Detection of Comic Character in Digital Image Based on HOG Features

Imai Satoshi[1], Naoya Wada[2], Atsushi Koike[3]

[1,2,3]Faculty of Science and Technology, Seikei University, Tokyo, Japan

ABSTRACT

This paper explores the detection of characters in comic images for providing readers beneficial information about them on e book. Although Haar-like features are widely used for the face detection from natural images, it is difficult to apply the same way to comic images because faces in comics are comically manipulated in general and are often not full-face. Our method is intended to detect eye regions which are very characteristic in comic images. Feature detection of local image is realized by using HOG features and the extracted data is given to SVM to classify the local image. Experimental result showed more than80% detection accuracy for the detection of characters with naked eyes.

Keywords: Character detection, comic image, HOG features, support vector machine

HOW TO CITE THIS ARTICLE

Imai Satoshi, Naoya Wada, Atsushi Koike, "Detection of Comic Character in Digital Image Based on HOG Features", International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 7 Issue 4, April -2018.

1. INTRODUCTION

An e-book is a digital form of a printed book. As the progress of mobile device conventional printed books get to be enjoyed as e-book. E-books are available on electronic devices and it produces an advantage that not only images of pages but also other information can be displayed on such devices simultaneously. Especially comic images include much no literal information about the characters and stories.

The goal of this research is to realize a new way to enjoy comics by making the most of advantages of e-book. One example is illustrated in Fig.1 where information about a character is displayed when a reader taps that character on a tablet. In order to realize it, regions of characters on the images have to be known to the application. Although the regions can be directed manually, automatic character detection should be applied to comic images to reduce costs. The use of Haar-Like features which compares the sum of intensity of two rectangular regions is well known method to detect face regions from natural images. However, characters in comics are drawn from many angles and are often comically manipulated. It means that the face detection from comic images cannot be dependent on the typical relationship of facial parts in natural images and it is hence difficult to detect face regions from comic images by using conventional algorithms.

Our approach is intended to detect eye regions because the character detection algorithms for comic images cannot be dependent on the relationship of facial parts and eyes are very characteristic components in comics. We present an eye feature analysis method using HOG features and show a result of evaluation using classification with SVM.

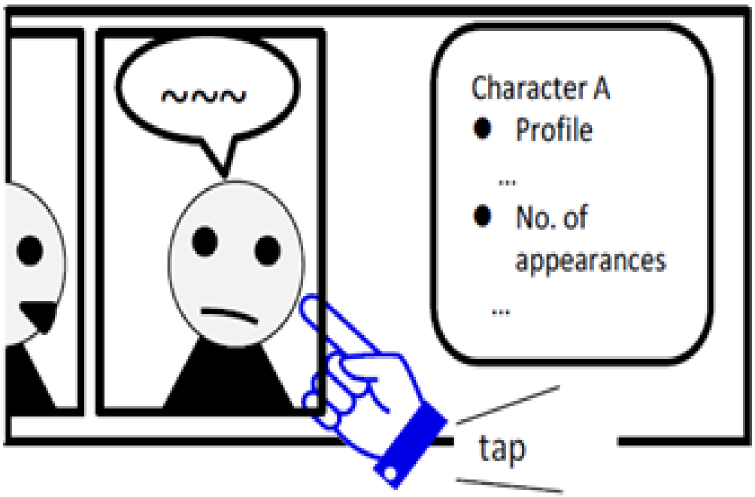


Figure 1: Example of enjoying comics as e-book

2. HOG FEATURE AND EYE REGION EXTRACTION

The Histograms of Oriented Gradients (HOG) features proposed by Dalal et al. [1] are based on local histograms of gradient orientations of intensity. HOG features characterize local object appearance and shape with the distribution of local intensity gradients of edge directions. It even does not require precise knowledge of the corresponding gradient or edge positions. Characteristics of a local object are represented by HOG features which are used to detect characters in comic images. For natural images, HOG features are often used to human detection [2][3]. The first step of calculation is gradient computation, where the magnitude m(x, y) and gradient orientation $m(x, y)$ of each pixel are calculated as follows.

$$\mathrm{m(x,y)} = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \quad (1)$$

$$\theta(x,y) = tan^{-1}\left(\frac{f_y(x,y)}{f_x(x,y)}\right) \quad (2)$$

$$\begin{cases} f_x(x,y) = L(x+1,y) - L(x-1,y) \\ f_y(x,y) = L(x,y+1) - L(x,y-1) \end{cases} \quad (3)$$

where (x, y) denotes position coordinate, $f_x(x, y)$ and $f_y(x, y)$ denote vertical gradient and horizontal gradientrespectively and $L(x, y)$denotes the brightness. Secondary,the gradient orientations are quantized into N groups and the sum of the magnitudes of quantized gradient orientations in cell region c is obtained as a histogram.

$$V_c = \{v_c(1), v_c(2), v_c(3), \cdots, v_c(N)\} \quad (4)$$

where $v_c(i)$ denotes the sum of the magnitude of the *i* the group. Finally, the cells are grouped into overlapping blocks ($q \times q$ cells) and the features are normalized as follows.

$$v_c'(n) = \frac{v_c(n)}{\sqrt{\sum_{k=1}^{q\times q\times N} v_c(k)^2 + \varepsilon}} \quad (5)$$

$$\varepsilon = 1$$

The process to extract HOG features is illustrated in Fig.2. Fig.2 (a) shows gradient orientations, where the real lines show the border of cells and the dashed lines show the border of pixels. Fig.2(b) shows a histogram of the magnitudes of quantized orientations. We use HOG features to detect eyes in comic images. Extracted data from each block is given to Support Vector Machine (SVM) which is intended to decide whether the block is an eye region or not.

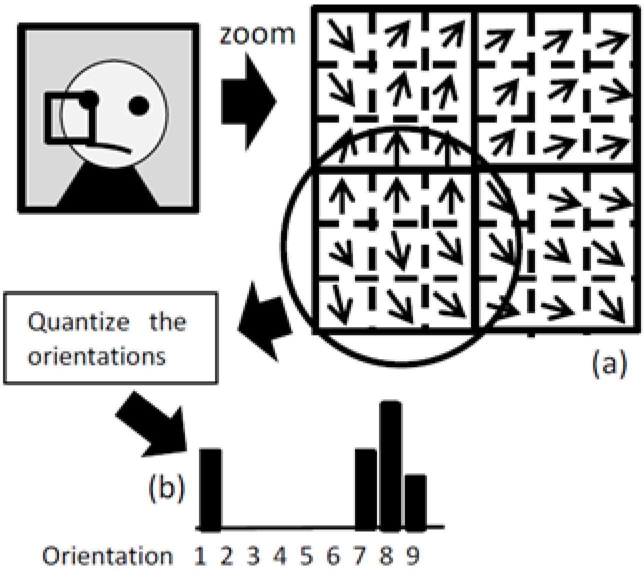


Figure 2: Extraction of HOG features

3. HOG FEATURE AND EYE REGION EXTRACTION

Support Vector Machine proposed by Vapnik [4] is a pattern recognition algorithm based on supervised learning model. SVM is generally supposed to classify some data points into two classes and is used in many other applications [5][6].One of the biggest problems in machine learning is that too much training makes the classifier lose flexibility to classify unknown data, and it is called over fitting. However, training in SVM can consider the tradeoff between classification accuracy for a given training dataset and flexibility for unknown data. It is realized by mapping the input data space into a higher dimensional space and deciding a hyper plane which separates two classes with considering the tradeoff between accuracy and flexibility.

Although many separating hyper planes can be thought, the optimal hyper plane among them gives the maximum margin, where the margin is defined as the distance between a separating hyper plane and the nearest training data. Now let a training dataset D as

$$D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\},\ x_i \in R^n \quad (6)$$

where $y_i \in \{-1, 1\}$denotes the binary classification. The goal is to find an optimal separating hyper plane

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad (7)$$

where where **w** and ***b*** denotes the weight and the bias. If two classes are linear separable, finding an optimal separating hyper plane is equivalent to solve the optimization problem expressed as

$$\text{Minimize:} Q(\mathbf{w},) = \frac{1}{2}\|\mathbf{w}\| (8)$$

$$\text{s.t.: } y_i(\langle \mathbf{w}, \mathbf{x} \rangle + b) \geq 1,\ i = 1,2,\cdots,n \quad (9)$$

A SVM obtained by solving this problem is called hard margin SVM and Fig.3 illustrates this separation. If two classes are non-linear separable, slack variable$\xi_i (\geq 0)$intended to relax hard margin is implemented, and added to Eq.(8) and Eq.(9) as

$$\text{Minimize: } Q(\mathbf{w},b,\boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\| + C\sum_{i=1}^{n}\xi_i \quad (10)$$

$$\text{s.t.: } y_i(\langle \mathbf{w},\mathbf{x}\rangle + b) \geq 1 - \xi_i,\quad \xi_i \geq 0,\quad i = 1,2,\cdots,n \quad (11)$$

where C denotes a variable to settle the tradeoff between maximization of a margin and classification error caused by the second term in Eq. (10). Figure 4 illustrates this situation. If samples whose slack variable is more than 0 are removed, the problem of the non-linear separable is equivalent to the problem of linear separable.

In order to resolve the optimization problem, Lagrange multipliersα_i and $\beta_i$are introduced and the problem isconverted to

$$\text{Minimize: } Q(\mathbf{w},b,\xi,\alpha,\beta) = \frac{1}{2}\|\mathbf{w}\| + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(y_i(\langle \mathbf{w},\mathbf{x}\rangle + b) - 1 + \xi_i) - \sum_{i=1}^{n}\beta_i\xi_i \quad (12)$$

Considering Karush-Kuhn-Tucker (KKT) conditions, Eq.(12) is converted to

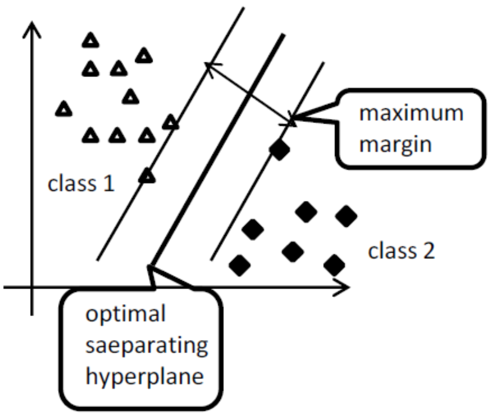


Figure 3: Data points and separating hyper plane in linear separable situation.

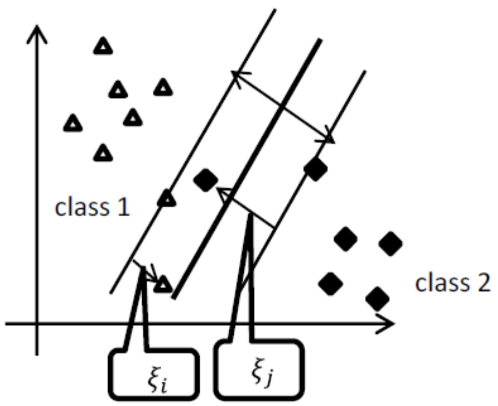


Figure 4: Non-linear separable situation.

$$\text{Maximize: } \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j) \quad (13)$$

$$\text{s.t. } \sum_{i=1}^{n}\alpha_i y_i = 0,\ 0 \leq \alpha_i \leq C,\ i = 1,2,\cdots,n \quad (14)$$

where$K(x_{i,}y_j)$ is a kernel function, and Eq. (13) and (14)must satisfy

$$\alpha_i^*(y_i(\langle \mathbf{w}, \mathbf{x} \rangle + b^*) - 1 + \xi_i^*) = 0,\ \ i = 1,2,\cdots,n \qquad (15)$$

where α_i^* and $\xi_i^*$are solutions' ones. Finally, therefore, the optimal decision function is obtained as

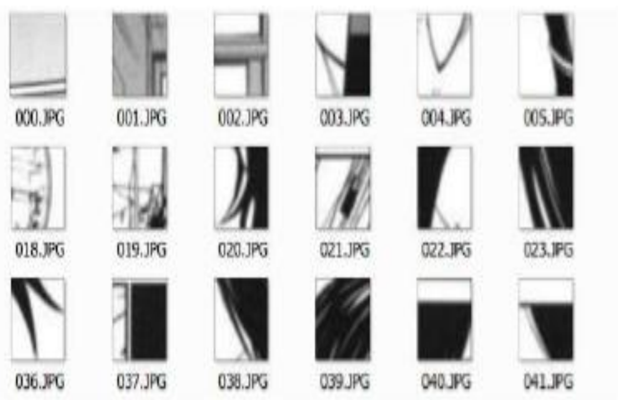
$$f(x) = \mathrm{Sgn}(\textstyle\sum_{i=1}^{n} y_i\, \alpha_i^* K(x_i, x_j)\ + b^* \qquad (16)$$

$$\mathrm{w}^* \ = \ \textstyle\sum_{i=1}^{n} \alpha_i^* x_i y_i \qquad (17)$$

4. EXPERIMENTS

We performed an experiment aimed to evaluate the eye region detection. A SVM was trained in advance to the experiment using 100 eye images as positive samples and 160non-eye images as negative samples which are extracted from several comics. A part of those samples are shown in Fig.5. The input dataset consisted of 215 images from a comic "Shigatsuwa Kimi no Uso" [7] and these images were normalized so the vertical size was 500 pixels (aspect ratio was kept). The block size for extracting HOG features was settled at 30 × 30 pixelsso too small eyes are excluded from the number ofappearance. The cell size to detect HOG features is settled at 55 pixels.

(a) Positive sample



(b) Negative samples

Figure 5: Part of samples for training SVM

Table I shows the statistical result. More than 80%detection rates were obtained about 3 characters. On the other hand, the result for Character A was degraded. This character wears glasses unlike the others and it disturbed the detection.

However, HOG features are still helpful for the character detection and more correctness will be obtained by combining it in other features.

Table I: Experimental Rsults.

Character	A	B	C	D
No. of apperance	76	43	15	45
No. of detection	8	36	13	36
detection rate [%]	10.53	83.72	86.67	80.00

SUMMARY

We have presented an eye region detection algorithm intended to detect characters from comic images. The algorithm extracts HOG features to obtain local information of an image. Extracted data is given to SVM and whether the block is an eye region or not is decided. Experimental result showed more than 80% detection rate for naked eyes and it indicates that the proposed algorithm can contribute character detection from comic images. Future work is to increase detection accuracy and reduce false detection. In order to realize it, more precise analysis of local image will be required. For example, SIFT and SURF techniques detect key points in a comic image. Our method will be improved by combining such techniques. Furthermore, the relationship of facial parts in natural images is very important in face detection from natural images. Although it is difficult to utilizing it in comic images because of comical manipulation, utilizing of facial parts other than eyes still probable to improve detection accuracy.

REFERENCES

[1]. N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection”, Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp.886–893.

[2]. T. Mitsui and H. Fujiyoshi, “Object Detection by Joint Features basedon Two-Stage Boosting”, Proc. of IEEE International Conference on Computer Vision Workshops (ICCV), pp.1169–1176, 2009.

[3]. M. Morita, M. Ding, H. Takemura, and H. Mizoguchi, “Human Detection Method Based on Feature Co-occurrence of HLAC and HOG”, Proc. of IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1689-169. 2011.

[4]. V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer, 1995.

[5]. Lei Li, Zhi-ping Gao and Wen-yan Ding, “Fuzzy multi-class support vector machine based on binary tree in network intrusion detection”, Proc. of International Conference on Electrical and Control Engineering(ICECE), pp. 1043-1046, 2010.

[6]. J. Susukh, S. Premrudeepreechacharn, T. Kasirawat, “Power quality problem classification using support vector machine”, Proc. Of International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON),pp. 58-61, 2009.

[7]. N. Arakawa, “Shigatsu wa Kimi No Uso”, vol.1, Koudansha, 2011.