Automated tracking and analysis of centrosomes in early Caenorhabditis elegans embryos

Steffen Jaensch\textsuperscript{1,2,*}, Markus Decker\textsuperscript{1}, Anthony A. Hyman\textsuperscript{1}, and Eugene W. Myers\textsuperscript{2}

\textsuperscript{1}Max-Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany and \textsuperscript{2}Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA

ABSTRACT

Motivation: The centrosome is a dynamic structure in animal cells that serves as a microtubule organizing center during mitosis and also regulates cell-cycle progression and sets polarity cues. Automated and reliable tracking of centrosomes is essential for genetic screens that study the process of centrosome assembly and maturation in the nematode Caenorhabditis elegans.

Results: We have developed a fully automatic system for tracking and measuring fluorescently labeled centrosomes in 3D time-lapse images of early C.elegans embryos. Using a spinning disc microscope, we monitor the centrosome cycle in living embryos from the 1- up to the 16-cell stage at imaging intervals between 30 and 50 s. After establishing the centrosome trajectories with a novel method involving two layers of inference, we also automatically detect the nuclear envelope breakdown in each cell division and recognize the identities of the centrosomes based on the invariant cell lineage of C.elegans. To date, we have tracked centrosomes in over 500 wild type and mutant embryos with almost no manual correction required.

Availability: The centrosome tracking software along with test data is freely available at http://publications.mpi-cbg.de/itemPublication.html?documentId=4082.

Contact: jaensch@mpi-cbg.de

1 INTRODUCTION

Genetic screens that produce and require analysis of massive amounts of image data are becoming a major tool in functional genomics. Here, we do so to study the assembly and maturation of centrosomes, an organelle in animal cells that serves as a microtubule organizing centre (Alberts \textit{et al}., 1994) during mitosis, in the nematode Caenorhabditis elegans. In principle, centrosomes can be of any size. The embryo, however, exhibits remarkable precision in setting centrosome size in relation to cell size that get progressively smaller through development. This is crucial because centrosome size in turn sets the length of the mitotic spindle (Greenan \textit{et al}, 2010) and is involved in signaling activities within the cell (Fletcher et al., 2008). Their goal is to trace the cell lineage up to hundreds of cells in one color channel and use a second color channel to monitor gene expression. Compared to our problem, the objects tracked are large, their phase easily distinguished by their shape, and their frame-to-frame movement is small, none of which are true for centrosomes. In this article we present a suite of algorithms that analyze these images (Fig. 1).

To robustly solve the problem of tracking centrosomes in C.elegans embryos, we use fluorescent markers such as GFP and reduce their levels by RNA interference (RNAi) in order to assay function. We then produce 3D time-lapse images of such labeled and/or genetically altered centrosomes. In this article we present a suite of algorithms that analyze these images (Fig. 1).

To visualize centrosomal proteins of interest in living C.elegans embryos, we use fluorescent markers such as GFP and reduce their levels by RNA interference (RNAi) in order to assay function. We then produce 3D time-lapse images of such labeled and/or genetically altered centrosomes. In this article we present a suite of algorithms that analyze these images (Fig. 1).

In the earliest related work, Bao \textit{et al}., (2006) report a system that identifies and tracks nuclei and meta-phase chromosomes in C.elegans via histone-GFP. Their goal is to trace the cell lineage up to hundreds of cells in one color channel and use a second color channel to monitor gene expression. Compared to our problem, the objects tracked are large, their phase easily distinguished by their shape, and their frame-to-frame movement is small, none of which are true for centrosomes. Recently, Kerekes \textit{et al}., (2009) present an algorithm for tracking centrosomes in neuron cells for the study of neuron migration. Short sequences of only 20 time points at 16 s intervals are recorded and the analysis cannot handle multiple centrosomes at the same x/y-location. These circumstances are thus considerably simpler than those considered here\textsuperscript{1}.

To robustly solve the problem of tracking centrosomes in developing C.elegans embryos, we have extended a bipartite...
that the centrosome is always constant or increasing in size up to cell size and embryo size. On the other hand, one can safely assume time. The mutational variants can further affect centrosome size, assume that the number of centrosomes is a power of 2 at any given stage and genetic factors. Due to the asynchrony of the cell divisions cytoplasm. As the nucleus breaks down (Gorjanacz recruit of proteins such as arranged, barrel-shaped centrioles, each 200-nm in length, located next to the nucleus. The centrosomes then move to opposite poles. As the nucleus breaks down (Gorjanacz et al., 2007), microtubules grow outward from the centrosomes and attach to the chromosomes and to the anterior and posterior cell cortex. Thereafter the centrosomes are pulled apart, toward opposite sides of the cell, by means of these fibers. As the centrosomes do so, they pull copies of each chromosome toward them to form the DNA content of each emerging daughter cell. Finally, the centrosomes come to a stop and are rapidly disassembled. This appears in the images to be analyzed as either an explosion into several pieces that then rapidly dissolve or as a dissolving of the entire centrosome.

We follow this process through the first four to five rounds of cell divisions of a C.elegans zygote (fertilized egg). The process becomes slower the higher the cell stage, the centrosomes become smaller, and they reach their maximum size earlier in the division process. Their maximum diameter varies from ~1 – 5 μm depending on the cell stage and genetic factors. Due to the asynchrony of the cell divisions and the fact that we wish also to track mutational variants, one cannot assume that the number of centrosomes is a power of 2 at any given time. The mutational variants can further affect centrosome size, movement, the number per cell and cell cycle duration as well as cell size and embryo size. On the other hand, one can safely assume that the centrosome is always constant or increasing in size up to beginning of its dissolution and they have a biophysical limit on their maximum velocity.

2 METHODS

2.1 Imaging

Using GFP-strains of any centrosomal protein such as γ-TUB-GFP: we image centrosomes in developing C.elegans embryos by live-cell fluorescence microscopy for up to ~ 2h with a spinning disc microscope. The GFP-signal attenuates significantly towards the bottom of the embryo and also decreases with time due to photo-bleaching. Although it takes only ~ 6s to acquire a 3D stack with 50 planes, we image only at time intervals between 30 and 50 s to minimize laser-induced damage of the embryo as well as said photo-bleaching. Each 3D stack is 512 × 512 × 50 pixels, and typically 100 time points are sampled resulting in roughly 2 Gb of data per experiment. The image resolution in x and y is 133 nm per pixel and the distance between two z-planes is set to 500 or 600 nm. The embryo is usually stationary, but small movements cannot be excluded and manual adjustments of the focus during the pause between two time point samplings may cause additional movement of the whole embryo in the z-direction. With the laser set to minimize exposure levels, the signal to noise ratio is quite low and autofluorescence as well as random particles further corrupt the images. Nonetheless, the ellipsoidal shape of centrosomes and the temporal coherence of their trajectories allow us to reliably analyze the resulting movies.

2.2 Centrosome tracking

Our tracking algorithm begins like many others by over-detecting spot-like objects. It is tuned to very high sensitivity, so that missing detections virtually do not occur. We then find the core sub-trajectories that we are confident are correct, in that the appearance and displacement vector sequence of the centrosome through the trajectory are mutually consistent and the individual spot calls used are clearly not noise. Our approach is further distinctive in that it adaptively determines the images. Nonetheless, the ellipsoidal shape of centrosomes and the temporal coherence of their trajectories allow us to reliably analyze the resulting movies.

2.2.1 Object over-detection. We first over-detect possible centrosomes in each 3D stack by finding all objects that are to first approximation a Gaussian spot of some radius. Each stack I is first smoothed slightly to give \( I_x \) by applying a 3D-Gaussian filter of radius ~6 pixels chosen as per Thomann (2003). We consider there to be an object at position \( p \) with radius \( \sigma \) if (i) the value at \( p \) in \( I_x \) is locally maximal, and (ii) the normalized cross-correlation (Haralick and Shapiro, 1992) of the raw image \( I \) with a Mexican hat filtered center at \( p \) is greater than \( \gamma \). Several discrete values of \( \sigma \) are used and we find that setting \( \gamma \) to 0.7 gives a very low false negative rate (i.e. it is rare to not detect a centrosome) while leaving us with a manageable number of false positive signals (e.g. noise and autofluorescent “spots”).

Unfortunately, the same centrosome will be detected multiple times at slightly different radii and different positions. To solve this problem, in a post-processing step we first compute for each detected object \( p \) a refined estimate of its (\( x,y \))-position, its radius \( \rho(p) \) and its mean-illumination-to-background ratio \( \sigma(p) \) as we will explain below. We then iteratively detect pairs of objects \( p \) and \( q \) that are closer together than the sum of their radii \( \rho(p) + \rho(q) \) and discard the dimmer object (i.e. the one with the smaller \( \eta \)-value). This eliminates on average 54% ±10% (SD) (in ~40 movies) of the initially detected objects and is clearly conservative as two centrosomes do not intersect in space.

Another problem that may cause tracking errors are objects outside the embryo such as the polar body or remains of the embryo’s parent. We remove them by first detecting the contour of the embryo in a maximum intensity
projection using Otsu thresholding (Otsu, 1979) and marker-based watershed segmentation (Meyer, 1994), and then discarding all object calls that are outside the convex hull of the detected embryo contour. This step eliminates on average 28% ± 19% (SD: n = 40) of the objects from the last step.

We now have a set of possible centrosome centres and an estimate of their radius. Empirically, it is exceeding rare for a centrosome to not be in this set, and anywhere from 0 to 90% (average: 34% ± 18% of the remaining spots are false positive signals (e.g. noise, etc.) depending on the time point and acquisition conditions.

2.2.2 Object refinement and estimation

To refine a putative centrosome’s centre (x(p),y(p)) and compute its radius r(p), its mean illumination-to-background ratio η(p) as well as its total illumination ι(p), we find the best least squares fit of the spot at p to an elliptical 2D-Gaussian model with a background offset (using the trust-region-reflective algorithm (Coleman and Li, 1996) implemented in Matlab). We do so only in the 2D plane with the same z-value as p, primarily because 3D statistics are less reliable due to the relatively poor z-resolution of confocal microscopy as well as movement of the centrosome while it is imaged. All parameters are free and real-valued so that in the end we have a sub-pixel centre (x(p),y(p)) and angle θ(p) of the principle axis, SD σx(p), σy(p) along the principal and minor axes, background level b(p) and peak height ι(p). Figure 2(a) and (b) show an object height map and the fitted Gaussian, respectively. From this Gaussian, we then have radius μ(p) = 2σ(p), ι(p), b(p), and x(p), y(p) = volume under the Gaussian - b(p).

Occasionally, when centrosomes are very small and near each other, they are recognized as a single object. The signature of these cases are that the ratio r(p)/σ(p) is large and σ(p) is small. In these cases, we try fitting the signal to two Gaussian models and if we have a better residual and both σ-ratios are smaller than r, then we accept the two spots as real and associated Gaussian models.

At the resolution of light microscopy there is no clear boundary between a centrosome and the background. Therefore, the radius of the fitted Gaussian is actually only a measure that is relative to an unknown but fixed scaling constant. Above, we empirically chose this scaling constant to be 2 based on human perception of centrosome size. Note, however, that the value of this constant plays no role when comparing the size of centrosomes relative to each other. Furthermore, while the model above is sufficient for the purposes of computationally inferring centrosome tracks, for the final biological analysis of the data one needs as accurate a measurement of centrosome size as is possible and so we undertake the following two refinements.

In the late stages of a centrosome’s trajectory, when the centrosome is effectively dissolving, the intensity profile of the spot is not Gaussian, typically having a dip in the centre as illustrated in Figure 2c. To more accurately measure the radius of the object in these cases, we assume to first approximation that the illumination of the spot is symmetric with respect to the sub-pixel center and elliptical iso-contours of the fitted Gaussian, and compute the radial intensity profile of the spot as a function of ellipse radius r. That is for each ellipse centered about (x(p),y(p)) with skew σx(p)/σy(p) and angle θ(p), we compute the mean intensity along this ellipse. This is illustrated in Figure 2d along with the radial intensity profile of the fitted Gaussian. The spot radius μ(p) is then estimated as that value for which the radial intensity of the raw image equals the radial intensity of the fitted Gaussian at μ(p).

Finally, we correct for image distortion. Due to the convolution of each light point with the point spread function of the microscope, objects appear bigger in the images than they actually are. An approximation is given by

\[ b(r) = \frac{r}{\text{PSF}(r)} \]

(1) where \( b(r) \) is the width of the point spread function and \( r \) is the radius. Our final estimate is within 1% of the true value when tested against beads with a diameter of 1100 nm injected into the embryo.

2.2.3 The bipartite graph matching paradigm

Let \( V_1 = \{ p_1, p_2, \ldots, p_m \} \) and \( V_2 = \{ q_1, q_2, \ldots, q_n \} \) be the set of detected objects at time point \( t = 1 \ldots n \) and \( V = V_1 \cup V_2 \) be the set of all objects. Note that for each time point the number \( m \) of objects can be different. We define a weighted directed graph \( G = (V, E, w) \) where the set of edges represents all possible matching hypotheses between objects detected at successive time points. There is an edge \((p, q)\) between objects \( p \in V_1 \) and \( q \in V_2 \) if and only if:

1. \( |p-q| < 4d_{\text{dS}} \); \( d_{\text{dS}} \) models the biophysical limit on centrosome velocity (300 nm/s) and also takes into account cell movement within the embryo as well as movement of the entire embryo.

2. If \( p \in V_1 \) and \( q \in V_2 \) then \( r(p) > r_{\text{tend}} \); Biologically the centrosome is always increasing or constant in size with time up to the beginning of its dissolution but due to measurement and estimation error, the computed radius varies. Empirically, we have observed that once the radius exceeds \( r_{\text{tend}} = 700 \) nm (well above the resolution limit), it never becomes smaller than \( r_{\text{tend}} = 600 \text{ nm} \). We call this criterion radius hysteresis and in what follows we require that it not only be true of edges but also of sequences of edges, i.e. once a centrosome becomes larger than \( r_{\text{tend}} \) on a trajectory it does not then become smaller than \( r_{\text{tend}} \) later in the trajectory (Fig. 3).

These two simple conditions eliminate a large number of potential correspondences that are certainly incorrect.

The weight \( w(p,q) \) of an edge of the graph is intended to reflect the chances that \( p \) and \( q \) represent the same centrosome. Biologically we know that centrosomes move in a relatively smooth manner and that facets of their appearance such as their radius and intensity also change gradually. Thus we expect \( p \) and \( q \) to be close together and for their appearance to be similar. We thus chose to define the weight or ‘distance’ between \( p \) and \( q \) as:

\[ w(p,q) = (1 - a)\text{dis}(p,q) + a\text{dis}(p,q) \]

where \( a \) is the degree of smoothness of the object in the image, \( \text{dis}(p,q) \) measures the distance between \( p \) and \( q \), and \( a \) is a scaling factor to ensure that the distance is between 0 and 1.
centrosome. Then our solution in this special case is the set of paths that do not require more structure than this. However, suppose that there were no data, but since we want solutions for mutants as well, we can in general a number of vertex-disjoint paths, each corresponding to a centrosome’s dynamically during the algorithm and so will be described later.

In our case, the set of edges between settings as a way of discovering these false positives. Moreover, as we solve positive objects and we will do so numerous times with different parameter a standard technique to use bipartite matching even when there are false positives in and involve objects smaller than \(r_{\text{est}} \) (600 nm) thus radius hysteresis edges will be removed from the matching hypothesis graph.

where 

\[ d_2(p,q) = \| p - q \|_2 \]

and 

\[ d_2(p,q) = \sum_{\alpha=1}^{d_A} \min_{\alpha} \frac{\| p(q)_\alpha - q(q)_\alpha \|}{\mu_\alpha} \]

where \( d_2 \) is a normalized spatial distance, \( d_2 \) is a normalized relative-appearance distance and \( w \in [0,1] \) is a free parameter (to be selected by the algorithm) that weights the contributions of these two asces of coherence. In our implementation we use two appearance features, \( \varphi(p) \) the radius of \( p \) and \( \varphi(p) \) the total intensity of \( p \). The distances are normalized using the concept of a z-score (http://www.statisticalminds.com/essentials/Standard-Score/Overview.php) wherein the measure is the number of standard deviations above or below the mean value of a distribution of ‘raw’ scores. In our case, the set of edges \( D \) contributing to the distribution is decided dynamically during the algorithm and so will be described later.

A solution \( S \) to the tracking problem is a set of edges in \( G \) that form a number of vertex disjoint-paths, each corresponding to a centrosome’s lifespan. We could add additional constraints on these paths for wild-type data, but since we want solutions for mutants as well, we can in general not require more structure than this. However, suppose that there were no false-positive objects in the graph, that is, every vertex truly represents a centrosome. Then our solution in this special case is the set of paths that involves the maximal number of edges, and if there are several, then we want the one with the set of edges of minimum weight. This is equivalent to finding the best bipartite matching between \( V_t \) and \( V_{t+1} \) for each time \( t \) and we can do so using the Hungarian algorithm (Munkres, 1957). It is a standard technique to use bipartite matching even when there are false positive objects and we will do so numerous times with different parameter settings as a way of discovering these false positives. Moreover, as we solve each bipartite matching for successive value of \( t \), we eliminate any edges between \( V_t \) and \( V_{t+1} \) whose addition to the trajectories up to time \( t \) would result in a trajectory with radius hysteresis (Fig 3).

Finding reliable core sub-trajectories The first step is to find the middle part of each centrosome’s trajectory where the centrosome is large and bright. We refer to this part of a trajectory as the core trajectory. While it is desirable to establish as much of each trajectory as possible, it is more important that the core trajectories be error-free.

For several thresholds \( \alpha \in [0,1] \) we consider the subgraph of \( G \) restricted to the objects that are brighter than \( \alpha \), i.e. \( V(\alpha) = \{ p \in V | \varphi(p) \geq \alpha \} \). Note that the larger the threshold \( \alpha \), the fewer false-positive objects there are in \( V(\alpha) \) until at some value, not yet known to us, practically no false-positives remain. We first apply the bipartite matching heuristic described above with \( \alpha = 0 \) and \( \alpha = 1 \) on \( V(\alpha) \) yielding a solution \( S(\alpha) \) that is the set of edges in the optimal matching at each time point. This solution is solely based on the spatial distances \( d_2 \) between the objects. Given \( S(\alpha) \), we let it be the set of edges \( D \) for computing the means and standard deviations of the spatial and feature differences that then determine normalized scores for \( d_2 \) and \( d_3 \). For a series of different values of \( \alpha \), we then apply the bipartite matching heuristic where the weights of each edge are determined by \( D = S(\alpha) \) and \( a \). We refer to this solution as \( S(a,\alpha) \) and observe that it takes object appearance into account as well.

Our key idea is that if \( S(a,\alpha) = S(a,\alpha) \) or nearly so, and also the larger \( a \), then the more likely it is that the edges in \( S(a,\alpha) \) are all correct. In the ideal case, \( S(\alpha) \) would equal \( S(\alpha) \) for some \( \alpha \). So for a set of choices \( \alpha \in [0,1] \), we compute a table of the number of conflicts between \( S(\alpha) \) and \( S(a,\alpha) \) as \( K(\alpha,a) = | S(a,\alpha) \setminus S(\alpha) | \) and find the smallest value \( \omega(\alpha,a) = \min(K(\alpha,a)) \). Almost always \( \omega(\alpha,a) = 0 \) and there are multiple choices of \( \alpha \) and \( a \) that give the value \( \omega(\alpha,a) \). Among these we select \( \alpha^* \) where \( \alpha^* \) is the largest \( \alpha \) that gives us \( \omega(\alpha,a) \) for some \( \alpha \). \( a^* = \max(\alpha | | \omega(\alpha,a)| = K(\alpha,a) | ) \).

\( \omega^* \) is the lowest possible brightness threshold \( \omega^* \), indicating that among those at blending factor \( \alpha^* \) that give \( K(\alpha,a) \), i.e. \( \omega^* = \min(\alpha) \left| V(\alpha) \setminus V(\alpha^*) \right| \). To conclude, we let the edges in \( \omega = S(a^*,\alpha^*) \) be our core trajectories and in most cases \( \omega^* \geq 8 \). In effect, we have found the level \( \omega^* \) that minimizes the possibility of having false positives in \( V(\omega^*) \) and hence the solution to the bipartite matching heuristic is most likely to be correct. The so established core trajectories include on average 86% ±11% (SD) (\( a = 40 \)) of the centrosome objects of the complete trajectories that we wish to establish.

2.2.5 Extending core trajectories to full trajectories. Given the core trajectory of each centrosome, we now go back to the full set of called objects and extend the trajectories forward and backward in time to obtain the full trajectories. To this end, we apply the same core procedure of our tracking algorithm with edge weights that combine spatial distance and appearance distance as described above with the weighting factor \( \omega = 0.3 \) and \( D = \omega \). The choice of \( \omega \) was empirically chosen and indicates a slight preference towards minimizing distance changes over appearance changes during the extension phase. Moreover, we compute the mean \( \mu_\omega \) and SD \( \sigma_\omega \) of all the edges in the core trajectories and eliminate from \( F = \omega \) all edges whose weight is >4 SDs above the mean, i.e. \( \mu_\omega + 4\sigma_\omega \), as being implausible. \( F \) is the set of all edges as defined in Section 2.2.3.

In the following, we will describe the track extension algorithm for the backward case. Forward extension is analogous. We extend core tracks backward iteratively starting with \( t = n \) and working backward to 1. Suppose we have already iteratively extended backward to time \( t \) and let \( B(t) \) be the set of edges in our current set of tracks where the induction starts with \( B(0) = \emptyset \), the set of core tracks computed in the previous subsection. Note that all edges \( (a,b) \in B(t) \) with \( b \in V_{t-1} \) are core trajectory edges.

To extend to time \( t-1 \) we first find a bipartite matching between \( V_{t-1} \) (i.e. the set of all objects at time \( t-1 \)) and \( U(t) = \{ b \in V_t | \omega(b) \in V_{t-1} \} \) (\( (a,b) \in B(t) \)) (i.e. the set all objects at time \( t \) that are adjacent to an edge in \( B(t) \)) using the Hungarian algorithm with the weights and edges described in the previous paragraph. We remove from consideration any edge whose addition to the established trajectories would result in a trajectory with radius hysteresis.

Suppose the algorithm repeats the set of edges \( H_t \) as the best matching between \( V_{t-1} \) and \( U(t) \). We must reconcile these edges against the core trajectory edges \( P_t = \{ (a,b) \in V_{t-1} \times V_t | (a,b) \in B(t) \} \) that are currently connecting time \( t-1 \) and \( t \). An edge \( e = (a,b) \in H_t \) is compatible with \( B(t) \) if either \( e \in B(t) \) (i.e. \( e \) is also a core edge) or there are no edges adjacent to \( a \) and \( b \) edges between any vertex at \( t-1 \) and \( b \) in \( B(t) \) (i.e. \( a \) is not in any core trajectory and \( b \) is not already matched to another object at \( t-1 \)). There is only one situation in which we will accept non-compatible edges in \( H_t \) as part of the new extension \( B(t+1) \). A pair of edges \( (a,b) \) and \( (c,d) \) in \( H_t \) is an exchange.
We then repeat the stitching process with less strict conditions for (i) trajectories that are shorter than a minimum of 10 minutes on the duration of the centrosome cycle and (ii) trajectories with initial radius >500 nm. For these rare cases of certainly broken trajectories we allow twice the respective maximum difference (condition 2) and relax the radius hysteresis to \( r_{max} = 1000 \) nm and \( r_{cmax} = 550 \) nm (condition 3).

### 2.3 Biological identities of the tracked centrosomes

It is often necessary to know the biological identities of the tracked centrosomes, that is, the identity in the *C. elegans* lineage (Schnabel et al., 1997; Sulston et al., 1983) of the cell it is located in and whether it is the anterior or the posterior associated pole of the spindle. We determine these identities by establishing a binary tree of trajectories representing the pair and parent/child relationships between the tracked centrosomes as follows.

1. For centrosome movies that start at the 1-cell stage the anterior/posterior orientation of the embryo is easily determined as the sperm always enters on the posterior side of the embryo (Goldstein and Hirsh, 1996). Thus, the pole closest to the centrosomes detected in the first frame of the movie is the posterior pole. For movies that start at later cell stages, we additionally take into account the stereotypical timing of the cell divisions to figure out where the posterior pole is. For example, in the 4-cell stage, the cells ABa and ABp divide before EMS, and EMS divides before P2.

2. The centrosomes are paired by finding an optimal bipartite matching between their trajectories. There is an edge between trajectory \( a \) and \( b \) if (i) they overlap in time by at least 6 min, (ii) the distance between the centrosomes is positively correlated with time (i.e. they tend to separate) and (iii) their minimum centrosome separation distance is \(< 12 \) µm. The weight \( w(a, b) \) between \( a \) and \( b \) is the product of (i) the minimum centrosome separation distance in the first half of the trajectories and (ii) the absolute difference in radius averaged over time. Intuitively, we expect a pair to be near each other and for their centrosomes to be at the same growth stage. For the pairs so established, we label the one closer to the posterior pole at the end of its trajectory as the posterior child and the other as the anterior child.

3. In the last step, we determine for each centrosome pair its (common) parent centrosome. In the 1-cell stage, we label the one closer to the posterior pole as the parent and the other as the child. In the 2-cell stage, the cell division is straightforward as the cell is divided in two. For movies that start at later cell stages, we additionally take into account the stereotypical timing of the cell divisions to figure out where the posterior pole is. Intuitively, we expect a pair to be near each other and for their centrosomes to be at the same growth stage. For the pairs so established, we label the one closer to the posterior pole as the parent and the other as the child.

Given the invariant cell lineage of *C. elegans* and its canonical naming scheme with respect to anterior/posterior localization, we then only need to traverse the resulting tree level by level to assign a cell name to each centrosome.

### 2.4 Detection of the nuclear envelope breakdown

In order to compare the time series of centrosome statistics with each other we need to register them in time. It is common practice in biology to define such a registered time axis relative to a cell cycle event. We use nuclear envelope break down (NEBD) (Gojanac et al., 2007) because it can be observed in our images without an additional marker. The nucleus is free of centrosomal proteins and so appears as a dark, circular region in the cytoplasm. As the nucleus breaks down, there is an inflow of GFP-labeled molecules from the cytoplasm into the nuclear region and its fluorescent intensity increases. The NEBD is the time point for which the rate of this inflow of luminosity is maximal.

As the dark nuclear region is difficult to delineate based purely on signal, we take advantage of prior knowledge. The nucleus is located roughly at the center of the embryo and its fluorescent intensity increases. The NEBD is the time point for which the rate of this inflow of luminosity is maximal.

![Diagram of an exchange during the backward extension phase](https://example.com/diagram.png)

**Fig. 4.** Illustration of an exchange during the backward extension phase (the edges are drawn in the opposite direction only to illustrate that we work backward in time). The core trajectory edge \((a, d) \in B(t)\) is replaced by the exchange edges \((a, b) \in H_t\) and \((c, d) \in H_t\).
We have analyzed over 500 centrosome movies that cover different errors. We use the following error categories that cover the space of all observed errors. For example, a completely missed centrosome is a more severe tracking error than a trajectory that misses only a couple of frames at the very beginning of the cycle or diffuse, disassembled centrosomes which are a little dim spot. We estimate the intensity of the nucleus as the average of the darkest half of the pixels in this circle. Plotting this value through time yields a curve illustrated in Figure 5. We first find the time of the maximum intensity of the curve and then set as the NEBD the time of the maximum rate of change of the intensity in the preceding 5 min.

3 EVALUATION

3.1 Tracking performance

We have analyzed over 500 centrosome movies that cover different stages of embryonic development between the 1- and the 16-cell stage. To quantify the tracking performance, we randomly selected 10 multi-cell stage movies for visual inspection. Due to the small number of movies that contain the 8- and 16-cell stage in our corpus, we additionally inspected three random examples of such movies. Many papers on tracking systems including (Bao et al., 2006) identify three types of errors: failure to incorporate an object, incorporation of a false-positive object and incorporation of an incorrect object in a given track. We use a finer error categorization because some kinds of errors are more severe than others. For example, a completely missed centrosome is a more severe tracking error than a trajectory that misses only a couple of frames at the beginning of the cycle when the centrosome is a tiny dim spot. We use the following error categories that cover the space of all observed errors.

• **Short track.** The track is correct but misses a few frames at either end.
• **Long track.** The track is correct but has a few extra frames at either end.
• **Broken track.** A centrosome is completely tracked but the trajectory is broken into two or more pieces. Stitching the involved tracks would correct the problem.
• **Fused track.** Two or more centrosomes are in a track but splitting the track would correct the problem.

The tracking results of randomly selected centrosome movies starting at different cell stages were analyzed.

![Fig. 6. Typical time versus nuclear intensity curve computed by the nuclear envelope breakdown detection algorithm.](image)

| Cell stage no. of tracks | Minor errors | Major errors |
|--------------------------|--------------|--------------|
| 1                        | 8            | 0 0 0 0 0 0 0 0 |
| 2                        | 32           | 0 3.1 0 0 0 0 0 0 |
| 4                        | 48           | 10.4 2.1 0 0 0 0 0 |
| 8                        | 48           | 10.4 2.1 0 0 0 0 0 0 |
| 16                       | 64           | 17.2 3.1 1.6 0 1.6 3.1 0 |

Table 1. Minor and major tracking error rates in percent of the total number of tracks

The tracking results of randomly selected centrosome movies starting at different cell stages were analyzed.

• **Chimera track.** The track contains more than one centrosome or non-centrosome object and it is not a long or fused track.
• **Missing track.** A centrosome is not tracked at all.
• **Particle track.** A non-centrosome object is tracked but involves no centrosomes.

Table 1 summarizes the evaluation results. Tracks that are a few frames too short are the most prevalent error but it is also the least severe because it only involves faint, tiny centrosomes at the very beginning of the cycle or diffuse, disassembled centrosomes at the very end of the cycle. Fusion errors and particle tracks never occurred. More severe errors occurred only in the 16-cell stage. For example, two missed centrosomes were due to their being so deep in the stack, and hence so dim, that they were not picked up in the core trajectory extraction step.

3.2 Accuracy of NEBD detection

To evaluate the performance of our automatic NEBD detection method, we manually determined the NEBD for 30 randomly selected movies containing a total of 95 cells and the difference in frames to that determined automatically. During manual NEBD detection, we experienced that it was often ambiguous as to the exact frame in which NEBD occurred. Therefore, we considered automatic NEBD detection to be correct if it was within one frame of the manually determined NEBD. We measured exact agreement for 64% and agreement within one frame for 34% of the cases, yielding 98% accuracy for NEBD detection.

3.3 Application to the study of centrosome size

To validate the utility of our system for studying centrosome size, we checked if it could detect previously described phenotypes. As an example, TAC-1 is a protein essential for microtubule assembly in vivo (Bellanger et al., 2003; Srayko et al., 2003). Depletion of TAC-1 causes centrosomes to get less strongly pulled apart during disassembly due to reduced microtubule forces. We knocked down TAC-1 by RNAi and plot in Figure 6(a) the mean centrosome radius with time versus wild type. As expected, centrosomes in tac-1(RNAi) embryos (red curve) are eventually smaller than in wild-type embryos (black curve). Moreover, spindles are shorter compared to wild type embryos (Fig. 6(b)) confirming previous results by (Srayko et al., 2003). Our time-resolved data additionally shows that TAC-1 has no significant effect on centrosome size before centrosome disassembly begins at metaphase (∼150 s after NEBD).
AIR-1. It is required for centrosome maturation and also involved in spindle assembly (Hannak et al., 2001). Partial depletion of AIR-1 leads to significantly smaller centrosomes compared to wild type (Fig. 7(a) and after nuclear envelope breakdown the centrosome radius over time in one-cell stage embryos (error bars indicate SEM). (b) Spindle length is plotted as the distance between centrosomes over time for the same embryos as in (a) (error bars indicate SEM).

A protein that actually affects the size of centrosomes is the kinase AIR-1. It is required for centrosome maturation and also involved in spindle assembly (Hannak et al., 2001). Partial depletion of AIR-1 leads to significantly smaller centrosomes compared to wild type [Fig. 7(a)] and after nuclear envelope breakdown the centrosome separation cannot be maintained [Fig. 7(b)].

4 DISCUSSION

We have described a system for tracking and precisely quantifying the size of GFP-labeled centrosomes in early C.elegans embryos. The strengths of our system are its low error rate and its ability to automatically extract and visualize all the data that is necessary to convert raw movies into aligned and correctly labeled time-series of centrosome life cycles ready for biological interpretation. This facilitates quantitative studies of centrosomes to an extent and precision that would not be possible with manual image analysis.

The robustness of our tracking algorithm relies on three things. First, in our layered approach we initially focus on sub-problems that can be solved with very high confidence. These reliable partial solutions are then incrementally refined until the overall tracking problem is solved. However, while the general idea of a layered approach could also be applied to other tracking problems, our concrete implementation is tailored to the problem of centrosome tracking. Second, we make extensive use of prior knowledge but rely only on properties that are also valid for mutants. Third, while our detection algorithm produces a number of false positive object calls, false negative calls are extremely rare and occur only at the very beginning and end of the centrosome cycle.

So far, we have successfully processed over 500 centrosome movies with minimal manual corrections. The most challenging movies were the ones starting at the end of the 4-cell stage / beginning of the 8-cell stage and ending somewhere during the 16-cell stage. In these movies, it was difficult to figure out the identities of the initial centrosomes (the ones with their parent centrosomes not in the movie) and the parent/child relations between the 8- and the 16-cell stage centrosomes.

In our ongoing work we have quantified the dynamics of centrosome maturation over development from the 1- to the 8-cell stage in wild type and mutant embryos and identified several genes that affect this process. We are currently investigating whether one of these genes has a key role in controlling centrosome size. Moreover, in an initial study, our system was also able to track centrosomes in human tissue cultures cells and, with some modifications, nuclei in early C.elegans embryos.

ACKNOWLEDGEMENTS

We greatly acknowledge the initial work on centrosome tracking in 1-cell stage embryos by Jacques Pecreaux. We also would like to thank Horatiu Fantana for injecting beads into embryos as well as Fujiu Long, Nathan Clark and Hanchuan Peng for helpful discussions on the tracking algorithm.

Conflict of Interest: none declared.

REFERENCES

Alberts, B. et al. (1994) Molecular Biology of the Cell, 4th edn., Garland Publishing, NY, 1030–1032.

Bao, Z. et al. (2006) Automated cell lineage tracing in Caenorhabditis elegans. Proc. Natl Acad. Sci., 103, 2707–2712.

Bellanger, J.M. and Gonczy, P. (2003) TAC-1 and ZYG-9 form a complex that promotes microtubule assembly in C.elegans embryos. Curr. Biol., 13, 1488–1494.

Coleman, T.F. and Muschel, R. (2006) The centrosome and the DNA damage induced checkpoint. Cancer Lett., 243, 1–8.

Goldstein, B. and Hid, S.N. (1996) Specification of the anteroposterior axis in Caenorhabditis elegans. Development, 122, 1467–1474.

Gorjanacz, M. et al. (2007) What can Caenorhabditis elegans tell us about the nuclear envelope? FEBS Lett., 581, 15, 2794–2801.

Greeman, G. et al. (2010) Centrosome size sets mitotic spindle length in Caenorhabditis elegans embryos. Curr. Biol., 20, 353–358.

Hannak, E. et al. (2001) Aurora-A kinase is required for centrosome maturation in Caenorhabditis elegans. J. Cell Biol., 145, 1109–1116.

Hazalka, R.M. and Shapiro, L.G. (1992) Computer and Robot Vision (Volume II), Addison-Wesley, Boston, MA, USA, 316–317.

Kulinski, Y. (2007) Intracellular objects tracking. Eur. J. Cell Biol., 86, 560–578.

Korelic, R.A. et al. (2009) Automated 3-D tracking of centrosomes in sequences of confocal image stacks. Conf Proc IEEE Eng Med Biol Soc., EMBC 2009 Annual International Conference of the IEEE, 6994–6997.
Kirkham, M. et al. (2003) SAS-4 is a C. elegans centrosomal protein that controls centrosome size. Cell, 112, 575–587.

Meyer, F. (1994) Topographic distance and watershed lines. Mathematical morphology and its applications to signal processing, 38, 113–125.

Munkres, J. (1957) Algorithms for the Assignment and Transportation Problems. Journal of the Society of Industrial and Applied Mathematics, 5, 32–38.

Otsu, N. (1979) A threshold selection method from grey level histograms. IEEE Trans. Syst. Man Cyber., 9, 62–66.

Sage, D. et al. (2005) Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. IEEE Trans. Image Proc., 14, 1372–1383.

Song, M. H. et al. (2008) The conserved protein SZY-20 opposes the PKA-Related kinase ZYG-1 to limit centrosome size. Dev. Cell, 18, 901–912.

Schabes, R. and Priess, J. R. (1997) Specification of the early embryonic cell fates in C. elegans. In (D. Riddle, T. Blumenthal, B. J. Meyer and J. Priess) (eds.). The Nematode C. elegans, Vol. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sulston, J. E. et al. (1983) The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol., 100, 64–119.

Szymkowiak, M. et al. (2003) Caenorhabditis elegans TAC-1 and ZYG-9 form a complex that is essential for long astral and spindle microtubules. Curr. Biol., 13, 1506–1511.

Thomann, D. M. (2003) Algorithms for detection and tracking of objects with super-resolution in 3D Fluorescence Microscopy. PhD Thesis, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland, pp. 12 and 58.