DISCRIMINATIVE SEMI-MARKOV MODELS FOR AUTOMATED MITOTIC PHASE LABELLING

*A. El-Labban, A. Zisserman**

Department of Engineering Science,
University of Oxford,
United Kingdom

Y. Toyoda, A.W. Bird, A. Hyman

Max Planck Institute of Molecular
Cell Biology and Genetics,
Dresden, Germany

ABSTRACT

With the widespread use of time-lapse data to understand cellular function, there is a need for tools which facilitate high-throughput analysis of data. We present a system for automated segmentation and mitotic phase labelling based on a wide margin discriminative Semi-Markov Model. This work takes the novel approach of using temporal features evaluated over the whole of the mitotic phases rather than over single frames, thereby capturing the distinctive behaviour over the phases. This approach extends and substantially improves on our previous approach of using dynamic time warping to align temporal feature signals to a reference.

1. INTRODUCTION

In the field of cell biology, there is an increasing use of time-lapse data to understand cellular function. Using automated microscopes, large numbers of images can be acquired, delivering 3D videos of cell samples over time. Analysing the images manually is extremely time consuming as there are typically thousands of individual images in any given sequence. Additionally, decisions made by those analysing the images, e.g. labelling a cell cycle stage, can be subjective, especially around transition boundaries between stages, leading to inconsistencies in the annotation. There is therefore a need for tools which facilitate automated high-throughput analysis.

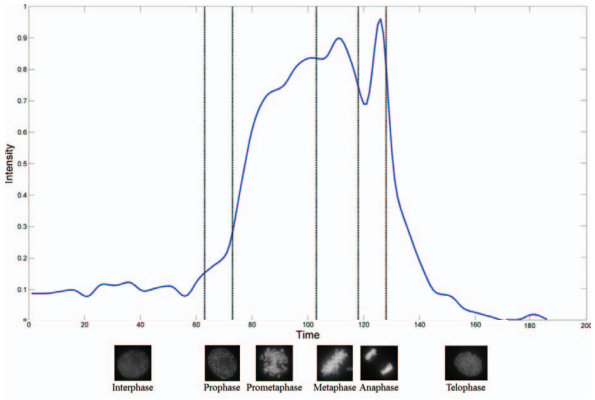
The objective of this paper is to automatically identify and track the individual cells throughout a time-lapse sequence, and to label the mitotic phase for each cell at every time point. Over large volumes of data, such outputs can be used to derive statistics on cellular function.

Existing approaches to mitotic phase labelling follow a number of common stages [1, 2, 3]: frames are first segmented into individual cells using image processing operations such as adaptive thresholding and watershed algorithms, and the cells are then tracked throughout the sequence. To classify the cell phase, each segmented cell is described by a feature vector which can include the shape, size, intensity (max, min, mean, variance) and texture. A prediction of the cell phase for each frame is then obtained by classifying the feature vector, e.g. by using a support vector machine (SVM). A final labelling can then be obtained using a temporal Hidden Markov Model (HMM) [4] across the track to correct the individual frame predictions. We follow the standard segmentation and tracking stages of this framework here (Section 2). The novelty in this work is that, instead of attempting to label the cell phase for each frame independently, the cell cycle stages are labelled using a discriminative Semi-Markov Model (SMM) using features evaluated over temporal segments of cell tracks. By evaluating features over segments rather than on a per frame basis, the SMM model is able to capture the distinctive behaviour during mitoic phases, as illustrated in Figure 1. This immediately produces segments with contiguous labellings without the need for the two-stage approach of individual frame labelling smoothed with an HMM. It also avoids the dependence on the unrealistic exponentially decaying duration model of HMMs. This is particularly useful as under many experimental conditions the duration of mitotic stages is altered, but the appearance of the chromatin is not. Detecting these changes can be key to elucidating insights into where a particular defect is occurring. This approach significantly improves on our previous work [5], which used dynamic time warping (DTW) to map temporal signals to labelled references to obtain labelled segments. The learnt SMM provides a richer model which outperforms the less general approach of DTW.

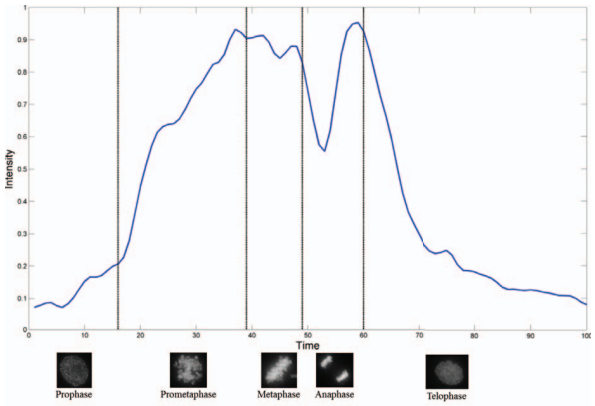
2. DATA & PREPROCESSING

Data. For the purposes of this paper, we use the MitoPhase dataset of [5]. This dataset consists of 54 3D time-lapse image sequences of HeLa (human epithelial adenocarcinoma) cells containing a fluorescently tagged chromatin marker, histone H2B-mCherry, acquired with a $60\times$ microscope objective. Of these sequences, 23 were obtained under conditions in which one of three proteins required for timely progression though mitosis (TACC3, CLTC, or GTSE1) were depleted from cells by RNAi. The videos have a temporal spacing of either 1

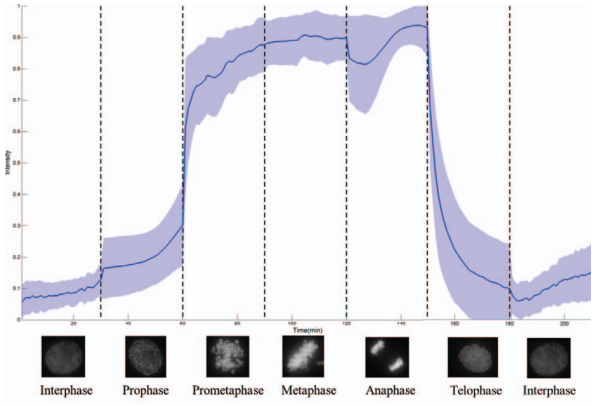
*This work was supported by EU ERC grant VisRec no. 228180.



(a)



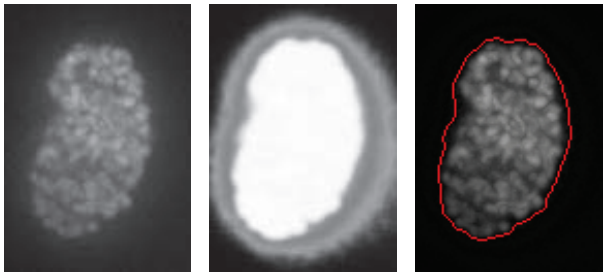
(b)



(c)

Fig. 1: (a)-(b) Maximum pixel intensity over time for 2 example cells from the dataset. Vertical black lines indicate transitions through mitosis. Corresponding stages of mitosis are shown below each image. (c) Maximum intensity signal averaged over the training data set. Signals are resampled such that interphase and each of the 5 mitotic stages has the same duration before averaging. The shaded region represents ± 1 standard deviation.

or 5 minutes, and last up to 300 frames. Frames have around 8–10 fully visible cells, with more partially visible around image boundaries, and there are up to 4 mitotic cells per video.



(a) (b) (c)

Fig. 2: (a) A sample image of a mitotic cell, cropped from a larger image. (b) Probability map output of the logistic regression classifier. White means high probability of being part of a cell. (c) Final segmentation boundary shown in red, overlaid on the original image.

There is range of 1–3 microns in the spatial resolution in the z-direction, with 5–14 planes imaged, and a resolution of approximately 0.2 microns in the xy-directions in each plane. For all the experiments and figures in this paper, each z-stack of 3D planes is reduced to a single image by maximum intensity projection. The 54 sequences are split into two equal partitions for training and testing.

Cell Detection by Segmentation. The approach taken here is to consider the segmentation as a two-class classification task where cells are one class (foreground) and background is the other. We do this in a two stage process which is illustrated in Figure 2. First, a logistic regression classifier [6] is used to generate a probability map of pixels being foreground. We then use graph cuts [7] to obtain the final, globally optimum, binary segmentation. The details are described in [5].

Tracking. After detection, cells are tracked using a nearest neighbour match on their centroids. This simple method assumes that the distance a cell is likely to move from one frame to the next is sufficiently small compared to the distance to other cells. In practice this is the case for the density of cells in our data, and there is almost no track confusion or track loss.

Ground Truth Annotation. The tracks are manually annotated, assigning each frame to interphase or one of the mitotic phases illustrated in Figure 1 (c). The tracks of all mitotic cells are collected to give a set of 119 tracks, 58 in the training set and 61 in the testing set.

Features. The maximum pixel intensity signal is extracted for each tracked cell at every time point. These signals are then normalised on a per-cell basis to lie in the range $[0, 1]$. Gradients are computed at two scales by filtering with the derivative of a 1×7, $\sigma = 1$ and a 1×13, $\sigma = 2$ Gaussian, then normalised on a per-cell basis to the range $[-1, 1]$. These three signals are concatenated to give a 3-vector at each time frame.

3. SEMI-MARKOV MODELS

A Semi-Markov model (SMM) [8] is an extension of a HMM, allowing the underlying process to be a semi-Markov chain with each state having a variable duration. By exploiting segment-level features, rather than just frame-level, this model overcomes the duration modelling limitation of HMMs, where the probability of remaining in a given state decays exponentially with time.

For our purposes, we use the wide margin discriminative SMM model of [8] which, for a given sequence $\mathbf{x}$, aims to find an optimal segmentation, into temporal segments of frames, $\mathbf{s} = [s_1 \dots s_n]$, given by: $s_j = [a_j, b_j, y_j]$ where a_j and b_j are the start and end positions of the segment respectively and $y_j \in [1 \dots 6]$ is label within the segment, interphase or one of the 5 mitotic phases.

By defining the model parameter $\mathbf{w}$ and feature mapping $\Phi(\mathbf{x}, \mathbf{s})$, the maximum likelihood optimal segmentation is then the one given by:

$$\mathbf{s}^* = \arg\max_{\mathbf{s}} \langle \mathbf{w}, \Phi(\mathbf{x}, \mathbf{s}) \rangle \tag{1}$$

which can be solved efficiently using a Viterbi-like algorithm.

Learning the model parameter $\mathbf{w}$ from a training set of manually labelled sequences is done by solving the regularised optimisation problem:

$$\min_{\mathbf{w}} \frac{||\mathbf{w}||^2}{2} + C \sum \xi_i \tag{2}$$

$$\text{s.t. } \langle \mathbf{w}, \Delta\Phi(\mathbf{x}, \mathbf{s}_i) \rangle \geq \Delta(\mathbf{s}_i, \mathbf{s}) - \xi_i \tag{3}$$

where the i subscript indicates ground truth labelling, $\Delta(\mathbf{s}_i, \mathbf{s})$ is the label loss, and $\Delta\Phi(\mathbf{x}, \mathbf{s}_i) = \Phi(\mathbf{x}, \mathbf{s}_i) - \Phi(\mathbf{x}, \mathbf{s})$. The set of constraints is over all possible segmentations $\mathbf{s}$.

We modify the label loss from the original of [8], which is given by the distance between the true and predicted sets of boundaries: $|\mathbf{s}| + |\mathbf{s}_i| - 2|\mathbf{s} \cap \mathbf{s}_i|$, to be the number of misclassified frames, ignoring segment boundary errors which are within 2 frames of the ground truth. Allowing this slack helps to overcome inconsistencies in the ground truth annotation which can arise due to the very gradual changes in appearance of the chromatin at some transitions. We train and test models using both loss functions and the performance improvement our loss function gives is shown in Table 1.

3.1. Feature mapping

In contrast to [5], where phase labelling is based on distance from a reference signal such as the one in Figure 1(c), the features described in Section 2 are used directly here. Each temporal sample of the feature is discretized into evenly spaced bins using "soft assignment", where the weight assigned to each bin is a linear function of distance to the bin centre, and represented by an L1-normalized *signal vector*. Bins of width 0.1 are used for the both the intensity and gradient features. This feature encoding is better suited to the linear kernel of the discriminative SMM. A sequence x consists of a concatenation of these temporal signal vectors. The feature mapping for a segment s_j, $\Phi(\mathbf{x}, s_j)$, is then formed from the concatenation of the following three features:

Model	Exact	±1	±2
Dynamic Time Warping	0.76	0.84	0.87
SMM with loss function of [8]	0.77	0.85	0.88
SMM with our loss function	**0.81**	**0.89**	**0.91**

Table 1: Phase Labelling Results. Scores shown are the mean per class accuracy i.e. the mean of the diagonal of the confusion matrix. Results are given for exact matching and allowing for 1 or 2 frames slack in detection of the transition boundary. Semi-Markov Models trained using both loss functions outperform the Dynamic Time Warping approach. Our approach of allowing the 2 frame slack in the loss function shows a more substantial improvement over DTW.

Segment level feature. This captures the characteristics of the signal over the segment as a whole. $\phi_1(\mathbf{x}, a_j, b_j, y_j) = f_1(\mathbf{x}, a_j, b_j) \otimes y_j$, where $\otimes y_j$ represents the tensor product with a vector of length equal to the number of classes and with the y_jth element set to one and the rest to zero (as used in [9]). $f_1(\mathbf{x}, a_j, b_j)$ consists of the signal vector at 5 evenly spaced points taken from the segment of $\mathbf{x}$ from a_j to b_j, linearly interpolating where necessary.

Segment boundary feature. This captures information about the start of a segment, obtained from a five frame local window centered on the boundary frame a_j. The feature is given by the mean of the signal vectors in the window and the tensor product with the label vector is taken.

Neighbouring segment level feature. This captures the interaction between neighbouring segments therefore allowing the model to learn which transitions between labels are feasible. The feature is the concatenation of the segment level feature evaluated over the current segment from a_j to b_j and over a subsequent segment from b_j to $b_j + d$, where d is a defined minimum duration. In this case the tensor product is taken with both the current and subsequent label vectors.

4. EVALUATION & RESULTS

The SMM is trained and tested using the data split described in Section 2 using both loss functions described in Section 3. Confusion matrices of frame classification into phases are generated for each test sequence after labelling all frames. The mean per-class accuracy is then calculated as the mean of the diagonal of the confusion matrix. This measure is preferred to the per-frame classification accuracy which is dominated by correct classification of interphase frames which appear approximately $10\times$ more than the next most frequent phase in the data.

Classification accuracy is also measured for the cases where transition boundaries are allowed to be within ±1 and ±2 frames of their true positions. Table 1 shows the results

for both versions of SMM and our previous DTW approach using the same features. There is a clear improvement in classification accuracy when using the SMM, which achieves very high accuracy within two frames of the true boundaries. The table also shows the improvement given by our change to the original loss function of [8].

It can be seen from the confusion matrices shown in Table 2, which shows results using our loss function, that allowing the slack during testing results in significant improvement in prophase, metaphase and anaphase. This is likely due to the fact that these phases have very short durations on average in the MitoPhase data (7, 10, and 4 mins respectively on average, compared with an average of 206 mins for interphase) and can appear for as little as a single frame in the sequences with the lowest temporal resolutions.

The significant errors occur at the end of prometaphase and telophase where, as seen in Figure 1, the change in the intensity signal is quite gradual and frequently affected by signal noise. The error is particularly clear for the case of telophase where the model consistently predicts a transition to interphase too soon, suggesting that intensity alone is not sufficient to distinguish between the two.

The mean per-class evaluation is very sensitive to the number of frames in the shorter phases. This is shown by the fact that the overall proportion of correctly classified frames only improves from 0.89 to 0.91 by allowing a 2 frame slack, whereas the mean per class score improves from 0.81 to 0.91.

5. CONCLUSIONS

The approach we have proposed equips feature based cell-cycle phase recognition with an additional discriminative Semi-Markov Model so that image information over temporal segments, rather than just at a specific point of time, is utilized for recognizing different phases of the cell cycle. The approach has been shown to outperform the Dynamic Time Warping approach of our previous work, using the same features and data. We expect further improvements by adding features such as the temporal variation in texture and area.

	Interphase	Prophase	Prometaphase	Metaphase	Anaphase	Telophase
Interphase	0.97					
Prophase	0.17	0.73	0.10			
Prometaphase			0.80	0.16		
Metaphase			0.12	0.81	0.07	
Anaphase				0.06	0.82	0.12
Telophase	0.22					0.75

(a) Exact boundary. Mean per-class accuracy: 0.81

Interphase	0.97					
Prophase	0.12	0.87				
Prometaphase			0.85	0.12		
Metaphase			0.06	0.92		
Anaphase					0.96	
Telophase	0.21					0.78

(b) ± 1 frame. Mean per-class accuracy: 0.89

Interphase	0.98					
Prophase	0.09	0.91				
Prometaphase			0.87	0.10		
Metaphase				0.95		
Anaphase					0.98	
Telophase	0.19					0.80

(c) ± 2 frames. Mean per-class accuracy: 0.91

Table 2: Confusion matrices of testing results allowing up to 2 frames of slack in the classification. Rows represent the true classes and columns are the predictions. Values less than 0.05 have not been shown.

References

[1] N. Harder, F. Mora-Bermdez, W. Godinez, J. Ellenberg, R. Eils, and K. Rohr, "Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences," in *MICCAI*, 2006.

[2] N. Harder, F. Mora-Bermúdez, W. J. Godinez, A. Wünsche, R. Eils, J. Ellenberg, and K. Rohr, "Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time.," *Genome research*, 2009.

[3] M. Held, M. H. A. Schmitz, B. Fischer, T. Walter, B. Neumann, M. H. Olma, M. Peter, J. Ellenberg, and D. W. Gerlich, "CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging.," *Nature methods*, Aug 2010.

[4] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–286, 1989.

[5] A. El-Labban, A. Zisserman, Y. Toyoda, A. W. Bird, and A. Hyman, "Dynamic time warping for automated cell cycle labelling," in *MIAAB*, 2011.

[6] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[7] Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in *Proc. ICCV*, 2001, vol. 2, pp. 105–112.

[8] Q. Shi, L. Cheng, L. Wang, and A. J. Smola, "Human action segmentation and recognition using discriminative semi-markov models.," *IJCV*, 2011.

[9] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, "Large margin methods for structured and interdependent output variables," *JMLR*, 2005.