 Gauss function used for weighting the image.

$$G(\rho) = \exp\left(\frac{-\rho^2}{\sigma^2}\right) \quad (4)$$

σ: The height of the image divided by four.

By this weighting function, lines that are close to the center of the image are assigned big weights, while those away for the center of the image are assigned small weights.

Using Eqs. (1), (2), (3), and (4), the values A(ρ, θ) determined by the line division process for straight lines L(ρ, θ) are given by Eq. (5).

$$A(\rho,\theta) = G(\rho) \sum_{(x,y)\in M(\rho,\theta)} \{g_\theta(x,y)C(x,y)\} \quad (5)$$

When θ is between 45 degrees and 135 degrees, the line is a candidate horizontal division line. If the above condition is not satisfied, then the line is taken to be a candidate vertical division line. In addition, since straight lines along the frame are multiplied by a large weighting C(x, y), the straight line around the center of the frame are detected as candidate frame division lines. By repeating this process, high precision frame decomposition is possible.

In reality, frame decomposition is performed as follows. First, with the whole image as input, among the candidate division lines detected by the above process, the one with the largest value of A(ρ, θ) is used to divide the image. At this point in time, whether or not the line encloses the entire image into a rectangular region is used to decide on division or non-division and the image is then bisected if division is decided. Otherwise, the process is terminated without division. For the regions that have been divided, the information bearing regions are once again enclosed in a rectangle and the process mentioned above is recursively applied to determine the dividing lines and then proceed to bisecting the comic image.

Fig. 5 shows the result of frame decomposition of the comic image in Fig. 1. The areas enclosed by the rectangular black frames show each of the frame regions. By eliminating the margins, it can be confirmed that extraction of the information carrying regions is possible.

The frame decomposition flowchart is shown in Fig. 6.

2.2 Feature Extraction Method

The-bag-of-keypoints model is used. By this method, ignoring the location information inside the image and collecting the set of local features, categorization of the image is realized. Here feature extraction is done by SIFT [5-6]. A summary of feature extraction by SIFT is given below.

- Determine the scale by maximizing the Gaussian difference;
- Set the major gradient direction as the local direction;
- Compute the gradient direction histogram as a 128-dimension vector ($4 \times 4 \times 8$);
- For robustness to change normalize descriptor.

These feature quantities are relatively robust to changes in scale, so similarities can be stably detected.

Fig. 5 The result of frame decomposition of the comic image in Fig. 1.

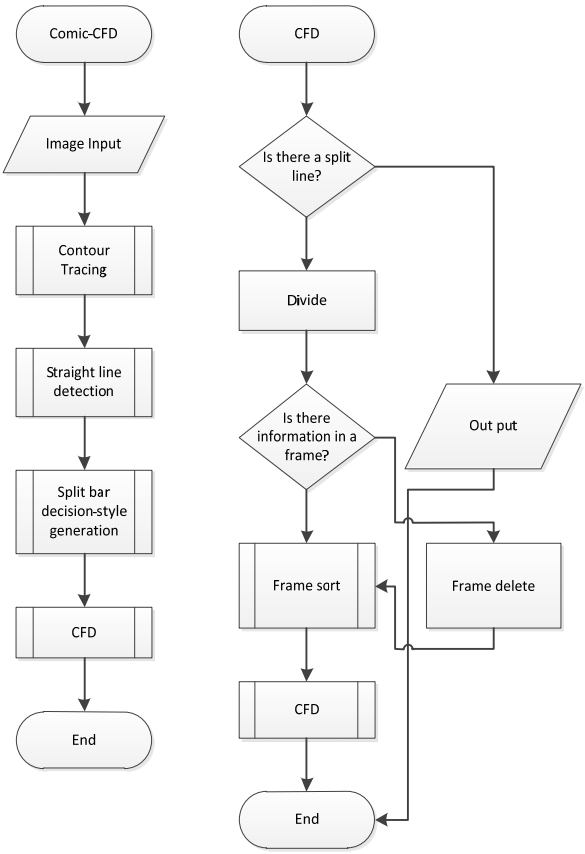


Fig. 6 The frame decomposition flowchart.

For quantization, the k-means is used, then feature extraction is performed by SIFT which employs clustering, followed by SIFT feature quantity description, codebook generation, class association, a series model for the feature quantity is generated. Fig. 7 shows the generation flow.

SIFT is divided into the feature detection part and the part for feature quantity description of points in the vicinity of the feature points.

The feature point detection part uses DOG to detect multiple points at which the change in shading gradient is large using a group of multiscale images obtained by reducing the size of the original image by reducing through several stages. The feature quantity description part further divides the local regions in the vicinity of the feature points in 4×4 regions, and expressed as 128-dimension vector for the purpose of making a histogram for each of the 8 directions.

From the vector obtained by SIFT feature detection, clustering k-means was performed, and then a descriptor, called the codebook, was generated. For each cluster, a histogram was taken according the occurrence rate. In addition, since the number of feature points extracted varies with the image, the occurrence rate was normalized by the number of feature points. The codebook was represented as a k-dimension histogram and here k = 1,000 were used.

2.3 Discrimination Method

Based on the codebook, the feature vectors are extracted from the training image and then trained by classifiers. Similarly, the feature vectors are extracted from the categorization image. The classifier then determines the category from the values obtained above. The multi-class SVM [7-9] classifier is used.

3. Results and Evaluation of the Proposed Method

3.1 Experiment Database

In the database for the experiment, 942 pages drawn from 10 different comics that are currently widely read in the Japanese comic market were used. For the comics, using the frame decomposition method described above, 4,145 image frames were obtained.

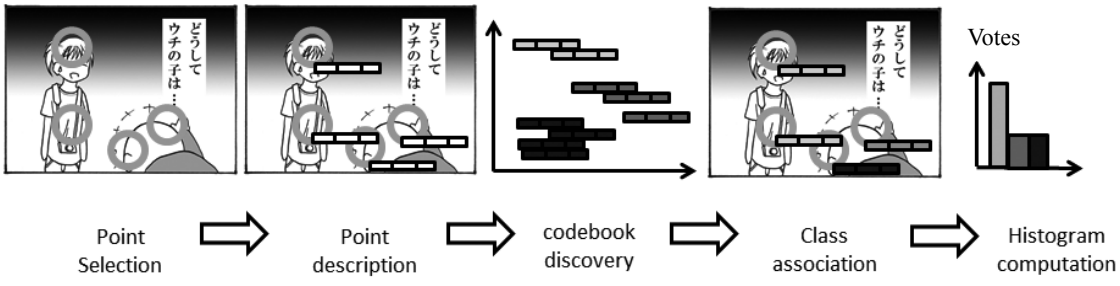


Fig. 7 Bag-of-keypoints model flow.

At this point, it is confirmed that 100% of frame decomposition is correctly performed.

3.2 Evaluation Method

We investigated the precision of category classification of comic images. The experimental dataset was divided into groups of 10, and taking each group as training data, the classifier was made to train the data.

For evaluation, we used Eq. (6) for Precision, Eq. (7) for Recall and Eq. (8) for the F-measure.

$$Precision = \frac{Correctly\ classified\ datasets}{Classified\ datasets} \quad (6)$$

$$Recall = \frac{Correctly\ classified\ datasets}{Datasets\ that\ ought\ to\ be\ classified} \quad (7)$$

$$\mathrm{F-measure} = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \quad (8)$$

3.3 Experimental Results

Fig. 8 shows category classification success rate for the comic images.

3.4 Discussion

In the case of per page basis, the F-measure has a highest value of 31% and an average classification rate of 19%. The frame decomposition method improved the F-measure's highest value to 50% and the average classification rate to 30%. In the case of frame decomposition, an improvement in the classification rate of 10% to 20% was achieved. The reason for the difference could be that for the per page basis, quantization error on visual words generation, errors in the extraction of feature points, etc., occurred.

As for computation speed, the frame decomposition method was about 25% faster than the per page basis method. The increase in processing steps for the frame decomposition method led to the decrease in input image size, which resulted in a faster feature extraction process.

4. Comparison with Subjective Evaluation Results by Semi-Professionals

To evaluate the degree to which the proposed frame decomposition method compares to classification by human beings, members of Seikei University's Comic Research group performed classification. The evaluators were experts who had sufficient understanding and experience in comics. For this reason, it was expected that the results obtained by these human evaluators would be the best result. For the experiment, from the results obtained by per page and frame decomposition, three were chosen from each and 10 people were asked to perform evaluations.

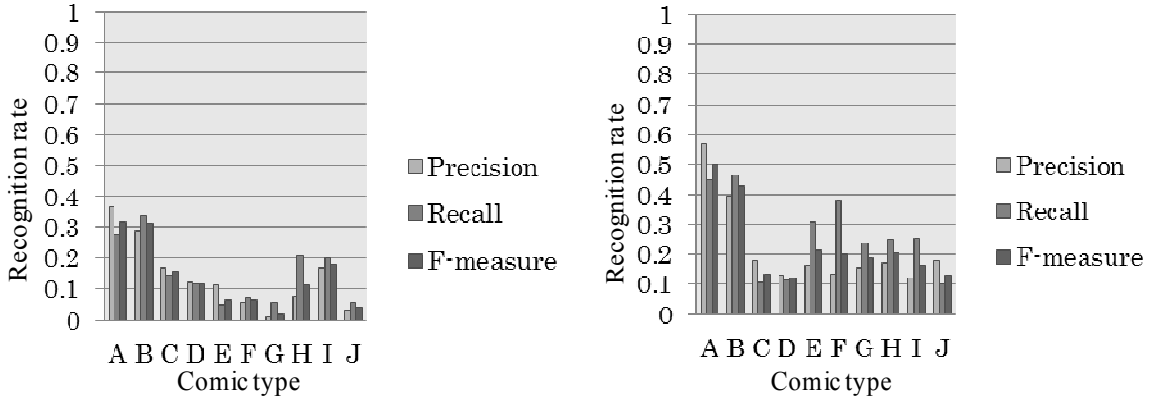
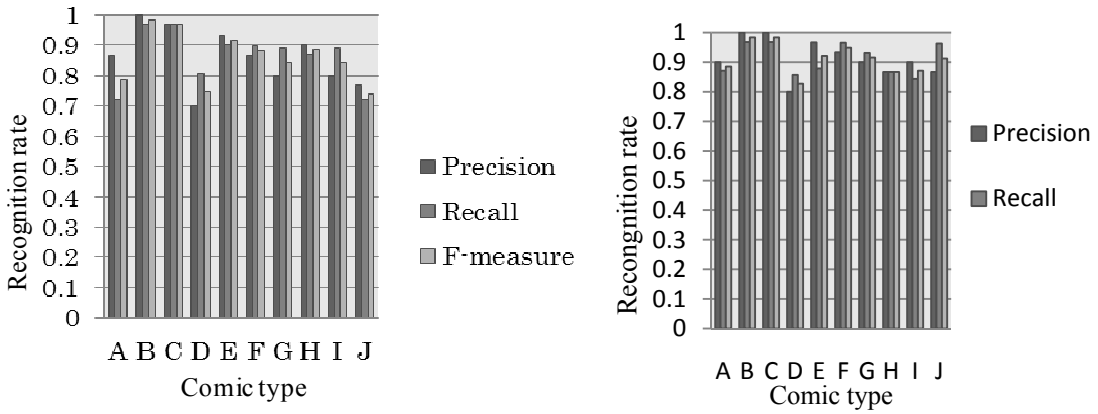


Fig. 8 Success rate of category classification by the per page and frame decomposition methods. (a) Per page basis category classification success rate. (b) Frame decomposition category classification success rate.



(a) (b)

Fig. 9 Subjective evaluation of classification success rate. (a) Subjective evaluation of classification success rate of the per page basis method. (b) Subjective evaluation of classification success rate of the frame decomposition method.

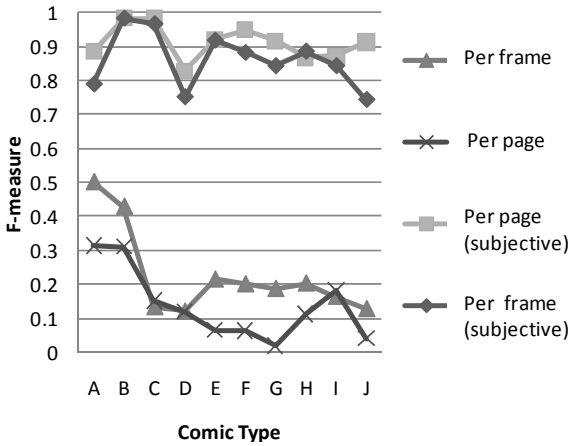


Fig. 10 Comparison of proposed methods with subjective evaluation results.

4.1 Experimental Results

Figs. 9 and 10 show the subjective evaluation results for category classification by each method.

4.2 Discussion

For the images used, since the evaluation was carried out by semi-professionals, there was no big difference between the per page basis and frame decomposition with success rate above 80% being obtained for both methods. This was expected because human beings perform classification by understanding the contents of the comics. Specifically, it is not just the entire picture but text in speech balloons and portraits, the evaluators can understand the content [10-11].

Additionally, for the F-measure value, there was a grin difference of about 60% when compared to the proposed method.

5. Conclusions

In this paper, we proposed a system of comic classification based on frame decomposition using database images and also performed experiments on comic classification.

In the experiments, 942 pages drawn from 10 different comics were used as the data set and category classification was performed. Using the comic page as it is gave a classification rate of 19% while the proposed method was able to increase the classification rate to 30%.

Compared to subjective classification by semi-professionals, the F-measure was 60% lower on the average. Against subjective evaluation by semi-professionals, a grim difference exists but for comic images, the effectiveness of frame decomposition has been confirmed.

A part of this work was supported by MEXT Grant-in-Aid for Building Strategic Research Infrastructures. We are very grateful for their support.

References

[1] G. Csurka, C. Bray, C. Dance, L. Fan, Visual categorization with bags of keypoints, Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision, 2004, p. 122.

[2] Y. In, A. Takahashi, K. Otsuka, K. Hirano, M. Higuchi, S. Kawasaki, A. Koike, H. Murakami, Fast frame decomposition and sorting by contour tracing for comic images, ITE Technical Report 34 (2010) 73-76.

[3] Yusuke In, et al., Using fast frame decomposition and sorting by contour tracing mobile phone comic imaging system, International Journal of Systems Applications, Engineering & Development 5 (2010) 216-223.

[4] A. Leandro, F. Fernandes, M.M. Oliveira, Real-time line detection through an improved Hough transform voting scheme, Pattern Recognition (PR), Elsevier, 2008. pp. 299-314.

[5] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91-110.

[6] A. Vedaldi, available online at: http://vision.ucla.edu/~vedaldi/code/siftpp/siftpp.html.

[7] T. Joachims, SVM multiclass, available online at: http:// www.cs.cornell.edu/People/tj/svmlight/svmmulticlass.html.

[8] H.Burdick, Digital Imaging Theory and Application, McGraw-Hill, New York, 1997.

[9] T. Joachims, SVM light Support Vector Machine, available online at: http://www.cs.cornell.edu/People/tj/svm%5Flight/.

[10] Yusuke In, et al, Similarity Detection of Comic Images: An Application of Image Local Features for Decomposed Comic Images, ITE Annual Convention Report, 2011, pp. 10-2.

[11] Yusuke In, et al, Comic image category classification using local features, in: Proceeding of the 2nd International Conference on Circuits, Systems, Control, Signals, 2011, pp. 55-60.