Dynamic Time Warping for Automated Cell Cycle Labelling

A. El-Labban¹, A. Zisserman¹, Y. Toyoda ², A. Bird², A. Hyman²

¹Department of Engineering Science, University of Oxford, United Kingdom
²Max 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 cell cycle phase labelling based on aligning temporal signals of simple features directly to a reference signal using Dynamic Time Warping (DTW). This is shown to result in a very accurate temporal labelling, significantly outperforming an approach based on independent frame classification. The method is evaluated on two datasets, acquired under different imaging conditions, and is found to perform well on both. We also describe a Hidden Markov Model implementation of the DTW which has comparable results.

I. 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 cell cycle 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 cell phase labelling follow a number of common stages [1], [2], [3], [4]: 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 (Haralick [6] and statistical geometric features [7]). 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) [8] across the track to correct the individual frame predictions. We follow the standard segmentation and tracking stages of this framework here (Section III). 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 by aligning temporal signals of simple features directly to a reference signal (learned from a training set of annotated cells). This is described in section VII. As will be seen in section VII this alignment results in a very accurate temporal labelling, so that cell cycle duration can be measured precisely. Figure II shows an example of an individual cell signal and the reference signal.

II. DATA

For the purposes of this paper, we created the MitoPhase dataset. This dataset consists of 54 3D time-lapse image sequences of cells containing a fluorescently tagged chromatin marker, histone H2B-mCherry, acquired with a 60x microscope objective. The videos have a temporal spacing ranging from 1 to 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. 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 x-y-dimensions 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. Some examples of typical frames from the dataset can be seen in Figure A.

The 54 sequences are split into two equal partitions for training and testing. After applying the segmentation and tracking process described in Section III, there are 650 resulting tracks. These tracks are manually annotated, assigning each frame to one of the stages of the cell cycle illustrated in Figure II (b).

The tracks of all mitotic cells showing at least 3 of the phases of Figure II are collected to give a set of 119 tracks, 61 in the training set and 58 in the testing set. Within each of these a set of temporal features are computed to give signals such as the one shown in Figure II (a).

III. CELL SEGMENTATION & TRACKING

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 A. First, a logistic
regression classifier [9] is used to generate a probability map of pixels being foreground:

\[
P(\text{foreground}|x, w) = \frac{1}{1 + e^{-wx}} \tag{1}
\]

where \( x \) is a feature vector evaluated at every pixel in the image, and \( w \) a weight vector learned by training on a manually segmented subset of the training data. The feature vector is described in Section VII.

We then use graph cuts [10] to obtain the final, globally optimum, binary segmentation by minimising the cost:

\[
E(f) = \sum_p D_p + \lambda \sum_{p,q} V_{pq} \tag{2}
\]

where \( D_p \) is logistic regression output and \( V_{pq} \) is a contrast dependent pairwise cost, which penalises differing labels in neighbouring pixels unless there is an intensity gradient between them.

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, and there is almost no track confusion or track loss.

IV. TEMPORAL ALIGNMENT

Our approach to labelling frames in a novel track into stages of the cell cycle consists of aligning a track signal to a reference signal with known transition times using dynamic time warping (DTW) [11]. DTW optimally warps (stretches) one sequence onto another to minimize an alignment cost under a set of allowed deformations which consist of repeating or removing samples in either the novel signal or the reference signal.

As a reference signal, the mean of the signals in a training set is used. Before taking the mean, the individual signals are resampled such that each cell cycle phase has a fixed duration, thereby aligning the stages and fixing the transition boundaries.

Fig. 1: (a) Maximum pixel intensity over time for a single cell. Vertical black lines indicate transitions through the cell cycle stages. (b) Maximum intensity signal averaged over the training data set. Signals are resampled such that each stage has the same duration before averaging. Corresponding cell cycle stages are shown below. (c) Maximum intensity signal averaged with stages resampled to the mean duration of all instances in the data. Shaded regions in (b) and (c) represent ±1 standard deviation.

Fig. 2: Four typical time frames from one of the sequences in the MitoPhase dataset. Each frame is a maximum projection across the z-stack of 3D planes. Two cells can be seen undergoing mitosis over the course of the sequence.
The first step of DTW is to compute a cost matrix. For a single feature \( f \), a test signal \( t(i) \) of length \( I \) and a reference signal \( \mu(j) \) of length \( J \), this is an \( I \times J \) matrix with elements chosen as:

\[
C(i, j)_f = \frac{(t(i) - \mu(j))^2}{\sigma^2(j)}
\]

where \( \sigma^2 \) is the variance of the training signals. For multiple features, the cost matrix is computed as:

\[
C = \sum_f C_f
\]

Dynamic programming is then used to find the minimum cost path across the matrix which passes through every point in the test signal, but can start and end anywhere on the reference. An example of aligning the sequence of Figure 3 is shown in Figure 4. The time complexity of the DTW alignment is \( O(IJn) \), where \( I \) and \( J \) are the lengths of the signals and \( n \) is the number of allowed moves in the alignment. In practice \( n \) is at least an order of magnitude smaller than \( I \) or \( J \).

The features used for the temporal signals consist of simple pixel intensity measures within the detected cells, specifically, max, mean and variance are used. These have distinctive patterns during certain phases of the cell cycle e.g. the dip then increase in maximum intensity during anaphase seen in Figure 1. Simple measures of shape are also used: bounding box area (as it is more robust than number of segmented pixels), ratio of minor and major axis lengths and compactness (the ratio of area to perimeter\(^2\)). Gradients of all temporal signal features are also computed at two different scales, with the smaller scale giving a measure of local gradient and the larger scale giving a measure of the longer term evolution of the signal. Different combinations of the features are tested in Section VII.

V. EXTENSION TO HIDDEN MARKOV MODELS

Dynamic Time Warping can be thought of as a special case of an HMM, with each point along the reference signal representing a hidden state and transition probabilities restricted to prevent backward moves in time. With this in mind we extend our approach to an HMM, using the same features.

This is done by dividing each phase into a number of shorter sub-phases and computing the mean and standard deviation for each feature within these. Emission probabilities can then be given as:

\[
P = \exp \left( -\frac{(t(i) - \mu(k))^2}{2\sigma^2(k)} \right)
\]

where, as in (3), \( t(i) \) represents the value of the feature signal at point \( i \), but \( \mu(k) \) and \( \sigma(k) \) represent the mean and standard deviation of sub-phase \( k \) respectively.

This approach has a time complexity of \( O(IK^2) \) where \( K \) is the total number of sub-phases. As this is significantly smaller than the length of the reference signal in DTW, the computational cost of the HMM is significantly lower. Although, as shown in VII both are fast enough to be suitable for high throughput analysis.

The advantage of the HMM over the DTW is that with an HMM a choice of sequences can be included, whereas DTW encourages alignment to a single guiding sequence of (vector) values. So, for example, cell death or other morphological classes such as those in [4], can also be included with an HMM as a choice at any time point.

VI. IMPLEMENTATION DETAILS

A. Segmentation

Before segmentation, each image sequence is individually normalised such that all pixel intensities lie in the range \([0, 1]\). The feature vector \( x \) is a pixel intensity histogram computed...
Fig. 5: Maximum intensity signal (blue) and its gradient (green). Despite signal noise, the gradient still shows distinctive characteristics during certain phases e.g. slow decrease over prometaphase or sharp change from positive to negative at anaphase–telophase boundary.

For the pairwise term in the graph cut cost function, an 8-connected pixel neighbourhood is used to ensure smooth segmentation boundaries. The image intensity gradient is computed by filtering with the derivative of a Gaussian rotated to 4 evenly spaced orientations at $0^\circ$, $90^\circ$ and $45^\circ$ degrees from horizontal to correspond to the directions of connectivity in the neighborhood.

B. Dynamic Time Warping

Using the mitotic tracks collected as described in Section II, the temporal signals listed in Section IV are computed. All signals are then normalised to lie in the range $[0, 1]$. Gradients are computed by filtering with the derivative of a $7 \times 7$, $\sigma = 1$ Gaussian rotated to 4 evenly spaced orientations at $0^\circ$, $90^\circ$ and $\pm 45^\circ$ degrees from horizontal to correspond to the directions of connectivity in the neighborhood.

For testing, mean signals and associated standard deviations are calculated from the training data as described in Section IV. Phase durations are resampled to their mean duration in the training data, either by sub-sampling in the case of long phases or by bicubic interpolation for shorter phases.

C. Hidden Markov Model

All phases apart from interphase and anaphase are divided into two equally sized sub-phases. Anaphase is divided into four, due to the large variation of the features within this phase in the training data, as seen in Figure IV. Interphase is left as a single phase as there is little variation. Transition probabilities are learned from the training sequences.

VII. RESULTS AND EVALUATION

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.

A number of combinations of features are tested, with results presented in Table IV. Classification accuracy is also measured for the cases where transition boundaries are allowed to be within $\pm 1$ and 2 frames of their true positions. It can be seen from the confusion matrices shown in Table IV that this 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 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 $0.83$ to $0.91$ using the best feature combination.

As a baseline for comparison, frames are classified independently using a RBF-SVM with the same features used to obtain
to be suitable for high throughput analysis with the MATLAB DTW, although both labelling approaches are fast enough to take advantage of the full temporal signal. Using DTW, the approach we have proposed equips feature based cell-cycle phase recognition with an additional dynamic time warping layer so that image information across time ranges, 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 an independent frame classification approach, and to successfully label cell phases on image sequences acquired under different conditions to the training data. The same features used for DTW are also shown to work within a Hidden Markov Model framework, giving similar results.

We also test the DTW approach, without retraining, on a sample of the data used by Held et al. [3], with qualitative results shown in Figure 4. The results are comparable to those on our own data, despite the significant difference in image acquisition conditions (10× microscope objective as opposed to 60× in the training data).

The HMM method described in Section V is tested only with the final set of features used for DTW, with performance comparable between the two methods. Results are shown in Table I. The HMM approach is approximately 10× faster than DTW, although both labelling approaches are fast enough to be suitable for high throughput analysis with the MATLAB implementations used here taking on average 7ms (HMM) and 51ms (DTW) per track to label on a 3GHz processor (not including data input/output, segmentation and tracking steps).

VIII. CONCLUSIONS

The approach we have proposed equips feature based cell-cycle phase recognition with an additional dynamic time warping layer so that image information across time ranges, 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 an independent frame classification approach, and to successfully label cell phases on image sequences acquired under different conditions to the training data. The same features used for DTW are also shown to work within a Hidden Markov Model framework, giving similar results.

Acknowledgments. This work was supported by EU ERC grant VisRec no. 228180. We are grateful for comments from Michael Isard and from the anonymous reviewers.

REFERENCES

Fig. 7: Sample detected and annotated tracks. Frames subsampled to show whole of mitosis. (a) Tracks from MitoPhase data. (b) Tracks from the data of Held et al. [3]. Phases are labelled as: Grey: Interphase, Green: Prophase, Blue: Prometaphase, Magenta: Metaphase, Yellow: Anaphase, Cyan: Telophase.

TABLE III: 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. The best performing (11) set of features, consisting of the maximum pixel intensity, area and compactness with their gradients, significantly outperforms the baseline (1). The fact that the maximum (4) is the best performing pixel intensity measure is likely due to the variance (5) and mean (6) over-smoothing the signals, losing the distinctive patterns which distinguish the phases. Despite performing poorly as a single feature (3), the intensity gradient gives a significant boost in combination with the raw signal – from (2) to (4). The HMM (11) performs similarly to DTW using the same features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Exact</th>
<th>±1 frame</th>
<th>±2 frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Baseline (RBFSVM)</td>
<td>0.60</td>
<td>0.66</td>
<td>0.70</td>
</tr>
<tr>
<td>(2) Max Intensity</td>
<td>0.61</td>
<td>0.71</td>
<td>0.76</td>
</tr>
<tr>
<td>(3) Max Intensity gradient</td>
<td>0.44</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td>(4) Max Int w/gradants</td>
<td>0.76</td>
<td>0.84</td>
<td>0.87</td>
</tr>
<tr>
<td>(5) Intensity variance w/gradient</td>
<td>0.70</td>
<td>0.78</td>
<td>0.80</td>
</tr>
<tr>
<td>(6) Mean Int w/gradient</td>
<td>0.74</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td>(7) Area w/gradient</td>
<td>0.66</td>
<td>0.73</td>
<td>0.77</td>
</tr>
<tr>
<td>(8) Compactness w/gradient</td>
<td>0.38</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td>(9) Max Int+area+axis ratio w/gradients</td>
<td>0.82</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>(10) Max Int+area+compactness w/grad</td>
<td>0.83</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>(11) HMM with features of (10)</td>
<td>0.84</td>
<td>0.91</td>
<td>0.92</td>
</tr>
</tbody>
</table>


