



# Automatic detection and segmentation of sperm head, acrosome and nucleus in microscopic images of human semen smears

Fariba Shaker, S. Amirhassan Monadjemi <sup>\*</sup>,  
Ahmad Reza Naghsh-Nilchi

Department of Artificial Intelligence, Faculty of Computer Engineering, University of Isfahan, Isfahan, 81746, Iran

## ARTICLE INFO

### Article history:

Received 4 July 2015

Received in revised form

19 April 2016

Accepted 22 April 2016

### Keywords:

Sperm morphology

Sperm detection

Sperm head segmentation

Active contours

Hue histogram

Tail detection

## ABSTRACT

**Background and objective:** Manual assessment of sperm morphology is subjective and error prone so developing automatic methods is vital for a more accurate assessment. The first step in automatic evaluation of sperm morphology is sperm head detection and segmentation. In this paper a complete framework for automatic sperm head detection and segmentation is presented.

**Methods:** After an initial thresholding step, the histogram of the Hue channel of HSV color space is used, in addition to size criterion, to discriminate sperm heads in microscopic images. To achieve an improved segmentation of sperm heads, an edge-based active contour method is used. Also a novel tail point detection method is proposed to refine the segmentation by locating and removing the midpiece from the segmented head. An algorithm is also proposed to separate the acrosome and nucleus using morphological operations. Dice coefficient is used to evaluate the segmentation performance. The proposed methods are evaluated using a publicly available dataset.

**Results:** The proposed method has achieved segmentation accuracy of 0.92 for sperm heads, 0.84 for acrosomes and 0.87 for nuclei, with the standard deviation of 0.05, which significantly outperforms the current state-of-the-art. Also our tail detection method achieved true detection rate of 96%.

**Conclusions:** In this paper we presented a complete framework for sperm detection and segmentation which is totally automatic. It is shown that using active contours can improve the segmentation results of sperm heads. Our proposed algorithms for tail detection and midpiece removal further improved the segmentation results. The results indicate that our method achieved higher Dice coefficients with less dispersion compared to the existing solutions.

© 2016 Elsevier Ireland Ltd. All rights reserved.

<sup>\*</sup> Corresponding author. Department of Artificial Intelligence, Faculty of Computer Engineering, University of Isfahan, Isfahan, 81746, Iran. Tel.: +98 9130913010; fax: +98 3136631727.

E-mail address: [monadjemi@eng.ui.ac.ir](mailto:monadjemi@eng.ui.ac.ir) (S.A. Monadjemi).

<http://dx.doi.org/10.1016/j.cmpb.2016.04.026>

0169-2607/© 2016 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Infertility is affecting about one of every ten couples [1]. It can have a negative effect on the quality of a couple's life and causes social and psychological problems [2–5]. The cause of almost half of the infertility cases is male factor [6]. To diagnose male infertility or subfertility, and to determine the treatment therapy, semen analysis is conducted. Sperm morphology assessment is one of several steps in semen analysis, in which the shape and size of sperm parts are examined, and the percentages of normal and abnormal sperms are calculated. Sperm morphology assessment of spermatozoa, when performed accurately and according to the WHO 2010 (World Health Organization) manual for the analysis of human semen, can be a robust determinant of sperm function/dysfunction [7]. Moreover, it will indicate the source of the male infertility and help in predicting the outcome of assisted reproductive techniques [7]. However, manual morphology assessment, besides being time consuming, is error prone and subjective [8]. Research has shown high inter and intralaboratory variability between manual results [9]. Therefore, there have been numerous studies in automatic evaluation of spermatozoa condition, applied both to human and to animal spermatozoa.

A spermatozoon consists of three main parts: *head*, *midpiece* and *tail* (Fig. 1), in which the head is divided into *acrosome* and *nucleus*. Abnormalities can happen in each one of these parts, but the most important ones are the abnormalities of the head [10]. To automatically detect the abnormalities of the head, the first step is the segmentation of the head from the background and to its constituent parts, namely, acrosome and midpiece. It is shown that the contour information of the sperm head is important for improving description and classification of the sperm head [11]. Therefore, accuracy is an important factor in extracting the sperm head contour.

Automatic segmentation of sperm head has been considered in a number of previous works. Park et al. used Hough transform to locate and segment the sperm head [12]. The boundary of the sperm head was approximated with an ellipse with five parameters. Nafisi et al. used size, elongation and the existence of tail to detect the sperm heads [13]. Then they fitted an ellipse to the extracted head and considered that ellipse as the head. Since these methods eventually use a best fitting ellipse as the segmented sperm head, the subtle variations in the shape of the sperm head, which affect their classification, will be missed. Gonzalez-Castro et al. used a combination of Otsu thresholding and watershed transform to segment the boar spermatozoa's head [14]. However, they validated their results only by visual inspection, which makes the results unreliable and subjective.

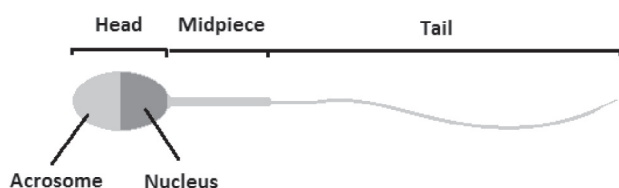


Fig. 1 – Different parts of a spermatozoon.

Works done by Carrillo et al., Bijar et al. and Chang et al. represent more precise segmentation of sperm head. Carrillo et al. used  $n$ 'th level thresholding followed by building growing masks [15,16]. These masks are created using prior information about the sperm's shape and help in the process of refining  $n$  different regions. Then these regions are fused together using some morphological operations. They used size and histogram analysis to distinguish the sperm from non-sperm particles. Bijar et al. used a method based on a Bayesian classifier to classify the image pixels into three classes of background, nucleus and acrosome-midpiece [17–19]. They used Adaptive Mixture Models (AMM) and Markov random field (MRF) to obtain and upgrade the class conditional probability density function (CCPDF), and a priori probability of each class. The method proposed by Bijar et al. suffers from a weakness in distinguishing the sperms from non-sperm particles. The only criterion in their algorithm is a minimum size for the particles. When there are artifacts and debris in images, the class distributions will be affected, which will be a source of error. Chang et al. provided a framework for detecting and segmenting human sperm heads based on clustering methods and combining  $L^*a^*b^*$  and RGB color spaces [20]. A method called *tail removal* is applied to the resulting binary image to remove sperm tails. They also proposed an algorithm to determine to which direction the sperm head points. This information is used to expand the initial head mask in a proper direction. One of the main contributions of Chang et al. was the introduction of the Gold-standard dataset [21]. This dataset has paved the way to make the evaluation of new algorithms more practical. To the best of our knowledge, Chang method is the state-of-the-art in detection and segmentation of sperm heads that is tested against a publicly available dataset.

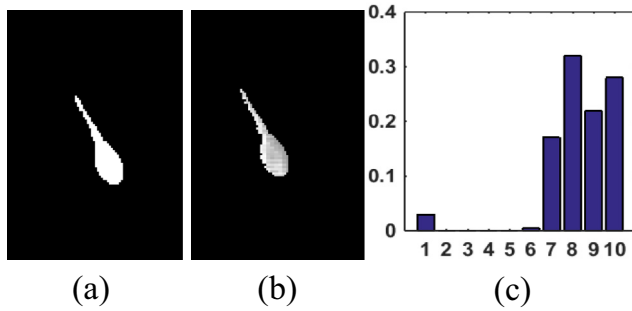
In this paper, we propose a segmentation method based on active contours. Our next contribution is a novel method for recognizing the position of tail, which uses a combination of curvature and edge clues. This tail point is used to locate and remove the midpiece from the segmented head. To recognize sperm heads from other materials present in microscopic images, we propose to use the histogram of Hue channel of HSV color space in addition to size criterion. The rest of the paper is organized as follows. In Section 2 our method is explained in detail. Then the evaluation criteria are presented in Section 3. The results and discussion are presented in Sections 4 and 5 respectively. Finally, we conclude the study in Section 6.

## 2. Materials and methods

Our method is composed of four major steps.

### 2.1. Extracting the initial head masks

Since the sperm heads are darker than the background, using a thresholding step on the grayscale images will extract the rough shapes of the heads. In this research, after converting RGB images to grayscale, Otsu method is used for thresholding. In this way, darker objects including sperms and non-sperms are extracted. As a result of staining, non-sperm particles have



**Fig. 2 – (a) Initial head mask. (b) Hue channel of the mask area. (c) Histogram of the Hue channel of the mask area.**

different color combinations from sperms; therefore, color seems to be an appropriate feature to distinguish them. Among different color spaces, HSV color space separates chrominance (Hue) from Saturation and Value, which results in robustness against lighting and shading. Moreover, the HSV is a typical perceptual color space and closer to our perception and feeling of colors. Therefore, the Hue channel of HSV color space is used to distinguish sperms from other particles. To this end, each image is transformed from RGB to HSV space, and then the histogram of their Hue channel is computed (Fig. 2). The extracted objects from the thresholding step are examined one by one using their normalized hue histogram (Fig. 2c), and the ones whose histogram has Euclidean distance of less than a threshold to the reference histogram of sperm heads ( $H_{ref}$ ) are accepted as possible heads. Fig. 2 shows examples of the initial head mask, the hue channel of the pixels inside the head mask and the corresponding normalized histogram. To create the reference histogram, 25% of sperm heads are randomly selected from the images of the dataset, and the  $H_{ref}$  histogram is calculated using Eq. 1. In this equation,  $H(i)$  is the normalized histogram of the Hue channel of one sperm head, and  $m$  is the number of selected heads. The Euclidean distance of the two histograms ( $HistDist$ ) is calculated using Eq. 2. In this equation,  $H_{ref}$  is the reference histogram, and  $H$  is the normalized histogram of the Hue channel of the object.

$$H_{ref} = \frac{1}{m} \sum_{i=1}^m H(i) \quad (1)$$

$$HistDist = (H - H_{ref})^T (H - H_{ref}) \quad (2)$$

Histogram distance and histogram intersection methods have been used in various applications such as image retrieval [22]. Also, Hue histogram has gained attraction due to its robustness to lighting and shading effects. But to the best of our knowledge, this is the first time this method is used for sperm detection.

The sperm head masks extracted by now include a part of their tails. Also, there are some sperms that are close to each other, and after thresholding, their extracted masks are connected to each other. To eliminate the tail and separate the connected head masks, two steps of erosion and dilation on the resulting masks are used. In the next step, a double threshold is applied on the size of the resulting masks.  $minThr$  is the

minimum size that will be accepted and  $maxThr$  is the maximum size. In the last step, sperms that are touching the image border are eliminated since they are not considered valid sperms.

## 2.2. Extracting the edges

Canny edge detector is used to extract the edges. To prevent the appearance of spurious edges, before applying the Canny edge detector, a low pass Gaussian filter of size  $9 \times 9$  with the standard deviation of 0.5 is applied twice to the grayscale image. Canny uses two levels of sensitivity thresholds; one is the minimum strength for the edge to be accepted as a valid edge and the other one is for the lowest strength for the edge to be accepted as a valid edge only if that edge is connected to a strong edge. In our experiments, we used 0.4 and 0.01 respectively for these thresholds. These thresholds were selected by trial and error as they presented the best tradeoff between preventing noisy edges and keeping useful edges. Fig. 4b shows an example of applying Canny, with the above mentioned thresholds, to one of the images from the dataset.

## 2.3. Applying Gradient Vector Flow (GVF) snakes

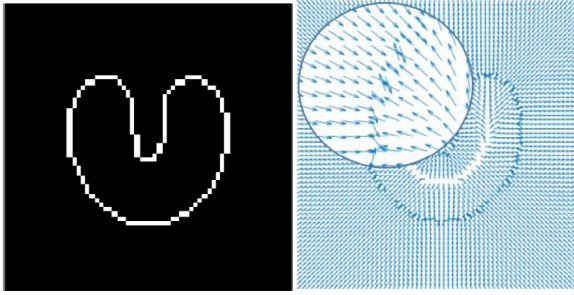
Gradient Vector Flow (GVF) snake is an edge based active contour model that is first proposed by Xu and Prince [23] to solve the problems with initialization and poor convergence to boundary concavities. A snake is a curve  $X(s) = [x(s), y(s)]$ ,  $s \in [0, 1]$  that moves in the space domain of an image to minimize the energy functional

$$E = \int_0^1 \frac{1}{2} [\alpha |X'(s)|^2 + \beta |X''(s)|^2] + E_{ext}(X(s)) ds \quad (3)$$

in which,  $\alpha$  and  $\beta$  are parameters controlling the tension and rigidity of the snake respectively,  $X'$  and  $X''$  are the first and the second derivatives of the snake, and  $E_{ext}$  is the external energy derived from the image.  $E_{ext}$  takes its minimum at the features of interest, which are usually the boundaries of the objects. Traditionally, the gradient of the edge map is used as the external energy, but this has two shortcomings. The first one is that for the smooth parts of the image, the gradient will be zero, resulting in no force to guide the snake, and the second problem is that the gradient has a large value only near the boundaries of the objects. GVF is therefore designed to alleviate these two problems. The GVF field is defined to be the vector field  $\mathbf{V}(x, y) = [u(x, y), v(x, y)]$  that minimizes the following energy functional.

$$e = \int \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |\mathbf{V} - \nabla f|^2 \quad (4)$$

This vector field has the desired properties: the vectors point toward the edges and also extend toward the flat areas of the image, resulting in extended energy field, far from the boundaries of the objects [23]. Therefore, this gradient field directs the initial contour toward the boundaries of the objects even when the contours are initialized far from the boundaries. Fig. 3 shows an example of an edge map and its corresponding GVF. The arrows indicate the gradient field, and as the figure shows,



**Fig. 3 – An example of an edge map (left) and its corresponding gradient vector flow field (right).**

the arrows start from far distance and point toward the closest edge. Apart from these properties, the GVF snake can capture boundary concavities too. In this research, the edges extracted by the Canny edge detector are used as the edge map and the contours of the head masks extracted in the previous step are used as the initial contours. These contours have the role of localizing the active contour model. In our experiments, we have noticed that the masks created by the experts are one pixel larger than the edges extracted by the edge detector algorithm, so after extracting the head mask by the active contour, we dilate them with a disk shape structuring element of size one pixel (the first step of Algorithm 2). Fig. 4 shows one of the images from the dataset (Fig. 4a) and its extracted

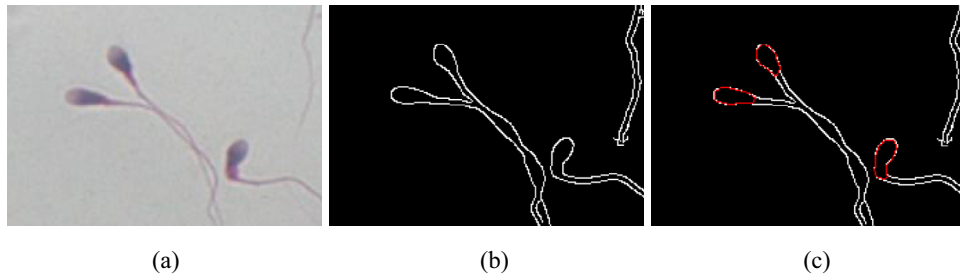
edges (Fig. 4b). Red lines in Fig. 4c show the result of applying GVF active contour.

#### 2.4. Separating acrosome and nucleus and removing midpiece

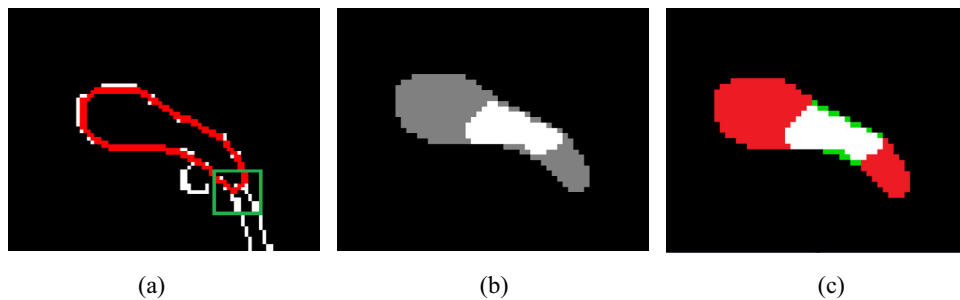
After extracting the head contours, the next step is the removal of midpieces, which are usually segmented with the heads, especially when there are swollen midpieces or cytoplasmic remnants. To do so, first the tail is located using a new method called Tail Detection using Curvature and Edge Energy (TDCEE). In this method, the maximum curvature on the head contour is located and then the energy of edges in a neighborhood of size  $(2n + 1) \times (2n + 1)$  pixels around that point is found (Fig. 5a). If the energy is greater than a threshold ( $\tau$ ), this point is considered as the start of the tail, otherwise, the next point with the highest curvature is selected as a new tail point candidate and the edge energy is calculated for this point again. For a two dimensional curve given parametrically as  $X(t) = [x(t), y(t)]$  the curvature is defined as

$$\kappa(t) = \frac{x(t)'y(t)'' - x(t)''y(t)'}{(x(t)'^2 + y(t)'^2)^{3/2}} \quad (5)$$

where  $\kappa$  is the curvature and the primes refer to derivatives with respect to parameter  $t$ . The energy of the edges inside the neighborhood  $W$  is calculated using the Eq. 6. In this equation,  $E$  is the energy,  $W$  is the neighborhood,  $i$  and  $j$  are the

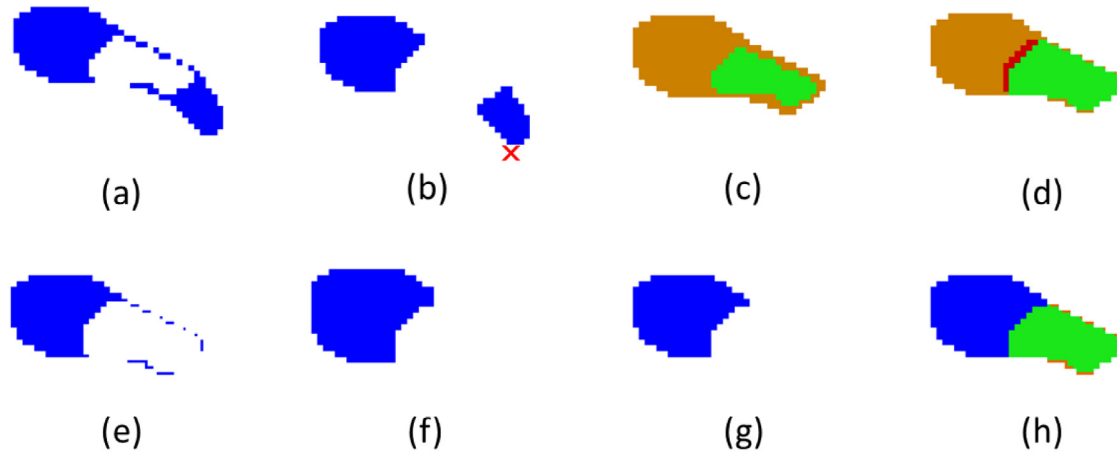


**Fig. 4 – Extracting the head masks using active contours. (a) The original image. (b) Edges extracted using the Canny edge detector. (c) Red contours extracted using active contours. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)**



**Fig. 5 – (a) Tail point with the highest curvature. The window around this point contains edges belonging to the tail. (b) Acrosome and midpiece are shown in gray and nucleus is shown in white. (c) Green pixels belong to the nucleus but are classified as class acrosome-midpiece. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)**





**Fig. 6** – This figure shows the steps of Algorithms 1 and 2. (a) Pixels belonging to class acrosome and midpiece; (b) acrosome and midpiece separated after erosion and dilation, red cross shows the tail point; (c) head mask after removing the midpiece and one pixel dilation, green color shows the pixels classified as nucleus; (d) nucleus after two pixels dilation (green). The red pixels belong to acrosome but as a result of nucleus dilation they are now part of nucleus. (e) Nucleus removed from the head mask; (f) acrosome after one pixel erosion and three pixels dilation; (g) acrosome after removing pixels outside the head mask. (h) Blue color shows the resulting acrosome, and green color shows the resulting nucleus. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

indices of the pixels inside  $W$ , and  $m$  and  $n$  are the height and width of the neighborhood.

$$E = \sum_{i=1}^m \sum_{j=1}^n W(i, j)^2 \quad (6)$$

It should be noted that Alegre et al. [24] used the position of the tail in their proposed method for determining the state of the acrosome. However they didn't clearly state their method for detecting the location of the tail.

To separate the acrosome and nucleus and to remove the midpiece, k-means clustering with two classes is used on the intensity channel of smoothed sperm head segments. To smooth the gray image, a hybrid filter consisted of a fourth order partial differential equation followed by a relaxed median filter is used [19]. After clustering, acrosome, midpiece and the area around the head, that have similar intensities, will be placed in the same class, and nucleus, which is darker than the rest, will be placed in a separate class. As Fig. 5b shows, the area around the head usually has brightness comparable to acrosome, because of the lighting effects, although it is part of the nucleus. To separate the acrosome and nucleus and remove the midpiece, some morphological operations are used based on prior knowledge about the sperm shape. Our method for locating and removing the midpiece is summarized in Algorithm 1, and our method for separating the acrosome and nucleus is presented in Algorithm 2. In these algorithms the distance from point  $(x_i, y_i)$  to a patch  $P$  is defined as

$$D_i = \min \{d_i | d_i = \sqrt{(x_t - x_i)^2 + (y_t - y_i)^2}, (x_i, y_i) \in P\} \quad (7)$$

**Algorithm 1.** Locate and remove the midpiece.

1. Smooth the intensity channel of the head segment.

2. Cluster the pixels of the head segment using k-means clustering with two classes (Fig. 5b).
3. Select the pixels belonging to midpiece and acrosome (the cluster with higher intensity) (Fig. 6a).
4. Erode and then dilate acrosome and midpiece with a disk shaped structuring element of size one pixel to disconnect the acrosome and the midpiece (Fig. 6b). The size of the structuring element was chosen empirically.
5. Find the distance from the tail point to the patches remained from the last step.
6. Remove the patch that is closer to the tail point, from the head mask (Fig. 6b)

After removing the midpiece the nucleus and the acrosome are separated using Algorithm 2:

**Algorithm 2.** Separate nucleus and acrosome.

1. Dilate the head mask with a structuring element of size one pixel (Fig. 6c). (Refer to Section 2.3)
2. Dilate the nucleus with a disk shape structuring element of size two pixels to include the border pixels (Fig. 6d). Two pixels is the maximum enlargement needed to affix the border pixels to the nucleus. This size was selected empirically.
3. Compare the nucleus mask with the head mask and discard any nucleus pixels located outside the head mask area as a result of dilation.
4. Find the acrosome mask by subtracting nucleus from the head mask (Fig. 6e).
5. Erode acrosome with a structuring element of size one to remove any remaining perimeter pixels wrongly classified as acrosome. Size one is chosen empirically.
6. Dilate the acrosome by a disk shape structuring element of size three pixels (Fig. 6f). (One pixel to compensate the

erosion in the last step, and two pixels to get back the acrosome pixels from the nucleus, which was the result of dilation in step 2 (red pixels in Fig. 6d)).

7. Compare acrosome with the head mask and remove any pixels located outside the head mask area.
8. Compare the acrosome mask with the nucleus and remove any nucleus pixels that overlap the acrosome, from the nucleus mask (red pixels in Fig. 6d).

### 3. Evaluation criteria

To evaluate the proposed method, we used Gold-standard dataset provided by Chang et al. [20]. This dataset consists of 19 images of stained sperm samples, with a total of 264 sperm cells from which 210 are considered valid. The images are  $780 \times 580$  pixels in RGB format. For each image, hand-segmented ground-truths are provided for sperm heads, nuclei and acrosomes. More details about the dataset and how the images were acquired can be found in Chang et al. [20]. The evaluation was done with respect to tail point detection, sperm detection and segmentation. The results were compared with those of Chang method and Carrillo method. The reason is twofold: (1) these methods have been previously compared together using the Gold-standard dataset, and therefore, the comparison would be more reliable; (2) to the best of our knowledge, the work done by Chang et al. is the state-of-the-art in automatic detection and segmentation of human sperm cells that is evaluated using a publically available dataset. We have compared our results with the results reported by Chang et al. [20]. In order for our results to be comparable with Chang et al.'s results, we have used the same criteria they used in their paper.

#### 3.1. Tail detection

The tail detection is considered successful if the detected tail point is closer to the midpiece than the acrosome. The distances from the tail point to the midpiece and the acrosome are calculated using Eq. 7. The evaluation of the tail detection method is done with respect to the valid sperms. Therefore, *True Detected* (TD) is defined as the number of the tail points located closer to the midpieces than the acrosomes, among the valid heads. *False Detected* (FD) is the number of tail points located somewhere closer to the acrosomes than the midpieces, and *Not Detected* (ND) is the number of sperms for which a tail point is not detected. *True Detection Rate* (TDR) is defined by Eq. 8. In this equation, the denominator adds up to the total number of sperm heads.

$$TDR = \frac{TD}{TD + FD + ND} \quad (8)$$

#### 3.2. Sperm detection

There are four parameters that can be used to evaluate the sperm detection: *True Positive* (TP) as the number of sperms that are detected correctly as being sperm; *False Positive* (FP) as the number of invalid sperms or other particles that are falsely detected as sperms; *False Negative* (FN) as the number of valid

sperms that are not detected as being sperms; and *True Negative* (TN) as the number of invalid sperms or non-sperm particles that are not detected as sperms. TN is not considered here because there is not any ground truth about the number of invalid sperms plus other particles. To evaluate the detection accuracy, we use *True Positive Rate* (TPR), *False Positive Rate* (FPR) and *Precision* as defined by the following formulas:

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

$$FPR = \frac{FP}{FP + TN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

In practice, FP is used instead of FPR. Since the denominator of FPR is constant, they will differ with only a constant coefficient. This also conforms to Chang method of evaluation and therefore makes the comparison of the results more reliable.

To evaluate the overall performance of a classifier, F-Score is used, which is the harmonic means of Precision and Recall (TPR). When Precision and Recall are equal, the F-Score is equal to their average, but when they are different F-Score will be less than their average. This property of F-Score favors a balanced classifier over an imbalanced one.

$$F\text{ Score} = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

#### 3.3. Segmentation of head, nucleus and acrosome

To evaluate the accuracy of segmentation, the results are compared to the hand segmented ground truth for head, acrosome and nucleus. To make our segmentation results readily comparable with the results reported by Chang et al., we use Dice coefficient defined by Eq. 13 [20].

$$D = \frac{2|S \cap G|}{|S| + |G|} \quad (13)$$

where S and G are the sets of pixels segmented automatically and manually respectively. This equation measures the amount of overlap between the automatic segmentation and the ground truth. The result is a number between zero and one, in which zero means no overlap and one means complete overlap.

## 4. Results

### 4.1. Sperm detection

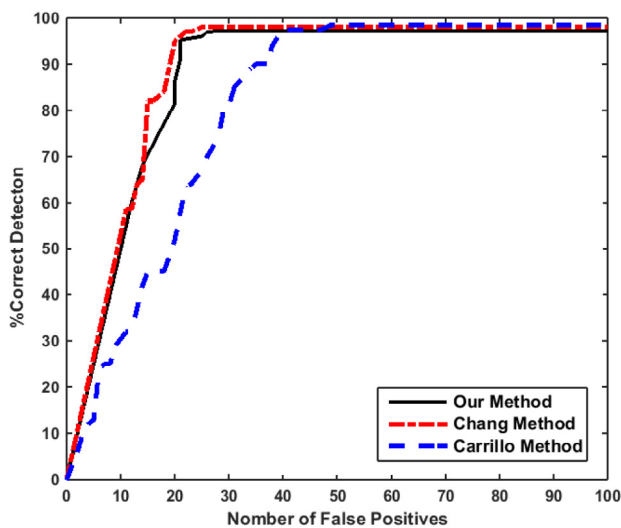
We have tested different thresholds for size and histogram distance. The thresholds for minimum size, maximum size, and histogram distance are changed according to Table 1 and the resulting TPR and FPs are used to plot the ROC (receiver op-

**Table 1 – The variation of the threshold values and the best performing value for each one.**

	Range	Best value
Min size threshold (pixels)	60:10:150	90
Max size threshold (pixels)	250:50:550	350
Histogram distance threshold	0.08:0.02:0.16	0.14

erating characteristic) curve (Fig. 7). As the curve shows our method has a performance close to Chang method and is much better than Carrillo method. In ROC curves, the area under the curve (AUC) indicates the performance of the system. The higher the AUC, the better the performance is. Our method achieved the AUC of 87%, which is close to Chang method with the AUC of 88% and significantly better than Carrillo method with AUC of 81%. The thresholds that achieved the best result are summarized in Table 1 and the best values for Precision and TPR for our method and those of Chang and Carrillo methods are presented in the first three rows of Table 2. As that table shows, our results are again very close to Chang method and significantly better than Carrillo method. It should be noted that these results have been achieved using only the Hue and size criteria.

Since Chang and Carrillo methods had used the whole dataset to optimize the parameters, we followed the same procedure to make our results comparable. However, to validate the results further, we repeated the test, this time using half of the dataset for optimizing the parameters and the other half

**Fig. 7 – ROC curves for sperm head detection, comparing our method with Chang method and Carrillo method.****Table 2 – Detection accuracy.**

	Precision	TPR	F-Score
Our method	88.3%	97.1%	92.5
Chang method	89.9%	97.6%	93.6
Carrillo	83.7%	95.7%	89.3
Our method, 50/50 (mean)	89.3%	93.3%	91.3
Our method, 50/50 (best)	88.9%	98.0%	93.2

for testing. The dataset was randomly divided into train and test subgroups. We repeated the experiments six rounds and each time obtained the parameter set that presented the best tradeoff between Precision and TPR. The testing was done using the optimum parameters, and then the average of the resulting TPRs, as well as Precisions, was calculated. Also the best combination of Precision and TPR, in terms of their F-Score, is reported (Table 2, last two rows). As the best results show, in some instances, the overall performance is improved compared to when we had used the whole dataset. The reason is, when the number of images reduces, the probability of having difficult cases will decrease. Therefore we could end up having better Precision and TPR. We obtained average Precision of 89.3% with the standard deviation of 3.9 and average TPR of 93.3% with the standard deviation of 3.8. As Table 2 (50/50 (mean)) shows, the average Precision is improved but the average TPR is reduced compared to when the whole dataset was used. This is a natural outcome since the Precision and TPR normally change in opposite directions. To gain a better insight for the results, we provided the F-Scores in the last column of Table 2. As the F-Scores show, even when two separate sets of images are used for training and testing, our overall performance is very close to the state-of-the-art where they used the whole dataset for parameter setting, which definitely can increase the accuracy.

#### 4.2. Sperm head segmentation

Fig. 8 (left) shows the boxplot corresponding to the Dice coefficients for head, nucleus and acrosome, and compares our results with the results of Chang and Carrillo methods. As the boxplot shows, in our method the average Dice coefficients are higher than Chang method. Also the height of the boxes for our method is less than that of Chang method, which indicates less dispersion in our results. This is a desirable outcome since it shows that most of the segmentation results are also close to the average performance. Table 3 summarizes the mean Dice coefficients (MDC) for head, nucleus and acrosome as well as the overall Dice coefficient. As the table shows, we have achieved an average of 92% overlap with the hand segmented heads, 87% overlap with the hand segmented nuclei and 84% overlap with the hand segmented acrosomes, which all show improvement compared to Chang method. Our overall mean Dice coefficient is 88%, which again shows 3% improvement compared to Chang method. It can be seen that the standard deviation for our results is also less than that of the other two methods. The statistical significance of our results compared to Chang's and Carrillo's results were evaluated using Wilcoxon rank-sum test with  $\alpha = 0.05$  significance level. According to this test, the improvements to head segmentation and nucleus segmentation were statistically significant. Fig. 8 (right) illustrates the probability density functions (PDF) for Dice coefficient graphically, and compares it with the results obtained by Chang and by Carrillo. The results of head, acrosome and nucleus segmentations are aggregated and shown together. As the figure shows, the PDF for our method is shifted towards the right, which suggests more accurate segmentation. Also the width of the curve is less, which is the sign of less dispersion in the results. As an example, Fig. 9 shows the results of our method for one of the dataset images.

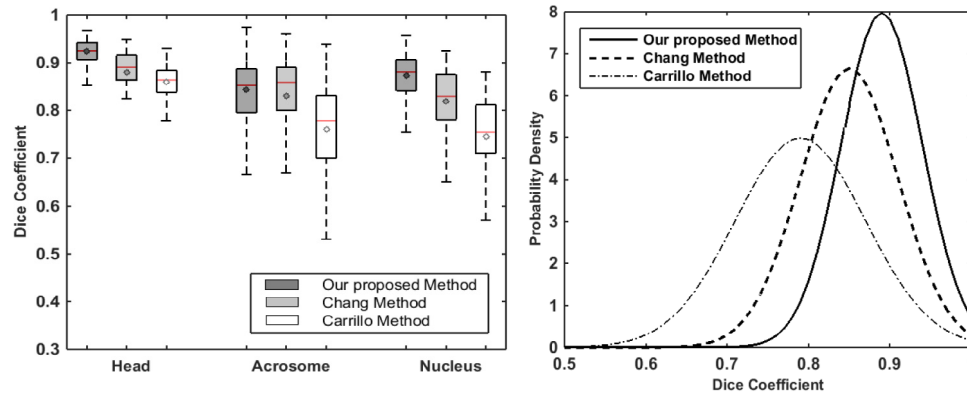


Fig. 8 – Left: boxplots show the Dice coefficients for head, acrosome and nucleus resulted from our method, Chang method and Carrillo method. Top and bottom edges of the boxes indicate the 75 and 25 percentiles and whiskers extend to the most extreme coefficients that are not outliers. The medians and means are shown with horizontal lines and circles respectively. Right: comparing the probability density functions for Dice coefficients. The results from head, acrosome and nucleus segmentations are aggregated and shown together. Continuous line shows the results from our method ( $\mu = 0.88$ ,  $\sigma = 0.05$ ) compared to Chang ( $\mu = 0.85$ ,  $\sigma = 0.06$ ) and Carrillo ( $\mu = 0.79$ ,  $\sigma = 0.08$ ) methods.

#### 4.3. Tail detection

We performed a number of experiments to evaluate TDCEE method. The size of the neighborhood was changed from  $7 \times 7$  to  $19 \times 19$  ( $n = 3, 5, 7, 9$ ) pixels, with increments of 4 pixels, and also the edge energy thresholds of  $2n$ ,  $3n$  and  $4n$  were tested. If the window is truly located at the tail point, then the energy of the edges depends on the size of the window; when the size of the window increases, a larger part of the tail will be inside the window and therefore the energy of the edges located inside

the window increases. Therefore the threshold values are chosen as multiples of the window size. The best performance was achieved when a neighborhood of size  $15 \times 15$  ( $n = 7$ ) pixels was used and also a threshold of 21 ( $3n$ ) was considered for the edge energy. Our experiments showed that a maximum of three trials is enough to reveal all the tail points. It means that after a maximum of three trials, all tail points were either correctly or incorrectly detected. Using these thresholds, TD, FD and ND were 196, 9 and zero respectively. Using Eq. 8, TDR will be 0.96, which means that from all of the sperms

Table 3 – Mean Dice coefficient for head, nucleus and acrosome as well as overall mean Dice coefficient and its standard deviation.

	Head MDC	Nucleus MDC	Acrosome MDC	Overall MDC	Overall STD of DC
Our proposed method	0.92	0.87	0.84	0.88	0.05
Chang method	0.88	0.82	0.83	0.85	0.06
Carrillo method	0.86	0.75	0.76	0.79	0.08

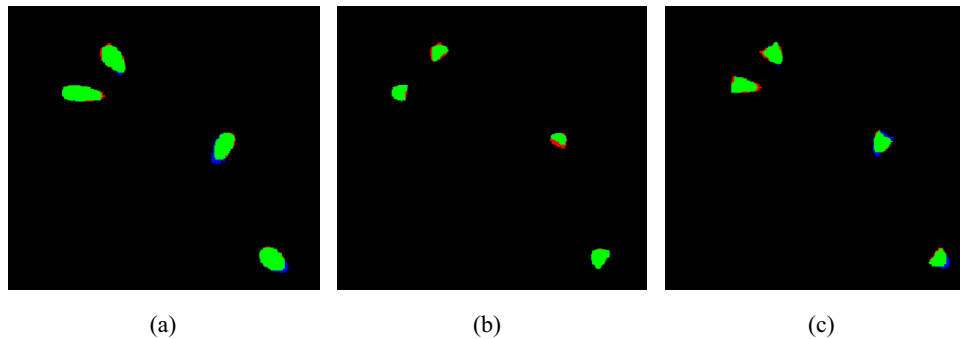


Fig. 9 – Comparing our automatic segmentation results with the hand segmented ground truth. Green color shows the overlap between automatic segmentation and the ground truth. Red pixels are those belonging to ground truth, which are not segmented with our method, and blue pixels are those segmented with our method, which does not belong to the ground truth. (a) Heads, (b) acrosomes and (c) nuclei. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



that were detected correctly as sperm, 96% of tail points were detected in the right place.

To validate TDCEE further, we divided the dataset into half randomly and repeated the above experiments on each division separately. We repeated the experiment ten rounds with different combination of images. Again the best performance was achieved using a neighborhood of size  $15 \times 15$  ( $n = 7$ ) and the threshold of 21 ( $3n$ ) consistently for each subset of images. The average TDR was 0.96 with the standard deviation of 0.03. This experiment has proven the robustness of our tail detection method.

## 5. Discussion

In this research we have proposed a method for sperm head segmentation, which outperforms the current state-of-the-art. We believe that the improvement was achieved as a result of combining active contours, Canny edge detector and TDCEE.

Using GVF active contour model has improved the segmentation of the sperm head by extracting the outer contour of the head more precisely than the methods used previously by Chang and Carrillo. Active contours tend to follow the edges of the objects and at the same time avoid sharp angles and fill in the gaps caused by weak edges. These properties are extremely important in extracting sperm heads in microscopic images since these objects have smooth contours, which are followed by active contours very well. Also there are weak edges that can be filled out properly by active contours. It should be noted that for one of the sperm heads, GVF was not able to segment the sperm head because of the relatively small nucleus. Since the starting contour was located too much off the center of the head, all the GVF forces pointed to one side and therefore the final contour was collapsed into one line. It was the reason behind one True Positive missing from our results compared to Chang method. The other observation was that for the sperm segmentation, the edges of the sperms seem to be the best features to be used in segmentation. The combination of Canny edge detector and active contours has produced good results. The reason is that Canny is able to extract the weakest possible edges, as long as they are connected to a strong edge. This is desirable because in the microscopic images of sperms, the gradient around some true edges are comparable with the gradient caused by noise, and Canny is able to distinguish them using their proximity to a strong edge.

In our algorithm for segmentation we used a number of erosions and dilations with different sizes. These sizes were chosen by visual inspection and experiment. We used less than half of the dataset images to find the optimum sizes and then we used them in the segmentation of the whole dataset, which resulted in better performance than Chang method. It should be emphasized that in Chang method, the parameters were tuned over the whole dataset to optimize the Dice coefficient.

TDCEE is an important step that will refine the segmentation results obtained by active contours. In this paper we used the position of the tail to remove the midpiece. If the position is detected correctly, the midpiece removal will be successful; otherwise the acrosome might be removed instead

of the midpiece, which will have an opposite effect on the segmentation results. With 96% success in tail detection, our tail detection method had an important effect on the improvement of the results. To the best of our knowledge this algorithm has not been used before in any of the previous works.

Separating acrosome and nucleus constitutes some difficulties since the lighting effects cause the pixels around the head to have intensities similar to acrosome even though they belong to nucleus. So clustering alone will not identify those pixels as being part of nucleus. Therefore, we should use the prior knowledge about the shape of the sperm head, in addition to clustering methods, to separate them. Our method is designed with regard to this prior knowledge. We have observed that the segmentation of acrosome by the experts is done for each sperm based on the difference in the intensity for that sperm. It means that the acrosome in one sperm consists of darker colored pixels compared to another sperm. This is one reason behind lower performance in segmentation of acrosome in our method as well as Chang and Carrillo methods. To improve the result of acrosome segmentation, we believe that this phenomenon should be investigated further.

We have also proposed the *HistDist* measure to discriminate sperms from non-sperm particles. We decided to use the Hue channel of HSV color space for two reasons. First, in HSV color space the chrominance (Hue) is separated from Saturation and Value, which makes the method more robust to changes in lighting. Second, our experiments with RGB color space did not produce satisfactory results. We used three dimensional histogram of RGB image with various bin sizes but the histogram distance was not able to discriminate between different kinds of objects. In RGB color space, the representation of colors is not uniform. Also the information about the chrominance is mixed with the Intensity and Saturation, which makes them sensitive to changes in lighting. These could be the reasons behind poor performance in this color space.

We observed that almost all of the false positives in sperm detection were due to the invalid sperms detected as valid. Some of the invalid sperms are relatively far from the borders of the image but since parts of their tails are outside of the image, they are considered invalid by the field experts. To distinguish these sperms, an improved tail tracking method should be used. We believe that this will improve the sperm detection results significantly.

## 6. Conclusion

In this paper, we have presented a complete framework for sperm detection and segmentation, which is totally automatic. To evaluate our method, we used a publicly available dataset, Gold-standard dataset, and compared our results with those of Chang method which, to the best of our knowledge, is the state-of-the-art in detection and segmentation of sperm heads. To detect sperm heads, after extracting the initial masks using Otsu method, we used a combination of size and histogram of Hue channel to distinguish sperms from non-sperm particles. To differentiate sperms from other particles based on their colors, the Euclidian distance between the learned histogram of sperm heads and the histogram of ex-

tracted regions were used. This way we achieved a detection rate comparable to the state-of-the-art solution.

To segment sperm heads, we combined Canny edge detector with GVF active contour. The result was an improved head contour extraction. We proposed a novel tail detection method using curvature and edge energy (TDCEE) with 96% TDR, which is used to refine segmented heads by locating their midpieces. With the combination of these three methods and by removing the midpiece, we were able to achieve 92% overlap with the hand segmented ground truth, which outperforms the state-of-the-art solution. Also our results show less dispersion compared to the existing solutions.

To segment the head into its constituent parts, namely, acrosome and nucleus, we proposed an algorithm based on prior knowledge about the shape of the nucleus and acrosome, which uses k-means clustering and morphological operations. Our method resulted in 84% and 87% overlap with the hand segmented acrosome and nucleus respectively, which again outperforms the existing solutions.

We believe that the True Positive Rate of sperm head detection can be improved significantly by improving the tail tracking methods and also studying the exact criteria used by field experts in defining valid and invalid sperms. Also we noticed that segmentation of acrosome and nucleus by the experts is based on the intensity variation within each sperm. So using local clustering may be more appropriate than global clustering.

## Acknowledgment

We would like to thank Ms. Violeta Chang for providing us with the results of their algorithms.

## REFERENCES

- [1] J. Boivin, L. Bunting, J.A. Collins, K.G. Nygren, International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care, *Hum. Reprod.* 22 (2007) 1506–1512.
- [2] L. Schmidt, Psychosocial consequences of infertility and treatment, in: D.T. Carrell, C.M. Peterson (Eds.), *Reproductive Endocrinology and Infertility*, Springer, New York, 2010, pp. 93–100.
- [3] L.A. Peronace, J. Boivin, L. Schmidt, Patterns of suffering and social interactions in infertile men: 12 months after unsuccessful treatment, *J. Psychosom. Obstet. Gynecol.* 28 (2007) 105–114.
- [4] N. Wiersema, A. Drukker, M. Dung, G. Nhu, N. Nhu, C. Lambalk, Consequences of infertility in developing countries: results of a questionnaire and interview survey in the South of Vietnam, *J. Transl. Med.* 4 (2006) 1–8.
- [5] G. Onat, N.K. Beji, Marital relationship and quality of life among couples with infertility, *Sex. Disabil.* 30 (2012) 39–52.
- [6] M.R. Maduro, D.J. Lamb, Understanding new genetics of male infertility, *J. Urol.* 168 (2002) 2197–2205.
- [7] D.R. Franken, How accurate is sperm morphology as an indicator of sperm function?, *Andrologia* 47 (2014) 720–723.
- [8] C. Brazil, Practical semen analysis: from A to Z, *Asian J. Androl.* 12 (2010) 14–20.
- [9] G. Barroso, R. Mercan, K. Ozgur, M. Morshedi, P. Kolm, K. Coetzee, et al., Intra- and inter-laboratory variability in the assessment of sperm morphology by strict criteria: impact of semen preparation, staining techniques and manual versus computerized analysis, *Hum. Reprod.* 14 (1999) 2036–2040.
- [10] World Health Organization, WHO Laboratory Manual for the Examination and Processing of Human Semen, fifth ed., 2010.
- [11] O. García-Olalla, E. Alegre, L. Fernández-Robles, P. Malm, E. Bengtsson, Acrosome integrity assessment of boar spermatozoa images using an early fusion of texture and contour descriptors, *Comput. Methods Prog. Biomed.* 120 (2015) 49–64.
- [12] K. Park, W. Yi, J. Paick, Segmentation of sperms using the strategic Hough transform, *Ann. Biomed. Eng.* 25 (1997) 294–302.
- [13] V.R. Nafisi, M.H. Moradi, M.H. Nasr-Esfahani, Sperm identification using elliptic model and tail detection, *World Acad. Sci. Eng. Technol.* 6 (2005) 419–421.
- [14] V. Gonzalez-Castro, E. Alegre, P. Morala-Arguello, S.A. Suarez, A combined and intelligent new segmentation method for boar semen based on thresholding and Watershed transform, *Int. J. Imag. Robot.* 2 (2009) 70–80.
- [15] H. Carrillo, J. Villarreal, M. Sotaquira, A. Goelkel, R. Gutierrez, Spermatozoon segmentation towards an objective analysis of human sperm morphology, in: *Proceedings of the 5th International Symposium on Image and Signal Processing and Analysis*, 2007, pp. 522–527.
- [16] H. Carrillo, J. Villarreal, M. Sotaquira, M.A. Goelkel, R. Gutierrez, A computer aided tool for the assessment of human sperm morphology, *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering*, 2007. BIBE 2007., 2007, pp. 1152–1157.
- [17] A. Bijar, M. Mikaeili, A.P. Benavent, R. Khayati, Segmentation of sperm's acrosome, nucleus and mid-piece in microscopic images of stained human semen smear, *Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, 2012 8th International Symposium on, 2012, pp. 1–6.
- [18] A. Bijar, M. Mikaeili, Sperm's tail identification and discrimination in microscopic images of stained human semen smear, *Image and Signal Processing and Analysis (ISPA)*, 2011 7th International Symposium on, 2011, pp. 709–714.
- [19] A. Bijar, A.P. Benavent, M. Mikaeili, R. Khayati, Fully automatic identification and discrimination of sperm's parts in microscopic images of stained human semen smear, *J. Biomed. Sci. Eng.* 5 (2012) 384–395.
- [20] V. Chang, J.M. Saavedra, V. Castañeda, L. Sarabia, N. Hitschfeld, S. Härtel, Gold-standard and improved framework for sperm head segmentation, *Comput. Methods Prog. Biomed.* 117 (2014) 225–237.
- [21] V. Chang, J.M. Saavedra, V. Castañeda, L. Sarabia, N. Hitschfeld, S. Härtel. Gold-standard [online]. Available from: <<http://morfologia.cedai.cl/public/>>.
- [22] S. Jeong, C.S. Won, R.M. Gray, Image retrieval using color histograms generated by Gauss mixture vector quantization, *Comput. Vis. Image Underst.* 94 (2004) 44–66, 4.
- [23] C. Xu, J.L. Prince, Snakes, shapes, and gradient vector flow, *IEEE Trans. Image Process.* 7 (1998) 359–369.
- [24] E. Alegre, M. Biehl, N. Petkov, L. Sanchez, Assessment of acrosome state in boar spermatozoa heads using n-contours descriptor and RLVQ, *Comput. Methods Prog. Biomed.* 111 (2013) 525–536.