IMAGE-BASED RECOGNITION OF INDIVIDUAL TROUTS IN THE WILD

Lingcong Zhao, Marius Pedersen, Jon Yngve Hardeberg

Department of Computer Science
Norwegian University of Science and Technology

Børre Dervo

Norwegian Institute for Nature Research

ABSTRACT

Individual fish recognition has potentials in applications as fish cultivcation and fishing tourism. Unlike previous research, which either based on physical marker or based on photograph comparison using observers, this paper propose an approach being able to identify individual brown trouts *(Salmo trutta)* automatically with images taken in the wild. Although big variation in illumination, poses of the trouts, and resolution we validated that just using a small patch taken from the head of the trout, which can minimize the variations, it's possible to recognize individuals automatically. Two methods were proposed based on a local density profile and on a codebook. Both of the methods gave modest recognition accuracy 64.9% and 74% respectively, which compared to random chance at 3.3% is significantly better.

Index Terms— Individual Fish Recognition, Recognition, Image in the Wild, SURF, melanophore Pattern, BoW

1. INTRODUCTION

The ability to recognize individual animals is regarded as a technique greatly needed in many fields. For biology, for example, it can effectively benefit studying of habitat use, migration timing, and physiological changes to individuals [1]. For fish farming, the ability of individual fish recognition can help to build a robust tracking system to monitor the motion and migration of fish. Such a system will provide valuable information about the weight growth pattern, health condition, social manner and so on for the fish, which is crucial for optimizing cultivation factors like temperature, fish density, breeding frequency and more. Also for fish tourist industry, if identification of individual fish is possible it can be used to estimate population, and furthermore provide an important foundation about how many fishing licenses to sell in an area.

Traditionally, research identification of individual animals has focused on so called Capture-Mark-Recapture (CMR) techniques, in which the animal has to be physically marked or tagged [2]. Typical techniques include fin-clipping, cold branding, tattoos, visible implant tags, and external tag identifiers attached by metal wire, plastic, or string [3, 4, 5]. Although these techniques achieved success in variable types of animal identification tasks, there are certain limitation and drawbacks. Firstly, when it comes to large scales of studies, the CMR method can be very time-consuming and may need many human operators to fulfill the marking tasks. Secondly, animals with small body size like insects or juvenile fish are hard to be tagged. In addition, after being tagged, the markers may not last long enough for long-term research since the marker can be destroyed, lost or vanished since animals are flexible. Apart from this, the main concern has been the physical and behaviorally influence it brings to the marked individuals [6]. Specially, when it comes to fish tagging, Persat [7] states that several techniques for individual fish tagging like jaw tagging, coded tags, fail to last longer than 9 months, and may cause wounds, infections and increase the morality and slow growth.

The limitations of traditional tagging unveil the need for non-invasive recognition of fish. Thus, the goal of this paper is to propose an automatic image-based method for individual recognition of brown trouts *(Salmo trutta)* in the wild.

First we introduce relevant background, then we present methodology and proposed recognition algorithms, which then followed by results and discussion. At last we conclude.

2. BACKGROUND

In order to overcome the limitations of CMR, non-invasive methods such as photograph based individuals identification methods has been explored. One of the earliest studies was in 1982, when Persat [7] took photographs of the left side of Ain graylings in a controlled set up. More than 400 graylings were captured in the beginning, but only forty were recaptured one year later. The ground truth data were provided by the traditional fin clipping method. The author then used the number and position of the spots as cues to identify the marked individuals. It concluded that the two cues, number and position of the spots, were well-defined for each individual and make it possible to identify each fish, but for individuals with few or none dots, other features such as general disposition of the scales, were needed. In the study, the author mentioned that the graylings must have a fork length longer than 17 cm. Furthermore, in 1993, Garcia de Leaniz et al. [8] used similar features (number of dots in a specific small area in salmonids' head) to recognize juvenile salmonids that were too small for conventional tagging methods. The author used three observers to do the identification task which ended up with 100% accuracy over 30 individual Atlantic salmonids that had been photographed every four months within a eight months total period. The authors also implemented the same experiment for juvenile brown trouts, where 84% (12 out of 14) individual brown trouts had been identified correctly by the observers via spot count in head region. Both of these two papers mentioned the concern of fork length of the fish, which was believed having firm relationship with the melanophore spot pattern. Merz [1] investigated this, where 295 juvenile Chinook salmon were photographed in the top head region (dorsal) in seven photo sessions over a 251-day period. Through the images, it could be clearly observed how the juvenile salmonids gradually grew melanophore spots on the dorsal region. They gave the very important conclusion that juveniles began developing spots, identifiable in images, between 167 and 197 days after conception. Once recognizable cephalic spots developed, with fork length around 140 mm, the pattern were 100 % recognizable with up to four trials over 106 individuals. The accuracy remained unchanged even till 55 days later. Gifford and Mayhood [9] carried out a two-year project on Westslope Cutthroat Trout, aiming at finding a way to protect this at-risk species. They found that for large adults the melanophore patterns appeared to be stable over at least two years. They implemented the identification via perception of the spot pattern throughout the whole side body. The spots patterns were not exactly the same for images before and after 2 years, the fish maintained the old spots pattern, but new spots were introduced in the time period. The weakness of this research is that the conclusion is only based on two individual trouts. Other related researches on individual animal recognition can be found in [2, 10].

When looking at the photograph based approaches, they together provide a strong proof that the melanophore pattern of salmonid, grayling and trouts are promising enough for individual identification, but with prerequisite that the fish should be juvenile-to-adult individuals (approximate fork length $>$ 140 mm). Most of the literature are mainly observer based, assisted by computers to fulfill tasks such as regression analysis between groundtruth and observation in [8] and sorting routines in a spreadsheet in [9]. In [1] the spot pattern of the head dorsal region was binarized into black and white in order to generate a spot pattern profile which comprised of the x, y coordinates of each spot centroid. But the processing and segmentation part was only briefly explained. The matching part was done by mathematically similarity computation with the normalized pattern profile. But again, details of the computation were not discussed. An additional unexplored point is recognition with images in the wild. All of the previous studies used controlled image acquisition systems in which variation of specular reflection, shadow, resolution, pose of the fish, illumination, occlusion and clutters were minimized. However, in many real applications, it will be too time-consuming since simply capturing the fish from the wild, posing them under a laboratory-like condition and maintain their motion during capturing is not easy work. Not to mention, for large scale analysis.

The use of computer vision have shown to be useful in classification of fish [11, 12, 13, 14].

So all these points mentioned above make the objective of the work in this paper unique and practically valuable, but at the same time very challenging since classification and identification with uncalibrated images from wild was still an unsolved problem in computer vision field. So more specifically, in this work, we aim at building an automatic algorithm for individual brown trout recognition with images taken under totally uncontrolled environment. We also want to quantify the importance of image enhancement when dealing with such uncalibrated images.

3. THE DATABASE

In this work the database was provided by NINA (Norwegian Institute for Nature Research), containing images of brown trouts mostly from Gudbrandsdalslågen, Norway. The images were formatted in JPEG and contained both close-up view and whole-body view of the target trouts.

The images were acquired by unknown devices under uncontrolled lighting conditions. It had vast variation in illumination, poses of the trouts and locations where the trouts were photographed. The variables in the database include specular reflection (images taken during night with exposure light), low resolution (images taken under water with motion artifacts), scaling (images taken in unknown distance), projection (different pose of the trouts), shadows (images taken under shelter), clutters and occlusion (from environment and human behavior). So it sufficiently represented a real world situation where all the above variables exist.

Inspired by literature, due to the rigid structure, the head region was proved quite effective. In addition, [8] and [9] provided another proof that the spot pattern in head region was sufficient enough for individual fish recognition. Therefore we will focus on a region of interest (ROI) of the trout, which is shown on Figure 1.

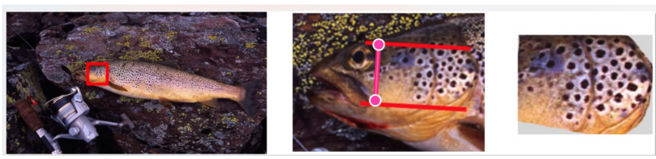

Fig. 1. The selection of ROI: side of the head.

As shown in Figure 1, the top line is a parallel line cross the top point of the eye, the bottom line passes through the deepest mouth point in x-axis. The right part of the ROI will follow the gill line (border between the head and the body). This step was done manually in Adobe Photoshop. The cropped ROI images were roughly rotated manually in order to let all the trouts more or less facing to one direction.

A total 175 images with 31 individual trouts have been used. Each individual trout had an unbalanced number of images, some individuals had 8-10 images while some only had 2-3 images. The original resolution of the images were around vertical (V): 2746, horizontal (H): 4134 pixels or V:4134, H:2746 pixels, but if only the target trouts were separated out,the resolution were around H:650-3200, V: 650-2800. Furthermore, if only taken ROI part, the resolution were down to V/H: 150-1500.

The cropped ROI region also have big intra-variation, inherited from the original images. Figure 2 shows the ROI-pairs of four individual trouts. It can be seen that noises and color infidelity (a), low-resolution, low-exposure (b), big affine/projection transformation (c) and highlight and washed out spots (d) were still present in the dataset.

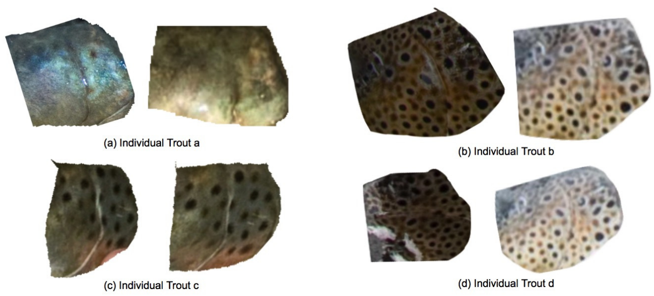

Fig. 2. The cropped ROI also have big intra-variation, from (a) to (d) are different trouts.

4. PROPOSED METHODS

Feature selection is vital for the work since it works as perception cues for human vision. In this work we propose two features: a binary feature and a grayscale level feature.

4.1. Proposed method 1: Local density

The first proposed method uses a straightforward local density feature. The feature extraction process is given below:

1. The input images (I_i) were segmented and converted to binary images (BW_i) where the melanophore spots were in white, other parts were black.

2. The binary images (BW_i) were then divided into small blocks for example $2*2$, $3*3$, this is a hyperparameter which can be customized for different tasks.

3. For each sub-region j, the number of spots was counted noted as $N_{i,j}$, where i stands for the i^{th} image, j stands for the j^{th} sub-region in the i^{th} image.

4. For each sub-region j, the percentage of white pixel was computed as the melanophore local density ($D_{i,j}$).

5. The final local density profile was the combination of $D_{i,j}$ and $N_{i,j}$. Figure 3 shows an example of a local density profile.

The success of the local density method depends highly on the segmentation step.

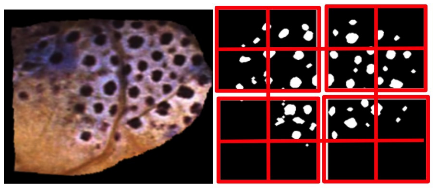

Fig. 3. Local Density profile feature extraction

4.1.1. *Enhancement*

As mentioned earlier the images are of varying quality, and many images were underexposed and with low contrast making the spot pattern to be barely visible. A local contrast enhancement method called AHE (adaptive histogram equalization [15] was used to increase the contrast. Unlike global contrast enhancement techniques like histogram equalization, which use the same transformation derived from the histogram for all the pixels, AHE performs in small local regions called tiles, where each pixel transforms based on the local tile transformation. Neighborhood tiles use bilinear interpolation to eliminate artificially induced boundaries. After this, a sharpening step to further highlight the spots using a high-pass filter was applied.

4.1.2. *Segmentation*

Segmentation was then done based on the enhanced images. In order to find the suitable color space to find the threshold T_{min} and T_{max}, multiple color spaces have been tested. The L channel from the CIELAB color space outperformed others, mostly since there were large intra-variation that made chromatic channel based segmentation difficult. The threshold $T_{\text{min}}(0)$ and $T_{\text{max}}(7.5)$ from L was chosen empirically. After that several morphological methods were used to refine the segmentation in order to filter out noise and have relatively smoother spot representation. The whole algorithm can be explained as following:

1. Define global contrast C_i of image i as :

$$C_i = std(L_{Ii})$$

where L_{Ii} standards for CIELAB L channel; std is the standard deviation of the L channel.The idea was since L channel was used for segmentation, images with low contrast will have low global contrast C_i in the L channel. Then the following steps will differ when dealing with low and high global contrast images.

$$\begin{cases} lowcontrastimage & C_i \leq T_{1i} \\ highcontrastimage & C_i > T_{1i} \end{cases}$$

where $T_{1\mathrm{i}}$, the global contrast threshold, was chosen empirically as 10.

2. Median filter and morphological opening were implemented in this step for denoising. The threshold $T2$ for opening was chosen differently for images with low and high global contrast. This was an observation obtained after multiple trials that high contrast images after the segmentation contained more noise than originally low contrast images. Thus high contrast images needed additional processing, using solidity and ratio of major and minor axes to filter out small regions which is not circle-shaped (like lines and polygons).

An example of enhancement and segmentation is shown in Figure 4:

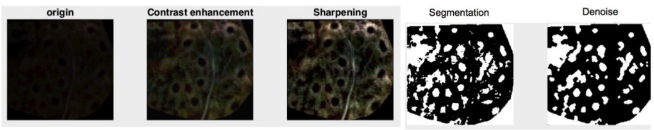

Fig. 4. Enhancement and segmentation example.

4.2. Proposed method 2: Codebook

The second method is using a codebook. It is based on SURF (Speeded Up Robust Features) [16], which is regarded as an approximated and fast version of SIFT (Scale Invariance Feature Transform) [17]. Both SIFT and SURF have been reported being invariant to uniform scaling, orientation, illumination changes, and partially invariant to affine distortion, which is promising for our application. Figure 5 shows it's potential for detecting spots. Before using SURF, only a contrast enhancement step was done. We will test two different methods; histogram equalization and histogram stretching.

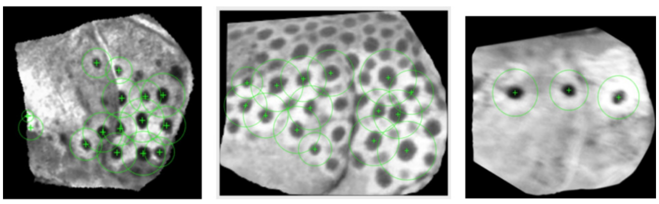

Fig. 5. SURF detector for localizing spots.

The SURF descriptor has high dimensionality. For a single image, more than 20,000 of SURF features were extracted, making directly use SURF feature usually yield poor performance. In addition, as images have different size, the extracted feature number varied between the images. The BoW (Bag of Words) [18] algorithm was used to deal with these issues. The pipeline of BoW is shown in Figure 6.

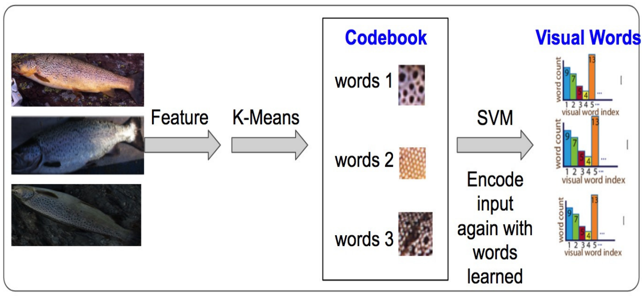

Fig. 6. Bag of Words (BoW) pipeline.

BoW was original a method for solving linguistic problems, but it is also widely used in the computer vision field. For a given image set features will be extracted, in this case SURF features. The high dimensional SURF features carries important information of the image set. It will then be used to train a codebook or vocabularies, which is made of different words, via K-means. Each word in the codebook stands for different information of the input image such as black spots, white spots, lines and so on. Next, each of the input images will be encoded again via SVM (Support Vector Machine) according to the codebook. Since this time, all the images will be encoded according to the same codebook, so the final output descriptors for each image will share the same dimension, which is a histogram describing the appearing frequency of each word in a image. If one test image needs to be classified, this image will go through the same process and match with one of classes in training set according to some similarity criteria (for example the euclidean distance of the histogram). As a similarity criteria we used Euclidean Distance.

5. RESULT

For the pattern recognition part we used five typical machine learning algorithm to see which algorithm worked better for this case. The five machine learning algorithms include: MLR (Multi-class Logistic Regression), KNN (K-Nearest Neighbor), RF (Random Forest), ANN (Artificial Neuron Network) and SVM (Support Vector Machine). The machine learning algorithms were built in R language, RStudio Version 1.0.143. The 175 images in the database were divided into training set (96 images) and test set (79 images), both with labels of 31 individual trouts. With this set of images the chance of recognizing a fish by chance is 1 in 31 (3.2%).

5.1. Proposed method 1: Local density

The binary feature method based on local density has parameters that will influence the performance of the method. We investigated first the influence of the hyperparameter K and the machine learning algorithms. The results as shown in Figure 7 indicate that a $4 * 4$ hyperparameter and random forest (RF) gives the highest accuracy.

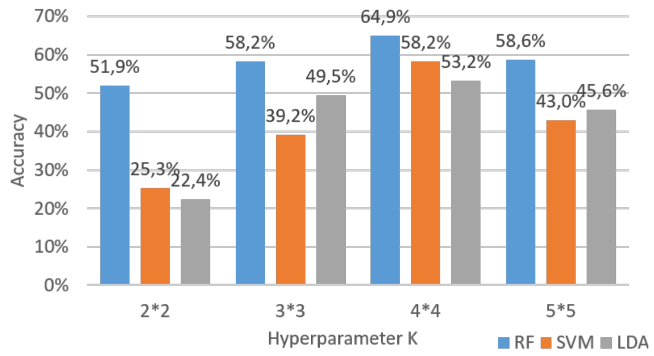

Fig. 7. Influence of the hyperparameter K and machine learning algorithms on the accuracy.

Figure 8 shows the result for the binary feature method based on local density. For comparison we include processing on an RGB image using Otsu [19] thresholding. We can see the results for the proposed method with local contrast enhancement, the proposed method with local contrast enhancement and sharpening, and the proposed method with local contrast enhancement, sharpening, hyperparameter selection and denoising. We can see from the results that the last method provide the best results (64.9%) on the test set.

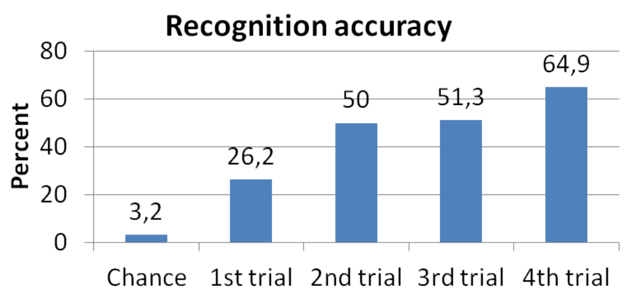

Fig. 8. Individual recognition accuracy in percent for the local density method. 1st trial: No enhancement + RGB Based Otsu threshold; 2nd trial: Local contrast enhancement + LAB based customized Threshold; 3rd trial: 2nd trial + Sharpening; 4th trial: 3rd trial+ Hyperparameter selection+ Morphology denoising. These results are based on the best hyperparameter $(4 * 4)$ and random forest.

This method has some drawbacks. It poorly provided spot distribution information of the melanophore patterns. So when two individual trouts both had very dense spots in the chosen ROI, this method might fail. In addition, the feature itself was not robust to scaling, big orientation and affine. Also, due to the huge variation in illumination, resolution, segmentation still needed improvement to deal with noise coming from unexpected non-spot areas.

5.2. Proposed method 2: Codebook

In the codebook method, two contrast enhancement methods were performed and compared, namely histogram equalization and histogram stretch. There is also one hyperparameter to define, the size of the codebook S, or in other words, how many visual words to learn from the training data. Normally a larger S will contain more information of the training set, but it must have a upper limitation where no new words can be learned. Also high S doesn't mean all the provided words are useful for the task, the noise information may also increase as S increase; if S is too small, on the other hand, the given codebook will not be sufficient enough to represent the training set, which will yield poor performance. The result for BoW method with different contrast enhancement methods and different sizes of the codebook is shown in Figure 9.

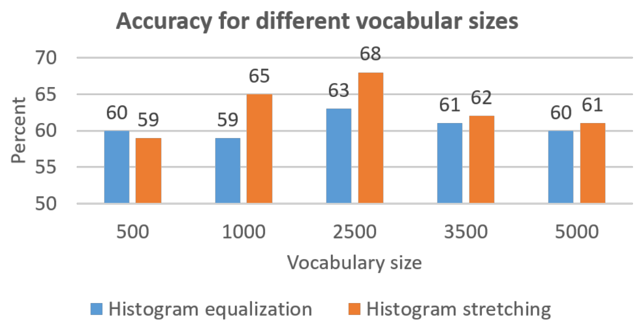

Fig. 9. Plot of Size of codebook and its effect on accuracy. Blue bars are histogram equalization method, Orange bars represents histogram stretch method.

The highest accuracy is around 68%, which is higher than the local density method. In order to find factors affecting the performance of the codebook method we carried out an analysis of cases where it fails. The main reasons for misrecognition can be classified into four categories. (1) The spot patterns can barely be seen (see Figure 10(a)). (2) For individuals with similar pattern on the chosen ROI, the accurate localization of all the spots become more important. Since the chosen ROI is not totally geometry free, so miss capturing of the boundary spots still happen which cause miss classification (see Figure 10(b)). (3) When the affine transformation is too large, it may cause inaccurate spot localization (see Figure 10(c)). (4) Influence from low-resolution or low contrast or both. For these kind of images, after enhancement, either there are unexpected spot like noise, which make the SURF extract them as relevant features (see Figure 10(d)).

We also calculated the results when images with a small size and low contrast have been filtered out, and we only have acceptable quality images. In this case we have 163 images (91 images for training, 72 images for testing) with 30 individuals (1 individual trout has been excluded due to the extreme low contrast). This increases the accuracy to 74%.

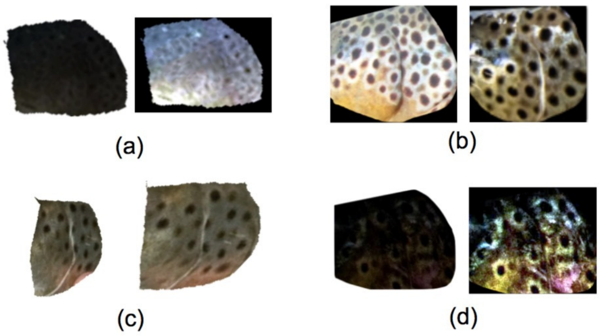

Fig. 10. Failure analysis for four categories. (a) left- original image, right - enhanced image; (b) left - one individual, right - the other individual; (c) left - one individual, right - the other individual;(d) left- original image, right - enhanced image.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed two methods for recognition of individual brown trouts *(Salmo trutta)* . There are three unique points in our approach that haven't been studied in literature before. First, our approach works by program automatically instead of using observers, which is promising for practical application, especially when it comes to large scale studies; second, our approach deals with uncalibrated images taken by unknown devices under unknown illumination. Last, we proposed two methods, both of them acquired modest accuracy, 64.9% and 74% respectively (chance 3.3%).

This is still an initial step for automatic recognition systems for individual fish. A larger dataset is naturally the next step to further evaluate the proposed methods.

References

[1] Joseph E Merz, Paul Skvorc, Susan M Sogard, Clark Watry, Scott M Blankenship, and Erwin E Van Nieuwenhuyse, "Onset of melanophore patterns in the head region of chinook salmon: a natural marker for the reidentification of individual fish," *North American journal of fisheries management*, vol. 32, no. 4, pp. 806–816, 2012.

[2] Lloyd Gamble, Sai Ravela, and Kevin McGarigal, "Multi-scale features for identifying individuals in large biological databases: an application of pattern recognition technology to the marbled salamander ambystoma opacum," *Journal of Applied Ecology*, vol. 45, no. 1, pp. 170–180, 2008.

[3] John D Koehn, "Why use radio tags to study freshwater fish," *Fish Movement and Migration.(Eds DA Hancock, D. Smith and JD Koehn.) pp*, pp. 24–32, 2000.

[4] JOSEPH E Merz, "Seasonal feeding habits, growth, and movement of steelhead trout in the lower mokelumne river, california," *California Fish and Game*, vol. 88, no. 3, pp. 95–111, 2002.

[5] Jason P Dietrich and Richard A Cunjak, "Evaluation of the impacts of carlin tags, fin clips, and panjet tattoos on juvenile atlantic salmon," *North American Journal of Fisheries Management*, vol. 26, no. 1, pp. 163–169, 2006.

[6] Roger A Powell and Gilbert Proulx, "Trapping and marking terrestrial mammals for research: integrating ethics, performance criteria, techniques, and common sense," *ILAR journal*, vol. 44, no. 4, pp. 259–276, 2003.

[7] H Persat, "Photographic identification of individual grayling, thymallus thymallus, based on the disposition of black dots and scales," *Freshwater Biology*, vol. 12, no. 1, pp. 97–101, 1982.

[8] Carlos Garcia de Leaniz, Neil Fraser, Victor Mikheev, and Felicity Huntingford, "Individual recognition of juvenile salmonids using melanophore patterns," *Journal of Fish Biology*, vol. 45, no. 3, pp. 417–422, 1994.

[9] Caitlin M Gifford and David W Mayhood, "Natural marks for identifying individual fish in small populations of at-risk westslope cutthroat trout," in *Wild trout symposium IX: sustaining wild trout in a changing world*, 2013, pp. 275–81.

[10] MJ Donaghy, AF Youngson, and PJ Bacon, "Melanophore constellations allow robust individual identification of wild 0+ year atlantic salmon," *Journal of fish biology*, vol. 67, no. 1, pp. 213–222, 2005.

[11] Thitinun Pengying, Marius Pedersen, Jon Yngve Hardeberg, and Jon Museth, "Underwater fish classification of trout and grayling," 2019, Under review.

[12] Espen Myrum, Simen Andre Nørstebø, Sony George, Marius Pedersen, and Jon Museth, "An automatic image-based system for detecting wild and stocked fish," 2019, Under review.

[13] SO Ogunlana, O Olabode, SAA Oluwadare, and GB Iwasokun, "Fish classification using support vector machine," *African Journal of Computing & ICT*, vol. 8, no. 2, pp. 75–82, 2015.

[14] S. Bermejo, "Fish age classification based on length, weight, sex and otolith morphological features," *Fisheries Research*, vol. 84, pp. 270–274, April 2007.

[15] Robert Hummel, "Image enhancement by histogram transformation," *Unknown*, 1975.

[16] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[17] David G Lowe, "Object recognition from local scale-invariant features," in *International Conference on Computer Vision (ICCV)*, 1999, vol. 99, pp. 1150–1157.

[18] Zellig S Harris, "Distributional structure," *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.

[19] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.