Automatic Tracking and Motility Analysis of Human Sperm in Time-Lapse Images

Leonardo F. Urbano, Puneet Masson, Matthew VerMilyea, and Moshe Kam*

***Abstract*— We present a fully automated multi-sperm tracking algorithm. It has the demonstrated capability to detect and track simultaneously hundreds of sperm cells in recorded videos while accurately measuring motility parameters over time and with minimal operator intervention. Algorithms of this kind may help in associating dynamic swimming parameters of human sperm cells with fertility and fertilization rates. Specifically, we offer an image processing method, based on radar tracking algorithms, that detects and tracks automatically the swimming paths of human sperm cells in timelapse microscopy image sequences of the kind that is analyzed by fertility clinics. Adapting the well-known joint probabilistic data association filter (JPDAF), we automatically tracked hundreds of human sperm simultaneously and measured their dynamic swimming parameters over time. Unlike existing CASA instruments, our algorithm has the capability to track sperm swimming in close proximity to each other and during apparent cell-to-cell collisions. Collecting continuously parameters for each sperm tracked without sample dilution (currently impossible using standard CASA systems) provides an opportunity to compare such data with standard fertility rates. The use of our algorithm thus has the potential to free the clinician from having to rely on elaborate motility measurements obtained manually by technicians, speed up semen processing, and provide medical practitioners and researchers with more useful data than are currently available.**

***Index Terms*— Computer assisted semen analysis (CASA), human sperm imaging, JPDAF, sperm motility, sperm tracking.**

I. Introduction

HUMAN sperm motility is of great interest to biologists studying sperm function and to medical practitioners evaluating and treating male infertility [1]–[3]. Approximately one in six couples in the United States has fertility problems [4], and many of them seek diagnostic semen analysis to help assess the cause. Common parameters evaluated during a typical semen analysis include sperm concentration, total sperm number, percentage of motile sperm, percentage of sperm exhibiting forward progression, and percentage of sperm having normal morphology. Animal sperm are also routinely examined and studied by theriogenologists and agriculturalists engaged in industrial animal husbandry.

Today, the prevailing method for analyzing sperm at fertility clinics and research laboratories is laborious and subjective [5]. Typically, technicians use microscopes to count sperm cells manually and to appraise the quality of sperm movement visually in accordance with elaborate standard protocols [6]. In some clinics, more objective computer-assisted semen analysis (CASA) instruments are used to trace the swimming paths of sperm automatically in time-lapse microscopy image sequences [7]. Using digital image processing algorithms, these instruments "connect-the-dots" (i.e., connect the coordinates of detected sperm cells) between video frames, enabling rapid automatic quantification of sperm motility parameters for hundreds of cells at a time.

The exact importance of parameters commonly measured by CASA is not yet known, and the relationship between sperm swimming patterns and fertility remains an open biological question. Progress in this area requires collection and analysis of sperm motility parameters over time and on a large scale. Such data are not currently provided by today's CASA instrumentation.

A. Problem Statement

It is widely recognized that the image processing and sperm tracking capabilities of today's CASA instruments can be enhanced [1], [8]–[10]. In particular, most CASA instruments cannot reconstruct reliably the paths of two or more sperm swimming in close proximity or whose paths intersect. This shortcoming requires substantial user invention and tuning. Often, the trajectories of a significant number of sperm involved in cell-to-cell collisions and near-misses are excluded from analysis by CASA algorithms. Since higher velocity sperm are more likely to be involved in apparent collisions, their exclusion tends to bias motility measurements toward slower sperm. To reduce the probability of cell collisions, samples are often diluted (typically to less than 20×10^6 sperm/mL) or analysis is limited to short video clips (typically less than 1–2 sec). Finally, CASA motility analysis can vary significantly when different system settings are used [11], [12], making it difficult to compare results obtained with the same instrument in different studies. These difficulties, combined with the prohibitively high cost of acquiring and maintaining a CASA instrument, have often outweighed the benefits of automation.

Robust multi-target tracking algorithms, developed originally for radar applications and video processing, have addressed similar challenges successfully in other

Manuscript received September 5, 2016; revised November 7, 2016; accepted November 8, 2016. Date of publication November 18, 2016; date of current version March 2, 2017. *Asterisk indicates corresponding author.*

L. F. Urbano is with the Department of Electrical and Computer Engineering and Data Fusion Laboratory, Drexel University, Philadelphia, PA 19104 USA (e-mail: leonardo.f.urbano@gmail.com).

P. Masson is with the Penn Fertility Care, Hospital of the University of Pennsylvania, Philadelphia, PA 19104 USA.

M. VerMilyea is with Texas Fertility Center, Austin, TX 78731 USA (e-mail: matthew@austinivf.com).

*M. Kam is with the New Jersey Institute of Technology, University Heights Newark, Newark, NJ 07102 USA (e-mail: kam@njit.edu).

Digital Object Identifier 10.1109/TMI.2016.2630720

domains [13], and could potentially help improve CASA. Over the years, interest in applying such algorithms to track viruses, bacteria, stem cells, sub-cellular organelles and other biological particles has increased [14]. In this paper, we discuss the adaptation and application of such methods to multi-sperm tracking and motility analysis.

B. Main Contribution– Multi-Sperm Ensemble Tracking

We have developed a fully automated, robust, multi-sperm tracking algorithm. It has the demonstrated capability to detect and track simultaneously hundreds of sperm cells in recorded videos while accurately measuring motility parameters over time and with minimal operator intervention. Sperm measurement-to-track association conflicts occurring during real and apparent cell-to-cell collisions were reconciled by adapting and applying the joint probabilistic data association filter (JPDAF) [15], representing a mature technology employed in air traffic control systems. This approach uses independent Kalman filters to estimate the position and velocity of each sperm tracked, and enables accurate tracking of sperm in undiluted specimens over periods significantly longer than 1 sec (typically 15–45 sec at 200× magnification).

Using the swimming paths reconstructed by our algorithm, we calculated values of eight important sperm motility parameters: straight-line velocity (VSL), curvilinear velocity (VCL), average path velocity (VAP), linearity of forward progression (LIN), curvilinear path wobble (WOB), average path straightness (STR), amplitude of lateral head displacement (ALH), and mean angular displacement (MAD). Unlike most CASA instruments which measure motility parameters only once per sperm, our algorithm continually measured all parameters along the entirety of each reconstructed sperm trajectory, revealing interesting temporal changes in individual sperm motility not evident in population statistics. The biological significance of the observed swimming pattern transitions is yet unknown, but our approach provides the opportunity to study them at low cost, using thousands of data points.

C. Previous Work

Katz and Davis pioneered automatic sperm tracking in the mid-1980s [16]. Although commercial CASA instruments often use proprietary tracking algorithms, it is evident that most are still broadly based on the original technique [7]. Specifically, a user-selected gray level threshold is globally applied to all video frames to identify pixels belonging to sperm, and the centroids of the resulting segmented blobs are taken as measured sperm positions. To track sperm, a circular measurement validation region (or "gate") centered at each measurement in one frame is used to select a path-continuing measurement from the next frame. The user chooses the gate radius (typically in proportion to the maximum expected sperm head displacement between any two video frames) which is applied globally across all frames and all tracks. This method works well for tracking a small number of well-separated targets in the absence of clutter, but its effectiveness rapidly degrades if targets become closely spaced and their overlapping validation gates contain multiple conflicting measurements. In these cases, CASA instruments often exclude the affected tracks from analysis or attempt to continue the track by selecting the nearest-neighbor (NN) measurement.

Beresford-Smith and Van Helden [17] first applied radar tracking algorithms to sperm tracking by adapting the probabilistic data association filter (PDAF) to track a single sperm in clutter, but reported no experimental data. Recent ad-hoc methods for tracking single sperm through collisions in low-density samples were reported by Shi et al. [18], and Liu et al. [19]. In [20], Tomlinson et al. described a CASA system based on multi-target tracking algorithms that can track multiple sperm and grade their motility using 1-sec video clips. However, the problem of tracking through collisions or over long durations was not addressed. In [21], Su et al. used a lens-free holographic imaging sensor to track the 3D swimming paths of sperm for 10–20 seconds in highly diluted sample preparations (less than 10×10^6 sperm/mL). Their method requires a significant computational cost (greater than 2 hours of post-processing). In [22], Berezansky et al. applied mean shift and optical flow for sperm detection and tracked sperm cells, but the run time was significant and the method cannot detect non-moving sperm. In [23], Sorensen et al.studied multi-sperm tracking using both Kalman and particle filters.

None of the studies cited above has presented a systematic validation of the proposed sperm-tracking algorithm, and therefore it is difficult to assess and compare performances. However, methods for validating multi-target tracking algorithms have been in development since the 1970s. The recent work of Ristic et al. [24] summarizes these methods along with suggestions for consistent assessment of tracking algorithm performance. We have used the approach suggested in [24], known as the optimal subpattern assignment (OSPA) distance metric for labeled tracks, in assessing our algorithm's performance.

D. Organization of the Paper

This paper is organized as follows. In Section II we describe the basic sperm segmentation and spatiotemporal tracking algorithm, and show representative images that demonstrate its operation. We also describe key motility parameters (VCL, VSL, LIN, ALH, VAP, WOB, STR, and MAD) and how they are extracted. In Section III we describe results from applying the algorithm to human sperm samples. Section IV describes scenario modeling and algorithm validation. The Appendix summarizes implementation details of the JPDAF algorithm used for our experiments. Sperm tracking videos produced by our algorithm are available at http://vimeo.com/123584621.

II. Methods

A block diagram of the algorithm is shown in Figure 1(a) (the implementation currently uses MATLAB). The input to the algorithm is a sequence of time-lapse images currently encoded either as an MP4 or AVI video file. The output of the algorithm is a database file containing the set of reconstructed swimming paths for each sperm tracked and corresponding values for a host of parameters of interest. These data are

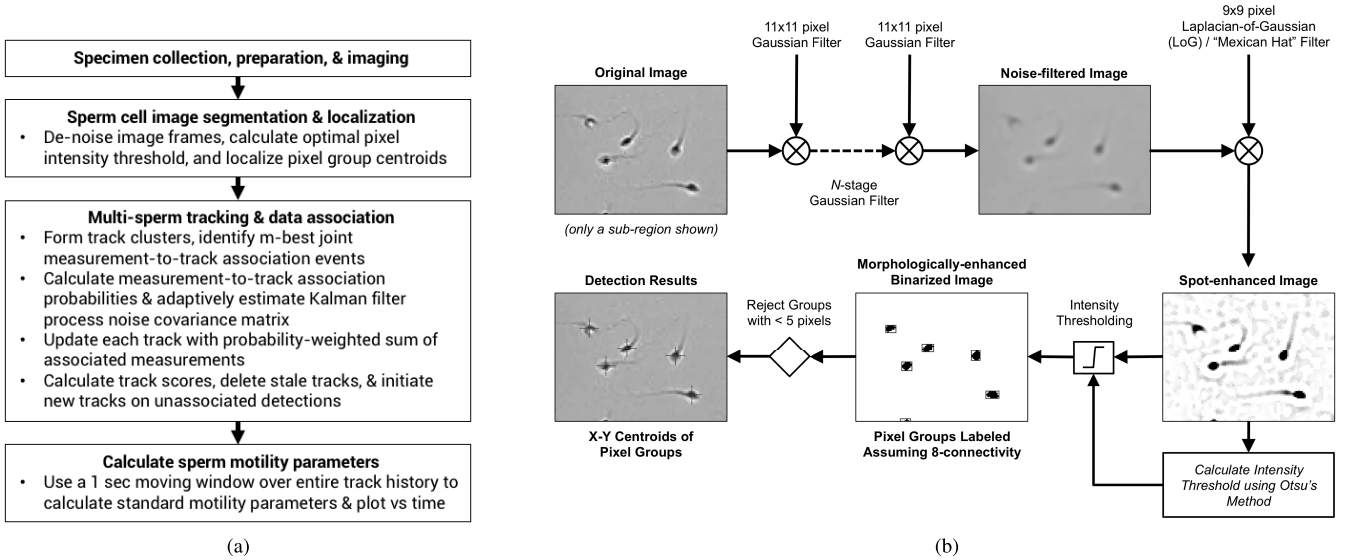


Fig. 1. (a) Algorithm block diagram. (b) Illustration of the sperm segmentation and localization algorithm. Each original image frame is successively convolved with a Gaussian filter to suppress image noise, and then a spot-enhancing Laplacian-of-Gaussian (LoG) filter is applied and binarized using an automatically-calculated intensity threshold. Spurious detections are reduced by morphological erosion and dilation and the centroids of pixel groups having <5 pixels are regarded as detected sperm heads. This process is repeated for all video frames and the set of resulting detections is the input to the tracking algorithm.

visualized as parameter scatter plots, trajectory path histories, and annotated movies.

A. Specimen Collection and Preparation

Human semen specimens for our study were collected and processed by the In-Vitro Fertilization laboratories at Penn Fertility Care in accordance with policies of the University of Pennsylvania. Each specimen was allowed to liquefy for 30-40 min at room temperature and was then washed in media. Washing separates sperm cells from seminal plasma and consists of three cycles of centrifugation. First, ≤ 3 mL of semen is layered onto a 1.0 mL preparation of density gradient media (0.9 mL PureSperm100, Spectrum Technologies and 0.1 mL Enhanced WG, modified human tubal fluid with HAS and Gentamicin, Conception Technologies) and is centrifuged for 20 min at 400 g. The pellet is resuspended in 2 mL Enhanced WG and centrifuged again for 5 min at 400 g. Lastly, the supernatant is discarded and the pellet is resuspended in 0.5 mL Enhanced WG and 5 μL is pipetted onto a 20 μm deep Vitrolife MicroCell chamber. A microscope stage warmer is used to regulate the temperature of each sample at 37 °C during phase contrast imaging.

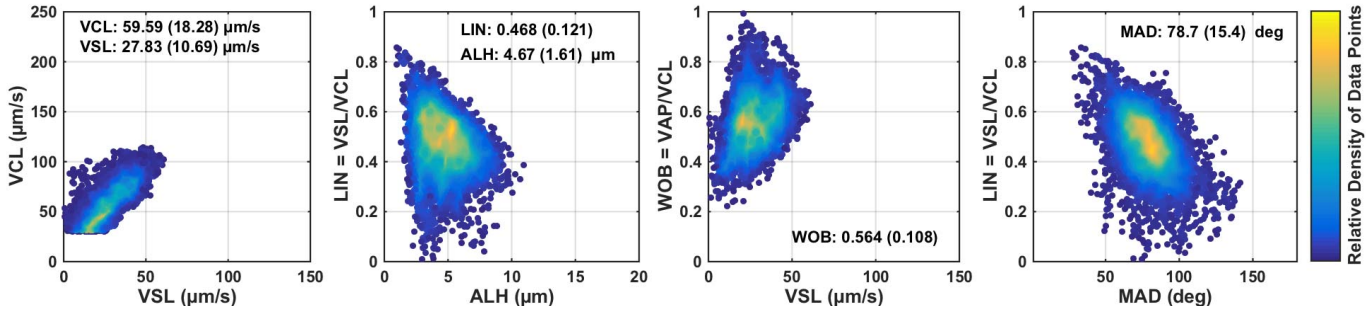
B. Sperm Segmentation and Localization

To identify and localize sperm cells automatically in phase contrast images, a custom-made image processing algorithm was used (it is currently written in MATLAB). Our laboratory video capture equipment records 640×480 pixel video frames at 15 frames per second (fps) at $200\times$ magnification (0.857μm/pixel). In these images, each sperm head appears as a dark oval spot surrounded by a white horseshoe-shaped halo interrupted where the tail joins the head. To identify pixels belonging to sperm heads, we use an algorithm based on intensity thresholding combined with a sequence of image-enhancing pre-processing steps (see Figure 2(b)).

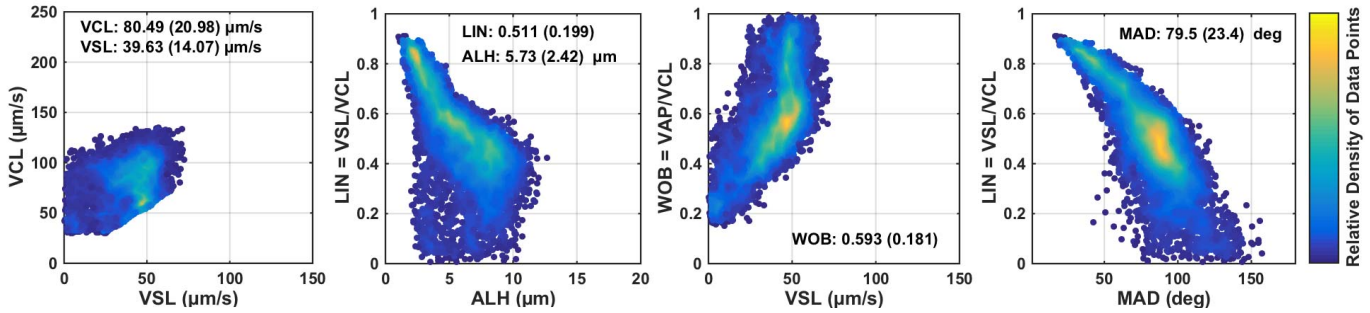
The algorithm first convolves successively the original image with an 11×11 pixel Gaussian filter (commensurate with the size of a sperm head) to reduce image noise and soften halo edge pixels. The Gaussian filter is applied 5 times, but the number of iterations can be controlled by the user. Spot-enhancement is achieved using a 9×9 pixel Laplacian-of-Gaussian (LoG) ("Mexican-hat") filter to increase the contrast between the sperm head and its surrounding halo.[1] Next, the spot-enhanced image is binarized using an intensity threshold calculated using Otsu's method [25] and multiplied by a user-specified weighting factor. This filtering is done for each frame separately to adapt to frame-to-frame variation in field illumination. To reduce spurious detections, the binarized image is morphologically eroded and dilated with a 5×5 and 3×3 diamond structuring element, respectively. Finally, groups of pixels are labeled assuming eight-connectivity, and any group having less than 5 pixels is considered a non-sperm particle and discarded. The centroid of each pixel group is calculated and regarded as a sperm position measurement.

Using this method, sperm cell detection accuracy is approximately 95% and false detection rate is less than 1% (estimates based on comparison with manual detection of sperm in five randomly chosen video frames). In total, the sperm segmentation algorithm has two parameters: the number of iterations of the convolution of the Gaussian kernel for low-pass filtering, and a user-specified weighting factor for the Otsu threshold on the de-noised data. These were set manually using a test video, and did not have to be readjusted for subsequent videos

[1]Since they are linear filters, the successive application of a Gaussian and LoG filter can be combined into a single filter stage to speed up processing.



(a) Sperm motility parameters for Specimen A (440 total sperm tracked, 409 motile) in 60 sec video clip



(b) Sperm motility parameters for Specimen B (277 total sperm tracked, 202 motile) in 60 sec video clip

Fig. 2. Sperm motility data for two Specimens (A and B) of human sperm. Measured sperm motility parameters VCL, VSL, VAP, LIN, WOB, ALH, and MAD for motile sperm (i.e., sperm with mean VCL $> 20\mu$m/sec). The colorbar encodes the relative density of the data points. Each motility parameter was measured continuously for each sperm tracked, but only the measurements in the first *5* sec of each track were used to populate the figures and calculate statistics. Mean and standard deviation of motility parameters are indicated in each scatter plot. Colorbar indicates relative density of data points, blue is low density and yellow is high density.

produced from the same microscope. A 1-minute long video at 15 fps (900 frames) is processed by this algorithm in 444.2 sec ($\approx$ 2 frames/sec) using a 1.3 GHz Intel Core i5 processor with 4 GB 1600 MHz DDR3 RAM.

C. Multi-Sperm Tracking and Data Association

Nearly all commerical CASA instruments use a nearest neighbor (NN) tracking scheme in which the nearest measurement to a track is assumed to be correct. Alternatively, the global nearest neighbor (GNN) scheme finds the optimal assignment of measurements to tracks that minimizes the sum of the pair-wise distances [13]. Both NN and GNN assume perfect measurement association [15]. The multi-hypothesis tracker (MHT) and the JPDAF [15] are more sophisticated and the most well-known algorithms for multi-target tracking. Both use independent Kalman filters to estimate the state and covariance of each target tracked. The JPDAF uses only the measurements in the current frame to evaluate measurement-to-track correspondence, unlike the MHT which uses the entire past measurement history. The JPDAF is essentially an approximation of the MHT with a time depth of one frame. Choosing an algorithm is a trade-off between run-time and tracking accuracy. For rapid analysis of sperm sample motility, the MHT is likely to be more robust but less practical than the JPDAF since its memory requirements grow rapidly with increasing tracking duration and number of targets. In this study, we implemented a custom-made adaptation of the JPDAF algorithm in MATLAB.

In the JPDAF, each track is updated using a probability-weighted sum of all its validated measurements at the current time. Unlike the PDAF, which does not consider the existence of other targets, the JPDAF calculates the association probabilities jointly across all measurements and targets and involves identifying all feasible joint association events (i.e., all possible permutations of assignments of measurements to tracks or to clutter)– which is NP-hard [26], [27]. To speed up processing, we group collocated measurements and tracks into clusters, using Kusiak's algorithm [28] which is then processed separately and independently by the JPDAF. Instead of identifying exhaustively all feasible joint association events required by the pure JPDAF, we use Cox's adaptation [29] of Murty's original method [13] to identify only the m-best (i.e., most highly probable) joint association events. Track clustering and m-best event ranking make the JPDAF practical for tracking sperm cells in near real-time.

Since sperm cells can exhibit abrupt maneuvers, each independent Kalman filter uses a continuous white noise acceleration (CWNA) target motion model [13] with an adaptive process noise covariance. The process noise covariance is estimated over time using a fading memory filter [30], and allows the Kalman filter track covariance to "breathe" (i.e., increase during maneuvers and shrink during quiescent swimming). This dynamic process noise helps reconcile measurement-to-track association conflicts between sperm having different swimming speeds and patterns.

Additional details of the JPDAF implementation, including track management, are provided in the Appendix.

D. Calculation of Motility Parameters

We calculate eight standard sperm motility parameters defined in [6]: 1) *Curvilinear velocity* (VCL, μm/sec) is the sum of the distances between each measured sperm position divided by the analysis time, 2) *Straight-line velocity* (VSL, μm/sec) is the straight-line distance between the first and last sperm position divided by the analysis time, 3) *Average path velocity* (VAP, μm/sec) is the time-averaged velocity of a sperm head along its average path, 4), *Linearity of forward progression* (LIN = VSL/VCL, dimensionless) is the linearity of the curvilinear path, 5) *Path wobble* (WOB = VAP/VCL, dimensionless) is a measure of oscillation of the actual path about the average path, 6) *Straightness* (STR = VSL/VAP, dimensionless) is a measure of the linearity of the average path, 7) *Amplitude of Lateral Head Displacement* (ALH, μm) is the average distance of the sperm head from the average sperm-swimming path where the average path is calculated using a 5-point moving average, and 8) *Mean angular displacement* (MAD, degrees) is the time-averaged absolute value of the instantaneous turning angle of the sperm head along its curvilinear trajectory.

Motility parameters are calculated using the set of position measurements associated to a track over its entire track history and are made available as a database file for post-processing and cluster analysis. The first and last 5 points of the trajectory are discarded from analysis to prevent track initiation and track termination artifacts from corrupting the motility calculations. A 5-point moving average is used to low-pass filter noisy signals when plotting individual parameters vs. time. In population statistics, motility analysis is limited to 5 sec per sperm.[2]

Despite our microscopic video recording equipment being limited to a maximum frame rate of 15 fps, our algorithm is not constrained to operate only at 15 fps. In fact, we have successfully applied our algorithm to track sperm in a limited number of 60 fps videos provided during the initial stages of another sperm tracking research project. To process these videos with 4× higher frame rate, we did not need to make any algorithm parameter modifications except to specify the higher frame rate value. This algorithm robustness to video frame rate is attributed to the fact that the continuous white noise acceleration (CWNA) target dynamics model used in the elemental Kalman filters of our tracker explicitly depends on the time period between consecutive video frames (see Eq. 3 and Eq. 4 in Appendix A).

E. Objective Tracking Evaluation Using the OSPA Metric

Evaluating the performance of a biological cell tracking algorithm is challenging because the ground-truth positions of real cells in each video frame are not known a-priori. The prevailing method to address this evaluation challenge is to apply a given tracking algorithm to synthetic image sequences or a set of synthetic detection sequences (bypassing the segmentation portion of the algorithm) where ground truth tracks are known. An objective metric such as the OPSA distance [24] can then be used to quantify performance in terms of three tracking errors: (1) cardinality errors (having too few or too many estimated tracks versus ground truth tracks), (2) localization errors (root mean square spatial distance between OSPA-paired estimated and ground-truth tracks), and (3) labeling errors (track number assignment swaps).

In this paper, we evaluated the tracking algorithm by applying it to a set of synthetic measurement sequences representing four different scenarios having multiple simulated moving objects in a 500 μm^2 field-of-view. Imperfect image segmentation was simulated by applying isotropic white Gaussian measurement noise with 2 μm standard deviation to the ground-truth object positions, and each simulated object was detected with a probability of $P_D = 0.95$. Random detections due to over-segmentation were simulated by injecting Poisson distributed random detections with a spatial density of $10^{-5}/\mu\text{m}^2$. Synthetic measurement sequences were used as input into the JPDAF sperm tracking algorithm, to produce estimated trajectories over time in 100 Monte Carlo replications per scenario. The OSPA distance from ground-truth tracks was calculated using standard OSPA parameters ($p = 1$, $l = 25$ μm, and $c = 50$ μm).

Using the OSPA metric, we compare our JDPAF implementation to a simple nearest-neighbor (NN) association scheme used in standard CASA, a global nearest neighbor (GNN) scheme [13] representing a typical deterministic association algorithm, and the PDAF. The results of the comparison are presented in Section III.

III. Results

In this section, we present sperm trajectories and motility analysis for sample populations and individual sperm cells obtained from real image sequences collected by a fertility clinic using the method described in Section II. We also present the results of an objective comparison of four different data association algorithms using synthetic detection sequences.

A. Motility Analysis of Sperm Populations

In Figure 2 we present measured sperm motility parameters for motile sperm (i.e., sperm with mean VCL > 20 μm/sec) for two specimens collected from two different human subjects. Each motility parameter was measured continuously for each sperm tracked. Parameter mean and standard deviations are indicated in each subplot and the colorbar encodes the relative density of the data points shown.

Manual analysis of the two specimens was performed by two fertility technicians following standard protocols. The technicians estimated an average sperm concentration of 36.75×10^6 sperm/mL with 92% motile sperm for specimen (a) and 18.25×10^6 sperm/mL with 97% motile sperm for specimen (b). In comparison, the automatic sperm tracker described in Section II estimated 93% (409 out of 440 total tracked) motile sperm for specimen (a) and 73% (202 motile sperm

[2]In general, track lengths are longer for slower sperm than for faster sperm which leave the field of view. To limit the impact of including a disproportionate number of data points belonging to slow moving sperm in population statistics, only measurements collected in the first 5 sec of each track are used.

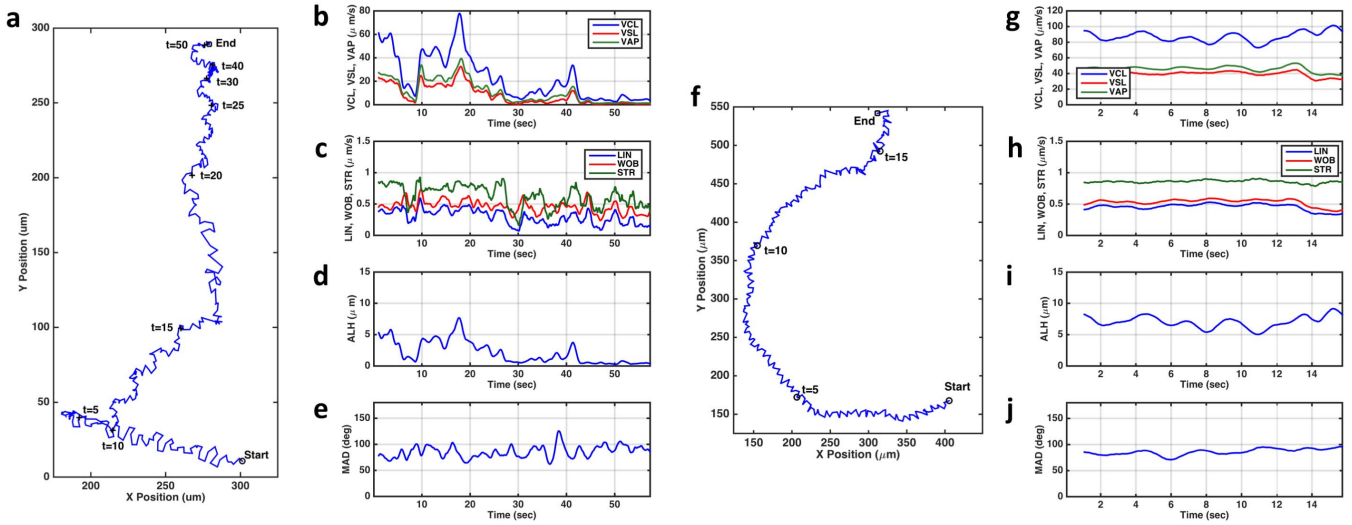


Fig. 3. Changes in individual sperm motility over time. Track histories (a, f) and motility parameters over time (b-e, g-j) for two different sperm with different swimming patterns. The sperm in (a) is tracked for 825 frames (55 sec) and its motility parameters are shown in panels (b-e). The sperm in (f) is tracked for 225 frames (15 sec) and its motility parameters shown in panels (g-j). The sperm in (a) exhibits a gradual but marked reduction in VCL, VSL, and VAP during the last 45 sec of swimming, while the sperm in (b) has relatively constant motility throughout the tracking duration.

out of a total of 277 tracked) motile sperm for specimen (b). The collected parameters suggest that sperm in specimen (a) are slower than sperm in specimen (b). Specifically, the sperm in specimen (a) have a mean VCL of 60 μm/sec compared to 80 μm/sec in specimen (b). Average LIN and ALH were also lower for specimen (a) at 47% and 4.7 μm compared to specimen (b) at 51% and 5.7 μm, suggesting sperm in specimen (b) exhibit more vigorous flagellar beating than (a). The mean WOB for both specimens lie within 5% of each other, but the spread– represented by the standard deviation– of WOB values for specimen (b) is twice that of specimen (a). Likewise, the mean MAD for both two specimens lie within 1% of each other, but the spread of MAD values for specimen (b) is about 30% larger than specimen (a).

B. Analysis of Sperm Swimming Transitions

The swimming patterns of individual sperm are known to change over time, but the significance of such transitions is not well understood. An advantage of an algorithm capable of tracking multiple sperm through collisions is the ability to collect long track histories (i.e., tens of seconds) which may reveal idiosyncracies in individual swimming patterns. We demonstrate this capability in Figure 3, which shows trajectories and motility parameter time histories for two different sperm tracked in specimen (a) of Figure 2.

Specifically, Figure 3 shows the spatiotemporal track histories (2D position coordinates and time) of two different sperm having different swimming patterns and their corresponding motility parameters over time. The trajectory of the sperm in (a) takes place over 825 frames (55 sec) and shows initially movement in the negative X direction for the first 5 to 10 seconds and then movement in the positive Y direction for the last 45 seconds. Figure 3 (b-d) show that during the last 45 seconds there is a gradual but very marked decrease in velocity parameters VCL, VSL, and VAP, in flagellar beating parameter ALH, in average path straightness parameter STR, and in linearity of forward progression parameter LIN. In fact, VCL peaks at 80 μm/sec at approximately 18.5 sec, and then decreases to 10 μm/sec at 55 sec; a brief burst of swimming is evident at the 40 sec mark before the cell becomes nearly immotile.

Another example is given in Figure 3 (f). The trajectory of the sperm in (f) takes place over 225 frames (15 sec) and traces a clockwise circular arc with a radius of approximately 150 μm. Figure 3(g-j) show that during this period the velocity parameter VCL is nearly sinusoidally modulated about 90 μm/sec with an amplitude of $\approx$ 10 μm/sec and a period of $\approx$ 2.1 sec. VAP and VSL are also nearly sinusoidally modulated between 40 and 50 μm/sec, average path straightness STR is relatively constant at a value of 0.85, and ALH oscillates between 6 and 9 μm. MAD is relatively constant at 90 deg.

C. Representative Examples of Multiple Sperm Tracking

Representative snapshots of the JPDAF described in this paper applied to multiple sperm tracking are shown in Figure 4 (a-d). Each snapshot is a 150 μm^2 intensity-inverted sub-region of its original phase contrast video image frame. In each snapshot, red dots indicate raw position measurements for each sperm over an interval of 1 sec connected by a green line. Blue lines represent corresponding Kalman-filter estimated trajectories. The number next to each sperm is the track number. Red ellipsoids represent the 3-sigma predicted position estimation error covariance for each sperm, which is also the validation gate used in the JPDAF. The four snapshot sequences illustrate different aspects of the tracking algorithm.

Sequence A in Figure 4 shows three sperm of interest (track #34, #192, and #313) over a period of 25 frames (1.6 sec). Sperm #34 and #192 are immotile, but sperm #313 is motile

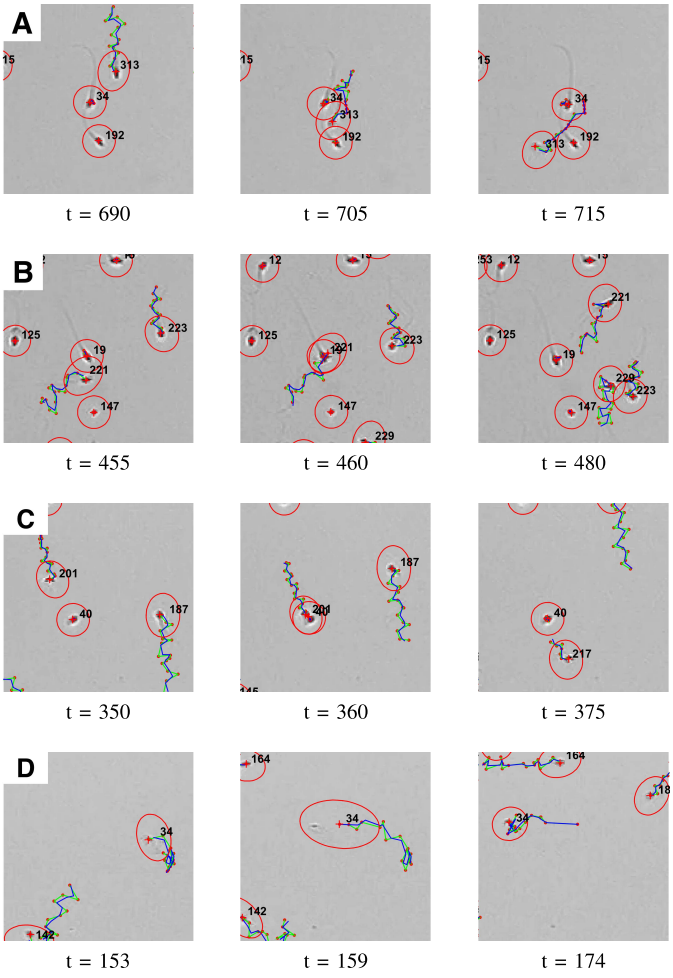


Fig. 4. Snapshots of multi-sperm tracking. (A) Tracking three sperm swimming in close proximity. (B) Tracking of motile sperm colliding with non-motile sperm. (C) Tracking failure (lost track) due to measurement coalescence. (D) Tracking an abruptly accelerating sperm by using adaptive process noise in a Kalman filter.

and moving in close proximity to both, and passes between them. During this close proximity swimming, no tracking errors are incurred (i.e., all sperm maintain their original track number labels throughout the tracking).

Sequence B in Figure 4 shows successful tracking through a collision between motile sperm #221 and immotile sperm #19 over a period of 25 frames. Despite complete occlusion of the two sperm at frame t=460, the JPDAF maintains distinct tracks of both sperm. Track maintenance is possible because the predicted position of motile sperm #221 extends past the occlusion, enabling continuation of the track when the two sperm eventually separate.

Sequence C in Figure 4 shows a tracking failure involving two sperm over a period of 25 frames. In this sequence, sperm #40 is immotile and is struck directly by sperm #201 at t=360. Between $360 \leq t \leq 365$, pixels from both sperm are coalesced and incorrectly detected as a single sperm. During this time, both tracks compete for the single measurement and track #201 is not updated by any measurement and eventually terminated by t=365. At t=375, a new track is initiated on the motile sperm (#217, previously #201).

Sequence D in Figure 4 shows a single motile sperm (#34) exhibiting an abrupt acceleration over a period of 21 frames (approximately 1.4 sec). During this time, its validation gate expands and elongates in the direction of the jump, enabling the track to continue past its prediction, and demonstrating the advantage of using adaptive process noise.

D. Comparison of Data Association Algorithms

Figure 5 compares four algorithms NN, GNN, PDA, and JPDA, applied to four different simulated scenarios.[3] Each tracking algorithm was applied to four simulated multi-object tracking scenarios illustrated in Figure 5 (a-d). The performance of each algorithm over time was computed by averaging the OSPA distance between true and estimated tracks over 100 Monte Carlo replications Figure 5 (e-h). A higher mean OSPA distance means worse tracking performance.

Scenario A in Figure 5 (a) is a control scenario in which three well-separated objects move at constant velocity over a period of 9 sec (135 frames) and there are no collisions. The mean OSPA distance for each algorithm in this scenario is plotted in Figure 5 (e). As expected, all of the algorithms have a small and relatively constant mean OSPA vs. time. The GNN and JPDAF algorithms have the best performance, followed closely by NN and PDAF.

Scenario B in Figure 5 (b) is more challenging and involves three objects whose trajectories intersect. It is clear from Figure 5 (f) that until the targets cross the mean OSPA distance for all algorithms resembles that of the control scenario, after which it increases for all algorithms due to increased tracking errors. Of the four algorithms, JPDAF performs the best while PDA performs the worst, and NN is nearly three times worse than the JPDAF.

Scenario C in Figure 5 (c) involves three objects initially approaching each other, then moving in parallel with a minimum separation of 20 μm, and eventually separating. Figure 5 (g) shows significantly increased tracking error for PDA and NN due to persistent measurement-to-track misassociation, while GNN and JPDAF tracking exhibit the smallest error.

Scenario D in Figure 5 (d) resembles Scenario C, except now the minimum separation distance is cut in half to 10 μm. Figure 5 (h) shows, again, that PDA and NN perform the worst when the objects approach each other, and performance of the JPDAF and GNN have degraded compared to Scenario C. In fact, the reduced initial separation between targets resulted in early track swaps for JPDAF and GNN, signified by the higher mean OSPA distance values in the first 40 frames. In his scenario, GNN performed the best, followed by JPDAF.

The JPDAF and GNN algorithms consistently outperformed the NN and PDAF algorithms in all four simulated scenarios. For well-separated and parallel targets, JPDAF gave comparable performance to GNN. In the crossing target scenario, JDPAF outperformed all the other algorithms. Visual inspection of the videos collected for our sperm tracking experiments suggest that crossing sperm trajectories are far more common

[3]In our implementation of the NN and GNN algorithms, a constant velocity motion model was used for target state prediction.

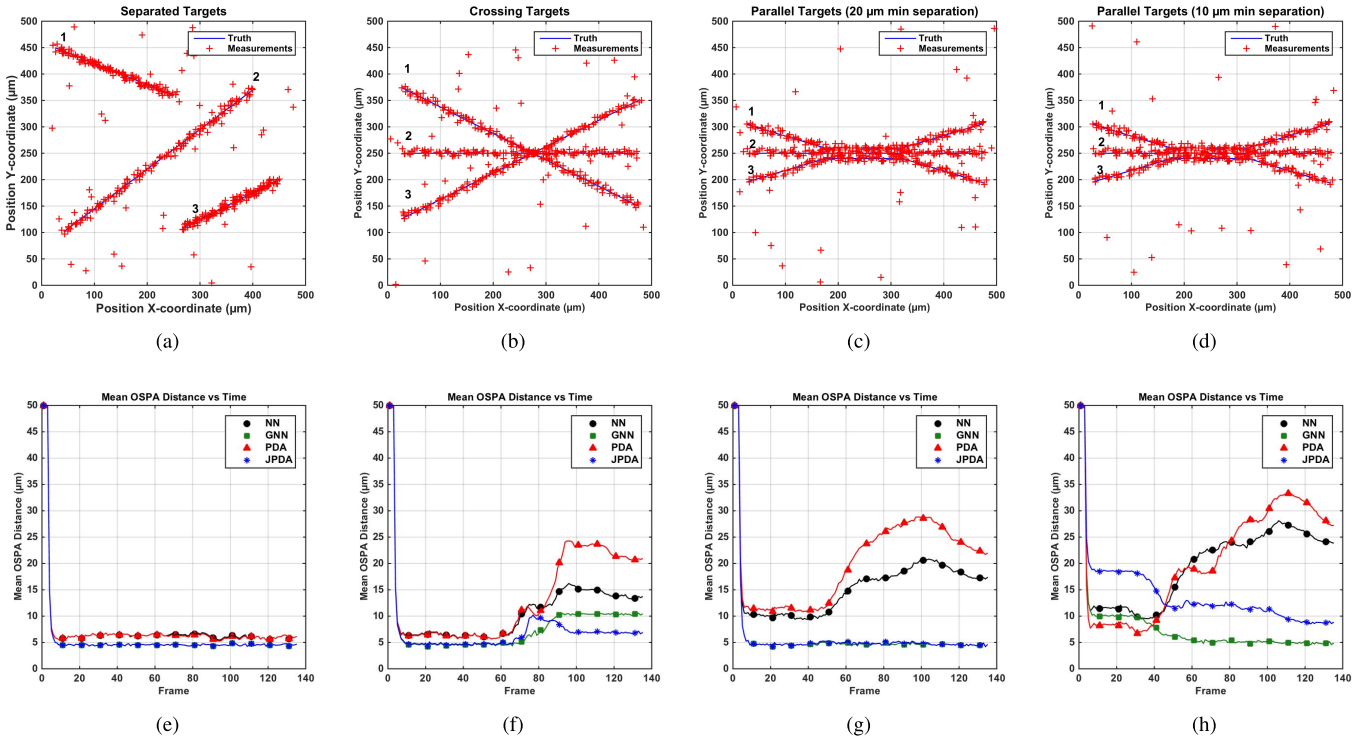


Fig. 5. Comparison of four data association algorithms using four simulated scenarios involving three targets in a 500 μm $\times$ 500 μm area at 15 frames per second, each 9 sec long (135 frames). The four scenarios are: (a) widely separated targets, (b) crossing targets, (c) parallel targets separated by 20 μm, and (d) parallel targets separated by 10 μm. True target trajectories are represented by blue lines, and measurements and random clutter by red crosses. For each algorithm, the OSPA distance was averaged over 100 Monte Carlo replications and plotted vs frame number in (e)–(h); larger OSPA distance means worse tracking accuracy. The four algorithms are NN = nearest-neighbor (standard CASA), GNN = global nearest neighbor, PDA = probabilistic data association, and JPDA = joint probabilistic data association.

than parallel sperm trajectories. Our conclusion is therefore that the JPDAF appears to be the preferred algorithm for sperm tracking in our architecture (Figure 1).

IV. Conclusion

We adapted the standard JPDAF to detect and track automatically the swimming paths of human sperm cells in time-lapse microscopy image sequences of the kind collected routinely by a fertility clinic. Using this algorithm, we tracked automatically hundreds of human sperm and measured their dynamic swimming parameters over time, with minimal user intervention, and without sample dilution. Unlike existing computer assisted semen analysis (CASA) instruments, our algorithm has the capability to track sperm swimming in close proximity and during apparent cell-to-cell collisions. The use of our algorithm frees the clinician from having to rely on elaborate motility measurements obtained manually by technicians, speeds up semen processing, and can potentially provide patients, medical practitioners, and researchers with more pertinent data and information than are currently available.

Appendix

A. Dynamic Model for Sperm Tracking

The state of each sperm is defined by its 2D position and velocity at time (frame) k and its time evolution is approximated by a continuous white noise acceleration (CWNA) target model [13], given by

$$x_t(k+1) = F x_t(k) + w_t(k), \tag{1}$$

where $x_t(k)$ is the 4×1 target t state vector at time k and $w_t(k)$ is the 4×1 zero mean random process noise vector. A measurement $z_t(k)$ (target position plus noise) is obtained every T seconds given by

$$z_t(k) = H x_t(k) + n_t(k), \tag{2}$$

where $n_t(k)$ is the 2×1 zero mean stationary white position noise vector with covariance matrix $N(k)$. The constant systems dynamics and measurement matrices F and H are given by

$$F = \begin{bmatrix} I_2 & I_2 T \\ 0_2 & I_2 \end{bmatrix}, \quad \text{and} \quad H = \begin{bmatrix} I_2 & 0_2 \end{bmatrix}, \tag{3}$$

where I_2 and 0_2 are the 2×2 identity and zero matrices, respectively. We assumed a stationary measurement noise covariance matrix $N(k) = \sigma_n^2 I_2$, with $\sigma_n = 2$ μm. The process noise covariance matrix is initialized according to

$$Q_t(0) = \begin{bmatrix} I_2 T^3/3 & I_2 T^2/2 \\ I_2 T^2/2 & I_2 T^2 \end{bmatrix} \tilde{q}_0, \tag{4}$$

where $\tilde{q}_0 = 20 \mu$m/sec.

B. JPDAF for Sperm Tracking

This section closely follows the JPDAF notation developed in [27]. We denote the set of m sperm position measurements at frame k by $Z(k) = \{z_j(k)\}_{j=1}^{m}$ and the set of all measurements up to and including frame k by $Z^k = \{Z(i)\}_{i=1}^{k}$. The predicted state, predicted covariance and predicted measurement of target t at time k are given respectively by

$$\hat{x}_{t_j}(k|k-1) = F\hat{x}_{t_j}(k-1|k-1), \tag{5}$$

$$P_{t_j}(k|k-1) = FP_{t_j}(k-1|k-1)F^T + Q_t(k), \quad \text{and} \tag{6}$$

$$\hat{z}_{t_j}(k|k-1) = H\hat{x}_{t_j}(k|k-1). \tag{7}$$

Here the index t_j denotes the target t to which measurement j is associated. To handle sperm maneuvers and adapt the measuremnt validation gate to sperm speed, we use a time-varying process noise covariance given by [30],

$$Q_t(k) = c_1 Q_t(k-1) + c_2 \nu_t \nu_t^T + c_3 Q_t(0), \tag{8}$$

where $\nu_t = \hat{x}_t(k|k-1) - \hat{x}_t(k-1|k-1)$ and the fading memory coefficients $c_1 + c_2 + c_3 = 1$ (we used $c_1 = 0.3$, $c_2 = 0.5$, and $c_3 = 0.2$). The residual $\nu_{jt}(k)$ and residual covariance matrix $S_{t_j}(k)$ are given by

$$\nu_{jt}(k) = z_j(k) - \hat{z}_t(k|k-1), \tag{9}$$

$$S_{t_j}(k) = HP_{t_j}(k|k-1)H^T + N(k). \tag{10}$$

In the JPDAF, the probability that measurement j is associated to track t is computed jointly across all targets. Calculating the association probabilities first requires finding the set of all feasible joint association events. A feasible joint event is an assignment of measurements to tracks or to clutter in which each track is assigned at most one or no measurements and no measurement is assigned to more than one track. The probability of the joint association event $\theta(k)$ given all measurements up to and including time k can be written as

$$P\{\theta(k)|Z^k\} = \frac{1}{c}\prod_{j=1}^{m}\{\lambda^{-1} f_{t_j}[z_j(k)]\}^{\tau_j(\theta)} \times \prod_{t=1}^{n}(P_D^t)^{\delta_t(\theta)}(1-P_D^t)^{1-\delta_t(\theta)}, \tag{11}$$

where $\tau_j(\theta) = 1$ if measurement j is associated with any target in the event θ, and $\delta_t(\theta) = 1$ if target t is associated with any measurement in the event θ. Here, λ is the known spatial density of the Poisson distributed clutter (an algorithm design parameter), c is a normalizing constant. In (11), $f_{t_j}[z_j(k)]$ is the Gaussian pdf with argument $z_j(k)$, mean $\hat{z}_{t_j}(k|k-1)$, and covariance $S_{t_j}(k)$, given by

$$f_{t_j}[z_j(k)] = |2\pi S_{t_j}(k)|^{-1/2} \exp\left(-\frac{1}{2}d_{jt}^2(k)\right), \tag{12}$$

where $d_{jt}^2(k)$ is the normalized statistical distance between measurement j and track t given by

$$d_{jt}^2(k) = \nu_{jt}(k)^T S_t(k)^{-1}\nu_{jt}(k). \tag{13}$$

Under the assumption that the states of the targets conditioned on past measurements are mutually independent, the marginal association probability $\beta_{jt}(k)$ that measurement j is associated to target t at time k is obtained by summing the probabilities of the joint events where this target-to-measurement association occurs. This can be written as

$$\beta_{jt}(k) = \sum_{\theta} P\{\theta(k)|Z^k\}\hat{\omega}_{jt}(\theta, k), \tag{14}$$

where $\hat{\omega}_{jt}(\theta, k) = 1$ if measurement j is associated with target t in event $\theta(k)$ and equal to zero otherwise. The filter gain $W_t(k)$, the probability-weighted combined residual $\nu_t(k)$, and the updated target state $\hat{x}_t(k|k)$ are given by

$$W_t(k) = P_t(k|k-1)H^T S_t(k)^{-1}, \tag{15}$$

$$\nu_t(k) = \sum_{j=1}^{m}\beta_{jt}(k)\nu_{jt}(k), \text{ and} \tag{16}$$

$$\hat{x}_t(k|k) = \hat{x}_t(k|k-1) + W_t(k)\nu_t(k). \tag{17}$$

The pseudo-measurement $\tilde{z}_t(k)$ is the probability-weighted sum of the measurements used to update track t, given by

$$\tilde{z}_t(k) = \nu_t(k) - \hat{z}_t(k|k-1). \tag{18}$$

The updated target covariance matrix $P_t(k|k)$ is then given by

$$P_t(k|k) = \beta_{0t}(k)P_t(k|k-1) + [1-\beta_{0t}(k)]P_t^c(k|k) + \tilde{P}_t(k), \tag{19}$$

where β_{0t} is probability that target t is not associated with any measurement at time k, $P_t^c(k|k)$ is the covariance that would be obtained with no measurement origin uncertainty given by

$$P_t^c(k|k) = P_t(k|k-1) - W_t(k)S_t(k)W_t(k)^T, \tag{20}$$

and $\tilde{P}_t(k)$ is the spread of the innovations given by

$$\tilde{P}_t(k) = W_t\left[\sum_{j=1}^{m}\beta_{jt}(k)\nu_{jt}(k)\nu_{jt}(k)^T - \nu_t(k)\nu_t(k)^T\right]W_t^T. \tag{21}$$

C. Track Clustering

The computational load of the JPDAF can be reduced significantly if the entire set of targets and measurements can be divided into independent clusters. A cluster is a subset of tracks that co-validate one or more measurements. In our implementation, measurement j is validated by track t if it simultaneously satisfies position and velocity gating criteria. The position gate is satisfied if measurement j falls within the two-dimensional predicted position error ellipsoid defined by the filter-calculated residual covariance matrix $S_{t_j}(k)$, i.e, if $d_{jt}^2(k) \leq \gamma_p^2$, and the velocity gate is satisfied if $\|\nu_{jt}(k)\|/T \leq \gamma_v$. We used $\gamma_p^2 = 11.6183$ (two degree of freedom Chi-square cdf evaluated at an assumed gate probability of $P_G = 0.997$) and $\gamma_v = 300$ μm/sec for the maximum sperm swimming speed. Applying these gates, the $m \times n$ binary validation matrix A at time k is given by

$$A(j,t,k) = \begin{cases} 1, & \text{if track } t \text{validates measurement } j \\ 0, & \text{otherwise,} \end{cases} \tag{22}$$

and by the corresponding matrix of pair-wise distances

$$D(j,t,k) = \begin{cases} d_{jt}^2(k), & \text{if track } t \text{validates measurement } j \\ \infty, & \text{otherwise.} \end{cases}$$

We used Kusiak's algorithm [28] on A to identify distinct clusters and applied the JPDAF to each cluster independently.

D. Track Initiation, Continuation and Termination

A tentative track is initiated on any measurement not belonging to a cluster, and a track is confirmed or deleted based on its track score. The track score $\ell_t(k)$ for track t at time k is the running sum (over the total life of the track) of the log-likelihood ratio that the pseudo-measurement $\tilde{z}_t(k)$ originated from track t rather than from clutter [13], given by:

$$\ell_t(k) = \sum_{l=k_0^t}^{k} \ln(\lambda^{-1} f_t[\tilde{z}_t(l)] P_D^t), \tag{23}$$

where k_0^t is the time when the track t was created. Equation (23) can be expressed recursively as

$$\ell_t(k) = \ell_t(k-1) + \Delta\ell_t(k), \tag{24}$$

where $\Delta\ell_t(k)$ depends on whether or not a track is updated by a pseudo-measurement according to

$$\Delta\ell_t(k) = \begin{cases} \ln(1 - P_D^t), & \text{if track } t \text{is not updated,} \\ \ln(\lambda^{-1} f_t[\tilde{z}_t(k)] P_D^t), & \text{if track } t \text{is updated.} \end{cases} \tag{25}$$

At the start of track life the score was initialized to $\ell_t(k_0^t) = \ln(\lambda^{-1}\lambda_n)$ where λ_n is the expected number of measurements from new targets per unit area per scan of data [31]. A tentative track was promoted to a confirmed track at time k if $\ell_t(k) > \eta_c$ and a confirmed track was deleted at time k if $\ell_t(k) - \ell_t(k_m) < \eta_d$ where $k_0^t \le k_m^t \le k$ is the index of the maximum score of track t. The confirmation and deletion thresholds η_c and η_d are given by $\eta_c = \ln(P_{DT}(1 - P_{CF})^{-1})$ and $\eta_d = \ln(P_{CF}^{-1}(1-P_{DT})) + \ln(\lambda^{-1}\lambda_n)$ where P_{DT} is the probability of deleting a true track and P_{CF} is the probability of confirming a false track, which are both algorithm design parameters [13]. If the pair-wise distance between the state estimates of any two tracks is less than 0.01 μm, then the track having the lower track score is deleted in order to eliminate redundant tracks.

Acknowledgment

The authors are grateful to Jeanne Ricci, Kristina Staab, and Kate Campbell of Penn Fertility Care at the Hospital of the University of Pennsylvania for sperm sample collection, preparation and imaging.

References

[1] D. Mortimer, *Practical Laboratory Andrology*. Oxford, U.K.: Oxford Univ. Press, 1994.
[2] E. A. Gaffney, H. Gadelha, D. J. Smith, J. R. Blake, and J. C. Kirkman-Brown, "Mammalian sperm motility: Observation and theory," *Annu. Rev. Fluid Mech.*, vol. 43, pp. 501–528, Jan. 2011.
[3] Z. Zhang *et al.*, "Human sperm rheotaxis: A passive physical process," *Sci. Rep.*, vol. 6, no. 23553, pp. 1–8, 2016.
[4] The Practice Committee of the American Society for Reproductive Medicine, "Diagnostic evaluation of the infertile male: A committee opinion," *Fertility Sterility*, vol. 98, no. 2, pp. 294–301, 2012.
[5] M. Tomlinson, S. Lewis, and D. Morroll, "Sperm quality and its relationship to natural and assisted conception: British fertility society guidelines for practice," *Human Fertility*, vol. 16, no. 3, pp. 175–193, 2013.
[6] *Laboratory Manual for the Examination and Processing of Human Semen*, 5th ed., World Health Org., Geneva, Switzerland, 2010.
[7] S. T. Mortimer, "CASA—Practical aspects," *J. Androl.*, vol. 21, no. 4, pp. 515–524, 2000.
[8] R. J. Sherins, "Clinical use and misuse of automated semen analysis," *Ann. NY Acad. Sci.*, vol. 637, no. 1, pp. 424–435, 1991.
[9] S. T. Mortimer, "A critical review of the physiological importance and analysis of sperm movement in mammals," *Human Reproduction Update*, vol. 3, no. 5, pp. 403–439, 1997.
[10] R. P. Amann and D. Waberski, "Computer-assisted sperm analysis (CASA): Capabilities and potential developments," *J. Theriogenol.*, vol. 81, no. 1, pp. 5–17, 2013.
[11] R. O. Davis and D. F. Katz, "Standardization and comparability of CASA instruments," *J. Androl.*, vol. 13, no. 1, pp. 81–86, 1992.
[12] H. Feitsma, M. Broekhujse, and B. Gadella, "Do CASA systems satisfy consumers demands? A critical analysis," *Reproduction Domestic Animals*, vol. 46, no. s2, pp. 49–51, 2011.
[13] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA, USA: Artech House, 1999.
[14] N. Chenouard *et. al.*, "Objective comparison of particle tracking methods," *Nature Methods*, vol. 11, no. 3, pp. 281–289, 2014.
[15] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Syst.*, vol. 29, no. 6, pp. 82–100, Dec. 2009.
[16] D. F. Katz, R. O. Davis, B. A. Delandmeter, and J. W. Overstreet, "Real-time analysis of sperm motion using automatic video image digitization," *Comput. Methods Programs Biomed.*, vol. 21, no. 3, pp. 173–182, 1985.
[17] B. Beresford-Smith and D. V. Helden, "Applications of radar tracking algorithms to motion analysis in biomedical images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Nov. 1994, pp. 411–415.
[18] L. Z. Shi, J. M. Nascimento, M. W. Berns, and E. L. Botvinick, "Computer-based tracking of single sperm," *J. Biomed. Opt.*, vol. 11, no. 5, p. 054009, Sep./Oct. 2006.
[19] J. Liu, C. Leung, Z. Lu, and Y. Sun, "Quantitative analysis of locomotive behavior of human sperm head and tail," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 390–396, Feb. 2013.
[20] M. J. Tomlinson *et al.*, "Validation of a novel computer-assisted sperm analysis (CASA) system using multitarget-tracking algorithms," *Fertility Sterility*, vol. 93, no. 6, pp. 1911–1920, 2010.
[21] T.-W. Su, L. Xue, and A. Ozcan, "High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 40, pp. 16018–16022, 2012.
[22] M. Berezansky, H. Greenspan, D. Cohen-Or, and O. Eitan, "Segmentation and tracking of human sperm cells using spatio-temporal representation and clustering," *Proc. SPIE*, vol. 6512, p. 65122M, Mar. 2007.
[23] L. Sorensen, J. Østergaard, P. Johansen, and M. De Bruijne, "Multi-object tracking of human spermatozoa," *Proc. SPIE*, vol. 6914, p. 69142C, Mar. 2008.
[24] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo, "A metric for performance evaluation of multi-target tracking algorithms," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3452–3457, Jul. 2011.
[25] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[26] D. F. Crouse, Y. Bar-Shalom, P. Willett, and L. Svensson, "The JPDAF in practical systems: Computation and snake oil," *Proc. SPIE*, vol. 7698, pp. 76981I-1–76981I-10, Apr. 2010.
[27] K. Romeo, D. F. Crouse, Y. Bar-Shalom, and P. Willett, "A fast coalescence-avoiding JPDAF," *Proc. SPIE*, vol. 8393, pp. 83930U-1–83930U-14, May 2012.
[28] A. Kusiak and W. S. Chow, "An efficient cluster identification algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 17, no. 4, pp. 696–699, Jul. 1987.
[29] I. J. Cox and M. L. Miller, "On finding ranked assignments with application to multitarget tracking and motion correspondence," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 486–489, Jan. 1995.
[30] A. Genovesio, Z. Belhassine, and J. C. Olivo-Marin, "Adaptive gating in Gaussian Bayesian multi-target tracking," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 147–150.
[31] Y. Bar-Shalom, S. S. Blackman, and R. J. Fitzgerald, "Dimensionless score function for multiple hypothesis tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 392–400, Jan. 2007.