Multi-object tracking of human spermatozoa

Lauge Sørensen[a], Jakob Østergaard[a], Peter Johansen[a], and Marleen de Bruijne[a,b]

[a] Department of Computer Science, University of Copenhagen, Denmark;
[b] Biomedical Imaging Group Rotterdam, Erasmus MC, The Netherlands

ABSTRACT

We propose a system for tracking of human spermatozoa in phase-contrast microscopy image sequences. One of the main aims of a computer-aided sperm analysis (CASA) system is to automatically assess sperm quality based on spermatozoa motility variables. In our case, the problem of assessing sperm quality is cast as a multi-object tracking problem, where the objects being tracked are the spermatozoa.

The system combines a particle filter and Kalman filters for robust motion estimation of the spermatozoa tracks. Further, the combinatorial aspect of assigning observations to labels in the particle filter is formulated as a linear assignment problem solved using the Hungarian algorithm on a rectangular cost matrix, making the algorithm capable of handling missing or spurious observations. The costs are calculated using hidden Markov models that express the plausibility of an observation being the next position in the track history of the particle labels. Observations are extracted using a scale-space blob detector utilizing the fact that the spermatozoa appear as bright blobs in a phase-contrast microscope. The output of the system is the complete motion track of each of the spermatozoa. Based on these tracks, different CASA motility variables can be computed, for example curvilinear velocity or straight-line velocity.

The performance of the system is tested on three different phase-contrast image sequences of varying complexity, both by visual inspection of the estimated spermatozoa tracks and by measuring the mean squared error (MSE) between the estimated spermatozoa tracks and manually annotated tracks, showing good agreement.

Keywords: Motion analysis, Statistical methods, Tracking, Human spermatozoa, Particle filter, Kalman filter, Hidden Markov model, Scale-space blob detection, Hungarian algorithm

1. INTRODUCTION

Increasing concern about the environment and its supposed effect on male fertility gives rise to search for new methods for semen analysis.[1] It is known that more than one in six couples has fertility problems and that more than 30% of these problems are clearly related to male infertility.[2] Sperm motility is considered one of the most important characteristics for evaluating fertility, which makes quantitative analysis of sperm motility by computer-aided sperm analysis (CASA) a valuable tool in semen analysis.

In general, sperm quality is assessed manually by inspection of sperm samples in a phase-contrast microscope. There has been done some work on standardizing the inspection procedures, for example the World Health Organization (WHO) has developed a manual depicting guidelines for examination of human semen.[1] Despite these guidelines and the existence of reference values to compare against it cannot be avoided, that the assessment is based on some subjective expert knowledge, and is thus subject to interobserver and intraobserver variability. Further, the manual inspection is a time consuming process. The three main reasons for using a CASA system are:

- To automate an otherwise time consuming process.
- To make the assessments objective.
- To get precise and reproducible assessments.

Further author information:
Lauge Sørensen: E-mail: lauges@diku.dk, Telephone: (+45) 35 32 14 33
Jakob Østergaard: E-mail: jakobnygren@gmail.com

Medical Imaging 2008: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6914, 69142C, (2008)
1605-7422/08/$18 · doi: 10.1117/12.771135

Also, a CASA system can be used to collect a vast amount of data that can form the basis for further spermatozoa research. The system does not necessarily need to have real-time performance, that is, being able to assess the quality almost instantly when being presented to a sample. An alternative, that is just as suitable, is to let the system work on digital recordings of the samples. This alternative is the one that we aim at, thus the overall goal is to develop a tracking algorithm that works on digital recordings.

Not much literature on digital image analysis for CASA is available. Berezansky et al. recently proposed to use mean-shift clustering in a feature space containing both spatial, temporal, and optical flow features on subsequences combined with a matching step for linking clusters across subseqences.[3] Several commercial manufacturers provide CASA systems, for example Hamilton Thorne and Hobson Sperm Tracker,[1] but not much is known on the methods and performance of these systems. For tracking other kinds of cells and intracellular components in bioimaging, current methods generally involve detection of objects by thresholding or model fitting, largely ignoring available temporal information and prior knowledge, and subsequently linking of detected objects across frames in some local or global manner.[4] Recently particle filters have become a popular tool for tracking in many fields, also in bioimaging. For example tracking of microtubule dynamics in fluorescence microscopy images using a specialized particle filter, that includes a data-dependent proposal distribution taking into account the most recent observations to cope with a peaked likelihood.[5]

Motion analysis in bioimaging requires tracking of large and time-varying number of spots in noisy images, which in the case of spermatozoa gives rise to several additional difficulties. The spermatozoa need to be detected in each frame, which is not trivial due to debris and other non-spermatozoan cells in the seminal fluid, generally termed round cells, that can have appearance similar to that of spermatozoa. The frame rate of 25 frames/second in the available data makes the spermatozoa movement very noisy, and this together with the fact that occlusions occur, both between individual spermatozoa and between spermatozoa and round cells, makes the motion tracking a very challenging problem. We propose a detection based on the Laplacian of the Gaussian operator (LoG) which has a shape that resembles the spermatozoan head including the halo which is a phase-contrast artifact. We use a particle filter for the motion tracking and handle the peaked likelihood by using Kalman filters within the particle filter for taking the most recent observations into account. Assigning observations to individual spermatozoa tracks, or particle labels, is done using the Hungarian algorithm together with hidden Markov models that take the track histories into account. The developed system is tuned to the problem of tracking spermatozoa by selecting a scale for the LoG operator, a dynamical model for the Kalman filter, and training the hidden Markov models on annotated spermatozoa tracks, but the system could in principle be tuned to other motion analysis problems in bioimaging.

2. METHODS

In this section we give a brief overview of the proposed system, for further details see our Master's thesis.[6] Figure 1 shows a schematic overview, and the interactions between the different parts, and their roles, are further described below.

One of the key challenges in multi-object tracking is the fact that the problem consists of both a continuous and a discrete aspect. Continuous, because the objects that are being tracked move through continuous space, and discrete because of the inherent data association problems of assigning observations and object labels to individual tracks. In our system the particle filter, in combination with the Kalman filters, handles the continuous aspect of the multi-object tracking problem, and the Hungarian algorithm, in combination with the hidden Markov models, handles the discrete aspect. Sections 2.1 and 2.2 describe how to deal with the continuous and discrete aspect respectively, and object detection is described in Section 2.3.

2.1 Motion Tracking

The spermatozoa tracks are to be augmented with a new position in each frame, utilizing the detections from 2.3 and the assignment of observations to lables determined in 2.2. Augmenting the tracks can also be seen as tracking the motion of the spermatozoa.

The motion tracking is done using a particle filter.[7] A particle filter is an algorithm based on Bayesian sequential importance sampling and resampling for approximating a distribution of arbitrary complexity by a multiset of

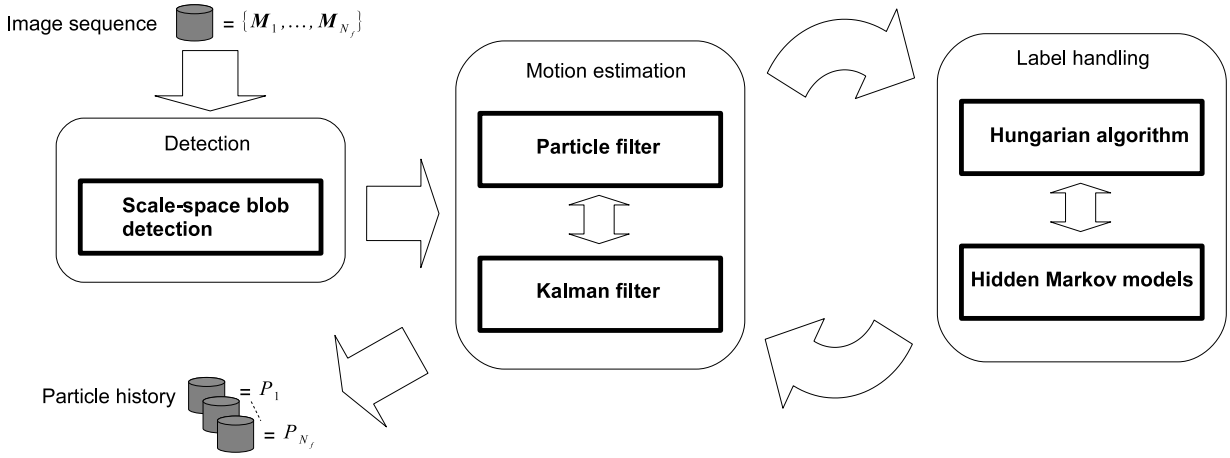


Figure 1. Overview of the proposed system. $\mathcal{P}_t$ is the particle set at time t, $\mathbf{M}_t$ is the image frame at time t, and N_f is the number of frames or time steps.

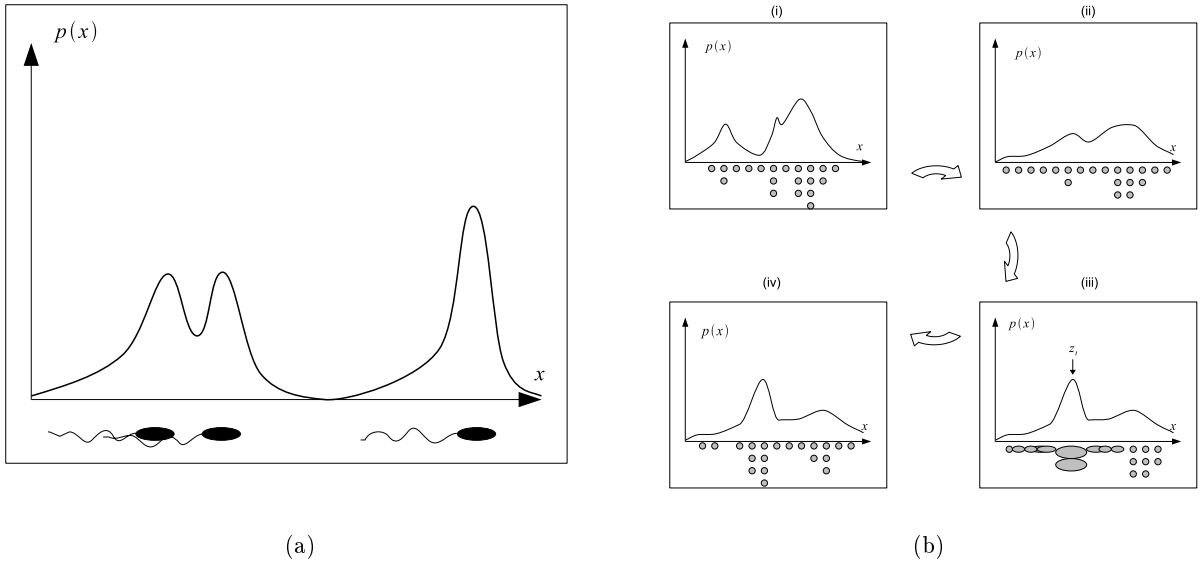
samples, and due to these two basic parts of the algorithm it is also commonly referred to as the sampling importance resampling (SIR) algorithm. It has already been successfully applied in many tracking problems with clutter.[5,8] In our case the distribution to approximate is on the positions of the spermatozoa, as illustrated in Figure 2(a), and by using a particle filter it is possible to maintain multiple hypotheses about the spermatozoa positions, making the system more robust towards clutter in the image sequences.

2.1.1 Particle filters

In the following let the state at time t of the object that we wish to track be denoted by $\mathbf{x}_t$ and the set of observations at time t by $\mathbf{z}_t$. The basic steps of the particle filter algorithm are illustrated in Figure 2(b) and detailed below. The algorithm maintains a multiset $\mathcal{P}_t = \{(\mathbf{x}_t^1, w_t^1), \ldots, (\mathbf{x}_t^{N_p}, w_t^{N_p})\}$ of N_p particles with associated weights. These particles represent hypotheses about the object state and are an approximation of the posterior distribution $p(\mathbf{x}_t|\mathbf{z}_t)$. Initially the particles are drawn from a prior distribution yielding the initial approximation. (i) At the beginning of each iteration the particle set approximates the posterior distribution from the previous time step $p(\mathbf{x}_{t-1}|\mathbf{z}_{t-1})$. (ii) Each particle is moved in state space by sampling form a proposal distribution $q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_t)$, where a popular simplification is to use a dynamical model as proposal distribution: $q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_t) = p(\mathbf{x}_t|\mathbf{x}_{t-1})$.[8] The dynamical model includes a noise term to reflect uncertainty in the object behavior. The effect is that the particles are moved and spread out, increasing the uncertainty resulting in a less peaked distribution. (iii) The moved particles are then given weights according to

$$w_t = \frac{p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_t)}, \tag{1}$$

where $p(\mathbf{z}_t|\mathbf{x}_t)$ is the observation likelihood. These weights are called importance weights. When the dynamical model is used as proposal distribution (1) reduces to $w_t = p(\mathbf{z}_t|\mathbf{x}_t)$. The effect of the weighting is that "good" particles, in the sense of being in high probability areas of the observation likelihood, are given high weight resulting in a more peaked distribution. Sampling from the proposal distribution and weighting the particles using (1) is generally called importance sampling in the literature. (iv) The last step is resampling from the current particle set where a new multiset is drawn with replacement from the current set of particles, according to their weights w_t, and given equal weights. This set is the final approximation of $p(\mathbf{x}_t|\mathbf{z}_t)$ in the current iteration and forms the basis for the next iteration, where the particles are again moved etc. This new multiset will most likely contain duplicate good particles, thus even though the particles have uniform weights at this point, areas of high probability in the underlying distribution are represented by more particles. This can be seen in (i) and (iv), where the particle filter implicitly represents the underlying distribution by placing more particles in areas of high probability.



(a) (b)

Figure 2. Motion tracking using a particle filter. (a) Probabilistic approach to tracking of spermatozoa. Instead of tracking the exact positions of the spermatozoa heads, the positions are expressed as a probability distribution. In this setting it is possible to model uncertainty about the positions which makes the tracking algorithm more robust. (b) Illustration of an iteration of the particle filter. (i) Particle set at time step $t-1$. (ii) A new particle set obtained by importance sampling using the old particle set and a dynamical model. (iii) Weighting using the most recent observation. (iv) The final particle set at time step t obtained by resampling from the current set.

2.1.2 Improved proposal distribution using Kalman filters

Initial experiments with a particle filter implementation using a dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ as proposal distribution for tracking the spermatozoa showed very bad results. The main reason for this is that the most current observations are not considered in the proposal distribution, and since the observation likelihood of the spermatozoa heads is very narrow and the dynamics of the spermatozoa are very noisy, very few samples are generated in high probability areas. Various methods for moving the particles in the direction of high probability regions exist, including prior editing, rejection methods, and auxiliary particle filters.[9] These methods incorporate an extra layer in the importance sampling step, that still uses $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ as proposal distribution, but seeks to move the particles to regions of high probability at the expense of additional computations, by searching "blindly" using the dynamical model. Another approach is to incorporate the current observations directly, for example using Kalman filters for forming Gaussian approximated proposal distributions that take the current observations into account.[9] Using this approach has given a significant performance gain. It should be noted that we use standard Kalman filters[10] as opposed to the unscented Kalman filters used in.[9]

Figure 3 illustrates the mechanisms involved in forming the proposal distribution using a Kalman filter. In the general Kalman filter setup the process is governed by a linear stochastic differential equation that with some simplifications is given by

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathbf{F}\mathbf{x}_{t-1} + v, \tag{2}$$

where $\mathbf{F}$ is a state transition model that advances the previous state $\mathbf{x}_{t-1}$ and $v \sim \mathcal{N}(0, \mathbf{Q})$ is the process noise with covariance $\mathbf{Q}$. This is also a common dynamical model in particle filters. It is a requirement that both the advanced object state and the observation are Gaussian distributions in order to apply the Kalman filter. Thus we associate an error covariance matrix $\mathbf{P}_t$ with each object state and also express observations as Gaussian distributions $p(\mathbf{z}_t)$. The Kalman filter applies the deterministic part of the dynamical model to advance the object state and the result is a Gaussian distribution centered on $\mathbf{F}\mathbf{x}_{t-1}$ with covariance $\mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^T + \mathbf{Q}$. This distribution is then combined with the observation distribution by the Kalman gain, resulting in a new Gaussian distribution used as proposal distribution in the particle filter. Several details about the Kalman filter have been left out in this explanation, thus for further details see.[10]

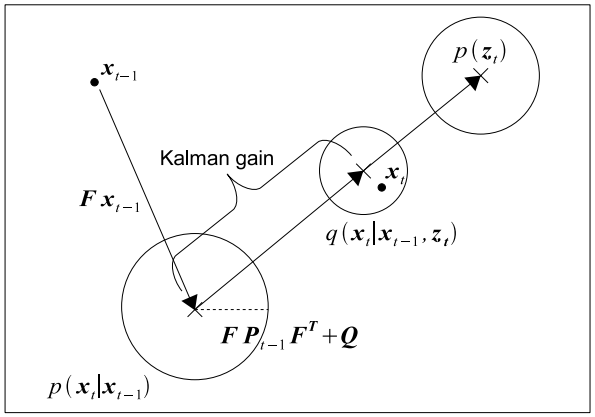


Figure 3. Illustration of using a Kalman filter for making a Gaussian approximated proposal distribution taking an observation into account. $p(\mathbf{z})$ is the observation expressed as a Gaussian distribution. $\mathbf{x}_{t-1}$ is the initial object state and $\mathbf{x}_t$ the final object state sampled from the Gaussian proposal distribution $q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{z}_t)$.

2.2 Label Handling

In each frame we have a set of spermatozoa candidate positions, also termed observations, and a set of particles with associated labels. We want to determine which observations stems from which spermatozoa tracks, and thus which particle labels they should be associated with. These associations are used in the motion tracking, more precisely the Kalman filters. The association of observations to object labels is achieved doing linear assignment of observations to labels, using the Hungarian algorithm[11] on a cost matrix, where the costs are the likelihood of observations being the next position in the particle filtered tracks. The costs can be Euclidean distances or probabilities using more sophisticated modeling. In this paper we have used hidden Markov models[12] to compute the likelihoods of the spermatozoa tracks when appending the observations, and used these probabilities as costs.

The hidden Markov model is a state space model that in contrast to the Kalman filter and particle filter, is discrete. Hence, the hidden Markov model can be interpreted as a graph where each node represents an object state with an associated probability distribution for the observation. We only considered continuous distributions for the observations and fully connected models. As continuous distributions we used a mixture of three isometric Gaussians in each state, to have some flexibility to deal with potential multi-modality in the distributions. The parameters of the model were trained on a set of hand annotated tracks from the data. Two different feature sets were used in the training process. The first set consisted of global features which were global second order curvature in the tracks to see how much they change directions, Fourier coefficients to see if there were sinusoidal frequency patterns in the movement, and track length to see how fast they swim. The purpose was to look at the tracks from a global view and see if the spermatozoa fell into different categories of movement behavior, and therefore forming a foundation for modeling based on data clusters - in this case using a hidden Markov model for each category of movement behavior. Using K-means for clustering with the Davies-Boulding validity index[13] as inter-intra cluster separation measure, we decided upon using five hidden Markov models, with each one trained and specialized on local features extracted from the tracks belonging to each of those behavioral clusters. The local features were velocity, acceleration, angle, and local second order curvature. The hidden Markov models were trained using the Baum-Welch algorithm[12] which is an instance of the Expectation-Maximization algorithm using the forward-backward algorithm. Since the particle filter and Kalman filers are using different features than the hidden Markov models, the observations, i.e. positions, were converted to the local hidden Markov model features before entering the model.

Each hidden Markov model was used to evaluate the likelihood of an observation $\tilde{\mathbf{z}}_t$ appended to a sequence of

hidden Markov observations $\tilde{\mathbf{Z}} = \{\tilde{\mathbf{z}}_1, ..., \tilde{\mathbf{z}}_{t-1}, \tilde{\mathbf{z}}_t\}$ given a set of model parameters $\boldsymbol{\theta}$

$$p(\tilde{\mathbf{Z}}|\boldsymbol{\theta}) = \sum_{i=1}^{P} p(\tilde{\mathbf{Z}}|\tilde{\mathbf{X}}_i, \boldsymbol{\theta}) p(\tilde{\mathbf{X}}_i|\boldsymbol{\theta}), \tag{3}$$

where P is the number of all possible paths $\{\tilde{\mathbf{X}}_1, ..., \tilde{\mathbf{X}}_P\}$ through state space given N observations, and model parameters $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{\phi}, \mathbf{B}\}$ consisting of state transitions $\mathbf{B}$, prior probabilities $\boldsymbol{\pi}$, and observation probabilities $\boldsymbol{\phi}$. Evaluating the likelihood in (3) for each observation $\tilde{\mathbf{z}}_t$ as the next position in the monitored tracks in each successive frame, results in a likelihood matrix $\mathbf{A}$ that can be used to decide the matching between observations and tracks. We used the Hungarian algorithm - a linear assignment algorithm - to make the assignment, using the likelihood matrix. Normally the Hungarian algorithm works with quadratic sized matrices, but we are also dealing with missing/spurious observations, and therefore need to work with rectangular matrices, and so the Hungarian algorithm was modified to work with rectangular matrices, for more details see the Master's thesis.[6] The formula for updating the likelihood matrix is

$$\mathbf{A}_{jk} = \max_m p(\tilde{\mathbf{z}}_1^k, ..., \tilde{\mathbf{z}}_{t-1}^k, \tilde{\mathbf{z}}_t^j|\boldsymbol{\theta}_m), \tag{4}$$

where A_{jk} is the likelihood of assigning the jth observation $\tilde{\mathbf{z}}_t^j$ to the kth track, and m iterates the five trained hidden Markov models, each having their own parameter set $\boldsymbol{\theta}_m$. The assignment is made more robust and faster by setting $A_{jk} = \infty$ for $||\mathbf{x}_{t-1}^k - \mathbf{z}_t^j||_2 > \nu$, where ν is a threshold indicating the maximal displacement per frame, derived from the velocity of the spermatozoa.

The advantage of using likelihoods modeled on track history instead of using for example Euclidean distance is apparent in occlusion situations and situations where the spermatozoa moves so close that their positions are hard to distinguish from each other.

2.3 Detection

Approximately 2/3 of the spermatozoa head appears as an almost circular bright blob in the phase-contrast images, and it is the position of these bright blobs that we wish to track. Based on these properties of the spermatozoa appearance, scale-space blob detection[14] seems a suitable choice for detecting where the spermatozoa reside in the current frame.

For each frame in an image sequence, a scale-space blob detector at a suitable scale is used for computing a filter response in each pixel. This response is converted into a probability of that pixel being the center of a spermatozoan head using an unnormalized Gaussian distribution, with parameters estimated from a set of typical spermatozoa heads. Based on the pixel probability map, likely spermatozoa positions are extracted from the image using a combination of thresholding and connected components. Both the extracted likely spermatozoa positions and the pixel probability map are used in the motion estimation process, the former as observations in the Kalman filter and the latter for forming the observation likelihood in the particle filter.

The scale-space representation of the image $\mathbf{M}(x, y)$ at scale σ^2 is defined by

$$L(x, y, \sigma^2) = g(x, y, \sigma^2) * \mathbf{M}(x, y),$$

where $*$ is the convolution operator and $g(x, y, \sigma^2)$ is a Gaussian kernel given by

$$g(x, y, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\Big(- \frac{(x^2 + y^2)}{2\sigma^2} \Big),$$

where the scale parameter σ^2 is the variance of the Gaussian kernel. A larger σ^2 means a broader kernel, which has the effect that the image is more smoothed by the convolution. Different objects in the image will only appear as meaningful entities at certain scale ranges.

A common operator for blob detection is the Laplacian of the Gaussian operator (LoG), given by

$$\nabla^2 g(x, y, \sigma^2) = \frac{1}{2\pi\sigma^6}(x^2 + y^2 - 2\sigma^2) \exp\Big(- \frac{(x^2 + y^2)}{2\sigma^2} \Big). \tag{5}$$

The operator has a deep negative cavity in the middle, surrounded by low positive hills. As a consequence of this shape, the result of convolving an image with the operator is a strong positive response for dark blobs and a strong negative response for bright blobs. There is not much variability in the size of the spermatozoa heads and therefore the detection is done only at one scale. Deriving the optimal scale can be done by monitoring the scale normalized operator response, also called the scale-space signature. For the LoG operator a normalized response is obtained my multiplying the operator response by the scale.[15]

Using the information provided by the scale-space signature, a filter having maximum response in the center of a blob $\mathbf{b} = [b_x, b_y]^T$ resembling a spermatozoan head can be expressed using an unnormalized Gaussian distribution

$$p_{blob}(\mathbf{b}|\mu_{head}, \sigma_{head}, \sigma^2_{max}) = \exp\Big(- \frac{(\sigma^2_{max}\nabla^2 L(b_x, b_y, \sigma^2_{max}) - \mu_{head})^2}{2\sigma^2_{head}}\Big), \tag{6}$$

where σ^2_{max} is the optimal scale, μ_{head} is the mean normalized response and σ_{head} is the standard deviation of the normalized response, both at the optimal scale. (6) is used when forming the observation likelihood in the particle filter described in 2.4.

2.4 Outline of our algorithm

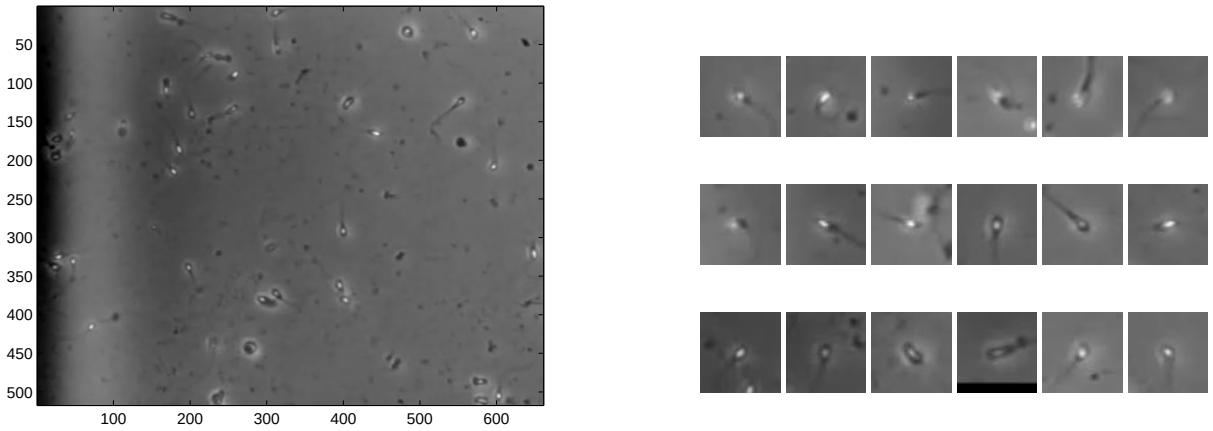
We model an object state by its position, velocity, and label, and let the particles be hypotheses about several objects at the same time. In each iteration, the spermatozoa positions, or observations, found in the detection step, see Section 2.3, are associated to object labels using the label handling step, which is further described in Section 2.2. We ensure that all particles sharing the same label are associated with the same observation. The estimated object states from the previous iteration are then updated using the three main steps in the particle filter algorithm described in Section 2.1. The particle filter is the main algorithm in our system that binds everything together, and therefore the outline of the algorithm follows the three main steps in the particle filter:

1. Importance sampling.
   (a) Sampling from proposal distribution. The extracted likely spermatozoa positions, or observations, are assigned to object labels by solving a linear assignment problem with (4) as cost matrix. Using Kalman filters with the particles, containing previous object states, and current associated observations, Gaussian proposal distributions are formed. Sampling from these proposal distributions yields updated object states for each of the particles. The effect is that the particles are both advanced in their current direction of movement and drawn in the direction of the observations. In cases where labels have no associated observation, only the prediction step in the Kalman filter is applied, resulting in the particles only being advanced in their current direction of movement.
   (b) Importance weights. The updated particles are weighted using (1), with the proposal distributions as described in (a) and the observation likelihood is formed by finding the maximal pixel probability and position within a local square neighborhood around the current object state position in the pixel probability map found using (6), and weighting the maximal probability exponentially by the distance to the object state position
2. Resampling. This step is just following the standard procedure. High/low weight particles are duplicated/suppressed and the resulting multiset of particles are given equal weight, yielding the final particle set for that iteration.

At any given iteration of the particle filter algorithm, we can estimate various quantities based on the particles. The most interesting is the mean of the state positions which are the spermatozoa tracks.

3. EXPERIMENTS AND RESULTS

The data set used in this study consists of digital images sequences of sperm samples recorded using a phase-contrast microscope at a magnification of $\times 100$ and a chamber of 10 μm depth. All the sequences used in this work are obtained at a frame rate of 25 frames/second. The image sequences stem from different subjects



(a) (b)

Figure 4. The data. (a) A frame from one of the image sequences in the data set. The spermatozoa are seen as bright blobs. Most often the tail can be seen as a thin, long, dark shadow extending from the dark part of head. (b) A closer look at some typical spermatozoa in the data set. approximately 2/3 of the head appears as very bright blob-like. The shape of the bright part varies from almost circular to elliptic.

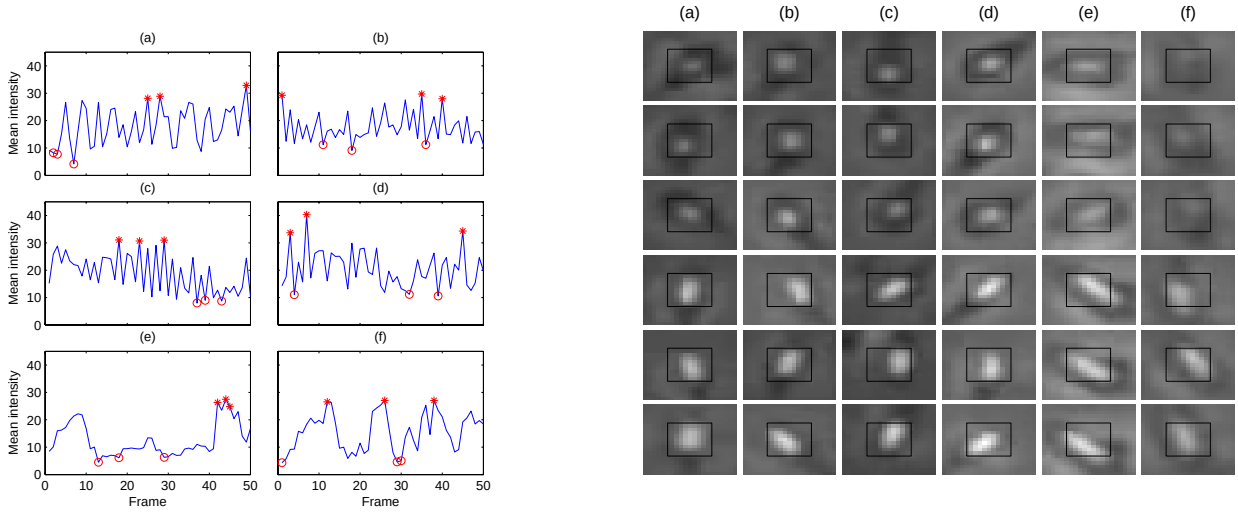
with varying sperm quality, thus the data represents great variety in the number of spermatozoa. Figure 4(a) shows a frame from an image sequence in the data set. The general intensity gradient seen in the frame appears in all frames, and is removed in our algorithms by simple background subtraction. Figure 4(b) shows typical spermatozoa appearances in the data set. In general the heads appear as bright blobs of varying shape, from almost circular to elliptic. There is also a bright halo around the head, which is an artifact due to phase-contrast microscopy, and the tail is not always visible. Our system tracks only the center of the bright part of the head, i.e. the center of the blob.

Apart from occlusions, interaction with other objects in the seminal fluid, and a generally noisy swimming behavior due to the fact that the spermatozoa whip their tail to make progress, which causes the head to oscillate about the main direction, the intensity of the spermatozoa head also has a tendency to drop significantly in some frames. A survey of this tendency was conducted by monitoring the mean intensity in a local neighborhood around the center of the head of six spermatozoa and the result is shown in Figure 5. In Figure 5(a) the intensity changes are clearly seen, and they also seem to be almost periodic. The appearance of the six spermatozoa heads at the three mean intensity highs and lows is shown in Figure 5(b). As seen, the heads almost attain the same intensity as the background in some cases. One explanation of this behavior is that the sperm samples were recorded in a 10μm chamber and therefore the spermatozoa can swim in and out of focus.

The optimal scale for detecting spermatozoa is determined by monitoring the mean of the normalized filter response in the center of the heads of typical spermatozoa, across different scales, as described in Section 2.3. The result is shown in Figure 7, where the mean response attains a maximum at approximately $\sigma^2_{max} = 1.3$ pixels. The mean response at this maximum is $\mu_{head} = 37$ and the standard deviation is $\sigma_{head} = 7$. In all tracking experiments 1000 particles were used, and $\nu = 20$ pixels. The settings for the matrices used in the Kalman filters, see Section 2.1.2, where the label is excluded from the state vector, i.e. $\mathbf{x}_t = [x_t, y_t, \dot{x}_t, \dot{y}_t]^T$, are

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & \tau & 0 \\ 0 & 1 & 0 & \tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{P}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where (x_t, y_t) is the state position of an object at time t, $(\dot{x}_t, \dot{y}_t)$ is the state velocity of an object at time t, and τ is the period of a time step, thus 1/25 seconds per frame.



(a) (b)

Figure 5. Semi-periodic intensity changes. (a) The mean intensity in a local neighborhood around the center of the head of six spermatozoa, monitored simultaneously in the same image sequence for 50 frames. The three highest mean intensities are marked with an asterisk and the three lowest with a circle. The corresponding appearances are shown in Figure 5(b). (b) The appearance of the six monitored spermatozoa when their mean intensity in a local window, here marked with black, is lowest/highest. The top-three rows show the three lowest mean intensities corresponding to circles in Figure 5(a) and the bottom-three rows show the three highest mean intensities corresponding to asterisks in Figure 5(a).

Three different image sequences from different sperm samples are used in evaluation of the system. The evaluation is done both by visual inspection of the output of the system and by comparing the estimated tracks to manually annotated tracks. Table 1 summarizes the characteristics of the sequences as well as the measured results. The mean squared error (MSE) is computed between the estimated tracks and the manually annotated tracks.

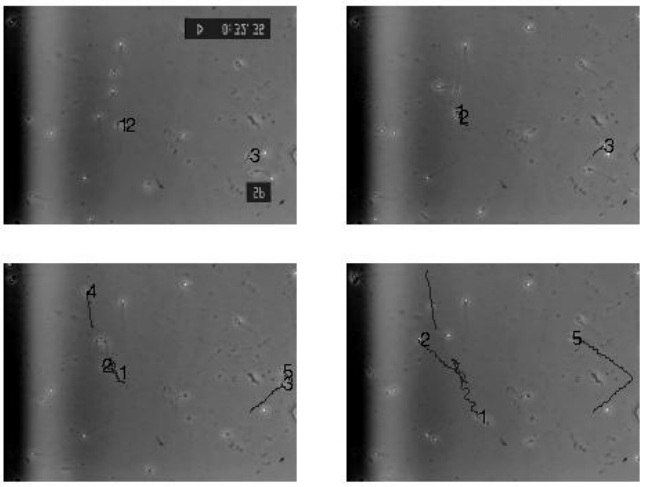
Sequence	Motile/non-motile spermatozoa frame 1	Num. frames	MSE (pixels)	Failed trackings (%)
IS1	19 (approximately)	50	2.0	10
IS2	33 (approximately)	50	1.2	4
IS3	45 (approximately)	50	1.6	16

Table 1. Characterization of the test data and measured results. The spermatozoa count is approximate due to the fact that it is sometimes problematic to distinguish between non-motile spermatozoa and round cells.

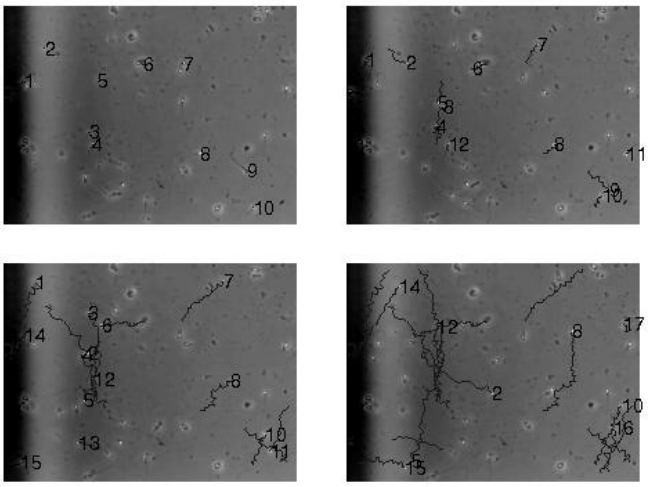
Positions in estimated tracks lying too far away from any positions in the manually annotated tracks (>5 pixels, the approximate radius of a spermatozoan head) do not contribute to the MSE, but are counted as tracking failures. The estimated tracks are also shown in Figure 6. All estimated tracks with a history below twenty-five successive frames are discarded, since the spermatozoa should be tracked for at least one second in order to give reliable measurements.[1]

4. DISCUSSION AND CONCLUSIONS

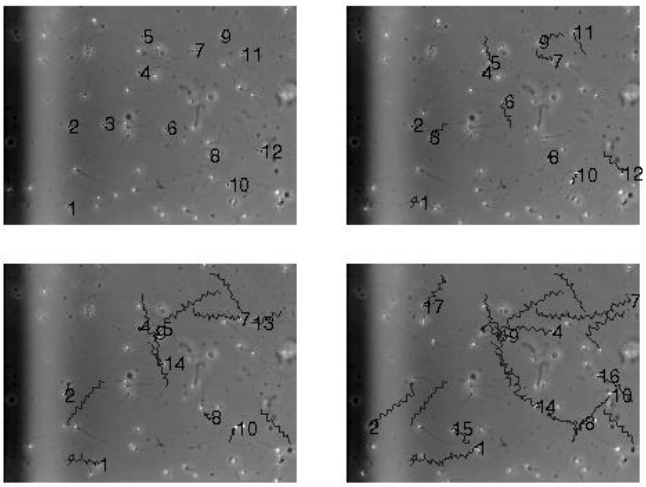
The estimated positions in all sequences are in general very close to the center of the heads (average MSE = 1.6 pixels) and therefore form a good foundation for computing common CASA motility variables like curvilinear velocity, straight-line velocity, average path velocity, and beat-cross frequency,[1] and in general other statistics. The systems ability to distinguish spermatozoa from background is also good. Only occasionally do background objects get detected as spermatozoa, giving rise to birth of objects with new labels, and these are always discarded due to short track histories. Not all occlusion situations are handled perfectly, and so there are situations where



(a) IS1.



(b) IS2.



(c) IS3.

Figure 6. Estimated spermatozoa tracks (solid lines) in IS1, IS2, and IS3, superimposed on the images. The numbers are the object labels. Frames 1, 10, 25, and 50 are shown.

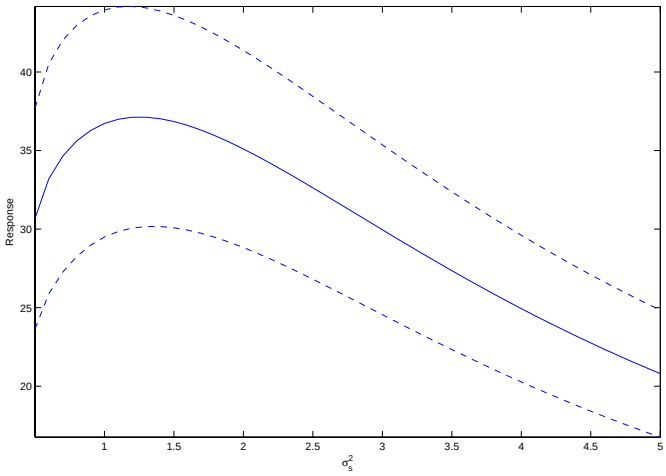


Figure 7. Mean responses at the center of the head of typical spermatozoa segments in the data. The solid line shows the normalized LoG operator response and the mean plus/minus one standard deviation of the normalized responses are shown as blue dashed lines.

label swaps occur, meaning that two or more particle labels are associated with the wrong spermatozoa track. These label swaps generally seem solvable when considering the track history prior to the occlusion. But there are also occlusions that are hard to resolve, even for the human eye, for example in IS3 shown in Figure 6(c), where the spermatozoa concentration is high. If only general characteristics about the velocity of the spermatozoa in a sample are needed, for example the mean curvilinear velocity of all motile spermatozoa in an image sequence, then label swaps are not a problem. However, for other characteristics it could be a problem.

To further improve the system we need to add some more intelligent postprocessing, or augment the proposed system with some higher-level inference for resolving label swaps. One idea is to use the hidden Markov models to evaluate the likelihood of the estimated tracks after occlusion and compare against the likelihoods obtained by switching the labels after the occlusion. If the resulting switched tracks have higher probability, then make the switch permanent.

To conclude we have presented a CASA system utilizing various advanced statistical models for multi-object tracking of human spermatozoa. The system was able to track most of the motile spermatozoa in three image sequences with the estimated position very close to the position indicated by a human observer (average MSE = 1.6 pixels).

ACKNOWLEDGMENTS

The authors would like to thank Visiopharm* for providing the contact to the University Department of Growth and Reproduction at Rigshospitalet and for helping in the early stage the project. We would also like to thank Niels Jørgensen and colleagues at the University Department of Growth and Reproduction at Rigshospitalet† for providing data.

REFERENCES

1. WHO, *WHO Laboratory Manual for the examination of human semen and sperm-cervical mucus interaction, fourth edition.* World Health Organization, 1999.

*http://www.visiopharm.com/
†http://www.reproduction.dk/

2. J. Verstegen, M. Iguer-Ouada, and K. Onclin, "Computer assisted semen analyzers in andrology research and veterinary practice," *Theriogenology* **57**, pp. 149–179, January 2002.
3. M. Berezansky, H. Greenspan, D. Cohen-Or, and O. Eitan, "Segmentation and tracking of human sperm cells using spatio-temporal representation and clustering," in *Proc. SPIE Medical Imaging 2007: Image Processing*, 2007.
4. E. Meijering, I. Smal, and G. Danuser, "Tracking in molecular bioimaging," *Signal Processing Magazine, IEEE* **23**(3), pp. 46–53, 2006.
5. I. Smal, K. Draegestein, N. Galjart, W. Niessen, and E. Meijering, "Rao-blackwellized marginal particle filtering for multiple object tracking in molecular bioimaging," *Inf Process Med Imaging* **20**, pp. 110–121, 2007.
6. J. Østergaard and L. Sørensen, "Multi-object tracking of human spermatozoa using a particle filter," Master's thesis, Dept. of Computer Science, University of Copenhagen, Denmark, 2007, ftp://ftp.diku.dk/diku/image/publications/oestergaard.060513.pdf.
7. M. Arulampalam, M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing* **50**, pp. 174–188, February 2002.
8. M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," *International Journal of Computer Vision* **29**(1), pp. 5–28, 1998.
9. R. van der Merwe, N. de Freitas, A. Doucet, and E. Wan, "The unscented particle filter," Tech. Rep. CUED/F-INFENG/TR380, Cambridge University Engineering Department, August 2000.
10. G. Welch and G. Bishop, "An introduction to the kalman filter," Tech. Rep. TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 2006.
11. R. A. Pilgrim, "Munkres' assignment algorithm. modified for rectangular matrices." Course Notes, Murray State University.
12. L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE* **77**(2), pp. 257–286, 1989.
13. D. Davies and D. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Analysis and Machine Intelligence* **1**(2), pp. 224–227, 1979.
14. T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Norwell, MA, USA, 1994.
15. T. Lindeberg, "On scale selection for differential operators," in *Proc. 8th Scandinavian Conference on Image Analysis*, pp. 857–866, (Tromso, Norway), May 1993.