# Automatic fish ageing from otolith images using statistical learning.

Ronan Fablet, Nicolas Le Josse, Abdesslam Benzinou

# Automatic fish age estimation from otolith images using statistical learning

Ronan Fablet[1], Nicolas Le Josse[1] and Abdesslam Benzinou[2]
IFREMER/LASAA ENIB/RESO
BP 70, 29280 Plouzane,France BP 70, 29280 Plouzane,France
{rfablet@ifremer.fr}, www.ifremer.fr/lasaa/rfablet

## Abstract

*In this paper, we investigate the use of statistical learning techniques for fish age estimation from otolith images. The core of this study lies in the definition of relevant image-related features. We rely on the characterization of a 1D signal summing up the image content within a predefined area of interest. Fish age estimation is then viewed as a multi-class classification issue using neural networks and SVMs. A procedure based on demodulation and remodulation of fish growth patterns is used to improve the generalization properties of the trained classifiers. We also investigate the combination of additional biological and shape features to the image-related ones. The performances are evaluated for a database of several hundred of plaice otoliths.*

## 1. Problem statement

Fish age-length keys are at the core of stock assessment and marine ecology issues. Acquiring fish age data is mainly performed by interpreting fish otoliths, which are the inner ears. These calcified structures are formed through an accretionary process, which results in alternated translucent (winter) and opaque (summer) rings, as illustrated by Fig.1 for a plaice otolith. Fish age is then determined by counting the number of winter rings. This tedious task is achieved by expert readers, who typically interpret several thousand of otoliths a year. Computer-assisted tools are then sought to help in the acquisition of age-length keys and in the storage of otolith interpretation [5].

In this paper, we investigate the use of statistical learning techniques, namely neural networks and Support Vector Machines (SVM), to infer fish age from otolith images. The results obtained in [7] demonstrated the interest of neural approaches for fish aging issues. However, it was shown that the proposed features extracted from otolith images did not bring significant improvement compared to using only biological data. This is somewhat in contradiction with the experience of expert readers, who mainly rely on otolith interpretation for fish aging. Biological information is only used for verification purposes.

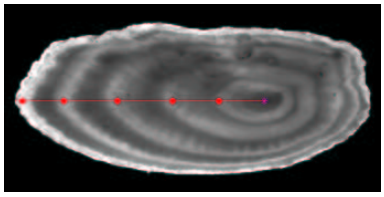

**Figure 1.** *Illustration of Plaice otolith interpretation for a 4 year old individual. The interpretation of the winter translucent rings is displayed by the markers set on the radial drawn on the main reading axis.*

Our goal is then to define more relevant image-based features. In addition to biological and geometrical features detailed in Section 2, a new set of image-related features is introduced in Section 3 to actually characterize the otolith content in terms of alternation of opaque and translucent rings. Section 4 briefly presents the application of neural networks and SVMs to our multi-class classification issue. In particular, a procedure targeted at increasing the generalization properties of the classifiers is proposed. Results are reported in Section 5 for real Plaice otolith data.

## 2. Biological and geometrical features

### 2.1. Biological features

Biological information are usually available for fish otolith databases. We will use fish length and sex, and the catch date. If available, other information such as the sexual maturity or otolith weight could also be considered. Though relevant for aging issues [7], these features are sufficient alone to provide accurate age estimates due to the huge growth variability among fish individuals.

### 2.2. Geometrical features

The second category of features we use relies on the characterization of the external shape of the otolith. We

typically compute the perimeter, the surface, and the rate between the length along the two principal axis.

This type of feature is obviously not sufficient alone to infer fish age as shown in [1]. However, we aim at evaluating whether or not it can bring additional cues in combination to other feature types to improve fish age estimation.

## 3. Image-related features

The core of our study is the extraction of relevant features to represent the content of otolith images, in terms of alternation of translucent and opaque rings. We will investigate the use of two different feature sets: frequency-based and peak-based features.

### 3.1. From 2D image content to 1D representation

Otolith interpretation is usually performed by experts along a predefined reading axis, as shown by Fig.1 for a Plaice otolith. The automatic analysis of the content of otolith images can then be restricted to a region around the main reading axis.

To this end, a template-based approach inspired from [9] extracts a one-dimensional signal depicting the 2D image content, relevant for interpretation purposes, within an angular sector around the considered reading axis. More precisely, given the otolith center 0 and an angular sector $\mathcal{S}$, the following 1D signal $s_{\mathcal{S}}$ is computed:

$$s_{\mathcal{S}}(\alpha) = med(I(O + \alpha * T_{\mathcal{S}})) \quad (1)$$

where $med()$ is the median operator, $T_{\mathcal{S}}$ is the template model used within $\mathcal{S}$, $I$ the image intensity function, and $\alpha$ a scaling factor between 0 and 1. As in [9], the template model $T_{\mathcal{S}}$ is given by the sampled external otolith shape within $\mathcal{S}$. $s_{\mathcal{S}}(\alpha)$ is then the median intensity value along the template $T_{\mathcal{S}}$ scaled by $\alpha$ w.r.t. the otolith center $O$. Since the scaling of the external otolith shape w.r.t. the otolith center provides a good approximation of the shape of the seasonal rings [9], $s_{\mathcal{S}}$ depicts oscillations corresponding to the alternation of translucent and opaque rings within $\mathcal{S}$.

### 3.2. Growth-adapted filtering

Otolith growth is not linear but rather exponential. Therefore, rings tend to be thiner when the fish gets older. In terms of frequency content, the signal $s_{\mathcal{S}}$ is then highly non-stationary. In addition, the otolith is usually thiner close to its edge, what results in a low-frequency component in $s_{\mathcal{S}}$. Due to these two major issues, the straightforward use of FFT or DCT features computed from $s_{\mathcal{S}}$ as in [7] does not seem a relevant choice.

Contrarily, we develop a growth-adapted filtering to analyze $s_{\mathcal{S}}$. Given a mean *a priori* growth model $L = \Phi(t)$, with $L$ the distance to the otolith center, and $t$ the time variable in years, $s_{\mathcal{S}}$ is initially demodulated w.r.t. $\Phi$:

$$s_{\mathcal{S}}^{DM}(t) = s_{\mathcal{S}}(\Phi(t)/L_{max}), \quad (2)$$

where $L_{max}$ is the median length between the center of the otolith and the edge within $\mathcal{S}$. Though non-stationary, the demodulated signal $s_{\mathcal{S}}^{DM}$ involves a more compact frequency content, what makes easier its analysis.

To remove the low-frequency component $s_{\mathcal{S}}^{T}$ from $s_{\mathcal{S}}^{DM}$, $s_{\mathcal{S}}^{T}$ is estimated using a convolution to a Gaussian kernel $g_{\sigma_T}$ with a large variance $\sigma_T^2$: $s_{\mathcal{S}}^{T} = g_{\sigma_T} * s_{\mathcal{S}}^{DM}$. High-frequency noise can also be filtered out using a Gaussian kernel with a low variance $\sigma_G$, such that the growth-adapted filtering finally comes to consider the filtered signal $s_{\mathcal{S}}^{G}$ computed as:

$$s_{\mathcal{S}}^{G}(t) = g_{\sigma_G} * \left[s_{\mathcal{S}}^{DM} - s_{\mathcal{S}}^{T}\right] \quad (3)$$

### 3.3. Frequency-based representation

Given the demodulated and filtered signal $s_{\mathcal{S}}^{G}$, one can use its frequency content as a descriptor of the alternation of ridges and valleys within $\mathcal{S}$. The DCT transform is then applied to $s_{\mathcal{S}}^{G}$, and the $N_{DCT}$ first DCT coefficients are stored to form the frequency-based feature vector.

### 3.4. Peak-based representation

Otolith interpretations performed by experts do not rely on the analysis of the image frequency content, but rather on the detection of ridge and valley structures corresponding to translucent and opaque rings. We then introduce a second kind of image-related features defined from the extrema of the signal $s_{\mathcal{S}}^{G}$. Given the positions of maxima (or equivalently the minima) of $s_{\mathcal{S}}^{G}$ $\{t_1, ....., t_n\}$, we define a new peak-based signal $s_{\mathcal{S}}^{PB}$ by:

$$s_{\mathcal{S}}^{PB}(t) = min\left\{\rho(t - t_k)\right\}_{k\in\{1,\ldots,n\}} \quad (4)$$

where $\rho$ is a distance kernel. The computation of $\{t_1, ....., t_n\}$ is based on the detection of the zero crossings of the first derivative of $s_{\mathcal{S}}^{G}$, estimated by finite differences. In practice, we use $\rho(u) = 2atan(u)/\pi$. The resulting signal $s_{\mathcal{S}}^{PB}$ only depends on the extrema position. In particular, compared to $s_{\mathcal{S}}^{G}$, $s_{\mathcal{S}}^{PB}$ is invariant to the variability of intensity ranges over images.

## 4. Statistical learning

Fish age estimation is regarded as a pattern classification issue. Given an otolith image $n$, we aim at mapping the associated feature vector $x^n$ to an age class $y^n$. To cope with this multi-class classification, we consider two different statistical learning techniques: neural networks and SVM.

### 4.1. Neural networks

Neural networks have been widely used for pattern recognition issues [2]. They are aimed at optimizing the interconnection weights between neurons to minimize the mean square error. Here, a back propagation neural network with one hidden layer is chosen. The decision function is a sigmoid. The number of neurons of the hidden layer has been empirically optimized, and set to fifty. In our experiments, we use the Netlab Matlab toolbox [6].

### 4.2. SVM

Kernel-based techniques such as SVMs [8] are among the most efficient approaches for pattern recognition. The key idea of SVMS, which are maximum margin classifiers, is to map the original feature space to a higher dimensional. The selected mapping $\Psi$ has to verify that the dot product between two points in the new feature space $\Psi(x_1) \cdot \Psi(x_2)$ can be rewritten as a kernel function $K(x_1, x_2)$. The binary SVM classification then resorts to determine the hyperplane separating the two classes within the feature space associated to the kernel function. We let the reader refer to [8] for further details.

In our study, we will use a Gaussian kernel, $K(x_1, x_2) = exp(-(x_2 - x_2)^2/2\sigma^2)$. The binary SVM classification is extended to a multi-class classification using a one vs. all strategy. We use the libSVM package [3] to test for two different SVM techniques: C-SVM and nu-SVM [3, 8].

### 4.3. Training issues

Given a training set $\mathcal{T}$ of otolith images and the associated feature vectors $\{x^k\}_{t\in\mathcal{T}}$ ($x^k$ possibly refers to frequency-based, peak-based, biological or geometrical features, or a combination of these features), we first apply a Principal Component Analysis (PCA) to retain the first $N_{PCA}$ components. The resulting feature set is denoted by $\{x^k_{PCA}\}_{k\in\mathcal{T}}$. Statistical learning is then performed within the associated feature space.

To increase the robustness of the learning stage, we virtually increase the number of elements within the training set $\mathcal{T}$ as follows. Each element $k$ within $\mathcal{T}$ is associated to a growth pattern $L = \Phi_k(t)$ provided by the interpretation of the otolith made by the expert. Given $k \in \mathcal{T}$, the template-based signal $s^k_S$ (cf. Eq.1) is demodulated w.r.t. the growth model $L = \Phi_k(t)$ and remodulated w.r.t. growth models randomly chosen within $\{\Phi_n\}_{n\in\mathcal{T}}$. The image-related features computed for these virtual signals are then combined to the original ones to create a larger training set. The remodulation is typically performed with five growth models. Thus, the size of the training is multiplied by a factor 6six. This is expected to improve the generalization performed by the statistical learning techniques.

## 5. Experiments

### 5.1. Training and test sets

The experiments are carried out for a database of 300 images of plaice otoliths from age group 1 to 6. Due to the lower number of samples in age groups 5 and 6, we merge both groups to end up with a five-class issue: 1 to 4 and 5+. This grouping can also be motivated by the fact that age groups 4 and less usually represent more than 85% of the commercial landings. The ground truth for fish age estimation was provided by the interpretation of an expert in plaice otolith readings. The training set is formed by selecting the three fourth of the images of each age class, while the remaining images are used for the test set. For each image, image-related features are extracted for five non-overlapping angular sectors of width $\pi/50$ around the main reading axis. Angular sectors, for which information required to interpret the otolith is missing due to ring discontinuities or local artifacts, are discarded from the training set, but no such checking is carried out over the test set. We resort to training and test sets including respectively 1500 and 600 elements. Using the virtual growth-based signal generation, the training set finally comprises 8000 elements.

### 5.2. Classification results over the test set

We evaluate the performances of the proposed approach using different feature sets and classifiers. The performances, reported in Fig.2 are analyzed in terms of mean correct classification rate w.r.t. the five age groups over the test set. The overall evaluation shows that the best results (86%) are obtained by C-SVMs [8] combined to the peak-based features and the remodulation procedure within the training set. Thus, this configuration is used as the reference in the subsequent.

Fig.2.a highlights the need for non-linear techniques to cope with the considered classification task, and shows that C-SVMs outperform the neural nets and nu-SVMs [2, 8].

In Fig.2.b, we compare different feature types: peak-based features, peak-based features (PB), frequency-based features combined to the remodulation procedure (FB), DCT coefficients of the signal $s_S$ as used in [7] (DCT), biological features (B) and geometrical features (G). We also analyze the relevance of the remodulation process (Rem) to synthesize new examples within the training set. Fig.2.b first indicates that biological and geometrical features are far too simple to provide alone relevant cues to infer fish age. The influence of the growth-adapted filtering is also demonstrated by the rather low performance achieved using the DCT features from the raw signal. Besides, it proves the superiority of PB features over FB features and validates the interest of the remodulation procedure, which brings a sig-

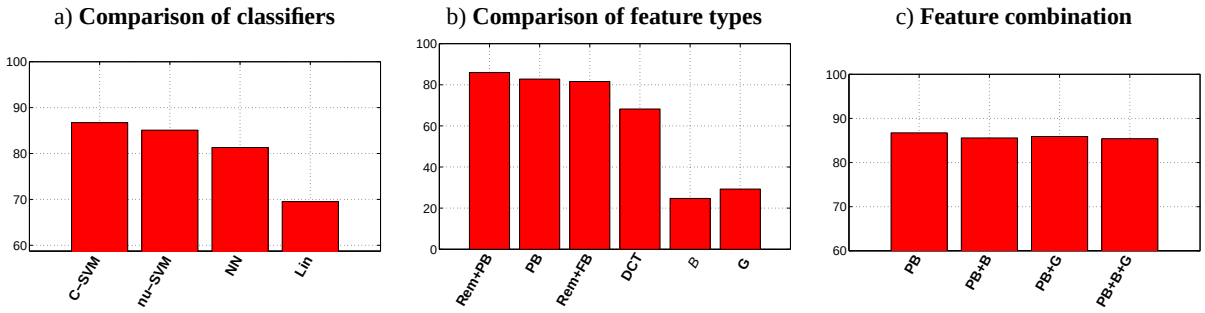


**Figure 2.** *Overall evaluation of fish age estimation for the plaice test set in terms of rate of correct classification. a) Given the peak-based feature vectors, we compare four different classifiers: C-SVMs, nu-SVMs, neural nets (NN) and linear SVMs. b) using C-SVM, we evaluate the relevance of four feature vectors: peak-based features (PB), frequency-based features (FB), biological features (B) and geometrical features (G). c) Using C-SVM, we compare combinations of peak-based, biological and geometrical features: PB, PB+B, PB+G, PB+B+G. In all cases, we use the remodulation procedure over the training set.*

| | 1 | 2 | 3 | 4 | 5+ | error in [-1,+1] |
|---|---|---|---|---|---|---|
| 1 | 91.6 | 8.3 | | | | 100 |
| 2 | | 91.1 | 5.8 | | 2.9 | 97 |
| 3 | | 4.1 | 91.6 | 4.1 | | 100 |
| 4 | | 4.3 | 13 | 78.2 | 4.3 | 95.5 |
| 5+ | | | 5.8 | 5.8 | 88.2 | 94.1 |

**Table 1.** *Confusion matrix of the results of fish aging for each image of the test set by combining the classifications of individuals angular sectors.*

nificant improvement from 82.8% to 86% of mean correct classification rate when using PB features.

We also manage to combine biological and geometrical features to PB features. Surprisingly as reported in Fig.2.c, this does not bring any improvement compared to PB features alone. This combination indeed tends to decrease the mean correct classification rate, while the mean-square error is slightly lowered. This is certainly due to the high overlapping in terms of length and shape features between the different age groups, which decreases the discriminant power of these features.

### 5.3 Extension to image classification

From this overall evaluation over the test set, the best configuration is to combine C-SVMs to PB features and to exploit the remodulation procedure. Besides, we can supply a more robust image classification by combining the classes assigned to several angular sectors. To this end, we use a voting procedure: given the age groups assigned to five angular sectors of the tested image around the main reading axis, this image is classified as belonging to the age group with the maximum number of votes. Tab.1 reports the confusion matrix of the resulting image-based classification. The mean correct classification rate is 88%, and only age group 4 has a classification rate just below 80%. These results are very promising for an application to routine aging, since they are in the range of the levels of agreement observed between expert readers.

Compared to previous work using statistical learning for fish aging [5], the reported classification rates are much greater due to the definition of more relevant image-based features. Besides, as reported in [4], previously proposed 1D and 2D techniques are both outperformed by the proposed statistical learning scheme with respectively 50% and 80% of correct classification. Besides, compared to 2D approaches which require about one minute per image, the SVM classification from PB features just needs a few seconds on a 2.4GHz PC in a Matlab implementation.

**Acknoledgments :** *The authors wish to thank André Ogor from for his help in the acquisition of otolith images, and Marie-Line Manten who provides the expert interpretation of the image of plaice otoliths.*

## References

[1] FAbOSA: Fish Ageing by Otolith Shape Analysis. FAIR CT97 3402 report.

[2] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[3] C.-C. Chang and C.-J. Lin. LibSVM: a library for Support Vector Machines. *www.csie.ntu.edu.tw/cjlin*.

[4] A. Guillaud etal. Autonomous agents for edge detection and continuity perception on otolith images. *Image and Vision Computing*, 20(13-14):955–968, 2002.

[5] A. Morison, S. Robertson, and D. Smith. An integrated system for production fish ageing: image analysis and quality insurance. *North American Journal of Fisheries Management*, 18:587–598, 1998.

[6] I. Nabney. *Netlab: Algorithms for Pattern Recognition* Springer, 2001.

[7] S. Robertson and A. Morison. Development of an artificial neural network for automated age estimation. Technical Report 98/105, Marine and Freshwater Resources Institute, 1998.

[8] B. Scholkopf and A. J. Smola. *Learning With Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2002.

[9] H. Traodec etal. Use of deformable templates for otolith 2D growth ring detection by digital image processing. *Jal of Fish. Res.*, 46(1-3):155–163, 2000.