MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara- ken New Public Hall, Nara, Japan

3—13 Enhancing Character Recognition by Optimizing Focal Length

Takahiko SHIMMURA[1] Yoshinao SHIRAKI[2], Naoki MUKAWA[2] Minako SAWAKI[2] Norihiro HAGITA[3]

NTT DATA CORPORATION NTT Communication Science Laboratories ATR Media information Science Laboratories

Abstract

This paper proposes to use zooming function of digital cameras for robust character recognition and to control focal length for improving the performance even in dim light conditions. For this purpose, we determine the optimum focal length vs. camera-object distance through several recognition experiments. Recognition performance for insufficient illuminated images can be enhanced by adopting the optimum focal length. Experiments involving multiple fonts show that our method improves the recognition rates over a wide range of camera-object distances and illumination variations.

1 Introduction

This paper proposes to use control function of digital cameras for character recognition in scenes. In recent years, the performance of digital video cameras and digital still cameras has been improved and such devices are now in wide use. It is natural to improve recognition performance by more fully utilizing the advanced camera control functions such as zoom and iris control offered by modern cameras.

A character recognition system is composed of an image capturing part and an image recognition part. The former part is the target of this study, while the latter has been examined in many literatures on character recognition. Some previous papers on recognition using digital cameras focused on novel utilization of camera functions. Nakajima et al. proposes a method for recognizing documents that joins adjacent image frames without using image-based mosaic technique [1]. Fujisawa et al. proposes a concept of language translator: a digital camera with character recognition and recognition functions. They review the possible technical approaches for camera control such as zoom and auto-focus [2].

In contrast, this paper aims at enhancing recognition performance by controlling the camera functions automatically. Zooming-in may improve the recognition performance, however it has not been confirmed that zooming-in always works well in various environmental conditions such as illumination conditions, distances between camera and object, and reflectance properties of the object's surface. In this paper, we consider a digital still camera to be an input part of character recognition system and aim to control focal length for recognizing characters under insufficient illumination conditions, such as twilight hours, cloudy days, because zooming-in may cause shading and blurring on the images. An example of binarized images, which were taken with zooming-in, is shown in Fig. 1. Fig. 1 shows that zooming-in degrades contrast and causes shading at the peripheral area of the image under insufficient illumination, thus its binarized image is degraded. This paper proposes a method to optimize the focal length (or zoom parameter) for accurate recognition under dim light conditions.

The recognition results show that our method improves the recognition rates of character images over a wide range of camera-object distances for different illumination levels.

This paper is organized as follows. Section 2 gives a brief review of the problems of zooming-in. Section 3 presents a recognition system that uses our proposed method. Section 4 describes how to determine the optimum focal length from a character image. Section 5 evaluates the proposed method. Our assessment of zoom control for character recognition is given.

2 Problems with Zoom Control in Recognition

As an example of character recognition in a 3D scene, imagine a tourist with a camera that has a built-in character recognition function. The camera is used to read small characters on a bulletin board far from the tourist. Since the characters are enlarged by zooming-in, it can be expected that the recognition rate would be improved. However, when the weather is bad and/or image brightness is insufficient, zooming-in may not improve the recognition rate. This section first analyzes the relationships between focal length and illumination level at the light-receiving element of the camera using an optical model. It then investigates the characteristics of recognition.

We assume a single lens camera model for simplicity so that the illuminance falling on the light-receiving element of the camera can be modeled as:

$$Ep = \frac{\pi}{4}\cos^4\theta\left(\frac{D}{Fp}\right)^2 L \quad , \qquad (1)$$

where Ep represents the illuminance at the light-receiving element, L is the irradiance of the viewed object's surface per unit area, θ is the angle between object surface and camera optical axis, D is the radius of the lens, and Fp represents the focal length of the camera [4]. From the equation, it is found that the illumination level decreases as the focal length increases (corresponding to zooming-in), and it is also found that this effect is most clearly seen at areas where θ is large (i.e. the corners of the image) because of the $\cos^4\theta$ characteristics. In the case of zoom lenses, Amano et al. [9] shows that shading around the edge of an image is induced by vignetting distortion rather than by the $\cos^4\theta$ characteristics. He

[1] Address: Kayabacho Tower Bldg., 1-21-2, Shinkawa, Chuo-ku, Tokyo 104-0033 JAPAN E-mail: shinmuratkh@nttdata.co.jp

[2] Address: NTT Corporation, 3-1, Morinosato, Wakamiya, Atsugi-shi, Kanagawa, 243-0198 JAPAN

[3] Address: ATR, 2-2-2,Hikari, Seika-cho, Soraku-gun, Kyoto, 619-0288 JAPAN

proposes the variable cylinder model that can describe vignetting distortion characteristics.

These two models indicate that zooming-in causes image shading and lowering brightness. Fig. 1 shows shading around the edge of an image. It is obvious that the shading is produced as the focal length (the degree of zoom-in) is increased. Fig. 2 shows the brightness variations of the character images of the cardboard shown in Fig. 1. Because the area of background is larger than that of characters, a peak of the frequency appears in higher gray-level side in Fig. 2. The figure shows that the longer the focal length, the darker the image. As to our preliminary experiments, however, both the $\cos^4\theta$ characteristics and vignetting distortion model cannot hold for accurately describing the phenomenon of the brightness lowering because image brightness is controlled automatically by digital still camera function.

From the viewpoint of character recognition, shading hinders the segmentation of characters from images; low brightness degrades binarization. That is, zooming-in under dim light conditions may lead to lower recognition accuracy. To maximize the recognition accuracy, a method that finds the optimum focal length is proposed.

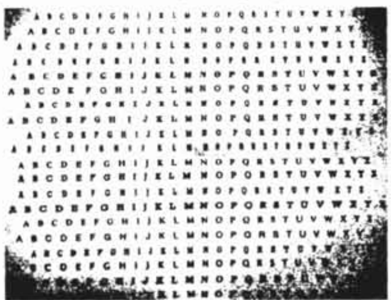

Fig. 1 An example of binarized image (focal length: 37.8 mm, threshold: 40, camera-distance: 454 cm, illuminance: 50 *lx*).

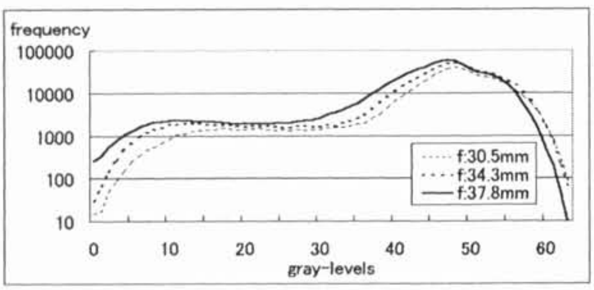

Fig. 2 Variation of 256 gray level histogram of Fig.1 by changing the focal length.

3 System Considered

Fig. 3 shows our proposed recognition system. Our research is directed towards how to estimate the optimum focal length for this system. First, the picture is taken at an initial focal length *f0* and binarized. Second, the system yields an average character size *s0* in the image and then the distance between camera and object is estimated from the character size. Then, the optimum focal length *f** is estimated by referring predetermined "focal length table". Finally, the characters are again captured using the optimum focal length, *f**, and the characters of the images are recognized

In this paper, instead of an information board, we use a sheet of cardboard on which the characters shown Table 1 are printed. The sheet is posted on a wall in a room. The sheet size and character printing size is known. The sheet is captured as a single frame image. The character images are manually clipped from the full picture image.

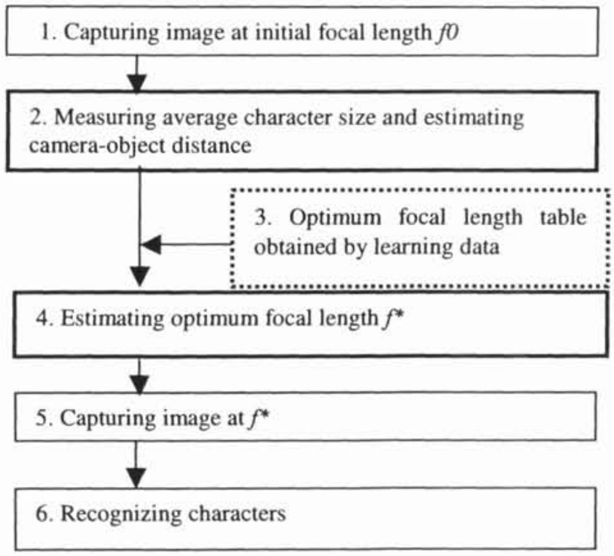

Fig. 3 Recognition system using zooming.

4 Method of Adjusting Focal Length

4.1 Focal Length Table

This section describes a method to obtain a "focal length table" representing a relationship between an optimum focal length and camera-object distance. To examine the characteristics of recognition rate and focal length, both illumination levels and camera-object distances were changed. The conditions of image acquisition are shown in Table 1. Then, the recognition rate was calculated using learning data of character images. The specification of recognition algorithm is shown in Table 2. To pick out some appropriate recognition processes on table 2, some conventional processes were tested for this digital still camera. Fig. 4 shows examples of fonts used in our experiments. The camera-object distances and illumination levels were chosen according to preliminary experiments.

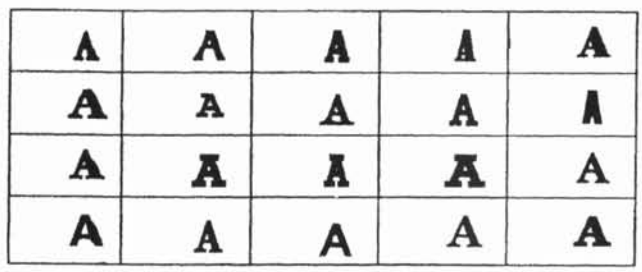

Fig. 4 Example of 20 fonts.

Fig. 5 illustrates the recognition performance at a camera-object distance of 454 cm. It shows that at 200 *lx*, zooming-in continually improves the recognition rate until zooming-in makes the camera finder full with the image. At 50 *lx* and 70 *lx*, on the other hand, a peak of recognition rate can be seen, and the fact shows that the optimum focal length exists to achieve the highest recognition rates for low-illuminated images. Through the experiments the optimum focal lengths at other camera-object distances can be found out as well. As shown in Fig. 5, the experiments showed that the optimum focal length for one camera-object distance is not sensitive to illumination changes. It seems appropriate to define one optimum focal length for one camera-object distance, regardless of illumination changes. Thus, the characteristic of optimum focal length is drawn simply by camera-object distance as shown in Fig. 6.

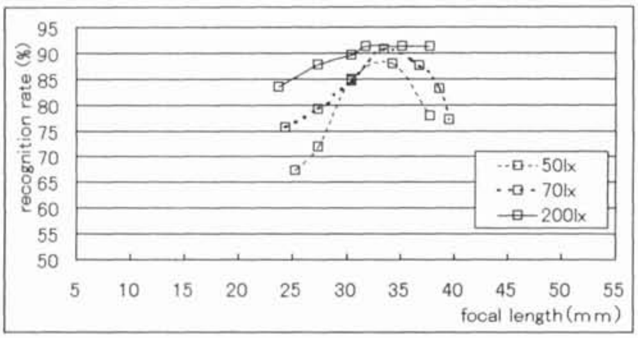

Fig. 5 Recognition rate at different illumination levels: 50, 70, 200 lx, camera-object distance: 454 cm.

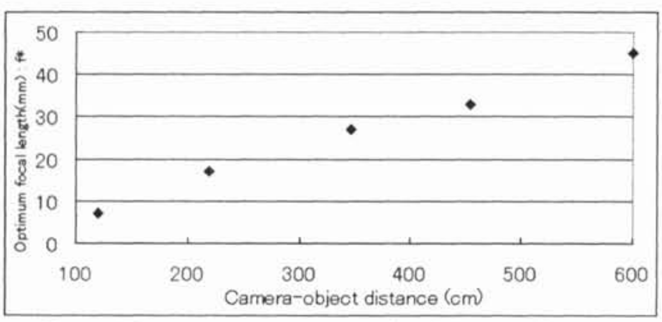

Fig. 6 Estimating optimum focal length from camera object distance at *f0* = 35.2 mm.

4.2 Estimating Optimum Focal Length

For obtaining one optimum focal length of an unknown image, we must determine the distance between camera and object. In this study we assume the font size is known and fixed. Therefore, we can estimate the distance from the character size in the captured image. (When the proposed method extended to variable font size characters, techniques of depth-from-focus and depth from stereo will be more appropriate for the distance estimation.) Table 3 shows the relationship between camera-object distance and average character size. The camera-object distance of an unknown image is estimated by interpolation from the table. Finally, using Fig. 6 we can obtain optimum focal length *f** for the captured image.

To take the above example, when the system takes a picture of characters from unknown distance at *f0*, it yields *s0* by averaging all of height and width in the image. And the system calculates the distance at *s0* from Table 3 using linear regression analysis. Then, with this calculated distance, the system estimates an optimum focal length, *f**, as 28.5 mm, using linear regression analysis through the data points on Fig. 6.

5 Evaluation of Proposed Method

The experiments to evaluate the proposed method were performed under the conditions shown in Tables 1 and 2, and its performance was compared with a case that the focal length is so large as to maximize the image within the camera finder. Some of the fonts in Table 2 are the same or similar to the fonts in reference [5]. 520 characters (26 alphabet capital characters of 20 fonts) were printed randomly on 20 lines on cardboard, and captured in a single frame image. The illumination levels were 60, 100 *lx* and camera-object distances are 283, 400, and 536 cm. The focal length was determined using three methods: ***A***. proposed method, ***B***. zooming-in so that the characters filled the viewfinder, ***C***. default (7 mm).

Character recognition part used in our experiments consists of input image binarization, character segmentation, character aspect size normalization, and feature vector calculation. Templates were made from a full-sized picture image in viewfinder, whose camera-object distance was 120 cm and illumination level was 200 *lx*. The distances between each input feature vector and a set of stored templates were calculated. The system output the category of the corresponding template whose distance is minimum. Input images include 20 font sets. For strict evaluation, leave-one-out procedure was used, this means the font used in the input character image was not included in the dictionary.

The experimental results are shown in Table 4. The results show that our method improves the recognition rates over a wide range of camera-object distances for different illumination levels by contrast with the way using zooming-in (see Table 4).

For example, at the illumination level of 60 *lx* and the camera-object distance of 400 cm, method ***B*** achieved the recognition rate of 76.5%; its focal length was 32.6 mm. Method ***A*** (proposed method) achieved the recognition rate of 80.6%; its focal length was 30.5 mm[*4]. Both the initial focal length, *f0* of method ***A*** and method ***B*** are the same. Method ***C*** (7 mm focal length) yielded the lowest recognition rate.

To confirm effectiveness of the proposed method against image shading and blurring caused by insufficient illumination, we evaluated local characteristics by dividing one image into 9 regions and calculated the recognition rate of each region (Table 5). It was shown that proposed method improved the recognition rate at the peripheral area of the image, while the recognition rate of the center area maintained compared with the case of maximum image enlargement.

6 Conclusion

We proposed a method that estimates the optimum focal length for enhancing character recognition performance using images captured under low-light conditions. The optimum focal length is estimated from calculated camera-object distance based on character size in an image. From the two experiments it was confirmed that our proposed method improves the recognition rate of degraded images caused by illumination variation. The results con-

*4 Although the estimated focal length using the proposed algorithm was 29 mm, the value of 30.5 mm was available to be chosen.

tribute to designing a camera system with a character recognition function. As the first step of our study, we assumed that the character size printed in the object is fixed and known. To find elementary problems of a camera system, conventional recognition processes were used. Our on-going study is to extend the method to handle the case wherein character size and reflection property of the object vary.

Table 1 Experimental data and camera specification

Characters	Capital characters: 26 of 20 fonts Print size: 48point
Digital still camera	CANON Power Shot-Pro90IS 1024x768pixels, optic zoom: 7-70 mm shutter speed: 1/25sec, diaphragm:f.2.8 no flash, no digital zoom
Image	256 gray levels

Table 2 Specification of recognition part

Size normalization	Adjustable aspect ratio, 48x48pixels Binarized image
Feature	Local Direction Contributivity
Distance measurement	Euclidian distance
Templates	Number of templates: 26 characters of 20 fonts Binarized image Camera-object distance: 120 cm
Evaluation procedure	Leave-One-Out

Table 3 Camera-object distance and average character size.

Camera-object distance (cm)	120	219	346	454	600
Character size (pixels) at $f0 = 35.2$ mm	48.2	26.5	16.5	12.9	9.8

Table 4 Results of evaluation (A: proposed method, B: maximum zooming, C. default (7 mm))

Illumination(*lx*)	100	100	100	60	60	60
Distance (cm)	283	400	536	283	400	536
Recognition rate (%)						
A.	**82.7**	**81.7**	**81.9**	**80.6**	**80.6**	**73.5**
B.	81.2	75.4	71.7	76.9	76.5	72.9
C.	38.1	3.3	3.7	29	3.8	3.7
Focal length (mm)						
A. estimation	22	28.5	40	22	29	39
A. manual work	22.4	31.1	39.6	23.7	30.5	41.8
B. zoom-in	24.8	35.2	47.8	25.2	32.6	46.5
C. default	7	7	7	7	7	7

Table 5 Recognition rates for nine region. The parentheses denote the number of characters at each region (Camera-object distance: 400 cm, illuminance: 60 *lx*).

(a) proposed method:

	Left	Middle	Right
Upper	**89.6**(48)	80.0(60)	75.0(48)
Middle	**84.4**(64)	75.0(80)	62.5(64)
Lower	**95.8**(48)	**83.3**(60)	**68.8**(48)

(b) maximum zooming:

	Left	Middle	Right
Upper	83.3(48)	90(60)	87.5(48)
Middle	82.8(64)	75.0(80)	62.5(64)
Lower	75(48)	60.0(60)	58.3(48)

7 Acknowledgments

We are grateful to Dr. H. Murase, Dr. K. Ishii, and Dr. N. Sugamura for encouraging our work. All research was performed at NTT Communication Science Laboratories.

References

[1] N. Nakajima, N. Tanaka and K. Yamada: "Document Reconstruction and Recognition from an Image Sequence," Proc. 14th Int. Conf. Pattern Recognition, pp.922-925, Brisbane, Australia, Aug. 1998.

[2] H. Fujisawa, H. Sako, Y. Okada, S-W. Lee: "Information Capturing Camera and Developmental Issues," Proc., 5st Int. Conf. On Document Analysis and Recognition (ICDAR'99), pp.205-208, 1999.

[3] W. Watanabe, Y. Okada, Y-B Kim and T. Takeda: "Translation Camera," Proc. 14th Int. Conf. Pattern Recognition, pp.613-617, Brisbane, Australia, Aug. 1998.

[4] Horn. B.K.P. and R.W. Sjoberg: "Calculating the reflectance map," Proc. DARPA IU Workshop, pp.115-126, Nov 1978.

[5] H.S. Baird and Reid F: "A 100-Font Classifier," Proc., 1st Int. Conf. On Document Analysis and Recognition (ICDAR'91), pp.332-340, St.-Malo, France, 30 Sep-2 Oct, 1991.

[6] N. Otsu: "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. SMC, VOL.SMC-9, NO.1, pp.62-66, Jan. 1979.

[7] T. Akiyama and N. Hagita: "Automated Entry System for Printed Documents," Pattern Recognition, Vol.23, No.11, pp.1141-1154, 1990.

[8] A. Guiducci: "Camera Calibration for Road Applications," Computer Vision and Image Understanding 79, pp.250-266, 2000.

[9] A. Amano, N.Asada and M.Baba: "Photometric Calibration of Zoom Lens Systems: Analysis and Correction of Vignetting Distortion by the Variable Cylinder Model,"(in Japanese), Trans. IEICE, Vol.J80-D-II, No.6, pp1458-1465, 1997.