COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 108 (2012) 873–881

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Texture and moments-based classification of the acrosome integrity of boar spermatozoa images

Enrique Alegre, Víctor González-Castro, Rocío Alaiz-Rodríguez, María Teresa García-Ordás*

Department of Electrical, Systems and Automatic Engineerings, University of León, 24071 León, Spain

ARTICLE INFO

Article history:
Received 13 May 2010
Received in revised form
28 December 2011
Accepted 11 January 2012

Keywords:
Acrosome integrity
Boar semen
Texture descriptors
Discrete Wavelet Transform
Invariant moments
Neural Networks
k-Nearest Neighbours

ABSTRACT

The automated assessment of the sperm quality is an important challenge in the veterinary field. In this paper, we explore how to describe the acrosomes of boar spermatozoa using image analysis so that they can be automatically categorized as intact or damaged. Our proposal aims at characterizing the acrosomes by means of texture features. The texture is described using first order statistics and features derived from the co-occurrence matrix of the image, both computed from the original image and from the coefficients yielded by the Discrete Wavelet Transform. Texture descriptors are evaluated and compared with moments-based descriptors in terms of the classification accuracy they provide. Experimental results with a Multilayer Perceptron and the *k*-Nearest Neighbours classifiers show that texture descriptors outperform moment-based descriptors, reaching an accuracy of 94.93%, which makes this approach very attractive for the veterinarian community.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Proper semen quality assessment is an important problem in medical and veterinarian research fields. It plays an important role in dealing with human fertility problems or with breed improvement of some species such as boars. The porcine industry is one of the most important fields where it is applied and it is focused on obtaining better individuals for human consumption in each generation.

In the last decade, several computer assisted approaches have been developed to evaluate the quality of semen samples. These approaches were first designed for human semen analysis [1], but they have been currently adapted to other species [2]. These systems are essentially based on parameters such as motility or morphometry [3–6], as they are directly related to the semen quality. However, up to our knowledge, the evaluation of the acrosome integrity of the spermatozoon heads is carried out manually, using stains and there are not any computer assisted tools for that analysis. This manual assessment has several drawbacks such as its high cost in terms of time, its lack of objectivity, or the requirement of specialized veterinarian staff and equipments. Hence, it would be very interesting to get an automatic classification of the acrosomes (intact or damaged). It would only require a high-featured digital camera and a computer, and it would let veterinarian experts save a lot of time when determining the proportion of damaged cells within a sample.

Texture analysis and classification have been applied to biology and medicine in the literature (e.g. to distinguish ulcer

* *Corresponding author*. Tel.: +34 987291989.
E-mail addresses: enrique.alegre@unileon.es (E. Alegre), victor.gonzalez@unileon.es (V. González-Castro), rocio.alaiz@unileon.es (R. Alaiz-Rodríguez), mgaro@unileon.es (M.T. García-Ordás).
0169-2607/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2012.01.004

regions from normal regions in capsule endoscopy images [7], to classify human embryonic stem cell nuclei [8] or to recognize leukocytes [9]), providing quite good results. Other examples are the works carried out by Morales et al., who apply texture analysis to the selection of human embryos for in vitro fertilization [10] or Perner et al., who classify Hep-2 cells into 6 different classes [11]. Sørensen et al. extract quantitative measures of emphysema severity computing Local Binary Patterns from CT images of lungs [12]. All of these works show accuracies around 90%.

One of the most powerful techniques is the multiresolution texture analysis. A successful example is the Discrete Wavelet Transform (DWT), which is widely employed for several applications. For instance, Zhou and Peng use the DWT to extract some features from fly embryo images in order to recognize various gene-expressed structures within them [13], providing accuracies near to 100%. Tsantis et al. [14] extracts some morphological and wavelet-based features from ultrasound images in order to evaluate malignancy risk of thyroid nodules. Quellec et al. proposed in [15] a content-based image retrieval method for diagnosis aid in medical image which compares signatures built from the wavelet transform of the images by means of a distance criterion. Wavelet textural properties are also used for breast cancer analysis applications [16], or in order to analyze the images captured by a wireless capsule endoscope, in the examination of diseases of small bowel [17].

Computer-based systems designed for semen analysis tasks should reliably segment the heads of the spermatozoa [18], extract the patterns which characterize them and finally classify those patterns in order to estimate how many damaged acrosomes contain each semen sample. There are few computer vision works that deal with boar sperm analysis. Furthermore, there are not any commercial tools at all that classify the spermatozoon heads in terms of their membrane integrity, although there are a few experimental works that address this problem.

A classification of the acrosome in terms of its integrity is performed in [19,20], using Learning Vector Quantization (LVQ). This work considers images of boar spermatozoa obtained with an optical phase-contrast microscope and tries to automatically classify single cells as acrosome-intact or acrosome-reacted. This approach uses the gradient magnitude along the outer contour of the sperm head as descriptor. The minimum error rate achieved in this work was 6.8%. Despite this reasonable result for semen quality control, it is very important to improve the accuracy of the classification, according to veterinary experts. Alaiz-Rodríguez et al. use texture descriptors with the aim of estimating the true – and unknown – proportion of damaged cells in a sample [21]. They quantify the unknown a priori probabilities of test sets using the outputs of a classifier trained with an image dataset with class prior probabilities that may not match those of the operational conditions.

Other related – although not similar – proposals are the works carried out by Sanchez et al. [22,23]. They classify the images of the spermatozoa according to its intracellular intensity distribution.

Our proposal is to describe the texture of the acrosomes by computing some first-order statistical and co-occurrence features proposed by Haralick in [24]. These features are computed from both the original image and the coefficients yielded by the DWT, to assess the power of multirresolution texture analysis. We hypothesized that descriptors based on the shape of the head or on its internal intensity would provide good results, but the experiments conducted classifying with Hu [25], Legendre [26] and Zernike [27,28] moments and histogram-based descriptors did not yield good results, as we will show.

The rest of the work is organized as follows: Section 2 describes how the different features have been obtained. The classification and the empirical results are evaluated and compared in Section 3. Finally, some concluding remarks are given in Section 4.

2. Methods

The goal of this work is to classify images of boar spermatozoa in terms of their acrosome integrity. The first step to achieve it is to extract the region of interest (ROI) – the head of each spermatozoon – from the whole image and then, some descriptors are extracted from it.

We have computed some features derived from the co-occurrence matrix proposed by Haralick et al. [24] and, additionally, some first order statistical descriptors both from the ROI and from its wavelet coefficients [29]. We have also extracted some moment-based descriptors in order to compare them with the texture descriptors in terms of the classifier performance. In particular, we have extracted the Hu, Legendre and Zernike moments.

Once the images are described, we classify them by means of two methods: the k-Nearest Neighbours and a Multilayer Perceptron with a log-sigmoid transfer function both in the hidden and in the output layer. Classification results are estimated taking the 70% of the images for training and the remaining 30% of the samples as the testing set.

2.1. *Preprocessing and segmentation*

Using a digital camera connected to a phase-contrast microscope, boar semen images are captured with a resolution of 780×580 pixels. The magnification of the microscope is $100\times$, so each image contains no more than 2 or 3 cells. Hence, most of the spermatozoa come from different takings, which means that illumination is not completely constant, and therefore the method will be robust to light changes. Information about the sample preparation can be found in [30].

The process of acquiring the images is the following: First, we take a snapshot in real color under the fluorescent illumination. Then, we capture another image in grey scale of the same spermatozoa under positive phase contrast illumination, with the sample in the same position. The images have been obtained in CENTROTEC, a veterinarian research centre interested in this problem.

The spermatozoa heads on the phase contrast images are then cropped and labelled as damaged or intact using the information provided by the snapshots taken under fluorescent illumination. Overlapped heads cannot be analyzed, so they are discarded. Fortunately, due to the conditions under

which the sample is obtained, overlapped heads are not common.

After they are labelled, each head is automatically segmented in order to obtain a mask of the region of interest. The segmentation of the intact acrosome heads is very straightforward but the segmentation of the damage acrosome heads is a more complex process. When the acrosome reacts, the membrane surrounding the acrosome fuses with the plasma membrane exposing the contents of the acrosome. The contrast between the spermatozoa head and the background is reduced when the enzymes released surround the acrosome. As a result, it is not easy to segment the damaged acrosome heads without errors. Several classical segmentation methods based on thresholding, region growing and the watershed transform have been tried. Even a new method [18] combining some of the classical ones has been proposed. But, in our case, the best results have been obtained with the following process.

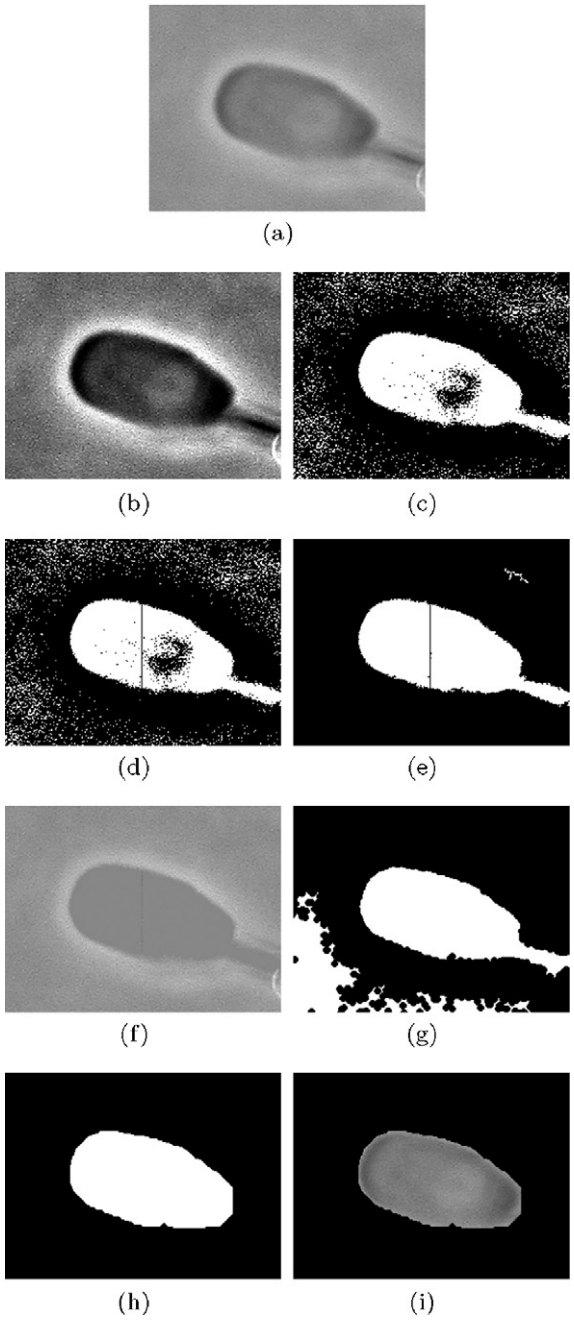

Fig. 1 – Segmentation steps for a single sperm head.

First of all, the image is converted to grey-level (Fig. 1(a)), and its contrast is increased by saturating a 1% of the pixel values at low and high intensities of the grey level image (Fig. 1(b)). Afterwards, it is binarized using a threshold obtained by means of the Otsu's method [31], which produces an image with white regions on a black background (Fig. 1(c)), which is split into two by a division line – horizontal if the number of rows is bigger than the number of columns, or vertical otherwise – yielding an image with at least two regions, each one on each half of the image (Fig. 1(d)). The holes are filled, and the regions whose areas are smaller than the third biggest one are removed (Fig. 1(e)). Later, pixels with value 1 are replaced by the mean of the original grey level pixels which are inside a rectangle centred at the centroid of the image with size $(numRows/2) \times (numCols/2)$ (Fig. 1(f)). Then, a new adjustment of the histogram and a thresholding is done – as it was done before – to obtain a binary image with white regions on a black background. Finally, after performing a dilatation, the removal of small areas, a filling of the holes and a connection of neighbour pixels, we obtain the segmented image (Fig. 1(g)).

Once the head is segmented, successive dilations and erosions are performed to separate it from the tail. Then, the tail and some noise are removed by deleting the regions which are touching the image boundaries and finally, the segmented head is obtained by leaving just the region with the biggest area (Fig. 1(h)). The detailed process is summarized in Algorithm 1.

Algorithm 1. Segmentation and mask of an image.
Require: Image with a boar spermatozoon, I.
Ensure: Segmented head, I_seg.
  if I is in RGB **then**
    Transform I to grey scale
  end if
  $I_{stretch} \leftarrow stretch\_histogram(I)$
  $IBin \leftarrow binarize\_otsu(I_{stretch})$
  $IBin_{divided} \leftarrow divide\_image\_in\_middle(IBin)$
  $IDiv_{fill} \leftarrow fill\_holes(IBin_{divided})$
  $IBin_{rem} \leftarrow remove\_small\_regions(IDiv_{fill})$
  $coordinates \leftarrow find\_white\_pixels(IBin_{rem})$
  $v \leftarrow mean\_central\_region\_size(I, n_{raws}/2, n_{cols}/2)$
  $INew \leftarrow I(coordinates) = v$
  $INew_{stretch} \leftarrow stretch\_histogram(INew)$
  $INew_{bin} \leftarrow binarize\_otsu(INew_{stretch})$
  $I_{dil} \leftarrow dilate\_with\_struct\_element(INew_{bin}, disk_4)$
  $ISeg_{rem} \leftarrow remove\_small\_regions(I_{dil})$
  $ISeg_{fill} \leftarrow fill\_holes(ISeg_{rem})$
  $I_{connected} \leftarrow connect\_neighbour\_pixels(ISeg_{fill})$
  $I_{aux} \leftarrow erode\_with\_struct\_element(I_{connected}, disk_1 8)$
  $I_{aux} \leftarrow erode\_with\_struct\_element(I_{aux}, disk_1 8)$
  $I\_seg \leftarrow remove\_regions\_in\_boundaries(I_{aux})$
  if I_seg have more than 1 region **then**
    $I\_seg \leftarrow remove\_small\_regions(I\_seg)$
  end if
  return I_seg

Fig. 2 – Segmented image of a sperm head with damaged (left) and intact (right) acrosomes.

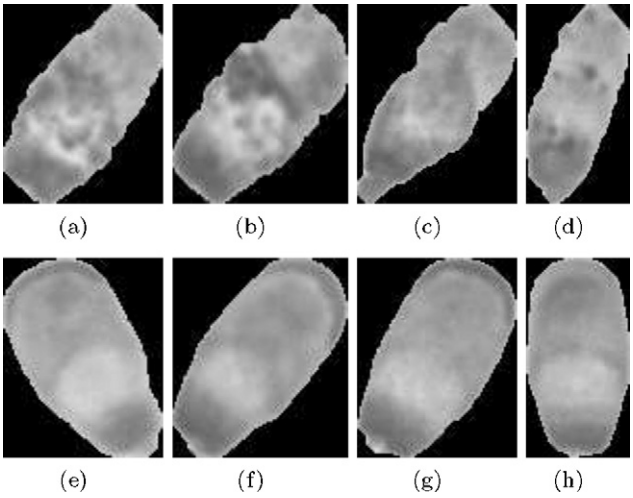

Fig. 3 – Cropped sperm heads with damaged (up) and intact (down) acrosome.

After this process is finished, some heads might be wrongly segmented, due to a fuzzy edge of the head or sample impurities in the image. Therefore, the next step is to discard the wrongly segmented images by using a semi-automatic process: First of all, the regions whose area is below the 30% of the average size of a boar spermatozoon head are eliminated. Afterwards, we calculate the eccentricity of an ellipse that has the same area as the region, and if it is not between 1.4 and 2.6, it is also rejected. These values have been obtained experimentally. Algorithm 2 shows the pseudocode of the automatic discarding method.

Algorithm 2. Automatic discarding of bad segmented images.

Require: Segmented head, *head_seg*. Average area of heads of the set, $Area_{avg}$.
Ensure: **true** if *head_seg* is well segmented, or **false** if not.
  if number of regions of *head_seg* is not 1 **then**
    return false
  else
    $proportion \leftarrow MajorAxis_{ellipse}/MinorAxis_{ellipse}$
    $A_{head} \leftarrow get\_area(head\_seg)$
    if $((proportion \leq 2.6$ and $proportion \geq 1.4)$ and $A_{head} \leq Area_{avg} * 0.3)$
    then
      return true
    else
      return false
    end if
  end if

In the next step, each segmented image is used to obtain a masked one with a black background and the original grey levels inside the head (Fig. 1(i)) [19].

After all this process, a set with 400 damaged and 400 intact spermatozoa heads is available.

Finally, all the images are cropped into its bounding box (Fig. 3), in order to reduce the proportion of black background pixels. We have also tried normalizing the mean grey-level of the whole set of images and rotating the heads so that they would be in horizontal position, with the tail insert point in the same side of the image, but it did not show any improvement. That worsened the results, what may be due to the grey level alterations introduced by the pixel interpolation performed in the heads rotation.

2.2. *Feature extraction*

As we have previously pointed out, texture analysis is used to classify images of boar sperm heads in terms of its membrane integrity. By looking at the aspect of the acrosomes (Fig. 2) it may seem that a morphological approach could be successful. However, the experiments we have carried out following this line did not show good results. Specifically, we have used Hu, Legendre and Zernike moments, among others, to characterize the shape of the acrosomes and we have compared the performance of both classes of descriptors.

According to the textural approach, we have extracted five different descriptors:

- Haralick features extracted from the co-occurrence matrix of the original image.
- First order statistical features extracted from the histogram of the original image.
- Contrast, correlation, energy and homogeneity extracted from the co-occurrence matrix of the original image and the coefficients of the first level of its Discrete Wavelet Transform (DWT).
- A selection of features from the vector of 65 features obtained after computing the Haralick descriptors extracted from the original image and the coefficients of the first sub band of the DWT of the image. This selection have been made by means of a Linear Discriminant Analysis (LDA).
- Mean and standard deviation of the coefficients of the three first sub bands of the DWT of the image.

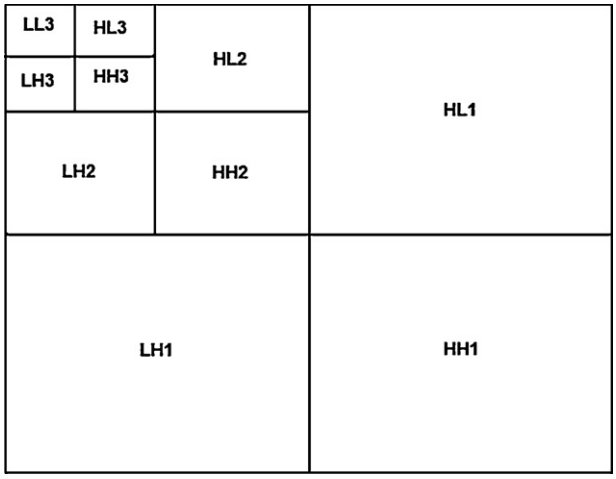

Fig. 4 – Names of the sub-bands of a 3-level wavelet transform.

2.2.1. *Discrete Wavelet Transform*

Information represented by spatial frequencies is often used for texture pattern recognition. Therefore, in this work we have applied the Discrete Wavelet Transform on the images. In particular, we have used the Haar family of wavelets which, in spite of being the simplest, outperforms the Coiflet and Daubechies [32]. The DWT is currently used effectively in signal and image processing because of its frequency domain localization capability. It extracts the high-frequency components of a signal, so that they can be analyzed separately. When the transform is applied on an image, four matrices of coefficients are obtained: approximations and horizontal, vertical and diagonal details (see Figs. 4 and 5). The first one holds almost all the energy of the image, while the other three hold the high frequency details.

2.2.2. *Linear Discriminant Analysis*

Linear Discriminant Analysis is a supervised classification method which assigns an object represented by a feature vector into one of n predetermined classes. It can also be used to find out how much each feature contributes to determine the class membership of the object, which allows us to reduce the length of the patterns with the least possible information loss.

We have used the Linear Discriminant Analysis with the whole set of WCF descriptors to rank the 65 original variables according to their weight in the separation of the classes. Then, the optimal features vector length has been selected for both the k-Nearest Neighbours (kNN) and the Multi-Layer Perceptron (MLP).

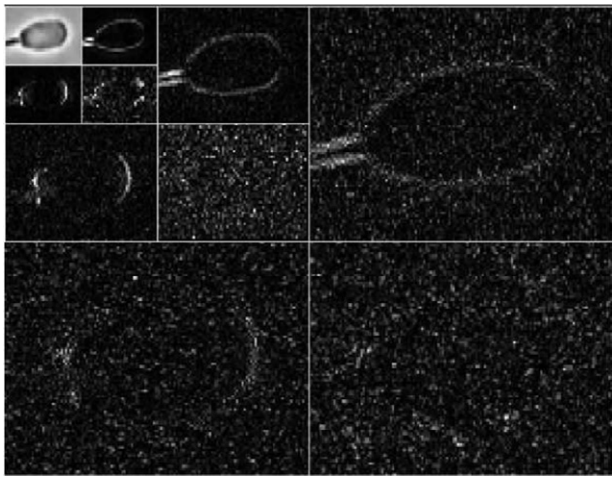

Fig. 5 – 3-Level wavelet transform of a spermatozoon head image.

2.2.3. *First order statistical descriptors*

As it has been previously pointed out, we have computed first order statistics both from the histogram of the original image and from the histograms of each one of the coefficients matrices of the first three DWT subbands.

The first order statistics computed are:

1. Mean:

$$m = \sum_{i=0}^{N-1} iP(x)(i) \tag{1}$$

2. Standard deviation:

$$\sigma = \sqrt{\frac{1}{N^2}\sum_{i=0}^{N-1}(P(x_i) - m)^2} \tag{2}$$

3. Energy:

$$U = \sum_{i=0}^{N-1} P^2(x_i) \tag{3}$$

4. Entropy:

$$e = -\sum_{i=0}^{N-1} P(x_i)\log_2 P(x_i) \tag{4}$$

Therefore, the vector which describes a single image is made up of four features, and we refer to it as *statistical descriptors* (SD).

We have also computed the mean and the standard deviation from the wavelet sub-bands obtained after three splits: LL1, LH1, HL1, HH1, LL2, LH2, HL2, HH2, LL3, LH3, HL3, HH3 (see Figs. 4 and 5). Hence, each image is represented by a 24 features vector. This descriptor has been called WSF (Wavelet Statistical Features) [29].

2.2.4. *Co-occurrence matrix descriptors*

The co-occurrence matrix combines both statistical and structural methods to extract descriptors from an image. The relative frequencies of grey level pairs of pixels separated by a distance d in the orientation θ are combined to make a relative displacement vector (d, θ), which is computed and stored in the so called *grey level co-occurrence matrix* (GLCM). It is used to extract some statistical texture features. We have computed the GLCM in four directions – 0°, 45°, 90° and 135° – averaging the features computed from each of them, and in different distances, in order to compare them.

First of all we have extracted 13 out of the 14 features proposed by Haralick et al. in [24] – all except the *Maximal Correlation Coefficient* – from the GLCMs of the original image. Therefore, we will have a vector of 13 features per image that we have named Haralick13.

We have also computed the same features from the GLCMs of the original image together with the coefficients of the first split of the wavelet transform – LL1, LH1, HL1, HH1 – which yield 65 features. Then, selecting the most relevant ones using the LDA, we have obtained a feature vector named WCF13.

Finally, we have extracted the Haralick features – *contrast*, *correlation*, *energy* and *homogeneity* – [33] from the same GLCMs as the ones used for the WCF13, as done in [29], yielding a vector of 20 features per image, naming this features vector WCF4.

2.2.5. *Moment-based descriptors*

A visual inspection of the images of the acrosomes (i.e. Figs. 2 and 3) suggests that the edge and some shape features would be enough to describe the different kind of acrosome images. In order to assess that, we have computed the Hu, Legendre and Zernike moments from the binary image of the ROI. We have chosen these moments because they are invariant to rotation, translation and scale, and they have been widely used in the literature [34,25–28].

3. Experimental results

In this section the accuracies of the proposed descriptors in the categorization of the acrosome state of the spermatozoon head are evaluated and compared. The two target classes are spermatozoon heads with damaged and with intact acrosomes. The whole dataset has 800 images, 400 of each class.

The classification has been carried out with the *k*-Nearest Neighbours (kNN) and a Multilayer Perceptron. The model accuracy has been estimated using 70-30 random sampling. Both with kNN and the Multilayer Perceptron, the data set has been split into two subsets. The first one takes randomly the 70% of the samples and the classifiers are trained with those descriptors. The second one takes the remaining 30% of the samples and the classifiers are tested with them. This process has been repeated 50 times and the presented results come from the average value for all the runs.

The classifier accuracy, expressed as the overall and for each class hit rate and the standard deviation were computed. For the best classifier, in our case the Neural Network, the area under ROC curve has also been obtained.

3.1. *Classification with k-Nearest Neighbours*

The classification of each image in the test set with kNN is based on finding the *k* nearest elements of the training set and assigning that image to the majority class of its *k* nearest images. The distance metric that has been used in our experiments is the euclidean one. Data were normalized to zero mean and standard deviation equal to one.

Table 1 reports the best accuracy rates of the classification estimated with 70-30 random sampling. The table also shows the number of neighbours, *k*, used for each descriptor.

According to these results, the second order statistical descriptors are the best ones, reaching accuracies of 87.12%, 89.35% and 90.53%, using Haralick13, WCF13 and WCF4, respectively. Therefore, moment-based descriptors, which yield accuracy rates between 61.99% and 76.98%, are clearly outperformed by them. Finally, statistical texture descriptors extracted from the histogram of the original ROI or from its DWT lead to poor results in the classification: 73.10% and 68.18% using WSF and SD, respectively.

These results prove that texture descriptors based on second order statistical features outperform moment-based or histogram-based descriptors, in particular when they are combined with the wavelet transform. It is quite remarkable that WCF4 provides quite balanced accuracy rates for both classes, whereas for other descriptors such as WCF13 or WSF, they are less balanced.

3.2. *Classification with a Multilayer Perceptron*

A Multilayer Perceptron (MLP) has been also used to classify the images. We evaluate a three layer network with a logistic sigmoid activation function both for the hidden and the output layers.

Several combinations of neurons in the hidden layer – 2, 3 or 5 – and different number of training cycles – 200, 300 and 400 – have been assessed with all the descriptors, in order to find the optimal network configuration. Results are estimated using 70-30 random sampling.

Table 1 – Classification accuracy rates for damaged and intact acrosomes with kNN. In brackets, the number of features for each descriptor or the standard deviation for each accuracy.

Descriptor	*k*	Accuracy (%) (overall)	Accuracy (%) (intact)	Accuracy (%) (damaged)
SD (4)	9	68.18 (3.08)	75.88 (4.44)	60.47 (5.19)
WSF (24)	7	73.10 (2.61)	90.52 (3.09)	55.68 (4.34)
Haralick13 (13)	7	87.12 (1.84)	90.17 (2.10)	84.07 (3.12)
WCF4 (20)	**7**	**90.53 (1.67)**	**92.92 (2.29)**	**88.13 (3.16)**
WCF13 (32)	19	89.35 (1.50)	95.05 (1.59)	83.65 (2.68)
Hu (7)	3	71.64 (2.61)	77.75 (3.83)	65.53 (3.99)
Legendre (9)	7	61.99 (2.37)	61.83 (4.23)	62.15 (3.68)
Zernike (9)	25	76.98 (2.36)	78.97 (3.44)	75.00 (3.50)

Rows in bold determine the best result.

Table 2 – Classification accuracy rates for damaged and intact acrosomes with a Multilayer Perceptron. In brackets, the number of features for each descriptor or the standard deviation for each accuracy.

Descriptor	Cycles	Neurons	Hit rate (%)	Hit intact (%)	Hit damaged (%)
SD (4)	300	3	70.68 (2.42)	74.13 (8.52)	67.23 (8.49)
WSF (24)	200	2	86.13 (2.43)	85.67 (4.04)	86.60 (2.85)
Haralick13 (13)	200	5	90.67 (1.68)	89.98 (2.44)	91.35 (2.64)
WCF4 (20)	**400**	**5**	**94.93 (1.56)**	**95.72 (1.87)**	**94.15 (2.76)**
WCF13 (19)	400	3	94.84 (1.37)	95.08 (1.88)	94.60 (2.22)
Hu (7)	400	5	77.81 (2.73)	82.48 (5.42)	73.13 (5.33)
Legendre (9)	400	5	59.19 (3.05)	58.62 (7.39)	59.77 (6.27)
Zernike (9)	400	2	78.30 (2.33)	80.62 (4.37)	75.98 (4.51)

Rows in bold determine the best result.

In Table 2 the best hit rates in the classification are shown, along with the number of hidden layers and training cycles used for each descriptor. In the case of the WCF13 features, we have used 19 features, selected from the ordered list provided by the LDA as shown in Section 2.2.4. Adding more features does not provide better results. These features are: *energy* of the original image and the four wavelet sub-bands, *correlation* of the original image and the sub-bands LL1, HL1 and LH1, *first information measure of correlation* of the original image and the sub-bands LL1 and HL1, *second measure of correlation* of the original image and the sub-band LL1, *homogeneity* of the sub-bands LL1, HL1 and LH1, *sum entropy* of the sub-band LL1 and the *entropy* of the same sub-band.

Once again, the hit rate provided by WCF4 is the best one – 94.93% – when the network has 5 neurons in the hidden layer and training it during 400 cycles. It is closely followed by WCF13, which yields an accuracy of 94.84% with a network with 3 neurons and the same training cycles. These two results show a reduction in the error rate from previous works [35], from 8% to 5.07%.

These accuracies also outperform those obtained using moment-based descriptors – 78.30% in the best case – or statistical descriptors extracted from the histogram. It proves that, opposite to what may be thought, moment-based descriptors or histogram-based features are not suitable for this problem.

In these classification problems the two wrong decisions have no the same cost. Categorizing a spermatozoon with damaged acrosome into the intact class has a higher cost than the opposite error in a semen quality evaluation system. Together with the hit rate, we evaluate the performance of the classifier with the Receiver Operating Characteristics (ROC) curve, which is a more powerful tool [36] for these situations.

Table 3 – AUCs of the descriptors classified with a Multilayer Perceptron.

Descriptor	AUC
WCF4	**0.985**
WCF13	0.980
Haralick13	0.961
WSF	0.938
Zernike	0.850
Hu	0.840
SD	0.781
Legendre	0.633

Rows in bold determine the best result.

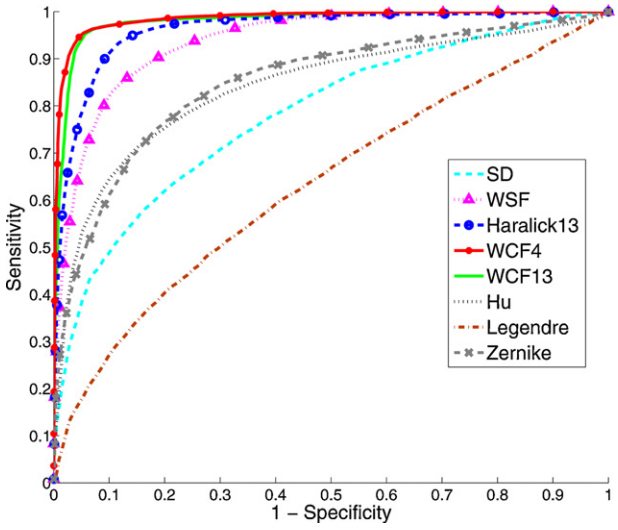

Fig. 6 – ROC curves of the Neural Network classifier with different descriptors.

Fig. 6 shows the ROC curve for each of the features evaluated and the Area Under the Curve (AUC) is shown in Table 3.

Features WCF4 and WCF13 achieve the best relative performance according to the AUC. The value of AUC (0.985) is also very close to the optimal value what makes this results very promising in the practical veterinarian field.

4. Conclusions

We have proposed a method to classify images of boar spermatozoa in terms of its acrosome integrity (damaged or intact). Different texture descriptors have been extracted from the head. We have computed both second order statistical descriptors directly from the original grey-level image – Haralick13 – or from its wavelet coefficients – WCF4 and WCF13 – and first order statistical descriptors – SD and WSF. The very first visual inspection of the samples (Fig. 2) suggests that they could be well characterized using shape features. Because of that we have extracted region descriptors based on moments – Hu, Legendre and Zernike moments.

Classification has been carried out using k-Nearest Neighbours and Neural Networks with a Multilayer Perceptron architecture. Results estimated by 70-30 random sampling show that the second order statistics outperform the first

order statistical descriptors, specially when they are combined with the DWT.

Classifying with *k*-Nearest Neighbours, WCF4 and WCF13 yield an accuracy of 90.53% and 89.35%. The best accuracy (almost 95%), however, has been achieved using Neural Networks. These results outperform the ones achieved in previous works [19,20]. The obtained values also show that moment-based descriptors do not provide good results – the best hit rate is lower than 79%, which proves that they are not suitable for this problem, opposite to what may be thought.

These results represent an improvement, and make this automatic approach based on texture analysis even more attractive for the veterinary community, that demands an automate and accurate discrimination system to replace the manual process. This work leads to further research on: (a) improve the present results, if possible, using different texture descriptors or classifiers, (b) finding new texture descriptors that automatically adapts to the variability of the texture and (c) finding texture descriptors to classify dead or alive sperm heads.

Conflict of interest

This work has been partially supported by the research project DPI2009-08424 from the Spanish Ministry of Education. This support does not involve any conflict of interest.

Acknowledgements

This work has been partially supported by the research project DPI2009-08424 from the Spanish Ministry of Education.

The authors would like to thank CENTROTEC for providing us the semen samples and for their collaboration in the image acquisition.

REFERENCES

[1] L. Ramos, J. Hendriks, P. Peelen, D. Braat, A. Wetzels, Use of computerized karyometric image analysis for evaluation of human spermatozoa, Journal of Andrology 23 (2002) 882–888.

[2] J. Auger, C. Lesaffre, A. Bazire, D. Schoevaert-Brossault, F. Eustache, High-resolution image cytometry of rat sperm nuclear shape, size and chromatin status. Experimental validation with the reproductive toxicant vinclozolin, Reproductive Toxicology 18 (2004) 775–783.

[3] M. Hidalgo, I. Rodriguez, J. Dorado, C. Soler, Morphometric classifcation of Spanish thoroughbred stallion sperm heads, Animal Reproduction Science 103 (2008) 374–378.

[4] J. Verstegen, M. Iguer-Ouada, K. Onclin, Computer assisted semen analyzers in andrology research and veterinary practice, Theriogenology 57 (2002) 149–179.

[5] M.E. Beletti, L.da F. Costa, M.P. Viana, A comparison of morphometric characteristics of sperm from fertile *Bos taurus* and *Bos indicus* bulls in Brazil, Animal Reproduction Science 85 (2005) 105–116.

[6] M.E. Beletti, L.da F. Costa, M.P. Viana, A spectral framework for sperm shape characterization, Computers in Biology and Medicine 35 (2005) 463–473.

[7] B. Li, M. Meng, Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments, Computers in Biology and Medicine 39 (2009) 141–147.

[8] R. Mangoubi, M. Desai, N. Lowry, P. Sammak, Performance evaluation of multiresolution texture analysis of stem cell chromatin, in: Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI, 2008, pp. 380–383.

[9] D. Sabino, L. da Fontoura, E. Rizzatti, M. Zago, A texture approach to leukocyte recognition, RealTimeImg 10 (2004) 205–216.

[10] D. Morales, E. Bengoetxea, P.L. Naga, Selection of human embryos for transfer by Bayesian classifiers, Computers in Biology and Medicine 38 (2008) 1177–1186.

[11] P. Perner, H. Perner, B. Muller, Texture classification based on the Boolean model and its application to hep-2 cells, in: Proceedings of the 16th International Conference on Pattern Recognition, vol. 2, 2002, pp. 406–409.

[12] L. Sørensen, S. Shaker, M. de Bruijne, Quantitative analysis of pulmonary emphysema using local binary patterns, IEEE Transactions on Medical Imaging 29 (2010) 559–569.

[13] J. Zhou, H. Peng, Automatic recognition and annotation of gene expression patterns of fly embryos, Bioinformatics 23 (2007) 589–596.

[14] S. Tsantis, N. Dimitropoulos, D. Cavouras, G. Nikiforidis, Morphological and wavelet features towards sonographic thyroid nodules evaluation, Computerized Medical Imaging and Graphics 33 (2008) 91–99.

[15] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, Wavelet optimization for content-based image retrieval in medical databases, Medical Image Analysis 14 (2010) 227–241.

[16] A. Karahaliou, I. Boniatis, S. Skiadopoulos, F.N. Sakellaropoulos, N.S. Arikidis, E. Likaki, G. Panayiotakis, L. Costaridou, Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications, IEEE Transactions on Information Technology in Biomedicine 12 (2008) 731–738.

[17] J. Bonnel, A. Khademi, S. Krishnan, C. Ioana, Small bowel image classification using cross co-occurrence matrices on wavelet domain, Biomedical Signal Processing and Control 4 (2009) 7–15.

[18] V. González-Castro, E. Alegre, P. Morala-Argüello, S.A. Suarez, A combined and intelligent new segmentation method for boar semen based on thresholding and watershed transform, International Journal of Imaging 2 (2009) 70–80.

[19] N. Petkov, E. Alegre, M. Biehl, L. Sanchez, LVQ acrosome integrity assessment of boar sperm cells, in: Computational Modelling of Objects Represented in Images – Fundamentals, Methods and Applications, Proceedings of the International Symposium on CompImage, Coimbra, Portugal, 20–21 October 2006, 2007, pp. 337–342.

[20] E. Alegre, M. Biehl, N. Petkov, L. Sánchez, Automatic classification of the acrosome status of boar spermatozoa using digital image processing and LVQ, Computers in Biology and Medicine 38 (2008) 461–468.

[21] R. Alaiz-Rodríguez, E. Alegre, V. González-Castro, 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, 2008, pp. 383–388.

[22] L. Sanchez, N. Petkov, E. Alegre, Classification of boar spermatozoid head images using a model intracellular density distribution, in: M. Lazo, A. Sanfeliu (Eds.), Progress in Pattern Recognition, Image Analysis and Applications: Proceedings of the 10th Iberoamerican Congress on Pattern Recognition, CIARP 2005, Lecture Notes in Computer Science, vol. 3773, Springer-Verlag, Berlin Heidelberg, 2005, pp. 154–160.

[23] L. Sanchez, N. Petkov, E. Alegre, Statistical approach to boar semen head classification based on intracellular intensity distribution, in: A. Gagalowicz, W. Philips (Eds.), Proceedings of the International Conference on Computer Analysis of Images and Patterns, CAIP 2005, Lecture Notes in Computer Science, vol. 3691, Springer-Verlag, Berlin Heidelberg, 2005, pp. 88–95.
[24] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics 3 (1973) 610–621.
[25] M. Husoy, Visual pattern recognition by moment invariants, IRE Transactions on Information Theory 8 (1962) 179–187.
[26] C. Chong, P. Raveendran, R. Mukundan, Translation and scale invariants of legendre moments, Pattern Recognition 37 (2004) 119–129.
[27] T. Lin, Y. Chou, A comparative study of zernike moments, in: Proceedings of the IEEE/WIC International Conference on Web Intelligence WI 2003, 2003, pp. 516–519.
[28] A. Ruggeri, S. Pajaro, Automatic recognition of cell layers in corneal confocal microscopy images, Computer Methods and Programs in Biomedicine 68 (2002) 25–35.
[29] S. Arivazhagan, L. Ganesan, Texture classification using wavelet transform, Pattern Recognition Letters 24 (2003) 1513–1521.
[30] L. Sanchez, N. Petkov, E. Alegre, Statistical approach to boar semen evaluation using intracellular intensity distribution of head images, Cellular and Molecular Biology 52 (2006) 38–43.
[31] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (1979) 62–66.
[32] L. Dettori, L. Semler, A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography, Computers in Biology and Medicine 37 (2007) 486–498.
[33] R. Haralick, Statistical and structural approaches to texture, Proceedings of the IEEE 67 (2007) 45–69.
[34] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd. ed., Prentice-Hall, 2008.
[35] M. Gonzalez, E. Alegre, R. Alaiz, L. Sanchez, Acrosome integrity classification of boar spermatozoon images using DWT and texture descriptors, in: Computational Vision and Medical Image Processing: VipIMAGE, 2007.
[36] F.J. Provost, T. Fawcett, R. Kohavi, The case against accuracy estimation for comparing induction algorithms, in: Proceedings of the 15th International Conference on Machine Learning, 1998, pp. 445–453.