NSERC USRA
Automatic Identification of Soniferous Behaviour

Tabitha Lee
under the supervision of Alexandra Branzan Albu

University of Victoria
August 2019

Contents

1 Introduction

A particular behaviour event that is of interest to biologists manually processing underwater data from Ocean Networks Canada (ONC) is the startle response of the sablefish. Remotely operated vehicles to observe soniferous fish behaviour, have been shown to disturb fish, as in [18]. A tool for the automatic detection of this behaviour pattern would help correlate the sound production associated with this event, as well as determine whether the startle event was due to visual or sonic cues. In [4], the startle response is a measure of reflex impairment and fish stress. In [1], the latency and contraction time in a startle response are measured to gauge the reactions and sound biology of herring.

Although an algorithm has not yet been developed for the automatic detection of startle responses, what follows are some methodologies that have been applied to the detection problem, as well as the results of the methodologies.

The startle response, also known as C-start and fast start, in teleost fish is well documented by [6]. The behaviour consists of:

- 'fast-body-bend'- consisting of a stereotyped displacement of the head and tail to one side
- 'return flip' - a non-stereotyped flip of the tail to the opposite side"

In [8], a fast start escape response is described, which allows fish to avoid an attacking predator. The startle response is a C-shape bend away from the stimulus, followed by a quick stroke of the tail in the opposite direction. [14] describes the startle motion as a "unique locomotive pattern characterized by high acceleration rates", an aperiodic movement comprising of the large amplitude tail-beat following a startle stimulus.

Although model-based approaches are effective and often employed in the context of anomalous motion detection, they require a great deal of data to train and also to test on. Therefore, we will mostly be focusing on online, sequential algorithms.

2 Requirements

The desired outcome is to detect the startle response in samples of sablefish. By detection, we wish to know that a startle motion has occured in the video sequence. Generally speaking, it is marked by a sudden change in motion. This includes the multiple frames spanning a sudden change in direction from stillness or a quick change in direction. As stated in [6], the initial frame is defined by the fast body bend. The end frame of the startle motion is taken to be the last frame of the motion characterized by a return flip[6]. The frames between the initial frame and the end frame will be considered the detected startle response.

Since we will measure the startle response in a length of frames, the tolerance for detection should also be in frames. According to [14], the length of startle response is 60 ms to 150 ms depending on the size of the fish. [19] states that an adult female sablefish can be up to 110 cm long, which means it is a large fish. We are also working with videos that are 15 fps, which means there is approximately 67 ms between each frame. Assuming a large fish with startle duration of 150ms, as well as allowing for some error, the tolerance should be no more than 3 frames before the start frame or 3 frames after the end frame.

3 Dataset

We use Ocean Networks Canada (ONC) videos. Our specific dataset has been given to us but are also on SeaTube. The videos were recorded in Barkley Axis in the Barkley Canyon in the North Pacific Ocean, spanning December 9, 2010 to July 14, 2011. The details of the videos in the dataset are in table 1.

Resolution	640 x 480
Aspect Ratio	1.333
Format	H264
Bitrate	298 kbps
Frame Rate	15 fps

Table 1: Video Data Information

The videos have been annotated by Guan Lu at ONC as containing one or more startle responses, and when in the video they occurred. Originally the videos were downloaded in .mp4 format. However, to ensure MATLAB compatibility on Linux, they have been converted by ffmpeg into the .ogg extension. It was not noted until the time of writing that .ogg is not a lossless format, so it may be more beneficial to work with the .mp4 files directly. The videos have also been further trimmed to approximately the length of the startling behaviour in the video (usually around 5-15 seconds). The naming convention of the trimmed videos keeps the original name of the videos, but have been appended with _1, _2, _3, etc. For example, if the video is the first startle in the video, the name has been appended with _1; _2 with the second trimmed startle in the video; _3 for the third, and so on. The videos do not contain any colour but they are still converted to black and white for ease of processing.

After trimming the ONC videos, the videos were sorted into high, medium, and poor quality videos. The videos were deemed "poor quality" if:

- Choppiness in transition from frame to frame, especially during fast movements like startle motion
- blurry frames
- Startle is only partially visible; most of it occurs out of frame
- startling behind an object
- Startle with very small, barely perceptible movements

Similarly, a video was deemed "medium quality" if:

- lots of marine snow
- barely perceptible startle
- startle motion occluded by other fish
- startling motion mostly out of frame
- startling motion blocks out entire camera frame

Only the videos deemed high quality were used with the algorithm described in figure 9.

Video Quality	Number
Good	29
Medium	6
Poor	11
Total Videos	46

Table 2: Number of Good, Medium, and Poor quality videos

The good videos were further sorted into Single Fish, Multiple Fish, Single Startle, and Multiple Startle. The numbers of each are as in figure 1. Furthermore, we will only look at the videos which contain a sablefish startlemotion, as sablefish motion is much more exaggerated than in the tiny eelpout. This reduces the number of good videos from 29 to 27.

Single Fish Single Startle (6)	Single Fish Multiple Startle (0)
Multiple Fish Single Startle (10)	Single Fish Single Startle (11)

Figure 1: Number of Fish and Number of Startles

3.1 Initial Qualitative Observations

It is noted that in many videos, there is a strong component of horizontal movement during startle motion, as in figure 2. Often after the startling motion, there is a plume of sand disturbance on the ocean floor near the location of sablefish disturbance (figure 3). Sometimes the startle response is exihibited by a single fish and in other occurrences it triggers startle responses in the sablefish nearby, as seen in figure 4.

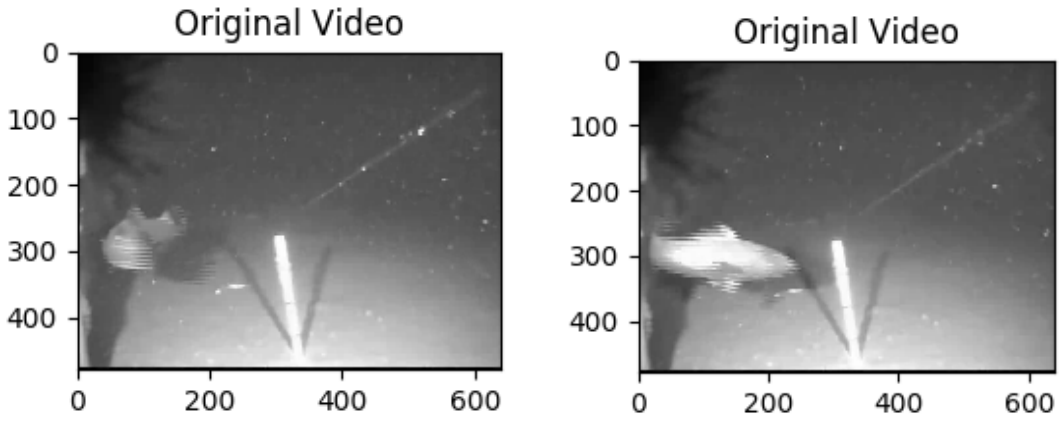

Figure 2: Fish movement has a large horizontal component during startle motion

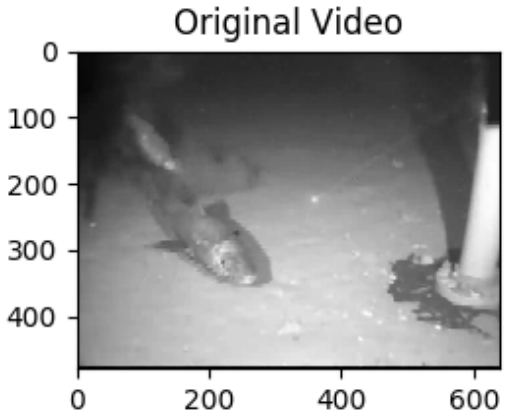

Figure 3: Plume of sand visible near the tail of the fish after startle

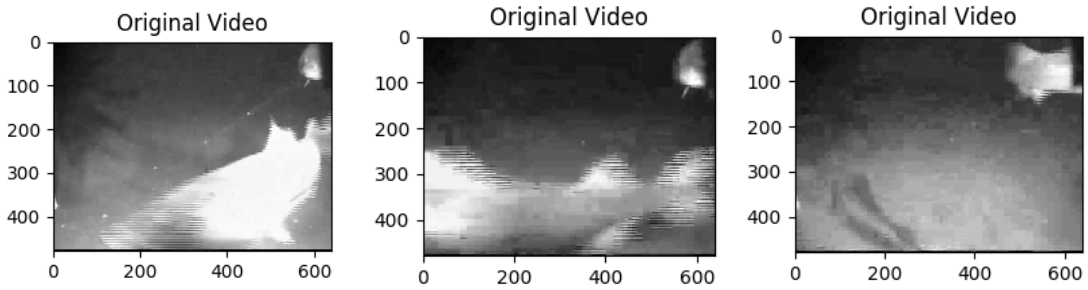

Figure 4: The startle of the fish in the foreground causes the fish at the top right hand corner to startle as well.

4 Attempted Approaches

4.1 Challenges

A large challenge in this project was the presence of marine snow in the foreground which were large enough to cause significant problems in any segmentation applications. The sablefish are often the same intensity as the background, making algorithms involving corner detectors difficult as more salient objects are picked up first. Finally, it is rarely the case in which there is a single fish in the frame with a startle motion. Detections based on changes of motion would be mistaken by other moving fish that are not exhibiting a startle motion. The extra startles present further challenges as it is much easier to reduce the effects of noise if the algorithm was only searching for one single startle response. In the current implementation, large noisy events could potentially be a startle response.

In figure 5 and figure 6 below, we can compare an ideal sequence to a non-ideal sequence. In an ideal sequence: there are many frames capturing the startle motion, the fish is clearly captured by optical flow, and the outline is quite clearly a fish. The presence of sea snow is very limited and the startle motion is quite apparent in the geometry of the fish. However, it would be more ideal if the fish startled more in-frame. In the non-ideal sequence, there are only a couple of frames to analyze, the fish is quite small and not as clearly defined. It is of similar size and shape as the sea snow also picked up by dense optical flow.

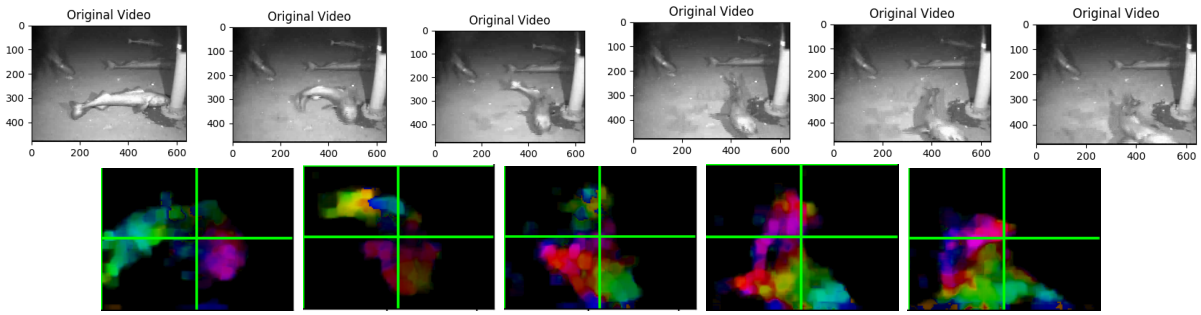

Figure 5: An ideal sequence (top) and dense optical flow (bottom) during startle motion. The first frame is the characteristic "C-shape" of startle motion and the transition of movement from frame to frame is smooth.

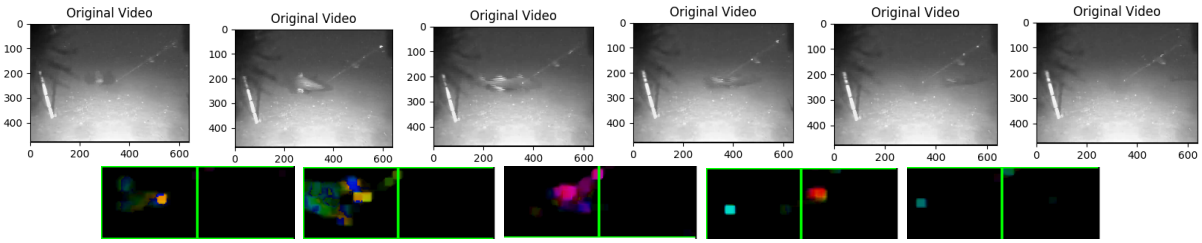

Figure 6: A non-ideal sequence (top) and dense optical flow (bottom) during startle motion. The characteristic "C-shape" is not clearly visible, the transition between frames is not smooth, and marine snow has been picked up by optical flow

4.2 Segmentation-based Approach

A segmentation-based approach was the initial attempt at the automatic detection of startle behaviour. We first detect the fish in the frame by segmentation, then track the centroid of the detected area. A Kalman filter will be used to reduce noise in movements. The centroid then becomes the sparse trajectory to be monitored. If there is a significant change in the trajectory of the fish, then a startle motion would be detected. This can be seen in figure 7.

This method did not work largely because the segmentation of fish and background. Due to the similarity of texture of background objects and fish, segmentation would often incorrectly segment the frame. For example, the algorithm would produce three fish segments when there was only one, as in figure 8, which cause the calculated trajectory of the fish to fluctuate greatly. Even if the fish detected in following frames were correct, errors in segmentation from frame to frame would cause the centroid of the frame to change much faster, and any large fluctuations in centroid position would result in an incorrect detection of startle motion.

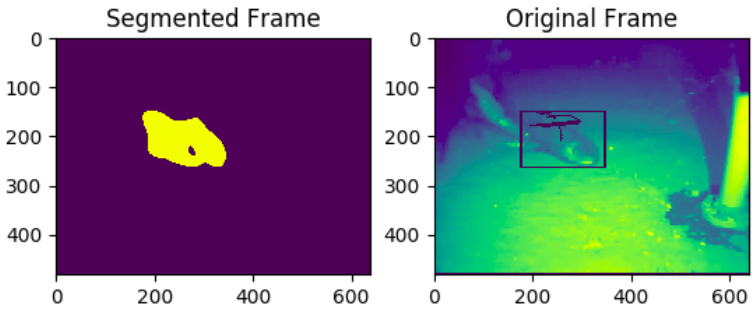

Figure 7: A segmentation-based approach and some trajectory tracking

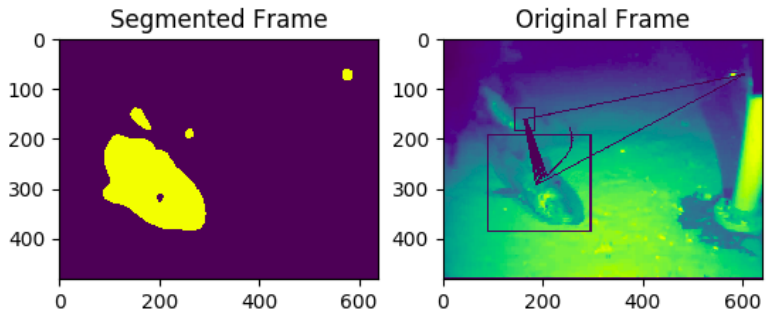

Figure 8: An incorrectly segmented image. The algorithm has placed 3 bounding boxes when there is only one fish. There are 3 blobs in the segmented image, resulting in incorrect trajectory tracking.

4.3 Dense Video Volumes

The approach in [12] presents an approach for online learning of spatially and temporally dominant and anomalous behaviours. In the context of our problem, the high acceleration startle motion of the fish would be anomalous behaviour compared to the other movements in the video: the marine snow drifting with the current in the same direction, the other mostly stationary fish, and the seaweed swaying back and forth. The algorithm is as follows: spatio-temporal volumes are densely sampled, gradients taken by the Sobel 5x5 3D Operator [10], and converted into Histogram of Oriented Gradients (HOG) descriptors [3]. These descriptors are furthered clustered into low-level codebooks which are compared with spatial-temporal ensembles of volumes for rare spatial or temporal behaviours. This approach is promising as there is no need for segmentation, and that startle motion appears to be a spatially and temporally anomalous event.

However, the naive approach of simply calculating dense video volumes requires a great deal of processing power - getting the gradients for each video volume for a single video of 463.7 kB and 268 frames, (each volume being 5 pixels x 5 pixels x 5 frames, overlapping) took longer than a day to only densely sample volumes and calculate gradients.After some optimization we decided that a different approach was necessary to detect startle motion. Nevertheless, the idea of densely sampling spatially and across time frames to capture motion has been incorporated into the adopted approach.

5 Main approach

5.1 Algorithm

We determined a much less processing-intensive approach was needed. The following approach is optical-flow based, which captures temporal motion information similar to the video volumes approach. However, the benefit of this approach is the built-in OpenCV method [15] for calculating dense optical flow based on Farneback's algorithm [7]. As well, not needing to capture the pixel information of each frame multiple times and not having to convolute each pixel with the 3D Sobel Operator as in [12] to find a gradient greatly reduces processing time.

The alogrithm can be visualized by the flowchart in figure 9.

5.1.1 Frame Pre-processing and filtering

There is very minimal pre-processing and filtering of each frame. Each frame is simply converted to grayscale, after which the optical flow is used to calculate between frames

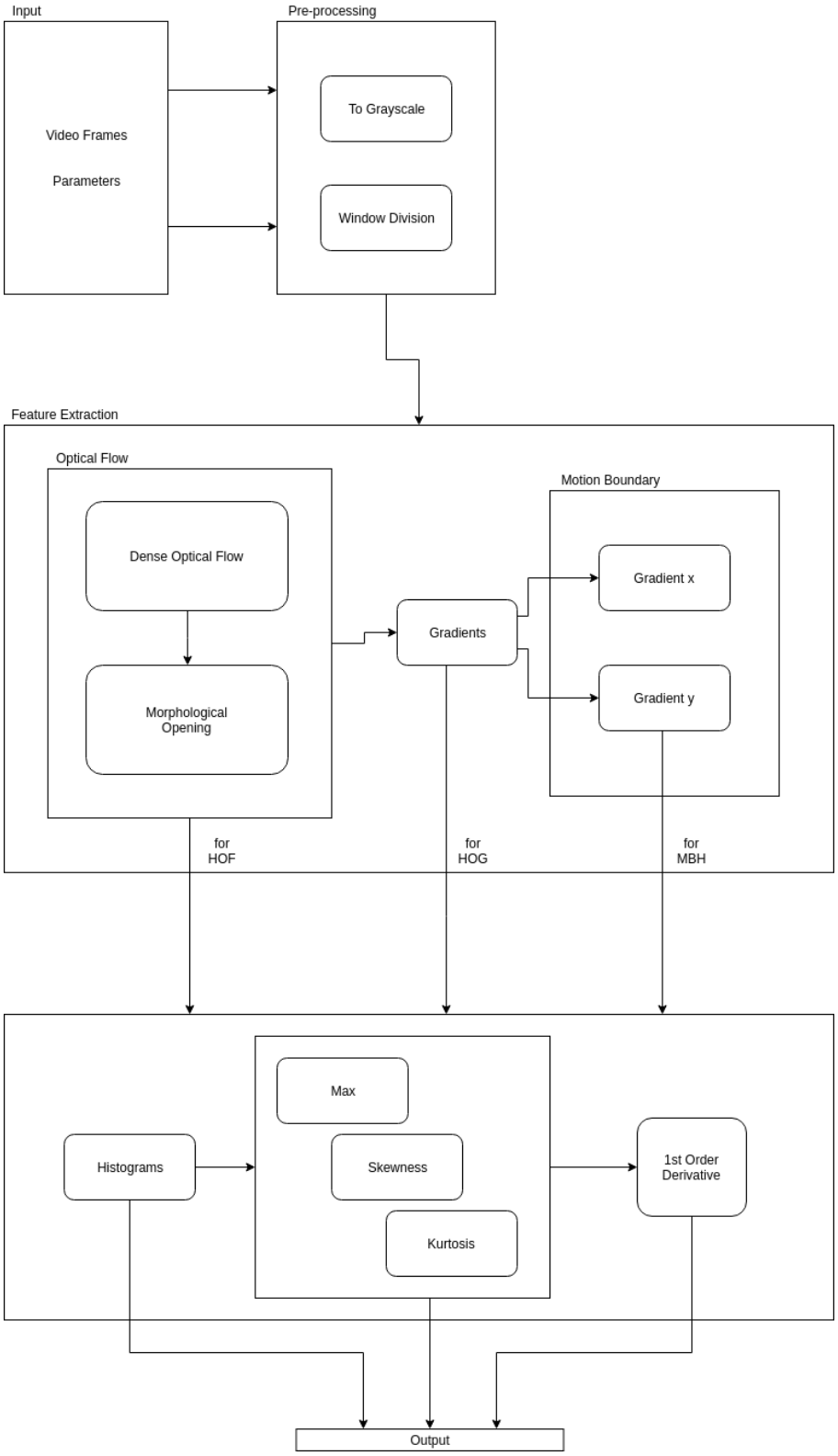

Figure 9: Flowchart for the Main Approach

5.1.2 Optical Flow

Optical Flow, according to [17], can be described as the pattern of an object moving between frames. It is calculated and represented as a 2-D vector field. Each vector is the movement of points between two consecutive frames. Its two main assumptions are:

1. The pixel intensities of the object do not change (much) between consecutive frames
2. Neighbouring pixels have similar motion

This yields the following equation:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{1}$$

Which can then be approximated via Taylor Series approximation and simplified to the following equation:

$$\frac{\partial f}{\partial x}u + \frac{\partial f}{\partial y}v + \frac{\partial f}{\partial t} = 0 \tag{2}$$

Where I is intensity, f is frame and

$$u = \frac{dx}{dt} \tag{3}$$

$$v = \frac{dy}{dt} \tag{4}$$

5.1.3 Lucas-Kanade Algorithm

There exists different algorithms for calculating Optical Flow. Two were examined in our algorithm. The first algorithm is the Lucas-Kanade algorithm [11]. The Lucas-Kanade algorithm in OpenCV is an implementation of [2]. This is a sparse algorithm which requires feeding a set of points to it in order to track motion. In order to obtain the set of points to track, each frame is fed through a corner detector. A Harris Corner detector was initially used, and can be seen in figure 10. As mentioned before, due to the texture and brightness of the fish, it is not naturally very salient so the Harris Corner detector detects points on other background objects in the frame such as a pipe or moving seaweed. Thus, it cannot track the movement of fish before or during a startle, so it is inept at detecting startle motion. The other corner detection method was that used by the OpenCV function cornerMinEigenVal, and can be seen in figure 11. The algorithm calculates the covariation matrix of derivatives, and the derivatives are computed using the Sobel operator.

$$M = \begin{bmatrix} \Sigma(\frac{dI}{dx})^2 & \Sigma\frac{dI}{dx}\frac{dI}{dy} \\ \Sigma\frac{dI}{dx}\frac{dI}{dy} & \Sigma(\frac{dI}{dy})^2 \end{bmatrix} \tag{5}$$

Then, it finds the eigenvectors λ_1, λ_2 for all the pixels and only stores the minimum, $\min(\lambda_1, \lambda_2)$ for corner detection. More about this algorithm can be found at [16]. However, the corner algorithm is still not strong enough to pick up any fish that may be moving in the background and so would not work for many of the training videos provided. Furthermore, both algorithms for corner detection seem to suffer greatly from frames that do not transition smoothly from one to another. Should this algorithm be attempted in the future, some further pre-processing would be largely beneficial.

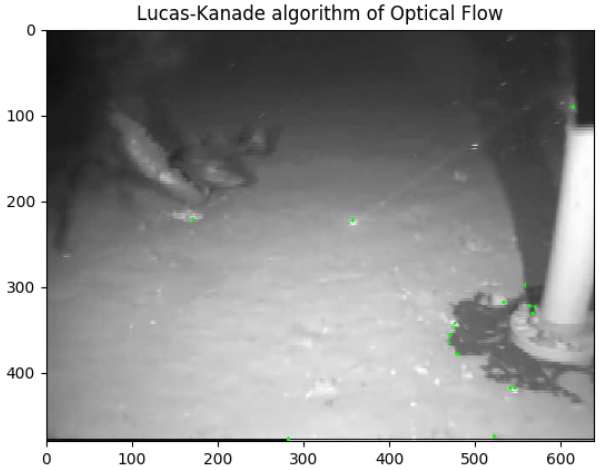

Figure 10: Lucas-Kanade algorithm of optical flow with Harris Corner Detection. Tracked points are in green. No points on the sablefish are tracked during the startle event.

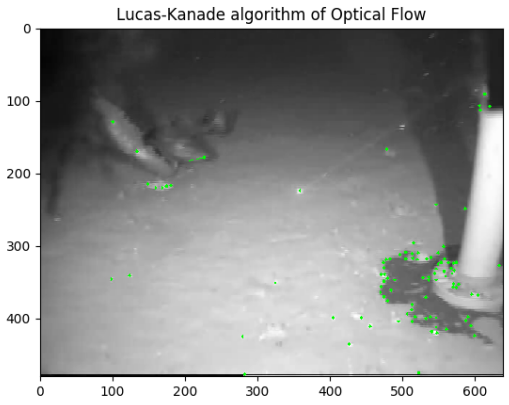

Figure 11: Lucas-Kanade algorithm of optical flow with minimal eigenvalue of gradient matrices for corner detection. Tracked points are in green. Some movement is captured from the fish during startle.

5.1.4 Gunnar Farneback Algorithm

The second algorithm examined and adopted is the Gunnar Farneback Algorithm [7]. It calculates dense optical flow between two frames, as in figure 12. The main advantage over Lucas-Kanade is that no points have to be detected, and the entire frame (or subsections of it) can be fed into the algorithm and the resulting movement captured in the 2-D array of flow vectors. This is much more effective than pre-detecting (often incorrectly) which points to track. More importantly, almost all moving objects are detected on the movement map, as shown below. However, the largest detriment to this algorithm is the fact that it is quite sensitive, and even smaller objects, such as marine snow, appear on the movement map. This adds more objects that are tracked, as well as presenting additional challenges. These include occluding a smaller fish to track in the background and making it harder to detect when something is sea snow or fish when both are similarly sized.

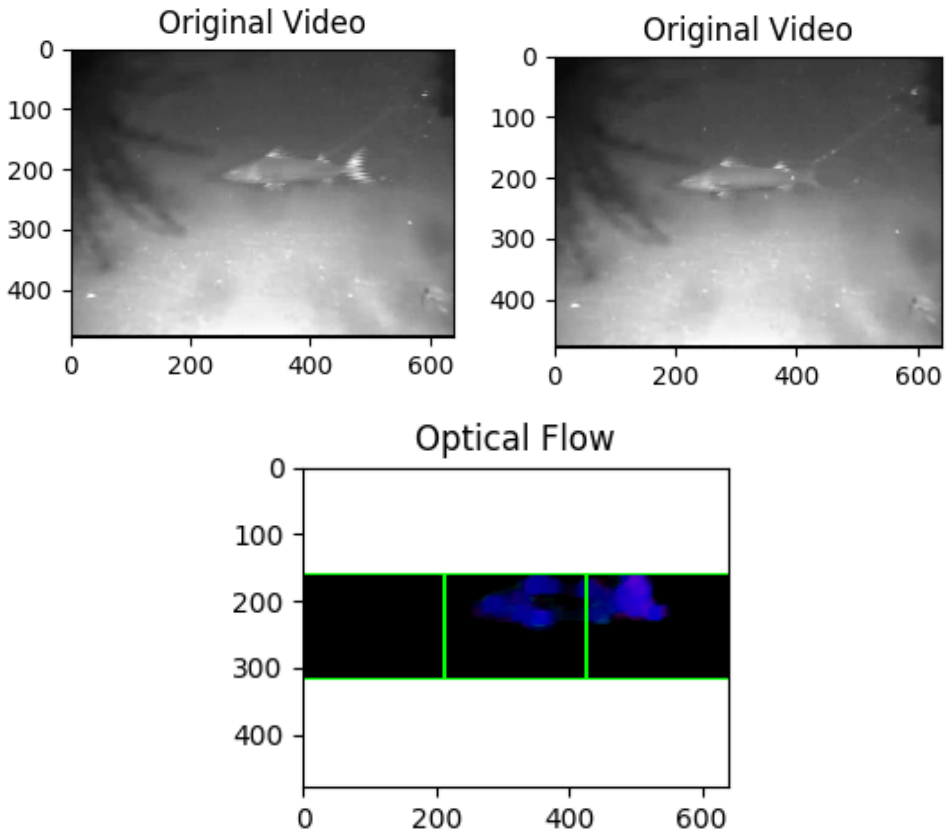

Figure 12: Two video frames required to calculate optical flow (top), and Gunnar-Farneback Dense Optical Flow for the two frames (bottom

5.1.5 Morphological Operations

The use of morphological operations after calculating dense optical flow is useful for filtering some effects of sea snow. More on the morphological operations can be found at [5].

The two operations used here are erosion followed by dilation. Erosion erodes the boundary of the foreground object through convolution, while dilation does the opposite and expands the boundary of the foreground object. Erosion followed by dilation is essentially "opening", which is very effective in removing small noise particles in the image. Unfortunately, the sea snow detected on the movement map is small but large enough that opening is not all that effective. In the future, in order to further reduce the effect of marine snow, a larger convolution kernel can be specified in OpenCV. However, the larger the convolution kernel, the longer the processing time. The dilation kernel of size 15 was chosen as it was the largest kernel size that did not seem to affect processing time too greatly, but as such, does not do much to filter out the effects of marine snow.

5.1.6 Sub-Dividing the Frames Into Smaller Windows

Looking at smaller, divided subsections of the frame has been greatly beneficial to the algorithms attempted. Generally, this is due to dampening the averaging effects of other moving elements when taking feature descriptor vectors. More specifically, the following reasons:

1. More accuracy in descriptors
2. Different optical flow calculated

More accuracy in descriptors by results in looking at only the windows containing the fish that will be startled. If we zoom in to only the windows containing the startle fish, then less marine snow will be in the window and therefore the effects of movement due to marine snow will be less accounted for when the descriptors are taken.

Essentially, a multiple fish single startle problem can be converted to a single fish single startle problem which makes automatic startle detection much easier. This can be seen in figure 13. Tracking the motions of one fish is much easier than multiple, and even more so when there are the added complications of occlusions.

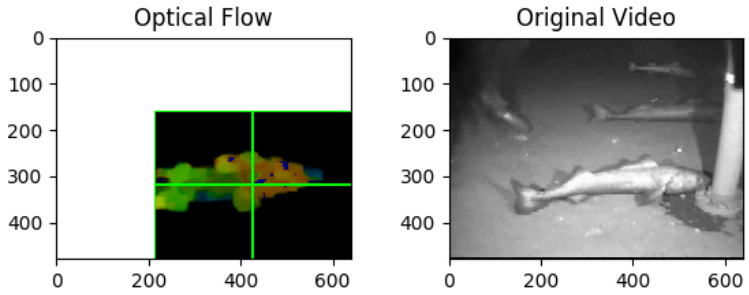

Figure 13: Smaller windows allows for conversion of a multiple fish problem to a single fish problem

A different optical flow will be calculated depending on how large the window size is, according to [21]. The spatial distributions of each object in each window versus the spatial distribution of the total frame of the video will be calculated differently, resulting in more localized and less averaged flow vectors.

5.1.7 Histogram-based Descriptors

The descriptors described below are all histograms. Motion and feature descriptors of this sort are commonly used, and are a nice way of compressing video information into a format that can be easily and usefully be compared against one another.

We will discuss Histogram of Optical Flow (HOF), Histogram of Oriented Gradient (HOG) [3], and Motion Boundary Histograms (MBH)[20]. Although HOG and MBH are implemented, only HOF is currently being used in the algorithm.

All vectors generated from the above algorithms, which are in Cartesian form, will be converted to polar form and binned into histogram bins according to their angles, with the magnitudes acting as weights. This histogram is now the feature descriptor used for comparison.

In order to present accurate feature descriptors, the vectors are normalized based on the number of pixels the feature vector was taken over. For example, if the window size was 100 x 200 pixels, then all histogram values will be divided by the product of the two dimensions, 20000.

- **HOF** We use the equations as in Farneback optical flow above; each u, v is converted to polar form and binned into histograms. This method is useful for gathering temporal information like motion but is not quite as useful for gathering textural information of individual objects.
- **HOG** To get gradient vectors, we use the Sobel operator. This means convoluting the frame with

the standard 3x3 Sobel kernel. For gradients in the x direction, G_x, and y direction, G_y, we use:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \tag{6}$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I \tag{7}$$

The gradients are then converted to polar form and again binned into 16 histogram bins. Where HOF does not capture textural information of objects, HOG is able to capture this information by taking the gradient in pixel intensity at each point. The resulting gradient is a vector, which is then converted to polar form, like HOF, and binned into histograms. While HOG is useful in computing textural information, it cannot capture temporal motion information. Both HOF and HOG are in-built OpenCV methods.

- **MBH** MBH is calculated by taking the horizontal and vertical gradients of the optical flow. There are no new equations. The result of calculating u, v are then each convoluted with the 3x3 Sobel kernel to get the gradients of an optical flow image. This results in 2 sets of oriented gradients, one for G_X and one for G_y. Both are then binned into histograms of 16 bins and concatenated, making a descriptor of length 32.
  MBH combines elements of both HOF and HOG, making it a useful feature descriptor. This algorithm does not track the locally constant camera movements and only significant changes in the optical flow field.

In our implementation, the histograms are then summed across all the smaller sub-divided windows, giving a total histogram for each frame.

5.1.8 Skewness

The skewness of a random variable X is defined by:

$$Skewness = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] \tag{8}$$

The skewness is the third standardized moment and describes the directional asymmetry of a distribution. The more positive the skew, the more the distribution appears to be distorted towards the right hand side; the mass of the distribution is more concentrated on the right. Similarly, the more negative the skew, the more the distribution appears to be distorted towards the left hand side.

We take the skew of each total histogram in each frame and plot it. We also take the first order derivative of skew and we can see that the values span positive and negative values.

The initial hypothesis was that the skew would reflect the curvature of a fish as it quickly bends its body to move in the quick startle motion. This would create stronger movement vectors in that direction, and be reflected in the histograms. Therefore, a sharp inflection in the first order derivative of the skew, either in the positive or negative direction, should reflect a startle motion.

5.1.9 Kurtosis

The kurtosis of a random variable X is defined by:

$$Kurtosis = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] \tag{9}$$

The kurtosis is the fourth standardized moment, but unlike skewness it is not a directional measure. Rather, it is a measure of the outliers in a distribution. All values in the distribution with a value close to the mean would not contribute to the kurtosis but a value further away from the peak would add to the kurtosis.

Similarly to the skew, we take the kurtosis of each total histogram in each frame and plot it. We have opted not to take the first order derivative of kurtosis as it seems quite noisy. It fluctuated between positive and negative values quite frequently and overall did not seem a very useful metric.

The initial hypothesis for using kurtosis was that the turning of the fish during startle motion, as well as any accompanying plumes of dust, would result in motion vectors being picked up in all directions. This would result in the histogram seeming more uniform and therefore the kurtosis should go up during a startle event.

5.1.10 Parameters

There are three parameters which could change the behaviour of running this script. These are outlined below:

- **Number of Height and Width Divisions**
  The number of height and width divisions control the granularity of window size. In general, the larger the number of divisions in each direction, the smaller the window and the better the detection as the frames can be selected to contain as much fish and as little marine snow as possible. However, without an algorithm for optimally dividing the frames into smaller windows, it was determined that a window size of 3 would produce good results.

- **Sampling Interval**
  The sampling interval between frames has an averaging effect. The frames between the sampling interval are not histogrammed, and therefore skew and kurtosis are not calculated for these frames. This reduces the effects of quickly changing values in skewness and kurtosis and acts as a sort of low-pass filter. It also allows the code to accomodate for more temporal information. However, the larger the sampling interval, the less information is being processed and will affect the accuracy of the results.

- **Number of Histogram Bins**
  The number of histogram bins simply affects the size of the feature descriptor vector obtained from the histograms. Since the result of skewness and kurtosis are calculated from the vectors themselves, the number of histogram bins do not affect the calculation of skewness and kurtosis. If there are too many histogram bins, then the feature vectors will be very large and the algorithm will be slower to run. If the number of histogram bins is too small, then the information in the vectors will not be represented sufficiently. For our purposes, it was determined that a histogram bin size of 16 was sufficient.

5.2 Results

5.2.1 Setup, Language and Frameworks

All work was performed on a PC operating with the following specifications:

Operating System	Linux Mint 18.3 Cinnamon 64-bit
Cinnamon Version	3.6.7
Processor	Intel Core i7-3770 CPU @ 3.40Ghz x 4
Graphics Card	NVIDIA Corporation GF116 [GeForce GTX 550 Ti]

Table 3: System Info

The approaches developed in this report were written in Python 3.5.2 with the OpenCV 2.4.1.0 framework. Another framework used is numpy 1.16.4. Both OpenCV and numpy should be installed in order to run any of the code in the repository.
The following tables are the result of running the video frames through the algorithm with the parameters below.

- Dividing the height and width of the frame into 3 width-wise and height-wise
- 16 histogram bins
- Sampling both every 3 frames and every 5 frames

The skewness, kurtosis, and their respective first derivatives are plotted against frame index in chronological order. An example of this is in figure 14. After, we analyze the properties of these plots.

The data in the following tables have been collected in the following ways:

- **Startle Offset** The frame number in which the startle motion begins have been manually annotated. They are based on the time of the original ONC annotations but are the first frame where the characteristic "C-shape" of the fish is observed

  After obtaining the histograms, the frame at which the global maximum or minimum (which ever is higher in magnitude) of histogram moment occurs is compared to the annotated startle. The comparison is done by measuring the number of frames between the annotated startle frame and the extremum frame.

  Startle Offset = Annotated Startle Frame - Maximum or Minimum Frame

  This is useful as a rough idea of when startle motions occur relative to when there are significant changes in their statistical moments.

- **Ground Truth Startle Duration** This is used as a comparison to how long the peak duration detected is.

  It has been manually annotated. We use the same startle frame from startle offset and we define the end of the startle motion, the End Frame, as the frame where the tail of the fish has moved in the opposite direction as when it was in the "C-shape" position. This is used, again as a very rough metric, to see how much of the motion was actually captured by the algorithm and how fast the startle motion event happened.

  Ground Truth Startle Duration = End Frame - Annotated Startle Frame

- **Peak Duration** First the maximum and its index in the data are found. Then, we find the left and right end points.

  For the left endpoint we first initialize a count and move to one index left of the maximum, taking the differences between the values of these two indexes. We again move to the left, and do the same. We compare the two differences, and if they have the same sign, then we increment a counter and continue in the same manner.

  This is inspired by derivatives, where leading up to a maximum or minimum, the derivatives on either side have gradually decreasing or increasing values, but have the same sign.

  The same process is done to find the right endpoint, except that we move to the right instead of to the left. Once we have both the left and right endpoint, the difference between the two we use as peak duration.

5.2.2 Skewness

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	87	9	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-60	6	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	-18	12	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	-45	9	3
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	21	6	3
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	12	6	3
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	72	3	3
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-39	6	3
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-30	3	3
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	-9	9	3
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	12	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	30	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	-54	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-96	6	3
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	-39	3	3
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	18	9	3
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-3	3	3
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	18	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	6	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-27	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	-51	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	54	12	3
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-33	3	3
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-33	3	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	-3	6	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-39	3	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	12	6	3
Average	-8.778	6.111	3
Standard Deviation	41.693	2.562	0

Table 4: Skewness of histograms, sampling every 3 frames

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	-35	9	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-60	6	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	18	12	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	-47	9	5
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	34	6	4
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	23	6	5
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	-38	3	5
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-36	6	5
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-18	3	5
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	32	9	5
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	-4	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	68	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	20	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-92	6	5
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	0	3	5
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	16	9	5
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-11	3	5
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	14	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	10	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-26	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	55	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	55	12	5
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-43	3	5
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-100	3	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	12	6	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-32	3	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	11	6	5
Average	-6.444	6.111	4.963
Standard Deviation	42.085	2.562	0.192

Table 5: Skewness of histograms, sampling every 5 frames

5.2.3 Kurtosis

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	-36	9	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-60	6	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	-18	12	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	45	9	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	21	6	3
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	12	6	3
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	72	3	3
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	33	6	3
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-24	3	3
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	30	9	3
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	12	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	30	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	-54	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-96	6	3
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	102	3	2
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	12	9	3
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-18	3	3
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	18	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	24	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-24	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	-51	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	39	12	3
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-33	3	3
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-33	3	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	-3	6	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-39	3	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	12	6	3
Average	-1	6.111	2.926
Standard Deviation	43.434	2.562	0.267

Table 6: Kurtosis of histograms, sampling every 3 framess

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	-35	9	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-60	6	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	13	12	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	-47	9	5
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	34	6	4
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	23	6	5
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	-28	3	5
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	19	6	5
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-13	3	5
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	32	9	5
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	11	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	18	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	20	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-97	6	5
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	-40	3	5
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	11	9	5
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-1	3	5
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	14	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	20	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-26	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	55	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	35	12	5
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-38	3	5
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	75	3	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	2	6	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-37	3	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	11	6	5
Average	-1.074	6.111	4.963
Standard Deviation	37.854	2.562	0.192

Table 7: Kurtosis of histograms, sampling every 5 frames

5.2.4 First Derivative of Skewness

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	-38	9	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-59	6	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	-20	12	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	43	9	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	22	6	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	10	6	2
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	73	3	2
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-41	6	3
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-32	3	3
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	-8	9	2
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	10	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	28	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	-56	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-98	6	2
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	-41	3	2
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	19	9	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-5	3	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	19	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	22	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-29	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	-50	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	34	12	2
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-32	3	2
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-32	3	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	-5	6	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-41	3	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	13	6	2
Average	-10.889	6.111	2.222
Standard Deviation	38.034	2.562	0.424

Table 8: First derivative of skewness of histograms, sampling every 3 frames

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	-37	9	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-62	6	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	16	12	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	-49	9	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	27	6	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	21	6	2
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	-40	3	2
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-33	6	2
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	0	3	2
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	25	9	2
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	-1	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	61	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	23	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-89	6	2
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	8	3	2
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	14	9	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-13	3	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	12	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	8	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-28	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	-47	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	58	12	2
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-40	3	2
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-102	3	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	15	6	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-34	3	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	14	6	2
Average	-10.111	6.111	2
Standard Deviation	40.355	2.562	0

Table 9: First derivative of skewness of histograms, sampling every 5 frames

5.2.5 First Derivative of Kurtosis

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	-35	9	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-62	6	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	-20	12	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	43	9	3
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	22	6	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	10	6	2
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	70	3	2
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-41	6	3
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-26	3	2
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	31	9	2
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	13	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	31	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	-53	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-101	6	3
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	46	3	3
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	19	9	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-20	3	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	19	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	25	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-23	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	-50	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	40	12	2
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-32	3	2
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-35	3	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	-5	6	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-38	3	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	13	6	2
Average	-5.889	6.111	2.296
Standard Deviation	40.049	2.562	0.465

Table 10: First derivative of kurtosis of histograms, sampling every 3 frames

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	-32	9	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-62	6	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	11	12	2
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	-49	9	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	17	6	2
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	26	6	2
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	-45	3	2
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-38	6	2
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	0	3	2
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	35	9	2
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	14	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	26	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	23	6	2
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-99	6	2
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	-42	3	2
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	9	9	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	2	3	2
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	12	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	18	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-28	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	58	6	2
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	33	12	2
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-35	3	2
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	78	3	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	0	6	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-39	3	2
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	14	6	2
Average	-3.444	6.111	2
Standard Deviation	39.473	2.562	0

Table 11: First derivative of kurtosis of histograms, sampling every 5 frames

5.2.6 Max

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	45	9	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-39	6	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	-3	12	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	21	9	3
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	-15	6	3
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	0	6	3
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	-57	3	3
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-108	6	3
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-6	3	3
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	96	9	3
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	-45	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	57	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	81	6	3
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	24	6	3
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	57	3	3
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	15	9	3
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	48	3	3
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	18	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	12	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-3	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	-15	6	3
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	48	12	3
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-39	3	3
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-90	3	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	18	6	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	6	3	3
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	12	6	3
Average	5.111	6.111	3
Standard Deviation	47.637	2.562	0

Table 12: Max of histograms, sampling every 3 frames

Video	Startle Offset (Frames)	Ground Truth Startle Duration (Frames)	Peak Duration (Frames)
BC_POD1_PTILTVIDEO_20110519T091755.000Z_1	100	9	3
BC_POD1_PTILTVIDEO_20110522T114342.000Z_1	-60	6	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_2	-2	12	5
BC_POD1_PTILTVIDEO_20110522T114342.000Z_3	-72	9	5
BC_POD1_PTILTVIDEO_20110522T173147.000Z_1	4	6	5
BC_POD1_PTILTVIDEO_20110522T173147.000Z_2	-2	6	5
BC_POD1_PTILTVIDEO_20110525T111402.000Z_1	-53	3	5
BC_POD1_PTILTVIDEO_20110527T063643.000Z_1	-56	6	5
BC_POD1_PTILTVIDEO_20110528T073304.000Z_1	-3	3	5
BC_POD1_PTILTVIDEO_20110528T165204.000Z_2	47	9	5
BC_POD1_PTILTVIDEO_20110528T165204.000Z_3	11	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_1	58	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_2	-40	6	5
BC_POD1_PTILTVIDEO_20110615T192950.000Z_3	-87	6	5
BC_POD1_PTILTVIDEO_20110616T171904.000Z_1	60	3	5
BC_POD1_PTILTVIDEO_20110616T171904.000Z_2	16	9	5
BC_POD1_PTILTVIDEO_20110618T185440.000Z_1	-6	3	5
BC_POD1_PTILTVIDEO_20110618T185440.000Z_5	9	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_1	10	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_2	-21	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_3	10	6	5
BC_POD1_PTILTVIDEO_20110624T174459.000Z_4	20	12	5
BC_POD1_PTILTVIDEO_20110625T183602.000Z_1	-38	3	5
BC_POD1_PTILTVIDEO_20110702T181655.000Z_1	-100	3	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_1	22	6	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_2	-37	3	5
BC_POD1_PTILTVIDEO_20110703T190647.000Z_3	11	6	5
Average	-7.370	6.111	4.926
Standard Deviation	46.654	2.562	0.385

Table 13: Max of histograms, sampling every 5 frames

5.2.7 Discussion

Generally, the extrema present in skewness and kurtosis are maximums resembling spikes. Sometimes the extrema is a minimum, but the majority are maximums.

In theory, the skewness and kurtosis should increase in the event of a startle motion as the vectors should form a more uniform distribution. The vectors are more distributed away from the center of the distribution, hence the increase in kurtosis. However, skewness in a directional measurement, and this indicates that more often than not, the optical flow vectors are oriented towards the down direction of the frame during a spike in skewness or kurtosis. When minimums are present, the optical flow vectors are oriented in the the up direction of the frame.

However, it should be noted that observing peaks in skew, kurtosis, or their respective first order derivatives may not be a useful metric as the spikes in these moments occur many frames before or after the event. The fact that the spikes can occur before or after indicates that it is not necessarily predictive of a startle. We can see this in figure 14 below. We can also see this in the high standard deviation values in the data.

It should also be noted that there are often both local and global extrema in the skewness a kurtosis plots like those in figure 14. Our measurements are taken to the global maximum or minimum, but often there exists a closer, local maximum or minimum to the vicinity of the annotated startle motion. The problem with tracking both global and all the local maxima and minima is that any of them could possibly indicate a startle motion, even if there was only one actual startle motion in the video frames analyzed.

If there are many local maxima and minima in the plots of skewness and kurtosis, it would be too noisy to use the data. However, looking at the first derivative of these measures presents a much more discernible spike in the skewness and kurtosis, usually occuring at the same video frame as the global maximum. This helps identify the peaks to focus on. Although we note that the magnitude of the average offset of skewness is slightly less than that of its first derivative, the standard deviation values for the first derivative are smaller.

Examining the peak duration row of the tables of the skewness, kurtosis, and their first derivatives' measurements, we note that the current method of finding peak duration does not offer much information. We can see that most spike events occur for about the same number of frames as the sampling interval, as there are many noisy fluctuations within the data. However, the values for peak duration do not differ too greatly from the ground truth average of 6.111 frames. A more global approach would be worthwhile to look into for determining the duration of a spike event. It should be noted that in this set, the average duration for ground truth startle duration, at 15 fps, is around 410 ms, which is much larger than the initial estimate of 150 ms, so the tolerance should be set higher accordingly.

We can also see that the offset from extrema to annotated startle is negative. We can see that for this set, on average, the extrema event in skew and kurtosis happen 1-10 frames before the annotated start of the startle motion. Kurtosis appears to give the best results at an average of -1.074 for a sampling interval of 5, although the standard deviation is quite high at 37.854. For such a low average startle offset, kurtosis seems to be the most promising.

The difference between the sampling interval of 3 and 5 are apparent in the averages and standard deviations. For skewness and its first derivative, a sampling interval of 5 has a much smaller frame distance from extrema to annotated startle, implying higher accuracy. However, the sampling interval of 5 has a larger standard deviation which implies that the sampling interval of 3 is more accurate. With kurtosis and its first derivative, the sampling interval of 5 performed better than the sampling interval of 3. This is evident by the negligible difference in average value between the two sampling intervals, but the much lower standard deviation of 37.854 for a sampling interval of 5 than with the standard deviation of 43.434 for a sampling interval of 3. Kurtosis is a measurement of the number of outliers from the mean vector direction, and its statistical moment has an even degree number, which makes it less sensitive to the effect of more positive and negative values in the data with a smaller or larger sampling interval. Skew, however, is more sensitive to sign differences since it is an odd degree moment.

Finally, we examine the maximum. It is the most sensitive to this global maximum or minimum approach, which is evident in the drastic change in average from 5.111 for sampling interval of 3 and -7.370 for a sampling interval of 5. This is further confirmed by the standard deviation values of 46 or higher, likely due to the fact that the fish are moving with much stronger optical flow vectors after a startle event but have smaller movements before and after.

The average startle offset for the maxiumum is the only one which is positive after being averaged, meaning it is after the annotated startle. This makes the most sense as it doesn't really make sense to be detecting startle movements before the actual annotated startle . However, with a sampling interval

of 5 its average startle offset value is negative once again. The maximum fluctuates often and so benefits more from a smaller sampling interval as we do not want to average out an event we are looking for. It is again noted that it fluctuates quite a bit, as seen by the huge standard deviation, and hence may not be a useful measure.

The main disadvantages to our algorithm is that there is no memory of events in the video prior to the sampling interval number of frames before. This does not help look in searching for a specific sequence of events. The other disadvantage is that it is not real-time and requires processing the entire video before estimating when the startle motion occurred.

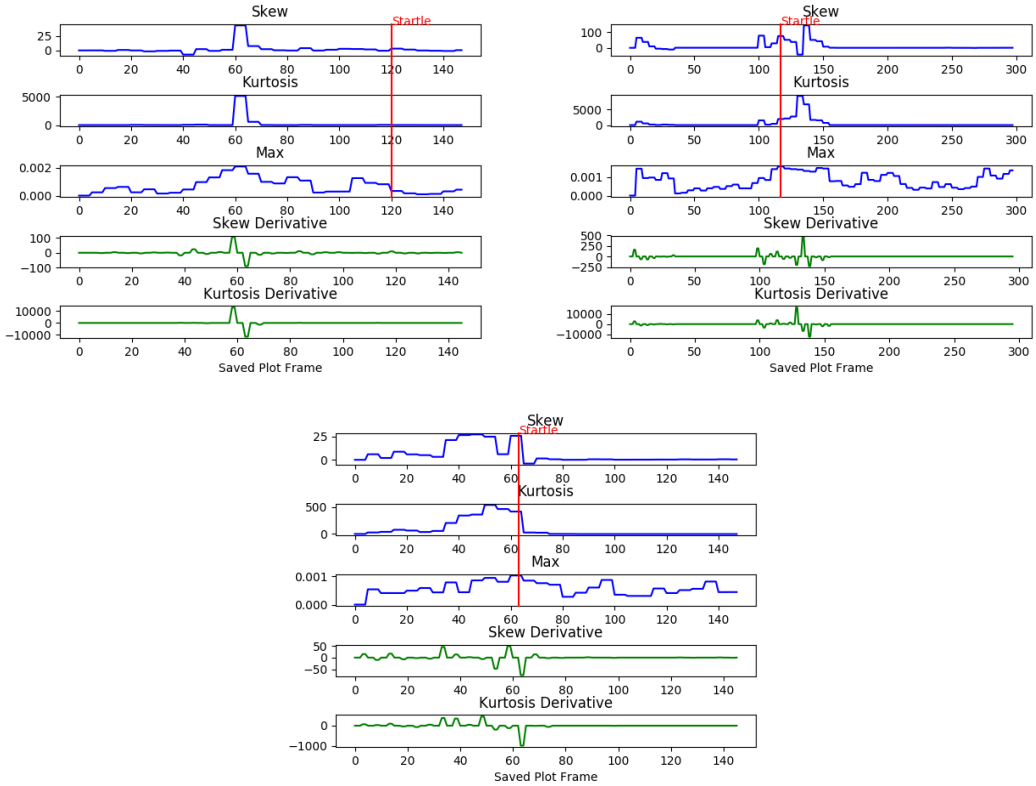

Figure 14: Clockwise from top left: Peak in statistical moment before, after, and at startle

6 Conclusion

6.1 Future Work

6.1.1 Automatic Window Detection

Currently, the windows to focus on are done manually, as in the user looks at a grid overlay on the video frame and selects the windows the fish will pass through in its motion, as in [12]. For proof of concept, this certainly gives better results than with the full frame. However, the manual selection of windows undermines the automatic detection of fish startle motion and should be removed in the future.

The problem of automatic detection of windows has been solved by methods such as individually calculating the optical flow of all windows in a frame and only using the windows with the summed movement above a specific threshold. Another would be to use the morphology of the object in the frame and only track the frame if the object is large enough. However, this presents problems as well, especially when objects span multiple frames or are of similar size to noise objects.

An alternative to the static division of each frame into a grid window would be a sliding window of a fixed size over each frame of the video. Again, the window selected for tracking motion would be optimized for most motion detected and overlapping windows would not be considered. Due to time constraints, automatic window detection was not attempted.

6.1.2 Correlation of Statistcal Data

If we look at any of the skew, kurtosis, first derivatives, and max, we can see that none of them particularly indicate a startle motion. However, one could look at a possible correlation of the extrema in all of the above to infer the location of the startle motion. In addition, if there was a an algorithm to correlate the local maxima / minima of statistical data instead of just using the global maxima over a broad period of time, the current approach would likely perform much better.

6.1.3 Normal and Anomalous Behaviour

Although the idea of normal and anomalous behaviour does not strictly apply to a fish startling, we are inspired by the methods, as in [12], applied to this idea. This is the most promising approach, but the problem is building and training the model of what is Normal Behaviour and what is Anomalous. Papers like [12] have gone into great detail for implementation. Assuming one had the time and computing resources to go about this method, it would be largely helpful to reduce the effects of normal behaviour. This could be very complex or as simple as just averaging, for example, the histograms, across the whole video instead of for just one frame. Then one could possibly compare other histograms to the average with tools like the Quadratic Chi Distance mentioned in [13]

6.1.4 Inclusion of Other Feature Descriptors

Histogram-based descriptors HOG and MBH are also calculated, but not used with the algorithm described. They can simply be swapped out for HOF or perhaps used together to infer startle motion frame location. Hu moments [9] would be another non-histogram-based descriptor worth looking into. It is image-moment-based and would be useful in finding the characteristic "C-shape" in the first frame of startle motion.

6.1.5 Re-examining Trajectories

Looking at trajectories again, but in the context of dense HOF or HOG vectors may be worth examining. One could exploit the mostly downward, and mostly horizontal movement the fish makes. If one were to keep track of the movement of all the pixels, the detection of a sudden downward but mostly horizontal movement may help with startle motion detection. The approach of dense trajectories is seen in [20].

6.2 Summary

In this study we have attempted to detect the startling motion of sablefish without the use of segmentation methods. Avoiding the use of segmentation methods allows us to work with conditions that may heavily contain marine snow, as well as with lower quality videos. Our approach uses histograms of optical flow as motion descriptors. These are then binned into histograms and analyzed by their statistical moments. In particular, we examine the third and fourth standaradized moments, the skewness and kurtosis. We have tested the results under 3 and 5 sampling intervals. Kurtosis appears to be the closest measure to detecting startle motion, with an average startle to maximum frame offset of -1 frames at sampling interval 3.

Although the results applying this algorithm were inconclusive, it offers some insight into possible areas to explore in the future.

References

[1] J.H.S. Blaxter and D.E. Hoss. "Startle Response in Herring: The Effect of Sound Stimulus Frequency, Size of Fish and Selective Interference with The Acousto-Lateralis System". In: *Journal of the Marine Biological Association of the United Kingdom* 61 (4 Nov. 1981), pp. 871–879. URL: https://www.cambridge.org/core/journals/journal-of-the-marine-biological-association-of-the-united-kingdom/article/startle-response-in-herring-the-effect-of-sound-stimulus-frequency-size-of-fish-and-selective-interference-with-the-acousticolateralis-system/3176F91D3ECFDEB498CB4C566F9460E6.

[2] J-Y. Bouguet. "Pyramidal Implementation of the Affine Lucas Kanade Feature Tracker Description of the algorithm". In: (2001). URL: https://pdfs.semanticscholar.org/aa97/2b40c0f8e20b07e02d1fd320bc7ebadfdfc7.pdf.

[3] N. Dalal and B. Triggs. “Histograms of Oriented Gradients for Human Detection”. In: *IEEE Xplore* (June 2005).

[4] M. Davis. “Fish stress and mortality can be predicted using reflex impairment”. In: *Fish and Fisheries* 11 (1 Mar. 2010), pp. 1–11.

[5] OpenCV 3.0.0-dev documentation. *Morphological Transformations*. Last accessed 21 August 2019. 2014. URL: https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_imgproc/py_morphological_ops/py_morphological_ops.html.

[6] R. Eaton, R. Bombardieri, and D. Meyer. “The Mauthner-Initiated Startle Response In Teleost Fish”. In: *Journal of Experimental Biology* 66 (1977), pp. 65–81. URL: https://jeb.biologists.org/content/jexbio/66/1/65.2.full.pdf.

[7] G. Farneback. “Two-frame motion estimation based on polynomial expansion”. In: *Lecture Notes in Computer Science* 2749 (2003), pp. 363–370.

[8] ME. Hale. “Startle responses of fish without Mauthner neurons: escape behavior of the lumpfish (Cyclopterus lumpus)”. In: *The Biological Bulletin* 2 (Oct. 2000), pp. 180–182. URL: https://www.journals.uchicago.edu/doi/pdfplus/10.2307/1542886.

[9] Z. Huang and J. Leng. “Analysis of Hu’s moment invariants on image scaling and rotation”. In: *IEEE Xplore* (June 2010). URL: https://ieeexplore.ieee.org/document/5485542?arnumber=5485542.

[10] H. B. Kekre and S. M. Gharge. “Image Segmentation using Extended Edge Operator for Mammographic Images”. In: *(IJCSE) International Journal on Computer Science and Engineering* 2 (2010).

[11] B. Lucas and T. Kanade. “An Iterative Image Registration Technique with an Application to Stereo Vision”. In: *Proceedings of Imaging Understanding Workshop* (1981), pp. 121–130. URL: http://cseweb.ucsd.edu/classes/sp02/cse252/lucaskanade81.pdf.

[12] M. Levine M. Roshtkhari. “Online Dominant and Anomalous Behavior Detection in Videos”. In: *IEEE Xplore* (2013). URL: https://pdfs.semanticscholar.org/bf14/a370f8676143456d648dbc20d3eeb5f693f6.pdf.

[13] O. Pele M. Werman. “The Quadratic-Chi Histogram Distance Family”. In: *ECCV* (July 2010). URL: http://www.cs.huji.ac.il/~werman/Papers/ECCV2010.pdf.

[14] Symposium on Mechanics and Physiology of Animal Swimming. *The Mechanics and Physiology of Animal Swimming*. Cambridge University Press, 1994.

[15] OpenCV. *calcOpticalFlowFarneback*. Last accessed 21 August 2019. 2014. URL: https://docs.opencv.org/3.0-beta/modules/video/doc/motion_analysis_and_object_tracking.html#calcopticalflowfarneback.

[16] OpenCV. *cornerMinEigenVal*. Last accessed 21 August 2019. 2019. URL: https://docs.opencv.org/3.0-beta/modules/imgproc/doc/feature_detection.html#cornermineigenval.

[17] OpenCV. *Optical Flow*. Last accessed 21 August 2019. 2019. URL: https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html.

[18] R. Rountree and F. Juanes. “First attempt to use a remotely operated vehicle to observe soniferous fish behavior in the Gulf of Maine, Western Atlantic Ocean”. In: *Current Zoology* 56 (2010), pp. 90–99.

[19] NOAA Fisheries Service. *Sablefish: Anoplopoma fimbria*. Last accessed 21 August 2019. 2010. URL: https://www.afsc.noaa.gov/Education/factsheets/10_Sablefish_fs.pdf.

[20] H. Wang et al. “Dense Trajectories and Motion Boundary Descriptors for Action Recognition”. In: *INRIA* (Jan. 2013). URL: https://hal.inria.fr/hal-00725627v2/document.

[21] S. Wang et al. “Anomaly detection in crowded scenes by SL-HOF descriptor and foreground classification”. In: *IEEE Xplore* (Dec. 2016). URL: https://ieeexplore.ieee.org/document/7900159/references#references.