AUTOMATED MONITORING OF HUMAN EMBRYONIC CELLS UP TO THE 5-CELL STAGE IN TIME-LAPSE MICROSCOPY IMAGES

Aisha Khan⋆   *Stephen Gould*⋆   *Mathieu Salzmann*†⋆

⋆ College of Engineering and Computer Science, The Australian National University, Canberra, AU
† Computer Vision Research Group, NICTA, Canberra, AU

{aisha.khan, stephen.gould}@anu.edu.au, mathieu.salzmann@nicta.com.au

ABSTRACT

Measurement of the proliferative behavior of human embryonic cells in vitro is important to many biomedical applications ranging from basic biology research to advanced applications, such as determining embryo viability during *in vitro* fertilization (IVF) treatments. Automated prediction of the embryo viability, by tracking cell divisions up to the 4-cell stage, improves embryo selection and may lead to increased success rates in IVF pregnancies. Recent research in cell biology has suggested that tracking cell divisions beyond the 4-cell stage further improves embryo selection. In the current state-of-the-art, later events (e.g., time to reach the 5-cell stage) can only be assessed manually. In this work we automatically predict the number of cells at every time point, and predict when the embryo divides beyond four cells in a time-lapse microscopy sequence. Our approach employs a conditional random field (CRF) that compactly encodes various aspects of the evolving embryo and estimates the number of cells at each time step via exact inference. We demonstrate the effectiveness of our method on a data set of 33 developing human embryos.

1. INTRODUCTION

Human assisted reproduction methods such as *in vitro* fertilization (IVF) are widely applied to treat infertility. The success of these methods relies on identifying the most viable embryos. In most cases embryologists select embryos by visual examination, which requires expert time and is prone to error. Recent advances in time-lapse microscopy technologies have led to the discovery of non-invasive prediction markers of embryo quality. Underlying these discoveries, the proper timing of embryo development has long been recognized as a critical factor in assessing embryo quality.

The first set of quantitative timing parameters reported by Wong et al. [12] have been confirmed to be correlated with the quality of human embryos and therefore can predict the viability of embryos as early as the second day post fertilization. Recently, additional parameters correlating with human embryo development have been added to complement these parameters. These additional developmental timing biomarkers include later events such as the time to reach the 5-cell

| (a) | (b) | (c) | (d) | (e) | (f) |
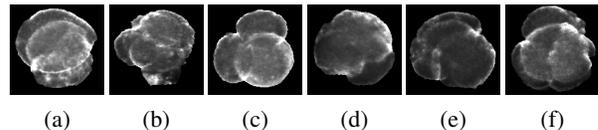
Fig. 1. Complexities of the developing embryo. (a) Fragmentation in the 2-cell stage. (b) Fragmentation and occlusion in the 4-cell stage. (c)–(d) Occlusion in the 4-cell stage. (e)–(f) Occlusion in the 5-cell stage.

stage, the length of the 5-8 cell interval and the timing of the fourth cleavage [1, 9].

Precise measurement of these timing parameters requires automated tools for tracking cell divisions. Automatic tracking of human embryonic cell divisions is currently limited to detecting cell divisions from the 1-cell stage to the 4-cell stage [1, 3, 10, 12]. This problem is challenging due to poor morphology of the embryos, self-occlusion, fragmentation and imaging limitations (e.g., see Fig. 1 (a)–(d)). Automated tracking of cell divisions beyond the 4-cell stage is a more difficult problem due to the increasing complexity of the developing embryo and the broad spectrum of biological phenomena that can occur (e.g., see Fig. 1 (e)–(f)). These complications demand more sophisticated techniques and impose a great technical challenge for an automated algorithm. Divisions to more than five cells become increasingly difficult to detect as cells tend to overlap more. Even human experts often disagree on the cell transition time [10]. To go beyond five cells it may therefore be necessary to keep track of the cells in several focal planes [9].

Most existing approaches to detect cells and cell divisions in microscopy images [2, 8, 13] rely on staining cell nuclei and cannot be applied to necessarily non-invasive human embryonic cells. Many other prior approaches [6, 7] are designed for cells other than the human embryonic cells and thus do not handle the various complexities involved in working with human embryonic cells.

The recent literature shows examples of applying proba-
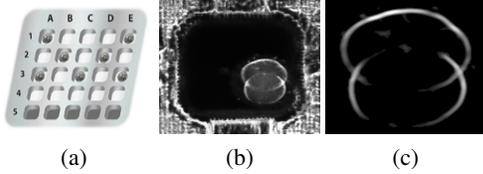
Fig. 2. Preprocessing a light microscopy image of a developing embryo. (a) Petri Dish. (b) Raw image. (c) Hessian image.

bilistic graphical models (PGM) to improve the selection of the viable embryo which may improve the success rates of IVF programs [4]. For example, Moussavi et al. [10] performed simultaneous segmentation and tracking in human embryonic cells up to the 4-cell stage. Their method uses geometric primitives based on cell boundary fragments in a conditional random field (CRF) model. However, embryos beyond the 4-cell stage are more complex, and cell segmentation and tracking becomes more challenging to achieve. In our previous work [3], we proposed a linear chain Markov model for detection and localization of cells in early stage embryo development (up to the 4-cell stage). This framework uses cell spatial information along with spatial continuity enforced over an entire sequence of frames. However, the main weakness of this approach is that the label space is exponential in the number of cells and thus does not easily go beyond the 4-cell stage. We experimentally compare our results to our previous approach [3] up to the 4-cell stage. These successful applications have motivated us to investigate PGMs for monitoring embryos beyond the 4-cell stage.

In this work we focus on automating measurement of the time to reach the 5-cell stage. We present a conditional random field to predict the cell stage of a human embryo (i.e., the number of cells) during the time-lapse imaging process up to five or more cells. Our framework integrates a rich set of discriminative features and visual cues. The proposed model avoids segmentation and tracking of individual embryo cells and can predict the number of cells and detect cell divisions directly in an image sequence.

2. METHODOLOGY

We perform automated monitoring of the embryonic cells beyond the 4-cell stage by infering the number of cells within each frame of a time-lapse microscopy sequence using a CRF model.

2.1. Image Pre-processing

Images of the developing embryos are acquired by the $Eeva^{TM}$ System developed by Auxogyn, Inc. Embryos are placed in a petri dish (see Fig. 2(a)) inside the incubators of IVF clinics and images are taken at five-minute intervals over a period of three to five days. We process the images to remove the dish boundary from the raw images (see Fig. 2(b)) by applying a rough embryo mask (see [3] for details).

2.2. Model Description

Given microscopy images of the evolving embryo, our goal is to predict the number of cells over time. We pose this problem in a CRF framework that integrates multiple visual cues and combines cell transition information of the evolving embryo. Formally, we represent the number of cells at time t with a discrete random variable N_t. Each variable N_t for $t \in \{1, \ldots, T\}$ can take on a label from the set $\mathcal{L}=\{1, \ldots, N^{max}\}$ where the last label corresponds to N^{max}-or-more cells in the embryo. As will be shown in the remainder of this section, our formalism admits efficient exact inference over a complete sequence of microscopy images.

We begin by performing independent frame-based prediction and then incorporate neighboring frame dependencies to the labeling via a smoothing pairwise term. We further improve the labeling by adding a cell transition predictor to the framework via a learned pairwise term.

2.2.1. Frame-Based Prediction

To perform frame-by-frame prediction of the number of cells we employed a boosted decision tree (BDT) classifier trained on a rich set of 88 handcrafted features. In particular, we used frame-based and cell evolution features [3]. Briefly, these features are derived from the intensity and Hessian image gray-level co-occurrence matrices (GLCM), convex hull, intensity variance, average intensity, circularity, concave regions, convex hull and eccentricity measures. The Hessian image is obtained from the raw image by applying a Hessian operator followed by eigenvalue analysis to highlight the cell membrane (see Fig. 2(c)). The frame-based features are computed from each frame, and cell evolution features are computed by taking the absolute difference between the current frame feature values and the corresponding feature values in the first frame (for an example of a frame-based feature see Fig. 3(a)).

After extracting these features for each frame a classifier is constructed by learning one-versus-all BDT classifiers for each cell cardinality separately. Once learned, their outputs are calibrated via a multiclass logistic classifier [11]. For prediction the classifier takes the feature vector of a frame as input and returns a probability distribution $P(N_t)$ for each frame $t \in \{1, \ldots, T\}$.

2.2.2. Adding Dependence to the Predictions

The frame-based prediction approach performs independent predictions using visual cues of the current frame only. In sequential data, as ours, predictions for the current frame are influenced by neighboring frames too. In some cases neighboring frame influence is vital. For example, sometimes a cell goes out of view by being completely occluded for a frame or two. Here neighboring frame influence can be captured by a CRF model with unary and pairwise terms.

Under the CRF formalism, instead of independent frame-based predictions, labeling/prediction corresponds to assign-
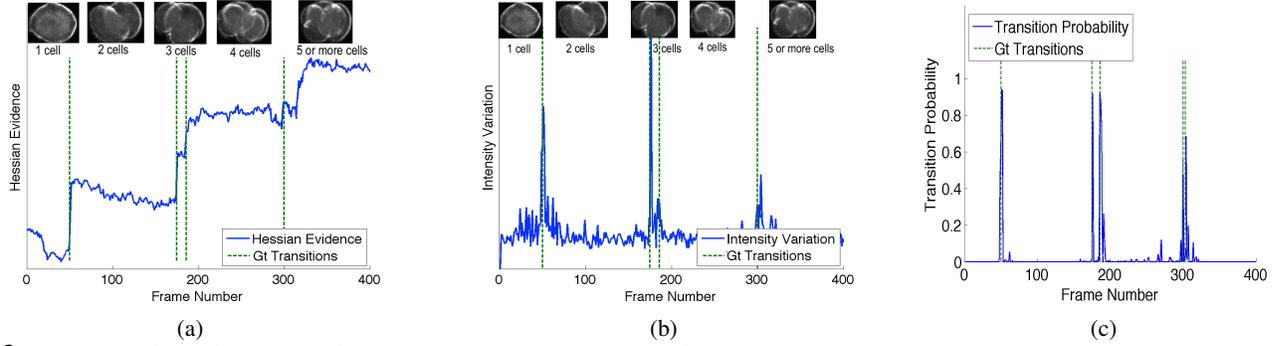
Fig. 3. Illustration of a (a) frame-based feature, (b) cell transition prediction feature. (c) Cell transition probabilities (cell transition prediction classifier) for a sequence. "Gt transitions" from human labeled grund truth.

ing a label to each variable to determine the best overall labeling for the complete sequence. Formally, let $N_t \in \mathcal{L}$ denote the label assigned to frame t. Then we can define the energy of a complete labeling over all frames $1, \ldots, T$ as

$$E(N_1, \ldots, N_T) = \sum_{t=1}^{T} \psi_t^{\text{U}}(N_t) + \lambda \sum_{t=1}^{T-1} \psi_{t,t+1}^{\text{P}}(N_t, N_{t+1}), \quad (1)$$

where the unary term (ψ_t^{U}) measures the consistency between the evidence of the frame and its label. We use the negative log-probability of the learned frame based classifier as a unary term.

The pairwise term (ψ_t^{P}) measures the consistency between neighboring frames by penalizing adjacent frames for taking different labels. The non-negative constant λ trades-off the strength of the pairwise term against the unary term and is chosen by cross-validation on the training set. Concretely, the pairwise term scores the compatibility of labels N_t and N_{t+1} for two consecutive frames. Since we wish to capture cell division events we use a simple model that enforces the number of cells not to decrease from time t to time $t + 1$,

$$\psi_{t,t+1}^{\text{P}}(N_t, N_{t+1}) = \begin{cases} 0, & \text{if } N_t \leq N_{t+1} \\ \infty, & \text{if } N_t > N_{t+1} \end{cases} \quad (2)$$

We seek the most likely number of cells for each frame, and ultimately the most likely sequence. This corresponds to the assignment that minimizes $E(N_1, \ldots, N_T)$, which can be obtained efficiently by dynamic programming.

2.2.3. Improving the CRF Model

Instead of simply encoding monotonicity of the number of cells we also want our pairwise term to influence the labels by scoring when cell division occurs. This requires prediction of cell divisions. To this end we use a BDT classifier that outputs the probability to have a transition. In addition to the features described above (Section 2.2.1), this classifier relies on a set of 46 features designed to capture transitions. These features are computed by taking the difference between the frame-based feature values of the consecutive frames. For example, the intensity variance between two consecutive frames is high when a transition occurs (for an example see Fig. 3(b)).

More specifically, for training, frames with manual annotation of cell divisions are used as positive examples and all frames in the sequences that do not contain transitions as negative examples. For testing, the BDT produces the probability p_D^t that any two frames t and $t + 1$ contain a cell transition (for an example of a complete sequence see Fig. 3(c)). Our model incorporates the classifier probability p_D^t on the cell transition for each time slice in the above-mentioned chain CRF via a pairwise term. It penalizes increase in the number of cells when a transition is not likely to happen as

$$\psi_{t,t+1}^{\text{P}}(N_t, N_{t+1}) = \begin{cases} 1 - p_D^t, & \text{if } N_t < N_{t+1} \\ p_D^t, & \text{if } N_t = N_{t+1} \\ \infty, & \text{if } N_t > N_{t+1} \end{cases} \quad (3)$$

Incorporating additional cues via the pairwise term does not increase the model complexity in terms of additional variables or tree width, and inference remains exact and efficient [5].

3. EXPERIMENTAL RESULTS

We evaluated the proposed approach on 33 time-lapse image sequences consisting of a total of 16,015 frames (with 21.7%, 28.4%, 1.8%, 28.7%, 19.5% of samples for 1 to 5-or-more cell cardinality, respectively). The sequences include embryos from eight different patients and show a certain degree of variation such as extra cellular material artifacts and cell reabsorption. We evaluated our method using leave-one-out cross-validation on the 33 sequences and compared it against our previous method [3] for up to the 4-cell stage on the same 12 sequences used there. For ground truth, the sequences were manually annotated for cell transitions. An important consideration for performance is interpanelist variation on the cell transition times [10]. So in our experiments a detected division frame is considered a true positive if it is within ± 3 frames to the ground truth.

We evaluated our method on the tasks of predicting the number of cells and localizing the cell division for $N^{\text{max}} = 4$ and $N^{\text{max}} = 5$. Results are reported on three variants of our method: i) Frame Based Prediction (FBP), ii) CRF with

		1-cell	2-cell	3-cell	4-cell	5-cell	Avg	Overall	Trans Acc (Avg)
$N^{max}=4$	FBP	98.4	96.0	59.4	97.2	—	87.8	96.5	16.4
	CRF	99.0	98.1	60.8	97.7	—	88.9	97.4	6.2
	CRF+TP	**100**	**99.7**	**88.5**	**99.8**	—	**97.0**	**99.6**	**0.7**
$N^{max}=5$	FBP	96.8	94.8	63.3	85.7	87.7	85.7	90.7	21.8
	CRF	98.2	96.5	66.1	87.2	91.3	87.9	92.7	9.9
	CRF+TP	**100**	**99.7**	**82.9**	**94.7**	**93.5**	**94.2**	**96.8**	**4.0**

Table 1. Number of cells prediction accuracy (%) and cell transition accuracy (number of frames).
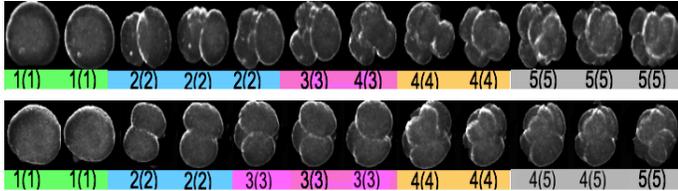
Fig. 4. Number of cell prediction results for (non-consecutive) frames of two sequences. Here, the number inside the parentheses is the ground truth and the number outside the parentheses represents the prediction.

Models	Cell Stg. Pred.	Cell Trans. Acc.	Exec.(s)
Khan et al. [3]	91.6	**1.2**	373.8
CRF+TP	**92.4**	2.5	**7.0**

Table 2. Baseline comparison: Cell stage prediction (%), cell transition accuracy (mean abs. diff.) and execution time per frame (sec).

monotonicity pairwise (CRF), and iii) CRF with learned cell transition probability pairwise term (CRF+TP).

We first evaluated our method on the task of predicting the number of cells in each frame (see Table 1). For both $N^{max}=4$ (99.6%) and $N^{max}=5$ (96.8%) the CRF+TP variant attained the highest overall accuracy followed by the CRF and then FBP. Fig. 4 shows examples of predicted numbers of cells. We see slightly lower performance on the 3-cell stage due to the scarcity of the 3-cell stage in the dataset (1.8%).

We also evaluated how well we predict the time of the cell divisions over the sequences. Here we report the mean absolute difference between the predicted division time (i.e., number of frames) and our hand labeled ground truth transitions (see Table 1). The overall best accuracy of 0.7 and 4.0 is achieved by the CRF+TP variant for both $N^{max}=4$ and $N^{max}=5$, respectively. This equates to 3.5 and 20 minutes deviation between the algorithm and the human expert for $N^{max}=4$ and $N^{max}=5$, respectively. The frame based variant performs poorly here because of the absence of influence from neighbouring frames in the model. Fig. 5 illustrates the mean cell transition error for each of the sequences. As clearly evidenced by the figure, the 4-to-5-or-more transition error is dominated by three of the 33 sequences. One of these sequences has high inter panelist disagreement on cell transition time and the other two sequences are affected by noise (e.g., fragments) in their embryo masks (Section 2.1).
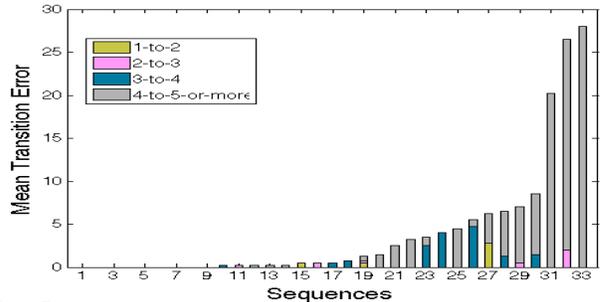
Fig. 5. Mean cell transition error (in number of frames) for each sequence. Sequences are sorted from lowest to highest error.

We see substantial improvement in performance on all metrics by adding a smoothness constraint (CRF) and further by adding additional information on the cell transitions (CRF+TP). In particular, for $N^{max}=5$ and for the two tasks (i.e., predicting the number of cell and the cell divisions), an improvement of 2.0% and 11.9 frames, respectively is obtained with the chain CRF over the FBP variant, and a further improvement of 4.1% and 5.9 frames respectively, is achieved by incorporating cell transition prediction in the CRF. Thus, for our problem the smoothness constraint is vital. Further addition of visual cues, such as cell transition information, enriches the underlying model and results in improving the overall performance.

In Table 2, we compare our results against our previous method [3] on the same subset of 12 sequences used in Khan et al. [3]. Note that our approach produces similar results, but is much faster. Also, our previous approach [3] is limited to predictions up to the 4-cell stage.

4. CONCLUSION

Recently reported embryo viability biomarkers include events that involve embryo monitoring beyond the 4-cell stage, such as time to reach five cells, which can only be assessed manually to date. Previous works focused on up to the 4-cell stage only and are not efficient or capable of going beyond the 4-cell stage. In this work we proposed a model that is capable of predicting beyond it. This will provide an objective, standardized embryo quality assessment free of human biases and hopefully will improve IVF outcomes. As future work we would like to incorporate cell location and cell tracking information for these later events.

References

[1] 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 and Sterility*, 2013.

[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] A. Khan, S. Gould, and M. Salzmann. A linear chain markov model for detection and localization of cells in early stage embryo development. In *WACV*, 2015.

[4] A. Kiessling. Timing is everything in the human embryo. *Nature Bio.*, 2010.

[5] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[6] K. Li, E. Miller, M. Chen, T. Kanade, L. Weiss, and P. Campbell. Computer vision tracking of stemness. In *ISBI*, 2008.

[7] A.-A. Liu, K. Li, and T. Kanade. Mitosis sequence detection using hidden conditional random fields. In *ISBI*, 2010.

[8] X. Lou and F. Hamprecht. Structured learning for cell tracking. In *NIPS*. 2011.

[9] M. Meseguer, J. Herrero, A. Tejera, K. M. Hilligse, N. B. Ramsing, and R. Jose. The use of morphokinetics as a predictor of embryo implantation. *Human reproduction*, 2011.

[10] F. Moussavi, W. Yu, P. Lorenzen, J. Oakley, D. Russakoff, and S. Gould. A unified graphical models framework for automated mitosis detection in human embryos. *IEEE Trans. Med. Imaging*, pages 1551–1562, 2014.

[11] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in large margin classifiers*, pages 61–74. MIT Press, 1999.

[12] 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.

[13] F. Yang, M. Mackey, F. Ianzini, G. Gallardo, and M. Sonka. Cell segmentation, tracking, and mitosis detection using temporal context. In *MICCAI*, 2005.