Proceedings of the Fourth IIEEJ International Workshop
on Image Electronics and Visual Computing
Koh Samui, Thailand, October 7-10, 2014

FACE DETECTION FOR COMIC IMAGES WITH DEFORMABLE PART MODEL

Hideaki Yanagisawa, Daisuke Ishii and Hiroshi Watanabe

*Department of Computer Science and Communication Engineering,
Graduate School of FSE, Waseda University*

ABSTRACT

Comic images include several kinds of picture elements, such as lines, dots, characters and sound effects. Therefore, they form quite complex structure compared with natural images. We have been trying to improve the convenience of e-comics by retrieving metadata elements, such as names of characters and positions of the characters. To extract characters from comic images, a method detecting characters' face using the Histograms of Oriented Gradients (HOG) features and discriminating them has been proposed. However, this method does not provide stable face detection. In this paper, Deformable Part Model (DMP), which is originally proposed to detect natural objects, is applied to comic images in order to improve accuracy of face detection. As a consequence, it is turned out that we can obtain 85.5 % detection rate for unknown images. Thus, DPM can be regarded an effective method to detect objects in comic images.

1. INTRODUCTION

According to a survey in 2011, sales of e-comics accounted for 81.7% of the e-book market [1]. It shows that e-comics have an important presence while the e-book market is expected to grow in the future.

In archived comic images, it is possible to provide more convenient services by providing metadata such as character names, and balloons and panel layout. For example, such services can be considered to change the image size to fit the screen of a terminal, to search target images from archived data based on the information of characters or particular scenes.

Currently, several approaches have been proposed for the extraction of the balloon and panel layout with high accuracy [2] [3].

On the other hand, to extract the characters, a method for detecting a face area of the characters using HOG features [4] and SVM have been proposed [5] [6]. Also, in this method, it is suggested that by limiting the iris portion of the face subjected to detection, it is possible to improve the accuracy of detection. However, the false positive rate is quite high, and does not provide stable detection.

In this paper, we apply Deformable Part Model, which is originally proposed to detect objects for a natural image, to comic images. Through the experiment of face detection to comic images, highly accurate detection rate can be obtained.

2. DETECTION METHOD

2.1 Histograms of Oriented Gradients (HOG)

Most of the comic images are painted by lines, which is basically binary representation. Comparing with a natural image, a comic image contains a lot of edge components. At the edge, the change of the intensity is large and at the flat area, the change of the intensity is small. Basically, HOG features use the information about the edge direction. Therefore, it is regarded to be a desired feature descriptor for comic images. HOG features can be calculated by the following procedure.

1. Gradient direction and gradient strength are calculated from the intensity of each pixel in the image. Next, the gradient direction is quantized to 9 discrete directions. Namely, separated into each 20° range of up to 160° from 0°.

2. The local area is split into cells, where each cells consists of 8×8 pixels.

3. In each cell area, create a gradient direction histogram of intensity. Then, obtain 9 dimensional vector.

4. The 2×2 cells are regarded as one block. To each block, normalization is performed by combining the vectors of the cells. The feature vector that is finally obtained is composed of multi-dimensional vector from the vectors of all blocks.

Fig. 1 Samples of face area.

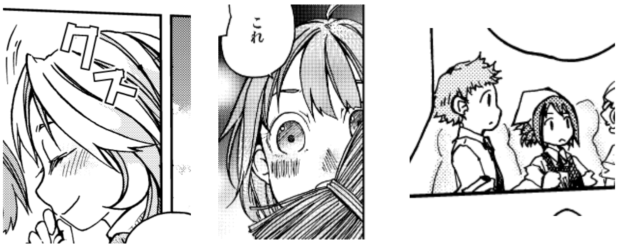

Fig. 2 Example of non-face area.

2.2 Deformable Part Model (DPM)

Deformable Part Model (DPM) is a method of object detection proposed by Felzenszalb et al. [7] [8]. This method expresses the object model as a set of parts, and evaluates it by the validity of each part and relative position relationship thereof. In the conventional method, part locations of the object are fixed. However, the part location is variable in DPM, and it is possible to respond to pose changes of the object. The score of detection window is calculated from the next equation.

$$\text{score} = \sum_{i=0}^{n} F_i \cdot \phi(H, p_i) - \sum_{i=1}^{n} d_i \cdot \phi(dx_i, dy_i) + b \quad (1)$$

In the first term of Eq. (1), it calculates the score of filters. The detection model of DPM is constructed form one route filter, which captures the entire image of object, and n-part filters, which capture several parts of the object. First, the score of the route filter is calculated form the inner product of route filter (F_0) and HOG feature map of the image ($\phi(H, p_0)$). Second, scores of part filters are calculated form the inner product of part filters ($F_1 \sim F_n$) and HOG feature map at the twice higher resolution ($\phi(H, p_1) \sim \phi(H, p_n)$).

In the second term of Eq. (1), it calculates the distortion of each part placement. Then, $\phi(dx_i, dy_i)$ suggests relative positions of the route filter and part filters, and d_i is distortion parameter.

By training, DPM sets the value of F_i and d_i, and it creates object detector. Detection by the DPM has been applied to various natural objects so far. However, detection results by DPM to objects represented by line drawings, such as the comic images, have not yet been shown. Therefore, we try to apply DPM to comic images, and check the performance of this approach through the experiment described in the next section.

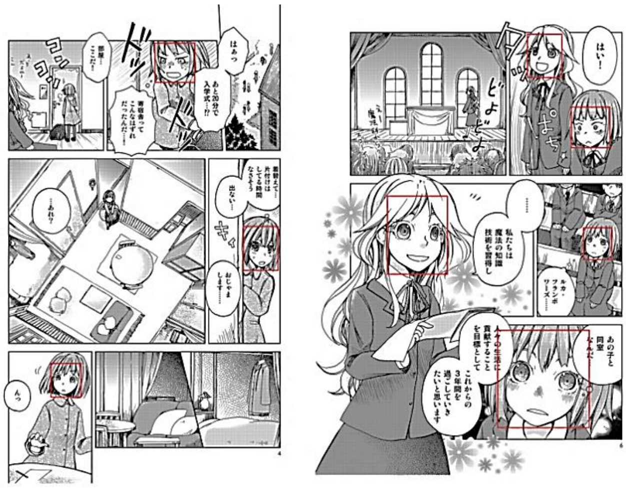

Fig. 3 Positive samples of learning images.
(Frames represent points specified in the bounding box as face areas.)

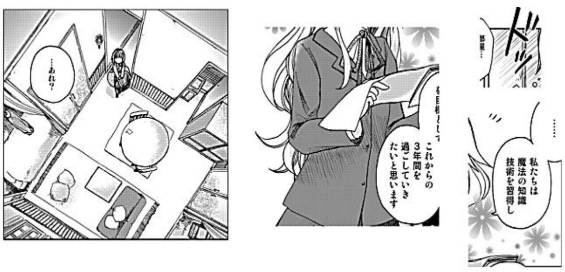

Fig. 4 Negative samples of learning images.
(Does not include the face area that defined in positive samples.)

3. FACE DETECTION

By giving the training data in DPM, we create a detector for the face area of the comic image. We also create a detector by HOG features and SVM. Two detectors are used to detect face area from comic images, and these performances are compared. In this experiment, we used the algorithm in the reference [9] for the learning model of DPM and detection.

Comic images, used in this experiment, are 26 pages, size of 1342 × 1877 pixels. We use 9 of these as a training sample. Face areas, which should be detected and learned, are assumed to satisfy the three conditions below.

1. Both eyes are included.
2. Size is 60×60 pixel or more.
3. Area from the top of eyes to chin is included.

Fig. 1 shows the example images of face area assumed by above conditions. Fig. 2 shows the example images of non-face area.

Positive samples, specifying the face areas, are taken and shown by bounding box from the learning image. Fig 3 shows the example of positive samples.

Negative samples are those to cut out a region that does not include the face area from learning images. Fig 4 shows the example of negative samples. As a result, positive samples are 28 locations in 9 images, and negative samples are 114 sheets of images.

4. RESULT

Fig. 5 shows the detection model of the face region generated by DPM after learning. From these figures, relative locations can be seen. Further, Fig. 6 shows an example image of the result of performing the detection of the face area from a comic image using two detectors. Table 1 shows the result of performing detection from known and unknown images using the two detectors. From the experiment, it is turned out that DPM greatly outperforms HOG from 17.4% to 61.3%.

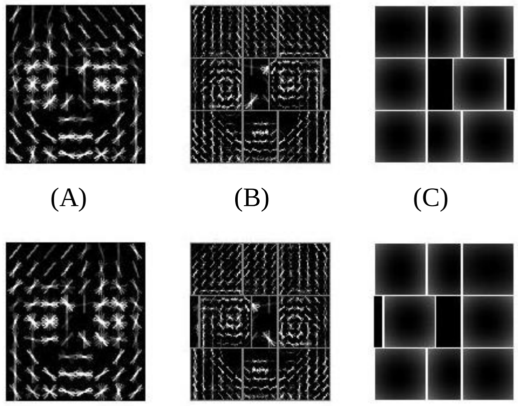

Fig. 5 Face detection model generated by DPM.
(A) HOG features of root filter. (B) HOG features of part filters. (C) Location of part filters.

Table. 1 Results of face detection from known and unknown images with DPM and HOG.

	Known images			Unknown images		
	precision	recall	F-measure	precision	recall	F-measure
DPM	92.9%	100%	96.3%	85.5%	100%	92.2%
HOG	75.0%	82.6%	78.6%	32.7%	38.7%	35.4%

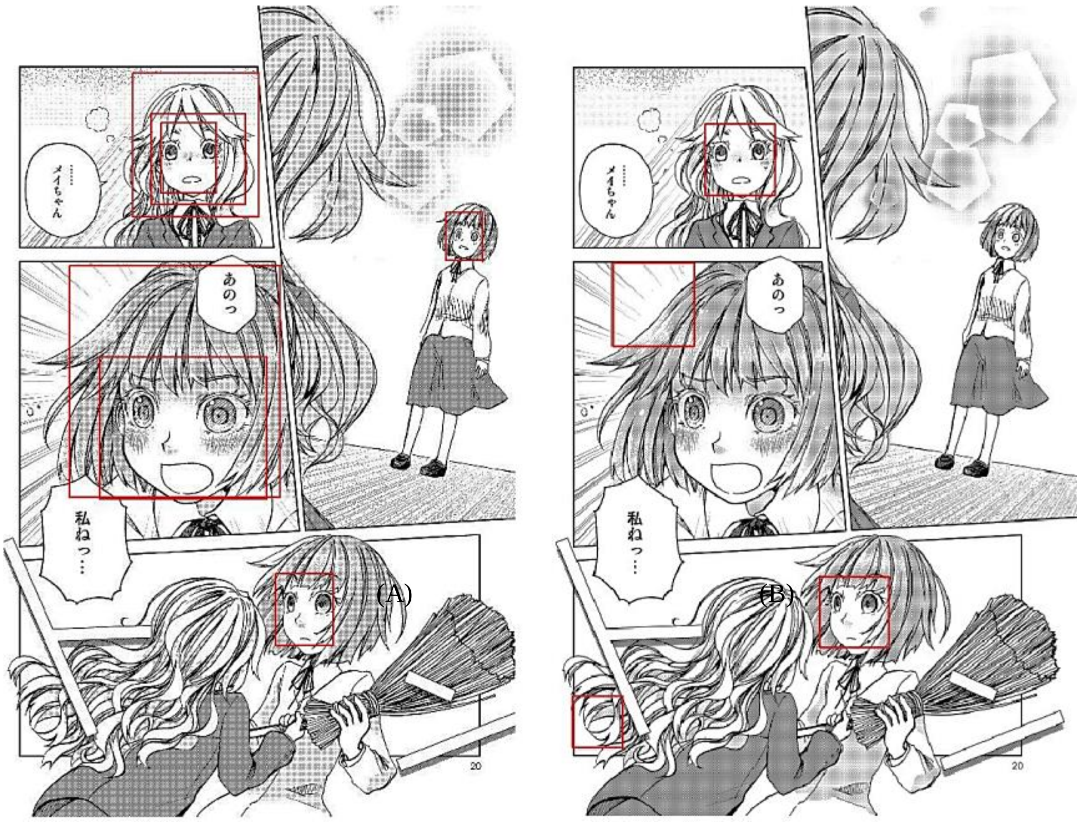

Fig. 6 (A) Result of face detection with DPM. (B) Result with HOG from same image. (Frames suggest detected area as character’s face).

5. CONCLUSION

In this study, we have tried to improve the detection accuracy of the character's face area in comic images by using the DPM. The experimental results showed that the DPM improves both precision and recall for unknown images compared with the method that uses HOG only. Therefore, we could conclude that DPM is valid for detecting objects in comic images.

On the other hand, DPM could not detect radically distorted face images as shown in Fig. 7 It seems difficult to detect all of such images by using only DPM. To solve this problem, combination of DPM and some other features such as texture pattern may be effective.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 25330137. All comic images used in Fig.1-Fig.7 except for Fig 5 in this paper are provided by Hinata Kino　(http://www.etheric-f.com/).

REFERENCES

[1] Internet Media Research (Ed), "e-Comic Business Report 2012", Impress Business Media Corporation, 2012.

[2] S. Nosaka, T. Sawano and N. Haneda, "Development of "GT-Scan", the Technology for Automatic Detection of Frames in Scanned Comic", FUJIFILM RESARCH & DEVELOPMENT, No. 57, pp. 46-49, 2012.

[3] T. Tanaka, F. Toyama, J. Miyamichi and K. Shoji, "Detection and Classification of Speech Ballons in Comic Images," ITE Vol. 64, No. 12, pp. 1933-1939, 2010.

[4] N. Dalal, and B. Triggs,"Histograms of Oriented Gradients for Human Detection," IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp.886-893, 2005.

[5] D. Ishii and H. Watanabe "A Study of Automatic Human Detection for Comic Image," IPSJ SIG Technical Report Vol.2012-AVM-76, No.1, pp.1-5, Feb, 2012.

[6] T. Tanaka, K. Shoji, F. Toyama, and J. Miyamichi, "Layout Analysis of　Tree-Structured Scene Frames in Comic Images," Proc. 20th International Joint Conference on Artificial Intelligence , pp.2885-2890, 2007.

[7] P. Felzenszalb, R. Girshick, D. McAllester, D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol.32, No.9, pp.1627-1645, 2010.

[8] P. Felzenszalb, D. McAllester, D. Ramanan, "A Discriminatively Trained, Multiscale, Deformable

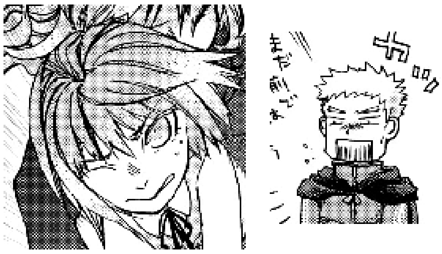

Fig. 7 Samples of face images that DPM could not detect faces.

Part Model," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[9] R. B. Girshick, P. Felzenszwalb, and D. McAllester, " Discriminatively Trained Deformable Part Models, Release 5", http://people.cs.uchicago.edu/~rbg/latent-release5/, <accessed in 7.2.2014>.