Article

Application of Entropy and Fractal Dimension Analyses to the Pattern Recognition of Contaminated Fish Responses in Aquaculture

Harkaitz Eguiraun [1,*], Karmele López-de-Ipiña [2] and Iciar Martinez [1,3,4]

[1] Research Center for Experimental Marine Biology and Biotechnology—Plentziako Itsas Estazioa (PIE), University of the Basque Country UPV/EHU, Plentzia 48620, Spain;
E-Mail: iciar.martinez@ehu.es

[2] Department of Systems Engineering and Automatics, University of the Basque Country UPV/EHU, Donostia 20018 Spain; E-Mail: karmele.ipina@ehu.es

[3] IKERBASQUE Basque Foundation for Science, Bilbao 48013, Spain

[4] Norwegian College of Fishery Science, Faculty of Biosciences, Fisheries and Economics, University of Tromsø, Tromsø N-9037, Norway

* Author to whom correspondence should be addressed; E-Mail: harkaitz.eguiraun@ehu.es;
Tel.: +34-946-018-448.

Abstract: The objective of the work was to develop a non-invasive methodology for image acquisition, processing and nonlinear trajectory analysis of the collective fish response to a stochastic event. Object detection and motion estimation were performed by an optical flow algorithm in order to detect moving fish and simultaneously eliminate background, noise and artifacts. The Entropy and the Fractal Dimension (FD) of the trajectory followed by the centroids of the groups of fish were calculated using Shannon and permutation Entropy and the Katz, Higuchi and Katz-Castiglioni's FD algorithms respectively. The methodology was tested on three case groups of European sea bass (*Dicentrarchus labrax*), two of which were similar (C1 control and C2 tagged fish) and very different from the third (C3, tagged fish submerged in methylmercury contaminated water). The results indicate that Shannon entropy and Katz-Castiglioni were the most sensitive algorithms and proved to be promising tools for the non-invasive identification and quantification of differences in fish responses. In conclusion, we believe that this methodology has the potential to be embedded in

online/real time architecture for contaminant monitoring programs in the aquaculture industry.

1. Introduction

Biological systems, such as a group of animals, are regulated by interacting mechanisms that operate across multiple spatial and temporal scales. When studying such a biological system, we are interested, as defined by Kitano [1], in how the large numbers of functionally diverse and multifunctional set of elements (*i.e.*, the individual animals in our case) interact selectively and nonlinearly to produce coherent rather than complex behaviours.

The output variables from biological systems usually have complex fluctuations, which contain valuable information about their intrinsic dynamics. Signals and noise are the basic components of signal and data analysis, but most of the time series generated by biological systems contain both deterministic and stochastic components [2]. Classical approaches are not able to quantify the complexity of these systems, this is why real world applications and noisy environments often require alternative techniques leading to improved systems. Thus, the combination of nonlinear and linear modelling and/or features has led to higher and more robust performance, something particularly promising for solving complex tasks in real environments [3].

One of the main goals of applied mathematics and computing in modern biology is to capture all the richness of complex biological systems into theoretical models [1,4]. Fractal analysis has found a wide area of applications in a variety of scientific fields from medicine [5–7] to speech recognition [8], walking pattern recognition [9], stress indicators in goats [10] and in fish motion studies [11–13]. The Fractal Dimension (FD) of a system is one of the most significant features to describe its complexity. Fractal systems have a characteristic called self-similarity, *i.e.*, a close-up examination of the system reveals that it is composed of smaller versions of itself. Self-similarity can be quantified as a relative measure of the number of basic building blocks that form a pattern, and this measure is defined as the FD, which is rarely an integer number. Usually, the more complex the signal, the higher its FD value [14]. FD analysis has been successfully applied by López-de-Ipiña [15] to speech analysis and by Nimkerdphol and Nakagawa [16] to show quantitative differences in the swimming behaviour of zebrafish provoked by the presence of hypochlorite in the water.

The Entropy of a system is also a nonlinear measurement that has found application in complex biological systems and has occasionally been decisive to understand the nonlinear nature of a problem. For instance, Kulish *et al.* [17] analyzed brain activity using the spectra of the FD based on the Renyi entropy: combined with a visualization tool, these authors showed an intrinsic asymmetry of the brain activity. Permutation entropy has been used in a wide range of applications where measurement of complex time series where needed. As an example Li [18] measured the effects of sevofluore on the the complexity in electroencephalographic series, Liu *et al.* [19] analysed the movement of the fruit fly and Bandt and Pompe [20] studied the complexity of chaotic time series.

Non-invasive pattern analysis of the responses of animals to unexpected events may have application in several fields, including monitoring of animal welfare and detection of contaminants, air pollutants or oxygen availability in closed environments. In contrast to analyzing the behavior, which is a very complex and species-specific attribute, the immediate response to a stochastic event is simpler, requires less computational power and once a suitable methodology is developed we believe that it will require less modifications in order to be applicable to other species and settings.

The animal case selected for this work was fish due to the worldwide spectacular increase in aquaculture production. This increase has created a need to non-invasively monitor and control relevant variables during the production, in particular those relating to seafood safety and to the welfare and health of the fish. Some studies have been conducted to build up machine-based systems for fish disease diagnosis. In those studies the main challenge is to develop the knowledge database, a task which is time and resource consuming and focuses mainly on water quality [21] and/or on nutritional problems, parasites, viruses, bacteria and fungal agent diseases [22]. Other studies aim at developing early warning systems [23], including the use of sms alert procedures [24]. The main challenge for these fish-health oriented studies is the development of the decision component rather than the warning systems itself.

In addition, suboptimal environmental conditions, such as hypoxia, feeding regime [25], high fish density [26,27], the presence contaminants in the water including human drugs [28] and hypochlorite, and the water's pH [16] have been shown to alter fish behaviour. This opens the possibility of using fish behaviour itself as a biomarker for environmental monitoring and in aquaculture settings. Fish behaviour is very complex [29] and the response to a stochastic event may be considered one aspect of it. Following Kitano's idea [1], we were interested in studying the coherent response of the group rather than the individual response of each fish, which in addition to requiring much more computational effort it may be impractical in real-life settings where there may be several thousand individuals in the same area or cage.

The objective of the present work was to develop a methodology for image acquisition, processing and nonlinear trajectory analysis of the collective fish response to a stochastic event. The acquisition of the information of the movement of the fish can be achieved by images, typically video recording [30–32] and/or by echo-sounds [25]. The information thus obtained may then be processed by different nonlinear algorithms as indicated above. Video recording was chosen for its simplicity and FD and Entropy for their proven suitability to identify nonlinear features.

2. Materials and Methods

2.1. Experimental Cases

Three experimental cases were used to test the methodology: C1 and C2, consisting of 81 fish each and differing only in that C2 fish were tagged with Visual Implant Elastomer by Northwest Marine Technology [33] and C3 with 41 fish that had been treated for 9 days with 4 µg methylmercury/L. During all the experimental period the fish were subject to a 12h /12h dark/light photoperiod and they were fed once a day INICIO Plus feed from BioMar (56% crude protein, 18% crude fat). The fish were placed in tanks (100 cm × 100 cm × 90 cm) filled up to 80.5 cm of height with 810 L of aerated seawater under

direct light (2 × 58 W and 5200 lumen) avoiding shadows as much as possible. To record the fish response one camera was placed in each tank positioned exactly in the same place.

The experimental procedure was approved by The Ethical Committee for Animal Welfare No. CEBA/285/2013MG. The fish were European sea bass (*Dicentrarchus labrax*, 4 ± 2 g, 8 ± 1 cm) generously provided by Grupo Tinamenor (Cantabria, Spain).

2.2. Image Acquisition and Pre-Processing

The schematic diagram of the working procedure is described in Figure 1. The video sequences and images were acquired using a GoPro Hero3 high definition camera in its GoPro Underwater housing attached to the tank by GoPro Side and Flat mounts placed in the top right corner of the tank just 3.8 cm from the right wall, 15 cm from the top of the tank and 5.5 cm below water level. This camera was used for convenience: its size is small, thus minimizing the effect of introducing foreign objects in the tank, and it has a water-proof protective case very convenient because it has to be submerged in contaminated water, so that at the end of the experiment the camera can be reused while the case is discarded as contaminated material. The recordings were made in high definition RGB scheme at 1440p, 24 frames per second (fps) and in 16:9 picture size. For the recording a SanDisk 32 Gb Ultra microSDHC™ (Class 10) secure card was used and for post-storing a 2 Tb Hard Disk. In order to minimize stressing factors, continuous recordings were done until the batteries of the cameras run out, which was about 1 h and 30 min. Within this period, a stochastic event consisting of a sudden hit in the tank was introduced and the 30 s pre- and 3 min post-event were processed (Figure 2).

Figure 1. Experiment's workflow showing the steps from image acquisition and processing to data treatment.
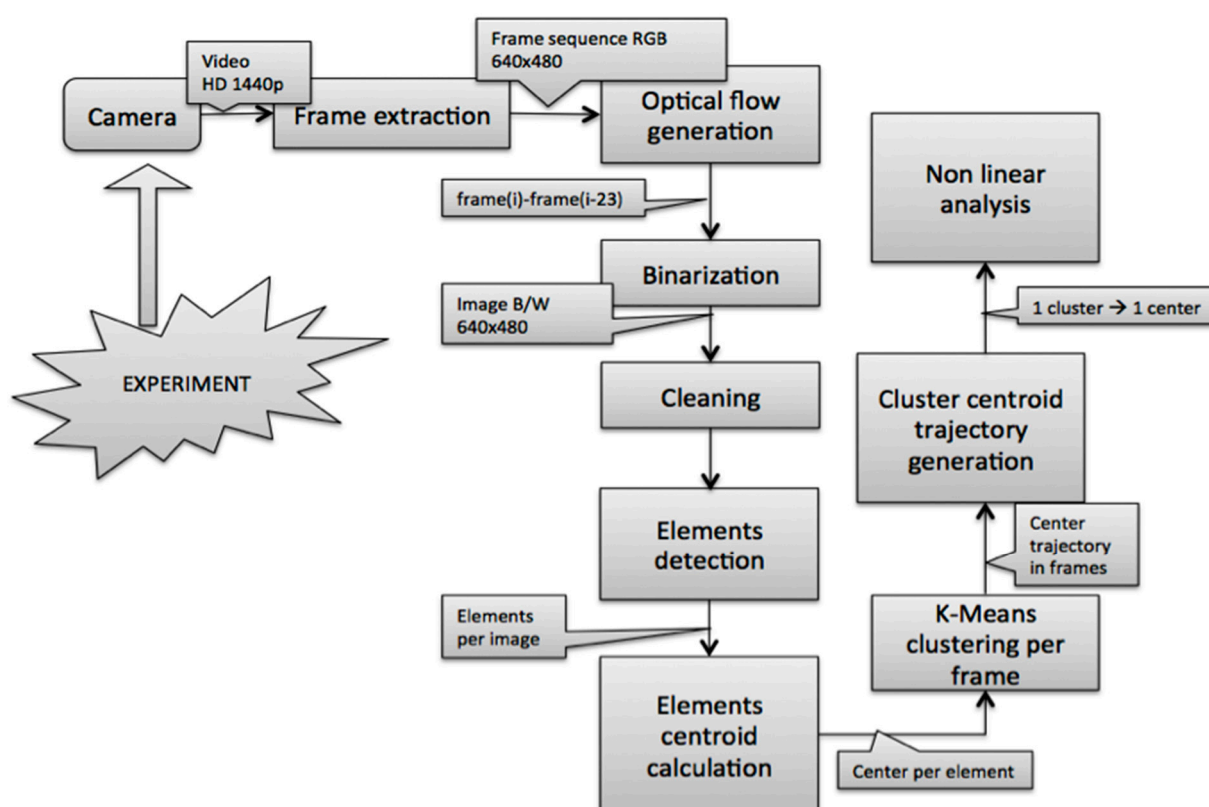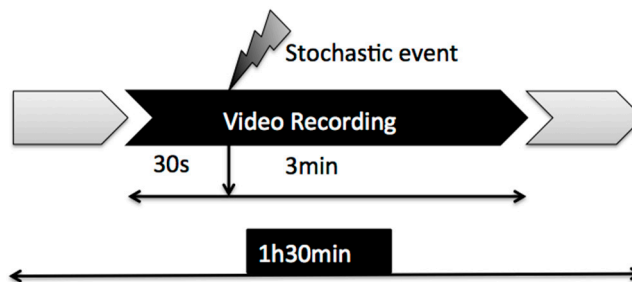
Figure 2. Recording procedure. The total recording time was 1h 30 min. The results presented here correspond to the analysis the 30 s pre- and 3 min post-event.

The 3 min 30 s of interest were located using the sound clip of each video analyzed with Audacy free software to determine exactly where the event happened. The 3 min 30 s clip was cut from the main video and converted into a sequence of images. Since the video was recorded at 24 fps it was converted also to 24 fps. The images were compressed from the 1440p HD format to the more convenient 640 pixel × 480 pixel format. Frame extraction and format conversion were made with the commercial iMovie software. After the video was converted into a color image sequence, the background and noises were eliminated and it was then converted to black and white.

2.3. Detection of Objects and Motion Estimation

From the point of view of image segmentation and object detection, and due to the nature of the set up, (a biological experiment in a real, small-scale industrial environment) there were three main problems: noise, artefacts and occlusions.

Two main sources of noise were identified: air bubbles and shadows. The main noise was generated by the air bubbles moving towards the surface. This creates a little wave system on the water surface, which makes the light penetrate the water in a nonlinear manner. The second noise source was the shadows of the fish on the bottom and the tank's walls. Although the lighting was placed on the ceiling above the tank to avoid this issue, the generation of some shadows was unavoidable (see Figure 3).

Figure 3. Example of an artefact indicated by the arrow, caused by a feed pellet. Occasionally, the system also had difficulties to discriminate some fish from their shadows (circle).

There were three main types of artefacts (anomalies introduced in the signal or in the data by the equipment or the technique): the first was caused by the pellets used for feeding the fish. Some food pellets were suspended in the water and when fish swam around them, they spread off forming black holes in the images (see Figure 3). The second (very similar to the first but smaller in size) was caused by the faeces of the fish, and the third by the light's reflections on the skin of the fish.

Occlusions are a well-known issue in tracking that occur when two or more tracked target images become one during a time period in the sequence. Occlusions are more frequent when target objects are similar to each other, as it happens in animal groups and fish [34]. Occlusions in fish tracking may lead to two types of misidentification: loss of fish identity and swapping identity between individuals [31]. Tracking problems take place both while the occlusion occurs and when the occlusion ends and the fish appear separately in the image. To solve this problem different solutions have been developed such as the use of 3D information [35,36], using the animals' characteristics, such as its shape [37–39] or analyzing the special topology of the shape [40–43]. Automatic scene calibrating systems are also very helpful tools and many approaches have been made in this field, for example an automatic calibrating camera system for tracking people [44].

Given that the robustness of the system depends on how the motion detection takes place and all the problems listed above severely limit the election of the motion estimation algorithm, we decided to use an algorithm based on optical flow in order to eliminate noise, artefacts and occlusions in one step. This conventional approach is based on the calculation of the local relative motion [45] which is also used for space segmentation [46], and it has the additional advantage that it can work with a moving camera and/or with moving backgrounds. We applied this method to detect objects and estimate their motion by a simple process which consisted in identifying the differences between one of the images and the image obtained in the previous second, *i.e.*, resting each frame from its 24th predecessor (since we work at 24 fps). This made it possible to delete background, noise and artifacts common to all images, including objects that had not changed position in the previous second while keeping those objects whose position changes in 1 second intervals, *i.e.*, the moving fish. Methods based on optical flow provide very valuable information but they are computationally intense and sometimes require specific hardware.

After the optical flow, or motion, was calculated, the images were binarized using standard morphologic operations in order to be able to detect the elements in the image and their centers in each frame.

2.4. Clustering and Trajectory Generation

In order to work in the most reliable way possible, knowing that our system, experiment and conditions have limitations and being particularly concerned about a potential loss of information due to the image segmentation and processing methods, we decided to use a clustering method to identify the fish group and calculate the group's centroid.

The centroids' positions were estimated by k-means because this algorithm is robust, with a good relationship between speed and stability and it works well with large amounts of data. Thus, once the centers of the objects were calculated, and knowing their coordinates in the two axes within each frame, k-means was applied to find the center of the entire group. In our particular case, the dataset were the objects' centers in each frame, from the first frame to the last one. K-means clustering operates on actual observations (rather than on a larger set of dissimilarity measures), and creates a single level of clusters.

K-means then treats each observation in the data set as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other, and as far from objects in other clusters, as possible. Depending on the kind of data to cluster, it is possible to use up to five different methods to calculate the distances. The best results for the current problem were obtained using the Euclidean distance [47,48].
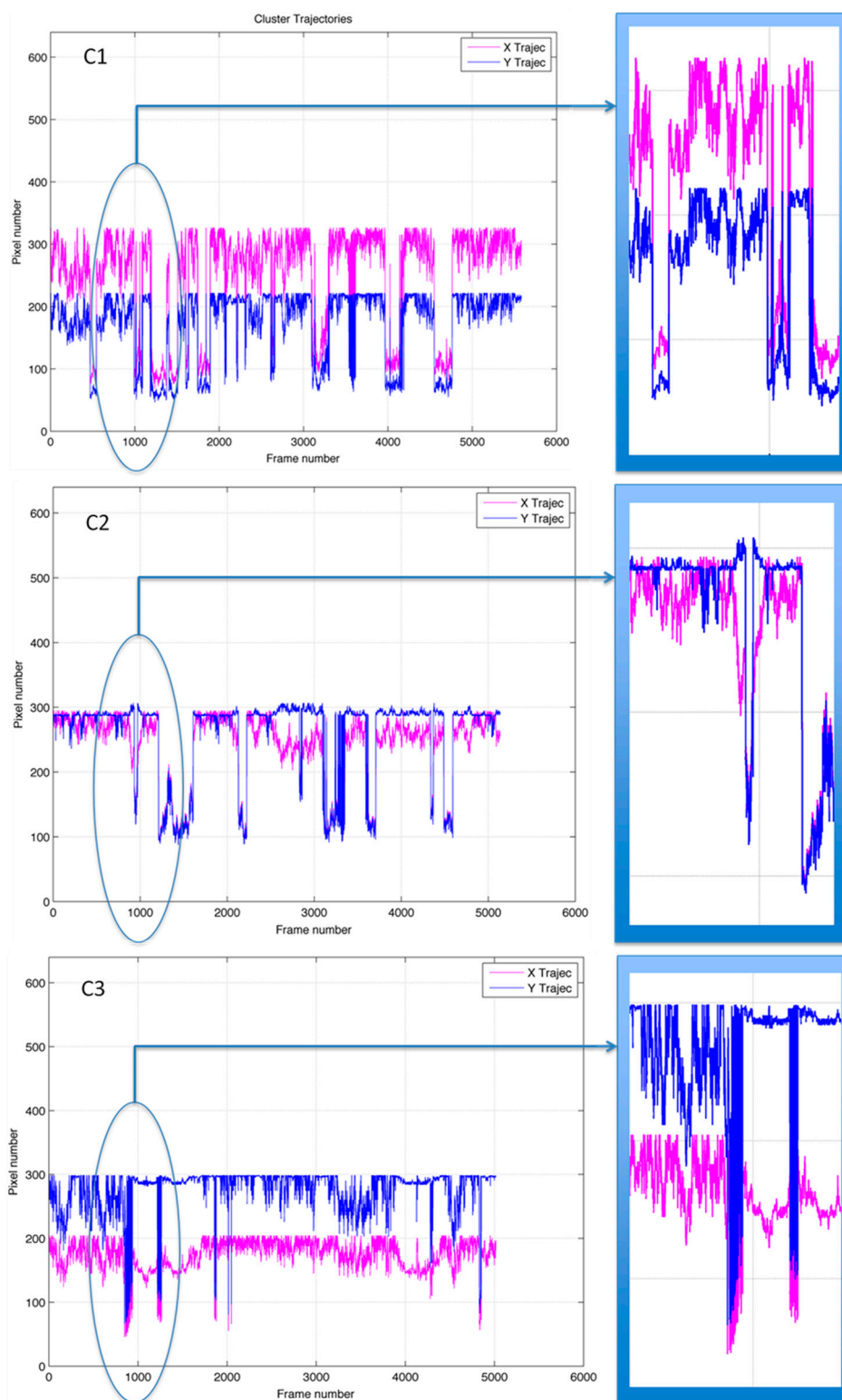
Each cluster in the partition was defined by its member objects and by its centroid, or center: the point where the sum of distances from all objects in the cluster is minimized. K-means computes the cluster's centroids differently for each distance measured in order to minimize the sum with respect to the distance specified. This is done using an iterative algorithm that minimizes the sum of the distances from each object to its cluster's centroid, over all clusters. The algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and separated as possible. It is possible to control the details of the minimization using several optional input parameters to k-means, including the initial values of the cluster centroids, and the maximum number of iterations. This algorithm, often applied for image segmentation, has successfully been used to detect animals [48]. The centers of the clusters are calculated in a two space dimension (2D), using both coordinates (X and Y) and the frame number, which goes from 1 to n in the image sequence, and then a trajectory table is built (Table 1).

Table 1. The cluster's centroid coordinates are calculated for each frame and from them the trajectory of the cluster is calculated.

Frame Number	X Coordinate	Y Coordinate	Centroid's Coordinates
1	x_1	x	x_1,y_1
2	x_2	y_2	x_2,y_2
...	...	...	...
n	x_n	y_n	x_n,y_n

A new figure (Figure 4) was then created by plotting the values obtained from Table 1: in the vertical axes the values corresponding to the pixel number of the *x*-coordinate of the centroid (in red) and to the pixel number of the *y*-coordinate of the centroid (in blue) and in the horizontal axis the frame number to which each *x* and *y* values correspond. Since the images have a 640 pixel × 480 pixel format, the scale of the vertical axis in Figure 4 goes from 0 to 640 pixel number (the lowest and highest theoretical possible value for the *x*-pixel number in Table 1, when the centroid is placed either in the border right or border left of the frame). The lowest and highest theoretical possible value for the *y*-pixel number in Table are 0 and 480 respectively. The horizontal axis, representing the frame number varies in the three cases due to the processing of the video sequences. In all three cases there was a sharp change in the trajectory of the centroid in response to the stochastic event, which occurs around frame number 720. The panels to the right in Figure 4 show a magnification of the region of the plot corresponding to the response to the event.

Figure 4. Plot of the values obtained from Table 1 for the three cases examined: C1 (top), C2 (middel) and C3 (lower pannel). The horizontal axis represents the Frame Number (from 0 to about 6000 frames processed) and the vertical axis represents the Pixel Number in each frame for the *x* (red line, from 0 to 640) and *y* (blue line, from 0 to 480) coordinates of the centroid. A stochastic event took place around frame number 720, indicated by a circle and amplified in the right panel, which resulted in a sharp alteration of the centroids' trajectories in the three.

2.5. Fractal Dimension (FD)

Among the various algorithms available to measure the FD, we selected those specially suited to time series analyses that don't need previous modelling of the system. Two of those algorithms are Higuchi [49] and Katz [50] named after their authors. We used the former as the main reference, and a modification proposed by Castiglioni [51] on the original version developed by Katz [50]. Higuchi was our first choice because it has been reported to be more accurate [52] but in most of these studies, the algorithm compared to was the original one developed by Katz himself. Castiglioni's improvement is theoretically sound and has not been tested in many experiments, so we considered interesting to test as an alternative.

The algorithm proposed by Higuchi [49] measures the FD of discrete time sequences directly from the time series $\{x_1, x_2, \ldots, x_n\}$. The algorithm calculates the length $L_m(k)$ for each value of m and k, where m is initial time $\{m = 1, 2, \ldots, k\}$ and k is time interval $\{k = 1, 2, \ldots, k_{max}\}$. N is the lenght of the sampled signal.

$$L_m(k) = \frac{\sum_{i=1}^{\frac{N-m}{k}} |x(m + i_k) - x(m + (i-1)k)|(n-1)}{\left\lfloor \frac{N-m}{k} \right\rfloor k} \tag{1}$$

After that, a sum of all the lengths $L_m(k)$ for each k is determined by:

$$L(k) = \sum_{m=1}^{k} L_m(k) \tag{2}$$

And finally, the slope of the curve $\ln(L(k))/\ln(1/k)$ is estimated using the best fit by linear least squares. The result is the Higuchi FD.

On the other hand, Katz [50] proposed a normalized formula of the FD:

$$FD = \frac{\log(n)}{\log(n) + \log(d/L)} \tag{3}$$

where the length L and the extension d of the curve are normalized using the average step $a = L/n$ using (4) and (5).

$$L = \sum_{i=1}^{n} l_{i,i+1} \tag{4}$$

$$d = \max\{l_{i,j}\} \tag{5}$$

Nevertheless, given that the input signal is a mono-dimensional waveform, the length and the extension can be rewritten using Mandelbrot's approach [53]. A simple and efficient way to do this is directing these two magnitudes in their own dimension as it was done by Castiglioni [51]. One by one, the extension on the Y axis is the range of y_k as seen in:

$$d = \max\{y_k\} - \min\{y_k\} \tag{6}$$

and the length L is the sum of all the increments in modulus as in:

$$L = \sum_{k=1}^{n} |y_{k+1} - y_k| \tag{7}$$

This latter calculation is what is called the Castiglioni's variation of Katz's algorithm [51].

Once the FD algorithms were selected, it was extremely important to choose the window size to be used for the calculations. The fractal window goes through the signal, *i.e.*, the object to analyze. This sliding window has a fixed size during each analyzing period. In order to calculate the FD's features of the signal there are no restrictions other than the total waveform length to the window size to be used. Each signal was analyzed using three fixed, but configurable, window lengths: 320 points, 640 points and 1280 points. The third window size of 1280 points was tested because previous studies that take into account the window size of similar dimension estimations [52,54] suggest that a bigger window could be useful in some cases. Since the FD is a tool intended to capture the dynamics of the system, with a short window the estimation would be very local and adapting fast to the changes in the waveform. When the window is longer, some details will be lost but the FD anticipates better the characteristics of the signal [8].

Each of the three experimental cases, C1, C2 and C3, were treated by Higuch, Katz, and the Castiglioni's variation of Katz's algorithm using window lengths of 320, 640 and 1280 points.

2.6. Shannon Entropy

Entropy is the measure of disorder in physical systems. The entropy $H(X)$ of a single discrete random variable X is a measure of its average uncertainty. Shannon entropy [55] is calculated by the equation:

$$H(X) = -\sum_{x_i \in \Theta} p(x_i) \log p(x_i) = -E[\log p(x_i)] \tag{8}$$

where X represents a random variable with a set of values Θ and probability mass function $p(x_i) = P_r\{X = x_i\}, x_i \in \Theta$, and E represents the expectation operator. Note that $p \log p = 0$ if $p = 0$.

For a time series representing the output of a stochastic process, that is, an indexed sequence of n random variables, $\{Xi\} = \{X_1, \dots, X_n\}$, with a set of values $\theta_1, \dots, \theta_n$, respectively, and $Xi \epsilon \theta_i$, the joint entropy is defined by:

$$H_n = H(X_1, \dots, Xn) = -\sum_{x_i \in \Theta} \dots \sum_{x_n \in \theta_n} p(x_1, \dots, x_n) \log p(x_1, \dots, x_n) \tag{9}$$

where $p(x_1, \dots x_n) = P\{X_1 = x_1, \dots, X_n = x_n\}$ is the joint probability for the n variables $X_1, \dots, X_n$.

By applying the chain rule to Equation (9), the joint entropy can be written as a summation of conditional entropies, each of which is a non-negative quantity:

$$H_n = \sum_{i=1}^{n} H(X_i | X_{i-1}, \dots, X_1) \tag{10}$$

Therefore, it concludes that the joint entropy is an increasing function of n. The rate at which the joint entropy grows with n, *i.e.*, the entropy rate h, is defined as:

$$h = \lim_{n \to \infty} \frac{H_n}{n} \tag{11}$$

For stationary ergodic processes, the evaluation of the rate of entropy has proved to be a very useful parameter [56–61]. The Shannon entropy was calculated from the same trajectory signals of the clusters' centroid for both axis, X and Y, which had been used to estimate the FD shown in Figure 4.

2.7. Permutation Entropy

As Shannon entropy, permutation entropy quantifies the disorder of a system. It has shown a good ability to measure complexity for large time series and basically this method converts a time series into an ordinal patterns series where the order of relations between the present and a fixed number of equidistant past values at a given time are described [19,62].

Following this idea, Bandt and Pompe [20] proposed a permutation entropy method based on the Shannon entropy measurement with the purpose of visualizing and quantifying changes in the time series. The permutation entropy is calculated for a given time series $\{x_1, x_2, \ldots, x_n\}$ as a function of the scale factor s. In order to be able to compute the permutation of a new time vector X_j, $S_t = [X_t, X_{t+1}, \ldots, X_{t+m-1}]$ is generated with the embedding dimension m and then arranged in an increasing order: $[X_{t+j_1-1} \leq X_{t+j_2-1} \leq \cdots \leq X_{t+j_n-1}]$. Given m different values, there will be $m!$ possible patterns π, also known as permutations. If $f(\pi)$ denotes its frequency in the time series, its relative frequency is $p(\pi) = f(\pi)/(L/s - m + 1)$. The permutation entropy is then defined as:

$$PE = -\sum_{i=1}^{m!} p(\pi_i) \ln p(\pi_i) \tag{12}$$

Summarising, permutation entropy refers to the local order structure of the time series, which can give a quantitative measure of complexity for dynamical time series. This calculation depends on the selection of the m parameter, which is strictly related with the length N of the analysed signal. For example Bandt and Pompe [20] suggested the use of m = 3, ... ,7 following always this rule:

$$m! < N \tag{13}$$

If m is too small (smaller than 3) the algorithm will work wrongly because it will only have few distinct states for recording. When using long signals, a large value of m is preferable but it would require a larger computational time.

In our particular case and due to the computational cost derived from analysing signals composed from 5000 samples, the m parameter was fixed at a value of m = 4.

3. Results and Discussion

The non-invasive tool we wished to develop targeted the responses of the fish groups rather than that of individual fish both to reduce the computational effort and because the response of the group may be considered the result of integrating all the responses each individual fish, which may be influenced by the physiological status of the individual, its size, status in the school's hierarchy and other factors that are usually unknown in monitoring purposes. Also, the response to a stochastic event was measured instead of other behavioural aspects (swimming pattern, daily activity, feeding, aggressiveness, *etc.*) because it permits to restrict the computational analysis in time to the duration of the response (three minutes was sufficient in our case, rather than observing the animals for longer periods of time where more variables

may play a role) and to reproduce the event at will in other settings for comparison purposes. Longer periods of time were also analyzed but the discriminative power of the analyses did not improve (results not shown).

Of the three cases examined, we expected C1 and C2 to behave similarly to each other and to be clearly different from C3 due to the methylmercury contamination in C3. The neurotoxic agent methylmercury was selected because of its increasing relevance as an environmentally ubiquitous pollutant that accumulates and biomagnifies in the trophic chain [63]. Although differences in fish stocking density between C1 and C2 (81 fish) vs C3 (41 fish) might have had an effect on the responses to the event, most probably they did not [26,27].

Two of the three FD algorithms used, Higuchi and Katz's variation proposed by Castiglioni were able to differentiate C1, C2 and particularly C3, for the three sampling window lengths in both coordinates of the clusters' centers, X and Y in a 2D analysis, as shown in Figure 5. Since there was a correlation between the X and Y coordinates, the rest of the calculations were performed on only one of them, the X values. The almost constant and close to 1 value of the FD obtained by using the Katz algorithm is in agreement with the results of Raghavendra [64] and confirms Castiglioni's note [51].

Indeed the modification proposed by Castiglioni was more sensitive to detect differences between the three cases than Higuchi, suggesting that for our particular application, it may be the most suitable algorithm.

The median and the standard deviation of the FDs obtained for each case and window length for the Katz-Castiglioni algorithm on the X values are plotted in Figure 6 and shown in Table 2. C3 showed the highest FD median value for all three window lengths, with very similar values that were close to 2.5 (Table 2) but it also displayed the largest dispersion of values (Figure 6). C1 had the smallest dispersion of values (Figure 6) and they had a tendency to increase with increasing window length. The FD of C2 varied between 1.9 and 2.1. Increasing window length seemed to diminish the dispersion of the values in C1 and C2, but did not affect the degree of dispersion in C3 (Figure 6). Interestingly our results agree with those of [16] in spite of these authors using a different species (zebrafish, which is a freshwater species that prefers warmer water), only one individual and a different contaminant: the FD of the swimming trajectory of their zebrafish also trended to increasing concentrations of sodium hypochlorite contaminating its water, obtaining FD values of the swimming trajectories between 2.11 and 2.14. The value of the FD varies depending on the algorithm used for its calculation, on the different units composing the time series whether normalization has taken place or not as adressed by Fuss [65] and on the window length; with lower values of window length producing lower FD values. Optimization of the window length for our particular case produced FD values higher than 2, which is also in accordance with the results of Nimkerdphol and Nakagawa [16].

The Shannon Entropy values calculated on the same data as the FDs, *i.e.*, the trajectories displayed in Figure 4 are shown in Table 3. There was no difference between the entropy calculated on the X or the Y values. It is noteworthy the large difference between the entropy of C3 and those of C1 or C2, which only differed slightly from each other (Table 3). This seems to indicate that the Shannon entropy decreases with increasing perturbation of the fish: tagging having only a minimal and possibly non-significant effect, but the presence of the contaminant drastically diminished the entropy of the system by an entire unit.

Figure 5. Comparison of the three fractal dimension algorithms (Higuchi, blue lines; Katz, green lines and Katz-Castiglioni, red lines) evolution for each case signal (X, broken lines; and Y, full lines) and for each window length (left column 320, middle column 640 and right column 1280) for C1 (top row), C2 (middle row) and C3 (lower row). In each graphic the vertical 0Y axis is the fractal dimension of the cluster's centroids and the horizontal 0X axis is the evolution of the fractal dimension per sliding window length. As mentioned in the text, the fractal dimension for the Katz algorithm gave a value of 1 regardless of the signal case or the window length.
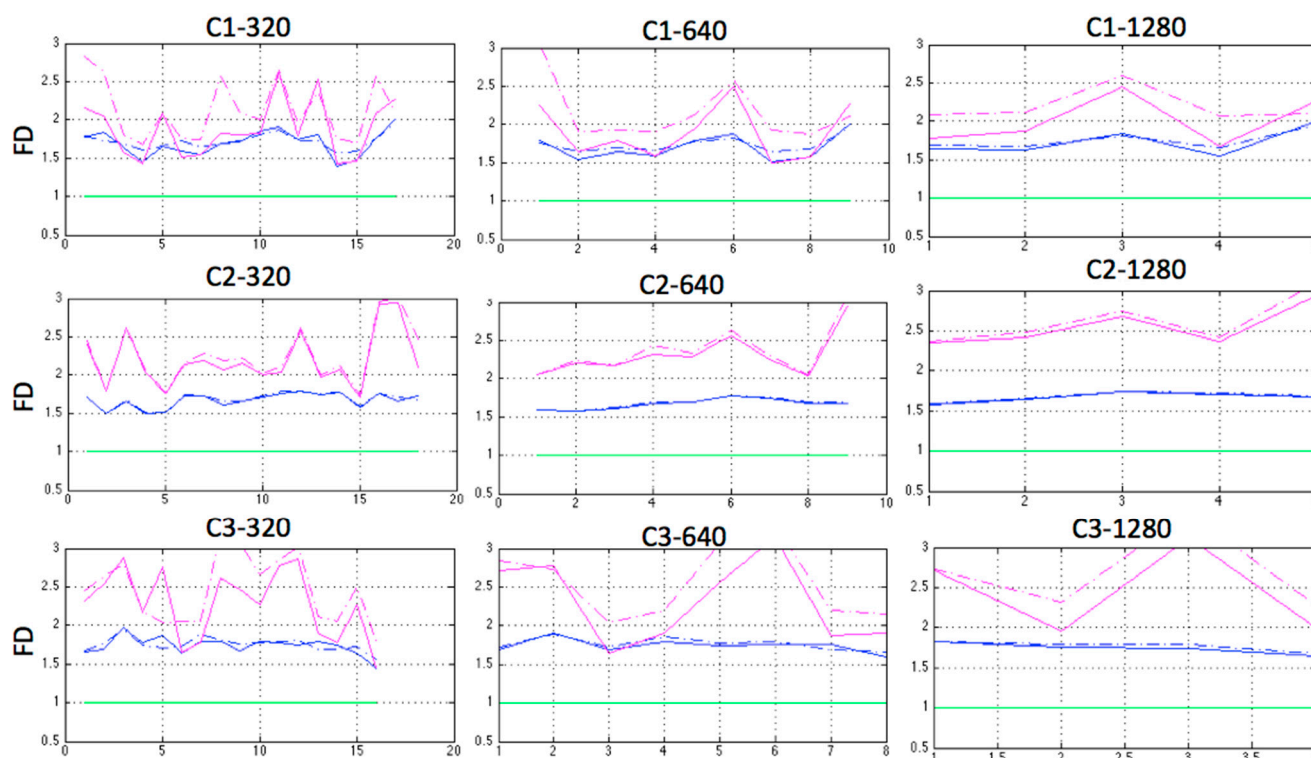

Table 2. Medians of the Fractal Dimensions Obtained for Each Sliding Window Lenght (320, 640 and 1280) in C1, C2, and C3. Note how close are the medians for C3.

Case	Sliding Window Length		
	320	640	1280
C1	2.1415	2.2624	2.4420
C2	2.0449	1.9188	2.1043
C3	2.4752	2.4618	2.5207

The permutation Entropy values calculated for the three experimental cases are also shown in Table 3. The results differ very slightly for X and Y signals and, in contrast to the results obtained using Shannon Entropy, the three analysed cases presented very similar permutation entropy values, although, in agreement with it, the permutation entropy values for C3 were smaller than for C1 and C2.

Technically, the method presented here demands a relatively large computation capability, particularly for the image processing step, which is of course susceptible of improvement. It must also be kept in mind on one hand that the analysis of the fish clusters trajectories does not depend on an image, they can

also be obtained from echo-sounds or infrared images and, on the other hand, that the methodology is not exclusive for fish and that with some modification may be applicable to other species.

> **Figure 6.** Boxplot representation of the fractal dimension data obtained by the Katz-Castiglioni algorithm, for each case (C1 upper, C2 middle and C3 lower row) and window length (320 left, 640 middle and 1280 right column). The upper and lower lines in each box represent the 75th and 25th percentile of the sample data respectively. The distance between those lines is the inter-quartile range and the red line inside the box is the sample data median. The whiskers (black lines) extend to the most extreme data points not considered outliers, while outliers (indicated by red plus signs) are data point whose value is outside 1.5 times the inter-quartile range.
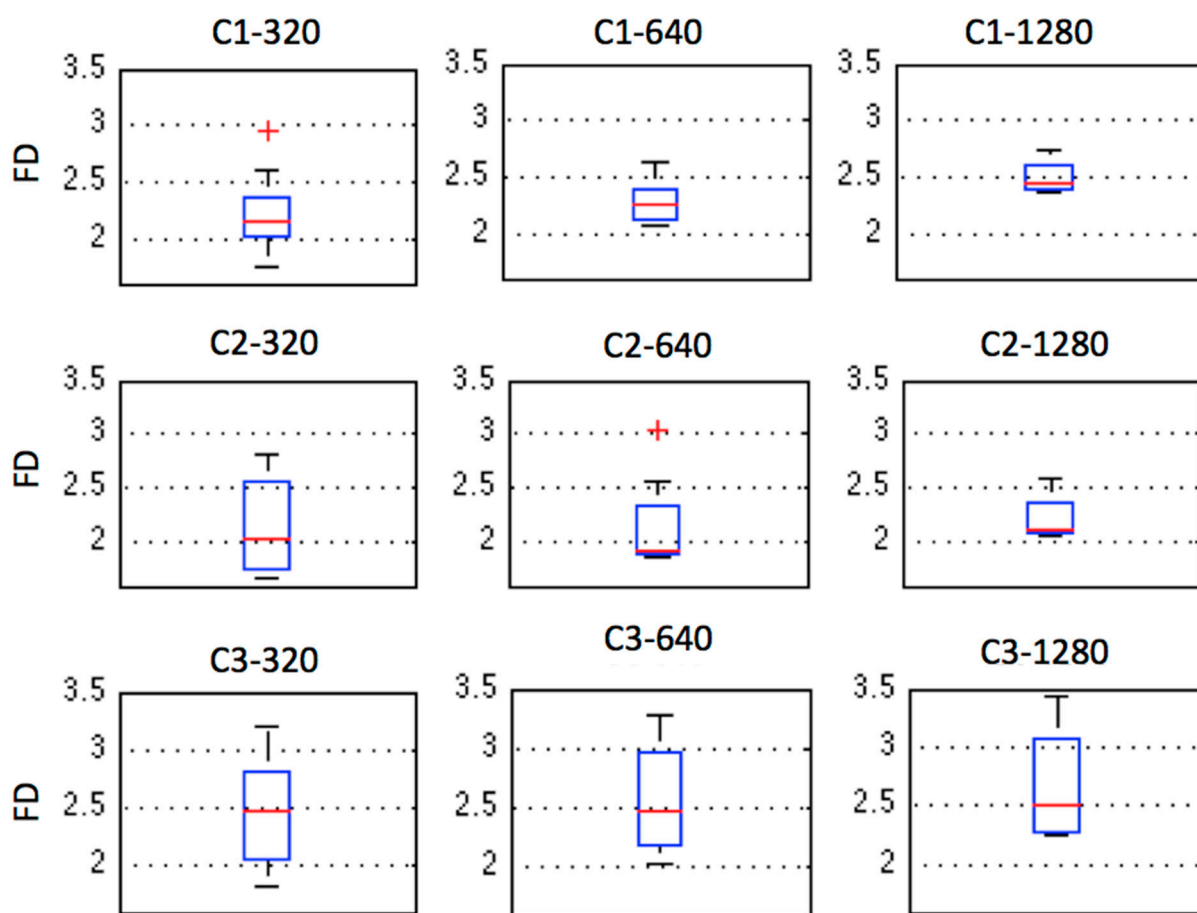

Table 3. Shannon Entropy Values Calculated for the X and Y Coordinates for C1, C2, and C3.

Case	Shannon Entropy Values		Permutation Entropy Values	
	X Coordinate	**Y Coordinate**	**X Coordinate**	**Y Coordinate**
C1	6.3016	6.3016	3.0881	3.0950
C2	6.2861	6.2861	3.1049	3.1250
C3	5.3628	5.3628	3.0413	3.0618

Of the tested methodologies, two of them, namely those based on the analysis of the Katz-Castiglioni FD and the Shannon entropy of the trajectories have been shown to be useful tools for non-invasive identification and quantification of changes of fish responses due to a highly relevant environmental

contaminant. We believe that it will be possible to embed this methodology in an on-line/real time architecture to monitor fish schools in a farm and in the wild, and that this kind of approach will find an application to identify contaminated waters in environmental monitoring programs.

4. Conclusions and Future Work

In conclusion, the present work describes a method for image acquisition, processing and nonlinear trajectory analysis suitable to identify variations in the response to an event of fish groups. The methodology here proposed shows a clear potential to aid implementing intelligent aquaculture systems but it needs to be tested with more contaminants, stressors, fish species and validated, prior to be embedded in real-time automatic systems using artificial intelligence methods.

Acknowledgments

Author Contributions

H.E. contributed to the design of the experiment, planned and performed the data analysis and interpretation of the results, and wrote the manuscript; K.L.I. contributed to the interpretation of the results and the drafting of the manuscript; I.M. contributed to the designed the experiment, interpretation of the results and drafting of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kitano, H. Computational systems biology. *Nature* **2002**, *420*, 206–210.
2. Costa, M.; Goldberger, A.; Peng, C.-K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 021906:1–021906:18.
3. Travieso, C.M.; Alonso, J.B. Special Issue on Advanced Cognitive Systems Based on Nonlinear Analysis. *Cogn. Comput.* **2013**, *5*, 397–398.
4. Spasic, S.; Savi, A.; Nikoli, L.; Budimir, S.; Janoševi, D. Applications of Higuchi's fractal dimension in the Analysis of Biological Signals. In Proceedings of 20th Telecommunications Forum TELFOR 2012, Belgrade, Serbia, 20–22 November 2012; pp. 639–641.
5. Accardo, A.; Affinito, M.; Carrozzi, M.; Bouquet, F. Use of the fractal dimension for the analysis of electroencephalographic time series. *Biol. Cybern.* **1997**, *77*, 339–350.

6.   Cáceres, J.H.; Sibat, Foyaca, H.; Hong, R.; Sautié, M.; Namugowa, A. Towards The Estimation of the Fractal Dimension of Heart Rate Variability Data. *Internet J. Cardiovasc. Res.* **2004**, *2*, number 1.

7.   Spasic, S.; Kesic, S.; Kalauzi, A.; Saponjic, J. Different anesthesia in rat induces distinct inter-structure brain dynamic detected by Higuchi fractal dimension. *Fractals* **2011**, *11*, 113–123.

8.   Ezeiza, A.; López de Ipiña, K.; Hernández, C.; Barroso, N. Enhancing the Feature Extraction Process for Automatic Speech Recognition with Fractal Dimensions. *Cogn. Comput.* **2013**, *5*, 545–550.

9.   Sekine, M.; Tamura, T.; Akay, M.; Fujimoto, T.; Togawa, T.; Fukui, Y. Discrimination of walking patterns using wavelet-based fractal analysis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2002**, *10*, 188–196.

10. Alados, C.L.; Escos, J.M.; Emlen, J.M. Fractal structure of sequential behaviour patterns: An indicator of stress. *Anim. Behav.* **1996**, *51*, 437–443.

11. Inada, Y.; Kawachi, K. Order and flexibility in the motion of fish schools. *J. Theor. Biol.* **2002**, *214*, 371–387.

12. Tikhonov, D.A.; Enderlein, J.; Malchow, H.; Medvinsky, A.B. Chaos and fractals in fish school motion. *Chaos Solitons Fractals* **2001**, *12*, 277–288.

13. Tikhonov, D.A.; Malchow, H. Chaos and fractals in fish school motion, II. *Chaos Solitons Fractals* **2003**, *16*, 287–289.

14. Kith, K.; Sourina, O.; Kulish, V.; Khoa, N.M. An algorithm for fractal dimension calculation based on Renyi entropy for short time signal analysis. In Proceedings of the 7th International Conference on Information, Communications and Signal Processing (ICICS), Macao, China, 7 December 2009; pp. 1–5.

15. López-de-Ipiña, K.; Alonso, J.-B.; Travieso, C.M.; Solé-Casals, J.; Eguiraun, H.; Faundez-Zanuy, M.; Ezeiza, A.; Barroso, N.; Ecay-Torres, M.; Martinez-Lage, P.; *et al*. On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. *Sensors* **2013**, *13*, 6730–6745.

16. Nimkerdphol, K.; Nakagawa, M. Effect of sodium hypochlorite on zebrafish swimming behavior estimated by fractal dimension analysis. *J. Biosci. Bioeng.* **2008**, *105*, 486–492.

17. Kulish, V.; Sourin, A.; Sourina, O. Human electroencephalograms seen as fractal time series: Mathematical analysis and visualization. *Comput. Biol. Med.* **2005**, *36*, 291–302.

18. Li, X.; Ph, D.; Cui, S.; Voss, L.J. Using permutation entropy to measure the electroencephalographic effects of sevoflurane. *Anesthesiology* **2008**, *109*, 448–456.

19. Liu, Y.; Chon, T.-S.; Baek, H.; Do, Y.; Choi, J.H.; Chung, Y.D. Permutation entropy applied to movement behaviors of Drosophila Melanogaster. *Mod. Phys. Lett. B* **2011**, *25*, 1133–1142.

20. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102.

21. Ma, H.; Tsai, T.-F.; Liu, C.-C. Real-time monitoring of water quality using temporal trajectory of live fish. *Expert Syst. Appl.* **2010**, *37*, 5158–5171.

22. Li, D.; Fu, Z.; Duan, Y. Fish-Expert: A web-based expert system for fish disease diagnosis. *Expert Syst. Appl.* **2002**, *23*, 311–320.

23. Li, N.; Wang, R.; Zhang, J.; Fu, Z.; Zhang, X. Developing a knowledge-based early warning system for fish disease/health via water quality management. *Expert Syst. Appl.* **2009**, *36*, 6500–6511.

24. Miaojun, X.; Jianke, Z.; Xiaoqiu, T. Intelligent Fish Disease Diagnostic System Based on SMS Platform. In Proceedings of the 3rd International Conference on Intelligent System Design and Engineering Applications, Hong Kong, China, 16–18 January 2013; pp. 897–900.

25. Polonschii, C.; Bratu, D.; Gheorghiu, E. Appraisal of fish behaviour based on time series of fish positions issued by a 3D array of ultrasound transducers. *Aquac. Eng.* **2013**, *55*, 37–45.

26. Di Marco, P.; Priori, A.; Finoia, M.G.; Massari, A.; Mandich, A. Marino, G. Physiological responses of European sea bass Dicentrarchus labrax to different stocking densities and acute stress challenge. *Aquaculture* **2008**, *275*, 319–328.

27. Papoutsoglou, S.; Tziha, G.; Vrettos, X.; Athanasiou, A. Effects of stocking density on behavior and growth rate of European sea bass (Dicentrarchus labrax) juveniles reared in a closed circulated system. *Aquac. Eng.* **1998**, *18*, 135–144.

28. Brodin, T.; Fick, J.; Jonssom, M.; Klaminder, J. Dilute concentrations of a psychiatric drug alter behavior of fish from natural populations. *Science* **2013**, *339*, 814–815.

29. Magnhagen, C.; Braithwaite, V.A.; Forsgren, E.; Kapoor, B.G. *Fish Behaviour*; Magnhagen, C., Braithwaite, V.A., Forsgren, E., Kapoor, B.G., Eds.; Science Publishers Inc.: Enfield, CT, USA, 2008; p. 646.

30. Kulish, V.; Sourin, A.; Sourina, O. Analysis and visualization of human electroencephalograms seen as fractal time series. *J. Mech. Med. Biol.* **2006**, *26*, 175–188.

31. Delcourt, J.; Becco, C.; Vandewalle, N.; Poncin, P. A video multitracking system for quantification of individual behavior in a large fish shoal: Advantages and limits. *Behav. Res. Methods* **2009**, *41*, 228–235.

32. Sourina, O.; Sourin, A.; Kulish, V. EEG Data Driven Animation and Its Application. In Proceedings of the International Conference Mirage 2009, Rocquencourt, France, 4–6 May 2009; pp. 380–388.

33. Brennan, N.P.; Leber, K.M.; Blankenship, H.L.; Ransier, J.M.; DeBruler, R. An Evaluation of Coded Wire and Elastomer Tag Performance in Juvenile Common Snook under Field and Laboratory Conditions. *North Am. J. Fish. Manag.* **2005**, *25*, 437–445.

34. Butail, S.; Chicoli, A.; Paley, D.A. Putting the Fish in the Fish Tank: Immersive VR for Animal Behavior Experiments. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, MN, USA, 14–18 May 2012; pp. 5018–5023.

35. Isard, M.; MacCormick, J. BraMBLe: A Bayesian multiple-blob tracker. In Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 2001), Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 34–41.

36. Zhao, T.; Nevatia, R. Tracking Multiple Humans in Crowded Environment. In Proceedings of the IEEE Conference on Computer Vision and pattern recognition (CVPR 2004), Washington DC, USA, 27 June–2 July 2004; pp. 406–413.

37. Isard, M.; Blake, A. Contour tracking by stochastic propagation of conditional density. In Proceedings of the European Conference on Computer Vision ECCV, Freiburg, Germany, 2–6 June 1996; pp. 343–356.

38. Maccormick, J.; Blake, A. A Probabilistic Exclusion Principle for Tracking Multiple Objects. *Int. J. Comput. Vis.* **2000**, *39*, 57–71.

39. Branson, K.; Belongie, S. Tracking Multiple Mouse Contours (without Too Many Samples). In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 1039–1046.

40. Rasmussen, C.; Hager, G.D. Probabilistic data association methods for tracking complex visual objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 560–576.

41. Sanchez, O.; Dibos, F. Displacement Following of Hidden Objects in a Video Sequence. *Int. J. Comput. Vis.* **2004**, *57*, 91–105.

42. Sigal, L.; Bhatia, S.; Roth, S.; Black, M.J.; Isard, M. Tracking loose-limbed people. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington DC, USA, 27 June–2July 2004; IEEE: Washington DC, USA, 2004; Volume 1, pp. 421–428.

43. Khan, Z.; Balch, T.; Dellaert, F. MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1960–1972.

44. Perdomo, D.; Alonso, J.B.; Travieso, C.M.; Ferrer, M.A. Automatic scene calibration for detecting and tracking people using a single camera. *Eng. Appl. Artif. Intell.* **2013**, *26*, 924–935.

45. Barron, J.L.; Fleet, D.J.; Beauchemin, S.S.; Burkitt, T.A. Performance of optical flow techniques. *Int. J. Comput. Vis.* **1994**, *12*, 236–242.

46. Ranchin, F.; Dibos, F. Moving Objects Segmentation Using Optical Flow Estimation. In Proceedings of the Workshop on Mathematics and Image Analysis, Paris, France, 6–9 September 2004; pp. 1–18.

47. Spath, H. *The Cluster Dissection and Analysis Theory FORTRAN Programs Examples*. Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1985.

48. Goncalves, W.N.; Monteiro, J.B.O.; de Andrade Silva, J.; Machado, B.B.; Pistori, H.; Odakura, V. Multiple Mice Tracking using a Combination of Particle Filter and K-Means. In Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007), Belo Horizonte, Brazil, 7–10 October 2007; pp. 173–178.

49. Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Physica D* **1988**, *31*, 277–283.

50. Katz, M.J. Fractals and the analysis of waveforms. *Comput. Biol. Med.* **1988**, *18*, 145–156.

51. Castiglioni, P. What is wrong in Katz's method? Comments on: "A note on fractal dimensions of biomedical waveforms". *Comput. Biol. Med.* **2010**, *40*, 950–952.

52. Tsonis, A. Reconstructing dynamics from observables: The issue of the delay parameter revisited. *Int. J. Bifurc. Chaos Appl. Sci. Eng.* **2007**, *17*, 4229–4243.

53. Mandelbrot, B. How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension Abstract. *Science* **1965**, *156*, 636–638.

54. Esteller, R.; Member, S.; Vachtsevanos, G.; Member, S.; Echauz, J.; Litt, B. A Comparison of Waveform Fractal Dimension Algorithms. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2001**, *48*, 177–183.

55. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.

56. Shaw, R. Strange Attractors, Chaotic Behavior, and Information Flow. *Zeitschrift Für Naturforschung A* **1981**, *36*, 80−112.

57. Takens, F. Invariants related to dimension and entropy. In Proceedings of the 13th Coloquio Brasileiro de Matematica; Instituto de Matematica Pura e Aplicada: Rio de Janeiro, Brazil, 1983.

58. Grassberger, P.; Procaccia, I. Measuring the strangeness of strange attractors. In *The Theory of Chaotic Attractors*; Hunt, B.R., Li, T.-Y., Kennedy, J.A., Nusse, H.E., Eds; Springer: New York, NY, USA, 2004; pp. 170–189.

59. Grassberger, P.; Procaccia, I. Estimation of the Kolmogrorov entropy from a chaotic signal. *Phys. Rev. A* **1983**, *28*, 2591–2593.

60. Eckmann, J.P.; Ruelle, D. Ergodic theory of chaops and strange attractors. *Rev. Mod. Phys.* **1985**, *57*, 617–656.

61. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2288–2297.

62. Bandt, C. Ordinal time series analysis. *Ecol. Model.* **2005**, *182*, 229–238.

63. Food and Agriculture Organization (FAO). *Report of the Joint FAO/WHO Expert Consultation on the Risks and Benefits of Fish Consumption*; FAO Fisheries and Aquaculture Report No. 978; FAO: Rome, Italy, 2010; Volume 978.

64. Raghavendra, B.S.; Dutt, D.N. A note on fractal dimensions of biomedical waveforms. *Comput. Biol. Med.* **2009**, *39*, 1006–1012.

65. Fuss, F.K. A robust algorithm for optimisation and customisation of fractal dimensions of time series modified by nonlinearly scaling their time derivatives: Mathematical theory and practical applications. *Comput. Math. Methods Med.* **2013**, *2013*, 178476.