# Modeling and Tracking of Heterogeneous Cell Populations via Open Multi-Agent Systems

A. Tramaloni[1], A. Testa[1], S. Avnet[2], S. Massari[2], G. Di Pompo[3], N. Baldini[2,3,#] and G. Notarstefano[1,#]

***Abstract*—Understanding cellular dynamics represents a critical challenge in biomedical research. Optical microscopy remains a key technique for observing live-cell behaviors *in vitro*. This paper introduces an enhanced cell-tracking algorithm designed to address dynamic changes in cell populations, including mitosis, migration, and cell-cell interactions, even within complex co-culture models. The proposed method involves three main steps: 1) modeling the movements and interactions of different cell types in co-culture experiments via tailored open multi-agent systems; 2) identifying parameters via real data for a multi-agent, multi-culture framework; 3) embedding the model within an Extended Kalman Filter, to predict the dynamics of heterogeneous cell populations across video frames. To validate the approach, we used a novel dataset involving the interplay between tumor and normal cells, namely osteosarcoma and mesenchymal stromal cells, respectively. This dataset offers a challenging and clinically relevant framework to track cell proliferation and study how cancer cells evolve and interact with stromal cells within their surroundings. Performance metrics demonstrated the effectiveness of the algorithm over state-of-the-art methodologies, highlighting its ability to track heterogeneous cell types, capture their interactions, and generate the estimated cell lineage tree.**

***Index Terms*—Multi-Agent Systems, Filtering, Tracking, Osteosarcoma, Mesenchymal Stromal Cells, Systems Biology**

## I. Introduction

In recent years, the study of cells has gained interest in the control community. Recent works include [1], [2]. The main idea behind these works is to leverage system theoretical approaches to model and control cell populations. These contributions span macro areas like cybergenetics [3] and multicellular systems modeling and simulation [4]. Relevant to our work is the application of multi-agent system paradigms to capture the emergent behaviors observed in heterogeneous cell populations, where individual cells can be modeled as interacting agents. Experimental data for these tasks are typically obtained from optical microscopy videos. Thus, to understand cell behaviors and their interactions, a preliminary task is to extract cell trajectories over time automatically. Biological cell tracking microscopy videos are thus becoming a fundamental task within broad application domains. Among the others, the tracking of tumor interaction with normal stromal cells *in vitro* is crucial to study tumor aggressiveness and drug resistance [5]. Among different techniques, optical microscopy is widely used to retrieve cellular images across different time instants [6].

### A. Related Works

We structure the literature review into three main parts. First, we examine existing cell tracking algorithms. Second, we explore multi-agent approaches in systems and synthetic biology, including how biological principles inspire the design of multi-agent robotic systems. Finally, we review system and control theory methods for analyzing and regulating cell population dynamics.

Multi-Cell Tracking (MCT) is a well-established problem in the literature, with applications ranging from cybergenetics to drug discovery [7], [8]. Given the significant interest in MCT, various datasets and challenges have been developed, such as the Cell Tracking Challenge [9], which serves as a benchmark for MCT algorithms on *in vitro* monoculture experiments. This study focuses on a novel co-culture dataset representing a new challenging scenario compared to other publicly available datasets.

Recent MCT algorithms [10], [11] combine convolutional neural networks (CNN) with recurrent units for segmentation and tracking in microscopy images. Deep learning-based methods include the use of CNNs with Viterbi algorithms [12], Faster R-CNN for multi-feature fusion [13], graph neural networks for temporal modeling [14], and single CNNs for joint segmentation and tracking [15]. Alternative approaches employ Markov random field models with spatial particle filters [16] or variational optical flow techniques [17]. Unlike these black-box methods, our approach combines predictive filtering with tailored multi-agent models that exploit cell-to-cell interactions to improve motion prediction. Moreover, to the best of the authors' knowledge, existing methods focus primarily on monoculture experiments, whereas co-culture settings accounting for cell population heterogeneity remain relatively uncommon.

Next, works connecting multi-agent system and systems biology are reviewed. Multi-agent modeling has been explored as a framework for designing emergent collective behaviors in synthetic biology [18], such as studying synchronized bacterial populations [19], and coordinating multi-robot systems inspired by chemotaxis principles [20]. For open multi-agent systems, the theoretical framework in [21] provides a basis for studying systems with time-varying state spaces. In our work, we build on similar concepts to model the dynamics of the

This work was supported in part by the Italian Ministry of Foreign Affairs and International Cooperation, grant number BR22GR01, and by AIRC under the IG 2018—project ID 21403—to N.B. ("Altered lipid metabolism as a stress reaction to acid tumor microenvironment and a driver of metastasis in osteosarcoma").

# Equal contribution.

[1] Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy. {andrea.tramaloni, a.testa, giuseppe.notarstefano}@unibo.it.

[2] Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy. {sofia.avnet3, nicola.baldini5}@unibo.it, stefania.massari96@gmail.com.

[3] Biomedical Science, Technologies, and Nanobiotechnology Lab, IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy. {nicola.baldini, gemma.dipompo}@ior.it.

considered "open" cell system, which is subject to population changes over time.

As for approaches that analyze cell behaviors and model biological interactions, several works often leverage concepts from systems theory, identification, and control. For example, [22] models tumor cell behavior using ordinary, partial, or stochastic differential equations, but does not address single-cell dynamics. Control-oriented approaches are prevalent in related literature, including feedback regulation of cellular populations [23], [24], control of populations endowed with synthetic toggle switches [1], [2], eco-evolutionary models of cancer under therapy [25], and adaptive control of nonlinear cancer-immune dynamics [26]. Overall, these works primarily emphasize control strategies. In contrast, our work focuses on model design, identification, and prediction of cell behaviors.

### B. Contributions

The contributions of this paper are as follows. We propose a novel algorithm, called Heterogeneous Open Multi-Cell EKF-based Tracking (HEOM-EKF), for tracking cells in time-lapse *in vitro* multi-culture experiments, with a specific focus on co-culture settings. HEOM-EKF addresses the following key limitations of existing MCT approaches.

**Modeling heterogeneous cell populations.** Most tracking algorithms are designed for monoculture settings and do not explicitly model interactions among different cell types. *Our contribution:* we model heterogeneous cell populations as an open multi-agent system [27] and propose a class of parametric interaction models that explicitly capture both inter- and intra-type dynamics (e.g., chemotaxis-like behaviors [28]) among cells such as MSC and osteosarcoma. The framework naturally handles switching dynamics in the population over time (e.g., due to mitosis). We focus on these two cell types because, based on several studies—both our own and those of others over the past decades—their interaction has been shown to play a crucial role in inducing aggressiveness and drug resistance in tumor cells [29]–[32].

**Interpretability of motion prediction models.** Many state-of-the-art methods rely on deep learning predictors, which often act as black boxes and provide limited insight into the underlying biological mechanisms. *Our contribution:* we design a structured open multi-agent dynamical model whose parameters are identified from real cell trajectories via tailored system identification procedures.

**Handling of population change events.** Many MCT methods do not robustly distinguish mitosis events from cells entering or exiting the field of view. *Our contribution:* we formulate an Extended Kalman Filter tailored to open multi-agent systems, equipped with switching label sets and dimensionality adaptation. This framework explicitly accounts for mitosis, entry, and exit events, enabling the reconstruction of the cell lineage over time.

In addition, we construct a new annotated dataset from raw time-lapse *in vitro* co-culture experiments involving normal mesenchymal stromal cells and transformed osteosarcoma cells. Unlike most publicly available benchmarks, which focus on monocultures, this dataset provides a challenging and clinically relevant testbed for tracking heterogeneous interacting cell populations and for validating the proposed modeling and identification framework.

A conference version of this work appeared in [33], where we proposed a monoculture cell-tracking algorithm based on a standard multi-agent formulation and evaluated it on a public dataset. That preliminary study assumed a constant number of cells and did not account for mitosis or cells entering and exiting the camera field of view. In contrast, this journal version addresses a more challenging co-culture scenario and leverages a private, unique dataset. Methodologically, we introduce an open multi-agent framework that handles multiple agent (cell) types and explicitly models dynamic population changes. Furthermore, the proposed algorithm reconstructs the cell lineage tree, enabling a deeper analysis of cell interactions and temporal evolution.

### C. Paper Organization

The paper is organized as follows. In Section II, we introduce the cell-tracking problem and outline the essential components of multi-object tracking algorithms. Section III formalizes the open multi-agent system, generalized for multi-culture scenarios, and presents the model identification strategy. In Section IV, we describe the proposed HEOM-EKF algorithm. Section V describes the dataset and its use for training and testing, analyzes the identified multi-cell model parameters, and presents numerical experiments that assess the performance of the proposed algorithm.

## II. Multi-Cell Tracking Problem Formulation

In this work, we consider a cell-tracking problem in a biological scenario in which the goal is to track multiple cells of different types in time-lapse sequences. Throughout the paper, we consider video frames in which the number of cells changes over time. These variations may result from biological processes such as mitosis, and/or from cells entering or exiting the microscope camera field of view (FOV). In this section, we formalize the fundamental concepts behind cell-tracking algorithms. We will delve into more details in Section III, in which we model the cell population as an open multi-agent system and propose suitable formulations for their dynamics. These dynamical models will be a key component of the proposed cell-tracking strategy in Section IV.

The proposed framework follows the widely adopted Track-by-Detection paradigm, in which object detections are first obtained independently on each frame and then linked over time by a dedicated tracking module. In this setting, the main building blocks are: $i)$ object detection, $ii)$ motion prediction, and $iii)$ data association. A schematic of this procedure is depicted in Figure 1. Object detection refers to the identification and localization of cells within images or video frames. Modern object detection methods often employ trained neural networks specifically for this task [34]. Motion prediction aims to estimate future positions of detected cells by combining their past trajectories with current detection measurements. This step is essential for ensuring

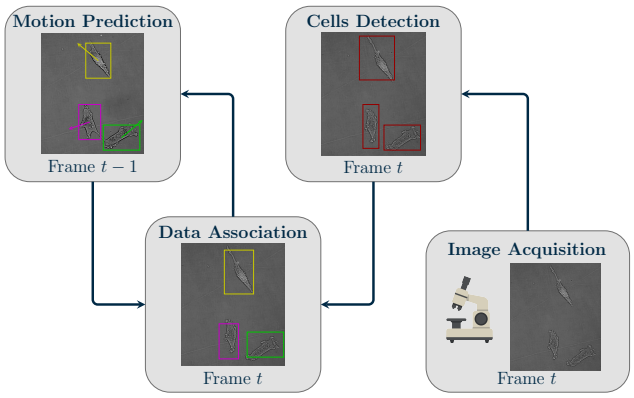



Figure 1: Scheme of a generic multi-cell tracking algorithm

track continuity and handling gaps or occlusions in the detections. Common techniques include Kalman filters, particle filters, and, most commonly, deep learning-based predictors. However, neural network predictors act as black boxes, thus limiting interpretability, and require large datasets, which are rare in this field. Instead, our approach models multi-cell behaviors explicitly and employs a closed-form solution for the identification of parameters, ensuring reliability and data efficiency. Data association involves linking detected cells across consecutive frames to preserve their identities over time. This process requires matching detections from different frames and assigning them to the correct cell track. Motion models and similarity metrics are frequently used to assess the likelihood of accurate associations.

## III. Heterogeneous Cell Populations as Open Multi-Agent Systems

In this section, we first formalize cell dynamics as an open multi-agent system. Then, we introduce a tailored class of parametric multi-agent terms to describe the interactions between heterogeneous cell populations within a general multi-culture scenario. These dynamical terms constitute the multi-agent model that will be then adopted in the proposed cell-tracking algorithm. The model design draws inspiration from a well-established line of research on bio-inspired multi-agent systems [28]. In particular, we extend to an open multi-agent framework neighboring terms motivated by swarm-like approaches, as in [35] and [36], that describe both positional and morphological cell state behaviors. The latter terms are weighted by a set of parameters to be identified through data collected from time-lapse experiments.

### A. Open Multi-Cell System Formalization

We model mitosis as an event where a parent cell splits into two new daughter cells. In this process, the parent cell is removed from the system and two new cells appear in its place. The latter cells retain the same type (i.e. agent type) as their parent. Moreover, to maintain distinct identities of cells over time, each cell is assigned a unique identifier label $i$. We consider a discrete-time framework, where $t \in \mathcal{T} \subseteq \mathbb{N}$ denotes the time index over an arbitrary horizon $\mathcal{T}$ (infinite or finite). In the context of a multi-culture experiment, we restrict attention to a finite time horizon, namely $\mathcal{T} = \{1, \dots, T\}$. Here, $t$ will be the discrete-time variable representing the frame at which the system is observed, and with $T \in \mathbb{N} \setminus \{\infty\}$ denoting the final time instant at which the experiment concludes. At any time $t \in \mathcal{T}$, we denote by $N^t \in \mathbb{N}$ the number of cells present within the microscope FOV. Additionally, we define the subset of time instants $\mathcal{T}_s \subset \mathcal{T}$ corresponding to switching times at which events related to cell population changes occur, i.e. when the number of cells within the observed frame changes, indicating a switch in the system dynamics. We denote the number of cells entering and exiting the heterogeneous cell population within the microscope FOV at time $t$ by $C^t_{\texttt{IN}} \in \mathbb{N}$ and $C^t_{\texttt{OUT}} \in \mathbb{N}$, respectively. Let $L^t \subset \mathbb{N}$ be the set of labels identifying all the cells present in the scene at time $t$. Consequently, we introduce its subsets $L^t_p \subset L^t$ and $L^t_d \subset L^t \times L^t$ containing the labels of parent and daughter cells, respectively. Based on that, mitoses occurrences can be described through time-dependent sets of label pairs, identifying parent cells splitting into daughter cells at the current time. Formally, mitoses occurring at frame $t$ are represented by

$$\mathcal{M}^t := \left\{(i_p, \boldsymbol{j}_d) : i_p \in L^t_p \subset \mathbb{N}, \boldsymbol{j}_d \in L^t_d \subset \mathbb{N}^2\right\}. \quad (1)$$

Instead, new cell entries into the camera field, at time $t$, are captured by the set $L^t_{\texttt{IN}} \subset \mathbb{N}$, such that

$$|\mathcal{M}^t| + |L^t_{\texttt{IN}}| = C^t_{\texttt{IN}}. \quad (2)$$

Similarly, cell exiting from the camera FOV at time $t$ are described by the label set $L^t_{\texttt{OUT}} \subset \mathbb{N}$, verifying

$$|L^t_{\texttt{OUT}}| = C^t_{\texttt{OUT}}. \quad (3)$$

Note that $L^t_p \cap L^t_d \cap L^t_{\texttt{OUT}} \cap L^t_{\texttt{IN}} = \emptyset$, but for different time instants these sets may have non-null intersections. Moreover, at least one of these sets is non-empty at times $t \in \mathcal{T}_s$. The total number of cells at any time $t$, previously introduced as $N^t$, evolves according to

$$\begin{aligned} N^{t+1} &= N^t + C^t_{\texttt{IN}} - C^t_{\texttt{OUT}} \quad \forall t \in \mathcal{T}_s \\ N^{t+1} &= N^t \quad \forall t \notin \mathcal{T}_s. \end{aligned}$$

Moreover, the dynamics of the label set $L^t$ can then be expressed as

$$L^{t+1} = \left(L^t \cup \bar{L}^t_d \cup L^t_{\texttt{IN}}\right) \setminus \left(L^t_p \cup L^t_{\texttt{OUT}}\right),$$

where $\bar{L}^t_d := \bigcup_{(a,b)\in L^t_d} \{a, b\}$ is the set of the individual elements from daughter cells label pairs.

Building upon the concepts introduced above, the following subsection presents the proposed open multi-agent parametric model.

### B. Open Multi-Agent Parametric Model

We now introduce the open multi-cell system that models the cell dynamics for the tracking purpose. In this framework, each generic agent $i$ represents a cell. The state of agent $i$ at time $t$ and its neighbor set are denoted by $\mathbf{x}^t_i \in \mathbb{R}^{n_x}$ and $\mathcal{N}^t_i \subset L^t$, respectively. To support a general multi-culture scenario in which different cell populations are interacting, we consider

$\Gamma \in \mathbb{N}$ distinct agent types, each representing a different cell type. Thus, each cell type will be associated a parameter $\boldsymbol{\theta}_\gamma \in \mathbb{R}^{n_\theta}$, $\gamma \in \{1, \dots, \Gamma\}$. We denote

$$\boldsymbol{\theta} := [\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_\Gamma^\top]^\top \in \mathbb{R}^{n_\theta \Gamma}, \tag{4}$$

the stack of the parameter vectors for each cell type. The constant parameters $\boldsymbol{\theta}$ will be learned from experimental data.

Let $L := \bigcup_{t=1}^{T} L^t$ denote the set of all agent labels across time. Then, each agent $i \in L$ is associated with a parameter vector $\boldsymbol{\theta}_{\gamma_i} \in \mathbb{R}^{n_\theta}$, determined by its type $\gamma_i \in \{1, \dots, \Gamma\}$. Based on that, agent dynamics are modeled according to the discrete-time formulation

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + f_i\left(\mathbf{x}_i^t, \{\mathbf{x}_j^t\}_{j \in \mathcal{N}_i^t}, \boldsymbol{\theta}_{\gamma_i}, t\right), \ \forall i \in L^{t+1} \cap L^t, \tag{5}$$

where the function $f_i$ captures the dynamics of agent $i$ based on its current state, the states of its neighboring agents, and its type-based parameters $\boldsymbol{\theta}_{\gamma_i}$. This formulation accounts for both cell state and type interactions, allowing the model to represent heterogeneous behaviors within a multi-culture experimental setting.

We propose coordinates of a Minimum Enclosing Rectangle (MER) as agent state within the proposed MCT algorithm. This representation of cell morphology is defined as

$$\mathbf{x}_{MER,i} := \left[\mathbf{p}_i^\top, s_i, r_i, \phi_i, \mathbf{v}_i^\top, v_{s,i}, v_{r,i}, \omega_i\right]^\top, \tag{6}$$

where $\mathbf{p}_i, \mathbf{v}_i \in \mathbb{R}^2$ denote the center position and velocity of the cell, $s_i, r_i \in \mathbb{R}$ its area and aspect ratio, and $v_{s,i}, v_{r,i} \in \mathbb{R}$ their respective rates of change. The orientation angle $\phi_i \in [0, 2\pi]$ specifies the rotation of the MER relative to a fixed reference axis, and $\omega_i \in \mathbb{R}$ is the angular velocity. This state representation is especially suitable for tracking cells with fusiform or anisotropic morphologies, as it captures orientation and aspect more faithfully than the standard bounding box format. Moreover, it leads to more informative outputs at the end of the tracking pipeline, supporting better discrimination between different types based on shape dynamics. However, standard performance metrics for tracking algorithms are typically designed for axis-aligned bounding boxes. As a result, evaluating tracking accuracy on the state format in (6) requires custom metrics that are not always readily available or comparable with prior work. To facilitate benchmarking and compatibility with existing evaluation frameworks, our algorithm also supports a simplified state representation based on axis-aligned bounding boxes. Formally,

$$\mathbf{x}_{BB,i} := \left[\mathbf{p}_i^\top, s_i, r_i, \mathbf{v}_i^\top, v_{s,i}, v_{r,i}\right]^\top, \tag{7}$$

where the state variables retain the same semantics as in (6), but omit orientation and angular velocity. This compact version is especially useful for performance evaluation and comparison with other methods, while still capturing essential information about cell shape and motion.

Based on these two representations, the neighbors of cell $i$ at time $t$ in (5) can be defined as

$$\mathcal{N}_i^t := \left\{j \neq i : \left\|\mathbf{p}_i^t - \mathbf{p}_j^t\right\| \leq R_s, \ \{i, j\} \subseteq L^t\right\},$$

where $R_s$ is a sensing radius that defines the region of space in which other cells are perceived. Here, we consider $R_s$ as a priori fixed quantity (not learnable).

Building on the generic formulation of the open multi-agent model in (5), we now specify the individual terms governing the dynamics of each state component. For better readability, we first define subgroups of the whole state

$$\boldsymbol{\pi}_i := [\mathbf{p}_i^{t\top}, \mathbf{v}_i^{t\top}]^\top \in \mathbb{R}^4, \ \ \boldsymbol{\sigma}_i := [s_i^t, v_{s,i}^t]^\top \in \mathbb{R}^2$$
$$\boldsymbol{\rho}_i := [r_i^t, v_{r,i}^t]^\top \in \mathbb{R}^2, \ \ \boldsymbol{\eta}_i := [\phi_i^t, \omega_i^t]^\top \in \mathbb{R}^2.$$

Then, for the MER state format, the dynamics of cell $i$ of type $\gamma_i$ are given by

$$\mathbf{p}_i^{t+1} = \mathbf{p}_i^t + \mathbf{v}_i^t \tag{8a}$$
$$s_i^{t+1} = s_i^t + v_{s,i}^t \tag{8b}$$
$$r_i^{t+1} = r_i^t + v_{r,i}^t \tag{8c}$$
$$\phi_i^{t+1} = \phi_i^t + \omega_i^t \tag{8d}$$
$$\mathbf{v}_i^{t+1} = \mathbf{v}_i^t + f_{i,1}\left(\boldsymbol{\pi}_i, \{\boldsymbol{\pi}_j\}_{j \in \mathcal{N}_i^t}, \boldsymbol{\theta}_{\gamma_i}\right) \tag{8e}$$
$$v_{s,i}^{t+1} = v_{s,i}^t + f_{i,2}\left(\boldsymbol{\sigma}_i, \{\boldsymbol{\sigma}_j\}_{j \in \mathcal{N}_i^t}, \boldsymbol{\theta}_{\gamma_i}\right) \tag{8f}$$
$$v_{r,i}^{t+1} = v_{r,i}^t + f_{i,3}\left(\boldsymbol{\rho}_i, \{\boldsymbol{\rho}_j\}_{j \in \mathcal{N}_i^t}, \boldsymbol{\theta}_{\gamma_i}\right) \tag{8g}$$
$$\omega_i^{t+1} = \omega_i^t + f_{i,1}\left(\boldsymbol{\eta}_i, \{\boldsymbol{\eta}_j\}_{j \in \mathcal{N}_i^t}, \boldsymbol{\theta}_{\gamma_i}\right), \tag{8h}$$

where

$$\begin{aligned} f_{i,1} := \ & \theta_{\gamma_i,1}\left(\frac{1}{|\mathcal{N}_i^t|}\sum_{j \in \mathcal{N}_i^t} \mathbf{p}_j^t - \mathbf{p}_i^t\right) \\ & + \theta_{\gamma_i,2}\left(\frac{1}{|\mathcal{N}_i^t|}\sum_{j \in \mathcal{N}_i^t} \mathbf{v}_j^t - \mathbf{v}_i^t\right) \\ & + \theta_{\gamma_i,3}\sum_{j \in \mathcal{N}_i^t} \frac{\mathbf{p}_i^t - \mathbf{p}_j^t}{\left\|\mathbf{p}_i^t - \mathbf{p}_j^t\right\|^2}, \end{aligned} \tag{9}$$

captures the tendency to cohesion, alignment, and repulsion behaviors, while the shape-related terms

$$\begin{aligned} f_{i,2} &:= \theta_{\gamma_i,4}\left(S_{\gamma_i} - s_i^t\right) \\ f_{i,3} &:= \theta_{\gamma_i,5}\left(R_{\gamma_i} - r_i\right), \end{aligned} \tag{10}$$

model the tendency to maintain a certain target area $S_{\gamma_i} \in \mathbb{R}$ and target aspect ratio $R_{\gamma_i} \in \mathbb{R}$, considering cells of type $\gamma_i$. Here, $\theta_{\gamma_i,j} \in \mathbb{R}$ with $j = 1, \dots, 5$ represent the learnable parameters associated with cells of type $\gamma_i$.

These multi-cell dynamics are inspired by engineering multi-agent models. Specifically, the terms in (9) follow swarm-based approaches and incorporate chemotaxis-inspired principles [35], [36]. In contrast, the morphological terms in (10), which govern area and aspect ratio dynamics are novel. These terms are designed based on our dataset analysis, where $S_\gamma$ and $R_\gamma$ are derived from morphological statistics (see Section V). Together, these components capture a biologically motivated tendency for each cell type to maintain its characteristic shape over time. Finally, we note that for the standard bounding box state defined in (7), the dynamics can be fully described using only equations (8a)-(8c) and (8e)-(8g).

### C. Identification of Model Parameters

We now present the proposed approach for identifying parameters within the multi-agent dynamics discussed in Section III-B. Here, we consider a generic formulation with $\Gamma$ agent populations captured by parameters $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta \Gamma}$, thus providing a flexible tool scalable to different experimental frameworks. Therefore, we propose an identification method tailored to open multi-cell systems, like those described in Section III-A.

The goal is to identify the parameters and obtain estimates that fit a training set of real cell trajectories extracted from *in vitro* co-culture experiments. Consider a dataset consisting of $T$ image frames with the relative cell segmentations and tracks. From the latter, for each frame $t$ we have access to $\mathbf{p}_i^t$, $s_i^t$, $r_i^t$ and $\phi_i^t$ for all cells $i$. From these, we compute the relative velocities $\mathbf{v}_i^t$, $v_{s,i}^t$, $v_{r,i}^t$ and $\omega_i^t$ using differentiation methods. We tested various differentiation methods commonly used in system identification from time series data, such as finite differences, smoothed finite differences, Savitzky-Golay filtering, spline interpolation, and Kalman-based filtering, which are also discussed in the context of data-driven model identification (e.g., in [37]). In the experimental tests, we adopted a Kalman-based differentiation approach, which provided smoother estimates and led to smaller residuals in the subsequent model identification phase. This approach allows us to approximate the necessary time derivatives despite the intrinsic stochasticity and noise in biological cell dynamics and aligns with established practices in the literature.

Since we defined the dynamics in equation (5) as a weighted sum of multi-agent state terms, and imposed sensing radius $R_s$ to be not learnable, we end up having a system with linear dependence in $\boldsymbol{\theta}$. Therefore, by properly redefining the system we can construct and solve a typical least squares (LS) problem. The solution to the LS problem is obtained by solving

$$\boldsymbol{\theta}^\star = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ \|\mathbf{Y} - \mathbf{D}\boldsymbol{\theta}\|^2, \tag{11}$$

where observation matrices are

$$\mathbf{Y}_i := [\mathbf{y}_i^1, \dots, \mathbf{y}_i^T]^\top \in \mathbb{R}^{n_x T}, \tag{12a}$$
$$\mathbf{Y} := [\mathbf{Y}_1^\top, \dots, \mathbf{Y}_N^\top]^\top \in \mathbb{R}^{n_x T N}. \tag{12b}$$

and design matrix

$$\mathbf{D} := [\mathbf{D}_1, \dots, \mathbf{D}_\Gamma] \in \mathbb{R}^{(n_x T N) \times \Gamma n_\theta},$$

in which each $\mathbf{D}_\gamma \in \mathbb{R}^{(n_x T N) \times n_\theta}$ represents the block corresponding to type $\gamma$ agents. The latter are defined as

$$\mathbf{D}_\gamma := [\mathbf{D}_{\gamma,1}^\top, \dots \mathbf{D}_{\gamma,N}^\top]^\top \ \ \forall \gamma \in \{1, \dots, \Gamma\},$$

where $N := |L|$ is the total number of cells in the dataset, considered here as a single time-lapse experiment.

For all the cells $i \in L$, we specify the entries of the observation matrices (12a) as

$$\mathbf{y}_i^t = \begin{cases} \left[\mathbf{x}_i^t - \mathbf{x}_i^{t-1}\right]^\top & \text{if } \ i \in L^t \cap L^{t-1} \\ \mathbf{0}_{n_x}^\top & \text{otherwise} \end{cases}.$$

Instead, the single-cell design matrices are constructed as

$$\mathbf{D}_{\gamma,i} = \begin{bmatrix} d_{i,1}^1 & d_{i,2}^1 & \dots & d_{i,n_\theta}^1 \\ \vdots & \vdots & & \vdots \\ d_{i,1}^T & d_{i,2}^T & \dots & d_{i,n_\theta}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{d}_i^1 \\ \vdots \\ \boldsymbol{d}_i^T \end{bmatrix} \in \mathbb{R}^{n_x T \times n_\theta},$$

verifying that for each $t$

$$\boldsymbol{d}_i^t = \mathbf{0}_{n_x \times n_\theta} \ \ \forall \gamma \neq \gamma_i \text{ or } i \notin L^t.$$

This condition, adding sparsity to the design matrix, ensures feasibility by limiting the data to cells present at the current frame $t$ and by selecting different weights based on type $\gamma_i$. Otherwise, the entries of cell $i \in L^t$ are functions of the considered cell and its neighbor states

$$\begin{aligned} d_{i,1}^t &= f_{i,1}\left(\mathbf{x}_i^t, \{\mathbf{x}_j^t\}_{j \in \mathcal{N}_i^t}\right) \\ &\vdots \\ d_{i,n_\theta}^t &= f_{i,n_\theta}\left(\mathbf{x}_i^t, \{\mathbf{x}_j^t\}_{j \in \mathcal{N}_i^t}\right). \end{aligned} \tag{13}$$

The solution of (11) that best fits cell trajectories extracted from a single multi-culture experiment is

$$\boldsymbol{\theta}^\star = \left(\mathbf{D}^\top \mathbf{D}\right)^{-1} \mathbf{D}^\top \mathbf{Y}. \tag{14}$$

Note that while $n_x T N \gg \Gamma n_\theta$ is necessary for the LS problem to be well-posed, invertibility of $\mathbf{D}^\top \mathbf{D}$ additionally requires full column rank. This depends on the informativeness of the collected cell trajectories. For instance, if cells exhibit minimal motion, certain dynamical terms may remain insufficiently excited, potentially causing rank deficiency despite abundant data. To verify these conditions, we analyzed the design matrix

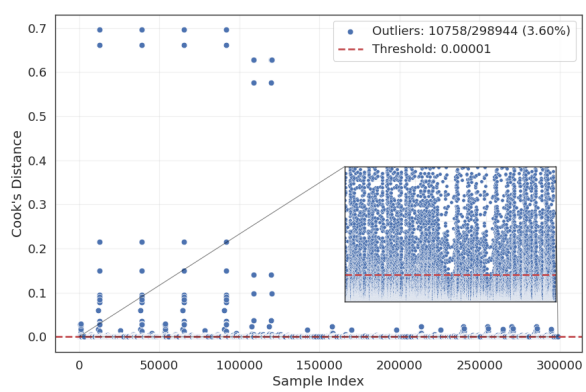



Figure 2: Cook's distance values across samples, highlighting detected outliers (blue dots) based on a threshold (dashed red).

constructed from our dataset. We confirmed that $\operatorname{rank}(\mathbf{D}) = \Gamma n_\theta$, ensuring full column rank. Specifically, we obtained a condition number of $\approx 4.76 \cdot 10^3$, a value that is generally regarded as numerically safe for least-squares computations in double-precision arithmetic. To further improve numerical stability and reliability, we applied Cook's distance metric [38] to identify and remove influential outliers, removing observations exceeding a threshold of $4/\left(n_x T N - \Gamma n_\theta\right)$. The latter is a common heuristic for outlier detection in regression analysis. This outlier removal reduced the sum of squared residuals

(SSR) from approximately 20.63 to 6.79, significantly enhancing both the fit quality and the robustness of the parameter estimates. Figure 2 illustrates the Cook's distance values across samples, highlighting the detected outliers (3.6% of total data). Since at least one cell is present in every frame ($L^t \neq \emptyset \; \forall t$), all rows of the design matrix correspond to valid observations, thus ensuring consistency of the identification procedure.

In the case of multi-experiment data, such as in our scenario, the process involves stacking observation vectors and design matrices for each data collection. Moreover, considering the scarcity of data in biological cell experiments, transformed versions of the same data sequences can be included to enhance robustness concerning invariance transformations, such as rotations, and horizontal and vertical flips. Details on the augmentation of the dataset are given in Section V on numerical experiments.

## IV. Multi Cell Tracking Algorithm

In this section, we detail our solution to the cell-tracking problem outlined in Section II. Our Heterogeneous Open Multi-Cell EKF-based Tracking (HEOM-EKF) algorithm introduces a novel motion prediction step employing a tailored Extended Kalman Filter for Open Multi-Agent systems that leverages the models described in Section III. Moreover, our algorithm includes a procedure to estimate cell population changes and reconstruct the cell lineage tree. The main steps of the algorithm are detailed in Algorithm 1.

Here, the set estimations $\hat{L}_p^t$, $\hat{L}_d^t$, $\hat{L}_{\texttt{OUT}}^t$, $\hat{L}_{\texttt{IN}}^t$ and $\hat{\mathcal{M}}^t$ refer to the previous definitions of true quantities introduced in Section III defined in (1)-(3). We also provide a graphical

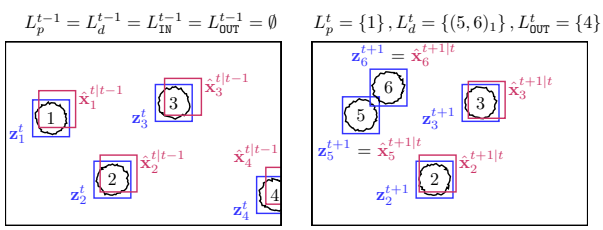



Figure 3: An illustrative example of label sets at different frames in a tracking scenario.

illustration (see Figure 3) of how HEOM-EKF behaves in the presence of a mitosis event and how these label set estimates change accordingly across frames.

In particular, the algorithm starts by initializing the state guess with detections at the first frame and setting up the EKF matrices. Then, throughout each frame, the algorithm iterates through three main steps. First, get the cell detections related to the current image. Second, it predicts the next state via the EKF prediction step leveraging the interaction dynamics learned from real data, whose identification of parameters is described in Section III-C. Third, it associates detections to predictions and updates the state estimate through the so-called EKF update step. Differently from standard filtering schemes, this step also takes into account changes in the number of cells. Specifically, when a cell population change is predicted at current frame, through MITOSIS-DETECT and

**Algorithm 1** HEOM-EKF

**Initialization**
Obtain $\mathbf{Z}^0$ from detector
Initialize EKF: $\hat{\mathbf{X}}^{0|0} = \mathbf{Z}^0$, $\mathbf{P}^{0|0}$, $\mathbf{Q}^0$, $\mathbf{R}^0$ and $\mathbf{H}^0$
Initialize mitosis set: $\hat{\mathcal{M}}^t = \emptyset \; \forall t \in \mathcal{T}$
**for** each frame $t \in \mathcal{T}$ **do**
**Object Detection**:
Obtain $\mathbf{Z}^t$ from detector
**EKF Prediction Step**:

$$\hat{\mathbf{X}}^{t|t-1} = \hat{\mathbf{X}}^{t-1|t-1} + f\left(\hat{\mathbf{X}}^{t-1|t-1}, \boldsymbol{\theta}^\star\right) \tag{15}$$

$$\mathbf{F}^{t-1} = \frac{\partial f}{\partial \mathbf{X}}\Big|_{\hat{\mathbf{X}}^{t-1|t-1}} \tag{16}$$

$$\mathbf{P}^{t|t-1} = \mathbf{F}^{t-1}\mathbf{P}^{t-1|t-1}\left(\mathbf{F}^{t-1}\right)^T + \mathbf{Q}^t \tag{17}$$

**Associate Detections to Predictions**:
Get $\mathbf{S}^t$ via (25) and $\mathbf{A}^t$ solving (26)
Reorder detections IDs in $\mathbf{Z}^t$ based on $\mathbf{A}^t$
**EKF Update Step**:

$$\mathbf{K}^t = \mathbf{P}^{t|t-1}\mathbf{H}^{t\top}(\mathbf{H}^t\mathbf{P}^{t|t-1}\mathbf{H}^{t\top} + \mathbf{R}^t)^{-1} \tag{18}$$

$$\hat{\mathbf{X}}^{t|t} = \hat{\mathbf{X}}^{t|t-1} + \mathbf{K}^t(\mathbf{Z}^t - \mathbf{H}^t\hat{\mathbf{X}}^{t|t-1}) \tag{19}$$

$$\mathbf{P}^{t|t} = (\mathbf{I} - \mathbf{K}^t\mathbf{H}^t)\mathbf{P}^{t|t-1} \tag{20}$$

**Search for Population Change Event**:
$\hat{\mathcal{M}}^t, \hat{L}_p^t, \hat{L}_d^t = \textsc{mitosis-detect}(\mathbf{Z}^t, \hat{\mathbf{X}}^{t|t}, \mathbf{S}^t, d_{\max}, IoU_{\min})$
$\hat{L}_{\texttt{IN}}^t, \hat{L}_{\texttt{OUT}}^t = \textsc{in-out-detect}(\mathbf{Z}^t, \hat{\mathbf{X}}^{t|t}, \mathbf{S}^t, t_{\max}, d_{\max}, IoU_{\min}, \epsilon)$

**if** $\hat{L}_p^t \cup \hat{L}_d^t \cup \hat{L}_{\texttt{OUT}}^t \cup \hat{L}_{\texttt{IN}}^t \neq \emptyset$ **then**
$\hat{\mathcal{T}}_s \cup \{t\}$
Update $\hat{\mathbf{X}}^{t|t}$, $\mathbf{P}^{t|t}$, $\mathbf{Q}^t$, $\mathbf{R}^t$ and $\mathbf{H}^t$ via Algorithm 4
**end if**
**end for**
**return** Cell tracks $\hat{\mathbf{X}}^{t|t} \; \forall t$ and cell lineage tree $\hat{\mathcal{M}}^t \; \forall t$

IN-OUT-DETECT procedures (see Algorithms 2-3), the EKF update variables are adapted to account for dimensionality variations (see Algorithm 4). We now detail all the steps in the algorithm.

### A. Object Detection

The object detection phase serves as the initial step in our algorithm, with the primary goal of localizing cells in consecutive frames of the input video. All quantities and computations in our algorithm are directly based on the pixel coordinates of the detected cells. This ensures consistency with the spatial information provided during the object detection step. In particular, the measurement of cell $i$ obtained via object detection at frame $t$ is an oriented bounding box. Formally

$$\mathbf{z}_i := \left[\mathbf{p}_i^\top, s_i, r_i, \phi_i\right]^\top. \tag{21}$$

Here, $\mathbf{p}_i$ denotes the center coordinates of the bounding box for the $i^{\text{th}}$ cell. The terms $s_i, r_i \in \mathbb{R}$ represent the area and aspect ratio of the detected region, respectively, while $\phi_i \in [0, 2\pi]$ indicates the orientation angle.

It is worth noting that the object detection is performed independently on each frame, producing frame-specific cell detections and measurements, denoted as $\mathbf{z}_i^t$. The identity labels $i$ are assigned arbitrarily by the detector, regardless of the method used. To ensure consistent tracking across frames prediction and data association steps (see Sections IV-C and IV-B) are needed to reorder and match cell identities over time.

### B. Motion Prediction

At each time $t$, the object detection phase yields a collection of measurement vectors $\mathbf{z}_i^t \in \mathbb{R}^{n_z}$ (cf. Section IV-A). Based on that, we can now detail the motion prediction step.

Let $\mathbf{x}_i^t \in \mathbb{R}^{n_x}$ denote the true, unknown state of cell $i$ at time $t$, represented in the format of equation (6). Accordingly, the detection process at each frame $t$ can be modeled by the following measurement equation

$$\mathbf{z}_i^t = \mathbf{I}_{n_z \times n_x} \mathbf{x}_i^t + \mathbf{w}_z, \qquad \forall\, i \in L^t, \tag{22}$$

in which $\mathbf{I}_{n_z \times n_x}$ is the observation matrix, which is assumed to be constant and identical across all measurements. The term $\mathbf{w}_z \in \mathbb{R}^{n_z}$ denotes zero-mean Gaussian noise with covariance $\mathbf{R}^t$.

Let us introduce $\hat{\mathbf{x}}_i^{t|t-1} \in \mathbb{R}^{n_x}$ as the predicted state of $i^{\text{th}}$ cell at time $t$, obtained after a prediction step in which the measurement $\mathbf{z}^{t-1}$ is known. Instead, $\hat{\mathbf{x}}_i^{t|t} \in \mathbb{R}^{n_x}$ represents the updated estimate through the measurement $\mathbf{z}_i^t$ obtained after an estimation update step. To retrieve the aforementioned predictions, our tracking algorithm adopts an Extended Kalman Filter [39] incorporating as nonlinear dynamics the open multi-agent model introduced in Subsection III-B.

We first define

$$\mathbf{X}^t := [{\mathbf{x}_1^t}^\top, \dots,\ {\mathbf{x}_{N^t}^t}^\top]^\top \in \mathbb{R}^{N^t n_x}$$
$$\mathbf{Z}^t := [{\mathbf{z}_1^t}^\top, \dots,\ {\mathbf{z}_{N^t}^t}^\top]^\top \in \mathbb{R}^{N^t n_z},$$

as the stack of cell population states and measurements. Similarly, we denote by $\hat{\mathbf{X}}^{t|t-1}$ and $\hat{\mathbf{X}}^{t|t}$ the stacks of predicted and updated state estimates, respectively. According to expression (22), the stack measurement vector is given by

$$\mathbf{Z}^t = \mathbf{H}^t \mathbf{X}^t + \mathbf{W}_z^t, \tag{23}$$

where $\mathbf{H}^t \in \mathbb{R}^{(N^t n_z) \times (N^t n_x)}$ and $\mathbf{W}_z^t \in \mathbb{R}^{N^t n_z}$ are the stacks $\forall i \in L^t$ of observation matrices and measurement noise vectors, respectively.

The multi-agent dynamics used to predict the $i^{\text{th}}$ cell state belong to the class of models introduced in Section III-B, i.e.

$$\hat{\mathbf{X}}^{t|t-1} = \hat{\mathbf{X}}^{t-1|t-1} + f\left(\hat{\mathbf{X}}^{t-1|t-1}, \boldsymbol{\theta}^\star, t\right) + \mathbf{W}_x^t,$$

which is the predicted version of equation (5) written in compact form. In particular, $f : \mathbb{R}^{n_x N^t} \times \mathbb{R}^{n_\theta \Gamma} \times \mathbb{R} \mapsto \mathbb{R}^{n_x N^t}$ suitably stacks the dynamics in (8) for all the agents present at current time, and $\mathbf{W}_x^t \in \mathbb{R}^{n_x N^t}$ stacks the noise vector $\mathbf{w}_x$ for all the cells. The noise $\mathbf{w}_x \in \mathbb{R}^{n_x}$ is a Gaussian random variable with zero mean and covariance $\mathbf{Q}$. The optimal parameters in (15) are learned by solving (14) with properly constructed $\mathbf{D}$ and $\mathbf{Y}$ (see Section III-C) on real data. The step in (17) is the update of the so-called estimation error covariance matrix $\mathbf{P}^t \in \mathbb{R}^{n_x N^t \times n_x N^t}$. The steps in (18)–(20) instead perform the update steps of the EKF. Here, $\mathbf{K}^t \in \mathbb{R}^{n_x N^t \times n_z N^t}$ is the Kalman gain matrix.

Thanks to this motion prediction design, our tracker can operate on a unified state representation for the heterogeneous cell populations taking advantage of the sparse structure of the dynamics. Indeed, the motion prediction of each cell is performed considering interactions between neighboring cells rather than independent dynamical entities.

### C. Data Association

The data association step bridges the gap between detected cells across consecutive frames, preserving their identities over time. Indeed, given two consecutive time frames $t-1$ and $t$, the $i^{\text{th}}$ cell may be detected and associated with different labels. At the generic time $t$, the EKF predictions in (15) provide the estimate of a novel bounding box $\hat{B}_i$ for cell $i$. This predicted box is centered in $\hat{\mathbf{p}}_i^{t|t-1}$ and has area and aspect ratio $\hat{s}_i^{t|t-1}$ and $\hat{r}_i^{t|t-1}$, respectively. Considering the cell state as in (6), also the predicted box orientation $\hat{\phi}_i^{t|t-1}$ is present.

Similarly, let $B_j$ denote the $j^{\text{th}}$ bounding box found by the detection step. Let us introduce the similarity matrix $\mathbf{S}^t \in \mathbb{R}^{N^t \times N^t}$ at image frame $t$, with single entry

$$\left[\mathbf{S}^t\right]_{ij} = \mathcal{S}\left(\hat{B}_i, B_j\right), \tag{24}$$

where $\mathcal{S}$ is a similarity function computed between the predicted bounding box of $i^{\text{th}}$ cell and the detected one of $j^{\text{th}}$ cell at time $t$. In this work, we use the so-called Intersection Over Union (IoU) metric [40], so that

$$\left[\mathbf{S}^t\right]_{ij} = \frac{|\hat{B}_i \cap B_j|}{|\hat{B}_i \cup B_j|}. \tag{25}$$

Given $\mathbf{S}^t$, the data association problem regards finding an optimal assignment matrix $\mathbf{A}^t$ at each frame of the video that maximizes the total similarity across all associations. This means solving at each time $t$ the following linear-assignment problem

$$\begin{aligned} &\max \sum_{i,j} \left[\mathbf{S}^t\right]_{ij} a_{ij} \\ &\text{s.t.} \ \sum_j a_{ij} = 1\ \forall i, \ \ \sum_i a_{ij} = 1\ \forall j, \ \ a_{ij} \in \{0,1\}, \end{aligned} \tag{26}$$

where each entry $a_{ij} = 1$ in $\mathbf{A}^t$ indicates that $j^{\text{th}}$ detection is assigned to $i^{\text{th}}$ prediction, and $a_{ij} = 0$ otherwise. This problem can be solved in polynomial time leveraging, e.g., the Hungarian algorithm [41]. To reject associations with a poor similarity index, we enforced a threshold for the IoU function in (25), denoted as $IoU_{\min}$.

We now define the set $\mathcal{U}$ containing unmatched elements at time $t$, which can be further divided into unmatched

predictions $\mathcal{U}_{\hat{x}}^t$ and unmatched detections $\mathcal{U}_z^t$. Specifically, the set of unmatches is defined as

$$\mathcal{U}^t := \left\{(i,j) : \left[\mathbf{S}^t\right]_{ij} \leq IoU_{\min}\right\}.$$

Then, we derive the subsets for unmatched predictions and unmatched detections

$$\mathcal{U}_{\hat{x}}^t := \bigcup_{(i,j)\in\mathcal{U}^t} \{i\}, \ \mathcal{U}_z^t := \bigcup_{(i,j)\in\mathcal{U}^t} \{j\}. \tag{27}$$

To provide better context, unmatched predictions and detections arise within a tracking algorithm where associations between predicted states and detections may not always be successful. In such cases, elements that fail to match are categorized into these sets to facilitate further analysis or handling in the tracking process.

### *D. Cell Population Changes and Cell Lineage Tree*

In this subsection, we define three routines used in our cell tracking strategy HEOM-EKF. We first introduce a method for estimating cell mitosis events (Algorithm 2). Then, we describe how to detect the entry and exit of cells within the microscope FOV (Algorithm 3). Lastly, we present the procedure to update the dimensions of the EKF matrices in response to the changes in the cell population (Algorithm 4).

We begin by introducing some quantities used within Algorithm 2 and 3. Let $N_z^t \in \mathbb{N}$ be the number of detected cells and $N_{\hat{x}}^t \in \mathbb{N}$ the number of predicted cell states. Also, let $\bar{\mathcal{U}}^t \subset \mathcal{U}^t$ be the subset of unmatched predicted and detected cells whose bounding boxes have no intersection and whose distance exceeds the mitosis threshold $d_{\max}$. This distance is determined based on real observed behaviors within the co-culture experiments, and it represents the maximum center distance at which a mitosis event is usually detected. Formally,

$$N_z^t := \frac{\dim\left(\mathbf{Z}^t\right)}{n_z}, \quad N_{\hat{x}}^t := \frac{\dim\left(\hat{\mathbf{X}}^{t|t-1}\right)}{n_x} \tag{28}$$

$$\bar{\mathcal{U}}^t := \left\{(i,j) \in \mathcal{U}^t : \left[\mathbf{S}^t\right]_{ij} = 0, \left\|\hat{\mathbf{p}}_i^t - \mathbf{p}_j^t\right\| \geq d_{\max}\right\} \tag{29}$$

$$\bar{\mathcal{U}}_{\hat{x}}^t := \bigcup_{(i,j)\in\bar{\mathcal{U}}_{\hat{x}}^t} \{i\}, \quad \bar{\mathcal{U}}_z^t := \bigcup_{(i,j)\in\bar{\mathcal{U}}_z^t} \{j\}. \tag{30}$$

The ability to accurately determine mitosis events is important for extracting relevant information about the evolution of heterogeneous cell populations over time. A possible approach consists of constructing the cell lineage tree for each cell. To this end, in the following, we present Algorithm 2.

The proposed method estimates mitosis occurrences at a given frame $t$ by computing the set estimates $\hat{L}_p^t$, $\hat{L}_d^t$, and thus $\hat{\mathcal{M}}^t$. In particular, the algorithm leverages the sets in (27) and (30), together with the quantities in (28), to distinguish new detections due to mitosis, cells entering the camera FOV or unmatched detections. After tracking is completed across the entire video sequence, the lineage of each cell is constructed using the parent-child relationships stored in the sets $\hat{\mathcal{M}}^t$ for all $t \in \mathcal{T}$. Note that $\hat{\mathcal{M}}^t = \emptyset$ at frames where no mitosis events are detected.

We now present Algorithm 3 to estimate the occurrence of cells entering or exiting from the camera FOV.

**Algorithm 2** MITOSIS-DETECT : Get set estimate $\hat{\mathcal{M}}^t$

**Require:** $\mathbf{Z}^t$, $\hat{\mathbf{X}}^{t|t}$, $\mathbf{S}^t$, $d_{\max}$, $IoU_{\min}$

**if** $\mathcal{U}_z^t \neq \emptyset$ **and** $N_z^t > \left(N_{\hat{x}}^t - |\mathcal{U}_{\hat{x}}^t| + |\bar{\mathcal{U}}_{\hat{x}}^t|\right)$ **then**

  **for** $i \in \mathcal{U}_z^t \setminus \bar{\mathcal{U}}_z^t$ **do**

    Find $j = \underset{j}{\operatorname{argmin}} \left\|\mathbf{z}_i^t - \mathbf{z}_j^t\right\|$

    New prediction labels: from $(i,j)$ to $(j_{d_1}, j_{d_2})$

    Update daughters set $\hat{L}_d^t = \hat{L}_d^{t-1} \cup \{j_{d_1}, j_{d_2}\}$

    Update parent set $\hat{L}_p^t = \hat{L}_p^{t-1} \cup \{j\}$

  **end for**

**end if**

Update $\hat{\mathcal{M}}^t$ based on definition (1)

**return** $\hat{\mathcal{M}}^t$, $\hat{L}_p^t$, $\hat{L}_d^t$

**Algorithm 3** IN-OUT-DETECT : Get set estimates $\hat{L}_{\mathtt{IN}}^t$, $\hat{L}_{\mathtt{OUT}}^t$

**Require:** $\mathbf{Z}^t$, $\hat{\mathbf{X}}^{t|t}$, $\mathbf{S}^t$, $t_{\max}$, $d_{\max}$, $IoU_{\min}$, $\epsilon$

**if** $\mathcal{U}_z^t \neq \emptyset$ **and** $N_z^t > \left(N_{\hat{x}}^t - |\mathcal{U}_{\hat{x}}^t| + |\bar{\mathcal{U}}_{\hat{x}}^t|\right)$ **then**

  **for** $i \in \bar{\mathcal{U}}_z^t : \min(p_{x,i}^t, \bar{x} - p_{x,i}^t, p_{y,i}^t, \bar{y} - p_{y,i}^t) \leq \epsilon$ **do**

    New prediction label $i \to j_{\mathtt{IN}}$

    Update entry set $\hat{L}_{\mathtt{IN}}^t = \hat{L}_{\mathtt{IN}}^{t-1} \cup \{j_{\mathtt{IN}}\}$

  **end for**

**end if**

**for** $i \in \left\{\mathcal{U}_{\hat{x}}^{t-t_{\max}} \cap \ldots \cap \mathcal{U}_{\hat{x}}^t\right\}$ **do**

  **if** $\min(p_{x,i}^t, \bar{x} - p_{x,i}^t, p_{y,i}^t, \bar{y} - p_{y,i}^t) \leq \epsilon$ **then**

    Remove prediction label $i$

    Update exit set $\hat{L}_{\mathtt{OUT}}^t = \hat{L}_{\mathtt{OUT}}^{t-1} \cup \{i\}$

  **end if**

**end for**

**return** $\hat{L}_{\mathtt{IN}}^t$, $\hat{L}_{\mathtt{OUT}}^t$

Analogously to Algorithm 2, this method begins by checking for unmatched detections at the current frame $t$, using the sets defined in (27) and (30), along with the values in (28). It then examines the set $\bar{\mathcal{U}}_z^t$ to identify detected cells that are $\epsilon$-close to the image boundaries $(\bar{x}, \bar{y})$, i.e., cells that have recently entered the field of view. To detect cells that are leaving the microscope FOV, the algorithm instead looks for predictions that remained unmatched over the past $t_{\max}$ time steps. If these predictions are also $\epsilon$-close to the image boundaries, they are considered as outcoming cells.

Algorithm 4 takes care of adapting the dimensions of EKF matrices. Specifically, it reorganizes the terms $\hat{\mathbf{X}}^{t|t-1}$, $\hat{\mathbf{P}}^{t|t-1}$, $\mathbf{H}^t$, $\mathbf{R}^t$, and $\mathbf{Q}^t$ to ensure feasible tracking. Especially, let us define the functions used for including new daughter cells into the tracking procedure as

$$g_d\left(\boldsymbol{a}, \boldsymbol{b}\right) := \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} \in \mathbb{R}^{(N+1)n_x},$$

in which $\boldsymbol{a} \in \mathbb{R}^{Nn_x}, \boldsymbol{b} \in \mathbb{R}^{n_x}$. Also, let

$$G_d\left(\boldsymbol{A}, \boldsymbol{b}\right) := \begin{bmatrix} \boldsymbol{A} & \boldsymbol{b}_{1:n_x-1} \\ \boldsymbol{b}_{1:n_x-1}^\top & b \end{bmatrix} \in \mathbb{R}^{(N+1)n_x \times (N+1)n_x},$$

in which $\boldsymbol{A} \in \mathbb{R}^{Nn_x \times Nn_x}$, $\boldsymbol{b} \in \mathbb{R}^{n_x}$, $\boldsymbol{b}_{1:n_x-1} \in \mathbb{R}^{n_x-1}$ and $b \in \mathbb{R}$. Instead, the functions used to remove a parent cell within HEOM-EKF are

$$g_p : \mathbb{R}^{Nn_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^{(N-1)n_x}$$
$$(\boldsymbol{a}, \boldsymbol{a}_i) \mapsto [\boldsymbol{a}_1^\top, \ldots, \boldsymbol{a}_{i-1}^\top, \boldsymbol{a}_{i+1}^\top, \ldots, \boldsymbol{a}_N^\top]^\top,$$

in which $\boldsymbol{a}_i$ belongs to $\boldsymbol{a} = [\boldsymbol{a}_1^\top, \ldots, \boldsymbol{a}_N^\top]^\top$, and

$$G_p : \mathbb{R}^{Nn_x \times Nn_x} \times \{1, \ldots, N\} \to \mathbb{R}^{(N-1)n_x \times (N-1)n_x}$$
$$(\boldsymbol{A}, i) \mapsto \boldsymbol{A}_{\setminus i},$$

with $\boldsymbol{A}_{\setminus i}$ being the matrix $\boldsymbol{A}$ with the $i^{\text{th}}$ row and column removed.

**Algorithm 4** Update EKF dimensionality

**Require:** $\hat{L}_p^t$, $\hat{L}_d^t$, $\hat{L}_{\texttt{OUT}}^t$, $\hat{L}_{\texttt{IN}}^t$, $\hat{\mathbf{X}}^{t|t}$, $\mathbf{P}^{t|t}$, $\mathbf{Q}^t$, $\mathbf{R}^t$ and $\mathbf{H}^t$

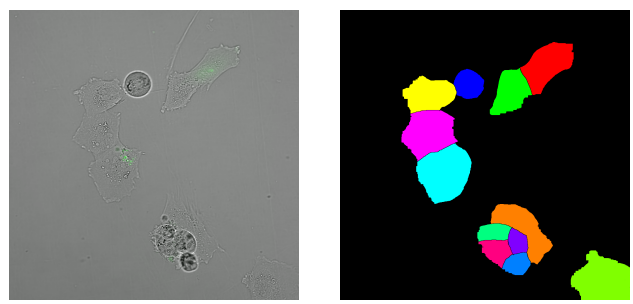
**for** $i_p \in \hat{L}_p^t$ and $\boldsymbol{j}_d \in \hat{L}_d^t$ **do**
$\hat{\mathbf{X}}^{t|t} = g_p\left(\hat{\mathbf{X}}^{t|t}, \hat{\mathbf{x}}_{i_p}^{t|t}\right)$
$\hat{\mathbf{X}}^{t|t} = g_d\left(\hat{\mathbf{X}}^{t|t}, \mathbf{z}_j^t\right) \ \forall j \in \boldsymbol{j}_d$
$\mathbf{P}^{t|t} = G_p\left(\mathbf{P}^{t|t}, i_p\right)$
$\mathbf{P}^{t|t} = G_d\left(\mathbf{P}^{t|t}, [\mathbf{0}_{n_x-1}^\top, 1]^\top\right) \ \forall j \in \boldsymbol{j}_d$
**end for**
**for** $i_{\texttt{OUT}} \in \hat{L}_{\texttt{OUT}}^t$ **do**
$\hat{\mathbf{X}}^{t|t} = g_p\left(\hat{\mathbf{X}}^{t|t}, \hat{\mathbf{x}}_{i_{\texttt{OUT}}}^{t|t}\right)$
$\mathbf{P}^{t|t} = G_p\left(\mathbf{P}^{t|t}, i_{\texttt{OUT}}\right)$
**end for**
**for** $i_{\texttt{IN}} \in \hat{L}_{\texttt{IN}}^t$ **do**
$\hat{\mathbf{X}}^{t|t} = g_d\left(\hat{\mathbf{X}}^{t|t}, \mathbf{z}_{i_{\texttt{IN}}}^t\right)$
$\mathbf{P}^{t|t} = G_d\left(\mathbf{P}^{t|t}, [\mathbf{0}_{n_x-1}^\top, 1]^\top\right)$
**end for**
Redefine $\mathbf{Q}^t$, $\mathbf{R}^t$ and $\mathbf{H}^t$ according to $N_z^t$ and $N_{\hat{x}}^t$

## V. Numerical Experiments

In this section, we first introduce the dataset used in our simulations. Then, we present the results of the proposed cell-tracking algorithm, analyze the optimal multi-agent model parameters that confirm cell type discrimination, and discuss the extracted cell lineage tree.

### A. Dataset

The dataset used in this study was derived from raw time-lapse image sequences captured using a confocal microscope (A1R MP microscope, Nikon) at IRCCS Istituto Ortopedico Rizzoli. The human 143b line were obtained from the American Type Cell Culture Collection (ATCC) and were cultured in Minimum Essential Medium Eagle Alpha modification (Sigma) with 10% fetal bovine serum (FBS, Sigma) and 1% penicillin/streptomycin (P/S) (Euroclone) and used for co-culture experiments after trypsinization and semi-confluence. Human adipose-derived MSC (ATCC) were maintained in complete alpha-MEM with 10% fetal bovine serum (FBS, Sigma) and 1% P/S and used for co-culture experiments after trypsinization and semi-confluence. Cells were incubated at 37 °C in 5% $CO_2$. GFP-transfected MSC cells, obtained as previously described [31], were used in co-culture experiments. For co-cultures, 143b cells and GFP+ MSC were mixed 1:3 and seeded at low density in complete alpha-MEM in petri dishes for time-lapse imaging. The transfection with GFP allowed us to clearly distinguish MSC from 143b cells and, consequently, accurately assign the measured features between the two cell types. The set of experiments carried out consists of 4 videos with 172 frames each with images captured every 4 minutes (25x immersion objective, NA 1.1, RI 1.333). Initial processing of the raw images was performed by using NIS Elements AR 5.40.01 software (Nikon), employed on all the images to perform semantic segmentation, labeling, and classification by cell type. We performed these tasks by combining manual adjustments with automated image processing tools. In particular, we first manually segmented representative cell regions and then used the software to automatically propagate and refine the contours across subsequent frames by tuning intensity thresholds and edge-detection parameters. When cells overlapped or formed dense clusters, segmentation was instead carried out entirely by manual annotation. Since the process could be subject to human and/or software automation errors, we refer to these annotations as the gold standard, rather than ground truths. Following segmentation, the software extracted



(a) (b)

Figure 4: Comparison between the original microscope image (a) and its final segmented version (b).

around 70 features per segmented object, including morphological and positional properties, allowing for detailed temporal analysis of each cell. These features, after filtering only the ones relevant to our purposes, provided a comprehensive characterization of the cells' dynamics and their interactions over time. To refine the dataset further, additional image processing techniques were applied, focusing on denoising and fixing any labeling or segmentation errors. The latter post-processing corrections ensured enhanced data quality. Before presenting the final results, we report summary statistics of key cell features extracted from the dataset (Figure 5). These metrics reveal clear morphological differences between MSC and 143b cells and motivate the design of the proposed open multi-agent dynamics. Although perimeter, circularity, and elongation are not explicitly modeled as state variables, their morphological information is implicitly encoded in the bounding box representation. In particular, the aspect ratio $r_i$ provides a proxy for circularity and elongation, capturing type-specific shape characteristics. The target aspect ratio $R_{\gamma_i}$ for

each cell type is derived from the mean circularity observed in the training data, thereby embedding morphological priors into the tracking framework. As expected, MSC display a more elongated, fibroblast-like morphology, whereas osteosarcoma (143b) cells tend to be more rounded. On average, MSC are larger than 143b cells in terms of both area and perimeter.

Finally, we want to highlight that this dataset is a unique resource to study cellular behavior in co-culture frameworks, setting it apart from most publicly available datasets, such as the ones from the Cell Tracking Challenge, which only focus on monoculture experiments.

### *B. Cell Detection in Video Frames*

In our experiments, we adopt a semi-automatic approach using the proprietary image-analysis software included with our microscope. This choice offers several advantages: (i) it simplifies the detection pipeline, allowing us to focus on assessing the performance of our novel motion prediction module; (ii) it avoids the need for large, annotated datasets typically required for supervised training; and (iii) it ensures compatibility and consistency with the data acquisition process, as the detection tool is tailored to the imaging system. For each video frame, the software provides region proposals (cell segmentations), from which axis-aligned or oriented bounding boxes can be extracted using basic image processing techniques. State-of-the-art cell segmentation tools rely on training and deploying a dedicated deep learning model for cell detection, as in [42], [43]. However, these approaches typically require a large amount of data and do not discriminate between different cell populations. Our multi-agent approach to cell tracking aims to introduce dynamical system knowledge into the process. This improves tracking performance even when the object detection component cannot exploit a large amount of data.

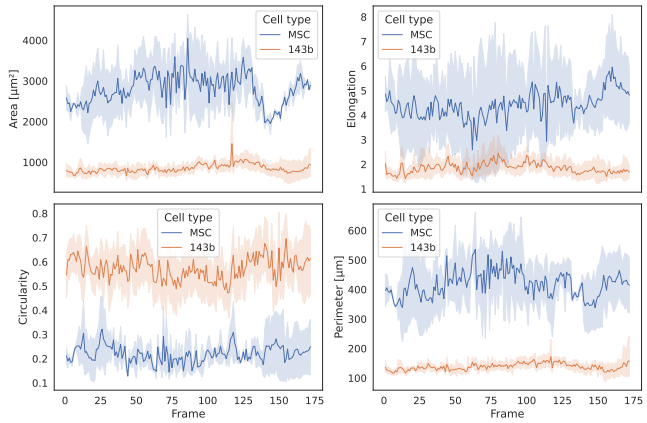



Figure 5: Mean and variance over time for the area, outer perimeter, circularity, and elongation, categorized by cell type.

### *C. Experimental Results*

We implemented the proposed methodology via the Python 3 programming language, testing our algorithm on co-culture videos from *in vitro* experiments involving MSC and osteosarcoma cells. Given the limited dataset of only four videos, we employed a cross-validation approach [44]. Specifically, we trained the models described in Section III, by computing the optimal solution $\boldsymbol{\theta}^*$ as in (11) considering three videos regarding a co-culture scenario ($\Gamma = 2$). Then we evaluated the tracking performances on the remaining one. We repeated this procedure for all possible combinations of training and test videos. To improve the robustness of our estimation and mitigate overfitting due to the small dataset, we adopted data augmentation techniques [45]. Specifically, we applied rotations of angles $\alpha \in 90°, 180°, 270°$ and orthogonal flips to each segmented image to then extract the corresponding rotated versions of $\mathbf{p}_i^t$, $s_i^t$, $r_i^t$, and $\phi_i^t$ for all cells $i$ at each frame $t$. After computing the required time derivatives, these transformed values were used to enlarge $\mathbf{Y}$ and $\mathbf{D}$ on the LS problem. Moreover, we worked on normalized data sequences to ensure better-conditioned matrix inversion, prevent feature predominance, and enhance the interpretability of parameters. The data augmentation and the subsequent identification process were performed offline. The model constants in equations (10), specifically the target area $S_{\gamma_i}$ and aspect ratio $R_{\gamma_i}$ with $\gamma_i \in \{\text{MSC}, \text{143b}\}$, were derived directly from dataset statistics regarding mean area and elongation observed across cell trajectories over time (refer to Figure 5). We further analyzed the learned model parameters by examining their distributions across different training sets. This allowed us to evaluate if our open multi-agent model could distinguish between the two cell types (MSC and 143b cells) by capturing their distinct behaviors. The ability to automatically distinguish cells in a co-culture model, such as MSC and osteosarcoma cells, is crucial, as both originate from the same mesenchymal progenitor line and appear morphologically similar under direct microscope observation. Consequently, identifying and analyzing cellular behaviors and interactions between these two cell types to gain new insights into cancer progression and aggressiveness is currently limited to systems that are expensive, complex, and often impractical to use. Specifically, we examined the parameter sets obtained from training on all possible combinations of the four videos, i.e., the power set of the dataset, which consists of $2^4$ different training sets. For each training set, we computed the identified parameters and analyzed their statistical distribution using kernel density estimation (KDE) to obtain a smooth approximation of the underlying data distribution. Figure 6 illustrates the KDE-based distribution of the identified parameters for the two cell types across different training sets. The analysis reveals distinct parameter distributions, supporting the hypothesis that the proposed model effectively differentiates MSC from 143b cells.

From a biological perspective, the identified parameter weights can be interpreted as follows. The cohesion parameter $\theta_{\gamma_i,1}$ quantifies the tendency of cells to remain in proximity to their neighbors. It reflects the propensity of cells to lose cohesion, detach from the tissue, and migrate to distant sites, a phenotype commonly associated with tumor cells and increased aggressiveness and metastatic potential. In line with this interpretation, our results indicate that mesenchymal stromal cells (MSCs) exhibit a lower tendency to lose cohesion

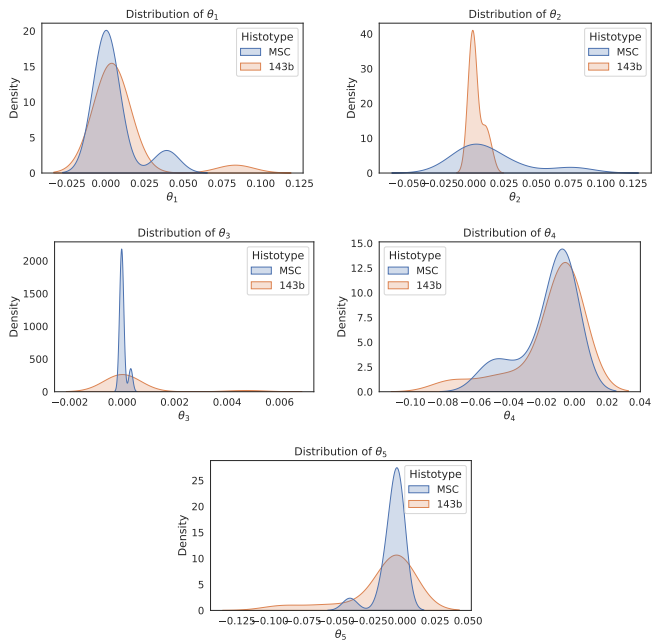



Figure 6: Distributions of the identified model parameters for MSC and 143b cells across diverse training sets.

than tumor cells. The alignment parameter $\theta_{\gamma_i,2}$ measures the tendency of cells to migrate in the same direction as their neighbors. The lower values with smaller variance observed in tumor cells suggest a more heterogeneous and less coordinated migratory behavior compared with MSCs, consistent with the disorganized migration typical of cancer cells that have lost responsiveness to local cues required for coordinated tissue dynamics and structural integrity. Further support for these findings is provided by the repulsion parameter $\theta_{\gamma_i,3}$. Tumor cells display a greater tendency to distance themselves from neighboring cells than normal cells, indicating a disruption of the interaction mechanisms characteristic of organized tissues. Finally, the morphological parameters $\theta_{\gamma_i,4}$ and $\theta_{\gamma_i,5}$, which regulate the tendency to maintain the cell-type-specific target area $S_{\gamma_i}$ and aspect ratio $R_{\gamma_i}$, show similar distributions between the two cell types. MSCs maintain elongated and relatively stable morphologies around their characteristic shape, whereas 143b cells exhibit increased variability around their own, more rounded morphology, consistent with their enhanced motility and well-documented morphological plasticity. Overall, these results demonstrate that the identified parameters not only distinguish MSCs from 143b cells statistically, but also capture impaired regulation of cell-cell interactions. As such, they represent informative indicators of phenotypic differences between normal and tumor cells and may serve as indirect measures of tumor aggressiveness in biological assays, even in the presence of normal cells capable of modulating tumor behavior.

To further detail the tracking algorithm, we now describe the initialization and tuning of the EKF terms used within Algorithm 1. In our implementation of the Extended Kalman Filter, the error covariance matrix $\mathbf{P}^{0|0}$ is initialized by assigning higher variance values to state components that are initially unobserved, reflecting our uncertainty in these estimates. Instead, the measurement covariance matrix $\mathbf{R}$ and the process covariance matrix $\mathbf{Q}$ are set based on baseline estimates obtained from preliminary noise analyses of the dataset. To account for different noise conditions, we further explored scaled versions of these baselines using factors of 0.5, 1.0, and 2.0. This systematic exploration enabled us to quantify the impact of these initialization choices on tracking performance and so identify the optimal settings for our application.

We want to underline that, since our focus is on evaluating the performance metrics of our tracker rather than the detector, we conducted experiments by simulating the detection process using gold standard annotations with injected Gaussian disturbances. Specifically, the measurement errors were simulated considering random noise with zero mean and defined covariance for each annotation. In the following, we present some results considering the state format (7) since, being the standard detection format, it enabled us to perform a comparison with other methodologies. Figure 7 shows a video snapshot regarding the numerical experiments on one of the co-culture videos. Here, we display both the bounding boxes predicted by our algorithm and the original gold standard annotations.

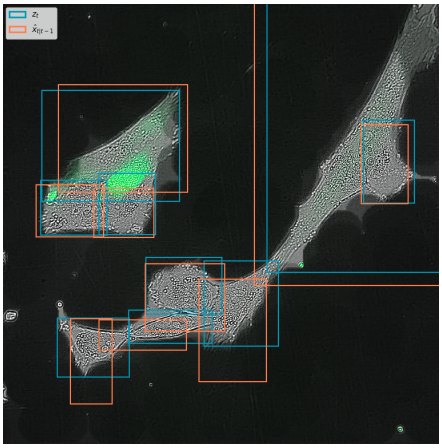


Figure 7: Predicted bounding boxes (orange) and gold standard annotations (blue) in a co-culture video frame.

We compared our scheme against two baseline multi-object tracking methods: SORT [46] and DeepSORT [47]. Both use a Kalman filter for motion prediction and the Hungarian algorithm for data association. SORT performs association using IoU only, whereas DeepSORT employs a neural network trained for re-identification to extract appearance features to improve association. For a fair comparison, the DeepSORT appearance model was fine-tuned on the images from our co-culture dataset. We initialized SORT and DeepSORT with the same matrices as in Algorithm 1 and simulated the detection process for our proposed solution. The chosen metrics are the Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP) [48] and Identity F1 Score (IDF1). We also report the average per-frame processing time, denoted as $T_{\text{avg}} \in \mathbb{R}_{>0}$ for each algorithm. MOTA depends on the number of misses, false positives, and mismatches.

MOTP compares the predicted bounding boxes against the real ones. IDF1 measures the accuracy of identity assignment by comparing correctly matched detections to the average number of ground-truth and predicted detections. Higher values of these metrics indicate better tracking performance. Table I summarizes the results of performing the cross-validation procedure on different experiments, referred to as "Well". This comparison highlights the effectiveness of our approach, as evidenced by a good MOTA score, which reflects the algorithm ability to track cells accurately over time. Furthermore, the MOTP metric demonstrates that the proposed tracking model outperforms SORT and DeepSORT, providing more precise predictions. Note that, DeepSORT does not improve over SORT. We attribute this to scarcity of data: the appearance network in DeepSORT is a relatively large model, whereas our total dataset consists of only four videos. Fine-tuning such a network on so few sequences is unlikely to generalize and can even degrade association performance compared to a purely geometry-based method. Moreover, Table I reports the results of an ablation study that isolates the impact of the motion-prediction component. Indeed, HEOM-EKF, SORT, and DeepSORT share the same Track-by-Detection structure and a KF-based tracking scheme; the key difference lies in the open system model and relative enhanced version of EKF. In HEOM-EKF, this model is identified from real co-culture data and explicitly tailored to heterogeneous, time-varying cell populations, whereas SORT and DeepSORT rely on standard single-object kinematic models. Keeping detection and association fixed isolates the contribution of the proposed multi-agent motion model within the EKF prediction step. In terms of per-frame computation time, our method is slower than SORT but substantially faster than DeepSORT. This outcome is expected since, relative to SORT, HEOM-EKF employs a more sophisticated motion model and an EKF whose state dimension scales with the number of cells (while avoiding the computationally intensive appearance feature extraction required by DeepSORT). Despite this increased complexity, the per-frame runtime remains on the order of tens of milliseconds, which is negligible relative to the minute-scale image acquisition intervals typical of cell culture experiments, thereby ensuring compatibility with real-time microscopy.

Additionally, the proposed algorithm predicts changes in cell population over time and, at the end of the time-lapse sequence, returns the estimated cell lineage tree $\hat{\mathcal{M}}^t$ for the frames $t \in \{1, \ldots, 172\}$ of the current *in vitro* experiment. Figure 8 shows a prediction result of the latter. Here, the state format in (6) was preferred since it allowed for more precise type discrimination and parent-daughter relation prediction. The depicted example of an estimated cell lineage tree is a reliable result despite the absence of a true reference family tree for direct comparison. In fact, MSC cells rarely divide (considering the duration of the single experiment of about 10 hours), while 143b osteosarcoma cells, due to their oncogenic nature, exhibit a significantly higher mitosis rate. The predicted lineage tree captures these expected proliferation patterns, reinforcing its plausibility.

The proposed framework shows promising performance, but the limited dataset size (4 co-culture videos) must be acknowledged. Despite cross-validation and data augmentation, overfitting risks remain. This limitation reflects a broader challenge in cancer-biology: high-quality time-lapse datasets of co-cultures are scarce due to experimental complexity and cost. Thus, the proposed work represents a proof-of-concept in this data-scarce context and highlights that our proposed method can represent a valuable solution as opposed to Machine Learning approaches that typically require large datasets for training. Moreover, this study provides a unique dataset for investigating MSC-143b interactions and evaluating tracking in co-culture experiments.

Table I: Performance comparison

| Metric | Experiment | SORT | DeepSORT | Alg. 1 |
|---|---|---|---|---|
| MOTA | Well 1 | 74.23 % | 55.30 % | **83.56 %** |
| | Well 2 | 40.78 % | 52.77 % | **67.58 %** |
| | Well 3 | 76.38 % | 54.09 % | **87.02 %** |
| | Well 4 | 79.97 % | 55.63 % | **96.17 %** |
| MOTP | Well 1 | 93.68 % | 85.33 % | **96.72 %** |
| | Well 2 | 91.45 % | 84.04 % | **95.51 %** |
| | Well 3 | 93.34 % | 83.56 % | **96.03 %** |
| | Well 4 | 94.33 % | 85.38 % | **97.18 %** |
| IDF1 | Well 1 | 62.61 % | 50.31 % | **69.32** % |
| | Well 2 | 54.20 % | 38.79 % | **66.67** % |
| | Well 3 | 65.59 % | 47.76 % | **71.07** % |
| | Well 4 | **82.02** % | 48.12 % | 80.88 % |
| $T_{\text{avg}}$ [ms/frame] | Well 1 | **2.33** | 204.98 | 59.73 |
| | Well 2 | **0.94** | 94.07 | 5.33 |
| | Well 3 | **3.32** | 184.42 | 60.72 |
| | Well 4 | **1.76** | 170.57 | 40.27 |

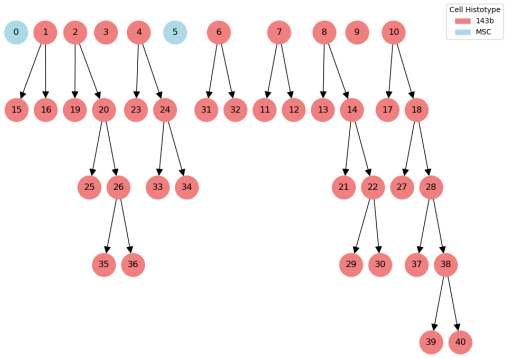



Figure 8: Example of a predicted cell lineage tree.

## VI. Conclusions

In this paper we developed a novel algorithm for tracking cells in optical microscopy videos through open multi-agent modeling. System identification techniques applied to *in vitro* experiments enabled to learn cell type-specific behaviors, accounting for cell-to-cell interactions even in multi-culture scenarios. The open multi-agent system framework is the building block of the proposed cell tracking algorithm that is based on a generalized EKF. Cross-validation tests with proper performance metrics were shown together with a comparison with two baseline algorithms in multi-object tracking. Results demonstrate that the proposed method represents a valuable solution in a data-scarce context as cancer biology for tracking

heterogeneous cell types, capturing cell-to-cell interactions and generating the cell lineage tree. This study establishes a unique testing framework using a dataset obtained from *in vitro* co-culture experiments with tumor and normal cells.

## REFERENCES


[1] D. Salzano, D. Fiore, and M. di Bernardo, "Ratiometric control for differentiation of cell populations endowed with synthetic toggle switches," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 927–932.

[2] D. Fiore, D. Salzano, E. Cristòbal-Cóppulo, J. M. Olm, and M. di Bernardo, "Multicellular feedback control of a genetic toggle-switch in microbial consortia," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 151–156, 2020.

[3] M. Khammash, M. Di Bernardo, and D. Di Bernardo, "Cybergenetics: Theory and methods for genetic control system," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 916–926.

[4] T. E. Gorochowski, A. Matyjaszkiewicz, T. Todd, N. Oak, K. Kowalska, S. Reid, K. T. Tsaneva-Atanasova, N. J. Savery, C. S. Grierson, and M. Di Bernardo, "Bsim: an agent-based tool for modeling bacterial populations in systems and synthetic biology," *PLoS ONE*, 2012.

[5] J. Mason and D. Öhlund, "Key aspects for conception and construction of co-culture models of tumor-stroma interactions," *Frontiers in Bioengineering and Biotechnology*, vol. 11, p. 1150764, 2023.

[6] E. J. Sutton, T. D. Henning, B. J. Pichler, C. Bremer, and H. E. Daldrup-Link, "Cell tracking with optical imaging," *European Radiology*, vol. 18, pp. 2021–2032, 2008.

[7] H.-J. Cheng, C.-H. Hsu, C.-L. Hung, and C.-Y. Lin, "A review for cell and particle tracking on microscopy images using algorithms and deep learning technologies," *Biomedical Journal*, vol. 45, no. 3, pp. 465–471, 2022.

[8] V. Ulman, M. Maška, K. E. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic *et al.*, "An objective comparison of cell-tracking algorithms," *Nature Methods*, vol. 14, no. 12, pp. 1141–1152, 2017.

[9] M. Maška, V. Ulman, P. Delgado-Rodriguez, E. Gómez-de Mariscal, T. Nečasová, F. A. Guerrero Peña, T. I. Ren, E. M. Meyerowitz, T. Scherr, K. Löffler *et al.*, "The cell tracking challenge: 10 years of objective benchmarking," *Nature Methods*, pp. 1–11, 2023.

[10] Y. Chen, Y. Song, C. Zhang, F. Zhang, L. O'Donnell, W. Chrzanowski, and W. Cai, "Celltrack r-cnn: A novel end-to-end deep neural network for cell segmentation and tracking in microscopy images," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 779–782.

[11] J. C. Kimmel, A. S. Brack, and W. F. Marshall, "Deep convolutional and recurrent neural networks for cell motility discrimination and prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 2, pp. 562–574, 2019.

[12] D. E. Hernandez, S. W. Chen, E. E. Hunter, E. B. Steager, and V. Kumar, "Cell tracking with deep learning and the viterbi algorithm," in *IEEE International Conference on Manipulation, Automation and Robotics at Small Scales*, 2018, pp. 1–6.

[13] H. Hu, L. Zhou, Q. Guan, Q. Zhou, and S. Chen, "An automatic tracking method for multiple cells based on multi-feature fusion," *IEEE Access*, vol. 6, pp. 69 782–69 793, 2018.

[14] T. Ben-Haim and T. R. Raviv, "Graph neural network for cell tracking in microscopy videos," in *European Conference on Computer Vision*. Springer, 2022, pp. 610–626.

[15] K. Löffler and R. Mikut, "Embedtrack—simultaneous cell segmentation and tracking through learning offsets and clustering bandwidths," *IEEE Access*, vol. 10, pp. 77 147–77 157, 2022.

[16] O. Hirose, S. Kawaguchi, T. Tokunaga, Y. Toyoshima, T. Teramoto, S. Kuge, T. Ishihara, Y. Iino, and R. Yoshida, "Spf-celltracker: Tracking multiple cells with strongly-correlated moves using a spatial particle filter," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 6, pp. 1822–1831, 2017.

[17] F. Boukari and S. Makrogiannis, "Automated cell tracking using motion prediction-based matching and event handling," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 3, pp. 959–971, 2018.

[18] T. E. Gorochowski, S. Hauert, J.-U. Kreft, L. Marucci, N. R. Stillman, T.-Y. D. Tang, L. Bandiera, V. Bartoli, D. O. Dixon, A. J. Fedorec *et al.*, "Toward engineering biosystems with emergent collective functions," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 705, 2020.

[19] P. Mina, K. Tsaneva-Atanasova, and M. d. Bernardo, "Entrainment and control of bacterial populations: an in silico study over a spatially extended agent based model," *ACS Synthetic Biology*, vol. 5, no. 7, pp. 639–653, 2016.

[20] S. Izumi, S.-I. Azuma, and T. Sugie, "Multi-robot control inspired by bacterial chemotaxis: coverage and rendezvous via networking of chemotaxis controllers," *IEEE Access*, vol. 8, pp. 124 172–124 184, 2020.

[21] A. Mironchenko, "Live systems of varying dimension: modeling and stability," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 3956–3961.

[22] H. Enderling and M. AJ Chaplain, "Mathematical modeling of tumor growth and treatment," *Current Pharmaceutical Design*, vol. 20, no. 30, pp. 4934–4940, 2014.

[23] V. Martinelli, D. Salzano, D. Fiore, and M. di Bernardo, "Multicellular pi control for gene regulation in microbial consortia," *IEEE Control Systems Letters*, vol. 6, pp. 3373–3378, 2022.

[24] A. Guarino, D. Fiore, D. Salzano, and M. di Bernardo, "Balancing cell populations endowed with a synthetic toggle switch via adaptive pulsatile feedback control," *ACS Synthetic Biology*, vol. 9, no. 4, pp. 793–803, 2020.

[25] C. Romano, A. Borri, and M. D. Di Benedetto, "Adaptive responses of cancer cells to tumor microenvironment changes: an evolutionary model perspective," in *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2024, pp. 2079–2084.

[26] H. Jiao, Q. Shen, Y. Shi, and P. Shi, "Adaptive tracking control for uncertain cancer-tumor-immune systems," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 6, pp. 2753–2758, 2020.

[27] L. Von Bertalanffy, "The theory of open systems in physics and biology," *Science*, vol. 111, no. 2872, pp. 23–29, 1950.

[28] H. Oh, A. R. Shirazi, C. Sun, and Y. Jin, "Bio-inspired self-organising multi-robot pattern formation: A review," *Robotics and Autonomous Systems*, vol. 91, pp. 83–100, 2017.

[29] M. Cortini, S. Avnet, and N. Baldini, "Mesenchymal stroma: role in osteosarcoma progression," *Cancer Letters*, vol. 405, pp. 90–99, 2017.

[30] S. Avnet, G. Di Pompo, T. Chano, C. Errani, A. Ibrahim-Hashim, R. J. Gillies, D. M. Donati, and N. Baldini, "Cancer-associated mesenchymal stroma fosters the stemness of osteosarcoma cells in response to intratumoral acidosis via nf-$\kappa$b activation," *International journal of cancer*, vol. 140, no. 6, pp. 1331–1345, 2017.

[31] S. Avnet, S. Lemma, M. Cortini, G. Di Pompo, F. Perut, M. V. Lipreri, L. Roncuzzi, M. Columbaro, C. Errani, A. Longhi *et al.*, "The release of inflammatory mediators from acid-stimulated mesenchymal stromal cells favours tumour invasiveness and metastasis in osteosarcoma," *Cancers*, vol. 13, no. 22, p. 5855, 2021.

[32] C. Massaro, H. N. Sensoy, M. Mulders, C. De Schrijver, C. Gómez-Martín, J. Simon Nieto, T. Lagerweij, A. Atmopawiro, J. Pérez-Boza, M. Bebelman *et al.*, "Tumor-secreted extracellular vesicles counteract therapy response by triggering inflammatory mesenchymal stem cell development," *Clinical Cancer Research*, vol. 30, no. 20, pp. 4714–4728, 2024.

[33] A. Tramaloni, A. Testa, S. Avnet, N. Baldini, and G. Notarstefano, "Biological cell tracking via multi-agent identification and filtering," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*. IEEE, 2024, pp. 1338–1343.

[34] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, "Deep learning in video multi-object tracking: A survey," *Neurocomputing*, vol. 381, pp. 61–88, 2020.

[35] H. Sayama, "Swarm chemistry," *Artificial Life*, vol. 15, no. 1, pp. 105–114, 2009.

[36] M. Eyiyurekli, L. Bai, P. I. Lelkes, and D. E. Breen, "Chemotaxis-based sorting of self-organizing heterotypic agents," in *ACM Symposium on Applied Computing*, 2010, pp. 1315–1322.

[37] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the national academy of sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[38] R. D. Cook, "Detection of influential observation in linear regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.

[39] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[40] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.

[41] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[42] E. Pedone, I. De Cesare, C. G. Zamora-Chimal, D. Haener, L. Postiglione, A. La Regina, B. Shannon, N. J. Savery, C. S. Grierson, M. Di Bernardo *et al.*, "Cheetah: a computational toolkit for cybergenetic control," *ACS Synthetic Biology*, vol. 10, no. 5, pp. 979–989, 2021.

[43] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature methods*, vol. 18, no. 1, pp. 100–106, 2021.

[44] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[45] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, 2022.

[46] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *IEEE International Conference on Image Processing*, 2016, pp. 3464–3468.

[47] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 3645–3649.

[48] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.

**Andrea Tramaloni** (S'23) received a master's degree in automation engineering from the University of Bologna in 2022. Currently, he is a Ph.D. student at the Center for Research on Complex Automated Systems (CASY), Department of Electrical and Information Engineering (DEI), University of Bologna, Bologna, Italy. He was a visiting scholar at the University of Bristol in 2025. His research interests include systems biology, cybergenetics, multi-agent systems, and optimal control.

**Andrea Testa** received the Laurea degree "summa cum laude" in Computer Engineering from the Università del Salento, Lecce, Italy in 2016 and the Ph.D degree in Engineering of Complex Systems from the same university in 2020. While working on this paper, he was a Junior Assistant Professor at Alma Mater Studiorum Università di Bologna, Bologna, Italy. He was a visiting scholar at LAAS-CNRS, Toulouse, (July to September 2015 and February 2016) and at Alma Mater Studiorum Università di Bologna (October 2018 to June 2019). His research interests include control of UAVs and distributed optimization. Dr. Andrea Testa served as an Associate Editor for the *19th* and *20th IEEE International Conference on Automation Science and Engineering (CASE)*.

**Sofia Avnet** received a Degree in Medical Biotechnologist, in 2000, and the Ph.D. degree, in 2005, a second Degree in cellular and molecular biology, in 2010, and the Specialization degree in pharmacology and toxicology, in 2023. During the first master's degree, she began research on bone tumors with the Kimmel Cancer Institute and the IRCCS Istituto Ortopedico Rizzoli. For the Ph.D. degree, she continued studying biological mechanisms of bone cancers with the Institute for Cancer Research, Candiolo. She also conducted research on bone pathophysiology as a Visiting Researcher with McGill University, in 2006, and a Research Fellow with the University of Turku, in 2007. From 2008 to 2019, she was a Biotechnologist with the Rizzoli Orthopedic Institute, focusing her research on tumor invasiveness, chemoresistance, microenvironment, and metabolism in bone cancers. Since 2019, she has been an Assistant Professor with the University of Bologna, where she develops 3D bone tissue and bone cancer models using spheroids, microfluidics, and bioprinting. Over her career, she has received several research grants as a Principal Investigator from the Italian Association of Cancer Research and the Ministry of Health. Dr. Avnet has been the President of the International Society of Cancer Metabolism (2015-2016).

**Stefania Massari** holds an MSc in Medical Biotechnology (2022) and a BSc in Biological Science and Technology (2018), both from the University of Trieste. She has been a biotechnologist and research fellow at the Rizzoli Orthopaedic Institute, where she has worked on molecular oncology and cancer metabolism. She has contributed to an AIRC-funded project investigating the effects of tumor microenvironment acidity on lipid metabolism in osteosarcoma. Currently, she is a PhD student at the Medical University of Vienna.

**Gemma Di Pompo** is a biologist (Master Degree in 2009 from University of Bologna), with a PhD in Biotechnology, Pharmacology, and Toxicology (2013) from the University of Bologna. In 2014, she has been a visiting researcher at Orofacial Development and Regeneration Unit (Institute of Oral Biology, Centre of Dental Medicine, University of Zurich, Switzerland). Currently, a senior researcher at the Rizzoli Orthopaedic Institute, her work focuses on bone biology, cancer microenvironment, and regenerative medicine. She has participated in multiple national and international research projects on bone tumors, cancer-associated pain, and advanced therapies using 3D bioprinting and tissue engineering. She was recently awarded a research grant as Principal Investigator by the AIRC Foundation to support her ongoing studies in targeting nerve-cancer crosstalk as a novel approach to improve osteosarcoma treatment. She has authored 31 peer-reviewed publications and has received recognitions, including the European Orthopaedic Research Society (EORS) Exchange Travel Grant and the "Top Junior Scientist" award.

**Nicola Baldini** received the Degree in medicine and surgery, the Certification degree in orthopaedics, and the Certification degree in oncology from the University of Bologna, Bologna, Italy, in 1981, 1987, and 1990, respectively. From 1988 to 1989, he was a Research Fellow with Harvard University. From 2013 to 2018, he was a Visiting Professor with Kyoto Prefectural University of Medicine. He is currently a Professor with the Department of Biomedical and Neuromotor Sciences, Alma Mater Studiorum, University of Bologna, and the Director of the Department of Research, Innovation and Technology, IRCCS Istituto Ortopedico Rizzoli, Bologna. His research interests include the microenvironment of bone tissue under physiological and pathological conditions, bone cancers, and regenerative medicine. Prof. Baldini has served as the President for the Connective Tissue Oncology Society, from 1999 to 2000, the European Orthopaedic Research Society, from 2008 to 2012, and the International Society for Cancer Metabolism, from 2019 to 2020.

**Giuseppe Notarstefano** (M'11) is a Professor in the Department of Electrical, Electronic, and Information Engineering G. Marconi at Alma Mater Studiorum Università di Bologna. He was Associate Professor (June '16 - June '18) and previously Assistant Professor, Ricercatore, (February '07 - June '16) at the Università del Salento, Lecce, Italy. He received the Laurea degree "summa cum laude" in Electronics Engineering from the Università di Pisa in 2003 and the Ph.D. degree in Automation and Operation Research from the Università di Padova in 2007. He has been visiting scholar at the University of Stuttgart, University of California Santa Barbara and University of Colorado Boulder. His research interests include distributed optimization, cooperative control in complex networks, applied nonlinear optimal control, trajectory optimization and maneuvering of aerial and car vehicles. He serves as an Associate Editor for IEEE Transactions on Automatic control, IEEE Transactions on Control Systems Technology and IEEE Control Systems Letters. He is also part of the Conference Editorial Board of IEEE Control Systems Society and EUCA. He is recipient of an ERC Starting Grant.