

Text Image Enhancement in Scenery Images for Degraded Character Recognition using DCT

Graduate School of Information Science & Engineering,
Tokyo Institute of Technology

Hiroki Takahashi and Masayuki Nakajima

rocky@img.cs.titech.ac.jp

Abstract: This paper proposes a method to enhance text images for a system which extracts, recognizes and translates multi-lingual characters in scenery images captured by a digital camera. The proposed method magnifies text images in frequency domain. It restores high-frequency components by a DCT(Discrete Cosine Transform) based approach with an estimated enlarged image, and reduces mosquito and block noises caused by JPEG(Joint Photographic Experts Group) compression in the enhanced process. The obtained enhanced text images are binarized and then recognized by a commercial character recognition software.

Experiments are performed for printed documents, sign boards and plates captured by a digital camera. In our experiments, texts are manually extracted from images because our goal in this paper is image enhancement. Compared with traditional approaches, the recognition ratio for our enhanced images by using a commercial character recognition software improves.

Keywords:

DCT(Discrete Cosine Transform), Image enhancement, Binarization, Degraded image, Character recognition

1 Introduction

Mobile computers have been made advanced in the functionality and been downsized rapidly. It becomes easier to write documents or take images anywhere. Moreover, as cellular phones have been widely spread, it becomes possible to send your current surrounding situation both in documents and images by connecting a computer with a cellular phone. On the other hand, video cameras and digital cameras have also been downsized and spread since it is enough cheap to get them. Recently, a digital camera has become one of media which record various information because it can record images and sound at the same time easily without leaking information. Furthermore, mobile computers or cellular phones with cameras are being sold though their cameras are with low resolution. In such situations, it is expected that digital cameras will play a role of new stationery by extracting various information from images, although the conventional functionality of them has been to record images. Such kinds of new instruments can be useful for image/video indexing and database management, tourist navigation, e-learning with translation systems and so on.

The authors have proposed a system which extracts texts from images and recognizes them[1, 2]. In the paper[1], candidates of text regions are extracted using two features of edge and color in images, respectively. Text regions are detected by integrating both of the candidates based on a few conditions regarding characteristics of texts. There-

after, the extracted text regions are enlarged by using bi-linear interpolation and are binarized by discriminant analysis method. 70.3% of 148 characters is recognized by using a commercial character recognition software which employs a new augmented cell method. On the other hand, the paper[2] extracts Hangul texts from scenery images by using characteristics of Hangul structure. Although Hangul has more than 3,000 kinds of characters, each character is composed of 24 basic Hangul alphabets which are 14 basic consonants and 10 basic vowels, and there are only 8 kinds of arrangements in those Hangul alphabets. The paper also recognizes Hangul alphabets from skeletons of the extracted patterns based on the simplicity of Hangul structures. As a result, 85.97% of characters are extracted from 100 images which are taken on arbitrary conditions. In the recognition process, it recognizes 71.96% of Hangul alphabets which are not concatenated each other. Both of these recognition ratios are not enough, because the images taken by digital cameras are not sufficient quality for character recognition. They are low resolution compared with conventional printed characters, lighting conditions are also ununiform. Moreover, the images have block noise and mosquito noise, since usual digital cameras compress them by using JPEG(Joint Photographic Experts Group).

In this paper, a method to enhance text images based on DCT(Discrete Cosine Transform) is proposed. The resolution of the images is low and they include noise caused by

JPEG compression. The purpose is to improve recognition ratio with conventional commercial character recognition softwares for enhancing text images in frequency space. The proposed method restores high-frequency components by a DCT based approach with an estimated enlarged image. Moreover, it reduces mosquito and block noises caused by JPEG compression in the enhanced process. In our experiments, texts are manually extracted from images because our goal in this paper is image enhancement.

The following section gives a literature of related researches. Section3 explains characteristics of scenery images. The overview of the proposed method is presented in section4. In section5, removal of mosquito noise as the pre-process and binarization method as the post-processing are discussed. The proposed method to expand character images is explained in detail in section6. The experimental results are shown in section7. Section8 concludes the paper with future work.

2 Related Work

Characters captured by digital cameras become low quality. Because they are caused by capturing with low resolution cameras, noise from uncleanness of dust, lighting conditions under different weather or time, internal and external camera parameters, and geometrical transformations of characters themselves, e.g. text on a cup. The quality of characters are affected by a variety of factors. Many robust recognition algorithms have been studied for low quality characters in a decade. The conventional approaches are divided into three categories[3]. The first one enhances degraded characters in pre-process which employs noise reduction, binarization and so on. The second one gives robust discriminant space for image degradation. The last one proposes adaptive recognition methods for degraded characters according to their quality. The target characters are also widely spreaded, e.g., title or telop in video[4], signboard or traffic sign in scenery images[5], back title on books[6], car number plates, number on containers, Web pages and so on.

Most of all previous methods for low resolution character recognition proposed robust discriminant criteria for noisy characters. In this paper, we focus on character enhancement for low resolution character recognition on signboards in scenery images. It has been proposed an interactive system to recognize and translate texts which represent attention, guidance or explanation included in images[5]. In this system, 91.7% of 971 characters are recognized. It is, however, hard to apply for low resolution images. Because most of all experimental characters are large enough to be recognized by a conventional character recognition software. On the other hand, some character recognition methods, especially for degraded characters in images taken by digital cameras or in video, are proposed[6][7][8][9]. In these degraded character recognition, robust binarization algorithm for low resolution and low contrast images are investigated and adaptively dic-

tionary selection method for low quality characters are proposed[6]. M. Mori et. al.[7] proposed a robust character recognition method for degradation of character edges caused by binarization process proposed by Kuwano et. al.[10] and background noise. These approaches aim to propose robust character recognition algorithm for degraded characters. They don't aim to enhance characters adapted for conventional character recognition algorithm. Our goal is to enhance an target degraded image. Especially in the researches related on texts extraction and recognition from video, recognition method using other information such as closed caption are proposed for video retrieval or video summarization[4]. R. Lienhart et. al. proposed automatic character segmentation in digital videos[11] and recognized the extracted characters by using their own OCR(Optical Character Reader) software based on a feature vector classification approach. They only deal with text added to the video artificially such as pre-titles, credit titles, closing titles and so on. In text enhancement method, J. Kosai et. al.[8] proposed to obtain directional features of characters from low resolution images. They uses images shifted by a half pixel in the low resolution and gain the feature values for high resolution images. It is, however, very difficult to capture appropriate images for this algorithm with a handy camera. In this paper we propose a text enhancement method in frequency domain with reduction of noise caused by JPEG compression.

Image magnification methods are roughly divided into two approaches. One is based on image registration[12], [13] which is the process of two or more overlaying images of the same scene taken at different times, from different viewpoints, and/or by different sensors. The other approach is single image interpolation[14], [15]. Although image registration can obtain much information from multiple images, these images must be shifted in sub-pixel. In single image interpolation, nearest neighbor, convolution of image samples with a single kernel, i.e. bilinear, bicubic, cubic kernel, directional method[16] and orthogonal transform methods[17],[18] are proposed. Convolution methods cannot recover the high-frequency components lost or degraded during the sampling process. Directional methods interpolates in the low-frequency direction(along the edge) rather than in the high-frequency direction(across the edge). Shinbori et. al. [17] employ iterative Gerchberg-Papoulis algorithm with DCT(Discrete Cosine Transform). The method restores high-frequency components lost in the sampling process using two constraints that the spatial extent of the image is finite and correct information is already known for the low-frequency components. It reduces error and gets high quality images compared with Nearest Neighbor, bilinear and cubic convolutions. Kawaguchi et. al.[19] extend the DCT-based image magnification method to reduce block noise and ringing effects by using overlapping blocks. The method mainly focuses on magnification of images with line components and improves Signal-Noise ratio. In this paper, we extend this method for degraded character image magnification.

3 The characteristic of scenery text images

There are distinctive problems different from those of character recognition of document images scanned by a scanner, in character recognition for scenery images taken by a digital camera. Light source quite changes, contrasts are ununiform and noise such as white noise sometimes occurs, compared with images taken by a scanner. It occurs image distortion caused by internal and external distortion of a camera. Images taken by a digital camera are mostly compressed by JPEG. As JPEG compression is lossy compression, high-frequency components of an original image is lost, and mosquito noise around edges or block noise caused by block encoding occur. Mosquito noise is distortion at the edges of objects and further characterized by its spatial characteristics. Sometimes, it appears pseudo-edge around the high-frequency components of the images or edges. Block noise leads to intensity jumps at neighboring block boundaries. It is necessary to deal with various problems as discussed above for character recognition of scenery images taken by digital cameras.

In this paper, a method to enhance text images taken by digital cameras is proposed. Especially, it reduces the noise occurred by JPEG compression and improves the recognition ratio for low resolution images by enlarging them in frequency space. In this paper, it assumes that experimental images are taken under the following conditions. Illumination on a text plane is ideally uniform. Texts and background are homochromatic. An optical axis of a camera is almost perpendicular to the text plane. There is few rotation of the text around the optical axis. Moreover, the camera system is assumed as ideal one. Besides, the experimental text images are segmented manually since the goal of this paper is enhancement of text images.

4 Process overview

The proposed procedure is illustrated in **Fig.1**. At first, text regions are manually segmented from an image. Then, the segmented text image is transformed into gray scale image by using this formula $I = 0.299R + 0.587G + 0.114B$. Maximum and minimum values of intensity are determined to perform contrast transformation for reducing mosquito noise as a pre-process. After that, the image is enlarged with DCT. In this image enlargement procedure, high-frequency component is restored by predicting high-frequency component which is lost by JPEG compression and sampling in digitizing the image. Later, an enlarged image is generated by maximizing the contrast with maximum and minimum values of intensity obtained in the pre-process. Finally, binarization is performed by using discriminant threshold analysis method, and the characters are recognized by a commercial software.

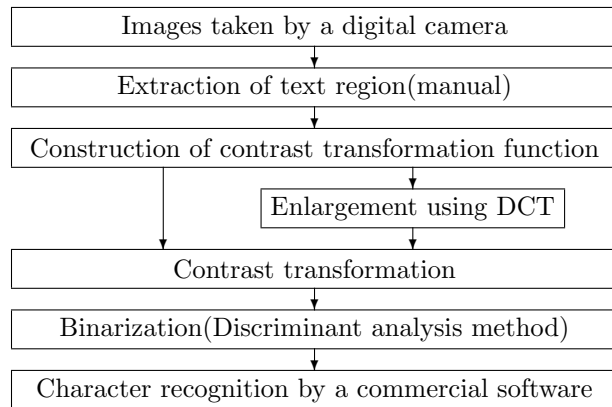


Figure 1: Process overview of image enhancement.

5 Pre- and Post-processing

In this section, a mosquito noise reduction process for JPEG compressed images using contrast transformation.

The target images of this paper are character images so that these images are finally binarized to input for a character recognition software. Therefore, contrast transformation is performed by using linear contrast transformation function to reduce the influence of mosquito noise. As shown in **Fig.2**, first of all, an image is segmented into character and background regions by using discriminant analysis method[20]. Next, distribution of intensities for each region is obtained. For example, if intensities of background region is bigger than that of character region, intensities are transformed by the following formula(1).

$$I = \begin{cases} 0 & \text{if } I_{org} \leq I_{min} \\ \frac{I_{org} - I_{min}}{I_{max} - I_{min}} & \text{if } I_{min} < I_{org} < I_{max} \\ 255 & \text{otherwise} \end{cases} \quad (1)$$

$$I_{min} = \mu_t + \sqrt{v_t}, \quad I_{max} = \mu_b - \sqrt{v_t} \quad (2)$$

where, I denotes an intensity after contrast transformation. I_{org} , I_{min} and I_{max} represent intensities of an original image, minimum and maximum intensities, respectively. μ_t and μ_b indicate average intensities of character and background region, respectively. v_t and v_b are variances of character and background region.

6 Enlargement of text images using DCT

In this section, a method to enlarge text images using DCT is explained. The proposed method modifies the Kawaguchi's method[19], whose method enlarges images including line drawings. The enlargement process also involves noise reduction such as block noise and mosquito noise.

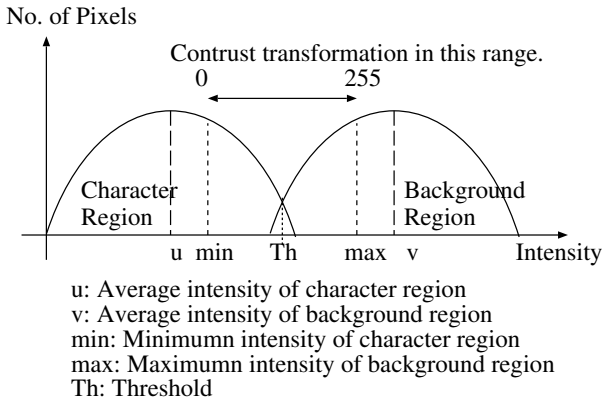


Figure 2: Contrast transformation for mosquito noise reduction.

The procedure of enlargement is illustrated in **Fig.3** and is described as followings;

1. **DCT:** DCT is performed on the blocks of $N \times M$ pixels and blocks of $N \times M$ DCT coefficients are obtained.
2. **Enlargement of frequency component:** Frequency bandwidth is expanded a times according to the power of enlargement. The high-frequency components in enlarged DCT coefficients are assumed to zero.
3. **IDCT(Inverse DCT):** IDCT is performed on the block of $aN \times aM$ DCT coefficients. The $n \times n$ enlarged blocks of images are combined into one block whose size is $anN \times anM$ pixels.
4. **Addition of region restricted images:** aw pixels around the block of $anN \times anM$ pixels are added to the block. Then we obtain the block of $a(w + nN) \times a(w + nM)$ pixels. The block is called region restriction image.
5. **DCT:** DCT is performed to the region restriction images.
6. **IDCT:** w pixels around the block of original $nN \times nM$ pixels are added to the block. The block of $(w + nN) \times (w + nM)$ pixels is obtained.
7. **DCT:** DCT is performed to the block of (6). The obtained DCT coefficients are represent to low-frequency components of the enlarged DCT coefficients of (5).
8. **Correction of low-frequency components:** The correct coefficients obtained in (7) are replace to the low-frequency components in enlarged block of (5).
9. **IDCT:** IDCT is performed to the DCT coefficients and enlarged image is obtained.

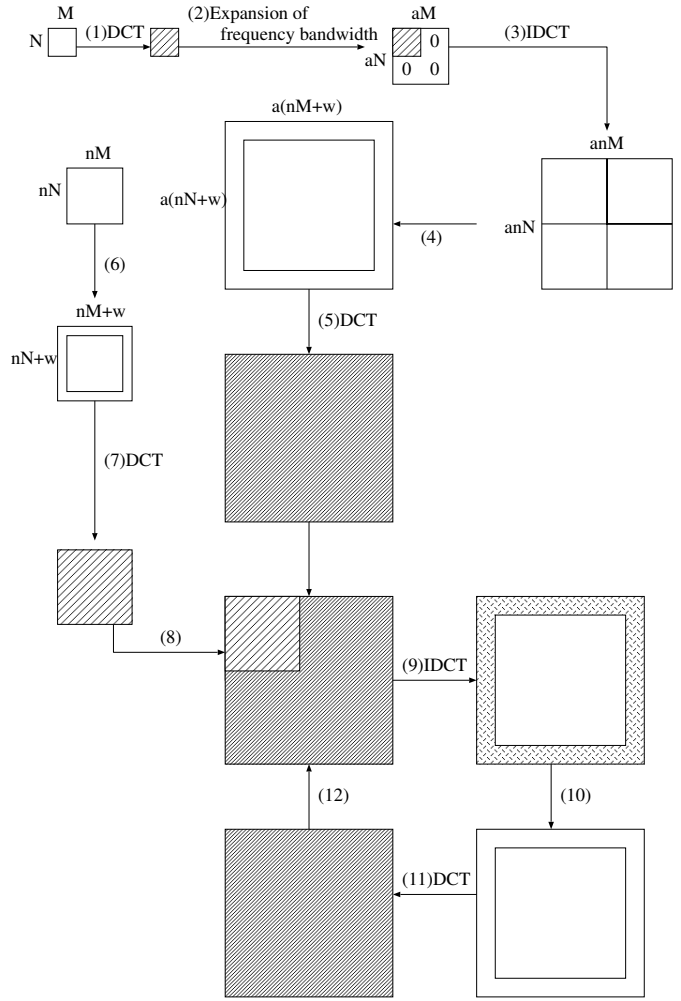


Figure 3: The proposed enlargement process.

10. **Region restriction:** The pixels of the block boundary are replaced to the original values used in (4).
11. **DCT:** DCT is performed to the block of (10).
12. **Iteration of the procedure:** The correct coefficients obtained in (7) are replace to the low-frequency components in enlarged block of (11). Repeat the procedures from (7) to (12), if necessary.

The procedures from step(9) to (12) are repeated several times. The enlarged block is obtained from the block in step(9) by removing the bounding pixels.

The differences from the previous methos[19] are as follows; In our method, estimated enlarge images are employed as region restricted images. It is possible for the proposed method to reduce mosquito noise occurred by the boundary between blocks in order to add bounding pixels. Mosquito noise has a feature that the longer the base length is, the wider it occurs. For that reason, the proposed method overlaps blocks when a block of $anN \times anM$ pixels is generated, i.e. in step(3). On the other hand, in

the restoration of high-frequency components, the blocks are not overlapped. This is because the bounding pixels play a role of the overlapped blocks. Moreover, the computation time of the proposed method is smaller than that of the previous method[19]. Because, the base length of the proposed method is shorter than that of the previous one in the restoration process of high-frequency components.

7 Experimental Results

In this section, text images taken by a digital camera are enlarged using the proposed method and then the enlarged images are recognized by a commercial character recognition software.

It is hard to recognize the text images extracted from the images taken by a digital camera correctly because most of all conventional character recognition algorithm for printed characters target texts in documents under ideal conditions such as uniform lighting, black characters on a white paper and so on. Moreover, recognition ratio of the conventional character recognition algorithms depends greatly on resolution of the target character images. In this paper, the availability of the current character recognition techniques is, therefore, discussed by enlarging a text images extracted from images taken by a digital camera.

The character recognition software used in this paper employs new augmented cell algorithm. As a preliminary experiment, the document shown in Fig.4 are recognized by the character recognition software. It is scanned by a scanner with varies resolution. The recognition results are illustrated in Table.1. The algorithm requires at least 18×18 pixels, even if the image is taken by a scanner.

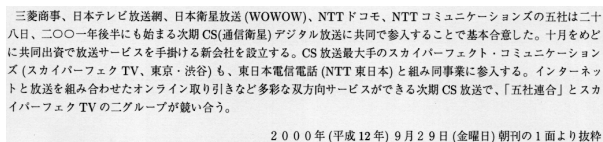


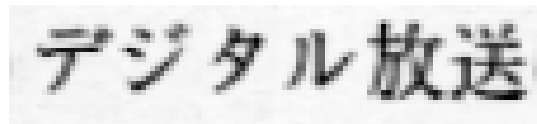
Figure 4: A text image taken by a scanner for preliminary experiments(18×18 pixels per character).

Table 1: Recognition ratio of text image taken by a scanner.

size per character	recognition ratio(%)
6×6 pixels	0%
12×12 pixels	0%
14×14 pixels	19.1%
16×16 pixels	55.2%
18×18 pixels	93.8%

7.1 Recognition of text images in documents

First of all, as a basic experiment for enlargement, the same documents used in the preliminary experiment are taken by a digital camera. Based on the results of the preliminary experiment, the sizes are set to 16×16 pixels, 20×20 pixels and 24×24 pixels per character. Fig.5 shows some parts of images taken by a scanner and a digital camera. Compared with the image taken by a scanner, the image taken by a digital camera is blurred and includes outstanding mosquito noise. An image taken by a digital camera is shown in Fig.6(a). The images illustrated in Fig.6(b) and (c) are enlarged 2×2 by bi-linear interpolation and the proposed method. The recognition ratios are shown in Table.2. Compared with the bi-linear interpolation, the proposed method improves the recognition ratios dramatically. In the experiments, block size of the original image is 8. N and M are set to 4, respectively. The placement of $N \times N$ pixels block is not related with the original blocks. The iteration number is set to one from the preliminary experiment.



(a) A text image taken by a scanner.



(b) A text image taken by a digital camera.

Figure 5: Comparison of image qualities(16×16 pixels per character).

Table 2: Recognition ratios for enlarged images(%)

size per character (pixels)	bi-linear		
	2×2 times	3×3 times	4×4 times
16×16	29.6	20.7	20.0
20×20	32.4	33.3	29.3
24×24	63.8	60.6	54.3
size per character (pixels)	proposed		
	2×2 times	3×3 times	4×4 times
16×16	48.1	54.1	57.0
20×20	71.4	74.1	74.1
24×24	89.8	88.2	90.6

7.2 Recognition of text images in scenery images

Some examples of images, of which character regions are segmented manually from scenery images, are illustrated in Fig.7. Table.3 shows the recognition results for enlarged 3×3 by bi-linear interpolation and the proposed method, and a binarized image with no enlargement. The proposed method gets superior recognition ratio compared with bi-linear interpolation.

Table 3: Recognition ratios.

(a) Enlarged Fig.7(a) by 3×3 .

	Recognition results	ratio
bi-linear	母が丘池区 Ma haeku 4m	30.0%
proposed	技が丘池区 Mnarbaaka A Q	45.0%
no enlargement	~Na~b~h~m	15.0%

(b) Enlarged Fig.7(b) by 3×3 .

	Recognition ratio	ratio
bi-linear	所牝地 月脱上申 X 脚 2-12-1 股針的 台口吉日!! 即控'l 三 股和 7 年'19321	48.6%
proposed	所神地 目照区大岡山 2 · 12-1 磁針著 谷口吉郎 滋数年 昭和 7 年 (1932)	77.1%
no enlargement	所 f1 三池 日似 K 火' 町 1 川 2-12-1 股す rh 谷口吉郎 軸 'l 三 昭和 7 年 (1932)	60.0%

(c) Enlarged Fig.7(c) by 3×3 .

	recognition ratio
bi-linear	26.9%
proposed	63.5%
no enlargement	19.2%

8 Conclusions

Text images taken by a digital camera are reduced noise caused by JPEG compression, enlarged by restoring high-frequency components in DCT, binarized by discriminant analysis method and then are recognized by a commercial software. The enlargement process involves noise reduction such as block noise and mosquito noise. As a result, it is possible to improve the recognition ratio by using a conventional character recognition algorithm compared with binarized images or enlarged images by bi-linear interpolation.

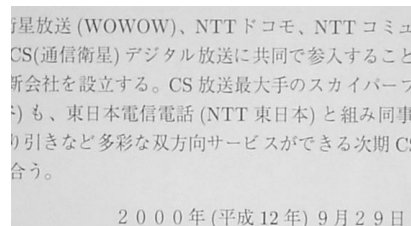
It is, however, not enough recognition performance. Character recognition scheme for degraded characters should be also investigated. Feedback from the character recognition module will be useful for text enhancement method. In this paper, target images are taken under comparatively ideal light condition. Images, however, are degraded by various effects in outdoor. We are planning to propose a robust enlargement method and binarization algorithm for various weather or shadows.

References

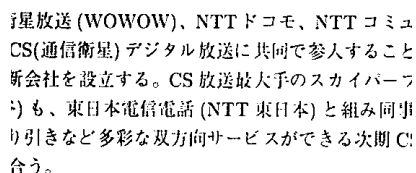
- [1] H. Takahashi, K. Kasai and M. Nakajima: “Extraction of text regions from a scenery image using color and edge features”, The journal of the institute of image information and television engineers, **56**, 6, pp. 979–985 (2002). (in Japanese).
- [2] D. Kim, H. Takahashi and M. Nakajima: “Extraction and recognition of hangul from scenery images”, Technical Report of IEICE, **102**, 55, pp. 65–70 (2002). (in Japanese).
- [3] M. Mori and M. Sawaki: “A survey of robust character recognition and its application”, IEICE Technical Report, **PRMU2001-275**, (2002). (in Japanese).
- [4] T. Sato and T. Kanade: “Contents extraction from news video by character recognition and associating of multimodal information”, Transactions of Information Processing Society of Japan, **40**, 12, pp. 4266–4276 (1999).
- [5] Y. Watanabe, Y. Okada, Y.-B. Kim and T. Takeda: “Character recognition and translation for texts in a scene”, The Journal of the Institute of Image Electronics Engineers of Japan, **26**, 6, pp. 670–676 (1997).
- [6] M. Sawaki, H. Murase and N. Hagita: “Recognition of characters in bookshelf images using automatic dictionary selection based on estimated degradation”, The journal of the Institute of Image Information and Television Engineers, **54**, 6, pp. 881–886 (2000). (in Japanese).
- [7] M. Mori, S. Kurakake, T. Sugimura, A. Shio and A. Suzuki: “Robust telop character recognition in video using background & foreground feature and dynamic modified classifier”, IEICE Trans. inf. & Syst., **J83-DII**, 7, pp. 1658–1666 (2000). (in Japanese).
- [8] J. Kosai, K. Kato and K. Yamamoto: “Character recognition at low resolution with video camera”, The journal of the Institute of Image Information and Television Engineers, **53**, 6, pp. 867–872 (1999). (in Japanese).
- [9] T. Hirano, T. Kameshiro, Y. Okada and F. Yoda: “Peripheral zero-crossing code and canonical discriminant analysis for recognizing character images of

poor quality”, IEICE Technical Report, PRMU-98-159, pp. 71–78 (1998). (in Japanese).

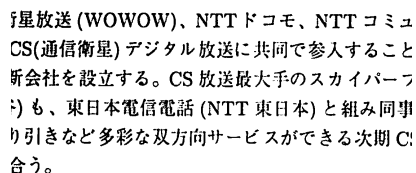
- [10] H. Kuwano, H. Arai, S. Kurakake and T. Sugimura: “Telop character extraction from video deta based on local binarization in each line”, D-12-42, p. 264 (1998). (in Japanese).
- [11] R. Lienhart and F. Stuber: “Automatic text recognition in digital videos”, Image and Video Processing IV, Proc. SPIE 2666-20 (1996).
- [12] S. C. Park, M. K. Park and M. G. Kang: “Super-resolution image reconstruction: A technical overview”, IEEE Signal Processing Magazine, **20**, 3, pp. 21–36 (2003).
- [13] B. Zitová and J. Flusser: “Image registration methods: A survey”, Image and Vision Computing, **21**, 11, pp. 977–1000 (2003).
- [14] T. M. Lehmann, C. Gönner and K. Spitzer: “Survey: Interpolation methods in medical image processing”, IEEE Transactions on Medical Imaging, **18**, 11, pp. 1049–1075 (1999).
- [15] F. M. Candocia and J. C. Principe: “Super-resolution of images based on local correlations”, IEEE Transaction on Neural Networks, **10**, 2, pp. 372–380 (1999).
- [16] K. Jensen and D. Anastassiou: “Subpixel edge localization and the interpolation of still images”, IEEE Transactions on Image Processing, **4**, 3, pp. 285–295 (1995).
- [17] E. Shinbori and M. Takagi: “High quality image magnification applying the gerchberg-papoulis iterative algorithm with dct”, IEICE Trans. inf. & Syst., **J76-D-II**, 9, pp. 1932–1940 (1993). (in Japanese).
- [18] S. A. Martucci: “Image resizing in the discrete cosine transform domain”, in Proc. Int. Conf. Image Processing, **2**, pp. 244–247 (1995).
- [19] Y. Kawaguchi and M. Nakajima: “High quality magnification method of printing image including line region”, Journal of Printing Science and Technology, **34**, pp. 38–46 (1997). (in Japanese).
- [20] N. Otsu: “An automatic threshold selection method based on discriminant and least squares criteria”, IEICE Trans. inf. & Syst., **J63-D**, 4, pp. 349–356 (1980). (in Japanese).



- (a) A text image taken by a digital camera (24×24 pixels per character).



- (b) An enlarged image of (a) by bi-linear interpolation.(2 × 2)

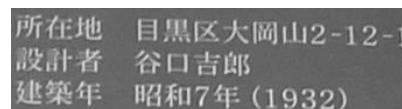


- (c) An enlarged image of (a) by the proposed method.(2 × 2)

Figure 6: Enhancement of document.



- (a) 28×28 pixels per a Chinese character, alphabet “o” is 6×12 pixels.



- (b) : 24×24 pixels per a character.



- (c): 24×24 pixels per Chinese character, alphabet “o” is 4×4 pixels.

Figure 7: Target images.