18 JUN 2002

LIBRARY AND INFORMATION CENTER
MAHIDOL UNIVERSITY

SEGMENTATION OF CHROMOSOME IMAGES USING AN ADAPTIVE THRESHOLDING TECHNIQUE AND A WATERSHED ALGORITHM

KANNIKAR KAMOLRAT

With compliments
of
บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2002

ISBN 974-04-1478-8

COPYRIGHT OF MAHIDOL UNIVERSITY

TH
K165
2002
C.2

Copyright by Mahidol University

4137840 SCCS/M : MAYJOR : COMPUTER SCIENCE ; M.Sc.(COMPUTER SCIENCE)
KEY WORDS : IMAGE SEGMENTATION / CHROMOSOME / THRESHOLDING / WATERSHED

KANNIKAR KAMOLRAT : SEGMENTATION OF CHROMOSOME IMAGES USING AN ADAPTIVE THRESHOLDING TECHNIQUE AND A WATERSHED ALGORITHM. THESIS ADVISOR : SUKANYA PHONGSUPHAP, Ph.D., DAMRAS WONGSAWANG, Ph.D., CHOMTIP PRONPANOMCHAI, Ph.D., 131 p. ISBN 974-04-1478-8

This research proposes an approach of segmenting chromosome images using an adaptive thresholding technique and a watershed algorithm, focussing on an automatic separation of touching chromosomes from one another. It is well realized that an analysis of laboratory animals' chromosomes is of great significance for genetic toxicology to analyze chromosomes' abnormality. An essential preliminary process in this regard is karyotyping. Manual karyotyping can be done by segmenting metaphase chromosomes one by one from a chromosome image and then making a pairing of these chromosomes based on their sizes, ranging from large to small, while a centromere position of chromosomes is also taken into account. Afterwards, a pairing of these chromosomes is made again on a standard-form paper. A problem found in manual karyotyping is that it takes a long time since in an analysis of chromosomes, a large number of cells (between 50 and 100) are studied. Furthermore, a chromosome analysis depends on an analyst's experience and skill, including his or her temper at the time. A lack of experience and skill in an analyst generally results in an erroneous or ineffective analysis.

Based on previous researches and studies, a system of a computerized chromosome analysis encounters a problem of a touching chromosome separation. If this problem arises, it still needs a human being to solve the problem. A current chromosome analytical system is, therefore, not completely automatic. With the above-mentioned problems, this research proposes an approach of segmenting a chromosome image using an adaptive thresholding technique so as to segment chromosomes from the background. This is carried out with a consideration of values of 2 thresholds : global and local. Nevertheless, segmented chromosomes may have some clusters consisting of touching chromosomes. So, there have to be steps to check which cluster consists of such chromosomes. In this research, checking is done with a principle of a multilayer perceptron (MLP) neural network. After checking of touching chromosomes, they are separated from each other with an iterative watershed algorithm.

The experiments in this research utilize 34 image samples of hamsters' chromosomes. The images are scanned and stored in a bitmap (.bmp) file. There were 3 experiments conducted in this research. The first one was chromosome segmentation using an adaptive thresholding technique. The result was a correct segmentation of 1,336 chromosomes, which is 89.3% of the total number, apart from an emergence of 8 broken and 69 touching chromosomes, which are 0.6% and 10.1% of the total number, respectively. The second experiment was a touching chromosome detection using a neural network. It was found that out of 1,016 samples of single and touching chromosomes, 741 samples were correctly detected, which is 72.43% of the total number, while 275 samples were erroneously detected, which is 27.07% of the total number. Finally, the third experiment was a touching chromosome separation using a watershed algorithm. The outcome was an accurate separation of 63 touching chromosomes out of the total 69 touching ones, which is 91.3% of the total number. There was also an erroneous separation of 6 touching chromosomes done in a wrong position, which was 8.7% of the total number.

Copyright by Mahidol University

4137840 SCCS/M : สาขาวิชา : วิทยาการคอมพิวเตอร์ ; วท.ม. (วิทยาการคอมพิวเตอร์)

กรรณิการ์ กมลรัตน์ : การแบ่งส่วนภาพโครโมโซมโดยใช้เทคนิคปรับค่าขีดแบ่งให้เหมาะและขั้นตอนวิธีสันปันน้ำ (SEGMENTATION OF CHROMOSOME IMAGES USING AN ADAPTIVE THRESHOLDING TECHNIQUE AND A WATERSHED ALGORITHM). คณะกรรมการควบคุมวิทยานิพนธ์ : สุกัญญา พงษ์สุภาพ, Ph.D., ดำรัส วงศ์สว่าง, Ph.D., ชมทิพย์ พรพนมชัย, Ph.D., 131 หน้า. ISBN 974-04-1478-8

ในงานวิจัยนี้ได้นำเสนอวิธีการแบ่งส่วนภาพโครโมโซมโดยใช้เทคนิคปรับขีดแบ่งให้เหมาะ (adaptive thresholding) และ ขั้นตอนวิธีสันปันน้ำ (watershed algorithm) โดยเน้นการแยกโครโมโซมที่ติดกันให้ออกจากกันได้อย่างอัตโนมัติ เนื่องจากการวิเคราะห์โครโมโซมของสัตว์ทดลองมีความสำคัญอย่างยิ่งสำหรับศาสตร์ทางด้านพิษวิทยาพันธุศาสตร์ (Genetic toxicology) เพื่อวิเคราะห์ดูความผิดปกติของโครโมโซม ซึ่งขบวนการพื้นฐานที่จำเป็นต้องทำคือ การจัดเรียงโครโมโซม (karyotype). การจัดเรียงโครโมโซมด้วยมือ (Manual karyotype) สามารถทำได้โดยนำภาพถ่ายของกลุ่มโครโมโซมระยะ metaphase แยกออกทีละตัว ๆ จัดเรียงกันเป็นคู่ ๆ โดยดูขนาดของโครโมโซมเป็นหลักจากขนาดใหญ่ไปหาขนาดเล็ก และดูตำแหน่ง centromere ประกอบ จากนั้นนำมาเรียงติดลงบนกระดาษตามแบบฟอร์มสากล การจัดเรียงโครโมโซมด้วยมือ ดังกล่าวพบปัญหาคือ เกิดความล่าช้า ต้องใช้เวลานาน เพราะในการวิเคราะห์โครโมโซมแต่ละครั้งต้องศึกษาจากเซลล์จำนวนมากอยู่ระหว่าง 50-100 เซลล์ และการวิเคราะห์โครโมโซมต้องขึ้นอยู่กับประสบการณ์ ความชำนาญของผู้ทำการวิเคราะห์และยังขึ้นอยู่กับอารมณ์ของผู้วิเคราะห์ในขณะนั้นด้วย ถ้าผู้วิเคราะห์มีประสบการณ์น้อยหรือมีความชำนาญไม่มากพออาจจะทำให้การวิเคราะห์ผิดพลาด นำมาซึ่งผลการวิจัยที่ขาดความถูกต้องแม่นยำ.

จากงานวิจัยที่ผ่าน ๆ มา ระบบที่ใช้วิเคราะห์โครโมโซมด้วยคอมพิวเตอร์ยังติดปัญหาในเรื่องการแยกโครโมโซมที่มีลักษณะติดกัน ซึ่งถ้าพบปัญหานี้จะต้องให้คนเข้าช่วยแก้ปัญหาในลักษณะนี้อยู่ ทำให้ระบบการวิเคราะห์โครโมโซมยังไม่สามารถทำได้อย่างอัตโนมัติโดยสมบูรณ์ ดังนั้นในงานวิจัยนี้ได้นำเสนอวิธีการแยกภาพโครโมโซมโดยใช้เทคนิคปรับขีดแบ่งให้เหมาะ เพื่อแยกโครโมโซมออกมา โดยพิจารณาค่าขีดแบ่ง 2 ค่า คือค่าขีดแบ่งส่วนกลาง (global threshold) และค่าขีดแบ่งเฉพาะที่ (local threshold) หลังจากนั้นโครโมโซมที่แยกออกมาได้อาจจะมีบางกลุ่มที่โครโมโซมอาจจะยังติดกันอยู่ จึงต้องมีขั้นตอนสำหรับตรวจสอบดูว่ากลุ่มไหนที่โครโมโซมติดกัน สำหรับการตรวจสอบนั้นงานวิจัยนี้จะใช้หลักการของ neural network แบบ multilayer perceptron (MLP) เมื่อตรวจสอบกลุ่มที่มีโครโมโซมติดกันได้แล้ว จะทำการแยกโครโมโซมที่ติดกัน โดยใช้ขั้นตอนวิธีสันปันน้ำ แบบวิธีวนซ้ำ (iterative methods)

สำหรับการทดลองในงานวิจัยนี้ใช้ภาพโครโมโซมของหนู (Hamster) เป็นภาพตัวอย่าง จำนวน 34 ภาพ ที่ scan เก็บในรูปแบบ bitmap (.bmp) ซึ่งการทดลองมีทั้งหมด 3 การทดลองคือ การทดลองที่ 1 : การแยกโครโมโซมโดยใช้เทคนิคปรับขีดแบ่งให้เหมาะ ผลลัพธ์คือสามารถแยกโครโมโซมได้ถูกต้อง 1,336 ตัว คิดเป็น 89.3%, มีโครโมโซมแตก 8 ตัว คิดเป็น 0.6% และมีโครโมโซมติดกัน 69 ตัว คิดเป็น 10.1% ; การทดลองที่ 2 : การตรวจหาโครโมโซมติดกันโดยใช้โครงข่ายประสาท ผลลัพธ์คือ จากโครโมโซมเดี่ยวและโครโมโซมติดกันจำนวน 1,016 ตัวอย่าง สามารถให้ผลลัพธ์ถูกต้อง 741 ตัวอย่าง คิดเป็น 72.93% และเกิดข้อผิดพลาด 275 ตัวอย่าง คิดเป็น 27.07%; การทดลองที่ 3 : การแยกโครโมโซมที่ติดกันออกจากกันโดยใช้ขั้นตอนวิธีสันปันน้ำ ผลลัพธ์คือจำนวนโครโมโซมติดกันทั้งหมด 69 ตัว สามารถแยกได้ถูกต้อง 91.3 % และเกิดความผิดพลาดจากการแยกผิดตำแหน่ง 8.7 %.

Copyright by Mahidol University