MULTI SIZE EYE DETECTION ON DIGITIZED COMIC IMAGE

Daisuke Ishii, Taichi Yamazaki and Hiroshi Watanabe

Graduate School of Global Information and Telecommunication Studies, WASEDA University

ABSTRACT

Metadata extraction on Comic: Manga, Japanese Style Comic is a key technology for comic distribution systems and user centric comic viewers. One of the most important metadata for that is a comic character's face position. However character's face are often deformed and occluded fatally. Thus, to become efficient face detection, eye position detection is focused on. Generally, Comic image is drawn by monochrome, and eye size has many variations. In this paper, we utilized reformed HOG feature to treat Manga image and to be able to handle size variation. HOG feature on digitized comic image is affected by image resolution condition. In our approach, the resolution of calculating HOG feature is fixed, and block size is determined according to input image. Thus, even if input image has resolution variation, HOG feature space is stable. Finally, the resolution fixed approach is compared to image resolution converting approach by experiment.

1. INTRODUCTION

Comic: Manga, Japanese Style Comic is now popularly diffused in not only japan but also many countries. In particular, digital distribution of comic has developed into a huge business in Japan.

Comic has story like novel, and that is consisted by text and picture elements: characters, backgrounds, visual effects, drawn texts, and etc. The picture of comic is depicted using black ink and dotted pattern called "Screentone" mainly.

We assume some functions: scene linking, character search, story summarization, auto translation and etc., as user centric application's features. For making user centric comic viewer, digital archive and etc., analyzing comics and utilizing metadata are needed.

For to analyze comic, we need to get much information about time line, characters, scripts, onomatopoeias, background contents and etc. Now, time line on one page can divide utilizing frame or structure analysis. Scripts and texts can be analyzed utilizing OCR technology, when text areas are successfully separated from image. Thus, now we aim to detect character presence and position.

In Comic, further elements are useful information as metadata, especially in that character position data is essential of automated story summarization, scene analysis, and so on.

General human detection method or image analysis

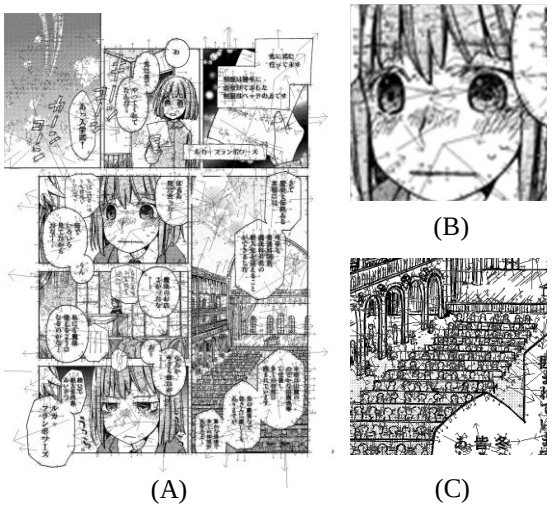
Fig. 1 Sample of comic images, that including text and picture elements: character, background, visual effect, drawn text. In particular eye area is drawn characteristic illustration.

method are hard to detect character in the comic, it because comic has different feature as compared to general movie or pictures i.e. the comic contents was depicted by black and white picture that have like frequent. For instance, SIFT feature [1] will be found whole area of the drawing. Furthermore, eye and many of other area has like frequent of black pixel. Thus, Haar-like feature [2] detector will return similar response.

The objective of this research is to detect eye position that is clue to the implicit presence of character. As described previously, comic image has characteristic feature not shared by natural picture. So, we attempt to utilize edge direction.

In this paper, to analyze comic image and detect eye position, we utilize an approach that is a variation of Histograms of Oriented Gradient (HOG) [3] feature. In that approach, block and cell number is fixed when calculating HOG feature. Then this method is compared with resolution conversion type HOG feature calculating approach. Those methods can detect eye position of comic character, where eye size is varied.

This paper is consisted as follows: Comic feature and calculated result of SIFT feature on comic image and conventional eye detection method that attends image rescaling were described in section 2. Eye detection method that utilized resolution fixed HOG method is given in section 3. In section 4, eye detection examination on digitized comic is demonstrated and



(A) (B) (C)

Fig. 2 (A) is a sample of SIFT feature on the comic. SIFT feature is depicted as arrow. (B) and (C) are expanded version of parts of (A.) SIFT feature was obtained beside of edges. However, as shown in figure, SIFT feature are obtained too many number to treat it.

consideration is described. Section 5 concludes this paper.

2. COMIC FEATURE AND EYE DETECTION

2.1. Comic image

Comic image is consisted by the black ink and screentone on white canvas mainly. A sample of comic image is shown in Fig. 1. In the comic image, some area's appearance frequency of black pixel is similar to each other, and the number of corner is too much. Therefore, Haar-like and SIFT feature are not usable to detect or represent the feature of comic image.

A calculated result of SIFT feature on the comic image is depicted in Fig. 2. In this result, SIFT feature is obtained too many to treat it, and some area like Fig.2 (C) is very complex.

2.2. Conventional Eye Detection Method for Comic

Eye region has less variation compared with face region. On the other hand, eye region size has large variation. Thus, to detect eye region, image resolution conversion is needed.

We have been proposed eye detection method utilized hog feature descriptor [4]. In that method, to conform the dimension of hog feature vector between whole learn and predict images, image resolution is normalized to given value.

Step of further method is as follows. First, training image set (eye image) is converted into normalized

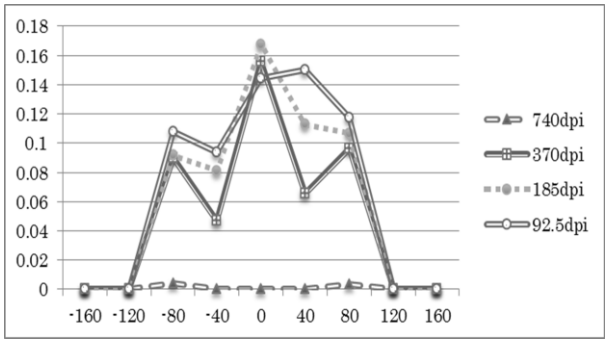


Fig. 3 Distribution of gradient function (page of comics), the trend of distribution of each resolution are different. Vertical value means appearance ratio and horizontal value means the quantized angles of gradient.

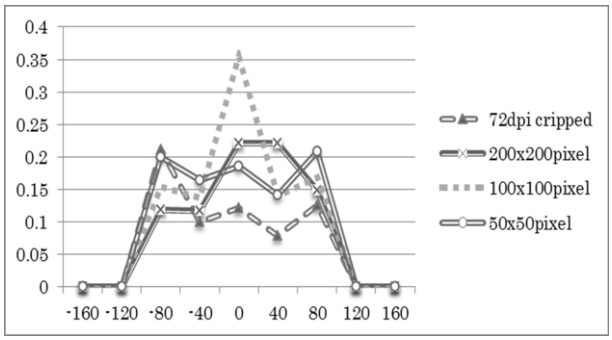


Fig. 4 Distribution of gradient direction (Face area). Vertical value means appearance ratio and horizontal value means the quantized angles of gradient. The distribution trend of each resolution is different too.

resolution. Next whole HOG feature is calculated on the same condition and learned by Support Vector Machine: SVM. Then, analyzed image is extracted by sliding window from comic image. Where, window size is e.g. 20x20, 40x40, 80x80 and 160x160 pixels respectively. Each analyzed image resolution is normalized and calculated HOG feature same as training set. Finally, all of HOG feature is discriminated by learned SVM. Certain amount of eye is determined on the digitized comic image using this method.

2.3. Detection Target

In research area of face tracking or detection, the target of detection is mainly whole face. However, we aim to detect eye on the comic to detect character (actor) existence. It because, in comic, character's face is often changing drastically compared with real picture.

Some examination to detect eye and face regions are expressed in our past studies [4] too. In the examination, discriminant model that learned face area derived many of mistakes than eye learned version.

Fig. 5 Positive sample of learning image, it is including only eye region. This set is constructed 89 image, and consisted of 42x42 to 150x150 pixel images.

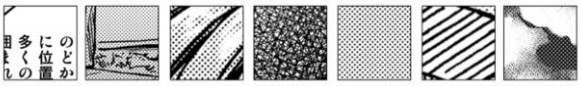


Fig. 6 Negative sample of learning image, it is including another of eye region. This set is constructed 178 image, and consisted of 32x32 to 212x212 pixel images.

Table 1 Detected positive number on one page

	resolution converting method	resolution fixed method
positive number (avg.)	131.2	276.1

3. MULTI SIZE EYE DETECTION

3.1. Effect of Resolution Conversion

Further eye detection method is including resolution conversion. However, gradient on digitized comic image is affected by image resolution conversion, and HOG feature is too.

A sample of a distribution of gradient direction is depicted as Fig. 3 and Fig. 4. It is confirmed that the trend of distribution of each resolution are different. Thus, it is assumed that conventional eye detection method affected resolution conversion.

3.2. Resolution Fixed Eye Detection

To remove the effect or resolution conversion, we utilize resolution stable HOG feature calculate method. The difference of this method with image resolution conversion method is the computation procedure of HOG feature.

In the proposed method, the block and cell number of HOG feature is fixed. On the other hand, pixel number that is included one block is changed adaptively. Another procedure is conformed to calculation of general HOG feature.

4. EXPERIMENT

4.1. Experiment for eye detection

Eye detection experiment on comic image is performed for to compare both methods. Experimental condition is as follows. HOG block number is 5x5 and cell size is 3x3 blocks. The direction of luminance gradient is quantized into 9. As a learning method, we utilized libsvm [5]. Input image is 1400x2000 pixels. That is scanned at 740dpi originally, and converted into 185dpi. Window size is four: 20x20, 40x40, 80x80, 160x160. Total window number on 1 page is 128787. The cell size of the resolution conversion method has normalized into 50x50 pixels.

Samples of positive learning image are shown in Fig. 5, and negative one is shown in Fig.6 respectively. Detection results are shown in Fig. 7 and Fig. 8. Fig. 7 is including learning images, and Fig. 8 is not including learning images i.e. unknown image. Detected average positive number is shown in Table 1.

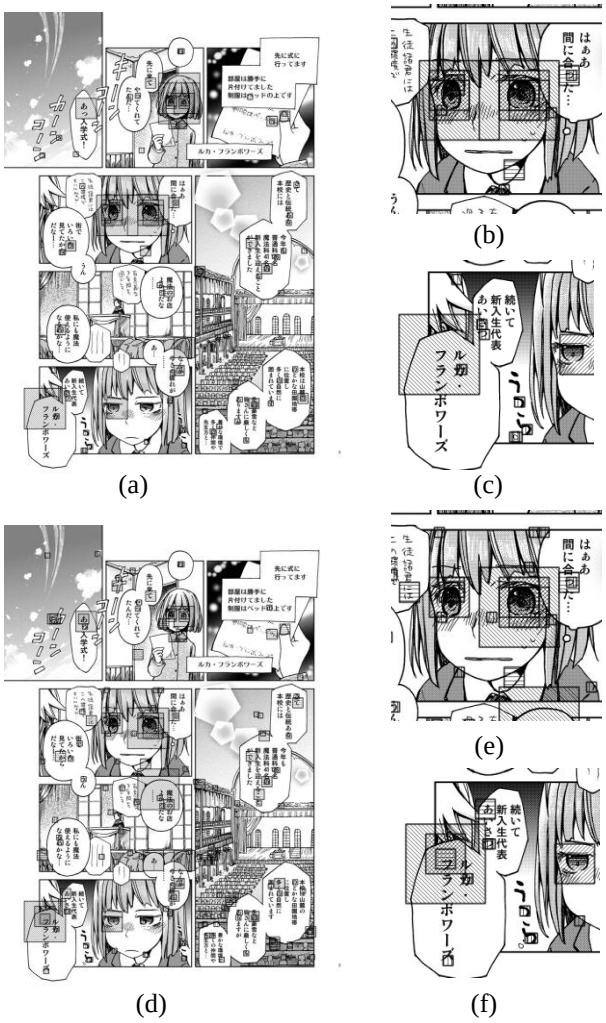


Fig. 7 is analyzed result of known images. Detected area is drawn by textured square. (A,) (B) and (C) are detected result of resolution conversion approach. (D,) (E) and (F) are detected result of resolution fixed approach. Part of learning image are derived this input image. Both methods can detect 5 eyes and 1 leak. Comparison on the image, resolution fixed method has many miss detection.

We evaluated detected and undetected number of eye. A total number of eye should be detected are 71. Could not detect eye number by both methods commonly are 6. In this experiment, resolution fixed method derived all eyes that detected by resolution conversion method. Furthermore the number of eye could detect resolution fixed method only is 4.

4.2. Consideration

By average positive number and subjective estimate on analyzed result image, following matter is suggested. In the resolution fixed eye detection method, both true positive and false positive numbers are increased compared to resolution converting method. Both methods can be obtained eye region regardless of leaned or not.

Discriminator's strictness is depends on application. The resolution fixed method makes little loose discriminator than the resolution conversion method. We assumed that loose discriminator has availability when it combined with pre or post processing utilized another aspect.

5. CONCLUSION

In this paper, to detect character's eye on the digitized comic image, first, we have showed the effect of resolution conversion for distribution of gradient direction, and proposed resolution fixed HOG feature. Then the conventional resolution converting method and proposal approach has been compared. Finally, it is suggested that resolution fixed approach makes little loose discriminator than resolution converting approach.

Fig. 8 is analyzed result of unknown images i.e. don't include learned eye. Detected area is drawn by textured square. (A,) (B) and (C) are detected result of resolution conversion approach. (D,) (E) and (F) are detected result of resolution fixed approach. Resolution fixed method can derived all eye area. On the other hand resolution converting method cannot detect one area.

ACKNOWLEDGEMENT

All Manga Image: used in fig.1 – fig. 8 on this paper are provided by Hinata Kino http://www.etheric-f.com/ .
This work was supported by JSPS KAKENHI Grant Numbers 12018464.

REFERENCES

[1] D. G. Lowe, "Object Recognition from Local Scale—Invariant Features," International Conference on Computer Vision, pp. 1150—1157, 1999.
[2] P. Viola, and M. Jonee, "Robust Real-Time Face Detection," International Journal of Computer Vision, Vol. 57, No. 2, pp. 137—154, 2004.
[3] Dalal. N, Triggs. "Histograms of Oriented Gradients for Human Detection," IEEE CVPR, pp. 886-893, 2005.
[4] Daisuke Ishii, Hiroshi Watanabe, "A Study of Automatic Human Detection for Comic Image," IPSJ-AVM, Vol. 2012-AVM76, No.2, pp.1-5, 2012.
[5] C. C. Chang and C. J. Lin. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1 – 27:27, 2011.