Electronics and Communications in Japan, Vol. 100, No. 10, 2017
Translated from Denki Gakkai Ronbunshi, Vol. 136-C, No. 9, September 2016, pp. 1367–1375

Proposal of an Encoded Marker for Working Robots: An Encoded Marker Easy to Detect in Various Positions and under Blur

NORIMASA KOBORI,[1] DAISUKE DEGUCHI,[2] ICHIRO IDE,[3] and HIROSHI MURASE[3]
[1]Partner Robot Division, Toyota Motor Corporation, Japan
[2]Information Strategy Office, Nagoya University, Japan
[3]Graduate School of Information Science, Nagoya University, Japan

SUMMARY

It is becoming important for working robots to be able to identify and pick objects in various tasks. As in the recent Amazon Picking Challenge, using a marker for the picking task is a more practicable approach. However, a common maker code for working robots does not exist so far. Conventional maker codes as represented by QR code or ARToolKit marker cannot be reliably detected from various viewpoints. Thus, in this paper, we propose a new encoded marker which is flexible to the marker's position and blur. The proposed marker can be detected by an approach based on the scale space theory independent from such conditions. In addition, the representation of data by M-sequence makes the encoded marker robust to blur. Experimental results showed the effectiveness of the proposed marker compared to the ARToolKit marker. Since the marker is more robust against ground clutter noise, various positions of markers and blur, it is more practicable. © 2017 Wiley Periodicals, Inc. Electron Comm Jpn, 100(10): 59–69, 2017; Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ecj.11987

Keywords: marker; scale-space theory; M-sequence; blur.

1. Introduction

In recent years, working robots such as Baxter [1] (Rethink Robotics) and NEXTAGE [2] (Kawada Robotics Corporation) have been becoming popular in factories, and so on. Among these robots, an interest in task for robots to pick up objects, such as Amazon Picking Challenge [3], has been increasing. In order that such robots can be adopted for all purposes, they need to ensure flexibility to distance and arrangement for targets. Namely, it is important that even if multiple objects as targets are located at any distances and arrangements, these situations can be detected and recognized, and these arrangements further can be accurately estimated. A method has been studied that objects are detected and recognized and their three-dimensional arrangements are estimated after registering or learning these models to this kind of problem setting [4, 5]. In this method, three-dimensional CAD model data of an object are prepared, its various arrangements are recorded as two-dimensional RGB-D image (or RGB image) templates, and then the object is detected and recognized and its arrangement is estimated by comparing it with templates. This method has practical disadvantages such that its computational cost is high because each object is compared with various templates and that higher detection accuracy cannot be realized without specific shapes.

Meanwhile, a method has been practically used that detection, recognition, and attitude estimation of objects are performed by attaching markers directly on targets. If a target is a small object and has curvatures, a marker may be attached on a box with a handle, in which the target object is placed [6, 7]. If the position relationship between the attached marker and the grip position of the object is confirmed in advance, a robot can act grip operation. Many of currently used markers have practical performance by limiting the distance and arrangement between a marker as an object to be shot and a camera. Therefore, such a system cannot be operated under a condition where markers cannot be easily detected. In the case that such a marker is applied, two serious disadvantages mainly arise if (1) markers with various distances and arrangements are placed under environment with disordered background and (2) blur is caused. In order that robots act various kinds of operations, (1) is a serious problem to be solved, and highly flexibility is demanded to distance to markers and their arrangements. It is also important to control the environment with blur of (2) so as to solve the problem (1). Blur is caused because a marker is situated out of the range of depth of field.

Considering the background explained above, in this paper, we propose a novel marker for the purpose to solve

the following two practical problems to a marker for working robots, assuming grip operation.

1. Marker easy to detect independent on blur
2. Marker easy to detect independent on background and arrangement

In order to solve them, we propose "detection" processing easy to find markers and a design pattern of markers necessary for that. A method focusing on "detection" has not been proposed before, and it is a specific task if robots deal with markers. In this paper, finding a marker is defined as "detection", and accurately identifying a marker among detected candidates is defined as "recognition".

The proposed marker can be detected under a condition where its spatial arrangement, position, and size (scale) are unknown in advance, and it is robust to its arrangement and scale change. Different from a conventional marker, it is not necessary to compare this proposed marker with multiple templates with different arrangements and sizes. Specifically, we propose an encoded marker utilizing scale-space theory [9] known as keypoint detection by SIFT [8]. Marker detection is efficiently performed by using a discrimination filter which extracts only marker shape [10] easy to calculate extremum of scale space and marker. Then, a marker can be recognized using M-sequence [11] even if its pattern is blurred.

Figure 1 shows an example where the proposed markers are detected. As illustrated in Fig. 1, some misdetections are observed with conventional ARTookKit markers. They are easily affected by light and shadow so that some markers cannot be recognized correctly (Fig. 1(b)). Meanwhile, it can be found that multiple proposed markers placed in different positions can be detected correctly in a cluttered environment (Fig. 1(d)).

Hereafter, associated studies are introduced in Section 2, a proposed marker is described in Section 3, experimental results and their discussion are described in Section 4, and conclusions and a challenge in the future are described in Section 5.

2. Associated Studies

In this section, conventional marker technologies are introduced in terms of "detect" to find markers. QR code [12], which is currently the most widely used, is the marker constructed by patterns such as edges and corners. In order to detect corner patterns of three corners as marks, it is required that the distance between a camera and a marker is short and a marker is shot from relatively front side. On capturing from distant places, modification of detection performance is aimed using encoded aperture [13] and super-resolution [14] since blurred patterns are observed. However, as it is assumed that this marker is utilized under a condition where it can be easily detected, it is difficult to apply this marker to working robots which need to detect markers situated in distant places.

Markers used for ARToolKit [15] are widely utilized to estimate position and arrangement of a camera in environment. A marker is represented by binary pattern, that is, black bold frame and inner part. Detection is performed by finding four points at corners of the frame and then comparing four sides. The relative position and arrangement between a camera and a marker is also estimated using the four points at the corners at the frame. This marker serves to estimate position and arrangement for any placement. However, there is a problem that misdetection often occurs if there are similar shapes to the marker such as a rhombus.

Meanwhile, many markers with color information have been proposed. As examples, ChamelonCode [16], ColorZip [17], ColorOCM [18], ColorBit [19], C-Band [20], HCCB [21], and so on, are given. These markers are represented by a set of patterns with colors, where features (information) are expressed based on pattern colors and color difference between adjacent colors. This kind of marker has been already adopted to manage books in libraries as practical use, where this system enables simultaneous detection and recognition of multiple markers by measuring markers attached on spines of books from a certain distance. Like QR code, the detection of these markers is classified into a method [?, [21]] to acquire specific patterns and a method [18, 19] to identify arrangement of colors. The former method has the same disadvantage as QR code has. In the latter method, a marker is difficult to be detected in the case that there are similar colors in their background and that color information is mixed or lacked by its blur because a marker far from a camera is shot by small size. Under a situation that colors of marker cannot be acquired precisely, while a method has higher flexibility to distortion and arrangement of markers, they cannot be detected.

As a marker which can be detected from a distant location, Nested Marker [22] and Bokode [23] are given. Nested Marker adopts fractal pattern, where it is designed to be similarly observed both in short distance and long distance with regard to its geometrical pattern. A marker Bokode is made of optical material, so that a floodlight is required for its detection. A marker receiving floodlighted light projects a larger pattern than itself in the air by the principle of pinhole camera. Both of the methods aim to solve the problem on distance and recognition (decoding performance of data), and the relationship between a camera and a marker is established only from front side. However, if a marker does not face to a front side, the same problem as QR code occurs. Meanwhile, as a marker robust to blur, Mono-Spectrum Marker [24] is given. In this method, an element of pattern is represented by single color in low-frequency region. Only recognition method is discussed. The detection procedure is the same as that of ARToolKit marker, so that it has similar problem.

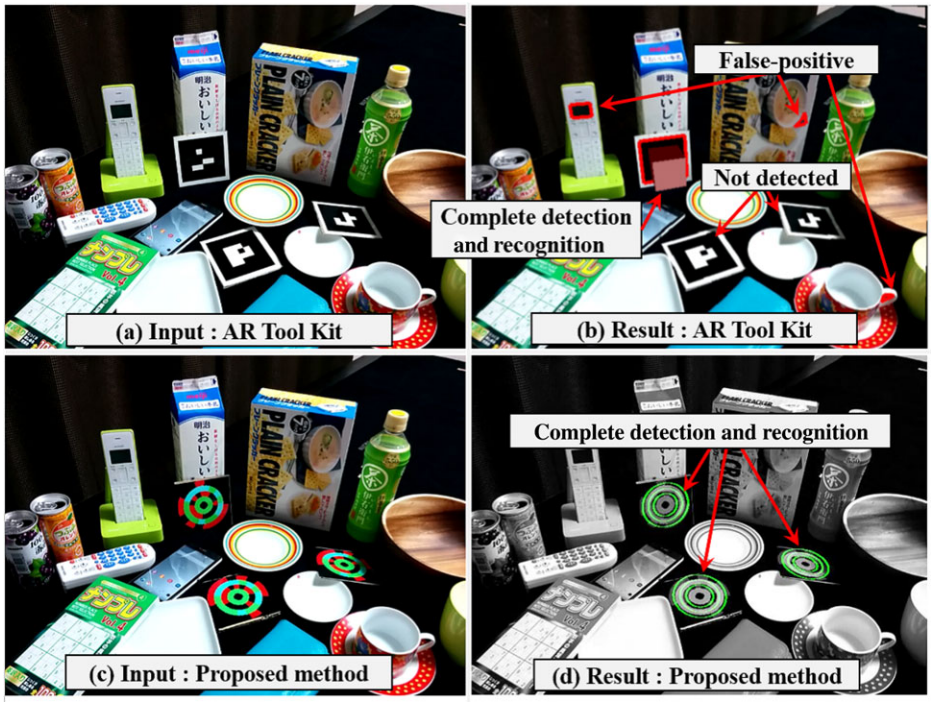


Fig. 1. Detection under a cluttered scene. [Color figure can be viewed at wileyonlinelibrary.com]

As described above, it is supposed that conventional markers are used under a condition where they can be easily detected. The detection under conditions is not considered well, where (1) markers are placed in various distances and by various arrangements in cluttered background, and (2) images are blurred.

3. Encoded Markers Easy to Detect Based on Scale-Space Theory

When obtaining information automatically from images, we cannot know in advance which scale should be used. In order to solve this problem, scale-space theory is proposed [9], where an image is represented as a set of scales with parameters so that all of the scales are processed simultaneously. Its feature is the marker detection applying scale-space extrema detection, which enables to easily detect marker's candidates independent on estimation of marker's size (scale) and its arrangement.

Detection

It is difficult to correctly detect markers placed by various arrangements. Meanwhile, a strong point of SIFT features is that they also utilize scale-space extrema detection (keypoint detection) and have invariableness to enlargement and reduction. If a marker is designed by applying this advantage, easy detection is expected to be possible without depending on distance from a camera to a marker. SIFT features also have invariableness to rotation and translation with description of information on gradient

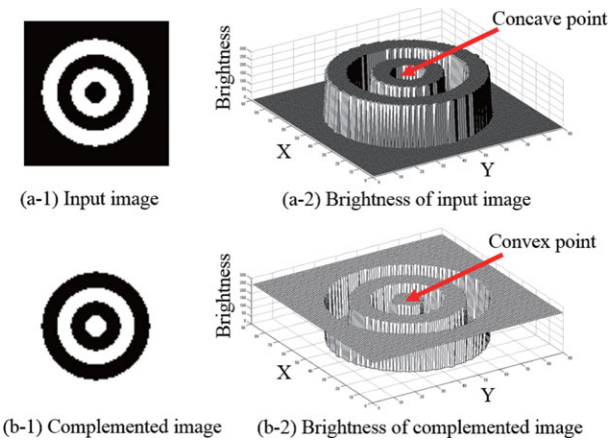


Fig. 2. Structure to detect by SIFT. [Color figure can be viewed at wileyonlinelibrary.com]

around keypoints. This is based on keypoints, so that they are also considered to have invariableness to rotation and translation. Namely, it is considered that easy detection can be realized independent on a marker's direction by applying scale-space extrema detection. Figures 2(a-1) and 2(b-1) show images of a detected marker and its brightness-inverted image, respectively. Further, Figs. 2(a-2) and 2(b-2) show three-dimensional drawings on brightness of (a-1) and (b-1), respectively. The scale-space extrema detection has a feature that it strongly responds to concave (local minimum) of brightness surrounded by convex of

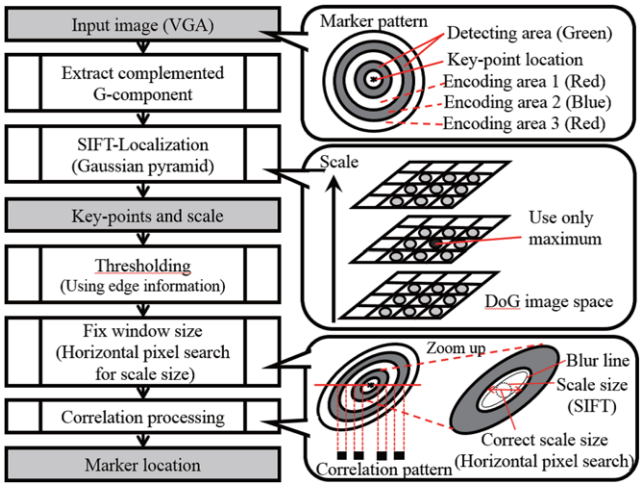
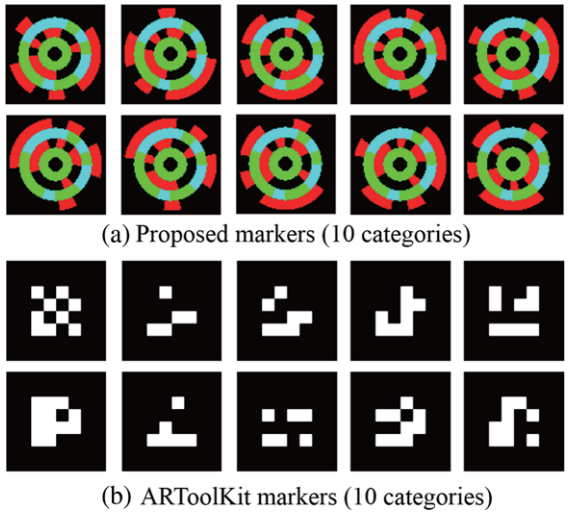


Fig. 3. Detection procedure. [Color figure can be viewed at wileyonlinelibrary.com]



(a) Proposed markers (10 categories)

(b) ARToolKit markers (10 categories)

Fig. 4. Comparison of marker's design pattern. [Color figure can be viewed at wileyonlinelibrary.com]

brightness shown in Fig. 2(a-1) and convex (local maximum) of brightness surrounded by concave of brightness shown in Fig. 2(b-1). Then, we design a marker using this feature. Specifically, it is a ring-shaped marker with a circle at its center. We develop a detection pattern such that its ring and center parts correspond to concave and convex of brightness, respectively. This marker has a pattern enabling easy scale-space extrema detection, so that the size of the marker can be estimated based on its calculated scale value.

Recognition

Pattern of marker is deteriorated if captured from far distance. Then, the encoding using M-sequence [11], which is known to have "high self-correlation and low mutual-correlation", is performed. With this encoding, a design is possible such that a steep correlation can be established between the measured encoded pattern and the designed encoded pattern, so that an encoded pattern on marker can be recognized (identified) precisely.

3.1 Marker detection

Detection process of marker is shown in Fig. 3. First, candidates of the marker are searched by the keypoint detection of SIFT. Second, the candidates of the marker are narrowed down for the acquired candidates through a discrimination filter. Finally, a decision process of marker is carried out.

3.1.1 Marker configuration

Figure 4(a) shows examples of proposed markers. As all of them have ring structures, therefore the keypoint detection of SIFT can capture the center of a marker. Further, a double-ring structure as a detection pattern is constructed using the G component in RGB color model, and its data pattern is represented with the R and B components. This double-ring structure by the G component is also used not only for the position detection of a marker, but also for correlation value calculation of pattern described in Section 3.1.4.

3.1.2 Narrowing of marker's candidates by keypoint detection

Only the G component is extracted from an input image, and then its brightness is inverted (a ring and a center of a marker correspond to concave and convex of brightness, respectively). After that, only convex of brightness (local maximum) is extracted from scale space (set of DoG images [8]). Then, a keypoint location and its scale are calculated. The calculation method is explained in a paper by Lowe [8]. At the beginning, we designed such that both of local maximum and local minimum were picked up as candidates, but only local maximum was selected as a candidate, so that image inversion is processed. Figure 5 shows candidate points of a marker by the keypoint detection of SIFT. While false candidates are also detected, it can be found that a keypoint is always at the center of a real marker in each image.

3.1.3 Narrowing of marker's candidates by a discrimination filter

Based on the information on gradient around keypoints, marker's candidates are further narrowed down according to the degree of gradient at edge in X and Y directions. Only the candidates with strong edge components in both of directions are extracted, and the others are removed.

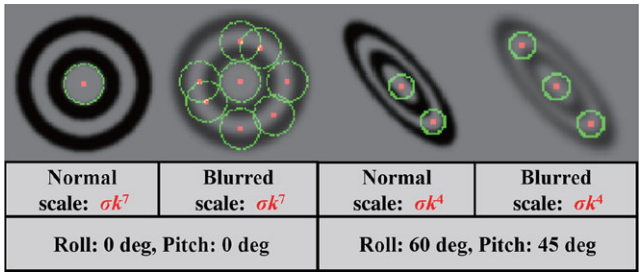


Fig. 5. Keypoint location and its accuracy. The red dot indicates a keypoint and the green circle indicates the scale size. σ is the initial scale and k is the step size of scale for SIFT. [Color figure can be viewed at wileyonlinelibrary.com]

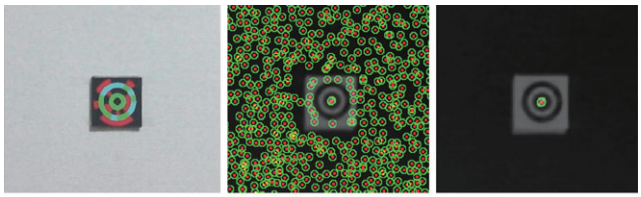


Fig. 6. Reduction of the candidates by edge information. [Color figure can be viewed at wileyonlinelibrary.com]

Its condition should be $|\mathrm{Det}(\boldsymbol{H})| > 1.0 \times 10\text{–}5$, where the equation for the discrimination is as shown in Eq. (1). Here, the matrix $\boldsymbol{H}$ is Hesse matrix for DoG images as shown in Eq. (2).

$$\mathrm{Det}(H) = D_{xx}D_{yy} - (D_{xy})^2 \tag{1}$$

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{2}$$

Figure 6 shows the result narrowed by a discrimination filter. It can be seen that only a marker is detected after a discrimination filter is applied, while multiple keypoints are detected on an uneven wall with concaves and convexes.

3.1.4 Marker judgment by pattern correlation

With locations of keypoints and their scales, a real marker is judged from candidates.

At first, scales are recalculated by the following procedure. The reason for the recalculation is because an error is caused between an estimated scale and an actual size of a marker if an arrangement of a marker is inclined for a camera. If a scale estimated in Section 3.1.2 is η, a mean brightness Φ is calculated in an area of 1.5 η square where its center is a keypoint location (x, y). After that, the numbers of successive pixels, larger than Φ, in the horizontal direction on left and right sides of the keypoint location (x) should be P_r and P_1, respectively. Then, a final scale value η is calculated by $(P_r+P_1)/2$. Here, the brightness of the location (x, y) is expressed by $I(x, y)$.

$$\begin{aligned} \hat{\eta} &= (P_r + P_l)/2 \\ P_r &= \sum_{i=1}^{0.75\eta} f(x+i, y) \quad P_l = \sum_{i=1}^{0.75\eta} f(x-i, y) \\ f(x,y) &= \begin{cases} 1 & (I(x,y) > \Phi) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned} \tag{3}$$

The decision as a maker is to compare a binarized pattern of a G component configured by a double ring with a designed ratio of a marker. A binarized pattern for the comparison has pixels equal to a marker size based on η with only one line in the horizontal direction with an estimated keypoint location as the center. Under this setting, a candidate satisfying Corr ≧ 0.8 should be decided as the marker. The correlation equation is as follows:

$$\mathrm{Corr} = \sum_{p=1}^{n} S_p O_p, \tag{4}$$

where S_p is the measurement pattern and O_p is the design pattern of the marker, and both of them are binarized to have the value +1 or –1. The symbol n [pixel] is the marker size calculated based on the estimated scale η. Because the marker shape is ring, the pattern ratio of a horizontal line of a shape passing through the center of ring is constant even if the shape is altered by the projective transformation. If the keypoint and the scale are correct, both of the pattern correlations agree with each other.

3.2 Marker recognition

Figure 7 shows a procedure of marker recognition. The decided marker is recognized as one of registered markers. A prior art [25] applies M-sequence encoding to detect angles of a turn table. It is used as a marker in this study.

3.2.1 Scanning of data area

The data on three ring parts (see Encoding area in Fig. 3) are encoded. In order to extract these parts, the fitting to ellipse is performed using the scale value η estimated in Section 3.1.

3.2.2 Data pattern generation using M-sequence

A pixel array of one round along a ring corresponds to one cycle of M-sequence. Using a biquadratic primitive polynominal, M-sequence is a row vector with $(2^4\text{-}1)$ elements of {1, –1}, and p, which is 16th element and –1 is added thereto, is utilized for encoding and decoding,

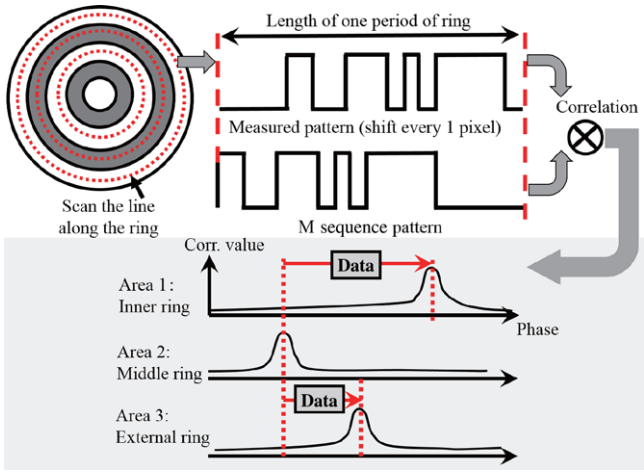


Fig. 7. Data decoding procedure. [Color figure can be viewed at wileyonlinelibrary.com]

where 1 and –1 correspond to the brightness 255 and 0, respectively.

3.2.3 Data representation by phase difference of M-sequence and decode processing

The phase-shifted patterns p equal to M-sequence are arranged to each of three rings (internal, middle, and external rings) shown in Fig. 7, and then data are represented by these phase differences. One round of the ring is set to 7 bit (128-bit resolution: about 2.8 degree/LSB). At decoding, convolution is operated between a measured pattern for one round and p with the same length of the measured pattern with regard to each ring by shifting by one pixel, where a shift amount is recorded at the point with the highest correlation value. Shift value at peak points for three rings is compared and they are calculated as phase differences. Uppermost 7 bit by the phase difference between the center ring and the internal ring, and lowermost 7 bit by the phase difference between the center ring and the external ring, namely, total 14 bit of data representation is possible.

4. Experiment and Consideration

4.1 Experimental condition

In order to confirm the effectiveness of the proposed marker, the proposed marker and the ARToolKit marker were compared with regard to the detection and recognition performances of marker. Specifically, the robustness to blur and various dispositions (distance and arrangement) under cluttered background for both of markers were compared. Further, in order to evaluate the robustness of the proposed marker to its geometrical disposition, the influence by distance and direction was analyzed. Markers corresponding to 14 bit (16,384 pieces) were registered for the proposed marker and the ARTookKit marker. In this experiment, 10 kinds of markers, as shown in Fig. 4, were utilized for both kinds of methods. The size of each marker is 9 cm square. This size adapted to the size of markers applied to a task to pick up objects from shelves by life support robots [6, 7]. A camera mounted in a smart phone (Nexus 5 by LG Electronics) was used to capture images and the resolution was 640×480 pixels. It is noted that a camera Cameleon by PointGrey, capable of controlling its open, was applied to evaluate the performance for blur. Similarly, the resolution was 640×480 pixels. The detection and recognition rates to the proposed marker and the ARToolKit marker realize 1 fps, respectively, using a PC mounting the CPU Corei5 by Intel, so that it is enough possible to start gripping operation after recognizing a marker.

In this experiment, if the distance between a detected point of the marker and its actual point was within 5 mm, the detection was regarded as successful detection, and if the data on the marker could be decoded correctly, it was regarded as successful recognition. In both the detection and recognition experiments, F value was used, which is the harmonic mean of precision and recall to the number of markers.

As parameters of scale-space extrema detection, the step size of scale was $k = 2^{1/3}$ and the initial scale was $\sigma = 1.2$, and 1 octave was stepped by three steps, then search was performed to three octaves (total nine steps).

4.2 Comparison on detection and recognition performances for blur

The experiment was carried out under the condition such that the degree of blur was varied by controlling the open of the camera. The positions of the camera and the markers were fixed, where the distance between them was 1 m. The measurements were implemented 100 times for

(a) Original image (b) Blurred image (c) Original image (d) Blurred image

Fig. 8. Blurred pattern used in the experiment. [Color figure can be viewed at wileyonlinelibrary.com]

Table 1. Performance against blur (*F*-measure)

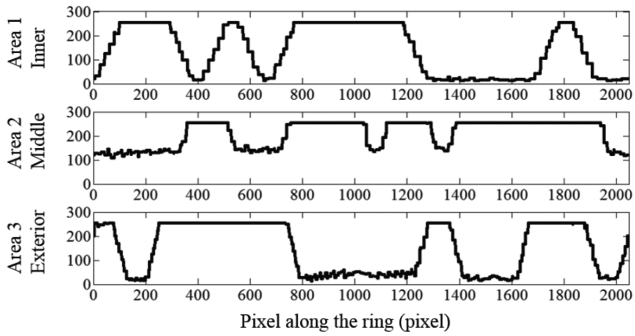
Marker	Detection	Decoding
ARToolKit	0.925	0.780
Proposed	1.000	0.978

each of 10 markers, and then total 1000 data were evaluated for each method. Figure 8 shows the blurred patterns. Table 1 further shows the result on detection and recognition. This table explained that the proposed marker has higher detection and recognition rates than the ARToolKit marker. It can be said from this result that the proposed marker is hardly influenced by blur. This consideration can be also seen from the points of (1) detection processing and (2) recognition processing.

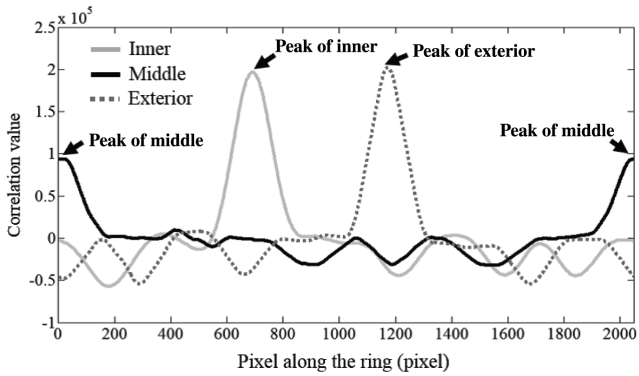
4.2.1 Detection performance at blur

Figure 5 shows the result that blurred images were made by simulation and then detected locations of keypoints were verified. It is obvious from these images that keypoints as marker candidates can be detected correctly on blurred images. It can be also found that markers with inclined arrangement can be detected correctly. As described above, it is considered that the detection performance of the proposed marker is deteriorated a little by blur. This is because extrema are searched in scale space. With regard to blur such that the point spread function can be approximated by the Gaussian function, no matter an image is blurred, a position of local maximum does not change, and thus the detection performance for the proposed marker is not influenced.

Meanwhile, ARToolKit detects a rectangle by extracting its outline, and then after measuring its length and breadth, if the rectangle is near a square, it is recognized as a marker. The reason why the detection performance of the ARToolKit marker is low is because the marker shape cannot be recognized as a rectangle after extracting its outline due to blur. It is considered that ARToolKit further searches a marker with a correlation value for brightness of its pattern so that the brightness variance causes its performance deterioration.



(a) M-squence signal in a blurred pattern.



(b) Decoded signal in a blurred pattern.

Fig. 9. Result of decoded signal against blurred pattern.

4.2.2 Recognition performance at blur

Data signals of marker and their convolutions to actual blurred images illustrated in Fig. 8 are shown in Figs. 9(a) and 9(b), respectively. The applied marker is the one on the lower right in Fig. 4(a), which is the blurred image. Referring to Fig. 9(a), it seems that the signals have waveforms generated by smoothing square waves due to blur of images. However, it can be found that three sharper peaks (external, middle, and internal rings) appear in Fig. 9(b). With the phase differences of these three peaks, decoding can be performed accurately. It is further obvious that encoding is not deteriorated by brightness applying M-sequence. From those described above, it is implied that the proposed marker is robust to blur with regard to the recognition performance.

4.3 Comparison of detection and recognition performances for various dispositions under cluttered background

Experiments were performed under five kinds of environments, where shadow of objects overlapped with markers under the cluttered background. Each marker (total 10 markers for one method) was disposed differently for every shooting by changing its distance from the camera

Table 2. Performance against ground clutter noise and placement (F-measure)

Marker	Detection	Decoding
ARToolKit	0.957	0.527
Proposed	0.990	0.921

and its arrangement, so that total 1000 images could be prepared for one method. The marker (the proposed and the ARToolKit markers) and the camera were fixed at the same positions, respectively. Table 2 shows the result. This table demonstrates that the proposed marker got significantly excellent results for both detection and recognition performances compared to the ARToolKit marker.

Figures 1(b) and 1(d) show examples of results on the ARTookKit marker and the proposed marker, respectively. Red frames are drawn to detected ARToolKit markers and three-dimensional objects are drawn if recognized. Meanwhile, if the proposed markers are detected and recognized, lines are drawn on the corresponding ring parts. As can be seen in these figures, the ARTookKit markers are misdetected in many points. The two ARToolKit markers in lower side of the figure are not detected. This is because patterns cannot be detected precisely due to the shadow by surrounding objects. The result in Table 2 reflects this cause remarkably. Therefore, it can be said that the ARToolKit marker is sensitive to brightness change. Meanwhile, the proposed marker can be detected accurately even in cluttered disposition, and it can be further recognized even if it is shadowed. From the next section, we consider the detection performance in cluttered background.

4.3.1 Detection performance in cluttered background

Figure 10 shows the marker candidates extracted through a discrimination filter explained in Section 3.1.2. Arrangements and scales of respective markers are different, but it can be found that the centers of these markers are accurately detected. The number of marker candidates through a discrimination filter could be controlled to between about 2 or 3 and about 10 per DoG image in scale space. The reason why the number of keypoints does not increase under such a cluttered background is because a discrimination filter is applied. With the proposed marker, the candidates can be narrowed down exactly. An excellent point to use the proposed marker is that the detection can be implemented in the same framework even if various arrangements and distances are applied to the marker. Namely, the comparison between a marker and a template is not required according to marker's arrangement and size.

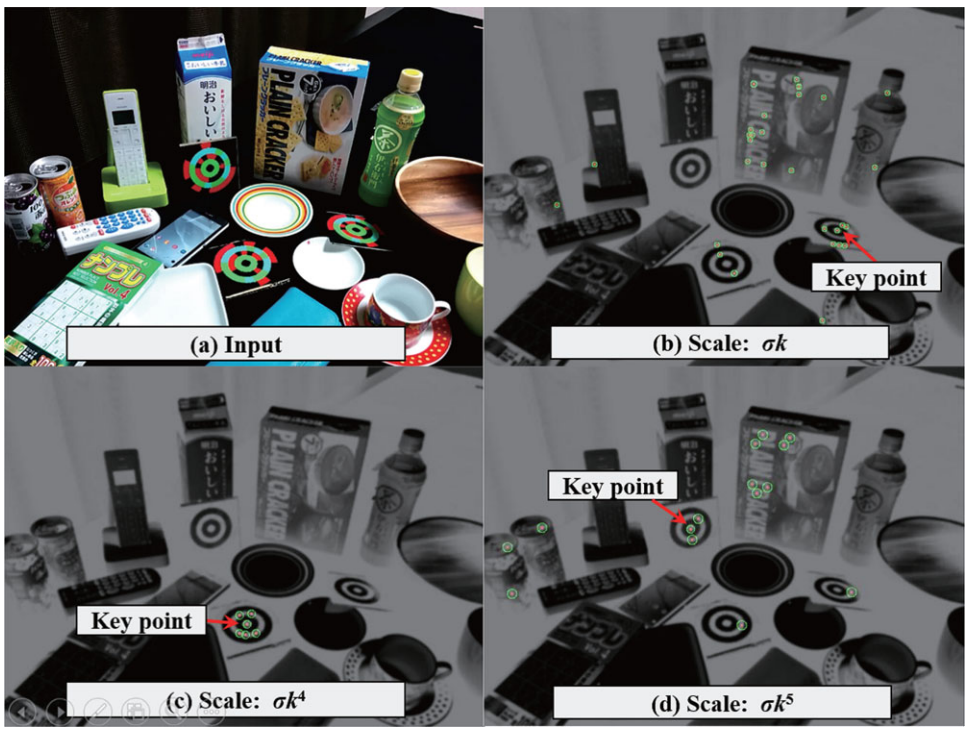
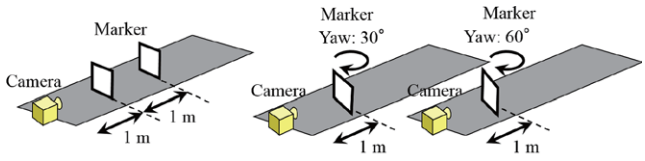


Fig. 10. Marker candidates by keypoint detection. The red dot indicates a keypoint and the green circle indicates the scale size. [Color figure can be viewed at wileyonlinelibrary.com]



(a) Distance: 1 m, 2 m (b) Angle: Yaw 30°, 60°

Fig. 11. Experimental environment. [Color figure can be viewed at wileyonlinelibrary.com]

Table 3. Evaluation by distance (F-measure)

Distance	Detection	Decoding
1 m	0.985	0.984
2 m	1.000	0.876

Table 4. Evaluation by angle at 1 m (F-measure)

Yaw angle	Detection	Decoding
0 degree	0.985	0.984
30 degree	1.000	0.970
60 degree	1.000	0.942

4.4 Analysis on detection and recognition performances to marker's distance

In order to investigate the robustness to distance of the proposed marker, performance evaluation was performed, where a marker was placed at 1 m and 2 m distance from a camera. Figure 11(a) shows an experimental environment. A marker was disposed such that it faced to front for a camera. A black nonreflective paper was set up so that nothing was shot except a marker. The size of the marker in image was 54 and 27 pixels for the distance of 1 m and 2 m, respectively. Table 3 shows its result on detection and recognition performances. For both of conditions, high detection and recognition performances could be obtained.

4.5 Analysis on detection and recognition performances to marker's angle

In order to investigate the robustness to angle of the proposed marker, performance evaluation was performed, where a marker was placed at 1 m distance from a camera and the marker was inclined by 0, 30, and 60 degree to the horizontal direction. Figure 11(b) shows an experimental environment. Table 4 shows its result on detection and recognition performances. The detection performance was not changed with larger angle. This is because the detection of keypoints is robust to the change of marker's direction [8]. As also shown in Fig. 5, the positions of keypoints can be detected precisely with inclined markers. Further from the result, a marker could be detected accurately even in the case of the inclination of 60 degree. From the results described above, we confirmed that marker's centers can be detected even if it is inclined to a camera.

5. Conclusions

In this paper, we proposed an encoded marker (1) easy to detect independent on blur and (2) easy to detect independent on background and disposition. Specifically, a pattern was designed which is easy to calculate as local maximum of scale space, and marker detection was realized by keypoint detection of SIFT. We further proposed a discrimination filter to narrow down marker's candidates and a method to correct scale error due to disposition. With these techniques, we demonstrated that, as the result of the experiment, a marker can be detected with extremely high accuracy robust to disposition and blur in cluttered background. We also confirmed that it is superior to the ARToolKit marker, which is widely used for robots and augmented reality.

Different from conventional markers, the comparison between this marker and multiple templates with different arrangements and sizes is not required. Therefore, we do not need to consider the scalability of computational cost due to disposition, and so on. The advantage is that marker's size can be estimated based on scale value of scale space, and therefore this is applied to the estimation of data decoding region of M-sequence code.

A challenge to be solved in the future is to enhance the recognition accuracy of marker with low resolution disposed at more distant location. At the detection experiment, the center of marker could be detected with high accuracy if its resolution was low. However, when a marker was inclined to a camera, recognition was failed in some cases because the fitting performance to ellipse and the accuracy to recalculate scales were not sufficient. Then, it is considered that an investigation of a method to represent data only with locations of keypoints is required. Specifically, we aim to investigate to combine RandomDot marker [26] with a marker structure proposed in this paper and M-sequence.

In Table 3, the detection accuracy of the proposed marker for 2 m was a little lower than that for 1 m, and in Table 4, that for 30 degree and 60 degree was a little lower than that for 0 degree. This is because the scale dealt in SIFT is discrete value and the size of marker with high sensitivity is changed according to scale step and initial scale value. Therefore, we need to investigate Spectral SIFT [28] which operates scale as continuous values. We further aim to compare the proposed method with the keypoint detection method without Gaussian kernel such as AKAZE

[27] and the keypoint detection method unchanged to affine transformation [29].

Acknowledgment

A part of this study was supported by JSPS KAKENHI Grant Number JP24240028.

REFERENCES

1. Rethink Robotics. Baxter http://www.rethinkrobotics.com/baxter/. Accessed on 09 January 2015.
2. Kawada Industries, Inc. NEXTAGE http://nextage.kawada.jp/ Accessed on 09 January 2015.
3. Amazon Picking Challenge. http://amazonpickingchallenge.org/ Accessed on 09 January 2015.
4. Hinterstoisser S et al. Gradient response maps for real-time detection of texture-less objects. IEEE Trans Pattern Anal Mach Intell 2012;34(5):876–888.
5. Brachmann E et al. Learning 6D object pose estimation using 3D object coordinates. 13th European Conf. Procs. Computer Vision, ECCV 2014, Part II, Lecture Notes in Computer Science, Vol. 8690, p 536–551, Springer, 2014.
6. Yamamoto T, Saito F, Hashimoto K, Ikeda K. 30th RSJ Conference Procs. on Development of Life Assistant Robot HSR, 3C2-1, 2012. (in Japanese)
7. Saito F et al. 30th RSJ Conference Procs. on Prototype of Life Assistant Robot HSR and Evaluation, 3C2-2, 2012. (in Japanese)
8. Lowe D. Distinctive image features from scale-invariant keypoints. Int J Comp Vis 2004;60(2):91–110.
9. Lindeberg T. Scale-space theory: A basic tool for analysing structures at different scales. J Appl Stat 1994;21(2):224–270.
10. Kobori N, Deguchi D, Ide I, Murase H. A proposal of encoding marker which is robust against blur and easy to detect. IEICE Tech Rep, IE2014-59, 2014. (in Japanese)
11. Matsuo K. Spectrum scattering technology. TDU Press; 2005. (in Japanese)
12. Denso Wave Magazine. QRcode.com http://www.denso-wave.com/qrcode/. Accessed on 09 January 2015. http://www.shift-2005.co.jp.
13. Takeda Y, Hiura S, Sato K. Integration of depth from defocus and stereo using coded aperture. IEICE Trans Inform Syst 2013;J96(8):1688–1700. (in Japanese)
14. Farsiu S, Robinson M, Elad M, Milanfar P. Fast and robust multiframe super resolution. IEEE Trans Image Process 2004;13(10):1327–1344.
15. Kato H, Billinghurst M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. Proceedings of 2nd IEEE/ACM International Workshop on Augmented Reality, p 85–94, 1999.
16. Shift, Inc. Chameleon Code http://www.shift-2005.co.jp/ Accessed on 09 January 2015. (in Japanese)
17. Color Code Laboratories, Inc. ColorZip http://www.colorzip.co.jp/. Accessed on 09 January 2015. (in Japanese)
18. Honda T, Kaneko S. Robust pattern matching technology using color feature and actual applications. OplusE 2013;35(12):1375–1380. (in Japanese)
19. Bcore, Inc. Color Bit http://www.colorbit.jp/. Accessed on 09 January 2015. (in Japanese)
20. Miyaoku K, Tang A, Fel S. C-Band: A flexible ring tag system for camera-based user interface. Proceedings of the 2nd International Conference on Virtual Reality, ICVR 2007, Lecture Notes in Computer Science, Vol. 4563, p 320–328, Springer, 2007.
21. Parikh D, Jancke G. Localization and segmentation of a 2D high capacity color barcode. Proceedings of 2008 IEEE Workshop on Applications of Computer Vision, p 1–6, 2008.
22. Tateno K, Kitahara I, Ohta Y. A nested marker for augmented reality. Proceedings of the 2nd International Conference on Virtual Reality, ICVR 2007, Lecture Notes in Computer Science, Vol. 4563, p 259–262, Springer, 2007.
23. Mohan A et al. Bokode: Imperceptible visual tags for camera based interaction from a distance. ACM Trans Graph 2009;28(3):Article No.98.
24. Toyoura M, Aruga H, Turk M, Mao X. Detecting markers in blurred and defocused images. Proceedings of International Conference on Cyberworlds 2013, p 183–190, 2013.
25. Ueda T. Measure position and angle by photographing experimental mark patterns. Interface, CQ Publishing; 2013. (in Japanese)
26. Uchiyama H, Saito H. Random dot markers. Proceedings of the 2011 Conference on IEEE Virtual Reality, p 35–38, 2011.
27. Alcantarilla P, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces. Proceedings of the 24th British Machine Vision Conference, p 1–11, 2013.
28. Koutaki G, Uchimura K. Applications to pattern matching using spectral theory and its performance evaluation. IEICE Trans Inform Syst 2013;J96(8):1664–1674. (in Japanese)
29. Hasegawa T et al. Multiple-hypothesis affine region estimation with anisotropic LoG filters. Proceedings of International Conference on Computer Vision 2015, p 585–593, 2015.

AUTHORS (from left to right)




Norimasa Kobori (nonmember) received a bachelor degree from the Department of Physics of the School of Science and Engineering of Waseda University in 2002. He completed the M.E. program at the Graduate School of the same university in 2004. He joined Sony Corporation in 2005 and then joined Toyota Motor Corporation in 2006. He is currently engaged in research on machine learning and image recognition at the Partner Robot Division of the same company. He completed the doctoral program at the Graduate School of Information Science of Nagoya University without degree in 2015.

Daisuke Deguchi (nonmember) received a bachelor degree from the Department of Information of the School of Engineering of Nagoya University in 2001. He completed the doctoral program at the Graduate School of Information Science of Nagoya University in 2006 (Doctor of Information Science). He was a postdoctoral fellow of Japan Society for the Promotion of Science from 2004 to 2006. He became a researcher at the Graduate School of Information Science of Nagoya University in 2006 and became a researcher at the Graduate School of Engineering of Nagoya University in 2006. He was a research associate at the Graduate School of Information and Science of the same university from 2008 to 2012. He became an associate professor at the Information Strategy Office of Nagoya University in 2012, where he currently holds the same position. He is a member of the Institute of Electronics, Information and Communication Engineers, and IEEE.

Ichiro Ide (nonmember) received a bachelor degree from the Department of Electronics Engineering of the School of Engineering of The University of Tokyo in 1994. He completed the M.E. program in information engineering at the Graduate School of Engineering of the same university in 1996. He completed the doctoral program in electric engineering at the same university in 2000 (D.Eng.), then he became a research associate at National Institute of Informatics. He became an associate professor at the Graduate School of Information Science of Nagoya University in 2004. He is a senior member of the Institute of Electronics, Information and Communication Engineers, and the Information Processing Society of Japan, and a member of the Institute of Image Information and Television Engineers, the Japanese Society for Artificial Intelligence, IEEE, and ACM.

Hiroshi Murase (nonmember) received a bachelor degree from the Department of Electricity of the School of Engineering of Nagoya University in 1978. He completed the M.E. program at the Graduate School of the same university in 1980 and then joined Nippon Telegraph and Telephone Public Corporation (currently NTT). He was a visiting researcher at Columbia University in the City of New York, USA from 1992 for 1 year. He has been a professor at the Graduate School of Information Science of Nagoya University since 2003 (D.Eng.). He is a fellow of IEEE, and the Institute of Electronics, Information and Communication Engineers, and a member of the Information Processing Society of Japan.