UNIVERSIDAD DE LEÓN

Departamento de Ingeniería Eléctrica y de Sistemas y Automática

Descripción Adaptativa de Texturas y Estimación de las Probabilidades a Priori de las Clases para el Control de Calidad Seminal

Tesis doctoral dirigida por

El Prof. Dr. Enrique Alegre Gutiérrez

y la Prof. Dra. Rocío Alaiz Rodríguez

y desarrollada por

Víctor González Castro

a fin de optar al grado de

Doctor por la Universidad de León

León, Junio de 2011

UNIVERSITY OF LEÓN

Department of Electrical, Systems and Automatic Engineering

Adaptive Texture Description and Estimation of the Class Prior Probabilities for Seminal Quality Control

A dissertation supervised by

Prof. Dr. Enrique Alegre Gutiérrez

and Prof. Dr. Rocío Alaiz Rodríguez

and submitted by

Víctor González Castro

in fulfillment of the requirements for the Degree of

PhilosofæDoctor (Ph.D.)
University of León

León, June 2011

Para mi familia. Para ti.

Agradecimientos

Una vez completado este trabajo puedo confirmar aquello que al principio solo sospechaba: una Tesis doctoral no es producto únicamente del trabajo del doctorando. Muchas son las personas e instituciones que han colaborado, de una u otra manera, para que esta aventura haya llegado a buen puerto. Quiero aprovechar estas páginas para agradecer de todo corazón la ayuda que me han prestado. En especial quiero agradecer:

Al Ministerio de Ciencia e Innovación, por haberme permitido disfrutar de una ayuda para la Formación de Personal Investigador durante los cuatro últimos años, sin la cual esta empresa hubiera resultado imposible.

A mis directores, Enrique Alegre y Rocío Alaiz, por darme la oportunidad de realizar mis estudios de doctorado bajo su dirección y supervisión. Ellos han puesto a mi disposición todos los recursos necesarios para facilitarme el trabajo, y se han mostrado disponibles en todo momento, dándome los conocimientos y consejos adecuados para poder sacar adelante esta tarea.

Asimismo agradezco a Maria Petrou por haberme acogido como visitante durante tres meses en el grupo de investigación de Comunicaciones y Procesamiento de Señales del *Imperial College London*, sus ideas, y la atención prestada, incluso después de finalizada dicha estancia.

A Centrotec S.A., por habernos facilitado las instalaciones y recursos para poder realizar la adquisición de imágenes de semen de verraco, y a las personas que allí trabajaban, por conseguir que nuestra labor fuese un poquito más llevadera. Quiero agradecer especialmente a David, mi compañero en esta tarea de adquisición de

imágenes, además de amigo, por todas las horas de conversaciones, apoyo mutuo y buenos consejos que pasamos entre el microscopio y el ordenador. Gracias también por su ayuda con la documentación de la preparación de las muestras, adquisición de las imágenes y demás conceptos veterinarios que, debido al desconocimiento, me hubiera sido imposible plasmar correctamente en esta Tesis.

A los profesores Manuel Castejón, Chema Foces y Joaquín Barreiro, cuya colaboración, consejos o presencia misma me han ayudado en un momento u otro de este proceso. A los compañeros doctorandos que me han hecho sentir acompañado en el camino: Laura, Manu, Dani, Sir Alexci, Diego, Maite, Oscar, ...

A mis Amigos, ya que en muchas ocasiones este trabajo me ha impedido prestarles la atención que yo hubiera deseado. Gracias por vuestra comprensión y apoyo.

A mis padres, por el amor, entrega y generosidad que han demostrado en el cuidado de sus hijos. Sé que lo han dado todo por mí, y me han ayudado siempre que lo he necesitado, sin siquiera tener que pedírselo. Sin duda, esta tesis es el resultado de la educación que me han proporcionado, y de los valores de trabajo, sacrificio y esfuerzo que ellos me han inculcado y que me acompañarán todos los días de mi vida. También incluyo en este agradecimiento a mi abuela, como copartícipe de la formación de la persona que he llegado a ser. Sé que se sentirá orgullosa. A mi hermano quiero agradecerle la simpatía que me ha transmitido siempre, y que me ha ayudado más de lo que he sabido transmitirle.

Por último, mi más sentido agradecimiento es para ti, Iria, porque has tenido la infinita paciencia de acompañarme en este viaje desde la primera jornada del mismo, has sabido disfrutar conmigo en las buenas etapas y has sido mi apoyo para no caer cuando las fuerzas flaquearon. Por tu amor, cariño, comprensión, valor y ánimo. Gracias también por tu asesoramiento sobre el formato del manuscrito y tu colaboración en alguna de las Figuras que se muestran en este trabajo. Por todo, sin ninguna duda, parte de esta Tesis es tuya.

Abstract

In this Thesis we have evaluated several approaches to describe digital image textures. In addition, we have proposed a new intelligent segmentation procedure, an original *adaptive* texture descriptor and two new methods for estimating class proportions (*quantification*) of unlabelled datasets.

The adaptive texture description method, called Adaptive Geodesic Pattern Spectrum (AGPS), is based on Mathematical Morphology and Geodesic distances. Experimental results with images from the VisTex database show that it is more efficient than the classical Pattern Spectrum (PS) when textures have similar texel shapes.

Additionally, we have proposed an iterative method, based on the posterior probalities (PP) estimates provided by a classifier, which is able to estimate the class distribution of a new unlabelled dataset. The other quantification proposal relies on measuring the distributional divergence with the Hellinger Distance (HD) between the new operational data and validation datasets generated from the available training dataset in a fully controlled way. They may work on the data itself (called HDx) or on the classifier outputs (called HDy). They have been evaluated using data from different real problems and compared with previous approaches based on the confusion matrix. Results show that using a quantification method greatly improves the naïve approach of counting the classifier predictions (CC), and HDy is the approach with the best average ranking.

Particularly, these methods have been assessed on a semen quality control application. This problem requires the segmentation of the sperm heads. Our proposal, which combines thresholding and the Watershed transform, achieved better results than simply using

thresholding, with a much lower computational cost than the latter. We have also assessed several texture (Curvelet and Wavelet transforms combined with first and second order statistics) and shape descriptors (Hu, Flusser, Legendre and Zernike moments) for this particular problem. Curvelet, in combination with co-occurrence matrix features, beat all of them in the detection of acrosome integrity. In addition, texture descriptors clearly outperformed shape-based ones.

In the semen assessment application, the most important task is to estimate class proportions rather than the classification of each individual example. Experimental results with the intact/damaged acrosome dataset show that HDy and PP are able to quantify the class distribution with the lowest mean absolute deviation. In addition, HDy showed a higher robustness to classifier performance. Regarding the estimations of class proportions of alive/dead spermatozoa described by means of AGPS, all proposed methods achieved similar results, greatly outperforming the baseline approach of classifying and counting (CC).

Resumen

En esta Tesis se han evaluado varias técnicas para describir texturas en una imagen digital. Además, hemos propuesto un nuevo método de segmentación inteligente, un descriptor de texturas *adaptativo* y dos nuevos procedimientos para estimar proporciones de clases (*cuantificación*) en conjuntos de datos no etiquetados.

El método de descripción adaptativa de texturas, llamado Espectro de Patrones Geodésicos Adaptativos (AGPS en sus siglas en inglés, *Adaptive Geodesic Pattern Spectrum*), está basado en Mofología Matemática y en distancias Geodésicas. Los resultados de los experimentos utilizando imágenes extraídas de la base de datos VisTex muestran que es más eficiente que el *Pattern Spectrum* clásico cuando los téxeles de las texturas tienen formas similares.

Adicionalmente se ha propuesto un método iterativo, basado en las estimaciones de las probabilidades a posteriori (PP) devueltas por un clasificador, capaz de estimar la distribución de las clases de nuevos conjuntos de datos no etiquetados. El otro método de cuantificación propuesto consiste en utilizar la distancia de Hellinger para medir la divergencia de la distribución entre datos nuevos y un conjunto de validación generado de manera controlada a partir de los datos de entrenamiento. Esta medida se puede realizar sobre los datos originales (HDx), o sobre las salidas de un clasificador (HDy). Estos métodos han sido evaluados utilizando datos de diferentes problemas reales y comparados con métodos previos basados en matrices de confusión. Los experimentos muestran que las estimaciones de los procedimientos de cuantificación mejoran los resultados obtenidos con el método simple, que es el generalmente utilizado, de clasificar y contar (CC), y que HDy es el que obtiene el mejor ranking medio.

Adicionalmente, estos métodos se han evaluado en una aplicación de control de calidad seminal. Como este problema requiere la segmentación de las cabezas de los espermatozoides, hemos realizado una propuesta que combina un método basado en umbralización con la transformada *Watershed*. De esta forma hemos obtenido mejores resultados que simplemente utilizando umbralización, con un coste computacional mucho más bajo que el método basado en *Watershed*. Además, para este mismo problema, se han evaluado varios descriptores de textura (las transformadas *Curvelet* y *Wavelet* combinadas con descriptores estadísticos de primer y segundo orden) y de forma (momentos de Hu, Flusser, Legendre y Zernike). La transformada *Curvelet*, combinada con características de la matriz de coocurrencia superó a todos los demás en la detección de la integridad acrosómica. Adicionalmente, se observó que los descriptores de textura claramente superaron a los descriptores de forma.

Un aspecto fundamental de las aplicaciones de control de calidad seminal es que resulta más importante estimar las proporciones de las clases que la clasificación de cada célula individual. En este contexto, los resultados experimentales con el conjunto de datos de acrosomas íntegros y dañados muestran que PP y HDy son capaces de cuantificar la distribución de clases con la menor desviación absoluta media y además, HDy mostró una robustez más alta con respecto al rendimiento del clasificador. En cuanto a las estimaciones de proporciones de las clases de espermatozoides vivos y muertos, descritos mediante AGPS, todos los métodos de cuantificación propuestos obtuvieron resultados similares, mejorando notablemente el método básico CC.

Contents

Índice general

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Motivation

Computer vision has become a powerful tool in industry. Security, manufacturing or quality control applications are just some examples where it is used successfully. Biomedicine is other field where the advantages of digital image processing are more and more useful.

Semen quality assessment is a crucial task in Artificial Insemination (AI) processes both in medicine, *e.g.* with therapeutic goals or for clinical analysis of infertility in male patients, and veterinary, *e.g.* with breed improvement purposes in food industry. In this particular case, there is a whole market around high-quality semen used for AI of animals like boars or cows. Owners of farms regularly purchase semen samples from breeding companies or production centres. These companies have to assure good quality standards and, thus, they have to carry out rigorous quality control procedures to guarantee that they sell samples with optimal fertilization potential.

There is a relation between sperm fertility potential and vitality and the structural integrity of the acrosome, since a semen sample which has a high proportion of spermatozoa with damaged acrosomes, or which are dead, will have a low fetilization potential. Currently, the assessment of these parameters is carried out visually, using staining techniques and counting the stained spermatozoa. This manual process is time consuming and may introduce errors, as it involves the subjectivity of a human observer. In addition, expensive equipment, *e.g.* a fluorescence microscope, is needed to observe samples, stains are sensitive to temperature variations, and they may even increase the number of dead spermatozoa.

It would be very useful to automatically detect damaged acrosomes and dead spermatozoa without using stains. This task could be carried out using digital image processing to analyse grey level images of the samples. It would make the process objective, faster, and would allow research labs and breeding companies to save money, as the requirements would just be a digital camera, a computer, and a conventional microscope.

A key factor for the success of a pattern recognition application based on computer vision is the feature set extracted from the image to describe it. In this case, it is possible to notice differences in the grey level distribution of the sperm head images between the cells whose acrosome is intact or damaged.

With regard to dead or alive spermatozoa, different kinds of textures also appear inside the head, so we have applied texture analysis for describing those two different classes. Traditional texture descriptors are often applied as-are on the whole region of interest (ROI), so it looked appropriate to propose an *adaptive* texture descriptor that would take into account local features inside the texture having no a priori knowledge about it.

Supervised learning techniques deal with the extraction of the best possible features from a set of labelled instances in order to train a classification model. Once the classifier is designed, it is applied as-is to new data in order to predict the class each individual belongs to. In supervised classification, it is often assumed that training and actual (test) data follow the same, although unknown, distribution (Duda et al., 2001). In particular, class prior probabilities of the training data set are considered to truly reflect the class distribution of the operational environment. However, time or space class stationarity cannot be assumed in many practical fields.

In particular, in a semen quality control application the class distributions cannot be considered stationary because of factors such as the animal/farm variability, the manipulation or the conservation conditions (Johnson et al., 2000). This is specially important, as semen cryoperservation is the most common approach for storing semen samples (Cardozo et al., 2009; Hernández et al., 2007; Watson, 1995), and it may affect the integrity of the membranes of the spermatozoa (increasing the number of damaged sperm cells). Remote sensing applications also suffer from that problem since a dataset collected in a given season from a region with different terrains (industrial, hay, wheat, corn, grass, ...) is employed to train the classifier. However, when that classifier is deployed, mismatches in terrain distribution may appear just because seasonal or location changes (Guerrero-Curieses et al., 2009).

It is well known that a mismatch between the actual class prior probabilities and those the classifier has been generated with, leads to suboptimal solutions (Duda et al., 2001). Whenever there is such a change, estimating this new class proportion, referred to as *quantification* (Forman, 2008), is fundamental to adapt the classifier to the new context with the goal of improving the individual classification performance (Saerens et al., 2002; Xue and Weiss, 2009). In other contexts, it is important by itself, since it is the main goal of the application.

In summary, artificial insemination techniques in the veterinarian field should guarantee that semen samples are optimal for fertilization. Therefore, the real interest in this domain is to assess the proportion of dead cells or damaged acrosomes, without any concerns about the individual classification of each cell. This task can be carried out using features automatically extracted from the texture of grey level images of the spermatozoa. This would allow to save time and money, and make more robust estimations of the proportion of undesirable spermatozoa in boar semen samples, thus, improving the AI process productivity in porcine industry.

1.2 Objectives

The work presented in this dissertation follows two research lines: Texture description and the estimation of class prior probabilities, *i.e. quantification*.

We have explored both of them, with the *main goal* of creating a system which allows to automatically assess the proportion of damaged acrosomes in a semen sample, by extracting information from grey-level images acquired using a phase contrast microscope.

Focusing on it, we defined the following *particular objectives*:

1. To design a method able to satisfactorily segment as many sperm heads as possible, which guarantees that each returned head is properly segmented, discarding the bad segmented ones.

2. To develop descriptors, either based on texture or shape, that capture the information of the spermatozoa heads necessary to characterise them.

3. To implement approaches that reliably estimate the class distribution of an unlabelled dataset in environments where the operational conditions are imprecise.

1.3 Main Contributions

The main contributions of this Ph.D. dissertation may be summarized as follows:

1. A new intelligent segmentation approach is proposed. This method combines histogram global thresholding (Otsu, 1979) and the Watershed transform (Meyer and Beucher, 1990). The proposed method is more accurate, *i.e.* segments better the sperm heads than the threshold-based approach, and considerably reduces the computational cost of the watershed-based one.

2. The Discrete Curvelet Transform (Candès et al., 2006) has been applied in combination with some statistical texture descriptors to characterise intact and damaged acrosomes. It has been compared with other texture and shape descriptors, outperforming all of them.

3. An original adaptive texture description approach based on Mathematical Morphology is presented in this Thesis. It is based on computing the Pattern Spectrum (PS) (Maragos, 1989) of a texture using structuring elements whose shape and size change at each pixel on the basis of a geodesic distance criterion. It has been called Adaptive Geodesic Pattern Spectrum (AGPS).

4. An approach to estimate the class proportions of unlabelled datasets that relies on the posterior probabilities provided by a classifier has been proposed. This posterior probability-based method (PP) is an iterative procedure based on the Expectation Maximization (EM) algorithm.

5. An original approach to estimate class proportions on the basis of measuring distributional divergences by means of the Hellinger Distance has been presented in this dissertation. Specifically, two proposals have been made, which measure these divergences between (a) the data itself (HDx) or (b) between the classifier outputs (HDy). Evaluated on several domains, this latter proposal achieved the best results, greatly outperforming the naïve approach of counting the classifier predictions, and it turned out to be very robust to classifier performance.

6. Two image databases containing the images of the intact/damaged acrosomes and the alive/dead spermatozoa have been built. In addition, some datasets containing the textural information of the intact/damaged acrosomes have been produced. It may be very useful for the scientific

community interested in computer vision and machine learning to work with datasets from these real domains.

1.4 Document Overview

In this section the structure of this doctoral Thesis is described. This first introductory chapter has been focused on motivating the work presented in this dissertation, its main objectives and original contributions. Now, the rest of the document is outlined as follows:

First of all, chapter 2 contains a review of texture description methods, adaptive preprocessing, segmentation and description approaches, as well as mathematical morphology and geodesic distance methods. There are not many works in digital image processing that focus on sperm images, so we have focused on literature that deal with cells and tissues, where possible. Likewise, works involving the problem of shifts in class prior probabilities are examined.

Next, the background methods that have been used to accomplish the goals of this Thesis are described in chapter 3. This chapter also contains a brief explanation about the basis of boar sperm quality assessment and the way the sperm images have been acquired.

The intelligent approach proposed for segmenting the heads and the method for detecting bad segmented heads are explained in chapter 4. The experimental results of this method with a set of heads with alive and dead images are shown in this chapter, as well.

The experiment about the description of intact and damaged acrosomes by means of the Discrete Curvelet Transform and its comparison with the Discrete Wavelet Transform and other shape descriptors is shown in chapter 5.

Then, the contribution of this Thesis with regard to the adaptive texture descriptor is explained in chapter 6. This proposal is tested with two different datasets: on the one hand, with textures of different materials extracted from the VisTex database and, on the other hand, with alive and dead sperm heads detection purposes.

The next two chapters deal with proportion estimation tasks. Firstly, chapter 7 describes the quantification methods proposed in this Thesis. In addition, an experiment to compare them with other previous approaches is presented in that chapter.

Afterwards, texture description of sperm images and their quantification are assessed together in chapter 8. The texture descriptors shown in this Thesis are extracted from both images of alive/dead spermatozoa and intact/damaged acrosomes. This data is then used with the quantification methods, on different scenarios.

Finally, a summary with the conclusions of this Thesis is shown in chapter 9. Possible future work lines are indicated, as well.

Appendix A has a brief description about each one of the datasets used in the quantification experiments of chapter 7, and a list of the publications derived from this work is shown in Appendix B.

Regulations about the Ph.D. studies at the University of León claim that if the doctoral Thesis is not written in Spanish, at least the table of contents, conclusions, and a résumé of each chapter must be written in Spanish. In order to comply with these regulations, we include a translation of the conclusions in chapter 10, and a summary of all chapters in part II.

CHAPTER 2

STATE OF ART

2.1 Digital Image Processing for Cells and Tissues Characterization

Digital image processing is widely used in microscopy (Bonnet, 2004) and biomedicine applications, and it turns out to be very useful in the automatic analysis of cells and tissues. Some examples where it has been used are detection of parts or artefacts within cells, identification of abnormalities in tissue sections or vessel segmentation.

Ruggeri and Pajaro compared the performance of Hu and Zernike moments in the recognition of images from different layers of the cornea, describing the shape of their cells and classifying them by a neural network classifier (Ruggeri and Pajaro, 2002). Liyun *et al*. proposed a spermatogonium recognition scheme that used Zernike moments to characterise the images (Liyun et al., 2009). It outperformed other approaches which used Hu and boundary moments. Osman *et al*. proposed an approach to detect tissue infected by tuberculosis bacilli in images of tissue sections, describing them by means of Zernike moments and using Multilayer perceptrons for the recognition stage (Osman et al., 2010). Marín *et al*. used a Neural Network and Hu moment invariants to characterise pixels in grey level retinal images with the goal of segmenting blood vessels.

Ong *et al*. summarised the elements that an image analysis of tissue sections application should include, as well as a survey of its stages and procedures (Ong et al., 1996). Amongst them, the feature extraction stage has singular interest, as Rodenacker and Bengtsson highlighted (Rodenacker and Bengtsson, 2003). They even showed a methodology to describe images with the goal of performing cytometric studies and proposed a taxonomy for the different kinds of features: *size and shape*, *intensity* – *i.e*. the pixel values –, *texture* and *structure*.

Many computer vision applications rely on texture analysis, as it is a very powerful tool in the characterisation of cells and tissues, as the many published works prove. However, despite its importance, there is not a formal definition of texture. Petrou and García Sevilla claim that "Texture is the variation of data at scales smaller than the scales of interest". There is usually a different definition depending on the method used to analyse it, as Gonzalez and Woods suggest (Gonzalez and Woods, 2002). Moreover, Castellano *et al*. carried out a review and description of some texture analysis techniques used with

medical images (Castellano et al., 2004), and divided these approaches in four categories:

- **Statistical approaches** consist in characterising textures using properties governing the distribution and relationships of grey level values within the image. This category includes first order statistical features computed from the histogram of the image (Sheshadri and Kandaswamy, 2007), second order statistical features, extracted from the grey level co-occurrence matrix (Mahmoud-Ghoneim et al., 2008; Tsang et al., 2005), run-length matrix (Chandraratne et al., 2006), or statistical moments (Morales et al., 2008; Patrizi et al., 2004).

- The **Model-based methods** try to predict the pixel values on the basis of a probability model previously assigned to the texture. Markov random fields (Luck et al., 2005; Suliga et al., 2008) are an example of this kind of descriptors.

- **Structural methods** try to find a hierarchical structure on textures. Although they can characterise textures, they actually are more useful at image synthesis.

- **Methods based on signal processing** make modifications to the image, either using filters – *e.g.* Laws masks (Laws, 1979; Poonguzhali and Ravindran, 2006) or Gabor filters (Pok et al., 2003) – or by means of transforms such as Fourier (Arof and Deravi, 1997), Wavelet (Mangoubi et al., 2008; Semler et al., 2005), Curvelet (Candès et al., 2006; Semler and Dettori, 2006), *etc.*

2.1.1 Statistical methods for texture description

One of the simplest approaches for describing texture is to compute statistical measures from the values of its pixels. These measurements can be either first order statistical descriptors, *i.e.* moments extracted directly from the grey-level pixels or histogram of the texture, or second order statistical descriptors, which consider not only the distribution of intensities, but also the positions of pixels with equal or nearly equal intensity values.

Albregsten *et al.* characterised the textures of images of cell nuclei from mice's liver by means of the invariant moments of Hu (Hu, 1962) with the aim

of classifying them according to their pathological state (*normal*, *proliferating*, *pre-cancer* and *cancer*) (Albregtsen et al., 1995) with good results. Bharathi and Ganesan compared the efficiency of the invariant moments of Legendre and Zernike applied to the characterization of Computer Tomography (CT) images of hepatic tissue, in order to classify it as normal or abnormal (Bharathi and Ganesan, 2008). Results showed that Zernike outperformed Legendre moments.

Subashini *et al.* used first order statistical descriptors to describe the density of breast tissue in digital mammographies with the goal of classifying it into fatty, glandular or dense tissue using SVM (Subashini et al., 2010), achieving quite good accuracy.

Since Haralick proposed them in his work (Haralick et al., 1973), features derived from the Grey Level Co-occurrence Matrix (GLCM) have been widely used in tissues and cells texture description and classification. An example is the work carried out by Tsang *et al.*, where a Content-Based Image Retrieval (CBIR) system for images of different organs is presented (Tsang et al., 2005). It implements several similarity measures using texture descriptors extracted from the co-occurrence matrix of both the whole image and the neighbourhood that surrounds each pixel. Sivaramakrishna *et al.* combined Haralick descriptors derived from the GLCM with posterior acoustic attenuation descriptors for the detection of cysts, benign solid masses and malignant solid masses in 2D breast ultrasound images (Sivaramakrishna et al., 2002). Raicu and her co-workers combined textual and textural information (some Haralick descriptors and other features extracted from the run-length matrix) to build a *texture dictionary* of CT images of some organs (liver, heart, backbone, kidneys, and spleen) (Raicu et al., 2004), with the aim of storing the relationships between image headers and texture information.

In the literature there are some works that explore how to increase the texture description performance of co-occurrence matrices. For instance, Mahmoud-Ghoneim *et al.* evaluated the impact of the image dynamic range (which has a direct influence on the size of the GLCM) on the description of magnetic resonance images (MRI) of the brain, in order to detect peritumoral white matter (Mahmoud-Ghoneim et al., 2008). This work showed that dynamic range has direct influence on classification accuracy, since both sensitivity and specificity were higher when the texture was rescaled to 128 grey levels. On the other

hand, Philips *et al.* analysed the directional features of 3D co-occurrence matrices extracted from a set of Computer Tomography (CT) liver images, in order to determine the orientations which better discriminated textures and, thus, reducing the number of computed GLCMs (Philips et al., 2008).

2.1.2 Signal processing-based texture description

Methods based on signal processing are widely used in texture description, usually in combination with other approaches. Texture analysis in the frequency domain is sometimes very useful, since it makes possible to characterise some details that would not have been possible otherwise. For instance, in (Ganeshan et al., 2009) a proposal for revealing abnormalities in images of apparently disease-free areas of the liver was shown. A Laplacian of Gaussian (LoG) spatial filter was used as non-orthogonal Wavelet filter in order to separate the fine, medium and coarse details of the image. Afterwards, some statistical features were computed from each one. The images were acquired without previously using any contrast-agents on the patients, which saves time, costs, and radiologic exposure for them.

In the literature there are some applications of Gabor filters for texture description. For instance, Grigorescu *et al.* compared the discriminative power of some features extracted from the power spectrum (which was obtained after applying a bank of Gabor filters on the texture) by means of a cluster analysis (Grigorescu et al., 2002). Ahmad *et al.* proposed a method to estimate the rotation variance with the goal of comparing two texture descriptors – based on the Discrete Fourier Transform (DFT) and Gabor filters –, in terms of their sensitivity to rotations (Ahmad et al., 2007). According to their experiments, the DFT performed better when images had no noise – both in terms of classification accuracy and rotation variance –, while it was outperformed otherwise.

Discrete Wavelet Transform (DWT) has been widely studied in the literature and used in multi-resolution analysis of tissues and cells texture, outperforming traditional analysis. Texture description using the DWT is not necessarily better than using Gabor filters, but they are computationally less expensive (Livens et al., 1997).

Karahaliou *et al.* compared the performance of several texture descriptors – Laws, and first and second order statistics extracted from the original image and from its Wavelet coefficients – on the analysis of tissue surrounding

micro-calcifications on mammographies, in order to classify them as malignant or benign (Karahaliou et al., 2008). Results showed that Wavelet-based descriptors performed better. Mala and Sadasivam analysed the texture of liver CT images in order to classify tissues as fatty or cirrhotic by means of a Neural Network. They used first order statistics from the horizontal, vertical and diagonal details of the DWT (Mala and Sadasivam, 2005). Tsantis *et al.* also applied the Wavelet transform (Tsantis et al., 2009) to characterise the texture of thyroid ultrasound images with the goal of determining the malignancy of the nodules (low-risk and high-risk).

Semler *et al.* carried out a comparison between different Wavelet families – Haar, Daubechies and Coiflets – for describing the texture of tissues from CT scans (Semler et al., 2005) by means of some features extracted from the GLCMs of the Wavelet horizontal, vertical and diagonal coefficients. Results showed that the best performance in classification have been achieved when the Haar Wavelet Transform was applied and, additionally, that the best results were achieved when the feature vectors were made up of the average attributes from the three components. Bonnel and his co-workers also used the Wavelet Transform on the analysis of color images of the small bowel, in order to classify them as normal or abnormal (Bonnel et al., 2009). Specifically, they used some Haralick features extracted from the cross co-occurrence matrices of the Wavelet coefficients.

Mangoubi *et al.* used multi-resolution texture analysis for the recognition of stem cell nuclei with similar differentiation levels. The DWT was applied on the images and a Gaussian model was created using its coefficients' statistical properties.Van de Wouwer and his co-workers also worked with cell nuclei, proposing in (Wouwer et al., 2000) a method to classify cells from breast tissue according to the degree of the malignant nuclei – benign and category I, II or III – (Wouwer et al., 2000). To perform this task, they combined features extracted from the co-occurrence matrix with the energy of Wavelet coefficients.

During the last years some transforms, more sophisticated than the DWT, have been developed besides it. Do and Vetterli presented the Discrete Ridgelet Transform and detailed its application for digital image processing in (Do and Vetterli, 2003). The Wavelet transform extracts just the horizontal, diagonal, and vertical details, whereas Ridgelet computes coefficients in multiple radial directions of the frequency domain and, therefore, extracts more details from

the image. Arivazhagan *et al.* carried out an experiment which consisted in classifying several textures from the VisTex database (Vision and Texture) which were described by first and second order statistical descriptors extracted from the coefficients of the Ridgelet transform (Arivazhagan et al., 2005). Results showed that textures were better characterised by means of the Ridgelet transform than when using the DWT (Arivazhagan and Ganesan, 2003).

The effectiveness of the Ridgelet transform lies in the detection and capture of information in radial directions. However, this is a limitation when working with medical images, where linear structures are not very likely to appear. Candès and Donoho developed the Curvelet transform (Candès and Donoho, 2000), an extension of the Ridgelet transform, which captures structural information in "wedges" in frequency domain, along multiple scales, locations and orientations. The discrete version of the Curvelet transform has been widely used in digital image processing since Candès *et al.* presented it (Candès et al., 2006). For instance, Arivazhagan *et al.* extracted several feature vectors based on it and compared their performance in the classification of images from several materials (Arivazhagan et al., 2006). On the one hand they used first order statistical descriptors and features extracted from the GLCM of the Curvelet coefficients both separately and combined, to characterise the textures. On the other hand a combination of the same features were extracted from the Wavelet sub-bands. Results showed that Curvelet features performed better than Wavelet ones. Eltoukhy *et al.* used the Curvelet transform in digital mammogram analysis. They classified mammographies into normal or abnormal by characterizing them with the biggest Curvelet coefficients from each decomposition level (Eltoukhy et al., 2010a). The same authors extended this work (Eltoukhy et al., 2010b) and compared the performance of this Curvelet-based descriptor with other features extracted from the Wavelet transform in the same problem. Results showed that the former yield higher accuracy than the latter. Semler and Dettori applied the Curvelet transform to CT images of different tissues and extracted some first order statistics from its coefficients, yielding very high hit rates (Semler and Dettori, 2006). This work was later extended by the same authors, who carried out a comparison between several descriptors based on first and second order statistics extracted from the Wavelet, Ridgelet and Curvelet transforms to describe images of several tissues (Dettori and Semler, 2007). Results showed that the descriptors extracted

from the Curvelet coefficients achieved better classification accuracy than the others.

2.1.3 Texture description for sperm cells

Despite texture descriptors are very widely used in medical imaging to analyse cells and tissues, there are not, up to our knowledge, commercial systems to characterise sperm cells by means of their texture. There are, however, some experimental works that use texture descriptors to classify sperm cells, according to their vitality, or their acrosome state. Sánchez and her co-workers (Sánchez et al., 2005a,b) classified the spermatozoa as "normal" or "abnormal" (alive or dead) according to the distance between the intracellular intensity distribution of the acrosome and an intensity distribution model which was built using a set of heads labelled by veterinary experts. Following this line, these authors tried to classify the spermatozoa as alive or dead (Sánchez et al., 2006), with quite satisfactory results for veterinary practice. González and her co-workers used first order statistics and Haralick features, combined with Wavelet coefficients and classified them using Neural Networks with the goal of detecting the spermatozoa with intact and damaged acrosomes (González et al., 2007), achieving hit rates of about 92%. Alegre and his co-workers also tried to classify the spermatozoa in terms of their acrosome status (Alegre et al., 2008). They characterised heads by means of the magnitude of the gradient along their outer contours. The hit rate in the classification with Learning Vector Quantification (LVQ) was about 93%. Finally, Alegre *et al.* compared several texture descriptors with the same goal – Haralick features, Laws masks, Legendre and Zernike moments – (Alegre et al., 2009). Both supervised – kNN and Neural Networks – and unsupervised – Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) – classifications were carried out. The best accuracy was achieved by Haralick features, classified by discriminant analysis (almost 94%).

2.2 Adaptive Methods in Digital Image Processing

Adaptive texture analysis methods make possible the study of the texture according to its local features, *i.e.* the method locally *adapts* to the texture.

2.2.1 Preprocessing

Sometimes it is necessary to carry out some operations on the image before starting its analysis – *e.g.* to remove the effects of noise, improve its contrast, to deblur the image if it is blurred. *etc.*–

Bastian *et al.* proposed a skeletonization algorithm for grey-scale images which used adaptive erosions and dilations whose degree depended on the shape of the edges (Bastian et al., 1995). These morphological operations used a non-linear Structuring Element (SE) which depended on the local gradient on the images, on the basis that steep edges should be eroded faster than gentle ones.

Yu and Bajaj present a method to enhance contrast on medical images (Yu and Bajaj, 2004). In comparison to other methods, which are based on equalizing the histogram of the whole image, their proposal consisted in varying the transfer function at each pixel, depending on the maximum, minimum, and average grey level of the neighbourhood.

Takemura and Ito proposed an adaptive smoothing filter for suppressing the noise component typical of medical ultrasound images. The shape of the filter depended on the local features of the noise and the contour of the tissue inside it (Takemura and Ito, 2003). Experiments showed that the segmentation was better when images were preprocessed using this filter. Stippel *et al.* proposed an adaptive algorithm for ultrasound image enhancement that used some features of the texture with the goal of enhancing its differences and, thus, improve boundary detection between healthy neonatal brain tissue and tissue affected by periventricular leukomalacia (Stippel et al., 2005).

2.2.2 Segmentation

Some authors have presented segmentation methods which somehow adapt to the local features of the image or texture.

Jiang *et al.* carried out an adaptive X-ray image segmentation of fruit pieces in quarantine in order to detect their internal status (Jiang et al., 2008). This method computed the threshold of the image at each pixel, depending on the distribution of grey levels of the neighbourhood pixels. Lee and Tanaka segmented images by using an adaptive mesh (Lee and Tanaka, 2002). The algorithm consisted in recursively dividing the image into triangles, using homogeneity and size criteria to stop divisions. Then, these divisions were repre-

sented on a binary tree, where leaves with similar properties were merged, thus obtaining the segmented image.

One of the applications of segmentation is the binarization of written documents. Valverde and Grigat used the Niblack's algorithm with the goal of finding several thresholds on the image – one per 15×15 pixels region – and removing noise afterwards using mathematical morphology (Valverde and Grigat, 2000). Gatos and his co-workers presented an adaptive approach for the binarization and enhancement of degraded documents, which achieved good results even with the presence of noise in the image (Gatos et al., 2006). This method had three steps: First, an adaptive Wiener filter was used to preprocess the image. Then, a rough estimation of the background and foreground regions was obtained by using the adaptive method proposed by Sauvola and Pietkäinen (Sauvola and Pietikäinen, 2000). Finally, the text segmentation was carried out by means of a distance criterion – between the preprocessed image and the background estimation –.

There are some approaches based on dividing the pixels of an image using cluster analysis, according to any feature. The description of the pixels may be carried out by means of Haralick features, computed from the GLCM of a window centred at each pixel. However, the problem of choosing the window size still exists. A solution is presented in (Lee and Lee, 1992), where the window of size W was divided into n smaller sub-windows of size $w \times w$. Some statistical features are extracted from each one, discarding those whose difference with the statistics computed from a center block is higher than a threshold. Tsao and his co-workers used a Fuzzy Kohonen neural network to carry out the recognition of those patterns (Tsao et al., 1994). However, the number of output neurons should be the same as the number of regions of interest within the image, which is a drawback, as this information may not be known a priori. Therefore, Lei and Feihu proposed an algorithm that automatically computed the number of regions the image had – which determine the number of output neurons – and also initial weights for the network, which made the convergence of the network faster (Lei and Feihu, 1999).

2.2.3 Description

Some authors have proposed adaptive methods focused on texture description.

In description problems, identifying which features have the most discriminative power for the images under study would be very suitable. Walker and his co-workers have tackled this problem. They proposed some approaches whose goal was to extract the areas with the highest discriminative power inside a multi-scale GLCM and then to compute features from these areas. One of them used a distance criterion between the matrices of the different classes (Walker et al., 1997), and the other one used genetic algorithms (Walker et al., 2003). Both found these areas using a training set of images. Further details can be found in (Walker, 1997). Following this line, Nielsen *et al.* presented a method to compute distance and difference matrices between classes, by using probability matrices (*e.g.* co-occurrence, run-length) extracted from a training set of images (Nielsen et al., 2004). They also proposed the extraction of two descriptors from them, which are able to discriminate images with good performance, according to the experiments.

These approaches require a training set of images in order to get the *best areas* in the co-occurrence matrices. However, there is not always such a priori knowledge about the images that have to be characterised.

Rohrmus proposed some geometrical texture features invariant to rotation, scale, and translation in order to detect defects in textures of materials (Rohrmus, 2005). He presented an adaptive improvement, which consisted in optimizing the feature space by means of a function which measures simultaneously the intra-class compactness and all inter-class distances. Hou and Parker proposed other approach for defect detection on textures of surfaces consisting in describing them by means of an adaptive variation of Gabor Wavelet filters which selects the best ones (Hou and Parker, 2005). Once again, this method needed a training set.

Huang *et al.* presented an adaptive approach to classify textures which separated the indeterministic – high frequency components – from the deterministic parts – low frequency components –of the texture by means of the Fourier transform (Huang et al., 2000). These parts are described separately by means of Gabor functions and Markov Random Fields, respectively.

2.3 Mathematical Morphology

Texture description has been used in many fields in order to automatically characterise and recognise objects. Describing a texture of a region of interest (henceforth, a ROI) means extracting some mathematical features to represent it. Although these techniques are applied as-are on the whole ROI, a texture may not be homogeneous, as sometimes there are areas with different properties inside, or even among different elements of the same class.

Mathematical morphology is a set of processes which can be applied to an image in order to remove details which are smaller than a reference set called *Structuring Element* (SE). Matheron (Matheron, 1975) and Serra (Serra, 1982) established its basis – focused on binary images –, which was later generalised to grey-scale images by Sternberg (Sternberg, 1986). Afterwards, Haralick *et al*. reviewed binary and grey-scale morphological operations as well as their properties and relations (Haralick et al., 1987).

From that moment on, Mathematical Morphology (MM) has been used with different purposes (*e.g.* image restoration, enhancing, segmentation or even texture description) in a wide range of applications; biomedicine (Bouraoui et al., 2010; González-Castro et al., 2009; Phukpattaranont and Boonyaphiphat, 2006; Said et al., 2006; Theera-Umpon and Dhompongsa, 2007), shape analysis (Yang and Li, 1995; Yu and Wang, 2005), industrial inspection (Qi and Yu, 2008), signal processing (Maragos and Schafer, 1987a,b) or geoscience (Pina et al., 2001; Soille and Pesaresi, 2002; Valero et al., 2010) are just some fields where this theory has been applied.

Regarding texture description, Maragos reviewed the basic concepts and operations of grey scale mathematical morphology and developed a shape-size descriptor called *Pattern Spectrum* (PS) (Maragos, 1989), generalizing the probabilistic size distributions proposed by Matheron for binary images (Matheron, 1975). The pattern spectrum of a texture is defined as a function of the size distribution of objects – "granules" – inside it, which is computed by making successive erosions and dilations with structuring elements of different sizes. It is also called *granulometric distribution function*, and the process of computing it is called *granulometry* (Petrou and Sevilla, 2006).

A considerable drawback for this method is that it is not clear which is the best structuring element for a certain problem or how to choose it (de Ves et al.,

2006), as some authors have disclosed. Even if it was decided which one to use, it has always the same shape, so possible variations within the ROI would not be considered. Huet and Mattioli presented a method to generate a minimal set of structuring elements that left the texture invariant and carried out some mathematical morphology transformations for texture defect detection tasks (Huet and Mattioli, 1996). Asano and his co-workers proposed to find the best Structuring Element to compute pattern spectrum of a texture (Asano, 1999; Asano et al., 2003), assuming that it is more uniform when the shape of the SE is similar to the form of its granules. Therefore, the shape of the Structuring Element was fixed under the criterion of reducing iteratively the variance of the size distribution.

These works deal with making a selection of the structuring element having an a priori knowledge about the types of images that will be described. Nevertheless, such knowledge is not always available in real computer vision problems. In addition, these approaches do not solve the problem of using fixed structuring elements on non-uniform textures.

Shih and Cheng presented an adaptive edge-linking method, based on mathematical morphology, which used an elliptical SE whose orientation and size were adjusted according to some local features (Shih and Cheng, 2004). Experimental results show the success of the method. Bouaynaya *et al*. presented in (Bouaynaya et al., 2006) an approach for image restoration and skeletonization which used spatially-variant mathematical morphology (Charif-Chefchaouni and Schonfeld, 1994), which involves varying the SE in size, orientation and shape within the image. This theory is presented in more detail in (Bouaynaya et al., 2008) and (Bouaynaya and Schonfeld, 2008) for binary and grey-scale images, respectively.

Cuisenaire defines morphological operators which use structuring elements whose size vary over the image, depending on a distance transform (DT) and provides a method to efficiently compute it (Cuisenaire, 2006). As an Euclidean DT is considered, the shape of these SEs is circular.

2.4 Geodesic Distance and Fast Marching Algorithm

Cuisenaire showed how common MM operators could be adapted so that the structuring element could vary across image pixels. He used an Euclidean distance transformation of binary images to make the SE vary across them. Cuisenaire himself presented and reviewed some distance transformations in his Ph.D. Thesis (Cuisenaire, 1999), one of which was the Geodesic Distance Transformation (GDT), studied in chapter 8 of that Thesis. He defined the geodesic distance between two pixels as the length of the shortest path among those which join them. This is very promising, as it captures the global non-linear structure and the intrinsic geometry of a surface, which is not the case of Euclidean distance, as Hamza and Krim pointed out in (Hamza and Krim, 2006). In this work, they proposed an object matching approach to characterise objects by means of a descriptor based on the distribution of geodesic distances between points on their surface. Berretti *et al.* presented a method for face recognition which enables direct comparisons between 2D face images against 3D face models (Berretti et al., 2007). Face representation has been carried out by measuring geodesic distances in 2D and 3D. Liang and Zhang used geodesic distances to develop a scheme to approximate the complex structure of the cingulum tracts and inspect possible tissue damages caused by regional micro structural changes in the White Matter of the brain (Liang and Zhang, 2008). Geodesic distance is used here with the goal of extracting measures of the pathways along the fibres.

Alternative geodesic distance transformations have also been proposed. For instance, Cárdenes and his co-workers presented one, based on a so-called occlusion points propagation (Cárdenes et al., 2003), which proved to be better – both in terms of accuracy and computational complexity – than the circular propagation proposed by Cuisinaire in (Cuisenaire, 1999). Later, Cárdenes *et al.* proposed other GDT approach, based on an algorithm for propagating what they call Locally Nearest Hidden Points (Cárdenes et al., 2010). Both methods were proposed for binary images.

On the other hand, Osher and Sethian developed some algorithms to follow fronts propagating on surfaces whose curvatures determine the propagation speed (Osher and Sethian, 1988). These algorithms were later used by Malladi

et al. for shape modelling purposes (Malladi et al., 1995). Sethian developed an algorithm for front propagation called Fast Matching, presented in (Sethian, 1996), which was later extended in (Sethian, 1999), where some applications of this algorithm were shown. Front propagation and, particularly, fast marching algorithm can be used to compute geodesic paths, as it was pointed out in (Cohen and Kimmel, 1996), where a boundary detection approach for shape modelling is presented. It is based on interpreting the snake as a path of minimal length in a Riemannian metric, which is computed using the Fast Marching algorithm. This method makes snake initialization an easier task, and it overcomes one of the active contour models' main drawbacks: being trapped by local minima.

Peyré and Cohen reviewed the numerical computation of geodesic distances on Riemannian manifolds, along with some of its applications (Peyré and Cohen, 2009). Amongst them, they show an approach of how to use the Fast Marching algorithm to compute both geodesic distance maps and shortest paths. Some years before, the same authors proposed method to segment surfaces which used fast marching algorithm to perform centroidal tesselation (Peyré and Cohen, 2004).

Cardinal and her co-workers proposed a segmentation approach for intravascular ultrasound images – which often have low quality and sometimes shadows produced by the catheter – of blood vessel walls (Cardinal et al., 2006). This approach used the fast marching algorithm for propagating regions and, therefore, segment the different areas of the blood vessel walls. Andrews and Sethian dealt with solving the travelling salesman problem in domains where the cost associated with passing through each point is not constant (Andrews and Sethian, 2007). If this cost was considered as a speed function, the problem could be seen as a front propagation one, as Osher and Sethian stated, so fast marching algorithm has been used to solve it. This approach was compared by a previous optimal method, and results showed that the fast marching based method was more efficient in terms of cost. Wu *et al.* presented an improved fast marching method to perform the contour tracking of small intestine segments in the frames of a sequence of cinematic-magnetic resonance imaging (Wu et al., 2009). This improvement consisted in using a Gaussian filter, as the images sometimes have low contrast in the borders of the segments.

2.5 The Problem of Shifts in Class Prior Probabilities

A typical supervised learning problem deals with the extraction of the best possible features from a set of labelled instances in order to train a classifier. Once it is designed, it is applied as-is to new data in order to predict the class to which each individual belongs to. This process has been widely studied and is called *supervised classification*. It is often assumed that training and future (test) data follow the same distribution (Duda et al., 2001), *i.e.* class prior probabilities estimated from the training data set are considered to truly reflect the class distribution of the operational environment.

However, this assumption does not always hold, as time or space stationarity is not guaranteed in many practical fields. For example, if a word sense disambiguation system is trained using words from a certain domain (*i.e.* sports news), but it is then used with instances from a different domain (*i.e.* political news), where the sense priors are different, the accuracy will be affected (Chan and Ng, 2005, 2006). Remote sensing applications also suffer from that problem. For instance, think of a dataset collected in a given season from a region with different terrains (industrial, hay, wheat, corn, grass, ...) which is employed to train the classifier, but when that classifier is deployed, mismatches in terrain distribution may appear just because seasonal or location changes (Guerrero-Curieses et al., 2009). Another illustrative example is direct mail marketing as the target costumers proportion or customer profile may vary between different areas.

It is well known that a mismatch between the actual class prior probabilities and those the classifier has been optimised for, leads to suboptimal solutions. Whenever there is such change, some authors rely on an eventual perfect knowledge of the new conditions by the end user (Drummond and Holte, 2006), but when this is not possible, estimating new class proportions is important to adapt the classifier to the new context (Saerens et al., 2002). Classifier adaptation to new operating conditions is a problem that has recalled high attention lately from several perspectives, with the goal of improving the individual classification performance. This problem becomes clearer in the case of *online* classification tasks – where data is provided in a stream –. Yang and Zhou propose an online incremental variation of the EM algorithm to estimate

class priors along a data sequence, once again with the goal of readjusting the classifier (Yang and Zhou, 2008). This method is computationally efficient and is quite effective in improving the classification accuracy. This work has been later extended in (Zhang and Zhou, 2010), where the authors point out that current algorithms do not perform well due to the lack of samples from the target distribution, so their approach makes use of data belonging not only to the target but also to similar distributions. Duch and Itert (Duch and Itert, 2002) proposed three types of procedures to correct the a posteriori probabilities provided by classifiers, with the aim of (a) increasing the overall accuracy, (b) restoring the balance between the training and test sets – if necessary – or (c) improving the confidence of classification. Xue and Weiss (Xue and Weiss, 2009) proposed some quantification methods that use estimates of the unlabelled data to adjust the original classifier. They also presented two semi-supervised classification approaches – which build a new classifier using the predictions made over the unlabelled data set – and finally, a hybrid method combining both of them.

Pérez and Sánchez-Montañés proposed in (Pérez and Sánchez-Montañés, 2007) a method to re-estimate the parameters of a statistical model when class proportions vary from the training to the test set, with the goal of using the new model to build a classifier and apply it to predict new samples. This re-estimation process is carried out by means of a variation of the EM algorithm (Dempster et al., 1977). In contrast, Xing *et al.* showed an algorithm which concentrates on refining the classification labels instead of the classification model, reducing the error rate (Xing et al., 2007). Saerens *et al.* proposed an iterative re-estimation process, based on the EM algorithm, for classifiers that provide estimates of the posterior probabilities of class membership (Saerens et al., 2002). Based on the new estimated conditions (priors), the classifier is adapted in order to minimise the error rate.

In order to overcome the problem of unknown class distributions of the test sets, Alaiz-Rodríguez and Cid-Sueiro used a minimax strategy on the learning of neural networks (Alaiz-Rodríguez and Cid-Sueiro, 2002). It consists in design a classifier that minimises the maximum error probability under the worst case conditions in a way that, even under changes on the class priors, the classifier guarantee an upper bound of the error rate. It is called *minimax* classifier. Guerrero-Curieses and his co-workers developed an algorithm

to train a minimax neural network (Guerrero-Curieses et al., 2004). Results showed that a network trained following this strategy produces a maximum error rate lower than standard networks do. Besides, the minimax network is almost invariant to changes in the distributions of the training sets, which is not guaranteed by the methods based on the re-estimation of a priori probabilities. Alaiz-Rodríguez *et al*. continued this study and proposed in (Alaiz-Rodríguez et al., 2005) two different training algorithms called *gradient-based* and *learning rate scaling*. However, the minimax approach may seem too conservative, as its goal is to optimise the performance under the *least* favourable conditions, *e.g*. following a pure minimax strategy in a marketing application can lead to solutions where minimizing the maximum loss could entail considering there are no potential clients. Therefore, Alaiz-Rodríguez and her co-workers proposed other method for training the neural network, called *minimax deviation* (Alaiz-Rodríguez et al., 2007).

Detecting failures in classifier performance due to shifts in the data distribution has been receiving attention in the machine learning community. In particular, Cieslak and Chawla (Cieslak and Chawla, 2009) have shown that the measure of the Hellinger Distance is very effective in detecting breakpoints in classifier performance due to shifts in class prior probabilities.

It should be noted that there are other applications where class proportions are subject to high variability and its estimation is itself valuable. To the best of our knowledge, there are few works which address with this problem of quantification. Some examples are news categorization (Forman, 2006, 2008), identify frequent issues from unstructured free-text fields of technical support logs (Forman et al., 2006), word text disambiguation systems (Chan and Ng, 2005, 2006), or artificial insemination techniques (Alaiz-Rodríguez et al., 2008; González-Castro et al., 2010; Sánchez et al., 2008).

Quantification techniques proposed in the literature are either based on the classifier confusion matrix (Forman, 2006, 2005; Vucetic and Obradovic, 2001) or on posterior probability estimations provided by the classifier (Alaiz-Rodríguez et al., 2008; Chan and Ng, 2006; González-Castro et al., 2010). Forman has also explored an approach based on the estimation of the class conditional probability densities (Forman, 2008), but it turned out to be outperformed by simple methods that rely on the confusion matrix.

To sum up, estimating the class prior probabilities of an unlabelled dataset plays an important role: (i) In supervised learning in order to detect fractures in classifier performance due to shifts in class prior probabilities (assuming that class conditional densities are fixed) and with the purpose of adapting the classifier to the new operational conditions whenever it is possible. (ii) In applications where the class distribution shows a high variability and its estimation has practical interest.

CHAPTER 3

METHODOLOGY

3.1 Boar Semen Quality Assessment

Semen quality assessment is a crucial task in Artificial Insemination (AI) processes both in medicine and veterinary. Focusing on the latter, AI provides several advantages. On the one hand, farmers can work with a reduced number of animals, as fertility of males and females can be controlled and, on the other hand, it allows to get better individuals each generation, since semen samples are optimal and, thus, genetic improvements can be maximized. In summary, thanks to Artificial Insemination it is possible for farms to maximize the quality of their animals saving money meanwhile.

Regarding boar semen, Rozeboom stated four factors related to sperm quality (Rozeboom, 2000): concentration, morphology, motility, and acrosome integrity. The first one is the number of spermatozoa per ejaculation. It is not just a component of semen quality evaluation procedures, but a tool to check the health and productive output of the boar. However, automatic procedures provide high errors due to debris and other particles which are usually confused with spermatozoa. Nevertheless, this parameter just assess the sperm number, but not their viability, *i.e.* a good number of spermatozoa per ejaculate is necessary, but not enough, to assure good fertilizing potential.

Motility has been a common parameter to assess sperm viability. Although there is disagreement between authors when they try to determine its relation with fertility (Rodríguez-Martínez, 2003), other works that study this parameter using CASA systems did find a correlation between motility features and the outcomes of in-vivo fertilization (Farrell et al., 1998; Holt et al., 1997).

Morphology is other relevant parameter in the evaluation of ejaculation quality. Once again, the correlation of this parameter with fertility vary hugely (Rodríguez-Martínez, 2003) when it is assessed manually. Therefore, there are some works that use digital image processing to make objective estimations of morphology in humans (Ramos et al., 2002), horses (Hidalgo et al., 2008), alpacas (Buendía et al., 2002), *etc*. Belletti *et al*. even characterized spermatozoa from several species by means of their morphometric parameters, computed using spectral methods (Beletti et al., 2005). A spermatozoon has usually three main parts: head, mid-piece and tail, as shown in Fig. 3.1.

Finally, veterinary experts believe that there is a relation between sperm fertility potential and the structural integrity of the acrosome, since it is essen-

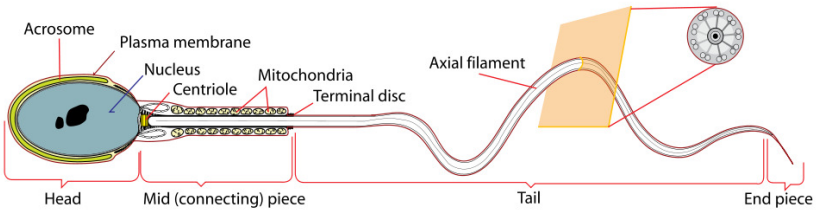


Figure 3.1: Diagram with the structure of a spermatozoon.

tial for the sperm to penetrate the ovum (Yanagimachi, 1994). The acrosome is basically a cap-like-structured vesicle that covers the anterior portion of the spermatozoon's head (see Fig. 3.2(a)) and contains a variety of enzymes that digest proteins and complex sugars. This material is secreted when the spermatozoon contacts the egg in a process known as capacitation, or acrosomal reaction. During this process the acrosome fuses with the sperm outer membrane (Fig. 3.2(b)), releasing its contents (Figs. 3.2(c), 3.2(d) and 3.2(e)). Fig. 3.2(f) shows a spermatozoon which has completely lost its acrosome. Acrosome reaction of a sperm sample can be assessed by flow cytometry or fluorescence microscopy, after staining it.

The enzymes released during capacitation break down the outer membrane of the ovum – called zona pellucida –, creating a thin path that allows the penetration of the spermatozoon into the egg and the joint of its haploid nucleus with the haploid nucleus of the egg. Fig. 3.3 shows a diagram of the stages of the acrosome reaction process as it happens in the Sea Urchin, which is analogous to mammals. We address the reader interested in further details about fertilization in mammals to (Wassarman et al., 2001).

In summary, the loss of a spermatozoon's acrosome integrity previous to the introduction into the female genital tract, or in an early stage of its transit involves that it will be unable to fertilize the egg (Silva and Gadella, 2006). Therefore, if a semen sample has a high fraction of spermatozoa with any damage in the acrosomal vesicle, it will have low fertilizing capacity and, thus, will be useless for artificial insemination.

The plasma membrane, which completely surrounds the spermatozoon, separates it from the external environment and develops many physiological functions which maintain the cell in optimal conditions and preserve its vitality.

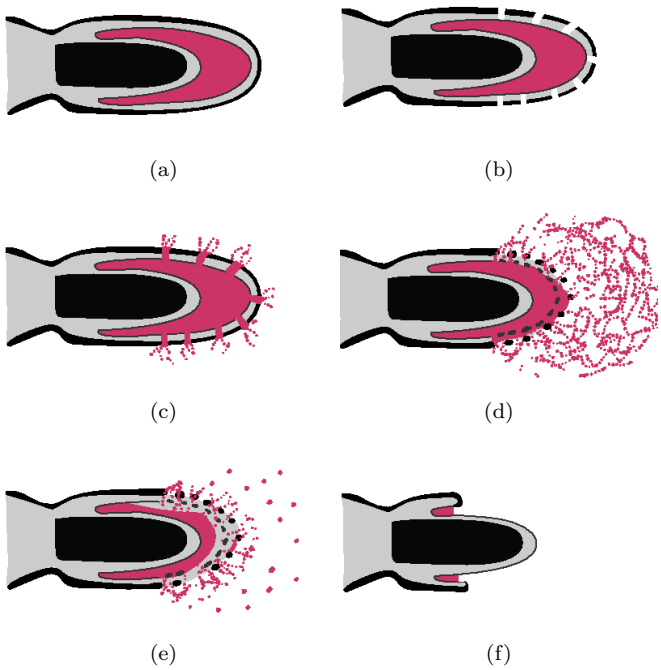


Figure 3.2: Acrosomal reaction of a spermatozoon.

By means of its semi-permeable features it is responsible for maintaining the chemical gradient of ions and other soluble components (Silva and Gadella, 2006) and for the transport of the glucose and fructose, which serve as energy source substrates, from the external environment into the cell. In addition, it is crucial in the fecundation stage, since it interacts with the female reproduction organs' cells and with the ovum. If the plasma membrane is not functionally intact the sperm cannot maintain its intracellular concentrations, nor produce the energy necessary for sperm movement. All these alterations would trigger the death of the spermatozoon and, thus, it would not be capable to fertilize in vivo.

3.1.1 Sample preparation for detecting acrosome integrity

In this section we are presenting a brief description about the preparation of ejaculated boar sperm to make a fluorescent stain with lectin FITC-PNA

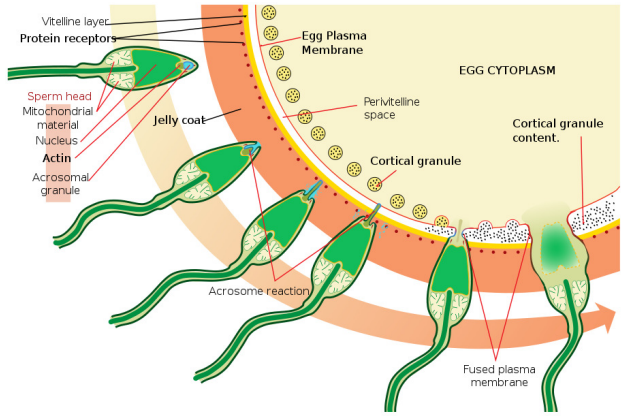


Figure 3.3: Diagram with the stages of the sperm penetration processs.

(*Arachis hypogea* marked with fluorescein isothiocyanate) whose goal is to determine if the sperm acrosome is structurally intact or damaged. When an acrosome is reacting its outer membrane and the plasma membrane fuse, thus releasing its contents and leaving the spermatozoon protected by just the internal acrosomal membrane.

When this happens, the *Arachis hypogea* lectin adheres the outer acrosomal membrane, specifically to the D-Galactose present on it. The fluorescein isothiocyanate (which, in this case, is attached to the lectin) is a fluorochrome that emits a green fluorescent light at a wavelenth of 521 nm when it is excited by a beam with a wavelength of 495 nm (de Carvalho Bessa, 2005).

Before carrying out this staining, diluted ejaculated doses must be preserved in a cold store, at a temperature of 15 °C, in order to keep all the spermatozoa physiological functions as good as possible (in terms of motility and vitality) and to preserve their acrosome and plasma membrane integrity in the best possible conditions.

First of all, an aliquot with 500 μl of diluted sperm is taken. Alfterwards, 5 μl of formaldehyde (< 0.3 %), 25 μl of FITC-PNA lectin, and the aliquot are added into an Eppendorf® tube. The formaldehyde paralyses the movement of the spermatozoa in order to make possible a better capture of the images.

The *Arachis hypogea* lectin marked with the fluorescein isothiocyanate (FITC-PNA) must be conserved under freezing and darkness, with the goal of preventing the fluorescein isothiocyanate to lose its capability of emitting fluorescence. Before it is used in the stain, this lectin must be tempered at 37 °C a few seconds. Once the sperm, the formaldehyde and the fluorescent lectin are together, an incubation must be carried out, in a bath tempered at 37 °C during 9 minutes. Afterwards, this sample must be stored in the darkness to avoid the fluorochrome from losing its fluorescence emitting capability (Cross and Meizel, 1989).

Then, a drop 3.5 μl of this incubated sample is put on a microscope slide and then a cover slip is put over it. It is necessary to previously clean both with alcohol, in order to remove all impurities from the glass. It is necessary to wait a while, to let the drop to distribute uniformly and the spermatozoa to be placed in just one plane under the cover slip. Thereafter, if we want to observe the sample with 100x magnification, a drop of immersion oil is put over the cover slip, and the phase plate condenser is rotated to position "Ph3". If the images are acquired with a magnification of 40x, the immersion oil is not necessary, and the position in the condenser must be "Ph2".

3.1.2 Sample preparation for detecting sperm vitality

This section briefly describes the preparation of semen samples to make a fluorescent stain. This allows to carry out an assessment of the boar spermatozoa plasma membrane integrity, *i.e.* the sperm vitality.

This staining is made using propidium iodide (PI) and carboxyfluorescein diacetate (CFDA). The PI is a fluorescent molecule used to stain cells DNA, and it is also a membrane-impermeable molecule, *i.e.* if the plasma membrane of an spermatozoon is intact the PI cannot penetrate it. On the contrary, if the membrane is somehow damaged, the propidium iodide goes through it and the cell DNA will by stained, thus emitting red fluorescence at a wavelength between 562 and 588 nm when it is excited with a 488 nm beam (Garner and Johnson, 1995).

The CFDA is a non-fluorescent molecule and is membrane-permeable, *i.e.* it goes through the plasma membrane even though it is completely intact. When the CFDA penetrates a spermatozoon with intact plasma membrane it suffers the action of intracellular esterases, becoming carboxyfluorescein, which is a

fluorescent molecule which emits green fluorescence, at a wavelength of 517 nm, when it is excited with a beam of 492 nm (Marti et al., 1998).

In summary, when the plasma membrane is intact (the spermatozoon is alive), just the CFDA penetrates into the sperm cell, emitting green fluorescence, but when the membrane is damaged (the spermatozoon is dead), both the PI and the CFDA penetrates it, thus fluorescing red, since due to the filter (B-2A EX 450-490, DM 505, BA 520) that is the predominant emitted radiance.

Once again, diluted ejaculated doses must be preserved in a cold store, at a temperature of 15 °C, before carrying out the staining, with the same purposes as mentioned in previous section.

First of all, an aliquot with 500 μl of diluted sperm is taken. Alfterwards, 5 μl of formaldehyde (< 0.3 %) to paralyse the movement of the spermatozoa, 10 μl of PI, 10 μl of CFDA, and the aliquot are added into an Eppendorf® tube.

Both the PI and the CFDA must be conserved frozen and in the darkness, with the goal of preventing them to lose its fluorescence-emitting capability under the excitation of a beam of 488 nm and 492 nm, respectively. Before it is used in the dye, they must be tempered at 37 °C a few seconds. Once the sperm, the formaldehyde, the PI and the CFDA are together, an incubation must be carried out, in a bath tempered at 37 °C during 9 minutes. Afterwards, this sample must be stored in the darkness to avoid the IP and CFDA from losing their fluorescence emitting capability.

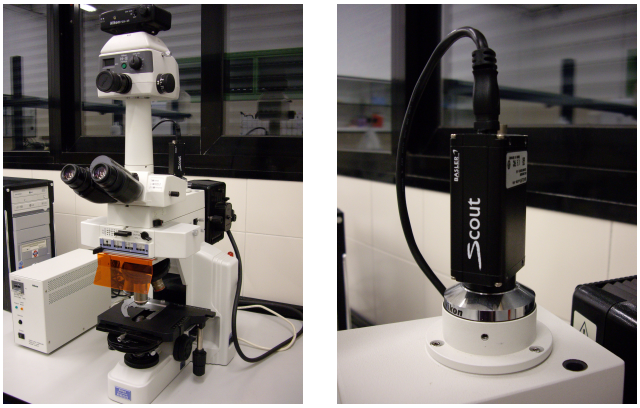
Then, the rest of the process is carried out the same way it was done with the intact-damaged preparation (section 3.1.1).

3.2 Image Acquisition

All the boar sperm images used in this Thesis have been captured in CENTROTEC, an Artificial Insemination Centre which is a spin-off of the University of León, under the guidance of veterinary experts. All semen samples were obtained from boars of three different breeds: Piyorker, Large White and Landrace.

These images have been acquired using a camera Basler Scout scA780-54fc (Fig. 3.4(b)), except for the ones used in the segmentation experiment (chapter 4), which were taken by a camera AVG Oscar F-810C, as this experiment was

designed and performed before the Basler camera was acquired. It is connected on the one hand to a computer with a specific software to control the functions of the camera and, on the other hand, to an epifluorescence microscope Nikon E-600 (Fig. 3.4(a)), which makes possible to view both fluorescence and phase contrast images of the samples.

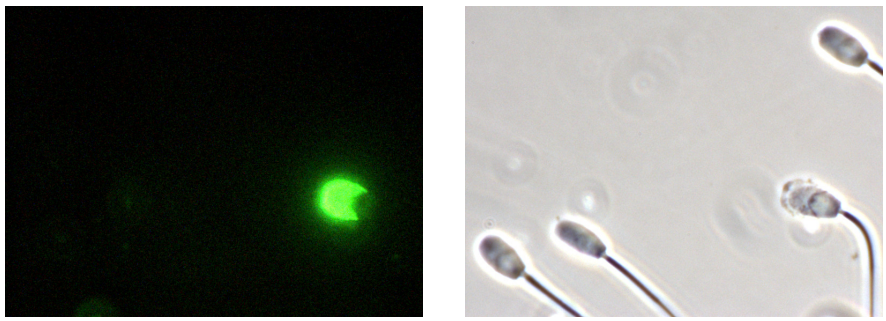


(a) (b)

Figure 3.4: Epifluorescence microscope (left) with digital camera Basler scA780-54fc used to capture sperm images (right).

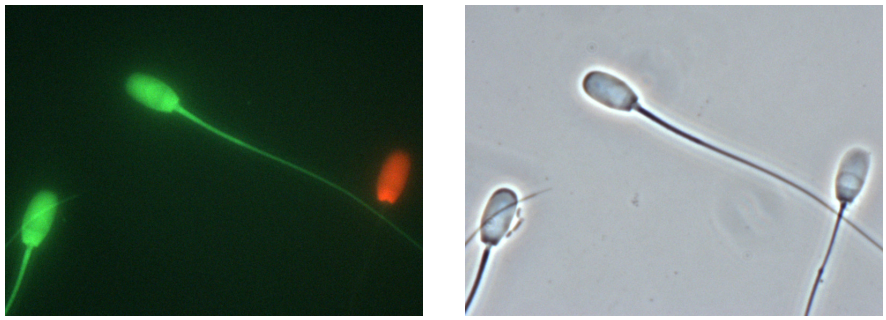
The magnification of the microscope has been set to 100x, so usually no more than three or four heads have been acquired each time. Before observing the sample under the fluorescent light, we turned on the visible light power supply and placed the visible light filter DIA-ILL to observe the spermatozoa in positive phase contrast. Once the lens was in focus an image was captured. Straight afterwards, the visible light was turned off, the fluorescence filter B-2A EX 450-490, DM 505, BA 520 was placed, and the fluorescent light was turned on to take another snapshot. Therefore, each sample produced two images with a resolution of 780×580 pixels.

The proposed approaches are intended to use just grey-level images, so only the phase-contrast images will be used in the experiments. However, we first need a ground truth that has been obtained using the colors of the fluorescent images for labelling the heads. Indeed, the reacting (acrosome-damaged) spermatozoa fluoresces bright green, while the intact ones emits no fluorescence (Gardón et al., 2001), as shown in Fig. 3.5, due to the preparation



(a) (b)

Figure 3.5: Sperm sample with intact and damaged acrosomes under fluorescence illumination (left) and phase contrast (right).



(a) (b)

Figure 3.6: Sample with alive and dead spermatozoa under fluorescence illumination (left) and phase contrast (right).

described in section 3.1.1. Also, using the fluorescence filter B-2A EX 450-490, DM 505, BA 520, both the alive and dead spermatozoa fluoresces green and red, respectively, as the beam excites the IP and the carboxyfluorescein at the same time. If the filter UV-2A EX 330-380, DM 400, BA 420 was used, only the alive spermatozoa would be fluoresced (green). If the used filter was the G-2A EX 510-560, DM 575, BA 590, only the dead spermatozoa could be perceived.

3.3 Texture Description

3.3.1 Second order statistical texture descriptors. Co-occurrence matrix

Histogram-based texture descriptors only use information about the distribution of grey level intensities. However, the relative position and relationship between the pixels of the texture are a very valuable information which is lost when first order statistics are used. In contrast, the grey level co-occurrence matrix – GLCM – (Haralick et al., 1973) combines both statistical and structural methods to extract features from the image. Let P be a position operator, defined by means of a distance d and an orientation θ, and let C be a matrix with size $N \times N$ (where N is the number of grey levels which are present in a texture), whose elements $C(k,l)$ are the number of pairs of pixels in the image which have grey levels k and l that are at a distance d apart and at the same time the line that connects them form an angle θ with the reference direction. Petrou and García-Sevilla defined the GLCM (Petrou and Sevilla, 2006) as:

$$C(k,l)_{d,\theta} = \sum_i \sum_j \delta(k - g(i,j))\delta(l - g(i + d\cos\theta, j + d\sin\theta)) \tag{3.1}$$

where $g(i,j)$ is the intensity of the pixel (i,j) of the image and $\delta(a-b)$ is a function which yields value 1 if $a = b$, or 0 otherwise.

An example of the computation of the co-occurrence matrix is shown in Fig. 3.7.

The normalized co-occurrence matrix, c is obtained by dividing the original GLCM by the number of pairs of pixels that satisfy P – which is the sum of its elements, n –. Therefore, each element in that matrix, $c(i,j)$, is the joint

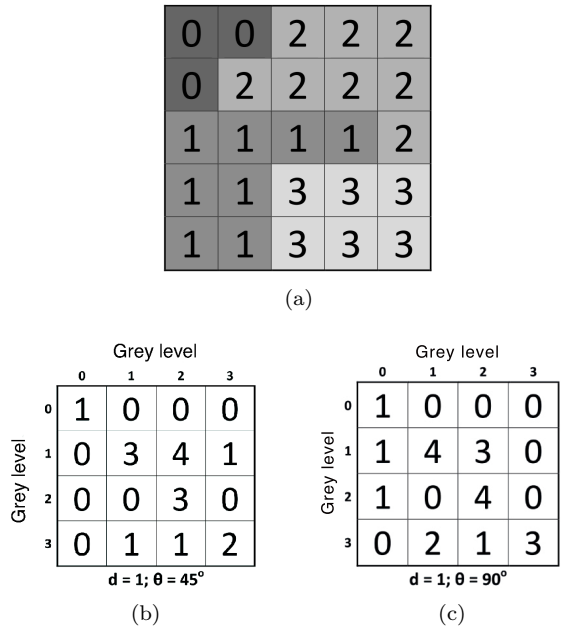


Figure 3.7: Co-occurrence matrices (down) extracted from an image (up).

probability for a couple of pixels in the image that satisfy the condition P to have grey level values g_i and g_j.

Since c is a joint probability density function, some statistics can be computed from it to characterize an image. Haralick *et al.* proposed 14 features, which can be found in the appendix I of (Haralick et al., 1973).

The GLCM is usually computed in four directions – 0°, 45°, 90° and 135° –, and then the features are extracted from each one. Hence, a set of four values are obtained for each measure, which are finally averaged in order to make them invariant to rotation.

3.3.2 Curvelet transform

The continous Curvelet transform can be defined by means of a pair of windows $W(r)$ and $V(t)$, which are a radial and an angular window, respectively. Considering W as a frequency-domain variable and r and θ as polar coordinates in the frequency domain,

$$\sum_{j=-\infty}^{\infty} W^2\left(2^j r\right) = 1, \qquad r \in \left(\frac{3}{4}, \frac{3}{2}\right) \tag{3.2}$$

$$\sum_{l=-\infty}^{\infty} V^2\left(t - l\right) = 1, \qquad t \in \left(\frac{-1}{2}, \frac{1}{2}\right) \tag{3.3}$$

Now, let U_j be a window in Fourier domain. It is a polar "wedge" supported by W and V whose size depends on the scale in each direction. It is defined by:

$$U_j(r,\theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right) \tag{3.4}$$

Let $\varphi_j(x)$ be the "mother" Curvelet, defined by means of its Fourier transform $\hat{\varphi}_j(\omega) = U_j(\omega)$. Let $\theta_l = 2\pi \cdot 2^{-\lfloor j/2 \rfloor} \cdot l$ the angles of orientation where $0 \leq \theta_l < 2\pi$. And let $k = (k_1, k_2) \in \mathbb{Z}^2$ be the position parameters. The Curvelet transform at scale 2^{-j}, orientation θ_l and position $x_k^{(j,l)} = R_{\theta_l}^{-1}\left(k_1 \cdot 2^{-j}, k_2 \cdot 2^{-j/2}\right)$ can be defined by:

$$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_l}(x - x_k^{(j,l)})\right) \tag{3.5}$$

where R_θ is the rotation in an angle θ, in radians.

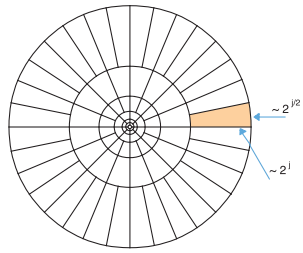


Figure 3.8: Representation of the continous Curvelet transform tiling.

A representation of the polar "wedges" U_j is illustrated in Fig. 3.8. We address the reader interested in further details to (Candès and Donoho, 2000).

However, we are working with digital images, so we need to use the discrete version of the Curvelet transform. This variant is linear and it takes cartesian

matrices as an input. These matrices have the form $f[t_1, t_2]$, where $0 \leq t_1, t_2 < n$.

The implementation of the Discrete Curvelet Transform (DCT) has been carried out using the "wrapping" algorithm, described in (Candès et al., 2006). This algorithm translates the curvelets at each scale and angle into a rectangular grid (see Fig. 3.9).

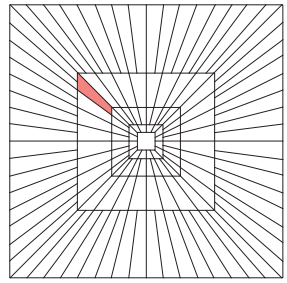


Figure 3.9: Representation of the discrete Curvelet decomposition.

The rotations and coordinates defined for the continuous Curvelet transform cannot be applied as is to Cartesian matrices. Therefore, it is necessary to replace those to their "Cartesian" equivalents. For example the radial window $(W_j)_{j\geq 0}$, $W_j(\omega) = W(2^{-j}\omega)$ would be:

$$\tilde{W}_j(\omega) = \sqrt{\Phi^2_{j+1}(\omega) - \Phi^2_j(\omega)}, \qquad j \geq 0 \tag{3.6}$$

Assume that V still follow the rule of equation (3.3). Then, it is redefined as:

$$V_j(\omega) = V(2^{\lfloor j/2 \rfloor}\omega_2/\omega_1) \tag{3.7}$$

Now, both $\tilde{W}_j$ and V_j can be used to define the cartesian window $\tilde{U}_j$ by:

$$\tilde{U}_j(\omega) = \tilde{W}_j(\omega)V_j(\omega) \tag{3.8}$$

The wrapping algorithm that computes the discrete Curvelet transform follows 4 steps:

1. The 2D Fast Fourier Transform (FFT) is applied on the image, so $\hat{f}[n_1, n_2]$ where $\frac{-n}{2} \leq n_1, n_2 < \frac{n}{2}$ are obtained.

2. For each scale j and angle l the product $\tilde{U}_{j,l} \cdot [n_1, n_2]\, \hat{f}\,[n_1, n_2]$ is computed.

3. This product is wrapped around the origin, obtaining $\tilde{f}_{j,l}\,[n_1, n_2] = W(\tilde{U}_{j,l}\hat{f})\,[n_1, n_2]$, where $0 \leq n_1 < L_{1,j}$ y $0 \leq n_2 < L_{2,j}$, and $L_{1,j}$ y $L_{2,j}$ are the dimensions x and y of the rectangle which supports $\tilde{U}_j$.

4. Finally the inverst 2D FFT to each $\tilde{f}_{j,l}$ to obtain the discrete Curvelet coefficients $c^D(j, l, k)$.

Once again, we address the reader interested in further details about the discrete Curvelet transform to (Candès et al., 2006).

3.3.3 Mathematical morphology

Some of the basic operators of mathematical morphology and their main properties will be reviewed in this section. We address the readers interested in further details about these concepts to (Sternberg, 1986), (Haralick et al., 1987) and (Maragos, 1989).

Let $f(x, y)$ be a finite-support grey-scale function on $\mathbb{Z}^2$, and let $G(x, y)$ be a grey-scale structuring element. Then, the erosion and dilation definitions are:

Definition 1 (Grey-scale dilation)**.** *Dilation of a grey-scale image is the result of the process of placing the origin of the reflected structuring element at each pixel of the image and assigning the maximum value among the pixels covered by its support each time.*

$$(f \oplus g)(x, y) = \max_{(i,j)} \{f(x - i, y - j) + g(i, j)\} \tag{3.9}$$

Note that if the structuring element is symmetric with respect to its origin, the word reflected is meaningless.

The dilation is both commutative and associative:

- Commutative:

$$A \oplus B = B \oplus A \tag{3.10}$$

- Associative:

$$A \oplus (B \oplus C) = (A \oplus B) \oplus C \tag{3.11}$$

A structuring element of size n can be obtained by successively dilating a structuring element. Let B be a finite connected subset of the discrete plane $\mathbb{Z}^2$, whose size is one by convention, then the finite set

$$nB = \underbrace{B \oplus B \oplus \ldots \oplus B}_{n \text{ times}} \tag{3.12}$$

define a structuring element of size n with the same shape as B (see Fig. 3.10). If $n = 0$, then $nB = (0, 0)$ by convention. Note also that $nB \oplus mB = (n+m)B$ for any set B and any non-negative integers m, n.

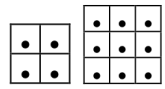


Figure 3.10: Square-shaped structuring elements with sizes 1 (left) and 2 (right).

Definition 2 (Grey-scale erosion)**.** *Erosion of a grey-scale image is defined as the process of placing the origin of the structuring element at each image pixel and assigning the minimum value of the pixels covered by its support each time.*

The erosion of f by g is the function:

$$(f \ominus g)(x, y) = \min_{(i,j)} \{f(x + i, y + j) - g(i, j)\} \tag{3.13}$$

The erosion obeys the following property:

$$(A \ominus B) \ominus C = A \ominus (B \oplus C) \tag{3.14}$$

Erosions and dilations are used together to compute the openings and closings of the images.

Definition 3 (Opening)**.** *Opening an image is the operation by which the image is dilated after having been eroded. Let I be an image, and nB be a structuring element of size n. The opening is denoted by:*

$$I \circ nB = (I \ominus nB) \oplus nB \tag{3.15}$$

Definition 4 (Closing)**.** *Closing an image is the operation by which an image is eroded after having been dilated. Using the same notation as in the definition of opening, the closing is denoted by:*

$$I \bullet nB = (I \oplus nB) \ominus nB \tag{3.16}$$

Both the opening and closing operations are idempotent:

$$(f \circ k) \circ k = f \circ k \tag{3.17}$$

$$(f \bullet k) \bullet k = f \bullet k \tag{3.18}$$

Let us show in Figure 3.12 an example of erosion, dilation, opening and closing of the image in Fig. 3.11 with the SE shown in Fig. 3.10(b).

15	17	22	17	15	16	18	21
15	46	50	55	46	25	20	15
15	58	65	60	43	25	20	18
17	39	47	61	38	23	18	19
15	19	24	34	25	19	15	14
10	17	19	28	35	15	17	20
13	18	21	33	38	39	45	42
13	17	16	34	31	37	16	16

Figure 3.11: 8×8 original image

46	50	50	55	55	46	25	21
58	65	65	65	60	46	25	21
58	65	65	65	61	46	25	20
58	65	65	65	61	43	25	20
39	47	61	61	61	38	23	20
19	24	34	38	39	45	45	45
18	21	34	38	39	45	45	45
18	21	34	38	39	45	45	45

(a)

15	15	17	15	15	15	15	15
15	15	17	15	15	15	15	15
15	15	39	38	23	18	15	15
15	15	19	24	19	15	14	14
10	10	17	19	15	15	14	14
10	10	17	19	15	15	14	14
10	10	16	16	15	15	15	16
13	13	16	16	31	16	16	16

(b)

15	17	17	17	15	15	15	15
15	39	39	39	38	23	18	15
15	39	39	38	38	23	18	15
15	39	39	39	38	23	18	15
15	19	24	24	24	19	15	15
10	17	19	19	19	15	16	16
13	17	19	31	31	31	16	16
13	16	16	31	31	31	16	16

(c)

46	46	50	55	46	25	21	21
46	46	50	55	46	25	20	20
58	58	65	60	43	25	20	20
39	39	47	61	38	23	20	20
19	19	24	34	38	23	20	20
18	18	21	34	38	23	20	20
18	18	21	34	38	39	45	45
18	18	21	34	38	39	45	45

(d)

Figure 3.12: Examples of dilation (b), erosion (c), opening (d) and closing (e) of the image in Fig. 3.11 with a 3×3 flat SE (a)

Pattern Spectrum

The *Pattern Spectrum* (PS) of an image is a morphological approach of measuring its size distribution and, thus, characterizing it. Different authors have given different definitions and ways to compute it. According to Maragos (Maragos, 1989), the PS is defined as:

$$PS_{f,G}(n) = A\left[(f \circ nG)(x,y) - (f \circ (n+1)G)(x,y)\right], \quad 0 \leq n \leq N \tag{3.19}$$

Where f is an image, G is a structuring element, and $A(f) = \sum_{(x,y)} f(x,y)$ and $(a-b)(x) = a(x) - b(x)$ denotes the point-wise algebraic difference between functions $a(x)$ and $b(x)$. N is the maximum size such that $f \ominus nG$ is not all empty.

Gonzalez and Woods defined granulometry as "the difference between the original image and its opening", as long as the particles in the image are lighter than the background. Otherwise the operation would be closing (Gonzalez and Woods, 2002). Following the same notation as in equation (3.19) this definition would be:

$$PS_{f,G}(n) = A\left[f(x,y) - (f \circ nG)(x,y)\right] \tag{3.20}$$

On the other hand, Petrou and García Sevilla exposed the so-called blanket method (Petrou and Sevilla, 2006) to compute the pattern spectrum, which consists in dilating and eroding the original image with a flat structuring element of size n and taking the point-wise algebraic difference of the two results.

$$PS_{f,G}(n) = A\left[(f \oplus nG)(x,y) - (f \ominus nG)(x,y)\right] \tag{3.21}$$

No matter which method is used, these results should be normalized:

$$PSNorm_{f,G}(n) = A\left(PS_{f,G}(n)\right) / A\left(f\right) \tag{3.22}$$

3.3.4 Geodesic distance

The geodesic distance between two pixels p_1 and p_n in a domain M is defined as the length of the shortest path which connects both pixels. Let $P = p_1, p_2, \ldots, p_n$ be the path between them, where p_i and p_{i+1} are 8-connected

neighbours $\forall i \in \{1, 2, \ldots, n-1\}$. The length of the path is defined as the sum of neighbour distances between successive points in the path, shown in equation (3.23).

$$l(P) = \sum_{i=1}^{n-1} d_N(p_i, p_{i+1}) \tag{3.23}$$

The term d_N can be any distance metric (*e.g.* city-block, euclidean, Chamfer, etc.). Note that an image can be considered as a Riemannian manifold (Jost, 2008), as Peyré and Cohen pointed out in (Peyré and Cohen, 2009). In that case, these metrics are not suitable, since they are not able of capturing the non-linear geometric structure of the data (Hamza and Krim, 2006).

A Riemannian manifold can be defined as an abstract parametric space $\mathcal{M} \subset \mathbb{R}^s$ (in practice, $s = 2$ for surfaces and $s = 3$ for volumes), equipped with a metric $x \in \mathcal{M} \mapsto H(x) \in \mathbb{R}^{s \times s}$ positive definite. Using this metric, the length of a curve γ can be computed by means of equation (3.24).

$$L(\gamma) = \int_0^1 \sqrt{\gamma'(t)^T H(\gamma(t))\gamma'(t)}\mathrm{d}t \tag{3.24}$$

The geodesic distance between two points x and y in a Riemannian space $(\mathcal{M}, H)$ is then defined as shown in equation (3.25).

$$d_{\mathcal{M}}(x, y) = \min_{\gamma \in \mathcal{P}(x,y)} L(\gamma) \tag{3.25}$$

where $\mathcal{P}(x, y)$ is the set if curves joining x and y, $\mathcal{P}(x, y) = \{\gamma \setminus \gamma(0) = x \ \ and \ \ \gamma(1) = y\}$.

The distance map in a Riemannian manifold to a set of starting points $\mathcal{S} = (x_k)_k \subset \mathcal{M}$ is defined in equation(3.26).

$$U_{\mathcal{S}}(x) = \min_k d_{\mathcal{M}}(x, x_k) \quad \forall x \in \mathcal{M} \tag{3.26}$$

If the metric H is continuous, then for any $\mathcal{S} \subset \mathcal{M}$, the distance map $U_{\mathcal{S}}$ is the unique solution of the Hamilton-Jacobi equation, shown in (3.27), also known as Eikonal Equation.

$$\|\nabla_x U_{\mathcal{S}}\|_{H(x)^{-1}} = 1, \quad where \ \|v\|_A = \sqrt{v^T A v} \tag{3.27}$$

Now, imagine the two-dimensional case in which an interface Γ, which is a closed curve in $\mathbb{R}^2$, propagates in its normal direction with a speed given by the function $F(x, y)$. Let $\phi(x, t)$ be a function that gives the distance d from the point x to Γ in the instant t (d may be negative if x is within Γ). Then an evolution equation for the interface may be produced, shown in (3.28).

$$\phi_t + F(x, y)\|\nabla\phi\| = 0 \tag{3.28}$$

Tracking the evolution of the surface can be very complex, for instance, if the points of the evolving interface try to cross over themselves, or even if the shape tries to break into two. The task becomes simpler if a stationary approach to the problem is considered instead. It consists in measuring the instants when the interface crosses the points (x, y) of the surface, obtaining a timing function $T(x, y)$. As an example, suppose that Γ (the initial curve) is a circle which propagates outwards, crossing over each of the timing spots. As can be seen in Fig. 3.13, a cone-shaped surface is built as Γ propagates, therefore, at any height T the *level* gives the set of points reached at time T.

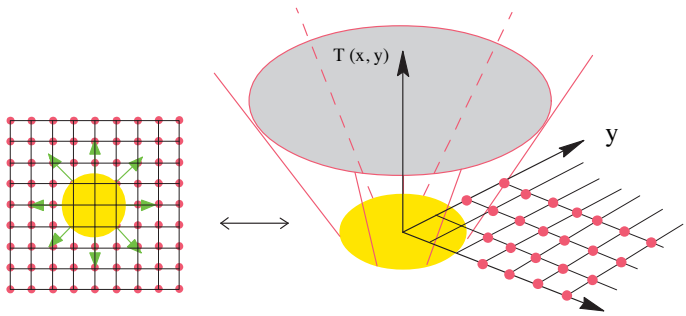


Figure 3.13: Example of the upwind scheme for front propagation.

Additionally, the gradient of this arrival time surface is inversely proportional to the speed of the propagation, as it is shown in equation (3.29).

$$\|\nabla T\|F = 1 \tag{3.29}$$

Equation (3.29) is a form of the Eikonal equation. This gradient can be approximated by equation (3.30). (Sethian, 1996).

$$\left[\begin{array}{l}\max\left(\max\left(D_{ij}^{-x}T,0\right),-\min\left(D_{ij}^{+x}T,0\right)\right)^2+\\+\max\left(\max\left(D_{ij}^{-y}T,0\right),-\min\left(D_{ij}^{+y}T,0\right)\right)^2\end{array}\right]=1/F_{ij}^2 \tag{3.30}$$

where finite difference notation have been used that, for instance $D_{ij}^{+x}T = \frac{T_{i+1,j}-T_{i,j}}{\triangle x}$ and $D_{ij}^{-x}T = \frac{T_{i,j}-T_{i-1,j}}{\triangle x}$.

Fast marching algorithm rests on "solving" equation (3.30) by building the solution outwards from the smallest T value. The idea is to sweep the front ahead in an upwind fashion by considering a set of points in a narrow band around the existing front, and march it forward, freezing the existing points, and bringing new ones into the narrow band structure. This approach is partially based on a method introduced by Chopp in (Chopp, 1993). Thus, the Fast marching method work as follows. First of all, the points must be initialized and tagged as belonging to one of these groups (see Fig. 3.14):

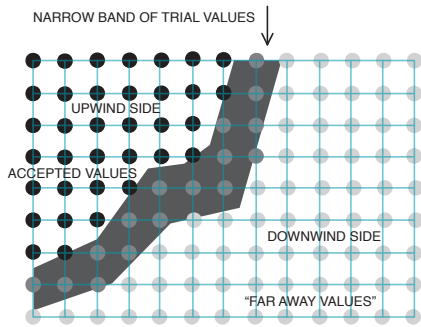


Figure 3.14: Upwind construction of accepted values in the narrow band approach.

- The *alive* points (set A) are the set of all grid points where the propagation has been made. Before the marching starts, they are initialized with the value $T_{i,j} = 0$.
- The *narrow band* points are those which are in the narrow band. They are initialized with the value $T_{i,j} = dy/F_{i,j}$. This set is called Nb.
- The *far away* points are all the other ones in the grid. We call this set F. They have a value $T_{i,j} = \infty$.

Afterwards, an iterative procedure is executed:

1. Let *trial* be the point in Nb with the smallest T value.
2. Add the point *trial* to A and remove it from Nb
3. Tag as *close* the 4-connected neighbours that are either in Nb or in F. In this case, remove it from F and add it to Nb.
4. Recompute the values of T at all neighbours, according to equation (3.30) and select the largest possible solution to the quadratic equation.
5. Return to the first step.

This algorithm works because the process of recomputing the T values at upwind neighbouring points cannot yield a value smaller than any of the *Alive* points. Therefore, we can march the solution outward, always selecting the narrow band grid point with minimum T value, and readjusting neighbours without having to go back and correct an accepted value.

Assuming that it takes no work to determine the member of the narrow band with the smallest value of T, the work required to compute the solution at all grid points – *i.e.* solving the Eikonal equation on a rectangular orthogonal mesh – is $O(N \log N)$, where N is the total number of grid points (Kimmel and Sethian, 1998).

3.4 Statistical Tests

There is not a established procedure to compare two or more algorithms over multiple problems, mainly because their behaviour is not deterministic, so the difference between their results might be due to random factors and not to a real improvement.

In order to determine whether the differences between two algorithms are significant or not, researchers may apply some statistical techniques, which are usually divided in two groups:

The *parametric tests* are the most common statistical tests. They are used on each algorithm, using the mean error and standard deviation over multiple executions to identify if the difference between two algorithms is statistically significant.

They assume some requirements to ensure that all components are compatible with each other:

1. The observations must be independent.
2. The sample data must have a normal distribution
3. The scores in the different groups must have homogeneous variances – homoscedasticity –.

On the other hand, the *non-parametric tests* use an ordinal way of ranking the results of the algorithms under study for each one of the problems. The data do not have to follow any particular distribution when using non-parametric tests.

It is very likely that the benchmarking datasets do not fulfil the required conditions to use parametric tests (Demšar, 2006) – *i.e.* there is no guarantee that the distributions are normal –. Therefore, we will use non-parametric tests in order to determine if there are statistically significant differences between the proposed algorithms. To carry out this task we will use the results of the algorithms assessed on different datasets, or the results on different iterations, when just one dataset has been used, as it is done in some works (Moreno-Torres et al., 2010).

The Wilcoxon signed-rank test (Wilcoxon, 1945) is the non-parametric alternative to the paired t-test. Therefore, it is a pairwise test that aims to detect significant differences between two algorithms. Let a_i be the difference between the performance scores of the two algorithms on the i^{th} of N datasets. These differences are ranked according to their absolute values (average ranks are assigned in case of ties). Let R^+ be the sum of the ranks where the difference is positive – *i.e.* the second algorithm outperforms the first one –, and R^- the sum of ranks for the opposite. Ranks of $a_i = 0$ are split evenly among the sums; one of them would be ignored if there was and odd number of them. It can be seen in (3.31) and (3.32).

$$R^+ = \sum_{a_i>0} rank(a_i) + \frac{1}{2}\sum_{a_i=0} rank(a_i) \tag{3.31}$$

$$R^- = \sum_{a_i<0} rank(a_i) + \frac{1}{2}\sum_{a_i=0} rank(a_i) \tag{3.32}$$

Let T be the smallest of the sums, $T = \min(R^+, R^-)$. If it is less than, or equal to the value of the Wilcoxon distribution with N degrees of freedom – whose table may be included on most books on general statistics, such as (Zar, 2007) –, the null hypothesis of equality of the algorithms is rejected. When the number of datasets, N, is large (such as $N > 25$), the statistics shown in (3.33) is distributed approximately normally. The null hypothesis can be rejected if z is smaller than -1.96 with $\alpha = 0.05$ (Demšar, 2006).

$$z = \frac{T - \frac{1}{4}N(N+1)}{\sqrt{\frac{1}{24}N(N+1)(2N+1)}} \tag{3.33}$$

CHAPTER 4

INTELLIGENT BOAR SPERM SEGMENTATION USING THRESHOLDING AND WATERSHED

An approach to improve the segmentation process and to reliably detect whether a head is well segmented or not is presented in this chapter. The goal is to minimise the number of bad images that will be further processed, either improving the success of the segmentation method, or automatically detecting and discarding bad segmentations. This approach consists in segmenting the images with a basic method based on thresholding in a first step. Then, the heads which are not well segmented are automatically detected, and a method which uses the Watershed transform (Meyer, 1994) is applied on them. Finally, the images which are still not well segmented are automatically discarded, and the others are masked (see Fig. 4.1).

(a) (b) (c)

Figure 4.1: Original head and its corresponding segmented and masked image.

The rest of the chapter is organised as follows: in section 4.1 the image set is briefly described. Segmentation methods are detailed in section 4.2, and the way they are combined to improve accuracy is shown in section 4.3. Afterwards the experiments and results are shown in section 4.4, and finally we discuss the conclusions in section 4.5.

4.1 Image Set

A set of alive and dead images has been used in the assessment of the proposed segmentation approach. The sample preparation and the image acquisition has been carried out as it was explained in sections 3.1.2 and 3.2, respectively, but this time a camera Oscar F-810 has been used[1]. All snapshots have been taken with a resolution of 3272×2469 pixels.

The set of images that will be used in these experiments have 422 and 341 images of alive and dead heads, respectively.

[1]The camera Basler scA780-54fc was acquired after developing this experiment

4.2 Segmentation Methods

Our proposal has been to combine two different segmentation techniques. One of them consisting of some morphological operations and Otsu's thresholding (Otsu, 1979), while the other one performs the segmentation by means of the Watershed transform (Meyer, 1994). The steps of the process have been:

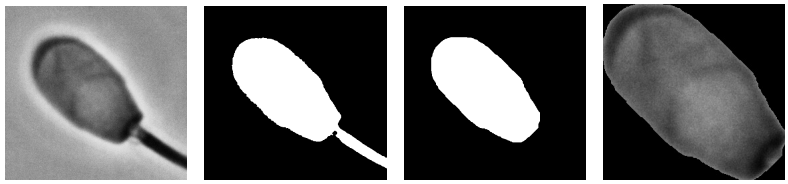
1. Segment all images by means of the thresholding-based approach.
2. Automatically detect the heads that are not well segmented, using the criteria that will be shown below (see section 4.3.1).
3. Segment those images by means of the Watershed-based method.
4. Detect and discard the heads which are still bad segmented.

4.2.1 Thresholding-based segmentation

The first segmentation approach is carried out by preprocessing the image with some morphological operations and then thresholding it. First of all, the image is converted to grey-scale and its contrast is increased by saturating a 1% of the pixel values at low and high intensities of the grey level image (Fig. 4.2(a)). Then, the image is binarised by means of the Otsu's thresholding method (Otsu, 1979). After that, we dilate that binary image using a disk-shaped structuring element, and then all objects except the biggest one are removed. Finally the complement of this image is obtained, so the spermatozoon should be segmented (Fig. 4.2(b)).

Afterwards, an opening operation using a disk-shaped SE is performed in order to automatically remove the tail, or at least separate it from the head. To make sure it is removed, all regions which are in contact with the border of the image are cleaned, so we obtain just the region with the head of the spermatozoon (Fig. 4.2(c)).

Finally the head is masked by making the pixel-wise multiplication of the binary and grey-level images, obtaining the original texture of the ROI over a black background. Then, the image is cropped in order to get the head contained into its bounding box (see Fig. 4.2(d)).

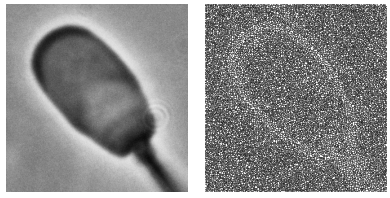


(a) (b) (c) (d)

Figure 4.2: Thresholding segmentation process.

4.2.2 Segmentation using Watershed transform

The Watershed transform (Meyer, 1994) considers the magnitude of the gradient of a grey-scale image (Fig. 4.4(b)) as a topographic surface, where pixels with the highest values are its top points. This surface is flooded from the regional minima at a constant speed and, when two regions are going to be merged by the water, a *wall* – which is called the Watershed line – is *built*. Unfortunately the result is usually an over-segmentation (Fig. 4.3), due to the high sensitivity of the gradient to noise.



(a)

Figure 4.3: Example of over-segmentation produced by the Watershed transform.

In order to avoid this unwanted effect, Meyer and Beucher proposed the use of markers to indicate the minima of the topographic surface where the flooding should begin from (Meyer and Beucher, 1990). These markers correspond to the foreground and the background of that image.

The foreground corresponds to the region of interest – in this case, the spermatozoon's head –. Therefore, we use the result of the thresholding-based segmentation approach (Fig. 4.4(d)). In case it yielded a completely black

image, a small square placed at the center of the image is used as foreground marker.

Four lines, one at each corner of the image, are used as background markers. Each one is obtained by drawing a white little square on a copy of the image with the foreground marker, at the corresponding corner. Then the Watershed transform is applied over its binary Euclidean distance transform (Breu et al., 1995). Repeating this process once for each corner, the four background markers (Fig. 4.4(c)) are obtained.

Afterwards, the pixels indicated by all markers are set to zero in the gradient image, so that they become its regional minima (Fig. 4.4(e)), and then the Watershed transform is computed. Once the head is segmented (Fig. 4.4(f)), it is opened with a disk-shaped structuring element to smooth it. Finally the original head is masked.

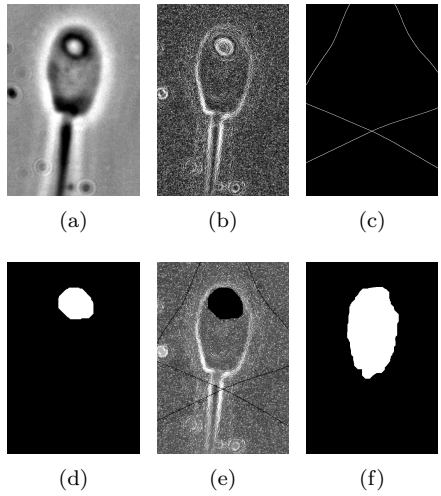


Figure 4.4: Watershed segmentation process.

4.3 Hybrid Segmentation

4.3.1 Detection of bad segmented heads

The main contribution in this approach is the decision of whether an image is well segmented or not. Each head is checked, and it is discarded if it does not

fulfil the following criteria:

1. Surface factor: The number of pixels of the head is greater or equal than the 70% of the average area of the heads of the whole set. This value has been obtained empirically, and it should not change even though the optics of the microscope did – *i.e.* the magnification changed –.

2. Eccentricity factor: The value of the ratio between the major and the minor axes of the ellipse that has the same normalized second central moments as the head must be between 1.4 and 2.6. These values have also been obtained experimentally.

4.3.2 Proposed approach

The approach proposed in this Thesis consists in combining the thresholding-based segmentation (see section 4.2.1) and the method based on the Watershed transform (section 4.2.2).

First of all, the images are segmented using the first method. Afterwards the images which are not well segmented are automatically detected, following the criteria of section 4.3.1 and the same method is applied over them using a 20% higher threshold. Then, the images which are still bad segmented are automatically detected and the Watershed-based approach is applied on them.

Finally, a new automatic inspection is carried out in order to remove the bad segmented heads from the final set.

The method based on the Watershed transform is carried out only over the images which are discarded from the thresholding process because its computational cost is higher than the former.

4.4 Experiments and Results

The performance of the hybrid approach has been assessed using a set of images of alive and dead heads of boar spermatozoa (see section 4.1), which has also been segmented with both the thresholding-based method and the Watershed transform so that these three approaches can be compared. Afterwards, a visual inspection has been made to identify possible false positives or negatives in the automatic detection of bad segmentations. When carrying out this manual inspection we have compared the segmented and masked region with the head

in the original image. When the region matches the head (which is the area "inside" the white halo in the original image), then it is considered as a well segmented region. It is important to point out that we have not made any kind of previous selection of the images.

The detected and the actual number of well and bad segmented images for both the alive and dead spermatozoa are shown using confusion matrices.

Some measures have been computed from them to assess the performance of each method:

- **The segmentation accuracy (A_s)**, which is the rate of images that are well segmented.
- **The detection accuracy (A_d)** refers to the quality of the detection method and corresponds to the rate of correctly detected images.

4.4.1 Results of the thresholding-based segmentation

Using the thresholding-based method we have obtained an overall A_s of 88.99%, although differences between classes are very remarkable: It is 97.16% in the case of the alive heads, while in the dead ones, it is 78.89%.

Regarding the accuracy of the detection (A_d), it has also been higher for the alive heads (97.87%) than for the dead ones (87.98%). The confusion matrices that hold the results for the alive and dead spermatozoa are shown in Tables 4.1 and 4.2, respectively.

Table 4.1: Results of the segmentation based on thresholding of the alive spermatozoa.

		Detection	
		Good	Bad
Real Segm.	Good	410	0
	Bad	9	3

Some examples of bad segmented heads with the thresholding-based approach are shown in Fig. 4.5.

Table 4.2: Results of the segmentation based on thresholding of the dead spermatozoa.

		Detection	
		Good	Bad
Real Segm.	Good	269	0
	Bad	41	31

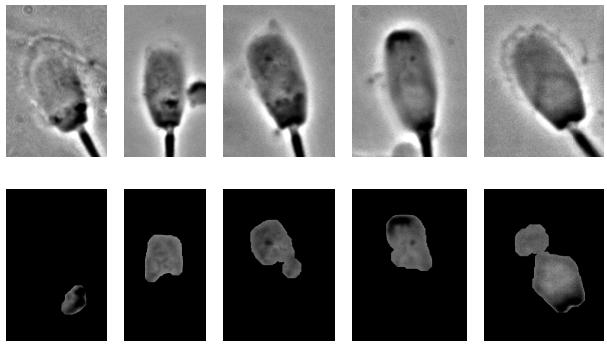


Figure 4.5: Examples of bad segmentations produced by the thresholding-based segmentation approach.

4.4.2 Results of the Watershed-based segmentation

The segmentation based on the Watershed transform is better than the previous one, getting an overall accuracy (A_s) of 90.30%. The improvement is greater in the case of the dead class (82.11%), while in the case of alive ones the A_s is almost the same (96.92%). This method performs better when segmenting dead sperm heads, because it is more robust to low contrast between the region of interest and the background.

It is remarkable that despite the accuracy of the segmentation is higher, the A_d of bad segmented heads is a little bit worse than in the previous approach, as a 96.92% of the alive heads and a 87.39% of the dead ones have been well detected. The confusion matrices these results have been computed from are shown in Tables 4.3 and 4.4, for the alive and dead spermatozoa, respectively.

Table 4.3: Results of the Watershed-based segmentation of the alive spermatozoa.

		Detection	
		Good	Bad
Real Segm.	Good	409	0
	Bad	13	0

Table 4.4: Results of the Watershed-based segmentation of the dead spermatozoa.

		Detection	
		Good	Bad
Real Segm.	Good	280	0
	Bad	43	18

It is worth to highlight the computational cost of this approach. The execution time for segmenting this image set has been 1595.5 seconds.

Fig. 4.6 depicts some bad segmented heads produced by this Watershed-based method.

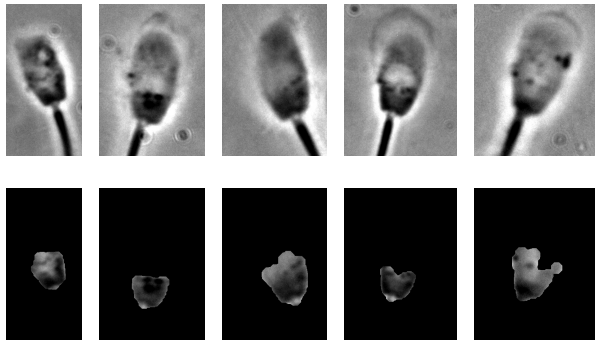


Figure 4.6: Examples of bad segmentations produced by the Watershed-based approach.

4.4.3 Results of the hybrid segmentation

The proposed approach is more efficient than the other two methods, yielding an overall segmentation accuracy of 90.96%: 97.39% of the alive heads and a 82.99% in the case of the dead ones.

As usual, the rate of images that are well detected is a little bit lower in the dead class (86.51%), in spite of the improvement in the A_s. All these results are shown in Tables 4.5 and 4.6.

Table 4.5: Results of the hybrid segmentation of the alive spermatozoa.

		Detection	
		Good	Bad
Real Segm.	Good	411	0
	Bad	11	0

Table 4.6: Results of the hybrid segmentation of the dead spermatozoa.

		Detection	
		Good	Bad
Real Segm.	Good	283	0
	Bad	46	12

The improvement in terms of computational cost of this mixed method is

noticeable, as the execution time has been 324.25 seconds: almost 5 times faster than the Watershed-based approach.

Once again, we show some examples of bad segmented heads obtained by this mixed process in Fig. 4.7.

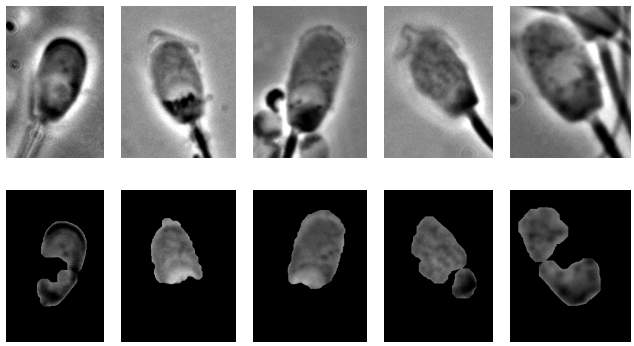


Figure 4.7: Examples of bad segmentations produced by the Watershed-based approach.

4.5 Discussion

We have presented a segmentation process that combines a method based on thresholding with other one based on the Watershed transform, which is applied to the images which the first one did not segment well. The bad segmented images are automatically detected by means of a method that we have proposed and assessed.

This proposal allows to get a larger image set for a further analysis, due to its higher efficiency with heads of both alive and dead spermatozoa (90.96% against 88.99% and 90.30% using thresholding and Watershed, respectively). It is true that the difference between this approach and the Watershed-based one is not very high, but the difference in computational cost (the former is almost five times faster that the latter) makes this new approach very appealing. These approaches show similar ratios of precision in the detection – around 97% and 86% for the alive and dead heads, respectively –. This imbalance is because the dead spermatozoa have usually lower contrast with the image background than the alive ones, which make the segmentation more difficult.

It is remarkable that neither the proposed method nor the other two produce false negatives, which means that the discrimination criteria do not discard any properly segmented images.

CHAPTER 5

CURVELET TEXTURE DESCRIPTORS FOR ACROSOME INTEGRITY ASSESSMENT

As it has been already pointed out in section 2.1.2, multi-resolution texture analysis is a powerful tool in tasks involving texture description (Ahmad et al., 2007; Grigorescu et al., 2002). Specifically, the Discrete Wavelet Transform (DWT) has been widely used in the literature in tissues and cells analysis (Bonnel et al., 2009; Tsantis et al., 2009), or even in sperm integrity assessment tasks (González et al., 2007), with quite successful results. Texture analysis using the Discrete Curvelet Transform (DCT) has achieved good performance in the literature, both used as a descriptor by itself (Eltoukhy et al., 2010a), and combined with first or second order statistics (Arivazhagan et al., 2006; Semler and Dettori, 2006). In addition, comparisons carried out between both transforms show higher performance of the latter in texture characterization, *e.g.* with CT images of tissues (Dettori and Semler, 2007), or in mammogram images (Eltoukhy et al., 2010b).

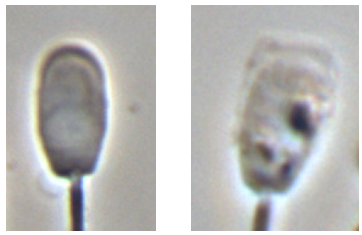
Regarding the success of the DCT, in this chapter we will present our proposal, that consists in applying it to the texture of sperm heads in order to recognise the integrity of their acrosomes and then to calculate first and second order statistics from the coefficients produced by the Curvelet transform. Performance in the classification of these two Curvelet-based descriptors will be compared with other texture features extracted from the DWT and with some shape descriptors based on moments.

The rest of the chapter is organised as follows: The image set used in this experiment is described in Section 5.1. Then, the descriptors that characterise the images are detailed in section 5.2, and the results of their classification are shown and discussed in section 5.3. Finally, conclusions are summarized in section 5.4.

5.1 Image dataset

The boar intact-damaged sperm image set that have been used in this experiment has been acquired as explained in section 3.2, from semen samples which had been prepared as it was stated out in section 3.1.1. Next, each head is automatically cropped (see Fig. 5.1) and labelled at the same time, thanks to the ground truth given by the fluorescence images.

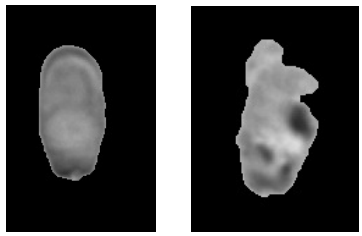
Afterwards, they were segmented by means of the approach presented in chapter 4. The heads that could not be properly segmented were automatically



(a) (b)

Figure 5.1: Examples of cropped heads.

recognised and removed from the set. Finally, 1849 images have been used in this experiment: 945 heads whose acrosome is damaged (either reacted or reacting) and 904 with an intact membrane. Fig. 5.2 shows an intact and a damaged acrosome after the segmentation.



(a) (b)

Figure 5.2: Images of intact (left) and damaged (right) acrosomes after the segmentation.

5.2 Characterization of the acrosome integrity

The goal of this experiment is to characterise and classify boar spermatozoa in terms of their acrosome integrity using texture analysis. By looking at the aspect of the segmented heads (Fig. 5.2), it seems that a morphological approach could be successful to describe the images. In order to check whether or not this statement is right, Hu, Flusser, Legendre and Zernike moments have been assessed for this task. According to texture descriptors, first and second order statistics, computed from both the original image and the Wavelet and

Curvelet transforms, have been extracted. Table 5.1 shows a summary of the descriptors that have been assessed, along with the number of features that each one has.

Table 5.1: Descriptors used in the intact and damaged acrosomes classification experiment

Descriptors	**Num. features**
WSF	24
WCF	20
CSF	52
CCF	108
Hu	7
Flusser	6
Legendre	9
Zernike	9

Classification results obtained by shape descriptors will be compared with those achieved by texture descriptors, as well.

5.2.1 Texture descriptors extracted from the Wavelet Transform

Information represented by spatial frequencies is often used for texture pattern recognition with satisfactory results, because of its frequency domain localization capability. Therefore, Discrete Wavelet Transform (DWT) has been applied on the images with the goal of characterising their textures. Specifically, we have used the Haar family of Wavelets which, in spite of being the simplest, outperforms the Coiflet and Daubechies, as Dettori and Semler pointed out in (Dettori and Semler, 2007). The DWT extracts the high-frequency components of a signal, so that they can be analysed separately. When the transform is applied on an image, four matrices of coefficients are obtained: approximations and horizontal, vertical and diagonal details. The first one holds almost all the energy of the image, while the other three hold the high frequency details.

We have computed two descriptors from the Wavelet coefficients of the image. The first one consists in calculating the mean and the standard deviation from the histograms of the Wavelet sub-bands obtained after three splits: LL1, LH1, HL1, HH1, LL2, LH2, HL2, HH2, LL3, LH3, HL3, HH3 (see Figs. 5.3 and

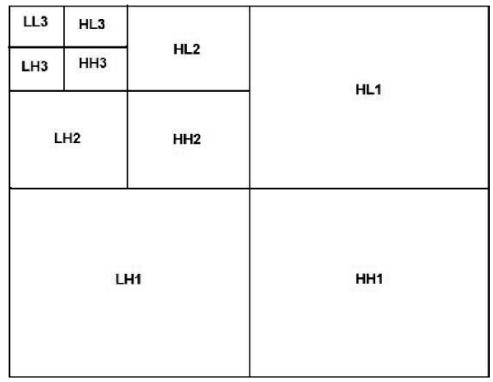


Figure 5.3: Names of the sub-bands of a 3-level Wavelet decomposition.

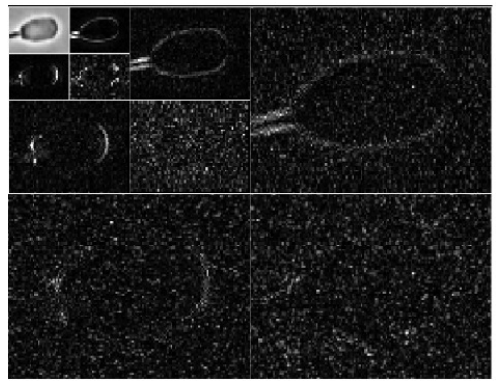
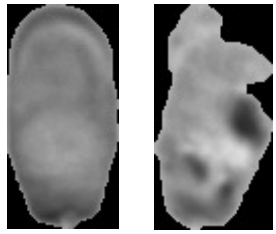


Figure 5.4: 3-level Wavelet decomposition of spermatozoon head image.

5.4). Therefore, each image is represented by a vector of 24 features, which is called *Wavelet Statistical Features* (WSF) (Arivazhagan and Ganesan, 2003).

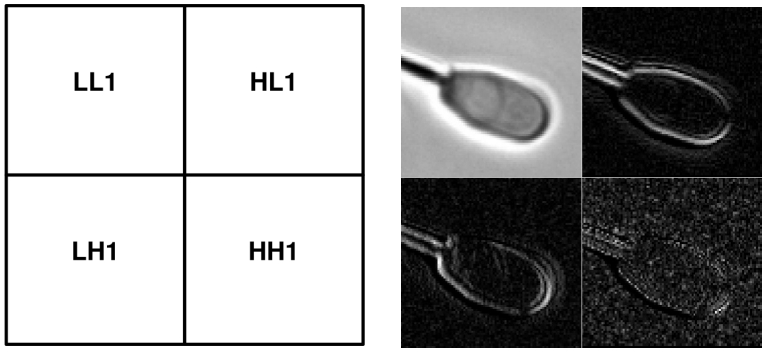
On the other hand, we have extracted the Haralick features *Contrast*, *Correlation*, *Energy* and *Homogeneity* from the GLCMs of the original image and from the coefficients of the first sub-band (LL1, HL1, LH1 and HH1 in Fig. 5.6(a)), as it was done in (Arivazhagan and Ganesan, 2003). Finally, the texture is characterised by a vector of 20 features, which is called *WCF* (Wavelet Co-occurrence Features). The co-occurrence matrices were computed with distances 1, 2, 3 and 5 and the best results were achieved when $d = 1$. All features have been averaged over the orientations 0°, 45°, 90° and 135° to make them somehow invariant to rotation.

Pixels from the background of the image do not hold any practical information, so each head is cropped into its bounding box (see Fig. 5.5) before describing it.



(a) (b)

Figure 5.5: Intact (left) and damaged (right) heads cropped into its bounding box.



(a) (b)

Figure 5.6: Names (left) and sub-bands (right) of a 1-level Wavelet decomposition of an image.

5.2.2 Texture descriptors extracted from the Curvelet Transform

The Discrete Curvelet Transform (Candès et al., 2006) has also been applied to describe the textures of the acrosomes, by means of the "wrapping" algorithm – which has already been used in (Dettori and Semler, 2007) with successful results –. As it has been stated out in section 3.3.2, a Fourier transform is

applied to the image yielding, for each scale and orientation, a product U_j, which is "wrapped" around the origin. Finally, the Curvelet coefficients are obtained by applying an inverse Fourier transform and they are represented in each scale and orientation by "wedges" (*i.e.* the shaded area in Fig. 3.9).

Two descriptors based on the DCT have been computed as well, called Curvelet Statistical Features (*CSF*) and Curvelet Co-occurrence Features (*CCF*). The first one consists in calculating the mean and standard deviation from the histogram of each wedge, analogously to WSF. The second one consists in computing the co-occurrence matrices from the original image and each one of the matrices of Curvelet coefficients and averaging the *Contrast*, *Correlation*, *Energy* and *Homogeneity* over the GLCM orientations 0°, 45°, 90° and 135°, as it was done to extract the WCF descriptor. Once again, the best results have been achieved when the distance of the co-occurence matrix was $d = 1$.

Both of them have been extracted from the coefficients in several combinations of scales (3 and 4) and angles in the second scale (8, 12 and 16), resulting that the best classification results were achieved when combining 4 scales and 8 angles, which yielded 26 "wedges" per image. Therefore, CSF and CCF have 52 and 108 features, respectively.

5.2.3 Shape descriptors

In order to assess whether the recognition of the acrosome integrity can be addressed by a shape-based approach, four different shape descriptors have been extracted from the heads (Fig. 5.7).

(a) (b)

Figure 5.7: Image of the shape of intact (left) and damaged (right) heads.

The region of each head has been described by means of the Flusser and Suk affine moment invariants (Flusser and Suk, 1994), Hu (Hu, 1962), Legendre

(Shu et al., 2000) and Zernike moments (Liao and Pawlak, 1997), as they have been broadly used in the literature (Chong et al., 2004; Lin and Chou, 2003; Ruggeri and Pajaro, 2002).

5.3 Experiments and Results

All images, characterised by means of their texture and their shape, have been classified with a Neural Network (NN) with a Multilayer Perceptron architecture. A list with all descriptors, along with the number of features that each one has, can be found in Table 5.1. The neural network has one hidden layer and a logistic sigmoid activation function (Equation 5.1) for the hidden and output layers. Learning was carried out with a momentum and adaptive learning rate algorithm. Data were normalized with zero mean and standard deviation equal to one.

$$o(n) = \frac{1}{1 + e^{-n}} \tag{5.1}$$

Several combinations of training cycles – 200, 300 and 400 – and neurons in the hidden layer – 2, 3 or 5 – have been assessed with all descriptors, in order to find the optimal network configuration. Classification has been carried out by stratified k-fold cross validation, which consists in dividing the data into k groups (which have the same distribution as the whole set), taking $k - 1$ to train the network and the other one for the test each time, repeating this process once per fold. Finally, the accuracy is obtained by averaging the results of the k folds. In order to avoid possible random effects, this procedure has been repeated 10 times, and results that we are presenting are an average of these 10 runs.

The cost of a wrong decision in this kind of problems is not equivalent. It means that considering a spermatozoon to have an intact acrosome when it is actually damaged has a higher cost than on the contrary. Therefore, accuracy used as a single metric is not suitable to illustrate the performance of the classifier, and ROC (Receiver Operating Characteristics) analysis is a more powerful tool, as some voices from the machine learning community have been claiming (Provost et al., 1998). Therefore, ROC curves of the descriptors and their AUC (Area Under the Curve) are shown in Fig. 5.9 and Table 5.3, respectively.

Classification accuracy for each descriptor along with the configuration of the NN they have been achieved with are shown in Table 5.2.

Table 5.2: Accuracy (in %) in the NN classification of intact and damaged acrosomes

Descriptor	Cycles	Neurons	Accuracy (%)		
			Overall	Intact	Damaged
WSF	400	5	87.26	89.25	85.37
WCF	400	3	96.43	96.30	96.56
CSF	200	2	96.42	96.48	96.36
CCF	200	5	97.00	97.29	96.73
Hu	400	5	82.76	85.39	80.24
Flusser	400	5	81.31	83.50	79.22
Legendre	400	5	71.74	76.93	66.76
Zernike	400	5	65.86	67.79	64.02

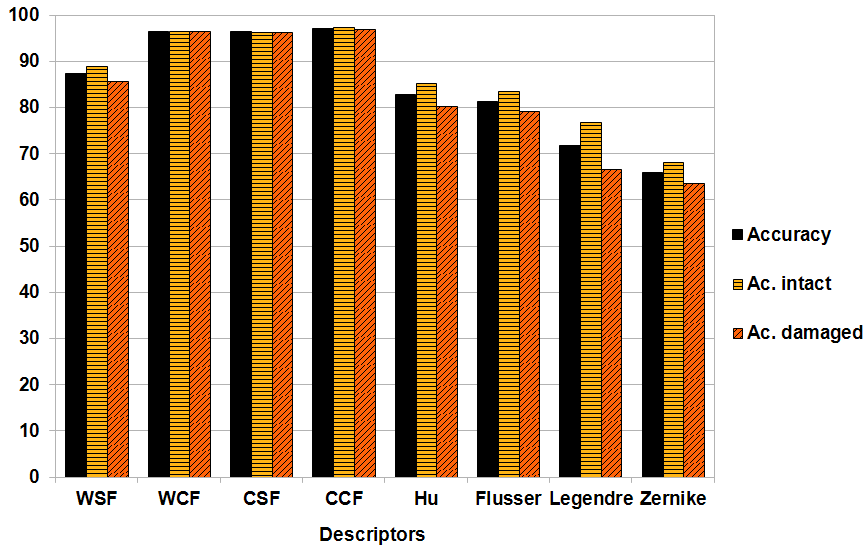


Figure 5.8: Graph bar with the accuracy in the classification of intact and damaged acrosomes.

First of all, regarding shape descriptors, it is remarkable that none of them show better performance than any of the textural features. Hu moments outperform the others (with a hit rate of 82.76%), but they are still worse than the

worst of the texture descriptors (WSF), whose accuracy is 87%. This proves that, contrary to what may be thought, shape descriptors are not suitable for this problem, while texture analysis is much more accurate.

The best results have been achieved by CCF with accuracy of 97%, closely followed by WCF (96.43%) and CSF (96.42%), respectively. It is very remarkable that they achieve very balanced hit rates, which is very interesting for the veterinary community, while the others do not (*i.e.* Legendre is the most imbalanced, with hit rates of 76.93% and 66.76% in the intact and damaged class, respectively). This can be more clearly seen in Fig. 5.8.

Fig. 5.9 depicts ROC curves, and the area under them is shown in Table 5.3. Both of them confirm previous results, since the AUC of CCF is the highest, followed by WCF and CSF. Once again, shape descriptors show the worst performance, as the best AUC is lower than 0.90.

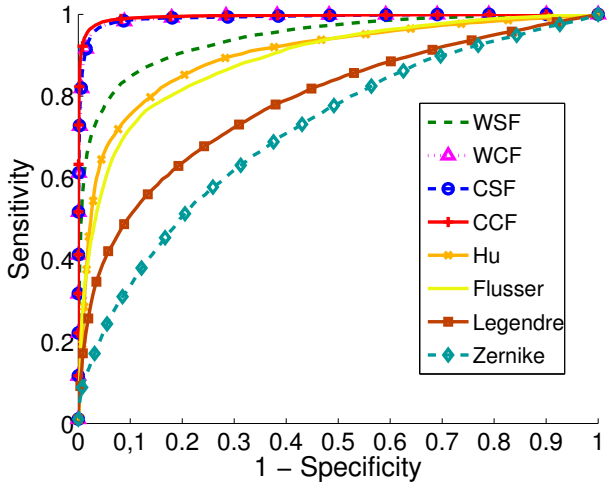


Figure 5.9: ROC curves of the NN when classification of damaged and intact acrosomes

Additionally, a Wilcoxon signed-rank test between CCF (the descriptor which obtained the lowest AUC) and the others has been performed. The observations that have been taken as scores in the test are the Area Under the ROC Curves of each test fold at each iteration (Moreno-Torres et al., 2010), so 100 scores per algorithm have been used. According to these tests, the differences between the AUCs of CCF and the others are statistically significant.

Table 5.3: AUC of the descriptors of the intact and damaged acrosomes

Descriptor	AUC
WSF	0.942
WCF	0.993
CSF	0.992
CCF	0.995
Hu	0.899
Flusser	0.888
Legendre	0.785
Zernike	0.716

5.4 Conclusion

In this chapter we have assessed the performance of multi-resolution texture analysis in the frequency domain for recognising boar spermatozoa acrosome integrity. In particular, we have compared the performance of Curvelet-based descriptors with other well known features based on the Wavelet transform. Therefore, both transforms have been applied to the image, and then first and second order statistics have been extracted from their coefficients to characterise the images. A glance at the segmented acrosomes would make the reader think that characterising them by means of their shape is a better approach. Therefore, some region descriptors based on moments have been calculated in order to check whether this assumption is right.

Classification has been carried out by means of a backpropagation Neural Network, using cross-validation. Results show on the one hand, that characterising the acrosomes by their shape is useless when trying to assess their integrity, as the best hit rate (around 83%), achieved by Hu moments, is far from the worst of the texture descriptors assessed in this work (WSF), which achieved an accuracy of 87.26%. This is also noticeable if their AUC are compared (0.899 against 0.942). On the other hand, results also show that the best performance is obtained when using Curvelet transform, yielding a hit rate of 97%. It is also remarkable that the best texture descriptors – CCF, WCF and CSF – have also produced quite balanced hit rates, which is very appealing for the veterinary community.

CHAPTER 6

ADAPTIVE GEODESIC PATTERN SPECTRUM (AGPS)

Texture description approaches have been widely used to extract mathematical features to represent a region of interest (ROI) with the goal of automatically recognising it. Conventional texture description techniques are often applied as-are on the whole ROI, obviating the fact that textures may not be homogeneous *i.e.* sometimes there are areas with different properties within, or even among different elements of the same class. When that happens, these approaches might be suboptimal, as they cannot capture all variations inside the images.

Mathematical Morphology (MM) is a theory which provides techniques to perform analysis of geometrical structures within an image. It comprises a set of processes which can be applied in order to, for instance, remove details smaller than a reference set called *Structuring Element* (SE). This theory was first developed for binary images (Matheron, 1975; Serra, 1982), but was later generalised to grey-scale images (Sternberg, 1986).

One of its applications is texture description. Maragos reviewed the basic concepts and operations of grey-scale MM and developed a shape-size descriptor called *Pattern Spectrum* (PS) (Maragos, 1989). It is defined as a function of the size distribution of objects – "granules" – inside a texture, which is computed by making successive erosions and dilations with structuring elements of different sizes. It is also called *granulometric distribution function*, and the process of computing it is called *granulometry* (Petrou and Sevilla, 2006).

However, it is not clear which is the best SE for a certain problem (de Ves et al., 2006). There have been some attempts for finding the best structuring element for certain problems (Asano et al., 2003; Huet and Mattioli, 1996), but all of them need a priori knowledge about the types of images to be described. Other works carry out mathematical morphology operations where the structuring element varies in size and shape according to some local features (Bouaynaya et al., 2006), or even on the basis of Euclidean distance (Cuisenaire, 2006)

Our proposal is to characterize textures by means of an adaptive pattern spectrum, which is calculated by using structuring elements which *best* fit the image at each pixel under a distance criterion. We are considering the texture as a topographic surface, whose height is its grey-level value, so the Euclidean distance is not suitable, as it is not able to capture the geometric structure of the surface. Nevertheless, the geodesic distance is more appealing, as it does

capture the geometry of the surface (Hamza and Krim, 2006). This approach does not have any a priori knowledge about textures and takes their possible variations into account, overcoming one of the main drawbacks of texture description methods.

Recognizing grey-scale images of alive and dead boar spermatozoa using digital image processing is a challenging task for which, up to our knowledge, there are not commercial tools neither other experimental approaches but the ones published by our research group. According to the veterinary experts, predicting whether a spermatozoon is alive or dead without using stains is extremely difficult and still unresolved.

In this chapter we present a new adaptive texture descriptor based on computing the Pattern Spectrum with a Structuring Element whose shape is different at each pixel on a basis of a Geodesic Distance criterion which provides better performance than the classical PS when describing regions whose texels have similar shapes, as it is able to better capture the intrinsic geometry of textures. This adaptive texture descriptor has been called Adaptive Geodesic Pattern Spectrum (AGPS). It has been assessed firstly for characterising and classifying images of different materials extracted from the VisTex database, with the goal of assessing its performance with different kinds of textures. Next, it has been used to characterise and classify phase contrast images (see section 3.2) of alive and dead sperm cells. Results have been compared with the performance obtained by the WCF descriptors on the same images.

The rest of the chapter is structured as follows: The image sets that have been used in the experiments are shown in section 6.1. Section 6.2 shows how the Pattern Spectrum has been extracted from images and our proposed adaptive alternative to it. Then the results on the Vistex database and on the sperm dataset are shown in sections 6.3.1 and 6.3.2, respectively. Finally, some concluding remarks are given in section 6.4.

6.1 Image Datasets

Different image sets have been used in this chapter. Firstly, reference textures of different materials found in the MIT Medialab Vision Texture (VisTex) image database (Vision and Texture) have been used. Several works where this database has been used can be found in the literature (Do and Vetterli, 2003;

Kim and Hong, 2009; Kokare et al., 2005). The goal of using this set has been to assess this approach with a more general image dataset which has different kinds of textures, what allowed us to find out when our adaptive method outperforms the classical Pattern Spectrum.

On the other hand, a set with dead and alive sperm images has been used, in order to assess the proposed adaptive texture descriptor in a real and, up to our knowledge, yet unexplored problem.

6.1.1 VisTex images

We have used 75 images (whose size is 512×512 pixels) from the MIT Vision Texture (VisTex) database which are taken from different materials and divided by categories. Each one may contain different types of textures, so we have separated them into sub-categories.

Each category has not enough images to consider results of the classifications reliable. Thus, we have divided each image into 25 sub-images of 102×102 pixels, which are not overlapped between themselves. However, this did not provide enough images yet, and making the subdivisions smaller would not be useful to assess the performance of the descriptor. Thus, the image was splitted up again into 16 new sub-images of the same size, by means of a grid with 4 rows and 4 columns. In order to make them as different to the others as possible, the "origin" of the grid was placed at pixel $(51, 51)$ (see Fig 6.1(b)). Therefore, 41 sub-images 102×102 pixels were obtained from each original image. An illustration of the division process is shown in Fig. 6.1.

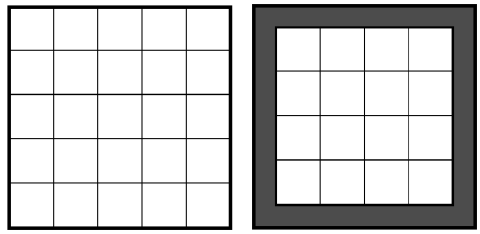


Figure 6.1: Division grids in the splitting process. The image is first divided into 25 (left) and then into 16 (right) sub-images.

Table 6.1 summarizes the VisTex categories and their corresponding sub-categories, how many images each one has, and the number of sub-images available after the division. The sub-classes that have been used in this chapter are shown in bold type.

Table 6.1: Summary of the categories and sub-categories of the VisTex database.

Category	Sub-category	Images	Sub-images
Bark	–	6	246
Fabric	**Basket (FaBa)**	4	164
	Fine (FaFi)	2	82
	Hair (FaHa)	3	123
	Rope (FaRo)	3	123
	Wicker (FaWi)	2	82
	Wool (FaWo)	2	82
Flowers	–	4	164
Food	**Beans (FoBe)**	4	164
	Sweets (FoSw)	4	164
Grass	–	2	82
Leaves	–	7	287
Metal	–	6	246
Misc	(MiCo)	2	82
	(MiGr)	2	82
Sand	–	5	205
Stone	–	2	82
Tile	Holes (TiHo)	4	164
	Tile (TiTi)	3	123
Water	–	6	246
Wood	Wood 1 (Wo00)	1	41
	Wood 2 (Wo01)	1	41

6.1.2 Alive and dead sperm images

The alive and dead heads have been cropped and segmented in the same way the intact and damaged acrosomes were (see section 5.1). Once that process was finished, the original grey level heads were cropped into their bounding boxes, extracted from the binary images. Later, they were registered to 63×108 pixels and (thanks to a previous detection of the tail) rotated in order to allow the major axis be in vertical position with the apical part placed in the upper side.

An example of the resulting alive and dead sperm heads images is shown in Fig. 6.2. Finally, the whole set has 470 alive and 375 dead spermatozoon heads.

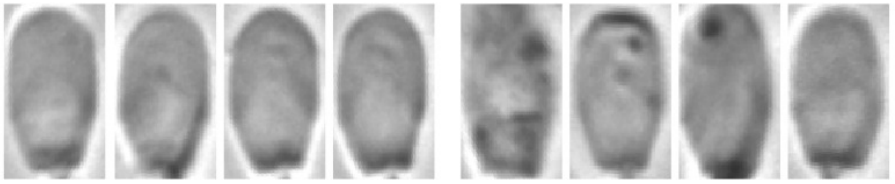


Figure 6.2: Images of alive (left) and dead (right) spermatozoa cropped and resized.

6.2 Description Methods

6.2.1 Pattern Spectrum

The basis of the Pattern Spectrum along with three possible approaches to compute it found in the literature were shown in section 3.3.3. These approaches were proposed by Maragos (Maragos, 1989), Gonzalez (Gonzalez and Woods, 2002) and Petrou and García Sevilla (Petrou and Sevilla, 2006), and they are defined in equations (3.19), (3.20), and (3.21), respectively.

We have computed the PS of several images using these definitions, measuring the average computation times in order to find out which one was the most computationally efficient. Specifically, the size of the pattern spectra has been set to 10, and they have been extracted from 200 images – with size 129×130 – of wood (see Fig. 6.3) using the flat, 3×3 square-shaped structuring element shown in Fig. 3.10(b) (with the origin in its central point). A comparison of the average computation times for each method is shown in Table 6.2.

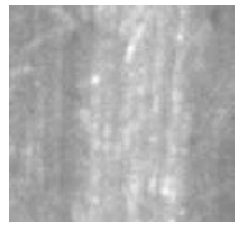


Figure 6.3: Example of wood image used in the measurement of the computation time of the PS.

Results show that the approach of Petrou and García Sevilla is clearly the most efficient, so this is the method that we have used in these experiments.

Table 6.2: Comparison of average execution times using the different PS extraction methods.

Approach	Time (s)
Petrou	0.77
Gonzalez	13.77
Maragos	18.28

6.2.2 Proposed approach: Adaptive Geodesic Pattern Spectrum

The structuring element (SE) used in the computation of the conventional pattern spectrum has the same shape on the whole image. However, using such SE does not allow to capture the geometrical variations of the texture.

Therefore, our proposal is to extract a SE whose shape and size change at each pixel where its origin is placed in a way that it fits the texture as well as possible. Considering the texture $f(x, y)$ to be a surface (see Fig. 6.4) whose pixels p_i are defined by the coordinates (x_i, y_i, z_i), where z_i is its grey level value $f(x_i, y_i)$ $\forall i$, then the support of a structuring element G_σ whose origin is at the point p_0 is made up of the elements which are within a distance of n units from p_0 less or equal than a threshold σ, as it is shown in equation (6.1).

$$Supp_{G_\sigma} = \{p_s = [x_s, y_s] \ ; \ d\,(p_0, p_s) \leq \sigma\} \tag{6.1}$$

When this condition is met, the SE that *best* fits the texture surrounding p_0 is obtained. The distance function d can be measured by means of any metric, *e.g.* Cityblock, Chessboard, Euclidean, *etc.* However, the non-linear structure of a surface cannot be captured by these distances, but Geodesic Distance can (Hamza and Krim, 2006).

Let us present a straightforward example to illustrate that. Fig. 6.5 shows a grey-scale image and its corresponding interpretation as a surface. Suppose we would like to find the shortest path that goes from $p = (100, 150)$ to $q = (160, 125)$, marked by a red and a blue dot, respectively, in Fig. 6.5(a). Taking

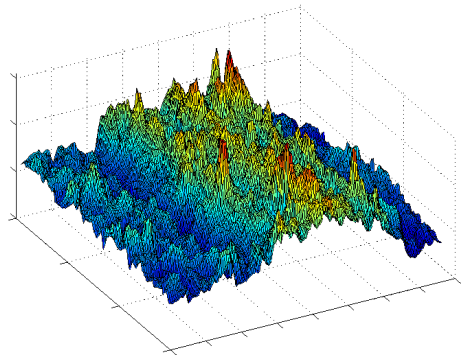
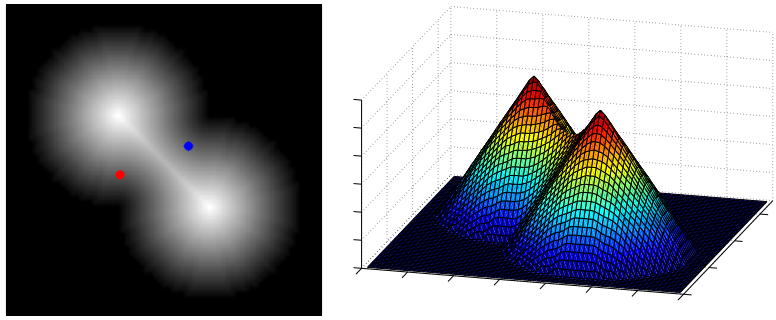


Figure 6.4: Texture in Fig. 6.3 seen as a surface.



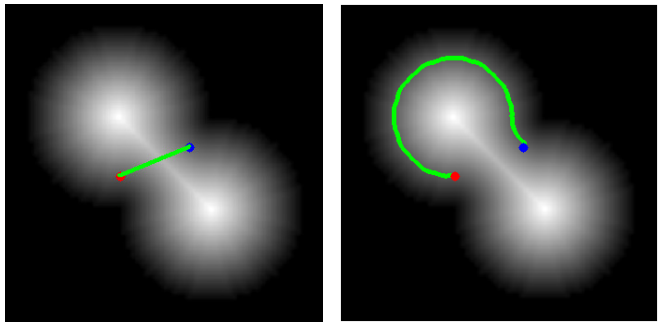
(a) (b)

Figure 6.5: Example of a grey-scale image (left) watched as a surface (right)

the minimum Euclidean distance path would mean to go through the surface, straight under the crest that joins the two peaks (see Fig. 6.6(a)).

The Geodesic path, however, does capture the intrinsic geometry of the surface, as it goes from p round the left "peak" until q is reached (see Fig 6.6(b)).

Therefore, our approach to extract the optimal structuring element at each point of the texture consists in computing a Geodesic distance map whose origin is at each pixel by means of the Fast Marching algorithm (Sethian, 1996). It makes possible to find the points which fulfil the condition of equation (6.1), which make up the support of each SE. Let $G_{i,j,n}$ be the structuring element whose origin is at point (i, j), and whose support is made of the points which are within a distance of n units in the map. The Adaptive Geodesic Pattern



(a) (b)

Figure 6.6: Euclidean (left) and Geodesic (right) paths that join p and q

Spectrum (AGPS) in the position n is then extracted by means of equation (3.21), but performing the erosions and dilations with the adaptive SE, $G_{i,j,n}$.

6.3 Experiments and Results

6.3.1 VisTex image description

In order to assess whether the proposed adaptive Pattern Spectrum is generalizable, *i.e.* it can be used as-is with any kind of texture, we have evaluated its performance with images of various types of materials. Specifically, some random pairs of VisTex sub-classes (see Table 6.1) have been considered as separated classification problems, and they have been characterized using the proposed AGPS and also the classical PS (computed using a 3×3 square-shaped SE), with comparison purposes. The sub-images used in this experiment are shown in Fig. 6.7.

Once again, central moments of orders 2 to 6 have been extracted from both functions so each texture has been characterized by 5 features. This time the best results have been achieved when the functions AGPS and PS have sizes 20 and 12, respectively.

Classifications have been carried out by means of Support Vector Machines (SVM) (Kim et al., 2002, 2003), and their results are shown in Fig. 6.8. These results are also shown in Table 6.3, where we have included the difference

Figure 6.7: Examples of 102 × 102 patches of VisTex sub-categories.

between the accuracies achieved by the Geodesic and the classical PS in the last column.

Regarding these results and making a visual inspection of images in Fig. 6.7, it can be noticed that the AGPS achieved better performance – with the greater difference over the classical PS – when texels were similar with each other. For instance, in problem FaRo *vs*. FoBe both textures had big and rounded texels. Regarding the textures FaWi and FoSw, both showed long and pseudo-rectangular shapes with slightly rounded ends, and in case of the FaRo *vs*. FaWi, both had big texels, in quite regular positions. Performance differences (in %) were 23.08, 7.84, and 7.04, respectively.

On the other hand, in the cases where the classical PS achieved better performance, texel shapes were quite different. For instance, in the problem FaBa *vs*. FaRo, the former had squared texels, while they are pretty much rounded in FaRo, and the performance difference (in %) was 15.53. Another example is the problem where the textures Bark and Metal were compared. The former showed thick and grouped grains, while the latter had fine grain, which is spread along the texture. In this case, the classical PS achieved a hit rate of 95.35%, against the 80% yielded by the AGPS descriptor.

Table 6.3: Classification accuracy of VisTex textures with adaptive and classic PS.

Classes	AGPS	PS	Difference
FaRo - FoBe	88.02	64.94	23.08
FaWi - FoSw	79.33	71.50	7.84
FaRo - FaWi	99.80	92.76	7.04
FaWi - Sand	98.64	93.34	5.30
Bark - FaWi	97.87	94.02	3.84
FaRo - Metal	92.04	88.78	3.25
Bark - FoSw	89.22	86.44	2.78
Bark - FaFi	92.26	90.79	1.46
FaBa - FaWi	95.77	94.63	1.15
FaFi - Sand	95.82	98.26	-2.43
FaWo - Sand	91.95	95.12	-3.17
Sand - Stone	92.91	96.10	-3.19
FaFi - Stone	95.72	99.34	-3.61
FoBe - Stone	92.27	98.37	-6.10
Metal - Stone	91.16	99.73	-8.57
FaBa - FoSw	77.99	90.64	-12.65
Bark - Metal	79.99	95.35	-15.35
FaBa - FaRo	84.12	99.65	-15.53

These results make sense under the point of view that the geodesic structuring element fits the region of the texture surrounding the pixel where its origin is placed, so that the pattern spectrum built using it discriminates similar textures better than the classical PS.

6.3.2 Sperm images description

The Geodesic Pattern Spectrum has been used to characterise images of alive and dead spermatozoa by means of their texture. We have assessed several sizes for the AGPS, but the best results have been achieved when the function has length 20. The PS and AGPS functions may be very large as-are, which is a drawback when they are classified due to the "curse of dimensionality". Therefore, five central moments (with orders 2 to 6) have been computed from them, so each texture is described by 5 features.

We have classified the data using a Neural Network with one hidden layer and a logistic-sigmoid activation function both in the hidden and in the output layer. The NN was trained using back-propagation, and learning was carried

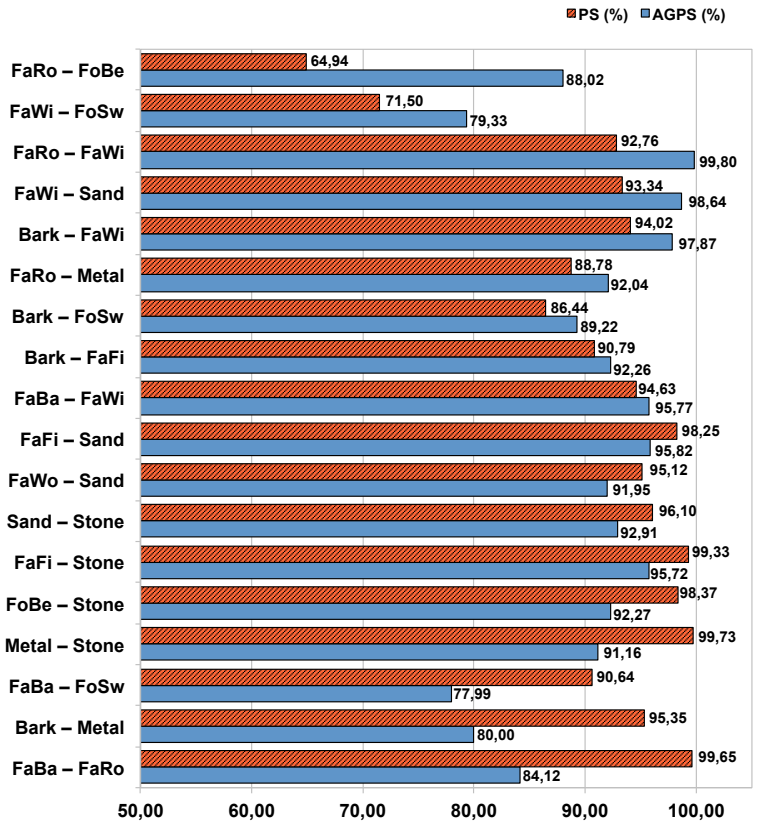


Figure 6.8: Results of the classification of VisTex textures.

out with a momentum and adaptive learning rate algorithm. Data were normalized with zero mean and standard deviation equal to one. Classification was carried out by means of 10-fold cross validation with several different combinations of neurons in the hidden layer and training cycles in order to find out the optimal configuration in terms of accuracy. We have also calculated the area under the ROC curves, since some authors claim that the hit rate is not the most suitable option to illustrate the performance of a classifier (Provost et al., 1998).

Performance in the classification of the AGPS has been compared with the classical Pattern Spectrum, computed with a 3×3 square-shaped structuring element whose origin is at the central point. In this case, the best hit rates

were achieved when the function has length 5. Tables 6.4 and 6.5 show the classification results with both the classical and the Geodesic Pattern Spectrum, respectively.

Table 6.4: Classification accuracy (in %) of alive and dead sperm described by the PS

Neurons	Cycles	AUC	Accuracy (%)		
			Overall	Intact	Damaged
2	200	0.646	62.64	83.89	36.01
2	300	0.661	63.90	80.06	43.63
2	400	0.665	64.17	81.57	42.35
3	200	0.659	63.59	78.83	44.48
3	300	0.671	64.24	80.28	44.13
3	400	0.668	64.30	81.17	43.17
5	200	0.666	64.54	79.28	46.06
5	**300**	**0.668**	**64.63**	**79.53**	**45.95**
5	400	0.670	64.33	80.00	44.68

Table 6.5: Classification accuracy (in %) of alive and dead sperm described by the AGPS

Neurons	Cycles	AUC	Accuracy (%)		
			Overall	Intact	Damaged
2	200	0.736	69.27	75.55	61.40
2	300	0.740	69.36	76.19	60.81
2	400	0.742	69.46	76.32	60.87
3	200	0.742	69.61	75.70	61.95
3	**300**	**0.744**	**69.75**	**76.21**	**61.64**
3	400	0.741	69.59	76.51	60.92
5	200	0.742	69.43	76.06	61.12
5	300	0.746	69.58	76.04	61.46
5	400	0.745	69.61	76.49	60.98

These results show that the classical pattern spectrum is outperformed by its adaptive variant, both in terms of hit rate (the AGPS achieves a hit rate of 69.58% against 64.63%, obtained by the conventional PS, both with their corresponding optimal NN configurations) and area under the ROC curve (0.746 against 0.668). Looking at the error rates of each class, it is also remarkable that they are much more imbalanced when using the PS than in the case of the AGPS (79% – 46% against 76% - 61%, respectively). We have carried out

a Wilcoxon signed-rank test using the scores of the AUC of each fold during all iterations (thus, 100 scores per descriptor), and results show that these differences are statistically significant. Therefore, these results confirm our hypothesis that adapting the shape of the structuring element to the texture at each point when computing the pattern spectrum is more suitable that using a fixed one.

The same image set has also been characterized by means of WCF descriptors (chapter 5), as they have provided good performance when classifying intact and damaged acrosomes. Results are shown in Table 6.6. CCF descriptors also achieved good results in that chapter. However, they needed a big number of features and our image set was not large enough to perform the experiment with them without being affected by the so-called curse of dimensionality (Donoho, 2000). Therefore, CCF has not been used in this chapter.

Table 6.6: Classification accuracy (in %) of alive and dead sperm described by WCF

Neurons	**Cycles**	**AUC**	**Accuracy (%)**		
			Overall	**Intact**	**Damaged**
2	200	0.763	68.21	73.68	61.36
2	300	0.761	68.14	73.04	61.99
2	400	0.758	68.86	74.47	61.83
3	200	0.760	68.52	73.28	62.55
3	300	0.757	68.17	72.49	62.77
3	400	0.753	68.36	73.57	61.82
5	200	0.752	68.38	72.74	62.90
5	300	0.751	67.49	71.57	62.37
5	400	0.741	67.39	71.13	62.69

Comparing the results achieved by AGPS and WCF, it is remarkable that the former outperforms the latter in terms of accuracy (69.58% against 68.21%), while the AUC is slightly better when WCF is used (0.763 against 0.746). Again, we have carried out a Wilcoxon signed-rank test using the scores of the AUC of the different folds and iterations (as it was done in chapter 5), and results show that there are not statistically significant differences between WCF and AGPS in this particular problem. When the hit rates have been used as scores, this test show that these differences are statistically significant.

6.4 Conclusion

In this chapter we have proposed an adaptive texture description approach based on mathematical morphology, which consists in computing the Pattern Spectrum using different structuring elements at each pixel whose size and shape changes to fit the surrounding texture with the basis of a distance criterion. The chosen metric has been the Geodesic distance, which has been computed by means of the Fast Marching algorithm proposed by Sethian.

This description approach is able to adapt itself to the textures without having any a priori knowledge, so possible variations within them are taken into account.

Firstly, this descriptor has been used with textures of several materials extracted from the VisTex public database to find out how general this method is. Also, it has been used to characterise images of alive and dead spermatozoa from a real semen quality application.

Results of the first experiment suggest that the superiority of the AGPS approach over the classical one strongly depends on the images under comparison. The former has more discrimination power when textures have similar texel shapes. Likewise, when texel shapes of each class are different between each other, the PS performs better than our approach. These results make sense, under the point of view that our adaptive proposal was designed with the purpose of better capturing the intrinsic geometry of textures. In the case of the alive/dead classification our adaptive Pattern Spectrum outperforms the classical one, and its performance was similar to the well known texture descriptor WCF. However, the obtained results – even with WCF – are not good enough to use this texture descriptor in a commercial system to detect alive and dead boar spermatozoa. Thus, this is still an open problem.

CHAPTER 7

ESTIMATING THE CLASS DISTRIBUTION: QUANTIFICATION

7.1 Problem formulation

Consider a classification problem with a labelled data set $S_t = \{(\mathbf{x}^k, d^k), k = 1, \ldots, K\}$ where $\mathbf{x}^k$ is the feature vector of the k^{th} element and d^k is its class label, which takes its value in $\Omega = \{d_0, d_1, \ldots d_{M-1}\}$.

Let us consider that all the samples $\mathbf{x}^k \in S_t$ have been independently recorded according to the class probability density function $p(\mathbf{x}|d_i)$ and the *a priori* probability of the class d_i in S_t is denoted by $P_t(d_i)$[1].

Let us also consider a classification model which has been generated using S_t. This classifier makes decisions in two steps: it first computes a soft output $\widehat{\mathbf{y}}^k$ and then, makes a hard decision $\widehat{d}^k \in \Omega$ based on it. It is well known that if the classifier is trained minimizing an appropriate cost function, the soft outputs $\widehat{y}_i^k$ will provide an estimation of the *a posteriori* probability of the observation $\mathbf{x}^k$ belonging to class d_i, denoted by $\widehat{P}_t(d_i|\mathbf{x})$ (Bishop, 1996).

Now, consider we have an unlabelled test data set $U = \{(\mathbf{x}^l), l = 1, \ldots, N\}$ from which there is no specific interest in knowing the class label of each instance, but the class distribution $P(d_i)$ needs to be estimated.

The naïve approach to estimate the actual class distribution is based on just counting the labels $\widehat{d}^k$ assigned by the classifier. This approach has been referred as Classify and Count (CC) in (Forman, 2008). The estimations made by this method will not be reliable since: (i) the classifier performance will drop if there is a difference between $P(d_i)$ and $P_t(d_i)$, and (ii) there is no guarantee that the errors for each class will compensate between each other.

There are some techniques based on the classifier confusion matrix (Chan and Ng, 2005; Forman, 2008), which will be briefly described in section 7.2, whose goal is to estimate the class prior probabilities. In this Thesis two new methods to accomplish this task are proposed. The first one, described in section 7.3, relies on the posterior probability estimates provided by a classifier. The second proposed method is based on measuring distributional divergences by means of the Hellinger distance, and it is described in section 7.4.

[1]The subscript t will be used for estimates based on the training set hereafter.

7.2 Previous approaches based on the confusion matrix

The confusion matrix summarizes the performance of a classifier. It is an observation of the number of elements classified as belonging to the class i when they actually belong to the class j. An example of a confusion matrix for a binary problem is shown in Table 7.1.

Table 7.1: Confusion Matrix for a binary classification problem

		Prediction	
		$\widehat{d}_1$	$\widehat{d}_0$
True class	d_1	TP	FN
	d_0	FP	TN

The count of positives P' assigned by the classifier includes both the true and false positives ($P' = TP + FP$), while the number of predicted negatives N' correspond to the sum of the true and false negatives ($N' = TN + FN$). Similarly, the number of real positive examples is $P = TP + FN$ while the number of actual negatives is $N = FP + TN$. Based on it, the following rates can be computed:

- True Positive rate: $tpr = \widehat{P}(\widehat{d}_1|d_1) = TP/P$
- False Positive rate: $fpr = \widehat{P}(\widehat{d}_1|d_0) = FP/N$
- False Negative rate: $fnr = \widehat{P}(\widehat{d}_0|d_1) = FN/P$
- True Negative rate: $tnr = \widehat{P}(\widehat{d}_0|d_0) = TN/N$

In a binary classification problem the probability that a classifier makes a positive prediction is:

$$\begin{aligned}
\widehat{P}(\widehat{d}_1) &= \widehat{P}(\widehat{d}_1|d_1) \cdot \widehat{P}(d_1) + \widehat{P}(\widehat{d}_1|d_0) \cdot \widehat{P}(d_0) = \\
&= \widehat{P}(\widehat{d}_1|d_1) \cdot \widehat{P}(d_1) + \widehat{P}(\widehat{d}_1|d_0) \cdot (1 - \widehat{P}(d_1)) = \\
&= tpr \cdot \widehat{P}(d_1) + fpr \cdot (1 - \widehat{P}(d_1)) = \\
&= tpr \cdot \widehat{P}(d_1) + fpr - fpr \cdot \widehat{P}(d_1) = \\
&= fpr + \widehat{P}(d_1) \cdot (tpr - fpr)
\end{aligned}$$

what leads to the estimation of the *a priori* probability of class d_1 as:

$$\widehat{P}(d_1) = \frac{\widehat{P}(\widehat{d}_1) - fpr}{tpr - fpr} \tag{7.1}$$

where

$$\widehat{P}(\widehat{d}_1) = \frac{P'}{P' + N'} \tag{7.2}$$

As it has been mentioned in the Introduction of this dissertation, the quantification process assumes that the within class densities $p(\mathbf{x}|d_i)$ do not change from the training to the new data sets (Saerens et al., 2002) and, therefore, there is no fundamental variation in the fpr and tpr between the train and the test set distributions. The confusion matrix can be estimated by techniques such as stratified k-fold cross validation where the value of k is recommended to be as high as possible, as suggested in (Forman, 2008).

This method has been referred to as *Adjusted Count* (AC) in (Forman, 2008) and it could be summarized as follows in a binary problem: First, a classifier is trained with the whole training set, and its performance (fpr and tpr) estimated via k-fold cross validation. Then, when the classifier is applied on a new unlabelled set, the probability of predicted positive elements $\widehat{P}(\widehat{d}_1)$ is computed according to (7.2) and finally, the estimation of the true percentage of positives is computed as (7.1).

If the problem we are dealing with has M classes, a system of M linear equations with respect to $P(\widehat{d}_j)$ should be solved in order to estimate the new class prior probabilities of all of them, as it can be seen in (7.3).

$$\widehat{P}(\widehat{d}_i) = \sum_{j=1}^{M} \widehat{P}_t(\widehat{d}_i|d_j)\widehat{P}(d_j), \quad j = 0, 1, \ldots, n \tag{7.3}$$

where $\widehat{P}(d_j)$ is the estimation of the *a priori* probability of class j and $\widehat{P}(\widehat{d}_i)$ is the observed class probability by looking at the classifier labels $\widehat{d}$.

The solution of (7.1) or (7.3), however, may be non consistent with the basic probability laws (*i.e.* values outside the interval $[0, 1]$). In a binary problem, Forman suggests (Forman, 2008) to clip the negative values to zero and fix the probability of the other class to one. In a multiclass problem, there is not an straightforward solution, though.

Based on Adjusted Count (AC), Forman also proposes the *Median Sweep* (MS) method (Forman, 2006). Briefly, it can be described as follows: first,

several confusion matrices are computed for different classification thresholds; then, the method AC is applied for each one and finally, the class distribution estimation is computed as the median of the estimations derived from each confusion matrix.

7.3 Quantification based on Posterior Probabilities

Based on a classification model whose outputs provide estimates of posterior probabilities, in this Thesis an algorithm is proposed in order to estimate the class distribution of a new dataset for a general multi-class problem (Alaiz-Rodríguez et al., 2008). It is inspired in an iterative procedure based on the EM (Expectation Maximization) algorithm proposed by Saerens *et al.* (Saerens et al., 2002) that adjusts the classifier outputs for the new deployment conditions without re-training the classifier as long as classifier outputs provide posterior probability estimates. This process indirectly computes the new class prior probabilities, what is the goal in this work.

Consider that the outputs $\widehat{\mathbf{y}}^k$ generated by the classifier for the set U are an approximation of the *a posteriori* probabilities of the classes, while the class frequencies in the training set are an estimation of the *a priori* probabilities. Thus, the prior and the posterior probability estimates are initialized with them as:

$$\widehat{P}^{(0)}(d_i|\mathbf{x}_k) = \widehat{y}_i^k \tag{7.4}$$

$$\widehat{P}^{(0)}(d_i) = \frac{|S_t^i|}{|K|} \tag{7.5}$$

where K is the total number of training examples and $|S_t^i|$ is the cardinality of the set of training examples from class i. Consider $\widehat{P}^{(r)}(d_i)$ the estimation of the new *a priori* probabilities and $\widehat{P}^{(r)}(d_i|\mathbf{x}^k)$ the new *a posteriori* probabilities at the r^{th} iteration of the algorithm. These estimations are given by (7.6) and (7.7), respectively.

$$\widehat{P}^{(r)}(d_i) = \frac{1}{N}\sum_{l=1}^{N}\widehat{P}^{(r-1)}(d_i|\mathbf{x}^k) \tag{7.6}$$

$$\widehat{P}^{(r)}(d_i|\mathbf{x}^k) = \frac{\dfrac{\widehat{P}^{(r)}(d_i)}{\widehat{P}^{(0)}(d_i)}\widehat{P}^{(0)}(d_i|\mathbf{x}^k)}{\displaystyle\sum_{j=0}^{M-1}\frac{\widehat{P}^{(r)}(d_j)}{\widehat{P}^{(0)}(d_j)}\widehat{P}^{(0)}(d_j|\mathbf{x}^k)} \quad . \tag{7.7}$$

This procedure is repeated during a certain number of iterations, or until the difference between two successive estimations is lower than a certain threshold. We refer to it as *Posterior Probability* method (PP) (Alaiz-Rodríguez et al., 2008; González-Castro et al., 2010).

7.4 Quantification based on the Hellinger Distance

As it has already been pointed out, we focus on problems where the within-class conditional densities $p(\mathbf{x}|d_i)$ are fixed, but the class prior probabilities $P(d_i)$ may shift after the classification model is generated. When this happens, the joint probabilities $P(\mathbf{x}, d_i)$ also vary and so the unconditional density $p(\mathbf{x})$ does. This also makes the posterior probabilities $P(d_i|\mathbf{x})$ change.

The effect of shifting class distributions on the data distribution $p(x)$ is illustrated in Fig. 7.1 and Fig. 7.2 for a binary classification problem where each class is defined by an univariate Gaussian distribution.

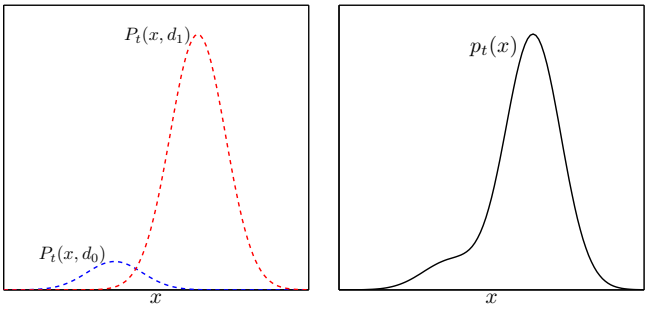


Figure 7.1: Training data. Joint probabilities $P_t(x, d_0)$ and $P_t(x, d_1)$ (left) and unconditional density $p_t(x)$ (right) for prior class probabilities $(P_t(d_0), P_t(d_1))$ equal to $(0.1, 0.9)$.

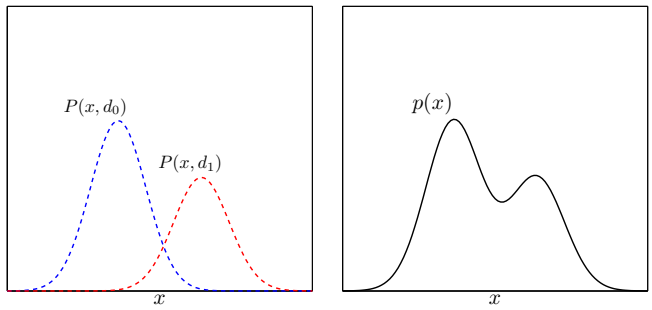


Figure 7.2: Test (future) data. Joint probabilities $P(x, d_0)$ and $P(x, d_1)$ (left) and unconditional density $p(x)$ (right) for prior class probabilities $(P(d_0), P(d_1))$ in the test set equal to $(0.6, 0.4)$.

The joint probabilities $P_t(x, d_0)$ and $P_t(x, d_1)$ for the training dataset with class priors $(P_t(d_0), P_t(d_1))$ and the data density $p_t(x)$ are shown in Fig. 7.1, while Fig. 7.2 plots the data distribution for the test set when the prior probabilities have changed $(P(d_0) \neq P_t(d_0), P(d_1) \neq P_t(d_1))$. This shift in class proportions may make the data distribution $p(x)$ significantly different from $p_t(x)$.

In a real practical problem we can estimate $p(x)$ from the test set U. Additionally, we are able to generate validation datasets V with any distribution $p_v(x)$ – using the training set S_t – to compute their differences with $p(x)$, in order to find the validation data distribution which is the most similar to the test data one. This process allows to estimate the actual class proportions, as the class distributions of the validation set that minimizes that difference.

When it comes to measure the difference between two probability distributions, the Kullback-Leibler divergence D_{KL} (Kullback and Leibler, 1951) becomes the most widely used option.

The KL divergence between probability distributions $p(x)$ and $q(x)$ on a finite set $\mathcal{X}$ is given by

$$D_{KL}(p||q)) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \tag{7.8}$$

This measure is always non negative, taking values in the interval $[0\ ,\ \infty)$, and $D_{KL}(p||q)) = 0$ if $p = q$. Strictly speaking, however, the KL divergence is

not a distance, since (a) in general it is asymmetric ($D_{KL}(p||q)) \neq D_{KL}(q||p)$) and (b) it does not satisfy the triangle inequality. The fact that it is not defined when $q(x) = 0$ limits also its use in certain applications where these situations arise.

The Hellinger Distance (HD) is other particular case of the family of f-Divergences (Csiszar and Shields, 2004), and it turns out to be very appealing for our purpose. Recently, it has been receiving attention in the machine learning community in order to detect failures in classifier performance due to shifts in data distributions. In particular, Cieslak and Chawla (Cieslak and Chawla, 2009) have shown that the measure of the HD is very effective in detecting breakpoints in classifier performance due to shifts in class prior probabilities. Here, we address the problem of class distribution estimation following a HD-based approach.

The Hellinger distance between two probability density functions $q(\mathbf{x})$ and $p(\mathbf{x})$ can be expressed as

$$HD(q,p) = \sqrt{\int \left(\sqrt{q(\mathbf{x})} - \sqrt{p(\mathbf{x})}\right)^2 dx} \tag{7.9}$$

which is non negative, bounded (it takes values from 0 to $\sqrt{2}$) and symmetric (*i.e.* $H(q,p) = H(p,q)$). Additionally, it is defined for whatever value of $p(\mathbf{x})$ and $q(\mathbf{x})$ and does not make any assumptions about the distributions themselves.

Similarity between two discrete data distributions can also be measured with the Hellinger distance by converting them into binned distributions with a probability associated with each of the b bins. Thus, the HD between the test data distribution (with unknown priors $P(d_i)$) and a validation data distribution with a given class distribution $P_v(d_i)$ can be estimated by measuring the HD between the unlabelled test dataset U and a validation dataset V (extracted from the available training data set according to $P_v(d_i)$) as

$$HD(V,U) = \frac{1}{n_f} \sum_{f=1}^{n_f} HD_f(V,U) \tag{7.10}$$

where n_f is the number of features and the distance between V and U according

to feature f is computed as

$$HD_f(V,U) = \sqrt{\sum_{i=1}^{b}\left(\sqrt{\frac{|V_{f,i}|}{|V|}} - \sqrt{\frac{|U_{f,i}|}{|U|}}\right)^2} \tag{7.11}$$

Note that b is the number of bins, $|U|$ the total number of test examples and $|U_{f,i}|$ the number of test examples whose feature f belongs to bin i. Likewise, $|V|$ and $|V_{f,i}|$ correspond to the validation dataset.

Estimating the test class distribution can be stated in a straightforward way as finding the class prior probabilities of the validation dataset $P_v(d_i)$ that minimize the HD with the test set. Generating validation datasets from the training dataset S_t with different prior probabilities can be conducted by subsampling and/or oversampling. However, this implies discarding and/or replicating instances and thus, losing or adding no information, which is specially serious when the samples are scarce, or when trying to generate datasets where some of the classes have very low prior probabilities.

Our approach deals with this by modelling the class-conditional probability density functions $p(\mathbf{x}|d_i)$ (assumed stationary), so that a validation data distribution $p_v(\mathbf{x})$ for a given class prior probability $P_v(d_i)$ with M classes can be computed as

$$p_v(\mathbf{x}) = \sum_{i=1}^{M} p(\mathbf{x}|d_i)P_v(d_i) \tag{7.12}$$

For the sake of clarity, we will illustrate the process of quantification based on the HD with a binary classification problem. Let us go back to the problem depicted in Figs. 7.1 and 7.2. A binned distribution for each of the class probability density functions ($p(x|d_0)$ and $p(x|d_1)$) can be obtained from the set of training examples that belong to class 0 (S_t^0) and class 1 (S_t^1), respectively, as depicted in Fig. 7.3.

Now, for any class distribution $P_v(d_i)$ we are able to model the data distribution $p_v(x)$ according to (7.12). Fig. 7.3(c) depicts the data distribution for class prior probabilities $P_v(d_0) = 0.6$ and $P_v(d_1) = 0.4$. This way, we use all the data available, and there is no need to replicate or discard samples.

Therefore, the HD between $p(\mathbf{x})$ and $p_v(\mathbf{x})$ according to feature f can be computed based on the available labelled data set S_t by means of (7.11) but making the substitution

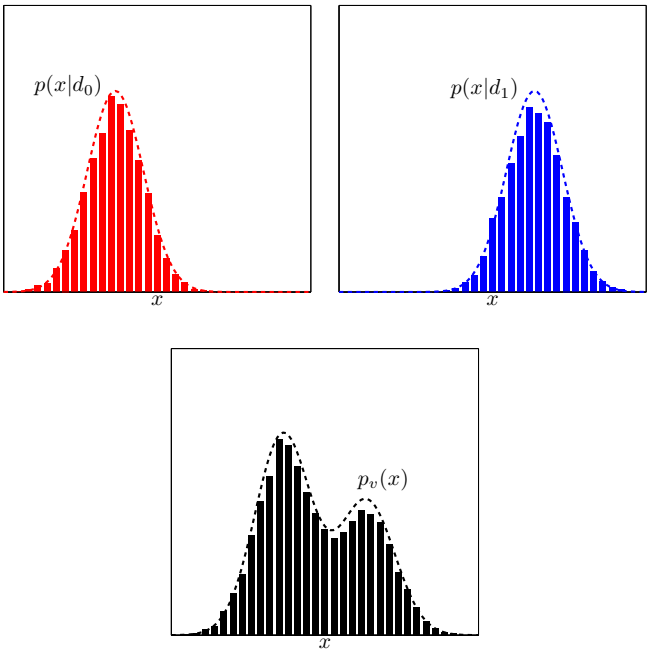


Figure 7.3: Binned distributions of the class probability density functions ($p(x|d_0$ and $p(x|d_1)$ used to model a data distribution $p_v(x)$ with $P_v(d_0) = 0.6$ and $P_v(d_1) = 0.4$.

$$\frac{|V_{f,i}|}{|V|} = \frac{|S^0_{t,f,i}|}{|S^0_t|} P_v(d_0) + \frac{|S^1_{t,f,i}|}{|S^1_t|} P_v(d_1) \tag{7.13}$$

where $P_v(d_0) = 1 - P_v(d_1)$ (in the binary case), $|S^0_t|$ is the total number of training examples that belong to class-0 and $|S^0_{t,f,i}|$ is the number of training examples from class-0 whose feature f belongs to bin i. Similarly, $|S^1_t|$ and $|S^1_{t,f,i}|$ are the equivalent measures for class-1.

Therefore, it is straightforward to simulate validation data distributions with any probabilities $P_v(d_i)$ and measure their HD with the unlabelled test data distribution according to (7.10), (7.11) and (7.13). Finally, through a search in the probability space, the estimated a priori probability of the test set is the one that minimizes this HD distance.

Let us consider that the test data follow the distribution depicted in Fig. 7.2(b) where the unknown class prior probabilities of the test set are $P(d_0) = 0.6$ and

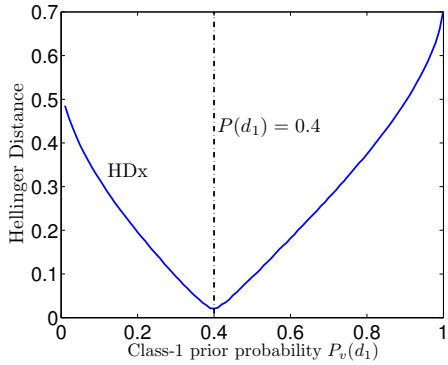


Figure 7.4: Hellinger distance between the test data distribution $p(\mathbf{x})$ and different validation data distributions $p_v(\mathbf{x})$ generated for class prior probabilities that vary from $P_v(d_1) = 0$ to $P_v(d_1) = 1$. The dashed vertical line represents actual the class-1 prior probability $P(d_1)$ of the test data set. Data are defined in a one dimensional space ($n_f = 1$).

$P(d_1) = 0.4$. Fig. 7.4 plots the Hellinger distance between the test set distribution and several validation distributions with $P_v(d_1)$ that ranges in the interval $[0, 1]$. It can be noticed that the minimum HD is achieved for the class prior probability ($P_v(d_1) = 0.4$) that matches the unknown test class distribution.

Data sparseness is a problem that usually have to be faced in real practical applications. Under operational conditions, training data sets are very likely not to be fully representative in all regions of the n_f dimensional space, in particular when the data dimensionality is high. When this happens, the estimated HD curve in Fig. 7.4 may be less reliable than that one obtained measuring the HD between the classifier output distributions. In this case, the problem is simplified because distributional divergences are measured with data (the classifier outputs) defined in a one dimensional space (for a two class problem) or in a $M - 1$ space for a general multi-class problem with M classes. These two approaches will be called HDx (HD between the feature vectors $\mathbf{x}$) and HDy (HD between the output vectors $\mathbf{y}$) (González-Castro et al., 2010), and a comparison of both is shown in Fig. 7.5 for a binary classification problem where data follow Gaussian distributions defined in a space with 20 dimensions ($n_f = 20$). Notice that HDy has a higher convexity so that estimating where its minimum lies is more reliable when the amount of instances is limited.

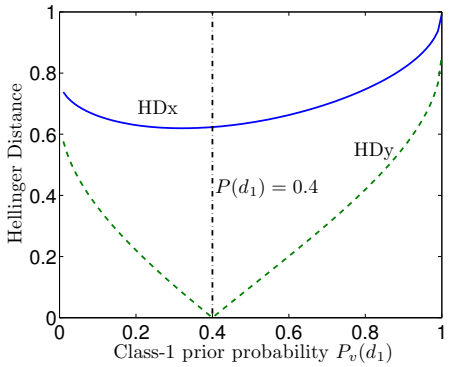


Figure 7.5: Hellinger distance between the classifier output distributions (curve HDy) and the data itself (curve HDx) of a test set and different validation settings. Data are defined in a twenty dimensional space ($n_f = 20$).

Our approach to estimate the new class distribution based on the Hellinger distance is summarized in Algorithm 1. The HDy alternative differs in the use of the classifier scores and the $M - 1$ dimensional space in which they are defined. Running the algorithm for a wide range of bins, the class prior probabilities are estimated as the median of the individual estimations.

Algorithm 1 ClassDistributionEstimate_HD (S_t, U, b)

for $P_v(d_1) = 0$ to 1 in small steps **do**
  for $f = 1$ to n_f **do**
    Compute HD_f according to (7.11), using (7.13) with $P_v(d_1)$
  end for
  $HD[P_v(d_1)] = \frac{1}{n_f} \sum_{f=1}^{n_f} HD_f(P_v(d_1))$
end for
return $\widehat{P}(d_1) = arg\ min(HD)$

7.5 Comparison of Quantification methods

7.5.1 Datasets

In this Section we have evaluated the performance of the quantification methods on 15 binary datasets, 14 of them extracted from the UCI (Frank and Asuncion,

2010) and other one taken from the ELENA project[1]. These datasets are described in Appendix A. They have a very wide range of size, number of features and class proportions, in order to provide diverse scenarios for the experiments (see details in Table 7.2).

A few classification problems were not binary and have been converted to two class datasets grouping classes as follows:

The Contraceptive Method Choice (CMC) dataset has originally three classes, which have been grouped into two: using contraceptive methods (class 0) or not (class 1).

Two datasets have been produced from the "Letter Recognition" one. The first one is called Letters (G), and it is aimed to recognize the letter G (class 1) against the others. We have called the other one Letters (H), and, similarly, its goal is to detect the letter H (class 1) amongst the others.

The original Page Blocks Classification dataset provides information about 5 types of blocks of a page layout, which have been put into two classes: Class 1 is formed by the pictures, and class 0 comprises the other blocks.

The Semeion Handwritten Digit dataset represents ten handwritten digits. In this case, the problem has been reduced to detect the digit 8 against the others.

In the case of the Red and White Wine Quality datasets (Cortez et al., 2009) the original target classes are the rate of the wine (from 0 to 10). We have reduced the target classes to *good quality wine* (interval $[0, 5]$) against *bad quality wine* (rates in $[6, 10]$).

Finally, the ten target classes in the Yeast dataset have been grouped into two: proteins in the nucleus or in other location inside the cell.

7.5.2 Performance Metrics

The mismatch between the real class distribution and the estimation provided by the different quantification methods assessed in this Thesis is measured by means of the Mean Absolute Error (MAE) and the Mean Relative Error (MRE).

[1] `http://www.dice.ucl.ac.be/neural-nets/Research/Projects/ELENA/elena.htm`

Table 7.2: UCI Datasets description

Dataset	Samples	Features	Class-1 prop. (%)
Breast Cancer Wisconsin	699	9	34.48%
CMC	1473	9	22.61%
Coil	9822	85	5.97%
Diabetes	770	8	38.57%
German Credits	1000	24	30.00%
Letters (G)	20000	16	3.87%
Letters (H)	20000	16	3.67%
Mammographic mass	830	5	48.55%
Page blocks (picture)	5473	10	2.10%
Phoneme	5404	5	29.35%
Semeion (8)	1593	256	9.73%
Spambase	4601	57	39.40%
Wine quality (red)	1599	11	3.94%
Wine quality (white)	4898	11	3.74%
Yeast	1484	8	28.91%

The absolute error, shown in (7.14) focuses on the class of interest – called class-1 in all our experiments – and is defined as the absolute value of the difference between its actual prior probability and the estimated one.

$$MAE(P(d_1), \widehat{P}(d_1)) = |P(d_1) - \widehat{P}(d_1)| \tag{7.14}$$

However, the MAE does not allow to evaluate the importance of the error. For example, if the MAE is 3, it can be produced by either an estimated distribution of 23% when the actual distribution is 20%, or an estimation of 4% when the real distribution is 1%. The MAE does not give this information, although in this second case the mismatch is more serious. The Mean Relative Error (MRE) metric – shown in (7.15) – does include such information.

$$MRE(P(d_1), \widehat{P}(d_1)) = \frac{|P(d_1) - \widehat{P}(d_1)|}{P(d_1)} \cdot 100 \tag{7.15}$$

7.5.3 Neural Network Classifier

In order to accomplish the quantification task, all methods except for the HDx require the use of a classifier. In addition, PP requires that the classifier outputs provide class posterior probability estimates. Classification was carried out

with a back-propagation Neural Network with one hidden layer and a logistic sigmoid transfer function for the hidden and the output layers. Learning was carried out with a momentum and adaptive learning rate algorithm. It is well known that when the training is carried out minimizing some loss functions such as the Mean Square Error used in this work (Bishop, 1996), the outputs provided by this model are estimates of class posterior probabilities.

Data were normalized with zero mean and standard deviation equal to one. The neural network architecture as well as the number of training cycles were determined by 10-fold cross validation.

7.5.4 Quantification based on the Hellinger distance: Empirical evaluation

The aim of the following experiment is to assess the performance of the quantification methods based on measuring the Hellinger distance between test and validation distributions – HDx and HDy –. Each dataset has been divided into two different subsets by stratified sampling: 30% of the elements are to be used as the test set, U, while the remaining 70% are planned to be used to train the classifier and generate the validation sets. From now on this set will be called S_t. Note that the test sets will have the same proportion of elements from the minority class (also called class-1) as the whole dataset has. Therefore, there will be 15 different scenarios where class-1 distributions vary from 2.10% to 48.55% (see Table 7.2).

In order to face a situation of uncertainty in the class distribution, the neural network classifier has been trained with a balanced dataset, which has all the instances of S_t that belong to the minority class and the same number of the majority class. The configuration of each network has been determined experimentally by 10-fold cross validation. Table 7.3 shows the number of training cycles and neurons in the hidden layer that lead to the optimal configuration in terms of misclassification rate for each database.

In this work, the approaches HDx and HDy require the computation of a binned distribution. Although preliminary experimental results did not show a high sensitivity to the number of bins b, in order to get a more robust estimation, b was chosen from 10 to 110 in steps of 10, and the final estimated a priori probability was taken as the median of these 11 estimations.

Table 7.3: Neural Network configurations with each classification problem

Datasets	Cycles	Nodes	Error rate (%)
Breast Cancer	200	2	4.64
CMC	200	2	37.04
Coil	200	2	30.39
Diabetes	200	2	20.43
German Credits	200	2	32.15
Letters (G)	400	5	7.67
Letters (H)	400	5	11.58
Mammographic mass	200	2	17.85
Page blocks (picture)	300	2	6.60
Phoneme	300	5	19.12
Semeion (8)	200	3	13.60
Spambase	300	3	7.34
Wine quality (red)	300	2	25.91
Wine quality (white)	400	2	28.40
Yeast	400	2	28.73

For each problem, results are the average of 50 randomly extracted test sets. Final results are the average of the estimations carried out for the 50 test sets. Tables 7.4 and 7.5 show the MAE and MRE, respectively, achieved by HDx and HDy. The former achieves good performance in all problems (with MAE in the order of 10^{-2}), but it is clearly outperformed by HDy, which achieves lower absolute errors, except in the case of Mammographic mass.

This improvement is specially clear, for instance, in the problem Page Blocks (picture), where the MRE is 16.28% when the quantification is made with the HDy, and 49.33% when using HDx. Regarding the MAE, HDy achieves 0.004 and HDx, 0.011. In this dataset $P(d_1) = 0.021$, which means that estimating class proportions with HDx would produce $\widehat{P}(d_1)$ between 0.010 and 0.032, while it would be more accurate (between 0.017 and 0.025) when using HDy. Other example is White Wine Quality, where the MRE is 28.97% with HDy, while HDx makes this error increase up to 56.20%. In this dataset $P(d_1) = 0.037$, so using HDx the estimations of $\widehat{P}(d_1)$ would be between 0.016 and 0.058, while HDy would be between 0.026 and 0.048. This is analogous to the results of the Red Wine Quality dataset.

We have also performed the Wilcoxon signed-rank test (Wilcoxon, 1945) with $\alpha = 0.05$ taking both the MAE and the MRE as scores, and their results

Table 7.4: MAE of the quantifications of the UCI databases with the methods HDx and HDy.

Dataset	HDx	HDy
Breast Cancer	0.014	0.012
CMC	0.064	0.038
Coil	0.017	0.012
Diabetes	0.050	0.028
German Credits	0.053	0.034
Letters (G)	0.005	0.002
Letters (H)	0.006	0.003
Mammographic mass	0.031	0.032
Page blocks (picture)	0.011	0.004
Phoneme	0.015	0.013
Semeion (8)	0.017	0.016
Spambase	0.013	0.006
Wine quality (red)	0.019	0.012
Wine quality (white)	0.021	0.011
Yeast	0.035	0.026

Table 7.5: MRE (in %) of the quantifications of the UCI databases with the methods HDx and HDy.

Dataset	HDx	HDy
Breast Cancer	4.04	3.58
CMC	28.16	16.81
Coil	27.78	19.78
Diabetes	12.87	7.37
German Credits	17.53	11.16
Letters (G)	13.59	4.07
Letters (H)	17.09	7.38
Mammographic mass	6.36	6.66
Page blocks (picture)	49.33	16.28
Phoneme	5.10	4.57
Semeion (8)	17.36	16.36
Spambase	3.32	1.61
Wine quality (red)	48.67	29.54
Wine quality (white)	56.20	28.97
Yeast	12.07	9.05

show that these differences are statistically significant, so it confirms the superiority of HDy over HDx. Results of this test when the MRE (Table 7.5) were taken as scores are shown in Table 7.6. The test produced the same results

when it was conducted with the MAE.

Table 7.6: Wilcoxon signed-rank test of the methods HDx and HDy.

Comparison	R^+	R^-	p-**Value**	**Null hypothesis of equality**
HDy *vs*. HDx	2	120	0.00018	Rejected (HDy outperforms HDx)

7.5.5 Comparison of Quantification Methods

In this section the method HDy (which outperforms HDx, as it has been shown in section 7.5.4) will be compared with CC, AC, MS and PP. This comparison has been carried out as it was done in section 7.5.4. The confusion matrices required for the methods AC and MS were estimated from the training set by 50-fold cross validation, as suggested in (Forman, 2008). The absolute and relative errors made by each quantification method on the assessed databases, as well as their ranks are shown in Tables 7.7 and 7.8, respectively. The final tier of the Table shows the average rank over the 15 problems.

Table 7.7: MAE of the estimations made by the quantification methods HDy, CC, AC, MS and PP.

Dataset	**CC**	**AC**	**MS**	**PP**	**HDy**
Breast Cancer	0.022 (5)	0.011 (1.5)	0.011 (1.5)	0.014 (4)	0.012 (3)
CMC	0.215 (5)	0.069 (2)	0.110 (4)	0.078 (3)	0.038 (1)
Coil	0.311 (5)	0.091 (3)	0.067 (2)	0.207 (4)	0.012 (1)
Diabetes	0.052 (5)	0.039 (3)	0.048 (4)	0.034 (2)	0.028 (1)
German Credits	0.142 (5)	0.090 (3)	0.097 (4)	0.083 (2)	0.034 (1)
Letters (G)	0.074 (5)	0.008 (2)	0.010 (3)	0.030 (4)	0.002 (1)
Letters (H)	0.105 (5)	0.008 (2)	0.011 (3)	0.038 (4)	0.003 (1)
Mamm. mass	0.026 (1)	0.039 (4)	0.044 (5)	0.032 (2.5)	0.032 (2.5)
Page (picture)	0.070 (5)	0.012 (3)	0.009 (2)	0.018 (4)	0.004 (1)
Phoneme	0.132 (5)	0.018 (3)	0.020 (4)	0.017 (2)	0.013 (1)
Semeion (8)	0.090 (5)	0.017 (3)	0.016 (1.5)	0.045 (4)	0.016 (1.5)
Spambase	0.015 (5)	0.008 (3)	0.010 (4)	0.007 (2)	0.006 (1)
Wine (red)	0.269 (5)	0.089 (3)	0.087 (2)	0.224 (4)	0.012 (1)
Wine (white)	0.231 (5)	0.052 (3)	0.037 (2)	0.153 (4)	0.011 (1)
Yeast	0.155 (5)	0.043 (3)	0.075 (4)	0.038 (2)	0.026 (1)
Avg. rank	4.733	2.767	3.067	3.167	1.267

Results highlight the importance of using an estimation method rather than relying on just counting the classifier predictions, as the method CC does (with

Table 7.8: MRE (in %) of the estimations made by the quantification methods HDy, CC, AC, MS and PP.

Datasets	CC	AC	MS	PP	HDy
Breast Cancer	6.41 (5)	3.24 (2)	3.22 (1)	4.18 (4)	3.58 (3)
CMC	94.9 (5)	30.52 (2)	48.45 (4)	34.64 (3)	16.81 (1)
Coil	521.1 (5)	152.73 (3)	111.71 (2)	345.97 (4)	19.78 (1)
Diabetes	13.54 (5)	10.05 (3)	12.38 (4)	8.92 (2)	7.37 (1)
German Credits	47.23 (5)	30.14 (3)	32.46 (4)	27.53 (2)	11.16 (1)
Letters (G)	191.36 (5)	21.58 (2)	25.82 (3)	76.24 (4)	4.07 (1)
Letters (H)	286.42 (5)	21.33 (2)	28.96 (3)	103.74 (4)	7.38 (1)
Mammog. mass	5.37 (1)	7.96 (4)	9.11 (5)	6.55 (2)	6.66 (3)
Page (picture)	327.33 (5)	56.78 (3)	43.81 (2)	82.16 (4)	16.28 (1)
Phoneme	44.78 (5)	6.21 (3)	6.91 (4)	5.79 (2)	4.57 (1)
Semeion (8)	91.81 (5)	17.41 (3)	16.16 (1)	45.93 (4)	16.36 (2)
Spambase	3.67 (5)	2.04 (3)	2.49 (4)	1.88 (2)	1.61 (1)
Wine (red)	678.61 (5)	224.23 (3)	218.62 (2)	564.58 (4)	29.54 (1)
Wine (white)	616.63 (5)	139.8 (3)	98.81 (2)	408.84 (4)	28.97 (1)
Yeast	53.6 (5)	14.91 (3)	25.91 (4)	12.96 (2)	9.05 (1)
Avg. rank	4.733	2.8	3	3.133	1.3

MAE up to 0.31, or MRE up to 678%). According to the results, the Hellinger Distance-based procedure is more reliable than the other quantification methods. It is noticeable that the MAE of HDy is never too high even in the case of imbalanced datasets. The MAE achieved by the method HDy is always lower than 0.040, and it outperforms the other methods in most cases. In general terms, *it has the smallest average rank*. The naïve approach CC is the worst method in all cases except in the case of the mammographic mass problem, where the a priori probability of the test set (0.49) is the most similar to the training set proportion (0.50).

The CMC dataset has a $P(d_1) = 0.23\%$, and the MRE is 94.9% when the estimations are made by CC, 34.64% with PP, 30.52% and 48.45% with AC and MS, while this error falls to 16.81% when the estimation method is HDy. With regard to the Coil dataset, whose natural proportion is 0.059%, the MRE of the naïve CC is 521%, 346% by PP and around 120% and 111% by AC and MS respectively, while this error is 19.78% when using HDy. The MAEs for this dataset are 0.311 (CC), 0.207 (PP), 0.091 and 0.067 (AC and MS), and 0.012, when using HDy. This means that while HDy estimates that the class-1 proportion is between 4.77% and 7.17%, it is between 0% and 15% when using

AC, or it goes up to 37% when counting the predictions of the classifier. With regard to a dataset where the class proportions are even more imbalanced, such as Page Blocks (picture), where $P(d_1) = 0.021\%$, the MRE is 327% when the naïve method CC is used, 82% with PP, 56.78% and 43.8% with AC and MS, and 16.28% when HDy is used. This means that the estimation of the class-1 proportion produced by HDy is 1.7% - 2.50%. Using MS, this estimation may be between 1.2% and 3%, 0.3% - 3.9% with PP, or up to 9.1% when CC is used.

In order to determine whether these differences are significant or not we have carried out a statistical comparison between the performance of the methods. We have specifically selected the Wilcoxon signed-rank test (Wilcoxon, 1945) with $\alpha = 0.05$. We have made comparisons between the algorithm which has achieved the best average rank (HDy) and the others. These tests, whose results can be seen in Table 7.9 show that there are statistically significant differences between HDy and the other methods.

Table 7.9: Wilcoxon signed-rank test for HDy against the other methods.

Comparison	R^+	R^-	p**-Value**	**Null hypothesis of equality**
HDy *vs*. CC	1	119	0.00012	Rejected (HDy outperforms CC)
HDy *vs*. AC	2	118	0.00018	Rejected (HDy outperforms AC)
HDy *vs*. MS	3	117	0.00031	Rejected (HDy outperforms MS)
HDy *vs*. PP	1	119	0.00012	Rejected (HDy outperforms PP)

7.6 Conclusion

In this chapter the problem of automatically estimating the proportion of data from each class on an unlabelled dataset (also known as quantification) has been addressed.

Specifically, we have presented and assessed the performance of the approaches based on Hellinger Distance (HDx and HDy) and Posterior Probability (PP) proposed in this Thesis. They have also been compared with Adjusted Count (AC) and Median Sweep (MS), as well as with the naïve approach Classify and Count. The experiment presented in this chapter has been carried out using 15 public databases taken from real domains, well known by the machine

learning community. These datasets provide a very wide range of class distributions and classification performances, thus, a wide range of scenarios. In addition, all the conclusions about the results are supported by non-parametric statistical tests.

Results show that using a quantification method is clearly better than the naïve approach, specially when the datasets are imbalanced. With regard to the quantification methods, HDy appears to be very appealing in this task, as it outperforms both HDx and the other methods, and it makes very good estimations, even if the datasets are very imbalanced.

CHAPTER 8

BOAR SEMEN QUALITY ASSESSMENT: AN EMPIRICAL STUDY

In this chapter, an empirical study has been carried out using real data from seminal quality control applications. Two experiments have been conducted with the goal of assessing the quantification approaches proposed in this Thesis (PP, HDx and HDy) and comparing them with the previous baseline methods.

Both experiments were designed following the same methodology. The first one uses the WCF descriptors, which characterise intact and damaged acrosomes by means of their texture (see chapter 5). Let us remind that this dataset contains 1849 instances, divided into 945 with damaged (class 1) and 904 with intact acrosome (class 0).

The second one uses the Adaptive Geodesic Pattern Spectrum proposed in this Thesis – extracted from the texture of alive and dead spermatozoa in chapter 6 –. Let us remind that the alive and dead datasets have 845 elements, 470 of them are alive (class 0) and 375 are dead (class 1).

These experiments are presented in sections 8.1 and 8.2, respectively.

8.1 Quantification of intact and damaged acrosomes

The performance of the quantification methods has been assessed on an application of semen quality control (see sections 3.2 and 5.1).

The training set has the 70% of the minority class examples from the whole dataset and the same number of elements from the majority class. Test sets have fixed size of 280 instances. Both sets are mutually exclusive and randomly extracted. We have evaluated ten scenarios where the proportion of damaged acrosomes in the test phase (operational environment) varies from 0.05 to 0.50. We have explored this wide range of deployment conditions in order to evaluate the algorithms. However, only the samples with proportion of damaged acrosomes equal or lower than 0.20 have interest from the point of view of veterinarian practice.

The neural network has 3 neurons in the hidden layer and is trained during 400 cycles (see Table 5.2), as it is the configuration which achieved the lowest error rate in the classification experiments of chapter 5 (3.57%). The rest of the design of the experiment is identical to the one described in sections 7.5.4 and 7.5.5.

Tables 8.2 and 8.1 show the MAE and MRE, repectively, of HDx and HDy, as well as CC, AC, MS and PP for each of the ten scenarios. A Wilcoxon signed-rank test between the method that achieves the lowest error at each training proportion and the others has been performed. To carry out the test, we used in this case the performance values of the 50 iterations (Moreno-Torres et al., 2010) resultant from the experiment. The best method, as well as the statistically equivalent ones for each scenario are underlined in those tables.

Table 8.1: Sperm cell data set. MRE of the quantification methods for ten different test scenarios.

$P(d_1)$	**CC**	**AC**	**MS**	**PP**	**HDx**	**HDy**
0.05	71.63	21.18	25.53	10.78	124.08	13.92
0.10	32.10	10.31	12.40	7.46	56.28	7.48
0.15	18.50	6.15	7.39	4.71	33.49	4.88
0.20	11.59	4.78	5.52	4.10	20.24	4.20
0.25	7.87	3.73	3.94	3.02	12.67	3.22
0.30	5.42	2.79	3.04	2.64	7.89	2.72
0.35	3.75	2.48	2.52	2.29	5.59	2.41
0.40	2.53	2.24	2.20	2.05	4.58	2.24
0.45	1.70	1.83	1.80	1.68	4.94	1.76
0.50	1.49	1.66	1.64	1.56	5.17	1.55

Table 8.2: Sperm cell data set. MAE of the quantification methods for ten different test scenarios.

$P(d_1)$	**CC**	**AC**	**MS**	**PP**	**HDx**	**HDy**
0.05	0.036	0.011	0.013	0.005	0.062	0.007
0.10	0.032	0.010	0.012	0.008	0.056	0.008
0.15	0.028	0.009	0.011	0.007	0.050	0.007
0.20	0.023	0.010	0.011	0.008	0.041	0.008
0.25	0.020	0.009	0.010	0.008	0.032	0.008
0.30	0.016	0.008	0.009	0.008	0.024	0.008
0.35	0.013	0.009	0.009	0.008	0.020	0.008
0.40	0.010	0.009	0.009	0.008	0.018	0.009
0.45	0.008	0.008	0.008	0.008	0.022	0.008
0.50	0.008	0.008	0.008	0.008	0.026	0.008

For test environments with proportions of damaged acrosomes between 0.05 and 0.25, HDy and PP outperformed the others, whereas they are statistically equivalent. What is more important, their estimations lead to very low MAE

in these scenarios (from 0.005 to 0.008) what are promising results in the application field. In contrast with the strategy Classify and Count (CC) with MRE up to 71.6% (in the case of a real proportion of 0.05), this deviation is lower than 14% using HDy and PP.

PP performance strongly depends on the quality of the estimates of the class posteriori probabilities provided by the classifiers. In order to get good a priori probability estimates it is necessary that the a posteriori probabilities relative to the training set are well approximated (Saerens et al., 2002). Only a low rate of misclassifications (in this case 3.57%) does not guarantee the success of the PP method, but the classifier outputs should be well calibrated. When this is not the case, the classifier scores can be scaled in the probability space by methods like Isotonic or Logistic regression (for a review of different methods and details see (Gebel and Weihs, 2007)). The method HDy does not require output calibration, though.

Looking at the whole picture, it can be observed that HDy and PP are always the best methods in any scenario of the boar semen application. We reconfirm that HDx is not suitable either for this application as the results with the UCI datasets have also pointed out. With respect to AC and MS, both have similar performance and are only competitive with HDy and PP for deployment conditions with relatively high proportion of damaged cells (0.30 and higher, what is out of the range of practical interest). Finally, once again results show the benefits of using an estimation method instead of relying on the naïve CC. For instance, in the estimation of a proportion of 0.10, the MRE can be reduced from 32.1% (CC) to 7.4% (HDy and PP).

8.1.1 Robustness

In this section, we study the robustness of the estimation methods with respect to the classifier performance. The aim of this experiment is to explore how the quantification is affected by the classifier performance underlying all the estimation methods, and, thus, how important is the tuning of the classifier in the estimation. In order to evaluate it, we use several neural networks, all with 3 neurons in the hidden layer but trained during different training cycles, in order to obtain classifiers with different accuracies (Table 8.3 shows the overall error rate estimated by 10-fold cross validation).

Table 8.3: Classifier Error rate of the sperm cell dataset for different number of training cycles.

Cycles	75	100	150	200	400
Error Rate (%)	32.10	15.40	6.47	4.15	3.57

Fig. 8.1 shows the evolution of the MRE (in %) for each of the configurations. All subfigures are plotted with the same axis limits, so that visually we can easily appreciate how the performance is affected.

The methods AC and MS have a maximum MRE around 20% for the optimal classifier (that one trained with 400 cycles) while they reach values of 120% when the network is trained with 100 cycles and higher than 160% for 75 cycles. Similarly, results show that the PP approach achieves the best estimations for an optimal classifier, but its performance strongly depends on the training conditions. Thus, when the network is trained during 75 cycles its MRE may be higher than 160%, while the maximum MRE when the number of training cycles is 400 is lower than 15%. With respect to the HDy method, its deterioration is lower than other methods when the base classifier performance worsens. For example, the maximum MRE of HDy is around 12% when the number of training cycles is 200, and it rises to 30% when the network is trained during 100 cycles, while in the case of PP, this increase goes from 19% to 118%.

HDy clearly outperforms the other quantification methods when the classifier performance is poor – *e.g.* 75, 100, 150 and 200 training cycles –. This is specially noticeable when the proportion of class-1 elements in the test set is low.

Graphically, it is observed that HDy is much more robust than the other methods. In order to illustrate this fact, the area under the curves (AU_MRE) in Fig. 8.1 have been computed and shown in Fig. 8.2.

Fig. 8.2 confirms low robustness for CC, PP, AC and specially for MS. The method HDy is more robust than the others with respect to changes in the classifier performance. Therefore, we can conclude that the classifier tuning is not particularly important when estimating the class prior probabilities of a dataset with a method based on the Hellinger distance.

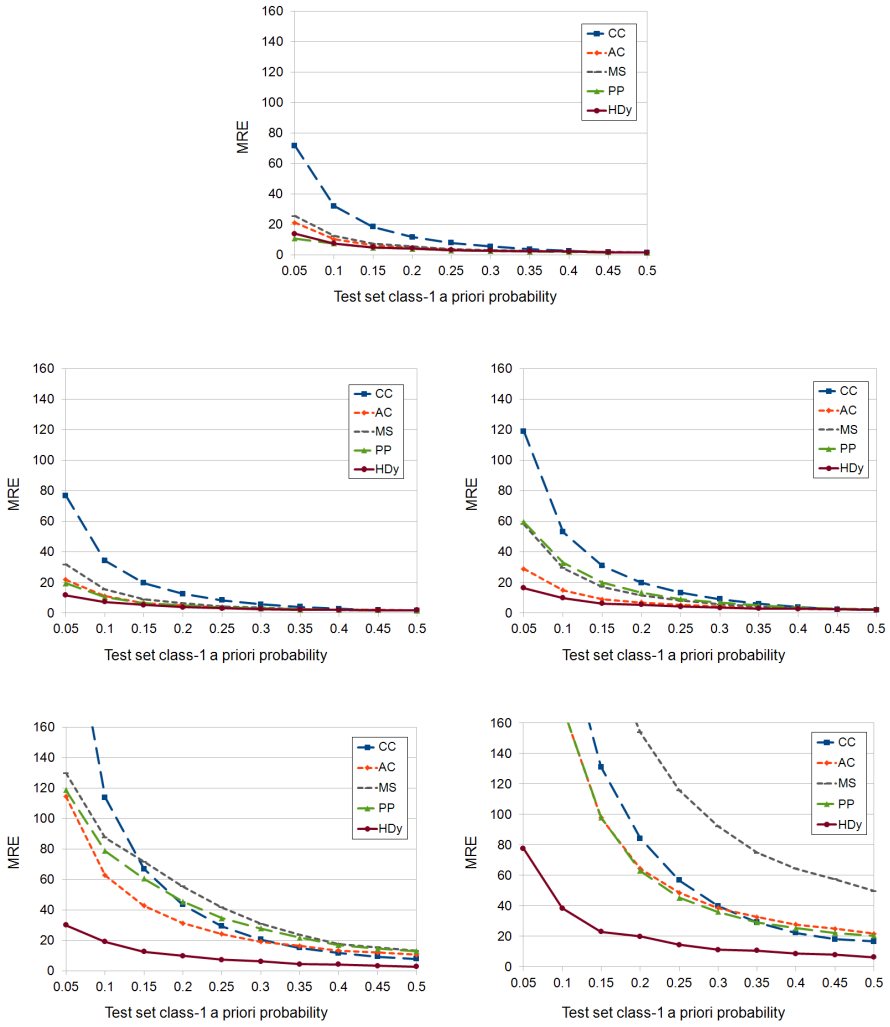


Figure 8.1: MRE for CC, AC, MS, PP and HDy using neural networks trained with 400, 200, 150, 100 and 75 cycles (from top to bottom and from left to right)

8.2 Quantification of alive and dead spermatozoa

These quantification approaches have also been used on real data with alive and dead sperm heads. Specifically, we have used a dataset containing the

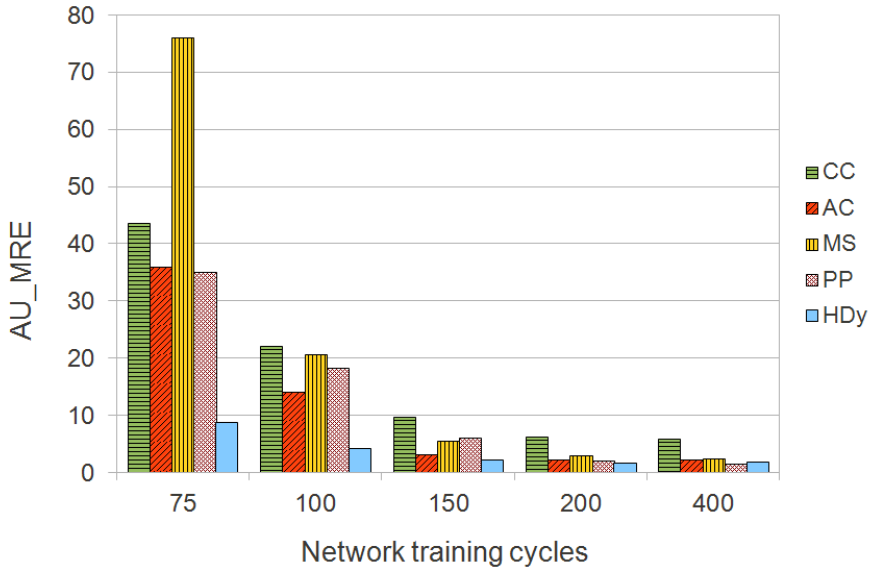


Figure 8.2: Area under the MRE curves (AU_MRE) of the quantification methods.

AGPS texture descriptors extracted from a set of alive and dead sperm images. Details about the image set, and how these descriptors have been computed can be found in chapter 6. The goal of this experiment is, on the one hand, to assess the performance of the quantification with PP, HDx and HDy, and, on the other hand, to assess whether the proposed Geodesic Pattern Spectrum is suitable for estimating the proportion of dead spermatozoa within a sample.

The dataset has 845 instances, where 470 correspond to alive spermatozoa and 375 to dead ones. The training set has the 70% of the dead samples (class-1) and the same number of elements from class-0 in order to make it to be balanced. Test sets have been fixed of 140 instances, and its proportion of dead spermatozoa varies from 0.05 to 0.50. Therefore, 10 different scenarios have been assessed. Once again, both sets are mutually exclusive and randomly extracted. The neural network had 5 neurons in the hidden layer and was trained during 400 cycles. This was the configuration which achieved the best hit rate with the AGPS descriptors, as it is shown in Table 6.5. The rest of the experiment has been designed as the intact and damaged quantification one,

shown in previous section.

The average MAE of AC, MS, PP, the naïve CC and the approaches based on the Hellinger distance for each scenario is shown in Table 8.4. Additionally, a Wilcoxon signed-rank test between the approach which achieved the lowest MAE (shown in bold) and the others was carried out using the results of the 50 iterations, as it was done in (Moreno-Torres et al., 2010). The best one is shown in bold type, and those which are statistically equivalent to it are underlined.

Table 8.4: MAE of the quantification methods for ten different test scenarios with the alive and dead AGPS data set.

$P(d_1)$	**CC**	**AC**	**MS**	**PP**	**HDx**	**HDy**
0.05	0.276	0.053	0.043	0.059	0.062	0.062
0.10	0.245	0.061	0.065	0.058	0.057	0.056
0.15	0.217	0.065	0.084	0.059	0.061	0.059
0.20	0.182	0.072	0.100	0.061	0.052	0.059
0.25	0.148	0.074	0.106	0.064	0.057	0.061
0.30	0.118	0.072	0.102	0.055	0.051	0.052
0.35	0.088	0.071	0.097	0.062	0.058	0.059
0.40	0.056	0.065	0.094	0.055	0.055	0.053
0.45	0.036	0.073	0.100	0.056	0.057	0.057
0.50	0.033	0.070	0.096	0.056	0.067	0.059

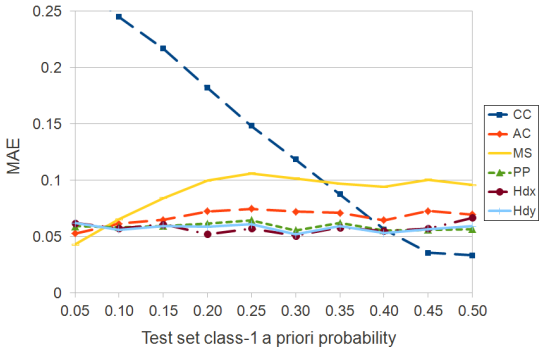


Figure 8.3: MAE for the quantification of the alive and dead sperm described by the Adaptive PS.

Results of the quantification with the Adaptive Geodesic Pattern Spectrum

(AGPS), shown in Table 8.4 and Fig. 8.3, evince the advantages of using a quantification approach instead of relying on the naïve approach of counting the predictions made by the classifier, except for actual proportions of 45%–50%, probably because the false positives and false negatives cancel each other out. It is quite remarkable that, in spite of the high error rate in classification (around 30%), the quantification results are reasonable. For instance, if the test set proportion of dead cells is 0.20. the absolute error is 0.05. Note that the MAE for CC is 0.18 for that scenario. The non parametric Wilcoxon signed-rank tests show that, generally, AC, PP, HDx and HDy show similar performance.

8.3 Conclusion

In sperm quality assessment tasks the final goal is not classifying each individual spermatozoon, but estimating the proportion of dead, or acrosome-damaged cells.

We have assessed the quantification methods proposed in this Thesis, as well as the naïve approach CC and the methods PP, AC and MS on two real semen quality control datasets. One of them contains data of intact and damaged acrosomes and the other one, alive and dead spermatozoa, both characterized by means of texture descriptors. This evaluation has been carried out in 10 different scenarios where the actual a priori probabilities changed from 5% to 50% damaged and dead cells, respectively. Non-parametric statistical tests have been carried out in order to validate the results.

Regarding the quantification of the damaged acrosomes (whose texture had been described by means of the WCF descriptors), HDy performed better than all the other methods, except for PP, which showed no statistically significant differences when the network configuration was optimal. The empirical study shows that Quantification with PP and HD can be carried out with MAE lower than 0.008 for the range of interest. The naïve approach, however, leads to MAE up to 0.036. We have also evaluated the robustness of the quantification methods when the performance of the neural network drops, and results show that HDy does not depend strongly on the network configuration, while other approaches do.

According to the quantification of the dead spermatozoa, whose texture has been described by the central moments of the Geodesic Pattern Spectrum function, proposed in this Thesis, results show once again that using a quantification method is much more reliable than directly counting the classification results. The results are promising (MAE around 0.05 with PP and HDy), specially considering that the classification error rate is high (around 30%). Nonetheless, the class distribution estimation of a dataset with dead/alive cells is still an open problem that deserves more research work in order to use it as a commercial system.

CHAPTER 9

CONCLUSION

9.1 Work summary

Two main work lines have guided the work presented in this dissertation: Texture analysis and description and class proportion estimation in environments where the class distributions are imprecise. These research lines have huge interest in many applications, specially those involving quality control. In particular, in this Thesis we have focused in the automatic semen quality assessment, specifically on the evaluation of the vitality and acrosome integrity of boar spermatozoa, since these two parameters are closely related to fertility (Yanagimachi, 1994).

The task of estimating the proportion of dead or reacting spermatozoon is usually carried out manually: Veterinary experts use stains on the samples, and then they perform a visual inspection under a fluorescent illumination to count the sperm heads. Generally, manual examination entails several drawbacks in any process that involves quality control. In this work we have explored the use of digital image processing techniques and supervised learning methods to automatically evaluate the sperm vitality or the acrosome integrity by means of just the grey-level images of the sperm cells, thus avoiding the use of traditional staining procedures. Computer vision systems are faster, more accurate and would make this process more affordable. Unfortunately, up to our knowledge, there are not commercial CASA systems that detect intact *vs*. damaged acrosomes or alive *vs*. dead cells, neither any published proposals about how to carry out these tasks.

We have proposed an adaptive texture descriptor based on Mathematical Morphology and Geodesic Distance which overcomes the drawbacks of traditional description techniques, such as being applied as-are on the whole texture, making them unable to grasp its variations. Our proposal consists in extracting the Pattern Spectrum (Maragos, 1989) using a structuring element whose size and shape vary at each pixel on the basis of a geodesic distance criterion.

Additionally, a supervised learning approach is adopted, in order to discriminate dead/alive or damaged/intact sperm cells. Unlike traditional supervised learning techniques that assume stationary data distributions, the class distribution in the semen quality control setting is imprecise. Estimating the actual class prior probabilities (quantification) of a semen sample is the main goal. We have proposed two quantification methods: One based on the posterior

probability estimations provided by the classifier and another based on measuring distributional divergences, either between the data itself or the classifier outputs.

In the rest of the chapter, the main conclusions of this work and future work lines are presented in sections 9.2 and 9.3, respectively.

9.2 General conclusions

As we have pointed out, we have dealt with the estimation of class proportion of damaged or dead cells following a supervised learning approach using information extracted from their texture in grey-level images. All methods have been assessed with images obtained in a semen production center, with a digital camera and a phase contrast microscope, which allows to evaluate whether these approaches will be successful under real operational conditions.

Results are very promising, and the contributions may help veterinary researchers to automatically assess sperm quality for artificial insemination purposes, saving time and money, specially in the case of the acrosome integrity classification and quantification.

There are other specific conclusions that can be extracted from this work, as well:

1. Regarding the segmentation process, the hybrid approach proposed in this Thesis achieved a rate of correct segmentations around 97% with the alive heads and 86% with the dead ones, outperforming both the thresholding and watershed segmentation methods. Moreover, the approach for detecting bad segmented heads did not make false negative detections at all, thus it is quite reliable. Although initially this process was not a goal in this Thesis by itself, regarding the semen quality control application it is crucial that the segmentation is accurate (*i.e.* it is able to segment correctly as many heads as possible) and reliable (*i.e.* it should prevent images that are not well segmented from being further processed).

2. Curvelet transform has never been used for assessing sperm acrosome integrity. In this Thesis, it has been used in combination with statistical descriptors, and its classification performance has been compared

with other texture descriptors based on statistics extracted from the coefficients of the Wavelet transform, and with some shape descriptors, as well. The classification was carried out by means of a backpropagation Neural Network (NN) and evaluated using cross validation. Curvelet co-occurrence features showed excellent results in terms of accuracy (97%) and area under the ROC curve (0.99). Performance of the Wavelet Co-occurrence Features and first order statistics extracted from the Curvelet coefficients are quite good, as well. On the contrary, shape descriptors showed very poor efficiency.

3. The proposed Adaptive Geodesic Pattern Spectrum (AGPS) is quite innovative in terms of development and implementation. It takes into account all possible variations within a texture without having any a priori knowledge about it. This proposal has been evaluated in two different image sets:

    - The Adaptive Geodesic Pattern Spectrum has been assessed with textures of diverse materials extracted from the VisTex database and compared with the conventional Pattern Spectrum (PS), in order to find out if this descriptor is generalisable. A careful analysis of the results shows that the superiority of the AGPS over the classical PS strongly depends on the characterized textures. The AGPS outperforms the PS when textures have similar texel shapes among each other. Likewise, when texels are different in all the textures, the PS performs better than our approach. These results make sense, under the point of view that our adaptive proposal was designed with the purpose of better capturing the intrinsic geometry of the texture, and they are very promising as a starting point for a research line.
    - It has been used to describe alive and dead sperm cells, and compared with both the classical PS and the Wavelet co-occurrence features. Classification based on NN shows that the AGPS outperforms the conventional PS in terms of both accuracy (69.87% against 64.77% respectively) and AUC (0.74 against 0.66). WCF yields a lower hit rate (68.76%), but better performance regarding ROC analysis (0.76). Classification results are not good enough to consider the use of these descriptors in a commercial system.

4. The design and development of the proposed approaches for estimating class priors are quite innovative. The first one is based on the posterior probability estimates provided by a classifier (PP). The other two methods are based on the measuring the Hellinger distance either between the data distributions themselves (HDx) or between the classifier outputs (HDy). They were compared with the naïve approach of counting the predictions made by a classifier and with other methods that rely on the classifier confusion matrix.

    - These methods have been first assessed on 15 real datasets extracted from public databases, which had different number of features and provided diverse error rates with a NN classifier. Results showed that HDy outperforms all the other estimation methods, according to the average rank of the absolute and relative deviations in the estimation. In addition, Wilcoxon signed-rank tests showed that these differences were statistically significant.
    - The quantification task was evaluated with the WCF descriptors computed from intact/damaged acrosomes on different test scenarios (with a priori probabilities ranging from 5% to 50%). The empirical study showed that HDy is still better than the others in terms of absolute error (between 0.007 and 0.009) and relative errors (between 1.55% and 13.92%), but without significant differences with PP.
    - In order to assess the robustness of the methods, the same dataset with different NN configurations (which made its performances fall) were tested. Results showed that HDy is more robust to these changes, as its absolute and relative quantification errors did not increase as much as other methods did. For example, when the damaged ratio is 0.15, HDy lead to a MRE around 20%, while it was around 100% in the case of AC and PP for the same particular NN. This is remarkable, as it shows that it is not crucial to perfectly tune the classifier to obtain good estimations, which makes this method very appealing for real operational conditions.

- The same quantification techniques have also been applied on the data of the alive and dead spermatozoa described by means of the adaptive texture descriptor proposed in this Thesis under the same conditions. The best absolute errors – between 0.05 and 0.06 –, were achieved by the HD-based methods in almost all scenarios (from 10% to 40%). These results are very promising, specially considering that the error rates in classification were about 30%, although in the case of sets with few dead samples we still do not consider possible to use these techniques in a commercial application for this particular problem.

5. The proposed quantification techniques have proved to be a powerful tool that is very appealing in applications where the class distributions are not stationary, and the final goal is to estimate the class prior probabilities. Morever, these methods, can be also applied to adapt the classifier to new operating conditions in imprecise scenarios, where the goal is the individual classification of each example.

6. The methods that we have presented and developed are completely automatic and allow to segment, describe and quantify sperm cells without user intervention. They could help veterinaries in semen quality control tasks, reducing the time they spend preparing samples and counting spermatozoa under fluorescent illumination, and saving money as well, as the required equipments are not too expensive.

9.3 Future work

In this section, we summarise the main research lines that remain open.

First of all, segmentation results suggest that there is still a chance to improve the detection process, reducing the bad segmented heads that are not detected. A classifier-based method may help minimise the false positives in the detection.

Discrete Curvelet transform has achieved quite impressive results in combination with statistical texture descriptors in the classification of intact and damaged acrosomes. However, the number of features when using first and second order statistics (52 and 108, respectively) make them useful only when

the image set is large. Therefore, dimensionality reduction techniques could be applied and assessed.

Classification results on materials with the adaptive texture descriptor have shown that its performance compared with the baseline Pattern Spectrum approach depends on the resemblance of the texels in both textures. Therefore, a method to find out whether to apply the adaptive or the classical PS could be used in order to improve the classification results. This work could be extended by designing a method to extract non-flat adaptive structuring elements, instead of just flat ones. This means that the structuring element could vary not only in terms of shape and size, but also in height along the texture.

Regarding the discrimination of alive/dead heads, the extraction of the features to describe them better is an open problem that has not been solved yet. Some works carried out in our research group suggest that there could be sub-populations within the classes, which would make this problem much more complex than the characterization of the intact and damaged acrosomes.

Quantification methods have shown very good performance, even when the classification accuracy is low, in different scenarios and using different datasets. The proposed approaches are aimed to problems where there are only two classes, therefore, future works could be addressed to the extension of these methods to multi-class problems. On the other hand, distributional divergence measures alternative to the Hellinger Distance could be assessed.

Other work lines that could be addressed at:

- To tackle the problem of detecting the damage degree in the membrane of the spermatozoa (instead of damaged or not), or even if the head is already reacted or is still reacting.
- To continue the work in order to confirm or discard the presence of sub-populations within the alive and dead spermatozoa, and propose texture descriptors that are able to detect them, if necessary.
- To assess and adapt, where necessary, segmentation and texture description methods to detect the acrosome integrity and vitality in images acquired with magnifications smaller than 100x.

As a final consideration, images of spermatozoa taken from thawed semen samples should be tested with the approaches presented in this Thesis, in order

to check if the proposals work or, if not, new approaches should be developed for them.

CAPÍTULO 10

CONCLUSIÓN

10.1 Recapitulación

Dos son las líneas fundamentales que han guiado el trabajo presentado en esta Tesis: por un lado el análisis y la descripción de texturas y por otro la estimación de proporciones de las clases de un conjunto de datos en entornos donde sus distribuciones son susceptibles de variación. Estas líneas de investigación tienen un gran interés en muchas aplicaciones, especialmente aquellas relacionadas con el control de calidad. En particular, nos hemos centrado en la evaluación automática de la calidad seminal, específicamente en la evaluación de la vitalidad y de la integridad acrosómica del esperma de verraco, ya que ambas están estrechamente relacionadas con la fertilidad (Yanagimachi, 1994).

En la actualidad la estimación de la proporción de espermatozoides muertos o dañados se suele llevar a cabo manualmente: los veterinarios utilizan tinciones en las muestras, y a continuación realizan una inspección visual de la misma bajo una iluminación fluorescente para contar las células que pertenecen a cada clase. Generalmente, este tipo de exámenes manuales llevan asociados varios inconvenientes en los procesos de control de calidad. En este trabajo se explora la utilización de técnicas de procesamiento digital de imágenes y métodos de aprendizaje supervisado con el objetivo de evaluar automáticamente la vitalidad espermática y la integridad acrosómica utilizando únicamente imágenes en nivel de gris, evitando la utilización de tinciones. Los sistemas de visión artificial son más rápidos y precisos, y permitirían que este proceso fuese más asequible. Desafortunadamente, no existen sistemas CASA que distingan acrosomas íntegros y dañados, o células espermáticas vivas y muertas, ni tampoco publicaciones de investigación sobre este tema.

En esta Tesis hemos propuesto un descriptor de texturas adaptativo basado en Morfología Matemática y distancia Geodésica que resuelve uno de los problemas de los descriptores de textura convencionales: ser aplicados tal cual en toda la textura, por lo que no son capaces de captar sus variaciones locales. Nuestra propuesta consiste en extraer el *Pattern Spectrum* (Maragos, 1989) utilizando un elemento estructurante cuya forma y tamaño varía en cada píxel en función de un criterio de distancia geodésica.

Además, se ha utilizado un método de aprendizaje supervisado para distinguir las células espermátivas vivas/muertas o íntegras/dañadas. A pesar de que los métodos tradicionales asumen que la distribución de las clases de los

conjuntos de datos son estacionarias, dicha distribución es imprecisa en el caso de aplicaciones de control de la calidad seminal, en las que estimar la verdadera probabilidad a priori de las clases (cuantificación) es el objetivo principal. Hemos propuesto dos métodos de cuantificación: uno basado en las estimaciones de las probabilidades *a posteriori* realizadas por el clasificador, y otro basado en medir diferencias entre distribuciones, ya sea entre los datos originales, o entre las salidas de un clasificador ante dichos datos.

En el resto del capítulo, las conclusiones principales de esta Tesis, así como las líneas de trabajo futuras se presentan en las secciones 10.2 y 10.3, repectivamente.

10.2 Conclusiones generales

Como se ha mencionado anteriormente, hemos tratado el problema de la estimación de la proporción de células dañadas, o muertas, en muestras de semen mediante métodos de aprendizaje supervisado. Para ello se ha utilizado información extraída de su textura en imágenes en escala de gris. Todos los métodos se han evaluado con imágenes obtenidas en un centro de producción con una cámara digital, y un microscopio de contraste de fases, lo que ha permitido comprobar el rendimiento de dichos métodos bajo condiciones reales.

Los resultados son muy prometedores, y las contribuciones pueden ayudar a evaluar automáticamente la calidad espermática para tareas de inseminación artificial, ahorrando tiempo y dinero, especialmente en el caso de la clasificación y cuantificación de la integridad acrosómica.

Asimismo, existen otras conclusiones específicas que se pueden extraer de este trabajo:

1. En cuanto al proceso de segmentación, el método híbrido propuesto en esta Tesis obtuvo una tasa de segmentaciones correctas del 97 % en espermatozoides vivos y del 86 % en muertos, superando a los métodos basados en umbralización y Watershsed. Además, el método para la detección de cabezas mal segmentadas no produjo falsos negativos, por lo que es bastante fiable. A pesar de que inicialmente este proceso en sí mismo no era un objetivo de la Tesis, teniendo en cuenta la aplicación de control de la calidad seminal es imprescindible que la segmentación sea precisa (que

pueda segmentar correctamente tantas cabezas como sea posible) y fiable (debería impedir que imágenes mal segmentadas pasen a futuras fases de procesamiento).

2. La transformada *Curvelet* nunca ha sido utilizada para evaluar la integridad acrosómica. En esta Tesis se ha utilizado en combinación con descriptores estadísticos, y su rendimiento en la clasificación se ha comparado con otros descriptores estadísticos de textura calculados en las matrices de coeficientes de la transformada *Wavelet*, así como con descriptores de forma. La clasificación se llevó a cabo con una red neuronal de retropropagación y la evaluación con validación cruzada. Los descriptores obtenidos de la matriz de coocurrencia y de la transformada *Curvelet* mostraron excelentes resultados tanto en tasa de aciertos (97 %) como en área bajo la curva ROC (0,99). El rendimiento de los estadísticos de segundo orden de los coeficientes *Wavelet* y de primer orden de los coeficientes *Curvelet* también fue bueno. Por el contrario, los descriptores de forma mostraron una eficacia muy pobre.

3. El *Pattern Spectrum* Adaptativo Geodésico (AGPS en sus siglas en inglés) es innovador en su diseño e implementación. Tiene en cuenta las posibles variaciones dentro de la textura sin necesidad de poseer conocimiento a priori sobre la misma. Esta propuesta se ha evaluado en dos conjuntos de imágenes diferentes.

    - El AGPS se ha evaluado con texturas de materiales diversos de la base de datos VisTex, y se ha comparado con el *Pattern Spectrum* (PS) convencional, con el objetivo de comprobar si es generalizable. El análisis de los resultados muestra que el rendimiento de AGPS frente al PS depende mucho de la naturaleza de la textura. El primero supera al segundo cuando los téxeles de las mismas tienen formas similares entre ellos. De la misma manera, cuando estos son diferentes el PS es mejor que nuestro método. Estos resultados tienen sentido desde el punto de vista que nuestra propuesta adaptativa se diseñó para capturar mejor la estructura geométrica intrínseca de la textura, y son muy prometedores como punto de partida para una línea de investigación.

- Nuestra propuesta se ha utilizado para describir células espermáticas vivas y muertas, y su rendimiento se ha comparado con el PS clásico y con las características de las matrices de coocurrencia de los coeficientes *Wavelet* (WCF). La clasificación con redes neuronales muestra que el AGPS supera al PS convenicional tanto en precisión (69,87 % contra 64,77 %) como en área bajo la curva ROC – AUC – (0,74 contra 0,66). WCF obtuvo una tasa de aciertos más baja (68,76 %), pero una mayor AUC (0,76). Los resultados en la clasificación no son suficientemente buenos para considerar el uso de estos descriptores en un sistema comercial.

4. El diseño y desarrollo de los métodos propuestos para estimar las probabilidades a priori de las clases son bastante innovadores. El primero se basa en las estimaciones de las probabilidades a posteriori realizadas por un clasificador (PP). Los otros dos se basan en calcular la distancia de Hellinger entre distribuciones bien de los datos como tal (HDx), o de las salidas de un clasificador (HDy). Ambas se han comparado con el método "ingenuo" de contar las predicciones realizadas por un clasificador, y con otros métodos basados en la matriz de confusión del mismo.

   - Estos métodos se evaluaron en primer lugar en 15 conjuntos de datos reales, extraídos de bases de datos públicas, con diferente número de características y con diversas tasas de error obtenidas con una red neuronal. Los resultados mostraron que HDy superó al resto, de acuerdo al ranking medio de las desviaciones absoluta y relativa de la estimación. Además, el test de los signos de Wilcoxon mostró que estas diferencias fueron estadísticamente significativas.
   - La cuantificación se evaluó en un conjunto de datos basado en los descriptores WCF de acrosomas íntegros y dañados en escenarios diferentes (con probabilidades a priori que varían entre 5 % y 50 %). Este estudio mostró que HDy es mejor que los otros métodos tanto en error absoluto (entre 0,007 y 0,009) como relativo (entre 1,55 % y 13,92 %), pero sin diferencias estadísticamente significativas respecto a PP.

- Con el objetivo de evaluar la robustez de estos métodos, se utilizó el mismo conjunto de datos con diferentes configuraciones en la red neuronal (que ocasionaban que su rendimiento cayese). Los resultados mostraron que HDy es más robusto a estos cambios, puesto que sus errores relativos y absolutos no se incrementaron sustancialmente, mientras que los de los otros métodos sí lo hicieron. Por ejemplo, cuando el ratio de dañados era 0,15, HDy obtuvo un error relativo de alrededor de 20 %, mientras que en el caso de AC y PP, este fue de alrededor del 100 % con la misma red neuronal. Esto es destacable, pues demuestra que no es crucial ajustar perfectamente el clasificador para obtener buenas estimaciones, lo que hace este método muy interesante para trabajar bajo condiciones reales.
- Los mismos métodos de cuantificación se aplicaron, bajo las mismas condiciones, con los datos correspondientes al descriptor adaptativo propuesto en esta Tesis extraído de imágenes de espermatozoides vivos y muertos. Los mejores errores absolutos (entre 0,05 y 0,06) se consiguieron con los métodos basados en la distancia de Hellinger en casi todos los escenarios (entre el 10 % y el 40 %). Estos resultados son muy prometedores, sobre todo si se tiene en cuenta que las tasas de error en la clasificación estuvieron en torno al 30 %, aunque en el caso de conjuntos con pocos espermatozoides muertos, todavía no se puede considerar su uso en aplicaciones comerciales.

5. Se ha probado que las técnicas de cuantificación propuestas son una herramienta muy interesante en aplicaciones en que la distribución de las clases no es estática, y el objetivo final es estimar su probabilidad a priori. Además, estos métodos se pueden utilizar para adaptar el clasificador a nuevas condiciones operacionales en escenarios imprecisos, donde el objetivo es la clasificación de cada elemento individualmente.

6. Los métodos que se han presentado y desarrollado son completamente automáticos, y permiten segmentar, describir y cuantificar células espermáticas sin intervención del usuario. Estos métodos pueden ser de ayuda para los veterinarios en tareas de control de calidad seminal, reduciendo el tiempo que requiere la preparación de las muestras y el conteo manual de los espermatozoides bajo iluminación fluorescente. También

permiten ahorrar dinero, puesto que el equipo requerido no es demasiado caro.

10.3 Futuras líneas de trabajo

En esta sección se resumen las líneas de investigación que continúan abiertas.

En primer lugar, los resultados de la segmentación sugieren que todavía es posible mejorar el proceso de detección, reduciendo el número de cabezas mal segmentadas que no se detectan. La introducción de un clasificador puede ayudar a minimizar los falsos positivos.

La transformada *Curvelet* discreta ha obtenido resultados impresionantes en combinación con descriptores de textura estadísticos en la clasificación de íntegros y dañados. Sin embargo, el número de características cuando se usan estadísticos de primer o de segundo orden (52 y 108 respectivamente) hacen que sólo sean útiles cuando el conjunto de imágenes es grande. Por ello, podrían aplicarse y evaluarse técnicas de reducción de la dimensionalidad.

Los resultados de la clasificación de materiales con el descriptor de textura adaptativo han mostrado que su rendimiento en comparación con el *Pattern Spectrum* depende del parecido de los téxeles en las texturas que se están describiendo. Por ello, se podría explorar un método para averiguar si es mejor usar el primero o el segundo y así mejorar los resultados en la clasificación. Además, nuestra propuesta se podría extender diseñando un método para extraer elementos estructurantes no planos, de forma que su altura también variase a lo largo de la textura.

En cuanto a los resultados de la discriminación entre células espermáticas vivas y muertas, su descripción es todavía un problema abierto. Algunos trabajos que se han llevado a cabo en nuestro grupo de investigación sugieren que existen subpoblaciones dentro de las clases, lo que haría este problema mucho más complejo que la caracterización de acrosomas íntegros y dañados.

Los métodos de cuantificación han mostrado un rendimiento muy bueno, incluso cuando la precisión de la clasificación es baja, en distintos escenarios y con conjuntos de datos diferentes. Nuestras propuestas están pensadas para problemas donde sólo hay dos clases, por lo que, de cara a trabajos futuros, se podría abordar la extensión de estos métodos a problemas con más clases.

Por otro lado, se podrían evaluar métricas alternativas de divergencia entre distribuciones.

Otras líneas de trabajo se podrían dirigir a:

- Abordar el problema de detectar el grado de daño en la membrana del espermatozoide (en vez de simplemente clasificarlo como íntegro o dañado), o incluso detectar si la cabeza ya ha reaccionado o todavía está reaccionando.
- Continuar trabajando para confirmar o descartar la presencia de subpoblaciones en los espermatozoides vivos y muertos, y proponer descriptores de textura que sean capaces de detectarlas, si fuera necesario.
- Evaluar y adaptar, donde fuera necesario, los métodos de descripción de texturas para detectar la integridad acrosómica y la vitalidad espermática en imágenes obtenidas con menos de 100 aumentos en el microscopio.

Como consideración final, se podrían probar los métodos propuestos en esta Tesis con imágenes de espermatozoides en semen descongelado, para comprobar si funcionan y, en caso contrario, desarrollar nuevos métodos para ello.

Bibliography

U. Ahmad, K. Kidiyo, and R. Joseph, "Texture features based on fourier transform and gabor filters: an empirical comparison," in *Proc. International Conference on Machine Vision ICMV 2007*, 28–29 Dec. 2007, pp. 67–72. 14, 69

R. Alaiz-Rodríguez and J. Cid-Sueiro, "Minimax strategies for training classifiers under unknown priors," in *Proc. 12th IEEE Workshop on Neural Networks for Signal Processing*, 4–6 Sept. 2002, pp. 249–258. 26

R. Alaiz-Rodríguez, A. Guerrero-Curieses, and J. Cid-Sueiro, "Minimax classifiers based on neural networks," *Pattern Recognition*, vol. 38, no. 1, pp. 29–39, January 2005. 27

——, "Minimax regret classifier for imprecise class distributions," *Journal of Machine Learning Research*, vol. 8, pp. 103–130, January 2007. 27

R. Alaiz-Rodríguez, E. Alegre, V. González-Castro, and L. Sánchez, "Quantifying the proportion of damaged sperm cells based on image analysis and neural networks," in *Proceedings of the 8th conference on Simulation, modelling and optimization*. World Scientific and Engineering Academy and Society, 2008, pp. 383–388. [Online]. Available: http://portal.acm.org/citation.cfm?id=1503955.1504029 27, 100, 101, 195

F. Albregtsen, H. Schulerud, and L. Yang, "Texture classification of mouse liver cell nuclei using invariant moments of consistent regions," in *Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, S. B. . Heidelberg, Ed., vol. 970, 1995, pp. 496–502. 13

E. Alegre, V. Gonzalez-Castro, S. Suarez, and M. Castejon, "Comparison of supervised and unsupervised methods to classify boar acrosomes using texture descriptors," in *Proc. Int. Symp. ELMAR ELMAR '09*, 2009, pp. 65–70. 17, 193

E. Alegre, M. Biehl, N. Petkov, and L. Sánchez, "Automatic classification of the acrosome status of boar spermatozoa using digital image processing and LVQ." *Computers in Biology and Medicine*, vol. 38, no. 4, pp. 461–468, April 2008. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2008.01.005 17, 193

J. Andrews and J. A. Sethian, "Fast marching methods for the continuous traveling salesman problem," *Proceedings of the National Academy of Sciences*, vol. 104, no. 4, pp. 1118–1123, 2007. [Online]. Available: http://www.pnas.org/content/104/4/1118.abstract 24

S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognition Letters*, vol. 24, no. 9–10, pp. 1513–1521, June 2003. 16, 72, 199

S. Arivazhagan, L. Ganesan, and T. G. Subash Kumar, "Texture classification using ridgelet transform," in *Proc. Sixth International Conference on Computational Intelligence and Multimedia Applications*, 16–18 Aug. 2005, pp. 321–326. 16

S. Arivazhagan, L. Ganesan, and T. G. S. Kumar, "Texture classification using curvelet statistical and co-occurrence features," in *Proc. 18th International Conference on Pattern Recognition ICPR 2006*, vol. 2, 2006, pp. 938–941. 16, 69

H. Arof and F. Deravi, "Concentric circular sampling for texture analysis," in *Proc. International Conference on Image Processing*, vol. 3, 26–29 Oct. 1997, pp. 190–192. 12, 193

A. Asano, "Texture analysis using morphological pattern spectrum and optimization of structuring elements," in *Proceedings of the International Conference on Image Analysis and Processing*, IEEE, Ed., 1999, pp. 209–214. 22

A. Asano, M. Miyagawa, and M. Fujio, "Morphological texture analysis using optimization of structuring elements," in *Geometry, Morphology, and Computational Imaging. Proceedings of the 11th International Workshop on Theoretical Foundations of Computer Vision*, ser. Lecture Notes in Computer Science, S. B. Heidelberg, Ed., vol. 2616, 2003, pp. 45–56. 22, 81, 194

W. Bastian, M. Petrou, and X. Leng, "Greyscale morphology with a non-linear structuring element," in *International Conference on Digital Signal Processing 95*, 1995, pp. 366–371. 18, 193

M. E. Beletti, L. da Fontoura Costa, and M. P. Viana, "A spectral framework for sperm shape characterization." *Computers in Biology and Medicine*, vol. 35, no. 6, pp. 463–473, July 2005. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2004.03.007 31

S. Berretti, A. Del Bimbo, P. Pala, and F. J. S. Mata, "Geodesic distances for 3d-3d and 2d-3d face recognition," in *Proc. IEEE Int Multimedia and Expo Conf*, 2007, pp. 1515–1518. 23

V. S. Bharathi and L. Ganesan, "Orthogonal moments based texture analysis of CT liver images," *Pattern Recognition Letters*, vol. 29, no. 13, pp. 1868–1872, October 2008. 13

C. M. Bishop, *Neural networks for pattern recognition*. Oxford University Press, 1996. 97, 110, 207

J. Bonnel, A. Khademi, S. Krishnana, and C. Ioana, "Small bowel image classification using cross-co-occurrence matrices on wavelet domain," *Biomedical Signal Processing and Control*, vol. 4, no. 1, pp. 7–15, January 2009. 15, 69

N. Bonnet, "Some trends in microscope image processing." *Micron*, vol. 35, no. 8, pp. 635–653, 2004. 11, 192

N. Bouaynaya and D. Schonfeld, "Theoretical foundations of spatially-variant mathematical morphology part ii: Gray-level images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 5, pp. 837–850, 2008. 22

N. Bouaynaya, M. Charif-Chefchaouni, and D. Schonfeld, "Spatially variant morphological restoration and skeleton representation," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3579–3591, 2006. 22, 81

——, "Theoretical foundations of spatially-variant mathematical morphology part i: Binary images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 5, pp. 823–836, 2008. 22

B. Bouraoui, C. Ronse, J. Baruthio, N. Passat, and P. Germain, "3D segmentation of coronary arteries based on advanced mathematical morphology techniques," *Computerized Medical Imaging and Graphics*, vol. 34, no. 5, pp. 377 – 387, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.compmedimag.2010.01.001 21, 193

H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, "Linear time euclidean distance transform algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 529–533, 1995. 58

P. Buendía, C. Soler, F. Paolicchi, G. Gago, B. Urquieta, F. Pérez-Sánchez, and E. Bustos-Obregón, "Morphometric characterization and classification of alpaca sperm heads using the sperm-class analyzer computer-assisted system." *Theriogenology*, vol. 57, no. 4, pp. 1207–1218, March 2002. 31

E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modelling & Simulation*, vol. 5, no. 3, pp. 861–899, January 2006. 6, 12, 16, 42, 43, 73, 196

E. J. Candès and D. L. Donoho, "Curvelets, multiresolution representation and scaling laws," in *Proc. SPIE. Wavelet Applications in Signal and Image Processing VIII*, vol. 4119, 2000, pp. 1–12. 16, 41

R. Cárdenes, S. K. Warfield, E. Macías, and J. Ruiz-Alzola, "Occlusion points propagation geodesic distance transformation," in *Proc. Int. Conf. Image Processing ICIP 2003*, vol. 1, 2003. 23

R. Cárdenes, C. Alberola-López, and J. Ruiz-Alzola, "Fast and accurate geodesic distance transform by ordered propagation," *Image and Vision Computing*, vol. 28, no. 3, pp. 307 – 316, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2009.05.013 23

M.-H. R. Cardinal, J. Meunier, G. Soulez, R. L. Maurice, E. Therasse, and G. Cloutier, "Intravascular ultrasound image segmentation: a three-dimensional fast-marching method based on gray level distributions," *IEEE Transactions on Medical Imaging*, vol. 25, no. 5, pp. 590–601, 2006. 24

J. A. Cardozo, P. Grasa, M. T. M. no, and J. A. Cebrián, "Adición de proteínas del plasma seminal ovino durante la congelación del espermatozoide y efectos sobre su motilidad y viabilidad," *Revista Corpoica - Ciencia y Tecnología Agropecuaria*, vol. 10, no. 1, pp. 51–59, 2009. 4

G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images." *Clinical Radiology*, vol. 59, no. 12, pp. 1061–1069, December 2004. 12, 192

Y. S. Chan and H. T. Ng, "Word sense disambiguation with distribution estimation," in *Proceedings of the IJCAI05*, 2005, pp. 1010–1015. 25, 27, 97

——, "Estimating class priors in domain adaptation for word sense disambiguation," in *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006, pp. 89–96. 25, 27, 194

M. Chandraratne, S. Samarasinghe, D. Kulasiri, and R. Bickerstaffe, "Prediction of lamb tenderness using image surface texture features," *Journal of Food Engineering*, vol. 77, no. 3, pp. 492 – 499, 2006, special Section: CHISA 2004 (pp. 379-471). 12, 192

M. Charif-Chefchaouni and D. Schonfeld, "Spatially-variant mathematical morphology," in *Proc. ICIP-94. Conf. IEEE Int Image Processing*, vol. 2, 1994, pp. 555–559. 22

C. Chong, P. Raveendran, and R. Mukundan, "Translation and scale invariants of legendre moments," *Pattern recognition*, vol. 37, no. 1, pp. 119–129, January 2004. 75

D. L. Chopp, "Computing minimal surfaces via level set curvature flow," *Journal of Computational Physics*, vol. 106, no. 1, pp. 77–91, May 1993. 49

D. Cieslak and N. Chawla, "A framework for monitoring classifiers' performance: when and why failure occurs?" *Knowledge and Information Systems*, vol. 18, no. 1, pp. 83–108, 2009. 27, 103, 194

L. D. Cohen and R. Kimmel, "Global minimum for active contour models: a minimal path approach," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition CVPR '96*, 1996, pp. 666–673. 24, 194

P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," *Decision Support Systems*, vol. 47, no. 4, pp. 547 – 553, 2009, smart Business Networks: Concepts and Empirical Evidence. 108, 179

N. L. Cross and S. Meizel, "Methods for Evaluating the Acrosomal Status of Mammalian Sperm," *Biology of Reproduction*, vol. 41, no. 4, pp. 635–641, October 1989. [Online]. Available: http://www.biolreprod.org/content/41/4/635 35

I. Csiszar and P. Shields, *Information Theory and Statistics: A Tutorial (Foundations and Trends in Communications and Information The)*. Now Publishers Inc, December 2004. 103

O. Cuisenaire, "Locally adaptable mathematical morphology using distance transformations," *Pattern Recognition*, vol. 39, no. 3, pp. 405 – 416, 2006. 22, 81, 194

——, "Distance transformations: Fast algorithms and applications to medical image processing," Ph.D. dissertation, Université catholique de Louvain, October 1999. [Online]. Available: http://www.tele.ucl.ac.be/PEOPLE/OC/Thesis.html 23

A. C. A. M. de Carvalho Bessa, "Influencia en la calidad espermática de la adición de distintas concentraciones de crioprotectores para la conservación del semen canino," Ph.D. dissertation, Facultad de Veterinaria. Universidad Complutense de Madrid, 2005. 34

E. de Ves, X. Benavent, G. Ayala, and J. Domingo, "Selecting the structuring element for morphological texture classification," *Pattern Analysis & Applications*, vol. 9, no. 1, pp. 48–57, Mayo 2006. 21, 81, 193, 202

A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistic Society (STOR)*, vol. 39, no. 1, pp. 1–38, 1977. 26

J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, December 2006. 51, 52

L. Dettori and L. Semler, "A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography." *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 486–498, April 2007. 16, 69, 71, 73, 198

M. N. Do and M. Vetterli, "The finite ridgelet transform for image representation," *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 16–28, Jan. 2003. 15, 82

D. L. Donoho, "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality," August 2000, lecture, American Math. Society: Challenges of the 21st Century. [Online]. Available: http://www-stat.stanford.edu/~{}donoho/Lectures/CBMS/Curses.pdf 93

C. Drummond and R. C. Holte, "Cost curves: An improved method for visualizing classifier performance," *Machine Learning*, vol. 65, no. 1, pp. 95–130, 2006. 25

W. Duch and L. Itert, "A posteriori corrections to classification methods," in *Neural Networks and Soft Computing*, L. Rutkowski and J. Kacprzyk, Eds. Berlin, Heidelberg, New York: Physica Verlag, Springer, 2002, pp. 406–411. 26

R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley and Sons, 2001. 4, 25, 194

M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two cad approaches that both emphasize an intelligible decision process." *Med Phys*, vol. 34, no. 11, pp. 4164–4172, Nov 2007. 178

M. M. Eltoukhy, I. Faye, and B. B. Samir, "Breast cancer diagnosis in digital mammogram using multiscale curvelet transform," *Computerized Medical Imaging and Graphics*, vol. 34, no. 4, pp. 269 – 276, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.compmedimag.2009.11.002 16, 69

——, "A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram," *Computers in Biology and Medicine*, vol. 40, no. 4, pp. 384 – 391, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2010.02.002 16, 69, 198

P. Farrell, G. Presicce, C. Brockett, and R. Foote, "Quantification of bull sperm characteristics measured by computer-assisted sperm analysis (casa) and the relationship to fertility," *Theriogenology*, vol. 49, no. 4, pp. 871 – 879, 1998. [Online]. Available: http://dx.doi.org/10.1016/S0093-691X(98)00036-3 31

J. Flusser and T. Suk, "Affine moment invariants: a new tool for character recognition," *Pattern Recognition Letters*, vol. 15, no. 4, pp. 433 – 436, 1994. [Online]. Available: http://dx.doi.org/10.1016/0167-8655(94)90092-2 74

G. Forman, "Quantifying trends accurately despite classifier error and class imbalance," in *Principles and Practice of Knowledge Discovery in Databases*, 2006, pp. 157–166. 27, 99

——, "Counting positives accurately despite inaccurate classification," in *Machine Learning: ECML 2005*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005, pp. 564–575. 27

——, "Quantifying counts and costs via classifcation," *Data Mining and Knowledge Discovery*, vol. 17, no. 2, pp. 164–206, October 2008. 4, 27, 97, 99, 113, 195, 207, 209

G. Forman, E. Kirshenbaum, and J. Suermondt, "Pragmatic text mining: Minimizing human effort to quantify many issues in call logs," in *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006. 27

A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: http://archive.ics.uci.edu/ml 107, 177

B. Ganeshan, K. A. Miles, R. C. D. Young, and C. R. Chatwin, "Texture analysis in non-contrast enhanced ct: impact of malignancy on texture in apparently disease-free areas of the liver." *European Journal of Radiology*, vol. 70, no. 1, pp. 101–110, April 2009. 14

J. Gardón, C. Matás, and J. Gadea, "Efecto del protocolo de preparación de los espermatozoides bovinos sobre el patrón de reacción acrosómica," *Anales de veterinaria*, vol. 17, pp. 19–26, 2001. 37

D. L. Garner and L. A. Johnson, "Viability Assessment of Mammalian Sperm Using SYBR-14 and Propidium Iodide," *Biology of Reproduction*, vol. 53, no. 2, pp. 276–284, 1995. [Online]. Available: http://www.biolreprod.org/content/53/2/276.abstract 35

B. Gatos, I. Pratikakis, and S. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognition*, vol. 39, no. 3, pp. 317–327, March 2006. 19, 193

M. Gebel and C. Weihs, "Calibrating classifier scores into probabilities," in *Advances in Data Analysis*, ser. Studies in Classification, Data Analysis, and Knowledge Organization, R. Decker and H. J. Lenz, Eds. Springer Berlin Heidelberg, 2007, pp. 141–148. 121

M. González, E. Alegre, R. Alaiz, and L. Sánchez, "Acrosome integrity classification of boar spermatozoon images using dwt and texture techniques," in *VipIMAGE -Computational Vision and Medical Image Processing.* Taylor and Francis Group London, 2007, pp. 165–168. 17, 69, 198

R. C. Gonzalez and R. E. Woods, *Digital image processing*, P. Hall, Ed. Tom Robbins, 2002. 11, 46, 85, 192, 203

V. González-Castro, E. Alegre, P. Morala-Argüello, and S. A. Suarez, "A combined and intelligent new segmentation method for boar semen based on thresholding and watershed transform," *International Journal of Imaging*, vol. 2, no. S09, pp. 70–80, Spring 2009 2009. 21, 193

V. González-Castro, R. Alaiz-Rodríguez, L. Fernández-Robles, R. Guzmán-Martínez, and E. Alegre, "Estimating Class Proportions in Boar Semen Analysis Using the Hellinger Distance," in *Trends in Applied Intelligent Systems*, ser. Lecture Notes in Computer Science, vol. 6096. Springer, 2010, pp. 284–293. 27, 101, 106, 195

R. González Urdiales, F. Tejerina, J. Domínguez, B. Alegre, A. Ferreras, J. Peláez, S. Bernal, and S. Cárdenas, *Manual de técnicas de reproducción asistida en porcino*. Universitat de Girona. Servei de Publicacions, 2006, ch. Técnicas de análisis rutinario de la calidad espermática: motilidad, vitalidad, concentración, resistencia osmótica y morfología espermática, pp. 19–38.

C. E. Green and P. F. Watson, "Comparison of the capacitation-like state of cooled boar spermatozoa with true capacitation." *Reproduction*, vol. 122, no. 6, pp. 889–898, Dec 2001.

S. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on gabor filters," *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1160–1167, 2002. 14, 69

A. Guerrero-Curieses, R. Alaiz-Rodriguez, and J. Cid-Sueiro, "A fixed-point algorithm to minimax learning with neural networks," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 34, no. 4, pp. 383–392, November 2004. 27

A. Guerrero-Curieses, R. Alaiz-Rodríguez, and J. Cid-Sueiro, "Cost-sensitive and modular land-cover classification based on posterior probability estimates," *International Journal of Remote Sensing*, vol. 30, no. 22, pp. 5877–5899, 2009. 4, 25, 194

A. B. Hamza and H. Krim, "Geodesic matching of triangulated surfaces," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2249–2258, 2006. 23, 47, 82, 86, 194

R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, November 1973. 13, 39, 40, 196

R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 532–550, July 1987. 21, 43

M. Hernández, J. Roca, M. A. Gil, J. M. Vázquez, and E. A. Martínez, "Adjustments on the cryopreservation conditions reduce the incidence of boar ejaculates with poor sperm freezability," *Theriogenology*, vol. 67, no. 9, pp. 1436 – 1445, 2007. [Online]. Available: http://dx.doi.org/10.1016/j.theriogenology.2007.02.012 4

M. Hidalgo, I. Rodríguez, J. Dorado, and C. Soler, "Morphometric classification of spanish throughbred stallion sperm heads," *Animal Reproduction Science*, vol. 103, no. 3–4, pp. 374–378, January 2008. 31

C. Holt, W. V. Holt, H. D. M. Moore, H. C. B. Reed, and R. M. Curnock, "Objectively measured boar sperm motility parameters correlate with the outcomes of on-farm inseminations: Results of two fertility trials," *Journal of Andrology*, vol. 18, no. 3, pp. 312–323, May/June 1997. 31

Z. Hou and J. M. Parker, "Texture defect detection using support vector machines with adaptive gabor wavelet features," in *Proc. Seventh IEEE Workshops on Application of Computer Vision WACV/MOTIONS '05 Volume 1*, vol. 1, 5–7 Jan. 2005, pp. 275–280. 20, 193

M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, February 1962. 12, 74

Y. Huang, K. L. Chan, and Z. Huang, "An adaptive model for texture analysis," in *Proc. International Conference on Image Processing*, vol. 1, 10–13 Sept. 2000, pp. 276–279. 20

F. Huet and J. Mattioli, "A textural analysis by mathematical morphology transformations: structural opening and top-hat," in *Proc. International*

Conference on Image Processing, vol. 3, 16–19 Sept. 1996, pp. 49–52. 22, 81, 194

J.-A. Jiang, H.-Y. Chang, K.-H. Wu, C.-S. Ouyang, M.-M. Yang, E.-C. Yang, T.-W. Chen, and T.-T. Lin, "An adaptive image segmentation algorithm for x-ray quarantine inspection of selected fruits," *Computers and Electronics in Agriculture*, vol. 60, no. 2, pp. 190–200, March 2008. 18, 193

L. A. Johnson, K. F. Weitze, P. Fiser, and W. M. C. Maxwell, "Storage of boar semen," *Animal Reproduction Science*, vol. 62, no. 1-3, pp. 143 – 172, 2000. 4

J. Jost, *Riemannian Geometry and Geometric Analysis*, 5th ed. Springer-Verlag, 2008, iSBN: 978-3-540-77340-5. 47

A. Karahaliou, I. Boniatis, S. Skiadopoulos, F. Sakellaropoulos, N. Arikidis, E. Likaki, G. Panayiotakis, and L. Costaridou, "Breast cancer diagnosis: Analyzing texture of tissue surrounding microcalcifications," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 6, pp. 731–738, November 2008. 15

J.-S. Kim and K.-S. Hong, "Color-texture segmentation using unsupervised graph cuts," *Pattern Recognition*, vol. 42, no. 5, pp. 735 – 750, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2008.09.031 83

K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1542–1550, 2002. 88

K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1631–1639, 2003. 88

R. Kimmel and J. A. Sethian, "Computing geodesic paths on manifolds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 15, pp. 8431–8435, 1998. [Online]. Available: http://www.pnas.org/content/95/15/8431.abstract 50

M. Kokare, P. K. Biswas, and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 6, pp. 1168–1178, 2005. 83

S. Kullback and R. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951. 102

K. Laws, "Texture energy measures," in *Proceedings of the DARPA Image understanding workshop*, 1979, pp. 47–51. 12

J. H. Lee and N. I. Lee, "A fast and adaptive method to estimate texture statistics by the spatial gray level dependence matrix (sgldm) for texture image segmentation," *Pattern Recognition Letters*, vol. 13, no. 4, pp. 291–303, April 1992. 19

S.-S. Lee and H. T. Tanaka, "Parallel image segmentation with adaptive mesh," *Systems and Computers in Japan*, vol. 33, no. 10, pp. 95–104, 2002. 18, 193

W. Lei and Q. Feihu, "Adaptive fuzzy kohonen clustering network for image segmentation," in *Proc. International Joint Conference on Neural Networks IJCNN '99*, vol. 4, 10–16 July 1999, pp. 2664–2667. 19

X. Liang and J. Zhang, "White matter integrity analysis along cingulum paths in mild cognitive impairment - a geodesic distance approach," in *Proc. 2nd Int. Conf. Bioinformatics and Biomedical Engineering ICBBE 2008*, 2008, pp. 510–513. 23

S. Liao and M. Pawlak, "Image analysis with zernike moment descriptors," in *Electrical and Computer Engineering, 1997. IEEE 1997 Canadian Conference on*, vol. 2, May 1997, pp. 700 –703 vol.2. 75

T.-W. Lin and Y.-F. Chou, "A comparative study of Zernike moments," in *Proc. IEEE/WIC International Conference on Web Intelligence WI 2003*, 2003, pp. 516–519. 75

S. Livens, P. Scheunders, G. van de Wouwer, and D. Van Dyck, "Wavelets for texture analysis, an overview," in *Proc. Sixth International Conference on Image Processing and Its Applications*, vol. 2, 14–17 July 1997, pp. 581–585. 14

W. Liyun, L. Hefei, Z. Fuhao, L. Zhengding, and W. Zhendi, "Spermatogonium image recognition using zernike moments," *Computer Methods and Programs in Biomedicine*, vol. 95, no. 1, pp. 10 – 22, 2009. 11

B. L. Luck, K. D. Carlson, A. C. Bovik, and R. R. Richards-Kortum, "An image model and segmentation algorithm for reflectance confocal images of in vivo cervical tissue," vol. 14, no. 9, pp. 1265–1276, 2005. 12

D. Mahmoud-Ghoneim, M. K. Alkaabi, J. D. de Certaines, and F.-M. Goettsche, "The impact of image dynamic range on texture classification of brain white matter." *BMC Medical Imaging*, vol. 8, p. 18, 2008. 12, 13, 192

K. Mala and V. Sadasivam, "Automatic segmentation and classification of diffused liver diseases using wavelet based texture analysis and neural network," in *Proc. Annual IEEE INDICON*, 11–13 Dec. 2005, pp. 216–219. 15

R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: a level set approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158–175, 1995. 24

R. Mangoubi, M. Desai, N. Lowry, and P. Sammak, "Performance evaluation of multiresolution texture analysis of stem cell chromatin," in *Proc. 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI 2008*, 2008, pp. 380–383. 12, 193

P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 701–716, July 1989. 6, 21, 43, 46, 81, 85, 131, 141, 193, 196, 202, 203

P. Maragos and R. Schafer, "Morphological filters–part i: Their set-theoretic analysis and relations to linear shift-invariant filters," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 8, pp. 1153–1169, 1987. 21

——, "Morphological filters–part ii: Their relations to median, order-statistic, and stack filters," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 8, pp. 1170–1184, 1987. 21

J. Marti, R. Pérez-PE, M. Fernández, J. A. Cebrián, and T. Muiño Blanco, "Inducción de la reacción acrosómica en semen ovino fresco. marcaje mediante la lectina de ricinus communis (rca)." *Producción Ovina y Caprina*, vol. XXIII, pp. 535–539, 1998. 36

G. Matheron, *Random Sets and Integral Geometry*, J. W. . Sons, Ed. John Wiley & Sons, 1975. 21, 81, 193

F. Meyer and S. Beucher, "Morphological segmentation," *Journal of Visual Communication and Image Representation*, vol. 1, no. 1, pp. 21–46, September 1990. 6, 57, 197

F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, 1994. 55, 56, 57

D. A. Morales, E. Bengoetxea, and P. Larrañaga, "Selection of human embryos for transfer by bayesian classifiers." *Computers in Biology and Medicine*, vol. 38, no. 11-12, pp. 1177–1186, 2008. 12, 192

J. G. Moreno-Torres, X. Llorà, D. E. Goldberg, and R. Bhargava, "Repairing fractures between data using genetic programming-based feature extraction: A case study in cancer diagnosis," *Information Sciences*, vol. In Press, Corrected Proof, pp. –, 2010. 51, 77, 120, 125, 201

B. Nielsen, F. Albregtsen, and H. E. Danielsen, "Low dimensional adaptive texture feature vectors from class distance and class difference matrices," *IEEE Transactions on Medical Imaging*, vol. 23, no. 1, pp. 73–84, 2004. 20, 193

S. H. Ong, X. C. Jin, Jayasooriah, and R. Sinniah, "Image analysis of tissue sections." *Computers in Biology and Medicine*, vol. 26, no. 3, pp. 269–279, May 1996. 11

S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12 – 49, 1988. [Online]. Available: http://dx.doi.org/10.1016/0021-9991(88)90002-2 23

M. K. Osman, M. Y. Mashor, and H. Jaafar, "Detection of mycobacterium tuberculosis in ziehl-neelsen stained tissue images using zernike moments and hybrid multilayered perceptron network," in *Proc. IEEE Int Systems Man and Cybernetics (SMC) Conf*, 2010, pp. 4049–4055. 11

N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979. 6, 56, 196

G. Patrizi, C. Manna, C. Moscatelli, and L. Nieddu, "Pattern recognition methods in human-assisted reproduction," *International Transactions in Operational Research*, vol. 11, no. 4, pp. 365–379, 2004. 12

O. Pérez and M. Sánchez-Montañés, "A new learning strategy for classification problems with different training and test distributions," in *Computational and Ambient Intelligence*, ser. Lecture Notes in Computer Science, F. Sandoval, A. Prieto, J. Cabestany, and M. Graña, Eds. Springer Berlin / Heidelberg, 2007, vol. 4507, pp. 178–185. 26

M. Petrou and P. G. Sevilla, *Image processing: Dealing with texture.* John Wiley & Sons, Ltd, 2006. 21, 39, 46, 81, 85, 196, 203

G. Peyré and L. Cohen, "Surface segmentation using geodesic centroidal tesselation," in *Proc. 2nd Int. Symp. 3D Data Processing, Visualization and Transmission 3DPVT 2004*, 2004, pp. 995–1002. [Online]. Available: http://dx.doi.org/10.1109/TDPVT.2004.1335424 24

——, *Geodesic Methods for Shape and Surface Processing*, ser. Computational Methods in Applied Sciences. Springer, 2009, vol. 13, ch. 2, pp. 29–56. 24, 47, 194

C. Philips, D. Li, D. Raicu, and J. Furst, "Directional invariance of co-occurrence matrices within the liver," in *Proc. International Conference on Biocomputation, Bioinformatics, and Biomedical Technologies BIOTECHNO '08*, June 29 2008–July 5 2008, pp. 29–34. 14, 192

P. Phukpattaranont and P. Boonyaphiphat, "Segmentation of cancer cells in microscopic images using neural network and mathematical morphology," in *Proc. Int SICE-ICASE Joint Conf*, 2006, pp. 2312–2315. 21

P. Pina, L. Ribeiro, and F. Muge, "A mathematical morphology contribution to study some aspects of hydrogeological systems," *Computers & Geosciences*, vol. 27, no. 9, pp. 1061 – 1069, 2001. 21

G.-C. Pok, J.-C. Liu, and K. H. Ryu, "New shape-based texture descriptors for rotation invariant texture classification," in *Proc. International Conference on Image Processing ICIP 2003*, vol. 3, 14–17 Sept. 2003, pp. III–533–6. 12, 193

S. Poonguzhali and G. Ravindran, "Performance evaluation of feature extraction methods for classifying abnormalities in ultrasound liver images using neural network," in *Proc. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS '06*, 2006, pp. 4791–4794. 12, 193

F. J. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," in *Proceedings of the* 15^{th} *International Conference on MachineLearning*, 1998, pp. 445–453. 75, 91, 200

D. Qi and L. Yu, "Application of omnidirectional structure element of mathematical morphology to wood computed tomography testing," in *Proc. Chinese Control and Decision Conf. CCDC 2008*, 2008, pp. 3702–3707. 21, 193

D. S. Raicu, J. D. Furst, D. Channin, X. Dong-Hui, A. Kurani, and S. Aioanei, "A texture dictionary for human organs tissues' classification," in *Proceedings of the 8th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2004)*, 2004, pp. 18–21. 13

L. Ramos, J. C. M. Hendriks, P. Peelen, D. D. M. Braat, and A. M. M. Wetzels, "Use of computerized karyometric image analysis for evaluation of human spermatozoa." *Journal of Andrology*, vol. 23, no. 6, pp. 882–888, 2002. 31

K. Rodenacker and E. Bengtsson, "A feature set for cytometry on digitized microscopic images." *Analytical Cellular Pathology*, vol. 25, no. 1, pp. 1–36, 2003. 11

H. Rodríguez-Martínez, "Laboratory semen assessment and prediction of fertility: still utopia?" *Reprod. Domest. Anim.*, vol. 38, no. 4, pp. 312–318, Aug 2003. 31

D. Rohrmus, "Invariant and adaptive geometrical texture features for defect detection and classification," *Pattern Recognition*, vol. 38, no. 10, pp. 1546–1559, October 2005. 20

K. J. Rozeboom, "Evaluating boar semen quality," *Animal Science Facts, Extension Swine Husbandry*, vol. ANS 00-812S, pp. 1–7, 2000. [Online]. Available: http://mark.asci.ncsu.edu/Publications/factsheets/812s.htm 31

A. Ruggeri and S. Pajaro, "Automatic recognition of cell layers in corneal confocal microscopy images." *Computer Methods Programs in Biomedicine*, vol. 68, no. 1, pp. 25–35, Apr 2002. 11, 75

M. Saerens, P. Latinne, and C. Decaestecker., "Adjusting a classifier for new a priori probabilities: A simple procedure," *Neural Computation*, vol. 14, pp. 21–41, January 2002. 4, 25, 26, 99, 100, 121, 194, 208, 209

E. H. Said, D. E. M. Nassar, G. Fahmy, and H. H. Ammar, "Teeth segmentation in digitized dental x-ray films using mathematical morphology," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 178–189, 2006. 21

L. Sánchez, N. Petkov, and E. Alegre, "Statistical approach to boar semen head classification based on intracellular intensity distribution," in *Computer Analysis of Images and Patterns (CAIP'05)*, ser. Lecture Notes in Computer Science, A. Gagalowicz and W. Philips, Eds., vol. 3691. Springer-Verlag Berlin Heidelberg, 2005, pp. 88–95. 17, 193

——, "Classification of boar spermatozoid head images using a model intracellular density distribution," in *10th Iberoamerican Congress on Pattern Recognition, CIARP 2005*, ser. Lecture Notes in Computer Science, M. Lazo and A. Sanfeliu, Eds., vol. 3773. Springer-Verlag Berlin Heidelberg, 2005, pp. 154–160. 17

——, "Statistical approach to boar semen evaluation using intracellular intensity distribution of head images." *Cellular and Molecular Biology*, vol. 52, no. 6, pp. 38–43, 2006. 17, 193

L. Sánchez, V. González, E. Alegre, and R. Alaiz, "Classification and quantification based on image analysis for sperm samples with uncertain damaged/intact cell proportions," in *Proceedings of the 5th International Conference on Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, vol. 5112, July 2008, pp. 827–836. 27, 195

J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, February 2000. 19

L. Semler and L. Dettori, "Curvelet-based texture classification of tissues in computed tomography," in *Proc. IEEE International Conference on Image Processing*, 2006, pp. 2165–2168. 12, 16, 69, 193

L. Semler, L. Dettori, and J. Furst, "Wavelet-based texture classification of tissues in computed tomography," in *Proc. 18th IEEE Symposium on Computer-Based Medical Systems*, 2005, pp. 265–270. 12, 15

J. Serra, *Image Analysis and Math. Morphology*, A. Press, Ed. Academic Press, 1982. 21, 81, 193

J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 4, pp. 1591–1595, 1996. [Online]. Available: http://www.pnas.org/content/93/4/1591.abstract 24, 48, 87, 194, 196, 204

——, "Fast marching methods," *SIAM Review*, vol. 41, no. 2, pp. 199–235, June 1999. [Online]. Available: http://www.jstor.org/pss/2653069 24

H. S. Sheshadri and A. Kandaswamy, "Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms." *Comput Med Imaging Graph*, vol. 31, no. 1, pp. 46–48, January 2007. 12, 192

F. Y. Shih and S. Cheng, "Adaptive mathematical morphology for edge linking," *Information Sciences*, vol. 167, no. 1-4, pp. 9 – 21, 2004. 22, 194

H. Shu, L. Luo, X. Bao, W. Yu, and G. Han, "An efficient method for computation of legendre moments," *Graphical Models*, vol. 62, no. 4, pp. 237 – 262, 2000. [Online]. Available: http://dx.doi.org/10.1006/gmod.2000.052375

P. Silva and B. Gadella, "Detection of damage in mammalian sperm cells," *Theriogenology*, vol. 65, no. 5, pp. 958–978, 2006, proceedings of IETS 2005 Satellite Symposium: Agricultural and societal implications of contemporary embryo-technologies in farm animals. 32, 33, 195

R. Sivaramakrishna, K. A. Powell, M. L. Lieber, W. A. Chilcote, and R. Shekhar, "Texture analysis of lesions in breast ultrasound images." *Computerized Medical Imaging and Graphics*, vol. 26, no. 5, pp. 303–307, 2002. 13

P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2042–2055, 2002. 21

S. R. Sternberg, "Grayscale morphology," *Computer Vision, Graphics, and Image Processing*, vol. 35, no. 3, pp. 333–355, September 1986. 21, 43, 81, 193

G. Stippel, W. Philips, and P. Govaert, "A tissue-specific adaptive texture filter for medical ultrasound images." *Ultrasound Med Biol*, vol. 31, no. 9, pp. 1211–1223, September 2005. 18, 193

T. Subashini, V. Ramalingam, and S. Palanivel, "Automated assessment of breast tissue density in digital mammograms," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 33 – 43, 2010. 13

M. Suliga, R. Deklerck, and E. Nyssen, "Markov random field-based clustering applied to the segmentation of masses in digital mammograms," *Computerized Medical Imaging and Graphics*, vol. 32, no. 6, pp. 502 – 512, 2008. 12, 192

A. Takemura and M. Ito, "Segmentation of ultrasonic images by using locally adaptive filter and wavelet analysis: Detection of superficial peripheral vein by a high-frequency ultrasonic equipment," *Electronics and Communications*

in Japan (Part III: Fundamental Electronic Science), vol. 86, no. 1, pp. 36–45, 2003. 18

N. Theera-Umpon and S. Dhompongsa, "Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 3, pp. 353–359, 2007. 21

W. Tsang, A. Corboy, K. Lee, D. Raicu, and J. Furst, "Texture-based image retrieval for computerized tomography databases," in *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems*, 2005, pp. 593–598. 12, 13

S. Tsantis, N. Dimitropoulos, D. Cavouras, and G. Nikiforidis, "Morphological and wavelet features towards sonographic thyroid nodules evaluation." *Comput Med Imaging Graph*, vol. 33, no. 2, pp. 91–99, March 2009. 15, 69

E. C.-K. Tsao, J. C.Bezdek, and N. R. Pal, "Fuzzy kohonen clustering networks," *Pattern Recognition*, vol. 27, no. 5, pp. 757–764, May 1994. 19

S. Valero, J. Chanussot, J. Benediktsson, H. Talbot, and B. Waske, "Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1120 – 1127, 2010. 21, 193

J. S. Valverde and R. R. Grigat, "Optimum binarization of technical document images," in *Proc. International Conference on Image Processing*, vol. 3, 10–13 Sept. 2000, pp. 985–988. 19

P. van der Putten and M. van Someren, "Coil challenge 2000: The insurance company case," Sentient Machine Research, Amsterdam. Also a Leiden Institute of Advanced Computer Science, Tech. Rep. 2000-09, June 2000. 177

J. Verstegen, M. Iguer-Ouada, and K. Onclin, "Computer assisted semen analyzers in andrology research and veterinary practice." *Theriogenology*, vol. 57, no. 1, pp. 149–179, January 2002.

M. Vision and M. G. V. Texture. http://vismod.media.mit.edu/vismod/imagery/visiontexture/. Available. [Online]. Available: http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html 16, 82

S. Vucetic and Z. Obradovic, "Classification on data with biased class distribution," in *Proceedings of the 12th European Conference on Machine Learning (ECML, Freiburg)*, 2001, pp. 527–538. 27

R. F. Walker, P. T. Jackway, and I. D. Longstaff, "Recent developments in the use of the co-occurrence matrix for texture recognition," in *Proc. DSP 97. 13th International Conference on Digital Signal Processing*, vol. 1, 2–4 July 1997, pp. 63–65. 20

R. F. Walker, P. T. Jackway, and D. Longstaff, "Genetic algorithm optimization of adaptive multi-scale glcm features," *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, vol. 17, no. 1, pp. 17–39, February 2003. 20, 193

R. F. Walker, "Adaptive multi-scale texture analysis with application to automated cytology," Ph.D. dissertation, University of Queensland, Brisbane, Australia, July 1997. [Online]. Available: http://www.koitsu.com/texture/thesis.pdf 20

P. M. Wassarman, L. Jovine, and E. S. Litscher, "A profile of fertilization in mammals," *Nature cell biology*, vol. 3, no. 2, pp. E59 – E64, February 2001. [Online]. Available: http://dx.doi.org/10.1038/35055178 32

P. F. Watson, "Recent developments and concepts in the cryopreservation of spermatozoa and the assessment of their post-thawing function." *Reprod Fertil Dev*, vol. 7, no. 4, pp. 871–891, 1995. 4

F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945. 51, 111, 115, 196, 213

G. V. D. Wouwer, B. Weyn, P. Scheunders, W. Jacob, E. V. Marck, and D. V. Dyck, "Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei." *Jorunal of Microscroscopy*, vol. 197, no. Pt 1, pp. 25–35, January 2000. 15

X. Wu, Q. Xi, Y. W. Chen, and S. S. Zhang, "Orientation adaptive fast marching method for contour tracking of small intestine," *Electronics Letters*, vol. 45, no. 23, pp. 1154–1155, 2009. 24

D. Xing, W. Dai, G.-R. Xue, and Y. Yu, "Bridged refinement for transfer learning," in *Knowledge Discovery in Databases: PKDD 2007*, ser. Lecture Notes in Computer Science, J. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenic, and A. Skowron, Eds. Springer Berlin / Heidelberg, 2007, vol. 4702, pp. 324–335. 26

J. C. Xue and G. M. Weiss, "Quantification and semi-supervised classification methods for handling changes in class distribution," in *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2009, pp. 897–906. 4, 26

R. Yanagimachi, *Mammalian fertilization*, 2nd ed. Raven Press, 1994, vol. 1, pp. 189–317. 32, 131, 141

C. Yang and J. Zhou, "Non-stationary data sequence classification using online class priors estimation." *Pattern Recognition*, vol. 41, no. 8, p. 8, August 2008. 26, 194

J. Yang and X. Li, "Boundary detection using mathematical morphology," *Pattern Recognition Letters*, vol. 16, no. 12, pp. 1277 – 1286, 1995. 21

L. Yu and R. Wang, "Shape representation based on mathematical morphology," *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1354 – 1362, 2005. 21

Z. Yu and C. Bajaj, "A fast and adaptive method for image contrast enhancement," in *Proc. International Conference on Image Processing ICIP '04*, vol. 2, 24–27 Oct. 2004, pp. 1001–1004. 18, 193

J. H. Zar, *Biostatistical Analysis (5th Edition)*. Prentice-Hall, Inc., 2007. 52

Z. Zhang and J. Zhou, "Transfer estimation of evolving class priors in data stream classification," *Pattern Recogn.*, vol. 43, no. 9, pp. 3151–3161, 2010. 26, 194

Part I

APPENDICES

APPENDIX A

DATASET DESCRIPTION

The performance of the quantification methods has been assessed on several different public datasets. All these databases are maintained by the UCI Machine Learning repository (Frank and Asuncion, 2010), except for the Phoneme dataset, which has been taken from the ELENA project[1].

1. The Wisconsin Breast Cancer dataset contains 699 samples divided in two classes (241 possitive and 458 negative). Each sample is defined by 9 features. The data was collected at the University of Wisconsin Hospitals, with the goal of distinguishing between benign (class 0) and malignant (class 1) breast tumours. Originally it had 16 missing values, which were filled by the average value of that feature of all the other elements.

2. The Contraceptive Method Choice (CMC) dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. The goal of this dataset is to predict the current contraceptive method choice of a woman based on her demographic and socio-economic characteristics. The original dataset has three classes: no use, long term or short term contraceptive method, and it has 1473 samples with 9 features.

3. The Coil dataset (van der Putten and van Someren, 2000) has 9822 samples and contains information about customers of an insurance company. Each sample is represented by 85 variables, including product usage and socio-demographic data. It has been supplied by the Dutch data mining company Sentient Machine Research. Its goal is to predict if a customer has a caravan insurance policy (class 1) or not (class 0). The number of class-1 instances is 586.

4. The Diabetes dataset has 768 instances with 8 features each. It has been provided by the National Institute of Diabetes and Digestive and Kidney Diseases, and it gives information about women of Pima Indian heritage. Its aim is to determine whether they have diabetes (class 1) or not (class 0). It has 268 elements in the positive class.

5. The German Credits dataset that has been used in this paper is a numerical version of the Statlog German Credit Data dataset produced by the Strathclyde University. The goal is to classify customers as good (class

[1] `http://www.dice.ucl.ac.be/neural-nets/Research/Projects/ELENA/elena.htm`

0) or bad (class 1) credit risks depending on 20 features about them and their bank accounts. The positive class has 300 instances.

6. The Letter Recognition dataset aims to identify digital images in black and white as one of the 26 capital letters in the English alphabet. The character images were based on 20 different fonts and each letter within these fonts was randomly distorted to produce 20,000 unique stimuli. Each stimulus was converted into 16 numerical attributes, which were scaled to fit into a range of integer values from 0 to 15.

7. The Mammographic Mass dataset (Elter et al., 2007) is used to predict the severity of a mammographic mass lesion (malignant or benign) by means of 5 features – such as BI-RADS assessment, or the patient's age –. This dataset was collected at the Institute or Radiology of the University Erlangen-Nuremberg between 2003 and 2006. It has two classes and 961 elements. We have removed the elements that had missing values, so the remaining dataset contains 830 instances: 427 benign and 403 malignant.

8. The Page Blocks Classification dataset is used to classify all the blocks of the page layout that has been detected by a segmentation process. The 5473 examples come from 54 different documents, and each observation concerns one block, which is described by means of 10 attributes. The type of blocks to be detected are: Text, Horizontal line, Graphic, Vertical line or Picture.

9. The Phoneme dataset is aimed to distinguish between nasal (class 0) and oral (class 1) sounds. It has 3818 samples in class 0 and 1586 in class 1, and each instance has 5 features.

10. The Semeion Handwritten Digit dataset was created by Tactile Srl and donated in 1994 to Semeion Research Center of Sciences of Communication. This dataset has 1593 handwritten digits (from zero to nine) from around 80 people which where scanned in a grey scale of 256 values, stretched in a rectangular box of 16×16 pixels and binarised with a threshold equals to 127. Each feature is the value of a pixel of the image.

11. The Spambase dataset contains information about 4597 e-mail messages. The task is to determine whether a given email is spam or not, depending

on its contents. This dataset was created in the HP labs by extracting 57 features from real emails.

12. The Red Wine Quality dataset (Cortez et al., 2009) is related to samples of red variant of the portuguese "Vinho Verde" wine. It has 1599 instances, and the data was acquired by researchers from the University of Minho and the Viticulture Commision of the Vinho Verde Region, in the north of Portugal. The goal is to model wine quality (which is the target class, in the interval $[0, 10]$) with the basis of 11 physiochemical features.

13. The White Wine Quality dataset has been acquired by the same researchers as the previous one which is related to the white variant of the Vinho Verde wine. It has are 4898 instances, and the same number of features and classes.

14. The Yeast dataset contains information about a set of 1484 Yeast cells described by means of 8 features. Its aim is to determine the localization site of the proteins on each cell. In the original dataset there were 10 target classes.

APPENDIX B

DERIVED PUBLICATIONS

Book chapters

Víctor González-Castro, Rocío Alaiz-Rodríguez and Enrique Alegre. *Perspectives on Pattern Recognition.* ISBN: 978-1-61209-118-1. Nova Publishers, 2011. ch. Class Distribution Estimation in Imprecise Domains Based on Supervised Learning. *To be published.*

Journal articles

V. González-Castro, R. Alaiz-Rodríguez and E. Alegre, "Class Distribution Estimation Based on the Hellinger Distance". *Submitted for publication.*

E. Alegre, V. González-Castro, R. Alaiz-Rodríguez and M. González, "Texture and Shape Based Classification of the Acrosome Integrity of Boar Spermatozoa Images", *Computer Methods and Programs in Biomedine*, 2011. *Submitted for publication.*

V. González-Castro, E. Alegre, P. Morala-Argüello and S.A. Suárez, "A Combined and Intelligent new Segmentation Method for Boar Semen Based on Thresholding and Watershed Transform". *International Journal of Imaging*, no S09, pp 70–80, Spring 2009.

International conferences

E. Alegre, M. T. García-Ordás, V. González-Castro and S. Karthikeyan, "Vitality assessment of boar sperm using NCSR texture descriptor in digital images", due to appear in *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2011)*, ser. Lecture Notes in Computer Science, vol. 6636, La Palma de Gran Canaria, Spain, 2011.

E. Alegre, O. García-Olalla, V. González-Castro and Swapna Joshi. "Boar spermatozoa classification using Longitudinal and Transversal Profiles (LTP) descriptor in digital images", due to appear in *14th International Workshop on Combinatorial Image Analysis*, ser. Lecture Notes in Computer Science, vol. 6669, Madrid, Spain, 2011.

V. González-Castro, E. Alegre, S. A. Suárez, O. García-Olalla and M. T. García, "Adaptive texture description for semen vitality assessment", in *Proceedings of the 6th International Seminar on Medical Image Processing and Analysis SIPAIM 2010*, 2010.

V. González-Castro, R. Alaiz-Rodríguez, L. Fernández-Robles, R. Guzmán-Martínez and E. Alegre, "Estimating Class Proportions in Boar Semen Analysis Using the Hellinger Distance", in *Trends in Applied Intelligent Systems*, ser. Lecture Notes in Computer Science, vol 6096 Springer, 2010, pp. 284-293

E. Alegre, V. González-Castro, S. A. Suárez and M. Castejón, "Comparison of supervised and unsupervised methods to classify boar acrosomes using texture descriptors", in *Proceedings ELMAR-2009*, 2009, pp. 65-70

R. Alaiz-Rodríguez, E. Alegre, V. González-Castro and L. Sánchez, "Quantifying the proportion of damaged sperm cells based on image analysis and neural networks", in *Proceedings of the 8th conference on Simulation, modelling and optimization*, World Scientific and Engineering Academy and Society, 2008, pp. 383-388

L. Sánchez, V. González, E. Alegre, R. Alaiz, "Classification and Quantification Based on Image Analysis for Sperm Samples with Uncertain Damaged/Intact Cell Proportions", in *Image Analysis and Recognition. 5th International Conference, Proceedings*, ser. Lecture Notes in Computer Science, vol. 5112, 2008, pp. 827-836

National conferences

V. González, E. Alegre, P. Morala-Argüello and S. A. Suárez Castrillón. "Segmentación de cabezas de espermatozoides de verraco mediante combinación de umbralización y transformada Watershed". *Actas de las XXIX Jornadas de automática 2008*, 2008.

E. Alegre, M. González, V. González-Castro and T. Alonso. "Evaluación de funciones Wavelet madre con descriptores de textura estadísticos en la clasi-

ficación del acrosoma de espermatozoides de verraco". *Actas de las XXXI Jornadas de automática 2010*, 2010.

Part II

RESUMEN DE LA TESIS EN CASTELLANO

En cumplimiento del punto 7º de la normativa complementaria del Real Decreto 778/1998, de 30 de Abril y de las normas para la aplicación del mismo, aprobadas por acuerdo de la Junta de Gobierno de fecha 10 de mayo de 1999, se adjunta un resumen en castellano de cada uno de los capítulos de esta tesis doctoral para que pueda admitirse a trámite.

1 Introducción

La evaluación de la calidad seminal es una etapa clave en la Inseminación Artificial (IA), tanto humana como animal. En este último caso, la IA permite que las granjas pueden trabajar con un número reducido de animales seleccionados. Esto supone por un lado, un ahorro en los costes de producción, y por otro, mejorar la calidad del producto, al obtener mejores individuos en cada generación. Por ello, los centros de producción que venden las muestras de semen deben realizar controles rigurosos para asegurar que dichas muestras cumplen unos altos estándares de calidad.

Existe relación entre el potencial fertilizador del esperma y la vitalidad espermática e integridad acrosómica: si una muestra presenta una alta proporción de espermatozoides con acrosomas dañados, o de muertos, la muestra tendrá un potencial fertilizador reducido. La evaluación de estos parámetros se realiza de manera visual, utilizando tinciones y microscopios de fluorescencia. Este proceso, sin embargo, es costoso en cuanto a tiempo y dinero – los microscopios de fluorescencia son equipos muy caros –, y adolece de falta de objetividad, pues depende de la pericia del observador. Por ello, sería muy útil poder detectar acrosomas dañados y espermatozoides muertos sin utilizar tinciones. El procesamiento digital de imágenes permitiría realizar esta tarea utilizando únicamente imágenes en escala de grises, por lo que no serían necesarias las tinciones ni el microscopio de fluorescencia. Con este objetivo, los métodos propuestos en esta tesis se han evaluado en esta aplicación concreta.

El análisis de texturas es una técnica muy potente, y muy utilizada en aplicaciones biomédicas. Los métodos convencionales de descripción de texturas se aplican tal cual en toda la región de interés, resultando imposible la captura de sus variaciones. Este problema se podría superar utilizando un descriptor adaptativo, que tuviera en cuenta las características locales sin poseer ningún

conocimiento a priori sobre la región que se va a analizar, y ese ha sido uno de nuestros objetivos en este trabajo.

En aprendizaje supervisado se asume que las probabilidades a priori de las clases de los datos de los nuevos datos (test) siguen la misma distribución que los datos con los que se entrenó el clasificador. Sin embargo, existen muchos casos prácticos en los que esto no se puede asumir. Cuando ocurren este tipo de cambios, el clasificador proporciona soluciones subóptimas, y es deseable adaptarlo al nuevo contexto para, de esta manera, mejorar su rendimiento en la clasificación. La estimación de la nueva proporción de las clases (proceso conocido como cuantificación) puede emplearse para esta adaptación. Existen dominios, sin embargo, en que esta estimación es interesante por sí misma, ya que ese es su objetivo. En particular, en aplicaciones de control de la calidad seminal las distribuciones de clases no son estacionarias, debido a cambios en las condiciones de conservación y manipulación (*e.g.* la congelación del semen incrementa el número de células dañadas y muertas). Además, el interés real es evaluar la proporción en una muestra de espermatozoides muertos, o de acrosomas dañados, sin importar la clasificación particular de cada célula.

Teniendo en cuenta esta problemática, las dos líneas de investigación seguidas en esta Tesis doctoral – descripción de texturas y estimación de probabilidades a priori de clases en un conjunto de datos – se han combinado con el propósito principal de crear un sistema que permita evaluar automáticamente la proporción de acrosomas dañados, o espermatozoides muertos, utilizando únicamente información sobre la textura en imágenes en escala de grises adquiridas mediante un microscopio de contraste de fases. Como paso previo, es necesario un método que permita segmentar tantas cabezas como sea posible, y que garantice que las cabezas devueltas están bien segmentadas. También hay que desarrollar descriptores que capturen la información necesaria para caracterizar células espermáticas. Por último, es necesario implementar métodos para estimar las distribuciones de clases de conjuntos de datos no etiquetados en dominios en que las condiciones operacionales son imprecisas.

Teniendo en cuenta estos objetivos y el trabajo realizado, las principales contribuciones son:

1. Un procedimiento inteligente para segmentar los espermatozoides que combina un método basado en umbralización y otro que utiliza la trans-

formada Watershed. Esta contribución permite segmentar más imágenes que el primero, y reduce considerablemente el coste computacional del segundo.

2. La transformada Curvelet discreta, en combinación con descriptores estadísticos de segundo orden, se ha utilizado para caracterizar acrosomas íntegros y dañados con mejor rendimiento que otros descriptores de textura y de forma.

3. Un método adaptativo de descripción de texturas basado en *adaptar* los elementos estructurantes usados para el cálculo del *Pattern Spectrum* a las características locales de la textura en función de distancias geodésicas. Este método se ha llamado Pattern Spectrum Adaptativo Geodésico (AGPS en sus siglas en inglés).

4. Un método de cuantificación que estima las probabilidades a priori de un conjunto a partir de las probabilidades a posteriori (PP) devueltas por un clasificador.

5. Dos métodos de cuantificación basados en medir divergencias entre distribuciones utilizando la distancia de Hellinger. El primero utiliza los datos originales (HDx), mientras que el segundo usa las salidas de un clasificador (HDy). La evaluación realizada en diferentes dominios muestra que los métodos de cuantificación presentan mejor rendimiento que el procedimiento simple de contar las predicciones del clasificador, siendo HDy el mejor de todos ellos.

6. Se han creado dos bases de datos con imágenes de acrosomas íntegros y dañados, y espermatozoides vivos/muertos. Asimismo se han construido conjuntos de datos con información de la textura de los primeros.

En las secciones 2 y 3 se expone un breve resumen de la revisión del estado de la técnica y de los métodos utilizados en el desarrollo de las propuestas de esta Tesis, respectivamente. A continuación, en la sección 4 se muestra un resumen del método inteligente de segmentación desarrollado para la evaluación de la calidad seminal. Los resultados experimentales de la aplicación de la transformada Curvelet a la descripción de la textura de acrosomas íntegros y dañados se muestra en la sección 5. Posteriormente, el desarrollo del descriptor

AGPS y resultados experimentales se resumen en la sección 6. Finalmente, en las secciones 7 y 8 se explican y comparan los métodos de cuantificación propuestos con otros ya existentes.

2 Revisión del Estado de la Técnica

2.1 Procesamiento digital de imágenes para la caracterización de células y tejidos

El procesamiento digital de imágenes es muy común en aplicaciones de microscopía (Bonnet, 2004) y biomedicina ya que es muy útil en el análisis automático de células y tejidos.

Muchas de estas aplicaciones se basan en el análisis de texturas. Sin embargo, a pesar de su importancia, no existe una definición formal de textura. De hecho, el concepto de textura cambia según el método que se use para analizarla (Gonzalez and Woods, 2002). Castellano *et al*. realizaron una revisión y una taxonomía de las técnicas de análisis de textura utilizadas con imágenes médicas (Castellano et al., 2004):

- Los **Métodos estadísticos** consisten en caracterizar las texturas por medio de propiedades que rigen la distribución y relaciones de los niveles de gris dentro de la textura. Estas pueden ser estadísticos de primer orden (Sheshadri and Kandaswamy, 2007), de segundo orden extraídos de la matriz de coocurrencia (Mahmoud-Ghoneim et al., 2008; Philips et al., 2008) o de la matriz *run-length* (Chandraratne et al., 2006), o momentos estadísticos (Morales et al., 2008).

- Los **Métodos basados en modelos** tratan de predecir los valores de los píxeles basándose en un modelo probabilístico previamente asignado a la textura. Un ejemplo de este tipo de descriptores son los modelos aleatorios de Markov (Suliga et al., 2008).

- Los **Métodos estructurales** tratan de encontrar modelos jerárquicos en las texturas. Sin embargo, aunque se pueden utilizar para describir texturas son más útiles para sintetizarlas.

- Los **Métodos basados en procesamiento de señales** realizan modificaciones sobre las imágenes, bien sea utilizando filtros como las máscaras de Laws (Poonguzhali and Ravindran, 2006) o los filtros de Gabor (Pok et al., 2003), o por medio de transformadas, como la de Fourier (Arof and Deravi, 1997), Wavelet (Mangoubi et al., 2008), Curvelet (Semler and Dettori, 2006), *etc*.

Además, existen algunos trabajos en los que se aborda la caracterización de células espermáticas por medio del análisis de texturas, ya sea para clasificarlas en función de su vitalidad (Sánchez et al., 2005a, 2006), o de su integridad acrosómica (Alegre et al., 2009, 2008).

2.2 Métodos adaptativos en el procesamiento digital de imágenes

Los métodos de análisis adaptativo de texturas permiten el estudio de la textura en función de sus características locales. En otras palabras, el método se *adapta* localmente a la textura.

Existen algunos ejemplos de métodos de preprocesamiento (Bastian et al., 1995; Stippel et al., 2005; Yu and Bajaj, 2004), segmentación (Gatos et al., 2006; Jiang et al., 2008; Lee and Tanaka, 2002), e incluso de descripción (Hou and Parker, 2005; Nielsen et al., 2004; Walker et al., 2003), si bien estos requieren algún tipo de conocimiento a priori sobre las texturas que se van a analizar.

2.3 Morfología matemática

Las bases de la morfología matemática fueron establecidas por Matheron (Matheron, 1975) y Serra (Serra, 1982) centrándose en imágenes binarias. Esta teoría fue generalizada posteriormente para imágenes en escala de grises por Sternberg (Sternberg, 1986). La morfología matemática se ha utilizado en numerosas aplicaciones y con diversos propósitos (Bouraoui et al., 2010; González-Castro et al., 2009; Qi and Yu, 2008; Valero et al., 2010).

Matheron propuso un descriptor de tamaño-forma, válido para describir texturas, llamado *Pattern Spectrum* (Maragos, 1989). Este descriptor tiene el serio inconveniente de que no está claro cuál es el mejor elemento estructurante para cada problema, ni cómo elegirlo (de Ves et al., 2006). Existen trabajos

con propuestas al respecto (Asano et al., 2003; Huet and Mattioli, 1996), aunque requieren conocimiento a priori de la textura que se va a describir. Otros trabajos tratan de hacer que el elemento estructurante varíe en función de características locales de las imágenes (Shih and Cheng, 2004), o de criterios de distancia Euclídea (Cuisenaire, 2006).

Sin embargo, como Hamza y Krim puntualizaron, utilizando la distancia Euclídea no es posible captar la estructura intrínseca de una superficie, mientras que la distancia Geodésica sí lo permite (Hamza and Krim, 2006). La textura de una imagen se puede considerar como una superficie, por lo que esta distancia es más adecuada. Sethian desarrolló un algoritmo, llamado *Fast Marching* (Sethian, 1996), que se puede utilizar para calcular caminos de distancia geodésica mínima (Cohen and Kimmel, 1996). Posteriormente, Peyré y Cohen revisaron los cálculos numéricos del algoritmo, y presentaron algunas de sus aplicaciones, entre las cuales destacaba cómo calcular mapas de distancias geodésicas (Peyré and Cohen, 2009).

2.4 El problema de los cambios en las probabilidades a priori de las clases

Como ya se indicó en la sección 1, los problemas de clasificación supervisada se centran en entrenar clasificadores con conjuntos de datos etiquetados posibles, para luego aplicarlos tal cual a nuevos patrones, con el objetivo de predecir la clase a la que pertenecen, asumiendo que las distribuciones de clases permanecen inalterables (Duda et al., 2001).

Sin embargo, esta suposición no siempre se cumple, y cuando las distribuciones de las clases de los conjuntos de test son diferentes respecto al conjunto con el que se generó el clasificador, su precisión se ve afectada. Algunos ejemplos de aplicaciones en las que esto ocurre son sistemas de eliminación de ambigüedades en palabras (Chan and Ng, 2006), o aplicaciones de teledetección (Guerrero-Curieses et al., 2009). En estos casos, es importante adaptar el clasificador al nuevo contexto (Saerens et al., 2002; Yang and Zhou, 2008; Zhang and Zhou, 2010). La detección de fallos en clasificadores debidos a cambios en la distribución de los datos ha sido estudiada recientemente por la comunidad de aprendizaje automático. La distancia de Hellinger se ha mostrado especialmente útil en este aspecto (Cieslak and Chawla, 2009).

Es necesario puntualizar que existen aplicaciones donde las distribuciones de las clases varían a lo largo del tiempo, y su estimación es interesante por sí misma (Forman, 2008). Algunos autores han estudiado este problema, y han propuesto algunos métodos de cuantificación (Alaiz-Rodríguez et al., 2008; González-Castro et al., 2010; Sánchez et al., 2008).

3 Metodología

La evaluación de la calidad seminal es una tarea crucial en Inseminación Artificial, para asegurar que una muestra de semen tiene altas posibilidades de fecundar un óvulo.

Existen cuatro parámetros que permiten evaluar la calidad seminal en verracos: concentración, motilidad, morfología e integridad del acrosoma. Centrándonos en este último, el acrosoma es una estructura que recubre el extremo apical de la cabeza del espermatozoide que contiene una serie de encimas que se liberan (en un proceso denominado capacitación) cuando el espermatozoide entra en contacto con el óvulo, permitiendo su penetración en él. Por ello, cuando un eyaculado presenta un alto porcentaje de espermatozoides capacitados antes de su introducción en el tracto genital de la hembra, o en una fase temprana de su tránsito por él, dicho eyaculado será inútil para su fertilización (Silva and Gadella, 2006). Por otro lado, la membrana plasmática separa al espermatozoide del medio externo, y desarrolla muchas funciones fisiológicas que lo mantienen en condiciones óptimas y preservan su vitalidad. Si dicha membrana no está funcionalmente intacta, el espermatozoide no podrá mantener sus concentraciones intracelulares ni producir la energía necesaria para su movimiento, lo que ocasionaría la muerte celular, dejándolo incapacitado para la fecundación in vivo.

Gracias a la preparación de las muestras con ioduro de propidio (PI) y diacetato de carboxifluoresceína (CFDA), o con la lectina FICT-PNA es posible distinguir, bajo la longitud de onda adecuada en un microscopio de fluorescencia, espermatozoides vivos/muertos, o con acrosomas dañados, respectivamente. Así, utilizando una cámara digital conectada a un microscopio de epifluorescencia, se fueron tomando las imágenes de las muestras de dos en dos: una en contraste de fases (escala de grises), y otra bajo una iluminación fluorescente (en color real). Las primeras se utilizarán en los experimentos para evaluar los

métodos propuestos, mientras que las segundas se usarán únicamente con el propósito de etiquetar las anteriores, y así poder comprobar los resultados de los métodos evaluados.

En la caracterización de las imágenes en escala de grises se utilizarán algunos de los descriptores estadísticos extraídos de la matriz de coocurrencia propuestos por Haralick *et al.* (Haralick et al., 1973) y la transformada Curvelet discreta, calculada por medio del algoritmo de *wrapping* descrito en (Candès et al., 2006). Las texturas también se describirán por medio del *Pattern Spectrum* (Maragos, 1989), calculado mediante el método propuesto en (Petrou and Sevilla, 2006), adaptándolo para que se ajuste a la textura lo mejor posible, según una transformada de distancia geodésica, calculada mediante el algoritmo *Fast Marching* (Sethian, 1996).

Por último, se realizarán tests no paramétricos que se usarán para detectar si las diferencias entre los resultados de los experimentos con los diferentes métodos propuestos son estadísticamente significativas, concretamente el test de los signos de Wilcoxon (Wilcoxon, 1945).

4 Segmentación Inteligente de Espermatozoides Mediante Umbralización y *Watershed*

Casi todas las aplicaciones que utilizan procesamiento digital de imágenes tienen como paso previo una segmentación. Cuando esta no es buena, los resultados del resto del análisis no serán completamente fiables, dado que se perdería información relevante de la textura y la forma. Por tanto, es necesario preservar tanta información sobre la región de interés original como sea posible.

Como consecuencia, se ha propuesto un método que permita, por un lado, mejorar el proceso de segmentación de los espermatozoides, y, por otro, detectar cuándo una cabeza no ha sido bien segmentada. Con ello, se intenta minimizar el número de imágenes malas que pasarán a posteriores etapas de procesamiento, bien porque el número de cabezas bien segmentadas sea mayor, o porque las cabezas mal segmentadas sean automáticamente detectadas y descartadas.

El método de segmentación propuesto combina dos procedimientos diferentes. El primero utiliza ampliación del contraste de las imágenes, binarización por medio del umbral proporcionado por el método de Otsu (Otsu, 1979) y algunas operaciones morfológicas para eliminar tanto el ruido de la binarización

como la cola. El segundo método está basado en la transformada Watershed y el método de marcado de Meyer y Beucher (Meyer and Beucher, 1990).

La transformada Watershed considera la magnitud del gradiente de una imagen como una superficie topográfica. Dicha superficie se va "inundando" desde los mínimos de la región a una velocidad constante y, en el momento en que el agua de dos regiones se va a mezclar, se construye una *pared*. Desgraciadamente, a menudo esto da lugar a una sobresegmentación, ya que la magnitud del gradiente es muy sensible al ruido. Por ello se utilizan marcadores que indiquen dónde están los mínimos de la superficie topográfica (desde los cuales comenzar la "inundación"). Como marcador del frente se utiliza el resultado de la segmentación con el primer método o, en caso de que este no devolviese nada, un cuadrado situado en el centro de la imagen. Como marcadores del fondo se utilizan unas pequeñas líneas, una en cada esquina de la imagen.

Para descartar las células que no han sido bien segmentadas se considera que una cabeza está mal segmentada cuando el área de una cabeza es menor que el 70 % del área media de las cabezas del conjunto de imágenes, o la proporción entre los ejes mayor y menor de una elipse que tenga el mismo segundo momento central que la cabeza no esté entre 1,4 y 2,6.

Así pues, en primer lugar se segmentan las imágenes mediante el primer método. A continuación se detectan las que no han sido bien segmentadas y se utiliza el mismo método, con un umbral un 20 % mayor que el detectado en primer lugar. Una vez terminado, se vuelve a comprobar qué imágenes están mal segmentadas, y se aplica el método basado en Watershed en ellas. Finalmente, las imágenes que todavía estén mal segmentadas se eliminan del conjunto final.

El conjunto de imágenes utilizado para evaluar estos métodos está formado por 422 cabezas de espermatozoides vivos y 341 de muertos. En esta evaluación se han tenido en cuenta tanto la precisión del método de segmentación (el número de imágenes bien segmentadas), como la precisión de la detección (el número de imágenes detectadas correctamente como bien o mal segmentadas).

El método basado en umbralización ha obtenido una precisión global en la segmentación del 88,99 % (97,16 % en las vivas y 78,89 % en las muertas), mientras que la precisión en la detección ha sido 97,87 % y 87,98 %, respectivamente. El método basado en Watershed ha resultado ser mejor que el anterior, obteniendo una eficiencia global de 90,30 % (96.92 % en vivos y 82,11 % en muertos), aunque la tasa de aciertos en la detección ha sido ligeramente inferior

que en el caso anterior (96,92 % y 87,39 % con vivos y muertos, respectivamente). Por último, el método propuesto mejora los anteriores en tasa de cabezas bien segmentadas, con un 90,96 % (97,39 y 82,99 % en vivos y muertos respectivamente), aunque la tasa de cabezas bien detectadas ha sido ligeramente menor en la clase de las muertas (86,51 %). Aunque estos resultados sean muy similares al método basado en Watershed, nuestra propuesta es mucho más eficiente computacionalmente, ya que se han obtenido tiempos de ejecución 5 veces menores.

5 Descriptores de Textura Basados en *Curvelet* para Evaluar la Integridad del Acrosoma

El análisis multi-resolución es una herramienta muy potente en tareas de análisis de texturas. La transformada Wavelet discreta (DWT en sus siglas en inglés) ha sido ampliamente utilizada en en el análisis de imágenes de células y tejidos, e incluso en la evaluación de la integridad acrosómica (González et al., 2007), con buenos resultados. Por otro lado, la transformada Curvelet discreta (DCT en sus siglas en inglés) también ha obtenido buenos resultados en análisis de texturas. Algunos trabajos han comparado ambas, resultando la DCT mejor que la DWT, por ejemplo en descripción de imágenes de tomógrafos (Dettori and Semler, 2007), o de mamografías (Eltoukhy et al., 2010b).

Por este motivo hemos aplicado la transformada Curvelet al reconocimiento de acrosomas íntegros y dañados en imágenes de semen de verraco, mediante el análisis de su textura, y hemos comparado su rendimiento con otros descriptores de textura basados en la DWT, así como con algunos descriptores de región.

5.1 Conjunto de imágenes

Una vez adquiridas las imágenes del microscopio, éstas se recortan automáticamente, de modo que cada imagen solo presente un espermatozoide. Posteriormente, se segmentan mediante el método presentado en la sección 4, descartando las que no se pudieron segmentar bien. Al finalizar este proceso, el conjunto de imágenes que queda para este experimento está compuesto por 1849 imágenes: 945 acrosomas dañados y 904 acrosomas íntegros.

5.2 Caracterización de la integridad de los acrosomas

El objetivo de este experimento es caracterizar y clasificar espermatozoides de verraco en función de la integridad de sus acrosomas mediante descriptores de textura. Al ver el aspecto que presentan, parece que los descriptores de forma pudieran ser adecuados para ello. Para comprobar si es así, de la región de cada cabeza se han extraído momentos de Hu (7 características), Flusser (6 características), Legendre y Zernike (9 características cada uno). En cuanto a los descriptores de textura, se han calculado descriptores estadísticos combinados con las transformadas Wavelet y Curvelet.

Cada nivel de descomposición de la transformada Wavelet de una imagen da lugar a cuatro matrices de coeficientes: aproximaciones, que almacenan casi toda la energía de la imagen, y los detalles de alta frecuencia horizontales, verticales y diagonales. Se han calculado dos descriptores:

- El primero consiste en calcular la media y la desviación típica de cada una de las 12 matrices de coeficientes obtenidas de aplicar los tres primeros niveles de descomposición a la imagen, por lo que tendrá 24 características. Este descriptor se ha llamado WSF (*Wavelet Statistical Features*) en sus siglas en inglés (Arivazhagan and Ganesan, 2003).

- El segundo consiste en extraer la características *Energía*, *Contraste*, *Correlación* y *Homogeneidad* de las matriz de coocurrencia de la imagen original y de los coeficientes obtenidos del primer nivel de descomposición Wavelet. Cada característica se ha promediado en las matrices de coocurrencia con las orientaciones 0°, 45°, 90° y 135°, para hacer a este descriptor invariante a la rotación. Este descriptor está compuesto por 20 características, y se ha llamado WCF en sus siglas en inglés (*Wavelet Co-occurrence Features*). Los mejores resultados se obtuvieron cuando el parámetro de la distancia en la matriz de coocurrencia fue $d = 1$.

En cuanto a los descriptores basados en la transformada Curvelet, se han extraído otros dos descriptores:

- De cada una de las "cuñas" obtenidas tras aplicar la transformada se ha calculado la media y la desviación típica. Estos descriptores se han llamado CSF en sus siglas en inglés (*Curvelet Statistical Features*).

- Análogamente a los descriptores WCF, se han calculado las matrices de coocurrencia de la imagen original y de las "cuñas" de la transformada, y de cada una se ha calculado la *Energía*, *Contraste*, *Correlación* y *Homogeneidad*, promediadas en las orientaciones de las matrices de coocurrencia 0°, 45°, 90° y 135°. Estos descriptores se denominan CCF (*Curvelet Co-occurrence Features*).

En ambos casos se han probado varias combinaciones de escalas (3 y 4) y ángulos en la segunda escala (8, 12 y 16), obteniendo los mejores resultados para los valores 4 y 8 respectivamente. El número de características de estos descriptores es 52 y 108, respectivamente.

5.3 Resultados experimentales

Los patrones con la información de la textura o de la forma de las imágenes han sido clasificados mediante una red neuronal con arquitectura de perceptrón multicapa. En las clasificaciones se han probado diversas combinaciones de ciclos de entrenamiento (200, 300 y 400) y neuronas en la capa oculta (2, 3 y 5), con el objetivo de encontrar la configuración óptima de la red. La clasificación se ha realizado utilizando validación cruzada con *k-folds*, que consiste en dividir el conjunto de datos en k subconjuntos del mismo número de datos, con la misma proporción de clases, y tomar $k-1$ para entrenar la red, y el subconjunto restante para el test. Este proceso se repite k veces, tomando un fold de test diferente cada una. La tasa de error será el promedio de la tasa de error de todos los folds. Por último, este proceso se ha repetido 10 veces, para evitar posibles efectos aleatorios. Las tasas de acierto presentadas son una media de estas 10 ejecuciones.

Existen voces en la comunidad de *Machine Learning* que indican que la precisión no es la métrica más adecuada para ilustrar el rendimiento de un clasificador, sino que el análisis ROC (*Receiver Operating Characteristics*) es más potente (Provost et al., 1998). Por lo tanto, se han calculado las áreas bajo la curva ROC (AUC en sus siglas en inglés) de cada descriptor.

En la Tabla 1 se muestran las tasas de acierto y las AUC de cada descriptor, junto con la configuración de la red con la que se obtuvieron.

En cuanto a los descriptores de forma, es notable que ninguno de ellos ha superado – ni siquiera igualado – en rendimiento a cualquiera de los de textura.

Tabla 1: Precisión (en %) de la clasificación de acrosomas íntegros y dañados

Descriptor	Ciclos	Neuronas	AUC	Precisión (%)		
				Global	Íntegros	Dañados
WSF	400	5	0,942	87,26	89,25	85,37
WCF	400	3	0,993	96,43	96,30	96,56
CSF	200	2	0,992	96,42	96,48	96,36
CCF	200	5	0,995	97,00	97,29	96,73
Hu	400	5	0,899	82,76	85,39	80,24
Flusser	400	5	0,888	81,31	83,50	79,22
Legendre	400	5	0,785	71,74	76,93	66,76
Zernike	400	5	0,716	65,86	67,79	64,02

Los momentos de Hu han sido los mejores (con una tasa de acierto de 82,76 %), pero todavía son peores que el peor de los descriptores de textura evaluados (WSF, con una tasa de acuerto del 87,26 %). Esto prueba que, al contrario de lo que se pudiera pensar, los descriptores de región no funcionan bien en este problema particular, mientras que los de textura son más apropiados. Los mejores resultados se han obtenido utilizando los descriptores CCF, que han obtenido una precisión en la clasificación del 97 %, seguidos por los WCF y los CSF, que obtienen 96,4 %. También resulta muy interesante que los errores de cada clase están muy equilibrados. Estos resultados se confirman en el análisis ROC, puesto que CCF alcanza la mayor área bajo la curva, seguida de WCF y CSF respectivamente, mientras que ninguno de los descriptores de forma no llegan a 0,90.

Adicionalmente, un test no paramétrico de Wilcoxon entre CCF y cada uno de los otros descriptores, utilizando las AUC de cada fold de test en cada iteración como puntuaciones, tal como se muestra en (Moreno-Torres et al., 2010). De acuerdo con estos tests, las diferencias entre los resultados son todas estadísticamente significativas.

6 *Pattern Spectrum* Adaptativo Geodésico

Los descriptores de textura se utilizan para extraer características cuantitativas que representan una región de interés (ROI en sus siglas en inglés), con el objetivo de, posteriormente, clasificarla. Las técnicas convencionales de análisis de texturas se aplican tal cual en toda la ROI, de manera que obvian el hecho

de que la textura de una región de interés puede no ser toda homogénea, por lo que pueden resultar subóptimas al no captar todas sus variaciones.

La Morfología Matemática (MM) proporciona una serie de procesos que se pueden aplicar a una imagen para analizar estructuras geométricas dentro de ella, *e.g.* para eliminar detalles más pequeños que una estructura denominada Elemento Estructurante (SE en sus siglas en inglés). Una de sus aplicaciones es al análisis de texturas, gracias a un descriptor denominado *Pattern Spectrum* (PS) (Maragos, 1989). El PS es una distribución del tamaño de objetos ("gránulos") dentro de una textura, que se calcula realizando sucesivas erosiones y dilataciones de la textura, utilizando para ello elementos estructurantes de diferentes tamaños.

Como ya puntualizamos, no está claro cuál es el mejor elemento estructurante para un problema determinado (de Ves et al., 2006). Utilizar un SE con la misma forma para toda la textura no solucionaría el problema de los descriptores de textura convencionales. Por lo tanto, nuestra propuesta es caracterizar la textura mediante un *Pattern Spectrum* calculado con elementos estructurantes cuya forma y tamaño varían en cada píxel en función de un criterio de distancia geodésica, sin requerir ningún tipo de conocimiento a priori sobre la textura.

En esta sección presentamos este descriptor adaptativo, que hemos denominado *Pattern Spectrum* Adaptativo Geodésico (AGPS en sus siglas en inglés), y lo evaluamos en la caracterización de texturas de diversos materiales, obtenidas de la base de datos VisTex, comparándolo con el *Pattern Spectrum* convencional. Por otro lado, estos descriptores también se han comparado, junto con los descriptores de textura WCF, en la tarea de caracterización de espermatozoides de verraco vivos y muertos. Hasta donde sabemos, no existen aplicaciones comerciales en los que se lleve a cabo esta tarea mediante procesamiento de imágenes digitales en escala de grises. De acuerdo a expertos veterinarios, determinar la vitalidad de un espermatozoide sin utilizar tinciones es una tarea extremadamente difícil, y todavía no resuelta.

6.1 Conjuntos de imágenes

Como indicamos anteriormente, se han realizado dos experimentos para evaluar el descriptor AGPS y compararlo con el PS convencional. En primer lugar se han utilizado imágenes texturas de diversos materiales de la base de datos

MIT Medialab Vision Texture (VisTex), con el objetivo de probar el método propuesto con diferentes tipos de texturas. Se han utilizado 75 imágenes de 512×512 píxeles, divididas en varias categorías que contienen muy pocas imágenes. Por ello, cada imagen se ha dividido en 41 imágenes de 102×102 píxeles cada una, en dos fases. Para la primera división se utiliza una rejilla de 5×5, con lo que cada imagen queda dividida en 25. A continuación la división se realiza mediante una rejilla de 4×4 con el mismo tamaño de celda que en el caso anterior. Para que estas divisiones sean lo más diferentes posibles a las anteriores, el origen de la rejilla se ha situado en el píxel $(51, 51)$. De este modo, se han extraído 41 imágenes de cada una de las originales.

En cuanto a las imágenes de espermatozoides vivos y muertos, el recorte y segmentación se han realizado de la misma manera que en el caso de los íntegros y dañados (sección 5.1). A continuación, cada cabeza fue recortada en la imagen en escala de grises por medio del mínimo rectángulo que limita su región, redimensionada a 63×108 píxeles y finalmente rotada de modo que su eje mayor quedase en posición vertical con la parte apical de la cabeza en la parte superior. Este conjunto de imágenes posee 845 cabezas de espermatozoides: 470 vivos y 375 muertos.

6.2 Descriptores

Diversos autores han propuesto formas diferentes de calcular el *Pattern Spectrum* (Gonzalez and Woods, 2002; Maragos, 1989; Petrou and Sevilla, 2006). Para determinar cuál de ellas utilizar se midieron los tiempos de ejecución de cada método en 200 imágenes de 129×130 píxeles de textura de madera. El tiempo medio de ejecución más bajo se obtuvo con el método propuesto por Petrou y García-Sevilla, mostrado en la ecuación (1), por lo que este será el que utilicemos.

$$PS_{f,G}(n) = A\left[(f \oplus nG)(x,y) - (f \ominus nG)(x,y)\right] \tag{1}$$

donde G es el elemento estructurante, f es la imagen y A es el área de la misma. Para normalizar el PS, se divide cada elemento de la función dividiéndolo entre el área de f.

La forma del elemento estructurante (SE en sus siglas en inglés) que se utiliza en dicho cálculo permanece inalterable para toda la imagen, por lo que

no logra captar las variaciones geométricas de la misma. Nuestra propuesta es calcular el PS mediante elementos estructurantes que varíen en tamaño y forma en cada píxel de forma que se adapte lo mejor posible a la textura. Si se considera la textura $f(x, y)$ como una superficie, en la que cada píxel p_i está definido por las coordenadas (x_i, y_i, z_i), donde z_i es su valor de nivel de gris, el soporte del elemento estructurante cuyo origen está en el punto p_0 estará formado por los elementos que se encuentran a una distancia máxima de n unidades del mismo, como se muestra en la ecuación (2).

$$Supp_{G_\sigma} = \{p_s = [x_s, y_s] \ ; \ d\,(p_0, p_s) \leq \sigma\} \tag{2}$$

La función de distancia se puede medir con cualquier métrica, *e.g. Cityblock*, *Chessboard*, Euclídea, *etc.* Sin embargo, estas métricas no pueden capturar la estructura no lineal de una superficie, mientras que la distancia Geodésica sí.

Por lo tanto, nuestra propuesta consiste en calcular, para cada píxel de la textura, un mapa de distancia geodésica cuyo origen esté en dicho punto por medio del algoritmo Fast Marching (Sethian, 1996), mediante el cual podremos encontrar los puntos que cumplen la condición de la ecuación (2), que constituyen el soporte del elemento estructurante. Por tanto, llamaremos $G_{i,j,n}$ al SE cuyo origen está en la posición (i, j), y cuyo soporte está formado por los puntos que se encuentran a una distancia menor o igual que n.

6.3 Experimentos y resultados

En primer lugar, se han tomado pares de clases aleatorios de la base de datos VisTex como problemas de clasificación separados. Estas imágenes se han descrito, por un lado, utilizando nuestra propuesta, AGPS, y por otro, con el PS convencional, calculado mediante un elemento estructurante de forma cuadrada y tamaño 3×3, con el objetivo de compararlos. Tanto PS como AGPS se han calculado con varias longitudes, y en ambos casos se han extraído los momentos centrales normalizados 2 a 6, por lo que todas las texturas han sido caracterizadas por medio de 5 características. Los mejores resultados se han obtenido cuando las longitudes de AGPS y PS fueron 20 y 12, respectivamente. Los resultados de las clasificaciones, junto con la diferencia de las tasas de acierto obtenidas con AGPS y con PS, realizadas con Máquinas de Vector Soporte (SVM en sus siglas en inglés) se muestran en la Tabla 2.

Tabla 2: Precisión en la clasificación de texturas VisTex con AGPS y PS convencional.

Clases	AGPS	PS	Diferencia
FaRo - FoBe	88,02	64,94	23,08
FaWi - FoSw	79,33	71,50	7,84
FaRo - FaWi	99,80	92,76	7,04
FaWi - Sand	98,64	93,34	5,30
Bark - FaWi	97,87	94,02	3,84
FaRo - Metal	92,04	88,78	3,25
Bark - FoSw	89,22	86,44	2,78
Bark - FaFi	92,26	90,79	1,46
FaBa - FaWi	95,77	94,63	1,15
FaFi - Sand	95,82	98,26	-2,43
FaWo - Sand	91,95	95,12	-3,17
Sand - Stone	92,91	96,10	-3,19
FaFi - Stone	95,72	99,34	-3,61
FoBe - Stone	92,27	98,37	-6,10
Metal - Stone	91,16	99,73	-8,57
FaBa - FoSw	77,99	90,64	-12,65
Bark - Metal	79,99	95,35	-15,35
FaBa - FaRo	84,12	99,65	-15,53

Teniendo en cuenta estos resultados, y realizando una inspección visual de las imágenes de las clases utilizadas en este experimento (ver Figura 1) se puede observar que AGPS presentó un mejor rendimiento (con mayor diferencia con respecto a PS) cuando los téxeles de las texturas analizadas son similares. Análogamente, cuando los téxeles de las texturas son muy diferentes entre sí, el rendimiento obtenido por PS supera ampliamente el de AGPS.

Esta conclusión tiene sentido desde el punto de vista que el elemento estructurante geodésico está adaptado a la región que rodea el píxel sobre el que está situado el origen, de modo que el *Pattern Spectrum* puede discriminar las texturas mejor que el PS clásico.

Por otro lado, en cuanto a la descripción de las imágenes de espermatozoides vivos y muertos, estas han sido descritas por medio de AGPS y PS. Una vez más, aunque las funciones se pueden utilizar tal cual como descriptores, debido a la "maldición de la dimensionalidad" se han extraído los momentos centrales normalizados de órdenes 2 a 6 para caracterizar las texturas con 5 características. Las mejores tasas de acierto se han obtenido cuando las lon-

Figura 1: Ejemplos de subimágenes 102×102 de las clases VisTex utilizadas.

gitudes de AGPS y de PS han sido 20 y 12, respectivamente. También se ha caracterizado este conjunto con los descriptores de textura WCF (Sección 5).

Los datos se han clasificado mediante una red neuronal con arquitectura de perceptrón multicapa, con una capa oculta y entrenada utilizando retropropagación. Los datos se normalizaron con media cero y desviación típica igual a uno, y la clasificación se llevó a cabo mediante validación cruzada con 10 folds, y repetida durante 10 iteraciones. En la Tabla 3 se muestran las tasas de error de la clasificación y las áreas bajo la curva ROC obtenidas, junto con las neuronas en la capa oculta y los ciclos de entrenamiento de la red.

Estos resultados muestran que el *Pattern Spectrum* adaptativo supera al clásico tanto en AUC como en tasa de acierto para todas las configuraciones de la red neuronal. Los tests de Wilcoxon realizados con las puntuaciones obtenidas para todos los folds durante las 10 iteraciones confirman que estas conclusiones son estadísticamente significativas. Si se comparan los resultados de AGPS y WCF, es notable que el primero supera al segundo en tasa de acierto, mientras que se da el fenómeno contrario si se compara el área bajo la curva ROC. En este caso, el test de Wilcoxon utilizando como puntuaciones las tasas de acierto muestran que las diferencias entre ambos descriptores son

Tabla 3: Precisión (en %) y AUC de la clasificación de los espermatozoides vivos y muertos utilizando AGPS, PS y WCF

Neuronas	Ciclos	PS		AGPS		WCF	
		AUC	Precisión	AUC	Precisión	AUC	Precisión
2	200	0,646	62,64	0,736	69,27	0,763	68,21
2	300	0,661	63,90	0,740	69,36	0,761	68,14
2	400	0,665	64,17	0,742	69,46	0,758	68,86
3	200	0,659	63,59	0,742	69,61	0,760	68,52
3	300	0,671	64,24	0,744	69,75	0,757	68,17
3	400	0,668	64,30	0,741	69,59	0,753	68,36
5	200	0,666	64,54	0,742	69,43	0,752	68,38
5	300	0,668	64,63	0,746	69,58	0,751	67,49
5	400	0,670	64,33	0,745	69,61	0,741	67,39

estadísticamente significativas, mientras que si se toman las AUC, no existen diferencias estadísticamente significativas.

7 Comparación de Métodos de Cuantificación

7.1 Enunciado del problema

En un problema de clasificación, sea $S_t = \{(\mathbf{x}^k, d^k), k = 1, \ldots, K\}$ un conjunto de datos donde $\mathbf{x}^k$ es el vector de características del elemento k-ésimo, y d^k la clase a la que pertenece. Los elementos $\mathbf{x}^k \in S_t$ han sido registrados independientemente de acuerdo a una función de densidad $p(\mathbf{x}|d_i)$, siendo $P_t(d_i)$ la probabilidad *a priori* de la clase d_i.

S_t se utiliza para generar un modelo de clasificación que, dado un $\mathbf{x}^k$, genera una salida $\widehat{\mathbf{y}}^k$, en base a la cual le asigna una clase d^k. Cuando el clasificador se entrena minimizando la función de coste adecuada, las salidas $\widehat{y}_i^k$ proporcionan una estimación de la probabilidad *a posteriori* de que la observación $\mathbf{x}^k$ pertenezca a la clase d_i ($\widehat{P}(d_i|\mathbf{x})$) (Bishop, 1996).

Considérese ahora un nuevo conjunto no etiquetado $U = \{(\mathbf{x}^l), l = 1, \ldots, N\}$ del cual se quiere estimar la distribución de las clases $P(d_i)$. El método ingenuo para llevar esto a cabo, llamado Clasificar y Contar (CC) (Forman, 2008), se basa en contar las predicciones $\widehat{d}^k$ asignadas por el clasificador. Sin embargo, si las probabilidades a priori del nuevo conjunto ($P(d_i)$) son diferentes de las

del conjunto de entrenamiento ($P_t(d_i)$), el clasificador se comporta de manera subóptima.

Existen métodos para estimar probabilidades a priori basadas en matrices de confusión, expuestas en la sección 7.2. Además, en esta tesis presentamos dos métodos originales para estimar dichas probabilidades. El primero se describe en la sección 7.3, y se basa en las probabilidades a posteriori proporcionadas por el clasificador. El segundo se basa en medir la diferencia entre distribuciones utilizando la distancia de Hellinger, y se describe en la sección 7.3.

7.2 Métodos previos basados en matrices de confusión

Las matrices de confusión resumen el rendimiento de los clasificadores. Muestran el número de elementos clasificados como pertenecientes a la clase i cuando en realidad pertenecen a la clase j. A partir de ellas se pueden extraer las siguientes métricas:

- Tasa de verdaderos positivos: $tpr = \widehat{P}(\widehat{d_1}|d_1) = TP/P$
- Tasa de falsos positivos: $fpr = \widehat{P}(\widehat{d_1}|d_0) = FP/N$
- Tasa de verdaderos negativos: $tnr = \widehat{P}(\widehat{d_0}|d_0) = TN/N$
- Tasa de falsos negativos: $fnr = \widehat{P}(\widehat{d_0}|d_1) = FN/P$

En un problema binario de clasificación la probabilidad de que un clasificador haga una predicción positiva es:

$$\begin{aligned}\widehat{P}(\widehat{d_1}) &= \widehat{P}(\widehat{d_1}|d_1)\cdot\widehat{P}(d_1) + \widehat{P}(\widehat{d_1}|d_0)\cdot\widehat{P}(d_0) = \\ &= tpr\cdot\widehat{P}(d_1) + fpr\cdot(1-\widehat{P}(d_1)) = \\ &= fpr + \widehat{P}(d_1)\cdot(tpr - fpr)\end{aligned}$$

por lo que la estimación de las probabilidades a priori de la clase d_i es:

$$\widehat{P}(d_1) = \frac{\widehat{P}(\widehat{d_1}) - fpr}{tpr - fpr} \tag{3}$$

Puesto que se puede asumir que la función de densidad de las clases $p(\mathbf{x}|d_i)$ no cambia a lo largo del tiempo (Saerens et al., 2002), se considera que no

hay cambios considerables en las fpr y tpr de las distribuciones de nuevos conjuntos de datos y las del conjunto de entrenamiento. Por ello, la matriz de confusión, y, por tanto, los valores de fpr y tpr, se pueden calcular mediante k-folds estratificado a partir del conjunto de datos de entrenamiento, utilizando un valor de k lo más grande posible (Forman, 2008). Una vez calculadas, y generado el clasificador, éste se aplica en los nuevos conjuntos de datos para calcular $\widehat{P}(\widehat{d_i})$. A continuación, gracias a la ecuación (3) se puede calcular la probabilidad a priori $\widehat{P}(d_i)$. El método se llama *Adjusted Count* (AC).

Basado en AC, Forman propuso el método *Median Sweep* (MS), que consiste, básicamente, en calcular varias matrices de confusión utilizando diversos umbrales de clasificación. Con cada una de ellas se aplica el método AC y finalmente, la estimación de la distribución de las clases será la mediana de las estimaciones derivadas de cada matriz de confusión.

7.3 Cuantificación basada en probabilidades a posteriori

El primer método de cuantificación consiste en un algoritmo iterativo basado en el algoritmo EM (*Expectation Maximization*) propuesto en (Saerens et al., 2002), que ajusta las salidas de un clasificador, calculando las nuevas probabilidades a priori como paso intermedio, lo cual es, precisamente, nuestro objetivo. Este método requiere que las salidas del clasificador sean estimaciones de las probabilidades a posteriori.

Las estimaciones de las probabilidades *a posteriori* y *a priori* de un conjunto U se inicializan con las salidas $\mathbf{y}^k$ generadas por un clasificador, y con las frecuencias de las clases en el conjunto de entrenamiento, respectivamente (Ecuaciones (4) y (5)).

$$\widehat{P}^{(0)}(d_i|\mathbf{x}_k) = \widehat{y}_i^k \tag{4}$$

$$\widehat{P}^{(0)}(d_i) = \frac{|S_t^i|}{|K|} \tag{5}$$

donde K es el número total de elementos del conjunto de entrenamiento, y $|S_t^i|$ es el número de instancias pertenecientes a la clase i en dicho conjunto.

Las ecuaciones (6) y (7) proporcionan la estimación de las probabilidades *a priori* y *a posteriori* en la iteración r-ésima, respectivamente.

$$\widehat{P}^{(r)}(d_i) = \frac{1}{N}\sum_{l=1}^{N}\widehat{P}^{(r-1)}(d_i|\mathbf{x}^k) \tag{6}$$

$$\widehat{P}^{(r)}(d_i|\mathbf{x}^k) = \frac{\dfrac{\widehat{P}^{(r)}(d_i)}{\widehat{P}^{(0)}(d_i)}\widehat{P}^{(0)}(d_i|\mathbf{x}^k)}{\displaystyle\sum_{j=0}^{M-1}\frac{\widehat{P}^{(r)}(d_j)}{\widehat{P}^{(0)}(d_j)}\widehat{P}^{(0)}(d_j|\mathbf{x}^k)} \quad . \tag{7}$$

Este procedimiento se repite durante un número predeterminado de iteraciones, o hasta que la diferencia entre dos estimaciones sucesivas sea menor que un umbral determinado. Este método se denomina *Posterior Probability* (PP).

7.4 Cuantificación basada en la distancia de Hellinger

Como ya se ha puntualizado, nos concentramos en problemas en que las funciones de densidad de las clases, $p(\mathbf{x}|d_i)$, se mantienen fijas pero en el que las probabilidades *a priori* de las clases, $P(d_i)$, pueden variar tras la generación del modelo de clasificación. Cuando esto ocurre, las funciones de densidad $p(x)$, así como las probabilidades *a posteriori* también varían.

En un problema real se puede estimar la $p(x)$ del conjunto U. Además, podemos generar conjuntos de datos de validación con una distribución $p_v(x)$ cualquiera, y calcular su diferencia respecto a $p(x)$, con el objetivo de encontrar la distribución de datos de validación más parecida a la de test. Este proceso permite estimar las proporciones reales de las clases, que corresponden a las distribuciones que minimizan dicha diferencia.

La distancia de Hellinger (HD) es una medida de esta divergencia que recientemente se ha utilizado para detectar cambios en las distribuciones de datos, que daban lugar a fallos en el clasificador. Dadas dos distribuciones discretas, la distancia de Hellinger entre ellas se puede calcular agrupándolas en *bins*, cada uno de los cuales presenta una probabilidad asociada. Por tanto, dados un conjunto de datos no etiquetados U con probabilidades a priori $P(d_i)$, y un conjunto de datos de validación V con probailidades a priori $P_v(d_i)$, extraído del conjunto de datos de entrenamiento, la distancia de Hellinger entre ellas viene dada por la Ecuación (8).

$$HD(V,U) = \frac{1}{n_f}\sum_{f=1}^{n_f} HD_f(V,U) \tag{8}$$

donde n_f es el número de características, y la distancia entre U y V respecto a la característica n_f se calcula de acuerdo a la Ecuación (9).

$$HD_f(V,U) = \sqrt{\sum_{i=1}^{b}\left(\sqrt{\frac{|V_{f,i}|}{|V|}} - \sqrt{\frac{|U_{f,i}|}{|U|}}\right)^2} \tag{9}$$

Nótese que b es el número de *bins*, $|U|$ es el número total de instancias en el conjunto U, y $|U_{f,i}|$ es el número de elementos de dicho conjunto cuya característica f pertenece al *bin* i. Análogamente, $|V|$ y $|V_{f,i}|$ corresponden al conjunto de validación.

Por lo tanto, se puede estimar directamente la distribución de las clases en el conjunto de test encontrando un conjunto de validación con unas probabilidades a priori $P_v(d_i)$ que minimice la distancia de Hellinger entre ambos. Se pueden generar conjuntos de validación mediante submuestreo del conjunto S_t, pero esto implica descartar instancias y, por lo tanto, perder información, lo cual no es deseable, especialmente si el número de instancias es bajo, o si el conjunto de datos está muy desequilibrado. Para evitar esto, en nuestra propuesta se modela la función de densidad $p(\mathbf{x}|d_i)$ (asumida estacionaria) de modo que la distribución de datos de validación $p_v(\mathbf{x})$ dada una probabilidad a priori de las clases $P_v(d_i)$ sea

$$p_v(\mathbf{x}) = \sum_{i=1}^{M} p(\mathbf{x}|d_i)P_v(d_i) \tag{10}$$

donde M es el número de clases.

Por lo tanto, la HD entre $p(\mathbf{x})$ y $p_v(\mathbf{x})$ respecto a la característica f se calcula mediante la Ecuación (9), pero realizando la sustitución

$$\frac{|V_{f,i}|}{|V|} = \frac{|S^0_{t,f,i}|}{|S^0_t|}P_v(d_0) + \frac{|S^1_{t,f,i}|}{|S^1_t|}P_v(d_1) \tag{11}$$

donde $P_v(d_0) = 1 - P_v(d_1)$ (en el caso binario), $|S^0_t|$ es el número de instancias del conjunto de entrenamiento que pertenecen a la clase 0, y $|S^0_{t,f,i}|$ es el número de ejemplos de entrenamiento de la clase 0 cuya característica f pertenece al *bin* i. Análogamente, $|S^1_t|$ y $|S^1_{t,f,i}|$ son equivalentes en la clase 1.

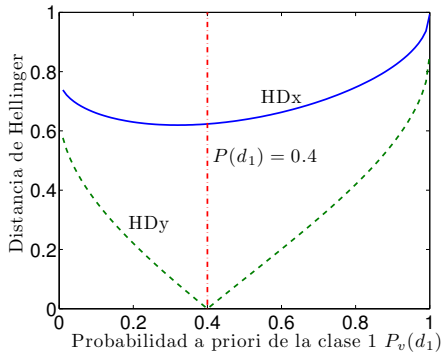


Figura 2: HD entre la distribución de salida del clasificador (curva HDy) y la de los datos originales (curva HDx) de un conjunto de test y diferentes conjuntos de validación. Los datos están definidos en un espacio 20-dimensional ($n_f = 20$).

En condiciones reales, la escasez de datos es un problema que ocasiona que el conjunto de datos de entrenamiento no sea representativo en todas las regiones del espacio n_f-dimensional. En estos casos la curva HD calculada puede no ser fidedigna. Una solución puede ser calcular la distancia de Hellinger entre las distribuciones de los datos de salida de un clasificador (ver Fig. 2). Esto simplifica el problema, al reducir su dimensionalidad a $M - 1$ (donde M es el número de clases). Estas dos propuestas se llaman HDx (HD entre vectores de características $\mathbf{x}$) y HDy (HD entre vectores de salida del clasificador $\mathbf{y}$).

7.5 Comparación de métodos de cuantificación

Para la comparación de los métodos de cuantificación expuestos se han utilizado 15 conjuntos de datos; 14 de ellos obtenidos del repositorio de la UCI, y otro del proyecto ELENA. Se escogieron con diferentes tamaños y proporciones de clases, por lo que se dispone de diversos escenarios para los experimentos. Todos los problemas son binarios, bien porque los conjuntos lo son originalmente, o porque han sido adaptados para que lo sean.

La diferencia entre la distribución real y la estimada por los métodos evaluados se ha medido por medio del Error Absoluto Medio y el Error Relativo Medio (MAE y MRE en sus siglas en inglés, respectivamente).

Como se ha expuesto, los métodos de cuantificación utilizan, de un modo u otro, un clasificador. Además, PP requiere que las salidas de dicho clasificador

proporcionen estimaciones sobre las probabilidades a posteriori. Por ello, la clasificación se ha llevado a cabo mediante una red neuronal entrenada mediante retropropagación, con una capa oculta. Los datos se normalizaron para que tuvieran media cero y desviación típica igual a uno. Tanto la arquitectura de la red como el número de ciclos de entrenamiento se determinaron para cada conjunto de datos mediante validación cruzada de 10-folds.

En primer lugar se han comparado los métodos HDx y HDy con las mencionadas bases de datos. Cada una se dividió en dos subconjuntos diferentes mediante muestreo estratificado con asignación proporcional: un 30 % se utilizó como conjunto de test, U, y el 70 % restante, llamado S_t, para entrenar el clasificador y generar los conjuntos de validación. La red neuronal en cada problema se entrenó con un conjunto equilibrado, que contenía todas las instancias de la clase minoritaria de S_t, y el mismo número de instancias de la otra clase, extraídas aleatoriamente. Las estimaciones con ambos métodos se calcularon con un número de *bins* que varió de 10 a 110 en pasos de 10. Para determinar la estimación final de la probabilidad *a priori* se tomó la mediana de estas 11 estimaciones. De este modo se evita tener que fijar el parámetro b. Para cada problema se realizaron 50 repeticiones (esto es, 50 conjuntos de test diferentes), con el objetivo de evitar posibles efectos aleatorios. Los resultados son la media de estas 50 iteraciones, y muestran que, mientras que HDx obtiene buenos MAE y MRE (en el orden de 10^{-2}), HDy lo superó claramente. Esta diferencia es estadísticamente significativa, de acuerdo con el test no paramétrico de la prueba de los signos de Wilcoxon (Wilcoxon, 1945).

Una vez que se ha mostrado que HDy supera a HDx, se ha comparado este con los métodos PP, AC, MS, y con el método CC. La metodología del experimento ha sido la misma que en el caso del experimento anterior. Las matrices de confusión para los métodos AC y MS se estimaron utilizando el subconjunto S_t mediante validación cruzada (50-folds). La Tabla 4 muestra los MAE para los 15 conjuntos de datos evaluados.

Se puede observar que las estimaciones son claramente mejores cuando se utiliza un método de cuantificación que cuando se confía en el método de simplemente contar las predicciones del clasificador. Por otro lado, HDy supera al resto de métodos de cuantificación en ranking medio, y, además, estas diferencias son estadísticamente significativas, según los tests de Wilcoxon que se han realizado. Si estos resultados se miden con el MAE la conclusión es la misma.

Tabla 4: MRE (en %) de las estimaciones realizadas por los métodos de cuantificación HDy, CC, AC, MS and PP.

Conj. datos	CC	AC	MS	PP	HDy
Breast Cancer	6,41 (5)	3,24 (2)	3,22 (1)	4,18 (4)	3,58 (3)
CMC	94,9 (5)	30,52 (2)	48,45 (4)	34,64 (3)	16,81 (1)
Coil	521,1 (5)	152,73 (3)	111,71 (2)	345,97 (4)	19,78 (1)
Diabetes	13,54 (5)	10,05 (3)	12,38 (4)	8,92 (2)	7,37 (1)
German Credits	47,23 (5)	30,14 (3)	32,46 (4)	27,53 (2)	11,16 (1)
Letters (G)	191,36 (5)	21,58 (2)	25,82 (3)	76,24 (4)	4,07 (1)
Letters (H)	286,42 (5)	21,33 (2)	28,96 (3)	103,74 (4)	7,38 (1)
Mammog. mass	5,37 (1)	7,96 (4)	9,11 (5)	6,55 (2)	6,66 (3)
Page (picture)	327,33 (5)	56,78 (3)	43,81 (2)	82,16 (4)	16,28 (1)
Phoneme	44,78 (5)	6,21 (3)	6,91 (4)	5,79 (2)	4,57 (1)
Semeion (8)	91,81 (5)	17,41 (3)	16,16 (1)	45,93 (4)	16,36 (2)
Spambase	3,67 (5)	2,04 (3)	2,49 (4)	1,88 (2)	1,61 (1)
Wine (red)	678,61 (5)	224,23 (3)	218,62 (2)	564,58 (4)	29,54 (1)
Wine (white)	616,63 (5)	139,8 (3)	98,81 (2)	408,84 (4)	28,97 (1)
Yeast	53,6 (5)	14,91 (3)	25,91 (4)	12,96 (2)	9,05 (1)
Rank medio	4,733	2,8	3	3,133	1,3

8 Evaluación de la Calidad del Semen de Verraco: Estudio Empírico

En esta sección se han aplicado los métodos de cuantificación propuestos, y se han comparado con los métodos previos (ver sección 7) utilizando datos reales de una aplicación de evaluación de la calidad seminal. En concreto, se han realizado dos experimentos: en el primero se utilizan datos de acrosomas íntegros y dañados caracterizados mediante los descriptores de textura WCF (ver sección 5). Recordamos que este conjunto tiene 1849 instancias: 945 de la clase "dañados" y 904 de la clase "íntegros". El segundo utiliza un conjunto de datos correspondientes a espermatozoides vivos y muertos cuya textura ha sido caracterizada mediante el descriptor adaptativo AGPS propuesto en esta Tesis (ver sección 6). Este conjunto está compuesto por 845 instancias: 470 de espermatozoides vivos, y 375 muertos.

8.1 Cuantificación de acrosomas íntegros y dañados

En este experimento el conjunto de entrenamiento está compuesto por el 70 % de las instancias de la clase minoritaria, y el mismo número de instancias de

la otra clase, para que esté equilibrado. El conjunto de test tiene 280 elementos, aunque variando las proporciones de la clase 1 entre 0.05 y 0.50. Ambos conjuntos son disjuntos, y se extraen aleatoriamente cada vez. La red neuronal tiene 3 neuronas en la capa oculta y se entrena durante 400 ciclos, pues es la configuración que mejores resultados dio en la clasificación (ver sección 5). El resto del diseño del experimento es igual que los de la sección 7.

En la tabla 5 se muestran los MAE de los métodos de cuantificación evaluados. Se han realizado, además, tests de Wilcoxon utilizando los valores del rendimiento en las 50 iteraciones. En dicha tabla se muestran subrayados los resultados de los métodos con el menor error, así como aquellos con los que el test de Wilcoxon no haya encontrado diferencias significativas en cada escenario.

Tabla 5: MRE de los métodos de cuantificación en los 10 escenarios de test.

$P(d_1)$	**CC**	**AC**	**MS**	**PP**	**HDx**	**HDy**
0,05	71,63	21,18	25,53	10,78	124,08	13,92
0,10	32,10	10,31	12,40	7,46	56,28	7,48
0,15	18,50	6,15	7,39	4,71	33,49	4,88
0,20	11,59	4,78	5,52	4,10	20,24	4,20
0,25	7,87	3,73	3,94	3,02	12,67	3,22
0,30	5,42	2,79	3,04	2,64	7,89	2,72
0,35	3,75	2,48	2,52	2,29	5,59	2,41
0,40	2,53	2,24	2,20	2,05	4,58	2,24
0,45	1,70	1,83	1,80	1,68	4,94	1,76
0,50	1,49	1,66	1,64	1,56	5,17	1,55

Se puede observar que HDy y PP son siempre los mejores métodos, que AC y MS presentan un redimiento similar, aunque sólo son competitivos con HDy y PP en proporciones de dañados a partir de 0.30. Estos resultados demuestran los beneficios de utilizar un método de estimación, en vez de confiar en el método CC.

También se ha evaluado cómo afecta el rendimiento del clasificador a la calidad de las estimaciones realizadas por los métodos de cuantificación. Para ello, se han utilizado redes neuronales con 3 neuronas en la capa oculta, pero variando el número de ciclos de entrenamiento, de modo que la tasa de error en la clasificación se incremente. En concreto se han utilizado 400 ciclos de entrenamiento (3,57 % de tasa de error), 200 (4,15 %), 150 (6,47 %), 100 (15,40 %) y 75 (32,10 %). Estas tasas de error se han obtenido mediante validación cruzada,

utilizando 10 folds. Uniendo los puntos de los errores relativos obtenidos en los 10 escenarios se puede formar una curva por cada método y con cada configuración de la red. La Figura 3 muestra las áreas bajo estas curvas (AU_MRE), donde se puede observar la baja robustez de CC, AC, PP y, especialmente, MS. El método HDy es mucho más robusto que el resto respecto a cambios en el rendimiento del clasificador. Esto es importante, ya que, utilizando este método, el ajuste del clasificador no es crítico.

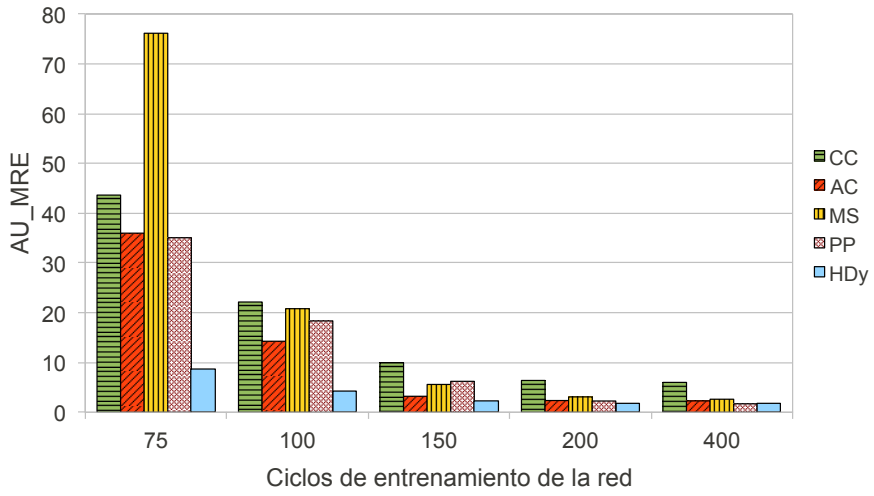


Figura 3: Área bajo las curvas de las MRE (AU_MRE) de los métodos de cuantificación.

8.2 Cuantificación de espermatozoides vivos y muertos

El diseño de este experimento es igual que en el caso de los acrosomas íntegros y dañados, excepto por el tamaño de los conjuntos de test, que en este caso ha sido de 140 elementos (también variando las proporciones de elementos de la clase "muertos" entre 0,05 y 0,50). En este caso, se fijaron 5 neuronas en la capa oculta de la red neuronal, y 400 ciclos de entrenamiento, ya que fue la configuración con la que se obtuvo la mejor tasa de acierto con el descriptor AGPS (ver sección 6).

Los MAE obtenidos por los métodos CC, AC, MS, PP y los métodos basados en la distancia de Hellinger se muestran en la tabla 6, en la que están subrayados

aquellos en los que el test no paramétrico de Wicoxon no encontró diferencias significativas.

Tabla 6: MAE de los métodos de cuantificación para 10 escenarios de test diferentes con el conjunto de datos AGPS de espermatozoides vivos y muertos.

$P(d_1)$	**CC**	**AC**	**MS**	**PP**	**HDx**	**HDy**
0,05	0,276	0,053	0,043	0,059	0,062	0,062
0,10	0,245	0,061	0,065	0,058	0,057	0,056
0,15	0,217	0,065	0,084	0,059	0,061	0,059
0,20	0,182	0,072	0,100	0,061	0,052	0,059
0,25	0,148	0,074	0,106	0,064	0,057	0,061
0,30	0,118	0,072	0,102	0,055	0,051	0,052
0,35	0,088	0,071	0,097	0,062	0,058	0,059
0,40	0,056	0,065	0,094	0,055	0,055	0,053
0,45	0,036	0,073	0,100	0,056	0,057	0,057
0,50	0,033	0,070	0,096	0,056	0,067	0,059

Estos resultados evidencian una vez más la ventaja de utilizar un método de cuantificación en vez de confiar en el método CC para estimar las probabilidades a priori de las clases en un conjunto de datos. Es muy notable que, a pesar de que en la clasificación la tasa de error es muy alta (alrededor del 30 %), los resultados de la cuantificación son razonablemente buenos (MAE alrededor de 0.05-0.06 en cualquier escenario, usando HDy).