Journal of Food Engineering 181 (2016) 84–91

Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

A machine vision system for automatic detection of parasites *Edotea magellanica* in shell-off cooked clam *Mulinia edulis*

Pablo A. Coelho [a, b, *], Sergio N. Torres [a, b], Wagner E. Ramírez [a, b], Pablo A. Gutiérrez [b], Carlos A. Toro [a, b], Jose G. Soto [b], Daniel G. Sbarbaro [a], Jorge E. Pezoa [a, b]

[a] Department of Electrical Engineering, University of Concepción, Concepción, Chile
[b] Center for Optics and Photonics (CEFOP), Concepción, Chile

ARTICLE INFO

Article history:
Received 4 November 2014
Received in revised form
20 October 2015
Accepted 8 March 2016
Available online 10 March 2016

Keywords:
Mulinia edulis
Edotea magellanica
Clams
Optoelectronic imaging
Machine vision
Image processing
Parasite detection
Real time classification

ABSTRACT

In this work, a machine vision system for the automatic online detection of the parasite *Edotea magellanica* inside of the clam *Mulinia edulis* is presented. The machine vision system uses a transillumination technique to acquire images of the clam-parasite tandem. To improve the light transmission properties of the clam and parasite tissues, a novel online flattening system to flatten the clam thickness was developed. The automatic detection of the parasite in a clam is accomplished by a binary decision tree classifier that analyzes clam images. The classifier was developed using a supervised pattern-recognition approach, and proper features were created using spectral, spatial, geometrical, and biological information, such as the relative transmittance, the location of the parasite in the clam, the shape of the parasite, and the anatomical regions of the clam. The prototype system was built and tested, achieving an average classification accuracy of 98%, for a laboratory training sample set of 200 clams, and an accuracy of 73% for another set of 200 clams in an industrial setting.

1. Introduction

It is known that the Chilean economy seeks to position among the global leaders in food production. The seafood industry is not an exception and, in fact, is one of the most important development clusters in Chile (Perez-Aleman, 2005). The seafood demand is significant throughout the World and is constantly increasing, consequently, the seafood industry is an industrial sector of great interest for both Chile and the World.

In the southern Chilean sea inhabits a clam called *Mulinia edulis* (Fig. 1), which is part of the Chilean seafood industry, and is being sold both domestically and abroad. This clam, however, is seasonally infected by a parasite called *Edotea magellanica* (Fig. 1). According to Gonzalez and Jaramillo (1991), between 1.5% and 4.5% of the clams are infected by the parasite, and the parasite neither affects the clam's meat nor human health. It is reported also by Gonzalez and Jaramillo (1991) that the parasite hides in a specific anatomical area of the clam, which is known as the clam's mantle (see Section 2.1). Due to the strictly quality control procedures required today in the food industry (Nollet and Toldra, 2010) this parasite must be removed for cosmetic reasons. Currently, in Chilean industries, the parasite extraction is carried out manually by human operators. As a consequence, this process is expensive, slow, and inaccurate.

In order to automate the process of detecting the parasite, some relevant detection techniques commonly used in the food industry are reviewed. Machine vision systems are widely used in the food industry to carry out automate tasks (Mathiassen et al., 2011). Online vision systems operating in the Visible (VIS) and the Near-Infrared (NIR) spectral ranges have been developed to carry out tasks involving food quality evaluation and the detection of diseases and infections.

With the aim of detecting hidden objects, many techniques have been successfully applied. The use of X-rays has shown good results, for example, in the detection of bones in poultry meat (Tao

* Corresponding author. Department of Electrical Engineering, University of Concepción, Concepción, Chile.
E-mail addresses: pcoelho@udec.cl (P.A. Coelho), sertorre@udec.cl (S.N. Torres), waramirez@udec.cl (W.E. Ramírez), pablogutierrez@udec.cl (P.A. Gutiérrez), ctoron@udec.cl (C.A. Toro), jsotop@udec.cl (J.G. Soto), dsbarbar@udec.cl (D.G. Sbarbaro), jpezoa@udec.cl (J.E. Pezoa).

http://dx.doi.org/10.1016/j.jfoodeng.2016.03.005
0260-8774/

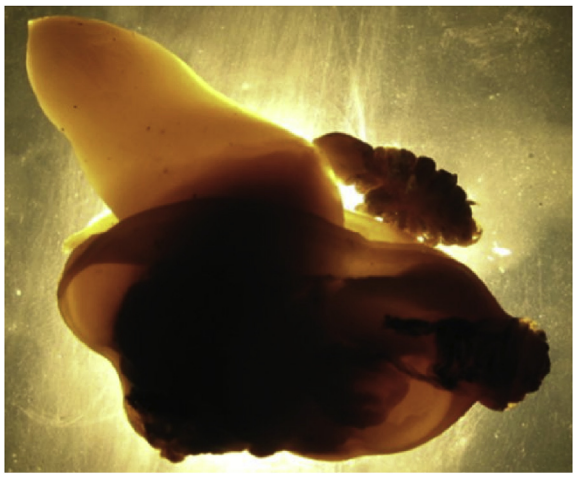


Fig. 1. The *Mulinia edulis* clam and the *Edotea magellanica* parasite. In this image, the parasite is placed outside the clam.

and Ibarra, 2000) and in fish fillets (Mery et al., 2011). These applications were successful because the foreign objects and the analyzed product had significantly different densities. Although this technology appears to be appropriate for our problem, it is discarded by two practical constraints. First, the clam's and parasite's tissues density are very similar so they can not be distinguished, as observed in practice and reported in Coelho (2011). Secondly, the cost of X-rays technology is out of reach for small Chilean seafood industries.

A different class of approaches to detect hidden objects uses an optical technique known as optical transillumination. This technique involves measuring the transmitted light through the sample and has shown good results in parasite detection problems associated to some seafood products. For instance, Sivertsen et al. (2011) developed a hyperspectral imaging system with an optical transillumination technique as a method for the automatic detection of parasites in cod fillets (*Gadus morhua*). This system reached an overall detection rate of 58% among all the nematodes in the cod fillet. Wold et al. (2001) investigated how to detect parasites hiding under the surface of fish's meat by using multispectral imaging and optical transillumination. The developed system was able to detect parasites at depths of up to 6 mm below the surface of the fish muscle. Heia et al. (2007) show that, by applying a white light transmission and a imaging spectroscopy setup to cod fillets, parasites can be detected at depths up to 0.8 cm below the surface of the fish fillet, that is 2–3 mm deeper than manual detection.

Our research group has previously reported in Coelho et al. (2013) a method, based on hyperspectral imaging analysis, for detecting the parasite *E. magellanica* hidden in the clam *M. edulis*. In that work, optical transmittance features were found for the clam-parasite tandem. The main result of that research is shown in Coelho et al. (2013). This result is the normalized average optical transmittance for the clam's stomach and mantle when a clam is infected by a parasite and when is parasite free. The hyperspectral analysis has shown that a better parasite detection can be achieved as one moves from the VIS to the Infrared (IR) spectral bands. It must be recalled that NIR, Mid-Wave Infrared (MWIR), and far IR technologies are expensive and, as in the case of X-Rays, they are also out of reach for small Chilean seafood industries. Thus, the idea is to focus on the use of the information in the red band of the VIS spectrum captured using inexpensive VIS technology. Along this line of work, our group reported a preliminary solution for the clam parasite detection problem in Soto et al. (2012). The main contribution reported in such work was a prototype solution to flatten the clam's width in order to homogenize the sample transmission thickness. The procedure to flatten the clam's width was performed manually and off-line.

The aim of this paper is to present a machine vision system that automatically detects the presence or absence of the parasite *E. magellanica* in the clam *M. edulis* by using an inexpensive camera in the VIS spectral range and an online flattening system. The prototype machine developed in this work, see Fig. 2, uses a visible RGB camera and a transillumination technique to acquire the clam's images. The constructed machine considers an embedded vision system with machine vision algorithms specially designed for the morphology and spectral transmission parameters of the clam-parasite tandem. A novel flattening system is also reported in order to homogenize the clam optical transmission properties. The automatic detection of the parasite in a clam was accomplished by means of a binary decision tree classifier that analyzes clam images. The classifier was developed using a supervised pattern-recognition approach, and proper features were created using spectral, spatial, geometrical, and biological information, such as the relative transmittance, the location of the parasite in the clam, the shape of the parasite, and the anatomical regions of the clam. The prototype system was built and tested in a Chilean seafood industry.

The rest of the paper is organized as follows. In Section 2 the clam sample preparation and the machine vision system main characteristics are described. The image preprocessing algorithms as well as the binary decision tree classifier are also presented in Section 2. The main results are discussed in Section 3. Finally, in the Section 4 the main conclusions of this work are summarized.

2. Material and methods

In this section, the clam sample preparation is explained. In addition, the optoelectronic components of the machine vision system and its associate control software are described. Finally, a new online flattening system to flatten the clam thickness is presented and tested.

2.1. Defining anatomical sections of interest in the clam

The first anatomical Region of Interest (RoI) used in this work is the mantle cavity, which was labeled as region A in Fig. 3, and the

Fig. 2. Prototype implementation of the machine vision system for the automatic detection of the parasite *Edotea magellanica* in shell-off, cooked *Mulinia edulis* clams.

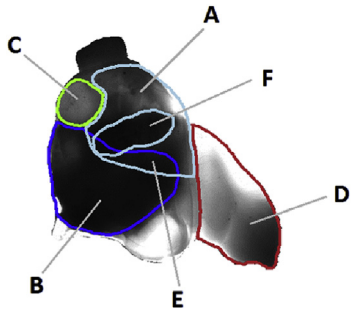


Fig. 3. *Mulinia edulis* main anatomical zones: A) The mantle, B) The stomach, C) The muscle, D) The foot, E) The mantle-stomach overlap zone, and F) The parasite zone.

area within the mantle where the parasite is most likely to be found, labeled as region F. Other anatomical RoI are the ones surrounding the mantle: the clam's stomach (region B), the clam's muscle (region C), and the clam's foot (region D). As stated in Coelho et al. (2013), the anatomical region where the mantle and the stomach overlap (region E) is the most ambiguous detecting zone, because of the high attenuation experienced by the optical transmittance.

2.2. Clam-parasite sample set preparation

The clam sample set was jointly selected by our research team and the staff with the Chilean seafood company associated to this research effort. In order to remove the clam's shell, *M. edulis* clams were vapor cooked at 123 °C for 3 min, and afterwards, placed on a vibrating screen for 30 s. The quality control staff of the seafood company manually inspected and labeled the shell-off cooked clams into two sets: the first set had 100 parasite-free clams, while the second set had 100 parasite infected clams. The average clams length was 50 ± 10 mm, while the average parasite length turned out to be 8 ± 2 mm.

2.3. Image acquisition

In the following subsections, the optics components and the vision system used for the image acquisition are described in detail. Also, the flattening system for the transillumination setup is presented as well.

2.3.1. Lighting, camera, and real-time, embedded vision system

The lighting unit used in the system consists in a 220 VAC, 180 W (at full light intensity) halogen light source (OSL1–EC, Thorlabs, USA), with a spectral emission in the range of 400–800 nm, at optical powers between 1 and 400 W. This light source emits a circular spatial pattern which is guided with an optical fiber (LLG0338-6, Thorlabs, USA) with a relative 97% optical transmission in the visible range. The optical fiber output-end is placed at a distance of 10 cm below the analysis region, in an angle of 90° with respect to the transparent conveyor belt, as depicted in Fig. 4. Under this setup, the relative optical transmission measured for the conveyor belt is 98%. It must be noted that this setup assures that the maximum light intensity not only reaches the sample at the analysis region but also covers the sample. Moreover, in Fig. 2 it is shown that an stainless steel dome has been placed above the conveyor belt, covering the camera, the cylinder, and the sample, thereby isolating the system from external light perturbations. The characteristics of the light source, optical fiber and conveyor belt are listed in Table 1.

The camera unit comprises an inexpensive, visible color Charge-Coupled Device (CCD) camera (scA640-70fc, Basler, Germany) and an 8 mm fixed, focal-length lens, both with C-Mount threads (male or female). The camera has a 1/3″ CCD sensor (Sony ICX424) with a resolution of 640 × 480 pixels and a frame rate of 30 frames per second (fps). The camera is placed inside the cylinder focusing on the analysis region, at a working distance of 10 cm, so that the clam samples are fully contained in the Field of View (FoV) of the camera when they are centered. During the image acquisition process, the clams are transported by the conveyor belt at a speed of 8 cm/s, while the camera continuously acquires several images of the flattened clam. By following a procedure described in Section 2.4.1, the vision system is able to recognize either if the sample is fully contained in the camera's FoV, at the center of the analysis region, or if it is just partially contained. In order to acquire non saturated VIS images, the lens aperture and the camera integration time were properly adjusted.

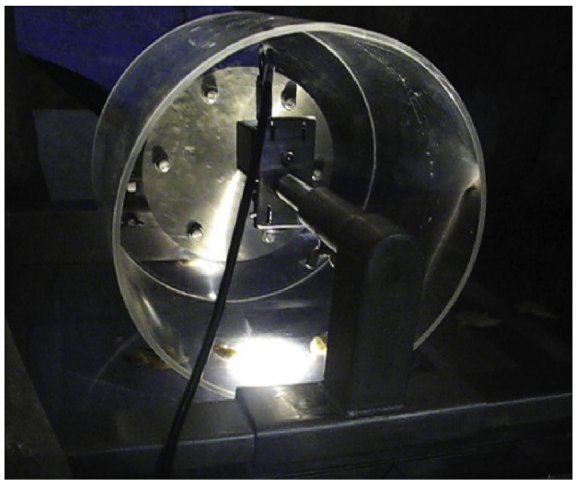


Fig. 4. Transparent acrylic cylinder used to flatten the clam thickness.

The vision system software for image acquisition and processing was coded using Labview and Vision Development Module (National Instruments, Austin, TX, USA). The vision system hardware comprises a real-time embedded processing unit NI EVS-1464RT (National Instruments, Austin, TX, USA), where the camera as well as all the output monitoring and control signals are connected to. Also, the processing unit runs a software where the image processing algorithm and the classifier are executed online.

2.3.2. Flattening system: homogenizing clam thickness

The main idea of the flattening system is to flatten the clam thickness in order to obtain effective images of the clam-parasite tandem. The developed flattening system reduces clam's internal cavities, thereby avoiding unwanted light reflection and dispersion as well as enhancing exclusively the light transmission properties of the clam and parasite tissue. The system comprises the transparent acrylic cylinder shown in Fig. 4, with the dimensions listed in Table 1. The cylinder was designed to protect the clam from the pressure used to even its thickness. Also, the cylinder seeks to avoid image distortion caused by an excessive curvature. The conveyor belt was selected to be flexible so it compensates the cylinder pressure when the clam is at the analysis region.

2.4. Image processing algorithm

An image processing algorithm was developed to clearly identify each clam's anatomical RoI depicted in Fig. 3. These regions are used in the classification stage (see Section 2.5) to define the

Table 1
Machine vision system: optical components and their characteristics.

Component	Material/Model	Size/Characteristics		
Light source	Thorlabs OSL1-EC	Light output	30,000	Foot–candles
		Spectral emission	400–700	nm
Optical Fiber	Thorlabs LLG0338-6	Belt distance	10	cm
		Relative transmittance	97	%
		Length	2	m
		Numerical aperture	0.48	
Conveyor belt	PVC transparent	Relative transmittance	98	%
		Width	0.4	m
		Thickness	0.003	m
		Belt speed	8	cm/s
Acrylic cylinder	Transparent acrylic	Relative transmittance	97	%
		Diameter	0.3	m
		Width	0.2	m
		Thickness	0.003	m

presence or absence of the *E. magellanica* parasite inside of the *M. edulis* clam. Fig. 5 shows a complete flowchart of the image processing algorithm and the parasite detection method devised in this work. The image processing algorithm considers four main stages, which are explained in the following subsections.

2.4.1. Image selection

As each clam is moving across the analysis region, a set of pictures are taken at different positions, as shown in Fig. 6a). Only one out of all the acquired images is selected for further analysis. The selected image is the one where the clam appears fully illuminated. Such selection is carried out automatically using the following procedure. First, solely the blue plane of each RGB image is considered. The reason to use only the blue plane is because, at such wavelengths, the tissue of the clam and the parasite is more opaque than at higher wavelengths (planes green and red). This observation is clearly shown in Coelho et al. (2013). Consequently, using only the blue plane of each image greatly simplifies the operation of distinguishing the clam from the rest of the image. Next, all the images are multiplied by a circular mask. After this operation, the histogram of the each image is calculated and the image having a histogram with the smallest average value is selected as the image completely illuminated. This idea is supported by the observation that when the clam is fully illuminated, its corresponding image is the darkest due to the low relative transmittance of the clam, as shown in Fig. 6b). Consequently, the average value of its histogram is the smallest of the set of images.

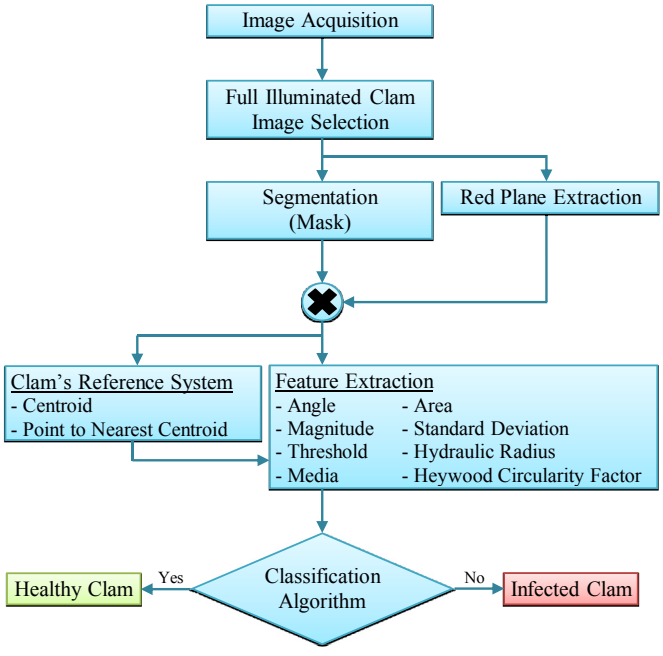


Fig. 5. Flowchart of the detection method.

2.4.2. Segmentation and mask generation

In this part of the image processing method, the goal is to generate a mask capable of isolating the clam from the rest of the objects in the image selected as fully illuminated. To do so, the color space of the selected image is first changed from RGB to HSL. The segmentation method for the selected image consists in generating three binary images from the intensity and saturation planes. The three binary images are: (i) the clam's edge image, (ii) the clam's area image, and (iii) an image where high intensity areas have been eliminated. Then, these three images are fused into a single image using the Hadamard product, thereby rendering an image with only the clam's region and other small objects. These small objects are generally associated with fractions of the shell, sand, or water particles. In order to remove such objects from the resulting image, solely the largest area in the image, which corresponds to the clam's area, is maintained. Finally, the mask drawing the contour convex hull is created as shown in Fig. 6c.

2.4.3. Clam's self-generated spatial reference system

With the aim of determining the position of the clam inner regions, a spatial reference system was designed. Since each clam may take any orientation on the conveyor belt, the designed reference system must account for clam's rotations. Thus, the spatial reference system is self-generated for each image and is based on polar coordinates (r,θ), where the system origin is at the mask centroid. The polar axis is set to be at the line defined by the centroid and the contour point closest to the mask centroid. It has been observed that such boundary point always lies between the anatomical zones D and F, at the specific location where the foot attaches to the mantle. It must be noted that such observation allows us to design a spatial reference system that is both common to all the clams and independent of clam's rotations. Fig. 6d shows a picture with the implementation of the reference system in polar coordinates. In the figure, the regions $z1$ and $z2$ are determined by the coordinates $(r1,\theta1)$ and $(r2,\theta2)$, respectively. It is noteworthy to mention also that the mask was reduced to avoid detecting objects around the clam, such as water particles. With this, the analysis focuses only on the location of areas inside of the clam.

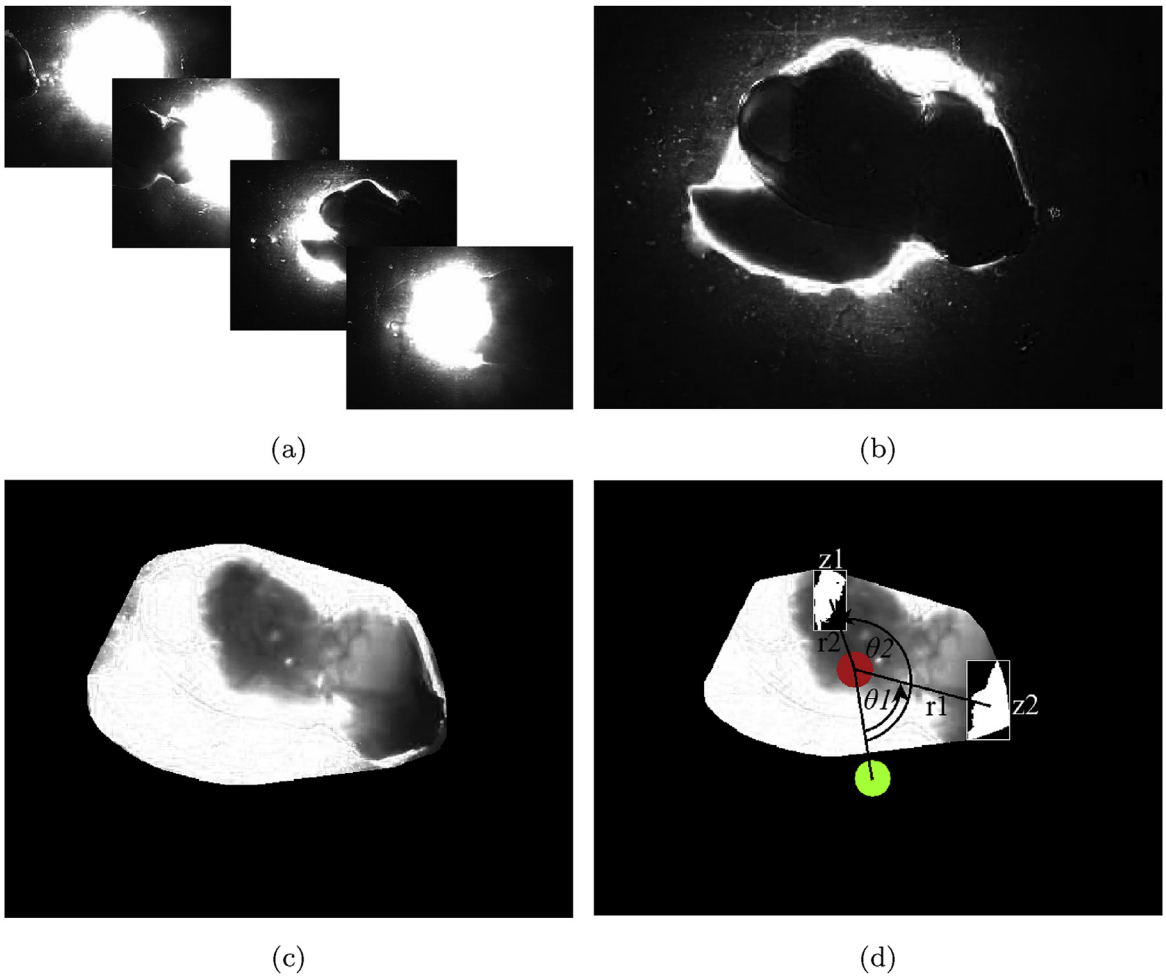


Fig. 6. Image processing sequence. (a) A set of images in the blue plane is taken. (b) Only the fully illuminated clam is selected for the detection process. (c) The mask applied to the clam image in the red plane. The picture is shown in grayscale. (d) The self-generated spatial reference system of the clam in polar coordinates. The system origin is denoted using a red circle (centroid of the mask) and the polar axis is the line connecting the red circle with the green circle, which is the contour point closest to the mask centroid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.5. Features extraction and database construction

Following the results found in Coelho et al. (2013), the candidate regions where a parasite can be hidden are generated segmenting the masked image in the red plane of the RGB color space. The segmentation is performed by thresholding the image at all the intensity values in the range [1255]. For each threshold value, a binary image is created where several regions may be generated, each with different characteristics. For example, Fig. 6d depicts an example of two regions ($z1$ and $z2$) that are generated by applying a threshold intensity value of 96 to the clam image. Next, the following features are extracted from each region: (i) their location (magnitude and angle), with respect to the spatial reference system; (ii) their normalized area; (iii) the mean and standard deviation of their intensities; (iv) their Heywood circularity factor; and (v) their hydraulic radius. Finally, for each generated area, a vector with the aforementioned features is generated. Table 2 lists the value of the features calculated for regions $z1$ and $z2$ shown in Fig. 6d.

The main idea of defining the aforementioned features is to create spectral, spatial, and geometrical information about the clam-parasite tandem. More precisely, the mean and standard deviation of the intensity values are features supplying information about the relative optical transmittance of the regions. The magnitude and angle features provide spatial information about the location of the RoI. The normalized area, the Heywood factor, and the hydraulic radius of the regions are features supplying geometrical information about the clam and the parasite. In particular, the Heywood circularity factor (defined as the ratio between the region perimeter and the perimeter of a circle with the same area) and the hydraulic radius (defined as the ratio between the region area and its perimeter) aim to quantify the similarity of the region with a cylinder, which roughly models the geometry of the parasite.

Using a supervised machine learning approach, each one of

Table 2
An example of the feature extraction for a clam.

Feature	Region z1	Region z2
Threshold	96.000	96.000
Angle (au)	0.532	0.181
Magnitude (au)	0.721	1.269
Mean	90.204	63.784
Standard deviation	4.155	18.751
Area (au)	0.027	0.044
Heywood factor	1.201	1.274
Hydraulic radius	8.856	10.714

these vectors in the database was manually labeled, and next, the entire database was used to train a classifier. The labels assigned to each class were '0,' which corresponds to a parasite-free region, and '1,' which represents a parasite-infected region. The classification criteria used to label an inspected clam is the following: if any of the regions composing the clam image is classified as a parasite infected, then the clam is announced as infected.

3. Results and discussion

In this section, the results achieved by the prototype machine vision system, when identifying the parasite *E. magellanica* inside of the shell-off cooked clam *M. edulis*, are discussed. First, the image acquisition and processing results are described, and next, the classification results are presented.

3.1. Image acquisition and processing results

With the frame rate of the camera set to 30 fps and with the conveyor belt moving at a speed of 8 cm/s, the total time taken by the system to analyze a clam sample is approximately two seconds. This time is measured since a clam enters to the FoV of the camera until is recognized by the system. Under this setting, and due to the high-intensity transillumination setup, no image blurring was observed during the image acquisition. This observation is important since it avoids extra processing routines accounting for this issue, and simultaneously, saves time for other processing and classification tasks. In addition, the classification time, measured since the time when the clam image is captured until a decision is announced, was approximately 500 ms. After the classification process is completed, a binary signal is sent to a red (correspondingly, green) LED indicating that the clam is infected (correspondingly, healthy). It is noteworthy to mention that, when the machine was operating, it was observed that a few number of samples stained both the conveyor belt and the cylinder, thereby affecting their light transmission. Although this could be an issue, the machine was designed to be washable by following hygiene standards. In our experiments the cylinder has required to be washed within a frequency of two operation hours. The flattening system effectiveness was corroborated by achieving uniformly transilluminated clam images, as those shown in Fig. 6b. Fig. 6 shows also the results for the acquisition and registration of the clam images.

3.2. Classification results in the laboratory

In this work, the database was used to train a binary decision tree classifier because such classifiers are fast, robust, accurate with large datasets, and computationally simple (Izenman, 2008; Breiman et al., 1984). The binary decision tree operates defining a set of multiples decisions, corresponding to tree nodes. Each node has a feature with an associated threshold, which was calculated in such a way that the feature space is partitioned, using a greedy approach, in order to separate the required classes.

To improve the performance of the classifier, prior probabilities for the classes and misclassification costs were also included as design parameters. First, it must be recalled that Gonzalez and Jaramillo (1991) reported that only between 1.5% and 4.5% of the clams are infected by the parasite. To account for such class imbalance, prior probabilities in the range [0.015,0.045] were associated to the parasite-infected class (Duda et al., 2001). Second, misclassification costs associated to both false positives (misclassifying a parasite-free clam) and false negatives (misclassifying a parasite-infected clam) were introduced. Note that it is key to minimize the number of false negative errors because they may result in unfordable economical and reputational costs associated with a final client finding a parasite in a clam. Additionally, to simplify the training of the decision tree classifier a maximum of 8 node-levels were defined and the Harsher pruning method was employed to reduce its size and computing complexity (Duda et al., 2001).

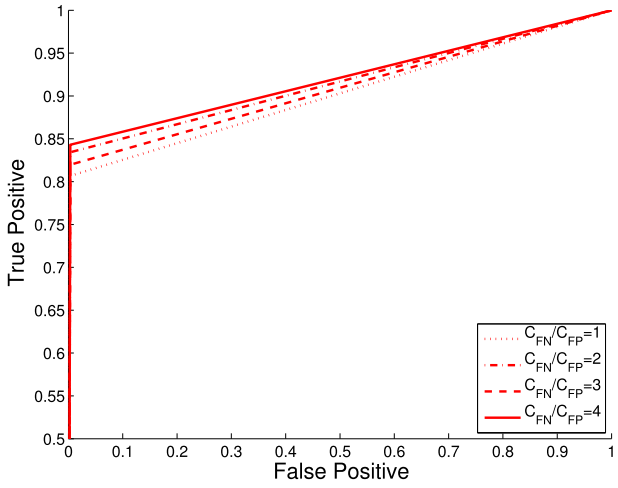
To train the classifier, a ten-fold cross validation methodology was used to statistically assess the performance of the classifier. The feature database was randomly separated into a training set containing the 90% of the features, and a test set containing the remaining 10%, and this procedure was repeated ten times (Duda et al., 2001). Results provided here correspond to average values over the ten folds. To train the classifier, two cases for the prior probabilities associated to the classes were considered. The "case 1" (correspondingly, the "case 2") assumed that the class parasite-infected has a prior probability of 0.015 (correspondingly, 0.045) and the class parasite-free has a prior probability of 0.985 (correspondingly, 0.955). In addition, to simplify the design of the classifier, misclassification costs were regarded as a single parameter, which was defined as the ratio between the cost of misclassifying a false positive, C_{FP}, and the cost of misclassifying a false negative, C_{FN}.

Fig. 7 shows the statistically performance of the binary decision tree classifier. Fig. 7a) and b) show the Receiver Operating Characteristic (ROC) curves for cases 1 and 2, respectively, parameterized by different misclassification cost rates. For the case 1, the best performance is achieved when the misclassification cost of a false negative is four times the misclassification cost of a false positive. For the case 2, however, the best performance is achieved when the misclassification cost rate is one. In both cases, the true positive classification probability, that is the probability of correctly classifying an infected clam, is approximately 0.85. Fig. 7c) shows average classification probability (that is, the probability of correctly classifying both parasite-infected and parasite-free clams), the false positive probability, and the false negative probability, as a function of the misclassification cost rate and for cases 1 and 2. The figure shows that the lowest average classification probability is 0.987, which is achieved for the case 1 and for a misclassification cost rate of 3.2. In addition, the figure also shows that the maximum false positive probability is very low: 0.004. These very good values for the performance metrics are expected due to the severe class imbalance. Since the parasite-free class is clearly dominant and the binary tree decision classifier exhibits a true negative probability of 0.99, it is expected that the average classification probability is close to such value, while false positives are extremely rare to find because, in the worst case, only 4.5% of the clams are infected. Finally, it must be commented that the false negative probability of the classifier is at most 0.193. Such value is considered very low, specially if one recalls that, at most, 4.5% of the samples are infected.

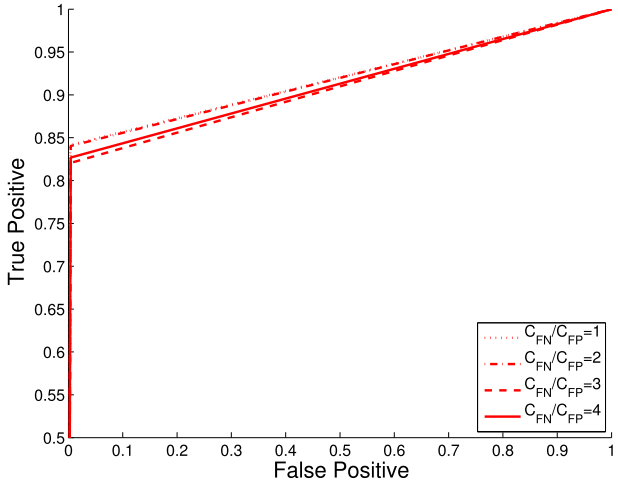
In order to specify a single operating point for the classifier, we consider the fact that in our problem false positive and false negative errors have different consequences. Note that it is key to minimize the number of false negative errors because they may result in unfordable economical and reputational costs associated with a final client finding a parasite in a clam. Thus, from Fig. 7c) we can upper bound the false negative probability to 0.150, achieving an average classification probability of 0.989 when the misclassification cost rate is 1.8 and the prior probabilities are 0.955 and 0.045 for classes 0 and 1, respectively.

3.3. Classification results in a Chilean seafood company

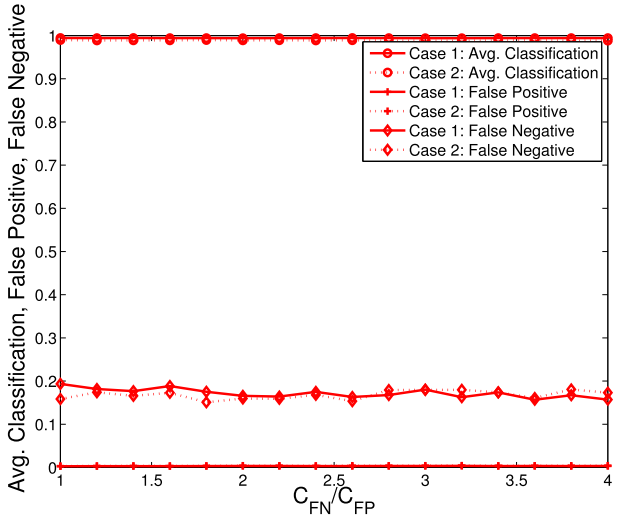
Using these parameters, a final classification test was conducted at a local seafood company using an additional sample set consisting of 100 infected clams and another 100 parasite free clams. For this new set the confusion matrix was: (i) True negative 48; (ii)



(a)



(b)



(c)

False positive: 52; (iii) False negative: 2; and (iv) True positive: 98. Note that, regardless of the lower average classification accuracy (73%), the false negative rate was only 2%, at expense of a high false positive rate: 52%.

Finally, it must be commented that, this last sample set was further examined and some unexpected conditions not registered in the feature database were found. For example, the clam stomach region overlaps the zone where the parasite usually hides. When the clam stomach is full of food, tissue becomes more opaque than usual making the classifier to identify such clam as infected. Another unexpected condition found was that, in some cases, the parasite was hiding deeper inside the clam, making the detection process nearly impossible for the system.

4. Conclusions

A machine vision system that automatically detects the presence or absence of the parasite *E. magellanica* inside the clam *M. edulis* was developed by using a transillumination system, a camera in the VIS spectral range, and an online flattening system. This innovative system is able to acquire uniform transilluminated clam images. The transillumination technique, together with the flattening system, was able to generate sharp clam images, in which the parasite was notably enhanced as a dark zone. The feature database and the classifier created in this work have shown that spectral, spatial, geometrical, and biological information can be effectively used as features for the parasite detection process. In addition, a novel and key step used in the classification process is the design of a self-generated spatial reference system that accounts for clam rotations. The system classification accuracy is controllable and can be selected according to the user needs. For example, an average classification accuracy of 98% can be achieved for the identification of infected clams. Finally, for the seafood industry, this technology in the VIS spectral range is economically feasible since it represents lower costs as compared to X-rays or hyperspectral imaging systems aiming similar goals.

Acknowledgments

The authors gratefully acknowledge Mr. Adan Lagos for manufacturing the machine and Mg. Miguel Soto for his assistance in this work. The authors would also like to thank the collaboration of Pesquera Tubul. The authors acknowledge the financial support from the INNOVA Biobio Grant 08-PC-S1292, the Center for Optics and Photonics PIACONICYT PFB0824, and CONICYT Ph.D. scholarship.

References

Breiman, L., Friedman, J., Stone, C.J., Olshen, R., 1984. Modern Multivariate Statistical Techniques. Chapman and Hall/CRC.

Coelho, P., 2011. Hyperspectral Study on *Mulinia edulis* Clam and *Edotea magellanica* Parasite for Optical Detection. Universidad de Concepción, Chile. Ms. Thesis (In Spanish).

Coelho, P.A., Soto, M.E., Torres, S.N., Sbarbaro, D.G., Pezoa, J.E., 2013. Hyperspectral transmittance imaging of the shell-free cooked clam *Mulinia edulis* for parasite detection. J. Food Eng. 117 (3), 408–416.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, second ed. John Wiley and Sons, Inc., New York.

Gonzalez, M., Jaramillo, E., 1991. The association between *Mulinia edulis* (Mollusca, Bivalvia) and *Edotea magellanica* (Crustacea, Isopoda) in southern Chile. Rev. Chil. Hist. Nat. 64, 37–51.

Fig. 7. Performance of the binary decision tree classifier. ROC curves, parameterized by the misclassification cost rate, for: (a) Case 1; and (b) Case 2. (c) Average classification probability, false positive probability, and false negative probability, as a function of the misclassification cost rate, for cases 1 and 2.

Heia, K., Sivertsen, A.H., Stormo, S.K., Elvevoll, E., Wold, J.P., Nilsen, H., 2007. Detection of nematodes in cod (*Gadus morhua*) fillets by imaging spectroscopy. J. Food Sci. 72 (1), E011–E015.

Izenman, A.J., 2008. Modern Multivariate Statistical Techniques. Springer, Philadelphia, Pennsylvania.

Mathiassen, J.R., Misimi, E., Bondo, M., Veliyulin, E., Ostvik, S.O., 2011. Trends in application of imaging technologies to inspection of fish and fish products. Trends Food Sci. Technol. 22 (6), 257–275.

Mery, D., Lillo, I., Loebel, H., Riffo, V., Soto, A., Cipriano, A., Aguilera, J.M., 2011. Automated fish bone detection using X-ray imaging. J. Food Eng. 105 (3), 485–492.

Nollet, L., Toldra, F., 2010. Seafood and Seafood Products Analysis. CRC Press.

Perez-Aleman, P., 2005. CLUSTER formation, institutions and learning: the emergence of clusters and development in Chile. Ind. Corp. Change 14 (4), 651–677.

Sivertsen, A.H., Heia, K., Stormo, S.K., Elvevoll, E., Nilsen, H., 2011. Automatic nematode detection in cod fillets (*Gadus Morhua*) by transillumination hyperspectral imaging. J. Food Sci. 76 (1), S77–S83.

Soto, M., Coelho, P., Soto, J., Torres, S.N., Sbarbaro, D.G., 2012. Automated parasite detection in clams by transillumination imaging and pattern classification. In: Proceeding of SPIE 8300, Image Processing: Machine Vision Applications V. San Francisco, California.

Tao, Y., Ibarra, J.G., 2000. Thickness - compensation X-ray imaging detection of bone fragments in deboned poultry – model analysis. Trans. ASABE 43 (2), 453–459.

Wold, J., Westad, F., Heia, K., 2001. Detection of parasites in cod fillets by using SIMCA classification in multispectral images in the visible and nir region. Appl. Spectrosc. 55 (8), 1025–1034.