LVQ acrosome integrity assessment of boar sperm cells

Nicolai Petkov[1], Enrique Alegre[2], Michael Biehl[1], Lidia Sánchez[2]
[1] *Institute of Mathematics and Computing Science, University of Groningen, The Netherlands*
[2] *Department of Electrical and Electronics Engineering, University of León, Spain*

We consider images of boar spermatozoa obtained with an optical phase-contrast microscope. Our goal is to automatically classify single sperm cells as acrosome-intact (class 1) or acrosome-reacted (class 2). Such classification is important for the estimation of the fertilization potential of a sperm sample for artificial insemination. We segment the sperm heads and compute a feature vector for each head. As a feature vector we use the gradient magnitude along the contour of the sperm head. We apply learning vector quantization (LVQ) to the feature vectors obtained for 152 heads that were visually inspected and classified by a veterinary expert. A simple LVQ system with only three prototypes (two for class 1 and one for class 2) allows us to classify cells with equal training and test errors of 0.165. This is considered to be sufficient for semen quality control in an artificial insemination center.

1 INTRODUCTION

Image processing techniques have been applied to assess the quality of semen samples for therapeutic and fertilization purposes (Verstegen, Iguer-Ouada, and Onclin 2002; Linneberg, Salamon, Svarer, and Hansen 1994). Techniques that had been originally developed for human semen samples have meanwhile been adapted to other species. Most works on boar semen evaluation focus on measuring concentration and motility of spermatozoa or on detecting sperm head shape abnormalities, such as double heads, macro heads etc. Image processing techniques are deployed to automate such analysis in Computer Aided Sperm Analysis (CASA) systems (Suzuki, Shibahara, Tsunoda, Hirano, Taneichi, Obara, Takamizawa, and Sato 2002). The use of sperm motility (Quintero, Rigaub, and Rodríguez 2004) presents several disadvantages, such as sensitivity to temperature changes and unclear relation to fertility. In the case of sperm morphology, assessment is based on the state of the sperm cell structure: head, middle piece and tail, on computing morphometric measures to detect head shape abnormalities (Rijsselaere, Soom, Hoflack, Maes, and de Kruif 2004; Beletti, Costa, and Viana 2005; Ostermeier, Sargeant, Yandell, and Parrish 2001) and on detecting droplets in tails. Analysis of intracellular density distribution has been used in (Sánchez, Petkov, and Alegre 2005b; Sánchez, Petkov, and Alegre 2005a; Biehl, Pasma, Pijl, Sanchez, and Petkov 2006).

Veterinary experts believe that sperm fertility is related to the state of the acrosome, a cap-like structure that develops over the anterior half of the spermatozoon's head. It has its own membrane and contains enzymes. As the sperm approaches the oocyte, an acrosome reaction takes place during which the anterior head plasma membrane of the sperm fuses with the outer membrane of the acrosome, exposing the contents of the acrosome. The released enzymes are required for the penetration of sperm through a layer of follicular (cumulus) cells that encase the oocyte. The acrosome reaction also renders the sperm capable of penetrating through the zona pellucida (an extracellular coat surrounding the oocyte) and fusing with the egg. For these reasons, veterinary experts believe that a semen sample with a high fraction of acrosome-reacted sperm has low fertilizing capacity and cannot be used for artificial insemination.

The traditional techniques to assess acrosome integrity, such as visual inspection by veterinary experts or staining, are time consuming and have relatively high costs. Despite the broadly recognized importance of acrosome integrity evaluation in semen quality assessment, we are not aware of any image processing work on this topic. In this work we propose a new method to assess acrosome integrity based on the automatic analysis of grey level images acquired with a phase contrast microscope. Our approach is based on the observation that there are some characteristic differences in the gradient magnitude profiles along

Figure 1: Boar semen sample images acquired using a phase-contrast microscope.

the contours of acrosome-intact vs. acrosome-reacted sperm heads. We extract a feature vector from the grey level image of a cell head and use this vector to classify the cell as acrosome-intact or acrosome-reacted by comparing it with prototype feature vectors. We determine these prototype vectors by applying Learning Vector Quantization (LVQ) to a training set of feature vectors obtained from the images of sperm heads that were marked by a veterinary expert by means of visual inspection as acrosome-intact or acrosome-reacted.

2 Vectorization

2.1 Pre-processing and segmentation

Boar sample images were captured by means of a digital camera Nikon Coolpix 5000 mounted on an optical phase-contrast microscope Nikon Eclipse. The magnification used was $\times 100$ and the dimensions of each sample image were 2560×1920 pixels. A boar semen sample image contains a number of spermatozoa which can vary widely from one sample to the next and also present different orientations (Fig. 1).

Sperm head images were obtained manually by cropping from a boar semen sample image. These images were visually inspected by a veterinary expert and the spermatozoa were classified as acrosome-intact or acrosome-reacted. For each obtained image (Fig. 2a), we segment the sperm head by converting the image to a binary image using Otsu's method (Otsu 1979) and applying several morphological operations (dilations and erosions), Fig. 2b. We use the contour of the sperm head binary mask in the following. We also localize the point where the middle piece, from which the tail develops, connects to the head. This point is used as a reference point in the following.

2.2 Scale-dependent gradient computation and vectorization

Let $f(x,y)$ and $g_\sigma(x,y)$ be a grey level distribution in an input image and a two-dimensional (2D) Gaussian function with standard deviation σ, respectively. The x- and y-components of the scale-dependent gradient of $f(x,y)$ are defined as convolutions of $f(x,y)$ with the x- and y-derivatives of $g_\sigma(x,y)$, respectively:

$$\nabla_{\sigma,x} f = f * \frac{\partial g_\sigma}{\partial x}\,, \qquad \nabla_{\sigma,y} f(x,y) = f * \frac{\partial g_\sigma}{\partial y}\,. \quad (1)$$

This approach to gradient computation (Schwartz 1950) has been shown to reduce noise and discretisation effects (Canny 1986; Tagare and de Figueiredo 1990; Grigorescu, Petkov, and Westenberg 2004). In the following we use the magnitude $M_\sigma(x,y)$ of the scale-dependent gradient:

$$M_\sigma(x,y) = \sqrt{\left(\nabla_{\sigma,x} f(x,y)\right)^2 + \left(\nabla_{\sigma,y} f(x,y)\right)^2}\,. \quad (2)$$

Fig. 3 shows intracellular density distribution images typical for acrosome-intact and acrosome-reacted boar spermatozoa, respectively, and Fig. 4 shows the corresponding gradient magnitude images.

Next we determine the gradient magnitude along the cell head boundary as a 1D function of the boundary curve length from the reference point to a given contour point in a clock-wise direction. For a given contour point we take the local maximum of the gradient in a 5×5 neighborhood of that point. The resulting discrete function is a vector that we re-size by interpolation to a uniform length of 40 elements. We also normalize this vector by dividing it by its largest element, Fig. 5. The vectors obtained for different sperm heads are used for LVQ.

3 Analysis by Learning Vector Quantization

3.1 LVQ training

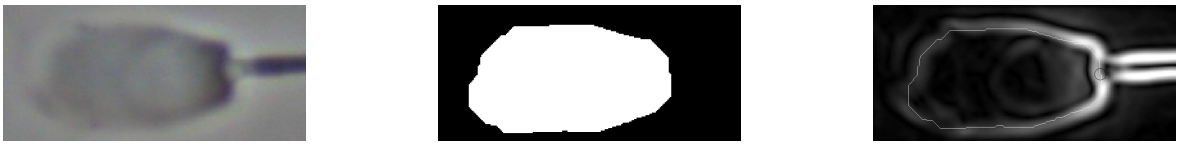
In the following we apply LVQ for the distance-based classification of the data. LVQ has been used in a variety of problems due to its flexibility and conceptual clarity (Kohonen 1995; Neural Networks Research Centre 2002).

We apply a heuristic training algorithm, so-called LVQ1 (Kohonen 1995), in order to determine typical representatives of each class from a (sub-) set of labeled training data $I\!D = \{\boldsymbol{\xi}^\mu, S_T^\mu\}_{\mu=1}^P$ Here, the $\boldsymbol{\xi}^\mu \in I\!R^N$ $(N = 40)$ are the vectors of gradient magnitudes along the contour as described in the previous section. The corresponding class membership provided by the veterinary experts is denoted as

$$S_T^\mu = \begin{cases} +1 & \text{if } \boldsymbol{\xi}^\mu \text{ was labeled as } \textit{acrosome-intact (class 1)} \\ -1 & \text{if } \boldsymbol{\xi}^\mu \text{ was labeled as } \textit{acrosome-reacted (class 2)} \end{cases} \quad (3)$$

Fig. 5 displays three example profiles from each of the two classes.

In the set of prototypes $\left\{\boldsymbol{w}^1, \boldsymbol{w}^2, \ldots, \boldsymbol{w}^M\right\}$ a vector $\boldsymbol{w}^j \in I\!R^N$ is supposed to represent data with class membership $S^j \in \{+1, -1\}$. These assignments, as well as the number of prototypes are specified prior to training and remain unchanged.



(a) (b) (c)

Figure 2: (a) Grey level image of a sperm head and a part of the middle piece protruding from the head. (b) Binary image of the sperm head obtained by thresholding and subsequent morphological processing. (c) Contour of the binary head mask (white line) superimposed on the gradient magnitude image. The reference point at the base of the middle piece is marked with a black circle.

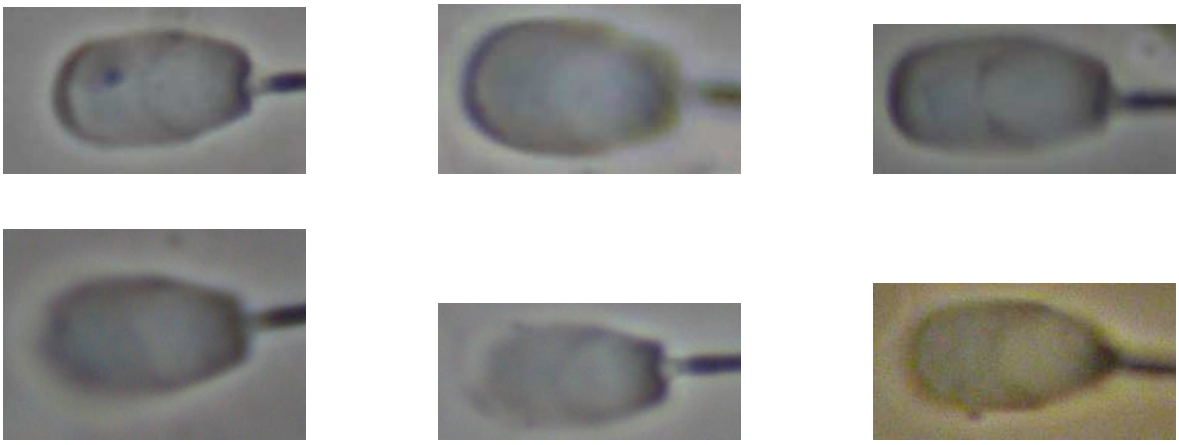


Figure 3: Grey level images characteristic of acrosome-intact (upper row) and acrosome-reacted (lower row) boar spermatozoa.

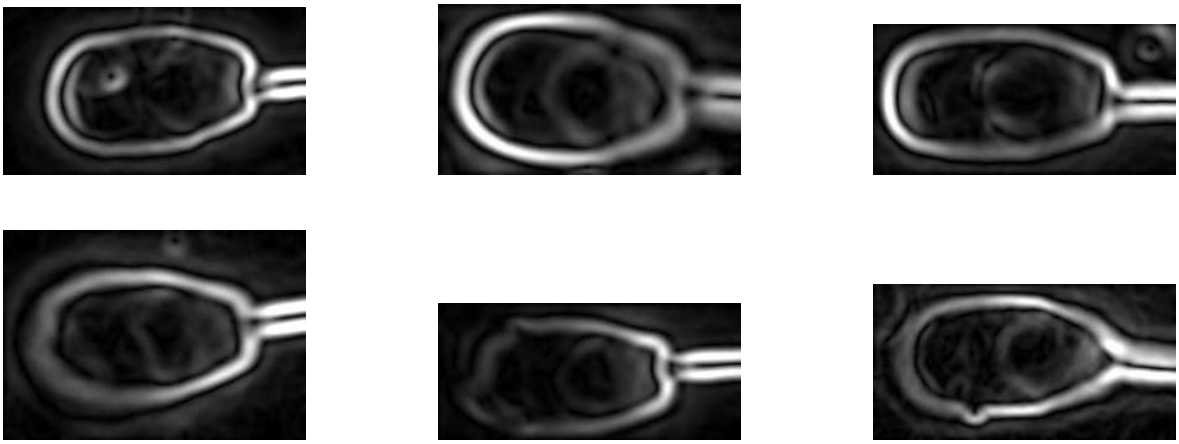


Figure 4: Gradient magnitude images characteristic of acrosome-intact (upper row) and acrosome-reacted (lower row) boar spermatozoa. The parameter σ used to compute the scale-dependent gradient is set to 0.03 of the cell head length.

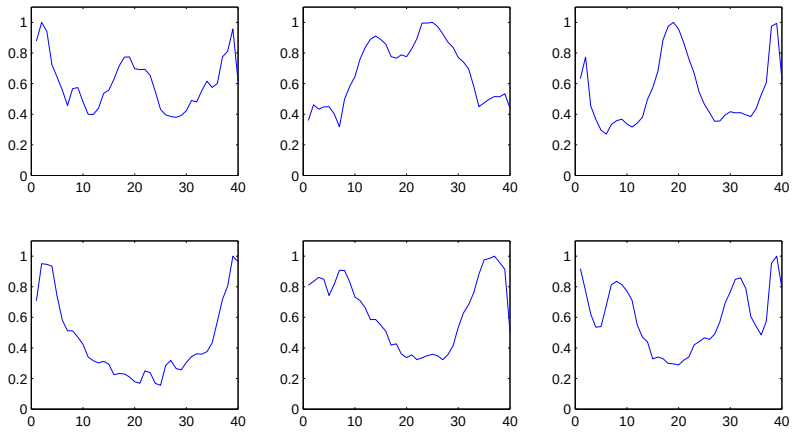


Figure 5: Example gradient magnitude profiles along the head boundary $\boldsymbol{\xi} \in I\!R^{40}$ from the class of acrosome-intact (upper row) and acrosome-reacted (lower row) spermatozoa. The displayed profiles correspond to the images shown in Figs. 3 and 4. These discrete 1D functions represent the vectors used for LVQ.

At each time step t of the iterative training procedure, one example $\{\boldsymbol{\xi}^\mu, S_T^\mu\}$ is selected randomly from $I\!D$. We evaluate the distances $d(j,\mu)$ of $\boldsymbol{\xi}^\mu$ from all current prototype vectors $\boldsymbol{w}^j(t)$. Here, we restrict ourselves to the standard (squared) Euclidean measure

$$d(j,\mu) \;=\; (\boldsymbol{\xi}^\mu - \boldsymbol{w}^j)^2 \;=\; \sum_i \left(\xi_i^\mu - w_i^j(t)\right)^2 . \qquad (4)$$

Next we identify the minimal distance $d(J,\mu)$ among all prototypes and the corresponding *winnner*

$$\boldsymbol{w}^J(t) \;\text{ with }\; d(J,\mu) = \min_k \{d(k,\mu)\} . \qquad (5)$$

In LVQ1, only this winner is updated according to

$$\boldsymbol{w}^J(t+1) = \boldsymbol{w}^J(t) + \eta(t)\, \left[S_T^\mu\, S^J\right]\, \left(\boldsymbol{\xi}^\mu - \boldsymbol{w}^J(t)\right) . \qquad (6)$$

Hence, the update is towards or away from the actual input $\boldsymbol{\xi}^\mu$, if the class labels of winner and example agree or disagree, respectively.

In the following studies, prototype vectors are initialized close to the mean of vectors $\boldsymbol{\xi}^\mu$ in the corresponding class. In order to avoid exactly coinciding $\boldsymbol{w}^j(0)$, small random displacements from the class conditional means are performed.

The learning rate $\eta(t)$ controls the step width of the iteration. It gradually decreases in the course of learning following a schedule of the form

$$\eta(t) = \eta_o\,/(1 + a\,t) \quad \text{with } \; a \; \text{ such that } \; \eta(t_f) = \eta_f . \qquad (7)$$

Results presented in the next section were obtained with a schedule that decreases the learning rate from $\eta_o = 0.1$ to $\eta_f = 0.0001$ in 1000 sweeps through the training set, i.e. $t_f = 1000\,P$.

After training, the system parameterizes a distance-based classification scheme: any data $\boldsymbol{\xi}$ is assigned to the class S^J which is represented by the closest prototype.

3.2 Cross–validation

To obtain estimates of the performance after training we employ eight-fold cross-validation: We split the set of 152 available training data (105 from class 1 and 47 from class 2) randomly into disjoint subsets $I\!D_i, i = 1, 2 \ldots 8$, of equal size. For a given number of prototypes, each of eight identically designed LVQ systems, $n = 1, 2 \ldots 8$, is trained from the set $\cup_{i \neq n} I\!D_i$ containing $P = 133$ examples. Then, $I\!D_n$ serves as a test set to evaluate the performance on novel data.

In the following, ϵ_{train} denotes the fraction of misclassified example data, obtained after training and on average over the eight systems. The test error ϵ_{test} quantifies the averaged performance with respect to the test set. Analogously we evaluate the class-specific test errors $\epsilon_{test}^{(1)}$ and $\epsilon_{test}^{(2)}$ as well as the training errors $\epsilon_{train}^{(1)}$ and $\epsilon_{train}^{(2)}$ with respect to only class 1 or class 2 data, respectively.

Although the training sets $I\!D_i$ overlap, the corresponding standard deviations provide a rough measure of the expected variation of the classifier with different realizations of the training set. The main purpose of the cross-validation scheme is to compare the performance of different LVQ schemes, i.e. systems with different numbers of prototypes.

4 Results

We have performed LVQ1 training following the above described scheme for systems with m and n prototypes representing class 1 and class 2 data, respectively. The table 1 summarizes the observed error

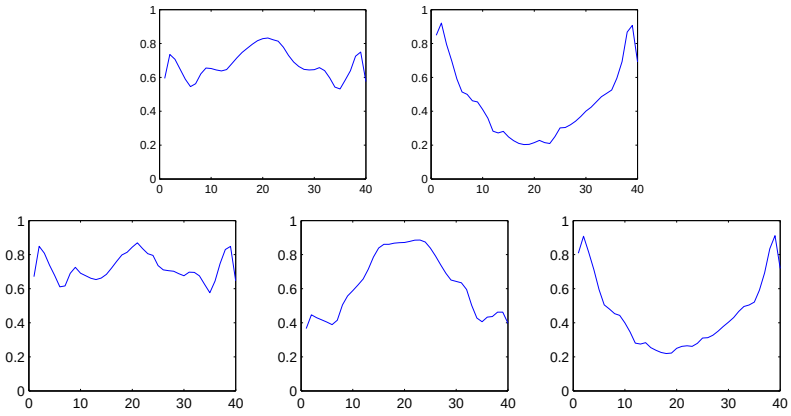


Figure 6: **Upper panel:** Two prototype profiles as obtained in LVQ1 from one set of 133 examples. The left and right prototype represent class 1 of acrosome-intact and class 2 of acrosome-reacted spermatozoa, respectively. **Lower panel:** Three prototype profiles as obtained in LVQ1 from one set of 133 examples. The leftmost and center prototypes represent class 1 (of acrosome-intact spermatozoa), whereas the rightmost profile corresponds to class 2 (of acrosome-reacted spermatozoa).

Table 1: Training and test errors observed using m and n prototypes of the class 1 and 2, respectively

(m,n)	training error	test error
$(1,1)$	$\epsilon_{train} = 0.167\,(0.01)$	$\epsilon_{test} = 0.165\ (0.09)$
	$\epsilon^{(1)}_{train} = 0.061\,(0.01)$	$\epsilon^{(1)}_{test} = 0.054\ (0.09)$
	$\epsilon^{(2)}_{train} = 0.404\,(0.02)$	$\epsilon^{(2)}_{test} = 0.393\ (0.13)$
$(2,1)$	$\epsilon_{train} = 0.165\,(0.01)$	$\epsilon_{test} = 0.165\ (0.07)$
	$\epsilon^{(1)}_{train} = 0.057\,(0.01)$	$\epsilon^{(1)}_{test} = 0.055\ (0.06)$
	$\epsilon^{(2)}_{train} = 0.408\,(0.04)$	$\epsilon^{(2)}_{test} = 0.423\ (0.24)$
$(1,2)$	$\epsilon_{train} = 0.160\,(0.01)$	$\epsilon_{test} = 0.177\ (0.05)$
	$\epsilon^{(1)}_{train} = 0.070\,(0.01)$	$\epsilon^{(1)}_{test} = 0.080\ (0.07)$
	$\epsilon^{(2)}_{train} = 0.359\,(0.04)$	$\epsilon^{(2)}_{test} = 0.398\ (0.23)$

measures in the simplest LVQ configurations. Numbers in parentheses give the corresponding standard deviations observed in the eight-fold cross-validation.

We would like to point out that even in the minimal setting ($m = n = 1$), the outcome of LVQ1 differs significantly from a naive representation of the two classes by their respective mean profiles. If we replace the two LVQ prototypes by class-conditional mean vectors, we obtain an average test and training error of $\epsilon_{test} \approx \epsilon_{train} \approx 0.184$. LVQ1 yields a better performance because the supervised training procedure detects and emphasizes the discriminative features in the data.

Figure 6 (upper panel) displays two LVQ prototypes as obtained in one of the training runs. The lower panel shows an example outcome of LVQ1 with two class 1 and one class 2 prototype ($m = 2, n = 1$). Note how the class 1 prototypes have *specialized* to represent the two predominant types of acrosome-intact profiles which are also apparent in Figure 5 (low or high intensity around ξ_1 and ξ_{40}).

In general, test and training error are lower with respect to class 1, which reflects greater fluctuations in the class 2 data. Employing more prototypes for the second class yields a more balanced classifier, however, the overall test performance remains unchanged or even degrades. Note, for instance, that a system with one class 1 but two class 2 prototypes $(m = 1, n = 2)$ clearly performs worse than the minimal configuration with $m = n = 1$.

When increasing the number of prototypes, i.e. the complexity of the LVQ system, we observe a decrease of the overall training error ϵ_{train}. However, it is accompanied by a (moderate) increase of the test error which signals *over-fitting*: While the particular training set can be represented in greater detail, the generalization ability deteriorates. At the same time, the variance of the outcome tends to be larger in the overly complex systems.

Finally, we obtained essentially the same results with vectors of 20 elements that were obtained by taking every second element of the vectors of 40 elements used above.

5 Summary and outlook

We extract feature vectors from boar sperm head images and use these vectors for LVQ training and classification as acrosome-intact (class 1) or acrosome-reacted (class 2). As a feature vector we use the gradient magnitude along the contour of the sperm head. A simple LVQ system with only three prototypes (two for class 1 and one for class 2) allows us to classify cells with equal training and test errors of 0.165. This

is considered to be sufficient for semen quality control in an artificial insemination center.

In future investigations we intend to increase the number of features by considering also the grey level distribution in the sperm head interior. We furthermore intend to apply more sophisticated cost function based schemes such as *Generalized Learning Vector Quantization* (GLVQ) as suggested in (Sato and Yamada 1995). Modified distance measures and *relevance learning*, see e.g. (Hammer and Villmann 2002), could be applied in order to obtain a better understanding of this classification task and to extract the most relevant features.

REFERENCES

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

Biehl, M., P. Pasma, M. Pijl, L. Sanchez, and N. Petkov (2006). Classification of boar sperm head images using learning vector quantization. In M. Verleysen (Ed.), *Proc. European Symposium on Artificial Neural Networks (ESANN), Brugge, April 26-28, 2006*, pp. 545–550. d-side, Evere, Belgium.

Canny, J. F. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence 8*(6), 679–698.

Grigorescu, C., N. Petkov, and M. A. Westenberg (2004). Contour and boundary detection improved by surround suppression of texture edges. *Image and Vision Computing 22*(8), 609–622.

Hammer, B. and T. Villmann (2002). Generalized relevance learning vector quantization. *Neural Networks 15*, 1059–1068.

Kohonen, T. (1995). *Self-organizing maps*. Springer, Berlin.

Linneberg, C., P. Salamon, C. Svarer, and L. Hansen (1994). Towards semen quality assessment using neural networks. In *Proc. IEEE Neural Networks for Signal Processing IV*, pp. 509–517.

Neural Networks Research Centre (2002). *Bibliography on the Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ)*. Helsinki University of Technology.

Ostermeier, G., G. Sargeant, T. Yandell, and J. Parrish (2001). Measurement of bovine sperm nuclear shape using Fourier harmonic amplitudes. *J. Androl. 22*, 584–594.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics 9*, 62–66.

Quintero, A., T. Rigaub, and J. Rodríguez (2004). Regression analyses and motile sperm subpopulation structure study as improving tools in boar semen quality analysis. *Theriogenology 61*, 673 – 690.

Rijsselaere, T., A. V. Soom, G. Hoflack, D. Maes, and A. de Kruif (2004). Automated sperm morphometry and morphology analysis of canine semen by the Hamilton-Thorne analyser. *Theriogenology 62*(7), 1292–1306.

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

Sánchez, L., N. Petkov, and E. Alegre (2005b). Statistical approach to boar semen head classification based on intracellular intensity distribution. In A. Gagalowicz and W. Philips (Eds.), *Proc. Int. Conf. on Computer Analysis of Images and Patterns, CAIP 2005, Lecture Notes in Computer Science*, Volume 3691, pp. 88–95. Springer-Verlag Berlin Heidelberg.

Sato, A. and K. Yamada (1995). Generalized learning vector quantization. In G. Tesauro, D. Touretzky, and T. Leen (Eds.), *Advances in Neural Information Processing Systems*, Volume 7, pp. 423–429. MIT Press.

Schwartz, L. (1950). *Théorie des Distributions*. Vol. I, II of Actualités scientifiques et industrielle. L'Institute de Mathématique de l'Université de Strasbourg.

Suzuki, T., H. Shibahara, H. Tsunoda, Y. Hirano, A. Taneichi, H. Obara, S. Takamizawa, and I. Sato (2002). Comparison of the sperm quality analyzer IIC variables with the computer-aided sperm analysis estimates. *International Journal of Andrology 25*, 49–54.

Tagare, H. and R. de Figueiredo (1990). On the localization performance measure and optimal edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence 12*(12), 1186–1190.

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