



Universiteit
Leiden
The Netherlands

Biological model representation and analysis

Cao, L.

Citation

Cao, L. (2014, November 20). *Biological model representation and analysis*. Retrieved from <https://hdl.handle.net/1887/29754>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/29754>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/29754> holds various files of this Leiden University dissertation

Author: Cao, Lu

Title: Biological model representation and analysis

Issue Date: 2014-11-20

Summary

Bioinformatics is an interdisciplinary field in which knowledge is derived through computational analysis of biological data. These biological data is acquired from a range of sources, such as genetic data, patient statistics and scientific literature. The goal of the research presented in this thesis is to develop methods for the analysis of microscopy images and extract useful knowledge from these images for biology. This thesis particularly focuses on the analysis of differences in the phenotype of biological specimens visualized in both 2D and 3D images and captured with different microscopes. Our results intend to support biology in analysing pathways and to generate a better representation of biological models for such phenotype analysis.

Chapter 2 focuses on the understanding of pathway development of epidermal growth factor receptor (EGFR) endocytosis. A model representation is provided so that a solution can be elaborated analysing the high-content cytomics screening for target discovery. The system can automatically extract the interesting objects (proteins) for the phenotype measurement and use pattern recognition methods to characterize the objects into characteristic developmental episodes of EGFR endocytosis.

Chapter 3 further elaborates on EGFR endocytosis research. In order to improve the phenotype identification process we devised and applied a hierarchical classification strategy. In addition, we introduced the wavelet-based texture measurements to generate extra prominent features for the classification. With these two refinements, the phenotype identification process significantly improved and can be employed for the discovery of regulators of the EGFR endocytosis process. In support of advanced representation of models that are derived from 3D images,

in Chapter 4 we introduce an analytical evaluation for the point based 3D surface reconstruction methods. In our study we used three analytical shapes, i.e. the sphere, the ellipsoid and the oval, so that a ground truth measurement is available. We studied these three shapes with the same volume size under different levels of noise and we evaluated three major surface descriptors: surface area, surface distance and curvature. The results revealed that within from the point cloud reconstruction methods, the Poisson reconstruction method performs best in surface preservation and noise suppression.

In Chapter 5 we have used the conclusions of Chapter 4 and further extend a system for 3D model representation and analysis from a stack of images. We developed a contour interpolation strategy to convert a 3D contour model to a uniformly distributed 3D point cloud. With this point cloud, subsequently, the Poisson reconstruction method is used to reconstruct an accurate surface model. Our representation is tested with two typical 3D models; in a study of zebrafish development using the proper 3D shape descriptors we have verified that the pre-processing and 3D representation works.

In Chapter 6 we have used more complex 3D models to investigate if these can be successfully analysed. The mammary gland in new-born mouse is a branched structure and this requires different strategies for analysis. Therefore we have employed centerline of the optimized 3D model in order to extract a good representation of the topological information. Features from the topology are used to analyse 3D models from mammary glands that developed under exposure of different potential endocrine disruptors. We presented that we are able to differentiate the different conditions and characterize the effect of the exposures. In addition, we modelled the branched structure using an L-system so as to obtain further evidence of the correct characterization of the different conditions using our measurement system.

In our studies, we have intended to find multiple ways to deal with image datasets in both 2D and 3D space. These image datasets are from high-throughput screening or from image stacks. All of the datasets have a relatively large volume. How to extract meaningful and crucial information from large quantities of biological images has been the major question that we intended to address. We have applied different methods for the extraction of objects of interest from the im-

SUMMARY

ages; we have used different shape descriptors to analyze the objects and we have utilized a range of pattern recognition strategies to categorize patterns that we suspected in the data. We have used these techniques in support of the further understanding of biology, i.e. pathways and development. Our work contributes to the field of bioinformatics, has shown to be meaningful and will show to be sustainable.

