International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 8 (2016) pp. 012–022

Feature Points based Fish Image Recognition

Takeshi Saitoh[1], Toshiki Shibata[1] and Tsubasa Miyazono[1]

[1]Dept. of Computer Science and Systems Engineering,
Kyushu Institute of Technology,
680-4 Kawazu, Iizuka, Fukuoka, Japan
saitoh@ces.kyutech.ac.jp

Abstract: **We are studying image-based fish identification. Most traditional approaches use a fish image wherein extraction is easy given that the fish region contrasts with a white or uniform background. This research introduces an approach to give several feature points by manual operation. In our proposed approach, we are able to accept fish image with complicated background, including rocky area. Further, to investigate efficient features for fish recognition in image, we define various features, including geometric features, bag of visual words (BoVW) models, and texture features. We collected images comprising 129 fish species under various photographic conditions and applied our proposed method to these images. From our results, we confirmed that a combination of geometric features and BoVW models obtained the highest recognition accuracy.**

Keywords: Fish image, feature points, geometric features, bags of visual word models, texture features.

I. Introduction

In our daily travels, we often encounter a large number of natural objects, such as plants, fishes, animals, and birds. With the rapid increase in the number of mobile terminals (e.g., tablets and smartphones), taking pictures of such natural objects has become increasingly easy. As such, there has been a number of recent works concerned with image-based recognition of natural objects. With respect to image-based recognition of natural objects, many researchers are interested in identifying flowers[1, 2], leaves[3, 4, 5], and birds[6, 7]; however, there is relatively little research on image-based fish recognition. We conclude that collecting large numbers of fish images is a difficult task because fish move quickly from one location to another. Furthermore, fish live in water, which is a more difficult medium for photographs.

From a social perspective, the habitat distribution of fish is changing due to the influence of global warming. Furthermore, poisonous fish that did not previously inhabit certain environments have been witnessed in those environments. Some poisonous fishes can appear quite similar to non-poisonous fishes. Differentiating fish species here is sometimes difficult for specialists and certainly for amateurs, thereby leading to accidents in which poisonous fish are eaten by mistake.

Belhumeur et al. report on a computer vision system they built to aid in the identification of plant species [3]. Their system requires the user to photograph an isolated leaf against a white background. The surprising contribution of their research is the use of over 200 species in their datasets. Moreover, their project team developed an electronic field guide application for iPhone and iPad called LeafSnap [4] that encouraged us to develop an image-based fish recognition system. Fish can be identified using our system.

The key contributions of our work, which are presented in this paper, are as follows.

- We propose a fish image recognition method using feature points for fish images with complicated backgrounds.
- We investigate efficient features for fish recognition.
- We collect fish image dataset larger-scale than other studies.

The remainder of our paper is organized as follows. In Section **II**, we introduce related research. Next, in Section **III**, we describe our dataset. In Section **IV**, we describe the necessary preprocessing before we are able to extract features. In Sections **V** and **VI**, we describe recognition features and recognition method, respectively. Experimental results are presented in Section **VII**. Finally, in Section **VIII**, we present our conclusions and future work.

II. Related research

In this section, we briefly introduce research related to our work. Table 1 summarizes related research, all of which is further described below.

Storbeck and Daan propose an image-based fish species recognition method [8]. Their vision system measures a number of fish features, as seen by a camera perpendicular to a conveyor belt. The specific features here are the widths and heights at various locations along the fish, which are then used as input values to a neural network. The number of species considered here is only six.

Chambah et al. propose an automatic color equalization model based on a color correction method [9]; they apply their method to an underwater fish image to segment fish regions. Their project focuses on developing an information system for aquariums. They calculate various features, including geometric, color, texture, and motion features. A

Table 1: Summary of related research.

	target images	# of species	# of samples	year
[8]	fish on conveyor belt	6	502	2001
[9]	underwater fish image	12	1,346	2003
[10]	unknown	20	610	2009
[11]	underwater fish image	10	3,179	2012
[12]	fish in a tank at the aquarium	20	200	2012
[13]	white background	30	900	2013
[14]	SQUID[15]	20	1,100	2015
[16]	underwater fish image (FishCLEF-31K)	10	**31,397**	2015
Ours	natural background	**129**	2,580	2016

feature reduction process is also applied to eliminate useless or redundant features. Further, a quadratic Bayes classifier is used to classify selected fish into one of the learned species. In their experiments, they collected 12 fish species and 1,346 samples.

Alsmadi et al. propose a system for recognizing isolated patterns of fish [10]. In their method, the given input fish image is first cropped to remove the ventral part of the pattern of interest, and then, a color histogram is calculated. From this histogram, three features (i.e., standard deviation, homogeneity, and energy) are extracted from the gray-level co-occurrence matrix (GLCM); further, two features of median and variance values are directly calculated. The multilayer feed forward neural network model with a back-propagation classifier is then employed for the classification task. The number of species in their research is 20.

Huang et al. propose a hierarchical classification method for live fish recognition in an unrestricted natural environment recorded by underwater cameras [11]. In this method, the Grabcut algorithm is first employed to segment fish from the background. Next, their method extracts 66 features, which consist of a combination of color, shape, and texture features from different parts of the fish. Their method also reduces the number of feature dimensions via forward sequential feature selection. The number of species in their research is ten, with 3,179 fish images.

Mushfieldt et al. study the foundations for an interactive system at an aquarium that can recognize various fish species and display instant information to users [12]. They propose a preprocessing procedure to first segment the fish. A support vector machine (SVM) is used to recognize 20 fish species based on shape and color information.

Pornpanomchai et al. report on a computer system they developed that is capable of recognizing some fish images; their system is called FIRS [13]. In their research, each fish image is taken against a white plastic plate with fluorescent bulbs below. Next, a thresholding method is applied to extract the fish region. Eight features are defined, and an artificial neural network is used for the recognition process. They conducted their research on 30 fish species.

Nasreddine and Benzinou propose shape matching and geodesics for fish recognition [14]. Their approach is independent of translation, scale, rotation, and starting point selection. In their work, they carried out performance experiments on a benchmark Shape Queries Using Image Databases (SQUID) [15].

Spampinato et al. present a large dataset for the fine-grained recognition problem of identifying fish species from images and videos [16]. They developed an effective nonparametric approach for automatic label propagation. The automatically labeled dataset was used for benchmarking fish species recognition approaches within the fish task of LifeCLEF2014. The number of species is only ten, though the number of samples is approximately 31,000.

(a) Non-target fish images (left: Japanese eel catfish, right: hairtail).

(b) Target fish images (left: mottled spinefoot, right: redspotted grouper).

Figure. 1: Fish images.

Most traditional approaches use a fish image in which it is easy to extract a fish region given a white or uniform background. In our research, we adopt an approach that presents several feature points based on manual operations by the user. Our proposed approach is able to accept fish images with complicated backgrounds, including against rocky backgrounds. Further, to identify efficient features for fish recognition, we defined various features, including geometric features, bag of visual word (BoVW) models, and texture features. Moreover, in related research, the number of species is consistently less than or equal to 30. Evaluation experiments using such small-scale datasets yields low reliability and is therefore not practical. Therefore, in our work, we constructed a large-scale dataset as compared with the related research and evaluated our proposed method.

III. Dataset

As described in Section **II** above, the number of species in related research is consistently less than or equal to 30. It is estimated that the number of fish species living in Japanese

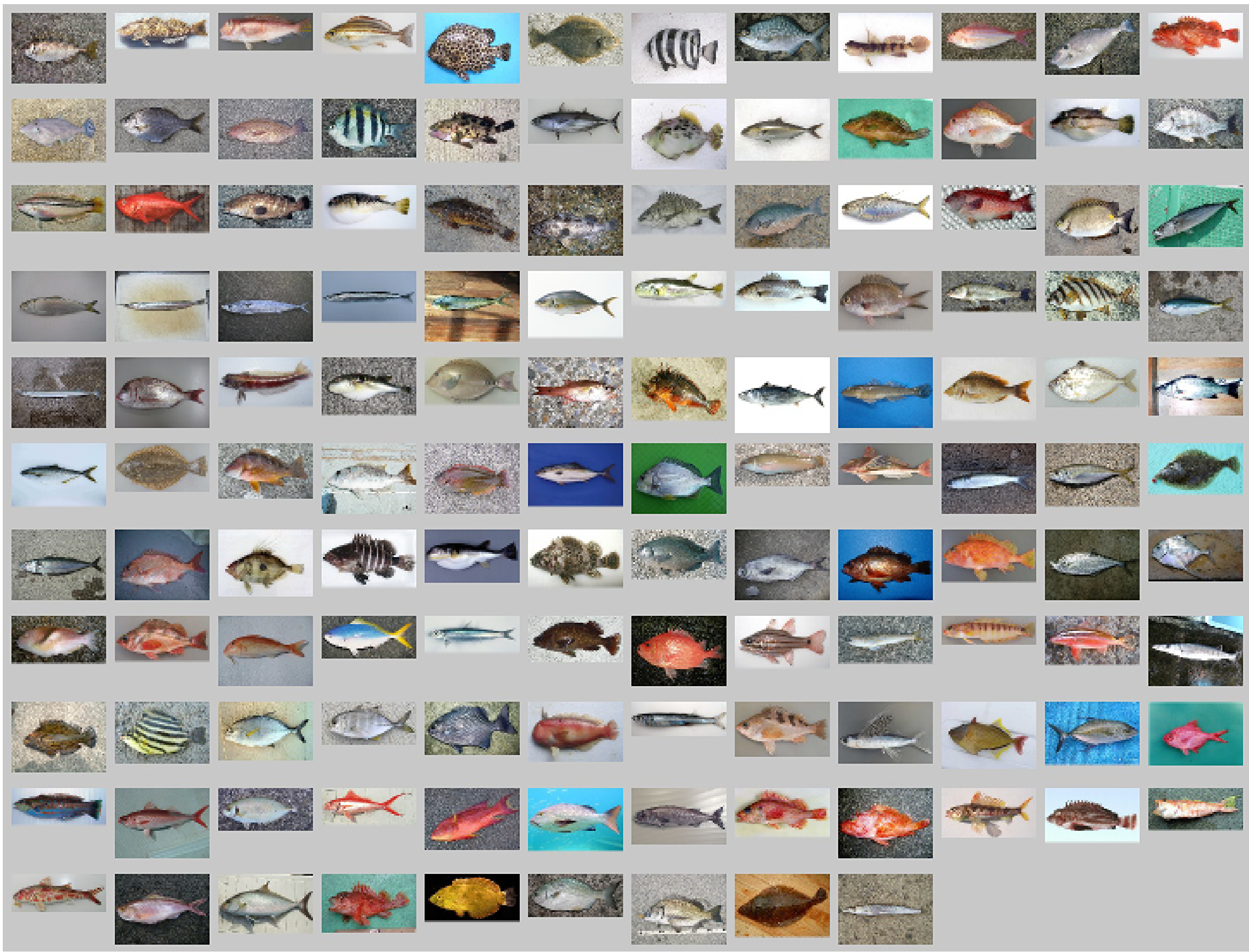


Figure. 2: 129 species of our dataset.

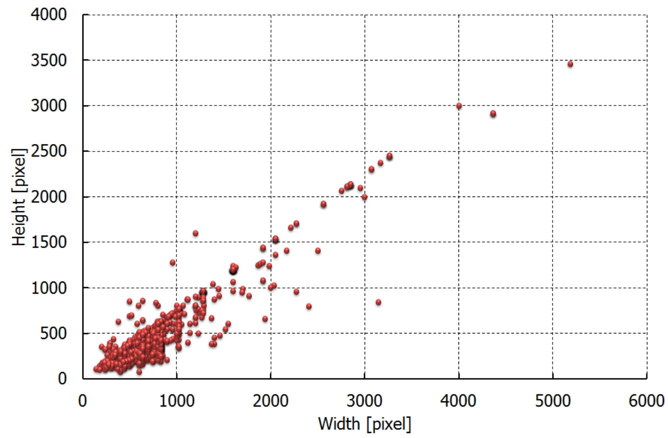
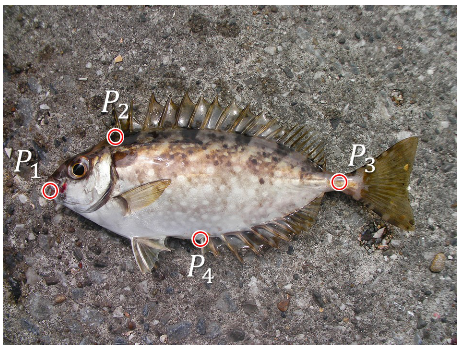
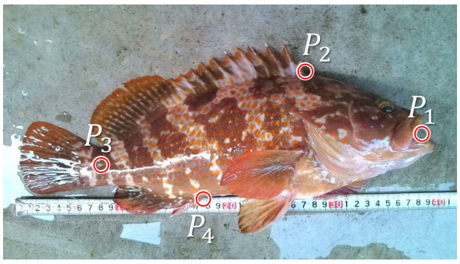


Figure. 3: Image size distribution of our dataset.



(a)



(b)

Figure. 4: Four feature points.

rivers and seas is approximately 3,300. When a system assumes a specific environment, such as an aquarium tank, a small number of species is not a problem; however, in our research, the target user is an angler, and thus we aim to include a large number of fish species in our dataset.

We therefore selected 129 species in the northern area of Kitakyushu in Japan as the recognition targets. Sample fish images of these 129 species are shown in Fig. 2. We collected 20 distinct samples for each of these species through the Web. Note that all fish images were checked and confirmed by experts. Here, since our proposed approach requires a variety of feature points, fish that are difficult to input feature points for are excluded, e.g., the elongated shape shown in Fig. 1(a). Figure 1(b) shows sample target fish images used in our research. We note here that the problem of identifying oddly shaped fish species is not considered in this paper and is an area for future work.

Since we collected fish images from the Web, the photographers and image sizes varied considerably. Among the 2,580 collected fish images, the size of the smallest image was 144×108 pixels, whereas the maximum image size was $5,184 \times 3,456$ pixels. The size distribution of our dataset is shown in Fig. 3. In this distribution, the horizontal axis represents width, while the vertical axis represents height. Thus, small sample images are plotted in the lower left, whereas large sample images are plotted in the upper right. The average size here was 692×425 pixels.

IV. Preprocessing

A. Feature points

Since target fish images used in our research are primarily pictures taken in natural conditions, the boundaries between the fish region and background is unclear, and it is difficult to automatically extract the fish region. Our proposed method adopts a feature points-based approach consisting of the following four points: mouth P_1; dorsal fin P_2; caudal fin P_3; and anal fin P_4. Each of these is shown in Fig. 4. Note that these points are manually provided by the user and are designed as characteristic locations to avoid incorrect input by users.

B. Normalization

As described in Section **III** above, the image sizes of our dataset are not the same; further, the orientation of the fish is not the same. These differences are inconvenient for calculating features. To address this, normalization processes for both size and orientation are applied.

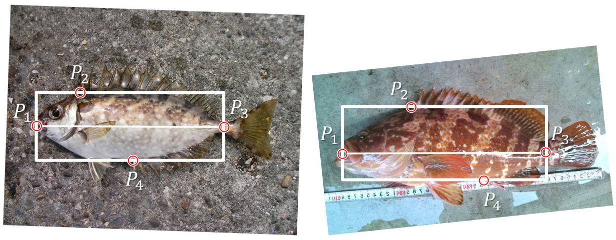
First, the direction of a fish image is arranged for feature extraction. The fish shown in Fig. 4(a) is leftward, whereas the fish shown in Fig. 4(b) is rightward; in this paper, all fish image are arranged leftward based on the positional relationship of P_1 and P_3. After arranging the direction, scaling and rotation processes are applied.

For the scaling, the image is scaled such that the distance between P_1 and P_3 is changed to L pixels. This means that a scale s is calculated by as $s = L/|P_1 - P_3|$. For the rotation, the image is rotated such that the P_1 and P_3 are on the same horizontal level. This means that orientation θ is calculated as $\theta = \tan^{-1}|P_{1,y} - P_{3,y}|/|P_{1,x} - P_{3,x}|$. By using s and θ, affine transformation is applied to the image

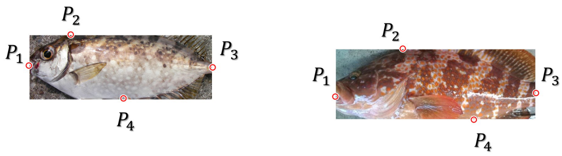
$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} s\cos\theta & -s\sin\theta & 0 \\ s\sin\theta & s\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

where (x, y) is an input coordinate and (x', y') is a corresponding output coordinate.

Figure 5(a) shows fish images after applying the above scaling and rotation transformations. Next, a rectangular region identified by the white box is extracted as the target fish image. Figure 5(b) shows example extracted fish images that are used for the feature extraction process that follows.



(a) After applying scaling and rotation transformations.



(b) Extracted fish image.

Figure. 5: Normalization process.

V. Features

In this paper, to identify efficient features, we define the following three feature groups: geometric features; BoVW model; and texture features.

A. Geometric features

Seven geometric features are defined based on the four feature points shown in Fig. 6; however, before describing geometric features, four lengths (L_{12}, L_{13}, L_{14}, L_{24}) between two feature points, as shown in Figs. 6(a) and (b), are defined as follows:

$$\begin{aligned} L_{12} &= |P_1 - P_2|, \\ L_{13} &= |P_1 - P_3|, \\ L_{14} &= |P_1 - P_4|, \\ L_{24} &= |P_2 - P_4|. \end{aligned}$$

Further, three ratios are defined as:

$$\begin{aligned} AR &= L_{24}/L_{13}, \\ R_{dorsal} &= L_{12}/L_{13}, \\ R_{anal} &= L_{14}/L_{13}. \end{aligned}$$

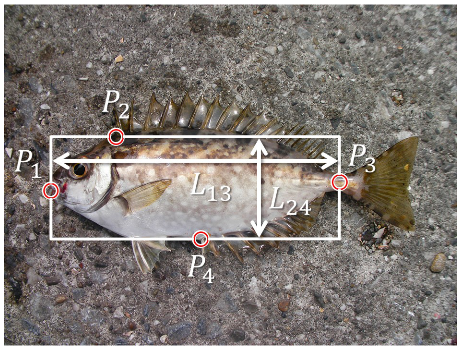
Here, AR is an aspect ratio and R_{dorsal} and R_{anal} are position ratios of the dorsal and anal fins, respectively.
The four angles shown in Fig. 6(c) are defined as:

$$\begin{aligned} \theta_1 &= \angle P_2 P_1 P_4, \\ \theta_3 &= \angle P_2 P_3 P_4, \\ \theta_2 &= \angle P_1 P_2 P_3, \\ \theta_4 &= \angle P_1 P_4 P_3, \end{aligned}$$

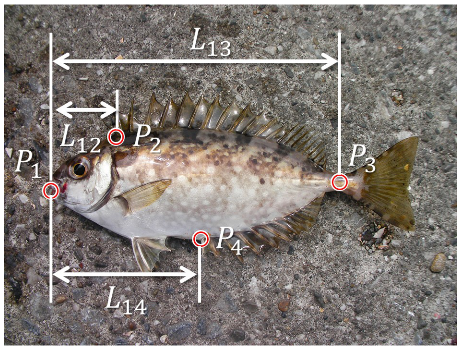
where θ_1, θ_3, θ_2, and θ_4 are the head, tail, upper, and lower side angles, respectively.

B. Bags of visual words model

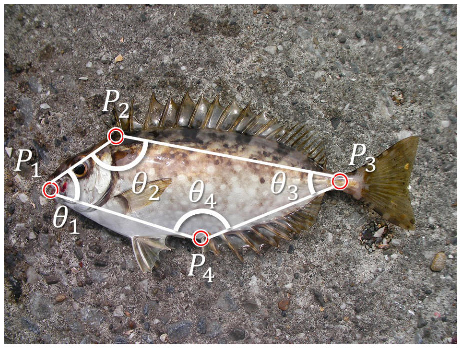
The BoVW model is widely used for object recognition, treating image descriptors as visual words. BoVW is a sparse vector of occurrence counts of a vocabulary of local features. This representation can also be described as a histogram of visual words. Here, the vocabulary is usually obtained by vector-quantizing image features into visual words.
The BoVW framework has two stages: one for training and one for testing. Each stage is further divided into two distinct steps. The first step in both stages is feature detection and representation. The second step is to train the model to predict the class label of new images in the testing stage. For each test image, a local classification problem is constructed by selecting only the nearest neighbors from the feature space.
Here, local features are represented by local descriptors. Popular local descriptors include SIFT [17], SURF, and HOG [18]; in this research, we use SIFT to represent local descriptors.
Several strategies of local feature detection are possible, including: (1) random sampling, which calculates local fea-



(a) Aspect ratio.



(b) Position ratios of dorsal and anal fins.



(c) Four angles.

Figure. 6: Geometric features.

(a) Spotted pattern.

(b) Horizontally striped pattern.

(c) Vertically striped pattern.

Figure. 7: Various fish patterns.

tures at random points within the image; (2) sparse sampling in which local patches are detected by interest point detectors able to select salient regions, such as edges, corners, and blobs; and (3) grid (or dense) sampling in which the image is segmented into sub-regions by horizontal and vertical lines according to a regular grid, extracting features in each fixed sub-region.

Fish have a wide variety of patterns, including spotted, horizontally striped, and vertically striped patterns, as shown in Fig. 7. When sparse sampling is applied, the number of detected feature points varies according to the pattern, and thus we apply dense sampling, which requires two parameters, i.e., scale G_s of each sample and distance G_d between two samples (note that these two parameters are discussed in Section **VII** below).

The most common quantization approach is k-means clustering, primarily because of its simplicity and convergence speed. In this paper, we apply k-means clustering to quantize the feature space.

C. Texture features

In our research here, we compute the following well-known texture features: local binary pattern (LBP) [19]; histogram of oriented gradient (HOG) [18]; discrete cosine transformation (DCT); gray-level co-occurrence matrix (GLCM) [20]; run length matrix (RLM) [21]; and shape-pass nonlinear filter (NF) [22]. Each of these is described below.

1) Local binary pattern (LBP)

LBP is a simple yet very efficient texture operator [19]. LBP captures the appearance of an image in a small neighborhood around a pixel. It consists of a string of bits, with bits corresponding to each of the pixels in the neighborhood. Each bit is turned on or off depending on whether the intensity of the corresponding pixel is greater than the intensity of the central pixel.

In this paper, we use a 3×3 pixel neighborhood, and the binary string of eight bits is quantized from zero to 255. LBP values are computed for all pixels in the fish image, and LBP values are pooled in a histogram with a dimension of 256. This histogram of LBP values is then used as a feature.

2) Histogram of oriented gradient (HOG)

HOG is a feature descriptor for detecting objects [18] in which images are divided into small connected regions called cells; for each cell, we compute a histogram of gradient directions or edge orientations for the pixels within the cell. Each cell is discretized into angular bins according to the gradient orientation. Each cell's pixel contributes a weighted gradient to its corresponding angular bin. Groups of adjacent cells are considered spatial regions called blocks. The grouping of cells into a block is the basis for the grouping and normalization of histograms. A normalized group of histograms represents a block histogram. The set of block histograms represents the descriptor. In our research, each target image is resized to 64×64 pixels, and the cell size is set to 8×8 pixels.

3) Discrete cosine transformation (DCT)

DCT is one of an extensive family of sinusoidal transformations that is a popular technique for image and video compression. The concept behind this transformation is to transform a set of points from the spatial domain into an identical representation in the frequency domain. The obtained DCT coefficients are ordered using zigzag scanning. In this paper, 64 lower-frequency coefficients are used as features.

4) Gray-level co-occurrence matrix (GLCM)

GLCM is a matrix of how often a pixel with a specific gray-level value occurs either horizontally, vertically, or diagonally in conjunction with adjacent pixels [20]. After calculating a GLCM, several statistics can be calculated. In this research, 14 statistics are calculated, i.e., angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, two information measures of correlation, and maximum correlation coefficient; further, four directions are considered (i.e., $0°$, $45°$, $90°$, and $135°$), and thus 56 features are calculated.

5) Run length matrix (RLM)

A run length represents the number of pixels in a run, while the run length value is the number of times such a run occurs in an image. RLM is a matrix in which each element $p(i, j, \theta)$ describes the total number of occurrences of runs of length j at gray level i, in a given direction θ [21]. Similar to the GLCM, five statistics are calculated, i.e., short runs emphasis, long runs emphasis, gray level non-uniformity, run length non-uniformity, and run percentage; further, four directions are considered (i.e., $0°$, $45°$, $90°$, and $135°$), and thus, 20 features are calculated.

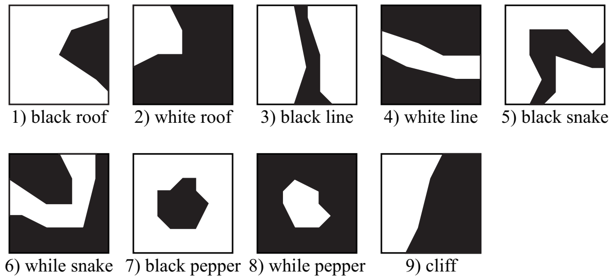


Figure. 8: Nine shape-pass nonlinear filters.

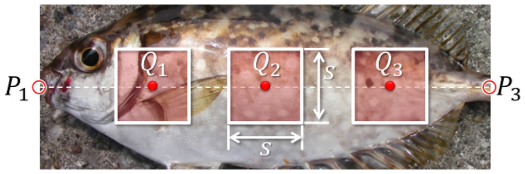


Figure. 9: Three regions for texture features.

6) *Shape-pass nonlinear filter (NF)*

Tamura and Atoda propose shape-pass nonlinear filters in which a natural texture is composed of a local shape; further, they propose the nine shape bases shown in Fig. 8 [22]. In this paper, nine shape-pass nonlinear filters are applied to each image, with the nine calculated values used as features.

7) *Regions of interest*

As noted above, fish have a number of various patterns, and these patterns can differ for each part of the fish, i.e., for its head, back, and abdomen. We assume these feature variations by calculating texture features from specific regions that are subsets of the entire region. Thus, to calculate texture features, three square regions with sizes equal to $s \times s$ pixels are defined around the following three points: head side Q_1; middle Q_2; and tail side Q_3. These ROIs are shown in Fig. 9.

VI. Recognition method

Proposed by Breiman, random forest (RF) is an ensemble training algorithm that constructs multiple decision trees [23]. This strategy has recently attracted increasing interest because it can be applied for classification, regression, and unsupervised learning. This approach has several advantages, including high levels of predictive accuracy delivered automatically, resistance to overtraining, rapid training (even with thousands of potential predictors), and diagnostics that pinpoint multivariate outliers. In our research, we apply RF as the recognition method with number of trees M experientially set to 200 and maximum depth of each tree D experientially set to 10.

VII. Experiments

Before performing recognition experiments, we first prepared four feature points for all fish images. Several students unfamiliar with fish completed this task in which all points were provided manually. The normalization parameter L was set to 640 pixels. From our results, the minimum, maximum, and average heights after the normalization process were 48, 421, and 213 pixels, respectively.
In our dataset, the sample number of each species was 20. We used the leave-one-out method, i.e., of the 20 samples for each species, 19 samples represented the training set, while the remaining sample was a test set. By varying one sample, the total number of recognition trials was 20 for each species. The average recognition rate of 20 results was used to calculate the resulting accuracy measures.

A. *Recognition experiments using each feature independently*

In our first experiment, we performed recognition experiments using each feature independently.
For the seven geometric features, the recognition accuracies of 48.5%, 75.5%, 84.1%, and 93.1% were obtained by considering only one, three, five, and ten candidates, respectively.
As for the BoVW models, four scale parameters $G_s = \{8, 12, 16, 32\}$ and six distance parameters $G_d = \{12, 14, 16, 18, 32, 64\}$ were prepared, and 24 recognition experiments were carried out. Here, the number of visual words was experientially set to 300. Experimental results are shown in Fig. 10, revealing that the highest recognition rate of 53.2% was obtained when $G_s = 12$ pixels and $G_d = 14$ pixels.
For texture features, we defined that h as the height of the clipped image, and $s = \alpha h$, where α was gradually increased from 0.1 to 0.9 by a step of 0.1; we experimented at three target locations (i.e., Q_1, Q_2, and Q_3) to identify efficient size and location. Experimental results are shown in Fig. 11, with numerical values on each bar indicating α in which the highest recognition accuracy was obtained. From our results, the maximum recognition rate of 44.8% was obtained when $\alpha = 0.9$, target location was Q_2, and HOG features were used.
Based on the above experimental results, we conclude that geometric features attained the best performance.

B. *Recognition experiments with combined features*

To evaluate the performance and effectiveness of feature combinations, we investigated the four combinations {G, B}, {G, T}, {B, T}, and {G, B, T}, where G indicates geometric features, B indicates BoVW models, and T indicates texture features. As for each feature, the optimal conditions that obtained the highest recognition performance in our previous experiments were used. Results are shown in Table 2. Here, since our objective is a fish identification system, we note that it is better to consider not only the first candidate, but also several candidates. Thus, Fig. 12 shows performance curves that indicate how often the correct classes for a query were placed among the top k matches, with k varying from one to 10.
From our results, we conclude that the best result was obtained when combination {G, B} was used. Under this condition, recognition accuracies of 66.6%, 86.2%, 92.0%, and 96.3% were obtained by considering one, three, five, and ten candidates, respectively. Moreover, even as k varied, this condition obtained the highest recognition accuracy under all

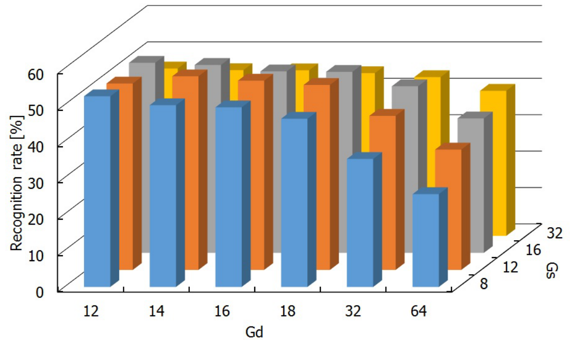


Figure. 10: Recognition results (BoVW models).

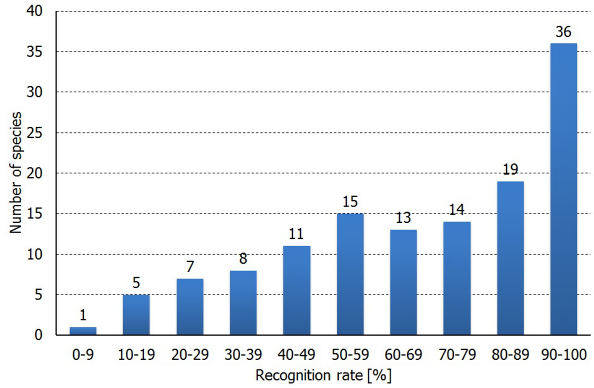


Figure. 13: Distributions of recognition rate.

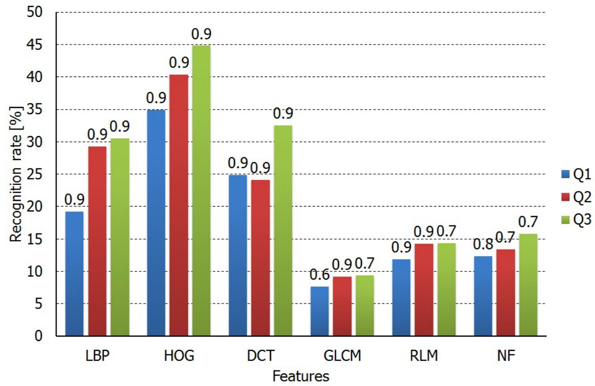


Figure. 11: Recognition results (texture features).

Table 3: Order-level recognition accuracy.

Order	# of species	# of samples	recognition rate [%]
Scorpaeniformes	17	340	64.1
Pieuronectiformes	4	80	55.0
Beryciformes	3	60	65.0
Perciformes	84	1,680	87.3
Beloniformes	4	80	82.5
Atheriniformes	1	20	30.0
Clupeiformes	3	60	45.0
Aulopiformes	2	40	57.5
Tetraodontiformes	9	180	85.6
Mugiliformes	1	20	70.0
Zeiformes	1	20	50.0

conditions. The two combinations {G, T} and {G, B, T} obtained the same performance curves. As for the remaining combination, {B, T}, which does not contain a geometric feature, the performance curve was lower than the performance curves using the independent feature conditions of G and B.

C. Discussion

Recognition results considering only the first candidate were analyzed in combination {G, B}, which obtained the best performance in the previous experiment. Figure 13 shows the distributions of recognition rates of all 129 species. From the figure, there were 55 species in which recognition accuracy exceeded 80%.

In our dataset, although the number of species was 129, the number of orders was 11. We assume that order-level recognition accuracy is higher than species-level recognition accuracy. To study this, we computed order-level recognition accuracy. Table 3 shows order-level recognition accuracy. Here, Perciformes occupied the most species, and its recognition accuracy was 87.3%. Scorpaeniformes had a large number of species, but its recognition accuracy was lower than that of Tetraodontiformes. These results indicate that the differences in Scorpaeniformes fish are larger than those in Tetraodontiformes. Typical examples are shown in Fig. 14. In particular, Fig. 14(a) shows Striped beakfish samples. Variations in the sample are small, and its recognition rate was 100%. Conversely, Fig. 14(b) shows Adjutant samples. Variations in the sample are large, and its recognition rate was 10%.

Next, we analyzed the RF-generated trees of RF in which the features is a were combination of {G, B}. Table 4 shows the ratios of the most selected features for the root node at each tree. From our results here, we found that all seven geometric features were selected in the first seven features, indicating that geometric features are more important than the BoVW model. Further, we found that the head side angle θ_1 was the most selected feature.

Considering all of our experimental results, for image-based fish recognition, shape information is certainly useful.

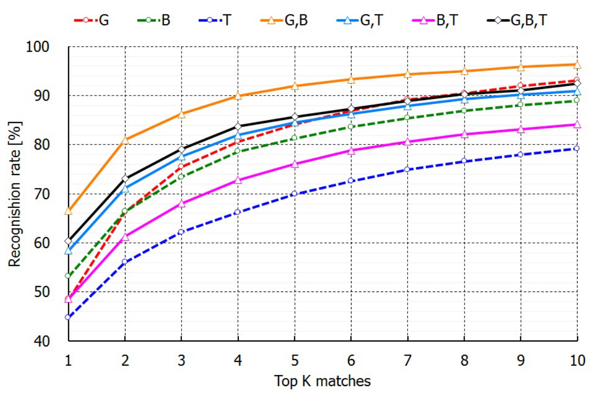
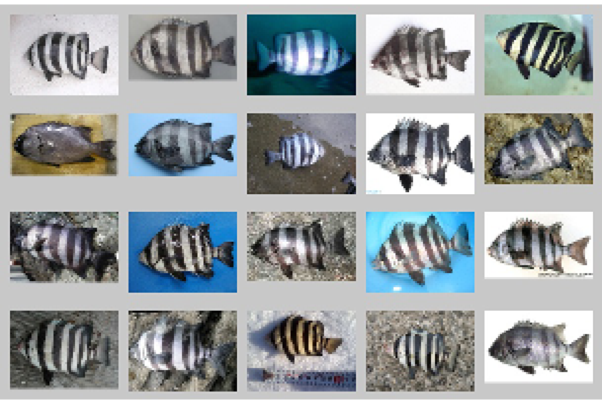


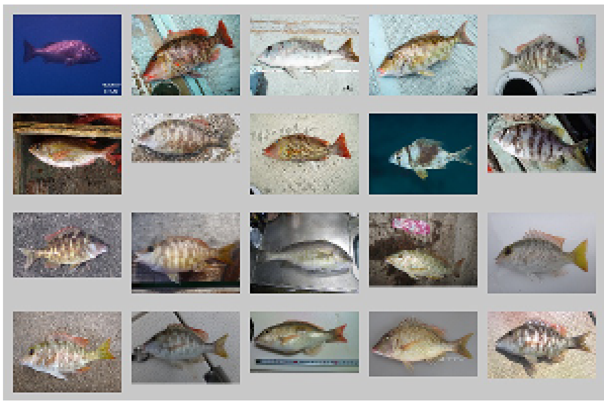
Figure. 12: Recognition results.

Table 2: Performance comparison with feature combinations.

candidates	G	B	T	{G,B}	{G,T}	{B,T}	{G,B,T}
# of dimension	7	300	1764	307	1771	2064	2071
1	48.5	53.2	44.8	66.6	58.5	48.6	60.4
2	66.2	66.4	56.0	81.0	71.1	61.3	73.1
3	75.5	73.4	62.2	86.2	77.6	68.1	79.1
4	80.6	78.6	66.2	90.0	81.9	72.8	83.7
5	84.1	81.3	69.9	92.0	84.6	84.6	85.7
10	93.1	89.0	79.2	96.3	90.9	84.1	92.5



(a) Striped beakfish.



(b) Adjutant.

Figure. 14: Variability in samples.

VIII. Conclusion

In this paper, we proposed a fish recognition method based on various feature points. Most traditional research requires a white or uniform background, which is limiting. In our approach, we relax this constraint and adopt a variety of feature points; as such, our proposed method is able to accept fish images in natural scenes. Moreover, we investigated efficient features for fish recognition, evaluating our proposed method with a large-scale dataset. We found that geometric features are most important versus other features that we studied.

We also note here that even if fish are from the same species, they differ in terms of appearance and shape based on the development stage (i.e., young, adult, and senility), gender, and photography time; however, in our study, we also considered species-level recognition. To improve performance, our future work includes the consideration of detail level. Finally, all feature points in this study were manually identified and input by several students, and thus, we did not consider feature point error, which we need to discuss in future work.

Table 4: Selected features for random forest.

order	features	ratio
1	θ_1	0.1375
2	θ_3	0.1075
3	θ_4	0.1050
4	AR	0.1000
5	R_{anal}	0.0825
6	R_{dorsal}	0.0825
7	θ_2	0.0750
8–307	BoVW	—

Acknowledgments

This research was partially supported by Kitakyushu Foundation for the Advancement of Industry, Science and Technology (FAIS). Making on a fish species list and a check of all collection images had cooperated by Keiesu Kikaku. We thank our collaboration to all staffs of Keiesu Kikaku.

References

[1] T. Saitoh, K. Aoki, and T. Kaneko, "Automatic recognition of blooming flowers," in *Proc. of 17th International Conference on Pattern Recognition (ICPR2004)*, vol. 1, 2004, pp. 27–30.

[2] A. Angelova, S. Zhu, and Y. Lin, "Image segmentation for large-scale subcategory flower recognition," in *Proc. of IEEE Workshop on Applications of Computer Vision (WACV2013)*, 2013, pp. 39–45.

[3] P. N. Belhumeur, D. Chen, S. Feiner, D. W. Jacobs, W. J. Kress, H. Ling, I. Lopez, R. Ramamoorthi, S. Sheorey, S. White, and L. Zhang, "Searching the world's herbaria: A system for visual identification of plant species," in *ECCV*, 2008, pp. 116–129.

[4] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *ECCV*, 2012, pp. 502–516.

[5] T. Saitoh, T. Iwata, and K. Wakisaka, "Okiraku search: Leaf images based visual tree search system," in *Proc. of IAPR International Conference on Machine Vision Applications (MVA2015)*, 2015, pp. 242–245.

[6] M. Das and R. Manmatha, "Automatic segmentation and indexing in a database of bird images," in *Proc. of*

the 8th International Conference on Computer Vision (ICCV2001), vol. 2, 2001, pp. 351–358.

[7] T. Berg, J. Liu, S. W. Lee, M. L. Alexander, D. W. Jacobs, and P. N. Belhumeur, "Birdsnap: Large-scale fine-grained visual categorization of birds," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2014)*, 2014, pp. 2019–2026.

[8] F. Storbecka and B. Daan, "Fish species recognition using computer vision and a neural network," *Fisheries Research*, vol. 51, pp. 11–15, 2001.

[9] M. Chambah, D. Semani, A. Renouf, P. Courtellemont, and A. Rizzi, "Underwater color constancy: Enhancement of automatic live fish recognition," in *Proc. of SPIE*, vol. 5293, 2003, pp. 157–168.

[10] M. K. Alsmadi, K. B. Omar, S. A. Noah, and I. Almarashdeh, "Fish classification based on robust features extraction from color signature using back-propagation classifiere," *Journal of Computer Science*, vol. 7, no. 1, pp. 52–58, 2011.

[11] P. X. Huang, B. J. Boom, and R. B. Fisher, "Underwater live fish recognition using a balance-guaranteed optimized tree," in *ACCV*, vol. 7724, 2012, pp. 422–433.

[12] D. Mushfieldt, M. Ghaziasgar, and J. Connan, "Fish identification system," in *Proc. of South African Telecommunication Networks and Applications Conference (SATNAC2012)*, 2012, pp. 231–236.

[13] C. Pornpanomchai, B. Lurstwut, P. Leerasakultham, and W. Kitiyanan, "Shape- and texture-based fish image recognition system," *Kasetsart Journal (Natural Science)*, vol. 47, pp. 624–634, 2013.

[14] K. Nasreddine and A. Benzinou, "Shape-based fish recognition via shape space," in *Proc. of 23rd European Signal Processing Conference (EUSIPCO2015)*, 2015, pp. 145–149.

[15] E. G. Petrakis, A. Diplaros, and E. Milios, "Matching and retrieval of distorted and occluded shapes using dynamic programming," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1501–1516, 2002.

[16] C. Spampinato, S. Palazzo, P. H. Joalland, S. Paris, H. Glotin, K. Blanc, D. Lingrand, and F. Precioso, "Fine-grained object recognition in underwater visual data," *Multimedia Tools and Applications*, pp. 1–20, 2015.

[17] D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. of IEEE International Conference on Computer Vision*, pp. 1150–1157, 1999.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893, 2005.

[19] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Proc. of 12th IAPR International Conference on Pattern Recognition (ICPR1994)*, 1994, pp. 582–585.

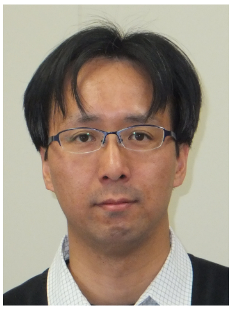
[20] R. M. Haralick, K. Shanmugam, and I.Dinstein, "Textural features for image classification," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 601–621, 1973.

[21] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 172–179, 1975.

[22] H. Tamura and O. Atoda, "A bank of "shape-pass" nonlinear filters extracting pictorial features of natural texture," *IEICE Trans. on Information and Systems*, vol. J82-D-II, no. 12, pp. 2260–2270, 1999.

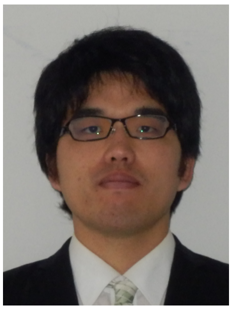
[23] L. Breiman, "Random forests," *Machie Learning*, vol. 45, no. 1, pp. 5–32, 2001.

Author Biographies



Takeshi Saitoh He was born in Shizuoka, Japan, in August 1976. He received the B.Eng., M.Eng., and D.Eng. degrees in Information and Computer Sciences from Toyohashi University of Technology, Japan, in 1999, 2001, and 2004, respectively. He became an Assistant Professor at Tottori University in 2004, and an Associate Professor at Kyushu Institute of Technology in 2010. He was a visiting researcher at the Center for Machine Vision Research, University of Oulu from September 2014 to January 2015. His research interests include image processing and pattern recognition. He is a member of the IEICE, IEEJ, IEEE, and the Society of Instrument and Control Engineers.

Toshiki Shibata He received his Bachelor of Information Engineering, and Master of Information Engineering degrees from Kyushu Institute of Technology, Japan, in 2013, and 2015, respectively. He joined ALPHA SYSTEMS INC. His research interests include image processing and pattern recognition.



Tsubasa Miyazono He received the Bachelor of Information Engineering degree from Kyushu Institute of Technology, Japan, in 2015. He is currently a master student with the Graduate School of Computer Science of Computer Science and System Engineering, Kyushu Institute of Technology, Japan. His research interests include image processing and pattern recognition.