Govocitos: a software tool for estimating fish fecundity based on digital analysis of histological images

J.M. Pintor[b], P. Carrión[b], E. Cernadas[a,*], E. González-Rufino[b], A. Formella[b], M. Fernández-Delgado[a], R. Domínguez-Petit[c], S. Rábade-Uberos[c]

[a] *CITIUS: Centro de Investigación en Tecnoloxías da Información da USC, Campus Vida, 15872, Santiago de Compostela, Spain*
[b] *Universiade de Vigo, Departamento de Informática, Campus Universitario As Lagoas, Ourense (Spain)*
[c] *Department of Fisheries Ecology, Instituto de Investigacións Mariñas-CSIC, Vigo (Spain)*

Abstract

To estimate productivity of a fish stock, the precise determination of fish fecundity is essential. The stereological method accurately estimates fecundity from histological images of a fish gonad. Traditionally, the fecundity is estimated using the stereological method, which overlays an hexagonal grid on the histological image and counts the number of grid points associated to each oocyte category (oocytes are reproductory cells) and the number of oocytes in each category. This process is done manually often using off-the-shelf software, but it is very time-consuming, requires specialized technicians, and does not allow to review the calculations. In this paper, we describe and evaluate the software Govocitos, which offers an easy and automatic way to estimate fecundity using the stereological method. Govocitos contains a module to automatically detects the matured oocytes in the slice (nearly 80%

*CITIUS: Centro de Investigación en Tecnoloxías da Información da USC, Campus Vida, 15872, Santiago de Compostela, Spain. E-mail: xoxe.cernadas@usc.es

of oocytes are correctly detected) and a module to automatically classify the oocytes according to the presence/absence of nucleus (with 84% of accuracy) and to three development stages (with 87% of accuracy). It also provides a user friendly GUI that allows the experts to modify the outlines and classifications of oocytes, to calculate diameters, areas and roundness, to build diameter frequency histograms, to count the points and objects inside the grid, to estimate partial and potential fecundity and to export the data to files and into a database. In addition, Govocitos provides the possibility of varying grid characteristics, it can be trained to work with different species and it allows to check and supervise the calculations whenever needed including in a later point in time. Govocitos is a free software that can be downloaded from `http://lia.ei.uvigo.es/daeira/software/govocitos`.

Keywords:
Histological image, fish fecundity, computer vision, edge detection, object recognition, texture analysis, classification.

---

1. Introduction

Fecundity is one of the most important parameters for estimating productivity of a fish stock, and thus, of interest to fishery scientists as a critical parameter of stock assessment, as well as a basic aspect of population dynamics, studied by Lasker (1985) and Hunter et al. (1992). To assess stock reproductive potential and its impact on recruitment variability (replacement rate of the population), it is necessary to study oocyte development

dynamics and estimate both the potential fecundity and the resorption rate (atresia) of oocytes in order to estimate the egg production that can be used as an indicator of stock productivity. Beside, Katsukawa (1997) states that it is a reference point for fisheries management to develop a precautionary approach to sustainable fisheries.

Determining fish fecundity in a precise and unbiased way is essential for stock assessment. Therefore, in the last decades many efforts have been undertaken to develop simpler, faster and more precise and accurate methods. The most common and widespread method to estimate fecundity is the gravimetric method, proposed by McGregor (1922) and Raitt (1933), that uses whole mounts samples of gonad tissue. Begenal and Braum (1978) summarizes most of these methods based on counting the number of mature oocytes in whole mounts preparations, assuming that all the oocytes larger than a certain diameter threshold are mature and recruited to be released. It is a basic, simple and inexpensive technique, however, it is not possible to determine the maturity stage of oocytes with the naked eye, especially at the beginning of the maturation process. This adds uncertainty to the fecundity estimates. Moreover, the gravimetric method can be a tedious and very time consuming procedure, especially for species with asynchronous oocyte development that simultaneously present a wide range of oocyte diameters within the ovary.

Assuming the basics of the gravimetric method and with the evolution of computerized image analysis,Witthames and Walker (1987) proposed alter-

native methods to estimate fecundity. Emerson et al. (1990), taking advantage of the precise classification of oocytes maturity stage from histological images under a microscope, developed a stereometric method based on the stereological approach proposed by Weibel and Gómez (1962) and Weibel (1979), that allows fecundity estimation guaranteeing at the same time the selection of mature oocytes by histology for fecundity estimation.

Stereology relates three-dimensional parameters (i.e., the volume) that define a structure with two-dimensional measures (i.e., the area) obtained from sections of this structure. Estimating the number of oocytes in each category (e.g., *previtelogenic*, *cortical alveoli*, *vitelogenic*, *atretic* or *hydrated*) and their relative area from histological sections is routinely done in many laboratories. Advantages and disadvantages of these methods are enumerated by Murua and Saborido-Rey (2003). Different interesting parameters can be estimated from a single sample using stereology: the spawning status of female, the atresia intensity and prevalence, the potential fecundity, the oocyte size distribution and maturation dynamics. For applying stereology to fecundity estimates some assumptions must be made — spherical shape of oocytes and homogeneous distribution of oocytes within the ovary — and correction factors have to be applied to adjust the oocyte frequency depending on their size (i.e., probability of being included in the histological section), because larger oocytes have a higher probability of appearing in the histological section than smaller ones. To estimate the relative area occupied by each oocyte category, usually a hexagonal grid of points is used. In

this grid, each point has the same associated area, so, counting the number of points that contact oocytes of one category and the number of oocytes of this category on the slide, the relative area and number of oocytes can be estimated and extrapolated to the entire ovary. Fine grids with a high number of points (i.e., small associated area) improves the accuracy of estimates but increases the time of processing the sample, because the counting is commonly done manually. Thus, there exists a trade-off between accuracy and time consumption of the method.

On the other hand, the oocytes' maturity stage has to be classified precisely, which requires experienced technicians and specific training for each species under study. The process is often not reproducible and different results are obtained by different technicians.

Applications of computer vision are recently playing a more and more important role in fisheries and fish farming industries. Mathiassen et al. (2011) provides a review of research and industrial solutions that use imaging technologies for the inspection of fish and fish products related to post-harvest operations like: automatic sorting and grading, estimate the chemical composition (e.g., salt, fat, proteins), weight estimation. Hu et al. (2012) use computer vision for fish diseases diagnose. Zion (2012) review the state of art of computer vision technologies in aquaculture, in which the main inspection task are: counting fishes, size measurement and mass estimation, gender identification and quality assessment, species and stock identification and monitoring welfare. The inspection tasks are employed in all stages of

production from hatcheries to harvest. All the above mentioned computer vision approaches are focused on the analysis of images of the whole fish or a fish product.

However, despite the progress of computing tools based on image analysis, the image analysis with commercial off-the-shelf software for fecundity estimation is still time-consuming, requires high technological imaging equipment and specialized technicians, and does not allow to review the calculation process. With the aim of giving an answer to all the needs and difficulties, we propose a multi-platform software (called Govocitos) that includes a specifically designed algorithm for the automatic recognition of oocytes, and that allows their maturity stage classification from histological images as well as fecundity estimation according to the stereological principle. This software fulfills the following requirements:

1. Provides a friendly graphical user interface (GUI) to interactively work with histological images.
2. Uses an algorithm to automatically recognize and classify the mature oocytes in a histological image.
3. Estimates automatically the oocyte diameter distribution facilitating the study of fish reproductive biology (even for less experienced technicians).
4. Provides data sharing among researchers from different laboratories and allows to review the results.

5. Works faster and more precisely than manual or semi-automatic traditional methods.
6. It is accurate and trustworthy.

Govocitos is being used by the Instituto de Investigacións Mariñas of CSIC (Spanish Research Center) since 2012. A version of Govocitos running locally in a personal computer can freely be downloaded from `http://lia.ei.uvigo.es/daeira/software/govocitos`. This paper is organized as follows: Section 2 first describes the architecture of the proposed system and then analyzes every module of the system and the underlying algorithms to automatically recognize and classify the mature oocytes in the histological image. Section 3 discusses the statistical evaluation of the automatic detection and classification algorithms and, finally, Section 4 draws the conclusions and proposes future work.

2. Methods

2.1. System architecture

The system is structured in three logical layers: the graphical user interface (GUI) layer, the application logic layer and the data storage and communication layer. The implementation of the layers uses currently four modules. Figure 1 shows the architecture of the system. The GUI provides a user friendly interface to draw, interact and visualize data (see Section 2.5). The application layer refers to the data processing of inputs (see Sections 2.2,

2.3 and 2.6). The first two layers are enclosed in the Govocitos software that runs on a personal computer. The third layer is implemented on a server machine with Internet connection. Govocitos is a multi-platform software (Windows/Linux) written in the C++ programming language. The division into modules is:

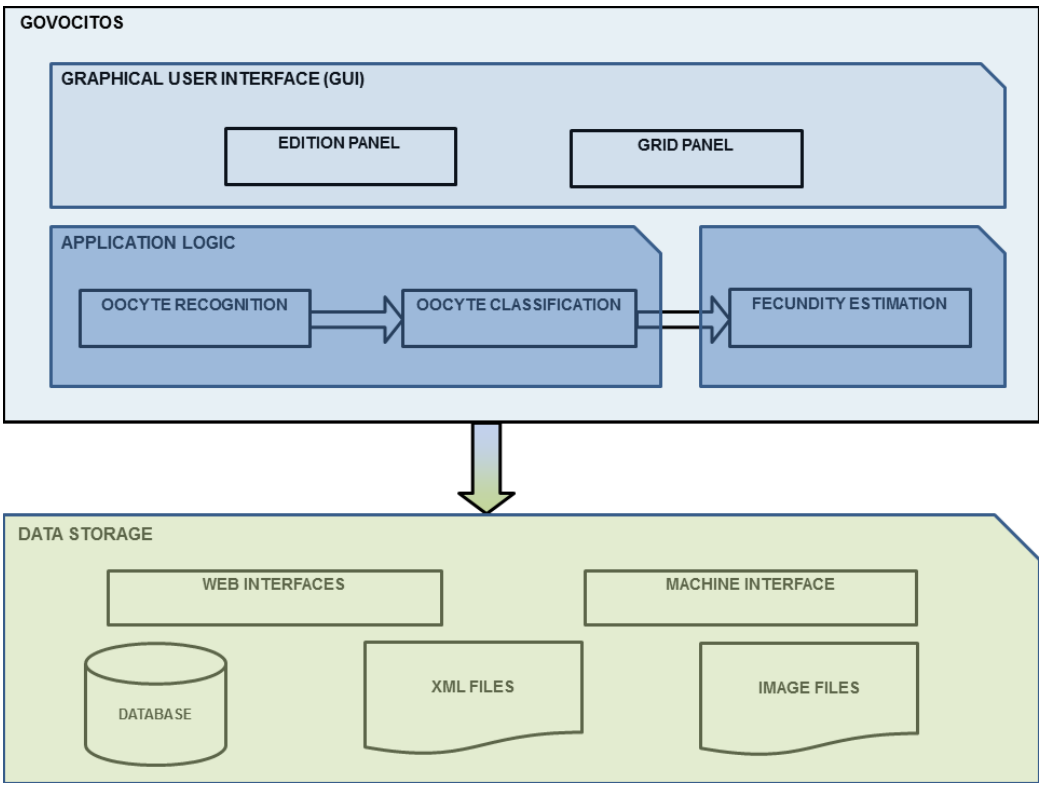

Figure 1: System architecture from the viewpoint of the developers.

GUI: an intuitive and friendly graphical user interface to manually draw the outline of matured oocytes and to specify its class (state of development and presence/absence of nucleus).

Unsupervised detection and classification: these modules process the histological image and automatically provide the outlines of matured oocytes and its classification. Afterwards, the experts can modify and supervise these results using the graphical interface.

Fecundity: it estimates the fecundity from the matured oocytes being recognized and classified by one of the two first modules and some information stored in the database like calibration, ovarian volume, fish weight, etc.

Data storage: the information required and calculated by Govocitos is supported by either local or web-based databases and XML files.

Since these modules are independent among them, new modules may be added or existing ones may be modified easily, facilitating the expansion, the improvement and the usage of the system for other purposes.

2.2. Oocyte recognition

From the point of view of computer vision, the recognition of matured oocytes in histological images is a segmentation process in which an image is divided in objects (mature oocytes) and background (rest of image). Many image segmentation methods have been proposed in the last four decades, being specially relevant the works of Sonka et al. (1999), Muñoz et al. (2003) and Papari and Petkov (2011)). These methods may be grouped in four families: thresholding, region-based, edge-based, and mixture methods. Some techniques enclosed in each group were tested by us in Alén et al. (2006), Anta

et al. (2007) and Cernadas et al. (2008)). The main conclusion at which we arrived is that histological images are textured and very complex, and that conventional methods are not robust and general enough to work satisfactory with different fish species and illumination conditions. Consecuently, for the segmentation of histological microscopic fish images we propose a multi-scalar Canny (MSC) filter followed by an information fusion step to remove edges detected by different runs of the Canny filter. Our method has two variants: an unsupervised and a supervised one, sharing their first step. There is a consensus in the scientific community about the superior performance of the Canny edge detector proposed by Canny (1986), in relation to other common differential operators like Sobel, Prewitt, Roberts, Laplacian, Laplacian of Gaussian (LoG) and difference of Gaussians (DoG).

The Canny edge detector is based on DoG performing the following steps: i) smooth the input image using a Gaussian kernel with a given smoothing parameter σ; ii) differentiate using a first difference operator; iii) perform the non-maxima suppression process, which finds local maxima in the direction perpendicular to the edge; and iv) perform the thresholding with hysteresis, using two thresholds (low and high, denoted by T_L and T_H respectively) in order to avoid spurious responses. The filter performance depends critically on the parameter values (thresholds and smoothing parameter σ).

The pixel neighborhood in the Canny filter is controlled by σ and its optimal value depends on the type and size of objects of interest in the image. High σ values smooth the inner regions of cells which are often textured.

When cells are very close to each other or edges are weak, high σ values blur interesting edges as well. In the latter case, finer scales (i.e., low σ values) must be used, although the Canny filter will detect more noisy edges. For a given σ value, the variation of thresholds T_L and T_H controls the strength in the gradient image. High thresholds detect the strongest edges (normally true edges), but they often miss other true edges. So, low thresholds are also necessary to detect weak true contours.

Let I be a histological image, S be the number of smothing values, $C_S = \{\sigma_i \mid i = 1, \ldots, S \text{ and } \sigma_i > 0.1\}$ be the set of scales, and let $C_T = \{(T_L^j, T_H^j) \mid j = 1, \ldots, T, \text{ and } 0 < T_L^j \leq T_H^j < 1\}$ be a set of threshold pairs. The MSC filter is applied to I in several steps. For all $\sigma_i \in C_S$: 1) smooth I with Gaussian filter using σ_i to obtain I_{Si}; 2) apply gradient operator to I_{Si}, providing the edge map image I_{Si}^M; 3) perform non-maxima suppression on I_{Si}^M; and 4) perform edge detection with hysteresis using all the threshold pairs in C_T. The threshold values initially in the interval $(0, 1)$ are scaled to an absolute threshold pair for the pixel values in the image, usually in the interval $(0, 255)$, so that their final values depend on the histogram of gradient magnitudes in the image (magnitude of I_{Si}^M). The output of the MSC filter is a set E_I with all the detected edges. Since the minimum diameter, $d_{\min}$, of matured oocytes is an intrinsic attribute of each fish specie, which can be specified by technicians, we apply a size filter to E_I removing all edges whose convex hull diameter is lower than $d_{\min}$.

In the *unsupervised algorithm*, the input is the histological image I and

the output are the detected oocytes. Let $E_I = \{e_j^I \mid j = 1, \ldots, M\}$ be the M edges detected by the MSC filter. We know that the oocytes of interest should have an elliptical shape for the majority of cells. We select a subset $E_I^R \subseteq E_I$ comprising the edges whose convex hull roundness is lower than $r_{\max}$, $E_I^R = \{e_j^I \in E_I \mid r(e_j^I) < r_{\max}\}$. Finally, we apply an *overlap test* in E_I^R to select the edges which are candidates to represent an oocyte: 1) if the centroid of an edge i lies inside the edge j and vice versa, the two edges are considered candidates for the same oocyte; and 2) if the centroid of i lies inside the edge j but the opposite is not true, edge i is considered a noisy oocyte. Among the candidate edges for the same oocyte, one of them is selected randomly.

In the *supervised version*, the output are also the detected oocytes, but the input also includes one point (x_o, y_o) inside a true oocyte, which is previously marked by the expert in the histological image using the Govocitos GUI. An edge $e_j^I \in E_I$ is a candidate to be the detected oocyte, if (x_o, y_o) is inside e_j^I. If there is more than one candidate, we randomly choose one of them.

2.3. Oocyte classification

Once the outlines of matured oocytes are recognized in the histological image, the software will classify the oocytes into classes according to the presence or absence of the nucleus and into states according to the developmental stage, i.e., *cortical alveoli*, *hydrated*, *vitelline*, and *atretic*. The work of González-Rufino et al. (2013) tested the performance of both classifications

using the most popular texture features and classifiers in the literature. The combination of a Support Vector Machine classifier (SVM) with first order statistics of the color RGB image and the uniform rotation invariant Local Binary Patterns (LBP) of a grey scale version of the image was included in Govocitos to provide the best trade-off between computation time and efficiency for both types of classifications. The texture feature vector has only 15 inputs.

The classifiers which discriminate between presence and absence of the nucleus and discriminate among the developmental stages must be trained before starting the classification process. To train a classifier, the system randomly selects a number of patterns (previously labeled manually by users). The patterns are stored in the database. Specifically, 1000 patterns (oocytes) including exemplars of each class (with and without nucleus for the classification according to presence/absence of nucleus, and *cortical alveoli*, *hydrated*, *vitelogenic* and *atretic* according to development stage). The maximum size of 1000 patterns allows to avoid slow training times. Given this information, the GUI of Govocitos includes a functionality to train the classifiers for each species. The classifiers can be trained as many times as needed, because a larger number of training samples may improve the classification accuracy.

2.4. Data storage

All data in the system are stored in an RDBMS (Relational Database Management System) using the SQL (Structured Query Language) for inter-

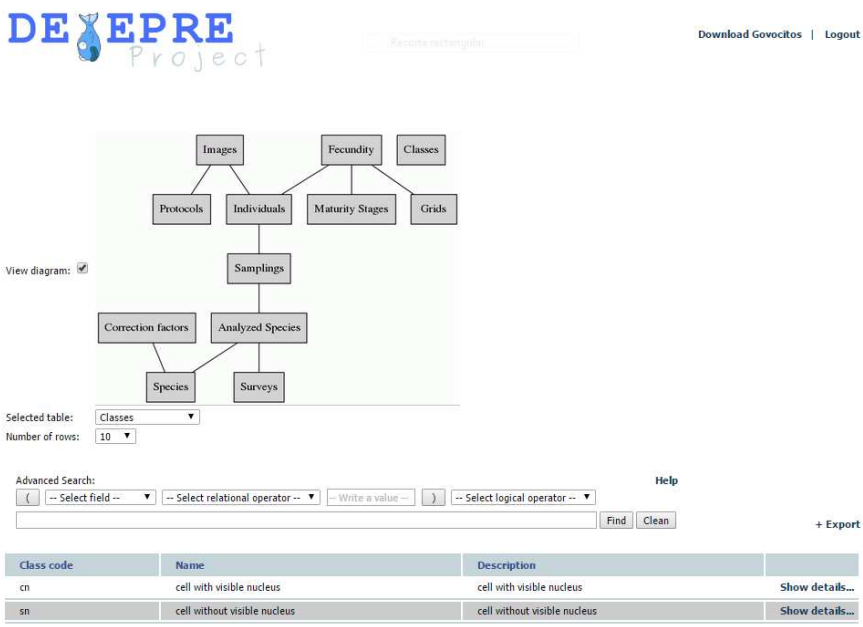

Figure 2: Screenshot showing the web interface to interact with the database.

action, XML (eXtensible Markup Language) files containing image attributes and the proper image files (usually in TIFF format, but other formats are supported as well). Access to the database is provided with a user interface and a machine interface. The expected way of access for users is the web interface, as it can be seen in the screenshot of Figure 2. The machine interface works with common network protocols. All ways of access have been designed to be both secure and easy-to-use. The image files contain the digitalized histological samples. The XML files store non-structured information like outlines and classes for each detected or annotated mature oocyte in the histological image. There is just one XML file per image. The RDBMS also stores authentication information.

2.5. Graphical user interface

The flowchart describing the functionality of Govocitos is shown in Figure 3. The graphical user interface offers different, so called, *views* depending on the current process the user carries out. Figure 4 shows the main interface window of Govocitos.

A typical working session for a user might be: first, the user opens all the processed images of an individual fish, just clicking in the specimen identification code to estimate fish fecundity, or a single colored histological image to recognize and classify its mature oocytes automatically (see Figure 3). Once a histological image is selected, the user performs the unsupervised detection or goes to the editing view to draw the outlines of target objects (i.e., mature oocytes). In this view, the user can delete or modify noisy/erroneous detected objects. Then, the user marks those oocytes that have not been recognized by the automatic algorithm. Re-starting with this new clue, the supervised algorithm (see Section 2.2) will try to provide the outlines of the annotated oocytes as well. After this, the user could again supervise the results using the editing tools. Figure 5 shows the editing view of Govocitos. Once the outlines of all mature oocytes are correct, Govocitos automatically classifies them into their class with or without nucleus and development stage. Classes and stages can be manually modified or re-assigned in the edition panel. Govocitos computes and displays the mean diameters and areas for the selected classes and stages. It also stores the data in files for further analysis, including the outlines and classifications of the matured oocytes in

the histological image.

The fecundity view encloses the functionality to estimate the individual fish fecundity. To perform this action, the XML files containing the outline and classification of oocytes for all images associated to a single fish had to be previously created, i.e., at least one image has to be processed and saved in the database. The fecundity is estimated using the Equation 1 with a previously overlaid hexagonal grid on the image. Traditionally, a 168-point grid covering only the central area of the image has been used to estimate fecundity (see Figure 6). However, Govocitos allows to configure the grid size, i.e., the number of points and their inbetween distance. Moreover, Govocitos allows to estimate the fecundity based on the entire image, i.e., considering all pixels as the Weibel grid area. Govocitos estimates the partial areas of each oocyte in all maturity stages counting all pixels occupied by each of them.

Independently of the grid size or the number of points, Govocitos estimates fecundity almost instantaneously. Figure 7 shows the fecundity view of Govocitos, which visualizes histograms of the distribution of diameters by class and stage, estimates the fecundity and exports all data in files and/or in the database.

2.6. Fecundity estimation

Stereology fundamentals have been applied to estimate the number of atretic, developing and hydrated oocytes, following the method developed by

Emerson et al. (1990), which is based on the Delesse principle and uses the equation proposed by Weibel and Gómez (1962) and presented by Murua et al. (2003). Specifically, the fecundity F is calculated by the following equation:

$$F = O_v \frac{k}{\beta} \frac{\sqrt{N_a^3}}{\sqrt{v_i}} \tag{1}$$

where β is the ratio between the longest and shortest axis of oocytes transected through the nucleus, O_v is the ovary volume, N_a is the number of oocyte transsections per unit area, v_i is the partial area of oocyte in the histological section, and k is a size distribution coefficient.

2.7. Sample preparation and image acquisition

Ovaries at different maturity stages from two fish species were selected to develop and test Govocitos: European hake (*Merluccius merluccius*) and pouting (*Trisopterus luscus*). One ovary lobe was fixed in 4% buffered formaldehyde and one slice per ovary of all females was embedded in paraffin. Then, 3 μm sections were cut and stained with haematoxylin and eosin for later microscopical analysis. Fecundity estimates were based on 4 microscope fields per ovary section using a Leica DMRE microscope connected to a Leica DFC320 FX digital video camera with 3.3 Mpixel and a spatial resolution of 1.09 μm. The exposure time and color balance are set automatically.

Figure 4 shows a typical image in Govocitos where the mature oocytes are already detected and classified. Different colors correspond to different oocyte development stage: *cortical alveoli* (black), *hydrated* (blue), *vitelline*

(red), and *atretic* (green); and the type of line corresponds to the oocyte class: continuous line for cells with nucleus and dashed lines for cells without nucleus.

3. Results

A network version of Govocitos was installed at the Instituto de Investigacións Mariñas (IIM) of the State Agency Consejo Superior de Investigaciones Científicas (CSIC) in 2012 with the aim to evaluate the software operating in a real environment. The IIM biologists used Govocitos in their daily work to calculate the fecundity, for which they need to recognize and classify all matured oocytes in the histological images. Their operations in the Govocitos GUI were logged in files for a later statistical evaluation. The following subsections summarize the results achieved for the automatic detection and classification algorithms, the issues related to the grid size used to estimate the fecundity, and the issues related to the time needed by experts to analyze each sample or entire fish.

3.1. Statistical evaluation of the detection and classification algorithms

A total of 61 images belonging to 8 fishes of European hake (31 images) and 8 to pouting (30 images) were analyzed, detecting and classifying 2405 oocytes (1186 for European hake and 1219 for pouting). The number of mature oocytes per image varied significantly ranging from 18 to 73 oocytes for European hake (with a mean value of 38 oocytes) and 23 to 70 for pouting (with a mean value of 40). Technicians performed the following tasks to

detect and classify the mature oocytes in every histological image using Govocitos: 1) load an image into Govocitos; 2) run the non-supervised oocytes detection algorithm; 3) delete the false positive (erroneous) oocytes; 4) run the supervised oocytes detection algorithm by clicking inside unrecognized oocytes; then the algorithm provides their outlines; 5) complete or modify manually the detection of oocytes; 6) run the automatic classification algorithm to discriminate among oocytes with/without visible nucleus and their development stage; and, finally, 7) modify manually the label of misclassified oocytes. All these steps were done with the GUI of Govocitos.

Table 1: Accuracy provided by Govocitos for oocytes detection and classification in histological images (in percentages).

	All species		European hake		Lane snapper	
	mean(std)	min-max	mean(std)	min-max	mean(std)	min-max
Detection algorithm						
Non-supervised	63.6(14.4)	19.5-88.5	61.6(9.9)	40.9-77.5	65.6(17.8)	19.5-88.5
Supervised	80.0(13.6)	27.0-98.0	79.7(9.3)	56.8-95.9	80.3(17.1)	27.0-98.0
False positive	19.7(14.8)	0.0-60.0	19.8(15.1)	0-51.5	19.7(14.7)	4.8-60.0
Oocyte classification						
With/without nucleus	83.8(11.3)		78.1(12.2)		81.4(18.6)	
Development stage	87.1(14.6)		89.4(6.8)		92.8(4.9)	

std: standard deviation, min: minimum, max: maximum.

For the statistical evaluation of the detection algorithms, we consider that an oocyte is correctly detected if the user does not modify the outline provided by the detection algorithm. We consider that an oocyte is a false positive whenever the user deletes the oocyte outline provided by the detection algorithm. We count the number of oocytes correctly detected after the non-supervised and the supervised detection algorithm (Steps 2 and 4 above, respectively) and the number of false positives (Step 3). For the classifica-

tion algorithm (Steps 6 and 7), we count the number of oocytes correctly classified by Govocitos.

Table 1 shows the results for both species. The classifier for each kind of classification was trained for each species using the same images as in González-Rufino et al. (2013).The detection algorithm does not need any training. The calibration parameters required are the minimum diameter of mature oocytes (100 μm for both species) and the spatial resolution of the image acquisition process (1.09 μm per pixels for both species). Although both parameters are not critical, including them decreases the number of false positives. The accurate detection rate is the ratio between the number of oocytes correctly detected and the total number of detected oocytes in the image. The false positive rate is the ratio between the number of oocytes detected erroneously and the total number of oocytes detected in the image by the software.

In relation to the detection algorithm, the mean number of oocytes correctly detected by the non-supervised algorithm is 63.6% and by the supervised version it is 80.0%. Hence, the supervised approach increases significantly the accuracy by approximately 16 points. There is a great variability among images for both species (about 9 points for European hake and 17 for pouting). Table 1 shows the minimum and maximum rates per image achieved for both species and steps of evaluation. Although the variability is higher for pouting (minimum correct detection of 27 and maximum of 98 for the supervised algorithm), it is still high for European hake (minimum of 56.8

to maximum of 95.9). We could not establish any correlation between the number of oocytes in the image and the correct detection rates. Visually, we can observe that the lowest detection accuracies correspond to images containing a high number of misshapen oocytes (e.g., hydrated oocytes) which makes the outline detection difficult, or atretic oocytes, which have the contour broken.

For some images, the automatic detection of the oocytes outline is almost perfect (with rates higher than 95%). Nevertheless, for some others the correct detection rate is rather low (only 27%). Therefore, we suggest that automatic detection of oocytes should be supervised by an expert before starting the fecundity calculation. Regarding the oocyte classification, the accuracy to classify the oocytes classes (with/without nucleus) is 83.8% on average for both species and 87.1% for development stage classification (joining the classes *viteline* and *atretic*). The performance is slightly higher for pouting (81.4% and 92.8% for classes and stages, respectively) than for European hake (78.1% and 89.4%, respectively).

3.2. Software validation

Validation is essential for any proposed software system. We have to ensure the correctness of the software comparing the fecundity results provided by both the traditional method and Govocitos (using the same parameters, e.g., a grid of 168 points). Already Dominguez-Petit et al. (2011) did a preliminary study comparing the fecundity estimated using grids of different

characteristics concluding that for the spatial resolution and species of this work, the fecundity estimates vary depending on grid characteristics (size, number of points and distance between points). The optimal level is reached for grids with interpoint distance of 30 pixels or lower, but the interpoint distance for the traditional grid of 168 points is 125 pixels for these images.

The estimation of the time needed to analyze each fish is really hard, essentially due to the performance variability of the automatic recognition algorithm among samples. The automatic recognition of oocytes takes less than one minute on a common personal computer for an image with 1550x2088 pixels. The automatic classification step is almost instantaneous. So, the time needed to analyze a sample is dominated by the time needed to modify and/or add the miss-detected and miss classified oocytes in the sample. Once the oocytes are detected and classified in the image, the diameters, areas and fecundity estimations are obtained almost instantaneously. As well, Govocitos integrates other processing tasks into a single program and it allows to review the intermediate results each time, hence, it avoids repetitions of complete analysis when the results are not adequate. Anyway, for most of the images the automatic computation of diameter, profile histograms and fecundity is adequate enough, alleviating the expert's daily work and making it easy to carry out fish reproductive studies by non-experts technicians.

4. Conclusions

Govocitos offers an easy and automatic way to estimate fish fecundity based on the traditional stereological method: use of a grid of points defined by the user, counting of points and objects inside the grid, estimation of stereological parameters, partial areas and volumes and estimation of potential fecundity and partial fecundity (for each development stage). Govocitos also classifies oocytes based on presence/absence of nucleus as well as developmental stage, calculates cell diameters, areas and roundness, builds diameter frequency histograms and exports data for later analyses, integrating all these tasks in a single application. In addition, Govocitos provides the possibility of varying the grid characteristics; even more, it allows using all pixels of the image as grid points increasing the accuracy of the calculation of partial areas of different objects within the image. Further, it allows checking and reviewing the calculations in every moment, automatic recognition and classification of oocytes and interactive supervision of the process by experts.

Govocitos includes a multi-scalar Canny filter to automatically detect the oocyte outline, achieving an accuracy of 64% in an unsupervised way, which increases up to 80% when the expert marks only one point on the unrecognized oocytes using the GUI. The classifier uses Support Vector Machines (SVM) combined with texture (Local Binary Patterns) and color features achieving an accuracy of 84% to discriminate between oocytes with or without nucleus and an accuracy of 87% to discriminate among three development stages (*cortical alveoli*, *hydrated* and *vitelline/atretic*).

It works both automatically and manually, or semi-automatically, as well as in local or network way. In summary, Govocitos is an easy-to-handle software (even for less experienced personnel) that facilitates laboratory routines for studying fish reproductive ecology for different fish species. It is free software that can be easily used after a short training period, offering the same or even better performance than other expensive and complex image analysis software tools.

Although some first tests using Govocitos with histological images of other laboratories provide good visual results, in future work, we plan to evaluate Govocitos more systematically in other real environments (i.e. in other laboratories) to test the robustness of the oocyte recognition and classification algorithms taking into account different sample preparations and adquisition conditions of the histological images. We also plan to extend Govocitos to other applications which require quantitative analysis on images. This analysis may eventually require to extend and adapt the detection and classification modules.

Acknowledgment

This investigation was partly supported by the Xunta de Galicia (regional government) project PGIDIT08MMA010402PR and partly executed at the Centro de Investigación, Transferencia e Innovación (CITI) en Ourense, Spain.

References

Alén, S., Cernadas, E., Formella, A., Domínguez, R., Saborido-Rey, F., 2006. Comparison of region and edge segmentation approaches to recognize fish oocytes in histological images. Lecture Notes in Computer Science 4142, 853–864.

Anta, P., Carrión, P., Formella, A., Cernadas, E., Domínguez, R., Saborido-Rey, F., 2007. Combining Region and Edge Information to Extract Fish Oocytes in Histological Images, in: 7th IASTED International Conference on Visualization, Imaging and Image Processing, pp. 82–87.

Begenal, T.B., Braum, E., 1978. Eggs and early life history, in: Method for assessment of fish production in fresh waters. Third edition. IBP Handbook No3. Blackwell scientific publication, pp. 165–201.

Canny, J.F., 1986. A computational approach to edge detection. IEEE Trans. on Pattern Analysis and Machine Intelligence 8, 679–698.

Cernadas, E., Carrión, P., Formella, A., Domínguez, R., Saborido-Rey, F., 2008. Recognize and classify fish oocytes in histological images, in: 8th IASTED International Conference on Visualization, Imaging and Image Processing, pp. 180–186.

Dominguez-Petit, R., Freire, J.M.P., Rábade, S., Fabeiro, M., Nieto, A.D., Carrión, P., Rufino, N., Cernadas, E., Fernández-Delgado, M., Dominguez-Vázquez, D., Saborido-Rey, F., Formella, A., 2011. New automatic soft-

ware tool to estimate fish fecundity based on image analysis, in: Fish Reproduction and Fisheries. Final Conference. Poster 2.4.

Emerson, L.S., Greer-Walker, M., Witthames, P.R., 1990. A stereological method for estimating fish fecundity. Journal of Fish Biology 36, 721–730.

González-Rufino, E., Carrión, P., Cernadas, E., Fernández-Delgado, M., Domínguez-Petit, R., 2013. Exhaustive comparison of colour texture features and classification methods to discriminate cells categories in histological images of fish ovary. Pattern Recognition 46, 2391 – 2407.

Hu, J., Li, D., Duan, Q., Han, Y., Chen, G., Si, X., 2012. Fish species classification by color, texture and multi-class support vector machine using computer vision. Computers and Electronics in Agriculture 88, 133–140.

Hunter, J.R., Macewicz, J., LO, N.C.H., Kimbrell, C.A., 1992. Fecundity, spawning, and maturity of female Dover Sole, Microstomus pacificus, with an evaluation of assumptions and precision. Fishery Bulletin 90, 101–128.

Katsukawa, T., 1997. Points of view: Introduction of spawning potential: improvement in the threshold management theory. Revie in Fish Biology and Fisheries 7, 2085–289.

Lasker, R., 1985. An egg production method for estimating spawning biomass of pelagic fish: Application to the Northern anchovy, Engraulis mordax. Technical Report 36. U.S. Dep. Comm.

Mathiassen, J.R., Misimi, E., Bond∅, M., Veliyulin, E., ∅stvik, S.O., 2011. Trends in application of imaging technologies to inspection of fish and fish products. Trends in Food Science and Technology 22, 257–275.

McGregor, E.A., 1922. Observations on the egg yield of Klamath river king salmon. Californi Fish and Game 8, 160–164.

Murua, H., Kraus, G., Saborido-Rey, F., Witthames, P., A.Thorsen, Junquera, S., 2003. Procedures to estimate fecundity of marine fish species in relation to their reproductive strategy. Journal of Northwest Atlantic Fishery Science 33, 33–54.

Murua, H., Saborido-Rey, F., 2003. Female reproductive strategies of marine fish species of the North Atlantic. Journal of Northwest Atlantic Fishery Science 33, 23–31.

Muñoz, X., Freixenet, J., Cufí, X., Martí, J., 2003. Strategies for image segmentation combining region and boundary information. Pattern Recognition Letters 24, 375–392.

Papari, G., Petkov, N., 2011. Edge and line oriented contour detection: State of the art. Image and Vision Computing 29, 79–103.

Raitt, D.S., 1933. The fecundity of the haddock. Scientific Investigations. Scotland Fishery Board 1, 1–40.

Sonka, M., Hlavac, V., Boyle, R., 1999. Image Processing, Analysis, and Machine Vision. International Thomsom Publishing (ITP).

564 Weibel, E.R., 1979. Steorological method: Practical methods for Biological
565 Morphometry. Academic Press.

566 Weibel, E.R., Gómez, D.M., 1962. A principle for counting tissue structures
567 on random sections. Journal of Applied Physiology 17, 343.

568 Witthames, P.R., Walker, M.G., 1987. An automated method for counting
569 and sizing fish eggs. Journal of Fish Biology 30, 225–235.

570 Zion, B., 2012. The use of computer vision technologies in aquaculture – A
571 review. Computers and Electronics in Agriculture 88, 125–132.

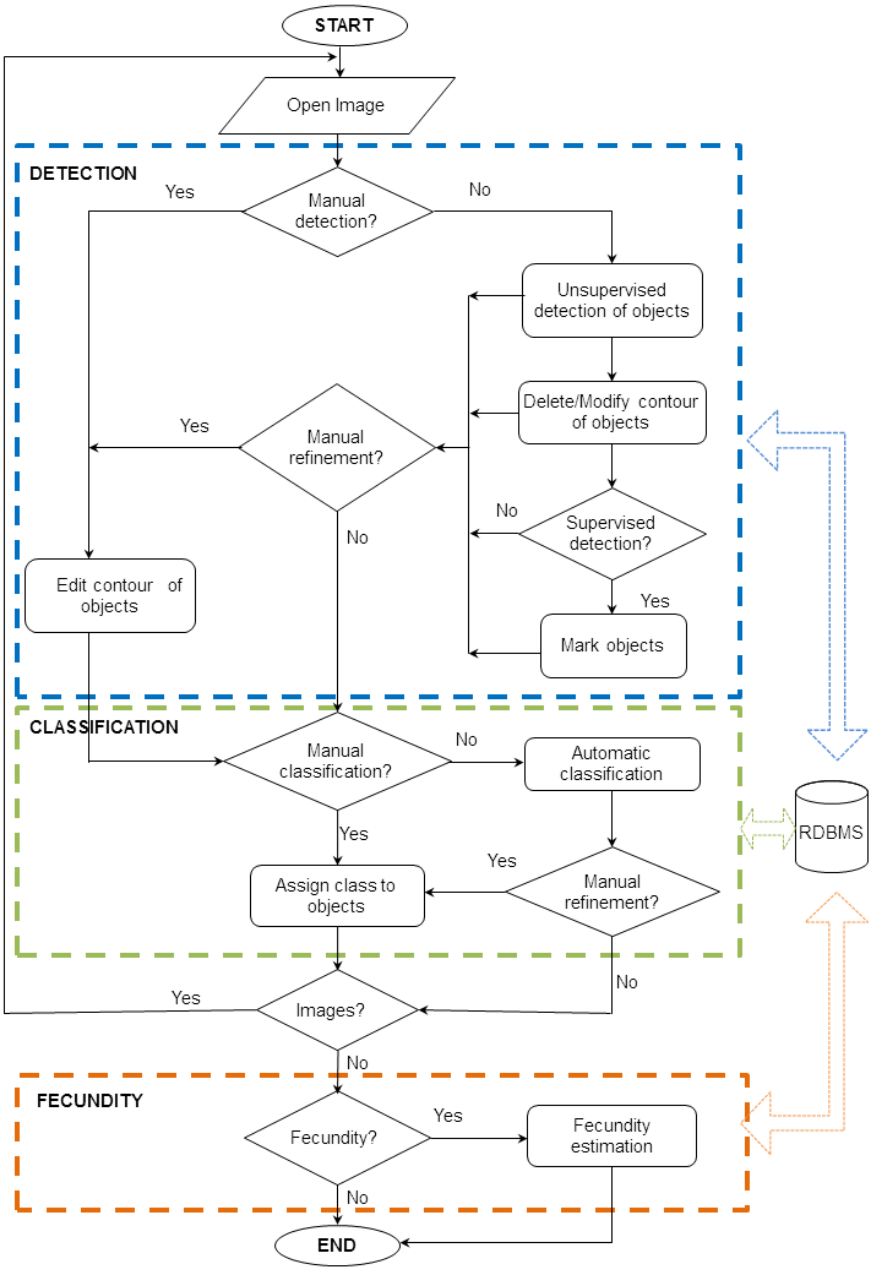

Figure 3: A flowchart containing the main task of Govocitos.

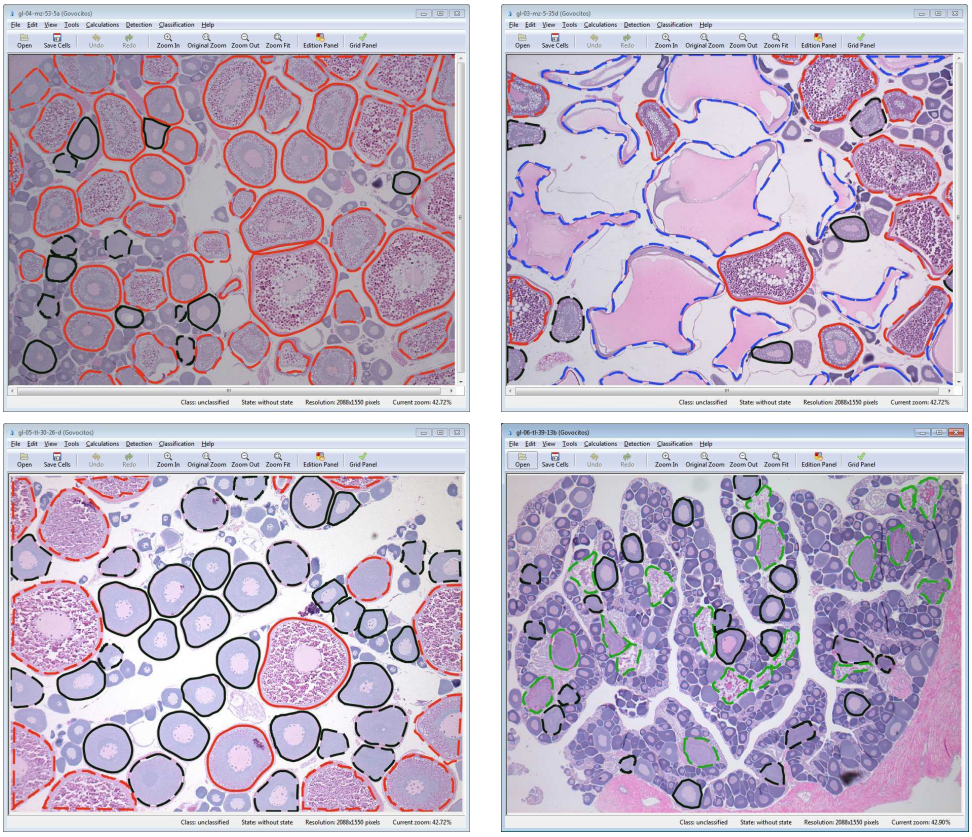

Figure 4: Govocitos software showing some typical histological images of fish ovaries from two species: European hake (upper panels) and pouting (lower panels). The true contours of mature oocytes are overlaid (the color and line type show the stage and class of each of them).

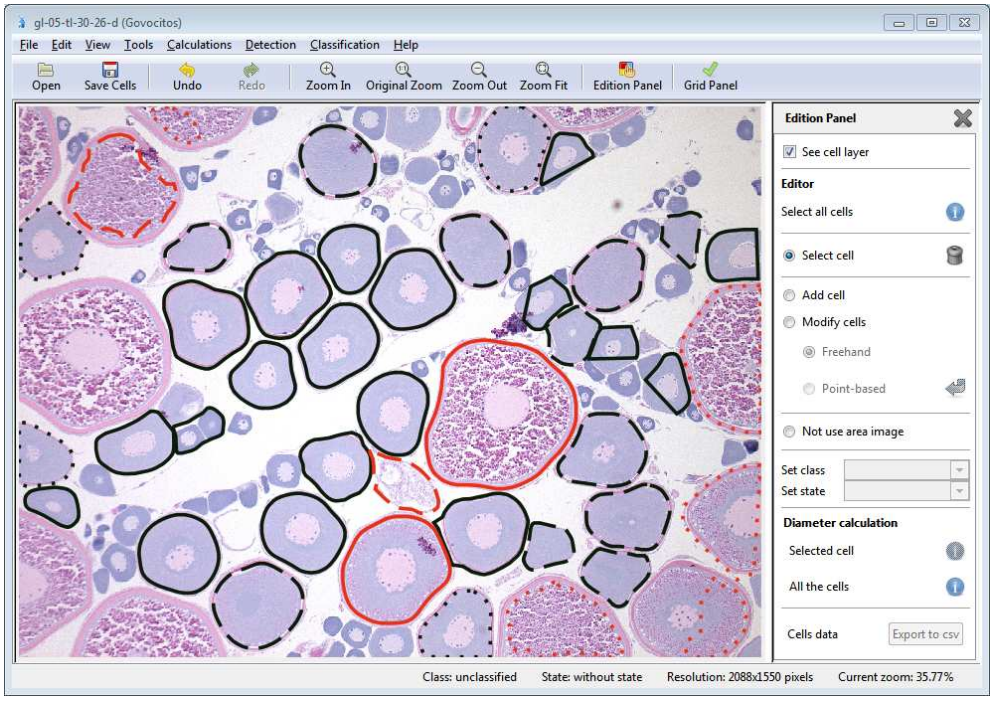

Figure 5: Govocitos software showing the editing view after automatic recognition and classification steps.

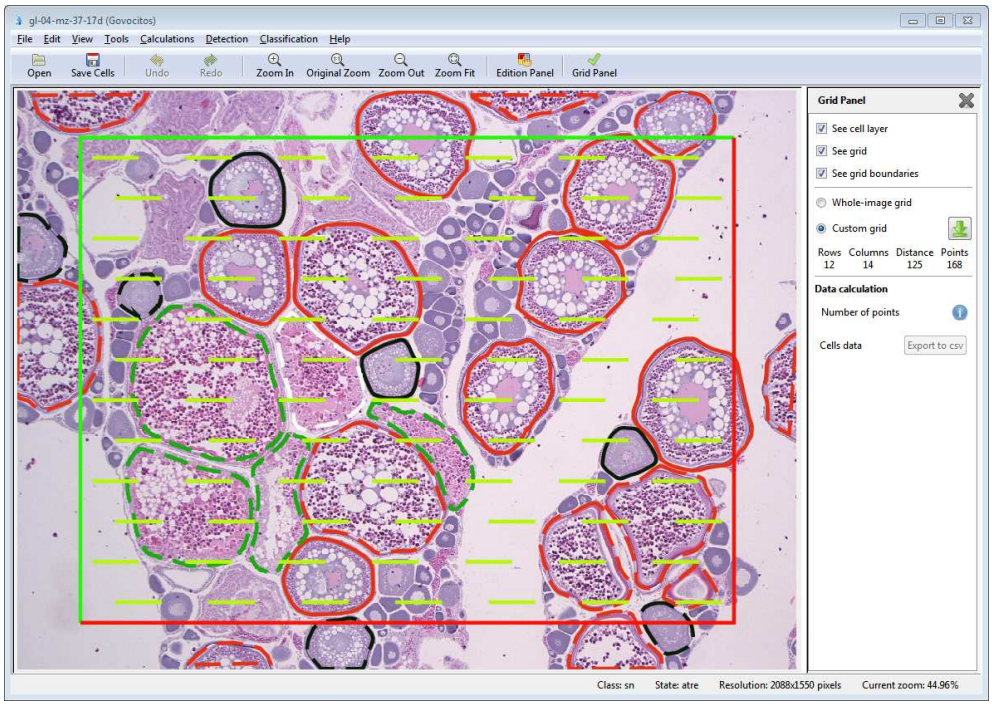

Figure 6: Govocitos software showing the grid view with the traditional grid of 168 points.

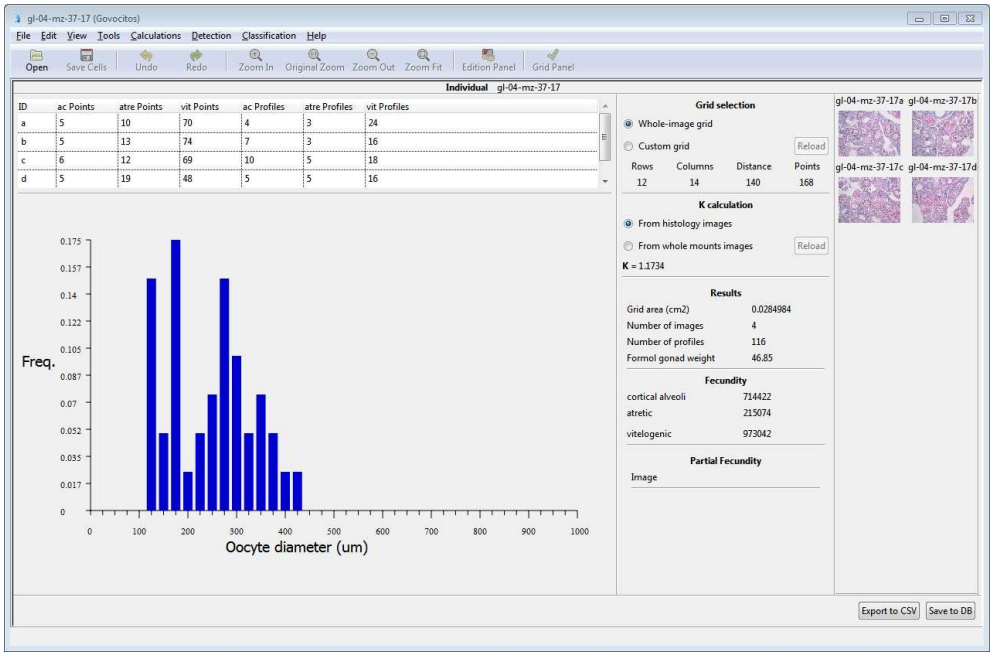

Figure 7: Govocitos software showing the fecundity view.