Navigation and Exploration System for Stem Cell Image Data

Category: Research

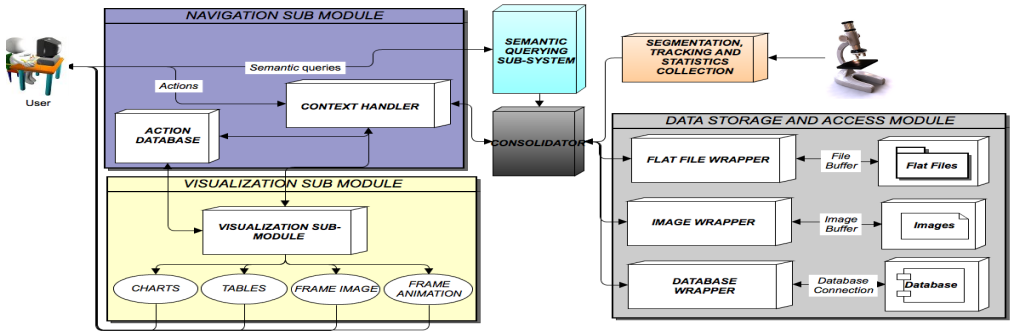

Figure 1. Block diagram of the system

1 Introduction

Cellular biology deals with studying the behaviour of cells. Current time-lapse imaging microscopes help us capture the progress of experiments at intervals that allow for understanding of the dynamic and kinematic behavior of the cells. On the other hand, these devices generate such massive amounts of data (250GB of data per experiment) that manual sieving of data to identify interesting patterns becomes virtually impossible. Our end-to-end system takes the available data, processes it and presents it to neurobiologists who frame and test hypothesis. The system design is motivated by many different yet unanswered exploratory questions pursued by our neurobiologist collaborators.

Some of those exploratory questions span the width of cellular biological experiments. For example, In stem cell research itself, various different types of pluripotent cells change via the process of mitosis and metabolism, and finally become specific cellular tissue (like neurons or skin cells), tracking back a cell to its parent cells several generation before is beneficial in analyzing propagation of various traits and abnormalities in the cells.

Our system is an effort in this direction, to analyze the image data from cellular experiments to give analytics tools to visualize and explore the data. In the next section we highlight our contribution to this system.

2 Our Contributions

We present an end-to-end system to automate different stages of the data capturing, management and usage in the context of experiments on human neural stem cells (hNSC). The complete set of contribution of the system is presented in our work [2] and [1]. Our system includes an image processing system to analyze the images and extract geometric and statistical features within and between stem cell images, a data management system to manage and handle queries on different types of data, a visual analytic system to navigate through and visualize the data and a visual query system to explore different relationships and correlations between the parameters. The main contributions are:

- *Segmentation and Tracking:* We present novel algorithms for image processing for cell identification, segmentation, and tracking. Our algorithm can also robustly find the boundary of the cells that enables accurate computation of other statistical parameters like area of the cell. Difference of gaussian convolution followed by a Watershed algorithm segments the cells robustly and a Hungarian bipartite graph matching solution on weighted graphs having cells as nodes and edges between cells of adjacent images can track the cells accurately. Our algorithms are also capable of detecting interesting events in cellular life processes like mitosis and cell death.

- *Data Management:* We propose a hybrid data representation, storage, and management technique that can handle statistical data, image data and semantic data. Our data management technique is specifically designed to provide fast query processing, efficient data integration for visualization, navigation, and exploration. In this, we have data organized as experiments, frames, cells and events such that each experiment can have multiple frames, each frame can have multiple cell and each cell can have multiple events. We also store the raw images and image cell masks that give pixel level segmentation information.

- *Navigation and Visualization:* We provide data navigation techniques that take semantics of the data into account and enable the users with contextual responses and navigation. The visualization and navigation techniques involve the user in a tight feedback loop using context menus, tool-tips and hyperlinked charts.

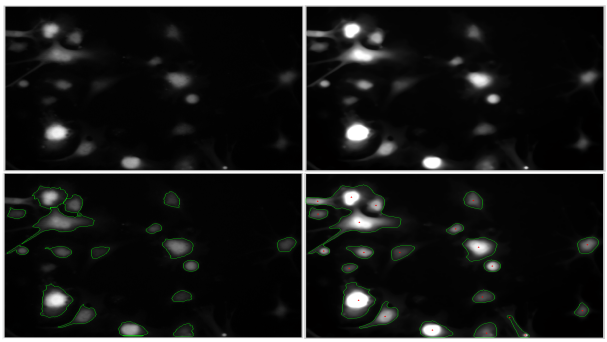

Figure 3. Segmentation: Clockwise from Top-Left: 1. Original Image, 2. Image after DoG contrast enhancement, 3. Cell centers and cell regions (after some morphological operations), 4. Final image after watershed.

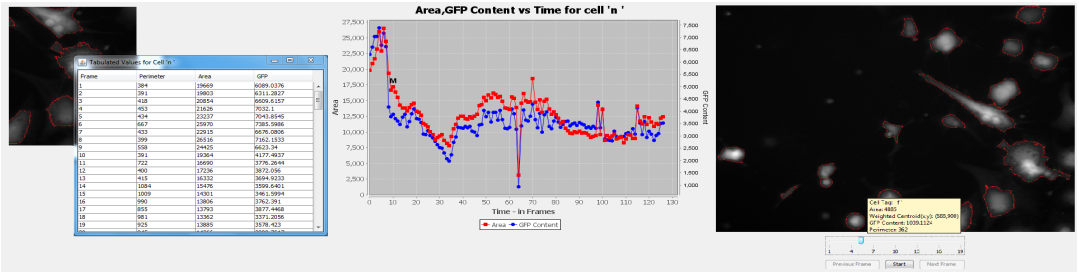

Figure 2. Visualization Entities: Tabular representation, Charts (showing how area and GFP content of a cell changes over time and a Mitosis event labelled M), Frame animation

- *Semantic Querying:* We introduce a new querying methodology that is designed based on the objects and attributes that are used in our application and use set operations to explore patterns and correlations in the data. Many biologically relevant queries can be converted to basic abstract query functions which are a powerful tool for hypothesis framing and testing by the biologists.

3 Evaluation of our System

The system helps neurobiologists perform experiments and analyze the results. The proposed visualization and querying methods help them derive higher level biologically relevant knowledge from experiments. For example, one biologically important statistical pattern that the biologists observed using our system is that when a cell is about to undergo mitosis, it contracts, loses its branches and it's protein content aggregates. It can be seen from the chart in Figure 2, that at the time of mitosis (marked by M), the area of the cell shrinks rapidly. From our system, the biologists also noticed the decrease in cell metabolism in terms of GFP content and production just before, during and just after mitosis, as seen in the chart in Figure 2. Our collaborators could assess the time required for the the rate of GFP production to return to normal, accurately. Such discoveries of cellular behavior will be later used in other experiments to control cell divisions, mobility, metabolism and other processes.

4 Future Work

The modular design of our system allows for it to be modified in multiple ways to achieve scalability in functionality.

The *segmentation and tracking* module could be changed to allow identification of different types of cells. From the data provided by the *segmentation and tracking* module, the visual analytics system can be extended to provide graphs that allow the user to see how different cells change thus allowing them to perform fate analysis.

The system can be extended to handle 3D+Time data where the images are captured at varying depths of substrate (z-planes) by confocal microscopes. These systems are implemented in some commercially available software such as Imaris. However, the detection of cells is done as blob detection in these systems, while our system uses a holistic approach that incorporates local features which can only improve in accuracy and performance as data from one more spatial dimension, namely the z-plane, is used for segmentation. Biologists capture various vital information regarding the conditions for proliferation of the stem cells such as volume and type of drugs introduced, CO_2 levels, etc. This information can be juxtaposed with the findings of our tracking results which helps convey a comprehensive model of the experiment. Data exploration by domain experts involves them making annotations regarding findings in the experiment. Our system can be enhanced to allow recording and consolidations of such annotatory notes that can lead to knowledge discovery.

We can also extend the existing system to allow recognition of cells other than the stem cells. The visual analytics system as it exists currently can take in such multi-cellular data and save them as different experiments. Due to our modular design, our system can be used in a different lab setup with very minor changes to only the Segmentation and Tracking parameters. Tracking can be made more robust by allowing for user interaction in the process of associating cells between frames.

The current query system can be extended to provide higher level queries. A more formal representation of the query language can allow for a wider range of biological queries to be answered. Further, we are also working on visual query system that is as powerful as a textual query system.

References

[1] Kulkarni, I. and Misrty, S. and Cummings, B. and Gopi, M. A visual navigation system for querying neural stem cell imaging data. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, 2011.

[2] Kulkarni, I. and Mukherjee, U. and Sontag, C. and Cummings, B. and Gopi, M. Robust segmentation and tracking of generic shapes of neurostem cells. In *Proceedings of the IEEE Symposium on Healthcare Informatics, Imaging, and Systems Biology*, 2011.