IEICE TRANS. INF. & SYST., VOL.E107–D, NO.10 OCTOBER 2024

PAPER

Reliable Image Matching Using Optimal Combination of Color and Intensity Information Based on Relationship with Surrounding Objects

Rina TAGAMI[†a)], ***Student Member***, **Hiroki KOBAYASHI**[†b)], ***Nonmember***, **Shuichi AKIZUKI**[†c)], ***and*** **Manabu HASHIMOTO**[†d)], ***Members***

SUMMARY Due to the revitalization of the semiconductor industry and efforts to reduce labor and unmanned operations in the retail and food manufacturing industries, objects to be recognized at production sites are increasingly diversified in color and design. Depending on the target objects, it may be more reliable to process only color information, while intensity information may be better, or a combination of color and intensity information may be better. However, there are not many conventional method for optimizing the color and intensity information to be used, and deep learning is too costly for production sites. In this paper, we optimize the combination of the color and intensity information of a small number of pixels used for matching in the framework of template matching, on the basis of the mutual relationship between the target object and surrounding objects. We propose a fast and reliable matching method using these few pixels. Pixels with a low pixel pattern frequency are selected from color and grayscale images of the target object, and pixels that are highly discriminative from surrounding objects are carefully selected from these pixels. The use of color and intensity information makes the method highly versatile for object design. The use of a small number of pixels that are not shared by the target and surrounding objects provides high robustness to the surrounding objects and enables fast matching. Experiments using real images have confirmed that when 14 pixels are used for matching, the processing time is 6.3 msec and the recognition success rate is 99.7%. The proposed method also showed better positional accuracy than the comparison method, and the optimized pixels had a higher recognition success rate than the non-optimized pixels.

key words: *image recognition, object detection, template matching, genetic algorithm, pattern matching*

1. Introduction

The demand for image processing technology has become even higher due to the increased production of semiconductors and the shift to labor-saving and unmanned operations in the retail and food manufacturing industries. With this situation, the objects to be recognized are becoming more diverse, and the color and design of objects vary widely. Therefore, image processing systems need to accurately recognize objects of various colors and designs.

The issue is to improve the versatility of the system and the diversity of the training data. For example, when considering reliable matching in object recognition, the versatility of image processing can be improved by using intensity information for the object, as shown in Fig. 1 (a), because of its black or gray design, and by using color information for the object, as shown in Fig. 1 (b) because of its green or red color. In particular, in the case of the object shown in Fig. 1 (c), since the design consists of black and pink, the recognition accuracy is expected to be the highest when both intensity and color information are used. However, most conventional image processing techniques convert color images to grayscale images before processing from the viewpoint of speed, which results in lower recognition accuracy when the object consists of colored objects. Although there are pattern matching methods that use information from color images, it is extremely difficult to achieve both high speed and reliability in matching. A possible solution is object recognition using deep learning, but this method has strong limitations in terms of computer hardware and memory at the production site. In the case of food package recognition, in particular, there are many types of objects to be recognized, and preparing a huge amount of training data is not cost-effective.

In this paper, we consider using template matching (TM) [1]–[3], which is often used in pattern matching and object recognition. TM is frequently used in production because it does not require a large amount of training data and has a certain degree of robustness. In addition, keypoint matching (KPM) [4]–[6] has low recognition accuracy depending on the frequency of the image. However, the lack of such limitations is an advantage of TM.

The purpose of the proposed method is to optimize the combination of pixels used for matching in accordance with the color and design of the target object and its surrounding objects, and we propose a fast and reliable method.

(a) (b) (c)

Fig. 1 Example of food packages. (a) is achromatic, (b) is chromatic colors, and (c) is object with design that includes both achromatic and chromatic colors.

Manuscript received February 15, 2024.
Manuscript publicized May 30, 2024.
[†]Chukyo University, Nagoya-shi, 466–8666 Japan.
a) E-mail: tagami@isl.sist.chukyo-u.ac.jp
b) E-mail: kobayashi@isl.sist.chukyo-u.ac.jp
c) E-mail: s-akizuki@sist.chukyo-u.ac.jp
d) E-mail: mana@isl.sist.chukyo-u.ac.jp
DOI: 10.1587/transinf.2024EDP7039

The idea of the proposed method is to select effective and unambiguous pixels from the target image and to optimize the combination of color and intensity information while carefully selecting a small number of pixels that are highly discriminative from the surrounding objects. The optimization uses a genetic algorithm, in which the combination of a pixel and the pixel's color and intensity is varied from generation to generation. This enables fast and reliable matching with high robustness against surrounding objects. This research enables stable matching even when the design colors of the target and surrounding objects are chromatic or achromatic, thus contributing to the automated recognition of objects with a wide range of color variations.

This paper is organized as follows: Sect. 2 describes related work and their problems. Section 3 describes the proposed method, and Sect. 4 describes the experimental results of the proposed method and a comparative method. Section 5 provides a conclusion of the proposed method. A preliminary version of this paper appeared in ISVC2023 [7].

2. Related Work and Problems

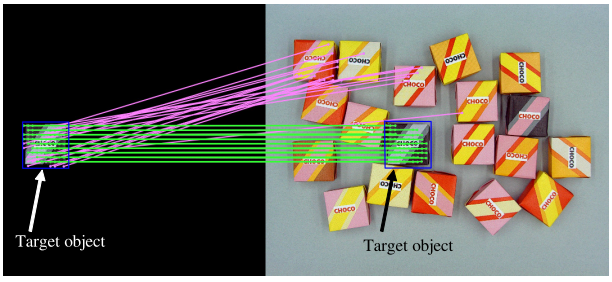
Various methods of object detection have been proposed, but in a production line, where computational resources are limited, methods that are understandable to the user, fast, and reliable are preferred. For this reason, KPM [8] and TM [3], which are both simple and practical, are often used on production lines. TM, in particular, is a relatively simple algorithm with low memory requirements, and a Field Programmable Gate Array (FPGA) can be used to speed up similarity calculations [9].

SIFT [10] is a well-known method for KPM. It is a rotation-invariant feature, but it is very costly in terms of generating Difference of Gaussian (DoG) images and calculating gradient information. AKAZE [11] is an improved method, but it detects feature points for each input image, which is fast but requires a certain amount of processing time. Various methods with higher speed and accuracy have been proposed for TM. There are methods [12] that use only the edge pixels of an object, methods [13] that detect edges that change little in each frame over time, and flexible matching ones [14] that use a segmented set of edges. However, they can only be used when the image frequency is high and edge information is sufficiently extractable.

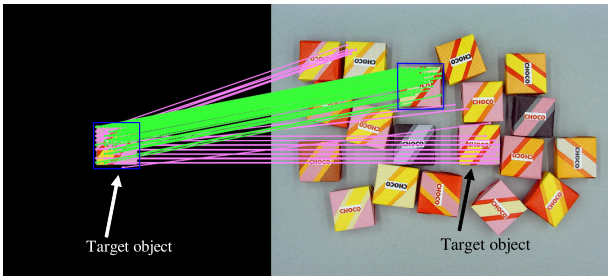
Other methods include Best-Buddies-Similarity (BBS) [15], [16], which counts the number of best buddies by splitting the target and search images into small patches for each RGB value and calculating similarities between the two patch images, and Deformable Diversity Similarity (DDIS) [17], which improves the similarity calculation of BBS [15] and reduces the computational cost, thus enabling faster processing. Occlusion Aware Template Matching (OATM) [18] transforms the TM problem from the original high-dimensional vector space to a search problem between two smaller sets and achieves high speed by using random grid hashing. We consider the number of similarity calculations to be a bottleneck for these methods [15]–[18], making it difficult to achieve a practical level of speed.

There are also methods [19], [20] that reduce the search area and ones [21], [22] that reduce the number of pixels used for matching to achieve high speed. Fast Affine Template Matching (FAsT-Match) [19] is fast because only pixels with smooth intensity values are used in the process, while Co-occurrence Probability Template Matching (CPTM) [21] achieves both high speed and high reliability by using only pixels with low pixel value co-occurrence. Color Co-occurrence of Multiple Pixels in Template Matching (CoPTM) [22], which extends CPTM [21] to color information, focuses on the co-occurrence of hue values and achieves high speed in color images. Co-occurrence based Template Matching (CoTM) [23], which uses co-occurrence histograms of quantized RGB values for similarity calculation, enables matching that is robust to deformations. However, it is not very versatile, because it is limited by the target object images: FAsT-Match [19] is not reliable for high-resolution images, FAsT-Match [19] and CPTM [21] are unreliable for color images, and CoPTM [22] and CoTM [23] are unreliable for grayscale images.

Recently, feature matching [24]–[27] that applies "Transformer" has also been proposed, in which global information in one image and potential matches are utilized on the basis of an analysis of the relationship between two images. And, recent methods [28] on deep learning for matching often focus on learning better features and descriptors from images using Convolutional Neural Networks (CNNs). When an object is different in color from the surrounding objects, it can be recognized as shown in Fig. 2 (a), but when the objects are similar, it is misrecognized as shown in Fig. 2 (b).



(a) Example of correct recognition



(b) Example of misrecognition

Fig. 2 Successes and failures in recognition by related method (LoFTR) [25]. (a) is the result of correct recognition, (b) is the result of misrecognition. The proposed method makes these two cases recognizable.

This is because learning is performed using only information from grayscale images.

The problems of the related work can be summarized into the following three main categories. First, the existence of limitations due to the color and design of the image. This means that the reliability of matching is reduced when the image is of a certain frequency or color. Second, even the state-of-the-arts (SOTA) method misrecognizes objects that are similar to the recognized target if they exist in the surroundings. Third, even if a method is capable of reliable matching, it lacks real-time processing. The proposed method that can solve these problems is explained in Sect. 3.

3. Proposed Method

3.1 Basic Idea

In this paper, we consider three requirements and corresponding ideas to solve the problems described in Sect. 2. The first requirement is general versatility in package design. The recognition success rate should not vary depending on the object design. For versatility, we use a combination of color and intensity information. The second requirement is to obtain robustness of the target object to its surrounding objects. A target object should not be misrecognized as its surrounding objects. For robustness, we use the information of the target object and its surrounding objects, and select pixels with low commonality between the target object and its surrounding objects. The third requirement is speed. The efficiency of the process can be improved by reducing the number of pixels used for matching. On the basis of these three ideas, we propose a method, in which an overview is given in the next section.

3.2 Overview of Proposed Method

Figure 3 shows an overview of the proposed method. The proposed method consists of three modules. First, a color image and a grayscale version of the image are prepared; the two images are input to Module A, which selects pixels that are effective for matching as color information and pixels that are effective for matching as intensity information from each images. By combining the respective pixels, there are discretely effective pixels in terms of color information and effective pixels in terms of intensity information, which improves the object's design versatility. Next, the combined pixels are input to Module B, where a genetic algorithm (GA) is used to select pixels that are highly robust to surrounding objects. Optimization methods such as greedy algorithm could be considered, but they are prone to local solutions, so GA was used, which is suitable for solving complex problems. Selected pixels are then used for matching in Module C to enable fast and reliable matching. The aforementioned process produces a small number of pixels that are robust to surrounding objects by combining color and intensity information. By optimally combining these pixels, we consider that the versatility of the object package design is increased

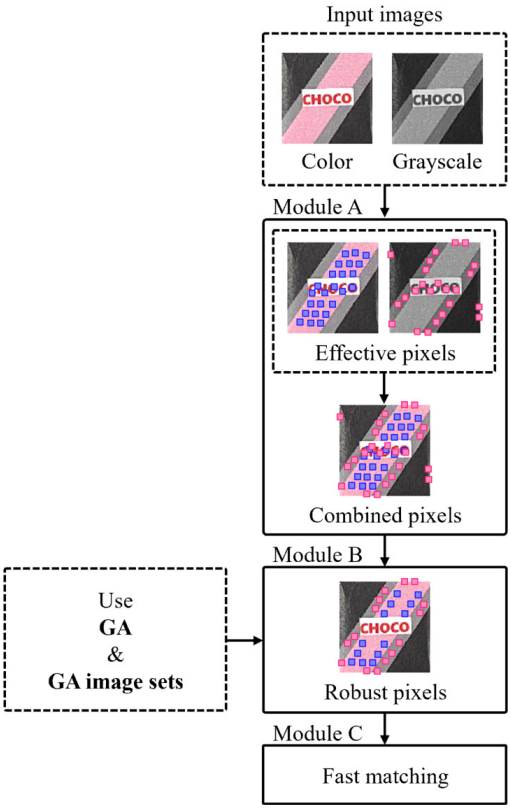


Fig. 3 Overview of proposed method.

and that matching only a small number of pixels enables fast processing.

Module A is described in detail in Sect. 3.3, Module B in Sect. 3.4, and Module C in Sect. 3.5.

3.3 Method of Combining Color and Intensity Information

The proposed method uses the co-occurrence of two pixels in the target image (starting pixel P and ending pixel Q) as an indicator of pixel distinctiveness. First, pixel pairs are applied to all locations in the image in a raster scan, and the values p and q of P and Q, respectively, are used as indexes to vote for the number of occurrences in a 2-dimensional matrix. After all pixel pairs have been voted on, a co-occurrence histogram is completed. Pixel pairs (displacement vectors d) can have several patterns of pixel distances, but in this case we used the patterns d = 1, 2, 4, 8, 16 in the horizontal and vertical directions. The more of these patterns there are, the more the spatial frequency of the image can be represented.

The specific process of the method is shown in Fig. 4. First, from the grayscale image of the target object, frequencies of the occurrence of intensity values per pixel pair are calculated, and from the color image, frequencies of the occurrence of hue values (Hue values of HSV, quantized to 256 levels) per pixel pair are calculated. Hue values can represent color information in a single channel and can be converted from RGB values at high speed. From each generated co-occurrence histogram, the pixel pairs with the lowest

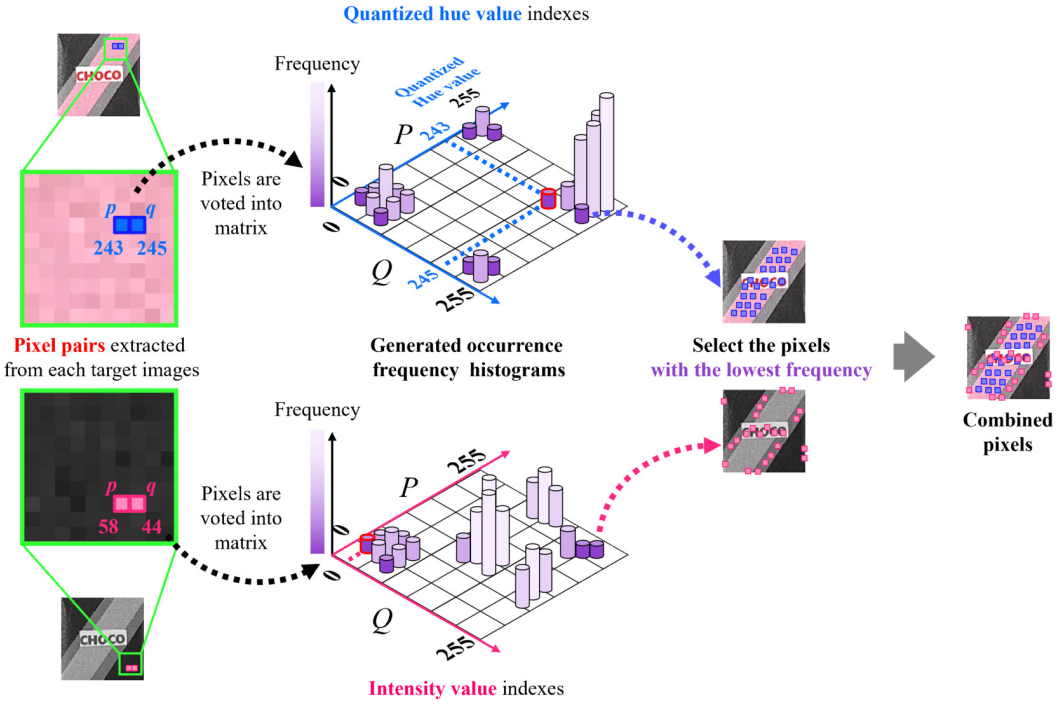


Fig. 4 Process of selecting effective pixels for matching in the target images.

occurrence frequency are selected.

In other words, the pixel pair occurrence frequency F_r is defined by Eqs. (1), (2), and (3), where P and Q are pixel pairs, p and q are pixel pair values, $\boldsymbol{v}_P = (x_P, y_P)$ and $\boldsymbol{v}_Q = (x_Q, y_Q)$, respectively, the displacement vector of Q relative to P is $d = (k, l)$, $\mathcal{P}$ is the position vector in the entire target object image, and $f(\cdot)$ is a value for a coordinate. Pixels with a high frequency of occurrence are patterns that often appear in the image, and by not selecting them, it is possible to carefully select only those pixels that are effective for matching. These pixels are disambiguated, and a certain reduction in mismatches can be expected. In addition, the selected pixels are a mixture of pixels that are valid as color information and pixels that are valid as intensity information, as shown in the left side of Fig. 4, leading to improved versatility in the design and colors of the object.

$$F_r(p,q) = \sum_{\boldsymbol{v}_P, \boldsymbol{v}_Q \in \mathcal{P}} \sum \delta(\boldsymbol{v}_P, \boldsymbol{v}_Q, p, q) \tag{1}$$

$$\delta = \begin{cases} 1 & if \quad \{f(\boldsymbol{v}_P) = p\} \wedge \{f(\boldsymbol{v}_Q) = q\} \\ 0 & otherwise \end{cases} \tag{2}$$

$$where, \quad \boldsymbol{v}_Q = \boldsymbol{v}_P + \boldsymbol{d} \tag{3}$$

3.4 Selection of Pixels that are Robust to Surrounding Objects

In the proposed method, the discriminative performance with surrounding objects is considered. Pixels that are not common to both the target object and the surrounding objects are selected, and should not be mismatched with the surrounding objects. The first step of the process is to evaluate the selected pixels using a set of assumed images prepared in advance, as shown in the upper right corner of Fig. 5. The group of assumed images consists of two types: the group of images for the target objects, and the group of images for the surrounding objects. Then, similarities between selected pixels and the two image groups are calculated. If pixels are selected in the process of Sect. 3.1 as pixels with infrequent intensity values, the similarities are calculated on the basis of their intensity value. If pixels are selected as pixels with infrequent hue values, the similarities are calculated on the basis of hue value. Using the calculated similarities, histograms are generated with similarity C_i on the horizontal axis and the frequency of similarity C_i on the vertical axis (bottom center of Fig. 5), and three evaluation indices D, S, and p_{max} calculated from these histograms are used to evaluate the discrimination performance of the selected pixels by Eq. (4).

The larger the evaluation value F, the higher the discrimination performance, meaning that the values of D and p_{max} should be large and the value of S should be small. w_1, w_2, and w_3 are weighing factors, and ϵ is a supplementary factor. The reselection of pixels in the GA and the evaluation of discrimination performance are repeated so that the evaluation value F becomes larger. This automatically determines the optimal combination of color and intensity information depending on the target object and surrounding objects.

$$F = w_1 \frac{1}{S + \epsilon} + w_2 D + w_3 p_{max} \tag{4}$$

Next, the evaluation indices are described in detail. The first evaluation index is the difference between the mean values of the histograms of the target and surrounding object

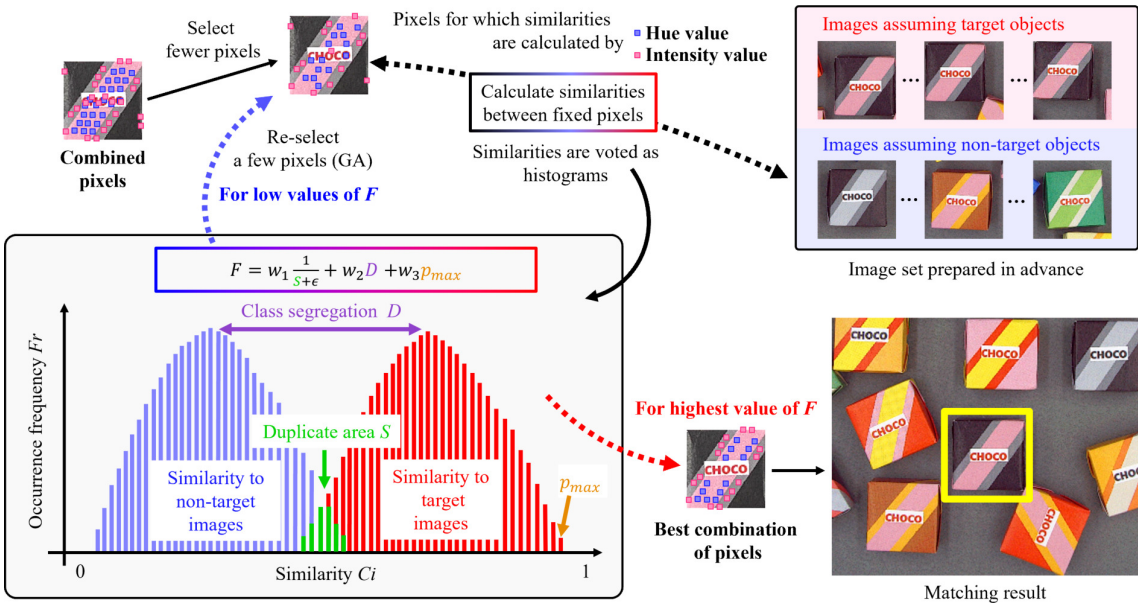


Fig. 5 Selection of robust pixels against surrounding objects and matching.

images, defined as the degree of class segregation D. The larger the class segregation D is, the further apart the histogram distributions of the two image groups are, and thus better discriminability can be expected from the thresholding process.

The second index is the overlap between the histograms of the two image groups, defined as the overlap area S. The smaller this area is, the smaller the risk of mismatching with surrounding objects.

The third evaluation index p_{max} is the value with the highest similarity among the target images of assuming. This value can suppress the stagnation of the evaluation value F when the histogram distribution of the surrounding object images has a higher similarity.

The pixel with the largest evaluation value F can be judged to be superior. The proposed method evaluates the discriminability of pixels using the aforementioned idea, and finally determines pixels with a certain level of goodness from a practical standpoint as an approximate solution.

3.5 Matching Method

The object detection is performed by calculating the similarity with the search image using the pixels selected by the process in Sect. 3.4. The similarity is calculated while raster-scanning the search image, and the best match position is the one with the highest similarity. Sum of Squared Differences (SSD) is used to calculate the similarity, and Sequential Similarity Detection Algorithm (SSDA) is used to speed up the calculation.

In the process flow, selected pixels $f(n)$ are stored as either intensity values or hue values in a 1D array $f_G(n)$ or $f_H(n)$. The i-coordinate and j-coordinate of the selected pixel are also stored in a 1-dimensional array as $f_i(n)$ and $f_j(n)$. The sum of the squares of the difference between the pixels of the target image and search images when they are shifted by (δ_x, δ_y) pixels and superimposed on each other is calculated using Eqs. (5) and (6). In this case, the value of the search image is $g(i, j)$, and it switches between intensity value and hue value depending on the information $f_R(n)$ of the object image. The information $f_R(n)$ is given to the target object image as 0 if the pixel has an intensity value and 1 if the pixel has a hue value. In this case, the number of selected pixels is M.

$$S_{SSD} = \sum_{n=0}^{M-1} (g(f_i(n) + \delta_x, f_j(n) + \delta_y) - f(n))^2 \tag{5}$$

$$f(n) = \begin{cases} f_G(n) & if \quad f_R(n) = 0 \\ f_H(n) & if \quad f_R(n) = 1 \end{cases} \tag{6}$$

4. Experiments and Discussion

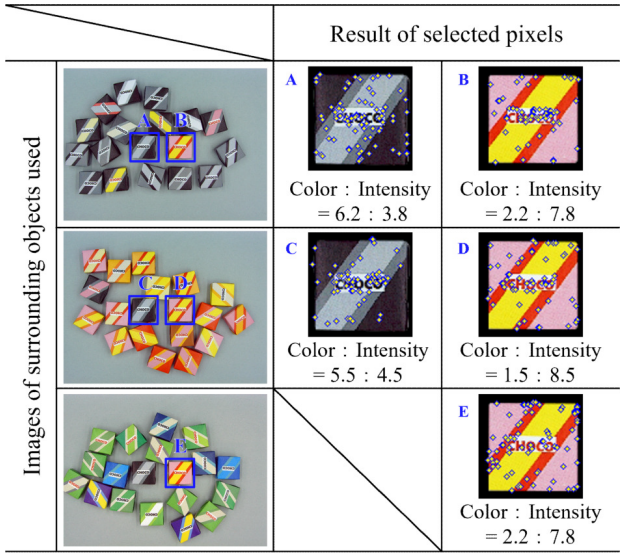
4.1 Implementation Details

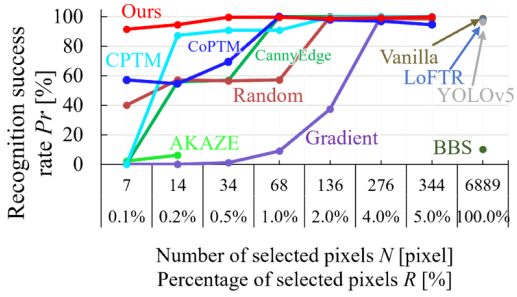
We conducted a verification experiment after discussing the pixels selected by the proposed method. The recognition performance of the proposed method was compared with that of the SOTA methods. In the GA of the proposed method, the number of genes in each individual is the number of combined pixels. The population size is 2500. The crossover, mutation and selection methods are uniform crossover, bit-flip mutation and roulette wheel selection, respectively. Furthermore, the crossover and mutation probabilities are 0.98 and 0.02, respectively. The number of max generations for the search is 50,000. The weights w_1, w_2, and w_3 of the evaluated value are 0.49, 0.49, and 0.02, respectively. All experiments are conducted on 64 GB RAM and AMD Ryzen 5 5600X.

Table 2 Relationship between the ratio of color and intensity information of selected pixels and the number of successful recognitions.

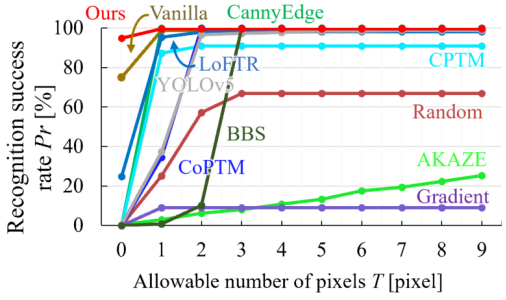
Color : Intensity	0:10	1:9	**1.5 : 8.5**	3:7	4:6	5:5	6:4	7:3	8:2	9:1	10:0
Recognition success rate	299	505	**598**	511	473	493	292	246	571	570	425

Table 1 Relation between target/surrounding objects and selected pixels.





(a) Recognition success rate for each number of selected pixels



(b) Recognition success rate for each tolerance pixel

Fig. 6 Comparisons matching performance of our method with other methods.

4.2 Result of Pixels Selected by Proposed Method

Table 1 shows pixels selected by the proposed method. The upper row of Table 1 shows the selected pixels of objects A to E (83 × 83 [pixels]), and the lower row shows the color and intensity ratio of the selected pixels. The vertical axis is the image of the surrounding objects used, and the rectangles in the image show the positions of targets A to E. We used 300 images (645 × 484 [pixels]) of such surrounding objects and selected the best combination of pixels using GA. In the case of target A, the intensity information of the target and surrounding objects is similar. Therefore, the intensity information is not effective for matching, and the color information in the selected pixels is used more frequently. Conversely, in the case of target D, the combination of intensity information was higher because the surrounding objects and target are particularly similar and the color information is not effective for matching. When the design of the target and the surrounding objects were not similar, as in the case of objects B and C, the combination of selected pixels was due to the design of the surrounding objects. In the case of target E, many pixels were selected from the red pattern. This is because red patterns do not exist in the surrounding objects. The percentage of color information in the selected pixels is higher for target E than for target D. This is because the surrounding objects of target E are not color-similar to target E. From the aforementioned, it can be said that the proposed method optimizes the selection of pixels on the basis of the mutual relationship between the target and surrounding objects.

4.3 Relationship between Optimal Combination of Pixels and Number of Successful Recognitions

As a verification experiment, we confirmed the relationship between the recognition success rate of the optimal combination of pixels obtained by the proposed method and pixels that are not optimal (pixels optimized by fixing the ratio of the combination). Table 2 shows the result. The pink cells show the ratio of color and intensity in the selected pixels. The blue cell shows the number of successful recognitions, and in the case of 600, the recognition success rate is 100%. The number of pixels used in this case was fixed at 68. The recognition success rate (99.6%/number of successful recognitions, 598) was the highest for the pixel with the best color/intensity combination (in red), indicating that the pixels obtained by the proposed method were optimal.

4.4 Performance Comparisons

Figure 6 (a) shows the recognition success rate for each number of selected pixels for the proposed method and the comparison methods [11], [15], [21], [22], [25], [29], [30]. The

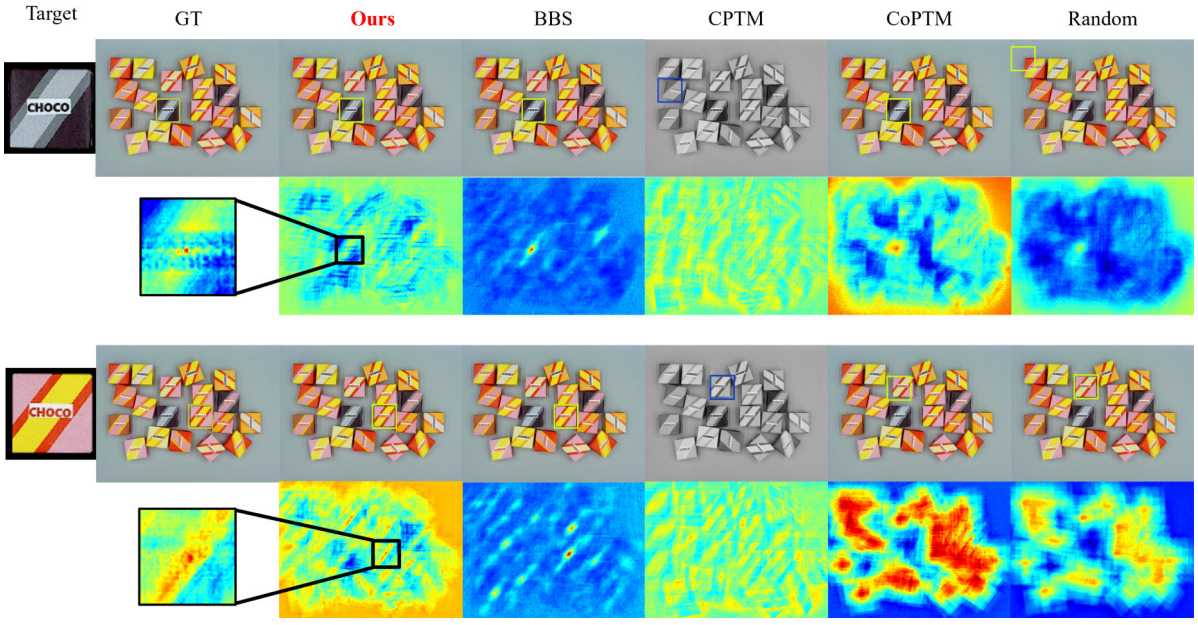


Fig. 8 Matching results and similarity heatmaps.

results shown in Fig. 6 use the target image of D in Table 1 and surrounding object images of the middle row. The horizontal axis N is the number of selected pixels in 83×83 pixels, which is the size of the target image, and the bottom row shows the ratio of the number of selected pixels in the target image, R. The vertical axis P_r is the recognition success rate (GT $\pm$ 2 pixels) when 600 search images were used. Even when the proposed method used only 34 pixels for matching, the recognition success rate was 99.7%. This indicates that the proposed method is capable of stable matching even with a small number of pixels, independent of the color and intensity of the target object and its surroundings. Figure 6 (b) shows the recognition success rates of the proposed method and comparison methods for each tolerance pixel. Tolerance pixels are the number of pixels that can be considered as successfully recognized no matter how many pixels are off from the GT. Compared with BBS and LoFTR, which is a SOTA feature assignment method, the proposed method had a higher recognition success rate even when the number of tolerance pixels was small. This indicates that the proposed method can match with high positional accuracy. Figure 7 shows the processing time t of the proposed method and comparison methods for each selected pixels N. (i) is the processing time for CPTM, CannyEdge, Gradient and (ii) is that for CoPTM, Random. The processing time of the proposed method was the shortest compared with the other methods. The proposed method shows that fast and reliable matching is possible even with a small number of pixels. Figure 8 shows the matching results and a similarity heatmap. The redder the color in the heat map, the higher the similarity. It can be seen that the proposed method had low similarity in the surrounding area and high similarity only in the area where the object was located. BBS, the SOTA for TM, has a similar heatmap, but the matching process took

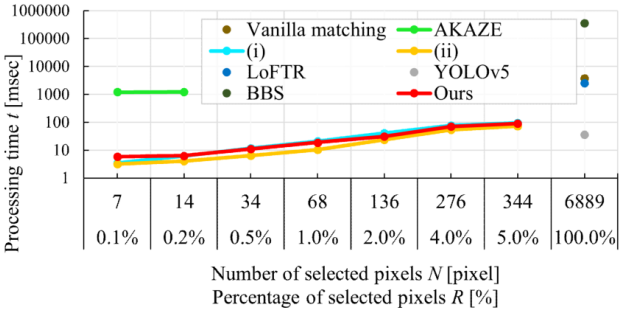


Fig. 7 Comparisons of processing time of our method with other methods.

about 300 s, so the proposed method has a better balance of speed and reliability.

4.5 Results on Real Data and Limitations

Because the proposed method targeted the recognition of products in production, we experimented with our own dataset, but to analyze the limitations of the proposed method, we experimented with a subset of the OTB dataset [31]. This dataset was annotated with bounding boxes for each frame and included challenging issues such as large deformations, occlusion, scale changes, illumination changes, blurring, and cluttered backgrounds. The subset consists of 105 images, and the image group used for pixel selection was images from frames not included in the subset. Figure 9 (a)–(h) show the target image in the upper left, the pixels selected by the proposed method in the upper right, the recognition results of the proposed method (green line) and the GT (red line) in the middle, and the BBS in the bottom. The proposed method was highly robust in the presence of similar target objects (in the case of this image,

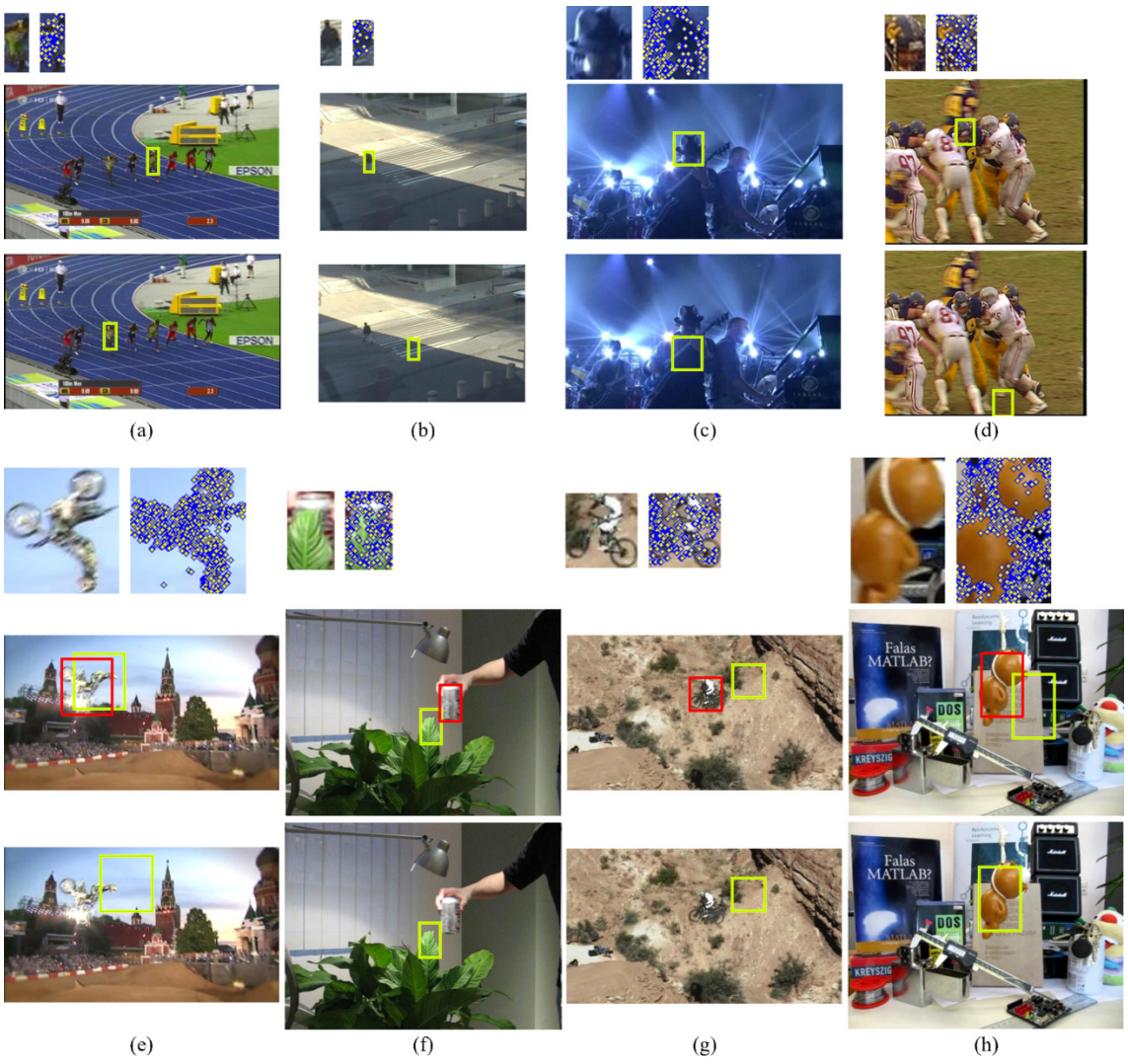


Fig. 9 Recognition success and failure. Upper left: target image, Upper right: selected pixels. Middle: recognition result (Proposed method), Bottom: recognition result (BBS). The green line is the output result, and the red line is the GT.

a person) as in (a), and in illumination changes as in (b) and (c). It was also robust to small deformations such as in (d). Although the proposed method misrecognized or misplaced in (e)–(g), in all of (a)–(g), BBS misrecognized the target, so the proposed method is more versatile than BBS.

5. Conclusion

We proposed a fast and reliable matching method that optimizes the combination of color and intensity of selected pixels. Experimental results showed that the optimal combination of pixels changed depending on the color and intensity of the target object and its surroundings. The optimal pixels had the highest recognition success rate compared with other combinations of pixels, and the average processing time was 6.3 msec when the number of selected pixels was 14. The method can contribute to the automation of recognition for objects with a wide range of colors because it reduces the misrecognition of peripheral objects and enables reliable matching.

Acknowledgements

This work was partially supported by JSPS KAKENHI, Grant-in-Aid for Scientific Research (C), Grant Number 21K03984.

References

[1] L. Cole, D. Austin, and L. Cole, "Visual object recognition using template matching," Australian Conference on Robotics and Automation, Dec. 2004.
[2] M. Sun, J. Xiao, E.G. Lim, B. Zhang, and Y. Zhao, "Fast template matching and update for video object tracking and segmentation," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.10791–10799, 2020.
[3] J. Cheng, Y. Wu, W. AbdAlmageed and P. Natarajan, "QATM: Quality-aware template matching for deep learning," Proc.

IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.11553–11562, 2019.
[4] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Lecture Notes in Computer Science, vol.3951, pp.404–417, 2006.
[5] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," Tenth IEEE International Conference on Computer Vision (ICCV'05), vol.2, pp.1508–1515, Oct. 2005.
[6] P.F. Alcantarilla, A. Bartoli, and A.J. Davison, "KAZE features," Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Lecture Notes in Computer Science, vol.7577, pp.214–227, Oct. 2012.
[7] R. Tagami, H. Kobayashi, S. Akizuki, and M. Hashimoto, "Reliable matching by combining optimal color and intensity information based on relationships between target and surrounding objects," Advances in Visual Computing, ISVC 2023, Lecture Notes in Computer Science, vol.14362, pp.163–176, Oct. 2023.
[8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," 2011 International Conference on Computer Vision, pp.2564–2571, 2011.
[9] Z. Chen, S. Li, N. Zhang, Y. Hao, and X. Zhang, "Eye-to-hand robotic visual tracking based on template matching on FPGAs," IEEE Access, vol.7, pp.88870–88880, 2019.
[10] D.G. Lowe, "Object recognition from local scale-invariant features," Proc. Seventh IEEE International Conference on Computer Vision, vol.2, pp.1150–1157, Sept. 1999.
[11] P.F. Alcantara, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," British Machine Vision Conference (BMVC2013), 2013. https://doi.org/10.5244/C.27.13
[12] M.-P. Dubuisson and A.K. Jain, "A modified Hausdorff distance for object matching," Proc. 12th International Conference on Pattern Recognition, vol.1, pp.566–568, 1994.
[13] J. Xiao and H. Wei, "Scale-invariant contour segment context in object detection," Image and Vision Computing, vol.32, no.12, pp.1055–1066, 2014.
[14] Q. Yu, H. Wei, and C. Yang, "Local part chamfer matching for shape-based object detection," Pattern Recognit., vol.65, pp.82–96, 2017.
[15] T. Dekel, S. Oron, M. Rubinstein, S. Avidan, and W.T. Freeman, "Best-buddies similarity for robust template matching," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2021–2029, 2015.
[16] S. Oron, T. Dekel, T. Xue, W.T. Freeman, and S. Avidan, "Best-buddies similarity—Robust template matching using mutual nearest neighbors," IEEE Trans. Pattern Anal. Mach. Intell., vol.40, no.8, pp.1799–1813, 2017.
[17] I. Talmi, R. Mechrez, and L. Zelnik-Manor, "Template matching with deformable diversity similarity," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.175–183, 2017.
[18] S. Korman, S. Soatto, and M. Milam, "OATM: Occlusion aware template matching by consensus set maximization," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2675–2683, 2018.
[19] S. Korman, D. Reichman, G. Tsur, and S. Avidan, "Fast-match: Fast affine template matching," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2331–2338, 2013.
[20] J. Lai, L. Lei, K. Deng, R. Yan, Y. Ruan, and Z. Jinyun, "Fast and robust template matching with majority neighbour similarity and annulus projection transformation," Pattern Recognit., vol.98, 107029, 2020.
[21] M. Hashimoto, T. Fujiwara, H. Koshimizu, H. Okuda, and K. Sumi, "Extraction of unique pixels based on co-occurrence probability for high-speed template matching," 2010 International Symposium on Optomechatronic Technologies, pp.1–6, Oct. 2010.
[22] R. Tagami, S. Eba, N. Nakabayashi, S. Akizuki, and M. Hashimoto, "Template matching using a small number of pixels selected by distinctiveness of quantized hue values," International Workshop on Advanced Imaging Technology (IWAIT) 2022, vol.12177, pp.662–667, April 2022.
[23] R. Kat, R. Jevnisek, and S. Avidan, "Matching pixels using co-occurrence statistics," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.1751–1759, 2018.
[24] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.4937–4946, 2020.
[25] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "LoFTR: Detector-free local feature matching with transformers," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.8922–8931, 2021.
[26] G. Bökman and F. Kahl, "A case for using rotation invariant features in state of the art feature matchers," Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.5110–5119, 2022.
[27] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "LightGlue: Local feature matching at light speed," arXiv preprint arXiv:2306.13643, 2023.
[28] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.224–236, 2018.
[29] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol.PAMI-8, no.6, pp.679–698, 1986.
[30] G. Jocher, K. Nishimura, T. Mineeva, and R. Vilariño, "yolov5," https://github.com/ultralytics/yolov5, 2020.
[31] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2411–2418, 2013.

Rina Tagami received her BE degree in Engineering from Chukyo University, in 2022. She is currently an ME student at Chukyo University under the supervision of Prof. Manabu Hashimoto. Her research interests include pattern matching, evolutionary algorithm and natural language processing. She is a member of JSPE.

Hiroki Kobayashi received his ME degree in Engineering from Chukyo University, in 2023. He is currently a PhD student at Chukyo University under the supervision of Prof. Manabu Hashimoto. His research interests include anomaly detection, pattern recognition and deep learning. He is a member of IEEE.

Shuichi Akizuki received his PhD degree in Information Science from Chukyo University, in 2016. He got research fellowship for young scientists, DC2 and PD, from JSPS, in 2016. From 2017 to 2018 he was a research associate in Faculty of Science and Technology, Keio University. He is currently a lecturer in School of Engineering, Chukyo University. His research interests include computer vision, robot vision and point cloud processing. He is a member of JSPE and RSJ.

Manabu Hashimoto received a ME and PhD degrees from Osaka University in 1987, 2000, respectively. In 1987, he began working at Mitsubishi Electric Corp. While working at the Advanced Technology R&D Center, he was engaged in the research and development of robot vision, three-dimensional object recognition, and the like. In 2008, he became a professor of Dept. of Mechanical and Information Engineering at Chukyo University. Since 2021, he is working as the director of Human Robotics Research Center and the dean of the Graduate School of Engineering. He received the RSJ Technical Innovations Award in 1998, the Odawara Award of JSPE, and the 2022 IWAIT Best Paper Awards. He is a member of Robotics Society of Japan, IEEE and others.