THE PHOTOGRAMMETRIC RECORD

The Photogrammetric Record 30(149): 46–62 (March 2015)
DOI: 10.1111/phor.12091

AUTOMATED FISH DETECTION IN UNDERWATER IMAGES USING SHAPE-BASED LEVEL SETS

MEHDI RAVANBAKHSH (mehdi.r@rmit.edu.au)
MARK R. SHORTIS (mark.shortis@rmit.edu.au)
RMIT University, Melbourne, Australia

FAISAL SHAFAIT (faisal.shafait@uwa.edu.au)
AJMAL MIAN (ajmal.mian@uwa.edu.au)
The University of Western Australia, Crawley, Australia

EUAN S. HARVEY (euan.harvey@curtin.edu.au)
Curtin University, Perth, Australia

JAMES W. SEAGER (jseager@seagis.com.au)
SeaGIS Pty Ltd, Bacchus Marsh, Australia

Abstract

Underwater stereo-video systems are widely used for the measurement of fish. However, the effectiveness of stereo-video measurement has been limited because most operational systems still rely on a human operator. In this paper an automated approach for fish detection, using a shape-based level-sets framework, is presented. Knowledge of the shape of fish is modelled by principal component analysis (PCA). The Haar classifier is used for precise localisation of the fish head and snout in the image, which is vital information for close-proximity initialisation of the shape model. The approach has been tested on underwater images representing a variety of challenging situations typical of the underwater environment, such as background interference and poor contrast boundaries. The results obtained demonstrate that the approach is capable of overcoming these difficulties and capturing the fish outline to sub-pixel accuracy.

KEYWORDS: fish detection, image segmentation, level sets, prior shape knowledge, registration, underwater image

BACKGROUND AND INTRODUCTION

SOUTHERN BLUEFIN TUNA (SBT – *Thunnus maccoyii*) is a highly migratory species of fish inhabiting open oceans at mid-latitudes in the southern hemisphere. SBT can live for up to about 40 years, reaching a weight of 200 kg and a length of 2 m, but are slow growing, maturing at around 8 to 12 years of age. In the past, the species has supported pole and longline commercial fisheries in Japan and Australia and, to a lesser extent, Taiwan, New

Zealand, Indonesia and Korea. However, SBT catches fell steadily from the 1960s due to over-exploitation, and in response to the formation of the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) and the formalisation of voluntary quotas in 1994, the focus has shifted from fishing to aquaculture. Farming operations based at Port Lincoln in South Australia now account for more than 98% of the 2014 Australian total catch quota of 5147 tonnes. The remainder are caught using longlines or by recreational fishing. SBT are exported almost exclusively to the Japanese "sashimi" markets where it is one of the most valuable fish supplied to the restaurant industry. The Australian SBT industry is valued at over A$1 billion and the export income is approximately A$300 million per annum.

The Australian industry operates by catching juvenile and sub-adult SBT in the Great Australian Bight, transporting the fish to floating grow-out cages at Port Lincoln and then fattening the SBT for several months on a diet predominantly comprised of pilchards (*Sardinops spp.*). The SBT are harvested for the Japanese markets when both the product and prices are optimal. The fish are approximately 1·0 to 1·4 m in length and weigh about 20 to 50 kg when caught. The aquaculture approach doubles the value of the exported fish.

Stereo-video Measurement

The quota has to be monitored to ensure that the catch of SBT transferred to the aquaculture cages does not exceed the tonnage limit imposed by the CCSBT. Traditionally, the SBT are counted during the transfers from the purse-seine nets into the grow-out cages to exclude mortality during the towing period (Fig. 1). Until recently, a sample of 40 fish has been universally used by the industry to estimate the total weight of up to 10 000 SBT in the largest cages. Concerns around stratification of fish in the water column and the statistical validity of the simple estimates have resulted in the introduction of underwater stereo-video systems by the Australian Fisheries Management Authority (AFMA) to provide an accurate measurement of the total weight of each transfer (Harvey et al., 2003). Underwater stereo-video is well established as an effective tool that is capable of making accurate, precise and non-invasive measurements of fish length (Naiberg et al., 1993; Harvey and Shortis, 1996; Steeves et al., 1998) and has the potential to measure 95% of the population in a cage. Based on the snout-to-tail lengths of the fish measured by stereo intersection, weight can be accurately deduced using a length–weight regression (Pienaar and Thomson, 1969; Harvey et al., 2003).

However, the effectiveness of the stereo-video measurement has been limited because most operational systems still rely on a human operator to identify and measure the snout and tail of the fish in order to determine the length by intersection. Whilst automation of the identification of objects and image measurement processes have been demonstrated in many other contexts, due to the uncontrolled underwater environment combined with the loss of contrast because of attenuation through the water, an automated solution for fish sizing has been elusive. Whilst automation of some aspects of the process has been established for at least 15 years (Lines et al., 2001), only recently have fully operational systems that identify, delineate, track and measure fish in an uncontrolled environment been reported (Shortis et al., 2013).

The aim of the current research is to develop a general approach to the automatic measurement of fish in uncontrolled underwater environments. The techniques and algorithms in the measurement approach should be capable of automated measurement of SBT during a transfer to ensure quota compliance, or any species in an aquaculture cage to manage quotas, monitor growth rates or optimise feeding regimes. Automated measurement

is also essential for marine science applications such as habitat monitoring to determine the sustainability and biodiversity of vibrant reef communities (Watson et al., 2009) or more dispersed deep water environments (Shortis et al., 2008). Whilst rapid, automated processing is needed for SBT quota management, long-term studies of management strategies for Marine Protected Areas or threatened species also require automated analysis to process the massive catalogues of video sequences that are captured for longitudinal or spatial studies. However, the initial priority for development of the automated processing is the SBT transfers, in order to rapidly develop an accurate and effective measurement process in response to increased use of underwater stereo-video systems mandated by AFMA.

Fish Identification and Outline Delineation

In the context of this research, automated detection methodologies comprise two steps: identification and subsequent delineation of the fish outline. The existing literature on fish detection has been mainly focused on the identification step only, where the presence of fish is recognised in the scene and followed by the estimation of the fish location (Evans, 2003; Walther et al., 2004; Morais et al., 2005; Zhou and Clark, 2006; Spampinato et al., 2008; Palazzo et al., 2013). In contrast, relatively few approaches have been reported that deal with both identification and the following delineation of the fish silhouette (Lines et al., 2001; Hariharakrishnan and Schonfeld, 2005; Khanfar et al., 2010).

In Palazzo et al. (2013), colour and texture features integrated into a covariance-based model are used to identify fish with an average success rate of 80%. Similarly, Spampinato et al. (2008) used colour and texture within an adaptive Gaussian mixture model with a performance of about 85%. The Gaussian mixture model has also been employed by Evans (2003) to integrate fish shapes in the image domain. In Zhou and Clark (2006), head and tail features of a particular fish species are extracted through Gabor filtering (designed for texture detection) and modelled by geometrical characteristics. This approach has assumed a certain orientation of the fish and the image being clear of occlusions. Morais et al. (2005) modelled a fish as an ellipse and proposed a methodology to identify and count fishes based on the application of a Bayesian filtering technique with a success rate of 81%. While this approach offers a number of flexibilities and is not constrained by the quantity, behaviour, size and pose of fish, it requires a large number of training samples for the filtering algorithm to perform successfully, which is time consuming to generate. The method proposed by Walther et al. (2004) to extract fish outperforms other methods with a success rate of 90% being reported. This technique employed an attentional selection algorithm to detect and track salient features across video frames. These features are then classified to identify the object of interest.

In contrast to those approaches aforementioned, the following three approaches produce the outline of fish in two separate steps. Hariharakrishnan and Schonfeld (2005) proposed an approach that identifies fish through a watershed-based segmentation and subsequently captures the fish boundary using motion vector information between frames. This approach requires a prior knowledge about the class of fish to process images automatically. Otherwise, the fish outline is extracted manually to be entered into the tracking module.

Khanfar et al. (2010) used combined frame differences and histogram thresholding to identify fishes in the images. Then, an edge-detection operator is applied to the image, segmented by thresholding to detect the initial fish. The fish outline is obtained by a

shrinking circle to locate its outer boundary more precisely. This method, however, can fail due to scene complexity of an uncontrolled marine environment and poor contrast of fish boundaries. Furthermore, the elastic curve employed can ultimately converge to the outer edge in which case the fish shape cannot be captured accurately.

Lines et al. (2001) proposed a pattern recognition method called the binary pattern classifier to identify objects in the image that are likely to be fish and, subsequently, a model-based point distribution model (PDM) method is used to fit a shape template, generated from a set of training images, to the target fish by minimising an energy-like function. This approach showed a relatively low success rate of 63% in precise delineation of fish outlines and is sensitive to the initial placement of the model. Furthermore, PDM cannot recover the fish shape when occluded by shadows. Although Lines et al. (2001) used shape information to move the curve to the expected positions, the underlying energy function is based on image gradient only, and thus is highly sensitive to the presence of noise and poor image contrast. A region-based energy term is necessary to overcome these limitations.

Most of these approaches use low-level image features such as colour, texture, intensity and motion to detect fish. However, in a real-life situation, the uncontrolled underwater environment produces images that are characterised by low contrast, background clutter and interference, partial occlusion caused by adjacent or foreground objects, varied illumination conditions and shadows. The aforementioned techniques fail to produce high-quality results mainly due to misleading low-level features resulting from image noise and occlusion, or lack of sufficient low-level features necessary for object modelling. High-level knowledge of fish can significantly aid in providing an efficient solution to these problems.

New Research Approach

In this paper, an automated approach for fish detection using a shape-based level-sets framework is presented. An example of underwater stereo-images used is shown in Fig. 1. The shape knowledge of fish is modelled by principal component analysis (PCA) (Leventon et al., 2000) that will be used to guide the level-set curves. PCA enables the representation of global shape variation of the object of interest through a training set of shape templates. The global shape information is incorporated into the Mumford–Shah functional, as reported by Chan and Vese (2001), which can detect objects in strongly cluttered scenes. A Haar-like detector method (designed for object recognition) is used to identify the existence of fish and determine their locations in the image. This information is vital to place the shape in close proximity to the object to be segmented, which increases the success rate and requires fewer iterations for convergence. The level-set algorithm is applied to one image at a time. Once the fish are independently identified on the left and right images, stereo intersections for the snout and tail are computed based on the well-established approach of a geometrically constrained epipolar search and template match between the two images.

The outline of this paper is as follows. In the next section a short review of level sets is given. The individual steps of the proposed detection strategy along with mathematical equations are described in the following section. Next, experimental results using underwater sample image sequences recorded in cages are presented and evaluated. The paper concludes with a discussion of the progress and results achieved, and an outlook for future work.

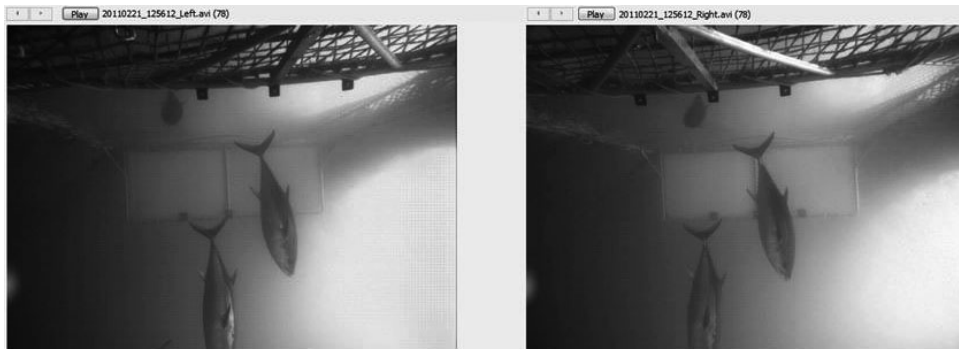


FIG. 1. Typical stereo-image pair captured during a transfer from the purse-seine net to the grow-out cage. The water surface is to the right of the images and the apparent vertical orientation of the fish is caused by the mounting of the stereo-video system on the side of the transfer gate.

LEVEL-SETS REPRESENTATION

Level sets share many properties with snakes. *Snakes* (Kass et al., 1988) have emerged as a powerful tool for semi-automated object modelling. They have been used in a photogrammetric context by authors such as Ravanbakhsh et al. (2008) and Yari et al. (2014) and are especially useful for delineating features like fish bodies that are difficult to model with rigid geometric primitives. Snakes are represented as explicit, parametric contours. As a result, they do not allow for automatic changes of topology. Several approaches were proposed to address this problem (Szeliski et al., 1993; McInerney and Terzopoulos, 1995) that propose heuristic procedures for detecting possible splitting and merging of the initial contour. In contrast, *level sets* (Osher and Sethian, 1988) are parameter-free, which allows for splitting and merging in a natural way. Snakes are known as an edge-based segmentation technique, which means the incorporation of region and shape information into their energy functional is not straightforward. In this work, a level-sets framework is adopted to introduce region and shape information into the segmentation process. Whilst the change in topology functionality is not used explicitly in this research, this capability will be incorporated in future work to improve the flexibility of the shape representation.

The core idea of level sets is to implicitly represent a contour C as the zero-level curve of a function φ of higher dimension (Fig. 2). An initialisation of φ can be constructed in the following way. Let C be a closed curve representing the boundary between two regions, one region inside the curve and another region outside the curve. Here φ is then defined as the signed distance $\pm d(x)$ to the curve, negative inside and positive outside. The definition is illustrated as:

$$\varphi(x) = \begin{cases} -d(x) & \text{if } x \text{ is inside } C \\ +d(x) & \text{if } x \text{ is outside } C. \end{cases} \tag{1}$$

While the use of the distance $d(x)$ is not mandatory when using level sets, it ensures that φ does not become too flat or too steep near C, and subsequently can be differentiated across the zero-level curve without running into numerical problems.

In order to combine the characteristics of the level-set function, image information and shape knowledge of the known object, an energy functional can be set up and

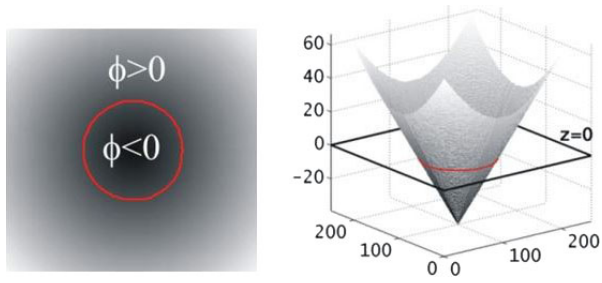


FIG. 2. Illustration of level sets. (a) The circular curve C is used to construct the level-set function φ such that φ is negative inside and positive outside the curve. Distance values d are grey-value coded. (b) A plane at zero level ($Z = 0$) intersects the level-set function φ and thus the zero-level curve C is obtained.

consequently minimised using the calculus of variations. Minimising the energy functional is performed in an iterative process moving the initial curve towards the object boundaries.

DETECTION STRATEGY

The fish detection strategy comprises three primary steps (Fig. 3). First, the presence of fish is recognised and the initial locations are determined using segmentation of a frame difference from an averaged background image. A Haar-like detector is then employed to

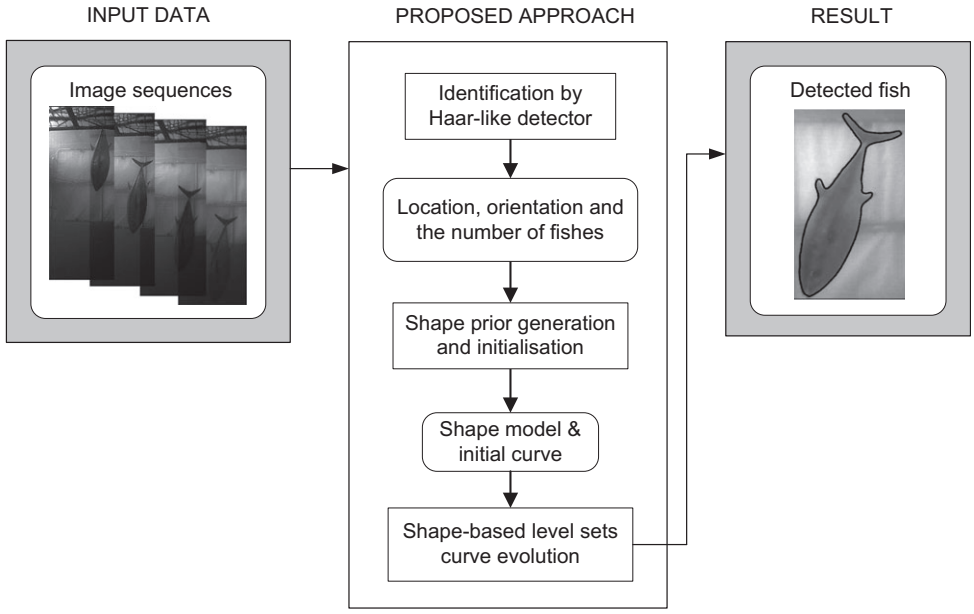


FIG. 3. Workflow of fish detection.

estimate the snout and tail locations, from which the initial position and orientation of each fish in the image can be derived. Subsequently, prior knowledge of the fish shape, the so-called shape-prior model, is constructed by PCA using a set of training samples. The level-sets curve is then initialised and evolved to locate the fish boundary. The result obtained consists of the detected fish.

Identification

In this stage, the location of the fish snout and tail in the image are determined. Precise localisation of the snout and tail leads to the estimation of pose parameters in 2D space, these being two rotations, two translations and one scale parameter.

Of the many approaches that can be employed to locate the fish snout and tail, template matching has shown a certain degree of success as reported in Mahmood and Khan (2010, 2012). In this approach, single representative templates, centred on the snout and tail mid-points, are extracted from sample videos. Template matching, using correlation rather than absolute differences to provide robustness in the presence of illumination changes, is used to locate the regions of interest in the video images. However, this approach does not account for perspective distortion, requiring either the use of multiple templates that capture appearance variations from different viewing angles, or using the more common solution of image matching based on affine or perspective transformation.

A more effective method of locating snout and tail is to use Haar-like features in a boosted classifier set-up (Viola and Jones, 2004) that has shown high object-detection accuracy, besides being able to operate in real time. The method is in widespread use for face detection, even being employed by low-cost consumer-grade cameras for real-time face-priority focus. Additionally, Haar-like classifiers are employed in this research for locating the snout and tail of fish in underwater image sequences because of their high detection speed and ability to perform a scale-space search.

To train the classifier, 200 manually cropped images of the target object (snout or tail) are used so that the classifier can learn which features (among a set of possibly thousands of features) can locate the target with high accuracy. These features, once learned, are then used to construct the object classifier that can locate the presence of the object in cluttered scenes. The training time depends on the amount of training data, but in general requires a few hours of processing time. Retraining is required only if the imaging conditions change substantially.

The results of independent detection of the snout and tail using Haar detectors are further improved using relationships between the detected snouts and tails. The search for the tail is based on the results of snout detection and vice versa, within the limits of separation and angle expected for the imagery. Fig. 4 shows an example of a Haar classifier used to identify the snouts and tails and of southern bluefin tuna (SBT) during a transfer.

Precise localisation of the tip of the snout and the valley point of the tail, used as reference points, are used to estimate rigid transformation parameters. These transformation parameters are then used to first generate the reference shape and subsequently initialise the shape model, two crucial steps in accurate and correct delineation of fish.

Shape-prior Generation

The generation of shape prior comprises two steps: first, the training samples need to be geometrically aligned; subsequently, the shape model is constructed from the aligned

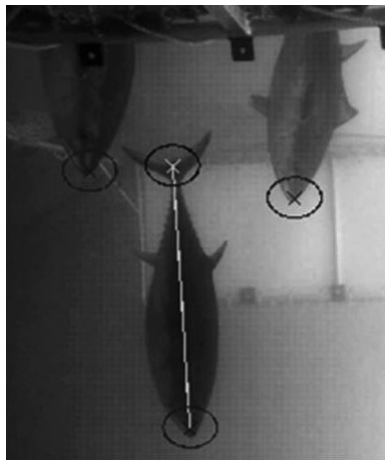


FIG. 4. The identification of SBT snouts and tails, marked by ellipses, using the Haar classifier.

shapes. The alignment involves matching shapes of training samples that differ in size, orientation and translation. In the literature, a large number of shape matching methods have been reported. A complete review of those methods is given in Veltkamp and Hagedoorn (1999).

In this paper, the alignment of training samples is realised using the method introduced in Chen et al. (2002). Suppose that the training set contains n given curves $C_1, \ldots, C_n$ with their corresponding interior regions $A_1, \ldots, A_n$ as subsets of R^2. The shape similarity measure of the shapes C_1 and C_2 are defined as:

$$a(C_1, C_2) = \text{area of } (A_1 \cup A_2 - A_1 \cap A_2). \tag{2}$$

In the alignment process, the so-called pose parameter of C_1 is considered to be fixed, and the remainder of the samples $(C_2, \ldots, C_n)$ are jointly aligned to C_1 through the solution of the rigid transformation:

$$C_j^{\text{new}} = s_j \mathbf{R}_j C_j + T_j \quad (j = 2, \ldots, n)$$

such that the area $a(C_1, C_j^{\text{new}})$ is minimised and where s_j is a scale factor, $\mathbf{R}_j$ is a rotation matrix and T_j is a translation. These values are obtained by a global optimisation algorithm called the genetic algorithm (Davis, 1991), which makes it less likely for the underlying function to be trapped in a suboptimal local minimum compared with purely local methods such as gradient descent.

The shapes are encoded in binary images to simplify the alignment task. Fig. 5 shows a set of 20 training samples manually digitised and the result of their alignment. The first sample (bottom row, leftmost), which is the scaled, shifted and rotated version of the corresponding manually digitised sample (top row, leftmost), is adopted as the reference. It has fixed pose parameters estimated in the identification process and to which the rest of the samples are registered. Figs. 6(a) and (b) illustrate the amount of shape variability depicted

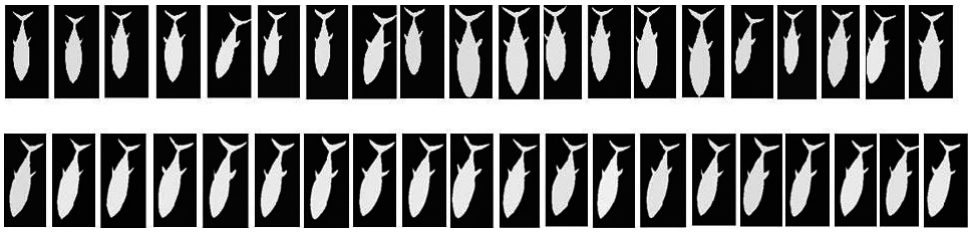
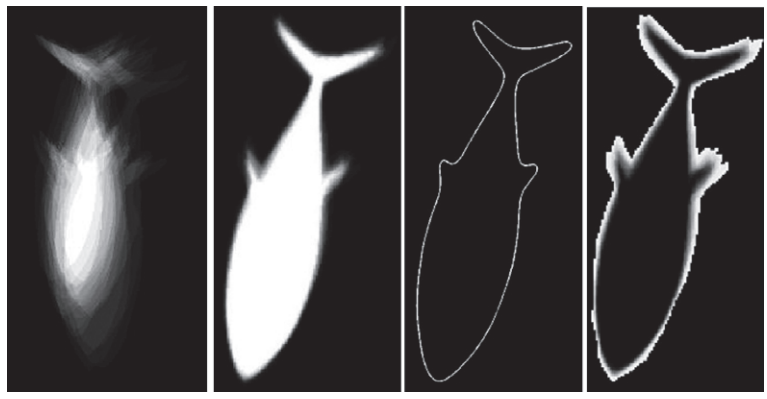


FIG. 5. Twenty manually digitised training samples. The top row shows the binary representation of the shape model of fish before alignment. The bottom row presents the training samples that were geometrically aligned.



(a) (b) (c) (d)

FIG. 6. (a) Overlaid training samples depicting varying degrees of overlap before alignment. (b) Aligned samples. (c) Average of aligned shapes. (d) Model variability (grey-value coded) with white representing the highest variability and black the lowest variability.

in the overlap images before and after the alignment. It can be seen that large shape discrepancies can often exist in real fish images. These shape differences can be removed successfully, which demonstrates the effectiveness of the alignment method (Fig. 6(c)). Furthermore, model variability is represented in Fig. 6(d), showing that the areas around the boundaries of the fish fin and tail experience the largest deformations in the fish body outline. It is interesting to note that key regions that could be used for species identification, such as the dorsal and anal fins together with the tail, are the profile sections which show the greatest variability.

In the next step, a shape model is constructed using the aligned shapes. The PCA method is selected to construct the shape model due to its efficiency at capturing the main variations of a training set while removing redundant information. In a similar manner to Leventon et al. (2000), the boundaries of each of the training shapes are represented in the training dataset as the zero-level set of n signed distance functions (SDFs) $\{\varphi_1, \ldots, \varphi_n\}$ with negative distances assigned to the inside, and positive distances assigned to the outside, of the shape boundary.

Suppose $\mathbf{M}$ is a matrix whose column vectors are the n aligned training SDFs $\{\varphi_i\}$. PCA is then applied to these SDFs to compute eigenvalues and eigenvectors of the covariance matrix:

$$\Sigma = \frac{1}{n}\mathbf{M}\mathbf{M}^{\mathrm{T}} \tag{3}$$

and the mean level-set function of the training set:

$$\bar{\varphi} = (1/n)\sum_{i=1}^{n}\varphi_i. \tag{4}$$

The eigenvectors are called principal components or eigenshapes. In practice, the first k principal components ($k \leq i$) are sufficient to model the major shape variations in the training samples. In Mika et al. (1999), a method is proposed for determining the value of k by examining the eigenvalues of the corresponding eigenvectors. This approach, however, cannot be adopted here as the value of k varies in different applications (Tsai et al., 2003). In this work, the value of k was set empirically based on extensive testing and visual evaluation. Then, shape can be represented as the zero-level set of the following function:

$$\varphi(\mathbf{w}) = \bar{\varphi} + \sum_{i=1}^{k} w_i\varphi_i \tag{5}$$

where $\mathbf{w} = \{w_1, \ldots, w_k\}$ denote the weights for the k eigenshapes, with the variances of these weights $\{\sigma_1^2, \ldots, \sigma_k^2\}$ given by the eigenvalues. In equation (5), the shape variability is restricted to the variability given by the eigenshapes. To accommodate a wider range of shape variability, a pose parameter $\mathbf{p}$ (this being the translation, scale and orientation) is incorporated into the level-set function of (5). With the addition of $\mathbf{p}$, the implicit description of shape is given by the zero-level set of the following function:

$$\varphi(\mathbf{w}, \mathbf{p}) = \bar{\varphi} + \sum_{i=1}^{k} w_i\varphi_i \tag{6}$$

where $\bar{\varphi}$ and each φ_i are now a function of $\mathbf{p}$.

Once the shape model is generated, an initial level-set function is constructed using a rectangle as the initial "curve" around the detected fish. Then, the zero-level set of the level-set function is evolved towards the fish boundary according to the energy functional. The energy functional is described in the following section.

Shape-based Level-sets Energy Functional

The energy functional is based on the segmentation model proposed by Chan and Vese (2001) in an effort to overcome the limitations found with the previous edge-based strategies. Unlike edge-based methods, where the provision of close initialisation to the object of interest and good contrast boundaries are necessary to locate those boundaries, the region-based methods used in this work are independent of image gradients and are less likely to converge to local minima if an undesirable feature or image noise is present.

Let I be a given image and C the evolving curve defined as $C = \{(x, y) \in R^2 : \varphi(x, y) = 0\}$, with u and v denoting two constants representing the averages of I inside and outside the curve C. Assume that the image I is formed by two regions of

approximately piecewise-constant intensities with distinct values of I_0^i and I_0^o, and that the object to be detected is represented by the region with value I_0^i and boundary C. Then, $I_0 \approx I_0^i$ inside the object (inside C) and $I_0 \approx I_0^o$ outside the object (outside C). By minimising the following energy equation, the boundary of the object of interest C is obtained (Chan and Vese, 2001):

$$E_{cv} = \int_{inside\ C} |I-u|^2 \,\mathrm{d}x\,\mathrm{d}y + \int_{outside\ C} |I-v|^2 \,\mathrm{d}x\,\mathrm{d}y \tag{7}$$

which is equivalent to the energy functional below (Tsai et al., 2001):

$$E_{cv} = -(u^2 A_u + v^2 A_v) = -\left(\frac{S_u^2}{A_u} + \frac{S_v^2}{A_v}\right) \tag{8}$$

where A_u and A_v denote areas, and S_u and S_v represent the sum intensity of areas, inside and outside C, respectively. Then, the gradient descent is employed to search for the parameters $\mathbf{w}$ and $\mathbf{p}$ that minimise E_{cv} to implicitly determine the segmenting curve C. The parameters A_u, A_v, S_u and S_v can be expressed in terms of $\varphi(\mathbf{w}, \mathbf{p})$:

$$S_u = \int_\Omega IH(\varphi(\mathbf{w},\mathbf{p}))\,\mathrm{d}A; \quad A_u = \int_\Omega H(\varphi(\mathbf{w},\mathbf{p}))\,\mathrm{d}A \tag{9}$$

and

$$S_v = \int_\Omega IH(-\varphi(\mathbf{w},\mathbf{p}))\,\mathrm{d}A; \quad A_v = \int_\Omega H(-\varphi(\mathbf{w},\mathbf{p}))\,\mathrm{d}A \tag{10}$$

where Ω defines a bounded and open subset of R^2 and H denotes the Heaviside function:

$$H(\varphi(\mathbf{w},\mathbf{p})) = \begin{cases} 1 & \text{if } \varphi(\mathbf{w},\mathbf{p}) < 0 \\ 0 & \text{if } \varphi(\mathbf{w},\mathbf{p}) \geq 0. \end{cases} \tag{11}$$

The energy function (8) is minimised with respect to $\mathbf{w}$ and $\mathbf{p}$ using gradient descent optimisation:

$$\nabla_{\mathbf{w}} E_{cv} = -2(u\nabla_{\mathbf{w}} S_u + v\nabla_{\mathbf{w}} S_v) + (u^2 \nabla_{\mathbf{w}} A_u + v^2 \nabla_{\mathbf{w}} A_v) \tag{12}$$

$$\nabla_{\mathbf{p}} E_{cv} = -2(u\nabla_{\mathbf{p}} S_u + v\nabla_{\mathbf{p}} S_v) + (u^2 \nabla_{\mathbf{p}} A_u + v^2 \nabla_{\mathbf{p}} A_v) \tag{13}$$

where the gradient parameters are given as:

$$\nabla_{w_i} A_u = -\nabla_{w_i} A_v = -\int_C \varphi_i(\mathbf{p})\,\mathrm{d}s \tag{14}$$

$$\nabla_{w_i} S_u = -\nabla_{w_i} S_v = -\int_C I\varphi_i(\mathbf{p})\,\mathrm{d}s \tag{15}$$

$$\nabla_{p_i} A_u = -\nabla_{p_i} A_v = -\int_C \nabla_{p_i} \varphi(\mathbf{w}, \mathbf{p})\, \mathrm{d}\, s \tag{16}$$

$$\nabla_{p_i} S_u = -\nabla_{p_i} S_v = -\int_C I \nabla_{p_i} \varphi(\mathbf{w}, \mathbf{p})\, \mathrm{d}\, s. \tag{17}$$

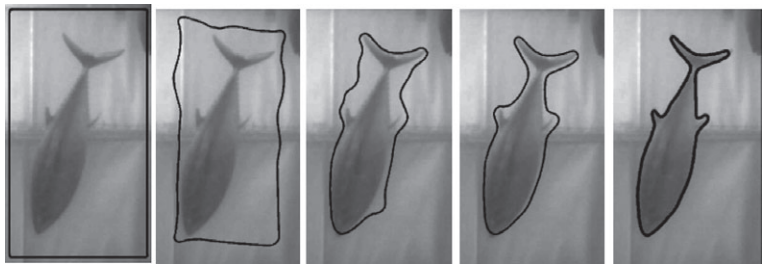
In these equations the segmenting curve C is given by the zero-level set of $\varphi(\mathbf{w}, \mathbf{p})$, and $\nabla_{p_i} \varphi(\mathbf{w}, \mathbf{p})$ is the gradient of $\varphi(\mathbf{w}, \mathbf{p})$ taken with respect to the ith component of the transformation matrix $\mathbf{p}$ that includes translation, rotation and scale. The gradient descent optimisation of equations (12) and (13) leads to the parameters $\mathbf{w}$ and $\mathbf{p}$. The updated $\mathbf{w}$ and $\mathbf{p}$ parameters, which are iteratively computed during the optimisation, are then used to implicitly determine the location of the segmenting curve C.

The curve evolution is terminated when the overall change in the evolving curve positions per iteration is less than 0·1 pixel. A smaller threshold considerably increases the computation cost, although the quality of the final result is similar.

Experimental Results and Evaluation

Underwater image sequences, recorded at the transfer gate between two cages, have been used to test the fish detection algorithm. From the large number of video samples recorded for eight transfers, 35 sample images have been chosen to represent the variable and uncontrolled nature of the marine environment. These images include a varying number of SBT fish with a range of illumination changes, background interference and occlusions caused by adjacent fishes. Moreover, SBT appear in the image sequences with missing or poor contrast boundaries which further exacerbates the challenging conditions.

In Fig. 7, an example of results is shown where the initial "curve" is placed as a rectangle around the fish of interest; this curve subsequently converges to the fish boundary by minimising the energy functional presented in the previous section. Further example results are shown in Fig. 8 where, in the four rightmost samples, SBT are partially occluded by other neighbouring fishes in the foreground and background. In almost all samples, fish boundaries are of low contrast, especially in the areas around the tail and fin. The detection results shown in Fig. 8 demonstrate that the approach is capable of overcoming those limitations typical of the underwater environment and capturing the fish outline accurately.



(a) (b) $n = 3$ (c) $n = 10$ (d) $n = 13$ (e) $n = 54$

Fig. 7. Fish detection result. (a) Initial curve; (b), (c) and (d) intermediate curves and (e) the final detection result. n denotes the number of iterations in the intermediate and the final result.

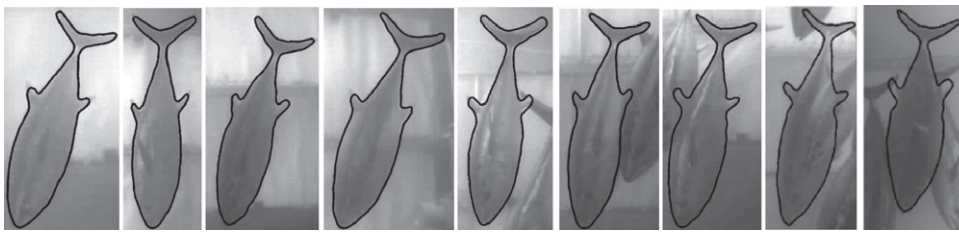


FIG. 8. Detection results of different fish in the presence of a range of background interference and foreground occlusions by other fish (two rightmost samples). *n* denotes the final number of iterations as shown for different fish samples.

In order to quantitatively evaluate the performance of the approach, the detection results were compared to manually plotted fish used as reference data. The comparison was carried out by matching the detection results to the reference data using the so-called *buffer method* (Heipke et al., 1998). A detected object is assumed to be correct if the maximum distance between the detected object and its corresponding reference does not exceed the buffer width. Furthermore, a reference object is assumed to be matched if the maximum deviation from the detected object is within the buffer width. Based on these assumptions the following quality measures were used in the present work:

(1) *Completeness*: the ratio of the number of matched reference objects to the total number of objects.
(2) *Correctness*: the ratio of the number of correctly detected objects to the number of detected objects.
(3) *Geometric accuracy*: the average distance between the correctly detected objects and its corresponding reference expressed as root mean square (RMS) value.

Table I shows the fish detection evaluation results. The buffer width can be defined according to the required detection accuracy for a specific application. In the tests in this research, the buffer was set to 3, 5 and 8 pixels according to the range of accuracy achievable at the identification step. Furthermore, this selection allows assessment of the relevance of the approach for applications that demand varying degrees of accuracy. Both the completeness and correctness have increased as the buffer width was increased from 3 to 8 pixels, implying that the results are more complete and correct for higher buffer-width values. However, the geometrical accuracy increases in inverse proportion to the buffer-width value, so that results obtained with a value of 3 pixels are more accurate than those obtained with a larger buffer-width value. There is thus a certain trade-off between correctness and completeness on the one hand, and geometric accuracy on the other.

As expected, the results are encouraging and sub-pixel geometric accuracy has been achieved in all experiments with high rates of completeness and correctness. Table I shows that the developed approach is, in principle, capable of extracting fish accurately under

TABLE I. Evaluation results for fish detection applied to 35 samples.

Buffer width (pixel)	*Correctness (%)*	*Completeness (%)*	*Geometric accuracy (pixel)*
3	89·6	91·4	0·7
5	94·3	94·3	0·8
8	100	100	0·9

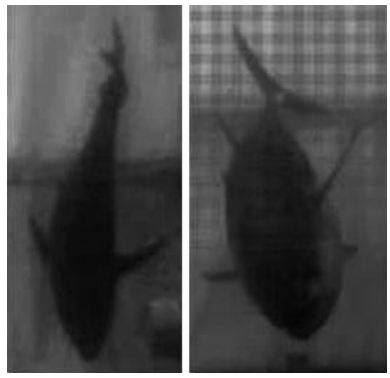


FIG. 9. Two samples where the approach fails to delineate the entire fish boundary.

occlusion and within variable underwater environments. Whilst no formal time comparison between manual measurement and the automatic approach has been carried out, due to the high curvature of the fish shape and the requirement for many points to be digitised, the automatic approach will certainly achieve much higher efficiency levels.

However, a few problems still remain. Fig. 9 displays two examples where large deviations from the prior-shape model can be observed, particularly in the upper half of the fish body and around the tail. As shown in Fig. 6(d), it is expected that most of the deformation takes place around the fins and the tail. Thus, such severe deformation cannot be absorbed with the current shape model.

CONCLUSION AND OUTLOOK

In this paper, an automated approach for the detection of fish from underwater images has been proposed, developed and tested. It comprises a region-based level-set method that enables the delineation of the fish outline. The shape information of fish is incorporated into the level-sets formulation through the PCA method to overcome such limitations as poor contrast boundaries, background clutter and occlusions caused by neighbouring fish. To provide a close initialisation for the shape model, the pose of fish in the image is determined using the Haar classifier. The results of the developed approach have been applied to 35 samples of varying quality and levels of occlusion, and include a quantitative evaluation of the results using three buffer-width values.

The presented results show that level sets can be used to delineate fish outlines from underwater images if the shape information of the fish species is incorporated into the level-sets energy functional. Furthermore, it was found that an energy function that is independent of image gradients and includes the shape model is able to overcome various kinds of disturbances and the problems related to low-quality images recorded in the underwater environment, such as poor contrast and uneven illumination.

The current approach is developed to detect SBT in a farming aquaculture environment. The techniques developed here have clear potential to be extended to wild habitats provided that the deformation of the fish body and movement information derived from image sequences is taken into account. In wild habitats, fish can move in any direction with large deformations occurring in the body, as shown in Fig. 9, causing the fish detection approach to break down.

For the technique to be successful in wild habitats, varying rates of deformation and fish orientation need to be modelled. The detection of different fish species in addition to SBT is another goal that will be pursued in future research, as in reef and other underwater habitats many fish species are present. Furthermore, investigation into the possibility of using colour information in the level-sets formulation will be carried out.

REFERENCES

CHAN, T. F. and VESE, L. A., 2001. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2): 266–277.

CHEN, Y., TAGARE, H. D., THIRUVENKADAM, S., HUANG, F., WILSON, D., GOPINATH, K. S., BRIGGS, R. W. and GEISER, E. A., 2002. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3): 315–328.

DAVIS, L. (Ed.), 1991. *Handbook of Genetic Algorithms*. Van Nostrand, New York, USA. 385 pages.

EVANS, F. H., 2003. Detecting fish in underwater video using the EM algorithm. *Proceedings of the IEEE International Conference on Image Processing*, 3: III – 1029–1032.

HARIHARAKRISHNAN, K. and SCHONFELD, D., 2005. Fast object tracking using adaptive block matching. *IEEE Transactions on Multimedia*, 7(5): 853–859.

HARVEY, E. and SHORTIS, M., 1996. A system for stereo-video measurement of sub-tidal organisms. *Marine Technology Society Journal*, 29(4): 10–22.

HARVEY, E. S., CAPPO, M., SHORTIS, M., ROBSON, S., BUCHANAN, J. and SPEARE, P., 2003. The accuracy and precision of underwater measurements of length and maximum body depth of southern bluefin tuna (*Thunnus maccoyii*) with a stereo-video camera system. *Fisheries Research*, 63(3): 315–326.

HEIPKE, C., MAYER, H., WIEDEMANN, C. and JAMET, O., 1998. External evaluation of automatically extracted road axes. *Photogrammetrie, Fernerkundung, Geoinformation*, 2: 81–94.

KASS, M., WITKIN, A. and TERZOPOULOS, D., 1988. Snakes: active contour models. *International Journal of Computer Vision*, 1(4): 321–331.

KHANFAR, H., CHARALAMPIDIS, D., IOUP, G., IOUP, J. and THOMPSON, C. H., 2010. Automated recognition and tracking of fish in underwater video. *Final Report, LA Board of Regents Contract*, NASA(2008)-STENNIS-08. 40 pages.

LEVENTON, M. E., GRIMSON, W. E. L. and FAUGERAS, O., 2000. Statistical shape influence in geodesic active contours. *IEEE International Conference on Computer Vision and Pattern Recognition*, 1: 316–323.

LINES, J. A., TILLETT, R. D., ROSS, L. G., CHAN, D., HOCKADAY, S. and MCFARLANE, N. J. B., 2001. An automated image-based system for estimating the mass of free-swimming fish. *Computers and Electronics in Agriculture*, 31(2): 151–168.

MAHMOOD, A. and KHAN, S., 2010. Exploiting transitivity of correlation for fast template matching. *IEEE Transactions on Image Processing*, 19(8): 2190–2200.

MAHMOOD, A. and KHAN, S., 2012. Correlation-coefficient based fast template matching through partial elimination. *IEEE Transactions on Image Processing*, 21(4): 2099–2108.

MCINERNEY, T. and TERZOPOULOS, D., 1995. Topologically adaptable snakes. *Proceedings of the Fifth IEEE International Conference on Computer Vision*, 840–845.

MIKA, S., SCHÖLKOPF, B., SMOLA, A., MÜLLER, K.-R., SCHOLZ, M. and RÄTSCH, G., 1999. Kernel PCA and de-noising in feature spaces. *Advances in Neural Information Processing Systems*, 11: 536–542.

MORAIS, E. F., CAMPOS, M. F. M., PADUA, F. L. C. and CARCERONI, R. L., 2005. Particle filter-based predictive tracking for robust fish counting. *18th IEEE Brazilian Symposium on Computer Graphics and Image Processing*, 367–374.

NAIBERG, A., PETRELL, R. J., SAVAGE, C. R. and NEUFELD, T., 1993. Non-invasive fish size assessment method for tanks and sea cages using stereo-video. *Techniques for Modern Aquaculture* (Ed. J. K. WANG). American Society of Agricultural Engineers, St Joseph, Michigan, USA. 372–381.

OSHER, S. and SETHIAN, J. A., 1988. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1): 12–49.

PALAZZO, S., KAVASIDIS, I. and SPAMPINATO, C., 2013. Covariance based modeling of underwater scenes for fish detection. *Proceedings of 20th IEEE International Conference on Image Processing*, 1481–1485. http://groups.inf.ed.ac.uk/f4k/PAPERS/ICIPcs13.pdf [Accessed: 24th November 2014].

PIENAAR, L. V. and THOMSON, J. A., 1969. Allometric weight–length regression model. *Journal of the Fisheries Research Board of Canada*, 26(1): 123–131.

RAVANBAKHSH, M., HEIPKE, C. and PAKZAD, K., 2008. Road junction extraction from high resolution aerial imagery. *Photogrammetric Record*, 23(124): 405–423.

SHORTIS, M. R., SEAGER, J. W., WILLIAMS, A., BARKER, B. A. and SHERLOCK, M., 2008. Using stereo-video for deep water benthic habitat surveys. *Marine Technology Society Journal*, 42(4): 28–37.

SHORTIS, M. R., RAVANBAKHSH, M., SHAFAIT, F., HARVEY, E. S., MIAN, A., SEAGER, J. W., CULVERHOUSE, P. F., CLINE, D. E and EDGINGTON, D. R., 2013. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. *Videometrics, Range Imaging, and Applications XII, SPIE* 8791: paper 0G.

SPAMPINATO, C., CHEN-BURGER, Y.-H., NADARAJAN, G. and FISHER, R., 2008. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *Proceedings of 3rd International Conference on Computer Vision Theory and Applications*, 2: 514–519.

STEEVES, G., PETERSON, R. H. and CLARK, L. D., 1998. A quantitative stereoscopic video system for visually measuring the linear dimensions of free-swimming fish. *Oceans '98 Engineering for Sustainable Use of the Oceans*, 3: 1405–1408.

SZELISKI, R., TONNESEN, D. and TERZOPOULOS, D., 1993. Modeling surfaces of arbitrary topology with dynamic particles. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 82–87.

TSAI, A., YEZZI, A., WELLS, W., TEMPANY, C., TUCKER, D., FAN, A., GRIMSON, W. E. and WILLSKY, A., 2001. Model-based curve evolution technique for image segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1: 463–468.

TSAI, A., YEZZI, A., WELLS, W., TEMPANY, C., TUCKER, D., FAN, A., GRIMSON, W. E. and WILLSKY, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2): 137–154.

VELTKAMP, R. C. and HAGEDOORN, M., 1999. *State-of-the-art in shape matching*. Technical Report UU-CS-1999-27. Utrecht University, the Netherlands. 24 pages.

VIOLA, P. and JONES, M. J., 2004. Robust real-time face detection. *International Journal of Computer Vision*, 57(2): 137–154.

WALTHER, D., EDGINGTON, D. R. and KOCH, C., 2004. Detection and tracking of objects in underwater video. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 544–549.

WATSON, D. L., ANDERSON, M. J., KENDRICK, G. A., NARDI, K. and HARVEY, E. S., 2009. Effects of protection from fishing on the lengths of targeted and non-targeted fish species at the Houtman Abrolhos Islands, Western Australia. *Marine Ecology Progress Series*, 384: 241–249.

YARI, D., MOKHTARZADE, M., EBADI, H. and AHMADI, S., 2014. Automatic reconstruction of regular buildings using a shape-based balloon snake model. *Photogrammetric Record*, 29(146): 187–205.

ZHOU, J. and CLARK, C. M., 2006. Autonomous fish tracking by ROV using monocular camera. *3rd Canadian Conference on Computer and Robot Vision*, Quebec City, Quebec, Canada. 8 pages (on CD-ROM).

Résumé

Les systèmes vidéo stéréoscopiques sous-marins sont largement utilisés pour la mesure des poissons. Cependant, la possibilité d'une telle mesure est limitée car la plupart des systèmes opérationnels s'appuient toujours sur un opérateur humain. Dans cet article, une approche automatisée pour la détection de poissons, utilisant une méthode par ensembles de niveau basée sur la forme, est présentée. La connaissance de la forme du poisson est modélisée par une analyse en composantes principales (ACP). Le classificateur de Haar est utilisé pour la localisation précise de la tête et du museau du poisson dans l'image, une information vitale pour l'initialisation du modèle de forme. L'approche a été testée sur des images sous-marines représentant une variété de situations difficiles typiques des environnements sous-marins, comme les interférences avec le fond et les limites peu contrastées. Les résultats obtenus montrent que la méthode permet de surmonter ces difficultés et de déterminer le contour du poisson avec une précision inférieure au pixel.

Zusammenfassung

Zur Vermessung von Fischen sind Unterwasser Stereo-Video Systeme weit verbreitet. Allerdings benötigen die im Markt erhältlichen Systeme in der Regel den menschlichen Operateur und sind somit hinsichtlich Automation in der Effektivität begrenzt. Dieser Beitrag stellt einen automatisierten Ansatz zur Erkennung von

Fischen vor, der auf Basis von Formkriterien arbeitet. Das Wissen über Fischformen wird durch eine Hauptkomponentenanalyse (PCA) modelliert. Mit Hilfe des Haar Klassifizierers wird eine genaue Lokalisierung von Kopf und Maul des Fisches im Bild erzielt, die entscheidend für die passende Initalisierung des Formmodells ist. Der vorgestellte Ansatz ist mit Unterwasseraufnahmen geprüft worden, die durchaus schwierige Situationen, wie Interferenz mit dem Hintergrund, oder schwache Kontrastgrenzen beinhalten. Es hat sich gezeigt, dass der Ansatz auch derartige Probleme bearbeiten kann und den Fischumriss mit Sub-Pixel Genauigkeit erfassen kann.

Resumen

Los sistemas de video estereoscópico subacuático se usan profusamente en la medida de peces. Sin embargo, en la mayoría de los sistemas operacionales de video estereoscópico la eficiencia de las medidas está limitada a su dependencia en un operador humano. En este artículo se presenta un procedimiento de detección automático de peces. La forma de los peces se modela mediante el análisis de componentes principales. El clasificador de Haar se usa para la precisa localización de la cabeza y el hocico del pez en la imagen, información vital para una inicialización del modelo de la forma del pez. Este procedimiento ha sido comprobado en una variedad de imágenes acuáticas representando diversas situaciones difíciles del entorno acuático, tales cómo interferencias del fondo de la imagen y fronteras con un pobre contraste. Los resultados demuestran que este procedimiento es capaz de superar dichas dificultades y capturar el contorno del pez con precisión subpíxel.

摘要

水下立体视频系统广泛用于鱼类的测量。然而,由于多数立体视频量测仍然依赖于人工操作,因此限制了立体视频测量的应用。本文提出了一种自动化鱼检测的方法,使用基于形状的水平集框架。其中鱼的形状的知识通过主成分分析模型采集,*Haar*分类器用于在图像中鱼的头、鼻子的精确定位,这是形状模型初始化前的重要信息。该方法已在各类具有挑战性的典型的水下环境中获取的水下图像上进行了测试,如背景干扰和低对比度边界。结果表明该方法能够克服以上难题,且捕获的鱼的轮廓可以达到亚像素级精度。