Proceedings Paper:
Van Raemdonck, L.E.M., Ur-Rehman, A., Davila-Garcia, M.L. et al. (2016) Human Sperm Morphology Analysis for Assisted Reproduction Techniques. In: Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015. , 06-08 Oct 2015, Bonn, Germany. IEEE.

https://doi.org/10.1109/SDF.2015.7347714

An Algorithm for Morphological Classification of Motile Human Sperm

Lore EM Van Raemdonck*, Ata-ur-Rehman*, M. Luisa Davila-Garcia*,
Lyudmila Mihaylova*, Robert F. Harrison*, Allan Pacey**
*Department of Automatic Control and Systems Engineering, University of Sheffield, UK
{lemvanraemdonck1, a.ur-rehman, mdavilagarcia1, l.s.mihaylova, r.f.harrison}@sheffield.ac.uk
**Academic Unit of Reproductive and Developmental Medicine, Department of Human Metabolism, University of Sheffield, UK
a.pacey@sheffield.ac.uk

***Abstract*—The analysis of human sperm as part of infertility investigations or assisted conception treatments is a labor intensive process reliant upon the skill of the observer and as such prone to human error. Therefore, there is a need to develop automated systems that can adequately assess the concentration, motility and morphology of live sperm. This paper presents an algorithm for analyzing the morphology of motile sperm. Techniques for eliminating the background, segmentation of the cells and template matching techniques are used to analyze the morphology in two stages: first stage eliminates the immotile cells and at the second stage the morphology of the motile cells is analyzed. Results are presented with real sperm samples recorded in the andrology lab at the University of Sheffield. The performance of the proposed algorithm is analyzed in terms of accuracy and complexity. The proposed algorithm demonstrates high accuracy under variable conditions.**

***Index Terms*—template matching, structural similarity measure, motility grading, sperm morphology**

I. Introduction

The assessment of sperm concentration, motility and morphology play an important role in the analysis of human semen [1]. The semen analysis can give an indication of fertility potential in males when appropriate quality-control measures are followed [2]. It has been found that the analyzing the percentage of sperm with normal morphology has great importance in male fertility diagnosis [3]. Compared to the sperm with abnormal morphology, sperm with normal morphology has greater ability to fertilize eggs [4]. However, a reliable, accurate and repeatable assessment of morphology analysis has not been achieved yet. Assessment of sperm morphology is a challenging task [5] and the World Health Organization requires it to be done by a trained scientist observing dead sperm which have been stained using histological dyes. This way of examining semen is prone to high variability [6] and the subjectivity of the laboratory staff [7]. The manual method to analyze semen depends on the natural ocular tiredness, the limits of visual perception and the experience of the health care professional.

The World Health Organization (WHO) has published guidelines and recommendation about how sperm morphology assessment must be conducted. However, following these guidelines to manually analyze the sperm morphology is difficult and therefore these guidelines are followed by only a few laboratories [8]. Moreover, because the sperm observed in the laboratory are dead the results obtained are not representative of what might be seen of live sperm observed in semen if that measurement were technically possible.

A sperm has a head, a middle piece and tail. Measures such as stain content, length, width, perimeter, area, and arithmetically derived combinations had been considered in the literature in order to classify sperm according to their morphology. A human sperm may contain different kinds of malformations. Areas of defects stated by WHO are: head defects, neck and mid piece defects, principal piece defects and excess residual cytoplasm (ERC) [1]. Nevertheless, there are only two classification groups: normal and abnormal.

Sperm is difficult to recognize using a computer system over video due to its fast movement and partial or full occlusion by other sperm cells. Besides sperm, a semen sample has other cells such as bacteria, epithelial cells, leukocytes, isolated sperm heads and tails which makes the analysis process even harder. Other challenges related to the manual sperm analysis are:

- repeatability of sperm analysis results.
- workload of the counting operators.

In order to get precision in sperm morphology assessment computer-assisted sperm morphometry analysis (CASMA) systems have been developed. However, currently used CASMA method are time consuming and require individual selection of each sperm [9]. Automated sperm analysis overcomes issues such as variability of the assessment results [7], subjectivity, precision and reproducibility [1].

This paper proposes an image processing algorithm for an automated morphology analysis of motile sperm. Morphology of the motile cells only is analyzed with the help of video frames from a recorded video of sperm samples.

Template matching techniques are developed that rely on several performance measures such as: the Bhattacharyya distance [10], the Structural Similarity Measure [11], and the correlation coefficients. Templates are formed for both normal an abnormal cells by using well established template shapes, from the World Health Organization [12]. These preliminary formed templates use image features such as intensity and edges and are compared with the current video frame, based

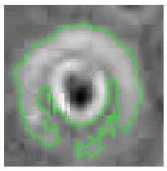
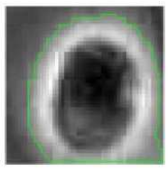
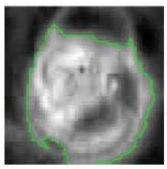
Template	Other 1	Other 2	Other 3
Image (blob)			
MaxInt	139	171	207
MinInt	80	83	129
MeanInt	117.62	126.56	168.16
Area	1354 pixels	1600 pixels	1787 pixels
MajorAxL	46.17 pixels	46.03 pixels	49.91 pixels
MinorAxL	43.10 pixels	45.28 pixels	46.36 pixels

Fig. 1. Templates of non-sperm cells.

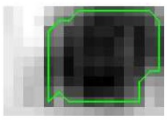
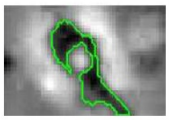
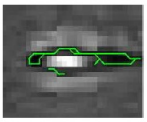
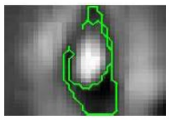
Template	Immobile 1	Immobile 2	Immobile 3	Immobile 4
Image (blob)				
MaxInt	96	96	96	96
MinInt	61	69	63	65
MeanInt	76.30	85.20	80.76	84.88
Area	108 pixels	247 pixels	42 pixels	129 pixels
MajorAxL	13.30 pixels	39.44 pixels	27.36 pixels	29.05 pixels
MinorAxL	10.59 pixels	12.59 pixels	3.87 pixels	13.03 pixels

Fig. 2. Templates of immotile sperm.

on the template matching. Inference is performed based on the similarity between the templates and the current video frames.

The remainder of the paper is organized as follows: Section II describes the proposed algorithm, Section III presents the extensive experimental results and finally, conclusions are presented in Section IV.

II. Proposed Algorithm

This paper proposes an image processing algorithm for an automated morphology analysis of motile sperm. The proposed method first distinguishes sperm from non-sperm cells (*e.g.* epithelial cells and isolated sperm heads or tails) present in semen. Three non-sperm cell templates are used, shown in Figure 1. Aim of the algorithm is to detect and classify motile and immotile cells. Four immotile cell templates are used for this distinction, shown in Figure 2. Finally, the morphology of motile cells is analyzed. The output of the proposed algorithm is a count of the number of motile cells with normal morphology.

A. Eliminating Immotile cells

The first stage of the algorithm eliminates the immotile cells and keeps only the motile cells for morphology analysis.

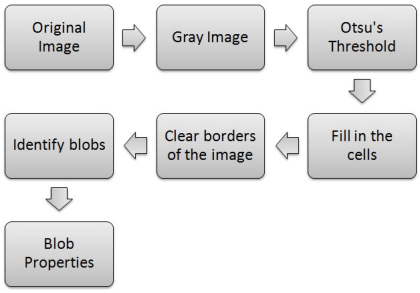


Fig. 3. Flowchart cells detection, where one blob represents one cell.

Figure 3 shows the flowchart of this stage of the algorithm. At this stage we select a frame from a recorded video for morphology analysis. The image is first converted to a gray scale format. To eliminate the background from the foreground *Otsu's threshold* [13] is used. After separating the background pixels we get foreground pixels which consist of multiple blobs (connected pixels). One detected blob is considered as one cell. Finally we measure the following properties of the detected blob:

- *Area* which is the number of pixels in the blob.
- *Bounding box* which is the smallest rectangle which can define the blob.
- *Filled image* which gives the binary image of the bounding box.
- *Major axis length* which calculates the length of the major axis of the ellipse around the blob.
- *Minor axis length* which calculates the length of the minor axis of the ellipse around the blob.
- *Pixel List* which gives a vector with pixel values of the blob.

Properties of the detected blobs are used to compare them with the templates of both motile and immotile cells. The library in Figure 2 represents the four cell templates that are used to match and classify cells in sample video frame. After getting the blobs of target cells and template cells we remove the background noise to improve the results. Each blob having area smaller than 25 pixels (based on experiments) is removed from the original frame. Properties of the each blob in the video frame are compared with the properties of motile, immotile and non-sperm cells. Figure 4 shows the flowchart of the cell classification procedure for classifying the cells into motile, immotile and non-sperm cells. We consider five different criteria to compare the target blobs with the template blobs. First we compare the areas, if the area of the selected blob is within $\pm 10\%$ of the area of the template, it is considered as a match and labeled as 1. If it does not satisfy the criterion it is not considered as a match and labeled as 0. The second criterion consists in the major and minor axes of the blob. The same principle is applied to the major and minor axes as in the case of the area.

Next, the following three criteria are used: the correlation (CORR) coefficient, the structural similarity measure (SSIM) and the Bhattacharyya distance (BHAT). These measures are used in binary classifiers in the literature to compare images [14], [15], [16], [17]. The CORR coefficient measures how two images resemble between -1 and 1, with 1 highly correlated

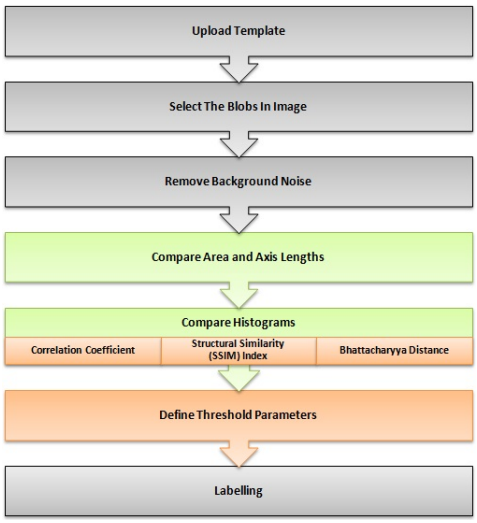


Fig. 4. Flowchart of the cell classification procedure.

and -1 non-correlated. Equation (1) shows the formula to calculate how two images are correlated:

$$CORR = \frac{\sum_m \sum_n (A_{mn} - \mu_A)(B_{mn} - \mu_B)}{\sqrt{(\sum_m \sum_n (A_{mn} - \mu_A)^2)(\sum_m \sum_n (B_{mn} - \mu_B)^2)}} \quad (1)$$

where A_{mn} represents one pixel of the template blob of size (m x n) and μ_A is the mean of the blob, similarly, B_{mn} is pixel of the target blob of size (m x n) and μ_B is its mean. The SSIM index measures how similar or dissimilar two images are based on the mean μ_a, the standard deviation σ_a and the covariance $\sigma_{a,b}$. Equation (2) shows the formula [17] how to calculate the SSIM index

$$SSIM(A,B) = (\frac{2\mu_A\mu_B}{\mu_A^2 + \mu_B^2})(\frac{2\sigma_A\sigma_B}{\sigma_A^2 + \sigma_B^2})(\frac{\sigma_{A,B}}{\sigma_A\sigma_B}) \quad (2)$$

where

$$\mu_a = \frac{1}{L}\sum_{k=1}^{L} a_k$$

$$\sigma_a = \sqrt{\frac{1}{L-1}\sum_{k=1}^{L}(a_k - \mu_a)^2}$$

$$\sigma_{a,b} = \frac{1}{L-1}\sum_{k=1}^{L}(a_k - \mu_a)(b_k - \mu_b)$$

with L pixels. The BHAT distance [15] measures the similarity between the normalized histograms of a template and a target between 0 and 1. Equation (3) shows the formula for calculating the Bhattacharyya distance $d(h_{temp}, h_{targ})$:

$$d(h_{temp}, h_{targ}) = \sqrt{1 - \rho(h_{temp}, h_{targ})} \quad (3)$$

where $\rho(h_{temp}, h_{targ})$ is the Bhattacharyya measure [15]. The larger the value ρ (≈ 1) or the smaller d (≈ 0) gets, the more similar are the histograms of the template and target.

The measured CORR and SSIM values are both between -1 and 1 and for the BHAT distance between 0 and 1. Next in the flowchart (Figure 4) is defining a threshold parameter to distinguish if a template matches the target or not. Each binary classifier is tested with a low ($= 0.50$), medium ($= 0.65$) and high ($= 0.79$) threshold parameter. Based on the experimental results (Table I) the best performing threshold parameter, equal to 0.65, is chosen for each binary classifier. If the measured classifier has a value above the threshold it is considered as a match and gets label 1, if it is below the threshold it is considered as a mismatch and gets label 0.

TABLE I
THE CLASSIFICATION METRIC RESULTS WITH VARYING THRESHOLD PARAMETERS FOR THREE CROPPED VIDEOS ($\mu \pm \sigma$).

Low Threshold Parameter = 0.50						
Metrics	Immotile			Motile		
ACC (%)	90 ± 1.6	89 ± 2.8	88 ± 2.6	96 ± 1.6	95 ± 4.2	97 ± 0.8
$F_{measure}$ (%)	81 ± 0.8	82 ± 0.8	82 ± 0.7	78 ± 12	78 ± 13	79 ± 18
PRE (%)	75 ± 1.0	75 ± 0.8	75 ± 0.8	65 ± 18	65 ± 18	67 ± 23
Recall (%)	89 ± 2.1	88 ± 2.3	90 ± 2.7	97 ± 1.4	97 ± 1.2	97 ± 1.3
SENS (%)	89 ± 2.1	88 ± 2.3	90 ± 2.7	97 ± 1.4	97 ± 1.2	97 ± 1.3
SPEC (%)	89 ± 1.4	88 ± 1.9	87 ± 1.3	95 ± 2.3	93 ± 1.9	96 ± 1.4
Medium Threshold Parameter = 0.65						
Metrics	Immotile			Motile		
ACC (%)	91 ± 0.8	91 ± 0.7	91 ± 0.9	99 ± 0.2	99 ± 0.2	99 ± 0.3
$F_{measure}$ (%)	81 ± 1.2	81 ± 1.4	82 ± 1.0	79 ± 11	81 ± 13	79 ± 15
PRE (%)	76 ± 1.5	75 ± 1.6	76 ± 1.4	66 ± 15	69 ± 17	67 ± 16
Recall (%)	88 ± 2.0	89 ± 1.9	89 ± 1.3	97 ± 0.5	97 ± 0.6	97 ± 0.4
SENS (%)	88 ± 2.0	89 ± 1.9	89 ± 1.3	96 ± 1.0	96 ± 1.7	96 ± 1.6
SPEC (%)	90 ± 0.8	91 ± 1.0	90 ± 1.6	99 ± 0.2	99 ± 0.2	99 ± 0.3
High Threshold Parameter = 0.79						
Metrics	Immotile			Motile		
ACC (%)	75 ± 14	72 ± 9.5	67 ± 11	96 ± 1.2	97 ± 1.3	97 ± 1.0
$F_{measure}$ (%)	81 ± 5.5	81 ± 2.4	82 ± 1.7	78 ± 15	83 ± 19	80 ± 17
PRE (%)	77 ± 3.5	76 ± 3.4	77 ± 2.1	66 ± 19	73 ± 21	69 ± 20
Recall (%)	86 ± 6.4	86 ± 3.0	87 ± 2.7	96 ± 1.7	96 ± 2.1	95 ± 2.3
SENS (%)	86 ± 6.4	86 ± 3.0	87 ± 2.7	96 ± 1.7	96 ± 2.1	95 ± 2.3
SPEC (%)	89 ± 1.6	89 ± 1.6	87 ± 3.5	99 ± 0.2	99 ± 0.9	99 ± 0.2

The blobs in the original image are now labeled with 1 or 0 and the number of motile, immotile and non-sperm cells are determined. Before analyzing the morphology of the sperm cells, only the motile cells are kept and all immotile and non-sperm cells are eliminated.

B. Cell Morphology Detection

Once the motile cells are selected, the shape of these cells is analyzed to classify them into normal and abnormal cells. A similar criterion which was used by WHO for classification of cells into motile and immotile is used. Again templates of normal and abnormal cells are manually selected and the properties of target and template blobs are compared to decide whether the cell is normal or abnormal. Seven motile cell templates are used, which are shown in Figure 5.

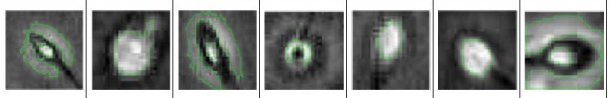
Template	Tapered	Round	Tapered	Small Acrosomal Area	Normal	Small Round	Pyriform
Image (blob)							
MaxInt	200	205	194	177	201	187	157
MinInt	76	144	123	133	147	139	121
MeanInt	143.87	169.19	138.29	152.56	170.1	165.58	134.14
Area	492 – 61	242	491 – 45	129	135	193	400 - 45
MajorAxL	37.45 – 12.73	21.04	39.68 – 10	17.52	15.4	17.9	33.93 – 8.81
MinorAxL	34.01 – 6.51	16.14	34.12 – 6.34	16.17	11.57	13.95	31.82 – 6.75

Fig. 5. Templates of motile sperm.

The same five criteria as explained in previous subsection are used here. To make the proposed algorithm more user friendly a graphical user interface in matlab is developed whose image is shown below in Figure 6.

III. Experimental Results

The proposed algorithm is tested on video frames taken from a recorded video which were recorded in the andrology lab of the Academic Unit of Reproductive and Developmental Medicine (University of Sheffield). The sample was obtained from a healthy volunteer who gave informed consent for his sample to be used for research purposes. All procedures had been approved by the University Research Ethics Committee (Ref: SMBRER293). The image resolution is set to 2040×1086. Results in Figure 7 show the classification of the cells into motile, immotile and non-sperm cells. For each classification the accuracy, precision, recall, $F_{measure}$, sensitivity and specificity are calculated.

A. Classification Metrics

By labeling the blobs in the image, the cells are converted in binary format and classified. To analyze the performance of the classification, metrics such as accuracy, precision, recall, $F_{measure}$, sensitivity and specificity are calculated. Those classification metrics are based on true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The TP is identified as the cells that are correctly labeled as 1 and TN is identified as the cells that are correctly labeled as 0. FP and FN are respectively identified as the cells that are incorrectly labeled as 1 and incorrectly labeled as 0. The accuracy of a measurement is the degree of closeness of the measurement of a quantity to that quantity's true value. The accuracy is calculated as [18]

$$Accuracy = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}. \quad (4)$$

where $P = TP+FP$ and $N = TN+FN$. Sensitivity is also known as the true positive rate (TPR) or recall. It measures the proportion of TP which are correctly identified as positives. The positives (P) are determined by FN and TP. The sensitivity is calculated as [18]

$$Sensitivity = \frac{TP}{P} = \frac{TP}{TP+FN} = Recall. \quad (5)$$

The false negative rate (FNR) is defined as $1 - Sensitivity$, that is the miss rate or the false alarm. The closer TPR to 1, the better it is and the closer FNR to 0, the better it is. Specificity is also known as the true negative rate (TNR). It measures the proportion of TN which are correctly identified as negatives. The negatives (N) are determined by the FP and TN. The formula to calculate the specificity is given in Equation (6)

$$Specificity = \frac{TN}{N} = \frac{TN}{FP+TN} \quad (6)$$

The false positive rate (FPR) is defined as $1 - Specificity$, the fall-out or fail to detect. The closer TNR to 1, the better it is and the closer FPR to 0, the better it is. The precision of a measurement is the degree of repeatability of a measurement. The precision is calculated as [19]

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

The $F_{measure}$ combines the recall and the precision [19]

$$F_{measure} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

B. Classification Metric Results

Accuracy, precision, recall, $F_{measure}$, sensitivity and specificity are calculated for each cell classification. The classification metrics are determined based on how well the classified sperm match the template. Three videos with around 130 frames are analyzed. The results are shown in Table II. The mean and standard deviation, $\mu \pm \sigma$, are given in percentage for each classification metric and the total number of sperm that are classified as immotile, motile or non-sperm cells.

TABLE II
THE CLASSIFICATION METRIC RESULTS AND TOTAL NUMBER OF CELLS FOR THREE DIFFERENT VIDEOS ($\mu \pm \sigma$).

Metrics	Immotile			Non-sperm Cells			Motile		
ACC (%)	91 ± 0.8	91 ± 0.7	91 ± 0.9	85 ± 2.3	84 ± 2.2	83 ± 2.1	99 ± 0.2	99 ± 0.2	99 ± 0.3
$F_{measure}$ (%)	81 ± 1.2	81 ± 1.4	82 ± 1.0	77 ± 2.9	76 ± 3.0	77 ± 2.3	79 ± 11	81 ± 13	79 ± 15
PRE (%)	76 ± 1.5	75 ± 1.6	76 ± 1.4	87 ± 2.8	86 ± 3.6	88 ± 1.9	66 ± 15	69 ± 17	67 ± 16
Recall (%)	88 ± 2.0	89 ± 1.9	89 ± 1.3	69 ± 3.5	69 ± 3.2	69 ± 2.6	97 ± 0.5	97 ± 0.6	97 ± 0.4
SENS (%)	88 ± 2.0	89 ± 1.9	89 ± 1.3	69 ± 3.5	69 ± 3.2	69 ± 2.6	97 ± 0.5	97 ± 0.6	97 ± 0.4
SPEC (%)	90 ± 0.8	91 ± 1.0	90 ± 1.6	92 ± 1.6	92 ± 1.4	91 ± 1.4	99 ± 0.2	99 ± 0.2	99 ± 0.3
TOTAL	113 ± 19	128 ± 22	130 ± 19	28 ± 5	37 ± 5	40 ± 5	255 ± 21	260 ± 22	260 ± 20

The closer the metric values to 100%, the better the discrimination of the cells. An arbitrary subdivision for interpreting the classification metric values is proposed:

$$\text{Classification Metric} = \begin{cases} 50\% - 60\% & \text{No Discrimination} \\ 60\% - 70\% & \text{Poor Discrimination} \\ 70\% - 80\% & \text{Acceptable Discrimination} \\ 80\% - 90\% & \text{Good Discrimination} \\ 90\% - 100\% & \text{Excellent Discrimination} \end{cases}$$

1) Immotile Sperm: Table II shows that the classification metric results can be considered as good - excellent for accuracy, $F_{measure}$, recall, sensitivity and specificity of immotile sperm. For the precision the result is around 76% which is an acceptable discrimination. This was expected looking at the results of the binary classifiers for immotile sperm in Table I.

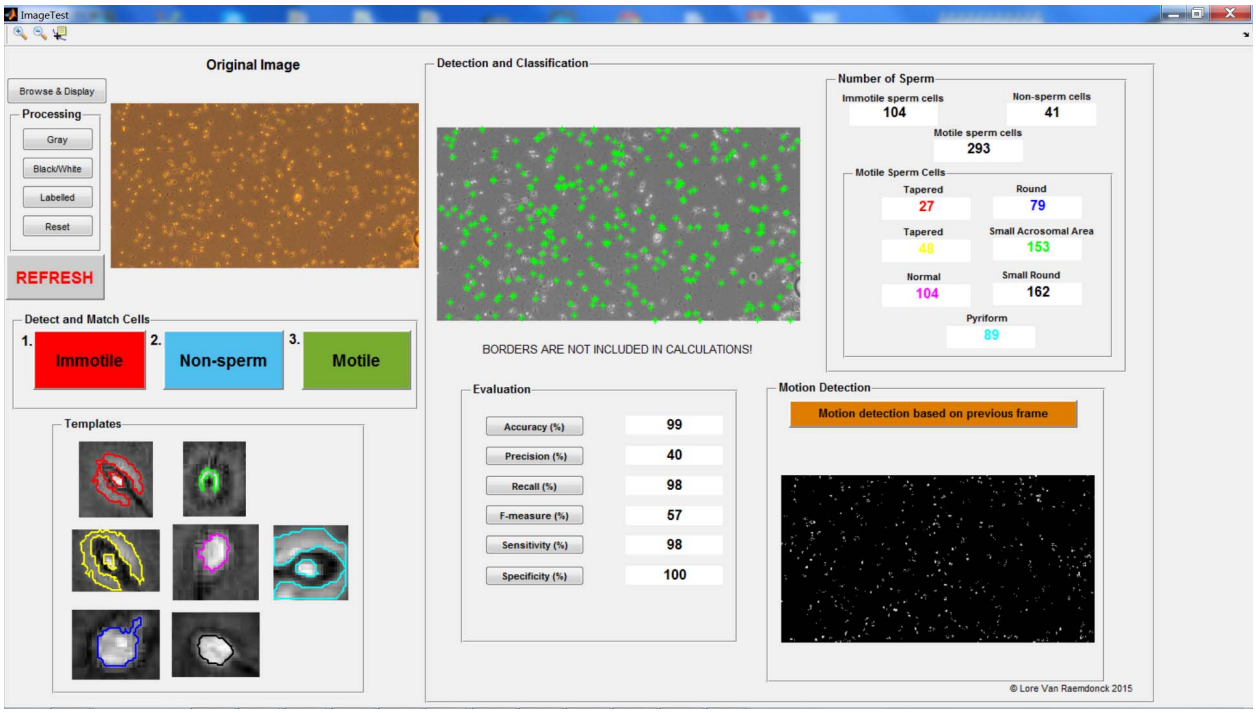


Fig. 6. Graphical user interface for the morphology analysis of motile sperm.

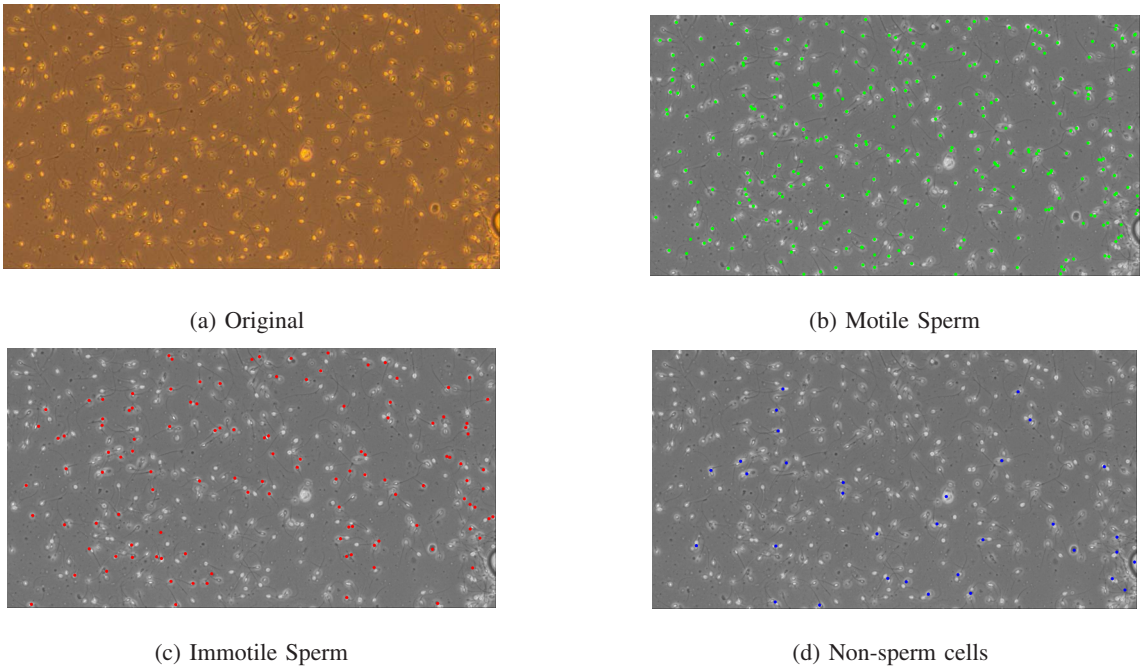
These results can be interpreted as; the template of immotile sperm has a high ability to match the actual immotile sperm cells in the image, which refers to the accuracy. However, the template of immotile sperm has a lower ability to consistently match those actual immotile sperm cells. This refers to a lower reproducibility and precision. Discriminating immotile sperm just based on their morphology is hard.

2) Non-sperm Cells: Classifying non-sperm cells is based on matching the area and axis lengths of the templates and thus not on binary classifiers. Table II shows that the classification metric results can be considered as good - excellent for accuracy, precision and specificity of non-sperm cells. For the recall and sensitivity it is rather considered as a poor discrimination. This means that 69% of the non-sperm cells are correctly labeled as non-sperm cells, thus 31% of the non-sperm cells are not detected. The specificity is around 91%, which means that 91% of the cells are correctly identified as not non-sperm cells, while 9% of those cells are incorrectly labeled as non-sperm cells. The library template exists of three non-sperm cells, by extending this library the results may improve.

3) Motile Sperm: Table II shows that the classification metric results can be considered as excellent, except for precision, which influences the $F_{measure}$. The mean percentage of precision for the three videos is around 67% with a larger standard deviation compared to other metric results in the table. The larger standard deviation can be explained by the use of seven templates to classify motile sperm into normal and abnormal. Depending on the image the classification within normal and abnormal sperm differs. As mentioned earlier in the paper, the borders are not included in the classification morphology algorithm. As the classification in this case is based on motile cells, it is possible that in one frame motile sperm cells are located in the center of the image and the next frame it is located at the border. This influences the reproducibility of the experiment, thus a lower precision and $F_{measure}$. Therefore, the template of motile sperm has a bad ability to consistently match the actual motile sperm cells. However, the closeness of the measured value to the true value is high an accuracy of 99%.

IV. Conclusions

A computer assisted sperm analysis algorithm is proposed for accurately and efficiently analyzing the morphology of sperm. Techniques for eliminating the background, segmentation of the cells and template matching techniques are successfully used to first classify the cells into motile and immotile cells and secondly the motile cells are classified into normal and abnormal cells. The performance of the proposed algorithm is evaluated and accurate classification of cells is shown.



(a) Original

(b) Motile Sperm

(c) Immotile Sperm

(d) Non-sperm cells

Fig. 7. Classification of cells into motile, immotile and non-sperm categories.

ACKNOWLEDGMENT

We are grateful to the support from the UK Engineering and Physical Sciences Research Council (EPSRC) for the support via the Bayesian Tracking and Reasoning over Time (BTaRoT) grant EP/K021516/1. We also acknowledge The National Council of Science and Technology (CoNaCYT) for their financial support to the project.

REFERENCES

[1] WHO, "WHO laboratory manual for the examination and processing of human semen," *World Health Organization*, vol. 30, 2009.

[2] A. Pacey, "Is quality assurance in semen analysis still really necessary? a view from the andrology laboratory." *Human reproduction (Oxford, England)*, vol. 21, no. 5, pp. 1105–9, May 2006.

[3] D. S. Guzick, J. W. Overstreet, P. Factor-Litvak, C. K. Brazil, S. T. Nakajima, C. Coutifaris, S. A. Carson, P. Cisneros, M. P. Steinkampf, J. A. Hill, D. Xu, and D. L. Vogel, "Sperm morphology, motility, and concentration in fertile and infertile men." *The New England journal of medicine*, vol. 345, no. 19, pp. 1388–1393, 2001.

[4] A. Morgentaler, R. D. Powers, M. Y. Fung, M. M. Alper, and D. H. Harris, "Sperm morphology and in vitro fertilization outcome: A direct comparison of world health organization and strict criteria methodologies," *Fertility and Sterility*, vol. 64, no. 6, pp. 1177–1182, 1995.

[5] D. Riddell, A. Pacey, and K. Whittington, "Lack of compliance by uk andrology laboratories with world health organization recommendations for sperm morphology assessment." *Human reproduction (Oxford, England)*, vol. 20, no. 12, pp. 3441–5, Dec. 2005.

[6] A. Agarwal and R. K. Sharma, "Automation is the key to standardized semen analysis using the automated sqa-v sperm quality analyzer." *Fertility and sterility*, vol. 87, no. 1, pp. 156–62, Jan. 2007.

[7] M. L. W. J. Broekhuijse, E. Soštarić, H. Feitsma, and B. M. Gadella, "Additional value of computer assisted semen analysis (casa) compared to conventional motility assessments in pig artificial insemination." *Theriogenology*, vol. 76, no. 8, pp. 1473–86.e1, Nov. 2011.

[8] W. Ombelet, H. Pollet, E. Bosmans, and A. Vereecken, "Results of a questionnaire on sperm morphology assessment," *Human Reproduction*, vol. 12, no. 5, pp. 1015–1020, May 1997.

[9] J. L. Yániz, S. Vicente-Fiel, S. Capistrós, I. Palacín, and P. Santolaria, "Automatic evaluation of ram sperm morphometry," *Theriogenology*, vol. 77, no. 7, pp. 1343–1350, 2012.

[10] A. Bhattacharayya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of Calcutta Mathematical Society*, vol. 35, pp. 99–110, 1943.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[12] W. H. Organization, "Laboratory manual for the examination of human semen and sperm-cervical mucus interaction," *Cambridge University Press*, pp. 14–17, 1999.

[13] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions On Systems And Cybernetics*, vol. 20, pp. 62–66, January 1979.

[14] E. Bailey, N. Fenning, S. Chamberlain, L. Devlin, J. Hopkisson, and M. Tomlinson, "Validation of sperm counting methods using limits of agreement." *Journal of andrology*, vol. 28, no. 3, pp. 364–73, 2007.

[15] P. Brasnett, L. Mihaylova, D. Bull, and N. Canagarajah, "Sequential Monte Carlo tracking by fusing multiple cues in video sequences," *Image and Vision Computing*, vol. 25, pp. 1217–1227, August 2007.

[16] P. Kathiravan, J. Kalatharan, G. Karthikeya, K. Rengarajan, and G. Kadirvel, "Objective sperm motion analysis to assess dairy bull fertility using computer-aided system - a review," *Reproduction in Domestic Animals*, vol. 46, no. 1, pp. 165–172, Feb. 2011.

[17] A. Loza, L. Mihaylova, and D. Bull, "Structural similarity-based object tracking in multimodality surveillance videos," *Machine Vision and Applications*, vol. 20, pp. 71–83, 2009.

[18] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, 1997.

[19] D. M. W. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *International Journal of Machine Learning Technology*, vol. 2, no. 1, pp. 37–63, 2011.