รายการอ้างอิง

1. R. Waranusast, B. Thipakorn and N. Covavisaruch. Automatic Measuring of Red Blood Cell Velocity in Modified Landis Technique Using Spatio-temporal Analysis. Proceedings of Image and Vision Computing New Zealand (2001): 357-362.

2. กฤษดา โชคสินอนันต์. การตรวจจับความเร็วรถยนต์แบบทันกาลโดยใช้การประมวลผลภาพวีดิทัศน์. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2542.

3. R. Lienhart, and A. Wernicke. Localizing and Segmenting Text in Images and Videos. IEEE Transactions on Circuits and Systems for Video Technology 12 (2002): 256-268.

4. A. K. Jain, and B. Yu. Automatic Text Location in Images and Video Frames. Proceedings Fourtennth International Conference on Pattern Recognition 2 (1998): 1497-1499.

5. T. Sato, T. Kanade, E. K. Hughes, M. A. Smith, and S. Satoh. Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions. Multimedia Systems 7 (1999): 385-395.

6. J. Xi, X. S. Hua, X.R. Chen, L. Wenyin, and H. J. Zhang. A Video Text Detection and Recognition System. 2001 IEEE International Conference on Multimedia and Expo (Tokyo) (2001): 1080-1083.

7. M. Cai, J. Song, and M. R. Lyu. A New Approach for Video Text Detection. Proceedings of International Conference on Image Processing 1 (2002): 117-120.

8 Pei Yin, Xian-Sheng Hua, and Hong-Jiang Zhang. Automatic Time Stamp Extraction System for Home Videos. IEEE International Symposium on Circuits and Systems (ISCAS) (2002): 73-76.

9. S. Mori, C. Y. Suen, and K. Yamamoto. Historical Review of OCR Research and Development. Proceedings of the IEEE 80 (1992):1029-1057.

10. R. P. W. Duin. A Note on Comparing Classifiers. Pattern Recognition Letters 17 (1995): 529-536.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

11. R. C. Gonzalez, and R. E. Woods. Digital Image Processing. U.S.A.: Prentice-Hall, Inc., 2002.

12. L. G. Shapiro, and G. C. Stockman. Computer Vision. U.S.A.: Prentice-Hall, Inc., 2001.

13. R. Keys. Cubic Convolution Interpolation for Digital Image Processing. IEEE Transactions on Acoustics, Speech, and Signal Processing 29 (1981): 1153-1160.

14. S. E. Umbaugh. Computer Vision and Image Processing: A Practical Approach Using CVIPtools. U.S.A.: Prentice-Hall, Inc., 1998.

15. R. Jain, R. Kasturi, and B. G. Schunck. Machine Vision. Singapore: McGraw-Hill, Inc., 1995.

16. G. W. Awcock, and R. Thomas. Applied Image Processing. Singapore: McGraw-Hill, Inc., 1996.

17. A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer. Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. IEEE Transactions on Medical Imaging 13 (1994): 716-724.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

โปรแกรมต้นแบบ

ภาคผนวก ก นี้อธิบายโปรแกรมต้นแบบที่ใช้ในการตรวจหาตำแหน่งและรู้จำเวลาของนาฬิกาตัวเลขในภาพวีดิทัศน์แบบอัตโนมัติ หัวข้อในส่วนนี้ประกอบด้วยเงื่อนไขของโปรแกรม และโครงสร้างหน้าจอของโปรแกรม

ก.1 เงื่อนไขของโปรแกรม

โปรแกรมต้นแบบที่พัฒนาขึ้นมีเงื่อนไขการทำงานดังต่อไปนี้

1) โปรแกรมทำงานบนระบบปฏิบัติการวินโดวส์ตั้งแต่รุ่น 98 ขึ้นไป และต้องติดตั้งโปรแกรม DirectX เวอร์ชัน 8.1 ขึ้นไป เพื่อใช้จัดการทางด้านวีดิทัศน์

2) ข้อมูลเข้าของโปรแกรมเป็นไฟล์ชนิด avi หรือ mpeg นอกจากนี้โปรแกรมสามารถรับข้อมูลภาพวีดิทัศน์จากเครื่องเล่นวีดิทัศน์ได้ แต่เครื่องคอมพิวเตอร์จะต้องมีแผ่นวงจรจับเฟรมติดตั้งอยู่

3) ผู้ใช้จะต้องสร้างแผ่นแบบของตัวเลขเพื่อใช้ในการรู้จำนาฬิกาตัวเลข โดยเก็บไว้ในโฟลเดอร์เดียวกัน แผ่นแบบเป็นได้ทั้งไฟล์ชนิด bmp และ jpg และต้องตั้งชื่อไฟล์ของเลข 0 ถึง 9 ในรูปแบบ 00 ถึง 09 ตามลำดับ

ก.2 โครงสร้างหน้าจอของโปรแกรม

โปรแกรมต้นแบบมีหน้าจอหลักแสดงดังรูปที่ ก.1 ประกอบด้วย

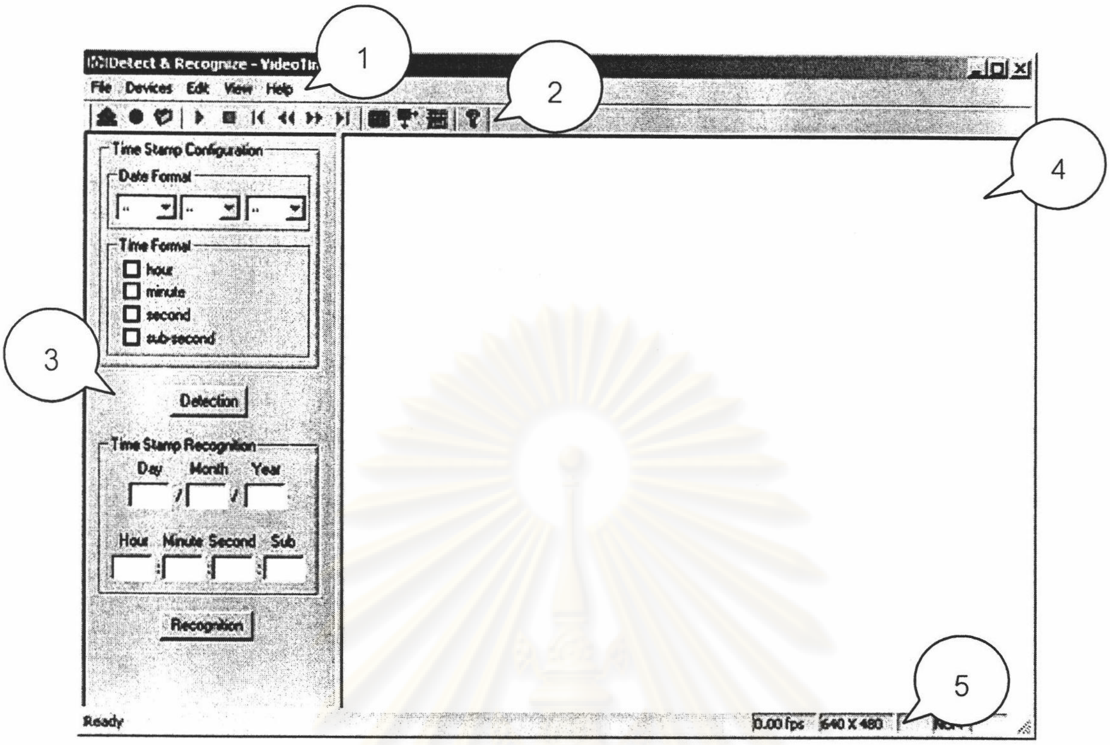
1. เมนู
2. แถบเครื่องมือ
3. หน้าจอสำหรับกำหนดรูปแบบของนาฬิกาและแสดงผลการรู้จำ
4. หน้าจอแสดงภาพวีดิทัศน์
5. แถบแสดงสถานะ

แต่ละส่วนประกอบมีรายละเอียดดังต่อไปนี้

<u>**เมนู**</u>

โปรแกรมประกอบด้วยเมนูหลัก 5 รายการ ได้แก่ File Devices Edit View และ Help แต่ละรายการมีเมนูย่อย ดังนี้

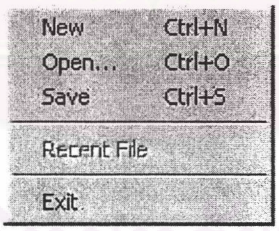
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ ก.1 หน้าจอหลัก

<u>File</u>

เมนู File ประกอบด้วยเมนูย่อยดังรูปที่ ก.2 โดยแต่ละเมนูย่อยมีหน้าที่ดังนี้



รูปที่ ก.2 เมนูย่อยของเมนู File

1. เมนูย่อย New ใช้สำหรับคืนค่าหน่วยความจำที่ใช้ประมวลผล เพื่อที่จะเปิดไฟล์วีดิทัศน์ไฟล์ใหม่
2. เมนูย่อย Open ใช้สำหรับเปิดไฟล์วีดิทัศน์ ซึ่งโปรแกรมรองรับไฟล์วีดิทัศน์ในรูปแบบ avi และ mpeg
3. เมนูย่อย Save ใช้สำหรับบันทึกภาพวีดิทัศน์จากเครื่องเล่นวีดิทัศน์เป็นไฟล์เก็บในเครื่องคอมพิวเตอร์
4. เมนูย่อย Exit ใช้สำหรับปิดโปรแกรม

<u>Devices</u>

เมนู Devices เป็นเมนูที่ใช้แสดงแผ่นวงจรจับเฟรมที่ติดตั้งในเครื่องคอมพิวเตอร์ที่โปรแกรมตรวจหาได้

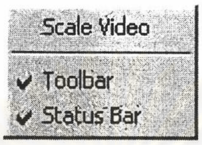
<u>Edit</u>

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เมนู Edit มีเมนูย่อย Setting Template ซึ่งใช้สำหรับให้ผู้ใช้เลือก Folder ที่เก็บแผ่นแบบของตัวเลข เพื่อใช้ในการรู้จำนาฬิกาตัวเลข

View

เมนู View เป็นเมนูที่จัดการเกี่ยวกับการแสดงผลประกอบด้วยเมนูย่อยดังรูปที่ ก.3



รูปที่ ก.3 เมนูย่อยของเมนู View

5. เมนูย่อย Scale Video ใช้สำหรับขยายขนาดภาพวีดิทัศน์ให้เต็มหน้าจอแสดงภาพวีดิทัศน์
6. เมนูย่อย Toolbar ใช้สำหรับปิดหรือเปิดการแสดงผลแถบเครื่องมือ
7. เมนูย่อย Status Bar ใช้สำหรับปิดหรือเปิดการแสดงผลแถบแสดงสถานะ

Help

เมนู Help มีเมนูย่อย About Video Time Stamp ซึ่งแสดงรายละเอียดเกี่ยวกับโปรแกรม

แถบเครื่องมือ

แถบเครื่องมือเป็นปุ่มที่ทำหน้าที่เดียวกับเมนูเพื่อให้ผู้ใช้สะดวกต่อการเรียกใช้ฟังก์ชันต่าง ๆ รวมถึงมีปุ่มสำหรับควบคุมการเล่นวีดิทัศน์ ปุ่มแต่ละปุ่มทำหน้าที่ดังตารางที่ ก.1

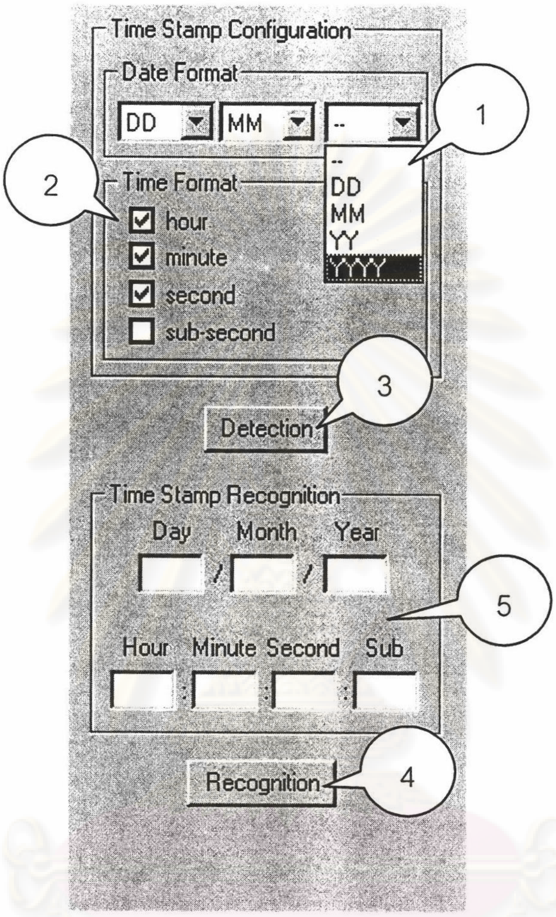
ตารางที่ ก.1 ปุ่มและหน้าที่ของแถบเครื่องมือ

	จับภาพวีดิทัศน์จากเครื่องเล่นวีดิทัศน์
	บันทึกภาพวีดิทัศน์จากเครื่องเล่นวีดิทัศน์เป็นไฟล์เก็บในเครื่องคอมพิวเตอร์
	เปิดไฟล์วีดิทัศน์ ซึ่งโปรแกรมรองรับไฟล์วีดิทัศน์ในรูปแบบ avi และ mpeg
	เล่นไฟล์วีดิทัศน์
	หยุดการเล่นไฟล์วีดิทัศน์
	ไปที่ตำแหน่งเริ่มต้นของไฟล์วีดิทัศน์
	ย้อนหลังจากเฟรมปัจจุบันไป 1 เฟรม
	เดินหน้าจากเฟรมปัจจุบันไป 1 เฟรม
	ไปที่ตำแหน่งสิ้นสุดของไฟล์วีดิทัศน์
	ตั้งค่าแผ่นวงจรจับเฟรม
	ขยายขนาดภาพวีดิทัศน์ให้เต็มหน้าจอแสดงภาพวีดิทัศน์
	ให้ผู้ใช้เลือก Folder ที่เก็บแผ่นแบบของตัวเลข
	แสดงรายละเอียดเกี่ยวกับโปรแกรม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หน้าจอสำหรับกำหนดรูปแบบของนาฬิกาและแสดงผลการรู้จำ

หน้าจอสำหรับกำหนดรูปแบบของนาฬิกาและแสดงผลการรู้จำอยู่ ณ ตำแหน่งทางซ้ายของโปรแกรม รายละเอียดของส่วนประกอบต่าง ๆ แสดงดังรูปที่ ก.4



รูปที่ ก.4 หน้าจอสำหรับกำหนดรูปแบบของนาฬิกาและแสดงผลการรู้จำ

8. ส่วนกำหนดรูปแบบวันที่ มีลักษณะเป็น Drop down box 3 กล่อง เพื่อให้ผู้ใช้กำหนดรูปแบบของวันที่ว่าประกอบด้วยวัน เดือน หรือปี และเรียงลำดับอย่างไร โดยแต่ละกล่องมีรายการดังนี้ - - หมายถึงไม่กำหนดรูปแบบ DD หมายถึง วัน MM หมายถึง เดือน YY หมายถึง ปี 2 หลัก และ YYYY หมายถึง ปี 4 หลัก

9. ส่วนกำหนดรูปแบบเวลา มีลักษณะเป็น Check box เพื่อให้ผู้ใช้กำหนดว่าวีดิทัศน์มีเวลาในรูปแบบใด ประกอบด้วย Check box ของชั่วโมง นาที วินาที และเศษของวินาที

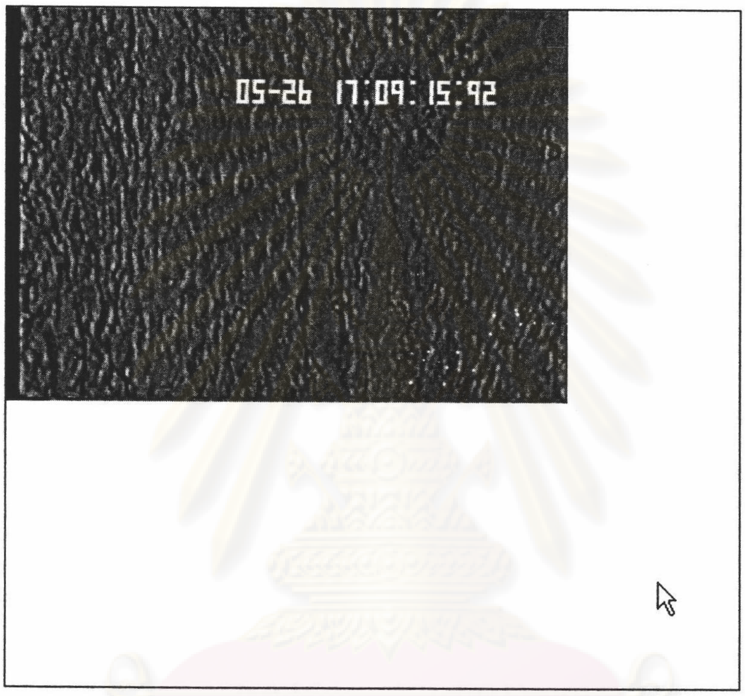
10. ปุ่มสำหรับตรวจหาตำแหน่งนาฬิกาตัวเลข เป็นปุ่มสำหรับประมวลผลวีดิทัศน์เพื่อตรวจหาตำแหน่งของนาฬิกาตัวเลข

11. ปุ่มสำหรับรู้จำนาฬิกาตัวเลข เป็นปุ่มเพื่อใช้รู้จำนาฬิกาตัวเลข ณ เฟรมปัจจุบัน หลังจากที่ทำการตรวจหาของนาฬิกาตัวเลขแล้ว

12. ส่วนแสดงผลการรู้จำ เป็น Edit box ที่ใช้แสดงผลการรู้จำนาฬิกาตัวเลข

<u>หน้าจอแสดงภาพวีดิทัศน์</u>

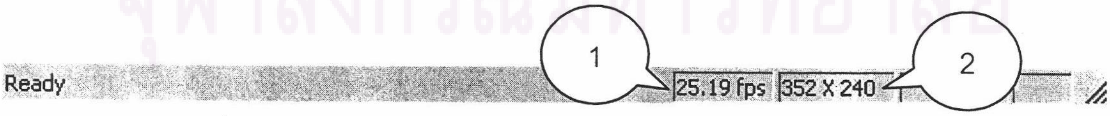
หน้าจอแสดงภาพวีดิทัศน์อยู่ ณ ตำแหน่งทางด้านขวาของโปรแกรม ทำหน้าที่แสดงภาพวีดิทัศน์ที่ผู้ใช้นำเข้าโปรแกรมทั้งจากไฟล์วีดิทัศน์และจากเครื่องเล่นวีดิทัศน์ รูปที่ ก.5 แสดงตัวอย่างการแสดงภาพวีดิทัศน์



รูปที่ ก.5 หน้าจอแสดงภาพวีดิทัศน์

<u>แถบสถานะ</u>

แถบสถานะอยู่ ณ ตำแหน่งด้านล่างสุดของโปรแกรม ทำหน้าที่แสดงข้อมูลของภาพวีดิทัศน์ว่ามีอัตราการจับเฟรมเท่าไร และภาพวีดิทัศน์มีขนาดเท่าไร จากรูปที่ ก.6 หมายเลข 1 แสดงอัตราการจับเฟรม และหมายเลข 2 แสดงขนาดของภาพวีดิทัศน์



รูปที่ ก.6 แถบสถานะแสดงอัตราการจับเฟรมและขนาดของภาพวีดิทัศน์

ภาคผนวก ข

บทความที่นำเสนอในงานการประชุมวิชาการ

1. บทความเรื่อง "การตรวจหาตำแหน่งตัวเลขบอกเวลาในภาพวีดิทัศน์" นำเสนอในงานประชุมวิชาการ The 7^{th} National Computer Science and Engineering Conference 2003 จัดที่มหาวิทยาลัยบูรพา จังหวัดชลบุรี ระหว่างวันที่ 28-30 ตุลาคม พ.ศ. 2546 บทความนี้ตีพิมพ์ไว้ใน Proceedings of The 7^{th} National Computer Science and Engineering Conference 2003 หน้า 13 – 18
2. บทความเรื่อง "Time Stamp Detection and Recognition in Video Frames" นำเสนอในงานประชุมวิชาการ The 2004 International Conference on Imaging Science, Systems, and Technology จัดที่ Las Vegas, Nevada, USA ระหว่างวันที่ 21-24 มิถุนายน พ.ศ. 2547 บทความนี้อยู่ในระหว่างรอตีพิมพ์ไว้ใน The 2004 International Conference on Imaging Science, Systems, and Technology

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การตรวจหาตำแหน่งตัวเลขบอกเวลาในภาพวีดิทัศน์
Time Stamp Detection in Video Images

เจษฎา แสงพานิชย์ และ นงลักษณ์ โควาวิสารัช
Chetsada Saengpanit and Nongluk Covavisaruch
Department of Computer Engineering, Faculty of Engineering
Chulalongkorn University, Bangkok 10330, Thailand
e-mail: chetsada.s@student.chula.ac.th and nongluk.c@chula.ac.th

บทคัดย่อ

การตรวจหาตำแหน่งตัวเลขบอกเวลาในภาพวีดิทัศน์เป็นขั้นตอนสำคัญในการทำระบบการทำดัชนีและการค้นคืนภาพวีดิทัศน์ โดยใช้เวลาเป็นดัชนีในการสืบค้น งานวิจัยนี้จึงนำเสนอขั้นตอนวิธีในการตรวจหาตำแหน่งตัวเลขบอกเวลาในภาพวีดิทัศน์โดยใช้สมบัติความเกี่ยวเนื่องกันตามเวลา (Temporal property) ของภาพวีดิทัศน์และใช้การวิเคราะห์เท็กซ์เจอร์จากลักษณะของตัวเลขซึ่งมีความแตกต่างจากส่วนอื่น ๆ ในภาพ จากผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธีที่นำเสนอสามารถตรวจหาตำแหน่งตัวเลขบอกเวลาแต่ละตัวได้อย่างถูกต้องโดยใช้จำนวนเฟรมวีดิทัศน์ที่น้อย

คำสำคัญ: การตรวจหาตำแหน่งตัวเลขบอกเวลา ภาพวีดิทัศน์

Abstract

Time stamp detection is a fundamental step to video indexing and retrieval systems which use time stamps as the indexes. In this paper, we propose an approach to detect time stamp in video images by applying temporal properties of video frames. Furthermore, we analyse texture features of time stamp digit font which are different from other objects' in the image. Experimental results show that our approach can detect time stamp accurately with only a few video frames.

Key-Words: Time stamp detection, video images

1. บทนำ

ตัวเลขบอกเวลาในภาพวีดิทัศน์เป็นข้อมูลที่ถูกบันทึกเพิ่มเติมลงในภาพและเป็นข้อมูลสำคัญที่ทำให้ทราบวันที่และเวลาที่ทำการบันทึกภาพ โดยได้ถูกนำไปใช้ประโยชน์ต่าง ๆ นอกจากจะเป็นหลักฐานที่แสดงถึงเวลาของเหตุการณ์ที่เกิดขึ้นในภาพแล้ว ยังสามารถใช้เป็นข้อมูลในการจับกลุ่มการถ่ายภาพแต่ละครั้ง (Shot clustering) ซึ่งนำไปทำระบบการทำดัชนีและการค้นคืนภาพวีดิทัศน์ได้ และการตรวจหาตำแหน่งของตัวเลขบอกเวลาเป็นขั้นตอนแรกที่มีความสำคัญต่อการสร้างระบบดังกล่าว

กล้องถ่ายวีดิทัศน์แบบดิจิทัลจะเก็บส่วนที่เป็นตัวเลขบอกเวลาแยกกับส่วนที่เป็นข้อมูลวีดิทัศน์ ดังนั้นจึงแยกข้อมูลที่เป็นตัวเลขบอกเวลาออกมาได้ง่าย ขณะที่ตัวเลขบอกเวลาของกล้องถ่ายวีดิทัศน์แบบแอนะล็อกจะถูกเก็บรวมไว้กับข้อมูลวีดิทัศน์ โดยจะนำตัวเลขบอกเวลาวางซ้อนทับ (Superimpose) บนภาพวีดิทัศน์ ทำให้ไม่สามารถแยกตัวเลขบอกเวลาออกจากวีดิทัศน์ได้โดยตรง [1] และในปัจจุบันยังนิยมเก็บภาพวีดิทัศน์ไว้ในม้วนเทป ทำให้ภาพวีดิทัศน์มีคุณภาพต่ำ เนื่องจากมีความละเอียดต่ำและมีสัญญาณรบกวนอยู่มาก นอกจากนี้พื้นหลังของภาพวีดิทัศน์จะมีความซับซ้อนและบางครั้งอาจมีสีใกล้เคียงกับสีของตัวเลขบอกเวลา ดังนั้นการตรวจหาตำแหน่งตัวเลข

บอกเวลาในภาพวีดิทัศน์นั้นจึงเป็นปัญหาที่ยากและน่าหาวิธีแก้ปัญหา

Pei Yin และคณะ [1] ได้คิดวิธีการหาตัวเลขบอกเวลาของภาพวีดิทัศน์โดยใช้ Spatial-Temperal Suppression (STS) เพื่อให้ตัวเลขซึ่งมีสีขาวเด่นออกมาจากพื้นหลัง จากนั้นหาตำแหน่งตัวเลขบอกเวลาอย่างหยาบโดยพิจารณาภาพเงาการฉาย (Projection profile) ทั้งแนวนอนและแนวตั้งของจุดภาพที่คาดว่าเป็นตัวเลขบอกเวลา นั่นคือจุดภาพที่มีค่ามากกว่าค่าขีดแบ่งค่าหนึ่ง ได้ผลลัพธ์เป็น Bounding box ของตัวเลขแต่ละตัว แล้วจึงปรับขนาด Bounding box ของตัวเลขแต่ละตัวโดยใช้ค่าเฉลี่ยของความกว้างและความสูงของ Bounding box ทุกอันมาเป็นตัวแทนความกว้างและความสูงของ Bounding box แต่ละอัน อย่างไรก็ตามวิธีดังกล่าวอาจมีปัญหาในกรณีที่พื้นหลังของตัวเลขบอกเวลาเป็นสีขาวหรือสีที่สว่างซึ่งจะทำให้หา Bounding box ของตัวเลขบอกเวลาผิดพลาดได้

งานวิจัยอื่นที่เกี่ยวข้องได้แก่งานวิจัยทางด้านการตรวจหาตำแหน่งข้อความในภาพวีดิทัศน์ [2-7] วิธีที่นิยมใช้แบ่งออกเป็น 2 ประเภทหลักคือใช้การวิเคราะห์เท็กซ์เจอร์ (Texture analysis) [2-6] และใช้การวิเคราะห์ส่วนประกอบที่เชื่อมต่อกัน (Connected component analysis) [7]

งานวิจัยนี้เสนอขั้นตอนวิธีการตรวจหาตำแหน่งตัวเลขบอกเวลาในภาพวีดิทัศน์โดยใช้สมบัติความเกี่ยวเนื่องกันตามเวลาของภาพวีดิทัศน์และใช้การวิเคราะห์เท็กซ์เจอร์จากลักษณะของตัวเลข โดยจะกล่าวถึงขั้นตอนวิธีการตรวจหาตำแหน่งตัวเลขบอกเวลาในหัวข้อที่ 2 การทดลองและผลการทดลองในหัวข้อที่ 3 และในหัวข้อที่ 4 จะแสดงการวิเคราะห์และสรุปผลการทดลอง

2. ขั้นตอนวิธีการตรวจหาตำแหน่งตัวเลขบอกเวลา

ภาพวีดิทัศน์ที่มีตัวเลขบอกเวลาในภาพโดยส่วนใหญ่จะถ่ายทำในลักษณะที่เป็น Home video ซึ่งพื้นหลังของเฟรมแต่ละเฟรมมีการเปลี่ยนแปลงอยู่ตลอดเวลา ในขณะที่ตัวเลขบอกเวลาจะอยู่ ณ ตำแหน่งเดิม ดังนั้นจึงสามารถใช้สมบัติความเกี่ยวเนื่องกันตามเวลาของภาพวีดิทัศน์ได้

เมื่อวิเคราะห์ลักษณะของตัวเลขบอกเวลาในภาพวีดิทัศน์ที่ใช้กันทั่วไปแล้วพบว่าตัวเลขมีสีโทนสีขาวและมีขอบเป็นสีดำ ดังนั้นงานวิจัยนี้จึงใช้เท็กซ์เจอร์ขอบภาพ [2-6] ในการหาตำแหน่งของตัวเลขบอกเวลาอย่างหยาบ โดยนำภาพวีดิทัศน์ที่เป็นเฟรมติดกันจำนวน 5 ภาพมาหาขอบภาพตามแนวตั้ง แล้วขยายขนาด (Dilation) ขอบภาพที่ได้ จากนั้นนำผลลัพธ์ที่ได้จากภาพทั้ง 5 ภาพมาเข้าตรรกะแอนด์ (Logical AND) ภาพที่ได้จะเหลือบริเวณที่เป็นตัวเลขบอกเวลาและส่วนที่เป็นขอบภาพของพื้นหลังอีกเล็กน้อย ทำให้สามารถหา Bounding box ของตัวเลขบอกเวลาอย่างหยาบโดยวิเคราะห์จากภาพเงาการฉายทั้งตามแนวนอนและแนวตั้ง

ขั้นตอนต่อไปเป็นการหา Bounding box ของตัวเลขบอกเวลาอย่างละเอียด เพื่อสร้างเป็นแม่แบบของตำแหน่งตัวเลขบอกเวลา (Template of time stamp location) โดยพยายามกำจัดพื้นหลังออกจากส่วนที่เป็นตัวเลข แล้วหาภาพเงาการฉายอีกครั้งเพื่อให้ได้ Bounding box ของตัวเลขแต่ละตัวอย่างถูกต้อง

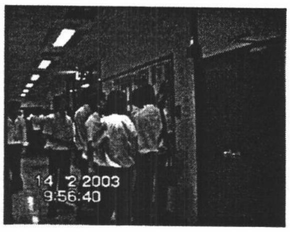
รายละเอียดของแต่ละขั้นตอนที่นำเสนอมีดังต่อไปนี้

2.1 การตรวจหาขอบภาพตามแนวตั้ง

ในงานวิจัยนี้ตรวจหาขอบภาพตามแนวตั้งเท่านั้นโดยไม่ได้ตรวจหาขอบภาพตามแนวนอนเนื่องจากขอบภาพตามแนวตั้งเป็นข้อมูลที่เพียงพอต่อการตรวจหาตำแหน่งตัวเลขบอกเวลา นอกจากนี้การตรวจหาขอบภาพตามแนวนอนนอกจากจะเพิ่มการประมวลผลแล้วยังทำให้เกิดขอบภาพในบริเวณที่เป็นพื้นหลังซึ่งยากต่อการกำจัดอีกด้วย

การหาขอบภาพตามแนวตั้งทำโดยใช้ตัวดำเนินการ Sobel กับภาพในแต่ละระนาบสี (R G และ B) หลังจากนั้นทำขอบภาพให้บางด้วยวิธี Non-maxima suppression [8] วิธีนี้เป็นการกำจัดจุดภาพที่มีค่าต่ำกว่าจุดภาพใกล้เคียง

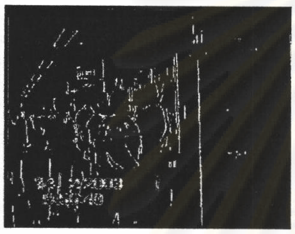
(Neighbors of a pixel) ซึ่งมีพิกัดในแนวทิศทางของเกรเดียนต์ (Gradient) จะได้ภาพขอบที่เป็นภาพลักษณ์ฐานสองและขอบภาพมีความหนา 1 จุดภาพ นำผลลัพธ์ที่ได้ในแต่ละระนาบมายูเนียน (Union) กัน รูปที่ 1 แสดงตัวอย่างการหาขอบภาพตามแนวตั้ง



(ก) ภาพต้นฉบับ

(ข) ตัวอย่างระนาบสีแดงของรูป (ก)

(ค) ภาพขอบแนวตั้งรูป (ข)

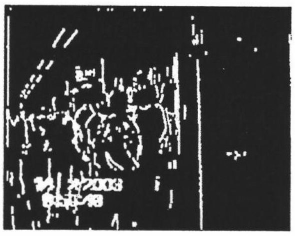


(ง) Non-maxima suppression ของรูป (ค)

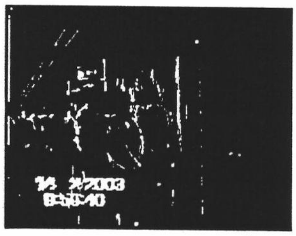
รูปที่ 1. ตัวอย่างการหาขอบภาพตามแนวตั้ง

2.2 การกำจัดขอบภาพที่ไม่ใช่ส่วนของตัวเลขบอกเวลา

การกำจัดขอบภาพที่ไม่ใช่ส่วนของตัวเลขบอกเวลาทำโดยนำภาพผลลัพธ์จากขั้นตอนที่ 2.1 มาขยายขนาด (Dilation) ด้วยสมาชิกโครงสร้าง (Structure element) แบบสี่เหลี่ยมขนาด 3x3 เนื่องจากต้องการนำผลลัพธ์ที่ได้จากขั้นตอนนี้ไปแอนด์ภาพขอบที่ได้มาจากภาพวีดิทัศน์ในเฟรมติดกันอีก 4 ภาพ ถ้าไม่ทำการขยายขนาด อาจทำให้บริเวณที่เป็นตัวเลขบอกเวลาหายไป การแอนด์เป็นการนำค่าของจุดภาพที่ตำแหน่งเดียวกันของภาพหลายภาพมาเข้าตรรกะแอนด์ จุดภาพที่เป็นภาพผลลัพธ์จะมีค่าเป็น 1 ก็ต่อเมื่อจุดภาพทุก ๆ จุดที่เป็นข้อมูลนำเข้ามีค่าเท่ากับ 1 ภาพผลลัพธ์ที่ได้หลังจากการขยายขนาดแสดงดังรูปที่ 2(ก) ส่วนรูปที่ 2(ข) เป็นภาพที่ได้จากการนำเฟรมติดกันทั้ง 5 ภาพที่ผ่านขั้นตอนการตรวจหาขอบภาพและการขยายขนาดแล้วมาแอนด์กัน



(ก) ภาพจากการขยายขนาด



(ข) ภาพจากการแอนด์

รูปที่ 2. ตัวอย่างการกำจัดขอบภาพ

2.3 การตรวจหาตำแหน่งของตัวเลขบอกเวลาอย่างหยาบ

การตรวจหาตัวเลขบอกเวลาอย่างหยาบเป็นการหาตำแหน่งของตัวเลขอย่างคร่าว ประกอบด้วย 2 ขั้นตอนดังนี้

2.3.1 การหาขอบเขตบนและล่างของตัวเลขบอกเวลา

การตรวจหาตำแหน่งตัวเลขบอกเวลาในภาพวีดิทัศน์แสดงให้เห็นได้โดยตีกรอบสี่เหลี่ยมล้อมรอบบริเวณที่เป็นตัวเลขบอกเวลา ขั้นตอนนี้เป็นการหาขอบเขตบนและล่างของตัวเลขบอกเวลาโดยหาภาพเงาการฉายตามแนวนอนกับภาพผลลัพธ์ในขั้นตอนที่ 2.2 ดังแสดงในรูปที่ 3

(ก) ภาพจากการแอนด์

(ข) ภาพเงาการฉายตามแนวนอน

รูปที่ 3. ตัวอย่างภาพเงาการฉายตามแนวนอน

โดยทั่วไป ค่าความถี่ในฮีสโทแกรมจากการทำภาพเงาการฉายของผลลัพธ์ในขั้นตอนที่ 2.2 จะมีค่าสูงในแถวที่มีตัวเลขบอกเวลา ดังนั้นการหาขอบเขตบนและล่างของตัวเลขบอกเวลาโดยใช้ค่าขีดแบ่งกับภาพเงาการฉาย [3, 4, 6] จึงเป็นวิธีที่นิยมใช้กัน

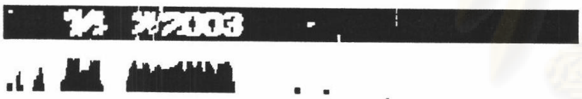
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

งานวิจัยนี้ปรับฮีสโทแกรมให้เรียบโดยใช้ตัวกรองมัธยฐานแล้วจึงหาค่าขีดแบ่ง T_{tb} โดยคำนวณจากค่าเฉลี่ย m ของค่าความถี่ในฮีสโทแกรมคูณด้วยค่าคงที่ค่าหนึ่ง k ดังสมการที่ 1 และกำหนดให้ตำแหน่งของความถี่ที่มีค่ามากกว่าค่าขีดแบ่งเป็นแถวของภาพที่มีตัวเลขบอกเวลาปรากฏอยู่ ทำให้สามารถหาขอบเขตบนและล่างของตัวเลขบอกเวลาได้

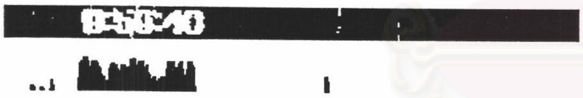
$$T_{tb} = k \times m \qquad (1)$$

2.3.2 การหาขอบเขตซ้ายและขวาของตัวเลขบอกเวลา

เมื่อทราบตำแหน่งบรรทัดของตัวเลขบอกเวลาแล้ว นั่นคือขอบเขตบนและล่างที่หาได้จากขั้นตอนที่แล้ว สามารถหาขอบเขตซ้ายและขวาของตัวเลขบอกเวลาได้โดยพิจารณาภาพเงาการฉายตามแนวตั้งในแต่ละบรรทัดของตัวเลขบอกเวลา ดังแสดงในรูปที่ 4 โดยรูปที่ 4(ก) แสดงบรรทัดของวันที่ ส่วนรูปที่ 4(ข) แสดงบรรทัดของเวลา และรูปล่างเป็นภาพเงาการฉายตามแนวตั้งของรูปบน



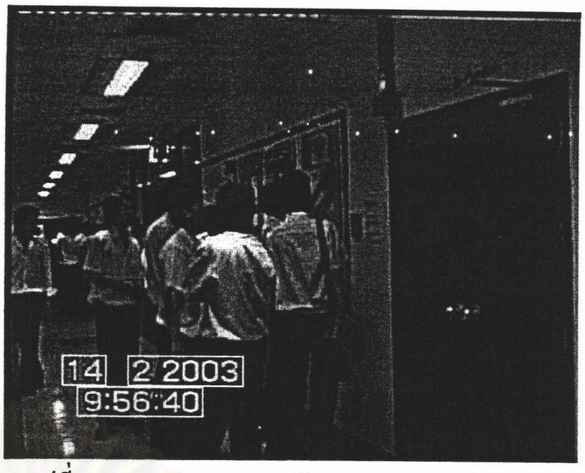
(ก) บรรทัดของวันที่



(ข) บรรทัดของเวลา

รูปที่ 4. ภาพเงาการฉายตามแนวตั้ง

การหาขอบภาพทำเฉพาะขอบภาพแนวตั้งทำให้ความสูงของความถี่ในฮีสโทแกรมจากภาพเงาการฉายของส่วนที่เป็นตัวเลขบางตำแหน่งมีค่าไม่มาก นอกจากนี้อาจมีส่วนที่ไม่ใช่ตัวเลขปรากฏอยู่ในบรรทัดของตัวเลขบอกเวลา ดังนั้นจึงกำหนดค่าขีดแบ่งให้มีค่าเท่ากับ 1 ใน 3 ของความสูงของแต่ละบรรทัด ผลลัพธ์ที่ได้จากขั้นตอนนี้เป็น Bounding box ที่ครอบคลุมตัวเลขบอกเวลาอย่างหยาบ ดังแสดงในรูปที่ 5



รูปที่ 5. Bounding box ของตัวเลขบอกเวลาอย่างหยาบ

2.4 การตรวจหาตำแหน่งของตัวเลขบอกเวลาอย่างละเอียด

การตรวจหาตำแหน่งของตัวเลขบอกเวลาอย่างละเอียดจะใช้ลักษณะของตัวเลขบอกเวลาซึ่งมีตัวเลขเป็นสีขาวมากกว่าค่าขีดแบ่ง T_{font} มีขอบของตัวเลขเป็นสีดำต่ำกว่าค่าขีดแบ่ง T_{edge} โดยสีขาวในภาพมีค่ามีค่าระดับเทามากกว่าสีดำ นอกจากนี้กำหนดให้ตัวเลขมีความหนาไม่เกินค่าขีดแบ่ง T_{stroke}

เนื่องจาก Bounding box ที่ได้จากขั้นตอนที่ 2.3.2 ซึ่งเป็น Bounding box อย่างหยาบ อาจครอบคลุมตัวเลขมากกว่า 1 ตัว ดังนั้น Bounding box แต่ละอันที่หาได้จะนำไปประมวลผลดังต่อไปนี้

- พิจารณาภาพต้นฉบับทั้ง 5 ภาพเฉพาะบริเวณที่เป็นขอบเขตของ Bounding box
- แปลงภาพสีเป็นภาพระดับเทาทั้ง 5 ภาพ
- ทำ Temperal Minimum Suppression [9] เพื่อลดความสว่างของพื้นหลังของภาพ โดยนำค่าระดับเทาของจุดภาพที่ตำแหน่งเดียวกันของภาพทั้ง 5 ภาพมาคำนวณหาค่าต่ำที่สุด แล้วแทนค่าไปในภาพผลลัพธ์
- พิจารณาจุดภาพแต่ละจุดว่าเป็นส่วนของตัวเลขบอกเวลาหรือไม่ตามรหัสเทียม (Psuedo code) ดังแสดงในรูปที่ 6 โดย *I(x, y)* คือค่าของจุดภาพ ณ ตำแหน่ง *(x, y)* ผลลัพธ์ที่ได้จะเป็นภาพของตัวเลขแต่ละตัวแยกจากกันเนื่องจากวิธีดังกล่าวสามารถกำจัดพื้นหลังได้ดี แม้พื้นหลังจะมีสีโทนสว่างคล้ายกับตัวเลขบอกเวลาก็ตาม

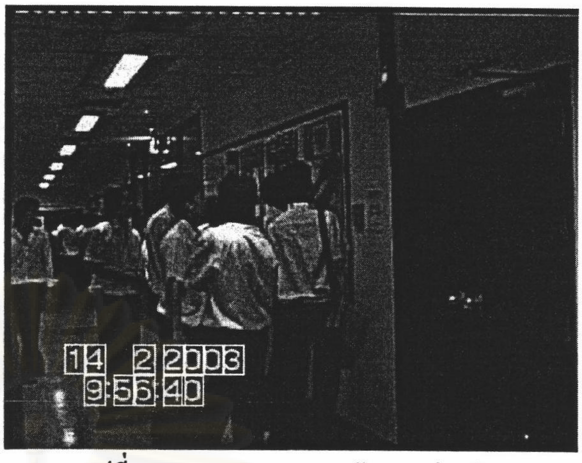
ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

- หา Bounding box ของตัวเลขแต่ละตัว โดยวิเคราะห์จากภาพเงาการฉายตามแนวตั้ง จะได้ Bounding box ของตัวเลขในแถวเดียวกันมีความสูงเท่ากัน แต่มีความกว้างต่างกัน ดังแสดงในรูปที่ 7

```
ถ้า I(x,y) < T_font แล้ว
  I(x,y) = 0;
มิฉะนั้น
  - นับจำนวนจุดภาพจากจุดนั้นไปทางด้านบนจนพบ
    จุดภาพที่มีค่าน้อยกว่า T_edge ให้จำนวนจุดภาพที่นับได้
    เท่ากับ Count_top
  - นับจำนวนจุดภาพจากจุดนั้นไปทางด้านล่างจนพบ
    จุดภาพที่มีค่าน้อยกว่า T_edge ให้จำนวนจุดภาพที่นับได้
    เท่ากับ Count_bottom
  ถ้า Count_top + Count_bottom < T_stroke แล้ว
    I(x,y) = 1;
  จบ
มิฉะนั้น
  - นับจำนวนจุดภาพจากจุดนั้นไปทางด้านซ้ายจนพบ
    จุดภาพที่มีค่าน้อยกว่า T_edge ให้จำนวนจุดภาพที่นับได้
    เท่ากับ Count_left
  - นับจำนวนจุดภาพจากจุดนั้นไปทางด้านขวาจนพบ
    จุดภาพที่มีค่าน้อยกว่า T_edge ให้จำนวนจุดภาพที่นับได้
    เท่ากับ Count_right
  ถ้า Count_left + Count_right < T_stroke แล้ว
    I(x,y) = 1;
  จบ
มิฉะนั้น
  I(x,y) = 0;
จบ
```

รูปที่ 6. รหัสเทียมเพื่อใช้พิจารณาจุดภาพว่าเป็นส่วนของตัวเลขบอกเวลาหรือไม่

จากนั้นจะปรับความกว้างของ Bounding box แต่ละอันให้เท่ากัน เนื่องจากต้องการสร้างแม่แบบของตำแหน่งตัวเลขบอกเวลาให้สามารถใช้ในภาพวีดิทัศน์ในเฟรมถัด ๆ ไปได้ โดยไม่ต้องคำนวณหาตำแหน่งใหม่ จะใช้ความกว้างของ Bounding box ที่มีความกว้างมากที่สุด เป็นตัวแทน Bounding box แต่ละอัน การปรับขนาดของ Bounding box แต่ละอันจะขยายจากตำแหน่งกึ่งกลางของ Bounding box ออกทางด้านข้าง ข้างละครึ่งหนึ่งของความกว้างที่เป็นตัวแทน ผลลัพธ์ที่ได้แสดงดังรูปที่ 8



รูปที่ 8. Bounding box ของตัวเลขแต่ละตัว (หลังปรับความกว้าง)

3. การทดลองและผลการทดลอง

งานวิจัยนี้ได้ทดสอบขั้นตอนวิธีที่นำเสนอกับภาพวีดิทัศน์ที่นำมาจากแหล่ง 3 แหล่ง ภาพทดสอบมีความละเอียด 288x352 จุดภาพ จำนวนทั้งสิ้น 20 ชุดภาพ ชุดภาพละ 5 ภาพ โดยภาพของแต่ละชุดภาพเป็นภาพที่ได้จากเฟรมติดกันในวีดิทัศน์ มีตัวเลขบอกเวลาที่ทำการตรวจหาทั้งสิ้น 225 ตัว ผลลัพธ์มีความถูกต้องก็ต่อเมื่อ Bounding box ครอบคลุมตัวเลขทั้งตัวและครอบคลุมตัวเลขเพียงตัวเดียวเท่านั้น มิฉะนั้นจะถือว่าเป็นการตรวจหาไม่พบ (Miss) นอกจากนี้ถ้ามี Bounding box อยู่ในตำแหน่งที่ไม่มีตัวเลขบอกเวลาปรากฏในภาพ จะถือว่า Bounding box นั้นเป็นสัญญาณเตือนที่ผิดพลาด (False alarm)

ผลการทดลองแสดงดังตารางที่ 1 การตรวจหามีความถูกต้อง 92.22% ตรวจหาไม่พบ 7.78% และมีสัญญาณเตือนที่ผิดพลาด 8.44%

ตารางที่ 1. ผลการทดลอง

ตำแหน่งตัวเลขที่ตรวจหา	ความถูกต้อง	การตรวจหาไม่พบ	สัญญาณเตือนที่ผิดพลาด
255	92.22%	7.78%	8.44%

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เมื่อสร้างแม่แบบของตำแหน่งตัวเลขบอกเวลาแล้วได้ทดลองนำแม่แบบที่ได้ไปใช้กับภาพวีดิทัศน์ในเฟรมถัด ๆ ไป โดยนำแม่แบบของตำแหน่งตัวเลขบอกเวลาที่หาได้ในรูปที่ 8 มาใช้กับภาพในรูปที่ 9

รูปที่ 9. การนำแม่แบบของตำแหน่งตัวเลขบอกเวลามาใช้

4. วิเคราะห์และสรุปผล

งานวิจัยนี้นำเสนอขั้นตอนวิธีการตรวจหาตำแหน่งตัวเลขบอกเวลาที่สามารถตรวจหาตำแหน่งได้ดีด้วยจำนวนภาพวีดิทัศน์ที่น้อย เหมาะสำหรับภาพที่ถ่ายทำในลักษณะที่เป็น Home video ซึ่งพื้นหลังมีการเปลี่ยนแปลงค่อนข้างมากในช่วงเวลาที่สั้น ๆ ขั้นตอนวิธีที่นำเสนอสามารถจัดการกับภาพที่มีสีโทนสว่างในบริเวณที่เป็นตัวเลขบอกเวลาได้ อย่างไรก็ตามขั้นตอนวิธีที่นำเสนออาจเกิดความผิดพลาดได้ในกรณีที่พื้นหลังมีรายละเอียดมากซึ่งทำให้มีขอบภาพจำนวนมาก เช่น ภาพต้นไม้ เป็นต้น นอกจากนี้การนำแม่แบบของตำแหน่งตัวเลขบอกเวลาที่หาได้มาใช้กับภาพวีดิทัศน์ในเฟรมถัด ๆ ไปอาจมีปัญหาถ้านาฬิกามีการเปลี่ยนแปลงแม่แบบ ซึ่งจะต้องมีการตรวจหาการเปลี่ยนแปลงนี้ด้วย

งานที่จะทำต่อไปในอนาคตเป็นการหาขั้นตอนวิธีในการแยกตัวเลขออกจากพื้นหลังเพื่อเข้าสู่กระบวนการรู้จำตัวเลขนาฬิกาต่อไป

5. เอกสารอ้างอิง

[1] Pei Yin, Xian-Sheng Hua, and Hong-Jiang Zhang, "Automatic Time Stamp Extraction System for Home Videos," IEEE International Symposium on Circuits and Systems (ISCAS 2002), 2002, pp. 73-76.

[2] Min Cai, Jiqiang Song, and Michael R. Lyu, "A New Approach for Video Text Detection," Proceedings of International Conference on Image Processing, 2002, vol.1, pp. 117-120.

[3] R. Lienhart, and A. Wernicke, "Localizing and Segmenting Text in Images and Videos," IEEE Transactions on Circuits and Systems for Video Technology, 2002, vol. 12, no. 4, pp. 256 -268.

[4] Jie Xi, Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, and Hong-Jiang Zhang, "A Video Text Detection and Recognition System," IEEE International Conference on Multimedia and Expo (ICME 2001), 2001, pp. 1080-1083.

[5] X. R. Chen, H. J. Zhang, "Text Area Detection form Video Frames," IEEE Pacific-Rim Conference on Multimedia (PCM2001), 2001, pp. 222-228.

[6] Y. Ariki, and K. Matsuura, "Automatic Classification of TV News Articles Based on Telop Character Recognition," IEEE International Conference on Multimedia Computing and Systems, 1999, vol. 2, pp. 148-152.

[7] A. K. Jain, and B. Yu, "Automatic Text Location in Images and Video Frames," Proceedings of International Conference on Pattern Recognition, 1998, vol. 2, pp. 1497-1499.

[8] J. Canny, "A Computational Approach to Edge Detection," IEEE Trans. Pattern Anal. And Machine Intelligence, 1986, vol. 8, no. 6, pp. 679-698.

[9] T. Sato, T. Kanade, E. Hughes, M. Smith, and S. Satoh, "Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions," ACM Multimedia Systems Special Issue on Video Libraries, 1999, vol. 7, pp. 385-395.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Time Stamp Detection and Recognition in Video Frames

Nongluk Covavisaruch and Chetsada Saengpanit
Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand
E-mail: nongluk.c@chula.ac.th and chetsada.s@student.chula.ac.th

Abstract

Time stamp superimposed on the video frames can be used as indices for video indexing and retrieval systems. Some applications use it for speed detection of objects. In this paper, we propose a time stamp detection approach based on edge detection method. Furthermore, we apply temporal properties of video frames so that time stamp can be located quickly. We also propose our digit recognition module to test the performance. Experimental results show that our approach can detect time stamp location accurately. The proposed approach can discriminate between date and time, and it can extract time with sub-second of precision. The recognition rate is also high enough to be used practically.

1. Introduction

Time stamp does not only show a recording time of the videos, it is also used beneficially in many applications. Shot clustering, for example, stores time stamps as indices for video indexing and retrieval systems. Moreover, in speed detection applications, time stamp can be utilized to time a moving object. Therefore, time stamp detection and recognition is an important step before developing these applications.

For digital video camcorders, we can acquire the time stamp directly because all information is saved in digital format. Although nowadays, digital video camcorders are used popularly because they are not so expensive, however, most of the media to store the video are still video cassettes. So, if we want to utilize the time stamp, we have to record by superimposing it on the video frames.

Extracting the time stamp from a video image is a difficult task. Video images are usually low quality, contain noises and have low resolution. It is mainly because recording the videos in video cassettes is an analog system. In addition, some frames might contain complex backgrounds or, sometimes the colors between the time stamp and the background are not significantly different.

Yin et al. [1] have proposed an automatic time stamp extraction system for home videos by using Spatial-Temporal Suppression (STS) in order to darken the background and to make time stamp remarkable. After rough segmentation, time stamp template is generated for recognition. They have applied a shot boundary detection to test the system, and the time stamps are extracted as indices at the boundary frames. It is reported that one-minute error is acceptable because it is precise enough for further video processing.

Other related works are about text detection and recognition from images and video frames. The approaches to detect text in video can be basically divided into three categories. The first one is connected-component-based method [2], which is efficient when the background is mainly composed of homogeneous color. The second category is texture-analysis-based method [3-5], which uses texture features in classifying between text and non-text. Generally, texture-analysis-based method is time-consuming and inaccurate to find text positions. The last one is edge-based method [6-8]. Since text regions are typically rich of edges, projection profiles are applied to analyze the edge density.

In this paper, we apply text detection method to locate time stamp based on edge detection. Besides, we utilize temporal properties of video frames so that time stamp can be located rapidly and accurately. Moreover, we propose an algorithm for segmenting time stamp digit from the background that is also useful in character segmentation. In recognition step, we propose our simple digit recognition module by using decision tree and similarity index [10] to compare the similarity between the ground truth and the unknown digit.

The remainder of this paper is organized as follows. Time stamp detection is explained in Section 2. In Section 3, we propose an algorithm for segmenting time stamp from the background whereas the time

stamp recognition is described in Section 4. Experimental results and conclusion are reported in Sections 5 and 6 respectively.

2. Time stamp detection

Background of a video changes frequently while the time stamp is always at the same position; therefore, the temporal frames are very helpful for detecting time stamp. In addition, we choose edge detection method to find time stamp positions because the time stamp region is rich of edges.

The time stamp detection is composed of 2 main steps which are coarse detection and fine detection. The coarse detection is to roughly detect the time stamp while the fine detection is to locate each digit of the time stamp. The detail procedure for each step is described as follows.

2.1 Coarse detection

The time stamp coarse detection procedure is shown in figure 1. First, we capture initial frame from video and find edges. The edge map is dilated by a 3×3 structure element since time stamp locations might shift around by one or two pixels in the frame [1]. The result from this step will be called "the initial image". Then, the next frame is captured and processed in the same way as the initial one. These two images are operated by logical AND to eliminate non-time stamp edges. We repeat this step until the ANDed result is good enough to coarsely locate the time stamp.

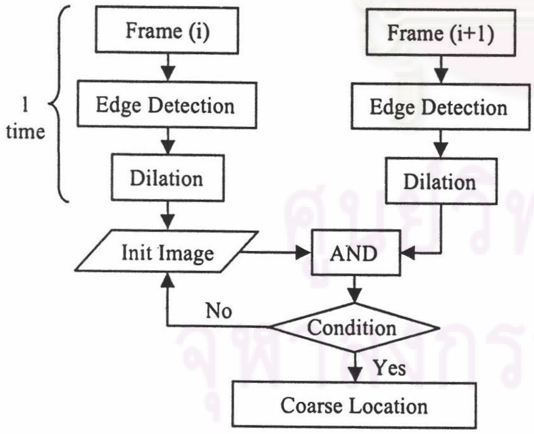
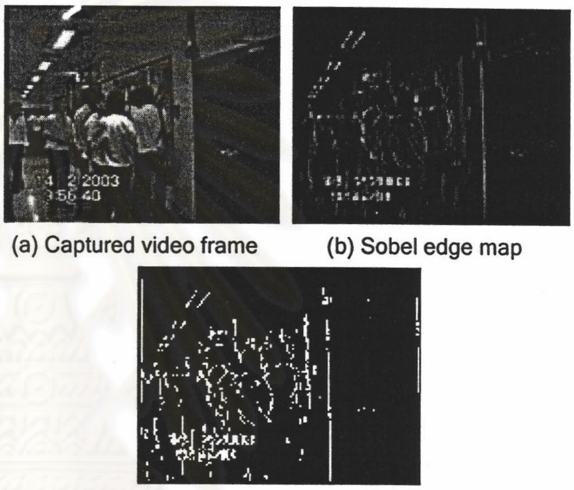


Figure 1. Time stamp coarse detection procedure

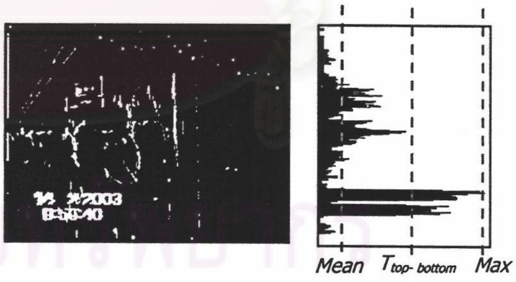
2.1.1 Edge detection. The captured video frame is converted from color to gray-scale image before detecting edges. In this paper, we want to find only vertical edges because digits are mostly composed of vertical strokes that are sufficient to detect the time stamp position. Besides, finding horizontal edges does not only increase processing time, but it also adds edges in the background that is hard to remove. Vertical mask of Sobel operator is used in this work and the edge map is thinned by non-maxima suppression method. Pixels that have value less than its neighbor pixels in gradient direction are removed. Example results from edge detection method are shown in figure 2.



(a) Captured video frame (b) Sobel edge map

(c) Non-maxima suppression of (b)

Figure 2. Edge detection



(a) ANDed image (b) Horizontal projection profile of (a)

Figure 3. Non-time stamp edge elimination

2.1.2 Non-time stamp edge elimination. To remove the non-time stamp edges from the edge map, temporal frames are utilized. Because the background of the video always changes while the time stamp is at the same position, so we eliminate the non-time stamp edges by operating a logical AND between the initial

edge map and the next frame's edge map. However, both edge maps have to be dilated before ANDing them because time stamp positions might shift around by one or two pixels, as mentioned earlier. We thus apply a 3×3 square structure element to dilate the edge maps. The result of the non-time stamp edge elimination is shown in figure 3(a). When non-time stamp edges are removed in each turn of the frame capturing, we examine whether the ANDed image is good enough for time stamp location. It is good enough when there are outstanding peaks in a horizontal projection profile of the ANDed image, as shown in figure 3(b), where the maximum value is higher than the mean value k_1 times. Then we can stop the AND operation.

2.1.3 Coarse location. We create bounding boxes surrounding the whole time stamp scan-line by scan-line because we cannot directly locate each time stamp digit due to the complex background. The top and bottom boundaries of the bounding box that encloses the time stamp are calculated first by analyzing the horizontal projection profile of the ANDed result; while the left and right boundaries are initially set to the left and right sides of the whole image respectively. An adaptive threshold, $T_{top-bottom}$ as shown in figure 3(b), is computed from Equation (1) as proposed in [4]. This threshold determines which lines of profile are time stamp positions.

$$T_{top-bottom} = k_2 \times Max_{profile} \qquad (1)$$

where k_2 is a proper constant derived from experiments and $Max_{profile}$ is the maximum value of the profile.

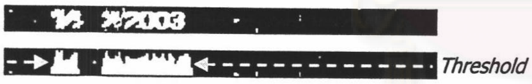


Figure 4. Time stamp bounding box and vertical projection profile

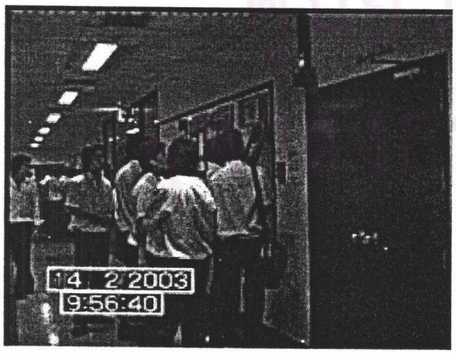


Figure 5. Coarse location result

We expand each bounding box at top and bottom by 10% of the located box's height to ensure that it covers the whole time stamp region. Next, we find the vertical projection profile of the bounding box to locate the left and the right boundaries of the box. Similar to finding the top and the bottom boundaries, a threshold is defined. However, unlike the previous process, the proper constant in Equation (1) is set to k_3 in place of k_2. The profile is scanned from left to right. If we find a position of the profile that has a value more than the threshold, that position will be set to the left boundary. The right boundary can be calculated in the same manner; however, the profile is scanned from right to left as shown in figure 4. Afterwards, we enlarge the left and the right boundaries by 50% of the box's height to ensure that the whole digit is inside the boundary. The coarse location result is illustrated in figure 5.

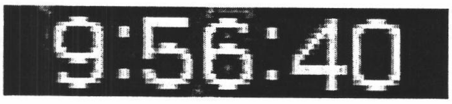
2.2 Fine detection

With our proposed method, digit color is required in order to finely locate time stamp. In general, time stamps have only two colors, white or black, but it is still difficult to identify the color of time stamp automatically. Although Lienhart and Wernicke [4] proposed a text-and-background color estimation, we have found that it can work well only in the text that is superimposed on different background colors. For the time stamp overlaid on the same background color, this method tends to fail. Therefore, in this paper, we require users to specify the digit color. For simplicity, we now assume in this paper that the time stamp digits are white.

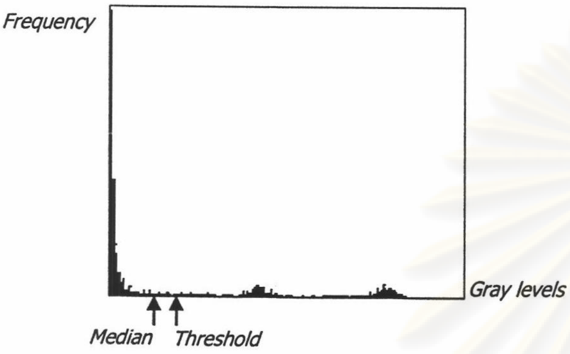
For this step, we utilize the temporal frames again. In addition, all processes are done within the time stamp location boundary derived from the previous step. Desiring white digits on the dark background, we apply a time-based minimum pixel value search method [9] that can reduce the affect of background complexity. Since we capture the video frame by frame for processing, stopping criterion has to be defined by examining a histogram of an output from the time-based minimum pixel value search method, see figure 6. Most bounding box regions consist of background that is preferred being dark; so if a median of histogram is less than a fixed threshold, the time-based minimum pixel value search will be ended.

The result from the time-based minimum pixel value search method is still a gray scale image, thus we binarize it by Otsu's method [11]. Then the actual height of time stamp digit is calculated by examining a horizontal projection profile of the image. Calculating

the top and bottom boundaries of the bounding box in this step must be very accurate because most background portions are eliminated.



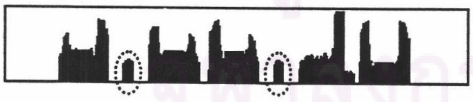
(a) time-base minimum pixel value search result



(b) Histogram of (a) shows median and threshold value

Figure 6. Stopping criterion of the time-base minimum pixel value search

Since the second and sub-second digits of the time-based minimum pixel value search output image might disappear or become imperfect, we locate them by utilizing the colon signs. Furthermore, with the assistance of colon signs, we can distinguish between date and time. Colon signs appear only in the time region while the date region has none. From the vertical projection profile, the colon signs are most likely at the positions where the histograms have low values and narrow bars; see figure 7(a). Since a characteristic of a colon sign is that they must have two small connected components that are equally distant from a center of the box, we also examine the connection component in the boundaries as shown in figure 7(b).



(a) Vertical projection profile

(b) Candidate colon signs

Figure 7. Colon signs detection

When colon signs are located, the hour, minute, second, and sub-second positions will be at the front or the back of the colon signs. Another region detected to be time stamp will be assigned to the date region. Each digit position can now be easily extracted by analyzing the vertical projection profile. A width of each digit is normalized based on the maximum width of all digits. Therefore time stamp template can be generated as shown in figure 8.

Figure 8. Fine detection result

3. Time stamp segmentation

The time stamp template from previous step will be used for the next other frames. Before proceeding to the recognition step, each digit has to be segmented from the background. In this paper, we propose an algorithm that uses connected component labeling and run-length method. Procedures are described in details as follows.

3.1 Image scaling and binarization

Due to low resolution of typical videos, we have to increase the resolution of the digit image by scaling. Bicubic interpolation method is used to enlarge the image. Width of a digit is normally less than or equal to its height, so we scale image to one fixed height, 60 pixels. Then scaled image is binarized by Otsu's method. There will be connected components of both digit and background, thus, in the next step, we have to label the connected components and determine which component is the digit.

3.2 Digit segmentation

First, we assume that the height of a digit component is about the height of the digit bounding box. However, when the background color is similar to the digit color, these criteria cannot identify the digit component. Figure 9(a) is an example for this case.

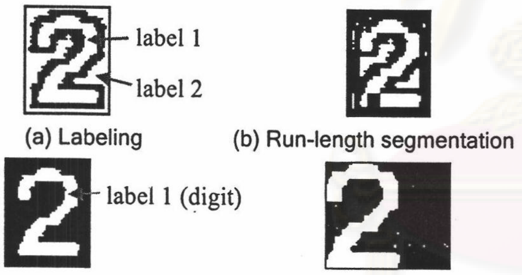
Therefore, we present a method to find the digit components using the following rules:

a) If there is only one connected component in the image, that component must be a digit.

b) If there is more than one connected component, run-length method is operated in both horizontal and vertical direction. The run-length lines that are longer than a stroke width threshold, T_{stroke}*, are eliminated, as well as the lines that have starting or ending points at a border of the bounding box. The result of this step is shown in figure 9(b). The label that contains most of the remainders is set to the digit label. Figure 9(c) shows the segmentation result.*

3.3 Normalization

We recognize digit by matching method; therefore, normalization is required to make images comparable. The digit component is initially cropped from the background. Next, it is resized to a fixed height by nearest neighbor interpolation so that the image is still a binary image. A black canvas that has the same height as a digit and the width equals 1.2 times of the height is created. The cropped digit is then placed on a leftmost side of the defined canvas as shown in figure 9(d).



(a) Labeling (b) Run-length segmentation

(c) Segmentation result (d) Normalization result

Figure 9. Digit segmentation

4. Time stamp recognition

The segmented digit is recognized by our simple recognition module using decision tree and conventional matching techniques because there are only ten classes of digit. First, we calculate number and positions of holes in a digit image. Decision tree is generated by dividing the tree into five branches; zero-hole, 1-hole at the top, 1-hole at the middle, 1-hole at the bottom, and 2-hole branches. Zero-hole branch, for example, consists of digits 1, 2, 3, 5 and 7, see figure 10. However, digit 4 might have either no hole or one hole depending on a font style as shown in figure 11. Using decision tree can significantly reduce matching error because sample space for matching is decreased. For matching technique, we choose a similarity index [10] to measure the similarity between two images.

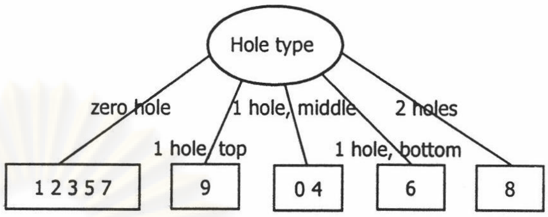


Figure 10. Decision tree of digits

(a) Digit 4 (no hole) (b) Digit 4 (1 hole)

Figure 11. Digit 4 in different font styles

4.1 Hole detection

Since each digit is composed of only one connected component, the number of holes in the digit can be found easily and rapidly. First, a contour of the digit component is detected. Then connected-component labeling is performed to the contour of digit image. The number of holes of the digit is the number of label minus 1. For example, see figure12, the contour of digit 9 consists of two connected components, so there is one hole in digit 9. In addition, the component that is the hole of the digit will not be the outermost component. Thus we can find the position of the hole in case of 1-hole digit whether it is at the top, middle, or bottom of the digit. The hole type of digit 9 shown in figure 12, for example, is 1-hole at the top.

(a) Digit 9 (b) Contour of digit 9

Figure 12. Hole detection

4.2 Similarity index

The similarity index uses size and location to measure the similarity between 2 images. The similarity index between two binary images, A_1 and A_2, is defined as a ratio of twice the

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

common area to the sum of individual areas as shown in Equation (2):

$$S = \frac{2n\{A_1 \cap A_2\}}{n\{A_1\} + n\{A_2\}} \quad (2)$$

where $n\{A\}$ is the number of digit pixels. The value of S is from 0 to 1. The more two images are similar; the closer S value is to 1.

5. Experimental results

Videos that are used in evaluating the performance of our approach must have time stamp superimposed on their frames. Tested videos are 288×352 resolution (PAL). We tested our approach with videos from three sources, two from two different camcorders and the other from a laboratory that uses a video timer with sub-second precision as the time stamp.

We separate the evaluation of time stamp detection from recognition. We tested our detection approach with 90 shots of the 30-minute videos. The average frames for detecting the time stamp is 31, and 77 shots out of 90 are correctly detected (85.6%). The coarse detection step works very well with a few frames, less than 12 frames in average cases, but it is found time-consuming because of the time-based minimum pixel value search procedure in the cases of the bright continuous frames.

The recognition method is evaluated with the digits cropped from videos, 90 cropped digits per class (0-9). Out of the total 900 tested digits, 895 are recognized correctly, and hence, the digit recognition rate is 99.4%.

6. Conclusions

In this paper, we present a time stamp detection and recognition approach. Detection is based on vertical edges of the time stamp and utilizes temporal frames. We propose a method for segmenting a digit from background by applying run-length method. The segmented digit is recognized by using decision tree and similarity index. Experimental results show that our method's accuracies are 85.6% in detection and 99.4% in recognition. However, if an actual automatic system is desired, the digit color has to be detected.

7. Future work

For our future work, we plan to apply some knowledge about the time stamp for post-processing in the recognition module. We also aim to apply our proposed method to some applications such as speed detection, shot boundary detection, etc.

8. Acknowledgement

The authors wish to express their grateful thanks to the Department of Computer Engineering - Industry Linkage Research Project Year 2004, Chulalongkorn University, for the funding support.

References

[1] P. Yin, X. S. Hua, and H. J. Zhang, "Automatic Time Stamp Extraction System for Home Videos," *IEEE International Symposium on Circuits and Systems (ISCAS 2002)*, 2002, pp. 73-76.

[2] A. K. Jain, and B. Yu, "Automatic Text Location in Images and Video Frames," *Proceedings of International Conference on Pattern Recognition*, 1998, vol. 2, pp. 1497-1499.

[3] D. Chen, J. M. Odobez, and H. Bourlard, "Text Detection and Recognition in Images and Video Frames," *Pattern Recognition*, 2003, vol. 37, no. 3, pp. 595-608.

[4] R. Lienhart, and A. Wernicke, "Localizing and Segmenting Text in Images and Videos," *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, vol. 12, no. 4, pp. 256 -268.

[5] X. Tang, X. Gao, J. Liu, and H. Zhang, "A Spatial-Temporal Approach for Video Caption Detection and Recognition," *IEEE Transactions on Neural Networks*, 2002, vol. 13, no. 4, pp. 961-971.

[6] M. Cai, J. Song, and M. R. Lyu, "A New Approach for Video Text Detection," *Proceedings of International Conference on Image Processing*, 2002, vol.1, pp. 117-120.

[7] J. Xi, X.S. Hua, X. R. Chen, L. Wenyin, and H. J. Zhang, "A Video Text Detection and Recognition System," *IEEE International Conference on Multimedia and Expo (ICME 2001)*, 2001, pp. 1080-1083.

[8] X. R. Chen, H. J. Zhang, "Text Area Detection form Video Frames," *IEEE Pacific-Rim Conference on Multimedia (PCM2001)*, 2001, pp. 222-228.

[9] T. Sato, T. Kanade, E. Hughes, M. Smith, and S. Satoh, "Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions," *ACM Multimedia Systems Special Issue on Video Libraries*, 1999, vol. 7, pp. 385-395.

[10] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric Analysis of White Matter Lessions in MR images: Method and Validation," *IEEE Transactions on Medical Imaging*, 1994, vol. 13, pp. 716-724.

[11] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems*, 1979, vol. 9, pp. 62-66.

ประวัติผู้เขียนวิทยานิพนธ์

นายเจษฎา แสงพานิชย์ เกิดวันที่ 10 กรกฎาคม พ.ศ. 2522 ที่จังหวัดอุตรดิตถ์ สำเร็จการศึกษาระดับปริญญาตรีวิศวกรรมศาสตรบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2543 หลังจากนั้นได้เข้ามาศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2544

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย