

# Improvement of Trademark Retrieval Using Pseudo Relevance Feedback

## ABSTRACT

As the number of registered trademarks become larger, the demand for trademark retrieval systems with high accuracy has grown to reduce the burden of human searchers. In particular, precision at high recall is important to avoid erroneous registration of newly submitted trademarks. In this paper we propose a method to realize this by using pseudo relevance feedback, which is a method to search again by using an expanded query formed based on top ranked results of the first search. From the experimental results with the database of about 620,000 trademarks, we have achieved the improvement of precision by 4% at 100% recall as compared to the method without the feedback.

## CCS CONCEPTS

• information systems → Image search;

## KEYWORDS

Trademark search, Local features, Large scale database, Reference point, Rejection processing

## ACM Reference Format:

. 1997. Improvement of Trademark Retrieval Using Pseudo Relevance Feedback. In *Proceedings of ICVGIP (ICVGIP'18)*. ACM, New York, NY, USA, 8 pages. [https://doi.org/10.475/123\\_4](https://doi.org/10.475/123_4)

## 1 INTRODUCTION

The workload of the trademark examination increases because more than 100,000 trademarks are applied each year, and naturally, the number of registered trademarks is increasing year by year. Trademark examination is the task of judging the similarity between a newly submitted trademark and registered trademarks in three points: similarities of appearance, appellation, and concept. Currently, it is impractical to evaluate similarity for all registered trademarks. The searchers narrow down registered trademarks by keyword search on the basis of elements associated with a newly submitted trademark and then judge similarity one by one visually. However, because of a large number of registered trademarks, visual comparison or inspection is the burden of the trademark examiner. Also, since the search word used in the keyword search is determined by the examiner subjectively, search omission occurs for the similar trademark not associated with the search word.

In order to solve this problem, it is necessary to make a trademark retrieval system with high accuracy. In particular, the apparent similarity is the most fundamental, which covers many cases. Therefore,

in this paper, we focus on the similarity of the appearance and proceed with the discussion.

Generally, the most important requirement for creating a trademark retrieval system is to prevent omission of search in order to avoid erroneous registration. In other words, high recall is required. Under this condition, it is necessary to reduce the retrieval of dissimilar images in order to save the workload of examination. In other words, high precision is preferable under high recall.

A lot of researchers have studied trademark retrieval systems of appearance similarity. Most of these systems are based on image information, and the methods of using local features are on the mainstream. The methods using local features focus on local information of images and find the correspondence between parts of images. However, since the correspondence does not completely represent the similarity of trademarks, retrieved images are sometimes incorrect. So recall of the methods using local features is high, but precision tends to be low. Therefore, various methods for improving precision have been proposed in recent years. One of them is "rejection of feature points based on reference point" proposed by Klinkigt et al. [11] (Hereinafter, the reference point is abbreviated as RP).

Klinkigt's method is the method to recognize a specific object in the database of natural images. In recent years, it has also been applied to the field of character recognition by Matsuda et al. [14]. In order to recognize characters in natural images, they used "rejection of feature points based on affine transformation" in addition to "rejection of feature points based on RP", so that precision is improved without decreasing recall and processing speed. However, Klinkigt's method has been used only for natural images and Matsuda's method has been used only for character images, and it is not known whether they work effectively for line graphics such as trademarks. In our previous method, We applied Klinkigt's method and Matsuda's method into the method using local features and they have proved that these methods work effectively with line graphics such as trademarks. Nevertheless, our previous method has a precision of 62.5% when the recall is 100%, and it still has room for improvement.

To further improve the precision, we apply a technique called pseudo relevance feedback [1][5][12][13] to our previous method, where pseudo relevance feedback performs automatic feedback by assuming that the top  $N$  cases of retrieval results are relevant to the query. From the experimental results with the database of about 620,000 trademarks, our proposed method with pseudo relevance feedback succeed in improving precision by 4 percent points at 100% recall as compared with the method without the feedback.

## 2 RELATED STUDIES

### 2.1 Trademark retrieval method

There are two kinds of methods to retrieve trademarks by using images. Some use global features and the others use local features.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted by ACM, provided that the fee of \$12.00 is paid directly to ACM. This permission is granted without fee for individuals and small organizations that are registered with ACM. This permission does not extend to other kinds of copying, such as for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
ICVGIP'18, December 2018, Hyderabad, India  
© 2016 Copyright held by the owner/author(s).  
ACM ISBN 123-4567-24-567/08/06.  
[https://doi.org/10.475/123\\_4](https://doi.org/10.475/123_4)

For example TRADEMARK system [10] proposed in 1992 uses global shape features such as the spatial frequency and contrast of the image. ARTISAN [2] proposed in 1996 is a prototype system that conducts trademark image search at the United Kingdom Patent Office. As image features, global shape features such as the area of a polygonal convex hull of boundary and the length of the longest boundary chord are used. As mentioned above, the trademark retrieval methods in the 1990s have mainly used global features. However, it is not possible to evaluate partial similarity only by measuring the shape similarity of a whole image. Therefore, recall and precision of methods using global features tend to be low. For this reason, in recent years, trademark retrieval methods with local features are on the mainstream because they can measure partial similarity. SIFT feature is the most popular local feature because it is robust against image rotation, scale change, and illumination change. Wang et al. [15] and Fu et al.[3] proposed search methods using SIFT features. These methods have contributed to the retrieval of similar images for trademarks. However, they have not achieved sufficient precision yet. In addition, the number of images in the dataset they used for evaluation was only 3, 621 and 12, 000, respectively. It is not a practical setting because the number of real trademarks is enormous.

## 2.2 Pseudo relevance feedback

Pseudo relevance feedback has been mainly used for improving accuracy of document retrieval [1][5][12][13]. In recent years it has also been used for improving the accuracy of image retrieval[4][7]. Among them, Joly's method [8] is one example of applying modified pseudo relevance feedback to a trademark retrieval and determined the images used as feedback with a threshold value. This method worked effectively for some queries, but there was a problem that it does not work well for queries with a small number of features. Their method was also experimented only on 10, 000 trademark images dataset, whereas we conduct on 620, 000 trademark images.

## 3 BASELINE METHOD

In this paper, we will verify the effectiveness of applying pseudo relevance feedback to our previous method (Hereinafter, our previous method is called the baseline method). In the baseline method, the flow of the whole process is divided into "storage phase" and "search phase" as shown in Fig. 1. "storage phase" is to store trademark images in the database. Local features extracted from stored trademark images (hereinafter referred to as database images) are registered in the database. "Search phase" is composed of four processes. Details of each process are described below.

### 3.1 Local feature extraction

We extract local features from trademark images both in Storage phase and Search phase as shown in Fig. 1. In this research, we use the SIFT features, which have the parameters of "orientation" and "scale" of a local region in addition to an information of the brightness value on the image. It is not necessary to use the information on "orientation" and "scale" in past researches, but it plays a central role in RP processing.

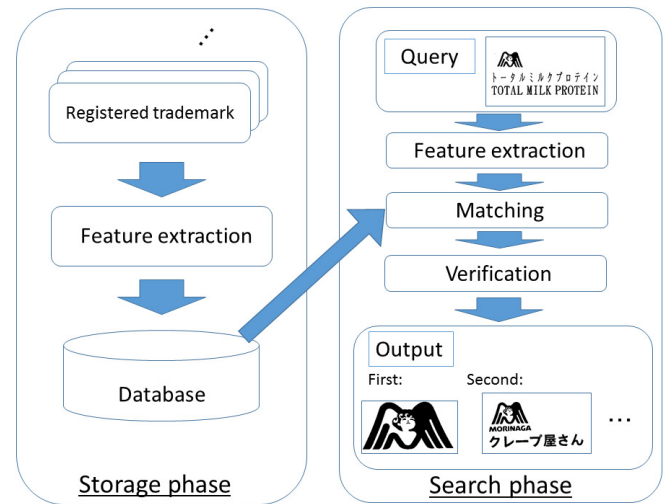


Figure 1: A flow of trademark retrieval processing

### 3.2 Matching process

We search for similar local features in the database to those obtained from a query image. The similarity between two local features is based on the distance between vectors of local features. However, when calculating distances to all the local features (the number of  $n$ ) obtained from the query image for all the local features (the number of  $m$ ) in the database, a long processing time of  $O(mn)$  is required. Therefore we use an approximate nearest neighbor search method to reduce processing time. The approximate nearest neighbor search method reduces the number of candidates which the distance is calculated by using approximate distance values. We use the Bucket Distance Hashing method [6], which is one of the most efficient approximate nearest neighbor search methods. This method selects the nearest neighbor candidates by using the multi-sequence algorithm without comparing data. See [6] for details. When we use this method, we need to determine parameters  $k$  and  $C$ , where  $k$  is the parameter of  $k$ -nearest neighbor and  $C$  is the number of local features which are roughly searched by approximate distances.  $C$  local features are searched by approximate distances, and the result is set up to the upper  $k$  neighborhood among them. We use these optimal parameters that are determined in preliminary experiments.

### 3.3 Verification process

In the matching process, the matching is not always suitable from the viewpoint of a human visual system because the matching is simply based on the distance of local features. For example, erroneous matching may occur when a simple pattern repeatedly appears on various trademarks. Figure 2 shows examples of erroneous matching in a simple pattern. As shown in Fig.2, the SIFT features in database images match the SIFT feature in the query image although they are not similar. To solve this problem, we consider the information of the arrangement of matched SIFT features in addition to individual matchings. A database image is recognized as a similar image to the query image if a sufficient number of

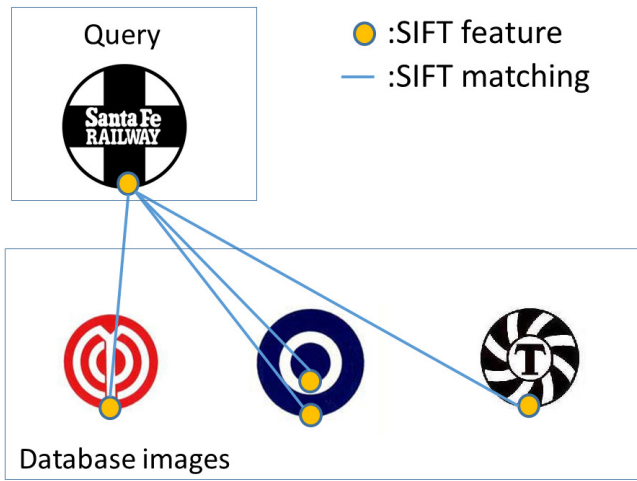


Figure 2: Examples of erroneous matching in the simple pattern

matched features remains even after rejection which is the process of removing low reliability of matching features.

In order to improve the accuracy of SIFT matching, “rejection of feature points based on RP” and “rejection of feature points based on affine transformation” are used. Each of these is described below.

3.3.1 *Rejection of feature points based on RP.* This process consists of the following four steps. Steps 1 and 2 are shown in Fig. 3, and step 3 is shown in Fig. 4.

- (1) First, we define the center of the database image as RP and find the direction and distance to the RP of each local feature.
- (2) Next, we project the local features of the database image on the query image.
- (3) After projection, we adjust the scale and the orientation of the all local features on the query image.
- (4) Finally, we calculate the coordinates of the RP on the query image from the vector to the RP.

When the above processing is applied to all the matched local features, the RP is calculated at one place on the query image if the database image has a similar part to a part of the query image. However, the calculated coordinates of RPs do not always indicate one place due to the estimation errors of scale and orientation. Therefore, when  $n$  points or more exist in a circle with radius  $r$ , it is assumed that they are concentrated in one place. Simultaneously, local features outside the circle are rejected.

By this process, correspondence relationships of similar parts are obtained between the query image and the database image. At this time, as shown in Fig. 5, the similar parts with different arrangement can be detected, where each similar part in the query image has individual RP. Therefore, by individually evaluating the local features belonging to each RP, it can be determined whether the part corresponding to the RP in a query image is similar to a part in a database image. While, when there is no RP, the database

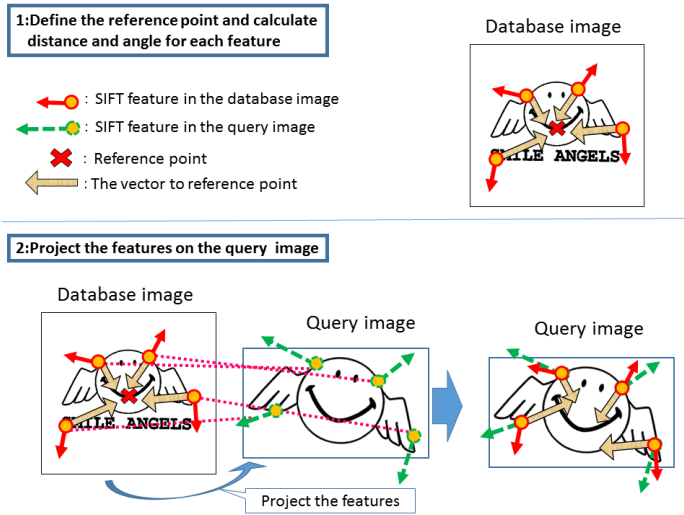


Figure 3: A method of removing low reliability of matching features with reference points in first step and second step

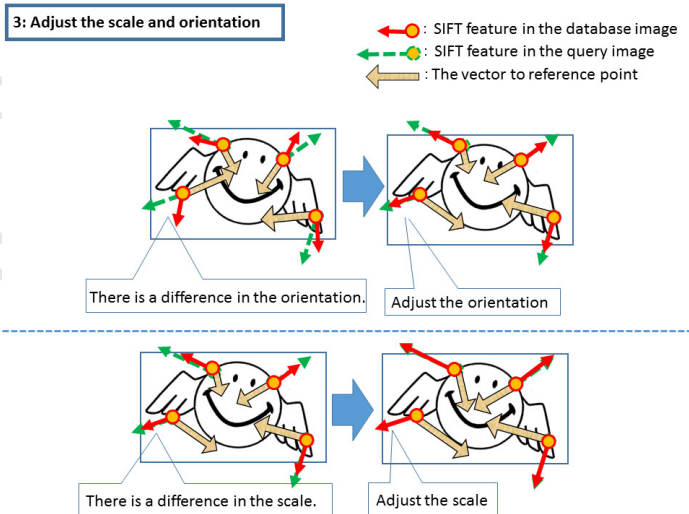
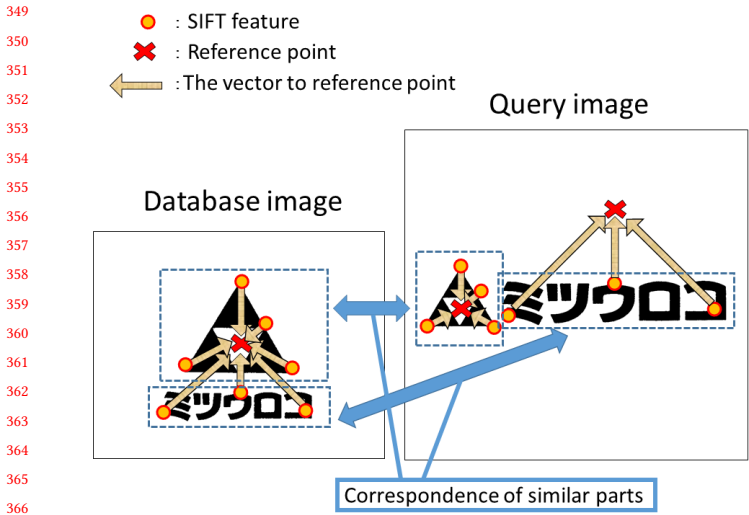


Figure 4: A method of removing low reliability of matching features with reference points in third step

image is rejected which means it is regarded as a low similarity image.

3.3.2 *Rejection of feature points based on affine transformation.* In order to investigate the reliability of the similar part for each RP obtained in the “rejection of feature points based on RP”, it is verified whether the positional relation of the local features is totally correct or not. First, an affine transformation matrix  $M$  is obtained for RP calculated on the query image, where the affine transformation is the transformation from the local features on a database image to those on the query image. Next, using the calculated affine transformation matrix  $M$ , the verified local features on the database image are transformed into the query image, and

291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348



**Figure 5: Detection of similar parts with different arrangement in an image by reference points**

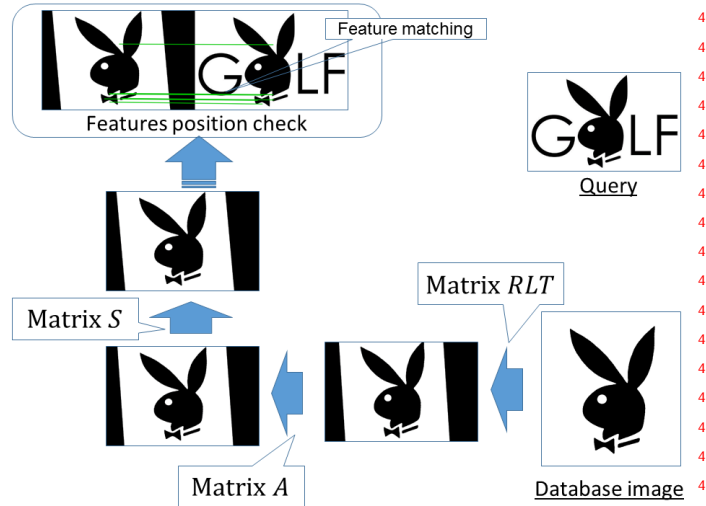
the positional deviation of each local feature is measured. The local feature is rejected if the positional deviation is larger than the threshold  $r'$  because it is regarded as an erroneous matching. In addition, after the rejection process, database images with less number of local features than  $n'$  are rejected.

While, the affine transformation matrix  $M$  can be decomposed into five matrices: translation  $T$ , scale  $L$ , rotation  $R$ , independent scaling  $A$  and shear  $S$ .

$$M = SARLT \quad (1)$$

The process of obtaining matrix  $M$  sequentially calculates the geometric transformation parameters in the order of translation  $T$ , scale  $L$ , rotation  $R$ , independent scaling  $A$  and shear  $S$ . The outline is shown in Fig. 6. will be described below. See [14] for details.

The translation  $T$  is obtained by RP correspondence. About the scale  $L$ , the distances between the RP and the local feature is obtained from both of the query image and the database image. And we obtain the scale  $L$  by taking the ratio of the distances. Since there are many pairs of RP and local features, the median value of the ratios obtained from the combination of all RP and local features is used as the estimated ratio. About rotation  $R$ , we obtain the line angles between RP and the local feature with the query image and the database image, and we calculate the difference of those angles. Likewise, the average of obtained rotation angles is used as the estimated rotation angle because a number of pairs of RP and local features are often obtained. Through the above, three components of translation  $T$ , scale  $L$ , rotation  $R$  are obtained. At this stage, these transformations are applied to the database image. The components of the geometric transformation remaining after this are independent scaling  $A$  and shear  $S$ . For independent scaling  $A$ , scales are measured separately in the horizontal direction and the vertical direction from the pair of RP and local features. Then, the ratios of the horizontal scales and the vertical scales between the query image and the database image are obtained, and



**Figure 6: Matrices used for affine transformation**

the median value of those is taken to be the estimated scale. About shear  $S$ , we focus on the horizontal shear since there are many deviations in the horizontal component. The variable  $s$  of shear  $S$  is calculated as  $s = -(x'_s - x_s)/y_s$  because the shear is represented as  $x'_s = x_s + sy_s$ , where  $(x_s, y_s)$  and  $(x'_s, y'_s)$  are the coordinates before and after shearing, respectively. The median value of the variables  $s$  for all combination of RP and the corresponding local features is used as the estimated variable of shear.

### 3.4 Output process

All the database images matched with a query image are retrieved. Basically, they are ranked by similarity, which is the number of matched features after rejection. We retrieve the rejected database images after the not rejected database images. The similarity  $p$  used for ranking database images that were not rejected is determined as follows:

$$p = f / \sqrt{q} \quad (2)$$

where  $f$  is the number of local features after “rejection of feature points based on affine transformation”,  $q$  is the total number of local features extracted from the database image. The ranks of database images with a larger number of local features tend to be higher. Therefore, we apply the normalization by  $\sqrt{q}$  to prevent the tendency. The ranking of rejected database images is also ranked by similarity shown below.

$$p' = f' / \sqrt{q} \quad (3)$$

where  $f'$  is the number of local features before “rejection of feature points based on RP”.

## 4 PSEUDO RELEVANCE FEEDBACK

Pseudo relevance feedback, also known as blind relevance feedback, is the retrieval method to expand queries. First, we retrieve the top  $N$  ranked images by using the original query image. And then we use them as queries as shown in Fig. 7. By using this method, it

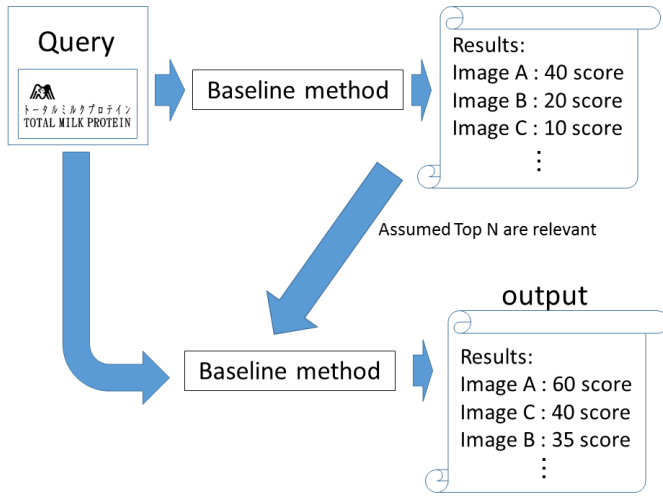


Figure 7: Pseudo relevance feedback

is possible to retrieve relevant images missed in the first retrieval in order to improve the performance. This method is easy to implement as it automatically determines the query image. But if the images of the top  $N$  are not relevant images, the performance may be degraded.

We apply this method to the baseline method and calculate the similarity  $p''$  as follows:

$$p'' = \alpha \frac{b}{\sqrt{q}} + (1 - \alpha) \frac{\sum_{j=1}^N d_j}{\sqrt{q}} \quad (4)$$

where  $b$ ,  $d_j$  and  $\alpha$  are the number of local features after "rejection of feature points based on affine transformation" by using original query image, the number of local features left without being rejected by using the top  $j$  ranked images, and the weight of the query image, respectively.

## 5 EXPERIMENT

In this paper, in order to verify the effectiveness of pseudo relevance feedback, we conducted a comparison with the baseline method. In the experiment, recall and precision were used as performance measures. When calculating recall and precision, 11 point interpolated average precision [9] was applied. 11 point interpolated average precision is the interpolated precision of looking at 11 recall levels (0.0, 0.1, 0.2, ..., 1.0). We averaged these scores across all of the different queries to find the score.

The experiment consists of the following two steps. First, we investigated the relationship between the accuracy and the number  $N$  of images used as the feedback, where  $N = 1, 3, 5$ . Next, we compare the proposed method with the best accuracy with the baseline method.

### 5.1 Normalization of images

Trademark images were normalized by the following procedure as shown in Fig. 8.

- (1) First, the binary image of a trademark image is generated.

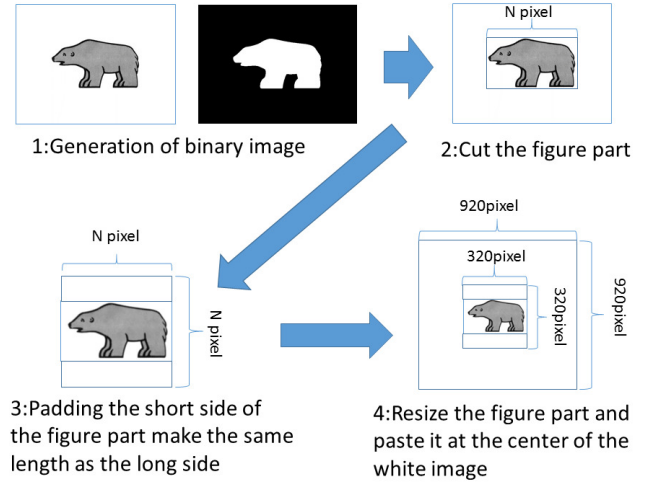


Figure 8: Image normalization

- (2) Next, we cut out the figure part of the binary image.
- (3) Then, in order to resize the clipped region without changing its aspect ratio, white pixels are padded so that the short side of the region may have the same length as the long side.
- (4) Finally, the normalized image is obtained by adding the appropriate margin surrounding the figure part.

The reason why the margin should surround the figure part is to extract reliable SIFT features. Using the normalization of images, an image of  $920 \times 920$  pixels was created with figure part at  $320 \times 320$  pixels. The resolution was determined based on the result of preliminary experiments.

### 5.2 Dataset

We collected 627, 326 trademark images registered in Japan Patent Office and used them as the dataset. 623, 101 images were used as database images, and the remaining 4, 225 images were used as query images, which have 59, 654 relevant images. Three trademark searchers and 28 students determined visually relevant images in database images to an image in query images one by one. In this work, they decided relevant images by judging whether there is the same shape on a part or whole of an image without considering differences in resolution or in color.

### 5.3 Parameters

Fixed parameters in the experiment are as follows.

- The number of local features which are roughly searched by approximate distances  $C = 10, 000$
- The parameter of  $k$ -nearest neighbor  $k = 50$
- Radius used in the rejection of feature points based on RP  $r = 100$
- Threshold of the positional deviation used in the rejection of feature points based on affine transformation  $r' = 100$
- Threshold of the number of the feature points used in the rejection of feature points based on RP  $n = 4$

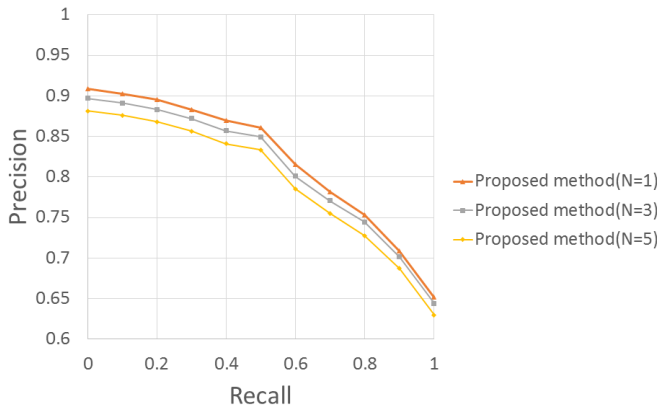


Figure 9: 11 point interpolated average precision by pseudo relevance feedback

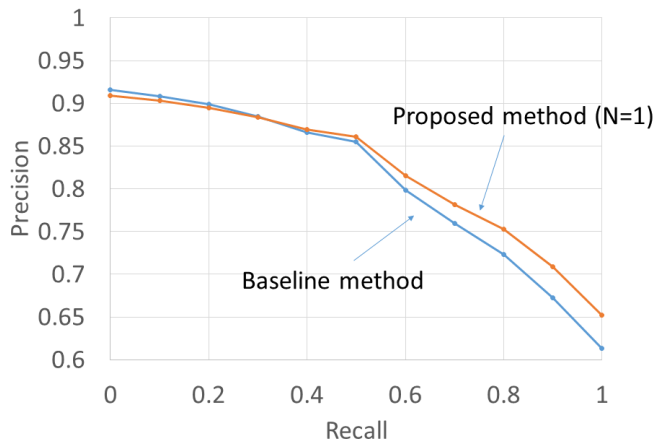


Figure 10: 11 point interpolated average precision for comparing proposed method (N=1) with the baseline method

- Threshold of the number of the feature points used in the rejection of feature points based on affine transformation  $n' = 4$
- The weight of pseudo relevance feedback  $\alpha = 0.5$ .

These parameters are fixed with the highest computational speed and accuracy based on the preliminary experiments.

### 5.4 Result

First, we investigated parameters with the highest accuracy by pseudo relevance feedback. The retrieval result of each method is shown in Fig. 9. The proposed method ( $N = 1$ ) was the highest precision at all recall. Then, most images used as feedback were



Figure 11: An example in which the accuracy was improved by feeding back a relevant image to the query image. The upper row shows the query image, the middle row shows the retrieval result of the baseline method and the lower row shows the retrieval result of the proposed method ( $N = 1$ ). An image surrounded by a red rectangle is the relevant image.

similar to the corresponding query images, which were 3523 out of 4225 top ranked images (83%). As shown in Fig. 9, the performance of the proposed methods ( $N = 3, 5$ ) is inferior to that of the proposed method ( $N = 1$ ). It seems that this result was caused by adding dissimilar images in feedback.

Second, we compared the proposed method ( $N = 1$ ) with the baseline method as shown in Fig. 10. The precision was improved by 3.90 percent points when the recalls of both methods were 100%.

An example of improvement by the proposed method ( $N = 1$ ) is shown in Fig. 11. In this example, the baseline method retrieved five relevant images, while the proposed method ( $N = 1$ ) retrieved

10 relevant images. Moreover, all the top eight retrieval images in the proposed method ( $N = 1$ ) were similar to the query image.

In addition, the precision of the baseline method was 0.6 percent points higher than the proposed method ( $N = 1$ ) when the recalls of them were 0%. This is because a dissimilar image was the top ranked image for some queries. As shown in Fig. 12, we can see that the ranking of relevant image went down from second to 21st because the feedback image (the top ranked image) was not a relevant image. Since it is necessary to prevent the omission of search at the trademark examination, precision at recall 100% is the most important. Therefore, we can show the good effects of the method with pseudo relevance feedback added to the baseline method.

In addition, we will show the good and bad retrieval examples in detail. Fig. 13 shows an example with the improved the retrieval accuracy. In this example, the top ranked image of the baseline method is similar to a part of the query image. Then, the proposed method ( $N = 1$ ) retrieved more relevant images. On the other hand, Fig. 14 shows an example of the worsened accuracy. In this example, the proposed method ( $N = 1$ ) could not retrieve images representing "aluminum" in Japanese in the upper part. The reason for this is that the top ranked image did not include the word "aluminum" although the query image included it. This lowered the rank of the pattern. However, it is not so common case. Therefore, it was found that the accuracy was improved by using pseudo relevance feedback.

## 6 CONCLUSIONS

In this paper, we add "pseudo relevance feedback" to our previous method. In order to confirm the effectiveness, we conducted two experiments. In the first experiment, we investigated how many images are needed as feedback in order to get the highest accuracy. As a result, the method using the top ranked image as feedback got the best performance. Next, we compared the proposed method ( $N = 1$ ) with the baseline method. In the second experiment, we have achieved the improvement of precision by 4 percent points at 100% recall as compared with the baseline method.

Future work includes to implement a retrieval method that can apply the similarities of appellation and concept to our trademark retrieval method.

## REFERENCES

- [1] Guihong Cao, Jian-Yun Nie, Jianfeng Gao, and Stephen Robertson. 2008. Selecting good expansion terms for pseudo-relevance feedback. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 243–250.
- [2] John P Eakins, JM Boardman, and K Shields. 1996. Retrieval of trade mark images by shape feature-the ARTISAN project. (1996).
- [3] Jianlong Fu, Jinqiao Wang, and Hanqing Lu. 2010. Effective logo retrieval with adaptive local feature selection. In *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 971–974.
- [4] Guo-Dong Guo, Anil K Jain, Wei-Ying Ma, and Hong-Jiang Zhang. 2001. Learning similarity measure for natural image retrieval with relevance feedback. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 1. IEEE, 1–1.
- [5] Djoerd Hiemstra, Wessel Kraaij, Renée Pohlmann, and Thijs Westerveld. 2000. Translation resources, merging strategies, and relevance feedback for cross-language information retrieval. In *Workshop of the Cross-Language Evaluation Forum for European Languages*. Springer, 102–115.
- [6] Masakazu Iwamura, Tomokazu Sato, and Koichi Kise. 2013. What is the most efficient way to select nearest neighbor candidates for fast approximate nearest



Figure 12: An example in which the accuracy was reduced by feeding back a dissimilar image to the query image.

- neighbor search?. In *Proceedings of the IEEE International Conference on Computer Vision*. 3535–3542.
- [7] Feng Jing, Mingjing Li, Hong-Jiang Zhang, and Bo Zhang. 2004. Relevance feedback in region-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology* 14, 5 (2004), 672–681.
- [8] Alexis Joly and Olivier Buisson. 2009. Logo retrieval with a contrario visual query expansion. In *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 581–584.



Figure 13: An example in which the accuracy was improved by feeding back an image almost similar to the query image.



Figure 14: An example in which the accuracy was reduced by feeding back a relevant image to the query image.

[9] K. Tuda K. Kita and M. Shishibori. 2002. *Jyohoukensaku arugorizumu*. Kyoritsu Shuppan, 20–21.

[10] Toshikazu Kato. 1992. Database architecture for content-based image retrieval. In *image storage and retrieval systems*, Vol. 1662. International Society for Optics and Photonics, 112–124.

[11] Martin Klunkigt and Koichi Kise. 2011. Using a reference point for local configuration of sift-like features for object recognition with serious background clutter. *IPSP Transactions on Computer Vision and Applications* 3 (2011), 110–121.

[12] Kyung Soon Lee, W Bruce Croft, and James Allan. 2008. A cluster-based resampling method for pseudo-relevance feedback. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 235–242.

[13] Yuanhua Lv, ChengXiang Zhai, and Wan Chen. 2011. A boosting approach to improving pseudo-relevance feedback. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 165–174.

[14] Takahiro Matsuda, Masakazu Iwamura, and Koichi Kise. 2014. Performance improvement in local feature based camera-captured character recognition. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on*. IEEE, 196–201.

[15] Zhenhai Wang and Kicheon Hong. 2012. A novel approach for trademark image retrieval by combining global features and local features. *Journal of Computational Information Systems* 8, 4 (2012), 1633–1640.