2011 International Conference on Document Analysis and Recognition

Similar Manga Retrieval Using Visual Vocabulary Based on Regions of Interest

Weihan Sun
Department of Computer Science and Intelligent Systems
Osaka Prefecture University
Osaka, Japan
sunweihan@m.cs.osakafu-u.ac.jp

Koichi Kise
Department of Computer Science and Intelligent Systems
Osaka Prefecture University
Osaka, Japan
kise@cs.osakafu-u.ac.jp

***Abstract*—Manga, a kind of graphic novels expressed by sequential line drawings, is an important document image publication which is invoking more and more attentions for their copyright protection. Because of simple constructions and abstract expression styles, similar copies are always applied in plagiarisms of mangas. In addition, the enormous volume of copyrighted manga publications issues a challenge to the task of detecting suspicious images. Considering only some regions of interests (ROIs) require copyright protection, we propose a bag-of-features method using visual words based on ROIs for similar manga retrieval. In the experiments, we applied real manga publications as data and proved the effectiveness of the proposed method.**

***Keywords*-copyright protection, similar image retrieval, bag-of-features, manga, similar copy**

I. Introduction

The term of manga comes from the Japanese language, which indicates a kind of narrative artwork expressed by sequential line drawings. Although the modern style of manga came to this world much later than other kinds of documents, it develops quickly and becomes an important publication over the world. Especially in Japan, manga is popular for people of all ages and have steadily become a major part of Japanese publishing industry. The manga market of Japan is over $6 billion, while the export amount is approximate $15 billion[1]. Recently, the developing digital technique drums up another booming business: e-manga (digital manga magazine) which can be easily downloaded and viewed by digital terminals such as PCs and cell phones. However, in contrast to the conveniences offered by the techniques, the problem of illegal copies becomes more serious, which is threatening the manga industry.

In reality, whether a drawing is an illegal copy or not is a controversy problem. It always depends on the judgment of professionals. However, because of the huge volumes of manga publications, it is impossible for human beings to check them page by page. The purpose of our research is to apply computer techniques instead of human beings to detect suspicious images for professionals' further judgments.

[1]based on the number from "2006 Publication Index Annual Report" by the All Japan Magazine and Book Publishers' and Editors' Association, the Research Institute for Publications

Since the objects in mangas are basically described by simple lines with few colors, it is easy to copy them by handwriting. Furthermore, by gripping some abstract features of mangas, similar copies can also be created, which challenge the illegal copy of detection of mangas.

In our previous research, we proposed applying the technique of image retrieval to detect partial copies of line drawings: copyrighted images are collected in a database; suspicious images are treated as queries; suspicious images are reported as results [1]. By applying a method of local feature matching, we achieved detecting both printed and handwritten partial copies of line drawings from complex backgrounds. However, because enormous volume of manga publications require copyright protection, the problem of database overload prevents the method to be applied in practice. Considering not the whole image of manga needs copyright protection, we have proposed applying Viola-Jones detection framework [2] to detect faces as regions of interests (ROIs) and only store ROIs into the database [3]. Although the method showed the scalability of database size and achieved the detection of similar ROIs, only faces of characters can be detected as ROIs.

As the first step of detecting illegal copies of mangas, we have to know which kind of manga (called a title of manga) is copied. In this paper, we propose a bag-of-features method using visual words based on ROIs for similar manga retrieval. For ROIs, since faces of characters are one important kind of ROIs in mangas, face ROIs detected by [3] are applied. Besides face ROIs, we also apply generic ROIs for overcoming the limitation of face ROIs and propose a method to detect more discriminative clusters of generic ROIs. In the part of experiments, we applied 20 titles of Japanese mangas as our data and proved the effectiveness of the proposed method. The experimental results also reveal that (1) besides face ROIs, there are also many other clusters of generic ROIs in mangas, (2) the effectiveness of face ROIs for similar manga retrieval with enhancement of clusters of generic ROIs.

The rest parts of this paper is arranged as follows: Section 2 provides an overview of the approach, section 3 introduces the region detectors and the feature descriptor applied in this research. Section 4 and section 5 describe the methods

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.217

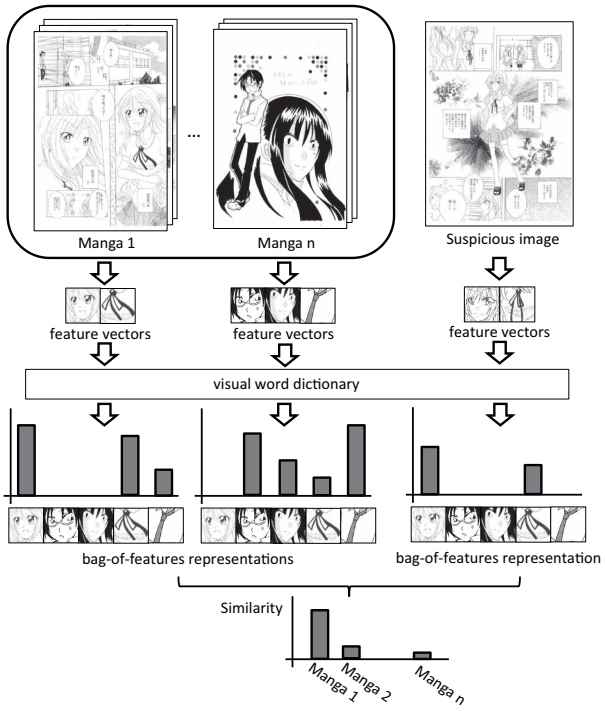


Figure 1. Outline of the proposed method.

of detecting clusters of generic ROIs and bag-of-features method for manga retrieval. Experiments and results are shown in Section 6. Finally, Section 7 is conclusions and future work.

II. Outline of the Approach

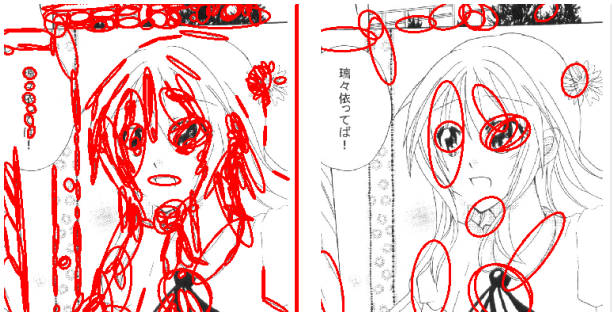
To retrieve similar mangas, we propose to apply a bag-of-features method by using the visual words based on face ROIs and clusters of generic ROIs. The bag-of-features method is a popular method for object classification and document retrieval, such as [10], [11] and [12]. It is analogous to the bag-of-words representation of text documents by an unordered set of words.

For our case, as shown in Fig. 1, copyrighted mangas categorized by their titles are collect in the database. By matching the feature vectors extracted from the database with a visual word dictionary, each title of manga get a bag-of-features representation. On the other hand, a suspicious image is treated as a query. By the same method, the query is also represented by a bag-of-features representation. Based on the comparison of bag-of-features representations between the query and the database, the title of manga with the maximum similarity is reported as the result.

III. Region Detector and Feature Descriptor

A. Region detector

In this research, two kinds of ROIs are applied: face ROIs and generic ROIs. For detecting face ROIs, we applied the region detector of [3].



(a) Ellipse MSERs. (b) Applied regions.

Figure 2. Examples of ellipse MSERs and candidate regions.

For detecting generic ROIs, we apply MSER (Maximally Stable Extremal Regions) [4] as the region detector. By the region detector, maximally stable regions (MSERs), in which the intensities of all pixels are greater or smaller than their boundaries, can be detected from the images. By diagonalizing the covariance matrices of MSERs, we can get some ellipse regions, as shown in Fig. 2(a). To make the regions contain more information, size of MSERs are magnified as k times (in this research we set $k = 1.2$). As shown in Fig. 2(b), considering the stability of MSERs, only the regions above a certain size are applied. Finally, we normalize the regions by rotating the long axis of ellipse parallel to the y axis of the image as generic ROIs.

B. Feature descriptor

Since similar line drawings contain many changes in detail, we propose applying HOG (Histogram of Oriented Gradients) [5] as the feature descriptor. In [1], we have proved that HOG outperform the SIFT (Scale-Invariant Feature Transform) [6] feature descriptor for describing handwritten copies of line drawings.

As shown in Fig. 3, first, we calculate gradient magnitude and orientation at each pixel and divide each ROIs into 8×8 cells evenly. Then, the gradient orientation are quantized into 6 bins. For each cell, we calculate the gradient orientation histogram based on the gradient magnitude. After that, cells are combined into overlapped blocks as 2×2 cells per block. By normalizing the features in blocks we obtain $6 \times 2 \times 2 \times 7 \times 7 = 1,176$ HOG features for each candidate region. These features are combined into a HOG feature vector which consists of $1,176$ dimension.

C. Matching

In this research, we define the similarity by Euclidean distance D between feature vectors F_i and F_j. If $D < T_1$ (T_1 is a threshold), F_i is similar with F_j, otherwise not. We will discuss the effectiveness of T_1 for similar manga retrieval in the part of experiments.

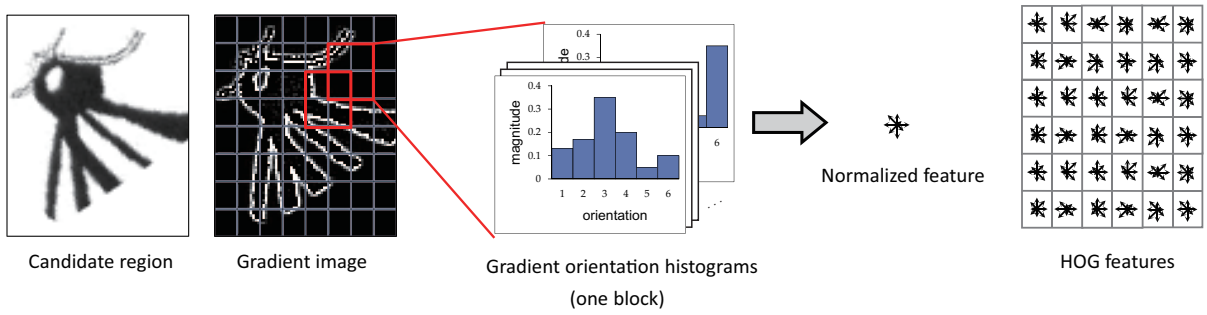


Figure 3. Calculation of HOG features.

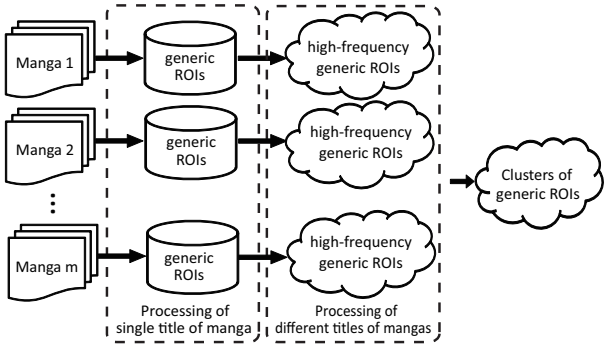


Figure 4. Extraction processing of clusters of generic ROIs

IV. Detection of Clusters of Generic ROIs

A. Overview

To select discriminative generic ROIs, we propose to detect clusters of generic ROIs according to the dependency between ROIs and mangas. Manga is a kind of graphic novel that describes a sequential narrative by drawings. Normally, one title of manga consists of multiple volumes which contain several manga pages. Because the main parts are usually drawn repeatedly in one title of manga, high-frequency parts should be important. However, some parts like word balloons, stripes and tones are employed frequently in all titles of mangas but outside the range of copyright. Therefore, we should consider the dependency between ROIs and mangas from the two scales: the scale of a single title of manga and the scale of different titles of mangas.

The processing of the proposed method is shown in Fig. 4. First, in the processing of a single title of manga, high-frequency generic ROIs are detected. Then, in the processing of different titles of mangas, high-frequency generic ROIs from different mangas are merged. By a further selection clusters of generic ROIs are outputted.

B. Processing of single title of manga

For each title of manga, we calculate in the feature space the neighbors of all generic ROIs and joint similar ones into clusters by a union-find algorithm. Clusters of high-frequency generic ROIs of manga d are defined as

$$\frac{N_{c,d}}{N_d} > T_2 \tag{1}$$

where $N_{c,d}$ is the number of generic ROIs in cluster c of manga d, N_d is the total number of generic ROIs from manga d, T_2 is a threshold that controls the number of clusters used for next stage. In this research, T_2 is set to be 5×10^{-5}.

To increase the speed of distance calculation, we propose applying ANN (Approximate Nearest Neighbor Searching) [7]. ANN is a method to find approximate nearest neighbors by using the k-d tree. While searching, the feature space shrunk by the factor $1/(1+\varepsilon)$ (ε is set to be 10 in this research).

C. Processing of different titles of mangas

In the processing of different titles of mangas, first, we merge the similar high-frequency generic ROIs from different titles of mangas. Then, all the merged clusters are evaluated by considering mutual dependency between clusters and different titles of mangas. We propose applying mutual information [8] to select them. Mutual information can be equivalently expressed as

$$I(X;Y) = H(X) - H(X|Y) \tag{2}$$

where X represents titles of mangas and Y represents merged clusters, $H(X)$ is the entropy of different titles of mangas, $H(X|Y)$ represents the conditional entropies of mangas given clusters in this method. Mutual information measures the information that X and Y share. In other words, it measures the certainty of X by knowing Y. The entropy of different titles of mangas is measured by

$$H(X) = -\sum_{j=1}^{n} P(x_j) \log P(x_j) \tag{3}$$

where $P(x_j)$ is probability for a titles of manga j, n is the number of different titles of mangas. All the probabilities P applied for the calculation of entropy are based on the number of ROIs in this method. $H(X|Y_i)$ is measured by

$$H(X|Y_i) = -\sum_{j=1}^{n} P(x_j|y_i) P(y_i) \log P(x_j|y_i) \tag{4}$$

where $P(x_j|y_i)$ is the probability of manga j given the merged cluster i, $P(y_i)$ is the probability of cluster i. By evaluating every merged clusters based on their mutual information with mangas, we get several clusters (Y_i) which satisfy $I(X;Y_i) > T_3$. The decision of T_3 depends on the distribution of mutual information of Y_i.

The advantages of the processing of this method are twofold: one is for considering frequency of generic ROIs in the scales of both one single title of manga and different titles of mangas, the other one is dominancy for calculations, including memory cost for distance calculations and needless recalculations of formal clusters of high-frequency generic ROIs for additional titles of mangas.

V. Manga retrieval

As visual words, we applied the HOG feature vectors of face ROIs and the centroid of feature vectors extracted from generic ROIs within one cluster. By gathering the visual words, we can build a visual word dictionary. Each title of manga d is represented by a bag-of-features representation $\boldsymbol{V}_d = (R_{1,d}, R_{2,d}, ..., R_{n,d}, W_{1,d}, W_{2,d}, ..., W_{m,d})$, where n and m are the total number of visual words based on face ROIs and clusters of generic ROIs, $R_{f,d}$ and $W_{c,d}$ represent weights for visual words based on face ROIs and clusters of generic ROIs, respectively. $R_{f,d}$ is set to be 1 if face ROI f belongs to manga d, overwise 0.

For the weight $W_{c,d}$, tf-idf (term frequency-inverse document frequency) weighting [9] is employed. In our case, the generic ROIs are terms and different titles of mangas are documents. Therefore, the weight $W_{c,d}$ is calculated by

$$W_{c,d} = \frac{n_{c,d}}{n_d} \log \frac{N}{n_c} \tag{5}$$

where $n_{c,d}$ is the number of generic ROIs of cluster c in manga d, n_d is the total number of generic ROIs in manga d, N is the total number of mangas, and n_c is the number of mangas containing the cluster c.

The query is also represented by the bag-of-features representation $\boldsymbol{Q}$. We calculate the similarities S_d between $\boldsymbol{Q}$ and $\boldsymbol{V}_d$ as

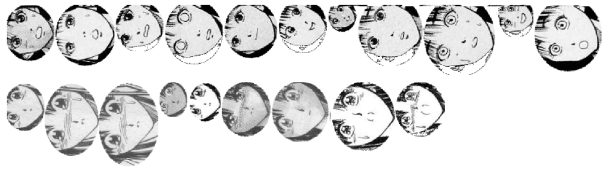
$$S_d = \frac{\boldsymbol{Q} \cdot \boldsymbol{V}_d}{\|\boldsymbol{Q}\| \|\boldsymbol{V}_d\|} \tag{6}$$

The manga with the maximum similarity is reported as the result.

VI. Experiments

A. Conditions

We collected 20 titles of Japanese mangas and employed vol. 1 and vol. 2 as our database. It contains $6,642$ full manga pages (about $800 \times 1,200$ pixels) in total. As queries, we utilized $1,884$ manga pages from vol. 3 of 10 titles of mangas whose vol. 1 and vol. 2 are stored in the database. Therefore, the queries are similar with images of the same manga but not exactly the same. All the pages of the manga magazines including cover pages were applied in our experiments without any selections.



(a) ROIs of face patterns.

(b) ROIs of kanji patterns.

(c) ROIs of mark patterns.

Figure 5. Examples of detected clusters of generic ROIs. The lines show ROIs in the scale of single title of manga. The rows show ROIs in the scale of different titles of mangas. (There are parts of real clusters of generic ROIs).

B. Detection of clusters of generic ROIs

First, we did an experiment to know what kinds of clusters of generic ROIs can be extracted by the proposed method.

By the region detector of MSER, we detected $1,376,571$ generic ROIs. For the processing of single title of manga, we got $216,614$ high-frequency generic ROIs. For the processing of different titles of mangas, we got $11,827$ clusters of generic ROIs. Examples of detected clusters of ROIs are shown in Fig. 5. The ROIs from different titles of mangas are shown in different lines, and ROIs in one line are from the same manga. From the examples, we can see that besides faces of main characters, many other patterns have high dependency with certain mangas. Such as Fig. 5(c) shows, they are a kind of special patterns only appeared in one manga. Also, we can see the effectiveness of the proposed method for clustering the similar regions. Besides scale transformations and rotations, the regions contain many changes in detail. Such as shown in Fig. 5(b), the regions are similar but not the same kanji character.

C. Similar manga retrieval

In this experiment, we tested the effectiveness of the proposed method for similar manga retrieval. As comparative methods, we employed the following 3 kinds of bag-of-features methods: (1) the method using all generic ROIs, in which feature vectors extracted from all generic ROIs were applied as visual words, (2) the method only using

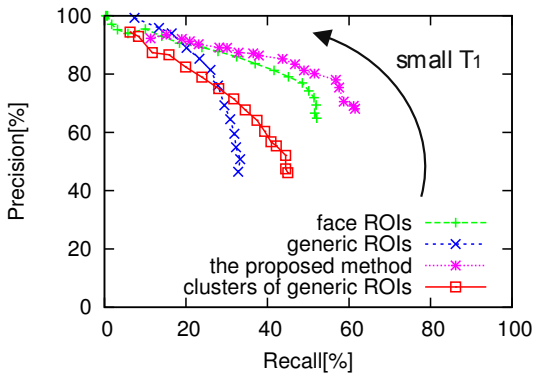


Figure 6. Precision and recall for similar manga retrieval.

clusters of generic ROIs, whose visual words are based on clusters of generic ROIs, (3) the method using only face ROIs, which applied feature vectors extracted from face ROIs as visual words. Their feature descriptor and matching method were the same. The same titles of mangas as queries were treated as the right answers. Since the matching of visual words is closely related to threshold T_1, we did the retrieval experiments by applying different threshold T_1 and got the results as shown in Fig. 6.

With the increase of T_1, because we can get more ROIs for both right and erroneous matchings, the precisions decreased with the increase of recall. For a small T_1, only ROIs with few differences were matched. Since the method based on clusters of generic ROIs applied the centroid of feature vectors extracted from ROIs, the visual words are different from original feature vectors. Therefore, for a small T_1 the method based on clusters of generic ROIs performed worse than the other three methods. Above a certain T_1, the matchings began to be untrusted and more erroneous visual words came out from queries, which decreased both precision and recall.

From the results, we can see

- The clusters of generic ROIs are more discriminative than generic ROIs without clustering. Comparing the method using all generic ROIs, the method using clusters of generic ROIs performed better.
- Face ROIs are one kind of important ROIs for mangas. Since face patterns are highly discriminative and almost always included in mangas, the method only using face ROIs showed high performance for similar manga retrieval. Although there are some face patterns in clusters of generic ROIs, the performance of clusters of generic ROIs did not outperform the method only using face ROIs, because of the limited number and less discriminative power of face patterns in clusters of generic ROIs.
- Besides face ROIs, other generic ROIs also offer us a clue for manga retrieval. Therefore, the proposed method performed best in the 4 methods for similar manga retrieval, even better than the method only using face ROIs for recall up to 10%.

The main reasons for failures are: (1) some manga pages, such as cover pages, do not contain discriminative ROIs, (2) limited database images for detecting the clusters of generic ROIs. Because of the complexities of mangas, we need more training data.

VII. Conclusion and Future Work

In this paper, we proposed a bag-of-features method using visual words based on ROIs for similar manga retrieval. Besides face ROIs, we proposed applying clusters of generic ROIs to enhance the effectiveness. From the experimental results, we proved the effectiveness of the proposed method. In addition, the results revealed that many kinds of generic ROIs can offer us discriminative features of mangas, which help us protect the copyright of manga.

Our future work includes (1) increasing the recall and precision of the suspicious manga detection, (2) extracting ROIs by other region detectors, (3) increasing the database of mangas.

Acknowledgment

This research was supported in part by the Grant-in-Aid for Scientific Research (B)(22300062) from Japan Society for the Promotion of Science (JSPS).

References

[1] W. Sun and K. Kise, “Detecting Printed and Handwritten Partial Copies of Line Drawings Embedded in Complex Backgrounds.”, International Conference on Document Analysis and Recognition, pp. 909–919, 2009.

[2] P. Viola, M. Jones. “Robust Real-Time Face Detection”, International Journal of Computer Vision vol. 57(2), pp. 137–154, 2004.

[3] W. Sun and K. Kise, “Similar Partial Copy Detection of Line Drawings Using a Cascade Classifier and Feature Matching”, International Workshop on Computational Forensics, pp. 121–132, 2010.

[4] J. Matas, O. Chum, M. Urban and T. Pajdla, “Robust Wide Baseline Stereo from Maximally Stable Extremal Regions”, British Machine Vision Conference, pp. 384–393, 2002.

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

[6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, vol. 60(2), pp.91–110, 2004.

[7] S. Arya, D. Mount, R. Silverman and A. Y. Wu, “An Optimal Algorithm for Approximate Nearest Neighbor Searching”, Journal of the ACM, 45, 6, pp.891–923, 1998.

[8] C. Manning and H. Schtze, “Foundations of Statistical Natural Language Processing”, MIT Press, ISBN: 0262133601, 1999.

[9] R. Baeza-Yates and B. Ribeiro-Neto, “Modern Information Retrieval”, ACM Press, ISBN: 020139829, 1999.

[10] J. Sivic and A. Zisserman, “Video Google: A Text Retrieval Approach to Object Matching in Videos”, IEEE International Conference on Computer Vision, pp. 1470–1477, 2003.

[11] G. Csurka, C. Bray, C. Dance and L. Fan, “Visual Categorization with Bags of Keypoints”, ECCV Workshop on Statistical Learning in Computer Vision, pp. 59–74, 2004.

[12] M. Rusiñol and J. Lladós, “Logo Spotting by a Bag-of-words Approach for Document Categorization”, International Conference on Document Analysis and Recognition, pp. 111-115, 2009.