Underwater fish-video images: Image quality and edge detection techniques

C.R. SAVAGE¹, R.J. PETRELL¹ and T.P. NEUFELD²

¹Bio-Resource Engineering Department, 2357 Main Mall, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; and ²Byssus Communications Ltd., P.O. Box 39179, Point Grey RPO, Vancouver, BC, Canada V6R 1G0. This paper was originally presented as ASAE Paper AQUA-92-107. Received 7 April 1993; accepted 31 May 1994.

INTRODUCTION

Fish health, condition, size, and number are important factors that have an impact on the economy of an aquaculture enterprise. Despite the economic impact, these factors are inconsistently monitored because the means of measurement are invasive, and/or difficult, costly, and time consuming to execute.

Considerable fish stress is associated with three conventional inventory methods: sample counts, total mass, and pilot-tank technique (Piper et al. 1986). The effect of stress on salmon is lower disease tolerance and increased mortalities (Strange and Schreck 1978; Schreck 1982; Maule et al. 1989). Another problem associated with the conventional methods is infrequent assessments. Frequent assessments are often not practical because these methods are coincident with operations such as harvesting and net changes (on a sea cage farm).

Most salmon sea cage farms update stock numbers by the difference method, which involves the daily or weekly collection of fish mortalities by a SCUBA (Self Contained Underwater Breathing Apparatus) equipped diver. Problems that limit the effectiveness of this procedure are: stress to fish; unaccounted losses by predators (i.e. herons, gulls, otters, and seals); escaped fish; and inaccurate stocking counts.

Problems in conventional inventory assessment have led to the development of several counting and sizing techniques. One class of techniques uses videotape equipment to record fish passing through a shallow or narrow tunnel (Spratt 1991). The videotaped fish are hand-counted from the video replay. Farmers have reported that the hand-counter method takes over 4 hours per sea cage. Spratt (1991) and Irvine et al. (1991) applied basic image processing techniques towards automating fish counts. The average computer automated count was 25% less than the actual number while the semi-automatic generated estimate varied 6.4% from the actual number of fish (Irvine et al. 1991).

Acoustic techniques have been viewed as potentially the most direct and effective means of estimating the abundance of fish in a net pen; however, cost and the inability to resolve problems relating to fish distribution and orientation limits use of hydroacoustic techniques to accurately determine fish biomass (Burcynski 1985).

Diseases and undesirable morphological changes in appearance should also be frequently monitored so prompt...
remedial action can be taken. These conditions are normally only detected when fish manifest gross symptoms or incur a high mortality rate. By this time the farmer has no choice but to accept fish losses. Some fish diseases (i.e. sea lice, _Lepeophtheirus sp._) and physical states (spinal curvature, fin necrosis, precocious maturation) can be visually detected noninvasively, provided a vision system is in place. The early detection that a vision system would provide could make the difference between financial loss or gain.

Natural-light videotaping with economical underwater video cameras can be used to noninvasively capture large quantities of fish images. If the images are of good quality they can be used to count, size, and detect visually obvious diseases and undesirable conditions of the fish. The quality of an image in the natural light of the underwater environment is affected by the amount and quality of sunlight, sun altitude, camera angle, light attenuation and scattering, and dome port distortion and aberrations (Frey and Tzimoulis 1968). Dome port distortion (typically radial distortion) and aberrations can be addressed through the use of good quality acrylics, polishing, and correction factors for distortion (Goshtasby 1989).

From an image processing perspective, a good quality image has a well distributed contrast range and well resolved features, i.e. the fish. Well-contrasted fish are characterized by uniform and large differences in gray-levels between the fish and its background. A well resolved fish has sharp and well-defined edges. A well-defined edge is a sharp gradient between two regions with relatively distinct gray-level properties (Gonzalez and Wintz 1987).

The purpose of this paper is to describe the quality of fish images as they relate to basic image processing techniques and to demonstrate the application of several edge detection algorithms to segment fish from their background. If a system can economically and quickly segment fish from the background, then the fish could be automatically counted, sized, or analyzed. This project is part of an ongoing research program aimed at the development of automated, noninvasive fish monitoring and inventory systems.

**MATERIALS AND METHODS**

The system hardware used to capture fish images included the Panasonic WVBD400 black and white camera with a sensitivity of 0.5 lux. The image sensor was a 17 mm interline CCD (Charge Coupled Device) with 768 (horizontal) x 493 (vertical) pixels. The camera was electronically shuttered to prevent blurring of moving targets in a video frame. A Cosmicar (Pentax Canada Inc., Vancouver, BC) 4.8 mm wide angle lens with a 95 degree diagonal field of view was used. With the extended depth of field of the lens, an object positioned 0.3 m to infinity from the camera was in focus. The camera was environmentally sealed by an underwater housing, which was made of an anodized aluminum case with a acrylic dome port and a PVC bulkhead (International Hard-suits, North Vancouver, BC). The dome port was corrected to the same magnification as the Cosmicar lens above water. The camera was connected to a 22.5 m umbilical cable which provided power, video signal, and Gen-lock (synchronization for stereo cameras). The dome port provided a 76° viewing angle in both the horizontal and vertical planes. The resulting viewing volume was a right regular pyramid. The height of the view volume, as measured perpendicular to the base of the pyramid, was considered to be the visibility (m) within the water column as defined by a Secchi disk reading.

A Panasonic AG-1960 S-VHS video cassette recorder accepted an RS170 video signal from the underwater camera unit (Fig. 1). The S-VHS VCR permitted approximately 400 lines of horizontal resolution thus allowing the recognition of finer detail and smaller fish.

Computing hardware consisted of an Intel 486DX 33 mHz CPU (8 MB RAM) with an 8k internal cache and a 64k external cache. Direct programming of DMA (Direct Memory Access) hardware allowed fast memory management.

The frame grabber card was an Imaging Technology OFG or Overlay Frame Grabber (Imaging Technology Incorporated, Bedford, MA). This card had four video inputs, an input LUT (Look Up Table), 16-triplet (red, green, and blue) output LUTs, and 1024 x 512 pixel frame memory. The input LUTs allowed the incoming pixel data to be transformed before they were stored in frame RAM. The transformation consisted of a programmable array of 256 gray-level values. The output LUTs allowed transformations to be viewed without altering frame RAM. The OFG was unique since 12 bits were available for manipulation. The lower eight bits were used for image data (providing 256 gray-levels of resolution) and the upper four were used for graphic overlay information. When the dynamic LUT mode was enabled, the value in the upper four bits (0 - 15) of each pixel location determined which of the 16 LUT transformations was selected. The underlying 8-bit pixel value determined the LUT entry. In static LUT mode, the currently selected LUT was used to transform the entire displayed image.

The Imaging Technology Vision-Plus AT subroutine library (Imaging Technology Incorporated, Woburn, MA) was used as the basic building block for the manipulation of OFG internal registers. The code for this application was written in Borland C++ 3.1 with some external calls to Borland Turbo Assembler 3.0 routines. Extensive modification to the Imaging Technology subroutine library was necessary to implement specific image analysis techniques. Assembly language routines were used for routines that required a high
execution speed. Protected mode was utilized to access an address space of $2^{24}$(16 mb) and was managed through a Phar Lap DOS extender (Phar Lap Software, Cambridge, MA). This was done because the completed software was larger than the regular DOS environment and the images were manipulated in memory.

Fish were videotaped in tanks and in sea cages. Video recordings of chinook salmon (Oncorhynchus tschawytscha) stocked at approximately 8 - 20 kg*m$^{-3}$ were taken in 3.3 m diameter tanks at the Marine Ecosystem Program, West Vancouver Fisheries and Oceans Laboratories, West Vancouver, BC. Fish ranged in size from 0.05 - 0.5 kg. The video camera was positioned at the bottom of the tank, oriented toward the centre standpipe, and tilted upward between 5-10°. Different types of screens were put over the tank so that the background was relatively brighter than the foreground. This procedure effectively backlit the fish within the tank, thereby producing fish silhouettes in medial or side, and anterior fish views.

Video recordings of fish in several sizes of sea cages were taken from an aquaculture research site at the Pacific Biological Station, Fisheries and Oceans Canada at Nanaimo, BC. The fish within these sea cages were of various age-classes and stocking densities. Two species, chinook and atlantic salmon (Salmo salar) were present in different cages in densities ranging from 4 to 8 kg*m$^{-3}$. The largest sea cage used was 7.5 m wide x 7.5 m long x 6 m deep. Secchi disc readings ranged from an average of 6 m in the summer to 11.5 m in the winter. Using different camera angles, fish images were recorded in a variety of medial, ventral, and anterior views. With the cameras oriented towards the water surface, ventral views of silhouetted or backlit fish were recorded.

At the sea cage site, the underwater camera unit was mounted onto a modified column. The column consisted of an aluminum base plate, a modified Manfrotto rotational tripod head, and an aluminum cylinder. The aluminum pipe-column was designed for negligible deflection in a 1.2 m/s current (100 mm outside diameter, 5 mm wall thickness). The column was painted black to appear less invasive. The column length was changeable using different pipe sections. The four sections, each 3 m in length and bolted together, gave a maximum length of 12 m. The entire unit floated vertically, when the bottom column section was flooded.

Video frames were digitized through the OFG Frame Grabber to 512 x 512 pixels. The image file names were coded for easy reference by 1) processing status, 2) tape number, and 3) tape time. Image contrast was enhanced by performing histogram equalization on subdivided portions or general areas of interest (GAOIs) of an image (Jain 1989). The histogram showed the number of pixels in the image at each gray-level over a gray-level range. If there were many pixels with small gray-levels then the histogram was skewed to the left, meaning the image was relatively dark and visa versa (Schalkoff 1989). A desirable image for processing contained pixels with gray-levels that were uniformly distributed over the entire gray scale. When the histogram of any GAOI was skewed, the pixel intensities were remapped over the entire dynamic range, ie from gray-levels 0-255. This process provided GAOI’s with increased contrast and is known as histogram equalization.

The GAOI’s were then analyzed using basic edge detection algorithms. Edge detection is fundamental for the segmentation of fish from its background, and it is defined as the set of procedures that identify the boundary between two regions with relatively distinct gray-level properties (Gonzalez and Wintz 1987). Both first and second derivative operators were chosen to represent changes in gray levels in a discrete plane. First derivative operators were the Roberts’ filter and some directional masks. The second derivative operator used was the Laplacian of Gaussian (LoG) (Marr and Hildreth 1980). Another operator that was used was the Beghdadi and Le Negrat algorithm (1987).

The Roberts filter is a 2 x 2 mask, which approximates the first derivative between discrete pixels. Most directional mask techniques search a pixel area for the occurrence of a gradient vector, which exceeds a predetermined or locally calculated gradient threshold. The value of this threshold gradient vector is used to represent the true direction of the gradient which is perpendicular to the edge. A compass mask is usually represented by eight, 3 x 3 kernels. The gradient search therefore occurs in 45° increments for a 360° rotation. The mask which produces the highest response is selected as the most likely edge orientation.

Marr and Hildreth (1980) developed the LoG or $\nabla^2 G$ operator as part of their work on image analysis and modelling of the human visual system. It is the Laplacian of the Gaussian function defined by:

$$\nabla^2 G = \frac{1}{\sigma^4} \left( 2\sigma^2 - r^2 \right) e^{-r^2/2\sigma^2}$$

in which

$$r^2 = x^2 + y^2$$

where:

- $x, y = \text{spatial coordinates}$,
- $\sigma^2 = \text{variance of the Gaussian function}$.

The two dimensional zero-crossing diameter $w$ was used to characterize the size of the function. It was given by Grimson (1981) as:

$$w = 2\sqrt{\sigma}$$

The keys to the success of the LoG operator over the other operators are the edge smoothing by the Gaussian function and the rotational invariance insured by the Laplacian portion. A disadvantage of the Marr-Hildreth operator is the large size of the kernels. Grimson (1981), in his implementation of the operator, used masks with $w$ values of 4, 9, 17, and 35 pixels. The disadvantage of the larger masks ($w=17-35$) is the enormous computational load placed on the computing system.

A connectivity algorithm (Davies 1990) was applied with an edge detection algorithm to determine if a central pixel followed a valid edge on a fish. The search strategy used to connect detected edges was to follow the edge vector perpendicular to the gray-level gradient. For instance, a directional mask was used to connect adjacent edge information by following edges of similar gradient characteristics.
If connected edges were multi-pixel wide, a minimally recursive algorithm was used to produce one pixel wide edges (Shou-Pyng Shu 1989). The one-pixel-wide edge detection algorithm used consisted of two parts. The first part involved convolving the image with an edge detection operator. Assuming numerous edges were detected, the image was operated upon by one of the three expressions:

\[ E_x(i, j) = E_x(i-1, j) + E_x(i, j) = E_x(i+1, j); \]

If \( E_x(i, j) \geq E_x(i-1, j); \)
and \( E_x(i, j) \geq E_x(i+1, j) \)
otherwise \( E_x(i, j) = 0 \) (4)

\[ E_y(i, j) = E_y(i, j-1) + E_y(i, j) + E_y(i, j+1); \]

If \( E_y(i, j) \geq E_y(i, j-1); \)
and \( E_y(i, j) \geq E_y(i, j+1) \)
otherwise \( E_y(i, j) = 0 \) (5)

\[ E_n(i, j) = \text{Output thinned edge image} \]

\[ E_n(i, j) = \sqrt{E_x^2 + E_y^2} \] (6)

where \( E_x(i,j) \) is the edge space function representing the input edge information for the pre-thinning edge operation, and \( E_y(i,j), E_n(i,j) \) range from 0-255 (assuming an 8-bit pixel intensity value). Any value which was greater than 255 was clipped to 255. Equation 4 was used to decrease the width of edges in the \( i \)-coordinate direction. This was accomplished by increasing intensity levels of edge pixels which had a local maximum intensity and then zeroing the edge pixels which did not have a local maximum intensity. In the same manner, Eq. 5 was used to decrease the edge width in the \( j \)-coordinate direction. If an edge detected at a pixel location was greater than neighbouring pixel intensities, the pixel intensity was replaced by the summation of the pixel intensity and its two neighbouring pixel intensities in order to increase the separation for the pixel (Shou-Pyng Shu 1989).

Detected edges were thresholded after edge detection algorithms were applied using a threshold value. A median filter of 3 x 3 pixels was used in selectivity remove areas of localized noise (Pratt 1991) and a spot noise filter was used in the binary image to eliminate single pixels that were not connected to any surrounding pixels.

RESULTS AND DISCUSSION

In both the tank and net pen environments, fish were apparently aware of the camera unit when it was introduced to them. Chinook salmon swam more quickly and avoided the camera, while Atlantic salmon swam slowly and would even approach the column and cameras. Videotaping of chinook salmon commenced when the fish appeared to return to a normal swimming behaviour after the introduction of the cameras. The adjustment period was approximately 20 minutes.

Several image files were processed. Two of those image files were chosen to be thoroughly discussed in this paper because they represent all of the problems encountered with both backlit and non-backlit fish images.

Backlit images

Backlit fish images usually had well-resolved and distinct edges (Fig. 2). The best fish silhouettes in a sea cage or tank were taken on overcast days because backgrounds had a more uniform gray-level. On sunny days, bright and shaded areas in the images were often evident. When the background consisted of a uniform gray level, gray levels of fish silhouettes video recorded in either a tank or sea cage were usually lower than the background gray-levels.

Fish-image file #A1010644 (Fig. 2) represents the case of a sunny day producing a non-uniform gray-level background. The image consists of chinook salmon video recorded with the cameras at a depth of 6 m and oriented toward the surface. The periphery of the image has dark regions representing net material and shadows. The straight, lower left, diagonal object in the image is the column that supports the camera. Parts of a fish and sometimes a complete fish image have the same gray-level as the almost white background. A fish image appeared white in the water if sunlight appeared to be reflecting off it. In the dark regions, the fish in the image have the same gray-level as their almost black background. Outside of the dark regions, most of the fish images consisted of gray-levels between 0-50 with gray-levels of over 125. This latter condition was typical of well-resolved and contrasted fish images.

One problem that arose with the use of a median filter was the removal of valid edge information. An unforeseen benefit of the median filter was the elimination of background net material which enhanced the fish-image segmentation. Histogram equalization was not necessary in this fish image because the contrast between the fish-image and the background was high in most GAOI’s. Edge detection was successfully accomplished.

The Robert’s filter effectively segmented fish in well backlit or non-shaded regions. The main advantage of this filter was the light computational load. Large image areas could be filtered at near real-time (video rates: 30 frames per second) using a commercial hardware convolver. The disadvantage of this filter was that the detected edges were several pixels wide; consequently, edge thinning (Eqs. 4–6) had to be implemented.

Edges were detected with a directional mask technique (Robinson 1977). The number of masks used increased the computational expense of the algorithm. A 512 X 512 image took twenty minutes to process on a 386SX and two minutes on a 486 DX/33 personal computer with an eight mask operator. Although the directional masks provided good edge detection, the time needed for processing could prohibit its use for automated segmentation of fish images using PC hardware or in real time.

The most connected fish edges on an individual fish (Fig. 3) resulted from a LoG edge detection operation and the use of the connectivity algorithm. The resulting edges were distinct and continuous throughout most of the image with the exception of those in surrounding dark areas represented by net material. A binary threshold of 5 was used to highlight edge information and single, isolated pixels were removed with a spot noise filter.

In initial experiments, \( \sigma \) was varied from 1 to 1.2 pixels
Fig. 2. Image file #A1010644 captured by a camera at 6 m depth within a sea cage. The camera faced toward the water surface.

(Eqs. 1-3) and a small kernel was selected to decrease computational time (Table I). Huertas and Medioni (1986) found a computational saving using the LoG operator, because of the separability of the LoG operator into x and y components (Table II). The separability reduces calculations per pixel from $n^2$ to $2n$, where $n$ is the width of the image window in pixels (King 1985). In our work, the coefficients in Table II were changed to match the coefficients in the center row and column of Table I to ensure that the coefficients in Table I summed to zero. A zero sum prevents edges from being displaced from their actual position.

The initial $\sigma$ values tried were sensitive to high spatial frequencies characterized by noise (Grimson and Hildreth 1985). To reduce the false edges produced by detecting noise, other values of $\sigma$ were tried on several images. Values of $\sigma$ between 1.5 and 3 reduced the detection of noise and a $\sigma$ of 2 appeared to optimize the signal to noise ratio. Those results are consistent with the results of Grimson and Hildreth (1985). Larger values of $\sigma$ were not used because the size of the kernel increased the burden on the CPU.

Unprocessed backlit video images can be used to manually count fish within a given view volume. Features within the bounds of the fish are not visible in a backlit image. Fish edges can be detected using a number of edge detection algorithms. The Roberts filter and directional masks are useful as long as the gray-level of an entire fish is greatly different from its background. Dark fish within a dark net area, for instance, would be very difficult to separate.

Non-backlit background

Backlit fish images, as mentioned above, can be used to count fish. To count or to inspect fish using different camera orientations and lighting conditions, the effect of a non-backlit background on image quality needs to be understood.

Table I: Kernel for the $V^2G$ (13x13) operator for $\sigma=2.0$

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>-4</td>
<td>-4</td>
<td>-4</td>
<td>-3</td>
<td>-3</td>
<td>-2</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-3</td>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-4</td>
<td>-3</td>
</tr>
<tr>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>-1</td>
<td>-2</td>
<td>-4</td>
<td>-2</td>
<td>6</td>
<td>15</td>
<td>20</td>
<td>15</td>
<td>6</td>
<td>-2</td>
<td>-4</td>
</tr>
<tr>
<td>-1</td>
<td>-3</td>
<td>-4</td>
<td>-1</td>
<td>9</td>
<td>20</td>
<td>28</td>
<td>20</td>
<td>9</td>
<td>-1</td>
<td>-4</td>
</tr>
<tr>
<td>-1</td>
<td>-2</td>
<td>-4</td>
<td>-2</td>
<td>6</td>
<td>15</td>
<td>20</td>
<td>15</td>
<td>6</td>
<td>-2</td>
<td>-4</td>
</tr>
<tr>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>0</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-3</td>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-4</td>
<td>-3</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-3</td>
<td>-4</td>
<td>-4</td>
<td>-4</td>
<td>-3</td>
<td>-3</td>
<td>-2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td>-2</td>
<td>-3</td>
<td>-2</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table II: Separable components of the LoG operator ($\sigma=2$). The row ($x$) and the column ($y$) filters have the same coefficients.

<table>
<thead>
<tr>
<th></th>
<th>-3</th>
<th>-4</th>
<th>-1</th>
<th>9</th>
<th>21</th>
<th>28</th>
<th>21</th>
<th>9</th>
<th>-1</th>
<th>-4</th>
<th>-3</th>
</tr>
</thead>
</table>

If fish images are videotaped in a sea cage with the camera oriented in any direction except toward the surface, the background will be composed of many gray-levels and the fish images will generally have poorly defined edges. In a tank, uniform background lighting conditions can be difficult to achieve, because lighting varies with the angle and brightness level of the sun and the effectiveness of the shading material on top of the tanks. Due to attenuation and scattering of light, contrast in any water column differs depending on depth, the side or forward lighting condition, the level of horizontal visibility, and the distance a fish is from the camera. Sunny days produce high spectral reflectance and shaded regions and therefore contribute to a non-uniform background gray-level. Backscattering of light off particles in the water (i.e. minerals, decayed organic particles, plankton) can also greatly reduce image quality by decreasing contrast and resolution.

Fish-image file #A1402441 (Fig. 4) represents nearly all the problems associated with a non-backlit background. The fish in the image are chinook salmon, approximately 1 kg in size. The camera was 6 m deep in the corner of a sea cage facing inward toward the centre of the sea cage on a horizontal plane. The underwater Secchi disk reading was 7.5 m, which made the underwater view volume approximately $600 \text{ m}^3$ (volume of a right square pyramid = $1.4 \text{ Secchi disk reading}^3$).

Fish image quality and gray-levels deteriorate as the fish become closer to the background gray-level and the distance of the fish from the camera increases. Details like eyes, skin coloration, lateral lines, shape, and fins are well defined in fish closest to the camera (Fig. 4). Distinct features in that fish image gave a full spectrum of gray-levels, 0 - 255 for the fish image. Dark background regions are numerous, especially at the lower periphery of this image. Dark regions are due to attenuation, sun angle, net material, or overhead obstructions shading the sea cage.

The unprocessed pixel intensity histogram of this file does not have a bi-modal distribution which is typical of a backlit fish image (Fig. 5a). Fish information was centred around a peak pixel intensity of 50 and the background information was centred around a peak pixel intensity of 125 in a backlit fish image. The unprocessed histogram of the non-backlit fish images produced a peak pixel intensity of 215 and the equalized version of the file had a similar peak. The pixel intensities representing the fish images covered the entire gray scale (Fig. 5b). Histogram equalization applied on GAOL's 100 x 100 pixels did enhance the contrast of fish in shadowed regions; however, it also led to the detection of false edge information. Histogram equalization on the whole image increased the contrast between the fish image and the background, but was not able to separate fish and background information.

The Robert's filter was applied to the enhanced image #A1402441. The threshold gradient vector magnitude value was increased from 20 to 30 to increase the number of correct edges. The threshold value selected describes the minimum magnitude of a gradient that will represent an edge response. However, increasing the magnitude also increased the number of false edge detections and amplified noise in the image. The Robert's filter was, therefore, consistently inadequate in extracting fish-im-

Fig. 3. The result of a LoG edge detection operation on image file #A1010644.
The camera was orientated horizontally toward the center of the cage.

The camera was orientated horizontally toward the center of the cage.

The edge detection algorithm developed by Beghdadi and Le Negrate (1989) partially detected edges on poorly-resolved images such as #A1402441 without the use of histogram equalization because the algorithm selectively enhanced and de-enhanced contrast in relation to all the pixel gray-levels within a windowed processing area. The Beghdadi and Le Negrate (1989) algorithm is implemented on a window of an odd number of pixels. In our experiments a 11 x 11 pixel window was used. A balance exists between selecting a window size which is large enough to minimize noise and yet small enough to allow a 512 x 512 image to be processed within a reasonable length of time. Unfortunately the algorithm was too computationally intensive to process a 512 x 512 image in the PC environment. As well, the Beghdadi and Le Negrate (1989) algorithm did not provide any smoothing of edges when compared with the LoG operator. Therefore, real-time, effective application of this algorithm was not considered plausible on an economical PC based image system.

The connectivity algorithm implemented on image #A1402441 resulted in many false fish-image edges. Both 3x3, 5x5 median filters, and spot noise filters were applied to remove false edges, nevertheless, valid edges were lost. Poor thinning (Eqs. 1-3) was obtained from image #A1402441 because of the lack of valid edge information. The results are directly coupled to the output of the edge detection operators. In general, when edge detection algorithms identified true edges the one-pixel-wide thinning algorithm produced results that enhanced the segmentation of fish images.

The unprocessed video images of non-backlit fish images can be used to manually count fish within a given volume and to inspect features within the fish. Fins, heads, eyes, body shape, and tails are easily viewed provided the fish is close to the camera. Complete and valid fish edge information can not be obtained using basic and simple edge detection algorithms, because detected noise and the poorly contrasting fish images, typical of the marine environment, work against those algorithms.

CONCLUSIONS

Underwater videotaping can be used to monitor and assess fish condition. Fish can be manually counted from these videotapes when it is considered that the fish are within a
Fig. 5. (a and b) Original and enhanced pixel intensities for image file #A1402441. Background and fish image information are not separable in non-backlit fish images.

specific view volume. Automatic or semi-manual systems require a means to segment the fish from their background. Images of backlit fish, that had relatively distinct edges, were more simply and completely segmented. The Robert’s filter provided good edge detection on backlit fish, however, the edges that this first derivative operator detected were several pixels wide; consequently, edge thinning had to be implemented. Regardless, the Robert’s filter has potential to be used on an economical PC-based system in near real-time by applying a hardware convolver.

More sophisticated edge detectors, such as the compass directional mask, the LoG operator, and the Beghdadi and Le Negrate algorithm, all produced well segmented fish, however, the computation time is not close to real-time on a PC-based system.

For non backlit and/or poorly-resolved fish images a first derivative edge detection algorithm, such as Robert’s filter, was consistently inadequate in extracting fish-image edge information. The application of the compass directional mask, the LoG operator, and the Beghdadi and Le Negrate algorithm did not produce enough significant edge information in a non backlit or poorly contrasted image to segment a fish from its associated background. The compass directional mask and the Beghdadi and Le Negrate algorithm, in particular, were too CPU intensive to employ in a PC-based system. The LoG operator, although slow on a PC-based system, may have potential in a real-time system. The operator is separable into $x$ and $y$ components thereby decreasing the convolution calculations from $n^2$ to $2n$ operations. If a hardware convolver were employed real-time edge detection and therefore segmentation of fish from the background may be possible on a PC-based system. Regardless, the need for further application of image processing strategies, both from software and hardware perspectives, has been demonstrated in this study.

ACKNOWLEDGMENT

This project was supported by the Science Council of British Columbia.

REFERENCES


Fig. 6. The result of a LoG edge detection operation on image file #A1402441.


