

Chromatin Decondensation: a Case Study of Tracking Features in Confocal Data

Wim de Leeuw*

Robert van Liere *

*Center for Mathematics and Computer Science, CWI, Amsterdam, Netherlands.

Abstract

In this case study we discuss an interactive feature tracking system and its use for the analysis of chromatin decondensation. Features are described as points in a multidimensional attribute space. Distances between points are used as a measure for feature correspondence. Users can interactively experiment with the correspondence measure in order to gain insight in chromatin movement. In addition, by defining time as an attribute, tracking problems related to noisy confocal data can be circumvented.

CR Categories and Subject Descriptors: I.3.3 [Computer Graphics]: Picture/Image Generation; I.3.6 [Computer Graphics]: Methodology and Techniques

Keywords: feature tracking, multidimensional visualization, biomedical imaging.

1 Introduction

In the quest for understanding biological processes that underlie control of gene expression, there is a strong need for methods to study the structural and functional organization of the cell nucleus. Recent progress in the luminescent labeling of cell components and the use of digital 3D microscopy allow biologists to generate time dependent volume data describing in detail specific processes of the living cell. Due to their complex nature, four dimensional structural analyses is difficult, if not impossible, using traditional analysis techniques.

In a previous case study we used standard visualization techniques, such as volume rendering and iso-surfaces, in a virtual environment for the exploration of the data [1]. From that study it became apparent that higher level visualization techniques are needed for better understanding of these processes.

Feature visualization is an attempt to present data to the user at a high level. The problem of feature tracking is the detection of features and the tracking features over time. There are two main difficulties of tracking luminescent labeled cell components in data taken from confocal microscopes. First, the definition of cell components in these data sets is not straightforward. This effects the detection of features. Second, the data is very noisy which interferes with detection and tracking.

In this paper we discuss an interactive feature tracking system which we use to analyze chromatin decondensation. The feature tracking algorithm resembles the method introduced by Sethi et al. in which a feature is described as a point in a multidimensional attribute space [2]. Distances between points are used as a correspondence measure between features. Our method can interactively scale attributes in order to assign relative weights to attributes. In contrast to other tracking methods, our method considers time as

an attribute of a feature. The main motivation of considering time as an attribute is that all information of the feature is treated in an uniform way. In this way, the problem of noisy data which exists in frame based feature tracking methods can be circumvented.

The paper is organized as follows. First, we discuss the biological background of the case study. After reviewing related work, we discuss our methods in some detail. Then, in section 5, we present our case. Finally, in section 6, we discuss the pros and cons of our approach.

2 Biological background

The regulation of gene expression at the level of DNA and regulatory proteins that bind to it is understood in quite some detail. However, the role of higher order chromatin folding in the interphase nucleus is less well understood. Using state of the art fluorescent DNA/chromatin labeling techniques in combination with confocal microscopy [3], biologists are able to produce time dependent 3D data sets of DNA/chromatin in the cell nucleus in living cells.

To this end, cells were used that express green fluorescent protein [4] tagged histone H2B¹. Histones are key components of chromatin. They constitute a protein octamere, around which the DNA is wrapped almost two full turns. This histone octamere plus the DNA together constitutes the nucleosome, the basic building block of chromatin in all eukaryotic cells. In the experiments discussed in this paper the GFP-labeled H2B allowed that chromatin was visualized in living cells and movement was followed in time and space using 3D confocal microscopy. Time series of 3D images were made of the process of decondensation of the chromatin after cell division (mitosis). During mitosis chromatin is densely packed in chromosomes. After mitosis, part of the chromatin decondensates to form a new nucleus.

The aim of these experiments was to analyze the movement of chromatin during formation of the cell nucleus of the newly formed daughter cell. Densely packed areas of chromatin are used to analyse the movement of the entire chromatin. The densely packed areas are represented as high intensity levels in the data. Hence, the goal of the feature tracking algorithm is to track positions of the high intensity levels in the data.

The data set consisted of a series of 134 3D data sets. Each data set is a stack of optical sections of 256×256 voxels. The number of optical sections in the stacks varied over time due to the flattening of the cell as the cell cycle progresses. The initial number of optical sections was 30, decreasing to 18 sections at later time points. To make processing easier extra optical sections were added, so that all sets contained an equal number of sections. Due to physical characteristics of a confocal microscope the optical resolution along the z-axis is four times less than in the x-y plane. The 3D images are corrected for this by scaling in the z-direction during rendering. The total amount of storage required for the data is 260 MB.

*CWI, Department of Software Engineering, P.O. Box 94097, 1090 GB Amsterdam, Netherlands. E-mail wimc@cwi.nl

¹HeLa cells expressing H2B-GFP[5] were kindly provided by H. Kimura of the University of Oxford

3 Related work

Detecting and tracking features in time dependent data has been studied by many researchers. Most algorithms consider feature tracking as a two step process. First, features are detected and extracted from each time step in the data. Then, in the tracking phase, the correspondence of features in successive time steps is used for the determination of tracks.

Sethi et al. compute tracks of objects in image sequences [2]. An object is characterized as a vector of attributes in a multidimensional attribute space. The distance in the attribute space is used as a correspondence measure between two objects. Tracks are constructed by maximizing a smoothness function over the set of possible feature sequences. The smoothness function takes both speed and angle into account. Phantom tokens are inserted if the number of detected features does not match between frames in the sequence.

Sametey et al. introduced the notion of evolutionary events, such as split, merge, birth and death events to describe the evolution of features [6]. Correspondence between features is determined by thresholds for each of the attributes. Additional correspondence criteria are formulated in case of evolutionary events. For example, in the case of track splitting, the size attribute of the feature is compared to the sum of the size attributes of the split features.

Reinders and Post use prediction methods for the evaluation of feature correspondences [7]. Gradients of attribute values in already found tracks are used as a prediction for the continuation of the track.

Instead of attribute correspondence between frames, Weigle and Banks take a different approach for finding features in time dependent flow data [8]. Here features are extracted directly from the four dimensional data. They calculated iso-values of expressions describing the feature directly in the four dimensional data.

Our approach resembles other methods in which features are expressed as points in a attribute space. It resembles Sethi's method in that correspondence is expressed as the euclidean distance between points. However, our approach differs in that it treats time as an attribute of the feature. Feature correspondence is based only on distances between points.

4 Methods

In this section we review the methods used for the tracking and visualization of features in a time dependent confocal data set. First, features are extracted from the each time step in the data set. A feature consisting of n attributes is described as an n -dimensional attribute vector, i.e. $F_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$. In our method, we reserve the last attribute a_{in} to denote the time step of the feature.

4.1 Tracking

Each feature is represented as a point in a multidimensional attribute space. Scaled attribute differences are used to define a distance function

$$S(F_i, F_j) = \sqrt{\sum_k (\lambda_k (a_{ik} - a_{jk}))^2} \quad (1)$$

in which λ_k is a user defined scaling factor for each attribute. The distance function S is used as a measure for the correspondence of features.

Two features are linked if their distance is smaller than a user defined threshold; i.e. $S(F_i, F_j) < T$ in which T is the threshold value. A track is the set of features which are linked.

Figure 1 illustrates the tracking algorithm in a two dimensional attribute space. The horizontal axis represents the time attribute,

and the vertical axis represents an intensity attribute. In the left image, 21 features (points) are shown. Dashed circles are drawn around two features to indicate the distance threshold. In the right image, lines that link points are shown to indicate three tracks.

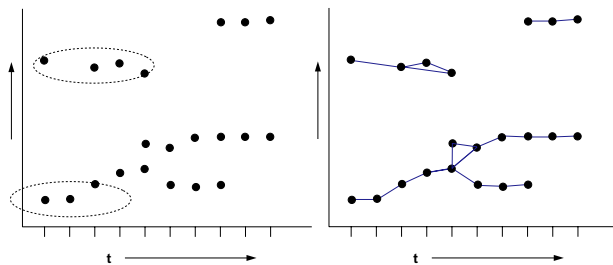


Figure 1: Illustration of the tracking algorithm. In the left image, dashed circles indicate linking criteria. Resulting tracks are shown on the right.

A number of remarks about the tracking algorithm can be made:

- The distances between points in attribute space are based on attributes of various types. There is no fixed way to compare differences in attributes of different types. For example, how does a certain change in intensity compare to a change in the position of a feature. Therefore, the attribute vector is scaled using the user controllable scaling vector, $\lambda = \{\lambda_1, \dots, \lambda_n\}$. In this way, the user can assign a relative weight to each attribute in the attribute vector.
- Time is not regarded as a special attribute and is treated as other attributes. Considering time as an attribute has the advantage that tracks can continue even if a feature is not found in a certain time step. This is often the case in noisy data.
- The distance criterion used can be geometrically interpreted as an hyper-ellipsoid in attribute space. All features within the hyper-ellipsoid are linked to the feature, including those features in the same time step. The user can, based on the value of the scaling vector and knowledge of the underlying data, judge if this is desirable or if λ -values have to be adjusted.
- The computation and storage of mutual distances of features is costly. For the storage of the mutual distance of N features, N^2 storage is required. The computation and storage is optimized by introducing a one dimensional sliding window of size T . Only those distances between features will be computed if the difference between the attribute value is in the sliding window; i.e. $\|a_{in} - a_{jn}\| < T$.

4.2 Visualization and Interaction

Figure 2 shows a schematic view of the tracks found in figure 1 (see also Reinders et. al [7]). The horizontal axis represents time, and the vertical axis represents feature. Tracks are shown as lines connecting features. The vertical ordering of features is chosen to minimize the number of edge crossings.

The track viewer can be linked to other data viewers. In our system we use a 3D viewer to show the positions of the tracks in the cell nucleus.

The track viewer is useful for the analysis of track structure. Splits and joins of tracks can easily be seen in the track viewer. Also, aspects of the lifetime of a track can be discerned in the track viewer, while these cannot be seen in the 3D viewer. For example, a track of a stationary feature is represented as a long line in the track

view while it is shown as a single point in the 3D viewer. Also, a fast moving feature with a short lifetime is a short line in the track viewer while it is a long line the 3D view.

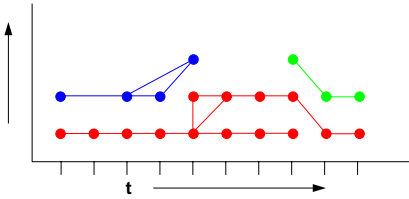


Figure 2: Track viewer showing the tracks in figure 1.

There are various possibilities for user interaction:

- The scaling vector $\lambda = \{\lambda_1, \dots, \lambda_n\}$ and distance threshold T can be adjusted interactively. Interactively changing the scaling vector results in re-shaping the hyper-ellipsoid. This allows the user to experiment with determining optimal feature correspondences for the given data.
- Interactive brushing allows the user to manually inspect the relation between tracks in the track viewer and three dimensional viewer.
- Interactive selection of features and tracks allows the user to simplify views and highlight tracks.

5 Results

In section 2, we discussed how the time dependent data set was acquired from the confocal microscope. To study internal nuclear motion, the movement of the most dense regions in the chromatin is analyzed. Features were used to represent these dense regions. Features are defined by local maximal intensities in the data. Each feature consisted of a 5 dimensional attribute vector, $\langle x, y, z, i, t \rangle$, in which x, y, z are the feature's 3D position, i denotes its intensity value and t is the time step. In total, 10,256 features were found.

Figure 3 and 4 give an overview of the data. The track view (Figure 3) shows all computed tracks for a particular scaling vector and distance threshold. The three dimensional viewer (Figure 4) shows the positions of the tracks in physical space. Only tracks with a length larger than 20 points are shown. Colors are used to label the tracks in both views.

Interpretation of the track view gives the following results:

- For the first 60 time steps the number of features and tracks gradually grows, and then fluctuates somewhat. The biological justification for this is that densely packed chromatin has a relatively homogeneous density, while large variations of chromatin density occur in decondensated chromatin.
- Sharp edges in the tracks are visible in the track view. These edges are due to the birth or death of many tracks in a time step. This occurs when the confocal microscope is recalibrated when scanning the cell.
- Despite the increasing number of tracks in time, track splitting is rare. A possible biological justification for this is that specific regions in the chromatin stay dense, while the chromatin in the neighborhood decondensates.

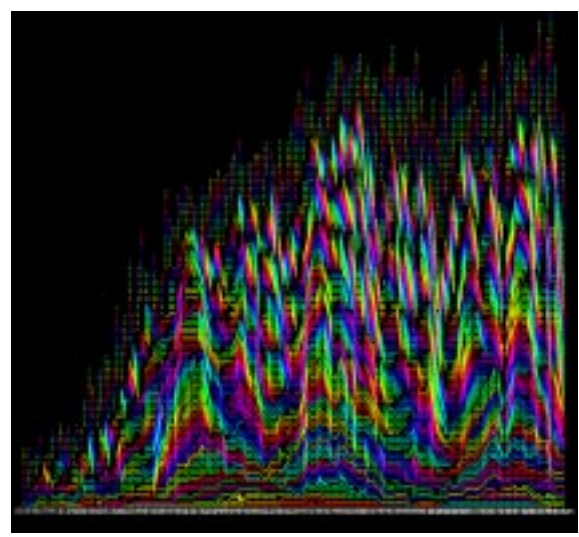


Figure 3: Track viewer showing the tracks found in the confocal microscope data.

Figure 5 shows the track view with a different threshold distance. In this case, tracks have many more branches. Although this view illustrates the interactive capabilities of the method, the found feature correspondences were not optimal given what the user knows about the process of chromatin decondensation.

Figure 6 is a zoomed in view of the track view. Some tracks continue although no corresponding features at given time steps can be found. For example, for the third track from the bottom no corresponding features are found at two time steps. The reason for not finding features is probably due to noise in the data. The reason for continuing the track is that the similarity of the attributes is such that, despite the time difference, the points are still in the hyper-ellipsoid.

6 Discussion

In our method the decision whether features belong to a track is based only on the distance between points; i.e. if points are in the vicinity of each other in attribute space. We believe that the main advantage of this approach is its simplicity. Feature attributes – including time – are treated uniformly and the method does not rely on phantom tokens (e.g. Sethi) or evolutionary events (e.g. Samtaney) in order to define tracks.

One of the main problems with the confocal data set is noise. Features are difficult to detect robustly in the data sets, resulting in inaccurate data for the tracking algorithm. For example, a feature may disappear for one or more time steps and then reappear. Since time is treated as an attribute of the feature, the user can adjust the λ value for the time attribute and in this way lower the threshold for time differences.

Interaction is important because no information of the chromatin movement (i.e. tracks in the data) is available. Only experimentation with the scaling vector can give insight to these movements. The track viewer and the three dimensional view has proven a valuable exploration aid for manually changing attribute weights. Users can inspect specific tracks in the track viewer and try to scale attributes in such a way that could, for example, increase the length a track. Based on the knowledge of the underlying problem, the user can interpret the views and judge if one scale vector is better than a different vector.

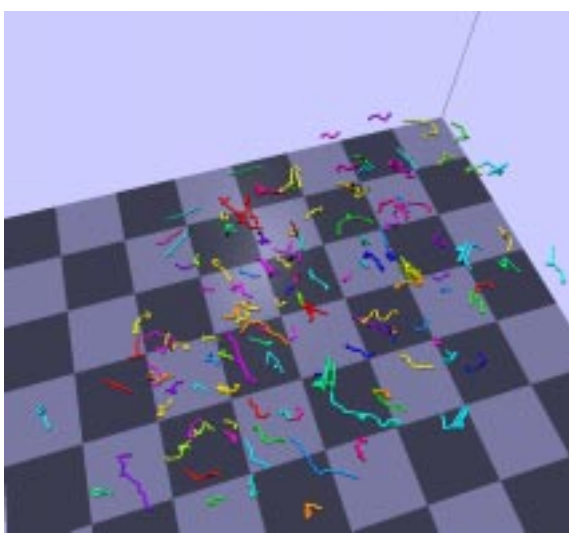


Figure 4: Three dimensional view showing the positions of features and tracks in the confocal microscope data.

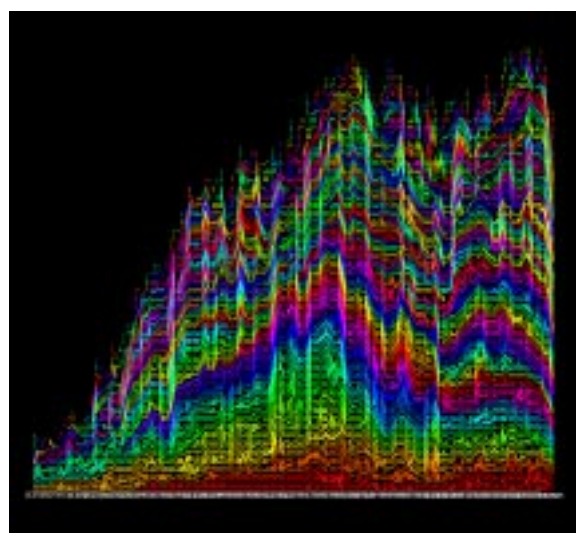


Figure 5: Track viewer showing tracks with a different threshold distance.

7 Conclusions

Using this system, our users were able to analyse the movement of chromatin during decondensation. We have discussed the implementation of an interactive feature tracking system. Features are described as points in a multidimensional attribute space. Distances between points are used as a correspondence measure. Users can interactively scale attributes in order to assign relative weights to attributes. By defining time as an attribute, tracking problems related to noisy confocal data can be circumvented.

In the future we will apply the system to other data sets. Initial studies have shown that the tracking and visualization methods can also be used to track the evolution of critical points in turbulent flow. In addition, we will study techniques that will automatically compute an optimal set of weights in order to find certain properties of the data.

References

- [1] W.C. de Leeuw, R. van Liere, P.J. Verschure, A.E. Visser, E.M.M. Manders, and R. van Driel. Visualization of time dependent confocal microscopy data. In Tomas Ertl, Bernd Hamann, and Amitabh Varshney, editors, *Proceedings IEEE Visualization 2000*, pages 473–477, Los Alamitos (CA), 2000. IEEE Computer Society Press.
- [2] I.K. Sethi and R. Jain. Finding trajectories of feature points in a monocular image sequence. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(1):56–73, 1994.
- [3] R. Rizzuto, W. Carrington, and R.A. Tuft. Digital imaging microscopy of living cells. *Trends in Cell Biology*, 8:288–292, 1998.
- [4] R.Y. Tsien. The green fluorescent protein. *Annual Review of Biochemistry*, 67:509–544, 1998.
- [5] T. Kanda K.F. Sullivan and G.M. Wahl. Histone-gfp fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cell. *Current Biology*, 7(8):377–385, March 1998.
- [6] R. Samtaney, D. Silver, N. Zabusky, and J. Cao. Visualizing features and tracking their evolution. *IEEE Computer*, 27(7):20–27, 1994.
- [7] F. Reinders, F. Post, and H. Spoelder. Attribute-based feature tracking. In E. Groller, H. Loffelmann, and W. Ribarsky, editors, *Data Visualization '99*, pages 63–72. Springer-Verlag, 1999.
- [8] C. Weigle and D. Banks. Extracting iso-valued features in 4-dimensional scalar fields. In B. Lorensen and R. Yagel, editors, *Proceedings IEEE Symposium of Volume Visualization '98*, pages 103–110, 1998.

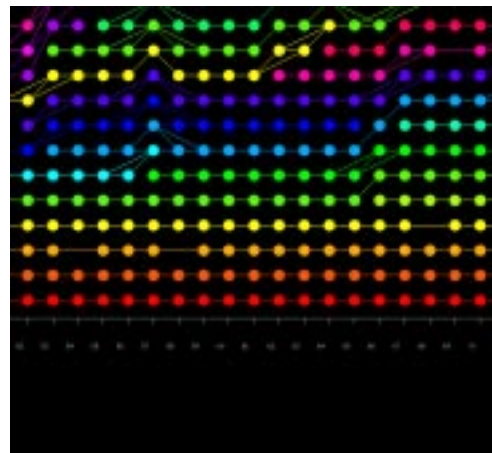


Figure 6: Track viewer showing details of tracks with missing features.