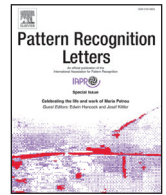




ELSEVIER

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Ranking cell tracking systems without manual validation[☆]


 Andrey Kan^{a,*}, John Markham^b, Rajib Chakravorty^b, James Bailey^a, Christopher Leckie^a
^a Department of Computing and Information Systems, Victoria Research Laboratory, National ICT Australia (NICTA), University of Melbourne, VIC, Australia

^b Department of Electrical and Electronic Engineering, Victoria Research Laboratory, National ICT Australia (NICTA), University of Melbourne, VIC, Australia

ARTICLE INFO

Article history:

Received 7 February 2014

Available online 22 November 2014

Keywords:

 Cell tracking
 Tracking quality
 Optimal assignment
 Tracker selection
 Bayes theorem

ABSTRACT

Automated cell segmentation and tracking can significantly increase the productivity of research in biology. In order to tune a tracking system for a particular video, researchers usually have to manually annotate a part of the video, and tune the algorithm with respect to this ground truth. However, large variability in cell video characteristics leads to different trackers and parameters being optimal for different videos. Therefore for any new video, manual annotation and tuning has to be performed again. Alternatively, suboptimal parameters have to be used which may result in a significant amount of manual post-correction being required. The challenge that we address in this paper is automated selection and tuning of cell tracking systems without the need for manual annotation. Given an estimate of the cell size only, our method is capable of ranking the trackers according to their performance on the given video *without the need for ground truth*. Our evaluation using real videos and real tracking systems indicates that our method is capable of selecting the best or nearly best tracker and its parameters in practical scenarios.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Over the past decades single cell tracking results combined with mathematical modeling are having an increasing impact on cell biology [1,2]. Automated cell segmentation and tracking can significantly increase the productivity of biological research. A major challenge in developing such a system is the large diversity of cell morphology and motility as well as variations in recording conditions (Fig. 1). This diversity has resulted in a large number of proposed cell tracking systems, where each system has parameters that need to be specified [3]. By a cell tracking system (CTS) we mean a combination of algorithms that is capable of both locating cells in video frames (cell segmentation) and maintaining cell identities throughout the video (data association). These two tasks can be approached separately [4] or within a single algorithm [5]. Furthermore, by a CTS we mean a combination of such algorithms with their parameters fixed to specific values. For example, we treat the same software with different parameters as two different CTSs. Occasionally by “CTS for the given video” we also mean “results of the CTS given the video as input”.

Due to the variability in experimental conditions, optimal combinations of algorithms and parameters can vary for different videos, even for the same cell type [7]. In order to find the best CTS for a given video, a practical solution is to annotate a part of a video and use the resulting ground truth to evaluate the performance of different CTSs. However, if a video from another experiment needs to be processed, the previously best found CTS may not be the best anymore, and ideally, a part of the new video needs to be manually annotated. Even within a single long video, a CTS tuned on one part of it may not be the best for another part. Manual annotation for each new video or different parts of the same video can severely compromise the effectiveness of an automated CTS.

Consider the following real example from an Australian medical research institution. A lab recorded a set of novel cell videos. The analysis of results required cell tracking, and multiple software packages appeared suitable for this task (e.g., [8,9]). Initial ground truth could have been created totally manually, but it was found beneficial to use a CTS with imperfectly guessed parameter values and then correct the automated results. It then took a researcher–biologist a few hours to choose and guess parameters and a few more hours to manually correct results in order to produce the ground truth (cell outlines and identities over 200 frames) for only three cells. This is a large amount of manual time, given that more cells are required for a representative ground truth and that the lab usually produces a few novel videos each year.

Here we address this challenge with a system for ranking CTSs *without the need for ground truth*. To the best of our knowledge, this is

[☆] This paper has been recommended for acceptance by A. Shokoufandeh.

* Corresponding author. Tel.: +61 3 9345 2596; fax: +61 3 9349 4596.

 E-mail addresses: andreyk@unimelb.edu.au (A. Kan), jmarkham@wehi.edu.au (J. Markham), Rajib.Chakravorty@nicta.com.au (R. Chakravorty), baileyj@unimelb.edu.au (J. Bailey), caleckie@csse.unimelb.edu.au (C. Leckie).

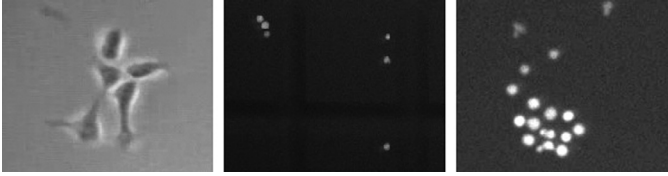


Fig. 1. Examples of a diversity in cell morphology and recording conditions. The images show neural progenitor cells (left, with permission from [6]) and B-lymphocytes in micro-grids (middle, right).

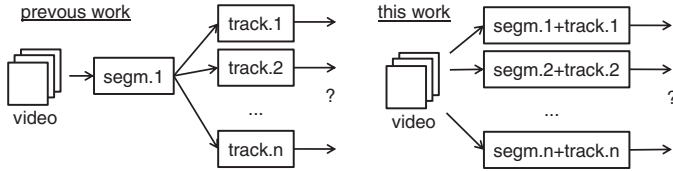


Fig. 2. Where segmentation and tracking steps are separated, the previous method is capable of choosing the best tracker, given a fixed segmentation [7]. In this work, we address a more general problem of choosing among a number of CTS, where a CTS can have a combined segmentation and tracking steps. Both methods operate in the absence of the ground truth.

the only system of its kind in cell tracking literature. Given a video (or a fragment) and a range of candidate CTSs, the user is only asked to provide an estimate of the cell size. Our system then ranks the CTSs according to their performance on the given video.

The problem of automated performance estimation has been previously addressed for medical image segmentation and tracking for surveillance cameras [10,11]. The previous methods rely on knowledge specific to their respective domains and are not directly applicable to CTSs. In the context of cell tracking, given a video and a fixed segmentation step, a previous method [7] is capable of comparing data association algorithms (Fig. 2). A more general problem of choosing among a number of CTS, where both segmentation and tracking steps can differ, remains an open research challenge that we address in this work.

Here we propose a novel method for ranking CTSs without the need for ground truth and using minimum user input. We design a scheme of pairwise comparisons (Section 2.2), and employ a special case of an optimal assignment problem to match detections (Section 2.3). Finally, we develop a general face validity test for a CTS (Section 2.4). We find that together these components based on minimum prior information can be effective in practical scenarios (Section 3).

2. Methods

2.1. Cell tracking preliminaries

A CTS is a function that takes a video as input and produces a detection and a set of links as output. A *detection* is a set of measurements, and a *measurement* is a vector that comprises a numerical description of each located object (tentative cell). A CTS can produce different numbers of measurements for different frames. Each measurement can contain information such as the cell centroid location, mean brightness, size, etc. While different CTSs may differ in the type of information they produce, we assume that cell centroid location is always present, and in this paper, we usually treat measurements as cell centroid locations. Finally, a *link* is a pair $\{\bar{m}_{i,j}, \bar{m}_{i+1,k}\}$ of measurements from consecutive frames i and $i+1$ (j, k index measurements). Links represent tentative moves or division events (Fig. 3).

Performance of a CTS is a measure of accuracy of detections and links. Given a video, a resulting detection and a set of links, performance is measured with respect to some manual annotation (*ground truth*). Based on the manual annotation, each measurement from

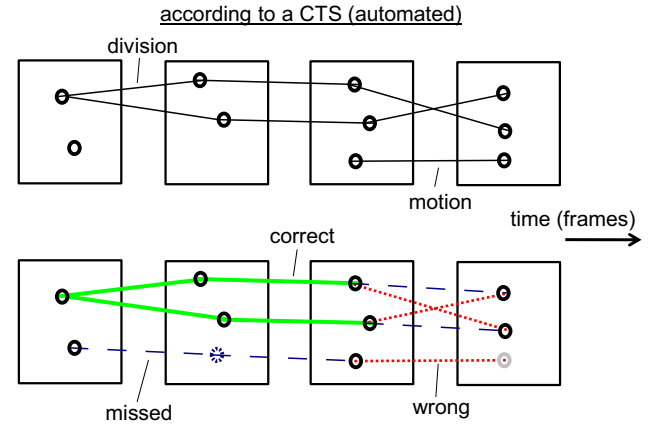


Fig. 3. A CTS produces a detection (set of circles) and a set of links, where each link can denote a tentative cell division or motion. After a manual validation, each link can be classified into correct (bold green, connects two correct measurements), wrong (dotted red, wrong association or incorrect measurements), and missed (dashed blue, absent in the output of the CTS). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the detection can be classified as either correct or spurious (Fig. 3), where the *spurious measurement* is a measurement that does not originate from any cell. Additionally there can be *missed* measurements when a cell is not represented with any matching measurement. Furthermore, each link can be classified as either correct, wrong, or missing (Fig. 3). The CTS performance is then computed as $F = N_{\text{corr}} / (N_{\text{corr}} + N_{\text{wrong}} + N_{\text{miss}})$. This equation defines an F-score, and it has been shown that such a performance measure adequately represents tracking accuracy [7]. Importantly, the above performance measure is defined with respect to a fixed input video. For example, one CTS can be better for one video, and another for a different video.

Finally, CTSs can be implemented as two sequential steps: cell segmentation that produces a detection, and data association that creates links over the detection [12,13]. Here, a cell segmentation algorithm can produce a detection first, and then different tracking algorithms can produce different versions of the links for the same detection. Given a fixed detection, each version of the links can be characterized by a quantity called the *ED-score* (see Appendix A). It has been shown that the ED-score correlates with the F-score in practical situations [7]. Note that ED-score does not require ground truth to be computed. In contrast, the computation of the F-score is based on manual annotation. Our CTS ranking system employs the ED-score in pairwise comparisons. Among other additions, we supplement the ED-score with a new method of detection matching presented below.

2.2. Method overview

The overall aim of this study is a system capable of comparing relative performance of CTSs while requiring minimal input from the user. While employing a previously proposed ED-score for this task, the three major challenges are (1) matching detections from different CTSs; (2) using minimum prior knowledge; and (3) turn pairwise comparisons into a ranking. We address the first challenge in Section 2.3. Furthermore, where possible we use rather general assumptions. Finally, the ranking challenge is addressed using cross-comparison based rankings with the exclusion of infeasible solutions.

A high level overview of our solution is presented in Algorithm 1. Given a set of results from candidate CTSs for the given video sequence, we first eliminate infeasible CTSs using the *FaceValidity* test (Section 2.4). We then perform pairwise comparisons of the remaining CTSs, and for each CTS compute the total number of wins across all comparisons. At the core of our method is a single comparison of two CTSs, in which one of the CTS results is estimated to be better

Download English Version:

<https://daneshyari.com/en/article/535281>

Download Persian Version:

<https://daneshyari.com/article/535281>

[Daneshyari.com](https://daneshyari.com)