

A Linear Chain Markov Model for Detection and Localization of Cells in Early Stage Embryo Development

Aisha Khan¹ Stephen Gould¹
¹College of Engineering and Computer Science
The Australian National University
Canberra, AU
{aisha.khan, stephen.gould}@anu.edu.au

Mathieu Salzmann^{1,2}
²Computer Vision Research Group
NICTA
Canberra, AU
mathieu.salzmann@nicta.com.au

Abstract

We address the problem of detecting and localizing cells in time lapse microscopy images during early stage embryo development. Our approach is based on a linear chain Markov model that estimates the number and location of cells at each time step. The state space for each time step is derived from a randomized ellipse fitting algorithm that attempts to find individual cell candidates within the embryo. These cell candidates are combined into embryo hypotheses, and our algorithm finds the most likely sequence of hypotheses over all time steps. We restrict our attention to detect and localize up to four cells, which is sufficient for many important applications such as predicting blastocyst and can be used for assessing embryos in in vitro fertilization procedures. We evaluate our method on twelve sequences of developing embryos and find that we can reliably detect and localize cells up to the four cell stage.

1. Introduction

Medically assisted reproduction technology such as *in vitro* fertilization (IVF) enables infertile couples to achieve successful pregnancies. Intensive research is being done in this field to improve the factors impacting results and welfare of the patients. Amongst these factors the decision of the embryologist in order to choose the most promising embryos is critical. In most cases, embryologists select them by visual examination based on embryo development and morphological characteristics, and their evaluation is subjective. These morphological markers have very limited predictive value in the IVF outcome [12]. Recent advances in time-lapse microscopy technologies have led to the discovery of robust and reliable non-invasive markers that predict embryo quality. For example, the time between cell divisions during early embryo development—from fertilization to the four cell stage—can be used to predict embryo viability and ultimately lead to increased success rates in

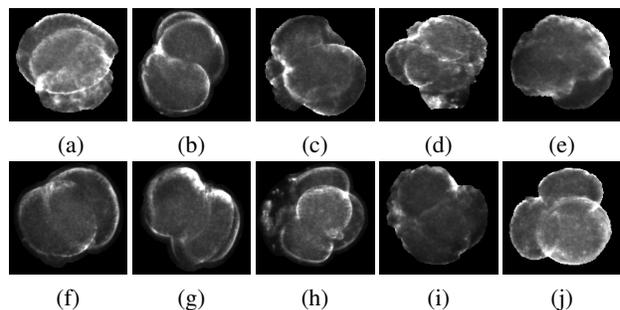


Figure 1. Complexities of the developing embryo. (a) Fragmentation in the 2-cell case. (b)–(c) Occlusion and fragmentation in the 3-cell case. (d)–(e) Fragmentation and occlusion in the 4-cell case. (f) Occlusion in the 2-cell case. (g)–(j) Morphological variation in the 4-cell case.

IVF pregnancies [5, 15].

Precise measurement of these timing parameters on a large scale requires automated tools for tracking cell divisions. This problem is challenging due to a diverse set of complexities such as poor morphology of the embryos, self-occlusion, deforming cell shapes, changing topology, missing data, poor visual features, and indistinctive and discontinuous cell boundaries (see Fig. 1 (a)–(e)). Moreover, the variability in morphologies within a cell stage (for example, for 4-cell cases see Fig. 1 (g)–(j)) and their similarity between cell stages (for example, for similarity between 2-cell and 4-cell case see Fig. 1 (f)–(g)) increases as the number of cells increases in an embryo. Additionally, due to imaging limitations the embryos are captured in a two dimensional image and lose information, such as depth, making events like cell occlusions hard to tackle. Furthermore, phenomena such as reverse cleavage (cell re-absorption), fragmentation, fragments re-internalization and other abnormal divisions may occur between time points, making it extremely difficult for an automated algorithm to deal with these critical biological processes. These challenges introduce uncertainty in the number and location of the cells and impose

difficulties in developing reliable algorithms for detecting cell divisions.

In this work, we utilize microscopy images of developing embryos taken at regular time intervals, and our goal is to detect cell divisions and locations up to the four-cell stage and detect and localize individual cells in these images.

In the context of the time-lapse microscopy images, many works have considered the problem of detecting cell divisions and classifying the cell development stages. These works can be categorized into tracking-based and tracking-free methods, each with and without explicit cell segmentation. We refer to methods that try to detect cell divisions by tracking cell boundaries or estimating trajectories and lineage as tracking-based, and to those that perform the detection directly from the image pixels and low level features as tracking-free. For example, Yang et al. [16] proposed a method that performs a spatio-temporal segmentation based tracking approach with level sets on human primary glioblastoma cells (U87-MG). Subsequently, Lou and Hamprecht [10] proposed a segmentation based cell tracking algorithm with a structured learning framework to learn associations between successive image frames. However, these approaches rely on staining cell nuclei to simplify detection. Our method is necessarily non-invasive due to ethical and legislative restrictions on studying them invasively, therefore simply detecting the cells is a challenge. A segmentation free cell tracking approach was proposed by Li et al. [8] to detect cell divisions in mouse myoblast stem cells using a cascaded classifier. However, divisions of stem cells are easy to detect compared to a human embryos where cells divide by cleaving in a relatively confined volume.

In the context of detecting cell divisions in human embryonic cells, a tracking-based method was proposed using a traditional particle filter by Wong et al. [15]. However, the traditional particle filter is challenged by the high dimensional search space of tracking multiple deforming shapes. State of the art cell stage prediction of human developing embryos reported by Wang et al. [14]. Their method is based on a 3-level classification scheme without explicit segmentation and tracking in human embryonic cells. In contrast, we present a machine learning approach with individual cell detection and segmentation/localization. Precise location of the cells can provide biologists with visual feedback and may aid in the discovery and characterization of novel biological phenomena. We experimentally compare our results to those of Wang et al. [14].

Recently, probabilistic graphical models have been used for detecting cell divisions in microscopy image sequences. For example [9], segmentation based mitosis phase labeling was performed in mesenchymal stem cells by applying a hidden conditional Markov random field. Subsequently, El-Labban et al. [2] presented a tracking-based method to label the mitotic phase for stained human epithelial adenocarci-

noma (HeLa) cells with a semi-Markov model. Their model is based on very simple features, such as intensity and gradient, whereas cell divisions in human embryonic cells are more challenging and require more powerful features.

As reviewed above, cell division detection usually involves cell segmentation and tracking. In the case of the developing human embryo it is very challenging to achieve this due to involved data complexity [14]. In this work to achieve our goal of detecting cell divisions and cell location we propose a framework that makes use of cell spatial information and combines two approaches that constrain the prediction over an entire sequence of frames: monotonicity enforcement (ME) and spatial continuity enforcement (SCE). The first approach (ME) enforces the number of cells to be monotonically non-decreasing over time and the second approach (SCE) enforces cells spatial continuity across frames. The framework takes a holistic approach by considering hypotheses up to four cells for each frame. These hypotheses are generated by a bottom-up data-driven method that attempts to find individual cells within the frame via an ellipse fitting procedure with a spatially diverse sampling procedure. We then track the evolution of these hypotheses over time using a linear chain conditional Markov random field that allows evidence from both past and future frames to influence our decision on the current frame.

2. Proposal-Based Embryo Localization

Our approach to tracking early stage embryo development makes use of a linear chain Markov model [6] to infer the most likely number and location of cells in the embryo. We begin by pre-processing the images to produce a set of ellipses representing candidate cells within the embryo. These candidates are then combined into hypotheses of one or more cells to describe the complete embryo. The hypotheses comprise the label space for each time slice in a linear chain Markov model with unary and pairwise potential functions for scoring hypotheses.

2.1. Image Pre-processing

We acquire images of embryos developing in a petri dish (see Fig. 2 (a)). A rough embryo mask is obtained by applying a shortest path algorithm to extract the embryo boundary in polar image space. We process the masked images by applying a Hessian operator followed by eigenvalue analysis to highlight the cell membrane. We call this the Hessian image (see Fig. 2 (b)). The same process has been applied in many interest point detectors to analyze local structure [1] and in many medical image-processing applications as a ridgeness measure [3].

2.2. Extracting Cell Candidates

Our method generates a set of cell candidates for each frame. We model cells as approximately elliptical in shape

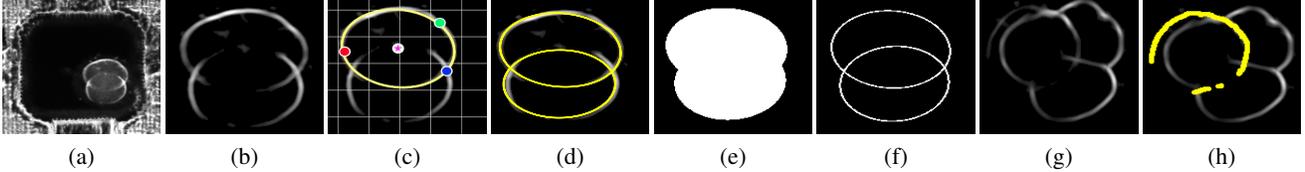


Figure 2. Illustration of preprocessing a light microscopy image I of a developing embryo (a) to produce an edge-enhanced ‘‘Hessian’’ image (b). The Hessian image H is used to propose cell candidates by ellipse fitting (c). Cell candidates are combined to propose embryo hypotheses (d). Features are extracted for each hypothesis using hypothesis mask M_h (e), and hypothesis perimeter P (f). An example of a cell with less evidence (g), highlighted in yellow (h).

and use a randomized sampling procedure in conjunction with an ellipse fitting method to generate candidates. These candidates are later combined to form hypotheses that completely describe an embryo consisting of one or more cells.

An ellipse can be parameterized as the set of 2D points $\mathcal{E} = \{(x, y)\}$ satisfying the quadratic constraint

$$a(x - x_0)^2 + b(x - x_0)(y - y_0) + c(y - y_0)^2 = 1, \quad (1)$$

where $b^2 - 4ac < 0$ for a valid ellipse. Here (x_0, y_0) represents the center of the ellipse and the parameters (a, b, c) encode its shape. Given a set of image points $\{(x, y)\}$ we apply the direct ellipse fitting method of Fitzgibbon et al. [4] to determine the five parameters (x_0, y_0, a, b, c) of the ellipse that best describes the points. Our method supports other ellipse fitting algorithms (e.g. [11]) but in early experiments we found that the method of Fitzgibbon et al. [4] applied to randomly sampled points worked best.

Generating a number of candidates for the one or more cells requires samples to be taken from the Hessian points to fit an ellipse. Selection of these sample points is not straightforward due to the presence of multiple cells and the non-uniform evidence for them (Fig. 2 (g)–(h)). Uniform sampling would result in taking more samples from the areas with dense Hessian points and ends up missing cell candidates for cells with less evidence. To address this problem we introduce a spatially diverse randomized sampling procedure. Briefly, we divide the Hessian image into an n -by- n grid (in our experiments we set n to 10). We then deterministically loop through all combinations of three grid squares (excluding squares that do not contain any Hessian points). In order to make our procedure robust we randomly sample a (Hessian) point within each grid square and fit an ellipse to all non-zero pixels within a small window around the sampled points. The procedure is illustrated in Fig. 2(c). Some of the grid squares will contain Hessian points (boundary fragments) from more than one cell (Fig. 2(c) blue point) or noise. This could result in either missing a cell candidate or a poorly fitted ellipse if the sample point is noise. To address this, we repeat this procedure a number of times using different random points. This generates a diverse set of cell candidates. Ellipses with high aspect ratio, small area or boundary exiting the field of view are discarded. We further reduce the number of candidates by clustering together

highly similar ellipses. Specifically, we use hierarchical agglomerative clustering [13] with the intersection-over-union (IoU) similarity measure, *i.e.*,

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (2)$$

where A and B represent the area of two ellipse candidates. We then select one representative candidate per cluster by choosing the representative with highest coverage ratio defined as the average distance between a point on the candidate ellipse and a non-zero Hessian image pixel. Precisely,

$$\text{CovRatio}(K) = \frac{1}{|\text{bd}(K)|} \sum_{p \in \text{bd}(K)} \min_{q \in H: H(q) \neq 0} \|p - q\|^2 \quad (3)$$

for ellipse K and Hessian image H . Here $\text{bd}(K)$ is the set of points along the boundary of ellipse K , *i.e.* its perimeter. This results in a small set of candidates (e.g., 30), which are combined to describe an embryo. With this sampling procedure we do not miss a cell candidate in any of the sequences considered in our experiments. Here we consider a cell candidate to be missing if none of the fitted ellipses overlap ($\text{IoU} > 0.7$) with the ground truth location of that particular cell.

2.3. Label Space

We wish to annotate each frame in the image sequence with the number and location of all cells within the frame. To this end, we consider all combinations of the candidates described in Section 2.2 as potential labels for a frame.

Formally, let $\mathcal{C}_t = \{c_i\}$ be the set of individual cell candidates for frame t . We generate a set of possible labels \mathcal{L}_t for frame t as

$$\mathcal{L}_t = \underbrace{\{(c_1) \in \mathcal{C}_t\}}_{\text{one cell}} \cup \underbrace{\{(c_1, c_2) \in \mathcal{C}_t^2\}}_{\text{two cells}} \cup \dots \cup \underbrace{\{(c_1, c_2, c_3, c_4) \in \mathcal{C}_t^4\}}_{\text{four cells}} \quad (4)$$

where \mathcal{C}_t^k is the set of all combinations of k ellipses from \mathcal{C}_t , and where we have assumed no more than four cells within the frame. Here, we define the set of combinations

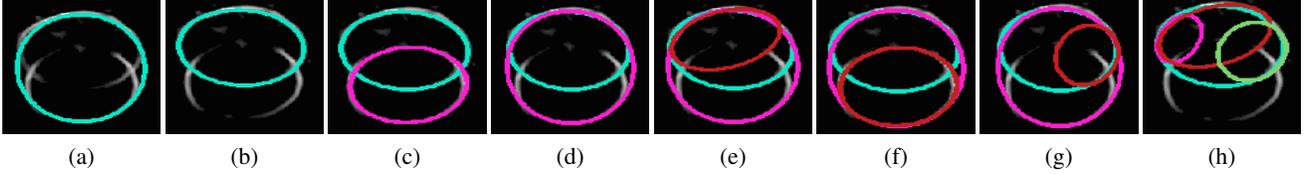


Figure 3. Label space examples for the 2 cell case of Fig. 2(b). (a)–(b) 1 cell Hypotheses, (c)–(d) 2 cell hypotheses, (e)–(f) 3 cell hypotheses, and (g)–(h) 4 cell Hypotheses.

\mathcal{C}_t^k to be unique up to permutation, *i.e.*, (c_1, c_2) is the same as (c_2, c_1) . For example, for the set of 30 cell candidates, the number of 4-cell hypotheses (\mathcal{C}_t^4) is 27405 (calculated as $\frac{30!}{26!4!}$). A label $x_t \in \mathcal{L}_t$ contains information about the number and location of cell candidates. Note that different frames will have different label spaces since these are generated from evidence for the cells within the image. For example, Fig. 3 ((a)–(h)) shows a few samples from the label space of the Hessian image shown in Fig. 2(b).

2.4. Linear Markov Chain Model

Given the set of possible labelings for each frame, our aim is to determine the best overall labeling, *i.e.*, the most likely number and location of cells throughout the entire time lapse microscopy sequence. We do this using a linear chain Markov model [7] with learned unary and pairwise potentials. Formally, let $x_t \in \mathcal{L}_t$ denote the label assigned to frame t . Then we can define the energy of a complete labeling over all frames $1, \dots, T$ as

$$E(x_1, \dots, x_T) = \sum_{t=1}^T \psi_t^U(x_t) + \sum_{t=1}^{T-1} \psi_{t,t+1}^P(x_t, x_{t+1}) \quad (5)$$

where the first summation is over unary terms (ψ_t^U), which score how well each label matches the image evidence, and the second summation is over pairwise terms ($\psi_{t,t+1}^P$), which enforce consistency between successive frames. For brevity we have omitted explicitly conditioning each term on image features (defined below).

Since our label space is comprised of complete embryo hypotheses for each frame, our graph is chain and inference is therefore easy. We score each assignment x_t using a learned unary function based on features derived from the image. We score joint assignments to adjacent frames (x_t, x_{t+1}) using the pairwise term, which can be encoded compactly. Finally, the most likely joint assignment over all frames can be determined efficiently via dynamic programming (or belief propagation [6]) as the assignment (x_1, \dots, x_T) that minimizes E .

2.5. Hypotheses Features

We encode each embryo hypothesis using 25 handcrafted and 7 learned features. Features are derived from the intensity image (I), the Hessian image (H), the thinned

Hessian image (H_{th}) (where the thinned Hessian image is obtained by reducing the Hessian pixels to single pixel connectivity), individual cell candidate masks (M_c), the hypotheses mask (M_h), individual cell candidate perimeters (P), and the hypothesis perimeter images (P_h). See Fig. 2 for an illustration of these images and masks. We categorize the features as: (i) Intensity Image Features, (ii) Shape Features, (iii) Hessian Image Features, (iv) Miscellaneous Features and (v) Number of Cells Predictor Features.

The features extracted from the intensity image measure the intensity distribution similarity between the embryo and the hypothesis, *e.g.*, similarity of the intensity histograms. The Hessian Image Features are derived from the Hessian image, the thinned Hessian image, and the hypothesis perimeter images. These features measure the boundary consistency between the hypothesis and the Hessian image and also capture the embryo geometry. We also derived features to capture the embryo shape and its similarity to the hypothesis based on, *e.g.*, solidity, orientation, convex hull, area, circularity and number of concave regions. The miscellaneous features capture other attributes of the hypothesis and the image frame, *e.g.*, hypothesis cardinality and frame number.

Number of Cells Predictor Features (PF):

We also learn a classifier to predict the number of cells (N) in each frame from global image features without performing any cell segmentation and tracking. The classifier takes the feature vector of a frame as input and returns a 4-dimensional probability vector for the number of cells. To this end, we used a boosted decision tree classifier trained on the 25 handcrafted features derived from the intensity variance, average intensity, circularity, concave regions, convex hull, eccentricity of the intensity and Hessian images. Knowing that the first frame of a sequence is always a 1-cell case, we compute additional features by taking the absolute difference between the feature value for the current frame and value of the corresponding feature in the first frame. This captures relative change of feature value for different sequences and accounts for variation in intensity or exposure across embryos. We refer to the resulting frame based prediction on the number of cells as FPN later in the experimental section. We defined additional features using the classifier probabilities to measure the consistency between the hypothesis cardinality and classifier predicted

Cell Cardinality	Number of Frames
1	1214
2	1588
3	74
4	1246
Total	4122

Table 1. Data set summary.

cardinality.

2.6. Learning the Unary Potential Function

The generated hypotheses are evaluated using a scoring function (unary potential) evaluated from the features described in Section 2.5. As scoring function, we use a boosted decision tree classifier learned from labeled training data. The negative log-probability returned by the classifier is used as our unary potential. From the complete label space of approximately 30,000 hypotheses for each frame we select the top 500 scoring hypotheses as labels in our linear chain Markov model. This makes pairwise computation and inference much faster.

2.7. Pairwise Potential Function

The role of the pairwise potential function is to model the evolution of cells over time. Concretely, the pairwise potential scores the compatibility of labels, x_t and x_{t+1} , for two consecutive frames. Since we wish to capture cell division events we use a simple model that enforces a monotonicity constraint, *i.e.*, the number of cells do not decrease from time t to time $t + 1$. More precisely, we define

$$\psi_{t,t+1}^P(x_t, x_{t+1}) = \begin{cases} 0, & \text{if } |x_t| \leq |x_{t+1}| \\ \infty, & \text{if } |x_t| > |x_{t+1}| \end{cases} \quad (6)$$

where $|x|$ denotes the number of cells represented by label x . We note that this is a naive biological assumption (in practice cell reabsorption does occur). More sophisticated pairwise terms can be developed but our simple approach works well for our problem.

Alternatively, we can enforce a cell to have spatial continuity across neighboring frames and our pipeline allows us to model this in a pairwise term. To this end, in addition to the above compatibility constraint we enforce cell boundary consistency constraint as

$$\psi_{t,t+1}^P(x_t, x_{t+1}) = \begin{cases} 1 - \text{IoU}^{1:1}(x_t, x_{t+1}), & \text{if } |x_t| = |x_{t+1}| \\ 1 - \text{IoU}^{1:M}(x_t, x_{t+1}), & \text{if } |x_t| < |x_{t+1}|, \end{cases} \quad (7)$$

where the function $\text{IoU}^{1:1}(\cdot, \cdot)$ measures the area of overlap between two hypotheses with the same number of cells. Here we consider all possible associations of cells in the first hypothesis with those in the second. For each association we compute the area of intersection of the two ellipses

divided by the area of their union. The function $\text{IoU}^{1:M}(\cdot, \cdot)$ is similar to $\text{IoU}^{1:1}(\cdot, \cdot)$, but allows a single ellipse in the first hypothesis to be associated with more than one ellipses in the second. This allows us to account for cell division.

3. Experiments

We evaluated the proposed method on twelve time-lapse image sequences consisting of a total of 4122 frames (see Table 1 for detail). The sequences were captured on embryos from two different patients using the EevaTM System developed by Auxogyn, Inc. Images were taken every five minutes over a period of two days. We evaluate our method using leave-one-out cross-validation and compared it to the method of Wang et al. [14] on the same subset of sequences for cell division accuracy. For ground truth these sequences were manually annotated for cell divisions and cell locations.

3.1. Experimental Details

We ran seven variants of our method and report results in Table 2. The methods are :

- Frame based predictor (FPN)
- Unary without prediction of the number of cells (Unary w/o PF)
- Unary with prediction of the number of cells (Unary + PF)
- Monotonicity enforcement (ME)
- Monotonicity enforcement with prediction of number of cells (PF+ME)
- Monotonicity with spatial continuity enforcement (ME+SCE)
- Monotonicity and spatial continuity enforcement with prediction of number of cells (PF+ME+SCE)

Monotonicity enforcement variants (*i.e.*, ME, PF+ME) are implemented with the pairwise term that enforces monotonicity constraint (Eqn. 6), and spatial continuity enforcement variants in addition enforces cell boundaries consistency constraint (Eqn. 7). The variants that include prediction of N (PF+ME, PF+ME+SCE) leverage the frame based prediction of the number of cells, which is incorporated as features (see Section 2.5 (V)) in the hypotheses scoring stage (Section 2.6). We evaluated our method for the tasks of predicting the number of cells, cell location and cell division time for all variants. In each case, we trained the model by holding out the sequence under test and sampling frames from the remaining sequences.

Our experiments were performed on a standard laptop with 8G memory with software implemented in Matlab. For each frame the computational cost (for sample generation, candidate generation and feature extraction) is of 373.8s, the inference does not add much to the computation cost.

Experiments	Cell Stage Prediction (%)						Cell Loc (%) IoU \geq 0.7	Cell Div. Acc. (No. Of Frames)			
	1-cell	2-cell	3-cell	4-cell	Avg.	Overall		1-2	2-3	3-4	Avg.
FPN	97.9	90.9	36.4	90.5	78.9	91.8	—	15.2	56.9	67.2	46.4
Unary w/o PF	97.6	93.7	5.4	93.4	72.5	93.2	76.9	12.0	30.0	32.9	24.9
ME	99.7	99.1	10.8	98.3	77.0	97.5	79.6	1.0	5.4	3.5	3.3
ME + SCE	99.7	99.3	6.7	99.0	76.2	97.7	80.1	1.0	5.0	3.0	3.0
Unary + PF	98.2	96.5	39.2	92.9	81.7	94.9	82.4	14.2	30.7	35.1	26.7
PF + ME	99.6	99.7	72.9	98.4	92.6	98.8	83.7	0.6	1.6	3.1	1.8
PF + ME + SCE	99.6	99.7	67.6	99.4	91.6	99.1	83.9	0.6	0.6	2.4	1.2
Wang et al. [14]	—	—	—	—	—	—	—	1.8	11.6	11.5	8.3

Table 2. Cell stage prediction, cell location accuracy, and cell division accuracy. For the cell stage prediction, the average is computed as the arithmetic mean of the one to four cell predictions, and the overall represents the fraction of correct instances.

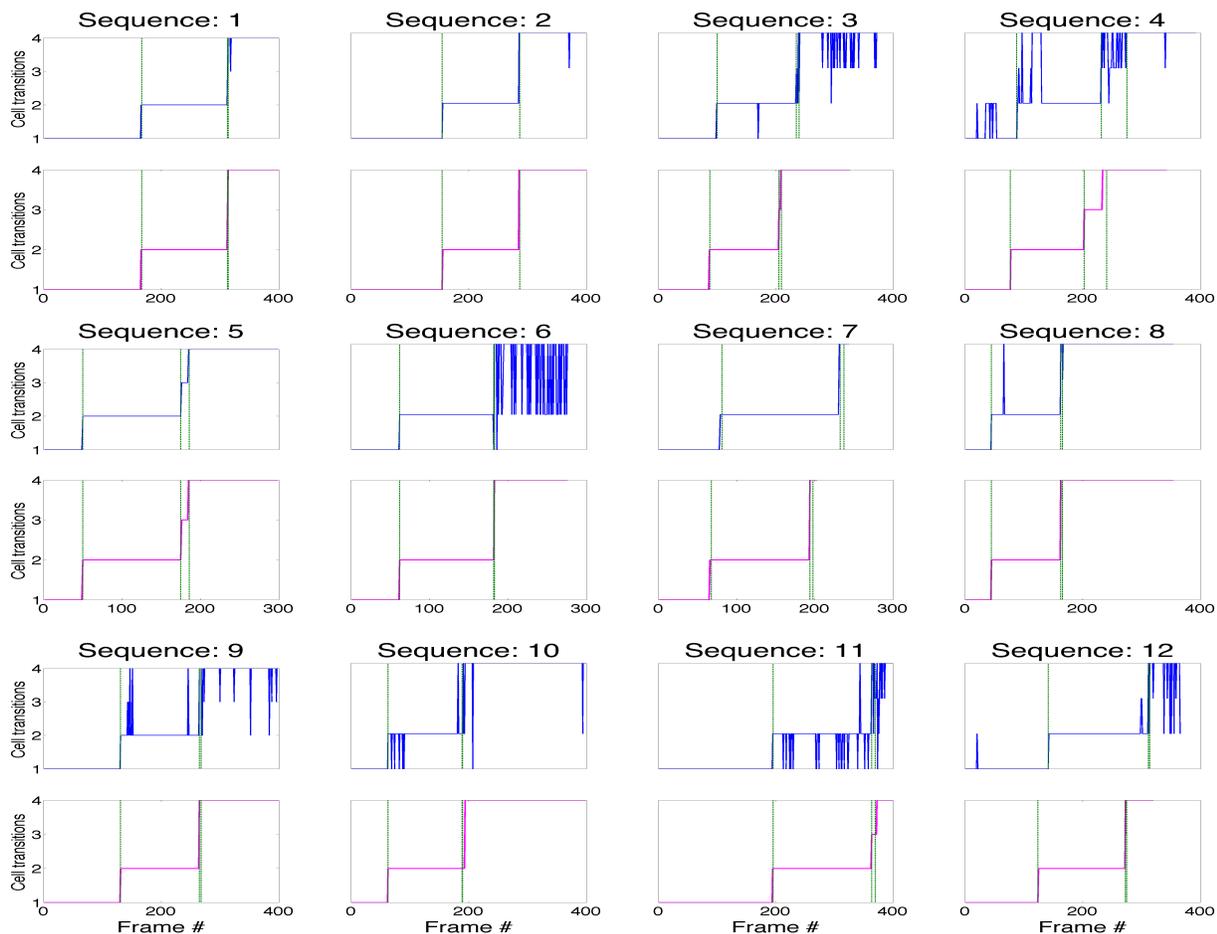


Figure 4. Blue lines: Cell divisions with unary potentials only. Magenta lines: Results with the Markov chain (PF+ME+SCE). The dotted green lines represents the ground-truth cell divisions.

With some optimisation our code can be made to run at frame rate (360s).

Task1: Predicting the number of cells: We first evaluate our method on the task of predicting the correct number of cells in each frame (see Table 2). The results show a higher overall accuracy of 99.1% with PF+ME+SCE variant, and a slightly lower accuracy of 98.8%, 97.7% and

97.5% with PF+ME, ME+SCE and ME, respectively. We analyzed the performance for each cardinality and the results show that our approach can reliably predict the correct number of cells up to the 4-cell stage (*i.e.*, 1-cell, 2-cell, 3-cell and 4-cell cases). The 3-cell case has comparatively low accuracy and this is because it is the most challenging cell stage in human developing embryos and lasts for a very

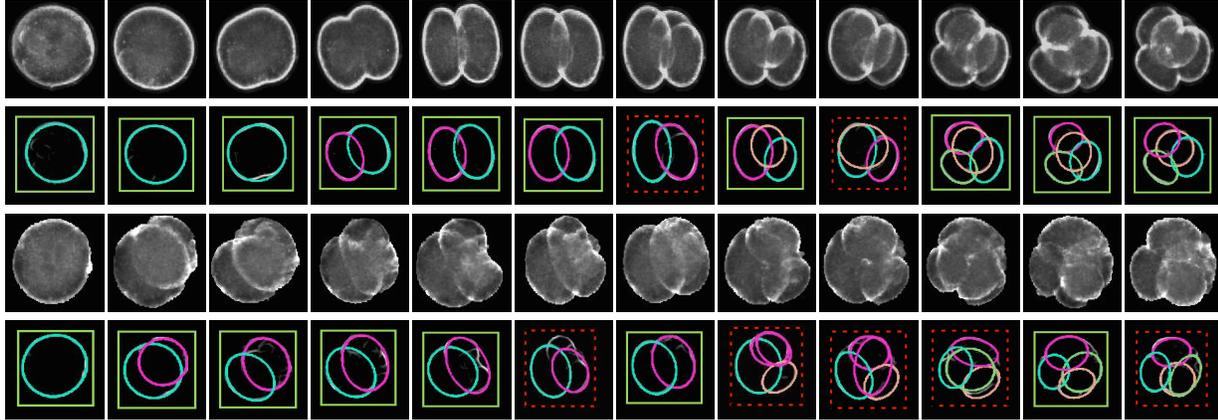


Figure 5. Cell detection and localization results for (non-consecutive) frames in two sequences along with their microscopy images. Figures with a green solid line rectangle show high accuracy results and red dotted line rectangles represent low accuracy ones (best viewed in color).

short time (only 1.79% of the frames in the dataset contains 3 cells). A performance boost for the 3-cell case (+62.1%) was achieved by combining the prediction of the number of cell and cell location information with the monotonicity enforcement approach. Accuracy dropped slightly (5.3%) when adding cell spatial continuity enforcement. This suggests that for embryos developmental stages with higher complexities, cell location and spatial continuity is difficult to model, and comparatively simpler approaches (*e.g.*, segmentation and tracking free) are more appropriate.

Task2: Predicting cell division: We also evaluated how well we predict the time at which cell division occurs (*i.e.*, the frame number) over the sequences (see Fig. 4). Here we calculated the cell division accuracy as the mean of the absolute difference between the predicted division time and our hand labeled ground truth divisions (see Table 2). Occasionally when we miss a division completely, we move on to predicting the next division (*e.g.*, from 2 to 4 cells directly). In such cases we consider the next cell division time as the time for the missed event. Overall the accuracy for detecting cell divisions is 3.3, 3.0, 1.8, and 1.2 with ME, ME+SCE, PF+ME and PF+ME+SCE, respectively. In terms of actual time our best accuracy of 1.2 (PF+ME+SCE) corresponds to a deviation of 6 minutes between the algorithm and the hand labeled cell divisions. In contrast, Wang et al. [14] cell division accuracy on these sequences is of 8.3, which corresponds to 41.5 minutes. Our method achieves state of the art accuracy for all three divisions and average is 7.1 higher than that of the method of Wang et al. [14] on these sequences. Performance improvement is seen for each cell division with the run that combines all approaches (PF+ME+SCE) together.

Task3: Predicting cell location: Finally we evaluate our ability to predict the correct location of each cell up to the 4-cell case. We compare the predicted cell locations

with our hand labeled ground truth using the intersection-over-union (IoU) similarity measure (Eqn. 2) and consider it to be a correct prediction if the IoU is above a pre-specified threshold. Table 2 shows that we can reliably predict correct location of the cells with the following overall accuracy: 79.6%(ME), 80.1%(ME+SCE), 83.7%(PF+ME) and 83.9%(PF+ME+SCE) for IoU of the predicted cell with ground truth > 0.7 . Fig. 5 shows examples of cell localization results.

3.2. Discussion

Comparison of the unary only (unary w/o PF, unary+PF) with Markov chain (ME, PF+ME, ME+SCE, PF+ME+SCE) variants shows superiority of the Markov chain approach on all three tasks. Also, two of the sequences (sequence 4 around time frame 91 and sequence 8 around time frame 67 face the biological phenomenon of cell re-absorption. Biologists do not consider an event as a cell division unless it stabilizes and disregard cell re-absorption. Cell re-absorption can be seen in the unary only results for these sequences (Fig. 4, plots with blue lines). All Markov chain variants (ME, ME+SCE, PF+ME and PF+ME+SCE) show their robustness to this phenomenon and smooth these occurrences correctly.

A significant improvement in the performance was attained by combining the prediction of N (PF) and cell location information along with the monotonicity enforcement (*i.e.*, on average 15.6% for Task1, 1.5 frames for Task2, and 4.1% for Task3, respectively). Incorporating spatial continuity further enhanced the performance by 0.6 frames, 0.2% for Task2 and Task3, respectively. For Task1 (predicting the cell stage) spatial continuity degrades the average performance by 1.0%. This is due to the complexity and scarcity of the 3-cell stage. Overall the best performance in all three tasks was obtained by combining the predictor of N with

cell location information in a framework with monotonicity and spatial continuity enforcement. This lets us conclude that gathering information from different aspects provides different insight and then combining this information enhances automated tasks such as cell detection and cell localization.

4. Conclusion

In this paper we have presented a novel approach to count and locate cells in time-lapse microscopy images of early stage human embryo development. Our model admits efficient inference over the entire sequence and allows us to exploit a rich set of features and pairwise compatibility terms. The main weakness of our approach is that the label space is exponential in the number of cells and thus only applicable to early stage development. Nevertheless, our experimental results show that the proposed technique is effective to detect and localize human embryonic cells. In future we plan to extend our model to handle the later stage development (*i.e.*, more cells).

Acknowledgement

We are grateful to Auxogyn, Inc. for their valuable support of this project. We also thank Farshid Moussavi and Peter Lorenzen for providing helpful comments on the manuscript. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy, as well as by the ARC through the ICT Centre of Excellence program.

References

- [1] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro. Principal curvature-based region detector for object recognition. In *CVPR*, 2007.
- [2] A. El-Labban, A. Zisserman, Y. Toyoda, A. W. Bird, and A. Hyman. Discriminative semi-markov models for automated mitotic phase labelling. In *ISBI*, 2012.
- [3] I. Ersoy, F. Bunyak, M. Mackey, and K. Palaniappan. Cell segmentation using hessian-based detection and contour evolution with directional derivatives. In *ICIP*, 2008.
- [4] A. Fitzgibbon, M. Pilu, and R. Fisher. Direct least-squares fitting of ellipses. *IEEE PAMI*, 1999.
- [5] A. Kiessling. Timing is everything in the human embryo. *Nature Bio.*, 2010.
- [6] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- [7] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, 2001.
- [8] K. Li, E. Miller, M. Chen, T. Kanade, L. Weiss, and P. Campbell. Computer vision tracking of stemness. In *ISBI*, 2008.
- [9] A.-A. Liu, K. Li, and T. Kanade. Mitosis sequence detection using hidden conditional random fields. In *ISBI*, 2010.
- [10] X. Lou and F. Hamprecht. Structured learning for cell tracking. In *NIPS*. 2011.
- [11] R. McLaughlin. Randomized hough transform: Improved ellipse detection with comparison. Technical report, The Uni. of Western Australia, 1997.
- [12] G. Paternot, A. M. Wetzels, F. Thonon, A. Vansteenbrugge, D. Willemen, J. Devroe, S. Debrock, T. M. D’Hooghe, and C. Spiessens. Intra-and interobserver analysis in the morphological assessment of early stage embryos during an IVF procedure: a multicentre study. *Reprod Biol Endocrinol*, 2011.
- [13] M. Steinbach, G. Karypis, and V. Kumar. A comparison of document clustering techniques. In *KDD Workshop on Text Mining*, 2000.
- [14] Y. Wang, F. Moussavi, and P. Lorenzen. Automated embryo stage classification in time-lapse microscopy video of early human embryo development. In *MICCAI*. 2013.
- [15] C. Wong, K. Loewke, N. Bossert, B. Behr, C. D. Jonge, T. Baer, and R. R. Pera. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature Bio.*, 2010.
- [16] F. Yang, M. Mackey, F. Ianzini, G. Gallardo, and M. Sonka. Cell segmentation, tracking, and mitosis detection using temporal context. In *MICCAI*, 2005.