Event Detection by Feature Unpredictability in Phase-Contrast Videos of Cell Cultures

Melih Kandemir[1],   Jose C. Rubio[1],   Ute Schmidt[2],   Christian Wojek[3],
Johannes Welbl[1],   Björn Ommer[1],   Fred A. Hamprecht[1]

1 Heidelberg University HCI/IWR, Heidelberg, Germany
2 Carl Zeiss Microscopy GmbH, Jena, Germany
3 Carl Zeiss AG, Oberkochen, Germany

Abstract. In this work we propose a novel framework for generic event monitoring in live cell culture videos, built on the assumption that unpredictable observations should correspond to biological events. We use a small set of event-free data to train a multioutput multikernel Gaussian process model that operates as an event predictor by performing autoregression on a bank of heterogeneous features extracted from consecutive frames of a video sequence. We show that the prediction error of this model can be used as a probability measure of the presence of relevant events, that can enable users to perform further analysis or monitoring of large-scale non-annotated data. We validate our approach in two phase-contrast sequence data sets containing mitosis and apoptosis events: a new private dataset of human bone cancer (osteosarcoma) cells and a benchmark dataset of stem cells.

Keywords: Event detection, mitosis, apoptosis, cell cultures, phase-contrast imaging

1   Introduction

Major activities in a cell's life cycle can be monitored from changes in its shape and appearance. Hence, visual monitoring of cell cultures has become a major trend in biomedical research (see [2] for a comprehensive review). With the development of advanced 3D (x, y, and z) and 4D (x, y, z and time) microscopy techniques such as lightsheet, dedicated screening and live cell imaging systems, the amount of recorded data has concomitantly increased to an extent beyond the feasible level for a human to perform handcraft analysis. Thus, there is growing necessity in life sciences for generic analysis tools that reliably find and trace changes of cells and organisms in large data sets.

As a framework for specialized analysis tools, we introduce a generic event detection method. We define an event as *a spatially and temporally local visual change that has a semantic meaning in biological processes.* Our goal is to provide a measure of the probability of such events occurring in a specific spatial and temporal location of a sequence of images, not only in terms of a localised detection but as a dense probability map for each timestep. We propose using a

multitask multikernel Gaussian process model [10] in the autoregression setting in a similar fashion to [8], that predicts a feature set of a region in the current frame from the feature set of the same region of the previous frame. Following the assumption that a large prediction error indicates a high confidence about the presence of an event, we construct event probability maps for each frame of a video sequence given those prediction errors.

Automation of event detection has been successfully applied to applications such as surveillance [1] and activity analysis [5]. Specifically for cell culture videos, this problem has first been studied by Yang et al. [11], who introduced a unified method for segmentation and tracking of live cell cultures that incorporates the temporal context information using level sets. The method is also able to detect mitosis based on the detected cell trajectories. Hand et al. [4] provide a benchmarking study on five live cell tracking methods applied to phase-contrast images, and report comparable tracking performance to the more invasive fluorescence imaging. Huh et al. [6] propose a mitosis detection method for the same setup, which depends on a heuristic-based hypothesis generation (selection of important image regions), and a time-smoothed conditional random field for classification of hypotheses as mitosis and other. In another study, Huh et al. [7] propose a method for apoptosis detection which depends on a more specialized hypothesis generation method and a support vector classifier.

All aforementioned methods i) are specialized for a single task, ii) require an application-dependent hypothesis generation mechanism, and iii) require supervision by ground-truth labels. The method we propose is an attempt for building a generic cell culture event detector that does not require event annotations and does not rely on hypothesis generation heuristics. Instead, it densely processes image features and outputs a dense map indicating probabilities of an event taking place.

2 Event detection from negative examples

We propose a novel event detection framework where a predictor is trained on local patches of a small set of event-free frames. For the remaining frames of the sequence, our model produces an event probability map based on the assumption that the predictor will make a larger prediction error during any type of event. Figure 1 illustrates the workflow of the proposed framework.

2.1 Feature Extraction

Given a sequence of T frames, we divide each frame in a single-scaled grid of P distinct local patches, from which we extract appearance features. We can summarize our features by stacking all patch features in a data set $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]$, where $N = P \times T$. Previous studies [9] have directed their efforts towards engineering powerful features to solve a specific application. Instead, we build a bank of generic appearance features, and learn which features best fit the event-free training data distribution. We extract five types of features: Difference of
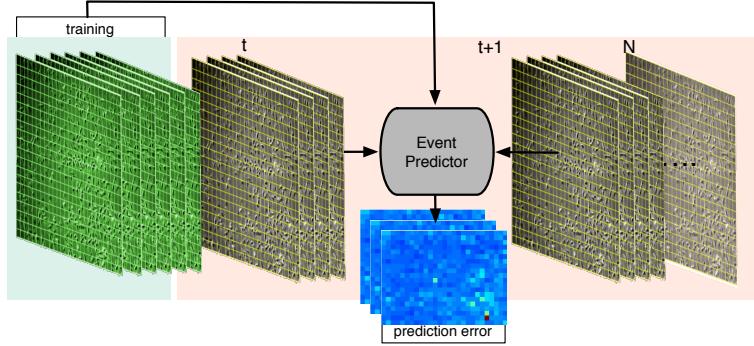
Fig. 1: The operation workflow of the event detection framework. The event predictor predicts the feature vector in a patch of frame $t + 1$ given the feature vector in the corresponding patch at frame t. It is trained on a few frames of a given query sequence where no events occur. For the remaining frames, the predictor outputs a score map of prediction errors in each patch. The prediction output can either be used as is for monitoring, or be thresholded for detection of specific events. This figure is best viewed in color.

local brightness histograms between consecutive frames (D-Hist), scale-invariant feature transform (SIFT), difference of SIFT descriptors between consecutive frames (D-SIFT), local binary patterns (LBP), and histograms of oriented gradients (HOG).

2.2 Multioutput Gaussian process autoregression

As the event detector, we propose to use autoregression (i.e. learning a model that takes a set of covariates as input and predicts the same set of covariates as output) for predicting the feature vector of an image region at the current frame from the feature vector of the same region from the previous frame. To capture non-linear dependencies across features of consecutive frames and intercorrelations of feature vectors, we propose using the spike-and-slab multitask multikernel Gaussian process model of Titsias et al. [10] in the autoregression setting, and refer to it as multioutput Gaussian process autoregression (MOGP). Let $\mathbf{x}_{(t,p)}$ be the feature representation of patch p at frame t, and $x^q_{(t,p)}$ be its q-th dimension. The generative process of the model is as below

$$x^q_{(t+1,p)} \sim \mathcal{N}(x^q_{(t,p)}|f_q(\mathbf{x}_{(t,p)}), \sigma^2_q), \qquad \forall t, p, q$$

$$f_q(\mathbf{x}_{(t,p)}) = \sum_{m=1}^{M} w_{qm}\phi_m\left(\mathbf{x}_{(t,p)}\right), \qquad \forall p, q$$

$$w_{qm} \sim \pi\mathcal{N}(w_{qm}|0, \sigma^2_w) + (1-\pi)\delta_0(w_{qm}), \qquad \forall t, q$$

$$\phi_m\left(\mathbf{x}^m_{(t,p)}\right) \sim \mathcal{GP}\left(\mu_m\left(\mathbf{x}^m_{(t,p)}\right), k_m\left(\mathbf{x}^m_{(t,p)}, \mathbf{x}^m_{(t',p')}\right)\right). \qquad \forall t, p, m$$

Here, $\phi_m(.)$ is a latent variable for kernel m, and it is governed by a Gaussian process prior. By $\mathbf{x}_{(t,p)}^m$, we denote the subset of features used by kernel m. Above, w_{qm} is the weight of kernel m for output dimension q, $f_q(.)$ is the noiseless regression output for dimension q, and $x_{(t+1,p)}^q$ is the noisy regression output of patch p at time point $t+1$. The term $\delta_0(.)$ is the Dirac delta function centered at zero, and π is a constant determining the mixing coefficients of the spike at zero and the normal distribution. This formulation serves as a spike-and-slab prior on the kernel weights, inducing sparsity, which allows effective selection of useful kernels and elimination of noisy ones. The mixing constant π and kernel weights w_{qm}, are learned by mean-field variational approximation. Noise covariances σ_q^2 and σ_w^2, the mean function $\mu_m(.)$ and the kernel hyperparameters are learned by Type II Maximum Likelihood. See [10] for further details.

We draw the hypothesis that the features of a patch become less predictable from its state at the previous time point under existence of an event. Hence, we use the prediction error $||\hat{\mathbf{x}}_{(t+1,p)} - \mathbf{x}_{(t+1,p)}||_2$ as a score proportional to the probability of the existence of an event at patch p at time $t+1$, where $\hat{\mathbf{x}}_{(t+1,p)}$ and $\mathbf{x}_{(t+1,p)}$ are the predicted and observed feature vectors, respectively.

We perform effective feature selection by assigning each feature set to a kernel and learning their contribution to the prediction process (i.e kernel weights w_{qm}) from training data. By plugging in a multiple heterogeneous set of features, the spike-and-slab prior on kernel weights will assign high weights to the features encoding richer information. This leverages the feature sets that best represent the variability of the event-free training data. Additionally, the multitask nature of the model allows incorporating intercorrelations of the features in the next frame, and its multikernel nature allows automatic feature selection.

3   Experiments

Devising our method as a generic event detector, we aim to appeal to multiple applications with different event types of interest. Due to practical reasons, we evaluate the correspondence of the event probability maps produced by our method to two types of biologically important events: apoptosis [7] and mitosis [6], even though other types of events such as imaging artifacts, atrophy, hypertrophy, and necrosis could also have been analyzed. We validate our model on two different types of cells: human osteosarcoma cells, and stem cells. To illustrate the event detection performance of our method, we provide precision-recall curve plots, and report the area under precision-recall curve (AUC) as a threshold-independent performance measure.

We compare the proposed MOGP method to OC-SVM (one-class SVM), which is used as a standard tool for novelty detection in non-biological applications [3]. We train the OC-SVM treating feature vectors of all training patches as i.i.d samples. This is meant to illustrate the effect of autoregression (i.e. making predictions across consecutive frames) on prediction performance. The static grid of local regions is built using 50 by 50 pixels sized patches for both data sets, so that the resolution of the frame determines the amount of image patches

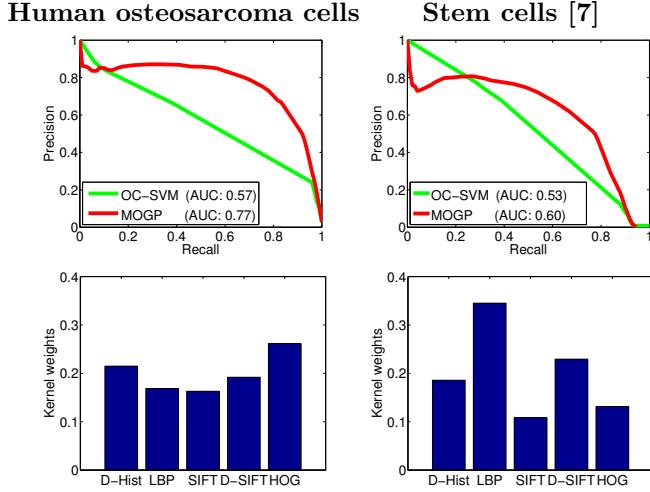**Human osteosarcoma cells   Stem cells [7]**

Fig. 2: Precision-recall curves of event predictors in comparison (top) and contribution of feature types to event prediction learned by MOGP (bottom) averaged over all human osteosarcoma cell (left) and stem cell (right) sequencs.

being sampled. We use the squared exponential function kernel as the covariance function $k(\cdot, \cdot)$ for all feature sets in the bank. We analyze each sequence independently and report the performance averaged over all sequences of a given cell type. We train both predictors on the first five frames of a given sequence, where there are no mitosis and apoptosis events.

We count a ground-truth event as correctly detected if the prediction score of a patch located less than 30 pixels away lies above the decision threshold, and its temporal location differs in three or less frames. In case of multiple detections within the temporal tolerance interval, we count only one of them as a true positive and the rest as false positives. This evaluation procedure is identical to that of [7].

3.1   Phase-contrast data

Human osteosarcoma (U2-OS) cell data set: This data set consists of human osteosarcoma cells on which apoptosis is induced by phototoxicity. When a sample is illuminated, the energy brought into the sample not only excites fluorophores but can also be transferred to other molecules. Subsequently, conversion of the energy leads to generation of radicals, for example reactive oxygen species such as hydrogen peroxide. These molecules are harmful for the cell as they damage proteins and DNA leading to misfunctional proteins or mutations. When cells are damaged to an extent where cellular defense mechanisms can no longer compensate for, cells trigger apoptosis (programmed cell death).

To observe apoptosis induced by phototoxicity, human osteosarcoma cells were seeded onto coverslips that contained an integrated grid to relocate cells.
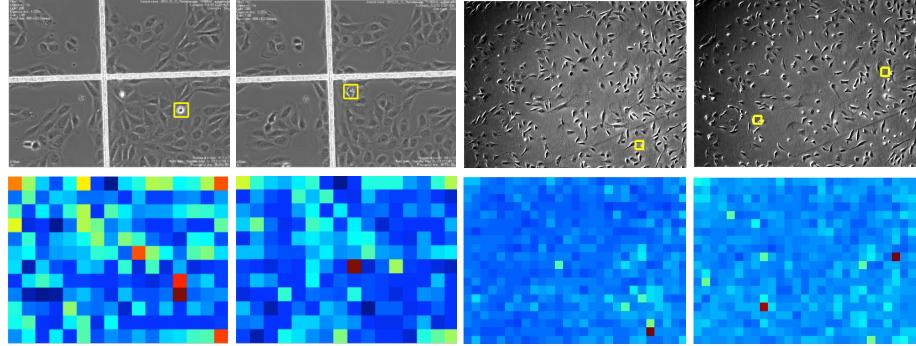
Fig. 3: Examples of frames (top) and prediction maps (bottom). First two columns show frames from the acquired sequence of human osteosarcoma cells. Last two columns show examples from the stem cell data set of [7]. Yellow bounding boxes in the top row show thresholded detections, while the color coding of the bottom row expresses the prediction error (increases as the color goes from blue through yellow to red).

Then, 5 to 10 cells in a field of view were irradiated with high intensity laser light. Phase contrast images of the living cells were acquired every 10 minutes for 24 hours with a $10\times$ phase contrast objective and under normal growth conditions. Following this procedure, 8 sequences of 134 frames were recorded with a resolution of 800×600 pixels. A sequence consists of 113.4 mitosis events and 4.4 apoptosis events on average.

Stem cell data set: For additional validation we choose a public data set introduced in [7] that consists of three sequences of 540 frames where apoptosis was induced in C2C12 myoblastic stem cells by applying Mitomycin C. The sequences contain 384.6 apoptosis events and 9.0 mitosis events on average.

3.2 Results and discussion

In the precision-recall curves of Figure 2, MOGP provides a reasonable performance in two distinct applications, 0.77 AUC for human osteosarcoma cells and 0.60 AUC for stem cells, which are significantly above the chance level of 0.006 AUC and 0.001 AUC, respectively. MOGP clearly outperforms OC-SVM as well on both data sets due to its autoregressive nature (i.e. ability to relate changes in consecutive frames).

The bar plots in Figure 2 show the weights MOGP assigns to feature sets, indicating the significance of their contribution to prediction. For the human osteosarcoma cell data set, in which 92.3% of the events are cell mitosis, HOG features achieve the largest weight, though with a marginal difference. This shows that MOGP discovers HOG as the most informative feature set, which is consistent with the hand-crafted feature set of the mitosis detector in [6]. For the

stem cell data set, LBP and D-SIFT features gain the largest weight, which is also parallel to the design of the apoptosis predictor proposed in [7].

Besides serving as a monitoring tool as they are, the scoremaps produced by our model can be used for event detection by determining a decision threshold. We choose the threshold from a small set of annotated validation data (6th to 10th frame of each sequence) that provides the greatest F_1-score (harmonic mean of precision and recall). We reach average 0.52 precision and 0.69 recall over all human osteosarcoma sequences. For the stem cell sequences, our predictor gives 0.68 precision and 0.61 recall, which is significantly lower than the heavily preprocessed and fully-supervised predictor of Huh et al. [7] that reaches 0.94 prediction and 0.89 recall. While supervising the predictor by events brings a large performance gain, it makes the predictor dependent on a particular application. We hereby demonstrate that a reasonable, though not perfect, performance can also be reached by much less modeling assumptions. Figure 3 shows examples of scoremap outputs together with thresholded detections of mitosis and apoptosis events in both data sets.

The fact that mitosis and apoptosis feature very similar morphological changes at the first stage of the event (rounding up, bright boundary) poses one of the main difficulties when performing apoptosis detection in the presence of mitosis. The observation of the prediction patterns produced by the model suggests that those responses encode discriminative information when tackling the problem of event categorization. Hence, supervised learning could be performed by these event *signatures*. Mitosis tends to manifest in two differentiated phases (round compact shaping, and separation) producing two distinguishable sharp peaks in the prediction error spanned across 5 to 6 frames. Apoptosis events however, show a single wide response (See Figure 4).

4   Conclusions

We propose a generic event detection framework that consists of an autoregression model that is able to characterize events in terms of their unpredictability given a set of *normal* data as training examples. We demonstrate that the model predictions correspond to biologically relevant cell events, as well as showing remarkable improvements over a well known approach (one-class SVM) in two different data sets. Additionally, our model maps each video frame into a probability map which sheds important information about the ongoing events, enabling further research on supervised approaches.

References

1. Borislav Antić and Björn Ommer. Video parsing for Abnormality Detection. In *ICCV*, 2011.
2. P. Antony, C. Trefois, A. Stojanovic, A.S. Baumuratov, and K. Kozak. Light microscopy applications in systems biology: opportunities and challenges. *Cell Communication and Signaling*, 2013.
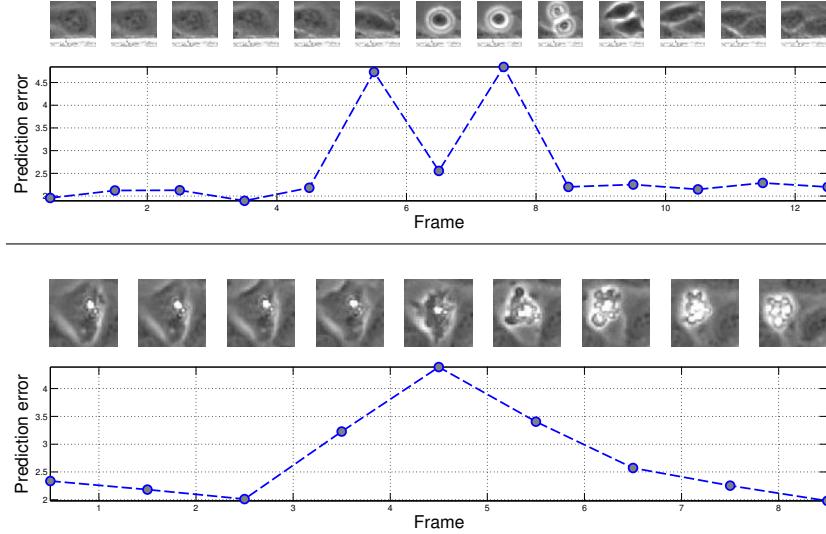
Fig. 4: Examples of time course of prediction error during mitosis (top), and apoptosis (bottom) on human osteosarcoma cells. Mitosis events show oscilating prediction errors due to two cell phases (blobbing and dividing). Apoptosis produces single wide error peaks.

3. A.B. Gardner, A.M. Krieger, G. Vachtsevanos, and B. Litt. One-class novelty detection for seizure analysis from intracranial EEG. *Journal of Machine Learning Research*, 2006.
4. A.J. Hand, T. Sun, D.C. Barber, D.R. Hose, and S. MacNeil. Automated tracking of migrating cells in phase-contrast video microscopy sequences using image registration. *Journal of Microscopy*, 2009.
5. M. Hoai, and F. de la Torre. Max-margin early event detectors. *CVPR*, 2012.
6. S. Huh, D.F. Ker, R. Bise, M. Chen, and T. Kanade. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *Medical Imaging, IEEE Transactions on*, 2011.
7. S. Huh, H. Su, and T. Kanade. Apoptosis detection for adherent cell populations in time-lapse phase-contrast microscopy images. In N. Ayach, H. Delingette, P. Golland, and K. Mori, editors, *MICCAI*. Springer, 2012.
8. J. Ma and S. Perkins. Online novelty detection on temporal sequences. In *SIGKDD*, 2003.
9. Y. Song, W. Cai, S. Huh, M. Chen, T. Kanade, Y. Zhou, and D. Feng. Discriminative data transform for image feature extraction and classification. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, editors, *MICCAI*. 2013.
10. M.K. Titsias and M. Lázaro-Gredilla. Spike and slab variational inference for multi-task and multiple kernel learning. In *NIPS*, 2011.
11. F. Yang, M.A. Mackey, F. Ianzini, G. Gallardo, and M. Sonka. Cell segmentation, tracking, and mitosis detection using temporal context. In J. Duncan and G. Gerig, editors, *MICCAI*. Springer, 2005.