IPSJ Transactions on Computer Vision and Applications Vol.6 48–52 (July 2014)

[DOI: 10.2197/ipsjtcva.6.48]

Express Paper

Recognition of Defocused Patterns

Masakazu Iwamura[1,a) Masashi Imura[1,b) Shinsaku Hiura[2,c) Koichi Kise[1,d)

Received: March 14, 2014, Accepted: April 24, 2014, Released: July 25, 2014

Abstract: The paper addresses the recognition problem of defocused patterns. Though recognition algorithms assume that the input images are focused and sharp, it does not always hold on actual camera-captured images. Thus, a recognition method that can recognize defocused patterns is required. In this paper, we propose a novel recognition framework for defocused patterns, relying on a single camera without a depth sensor. The framework is based on the coded aperture which can recover a less-degraded image from a defocused image if depth is available. However, in the problem setting of "a single camera without a depth sensor," estimating depth is ill-posed and an assumption is required to estimate the depth. To solve the problem, we introduce a new assumption suitable for pattern recognition; *templates are known*. It is based on the fact that in pattern recognition, all templates must be available in advance for training. The experiments confirmed that the proposed method is fast and robust to defocus and scaling, especially for heavily defocused patterns.

Keywords: pattern recognition, coded aperture, defocus, local feature

1. Introduction

Most recognition algorithms assume that the input images are focused and sharp, though physical focusing requires non-negligible time for lens movement. Therefore, recognition of defocused patterns is a significant but challenging task because their features for recognition are almost lost in degraded images. **Figure 1** shows an example of the effect of defocus; defocused images (Fig. 1 (b)) of characters "0" and "8" are much more similar than the original in-focus images (Fig. 1 (a)), and they are hard to distinguish by conventional image recognition algorithms.

One of the promising approaches to cope with defocus is to recover a less-degraded image using *coded aperture* [1], [4]. **Figure 2** shows camera models with and without coded aperture. Since circular aperture used in the conventional camera shown in Fig. 2 (a) does not preserve high-frequency components, it is difficult to recover an in-focus image from a blurred input image. In contrast, a camera with the coded aperture makes it possible to recover a less-degraded image [4], [8] because it avoids losing information on the input image. The reconstruction process, however, requires depth information because the size of defocus depends on the distance from the camera to the object. In other words, depth estimation is a key component of the image restoration algorithms of deblurring [1], [4]. Some solutions to estimate the depth have been presented. For example, depth from defocus [1], coded aperture pairs [9] and stereo method [7] are proposed. These methods, however, require two or more input images for depth estimation. In general, the use of multiple input images involves more cost for equipment, or requires a longer time for sequential image capture.

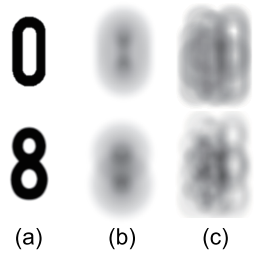


Fig. 1 Distinguishing "0" and "8." (a) Original patterns, (b) defocused with circular aperture and (c) defocused with coded aperture [8].

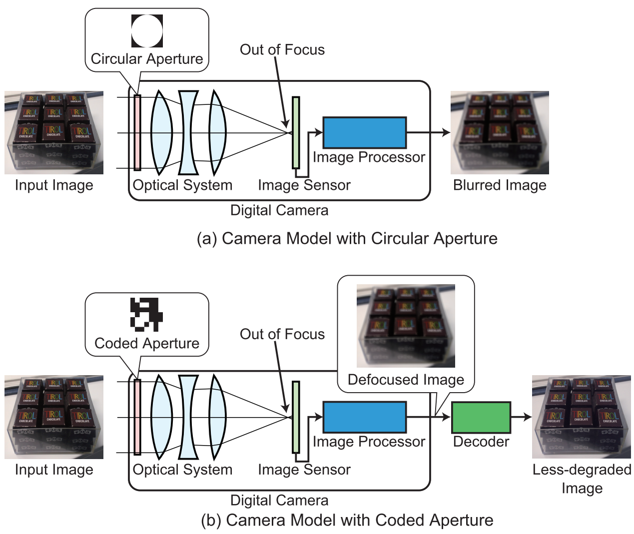


Fig. 2 Camera models with (a) circular and (b) coded apertures.

In this paper, we exclude them so as to pursue a solution for a single camera without a depth sensor. In such a case, estimating the depth is ill-posed and an assumption is required to estimate

1 Osaka Prefecture University, Sakai, Osaka 599–8531, Japan
2 Hiroshima City University, Hiroshima, 731–3194, Japan
a) masa@cs.osakafu-u.ac.jp
b) imura@m.cs.osakafu-u.ac.jp
c) hiura@hiroshima-cu.ac.jp
d) kise@cs.osakafu-u.ac.jp

Table 1 Comparison of recognition methods.

Methods	Aperture	Dictionary	Base recognition method	Translation invariance	Scale invariance
Conventional method	Circular	In-focus templates	Template matching	✓	-
Naive method	Coded	Defocused templates	Template matching	✓	-
Proposed method	Coded	Defocused templates	Arrangement of local features	✓	✓

the depth. For example, the depth of a barcode is estimated based on an assumption that a barcode consists of only black and white, and no color inbetween them exists [3]. This assumption is, however, not generally applicable to pattern recognition because it is not practical to limit the colors of a pattern.

Thus, we introduce a new assumption suitable for pattern recognition in general. That is, *templates are known*. We utilize the fact that in pattern recognition, all templates must be available in advance for training. By comparing templates and the (input) query image, not only depth but also the category of the query can be estimated. Based on this notion, we propose a recognition method of defocused patterns that is invariant to translation and scaling.

2. Proposed Framework

We present two methods in the proposed framework: the naive method and the proposed method. In contrast to the conventional method, their features are summarized in **Table 1**.

An overview of the proposed framework is presented in contrast to the conventional method. The conventional method in **Fig. 3** (a) uses the circular aperture and captures a defocused image as in Fig. 1 (b). The captured image can be far different from templates included in the dictionary. This may cause misrecognition. On the other hand, the proposed framework in Fig. 3 (b) uses a coded aperture and captures an encoded image as in Fig. 1 (c). In addition to the templates used in the conventional method, defocused patterns with different depths can be included in the dictionary. Finding the closest templates to the captured image reveals both category and depth of the captured image simultaneously.

A naive method in the proposed framework is to use the template matching method with the sliding window approach. In this way, the method can cope with the translation of a pattern but cannot cope with other degradation such as scaling. Thus, we propose to use a local-feature-based recognition method for camera-captured Japanese characters [6]. The recognition method is much faster than the template matching and can recognize Japanese characters including Kanji (Chinese characters), that may not be linearly aligned and may be printed with complex backgrounds. Though the method is robust to affine transformation, the robustness is limited to translation and scaling in the proposed method. This is due to the fact that using the coded aperture, a different order of rotation and defocus leads to different results (see **Fig. 4**) [*1]. Thus, a rotated pattern cannot be recognized with the rotated corresponding template.

The base method [6] is based on local features and their arrangement. As the local feature, Scale-Invariant Feature Transform (SIFT) [5] is used. We briefly present the recognition algorithm. In preparation, the SIFT features are extracted from all templates and they are stored in the database. For recognition, SIFT features are extracted. Then, the corresponding features in the database are found using the state-of-the-art approximate nearest neighbor search method [2]. Considering the arrangement of corresponding features in the query image and templates, the category and location of a pattern are determined simultaneously.

*1 More precisely, a camera-captured image is rotation sensitive unless the shape of the aperture is rotation invariant. The shape of the coded aperture used in this paper is rotation sensitive, while the circular aperture is rotation invariant.

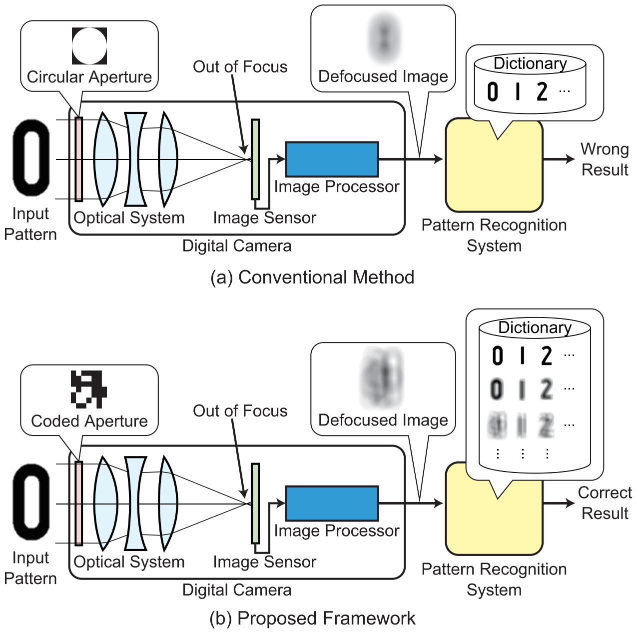


Fig. 3 Overviews of conventional method and proposed framework.

(a) Rotation after defocus (b) Defocus after rotation

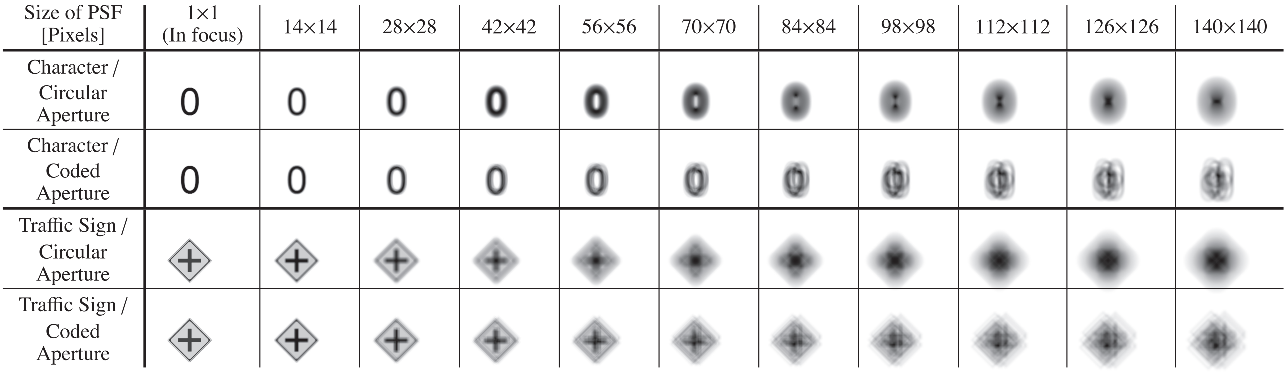
Fig. 4 Different order of operations leads to different results.

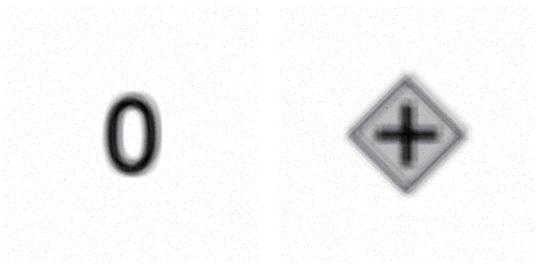
3. Experiments

Experiments with characters and traffic signs were conducted. For characters, 62 alphanumeric characters in MS Gothic font were used. For traffic signs, 97 traffic signs [*2] were used; 27 were for warning (slanted yellow squares), 58 were for regulation (square, triangular and rounded shapes drawn in red, blue and white) and 12 were for instruction (square and nearly triangle shapes in blue and white). The resolution of each template was 350×350 including white space surrounding a pattern. Without

*2 Extracted from http://www.mlit.go.jp/road/sign/sign/douro/ichiran.pdf.

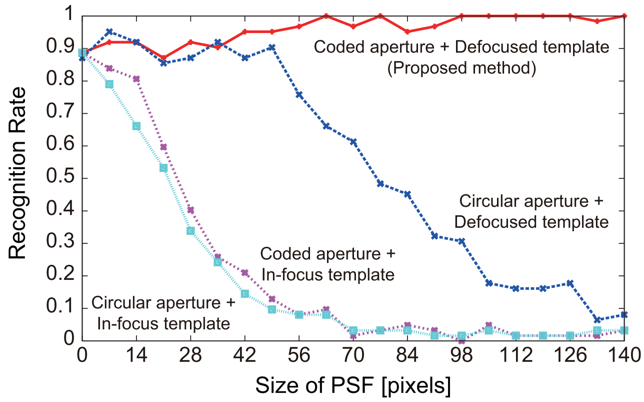
Table 2 Examples of templates.



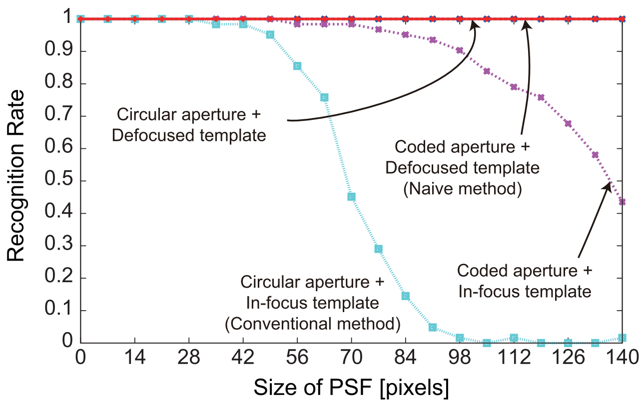


(a) Character (b) Traffic sign

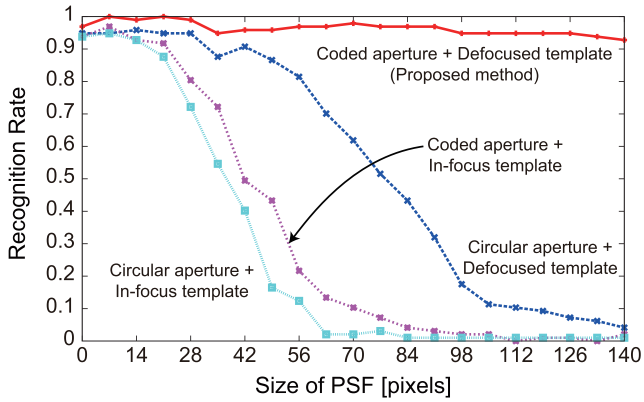
Fig. 5 Noisy query images.



(a) Character

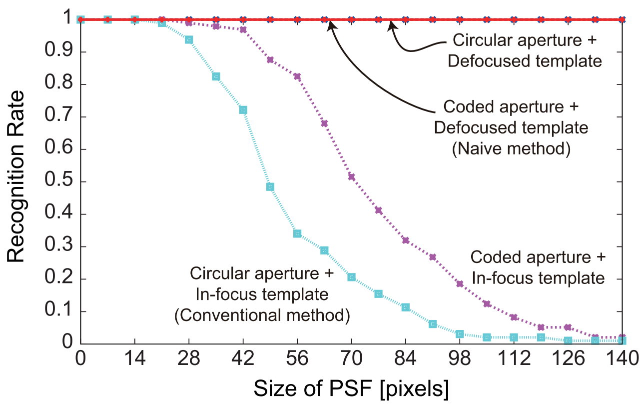


(a) Character



(b) Traffic sign

Fig. 7 Recognition rates of scale-invariant recognition methods.



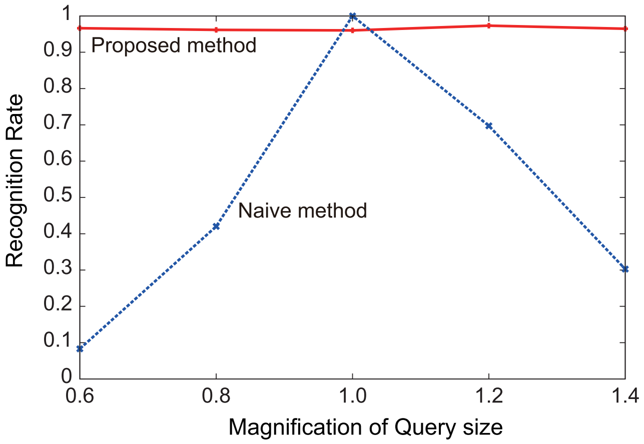
(b) Traffic sign

Fig. 6 Recognition rates of scale-sensitive recognition methods (recognition methods inbetween conventional method and naive method).

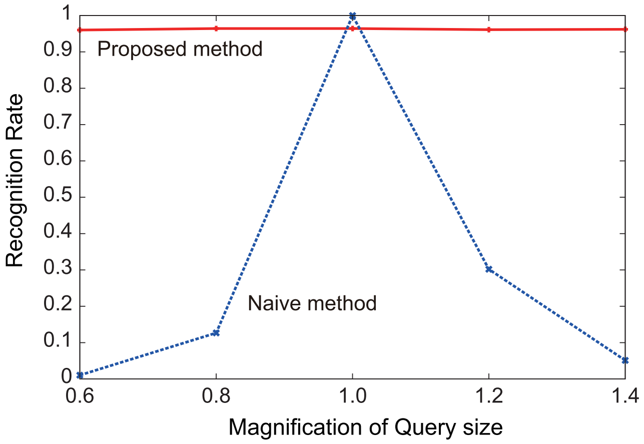
white space, the resolutions of non-defocused "0" and the sign shown in **Table 2** were 138 × 211 and 328 × 328, respectively. All the images were converted to gray scale in advance.

As examples are shown in Table 2, for each template, 20 levels of defocused images in addition to the original in-focus one were prepared as defocused templates. The 20 levels are described by the sizes of their point spread functions (PSFs) because a defocused image is obtained by convolution of the original image and a PSF. PSFs of from 7 × 7 to 140 × 140 pixels were used. Note that the size of the PSF is proportional to the distance between the focal plane and the pattern. The circular aperture and the coded aperture designed by Veeraraghavan et al. [8] were used.

As queries, the defocused templates where simulated CMOS noise was added were used. The noise was created as follows.



(a) Character



(b) Traffic sign

Fig. 8 Robustness to scale change.

A random number that follows the standard Gaussian was converted to the intensity of a pixel value. **Figure 5** shows examples of noisy query images.

Since "I," the upper-case letter of i, and "l," the lower-case letter of L, are quite similar in MS Gothic font, confusion of them was not regarded as misrecognition.

Let us present experimental results. First, four combinations of circular/coded apertures and in-focus/defocused templates on scale-sensitive recognition methods based on template matching were examined. They include the conventional and naive methods. **Figure 6** shows that recognition rates of methods with in-focus templates decreased as the size of the PSF increased, and methods with defocused templates achieved 100% regardless of the size of PSF. Comparing recognition rates of methods with in-focus templates, the one with a coded aperture outperformed the one with a circular aperture. Among the methods, the conventional method was the worst and the naive method was the best.

Secondly, four combinations of circular/coded apertures and in-focus/defocused templates on scale-invariant recognition methods based on arrangement of local features were examined. They include the proposed method. **Figure 7** shows the same tendency as Fig. 6; methods with defocused templates outperformed ones with in-focus templates; ones with a coded aperture outperformed ones with a circular aperture. Among the methods, the proposed method achieved the best.

Table 3 Query time for recognizing a pattern.

Methods	Query time [s]	
	Character	Traffic sign
Conventional method	6.0	6.9
Naive method	100.5	156.5
Proposed method	0.16	0.17

Finally, the robustness of the naive and proposed methods to scaling was examined. In this experiment, all the query images scaled from 0.6 times to 1.4 times were used. The recognition results in **Fig. 8** support the effectiveness of the proposed method; the proposed method was not affected by scaling. The naive method was almost useless for scaled query images, though it achieved 100% for non-scaled query images.

Query time is summarized in **Table 3**. It shows that the naive method took a far larger amount of time (about 20 times) than the conventional method to recognize defocused patterns. In contrast, the proposed method took a very short time (1/920 to 1/630 in comparison with the naive method) to realize scale-invariant recognition.

4. Conclusion

In this paper, we proposed a novel recognition framework to recognize defocused patterns with a single camera without a depth sensor. The framework is based on the coded aperture which can recover a less-degraded image from a defocused image if depth is available. However, in the problem setting of "a single camera without a depth sensor," estimating depth is ill-posed and an assumption is required to estimate the depth. To solve the problem, we introduced a new assumption suitable for pattern recognition; *templates are known*. It is based on the fact that in pattern recognition, all templates must be available in advance for training.

We presented two methods in the proposed framework. One is called the naive method which is based on the template matching technique with the sliding window approach. This is translation invariant. The other is called the proposed method which is based on the arrangement of local features. This is invariant to translation and scaling. Through the experiments, we confirmed that the proposed method is fast and robust to defocus and scaling, especially for heavily defocused patterns.

Acknowledgments This work was supported by JSPS KAKENHI, the Grant-in-Aid for Challenging Exploratory Research (23650090).

References

[1] Hiura, S. and Matsuyama, T.: Depth measurement by the multi-focus camera, *Proc. 1998 Conference on Computer Vision and Pattern Recognition* (*CVPR '98*), pp.953–959 (1998).

[2] Iwamura, M., Sato, T. and Kise, K.: What Is the Most Efficient Way to Select Nearest Neighbor Candidates for Fast Approximate Nearest Neighbor Search?, *Proc. 14th International Conference on Computer Vision* (*ICCV 2013*), pp.3535–3542 (2013).

[3] Kawamoto, Y., Hiura, S. and Asada, N.: Optimizing the Shape of Coded Aperture for Specific Subjects, *Poster of International Conference on Computational Photography* (*ICCP2013*) (2013).

[4] Levin, A., Fergus, R., Durand, F. and Freeman, B.: Image and Depth from a Conventional Camera with a Coded Aperture, *ACM Trans. Gr.*, Vol.26, No.3 (2007).

[5] Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91–

110 (2004).
[6] Matsuda, T., Iwamura, M. and Kise, K.: Performance Improvement in Local Feature Based Camera-Captured Character Recognition, *Proc. 11th IAPR International Workshop on Document Analysis Systems* (*DAS2014*) (2014).
[7] Takeda, Y., Hiura, S. and Sato, K.: Fusing Depth from Defocus and Stereo with Coded Apertures, *Proc. 2013 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR2013*), pp.209–216 (2013).
[8] Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A. and Tumblin, J.: Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing, *ACM Trans. Gr.*, Vol.26, No.3 (2007).
[9] Zhou, C., Lin, S. and Nayar, S.K.: Coded Aperture Pairs for Depth from Defocus, *Proc. 12th International Conference on Computer Vision* (*ICCV2009*), pp.325–332 (2009).

(Communicated by *Yasushi Makihara*)