

Development of a New Index to Evaluate Zooplanktons' Gonads: An Approach Based on a Suitable Combination of Deformable Models

M. Ramiro Pastorinho¹, Miguel A. Guevara², Augusto Silva³, Luis Coelho³,
and Fernando Morgado¹

¹ Biology Department, University of Aveiro, 3810 – Aveiro, Portugal
{pastorinho, fmorgado}@bio.ua.pt

² Computer Sciences Faculty, University of Ciego de Avila,
Ciego de Avila 69450, Cuba
mguevaral@yahoo.com

³ IEETA, University of Aveiro, 3810 -193 Aveiro, Portugal
{lcoelho, asilva}@ieeta.pt

Abstract. *Acartia tonsa* was used as model to establish an index for oocyte maturity determination in zooplankters based in citometry and histochemical evaluation of gonadic masses. Biometry was performed using an ocular micrometer and nucleus/cytoplasm ratios were obtained characterizing each of the three identified stages: Immature, Vitellogenic and Mature. This paper presents a novel approach since it joins (and, indeed, reinforces) the index framework with the evaluation of the same biological samples by a suitable combination of deformable models. Nucleus contour is identified through Active Shape Models techniques, and cytoplasm contour's detected through parametric Snakes, with prior image preprocessing based on statistical and mathematical morphology techniques. Morphometric parameters such as nucleus and cytoplasm area and ratio between them are then easily computed. As a result the dataset validated the applied methodology with a realistic background and a new, more accurate and ecologically realistic index for oocyte staging emerged.

1 Introduction

It's in the oceans that the gross majority of primary biomass is produced by phytoplankton (the world's largest plant crop). When grazing upon these primary producers, zooplankters constitute a crucial link to higher trophic levels which begin with fish and, most of the time, culminate in man [1]. Copepods often represent 80-95% of the mesozooplankton, a fact which considering continuously dwindling yields in fisheries catapults the understanding of their recruitment to an imperious necessity in order to characterize and quantify energetic flux in aquatic environments [1,2,3]. *Acartia tonsa* Dana (Copepoda: Calanoida) was used as model organism (given its dominance in zooplanktonic communities) to establish an index determining oocyte maturity stage (including inter-seasonal variance) in zooplankters [4]. Based in

citometry (measurements using an ocular micrometer and thus calculating N/C ratios = area of the Nucleus / area of the Cytoplasm in percentage) and histochemical evaluation of the gonadic masses (identification of chemical constituents) consisted in the division of the oocytes in three stages: Immature, vitellogenic and mature, and in a finer pattern recognizing differences between months of high (H) and low (L) abundance (were reproductive strategies differ [2,5]) within each stage [4]. Roughly by the same time a novel method based in deformable models was being developed to evaluate histological sections [6].

This paper presents a novel approach by fusing both concepts: re-evaluate the same biological material using a suitable combination of deformable models (Active Shape Models [7] and Snakes [8]) in order to either confirm or, by weight of evidence, build a new index. Nucleus contours are identified through Active Shape Models (ASM) techniques, and the cytoplasm contours are detected through parametric deformable models (Snakes), with a prior preprocessing based in statistical and mathematical morphology techniques. Smoothed instances of the final contours (nucleus and cytoplasm) are then obtained through ASM and spline approximation based on the detected cytoplasm edge points, respectively. Morphometric parameters such as nucleus and cytoplasm area and ratio between them are then easily computed.

The outcoming results are the amplification of the index to four development stages and enhanced capability of information gathering regarding oocyte biometry, leading to the clarification of inter-seasonal differences in the reproductive cells of these essential elements for energy transduction in aquatic ecosystems.

In section 2, we describe the details related with material and methods of our technique. Section 3 discusses the results and presents the new characterization index established for the determination of oocyte maturity. Section 4 concludes with a short summary of our work.

2 Material and Methods

For details on biological material collection and processing see Pastorinho et.al. [4].

2.1 Computer Technique

Our method to evaluate gonadic cell masses consists on four well differentiated stages: Initial Image Processing, Initial Segmentation, Final Segmentation and Feature Extraction, which are discussed below.

2.1.1 Initial Image Processing

Initial image processing is carried out to prepare images for objects differentiation and is divided in two steps: initial image preparation and image enhancing.

Initial Image Preparation. The presence of noise in images may represent a serious impairment for subsequent automated quantitative evaluation tasks. This median filter has been used extensively for image noise reduction and smoothing. The filter preserves monotonic image features that fill more than half the area of the transform window. Examples are constant regions, step, edges, or ramp edges of sufficient

extent. Median filters are especially good at removing impulse noise from images [9]. Initial image preparation consists on building the median image M (image denoising), which is achieved applying the median filter to the input image I (see Fig.1(B1)) (a 256 gray level image), with a window size of 9x9 pixels. Figure 1(B2) show the median image.

Image Enhancement. Mathematical morphology is a shape-oriented technique for image processing and analysis, based on the use of simple concepts from set theory and geometry [10]. Images under study contain gonadic cells of diverse shape and sizes, in which nucleus appear in different positions with respect to cytoplasm. Due to this, to increase the potential for future object discrimination was used a suitable combination of mathematic morphology (top-hat and bottom-hat) operations. We evaluated structuring elements of different shapes and sizes, obtaining the best results when an octagonal structuring element is used. A flat octagonal structuring element K was created computing the radius of the maximum diagonal diameter in the biggest cell's nucleus of the image under study (see Fig.1(B3)). The mathematical formulation to enhance image M to obtain image E is:

$E = A - B$ where E is the enhanced image

$A = M + T$

$T = M - \gamma(M)$ top - hat of M

$\gamma(M) = M \circ K = [(M \ominus K) \oplus K]$ open M with the structuring element K

$B = \phi(M) - M$ bottom - hat of M

$\phi(M) = M \bullet K = [(M \oplus K) \ominus K]$ close M with the structuring element K

where

M = Median image

K = Structuring element

\ominus = erode operator

\oplus = dilate operator

2.1.2 Initial Segmentation

The initial segmentation process is carried out to differentiate objects (gonadic cells) from background and includes two steps: primary image segmentation through thresholding and object's edge detection.

Thresholding. Histogram analysis reveals heuristically that the black pixels were in the interval (0,109) and the white pixels were in the interval (110,255). Mathematical formulation to achieve objects differentiation (image BW) is the following:

$$BW_{(i,j)} = \begin{cases} 255 & \forall E'_{(i,j)} > (\rho_2 - d) \\ 0 & \end{cases}$$

ρ_2 maximum local peak of white area in the frequency histogram of E'

$$\rho_2 \in [110, 255]$$

ρ_1 minimum local peak of black area in the frequency histogram of E'

$$\rho_1 \in [0, 109]$$

$$d = |\rho_2 - \rho_1|/10$$

$E' = E - E^C$; E^C complement of E' ; E' image enhanced

Edge Detection. This is a critical step, since edge information is a major driving factor in subsequent deformable models performance. Several techniques were tested such as a combination of noise suppression by average and median filtering, with different masks and thresholding, followed by binarization and edge tracking [11]. We've also tried with edge maps detectors [12], but these methods fail where a gonadic cell's edge is not completed and closed. However we found that applying a local median average, as we propose in [6], produces a more suitable set of nucleus and cytoplasm edges (see Fig.1(B4)). The mathematical formulation used to detect edges is:

E_{map} output image

$$E_{map}(i, j) = \frac{1}{(2M+1)^2} \sum_{(k,l) \in [-M,M]} |X(i_0+k, j_0+l) - \varpi|$$

ω $N \times N$ windows centered on pixel (i_0, j_0)

$$M = \frac{N-1}{2}$$

ϖ median of ω

BW input image

2.1.3 Final Segmentation (Deformable Models)

Mathematical foundations of deformable models represent the confluence of geometry, physics, and approximation theory. Geometry is used to represent object shape, physics inflict constraints on how the shape may vary over space and time, and optimal approximation theory makes available the formal underpinnings of mechanisms for fitting the models to measured data. We use a suitable combination of two kinds of deformable models: Active Shape Models (ASM) proposed by Cootes et. al. [7] to identify the nucleus contour and hereafter the Gradient Vector Flow (GVF) Snake proposed by Xu and Prince [8] to identify the cytoplasm boundary.

Active Shape Model. The ASM uses the point distribution model (PDM) and the principal component analysis (PCA) for modeling objects. The implementation of ASM is characterized by the following stages: labeling of training set; alignment of training set; capturing of statistics of a set of aligned shapes, and finally the application of the model to search shape in image. In order to model a nucleus, we represent it by a set of points. The labeling of the points (landmarking) is important, and these can be placed manually or semi-automatically. Each labeled point represents a particular part of the object or its boundary. We can describe each object

of the training set by the vector $X_i = [x_{i0}, y_{i0}, x_{i1}, y_{i1}, \dots, x_{in-1}, y_{in-1}]$, where n is the number of points that define each object and i is the object identifier. As a result of labeling the training set we have a set of N_s vectors. In order to study a variation of the position of each landmark throughout the set of training images, it's necessary to align all the shapes to each other. The alignment is done by changing the pose (scaling, rotating, and translating) of the shapes. That is made in order to minimize the weighted sum of squares of distances between equivalent points on different shapes. After the alignment of the training set, it there is a diffuse "cloud" around each landmark. These "clouds" represent the variability of object. We can use principal component analysis (PCA) to determine the variation modes of the object. If we assume that the first t principal components explain a sufficiently high percentage (96%) of the total variance of the original data, any shape in the training set can be approximated using the mean shape and the weighted sum of the deviations obtained from the first t modes:

$$X = \bar{X} + Pb$$

$$\bar{X} \text{ mean shape}$$

$$P = [p_1 \ p_2 \ \dots \ p_t] \text{ matrix of the first } t \text{ eigenvectors}$$

$$b = (b_1 \ b_2 \ \dots \ b_t)^T \text{ vector of weights}$$

The suitable limits for b_k are typically of the order of $-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k}$, where λ_k is the k th eigenvalue derived by PCA. Now we can apply this model to search a shape in image, but first we have to initialize the model over the image. Then, we examine the neighborhood of the landmarks of our estimate, trying to find the better locations for each landmark. Hereafter, we change the shape and the pose of our estimate to better fit the new locations for the landmarks. Each iteration produces a new acceptable shape. The system finishes the search when the shape has insignificant changes over successive iterations (when the desired movement for each landmark has a small value). In our study we assume this value equal to 8 pixels. The desired movement or adjustment of each landmark is obtained from modeling the gray level statistics around each landmark, in order to better fit of the object to the image.

Gradient Vector Flow Snake. For details on GVF model see our previous work Guevara et.al. [6]. But here we first use ASM to segment the nucleus, and then, based on the nucleus edge points the initial snakes for cytoplasm contours are automatically created. We use the parametric equation of the line formed with the nucleus centroid and the maximum radius between the centroid and the nucleus edges took in angles of 20° , in order to obtain the intersection points with the cytoplasm edges. Initial snakes were obtained with spline approximation over the set of this intersection points (see Fig.1 (A)), then the GVF snake deformation is carried out to produce the final snakes. In the snake deformation process to compute the cytoplasm edges we needed to increase rigidity and pressure force weighs, due to the shape variability of gonadic cells. The parameters used in the snake deformation process were: elasticity (0.05), rigidity (0.1), viscosity (1), external force weight (0.6) and pressure force weight (0.2). Figure 1(B5) show the final snakes representing nucleus and cytoplasm contour.

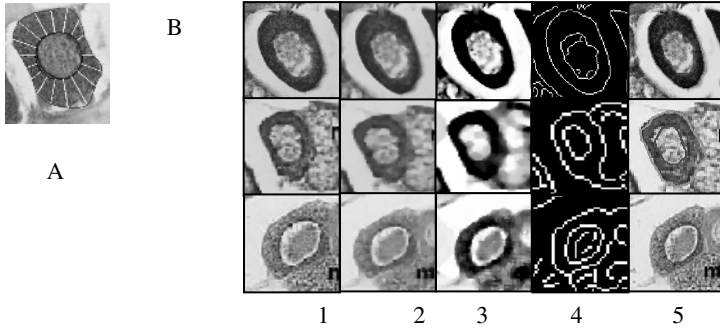


Fig. 1. (A) Spline approximation to create the cytoplasm initial snake; (B) 1-Original image, 2-Enhanced, 3-Segmented, 4-Edges, 5-Final cytoplasm and nucleus edges

2.1.4 Feature Extraction

Morphometric features express the overall size and shape of objects. For these features only the object mask O and its border ζ are needed, not the actual gray scale image [13]. We compute nucleus and cytoplasm areas and the ratio between them to evaluated gonadic cells. To do this we use as input the final snakes deformations (the arrays of edge points of nucleus and cytoplasm). Mathematical formulation and computational sequence of measurements are the following:

O_N nucleus pixels

O_C cytoplasm pixels

$\zeta_N \subset O_N$ set of edge pixels (final snake), contour of O_N

$\zeta_C \subset O_C$ set of edge pixels (final snake), contour of O_C

$A_N = |O_N|$ area = number of elements of O_N

$A_C = |O_C|$ area = number of elements of O_C

$ratio = (A_N / A_C) * 100$

3 Results and Discussion

This section will only be object of a brief set of considerations given the scope of this article. The original index [8] was divided in three maturity stages: Immature with $N/C=73.46$ ($H=73.02$; $L=73.91$), Vitellogenic with $N/C=46.98$ ($H=52.31$; $L=41.64$) and Mature with $N/C=20.78$ ($H=19.0$; $L=22.56$). The most significant outcome of this work was the unfolding (significantly different, $p < 0.001$, Table I) of the Vitellogenic stage in two: Primary ($N/C=62.39$; $H=59.61$; $L=65.17$) and Secondary ($N/C=29.04$; $H=23.01$; $L=35.07$). As it would be expected, substantial modifications occurred as well in Immature and mature stages: the latter with $N/C=12.29$ ($H=4.36$; $L=20.21$) and the former with $N/C=121.42$ ($H=102.38$; $L=140.45$) (significantly different, $p < 0.001$, Table 1). These results fit in ecological models that predict lower parental investment (less offspring, bigger in size) in low abundance epochs (more

severe environmental constraints, *e.g.* higher temperature, less available food, enhanced predatorial pressure [14]) in order to enhance viability of the spawned eggs and to assure their own survival [2,3,5].

Table 1. One-way ANOVA tests applied to Oocyte size in the months of September (I) and March (II) and to four stages of oocytical maturation (immature -III; primary vitellogenic -IV; secondary vitellogenic -V, and mature -VI) of *Acartia tonsa*. df = degrees of freedom; MS = Mean square; *F*s = Test Value; *P* = Probability value.

I- One way ANOVA of the oocytes size of *Acartia tonsa* distributed by maturation state, for the month of September. The null hypotheses is that the oocyte's size does not differ between maturity status.

II- One way ANOVA of the oocytes size of *Acartia tonsa* distributed by maturation state, for the month of March. The null hypotheses is that the oocyte's size does not differ between maturity status.

III- One-way ANOVA of the Immature (Im) stage of oocytes of *Acartia tonsa* during the period of the study. The null hypotheses is that all the Im cells do not register any variation between samples

IV- One-way ANOVA of the Primary Vitellogenic (PV) stage of oocytes of *Acartia tonsa* during the period of the study. The null hypotheses is that PV the cells do not register any variation between samples.

V- One-way ANOVA of the Secondary Vitellogenic (SV) stage of oocytes of *Acartia tonsa* during the period of the study. The null hypotheses is that SV the cells do not register any variation between samples.

VI- One-way ANOVA of the Mature stage of oocytes of *Acartia tonsa* during the period of the study. The null hypotheses is that all the M cells do not register any variation between samples.

Source of variation	Df	MS	<i>F</i> s	<i>P</i>
I-Sample	2	0.008	764.28	<i>P</i> < 0.001
II-Sample	2	0.011	752.33	<i>P</i> < 0.001
III-Sample	2	0.015	114.17	<i>P</i> < 0.001
IV-Sample	2	0.004	78.92	<i>P</i> < 0.001
V-Sample	2	0.003	46.23	<i>P</i> < 0.001
VI-Sample	2	0.002	36.86	<i>P</i> < 0.001

4 Conclusions

We presented an innovative method to segment histological sections based on a suitable combination of deformable models: Active Shape Models and Gradient Vector Flow Snakes, which allowed developing a new index to evaluate Zooplanktons' gonads. This approach is an extension of our previous work [6], but in this case was include an Active Shape Model to semiautomatically detect the gonad's cell nucleus (N). Afterward, using the set of point of the nucleus contour is built the initial snake to detect automatically cytoplasm (C) contour. The ability of our algorithm was demonstrated on an experimental representative dataset. For present and future biological studies the most significant outcome of this work was the unfolding of the Vitellogenic stage in two: Primary and Secondary. As it would be expected, substantial modifications occurred as well in Immature and mature stages: the latter with N/C=12.29 and the former with N/C=121.42.

Acknowledgment

The authors wish to thank to Research Institute of University of Aveiro Project CTS 2002/22 and to Research Unit 127/94 - IEETA for financial support.

References

1. Runge, J. A.: Should we expect a relationship between primary production and fisheries? The role of copepod dynamics as a filter of trophic variability. *Hydrobiol.*, 167-168 (1988) 61-71.
2. Miralto, A., Ianora, A., Butino, I., Romano, G. and Di Pinto, M.: Egg production and hatching success in north Adriatic sea populations of the copepod *Acartia clausi*. *Chemistry and Ecology* 18(1-2) (2002) 117-125.
3. Prokopchuk, I.: Mesozooplankton distribution, feeding and reproduction of *Calanus finmarchicus* in the western Norwegian Sea in relation to hydrography and chlorophyll *a* in spring. The United Nations University, Final Project (2003).
4. Pastorinho, R., Vieira, L., Ré, P., Pereira, M., Bacelar-Nicolau, P., Morgado, F., Marques, J.C. and Azeiteiro, U.: Distribution, production, histology and histochemistry in *Acartia tonsa* (Copepoda: Calanoida) as means for life history determination in a temperate estuary (Mondego estuary, Portugal). *Acta Oecologica* 24 (2003) S259-S273.
5. Hairston Jr., N.G. and Bohonak, A. J.: Copepod reproductive strategies: life history theory, phylogenetic pattern and invasion of inland waters. *Journal of Marine systems* 15 (1998) 23-34.
6. Guevara, M.A., Silva, A., Oliveira, H., Pereira, M.L., Morgado, F.: Segmentation and Morphometry of Histological Sections using Deformable Models: A New Tool for Evaluating Testicular Histopathology. *Progress in Pattern Recognition, Speech and Image Analysis, Lecture Notes in Computer Science*. Vol. 2905 (2003) 282-290.
7. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models – their training and application. *Computer Vision and Image Understanding*, Vol. 61, n.º 1 (1995) 38-59.
8. Xu, C. and Prince, J.L.: Gradient Vector Flow. A New External Force for Snakes. *Proc. IEEE Conf. on Comp. Vis. Patt. Recog. (CVPR)*, Los Alamitos: Comp. Soc. Press. (1997) 66-71.
9. Senel H.G., Peters R.C.: Topological Median Filters *IEEE Trans. Image Processing*, Vol.11 (2002) 89-104.
10. Soille, P. *Morphological Image Analysis*. Springer-Verlag, Berlin (1999).
11. Guevara, M. A., Rodríguez R.: Knowledge-based vision techniques to classify sugarcane somatic embryos. *Proceedings 13th ISPE/IEEE International Conference on CAD/CAM, Robotic and Factories of the Future (CARS & FOF'97)*. (1997) 388-396.
12. Canny J.F.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.* Vol. PAMI-8, (1986) 679-678.
13. Rodenacker K., Bengtsson E.: A feature set for cytometry on digitized microscopic images. *Analytical Cellular Pathology IOS Press*, Vol. 25 (2003) 1-36.
14. Maffei, C., Vagaggini, D., Zarattini, P., Mura, G.: The dormancy problem for crustacea Anostraca: A rigorous model connecting hatching strategies and environmental conditions. *Ecological Modelling* 185 (2005) 469-481.