

Creating Classification Features for Biological Images

Abstract

Computational Biology covers a broad spectrum of diverse applications within the field of molecular biology. Much of the computational work done so far has focused at the molecular level because of the strong computational characteristics of molecular biology. This tight focus has left relatively unexplored other aspects of biology such as cell biology that might benefit from computational techniques. This thesis examines the use of computational techniques in the field of cell biology. Towards this end, we develop a system based on image processing and machine learning to characterize the cellular events occurring in the cell division process meiosis and to classify images of cells exhibiting these events. The cellular events in question are the eight phases of Meiosis and the post-meiotic events that decide the phenotype of the resulting cells. The results of this thesis suggest the existence of significant application potential in computational techniques to the field of cell biology.

Acknowledgements

First, I would like to thank my advisor, Dr. Rich Maclin, for his persistent efforts in helping me finish this thesis. I would like to express my gratitude to Dr. Doug Dunham and Dr. Joe Gallian for their willingness to serve on my committee. I would also like to thank Dr. Qin Liu and her students, Kevin Wolfe and Jennifer Baumgardt, for providing images for this work and helping me with the biological aspects of this thesis. Finally, I would like to thank the CS faculty and my fellow students for their support during these two years of graduate study and making my stay in Duluth an enjoyable and memorable experience.

Contents

1	Introduction	1
2	Background	4
2.1	Meiosis	4
2.1.1	Phases of Meiosis	7
2.1.2	Wild-type and Mutant cell types	18
2.2	Image Analysis	21
2.2.1	Image Preparation	21
2.2.2	Image Histogramming and Segmentation	21
2.2.3	Image Normalization	22
2.2.4	Region Isolation	23
2.2.5	Region Identification	25
2.3	Machine Learning	27
2.3.1	Inductive Learning	28
3	Characterizing Cells during Meiosis	29
3.1	Features Examined	29
3.1.1	Feature: Number of Internal Regions	31
3.1.2	Feature: Main Region Occupancy by Internal Regions..	31
3.1.3	Feature: Area	31
3.1.4	Feature: Perimeter	32
3.1.5	Feature: Radius	32
3.1.6	Feature: Compactness	32
3.1.7	Feature: Smoothness	32
3.1.8	Feature: Texture	32
3.1.9	Feature: Concavity	34
3.1.10	Feature: Concave points	34
3.2	Initial Results	34
3.2.1	Result: Number of Internal Regions	35
3.2.2	Result: Main Region Occupancy by Internal Regions..	35
3.2.3	Result: Area	38
3.2.4	Result: Perimeter	38
3.2.5	Result: Radius	41
3.2.6	Result: Compactness	41
3.2.7	Result: Smoothness	41
3.2.8	Result: Texture	45
3.2.9	Result: Concavity	45
3.2.10	Result: Concave points	45
3.3	A Cell Classifier	49
3.4	Test Results	52

4	Future Work	58
	4.1 Event Classification	58
	4.2 Image Processing	59
	4.3 Machine Learning	60
5	Conclusions	61
6	References	63

List of Figures

2.1	Schematic drawing of the reproduction process in multicellular organisms	6
2.2	Appearance of homologous pair of chromosomes at the beginning Of Meiosis	6
2.3	Picture of a cell in prophase I	8
2.4	Picture of a cell in prometaphase	9
2.5	Picture of a cell in metaphase I	11
2.6	Picture of a cell in anaphase I	12
2.7	Picture of a cell in telophase I	13
2.8	Picture of a cell in prophase II	15
2.9	Picture of a cell in metaphase II	16
2.10	Picture of a cell in anaphase II	17
2.11	Picture of a cell in telophase II	19
2.12	Images of cells exhibiting (a) wild-type, (b) ms2 mutation and (c) po mutation	20
2.13	Histogram of a cell in telophase I	20
2.14	Image of cells before and after applying normalization and their corresponding histograms	24
2.15	A cell isolated from a group of cells using region isolation tool	26
2.16	Regions identified in a cell image using region extraction tool	26
3.1	Images of cells exhibiting (a) wild-type, (b) ms2 mutation and (c) po mutation	30
3.2	Images of cells exhibiting (a) prophase I, (b) prometaphase, (c) telophase I, and (d) telophase II	30
3.3	Radial region lines used for computing region smoothness	33
3.4	Region chords used for computing region concavity	33
3.5	Plots of number of internal regions to main regions	36
3.6	Plots of main region occupancy by internal regions	37
3.7	Plots of area of internal regions	39
3.8	Plots of perimeter of internal regions	40
3.9	Plots of radius of internal regions	42
3.10	Plots of compactness of internal regions	43
3.11	Plots of smoothness of internal regions	44
3.12	Plots of texture of internal regions	46
3.13	Plots of concavity of internal regions	47
3.14	Plots of number of concave points in internal regions	48
3.15	A classifier for wild-type, ms6 and po mutation cell images	50
3.16	A classifier for prophase I, prometaphase, telophase I and telophase II cell images	51

3.17	Plot of the number of internal regions to main regions in wild-type, ms2 and po mutation test cell images	53
3.18	Plot of main region occupancy by internal regions in wild-type, ms2 and po mutation cell images	53
3.19	Plot of texture of internal regions in wild-type, ms2 and po mutation test cell images	54
3.20	Plot of the number of internal regions in prophase I, prometaphase, telophase I and telophase II test cell images	56
3.21	Plot of main region space occupancy by internal regions in prophase I, prometaphase, telophase I and telophase II test cell images	56

List of Tables

3.1	Test results of classifier for wild-type cell, ms6 and po mutation Cells	52
3.2	Test results of classifier for prophase I, prometaphase, telophase I and telophase II cells	55