Extraction and Recognition of Shoe Logos with a Wide Variety of Appearance Using Two-Stage Classifiers

Kazunori AOKI\textsuperscript{(a)}, Nonmember, Wataru OHYAMA\textsuperscript{(b)}, and Tetsushi WAKABAYASHI\textsuperscript{(c)}, Members

SUMMARY A logo is a symbolic presentation that is designed not only to identify a product manufacturer but also to attract the attention of shoppers. Shoe logos are a challenging subject for automatic extraction and recognition using image analysis techniques because they have characteristics that distinguish them from those of other products; that is, there is much within-class variation in the appearance of shoe logos. In this paper, we propose an automatic extraction and recognition method for shoe logos with a wide variety of appearance using a limited number of training samples. The proposed method employs maximally stable extremal regions for the initial region extraction, an iterative algorithm for region grouping, and gradient features and a support vector machine for logo recognition. The results of performance evaluation experiments using a logo dataset that consists of a wide variety of appearances show that the proposed method achieves promising performance for both logo extraction and recognition.

key words: logo extraction, logo recognition, maximally stable extremal regions, gradient histogram features, support vector machine

1. Introduction

A logo is a symbolic presentation that is designed not only to identify a product manufacturer but also to attract the attention of potential buyers. Manufacturers carefully design their logos so that their characteristics, impressions and philosophies are expressed. Moreover, logos on a person’s belongings can play an important role in characterizing and identifying the person. Extraction and recognition of logos from images captured by multiple surveillance cameras could provide useful information for identification and tracking of individuals.

Automatic extraction and recognition of shoe logos using image analysis techniques is challenging because shoe logos have characteristics that distinguish them from the logos of other products, and their appearance can vary substantially. Figure 1 shows examples of shoe logos captured by standard still cameras. The logos shown in Fig. 1 (a) and (b), which belong to the same company, have the same shape but different colors. Figure 1 (c) and (d) are examples of logos consisting of multiple components. Figure 1 (e) and (f) show the most common appearance variations of shoe-logos images, i.e., rotation, occlusion, and perspective distortion. Automatic extraction and recognition techniques must handle these problems properly, because shoes are usually worn on feet and, therefore, move frequently with respect to stationary cameras. Additionally, since shoe logos are usually appeared as an integrated design component of shoe design, they have significant within-class variation due to the variation in color, fabric material, shape of shoe and dirt or aging of shoe.

In the present paper, we propose an automatic extraction and recognition method for shoe logos using a limited number of training samples. The proposed method employs maximally stable extremal regions (MSERs) \cite{1} for the initial region extraction, an iterative algorithm for region grouping and gradient features, and two-stage support vector machines (SVM) for logo recognition. For performance evaluation, we use the IEICE pattern recognition and media understanding (IEICE-PRMU) shoe logo dataset \cite{2}. This dataset consists of shoe logo images captured in uncontrolled condition.

The main contributions of this research are the following:

1. We propose a method which extracts and recognizes shoe logos which contains wide within-class variety due to the reasons mentioned above.
2. In order to improve the performance of extraction and recognition, two classifiers of which objectives are different are introduced.
3. For performance evaluation, IEICE PRMU shoe logo dataset is used. While the dataset is originally created for a competition, its property where real shoe logos captured in uncontrolled condition enables us to eval-

\begin{figure}[h]
\centering
\begin{subfigure}{0.3\textwidth}
\includegraphics[width=\textwidth]{a.png}
\caption{(a)}
\end{subfigure}
\begin{subfigure}{0.3\textwidth}
\includegraphics[width=\textwidth]{b.png}
\caption{(b)}
\end{subfigure}
\begin{subfigure}{0.3\textwidth}
\includegraphics[width=\textwidth]{c.png}
\caption{(c)}
\end{subfigure}
\begin{subfigure}{0.3\textwidth}
\includegraphics[width=\textwidth]{d.png}
\caption{(d)}
\end{subfigure}
\begin{subfigure}{0.3\textwidth}
\includegraphics[width=\textwidth]{e.png}
\caption{(e)}
\end{subfigure}
\begin{subfigure}{0.3\textwidth}
\includegraphics[width=\textwidth]{f.png}
\caption{(f)}
\end{subfigure}
\caption{Examples of shoe logos \cite{2}: (a) and (b) examples belonging to the same brand but with different colors, (c) and (d) examples containing multiple connected components, (e) rotation, and (f) occlusion and rotation.}
\end{figure}
ulate the method for logo detection and recognition in real scenario.

2. Related Works

The extraction and recognition of logos in images has attracted the attention of many researchers. Several studies on the automatic extraction of logos have been reported. In this section, we provide brief review for logo detection and recognition methods and discussion about relation between the present and conventional works.

Affine and non-rigid transformation occur frequently in real-world images. This makes logo detection and recognition complex, especially for model-based approaches, owing to the difficulty in collecting sufficient samples to obtain a robust model. To address this problem, Farajzadeh et al. [3] proposed an exemplar-based method for logo or trademark recognition. Their method uses new sample synthesizing which generates training logo images from standard logos with different tilts and rotations. This approach employs a two-stage strategy where initial candidate regions are extracted using efficient sub-window search (ESS) [5]. These regions are then recognized using a linear SVM trained using new synthesized samples. The dataset for classifier training includes logo images synthesized using gamma correction, difference of Gaussian filtering, equalization of variation, and size normalization. The main disadvantage of ESS is the significant tradeoff between recognition accuracy and false-positive detection. When the number of synthesized samples is increased to improve recognition accuracy, the number of false-positive detections drastically increases.

Chu et al. [4] proposed a method using visual patterns. This approach first extracts scale-invariant feature transform features [6] from both test images and a logo image. Features with high similarity in both the test images and logo image are found using locality sensitive hashing (LSH) [7]. The main purpose of their method is to improve computational efficiency by eliminating outliers in a test image obtained from an exhaustive sliding-window search. The extracted feature points are combined using the mean-shift algorithm [8] to extract local logo regions. The extracted regions are classified using visual word histograms (bag of words) and visual patterns. This method obtains high computational efficiency at the sacrifice of extraction accuracy. The authors reported that the method obtains only 19.0% recall and 30.0% precision.

As another application of LSH for logo detection, Romberg et al. proposed the bundle min-hashing [9], [10]. This method calculates features which are more robust to appearance variation than single visual word by bundling visual word and spacial neighborhood features. This approach outperformed existing another approaches in logo recognition performance by combining synthetic data augmentation.

While these three methods were carefully designed for handling appearance variations due to affine and non-rigid transformation which occur frequently in real-world images, they did not pay enough attention for large intraclass variations. Therefore, it is quite difficult for these methods to extend for detection and recognition of shoe logos which have significant appearance variations.

Logo detection and recognition can be categorized as a subproblem of object retrieval. In the literature of object retrieval, evaluating similarity between two images, a query and a gallery, is important. Shen et al. [11] proposed spatially-constrained similarity measure (SCSM) for large-scale object retrieval. SCSM could handle object rotation, scaling, view point change and appearance deformation. Furthermore, based on the retrieval and localization results of SCSM, they introduced a robust re-ranking method with the k-nearest neighbors of the query for automatically refining the initial search results. While high performance of this method on object retrieval and object categorization was confirmed by experiments, such high performance may not be expected for shoe logos due to its variation of design and colors.

Deep neural network architectures have been employed in image recognition because of their promising performance and high adaptivity. Girshick et al. [12] proposed a method called regions with convolutional neural network (R-CNN) for generic object recognition. R-CNN is an approach that combines selective search [13] and CNN. The method extracts initial regions using efficient graph-based image segmentation [14]. It then iteratively groups regions using similarity calculated from appearance features (color, texture, size and fill features) in the regions. Although R-CNN achieved a high performance score on PASCAL2010 [15], the method requires a large-scale, accurately annotated dataset for training. Creating such a dataset for shoe logos is difficult because they have a wide variety of appearances even in the same brand.

To overcome the problem of deep-learning based method where the performance depends on the size and quality of training dataset, synthesizing dataset is the most common approach. Su et al. [16] proposed an algorithm for generating synthetic context logo (SCL) training images for deep learning technique. SCL training data generation method synthesis logo images in context by overlaying a transformed logo exemplar at a random location in non-logo context images to deal with unknown background clutters. While this method is considered to be applicable for shoe logo detection and recognition, difficulties on preparing exemplar of shoe logo of which design is integrated with shoe itself will happen.

From the dataset perspective, several datasets for evaluating logo detection and recognition have been proposed. LOGO-Net proposed by Hoi et al. [17] is one of the largest dataset for logo detection and brand recognition. To facilitate deep-learning based logo detection, they built a large scale dataset by exhaustively collecting images from online retailer website. However, the inherent property of these images in which the background is simple and plane is some-
times different from real-world image. To the best of our knowledge, IEICE-PRMU shoe logo dataset used for evaluation in the present paper is one of the most difficult dataset which consists of real-world shoe images.

The authors also proposed shoe logo extraction and brand recognition method which employs MSER based region extraction and verification and recognition by single stage SVM [18]. This our previous report is the first published paper which employed IEICE PRMU dataset for quantitative evaluation for logo detection and recognition. In the present paper, the authors propose an extended version of the previous method, by introducing the second stage SVM which verifies that the recognition result by the first stage SVM is valid.

3. Proposed Method

Figure 2 shows the outline of the proposed method. In the proposed method, we input one still red–green–blue (RGB) color image. The method outputs regions in which a logo appears and the class (name of brand) corresponding to each region. The proposed method consists of main two stages: extraction of the shoe logos and recognition of the extracted logo regions. The following sections describe each stage in detail.

3.1 Extraction of Shoe Logos

The purpose of this stage is to extract all regions that might contain a logo from an input image. The extraction stage consists of an initial region extraction using MSERs and iterative region integration using histogram intersection.

3.1.1 Initial Region Extraction

The initial region extraction for shoe logos employs MSERs [1] for each color plane of the input RGB image. The MSER method is a region segmentation based on pixel dilation in the present paper is one of the most difficult dataset which consists of real-world shoe images.

The input and output of the region integration are a set of regions \( R^{0} \) that maximize histogram intersection \( H(r_{i}, r_{j}) \) such that the smallest logo is extracted correctly. Variations in another parameters contained in the algorithm of MSER do not influence the result of initial region extraction.

3.1.2 Iterative Region Integration

The above extraction process extracts single connected components as initial regions. Because some logos contain multiple connected components, we perform region integration using histogram intersection to extract these multiple connected components as one integrated candidate region. The basic concept of region integration is similar to hierarchical grouping algorithm in selective search [13]. The method continues grouping candidate regions until the number of MSER becomes one. Figure 3 shows an example of this process. The region integration works iteratively as follows.

The input and output of the region integration are a set of extracted initial regions \( R^{0} = \{ r_{1}, \cdots, r_{n} \} \) and a set of integrated regions \( R^{C} \), respectively. In the example shown by Fig. 3, \( r_{1} \) to \( r_{4} \) in (b) are the initial regions extracted from (a).

1. Initialize \( R^{C} \) as \( R^{0} \):

\[
R^{C} = R^{0}.
\] (1)

Let \( k = 0 \) and proceed to the next step.

2. Determine two different regions \( (r_{i}, r_{j}) \in R^{k} \), \( (i \neq j) \) that maximize histogram intersection \( H(r_{i}, r_{j}) \):

\[
(r_{i}, r_{j}) = \arg \max H(r_{i}, r_{j}).
\] (2)

The histogram intersection \( H(r_{i}, r_{j}) \) is calculated by

\[
H(r_{i}, r_{j}) = \sum_{l=0}^{L-1} \min(h_{l}^{(i)}, h_{l}^{(j)}),
\] (3)

where \( h_{l}^{(i)} \) and \( h_{l}^{(j)} \) denote the \( l \)-th value of histograms
region integration is repeated until adequate logo region is obtained. which provide the maximum histogram intersection at each iteration. The lines in each subfigure denote the combinations of connected components extracted regions and their grayscale histograms are shown. The connected components of multiple connected components (e.g. the Yonex logo). In each subfigure, Fig. 3

Iterative region integration (2.1.2) example for a logo consisting of multiple connected components (e.g. the Yonex logo). In each subfigure, extracted regions and their grayscale histograms are shown. The connected lines in each subfigure denote the combinations of connected components which provide the maximum histogram intersection at each iteration. The region integration is repeated until adequate logo region is obtained.

obtained in regions \( r_i \) and \( r_j \), respectively. The connected components which construct a multi-component logo have similar color distribution. The histogram intersection is expected to properly capture this similarity. In the example (Fig. 3 (b)), \( r_1 \) and \( r_2 \) are selected as the regions which maximize the histogram intersection value.

3. Subtract \( r_i \) and \( r_j \) from \( R^k \) and create \( R^{k+1} \) using \( R^{k+1} = R^k - \{r_i, r_j\} \). Create a new integrated region \( r^{(ij)} = r_i \cup r_j \) and add \( r^{(ij)} \) to \( R^{k+1} \) and \( R \).

4. Increment \( k \) and iterate Steps 2 to 4 until the number of elements in \( R^k \) equals one. In the example, sub figures (c) and (d) show the components and integration at each iteration.

5. Region \( R^C \) is the final integrated region result.

We obtain the color histogram in (3) from each RGB plane with 16 bins. For example, when regions are detected in the red color-plane image, we obtain a color histogram using the red value.

3.2 Recognition of Logos

The extracted candidate regions are evaluated and classified in the recognition stage. We employ a two-stage strategy using two SVMs, one for classification (SVM-C) and one for verification (SVM-V). These SVMs are used consecutively for logo recognition, as shown in Fig. 2. We employ the grayscale gradient histogram features [19] and an SVM in the recognition stage.

We first clip the input image using the detected logo candidate regions and convert them to a fixed size (50 × 50 pixels). We extract gradient histogram features from the grayscale of the original image using the process in [19] with 5 × 5 spatial sub-blocks and 16 quantization directions. The dimensionality of extracted feature vectors is 400. These parameters are determined by our initial investigation with the training dataset, where we maximized the recognition performance.

Figure 4 shows the result of our initial investigation. Here, we evaluated the classification performance of SVM by three-fold cross validation over the training dataset. From the results, it is confirmed that the classification accuracy is influenced by the dimensionality of feature vectors. Since the dimensionality of feature vectors are calculated by the product of sub-block number and quantization directions, we tested several combination of parameters for classification accuracy. The best accuracy (indicated by A in the figure) is obtained when the dimensionality of feature vector is 400.

The role of SVM-C in the first stage is to estimate the posterior probability where the input logo candidate clipped by the extracted candidate region belongs each logo category. When we have \( n \) categories to classify, SVM-C outputs \( n + 1 \) posterior probability values, where \( n \) logo categories and one “non-logo” category.

The SVM-V in the second stage verifies the input logo candidate region using the posterior probability values estimated by SVM-C in the first stage. When a logo candidate region accurately captures a shoe logo, the posterior probability value of the logo class corresponding to the captured logo is expected to become large while others are low. In contrast, if the candidate region does not contain the logo or the class to classify the logo is confusing, the values of the posterior probability is expected to show a unique distribution. The task of SVM-V is to input \( n + 1 \) posterior probability values as a feature vector and to classify the vector into logo or non-logo classes.

SVM-C and SVM-V employ radial basis function kernels and are trained as multiple-class and two-class classifiers, respectively. The training schemes of the SVMs are described in the following section.
4. Evaluation Experiments

We conducted experiments to evaluate the effectiveness of the proposed method.

4.1 Dataset

We used the IEICE pattern recognition and media understanding shoe logo dataset [2]. The dataset consists of 661 images and ground-truth annotations are given for each image. The logos of eight brands with a wide variety of appearances are contained in the dataset. Figure 5 shows the appearance variations of the shoe logos in the dataset. Some images were not captured under controlled conditions and the images contain blur, rotation, occlusion, and perspective distortion. Samples of the eight brands in the dataset are shown in Fig. 6.

We employed a three-fold cross-validation for performance evaluation. The dataset was divided into three groups at random. Two were used for training and the remaining one was used as a test set. We conducted this evaluation ten times and calculated the mean performance.

4.2 Training of Classifiers

As described above, we employ a two-stage method using two SVMs, and the objective of each is different. Each SVM is trained using independent training data.

SVM-C in the first stage is trained with logo images extracted using the ground-truth annotation. The training SVM-C dataset contains 11 logo classes. The dataset originally contains eight brands, but three of these, ASICS, Mizuno, and New Balance, have two types of appearance. Thus, we divide each of these into two separate classes. The images belonging to the “non-logo” class is extracted from detected regions that do not overlap with annotated logo regions.

For training SVM-V in the second stage, we collect the dataset containing posterior value vectors of logo and non-logo regions. Applying the algorithm for logo candidate region extraction described in 3.1 to training images, we obtain several candidate regions. For logo verification, in which the extracted candidate region is identified as containing or not containing a logo, negative samples are necessary for SVM training. The negative samples were generated from detected regions that do not overlap the annotated logo regions.

4.3 Evaluation

We evaluated the overall performance of the method using recall $R$, precision $P$, and $F$-measure $F$. These are defined by the following:

$$R = \frac{1}{N_{GT}} \sum_{i=1}^{N_{GT}} \delta(S^{(i)}_c, S^{(i)}_o)O(S^{(i)}_c, S^{(i)}_o),$$

(4)
Using average best overlap (ABO), calculated by

\[ P = \frac{1}{N_{\text{DET}}} \sum_{j=1}^{N_{\text{DET}}} \delta(S_c^{(j)}, S^{(j)}), \]

(5)

\[ F = \frac{2PR}{P + R}, \]

(6)

We also evaluated the detection performance of the method using average best overlap (ABO), calculated by

\[ ABO = \frac{1}{N_{\text{GT}}} \sum_{k=1}^{N_{\text{GT}}} O(S_c^{(k)}, S^{(k)}), \]

(7)

where the overlap rate between regions \( S_c \) and \( S_o \) (shown in Fig. 7) is determined by

\[ O(S_c, S_o) = \frac{A(S_c \cap S_o)}{A(S_c) + A(S_o) - A(S_c \cap S_o)} \times 100. \]

(8)

For the above calculation, \( N_{\text{GT}} \) and \( N_{\text{DET}} \) denote the number of ground-truth regions and detected regions, respectively. Furthermore, \( S_c^{(i)} \) in the recall calculation in (4) is the detected region with the minimum distance from the \( i \)-th ground-truth \( S^{(i)} \). Similarly, \( S_c^{(i)} \) is the selected ground-truth region with respect to the detected regions. Regions \( S_c^{(k)} \) and \( S^{(k)} \) in the overlap ratio calculation are determined as the highest overlap regions, and \( \delta(S_c^{(i)}, S^{(i)}) \) denotes

\[ \delta(S_c, S_o) \begin{cases} 1 & \text{(class labels of } S_c \text{ and } S_o \text{ are same)}, \\ 0 & \text{(otherwise)}. \end{cases} \]

(9)

After \( P, R \) and \( F \) are calculated for each test image, we calculate the evaluation value by averaging over all test images.

To compare the extraction and recognition performances of the proposed method with those of other techniques, we implemented R-CNN and adopted the same conditions as used for the proposed method. CNN features are computed by forward propagating a mean-subtracted 50×50 RGB image through two convolutional layers and two fully-connected layers. We then obtain a 400-dimensional feature vector using the CNN.

5. Results and Discussion

Table 1 compares the results of logo extraction performance. Each row in the table denotes the method employed for logo extraction. The naive MSER method listed in the first row is adopted for preprocessing the grayscale image. The second and third rows show the extraction performance of the proposed method. Although the initial region extraction using three color-plane images outperforms the naive MSER, iterative integration of the extracted initial regions further improves the region extraction performance. The fourth row shows the ABO obtained by a selective search based method proposed by Uijlings [13]. These results show that the MSER algorithm extracts logos more efficiently when it is applied to color-plane images. The ABO is increased by introducing the proposed iterative region integration method. This suggests that logos containing multiple connected components are successfully reconstructed by the region integration approach.

Table 2 shows ABO values for each brand of shoe logo obtained by the proposed method. It is observed that the ABO value of Syunsoku is relatively lower than that of other logos. The reason for this is that the logos of Syunsoku consist of multiple components and appear relatively small in most images.

Figure 8 shows examples of shoe logos extracted by the proposed method. In subfigures (a) to (f), the red boxes denote shoe logos which are correctly detected and recognized. Although the logos in the images are a wide variety of sizes and rotations, the proposed method successfully recognizes these logos. A number of false-positive regions obtained by the region-extraction stage were successfully eliminated by introducing a negative class into the classifier. In subfigures (g) to (j), the green boxes denote shoe logos which the proposed method could not detect properly. The most common reason of error is small color difference between logo and background due to shoe and logo design (g) and (h)). When the color difference is quite small, MSER may fail to extract these logo regions in the initial region extraction stage. While the proposed method is robust against affine transformation of logo, weakness for perspective and nonrigid distortion is still remaining (i). A few cases where the proposed method could not detect a large logo region fragmented by an object like shoelace are observed (j).

Tables 3, 4 and 5 compare the extraction and recognition performance among the proposed method and other previously reported methods. We conducted a three-fold cross-validation.
AOKI et al.: EXTRACTION AND RECOGNITION OF SHOE LOGOS USING TWO-STAGE CLASSIFIERS

Fig. 8 Examples of extracted and recognized logos, (a) to (f) shows examples successfully extracted and recognized, (g) to (j) shows examples failure extraction.

Table 3 Quantitative evaluation of the extraction and recognition performance of our previous method [18]

| criterion   | mean | std  | max  | min  |
|------------|------|------|------|------|
| Recall     | 19.10| 0.53 | 20.30| 18.25|
| Precision  | 64.11| 1.74 | 66.51| 60.28|
| F-measure  | 29.42| 0.68 | 30.89| 28.51|

Table 4 Quantitative evaluation of the extraction and recognition performance of the proposed method

| criterion | mean | std  | max  | min  |
|-----------|------|------|------|------|
| Recall    | 26.44| 1.99 | 28.25| 23.87|
| Precision | 62.64| 1.04 | 64.33| 61.71|
| F-measure | 36.71| 1.37 | 38.88| 34.43|

Table 5 Quantitative evaluation of the extraction and recognition performance of R-CNN

| criterion | mean | std  | max  | min  |
|-----------|------|------|------|------|
| Recall    | 7.25 | 0.79 | 8.21 | 5.45 |
| Precision | 46.61| 3.44 | 51.19| 40.81|
| F-measure | 12.54| 1.29 | 14.08| 9.64 |

is small compared with the appearance variety.

6. Conclusion

In this paper, we proposed an approach combining shoe logo extraction and recognition. Our approach achieves an F-measure of 36.71%. Shoe logo extraction employing MSERs and hierarchical region integration works effectively for logos consisting of different colors and sizes. Logo recognition using gradient histogram features and an SVM works for both recognition and false-positive elimination of logos. The proposed two-stage approach is effective for improving performance of both extraction and recognition.

Further study and investigation of the performance of the proposed method on a larger dataset are required in future research.

References

[1] J. Matas, O. Chum, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” British Machine Vision Conference, pp.384–396, 2002.
[2] K. Shirai and W. Ohyama, IEICE-PRMU shoe logo dataset: https://sites.google.com/site/alcon2015prmu/prmu-shoelogodataset
[3] N. Farajzadeh, “Exemplar-based logo and trademark recognition,” Machine Vision and Applications, vol.26, no.6, pp.791–805, 2015.
[4] W.-T. Chu and T.-C. Lin, “Logo recognition and localization in real-world images by using visual patterns,” IEEE International Conference on Acoustic, Speech and Signal Processing, pp.973–976, 2012.
[5] C.H. Lampert, M.B. Blaschko, and T. Hofmann, “Beyond sliding windows: Object localization by efficient subwindow search,” 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2008.
[6] D.G. Lowe, “Object recognition from local scale-invariant features,” International Conference Computer Vision, pp.1150–1157, 1999.
[7] M. Datar, N. Immorlica, P. Indyk, and V.S. Mirrokni, “Locality-sensitive hashing scheme based on P-stable distributions,” Annual Symposium on Computational Geometry, pp.253–262, 2004.
[8] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” IEEE Trans. Inf. Theory, vol.21, no.1, pp.32–40, 1975.
[9] S. Romberg and L. Rainer, “Bundle min-hashing,” International Journal of Multimedia Information Retrieval, vol.2, no.4, pp.243–259, 2013.
[10] S. Romberg and R. Lienhart, “Bundle min-hashing for logo recognition,” Proceedings of the 3rd ACM conference on International conference on multimedia retrieval, pp.113–120, 2013.
[11] X. Shen, Z. Lin, J. Brandt, and Y. Wu, “Spatially-constrained similarity measure for large-scale object retrieval,” IEEE Trans. Pattern Anal. Mach. Intell., vol.36, no.6, pp.1229–1241, 2014.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ”Rich feature hierarchies for accurate object detection and semantic segmentation,” IEEE conference on Computer Vision and Pattern Recognition, pp.580–587, 2014.

[13] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, ”Selective search for object recognition,” International Journal of Computer Vision, vol.104, no.2, pp.154–171, 2013.

[14] P.F. Felzenszwalb and D.P. Huttenlocher, ”Efficient Graph-Based Image Segmentation,” International Journal of Computer Vision, vol.59, no.2, pp.167–181, 2004.

[15] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, and A. Zisserman, ”The PASCAL Visual Object Classes (VOC) Challenge,” International Journal of Computer Vision, vol.88, no.2, pp.303–338, 2010.

[16] H. Su, X. Zhu, and S. Gong, ”Deep Learning Logo Detection with Data Expansion by Synthesising Context,” 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp.530–539, 2017.

[17] S.C. Hoi, X. Wu, H. Liu, Y. Wu, H. Wang, H. Xue, and Q. Wu, ”Logo-net: Large-scale deep logo detection and brand recognition with deep region-based convolutional networks,” arXiv preprint arXiv:1511.02462, 2015.

[18] K. Aoki, W. Ohyama, and T. Wakabayashi, ”Automatic Extraction and Recognition of Shoe Logos with a Wide Variety of Appearance,” 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), pp.211–214, 2017.

[19] M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura, ”Handwritten numeral recognition using gradient and curvature of gray scale image,” Pattern Recognition, vol.35, no.10, pp.2051–2059, 2002.

Kazunori Aoki is a Master course student at the Graduate School of Engineering, Mie University. He is involved in research on automatic detection and recognition of shoe logos.

Wataru Ohyama is an assistant professor at the Graduate School of Engineering, Mie University. He received his Ph.D. degree from Mie University in 2007. His research interests include image sensing, document analysis and pattern recognition. He is a member of IEICE (Japan), IPSJ, IEE (Japan) and IEEE.

Tetsushi Wakabayashi received his Master’s degree from Mie University, Japan, in 1987, and his Ph.D. degree from Nagoya University in 1997. Since 1991, he has been working at Mie University, where he is presently a professor in the Department of Information Engineering. During 1998–1999, he was a Visiting Scholar at the Rensselaer Polytechnic Institute, USA. His research interests include character recognition, image processing, computer graphics, and human-computer interaction. He is a member of the IEICE (Japan) and IPS (Japan).