Accuracy Improvement and Objective Evaluation of Annotation Extraction from Printed Documents

Tomohiro Nakai, Kazumasa Iwata, and Koichi Kise
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai, Osaka, 599-8531 Japan
{nakai, iwata}@m.cs.osakafu-u.ac.jp, kise@cs.osakafu-u.ac.jp

Abstract

There is an approach of annotation extraction from printed documents in which annotations are extracted by comparing the image of an annotated document and its original document image. In one of the previous methods, the image of an original document is actually printed and scanned in order to reproduce image degradations of the image of the annotated document. However such a method lacks convenience since users have to use the same printer and scanner to obtain images of an annotated document and its original document. In this paper, we propose an improved annotation extraction method in which the image degradations are compensated by image processing. In the proposed method, the difference between original and annotated document images due to image degradations is reduced by not only removal of the degradations in the annotated document images but also reproduction of the degradation in the original document images. The proposed method consists of three steps of processing which are for dithering, for color change, and for local displacement. We also propose an objective evaluation of extracted annotations to compare the experimental results accurately. Experimental results of the proposed method have shown that the recall of extracted annotations was 80.94% and the precision was 85.59%.

1. Introduction

It is common for us to write annotations such as notes and corrections directly onto documents. Since annotations reflect user's interest and knowledge, user profiles can be obtained by extracting and analyzing annotations.

There is an approach of annotation extraction in which annotations are obtained by subtracting a document image without annotations (original document image) from an image of an annotated document (annotated document image). For a method of such an approach, even small differences between those two images other than annotations can cause noises. Displacement of images and image degradations caused during printing and scanning are the major factors of noises. Therefore they have to be removed in order to extract annotations successfully.

Nakai et al. have proposed an annotation extraction method [1] which has an alignment process to solve the displacement of images. The problem of image degradations is avoided by use of original document images which are obtained by actual printing and scanning. However, it has a defect that a pair of an original and an annotated document image has to be obtained with the same printer and scanner to equalize image degradations. It also has a problem that small displacement remains in each local area since images are aligned globally based on a single parameter.

In this paper, we propose an annotation extraction method in which the problems of image degradations and displacement are solved by image processing. The problem of image degradations is dealt with equalization process of an original and an annotated document image. Since the image degradation caused by printing and scanning is compositive, several steps of image processing are performed with respect to each type of degradation. The problem of displacement is solved by a local alignment process. It deals with local displacements which remain after global alignment. These processes improve accuracy of annotation extraction.

We also propose a system of objective evaluation of extracted annotations. In the previous work [1], extracted annotations are evaluated visually and subjectively. However, annotation extraction methods cannot be compared fairly without objective evaluation criteria. In the proposed system, a ground truth image is prepared for each annotated document image. Extracted annotations are evaluated based on comparison with its ground truth image.

2. Related work

Methods of annotation extraction from paper documents can be classified into two types. One is methods which do not require the original document images to extract annotation [2, 3, 4, 5]. Annotations are extracted only from the annotated document images based on colors and shapes of annotations. The other is methods which require the original document images [1]. Annotations are extracted by subtracting the original document image from its annotated document image. In the methods which do not require the original document images, annotations are extracted based on their colors and shapes. Therefore they have restriction on color and shape of annotations while annotations can be extracted accurately. On the other hand, methods which require the original document images have no restriction on annotations. However, availability of the original electronic documents is necessary. Moreover, various kinds of image degradation on the annotated document images have to be dealt with in order to extract annotations without noises.

Researches of image degradations caused by printing and scanning include removal of displacements of images and modeling of the pixel value distortions. In [1], an alignment method of document images is proposed. In this method, images are aligned based on correspondences of feature points. Since feature points are matched using similarity transformation invariants, similarity transformed images can be aligned. However, the method also has a problem that slight misalignment can remain in local areas of an image since the whole image is aligned using a similarity transformation parameter. A method of modeling the variation of luminance values of pixels is proposed in [6]. The model is created based on observation of how each pixel value changes before and after printing and scanning. We consider the original and annotated document images can be equalized using the method.

3. Proposed method

In the proposed method, image degradations and displacements of gray-scale original and annotated document images are compensated using image processing to improve accuracy of annotation extraction. Figure 1 shows the overview of the proposed method. Before applying the procedures of the proposed method, global alignment is performed using the alignment method of [1]. Then three kinds of image degradations, dithering, pixel value distortions, and local displacements are compensated one by one. Annotations are extracted as difference between the compensated original and annotated document images. Extracted annotations are evaluated by comparing each pixel of ground truth and difference images.

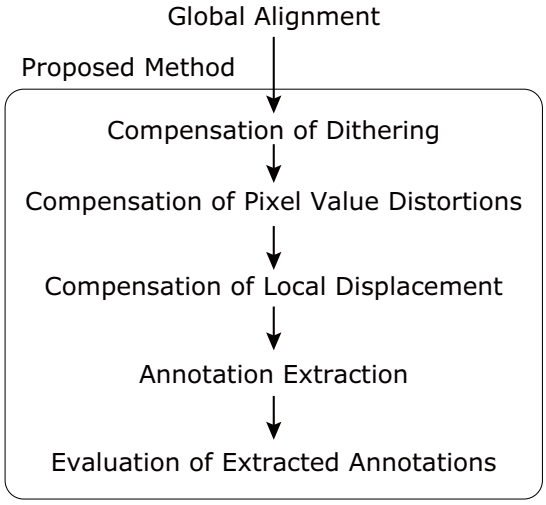


Figure 1. Overview of the proposed method.

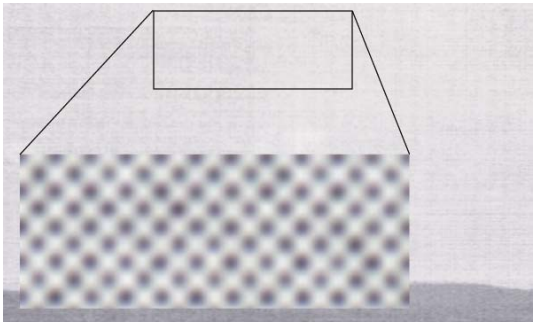
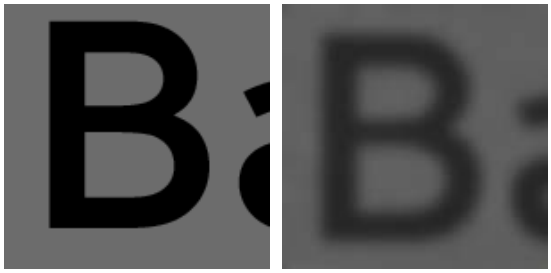


Figure 2. An example of dithering.



(a) Original document image. (b) Smoothed image.

Figure 3. A smoothed image differs from its original document image since it suffers from blurring.

3.1. Dithering compensation

Dithering is a technique to describe a full-color image with a limited color palette. Since a printer has a limited

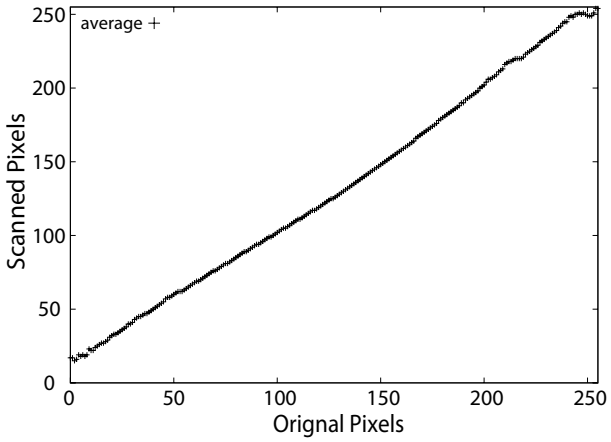


Figure 4. An example of pixel value distortions.

number of inks, medium colors of printed documents are represented as patterns of dots as shown in Fig. 2. In the pixel level, they are different from medium colors of the original document image. Therefore simply subtracting the original document image from the dithered annotated document image could result in noises in areas of medium colors. In order to eliminate noises, dithering of the annotated document images has to be compensated.

We introduce smoothing of the annotated document images to remove dithering. However, while smoothing blurs dot patterns, it also causes differences between the original and annotated document images. As shown in Fig. 3(a) and Fig. 3(b), contours of characters and patterns are also blurred. In order to equalize the original and annotated document images, comparable smoothing is also applied to the original document images.

3.2. Compensation of pixel value distortions

The annotated document images suffer from pixel value distortions. They are affected by printing conditions and sensitivity of scanners. Dithering also affects the pixel value distortions since pixel values of smoothed dot patterns are not necessarily equal to those of the original document images.

Due to the pixel value distortions, pixel values of the annotated document images are different from those of the original document images. It makes annotation extraction more difficult. In the proposed method, annotations are extracted by thresholding difference values of the original and annotated document images. Under more significant pixel value distortions, larger thresholds are needed in order to reduce noises. However, larger thresholds increase risk of losing annotations when annotations and their background

Figure 5. Each patch of an image is aligned separately.

have similar colors. Therefore the pixel value distortions have to be compensated in order to make annotation extraction easy.

In the proposed method, the pixel value distortions are compensated using a distortion model based on [6]. The pixel value distortion model shows how each pixel value changes before and after printing and scanning. Figure 4 shows an example of the model. It is created by investigating how pixel values change in pairs of the original and annotated document images. An average value of corresponding pixel values in the annotated document image is calculated for each pixel value of the original document image. Correspondences of pixels are determined as follows.

In the proposed method, the original document image and the annotated document image are aligned using the proposed method [1] in advance. A pair of pixels at the same coordinates in those images is considered to be corresponding. However, slight displacements can remain in local areas. In that case, pixel value changes caused by printing and scanning cannot be calculated since irrelevant pixels are corresponded. In the proposed method, paired pixels are not taken into account when they have difference larger than a certain threshold.

3.3. Compensation of local displacements

Since the annotated document images go through printing and scanning, they suffer from displacements. Displace-

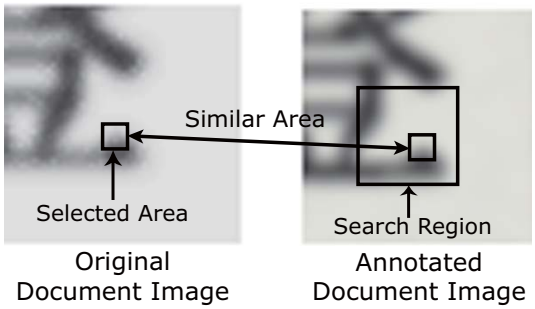


Figure 6. Alignment based on matching of selected areas.

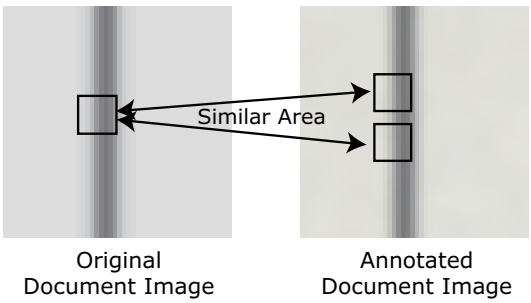


Figure 7. An example of erroneous corresponding.

ments make annotation extraction difficult since annotations are extracted by subtraction. In the previous method [1], the original and annotated document images are aligned based on features invariant to similarity transformation. However, aligned images still have slight displacements in local areas since whole images are aligned globally using a single similarity transformation parameter. That has harmful effects on accuracy of extracted annotations. In the proposed method, compensation of local displacements is introduced in order to improve extraction accuracy.

Alignments are performed by translating the original document images to cancel displacement. Due to distortion including curvature of paper, degrees of displacement vary by part of an image. Therefore it is inappropriate to use a common parameter of alignment for the whole image. In the proposed method, the input image is divided into several patches as shown in Fig. 5. An alignment parameter is calculated separately for each patch.

An alignment parameter of a patch is calculated based on pixel patterns of local areas. As shown in Fig. 6, a local area is selected from a patch of the original document image. Then the most similar area is retrieved from the search region set in the corresponding patch of the annotated document image. A difference of coordinates is calculated from the similar selected areas. An alignment parameter of a patch is a median of the differences.

Since an alignment parameter is calculated for each patch, a discontinuity of the parameter can occur at a boundary between patches. It can cause a gap of image pixels at the boundary. However, the translation of patches is limited because the original and the annotated document images are roughly aligned before this process. Therefore it is unlikely that the discontinuity of the parameter causes a serious problem. Flexible subtraction in annotation extraction described later also reduces the effect by the discontinuity.

Retrieving all local areas requires too much computational load. Besides, it is difficult to make correspondences of pixel patterns correctly in blank (white) or filled (black) areas due to indistinguishable patterns. Therefore only local areas of pixels at edges are retrieved in the proposed method. This is because pixels at edges have distinct pixel patterns in their neighbors.

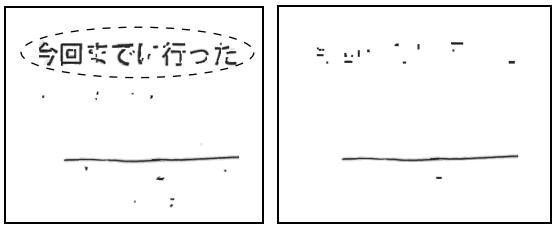
However, as shown in Fig. 7, pixels at edges can also have pixel patterns which are difficult to make correspondence. Therefore patches sometimes have erroneous alignment parameters. In the proposed method, erroneous parameters are detected and screened out as follows. Parameters are obtained for all patches shown in Fig. 5. It is unlikely that neighboring patches have greatly different parameters. Consequently, a parameter which is too much different from those of neighboring patches is considered to be irrelevant. Instead, the median of parameters of patches on the left, right, top and bottom is used. Thus irrelevant parameters caused by erroneous correspondences can be removed.

3.4. Annotation extraction

In the proposed method, annotations are extracted by subtracting the original document image from the annotated document image. A pixel of the annotated document image which has a larger difference than a threshold is extracted as a pixel of annotations.

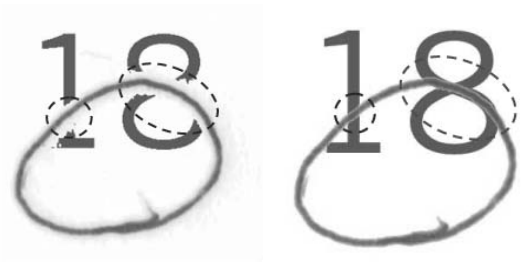
A major problem of annotation extraction is that images suffer from slight displacements even after the local alignment. In order to reduce noise caused by displacements as shown as dotted ellipse in Fig. 8(a), the pixel which corresponds to a pixel of the annotated document image is retrieved from a local region of the original document image. When the smallest difference of pixel values in the local region is larger than a threshold, the pixel of the annotated document image is extracted as that of annotations. As a result, annotations with reduced noise as shown in Fig. 8(b) can be extracted.

Such noise reduction also has a problem that not only noise but also annotations can be lost. When annotations



(a) Before compensation. (b) After compensation.

Figure 8. Compensation of noise caused by slight displacements.



(a) Before correction. (b) After correction.

Figure 9. Handling of overlapped pixels.

are close to printed characters, a pixel of annotations can correspond to that of printed characters and be lost since they have a smaller difference. In the proposed method, based on the idea that there are lost annotations in neighbors of detected annotations, neighboring pixels of annotations are also extracted in order to restore annotations.

3.5. Objective evaluation of extracted annotations

We also propose a system of objective evaluation of extracted annotations. In the proposed evaluation system, annotations are evaluated by comparing an image of extracted annotations and a ground truth image of annotations. Therefore ground truth images which include only annotations of the annotated document images are needed. If annotations were separated from the annotated document images, they would be used as a ground truth image. However, it is difficult to separate annotations precisely from the annotated document images without excessive amount of manual labor. Therefore we have to take an alternative approach for acquisition of ground truth images.

In the proposed evaluation system, we take an approach in which a ground truth image is not obtained from the annotated document image but a ground truth image is prepared in advance. The annotated document image is synthesized by overlaying the ground truth image to a printed and scanned original document image (scanned document image). Note that such a way to acquire the ground truth images and the annotated document images is taken only to evaluate annotation extraction methods. In the actual use of the proposed method, annotations are extracted from the annotated document images.

3.5.1. Ground truth image

As we mentioned above, annotations are separated in advance in the proposed system. The annotated document images are created by combining ground truth images and scanned images. Ground truth images and annotated document images have to be as natural as possible. While the use of tablets for annotation enables easy separation of ground truth images, it is difficult to simulate feathering and uneven ink. Writing annotations in a blank sheet also has a problem that it is difficult for annotations to reflect content of documents. In order to solve the above problems, not ordinary paper but half transparent tracing paper is used.

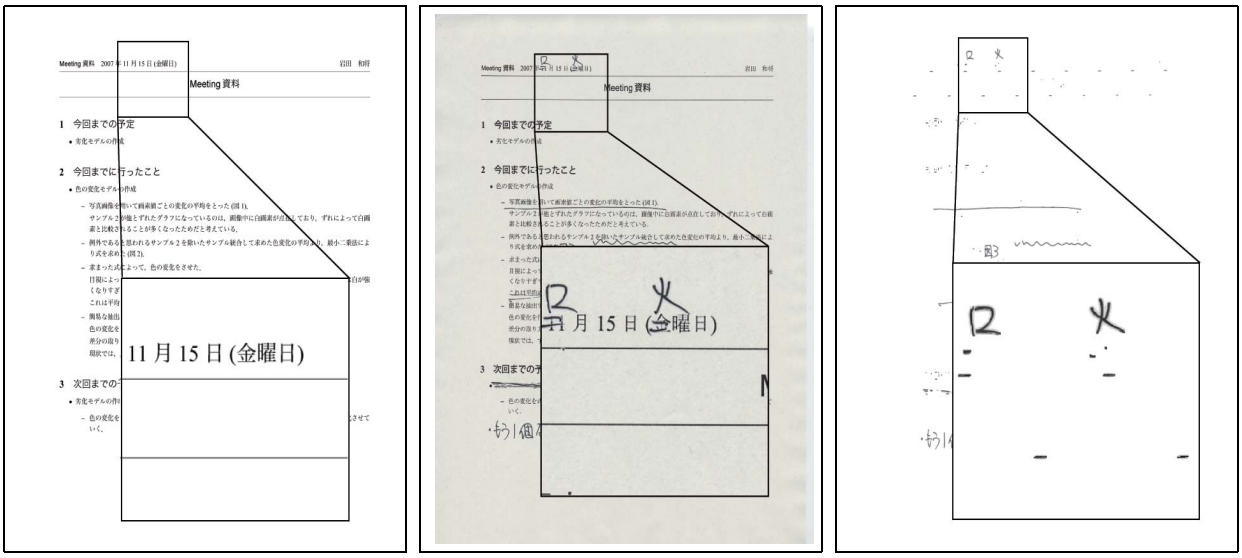
A ground truth image is created as follows. First, a sheet of tracing paper is lapped over a printed document. Annotations are written in the sheet of tracing paper. Then a ground truth image is created by scanning the annotated sheet.

3.5.2. Annotated document image

Annotated document images are synthesized from scanned images and ground truth images as follows. Each pixel of a ground truth image is classified into foreground (annotation) and background pixel. Then foreground pixels replace pixels of the scanned image. Since the ground truth image includes only annotations, an annotated document image can be obtained by writing positive pixels of the ground truth image over the scanned image. Such an approach involves two following problems.

The first problem is how to distinguish foreground pixels from background pixels. Since ground truth images are obtained by scanning, even background pixels have slight positive values due to contamination of paper and scanners. Therefore foreground and background pixels are determined by comparing them to a predetermined threshold. The threshold is set to filter out pixels of contaminations and not to lose annotations.

The second problem is how to deal with overlaps of printed characters of scanned images and annotations of ground truth images. As shown in dotted ellipses of Fig. 9(a), printed characters can be pruned away if foreground pixels of ground truth images displace pixels of



(a) Original document image. (b) Annotated document image. (c) Extracted annotations.

Figure 10. Examples of document images used in the experiments.

scanned images. Therefore weighted addition of foreground pixels of a ground truth image and pixels of a scanned image is used as the annotated document image. Figure 9(b) shows an example of corrected version of the annotated document image.

3.5.3. Objective evaluation of extracted annotations

Extracted annotations are evaluated by comparing to the ground truth images. Recall R and precision P are utilized for evaluation. R and P are calculated as

$$R = \frac{A}{C}, \quad P = \frac{A}{B},$$

where A is the number of pixels of extracted annotations which match those of the ground truth image, B is the number of pixels in the image of extracted annotations, C is the number of pixels in the ground truth image. Note that A, B and C do not include background pixels. Higher R means more sufficient annotations and higher P means less noise.

Pixels of extracted annotations and the ground truth image are considered to be matched when pixel values of all channels exactly match. Since all images have the same sizes and there is no factor of color change, exact matching can be used. Only foreground pixels are counted since most part of documents are occupied by background pixels. When background pixels are counted, they dominate recall and precision and near meaningless evaluation can be obtained.

Table 1. Experimental results for all images.

	Recall	Precision
Previous method	68.28%	73.28%
Proposed method	80.94%	85.59%

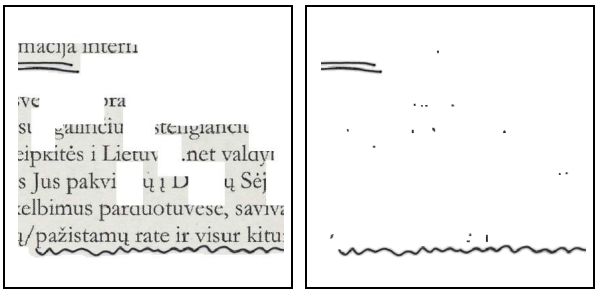
Table 2. Experimental results for successfully aligned images.

	Recall	Precision
Previous method	67.92%	79.16%
Proposed method	80.97%	91.74%

4. Experimental results

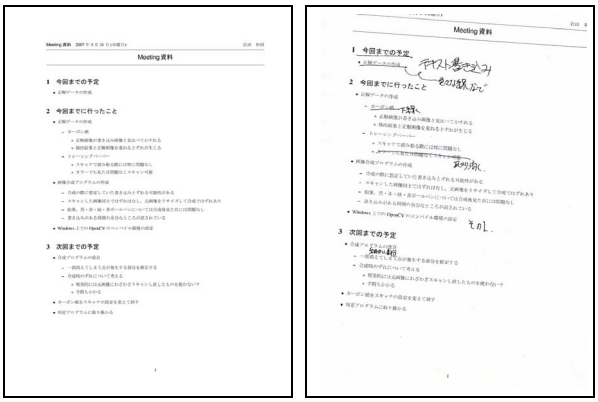
We compared extracted annotations using the proposed method and the previous method [1]. In the experiments, 100 pages of PDF documents were used. Ground truth images of annotations were obtained by scanning annotations for each document. Original document images were converted from the PDF files at 200 dpi using *convert* command of Image Magick. Annotated document images were synthesized from the ground truth images and images of the documents printed on A4 paper and scanned at 200 dpi. Figure 10 shows examples of the original document images, the annotated document images and extracted annotations.

The proposed method has several parameters. They af-



(a) Previous method. (b) Proposed method.

Figure 11. An example of accuracy improvement.



(a) Original document image. (b) Aligned annotated document image.

Figure 12. An example of unsuccessful global alignment.

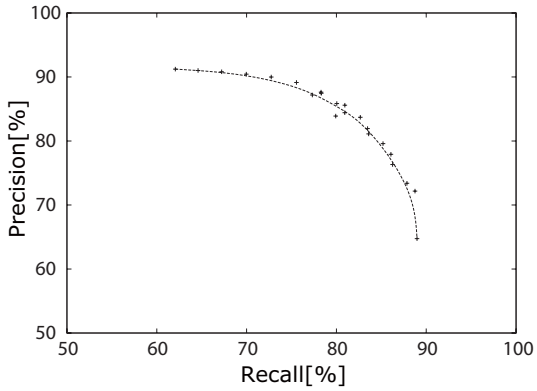
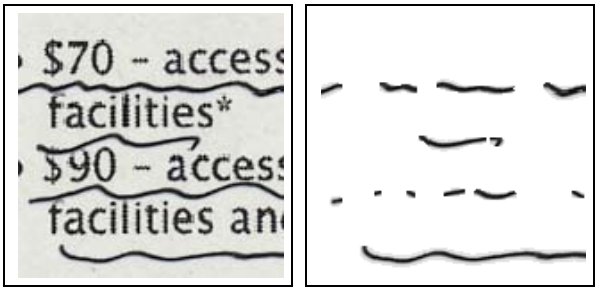


Figure 13. Relation between recall and precision.



(a) Annotated document image. (b) Extracted annotations.

Figure 14. An example of lost annotations.

fect recall and precision of extracted annotations. In this experiment, various values of two parameters mentioned in **3.4**, size of a local region for retrieval of a corresponding pixel and size of an extraction region for restoration of annotations, were examined. Table 1 shows experimental results of the proposed method with the highest F-value and those of the previous method. The proposed method realized 12.7% higher recall and 12.3% higher precision than the previous method. Figure 11 shows an example of extracted annotations with accuracy improvement. Reduced noise as shown in Fig. 11 realized the improvement of recall.

The results shown in Table 1 include those of erroneously aligned images as shown in Fig. 12. In the proposed method, the global alignment is assumed to be almost correct. Thus its failure ruins results of annotation extraction. In order to evaluate performance of the proposed method clearly, experimental results only for successfully aligned images are shown in Table 2. The global alignment was succeeded in 90 out of 100 images. In the case with these images, the recall of the proposed method was 91.74%. It indicates that annotations can be extracted accurately on the condition that the global alignment is succeeded.

The average processing time of the proposed method was about 30 seconds. Since that of the previous method was about 15 seconds, the proposed method took almost double time. This is because the proposed method includes additional procedures for compensation of image degradations.

Figure 13 shows transition of the recall and the precision when the two parameters were varied. As shown in Fig. 13, the recall and the precision are in trade-off relation. Therefore there can be two approaches to realize both high recall and high precision using the proposed method. One is to extract annotations using a recall oriented parameter and filter out noise afterward. The other is to extract annotations

using a precision oriented parameter and restore lost annotations afterward. We consider the latter approach promising. This is because restoration of annotations could be easier than removal of noise since extracted annotations can be used as clues to lost annotations. As shown in Fig. 14, a part of annotations can be lost when a precision oriented parameter is used. In this case, the lost annotations are close to the remaining annotations. Therefore the lost annotations can be retrieved by searching neighbors of the extracted annotations. Restoration of annotations from precise annotations enables both high recall and high precision.

5. Conclusion

In this paper, we proposed an improved annotation extraction method and objective evaluation system of extracted annotations. Accuracy improvement of annotation extraction is realized by compensation of image degradations. In the objective evaluation system, annotations are evaluated by being compared to a ground truth image pixel by pixel.

Experimental results of the proposed method have shown that the recall of extracted annotations was 80.94% and the precision was 85.59%. Compared to the previous method [1], accuracy was improved 12.7% at recall and 12.3% at precision.

Future work includes further improvement of recall. It can be realized by adding a procedure of restoration of lost annotations.

References

[1] T. Nakai, K. Kise, and M. Iwamura. A method of annotation extraction from paper documents using alignment based on local arrangements of feature points. In *Proceedings of 9th International Conference on Document Analysis and Recognition (ICDAR2007)*, pages 23–27, Sept. 2007.

[2] D. Möri and H. Bunke. Automatic interpretation and execution of manual corrections on text documents. In P. S. P. W. H. Bunke, editor, *Handbook of Character Recognition and Document Image Analysis*, pages 679–702. World Scientific, Singapore, 1997.

[3] J. Stevens, A. Gee, and C. Dance. Automatic processing of document annotations. In *Proceedings of 1998 British Machine Vision Conference*, volume 2, pages 438–448, 1998.

[4] J. K. Guo and M. Y. Ma. Separating handwritten material from macine printed text using hidden markov models. In *Proceedings of 6th International Conference on Document Analysis and Recognition*, pages 436–443, 2001.

[5] Y. Zheng, H. Li, and D. Doermann. The segmentation and identification of handwriting in noisy document images. In *Lecture Notes in Computer Science (5th International Workshop DAS2002)*, volume 2423, pages 95–105, 2002.

[6] Ching-Yung and S.-F. Chang. Distortion modeling and invariant extraction for digital image print-and-scan process. In *International Symposium on Multimedia Information Processing(ISMIP99)*, 1999.