

A Unified Graphical Models Framework for Automated Mitosis Detection in Human Embryos

Farshid Moussavi*, Yu Wang, Peter Lorenzen, Jonathan Oakley, Daniel Russakoff, and Stephen Gould

Abstract—Time lapse microscopy has emerged as an important modality for studying human embryo development, as mitosis events can provide insight into embryo health and fate. Mitosis detection can happen through tracking of embryonic cells (tracking based), or from low level image features and classifiers (tracking free). Tracking based approaches are challenged by high dimensional search space, weak features, outliers, missing data, multiple deformable targets, and weak motion model. Tracking free approaches are data driven and complement tracking based approaches. We pose mitosis detection as augmented simultaneous segmentation and classification in a conditional random field (CRF) framework that combines both approaches. It uses a rich set of discriminative features and their spatiotemporal context. It performs a dual pass approximate inference that addresses the high dimensionality of tracking and combines results from both components. For 312 clinical sequences we measured division events to within 30 min and observed an improvement of 25.6% and a 32.9% improvement over purely tracking based and tracking free approach respectively, and close to an order of magnitude over a traditional particle filter. While our work was motivated by human embryo development, it can be extended to other detection problems in image sequences of evolving cell populations.

Index Terms—Data driven Monte Carlo, embryo tracking, graphical models, mitosis detection.

I. INTRODUCTION

HUMAN embryo development remains poorly understood due to limited availability of human embryos, as well as the ethical and logistical challenges of studying them by conventional, invasive methods. Time-lapse microscopy (TLM) has recently emerged as a powerful noninvasive method to assess human embryo development. Quantitative image analysis on such images has also enhanced this modality's capabilities [2], [12], [16]. TLM can provide insights into how human embryos reach critical developmental milestones [38] and respond to chromosomal abnormalities [6]. Many recent studies have shown that cell division timings during early development (up

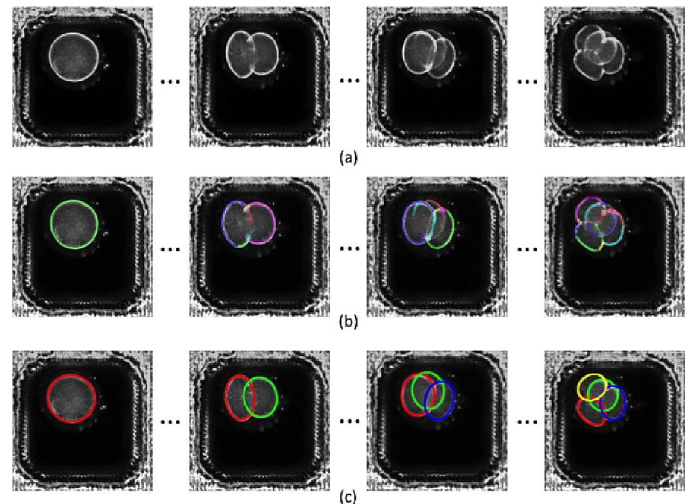


Fig. 1. Time lapse sequence of human embryo in 1-cell, 2-cell, 3-cell, and 4-cell stages: (a) raw images; (b) overlaid with boundary segment features; and (c) overlaid with inferred cell boundaries.

to four or five cells) can help predict downstream embryo fate [7], [9], [17], [27]. Early development—from fertilization to the four cell stage—typically lasts up to 48 h. Some examples of TLM images from a dark field microscope are shown in Fig. 1.

In recent years, a number of commercial TLM systems have been deployed and used to observe and manually measure division timing events. The Eeva Test is the only commercially available platform that is capable of performing embryo tracking with automated software. The first generation of Eeva has been clinically validated for outcome prediction in a prospective, multi-center clinical *in vitro* fertilization (IVF) study [8]. The framework used to provide the timing measurements that led to these results was first reported in [29], and is expanded upon in this paper.

The measurement of mitosis events is challenged by the manual localization of cells and detection of their divisions, which requires expert time and is prone to error. Embryologists in IVF clinics may be required to examine sequences of hundreds of embryos in a period of several days, from which they would select a potentially small number of cases that have certain desirable division timing parameters. This task significantly increases their already heavy workload. Expert measurement can also be prone to high inter-observer and intra-observer variability [31]. Automated mitosis detection algorithms have the potential to overcome this bottleneck and introduce a higher level of objectivity. In order to be

Manuscript received January 23, 2014; revised April 07, 2014; accepted April 11, 2014. Date of publication April 23, 2014; date of current version June 27, 2014. Asterisk indicates corresponding author.

*F. Moussavi is with Auxogyn, Inc., Menlo Park, CA 94025 USA (e-mail: fmoussavi@auxogyn.com).

Y. Wang and P. Lorenzen are with Auxogyn, Inc., Menlo Park, CA 94025 USA.

J. Oakley and D. Russakoff are with the Voxeleron, LLC, San Francisco, CA 94120 USA.

S. Gould is with the College of Engineering and Computer Science, Australian National University, Canberra 0200, Australia.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2014.2317836

successful, they would need to identify at least several embryos with the desired timing parameters. If they track the cell boundaries, they also have the benefit of providing potentially useful geometric and morphological information. We refer to methods that try to detect mitosis through tracking of cell boundaries as *tracking based*, and those that try the detection directly from the image pixels and low level features as *tracking free*.

Visual tracking has been studied extensively in computer vision [34]. The difficulty of any visual tracking task is determined by a number of key characteristics of the data, including the number and density of target objects and their mutual interaction, the strength of visual features, the clarity and distinctness of object boundaries, object motion, object shape and rigidity, object topology, and image quality. Many contemporary tracking methods used in microscopy address problems with a subset of these challenges. For example, active contours [4], level sets [5], [21], [30], [33], and registration based [32] methods address the tracking of a single object with a complex deforming shape but with relatively clear and continuous boundaries. However, they do not address the data association problem inherent with changing topologies or multiple interacting targets. Meanwhile, many point tracking methods exist that address data association for multiple interacting point targets with occlusions and false positives with the help of spatial and temporal context [19], [22], [28], but they do not deal with deforming shapes or changing topology. Human embryo tracking faces a diverse set of challenges, and to the best of our knowledge, has not yet been solved in a fully automated fashion. The targets are multiple interacting cells with changing shape and topology and stochastic motion. They divide by cleaving in a relatively confined volume. Boundary features have no color or texture, can be weak, incomplete, occluded, or be outliers. We can address these challenges by considering appearance features with three types of context: 1) geometric coherence between boundary pixels within a cell, 2) context of the shapes of different cells within an embryo, and 3) temporal context between image frames. Common approaches to such temporal tracking are rooted in probabilistic dynamic state space methods such as the Kalman filter [20], [37], hidden Markov models, or more generally dynamic Bayesian networks [3], [23]. These methods model the joint distribution over a target's trajectory as the product of a set of conditional probability distributions, known as motion models and observation models. In the case of human embryo tracking, the solution space is high dimensional and an exact solution is intractable. Thus, approximate inference is needed, which is often performed by sequential Monte Carlo methods [11]. Even then, the tracking results may not be conclusive in a significant portion of cases when using 2-D imaging. Recent work proposed and demonstrated a traditional particle filter for tracking human embryos on a select set of 14 examples [38]. However, the traditional particle filter is challenged by the high dimensional search space of tracking multiple deforming shapes. This and other significant challenges to the human embryo tracking problem served as the motivation for the work in this paper.

Prior efforts in embryonic and stem cell tracking have often focused on estimating trajectories and lineage, where large

numbers of cells are localized and tracked over time [25]. Many have successfully integrated segmentation with classification and target association to accomplish tracking and lineage reconstructions [1], [35]. Often fluorescent markers are deployed to help localize cells [39], but this is not an option for human embryos, which must not be tagged with contrast agents. Spatial temporal segmentation based tracking approaches with level sets have also been proposed [40]. Recently, probabilistic graphical models have been used in a tracking free approach for mitosis phase labeling in human adenocarcinoma cells [13] as well as automatic mitosis detection in mesenchymal stem cells, which are relatively 2-D in culture [26] and [18]. These studies suggest that tracking free approaches may be preferable to potentially intractable tracking based approaches if our primary interest is to detect mitosis events and not cell morphology.

While these two approaches are inherently different, they need not be mutually exclusive as they may provide complementary information. Here we propose a unified framework called augmented simultaneous segmentation and classification (ASSC) whose contributions to the human embryo mitosis detection are twofold. First, the framework combines tracking based and tracking free approaches through a conditional random field (CRF) [24] model. Second, the framework performs data driven approximate inference over this model to efficiently deal with the problem's high dimensionality and make the solution tractable. The CRF integrates relevant unknowns (such as shapes and number of cells) and a rich set of image sequence features, and encodes their mutual interactions across time and space via a set of compatibility functions, which are not restricted to be conditional probabilities. The observations (image features) are treated as fixed quantities instead of variables, and need not be modeled by a generative joint probability distribution. The data driven approximate inference is performed in two passes. The first pass is a tracking based simultaneous segmentation and classification (SSC), which aims to identify and track individual cells. Here, multiple hypotheses guided by the image data are maintained and represent high probability samples from the larger solution space. Separately, in the first pass, a tracking free portion classifies the number of cells and division events from a rich set of hand crafted and learned features. The second and final pass of inference augments the tracking based SSC inference with tracking free results to obtain a refined distribution on cell population size and ultimately the division events. While this framework was motivated by mitosis detection for human embryos under light microscopy, it can be applied to other problems which require early stage (4-8 cells) event detection of time lapse images of dividing cells with complex scenes, such as nonhuman embryos, stem cell populations, or cancer cell populations.

Our algorithm was trained on a development dataset containing 348 embryos and tested on a separate test dataset containing 417 embryos from multiple IVF clinics. Division events were labeled by an expert panel and we evaluated performance on sequences with low inter panelist variation. We observed that the number of sequences successfully measured to within 30 min of the expert panelists showed

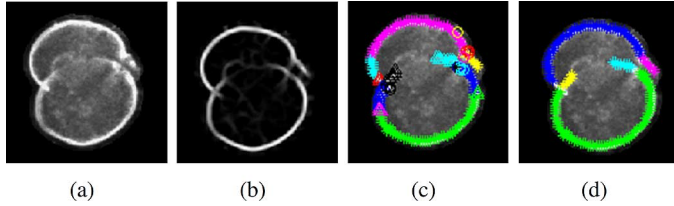


Fig. 2. Segment extraction: (a) raw image, (b) boundary feature image obtained with the Hessian operator, (c) standard contour feature detection step yields initial set of segments, and (d) segment merging step merges into fewer segments.

an improvement of 25.6% over a purely tracking based approach and a 32.9% improvement over a purely tracking free approach. Moreover, our results show close to an order of magnitude improvement over a traditional particle filter approach.

II. AUGMENTED SIMULTANEOUS SEGMENTATION AND CLASSIFICATION

Augmented simultaneous segmentation and classification (ASSC) seeks to determine the times of cell division events. It starts by extracting a rich set of features for use by the tracking based and tracking free components. It then performs approximate inference on a CRF model which incorporates these features, along with relevant latent variables.

A. Feature Extraction

The first step in the image sequence analysis is to extract image features. The tracking based portion uses contour features called boundary segments, which are designed for inferring cell boundaries and embryo geometry. The tracking free portion uses a rich set of 62 hand crafted lower level discriminative features which are designed for learning classifiers that detect the number of cells directly. We next briefly describe the features used for each portion.

1) *Tracking Based Features*: Tracking based features are specifically intended for determining cell boundaries and embryo geometry. The basic image features used by the tracking based portion are contour features called boundary segments, depicted in Fig. 1(b). Ultimately each segment will be assigned a label indicating which cell it belongs to, where multiple segments can belong to the same cell. Therefore it is desirable to have as few segments as possible to limit the label assignment space. For example, in a frame containing two cells and n segments, the segments can be labelled in 2^n ways. Ideally, a segment feature extraction would yield one segment per cell. In practice, however, a standard contour feature extraction would yield a larger number of smaller segments. We designed the segment extraction to be a standard contour feature detection followed by a merging inference step to reduce the number of segments, as depicted in Fig. 2.

We first create a boundary feature image using a Hessian operator [15]. This operator provides both magnitude and principal orientation for each pixel. We then extract contour features in a standard fashion by conducting a directed local search for coherent boundary pixels with hysteresis thresholding. A subse-

quent merging step combines the segments into a smaller set of larger segments, described next.

Consider at each frame t a set of K_t initial segments $s^{(t)} = \{s_k^{(t)}\}_{k=1 \dots K_t}$ where each segment $s_k^{(t)}$ is a collection of $m_k^{(t)}$ 2-D points $s_{k,i}^{(t)} \in \mathbb{R}^2$ with $i = 1 \dots m_k^{(t)}$. In words, $s^{(t)}$ is a set of initial segments, each segment composed of a set of points.

The segment merging step is posed as a clustering problem, where we would like to group together segments belonging to the same cell boundary. We define a pairwise compatibility function over any pair of segments $\nu(s_j, s_k)$ based on the following geometric properties: continuity, distance, adherence to a coherent underlying smooth curve, and mean curvature. This function has the form of a product of four smaller functions, each corresponding to one of the four geometric properties. That is, we have $\nu(s_j, s_k) = \prod_{i=1}^4 \nu_i(s_j, s_k)$ and where we normalize the output of each ν_i to be between 0 and 1. The higher the value of $\nu(s_j, s_k)$ the more likely segments s_j and s_k belong to the same cell boundary. Roughly speaking, our aim is to group segments so that all pairs of segments within a cluster have high compatibility, and all pairs of segments between clusters have low compatibility.

To achieve our aim we perform agglomerative clustering as follows. We start with each segment belonging to its own cluster. We then repeatedly merge any two clusters A and B where the minimum pairwise compatibility between any segment $s_a \in A$ and $s_b \in B$ exceeds some threshold. We stop when no such merger can be found. Note that a number of alternative merge criteria are possible, e.g., only merging clusters if the mean pairwise segment compatibility is above some threshold. However, in our experiments we found the simple maximum criterion to work well.

2) *Tracking Free Features*: The tracking free portion of our framework uses a two-level embryo stage classification model [36] that classifies each frame of an embryo video to one of four stages (i.e., 1-cell, 2-cell, 3-cell, and 4-or-more-cell). The features for this portion are intended for training a classifier that directly infers number of cells from groups of images. 62 standard hand-crafted embryo features are extracted from each frame of a human embryo video. The first level Adaboost classification model consists of four binary Adaboost classifiers, each trained for classifying one class from the rest of classes using the 62 features. The second level Adaboost classification model consists of another four Adaboost classifiers trained with an augmented feature set that includes additional features computed from the first level classification probabilities in a local temporal window. The first level Adaboost classification model is designed to perform classification on each frame independently, while the second level Adaboost classification model is designed to exploit local temporal context and refine the results from the first level.

The 62 embryo features include 22 gray-level co-occurrence matrices (GLCM) features, 10 local binary patterns (LBP) features, 15 Hessian features, 10 Gabor features, and five region properties. The GLCM, LBP, and Gabor features are well-known texture features for classification problems. The Hessian features are summary statistics (e.g., mean inten-

sity value) computed from Hessian filtered embryo images [Fig. 2(b)] in which the cell edges are enhanced. The region properties consist of area, number or convex hull points, solidity, eccentricity, and perimeter computed from detected regions of interest spanned by the segments obtained in Section II-A1.

The additional features for the second level Adaboost classifiers are temporal features computed from the probabilities output by the first level Adaboost classification model. At each frame, statistics (i.e., mean, median, max, min, and standard deviation) of the classification probabilities for each class are computed from within the local temporal window. This produces a total of 20 additional second level features that are added to the original feature set.

The one-vs-all scheme described above handles the multi-class classification of cell stage. Each binary Adaboost classifier (in each level of the two-level model) consists of N decision stumps (decision tree of depth one) and is trained to separate one class from all other classes. For an Adaboost classifier trained for class $i \in \{1, 2, 3, 4\}$, its output for a embryo image at frame t is

$$P(y_t = i | x_t) = \frac{\sum_{k=1}^N a_{ik} h_{ik}(x_t)}{\sum_{k=1}^N a_{ik}} \quad (1)$$

where x_t is the extracted feature vector for the image, a_{ik} is the weight of the base classifiers, $h_{ik} \in \{-1, 1\}$ is the output of the base classifiers, and $P(y_t = i | x_t)$ is the posterior classification probability normalized to $[0, 1]$.

B. CRF Model

The extracted features serve as a diverse set of observations from which we can estimate the numbers and shapes of cells in the embryo over time, as depicted in Fig. 1(c). We model a stochastic evolution of elliptical cells with the CRF in Fig. 3. At each frame t there are K_t segments, each with $m_k^{(t)}$ points, $k \in \{1, \dots, K_t\}$, and up to $N_{\max} = 4$ cells. The variables to be inferred are labels assigning segments to cells $l_k^{(t)} \in \{0, 1, \dots, N_{\max}\}$; ellipses $e_n^{(t)} \in \mathbb{R}^5$, $n \in \{1, \dots, N_{\max}\}$; and number of cells $N^{(t)} \in \{1, \dots, N_{\max}\}$. Each ellipse $e_n^{(t)}$ is associated with its parent from the previous time slice, $e_{Pa(n)}^{(t-1)}$. If $Pa(n)$ is dividing, it is called a mother cell, and its two daughter cells share the same parent. The observations are the set of segments $s^{(t)} = \{s_k^{(t)}\}_{k=1, \dots, K_t}$ where $s_k^{(t)}$ is a collection of points $s_{k,i}^{(t)} \in \mathbb{R}^2$ with $i = 1, \dots, m_k^{(t)}$; a classifier probability on the number of cells $c_N^{(t)} \in \mathbb{R}^{N_{\max}}$; and image similarity measure $\delta^{(t)} \in [0, 1]$. The image similarity measure serves as a noisy estimate of the transition probabilities for $N^{(t)}$, such that smaller values favor a division event for that frame. Many design choices can be considered. In this paper, we use normalized cross correlation (NCC) between a frame and a registered average of five previous frames, which provides good performance at a reasonable computational cost.

Compatibility functions are either over variables that: 1) are within one time slice (observation model Φ); or 2) span neigh-

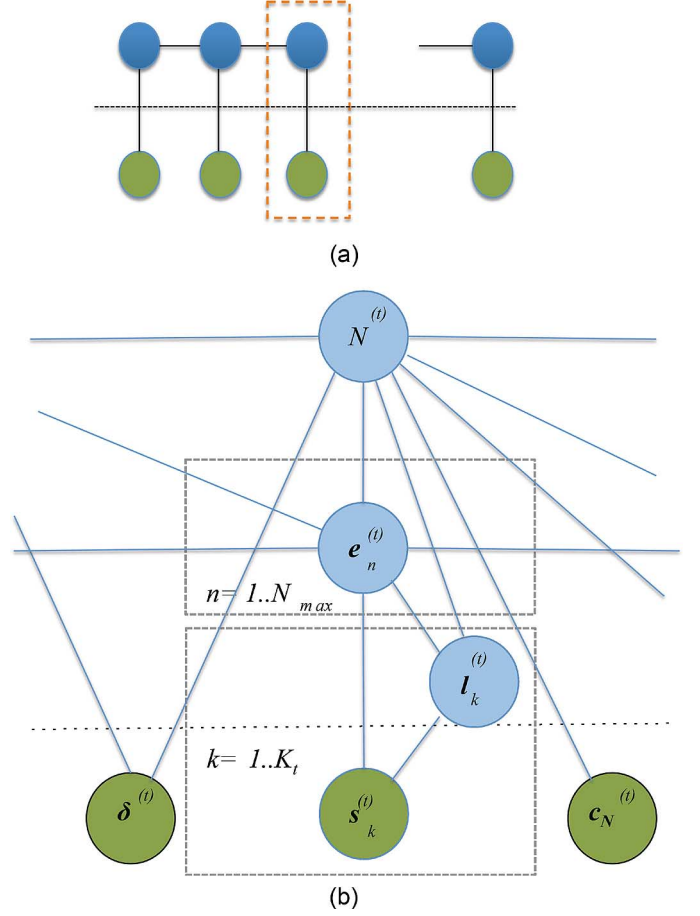


Fig. 3. CRF models: (a) an example CRF chain graph with known evidence nodes (shown in green), and unknown latent variables to be inferred (shown in blue), (b) our CRF model for one time slice— $N^{(t)}$ is number of cells at time t , $\delta^{(t)}$ is an image similarity measure, $c_N^{(t)}$ is an image based classifier, $e_n^{(t)}$ is the n th ellipse descriptor for time t , $s_{t,k}$ is the k th segment descriptor for time t , $l_{t,k}$ is the ellipse assignment for the k th segment at time t .

boring time slices (motion model Ψ). The CRF encodes the conditional probability distribution over all variables in a sequence of T frames as a scalar value proportional to the product of all compatibility functions

$$P \left(e^{(1:T)}, l^{(1:T)}, N^{(1:T)} \mid s^{(1:T)}, \delta^{(1:T)}, c_N^{(1:T)} \right) \propto \prod_{t=1}^T \underbrace{\Phi \left(e^{(t)}, l^{(t)}, N^{(t)}, s^{(t)}, c_N^{(t)} \right)}_{\text{observation model}} \cdot \prod_{t=2}^T \underbrace{\Psi \left(e^{(t-1:t)}, N^{(t-1:t)}, \delta^{(t)} \right)}_{\text{motion model}}. \quad (2)$$

We are interested at a minimum in the sequence $N^{(t)}$ which maximizes the marginal distribution $P(N^{(t)})$. When possible, we are also interested in the cell shapes (ellipses).

1) *Observation Model*: The observation model Φ is the product of three compatibility functions

$$\Phi \left(e^{(t)}, l^{(t)}, N^{(t)}, s^{(t)}, c_N^{(t)} \right) = \phi_0 \left(e^{(t)} \right) \phi_1 \left(e^{(t)}, l^{(t)}, N^{(t)}, s^{(t)} \right) \phi_2 \left(c_N^{(t)}(N^{(t)}) \right). \quad (3)$$

The function ϕ_0 is a binary indicator function which encodes limits on ellipse size and aspect ratio. These were learned cumulatively through a combination of experience by experts and observing many correctly identified and measured embryos and their associated statistics. It returns a zero if any of these values are out of an acceptable range and one otherwise. The function ϕ_1 encodes compatibility of ellipses, segments, and labels

$$\phi_1 \left(e^{(t)}, l^{(t)}, N^{(t)}, s^{(t)} \right) = \left(\prod_{i=1}^{N^{(t)}} f(\cdot, \cdot, \cdot)^{c_f} e^{-c_r r(\cdot, \cdot, \cdot)^2} \right)^{(1/N^{(t)})} + p_1 \quad (4)$$

where $f \left(e_i^{(t)}, l^{(t)}, s^{(t)} \right) \in [0, 1]$ is an ellipse coverage term, $r \left(e_i^{(t)}, l^{(t)}, s^{(t)} \right) \in \mathbb{R}_+$ is segment fitting error, and c_f and c_r are approximately learned using coordinate ascent starting from an empirically chosen starting point. This was necessary due to the long inference time for each iteration in any optimization used for learning, and the inherent resulting bottleneck. Here, p_1 is a constant prior term.

The function ϕ_2 encodes the tracking free portion in the form of the classifier $c_N^{(t)}$, and adds the same prior term p_1 for robustness

$$\phi_2 \left(c_N^{(t)} \right) = c_N^{(t)} + p_1. \quad (5)$$

For both ϕ_1 and ϕ_2 the constant prior term p_1 is added for robustness against noisy features. It is learned from training data by minimizing an objective function that estimates mean square deviation between our model prediction and expert measurements. This deviation is defined in Section III-C. The minimization is performed by a local search using coordinate ascent.

2) *Transition Model*: The transition model Ψ governs cell shape deformations and division

$$\Psi \left(e^{(t-1:t)}, N^{(t-1:t)}, \delta^{(t)} \right)_{t=2..T} = \prod_{n=1}^{N^{(t)}} \psi_1 \left(e_{Pa(n)}^{(t-1)}, e_n^{(t)}, N^{(t-1:t)} \right) \psi_2 \left(N^{(t-1:t)}, \delta^{(t)} \right). \quad (6)$$

The function ψ_1 encodes the underlying cell deformation and division process

$$\psi_1 \left(e_{i_1}^{(t-1)}, e_{i_2}^{(t)}, N^{(t-1:t)} \right) = \begin{cases} e^{-\rho(e_{i_1}^{(t-1)}, e_{i_2}^{(t)})}, & \text{if } i_1 \text{ is not the mother of } i_2 \\ e^{-\rho(h(e_{i_1}^{(t-1)}), e_{i_2}^{(t)})}, & \text{if } i_1 \text{ is the mother of } i_2 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where $\rho(e_{i_1}, e_{i_2}) = (e_{i_1} - e_{i_2})^T \Lambda (e_{i_1} - e_{i_2})$ with Λ a diagonal matrix of deformation costs, and h a nonaffine transform from a mother to daughter cell shape.

The function ψ_2 encodes influence of division on the image similarity measure, and constrains the number of cells $N^{(t)}$ to be nondecreasing

$$\psi_2 \left(N^{(t-1:t)}, \delta^{(t)} \right) = \begin{cases} \delta^{(t)}, & \text{if } N^{(t)} = N^{(t-1)} \\ 1 - \delta^{(t)}, & \text{if } N^{(t)} > N^{(t-1)} \\ 0, & \text{otherwise} \end{cases}. \quad (8)$$

C. Approximate Inference

We seek the most likely number of cells for each frame, $N^{(t)}$ from the CRF. As exact inference is intractable, we perform approximate inference in two passes—a tracking based inference, and an event inference that combines the tracking based and tracking free results.

1) *Tracking Based Inference—Simultaneous Segmentation and Classification*: The tracking based inference performs simultaneous segmentation and classification, and infers geometry from segments. It estimates $\hat{\Phi}_M(N^{(t)})$, the unnormalized max product marginal measure of $N^{(t)}$ by optimizing from time 0 to time t on a mutilated subgraph that excludes $c_N^{(t)}$ and $\delta^{(t)}$

$$\hat{\Phi}_M \left(N^{(t)} \right) = \max_{e, l} E(t) \quad (9)$$

where

$$E(t) = \prod_{\tau=1}^t \phi_0 \left(e^{(\tau)} \right) \phi_1 \left(e^{(\tau)}, s^{(\tau)}, l^{(\tau)}, N^{(\tau)} \right) \psi_1 \left(e^{(\tau-1:\tau)}, N^{(\tau-1:\tau)} \right) \quad (10)$$

$$= E(t-1) \phi_0 \left(e^{(t)} \right) \phi_1 \left(e^{(t)}, s^{(t)}, l^{(t)}, N^{(t)} \right) \psi_1 \left(e^{(t-1:t)}, N^{(t-1:t)} \right). \quad (11)$$

We maximize this recursion with data driven sequential Monte Carlo (DD-SMC). SMC is a sequential importance resampling algorithm in which particles are maintained as an empirical approximation of $E(t)$ for each time step in a sequence of time and measurement updates. A data driven particle refinement stage occurs between the time and measurement updates. Each particle is comprised of a set of cell ellipse descriptors and segment cell labels. Ultimately, after all particles are refined they are then scored estimated from the distribution represented by the particles. An example plot is shown in Fig. 5(a). We next describe the particle refinement stage.

2) *Particle Refinement Stage*: The main objective of the data driven refinement stage is to improve the likelihood of particles sampled from the proposal distribution. This is done right after the SMC time update, using boundary segment features, and results in an empirical distribution that better represents the observed data. In doing so, it additionally address two other core problems in the tracking: the missing boundary data, and the data association. The segments do not necessarily cover the entire cell perimeter, and may not provide sufficient statistics for shape fitting. Their associations to cells (labels) are also unknown in advance (Fig. 2). To address this issue, we use hard assignment expectation maximization (EM) [10] which iterates between solving two problems—estimating the missing boundary points and segment labels given a set of ellipses (E-step), and estimating an optimum set of ellipses given a completed set of data and labels (M-step). In this formulation, the observations are $s^{(t)}$, and the unobserved variables are $l_k^{(t)}$ and a set of missing points $z_{n,i}^{(t)} \in \mathbb{R}^2$. The points can be thought of as the outputs of the generative process for ellipse points in Fig. 4. The absence or presence of a point is determined by

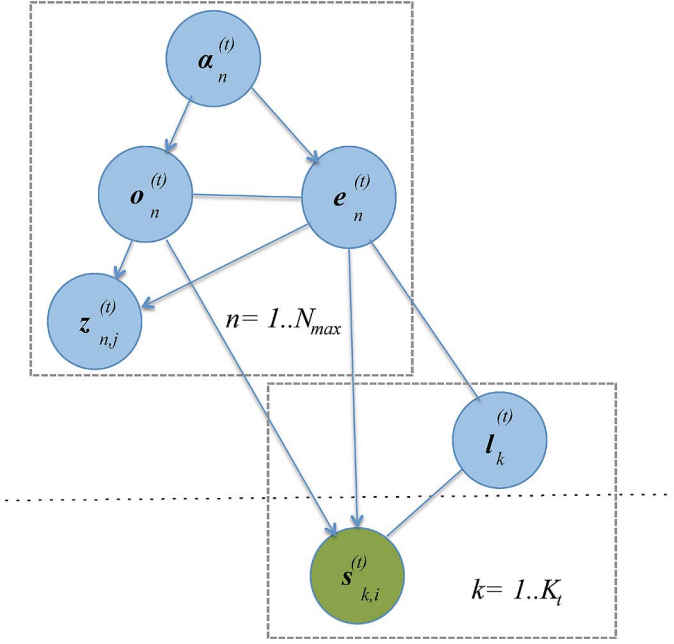


Fig. 4. Generative model for ellipse points. Observation variable $o_n^{(t)}$ controls whether a point appears or is missing, and is determined by angle range.

the boolean observation variable $o_n^{(t)}$ which is true for certain angle values $\alpha_N^{(t)} \in [0, 2\pi]$.

Let $\mathcal{S}_n^{(t)} = \cup\{s_k^{(t)}, l_k^{(t)} = n\}$ be the collection of all segments that are associated with cell n . Let $\mathcal{Z}_n^{(t)} = \{z_{n,i}^{(t)}, z_{n,i}^{(t)} \in \mathbb{R}^2\}$ be a complementary collection of points for cell n that lie in an angular range not covered by $\mathcal{S}_n^{(t)}$. Then $\mathcal{X}_n^{(t)} = \cup\{\mathcal{S}_n^{(t)}, \mathcal{Z}_n^{(t)}\}$ form the sufficient statistics for estimating $e_n^{(t)}$, the ellipse parameters for cell n . Also, let $g_n(\alpha) \in \mathbb{R}^2$ be the parametric function form of ellipse e_n . We now define the steps of our hard assignment EM refinement in standard fashion:

E-step: Estimate expected values of the unobserved variables

$$\begin{aligned} E[z_{n,\alpha}^{(t)} | e_n^{(t)}] &= g_n^{(t)}(\alpha), \forall \alpha, \text{ not covered by } s_n^{(t)} \\ \hat{l}_k^{(t)} | e_{1:N}^{(t)} &= \arg \min_n \bar{d}(s_k^{(t)}, e_n^{(t)}) \end{aligned} \quad (12)$$

where $\bar{d}(\cdot)$ is the average distance of the points in $s_k^{(t)}$ to ellipse $e_n^{(t)}$.

M-step: Combine with $s^{(t)}$ to get sufficient statistics and maximize the likelihood function

$$\begin{aligned} L(e_{1:N}^{(t)}; s_{1:K_t}^{(t)}, z_{1:N}^{(t)}, l_{1:K_t}^{(t)}) &= P(s_{1:K_t}^{(t)}, z_{1:N}^{(t)}, l_{1:K_t}^{(t)} | e_{1:N}^{(t)}) \\ &= \prod_{n=1:N} P(s_{1:K_t}^{(t)}, z_n^{(t)}, l_{1:K_t}^{(t)} | e_n^{(t)}) \end{aligned} \quad (13)$$

where the likelihood function decomposes into multiple smaller functions, one for each ellipse. This maximization therefore consists of fitting an ellipse to the points $\mathcal{X}_n^{(t)} = \cup\{\mathcal{S}_n^{(t)}, E[\mathcal{Z}_{n,\alpha}^{(t)} | e_n^{(t)}]\}$, where $\mathcal{S}_n^{(t)} = \cup\{s_k^{(t)}, \hat{l}_k^{(t)} | e_{1:N}^{(t)} = n\}$. We use the method in [14], which is a direct least squares formulation, and is efficient and robust.

3) *Event Inference:* The second and final pass of the approximate inference is the event inference. It combines the

tracking based and tracking free results to obtain the approximate marginal distribution on number of cells $\hat{P}(N^{(t)})$, and ultimately the most likely sequence $\hat{N}^{(1:t)}$. It is performed on a factor graph defined over the mutilated subgraph containing only $N^{(1:t)}$. The joint distribution over this mutilated subgraph factorizes over unary and pairwise factors defined in the factor graph

$$\begin{aligned} &\hat{P}(N^{(1:t)}) \\ &\propto \prod_{\tau=2}^t \hat{\Phi}_M(N^{(\tau)}) \phi_2(c_N^{(\tau)}(N^{(\tau)})) \psi_2(N^{(\tau-1:\tau)}, \delta^{(\tau)}). \end{aligned} \quad (14)$$

The unary factors are the products of $\hat{\Phi}_M(N^{(t)})$ in (11) and $\phi_2(c_N^{(t)})$ in (5). The pairwise factors are $\psi_2(N^{(t-1:t)}, \delta^{(t)})$ in (8).

We perform sum product belief propagation on this subgraph. The resulting final beliefs are the marginal distributions $\hat{P}(N^{(t)})$ [Fig. 5(b)]. The most likely transition times are taken to be the crossover points [Fig. 5(c)].

III. EXPERIMENTAL RESULTS

A. Imaging System and Acquisition

The images used for our development were acquired using the first generation of Eeva, which was enabled by digitally controlled time lapse microscopes with dark field illumination shown in Fig. 6. Eeva microscopes were designed to be used in incubators common to IVF clinics, and acquired images every 5 min. The microscope used a scientific optical objective with 4x magnification and a 5 Mpixel (2560 × 1920 pixels) CMOS image sensor array. The resulting resolution was 1 μm per pixel. A multi-well petri dish was also designed to hold multiple human embryos and their required media in group culture. This dish is placed into a platform directly on the microscope by an embryologist at the beginning of an imaging session that can last from three to five days. Each dish image contained 20 wells, not all of which were occupied. An initial preprocessing phase performed autofocus and autoexposure, and found the dish pose (rotation and translation relative to optical center). A rough region of interest for each well was then determined using the dish pose and the known dish and well geometries. This was then refined through a well segmentation step that found the boundary of each well using iterative closest points (ICP) on a predetermined template of an empty well. This refined region of interest allowed each embryo image to be cropped to 151 × 151 pixels. We applied our algorithm on sequences from such sessions taken from multiple IVF clinics. Our algorithm performed tracking of an embryo, across 500 image frames, on average in 1.5 h on one core of a standard Intel Xeon server using less than 4 GB.

B. Image Dataset

We had access to 765 labeled time lapse image sequences of human embryos obtained on the Eeva System (Auxogyn, Menlo Park, CA, USA) in seven IVF clinics over a period of one year. These were a superset of the data in the clinical study of [8]. Two expert embryologist panelists measured ground truth in the

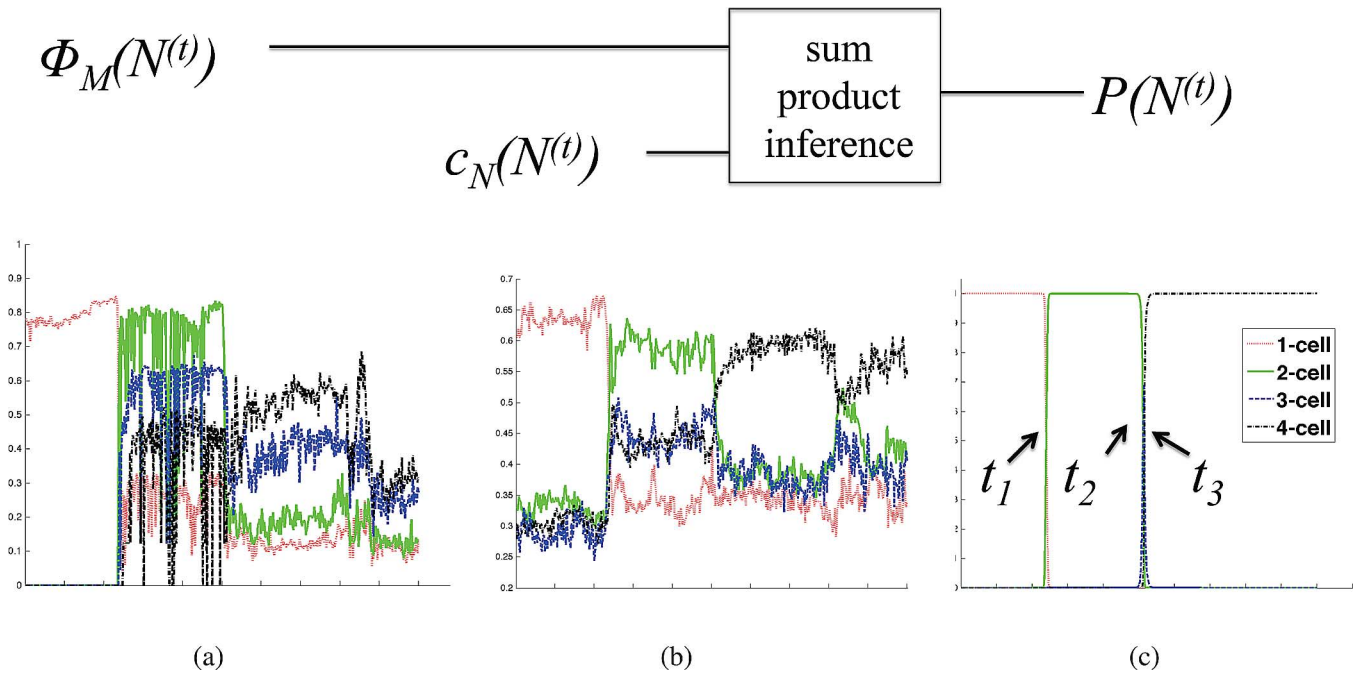
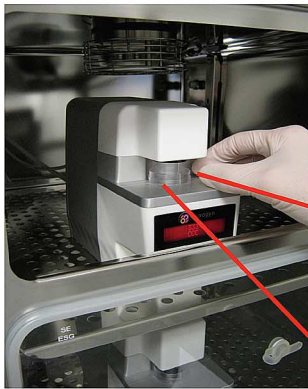


Fig. 5. Approximate inference results over CRF: (a) DD-SMC max product marginal measure $\hat{\Phi}_M(N^{(t)})$, (b) classifier posterior probabilities $c_N^{(t)}(N^{(t)})$, and (c) inferred marginal distribution over number of cells $\hat{P}(N^{(t)})$, from which the estimated division times are taken directly at the crossover points. In all graphs, time is shown in units of frames, each of which is 5 min.

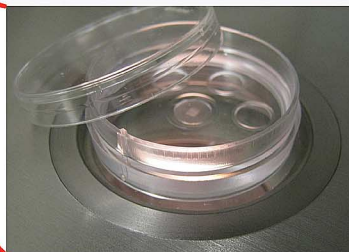
standard incubator



digital microscope
inside incubator



Images on front
panel display



Multi well petri dish
(group culture under
single media drop)

Fig. 6. Eva system comprising a uniquely designed dark field microscope (left) that holds a multi well dish customized for embryo cultivation under a single shared media drop (bottom, right). Microscope/dish combination fits within a standard incubator used by IVF clinics (top).

form of the times of first, second, and third mitosis t_1, t_2 , and t_3 , respectively. We trained our algorithm on 348 embryos and tested on a separate set of 417 embryos. An important consideration for detection performance is interpanelist variation. We observed that in a significant number of cases, panelists disagreed on the mitosis times. We measure the interpanelist disagreement

d_p as the rms disagreement between the two panelists assessed across all three transitions (t_1, t_2, t_3)

$$d_p = \sqrt{\frac{1}{3} \sum_{i=1}^3 d_{p,i}^2} \quad (15)$$

TABLE I
INTERPANELIST VARIATION FOR MEASURED MITOSIS TIMES AVERAGED ACROSS ALL TRANSITIONS FOR TRAINING AND TEST DATASETS

interpanelist disagreement	n	≤ 10 minutes	10 minutes to 1 hour	1 to 5 hours	5 to 10 hours	> 10 hours
training	348	250	28	26	41	3
test	417	312	29	11	57	8
total	765	562	57	37	98	11

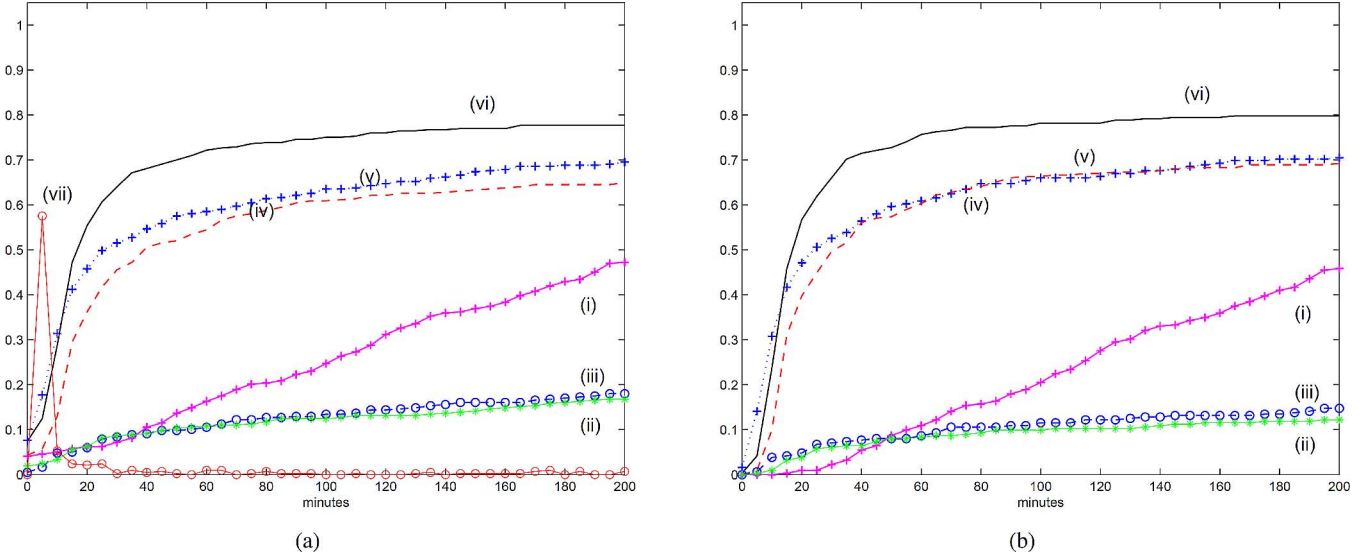


Fig. 7. Cumulative distribution functions of $\text{rmse}_{\text{tot}} - d_p$ in minutes for (a) all 417 test embryo sequences and (b) only test embryo sequences for which interpanelist was smaller than 10 min (312 embryos). Using (i) baseline fixed prior transition times averaged from training data, (ii) baseline transition times based on valley search of image similarity sequence, (iii) baseline traditional particle filter, (iv) DD-SMC tracking-based, (v) tracking-free, and (vi) all observables (ASSC). The distribution of d_p for all 417 embryos is also shown in (a), (vii).

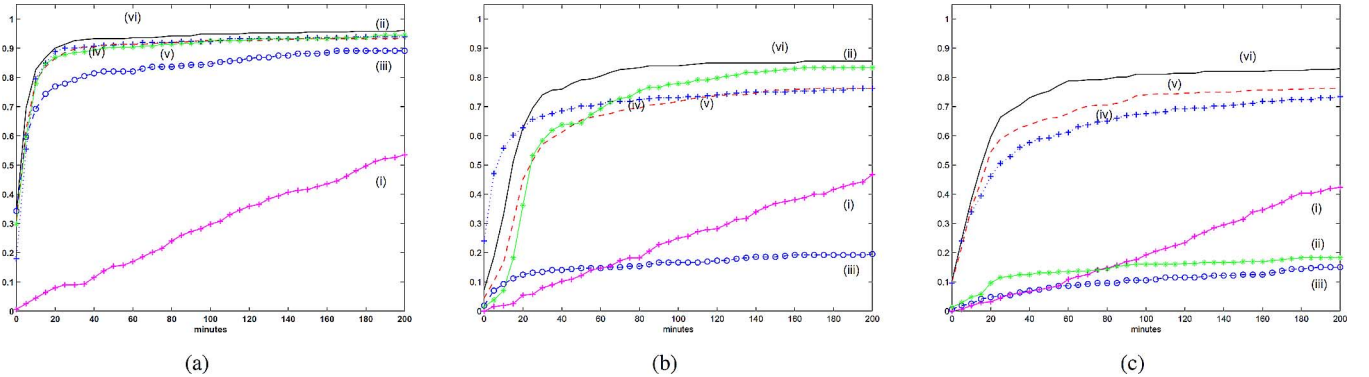


Fig. 8. Cumulative distribution functions of $\text{err}_i - d_{p,i}$ on all three transitions (a) t_1 , (b) t_2 , (c) t_3 for test embryo sequences for which interpanelist was smaller than 10 min (312 embryos). Using (i) baseline fixed prior transition times averaged from training data, (ii) baseline transition times based on valley search of image similarity sequence, (iii) baseline traditional particle filter, (iv) DD-SMC tracking-based, (v) tracking-free, and (vi) all observables (ASSC).

where $d_{p,i}^2 = E(t_i - E(t_i))^2$ is the variance of the panelist measurements for the i th transition times ($t_{i,1}, t_{i,2}$ for panelists 1 and 2, respectively, and $E(t_i) = (t_{i,1} + t_{i,2})/2$). We summarize this measure for all our datasets in Table I, and show its distribution in Fig. 7(a).

C. Overall Mitosis Detection Performance

We ran three versions of our algorithm on the test sequences: the complete ASSC algorithm, a tracking based version, and a tracking free version. These versions were implemented by modifying the event inference step. For the tracking based version, $c_N^{(t)}$ was set to one, and $\delta^{(t)}$ was set to 0.5, indicating no

prior on division events. For the tracking free version, $\hat{\Phi}_M$ was set to one, thus ignoring the DD-SMC tracking. In each of the cases, the prior term p_1 from (4) and (5) was relearned from the training data. We also implemented three baselines: a fixed prior, the times corresponding to local minima in the image similarity sequence, and a traditional particle filter as proposed in [38].

The fixed prior baseline is simply the mean of all transition times in the training data, and stays the same for all the test data. The second baseline is obtained by doing a simple valley search in the image similarity (NCC) sequences. The deepest three valleys in this sequence are taken to be the transition times. The

third baseline is a particle filter that generated particles but did not refine them, and used detected boundary points which were not grouped into segments. These runs were very long on the test dataset (more than a day on our cluster) as the number of particles required were over an order of magnitude higher than the DD-SMC due to the high dimensionality and not being data driven. Therefore, automated learning of parameters even with a simple local grid search was not feasible. A small number of critical parameters like c_r in (4) and Λ in (7) were empirically chosen by manual observation of a few cases and experimentation on the entire training set. This itself is an inherent disadvantage of traditional particle filters which do require extensive tuning and experimentation with model parameters.

We recorded the reported mitosis times in all six of these versions and evaluated performance. We calculated rmse_{tot} : the total rms error of the algorithm’s measurements

$$\text{rmse}_{\text{tot}} = \sqrt{\frac{1}{3} \sum_{i=1}^3 \text{err}_i^2} \quad (16)$$

where $\text{err}_i = t_{i,\text{alg}} - E(t_i)$ is the measurement error of the tracking algorithm compared to the average panelist for transition time t_i . We are interested in the incremental uncertainty incurred when using these methods relative to the panelists.

Fig. 7 shows the cumulative distribution function of the total deviation rmse_{tot} minus the average interpanelist disagreement d_p defined in Section III-B. We show this for all six algorithms tested. We show results for both the complete test dataset (417 embryos), as well as the portion of the test dataset which had low, ≤ 10 min, inter-panelist variation (312 embryos). On this dataset, the ratio of embryos that were tracked on all three transitions with an rmse_{tot} within 30 min of the interpanelist variation using the complete ASSC showed a relative improvement over tracking based and tracking free approaches of 25.6% and 32.9%, respectively. The relative improvement over any of the baselines was close to or greater than an order of magnitude. It can also be seen that the cdf of the fixed prior baseline is roughly a straight line, indicating a uniform distribution on the incremental uncertainty. In the case when d_p is small, it also reflects a roughly uniform distribution on the transition times themselves in the test dataset.

D. Per Transition Mitosis Detection Performance

We next performed a more detailed analysis on the performance for each transition for the low interpanelist variation portion of the test dataset. The results are shown in Fig. 8. The trend for the fixed prior baseline is similar for all transitions to that of the overall mitosis detection performance, and the transition times seem to have a fairly uniform distribution. By contrast the other two baselines start out with fair to good performance, and then deteriorate by the third transition. The particle filter baseline deteriorates fastest, followed by the valley search baseline. The valley search is understandably comparable to tracking free at finding the first and second transition. However, by the third transition it is very difficult to determine a 3–4 cell division based on image similarity alone, as the overall image change is more subtle and more scene details must be considered.

TABLE II
FRACTION OF DATASETS TRACKED TO WITHIN 30 MIN rmse_{tot} OF PANELISTS FOR EACH OF THE TRANSITIONS INDIVIDUALLY AS WELL AS ACROSS ALL TRANSITIONS, FOR BASELINE PRIOR MEAN TIMES, BASELINE IMAGE SIMILARITY DIPS, BASELINE PARTICLE FILTER TRACKING, DD-SMC TRACKING BASED, TRACKING FREE, AND ASSC. MEASURED ON DATASETS WITH LOW (≤ 10 MIN) INTERPANELIST DISAGREEMENT

<i>approach</i>	t_1	t_2	t_3	<i>all</i>
baseline prior mean times	0.0922	0.0698	0.0559	0.0196
baseline image similarity dips	0.8855	0.5866	0.1313	0.0642
baseline particle filter	0.7917	0.1346	0.0545	0.0705
DD-SMC tracking based	0.9006	0.6667	0.5288	0.5256
tracking free	0.8974	0.5705	0.6090	0.4968
ASSC	0.9263	0.7404	0.6859	0.6603

As far as the tracking based and tracking free components of ASSC, it can be seen that sometimes one approach works better. For example, in the t_3 transition (the most difficult), tracking free outperforms tracking based, as the scene is more complex and is no longer adequately modeled with simple shape and outlier assumptions. By contrast, in the t_2 transition, shape and structure of the two cell case is modeled better by the tracking based shape model than by the features in the tracking free approach. In all cases, combining the two approaches yields substantial improvement.

It should also be noted that overall tracking success is a higher bar to reach than each transition individually. This is why in some cases the overall success rate is lower than that of the worst transition.

The fractions of datasets for which transitions were measured with an rmse_i within 30 min of the interpanelist variation are summarized in Table II for each of the transitions.

E. Examples

To illustrate the behavior of our algorithms during successful and unsuccessful tracking, we show four examples of tracking results in Fig. 9. The first three are successful as they have a low rmse_{tot} , and the fourth one is not successful. For each case, we show the DD-SMC particle scores from the tracking based portion; and the classifier posterior probabilities from the tracking free portion. Together, these constitute the inputs to event inference, whose resulting final beliefs (the output marginal probabilities) are also shown. Corresponding image sequences for these examples are depicted as montages in Fig. 10. Videos of these sequences overlaid with labeled segments and inferred shapes are viewable in the Supplementary Materials. Each frame in these sequences is color coded based on the most likely number of cells inferred by ASSC for that frame.

IV. DISCUSSION

Our experiments confirm that a traditional model-driven approach is significantly challenged by the high dimensional search space, and that the search becomes more tractable if it is more data driven. This holds even if a purely-tracking based approach is used. The data in Fig. 7 and Table II show that DD-SMC tracking-based and tracking-free approaches have comparable performance and are each superior to the traditional particle filter baseline by close to an order of magnitude. By fusing the results of the tracking-based and tracking-free

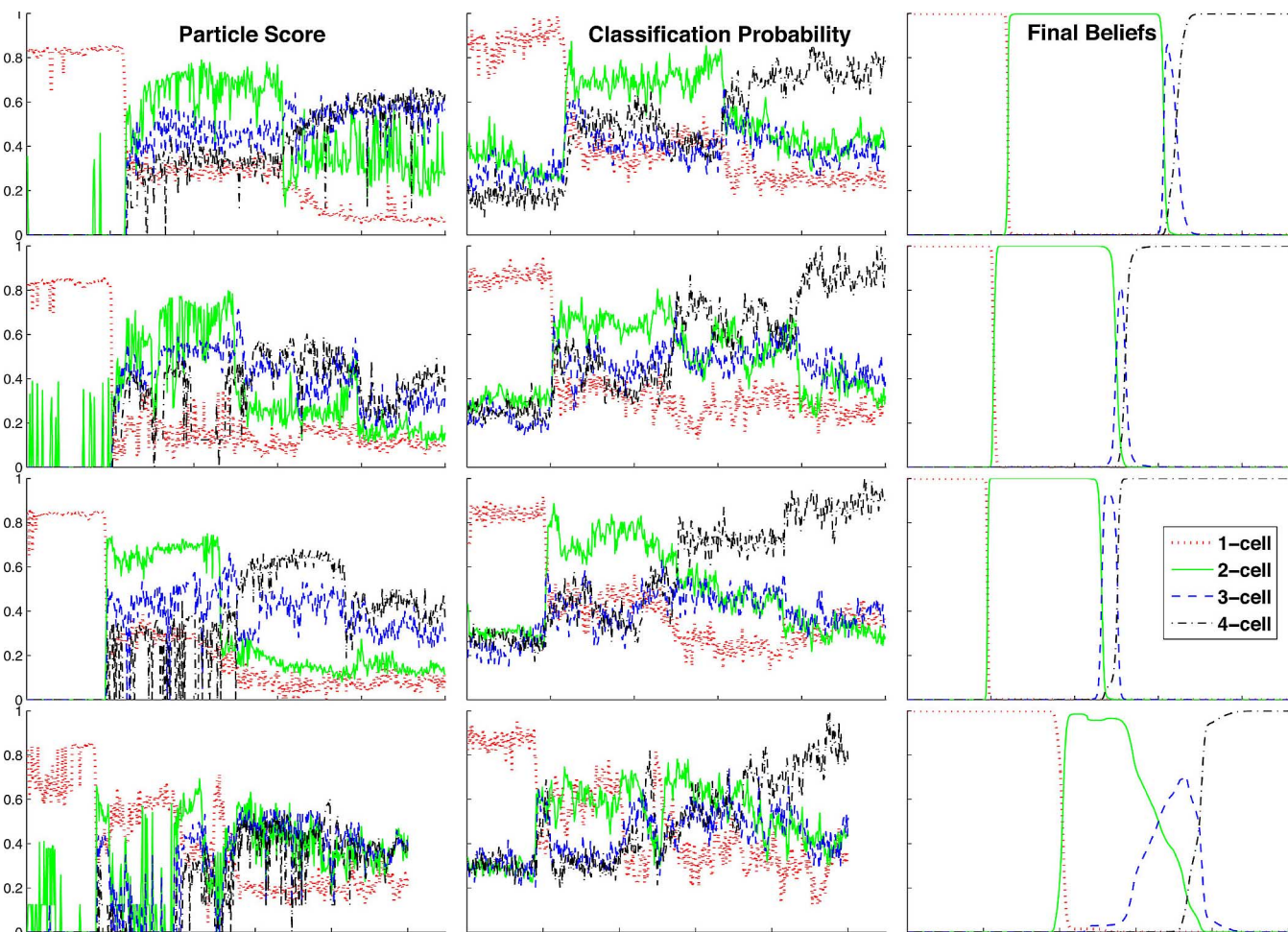


Fig. 9. Example output graphs for four examples. The first three detected mitosis events successfully and the last one did not. In each example, the DD-SMC particle scores, classifier probabilities, and final beliefs from belief propagation event inference are shown. While all graphs reveal some clues about the reliability of mitosis detection, the final beliefs summarize the situation best (as expected). Characteristics such as steady state period probabilities and transition slopes can potentially predict reliability.

approaches, ASSC achieves an improvement in performance of 25.6% over a purely tracking based approach and 32.9% over a purely tracking free approach.

The advantages of a more data-driven over a more model driven search are further highlighted at each transition. The results in Fig. 8 and Table II show that the model-driven approach diminishes rapidly as the number of cells increases in each transition. The performance drops by 83.0% from first to second transition and 59.5% from the second to third transition. The corresponding performance drops for ASSC are 20.1% and 7.4%, respectively. This suggests that while the search space becomes higher dimensional with each transition, the data-driven approach is more likely to keep the search more targeted to a reasonable subspace. Furthermore, more subtle scene details become more important at later transitions. This can be seen in the valley search baseline performance, in which the performance drops by only 33.8% from first to second transition but by 77.6% from the second to third transition which has more subtle scene details. It should also be noted that a valley search strategy cannot detect a synchronous division event in which two or more cells divide at once (e.g., $t_2 = t_3$), which is a common occurrence during healthy embryo development.

The relative behavior of tracking-based and tracking-free approaches can also be observed in Fig. 8. While the two to three cell transition is detected better by the tracking-based approach, the tracking-free approach overtakes it for the three to four-cell transition, which has more complexities in the scene. This suggests that the underlying model for our tracking-based portion can be improved for more than three cells, and at least some of the complexities not captured by this model are better captured by the classifier training. Such a trend would justify an alternative method of fusing of the two approaches in which the tracking free approach gets emphasized more at later transitions and stages. It also motivates the development of more sophisticated scene models, in which more complex shape and extra-cellular objects are included.

The probabilistic framework provides a collection of marginal distributions, which are more informative than the estimated transition times. This can be seen in the examples of successful and unsuccessful tracking in Fig. 9.

The successful tracking cases are associated with marginal distributions that are well behaved. Specifically, there are clear periods of one, two, and four cells which have high probabilities for sustained periods of time, and relatively sharp transitions.

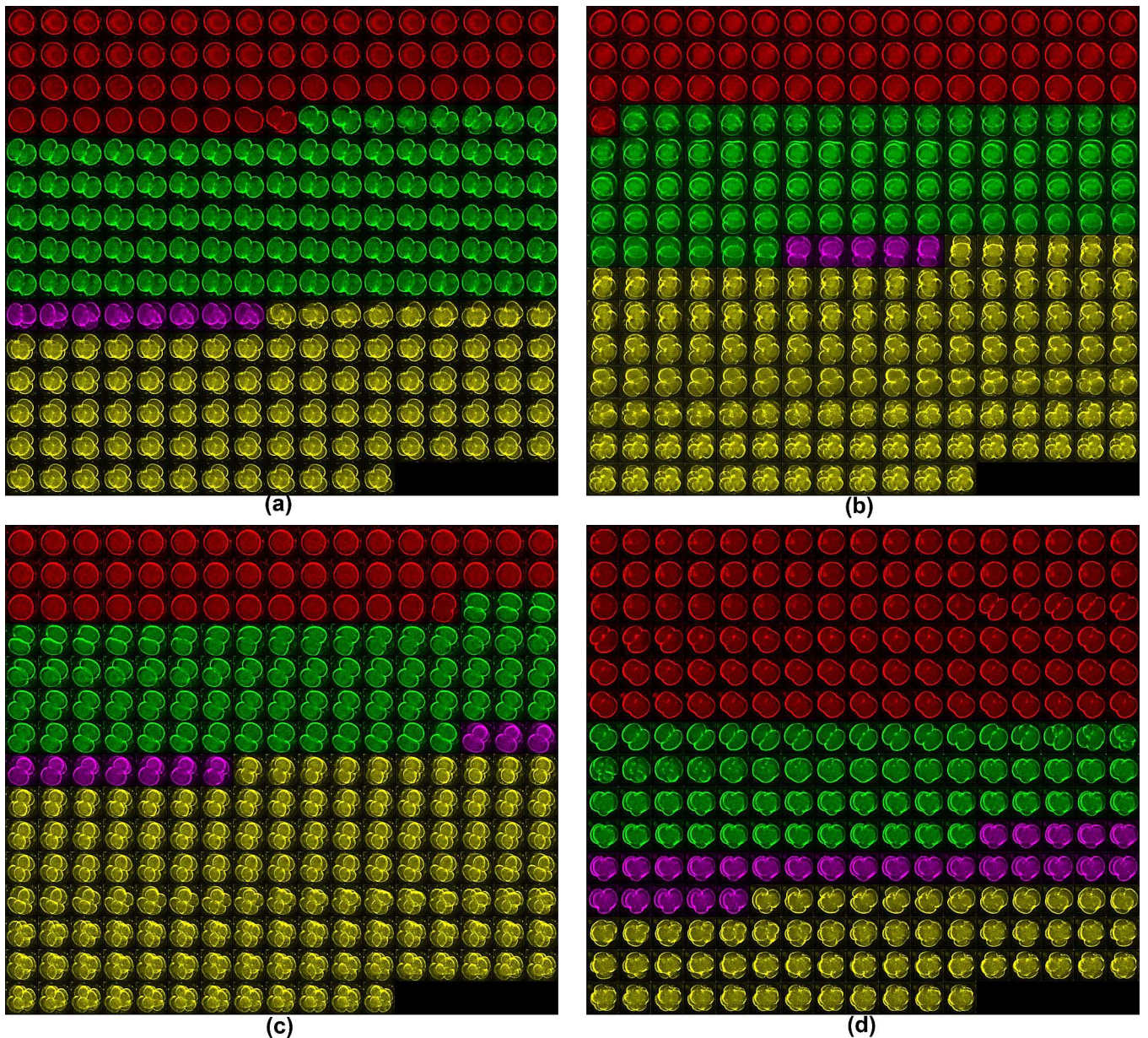


Fig. 10. Unrolled embryo image sequence (downsampled to 250 frames for illustration purpose) for the four cases represented in Fig. 9. Each frame is color coded based on the most likely number of cells inferred by ASSC for that frame.

The three cell periods tend to be short and not of high confidence due to their transitory nature. However, in the successful cases they still do have relatively clear transitions with other periods. The unsuccessful example deviates from this behavior. The two-cell period does not have a probability close to one during its dominant period. Also, the transitions between two- and three-cell periods are not sharp, reflecting uncertainty. We believe that the information in these graphs can be used to perform a self assessment of the quality of the inference, and identify cases where automated tracking was not successful so that an expert can manually intervene.

V. CONCLUSION AND FUTURE WORK

Automated tracking of early embryo development provides significant scientific and clinical benefits. Yet it remains difficult due to a multitude of challenges. Accordingly, our frame-

work takes a data-driven approach which combines multiple features and their spatial and temporal contexts in a unified CRF model that leverages tracking-based and tracking-free portions. We have demonstrated automated tracking comparable to manual expert measurements with a 32.9% improvement over a purely tracking-free approach and a 25.6% improvement over a purely tracking-based approach. This data-driven approach outperforms a traditional particle filter by close to an order of magnitude. The evidence suggests that the tracking-based approach works best in the early stages of one and two cells, and the tracking-free approach brings new improvement at three cells and beyond, when the geometric model is less applicable. With this insight, we can further improve overall performance by leveraging and learning from more labeled data as it becomes available, and incorporating more sophisticated models. We can also leverage the probabilistic framework's output in the form

of estimated marginal distribution over time to estimate a self assessment on the quality of the inference.

ACKNOWLEDGMENT

The authors would like to thank their colleagues in the engineering team in Auxogyn for their valuable technical contributions. The authors would also like to thank their external clinical partners and clinical team for making the imaging data and expert panelist measurements available to them, and biomedical research team for helpful discussions.

REFERENCES

- [1] O. Al-Kofahi, R. J. Radke, R. K. Goderie, Q. Shen, S. Temple, and B. Roysam, "Automated cell lineage construction: A rapid method to analyze clonal development established with murine neural progenitor cells," *Cell Cycle*, vol. 5, no. 3, pp. 327–335, 2006.
- [2] A. Beuchat, P. Thévenaz, M. Unser, T. Ebner, A. Senn, F. Uner, M. Germond, and C. Sorzano, "Quantitative morphometrical characterization of human pronuclear zygotes," *Human Reproduct.*, vol. 23, no. 9, pp. 1983–1992, Sep. 2008.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2007.
- [4] A. Blake and M. Isard, *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*, 1st ed. New York: Springer-Verlag, 1998.
- [5] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [6] S. Chavez, K. Loewke, J. Han, F. Moussavi, P. Colls, S. Munne, B. Behr, and R. R. Pera, "Dynamic blastomere behaviour reflects human embryo ploidy by the four-cell stage," *Nat. Commun.*, no. 3, 2012.
- [7] A. A. Chen, L. Tan, V. Suraj, R. R. Pera, and S. Shen, "Biomarkers identified with time-lapse imaging: Discovery, validation, and practical application," *Fertility Sterility*, vol. 99, no. 4, pp. 1035–1043, 2013.
- [8] J. Conaghan, A. A. Chen, S. P. Willman, K. Ivani, P. E. Chenette, R. Boostanfar, V. L. Baker, G. D. Adamson, M. E. Abusief, M. Gvakharia, K. E. Loewke, and S. Shen, "Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: Results from a prospective multicenter trial," *Fertility Sterility*, vol. 100, no. 2, pp. 412–419.e5, 2013.
- [9] M. Cruz, N. Garrido, J. Herrero, I. Prez-Cano, M. Muoz, and M. Meseguer, "Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality," *Reproductive Bio-Medicine Online*, vol. 25, no. 4, pp. 371–381, 2012.
- [10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
- [11] A. Doucet, N. D. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2005.
- [12] D. W. E. S. Filho and J. A. Noble, "A review on automatic analysis of human embryo microscope images," *Open Biomed. Eng. J.*, 2010.
- [13] A. El-Labban, A. Zisserman, Y. Toyoda, A. W. Bird, and A. A. Hyman, "Discriminative semi-Markov models for automated mitotic phase labelling," *Proc. IEEE Int. Symp. 9th Biomed. Imag.*, pp. 760–763, 2012.
- [14] M. Fitzgibbon, A. W. Pilu, and R. B. Fisher, "Direct least-squares fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, May 1999.
- [15] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," *Computer*, vol. 1496, pp. 130–137, 1998.
- [16] A. Giusti, G. Corani, L. Gambardella, C. Magli, and L. Gianaroli, "Blastomere segmentation and 3d morphology measurements of early embryos from Hoffman modulation contrast image stacks," in *Proc. 2010 IEEE Int. Conf. Biomed. Imag.: From Nano to Macro*, 2010, pp. 1261–1264.
- [17] D. Hlinka, B. Kalatova, I. Uhrinova, S. Dolinska, J. Rutarova, J. Rezacova, S. Lazarovska, and M. Dudas, "Time-lapse cleavage rating predicts human embryo viability," *Physiol. Res.*, vol. 61, pp. 513–525, 2012.
- [18] S. Huh, D. Fei, E. Ker, R. Bise, M. Chen, and T. Kanade, "Automated mitosis detection of stem cell populations in phase-contrast microscopy images," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 586–596, Mar. 2011.
- [19] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser, "Robust single-particle tracking in live-cell time-lapse sequences," *Nature Methods*, no. 5, pp. 695–702, 2008.
- [20] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng.*, ser. 82 D, pp. 35–45, 1960.
- [21] M. Kass, A. P. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.
- [22] Z. Khan, T. Balch, and F. Dellaert, "MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1960–1972, Dec. 2006.
- [23] D. Koller and N. Friedman, *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: MIT Press, 2009.
- [24] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," *Int. Conf. Mach. Learn.*, pp. 282–289, 2001.
- [25] K. Li, E. D. Miller, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell, "Computer vision tracking of stemness," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2008, pp. 847–850.
- [26] A. -A. Liu, K. Li, and T. Kanade, "Mitosis sequence detection using hidden conditional random fields," *Proc. IEEE Int. Symp. Biomed. Imag.*, pp. 580–583, 2010.
- [27] M. Meseguer, J. Herrero, A. Tejera, K. M. Hilligse, N. B. Ramsing, and J. Remoh, "The use of morphokinetics as a predictor of embryo implantation," *Human Reproduct.*, vol. 26, no. 10, pp. 2658–2671, 2011.
- [28] F. Moussavi, F. Amat, L. Comolli, G. Elidan, K. Downing, and M. Horowitz, "Markov random field based automatic image alignment for electron tomography," *J. Structural Biol.*, pp. 260–275, Mar. 2008.
- [29] F. Moussavi, Y. Wang, P. Lorenzen, J. Oakley, D. Russakoff, and S. Gould, "A unified graphical models framework for automated human embryo tracking in time lapse microscopy," presented at the IEEE Int. Symp. Biomed. Imag., Apr. 2014.
- [30] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.
- [31] G. Paternot, A. Wetzels, F. Thonon, A. Vansteenbrugge, D. Willemen, J. Devroe, S. Debrock, T. D'Hooghe, and C. Spiessens, "Intra and inter-observer analysis in the morphological assessment of early stage embryos during an IVF procedure: A multicentre study," *Reprod. Biol. Endocrinol.*, vol. 9, 2011.
- [32] D. Rueckert and P. Aljabar, "Nonrigid registration of medical images: Theory, methods, and applications," *Signal Process. Mag.*, vol. 27, no. 4, pp. 113–119, Jul. 2010.
- [33] J. A. Sethian, "Fast marching methods," *SIAM Rev.*, vol. 41, pp. 199–235, 1998.
- [34] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York: Springer, 2011.
- [35] R. Tomer, K. Khairy, F. Amat, and P. Keller, "Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy," *Nature Methods*, vol. 9, no. 7, pp. 755–763, 2012.
- [36] Y. Wang, F. Moussavi, and P. Lorenzen, "Automated embryo stage classification in time-lapse microscopy video of early human embryo development," *Proc. MICCAI*, 2013.
- [37] G. Welch and G. Bishop, An introduction to the Kalman filter Univ. North Carolina, Chapel Hill, NC, Tech. Rep., 1995.
- [38] C. C. Wong, K. E. Loewke, N. L. Bossert, B. B. A. Christopher, J. De Jonge, T. M. Baer, and R. A. Reijo Pera, "Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage," *Nature Biotechnol.*, vol. 28, pp. 1115–1121, 2010.
- [39] M. Wu, H. Y. Kwon, F. Rattis, J. Blum, C. Zhao, R. Ashkenazi, T. Jackson, N. Gaiano, T. Oliver, and T. Reya, "Imaging hematopoietic precursor division in real time," *Cell Stem Cell*, vol. 1, no. 5, pp. 541–554, 2007.
- [40] F. Yang, "Cell segmentation, tracking, and mitosis detection using temporal context," *Proc. MICCAI*, vol. 3749, pp. 302–309, 2005.