บทที่ 4

การบีบอัดข้อมูล

การเก็บข้อมูลแบบจุดภาพ (Raster) แฟ้มข้อมูลข้อมูลจะมีขนาดใหญ่โดยขึ้นอยู่กับขนาดของภาพ และความละเอียดในการสแกนภาพเข้าสู่ระบบคอมพิวเตอร์ ดังนั้นหน่วยความจำที่จะใช้ก็ต้องมากตาม ทำให้เกิดความไม่สะดวก และยุ่งยากในการพัฒนาโปรแกรม นี้จึงได้นำเทคนิคการบีบอัดข้อมูลมาประยุกต์ใช้

เทคนิคการบีบอัดข้อมูลมีหลายวิธี ซึ่งได้อธิบายไว้ในภาคผนวก ข. สำหรับงานวิจัยนี้ได้เลือกใช้วิธี Run-length compression (David และ John,1992)[1] เพราะเป็นวิธีที่ง่ายในการทำงาน ทั้งในขั้นตอนการบีบอัดข้อมูลและการคลายคืนข้อมูล นอกจากนั้นยังสอดคล้องกับวิธีการทำงานของโปรแกรมที่พัฒนา คือ การทำงานทั้งหมดจะกระทำทีละบรรทัด ให้เสร็จโดยเรียงบรรทัดจากบนลงล่าง ภายในบรรทัดก็จะทำงานเรียงจากซ้ายไปขวา การทำงานของวิธี Run-length compression ก็เช่นกันโดยปฏิบัติงานครั้งละหนึ่งบรรทัด

วิธีการทำงานของ Run-length compression คือ จะตรวจทีละจุดภาพ นับจำนวนจุดภาพที่มีค่าซ้ำกันและติดกัน แล้วแทนจุดภาพทั้งชุดนั้นด้วย จำนวนที่นับได้ แล้วตามด้วยค่าของของจุดภาพที่ซ้ำกัน เช่น ชุดของข้อมูลเป็น aaabbbbbccaaaaddddddd เมื่อผ่านการบีบอัดโดยวิธีนี้ ข้อมูลผลลัพธ์จะเป็น 3a5b2c4a6d

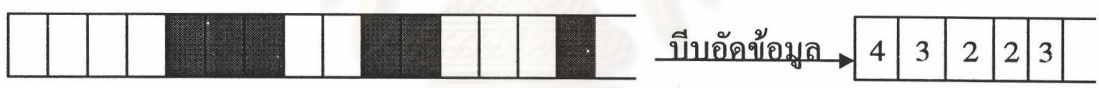
สำหรับงานวิจัยชุดนี้ได้นำหลักการวิธีของ Run-length compression มาประยุกต์ให้เข้ากับการทำงานโดยแฟ้มข้อมูลภาพแผนที่เส้นชั้นความสูงที่ได้จากการสแกนเป็นแฟ้มข้อมูลในรูปแบบ Gray scale โดยที่ 1 จุดภาพ มีค่า 1 ไบต์ (0-255) ข้อมูลส่วนที่เป็นเส้นชั้นความสูงจะมีค่าใกล้ 0 ส่วนที่เป็นพื้นของแผนที่เส้นชั้นความสูงจะมีค่าใกล้ 255 ดังนั้นค่าของจุดภาพ ทั้งภาพจะถูกแบ่งเป็น 2 ค่า คือ 0 กับ 255 เท่านั้น โดยส่วนที่มีค่าใกล้ 0 หรือมีค่าเป็น 0 จะถูกเปลี่ยนค่าเป็น 0 ทั้งหมด ในส่วนที่มีค่าใกล้ 255 หรือมีค่าเป็น 255 จะถูก

เปลี่ยนค่าเป็น 255 ทั้งหมด

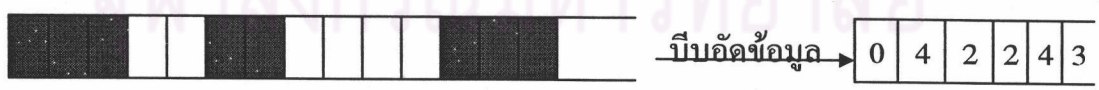
ในการนำวิธีการบีบอัดข้อมูลแบบ Run-length compression มาใช้ในงานวิจัยนี้ จะมีข้อแตกต่างหลัก คือ จะบันทึกจำนวนจุดภาพที่ซ้ำกัน ส่วนค่าของจุดภาพจะไม่บันทึกไว้ ดังนั้น จึงมีเงื่อนไขพิเศษเพิ่มเติม เพื่อจะได้สนับสนุนการทำงานในส่วนอื่น ๆ ที่จะกล่าวถึงต่อไป

เงื่อนไขเพิ่มเติมในการทำ Run-length compression สำหรับงานวิจัยชุดนี้คือ

1. การทำงานไล่ทำทีละบรรทัด แต่ทุกบรรทัดจะต้องบันทึกจำนวนของจุดภาพ ที่มีค่า 255 ก่อน เสมอ ถ้าบรรทัดใดไม่ได้ขึ้นต้นด้วยจุดภาพ ที่มีค่า 255 ก็ให้ค่านับชุดแรกเป็น 0 ดังแสดงในรูปที่ 4-1 และ รูปที่ 4-2



รูปที่ 4-1 ภาพแสดงการบีบอัดข้อมูล ในกรณีที่ขึ้นต้นด้วยจุดภาพที่มีค่า 255



รูปที่ 4-2 ภาพแสดงการบีบอัดข้อมูล ในกรณีที่ขึ้นต้นด้วยจุดภาพที่มีค่า 0

2. การบันทึกจำนวนนับของจุดภาพ ในการบันทึกแต่ละครั้ง จะต้องมีค่าที่จะบันทึกไม่เกิน 255 ถ้าเกิน 255 จะแบ่งบันทึกครั้งละ 255 จนครบจำนวน เช่น จำนวนที่นับได้มีค่า 300 การบันทึกจะแบ่งบันทึกเป็น 255 กับ 45 ดังแสดงในรูปที่ 4-3

3. ในการบันทึกจำนวนนับของจุดภาพ แต่ละครั้งจะต้องบันทึกจำนวนของจุดภาพ ค่า 0 และค่า 255 สลับกันเสมอ แสดงในรูปที่ 4-3

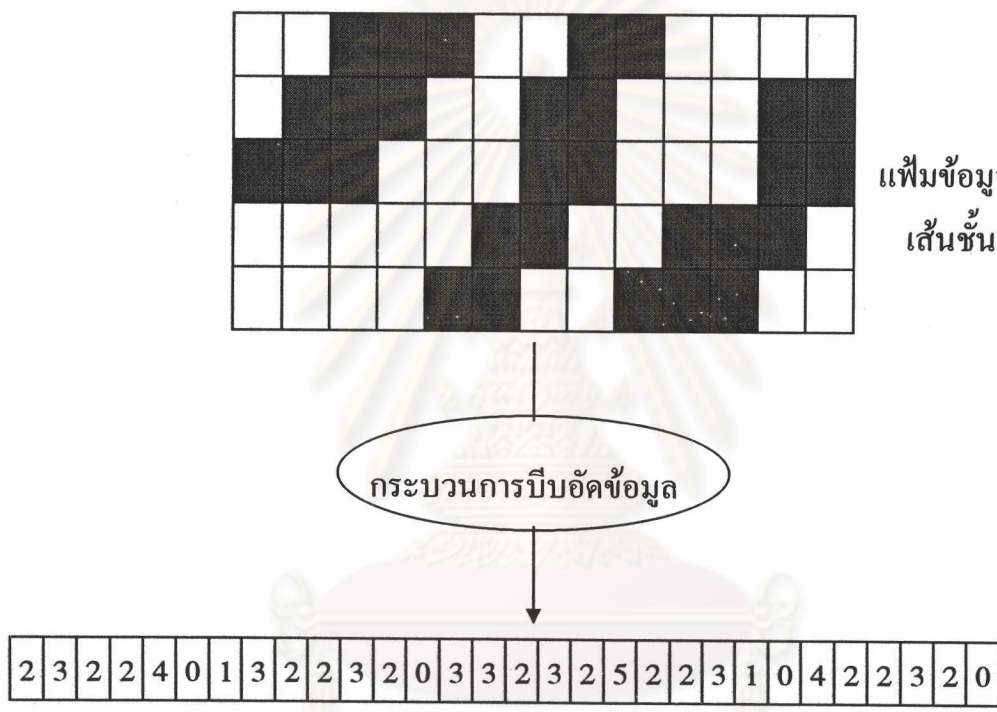
รูปที่ 4-3 ภาพแสดงการบีบอัดข้อมูลตามเงื่อนไขข้อ 2 และข้อ 3

4. การบันทึกจำนวนนับของจุดภาพ แต่ละบรรทัด ทุกบรรทัดจะต้องลงท้ายด้วยจำนวนนับของจุดภาพ ที่มีค่าเป็น 0 เสมอ ถ้าบรรทัดใดลงท้ายด้วยจุดภาพ ที่มีค่าเป็น 255 ก็จะบันทึกจำนวนนับสุดท้ายเป็น 0 แสดงในรูปที่ 4-4

รูปที่ 4-4 แสดงการบีบอัดข้อมูล ตามเงื่อนไขข้อ 4

แฟ้มข้อมูลผลลัพธ์ที่ได้หลังจากได้ทำการบีบอัดข้อมูลแล้วจะเป็นไบนารี่ โดยขนาดของแฟ้มข้อมูลจะขึ้นอยู่กับรายละเอียดบนแผนที่เส้นชั้นความสูง ถ้ารายละเอียดมาก แฟ้มข้อมูลก็อาจจะมีขนาดใหญ่ การบันทึกข้อมูลจะไล่เรียงกันจนหมดภาพ ดังนั้นการแบ่งบรรทัดของภาพจะได้จากการรวมค่าของแต่ละ 1 ไบต์ จนได้เท่ากับขนาดของภาพ

กระบวนการบีบอัดตามเงื่อนไขทั้งหมดนี้ ผลลัพธ์ที่ได้ จะได้แฟ้มข้อมูลบีบอัดข้อมูลและ มีผลพลอยได้คือ สามารถจำแนกเส้นชั้นความสูงจากพื้นแผนที่เส้นชั้นความสูงได้ระดับหนึ่ง



รูปที่ 4-5 ภาพแสดงการบีบอัดข้อมูล

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากวิธีการบีบอัดข้อมูลดังกล่าวมานี้ ผลการทดลองใช้กับพื้นที่ทดลองทั้ง 3 ประเภท ได้ผลดังตาราง 4-1

ตาราง 4-1 แสดงผลการบีบอัดข้อมูล

ประเภทพื้นที่	ขนาดแฟ้มข้อมูลเดิม (ไบต์)	ขนาดแฟ้มข้อมูลที่ถูกบีบอัด (ไบต์)	% การบีบอัด
พื้นที่ชันน้อย	240,000	5,614	97.66
พื้นที่ความชันปานกลาง	240,000	7,878	96.72
พื้นที่ภูเขา	240,000	11,204	95.33

ขนาดแฟ้มข้อมูลหลังจากทำการบีบอัดข้อมูลแล้ว จะมีขนาดเล็กลงมากหรือน้อยขึ้นอยู่กับความหนาแน่นของเส้นชั้นความสูง คือ ความหนาแน่นของเส้นชั้นความสูงมาก ขนาดของแฟ้มข้อมูลหลังจากทำการบีบอัดข้อมูลจะมีขนาดใหญ่กว่าแฟ้มข้อมูลที่มีความหนาแน่นของเส้นชั้นความสูงน้อย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย