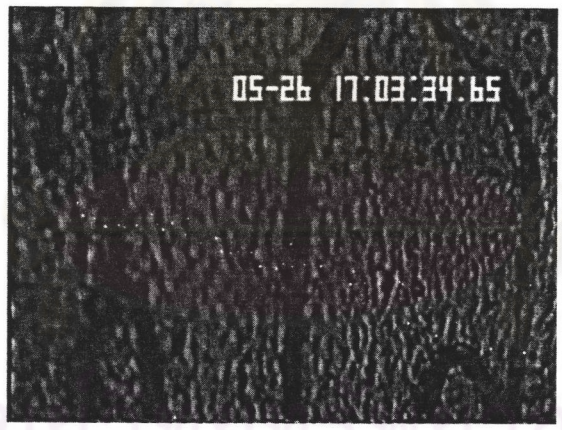
บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

นาฬิกาตัวเลขที่ปรากฏอยู่ในภาพวีดิทัศน์เป็นข้อมูลที่ถูกบันทึกเพิ่มเติมลงในภาพและเป็นข้อมูลสำคัญที่ทำให้ทราบวันที่และเวลาที่ทำการบันทึกภาพ โดยได้ถูกนำไปใช้ประโยชน์ต่าง ๆ นอกจากจะเป็นหลักฐานที่แสดงถึงเวลาของเหตุการณ์ที่เกิดขึ้นในภาพแล้ว ยังสามารถนำไปใช้วัดอัตราเร็วของวัตถุในภาพโดยเปรียบเทียบจากเวลาที่เปลี่ยนไประหว่างเฟรมที่วัตถุมีการเคลื่อนที่จากตำแหน่งหนึ่งไปยังอีกตำแหน่งหนึ่ง ตัวอย่างของภาพวีดิทัศน์ที่มีนาฬิกาตัวเลขแสดงดังรูปที่ 1.1 ซึ่งเป็นภาพแสดงการไหลของเซลล์เม็ดเลือดแดงในหลอดเลือด ข้อมูลนาฬิกาตัวเลขในรูปที่ 1.1 ประกอบด้วยเดือน วัน และเวลาตามลำดับ โดยที่เวลามีความละเอียดในระดับชั่วโมง นาที วินาที และเศษของวินาที



รูปที่ 1.1 ตัวอย่างนาฬิกาตัวเลขในภาพวีดิทัศน์ซึ่งแสดงการไหลของเซลล์เม็ดเลือดแดงในหลอดเลือด

นาฬิกาตัวเลขในภาพวีดิทัศน์สามารถนำไปใช้ในงานทางด้านต่าง ๆ ยกตัวอย่างเช่น งานทางด้านการวัดอัตราเร็วของวัตถุจากภาพ เช่น ทางด้านการแพทย์มีการวัดอัตราการไหลของเซลล์เม็ดเลือดแดงในหลอดเลือด [1] เพื่อใช้วิเคราะห์โรคบางโรคที่ทำให้เซลล์เม็ดเลือดแดงเกาะติดกับผนังหลอดเลือด และใช้วิเคราะห์ผลจากการทดลองใช้ยารักษาโรค โดยทดลองกับหนูทดลองว่ายาสามารถลดปริมาณเม็ดเลือดแดงที่เกาะผนังหลอดเลือดได้เท่าใด สามารถเพิ่มอัตราเร็วการเคลื่อนที่ของเซลล์เม็ดเลือดได้หรือไม่ และอีกตัวอย่างคืองานทางด้านการจราจร โดยได้ถูกนำไปใช้ในการวัดอัตราเร็วของรถยนต์ [2] สามารถนำไปวิเคราะห์สภาพการจราจรบนท้องถนน หรือนำไปใช้ตรวจจับรถยนต์ที่แล่นเกิน

อัตราเร็วที่กำหนด นอกจากนี้ยังมีงานทางด้านการทำดัชนีและการค้นคืนภาพวีดิทัศน์ [3] ซึ่งใช้นาฬิกาตัวเลขเป็นส่วนหนึ่งของดัชนีได้

ในงานวิจัย "การวัดอัตราการไหลของเซลล์เม็ดเลือดแดงในหลอดเลือด" [1] ได้ใช้แผ่นวงจรจับเฟรม (Frame grabber) จับภาพการเคลื่อนที่ของเซลล์เม็ดเลือดแดงจากภาพวีดิทัศน์ 30 ภาพในเวลา 1 วินาที แล้วจึงประมวลผลเพื่อหาตำแหน่งของเซลล์เม็ดเลือดแดงในแต่ละภาพ และนำตำแหน่งที่หาได้ไปวาดกราฟระหว่างระยะทางกับเวลา เพื่อนำไปคำนวณหาอัตราเร็วของเซลล์เม็ดเลือดแดง ณ เวลาใด ๆ โดยคำนวณเทียบจากเวลาระหว่างภาพแต่ละภาพห่างกัน 1 ใน 30 วินาที อย่างไรก็ตามวิธีดังกล่าวอาจเกิดความผิดพลาดได้ในกรณีที่แผ่นวงจรจับเฟรมมีคุณภาพไม่ดีพอและไม่สามารถจับภาพได้ครบทุกเฟรม ดังนั้นถ้ามีโปรแกรมที่อ่านนาฬิกาตัวเลขจากภาพแต่ละภาพได้โดยตรง ก็น่าจะทำให้การคำนวณความเร็วจากภาพวีดิทัศน์มีความถูกต้องและน่าเชื่อถือมากกว่าเดิม

1.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการตรวจหาตำแหน่งและรู้จำเวลาของนาฬิกาตัวเลขในภาพวีดิทัศน์แบ่งเป็น 2 ส่วนหลักได้แก่ การตรวจหาตำแหน่งข้อความในภาพวีดิทัศน์ และการรู้จำตัวอักษร

1.2.1 การตรวจหาตำแหน่งข้อความในภาพวีดิทัศน์

การตรวจหาตำแหน่งข้อความในวีดิทัศน์ได้รับความสนใจเป็นอย่างมากและมีการวิจัยกันอย่างแพร่หลายในปัจจุบันเนื่องจากเป็นพื้นฐานในการทำดัชนีและการค้นคืนข้อมูลที่อยู่ในวีดิทัศน์ ข้อความที่ปรากฏในวีดิทัศน์มีส่วนสัมพันธ์กับเนื้อหาโดยตรง อีกทั้งอาจแสดงข้อมูลบางส่วนที่เกี่ยวข้องด้วย เช่น ในรายการข่าวมักจะมีข้อความที่เป็นหัวข้อข่าว รายละเอียดโดยย่อ ชื่อผู้ประกาศข่าว สถานีโทรทัศน์ที่ถ่ายทอด และเวลาที่ออกอากาศ เป็นต้น นาฬิกาตัวเลขในภาพวีดิทัศน์ถือได้ว่าเป็นข้อความอย่างหนึ่ง ดังนั้นจึงสามารถประยุกต์แนวคิดในการตรวจหาตำแหน่งและรู้จำข้อความในภาพวีดิทัศน์มาใช้ในงานวิจัยนี้ได้ ตัวอย่างงานวิจัยที่ตรวจหาตำแหน่งข้อความในภาพวีดิทัศน์มีดังต่อไปนี้

1) Automatic Text Location in Images and Video Frames ปี ค.ศ.1998 โดย Jain และ Yu [4] เสนอวิธีการระบุตำแหน่งข้อความซึ่งเหมาะกับภาพหลายรูปแบบ เช่น ภาพโฆษณาจากหนังสือพิมพ์ ภาพจากหน้าเว็บ ภาพสีจากการสแกน ภาพวีดิทัศน์ วิธีที่เสนอใช้การวิเคราะห์องค์ประกอบที่เชื่อมต่อกัน (Connected component analysis) นอกจากนี้ได้เสนอวิธี Multi-valued image decomposition ซึ่งเป็นการแยกภาพออกเป็นหลาย ๆ ภาพตามค่าสีแต่ละสีภายในภาพ ข้อดีของวิธีนี้คือระบบใช้การคำนวณน้อย ทนทานต่อการเปลี่ยนแปลงขนาดตัวอักษร และมีความถูกต้องสูง ผลการทดลองกับภาพหลายรูปแบบมีความถูกต้องโดยเฉลี่ยร้อยละ 91

2) Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions ปี ค.ศ.1999 โดย Sato และคณะ [5] ได้วิจัยการทำดัชนีของห้องสมุดข่าวดิจิทัลโดยการรู้จำคำอธิบายภาพในวีดิทัศน์ การตรวจหาตำแหน่งข้อความในงานวิจัยนี้เริ่มจากการตรวจหาขอบภาพ

(Edge detection) จากนั้นจึงใช้ตัวกรองทำให้เรียบ (Smoothing filter) เพื่อกำจัดสัญญาณรบกวนและเชื่อมขอบภาพของส่วนที่เป็นตัวอักษรในข้อความเดียวกันให้ติดกัน แล้วจึงจัดกลุ่มบริเวณที่น่าจะเป็นข้อความโดยตีกรอบสี่เหลี่ยมล้อมรอบ ส่วนการปรับปรุงภาพจะเพิ่มความละเอียดของภาพโดยใช้วิธีการประมาณค่าจากจุดภาพย่อย (Sub-pixel interpolation) และแก้ปัญหาพื้นหลังที่ซับซ้อนด้วยตัวดำเนินการค่าต่ำที่สุด/ค่าสูงที่สุด (Minimum/maximum operator) กับเฟรมหลายเฟรมที่เรียงลำดับกัน แล้วจึงใช้ตัวกรองการดึงตัวอักษร (Character extraction filter) ซึ่งเป็นการสกัดเส้นแนวนอน แนวตั้ง และแนวทแยงที่เป็นส่วนของตัวอักษร รวมถึงเป็นการกำจัดส่วนของพื้นหลังที่ยังเหลือในภาพให้น้อยลง จากนั้นจึงวิเคราะห์โพรไฟล์การฉาย (Projection profile) เพื่อแยกตัวอักษรแต่ละตัวในข้อความออกจากกัน ส่วนการรู้จำตัวอักษรใช้เทคนิคการจับคู่แบบรูป (Pattern matching) ซึ่งใช้วิธีการหาค่าสหสัมพันธ์เพื่อเปรียบเทียบความคล้ายคลึงกันระหว่างตัวอักษรที่ต้องการนำมารู้จำกับตัวอักษรที่เป็นแผ่นแบบ และนำคำที่รู้จำได้ไปประมวลผลต่อโดยเปรียบเทียบกับคำที่ใกล้เคียงในพจนานุกรม

3) A Video Text Detection and Recognition System ปี ค.ศ.2001 โดย Xi และคณะ [6] ได้เสนอระบบตรวจหาตำแหน่งและรู้จำข้อความในวีดิทัศน์ การตรวจหาตำแหน่งข้อความทำโดยตรวจหาขอบภาพ แล้วแปลงเป็นภาพลักษณ์ฐานสอง (Binary image) จากนั้นนำภาพที่ได้เข้ากระบวนการทำขอบภาพให้เรียบ (Edge map smoothing) เพื่อหาส่วนที่น่าจะเป็นข้อความ แล้วจึงทำการกำจัดสัญญาณรบกวนขนาดเล็ก จากนั้นนำผลลัพธ์ที่ได้ไปแยกบริเวณที่น่าจะเป็นข้อความแต่ละข้อความออกจากกันโดยใช้โพรไฟล์การฉายทั้งในแนวตั้งและแนวนอน แล้วจึงระบุบริเวณที่ควรจะเป็นข้อความจริง ๆ ออกมาโดยกำหนดกฎแสดงลักษณะของข้อความเช่นมีความกว้างและความสูงอย่างต่ำเท่าใด สูงสุดเท่าใด เป็นต้น และเรียกบริเวณดังกล่าวว่ากล่องข้อความ (Text box) สำหรับการรู้จำนั้นจะให้ผลลัพธ์ที่ดีก็ต่อเมื่อภาพในส่วนนำเข้ามีความละเอียดสูง ดังนั้นจึงเพิ่มความละเอียดของกล่องข้อความโดยใช้วิธีการประมาณค่าในช่วงแบบเชิงเส้นคู่ (Bilinear interpolation) ก่อนที่จะแปลงเป็นภาพลักษณ์ฐานสองเพื่อนำไปเข้าโมดูลการรู้จำตัวอักษร ผลลัพธ์ของการตรวจหาตำแหน่งข้อความมีความถูกต้องเพียงพอที่จะนำไปใช้งานจริง ส่วนผลลัพธ์ของการรู้จำข้อความยังไม่ดีเท่าที่ควร เนื่องจากข้อความในวีดิทัศน์บางประเภท เช่น ข่าว มีตัวอักษรขนาดเล็ก

4) New Approach for Video Text Detection ปี ค.ศ.2002 โดย Cai และคณะ [7] ได้เสนอวิธีการใหม่เพื่อตรวจหาตำแหน่งข้อความในวีดิทัศน์ ในขั้นตอนแรกจะตรวจหาขอบภาพและหาค่าขีดแบ่งแบบครอบคลุม (Global thresholding) ซึ่งมีค่าต่ำเพื่อกรองขอบของส่วนที่ไม่ใช่ข้อความบางส่วนออก จากนั้นจึงหาค่าขีดแบ่งแบบท้องถิ่น (Local thresholding) เพื่อรักษาข้อความที่มีความเปรียบต่างต่ำให้คงอยู่และลดความซับซ้อนของพื้นหลังลง แล้วจึงใช้ตัวดำเนินการที่ได้เสนอใหม่นั่นคือ ESS (Edge Strength Smoothing) และ ECP (Edge Clustering Power) เพื่อพิจารณาหาบริเวณที่มีความเข้มและความหนาแน่นของขอบสูง สุดท้ายจึงระบุตำแหน่งของบริเวณที่เป็นข้อความโดยการตรวจหาแบบหยาบไปละเอียด (Coarse-to-fine) ซึ่งใช้โพรไฟล์การฉายเป็นตัวช่วย

นอกจากนี้ยังมีงานวิจัยที่เกี่ยวข้องกับการตรวจหาตำแหน่งนาฬิกาตัวเลขในภาพวีดิทัศน์โดยตรง คือ งานวิจัย Automatic Time Stamp Extraction System for Home Videos ปี 2002 โดย Pei Yin และคณะ [8] ได้คิดวิธีการตรวจหาตำแหน่งนาฬิกาตัวเลขในภาพวีดิทัศน์โดยใช้ Spatial-Temporal Suppression (STS) เพื่อให้ตัวเลขซึ่งมีสีขาวเด่นออกมาจากพื้นหลัง จากนั้นหาตำแหน่งนาฬิกาตัวเลขอย่างหยาบโดยพิจารณาโพรไฟล์การฉายทั้งแนวนอนและแนวตั้งของจุดภาพที่คาดว่าเป็นนาฬิกาตัวเลข นั่นคือจุดภาพที่มีค่ามากกว่าค่าขีดแบ่งค่าหนึ่ง ได้ผลลัพธ์เป็นกล่องสี่เหลี่ยมที่ครอบตัวเลขแต่ละตัว (Bounding box) แล้วจึงปรับขนาดกล่องสี่เหลี่ยมแต่ละอันให้มีขนาดเท่ากันโดยใช้ค่าเฉลี่ยของความกว้างและความสูงของกล่องทุกอันมาเป็นตัวแทนความกว้างและความสูงของกล่องสี่เหลี่ยมแต่ละอัน อย่างไรก็ตามวิธีดังกล่าวอาจมีปัญหาในกรณีที่พื้นหลังของนาฬิกาตัวเลขเป็นสีขาวหรือสีที่สว่างซึ่งจะทำให้หา ตำแหน่งนาฬิกาตัวเลขผิดพลาดได้

1.2.2 การรู้จำตัวอักษร

นาฬิกาตัวเลขมีลักษณะเป็นตัวเลขและสัญลักษณ์ตัวพิมพ์ ดังนั้นจึงได้ศึกษางานวิจัยที่เกี่ยวกับการรู้จำตัวอักษร โดยค้นคว้าจากวิธีที่ใช้กันอยู่ในปัจจุบันเพื่อเป็นข้อมูลช่วยในการวิจัยการรู้จำนาฬิกาตัวเลขต่อไป งานวิจัยที่เกี่ยวข้องได้แก่

1) Historical Review of OCR Research and Development ปี ค.ศ.1992 โดย Mori และคณะ [9] ได้ศึกษาความเป็นมาของงานวิจัยที่เกี่ยวกับการรู้จำตัวอักษร วิธีที่ใช้ในยุคแรกคือการจับคู่แผ่นแบบ (Template matching) ซึ่งเหมาะกับตัวอักษรพิมพ์และมีจำนวนตัวอักษรน้อย งานที่นำไปใช้ ได้แก่ การรู้จำตัวอักษรในเช็คธนาคาร ต่อมามีการใช้วิธีวิเคราะห์โครงสร้าง (Structure analysis) เพื่อรู้จำตัวอักษรลายมือเขียน เนื่องจากตัวอักษรมีรูปร่างและวิธีการเขียนแตกต่างกันเป็นจำนวนมาก ไม่เหมาะกับวิธีการจับคู่แผ่นแบบ วิธีวิเคราะห์โครงสร้างจะพยายามดึงลักษณะ (Features) ของตัวอักษรแล้วพยายามเขียนกฎแสดงความสัมพันธ์ระหว่างลักษณะเหล่านั้น เพื่อสามารถแยกแยะตัวอักษรแต่ละตัวได้ จากนั้นได้นำวิธีทั้งสองมาประยุกต์ใช้ร่วมกันเพื่อเพิ่มประสิทธิภาพการรู้จำให้สูงขึ้น ต่อมาได้มีการสอนให้คอมพิวเตอร์เรียนรู้ตัวอักษรซึ่งมีทั้งการใช้ข้อมูลทางสถิติ (Statistical learning) และใช้โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เข้ามาช่วย

2) A Note on Comparing Classifiers ปี ค.ศ.1995 โดย Duin [10] ได้ทำการเปรียบเทียบตัวแยกแยะระหว่างวิธีแบบใหม่กับวิธีแบบดั้งเดิม (Traditional methods) โดยจะเน้นการเปรียบเทียบระหว่างวิธีโครงข่ายประสาทเทียมซึ่งเป็นวิธีแบบใหม่กับวิธีกฎเพื่อนบ้านที่ใกล้ที่สุด (Nearest Neighbor Rule: NNR) ซึ่งเป็นวิธีแบบดั้งเดิม ประสิทธิภาพของการแยกแยะโดยโครงข่ายประสาทเทียมขึ้นอยู่กับการออกแบบสถาปัตยกรรมโครงข่าย กระบวนการการตั้งค่าเริ่มต้น ขนาดของขั้นการปรับค่าน้ำหนัก กฎของการหยุด เป็นต้น ซึ่งทั้งหมดที่กล่าวมานี้ขึ้นอยู่กับทักษะของผู้ใช้งานว่าสามารถออกแบบระบบได้เหมาะสมกับงานมากน้อยเพียงใด การแยกแยะแบบกฎเพื่อนบ้านที่ใกล้ที่สุดเป็นการแยกแยะแบบ

อัตโนมัติซึ่งง่ายและเร็วต่อการหาผลลัพธ์ Duin ได้ทำการทดลองกับเซตข้อมูลจำนวน 7 เซต พบว่าบางเซตข้อมูลโครงข่ายประสาทเทียมให้ผลดีกว่าวิธีแบบดั้งเดิม ในทางกลับกันบางเซตข้อมูลวิธีดั้งเดิมให้ผลดีกว่า อย่างไรก็ตามวิธีทั้งสองให้ผลไม่แตกต่างกันอย่างมีนัยสำคัญ

1.3 วัตถุประสงค์

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาขั้นตอนวิธีและพัฒนาโปรแกรมต้นแบบในการตรวจหาตำแหน่งและรู้จำเวลาของนาฬิกาตัวเลขในภาพวีดิทัศน์ เพื่อนำไปประยุกต์ใช้ร่วมกับโปรแกรมประยุกต์หรืองานที่เกี่ยวข้องอื่นต่อไป

1.4 ขอบเขตการวิจัย

1) สีของนาฬิกาเป็นสีระดับเดียว (Monotone color) ไม่มีลวดลาย
2) นาฬิกาเป็นตัวเลขอารบิก ซึ่งมีรูปแบบของตัวเลขที่มีเส้นเชื่อมต่อกันโดยไม่ขาดตอน
3) ตัวเลขของนาฬิกาแต่ละตัวต้องมีความสมบูรณ์ และมีความสูงอย่างน้อย 15 จุดภาพ
4) ผู้ใช้ต้องกำหนดรูปแบบของนาฬิกาก่อนทำการประมวลผล
5) ในกรณีที่ผู้ใช้ต้องการความรวดเร็วและความถูกต้องในการประมวลผล โปรแกรมมีทางเลือกให้ผู้ใช้สามารถกำหนดบริเวณของนาฬิกาในภาพ
6) การประมวลผลภาพในงานวิจัยไม่ได้เน้นให้เป็นการประมวลผลแบบทันกาล
7) การทดสอบทำกับภาพวีดิทัศน์อย่างน้อย 3 แหล่ง โดยแต่ละแหล่งใช้กล้องวีดิทัศน์ที่มีแบบอักษร รูปแบบ และตำแหน่งของนาฬิกาตัวเลขที่แตกต่างกัน

1.5 ขั้นตอนการดำเนินงานวิจัย

1) ศึกษาและวิเคราะห์ปัญหาวิธีการตรวจหาตำแหน่งข้อความในภาพวีดิทัศน์
2) ศึกษาและวิเคราะห์ปัญหาวิธีการรู้จำนาฬิกาชนิดตัวเลข
3) ศึกษาทฤษฎีและเทคนิคที่เกี่ยวข้องในการวิเคราะห์และประมวลผลภาพดิจิทัล
4) ทดลองนำความรู้ที่ได้ศึกษามาประยุกต์ใช้ในงานวิจัย
5) ออกแบบขั้นตอนวิธีการตรวจหาตำแหน่งและรู้จำเวลาของนาฬิกาตัวเลขในภาพวีดิทัศน์
6) พัฒนาโปรแกรมเพื่อตรวจหาตำแหน่งและรู้จำเวลาของนาฬิกาตัวเลขในภาพวีดิทัศน์จากขั้นตอนวิธีที่ออกแบบ
7) ทดสอบและวิเคราะห์ผล
8) สรุปและวิจารณ์ผลการทดลอง
9) จัดทำรายงาน

1.6 ประโยชน์ที่ได้รับ

1) ได้ขั้นตอนวิธีการและโปรแกรมต้นแบบเพื่อตรวจหาตำแหน่งและรู้จำเวลาของนาฬิกาตัวเลขในภาพวีดิทัศน์อย่างอัตโนมัติ

2) สามารถนำขั้นตอนวิธีที่ได้ไปใช้ร่วมกับงานที่เกี่ยวข้องด้านอื่น

1.7 โครงสร้างของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้มีทั้งหมด 6 บท แต่ละบทประกอบด้วยเนื้อหาดังนี้ บทที่ 1 กล่าวถึงความเป็นมาของปัญหา งานวิจัยที่เกี่ยวข้อง วัตถุประสงค์ ขอบเขต ขั้นตอนการวิจัย และประโยชน์ที่ได้รับ ดังที่ได้กล่าวไว้แล้วในส่วนก่อนหน้านี้ บทที่ 2 กล่าวถึงหลักการและทฤษฎีที่เกี่ยวข้องกับงานวิจัย สำหรับวิธีที่นำเสนอเพื่อแก้ปัญหาแบ่งออกเป็น 2 ส่วนหลักได้แก่ การตรวจหาตำแหน่งนาฬิกาตัวเลขในภาพวีดิทัศน์และการรู้จำนาฬิกาตัวเลข ซึ่งได้กล่าวไว้ในบทที่ 3 และ 4 ตามลำดับ ส่วนบทที่ 5 เป็นการทดลอง ผลการทดลอง และการวิเคราะห์และสรุปผลการทดลอง และบทสุดท้ายคือบทที่ 6 เป็นการสรุปผลการวิจัยและข้อเสนอแนะต่าง ๆ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย