เทคนิคการวิเคราะห์วีดีทัศน์เพื่อตรวจจับการเกิดอุบัติเหตุบนถนน
A Novel Technique for Accidental Detection in Motion Picture

ไกรศักดิ์ เกษร (Kraisak Kesorn)*

บทคัดย่อ

ระบบตรวจจับการเกิดอุบัติเหตุในวีดีทัศน์ช่วยให้การทำงานของเจ้าหน้าที่ตำรวจตรวจสอบเหตุการณ์บนท้องถนนได้ง่ายขึ้น โดยไม่ต้องเฝ้ามองวีดีทัศน์จากกล้องวงจรปิดตลอดเวลา บทความนี้นำเสนอเทคนิคในการตรวจจับการเกิดอุบัติเหตุในวีดีทัศน์ โดยจุดเด่นของงานวิจัยนี้คือการใช้คุณลักษณะของภาพแบบโลคอล (Local features) ร่วมกับคุณลักษณะอื่นๆ ได้แก่ ขนาดของยานพาหนะ ทิศทางการเคลื่อนที่ และความเร็วในการเคลื่อนที่ของยานพาหนะ เพื่อช่วยในการวิเคราะห์ภาพอุบัติเหตุให้มีประสิทธิภาพมากยิ่งขึ้น ซึ่งการทดลองพบว่าวิธีการที่นำเสนอมีประสิทธิภาพในการตรวจจับการเกิดอุบัติเหตุถึง 72.41% และพบว่าการใช้คุณลักษณะแบบโลคอลช่วยให้การรู้จำยานพาหนะได้ดียิ่งขึ้น ซึ่งส่งผลให้ประสิทธาภาพการตรวจจับอุบัติเหตุเพิ่มขึ้นด้วย

คำสำคัญ: การประมวลผลภาพวีดีทัศน์ การตรวจจับการเกิดอุบัติเหตุ วิชวลเวิร์ด ถุงของคำ

Abstract

Accidental detection system facilitates police officers to easier monitoring road eventswithoutcontinuouslykeep looking at CCTV's screen. This paper presents a novel technique for accidental detecting approach in videos. The main novelty of this work is the use of invariantlocal features (SIFT) to represent vehicle object incorporating with other significant factors for accidentalevent determinationsuch as vehicle size, moving direction, and speed to enhance accidental video analysis. The experimental results demonstrated that the proposed technique can increase the power of the detection system up to 72.41%. In addition, the use of local feature can improve vehicle recognition performance significantly.As a result, accidental detecting accuracy is improved.

Keyword: Image and Video Processing, Accidental Detection, Visual Word, Bag of Visual Word.

1. บทนำ

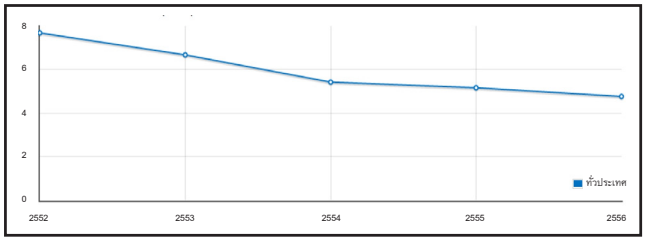
ด้วยความก้าวหน้าทางเทคโนโลยีทางด้านการประมวลผลภาพและวีดีทัศน์ในปัจจุบันทำให้นักวิจัยทางด้านวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศให้ความสนใจการทำวิจัยในการตรวจสอบการเกิดอุบัติเหตุบนท้องถนนที่ถูกบันทึกในรูปแบบวีดีทัศน์จากกล้องวงจรปิดเพื่อลดความสูญเสียชีวิต ทรัพย์สิน และเป็นประโยชน์ในการดำเนินการด้านกฎหมายของตำรวจได้อย่างมีประสิทธิภาพ การเกิดอุบัติเหตุบนท้องถนนเป็นความเสียหายต่อชีวิตและทรัพย์สินที่ใหญ่หลวงต่อคนในครอบครัวและชุมชนจากสถิติการเกิดอุบัติเหตุบันทึกโดยกรมการขนส่งทางบก พบว่าอุบัติเหตุจากการจราจรนั้นทวีความรุนแรงต่อเนื่องตลอดทุกปี จากรายงานพบว่าการเกิดอุบัติเหตุมาจากสาเหตุต่างๆ กันดังนี้ อันดับที่ 1 ขับรถเร็วเกินอัตราที่กฎหมายกำหนดร้อยละ 23.41 อันดับที่ 2 ขับรถตัดหน้ารถคันอื่นในระยะกระชั้นชิดร้อยละ 21.07 อันดับที่ 3 ขับรถตามหลังรถคันอื่นในระยะกระชั้นชิดร้อยละ 11.69 จากตัวเลขดังกล่าวหน่วยงานต่างๆ ทั้งภาครัฐ ภาคเอกชนและองค์กรอิสระจึงพยายามช่วยกันลดอัตราการเกิดอุบัติเหตุให้น้อยลง

ด้วยความก้าวหน้าทางเทคโนโลยีทางด้านการประมวลผลภาพและวีดีทัศน์ (image and video processing) ในปัจจุบันทำให้นักวิจัยทางด้านวิทยาการคอมพิวเตอร์ให้ความสนใจในการทำวิจัยด้านการตรวจเช็คการเกิดอุบัติเหตุบนท้องถนน

*ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

เพื่อลดความสูญเสียชีวิต ทรัพย์สิน และเป็นประโยชน์ในการดำเนินการด้านกฎหมายได้อย่างมีประสิทธิภาพเทคนิคทางด้านการประมวลผลภาพและวีดีทัศน์ได้ถูกนำมาช่วยในงานเฝ้าระวังอุบัติเหตุ (surveillance) แบบเดิมที่ใช้เจ้าหน้าที่เฝ้าดูที่หน้าจอคอมพิวเตอร์เพื่อคอยจับตาดูว่าถนนหรือสถานที่ใดเกิดอุบัติเหตุขึ้นบ้าง เรียกว่า การเฝ้าระวังแบบ Manual ซึ่งเป็นการยากที่จะให้พนักงานเพียง 1-2 คน เฝ้าจับตามองภาพวีดีทัศน์ที่มีเป็นร้อยหรือพันตัวในพื้นที่หนึ่งๆได้อย่างครอบคลุมและทั่วถึง หรือคงไม่คุ้มค่าที่จะจ้างพนักงานนับร้อยคนเท่ากับจำนวนกล้องวีดีทัศน์ที่ติดตั้งไว้ตามจุดสำคัญต่างๆ เพื่อคอยเฝ้าอุบัติเหตุที่จะเกิดขึ้นเวลาใดเวลาหนึ่งซึ่งไม่อาจทราบได้ จากข้อจำกัดดังกล่าวจึงทำให้ผู้วิจัยคิดค้นพัฒนาเทคนิคการตรวจจับการเกิดอุบัติเหตุบนท้องถนนขึ้นมา (monitoring and detecting) เพื่อช่วยการทำงานของตำรวจหรือเจ้าหน้าที่ที่เกี่ยวข้องให้มีประสิทธิภาพมากยิ่งขึ้น

จากข้อมูลของกรมทางหลวง กระทรวงคมนาคม พบว่าอัตราการเกิดอุบัติเหตุบนท้องถนนของประเทศไทย ระหว่างปี 2552-2556 มีอัตราการลดลงเล็กน้อย (แสดงดังภาพที่ 1) เมื่อเทียบกับจำนวนรถที่มากขึ้น สาเหตุประการหนึ่งคือมีการนำกล้องวงจรปิดมาตั้งไว้ตามจุดสำคัญต่างๆ ทำให้ประชาชนในพื้นที่มีความระมัดระวังในการขับรถมากยิ่งขึ้น



ภาพที่ 1 *สถิติการเกิดอุบัติเหตุบนทางหลวงแผ่นดินระหว่างปี 2552 ถึง 2556 จากกระทรวงคมนาคม*

ในหลายๆ ประเทศที่พัฒนาแล้ว เช่น อังกฤษ ญี่ปุ่น เกาหลี หรือสหรัฐอเมริกา ได้นำกล้องวงจรปิดหรือที่เรียกกันว่า CCTV มาติดตั้งเพื่อบันทึกวีดีทัศน์ และใช้เป็นหลักฐานสำคัญในการตามหาผู้กระทำผิดต่างๆ รวมถึงอุบัติเหตุบนท้องถนนด้วยเช่นกัน และในประเทศไทยได้มีการติดตั้งกล้องวงจรปิดตามเมืองสำคัญๆ เช่น กรุงเทพมหานคร เชียงใหม่ ภูเก็ตและพัทยา เป็นต้น แต่ปัญหาสำคัญคือวีดีทัศน์ที่บันทึกไว้นี้จะไม่ได้ถูกตรวจสอบโดยเจ้าหน้าที่ทั้งหมด แต่จะถูกเรียกขึ้นมาดูก็ต่อเมื่อมีผู้เสียหายมาขอดูเท่านั้น วีดีทัศน์ใดที่ไม่ได้มีการขอดูจะถูกเก็บไว้ในฮาร์ดดิสก์ (hard disk) เป็นเวลาประมาณ 45 วัน หลังจากนั้นจะถูกลบทิ้ง ทั้งนี้เนื่องจากปัญหาต่างๆ คือ 1) เจ้าหน้าที่ไม่สามารถที่จะดูคลิปวีดีทัศน์ทั้งหมดได้ว่ามีคลิปไหนที่มีเหตุการณ์อุบัติเหตุเกิดขึ้นบ้าง และ 2) เจ้าหน้าที่ไม่สามารถเก็บคลิปวีดีทัศน์ทั้งหมดไว้ได้เนื่องจากเสียพื้นที่ในการจัดเก็บข้อมูลจำนวนมากเนื่องจากมีจำนวนหลายร้อยคลิปวีดีทัศน์ดังนั้นจึงมีนักวิจัยที่จะพยายามทำการศึกษาค้นคว้าเทคนิคในการตรวจสอบภาพเหตุการณ์ในคลิปวีดีทัศน์ว่ามีอุบัติเหตุเกิดขึ้นหรือไม่ เพื่อลดข้อจำกัดของปัญหาทั้ง 2 ข้อที่ได้กล่าวมาแล้วข้างต้น

2. ทบทวนวรรณกรรม

2.1 เทคนิคในการวิเคราะห์ภาพเหตุการณ์อุบัติเหตุ

ในการวิเคราะห์ภาพเหตุการณ์อุบัติเหตุบนท้องถนนนั้นมีนักวิจัยหลายท่านทำการวิจัยในเรื่องนี้ ตัวอย่างเช่น Zou และคณะ [1] ทำการวิเคราะห์ภาพเหตุการณ์อุบัติเหตุโดยใช้ Hidden Markov Model (HMM) ในการจำแนกเหตุการณ์อุบัติเหตุต่างๆ Ikeda และคณะ [2] นำเสนอเทคนิคในการตรวจสอบเหตุการณ์ต่างๆ บนท้องถนน 4 ประเภท ได้แก่ รถจอด รถวิ่งช้า การหลุดของชิ้นส่วนรถยนต์และการเปลี่ยนช่องทางเดินรถ (lane) แต่อย่างไรก็ตามงานวิจัยของ Ikeda ยังไม่สามารถตรวจสอบหาเหตุการณ์อุบัติเหตุของรถยนต์บนท้องถนนได้ Lin et al. [3]เสนอแนวคิดใหม่ในการตรวจสอบภาพอุบัติเหตุบนท้องถนนโดยใช้เทคนิคที่ชื่อว่า “Image Tracking”Kamijoet al. [4] นำ HMM มาใช้ในการตรวจสอบการชนกันของรถยนต์ เช่นเดียวกับ Akozet al. [5] ซึ่งนำ HMM มาใช้ในการตรวจหาเหตุการณ์ไม่ปกติบนท้องถนน นอกจากนี้ยังมีเทคนิคต่างๆ หลายเทคนิคที่นำมาใช้ในการตรวจจับอุบัติเหตุบนถนนนอกจากนี้ยังมีการนำเทคโนโลยีอื่นๆ มาใช้ในการตรวจสอบการเกิดอุบัติเหตุได้แก่ GPS [6] เป็นต้น แต่ในงานวิจัยนี้จะนำเสนอแนวคิดใหม่ในการตรวจสอบการเกิดอุบัติเหตุบนท้องถนนโดยการใช้เทคนิคของการตรวจสอบวัตถุในคลิปวีดีทัศน์และการวิเคราะห์ทิศทางการเคลื่อนที่ของวัตถุดังกล่าวเพื่อหาข้อสรุปว่าเป็นฉาก (scene) ที่เกิดอุบัติเหตุหรือไม่

2.2 การตรวจสอบวัตถุในคลิปวีดีทัศน์ (Object detection)

ขั้นตอนแรกของการตรวจสอบการเกิดอุบัติเหตุใน

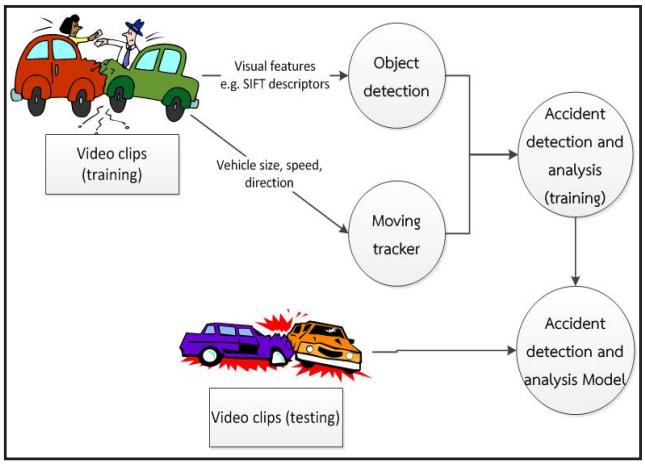
วีดีทัศน์นั้น ระบบจะต้องสามารถค้นหายานพาหนะหรือรถยนต์ในคลิปวีดีทัศน์ให้ได้เสียก่อน ทฤษฎีที่จะนำมาใช้ในงานวิจัยนี้คือทฤษฎีของ Tsaiet และคณะ [7] Bensrhairet และคณะ [8] และ Huh และคณะ [9] ซึ่งใช้คุณลักษณะของสี (color) ในการตรวจสอบวัตถุว่าเป็นรถยนต์หรือไม่ และนักวิจัยบางท่านใช้เส้นขอบ (edge) มาตรวจสอบและวิเคราะห์ภาพของวีดีทัศน์ แต่การใช้คุณลักษณะของสียังมีข้อจำกัดอยู่มาก เนื่องจากคุณลักษณะระดับต่ำนี้มีความน่าเชื่อถือน้อยนั่นเองในงานวิจัยนี้จึงเสนอการใช้คุณลักษณะแบบใหม่เพื่อใช้ในการตรวจสอบวัตถุในภาพวีดีทัศน์เรียกว่า “คุณลักษณะแบบโลคอล” ซึ่งเป็นคุณลักษณะระดับต่ำที่มีความน่าเชื่อถือมากกว่าคุณลักษณะพวก สีหรือเส้นขอบนั่นเอง ซึ่งจะขออธิบายเพิ่มเติมในหัวข้อที่ 3

2.3 การตรวจสอบทิศทางการเคลื่อนที่ของยานพาหนะในวีดีทัศน์ (Moving tracker)

จากข้อจำกัดของการใช้คุณลักษณะระดับต่ำของภาพมาตรวจสอบหาวัตถุในวีดีทัศน์ นักวิจัยยังนำเสนอการตรวจสอบทิศทางการเคลื่อนที่ของยานพาหนะในวีดีทัศน์ยังถูกนำมาใช้เพื่อลดข้อจำกัดดังกล่าวด้วยเช่นกัน โดยจะใช้ทฤษฎีที่สำคัญของ Tomasiet และคณะ [10] ซึ่งใช้เว็กเตอร์คุณลักษณะ (feature vector) ร่วมกับพิกัดของยานพาหนะ (coordinates) มาตรวจสอบ และคาดการณ์ทิศทางของยานพาหนะที่ควรจะเป็นในคลิปวีดีทัศน์ Kruger และคณะ [11] ใช้เทคนิคการเคลื่อนไหวของรถยนต์ในวีดีทัศน์ร่วมกับเวลาเพื่อเพิ่มประสิทธิภาพของการตรวจจับอุบัติเหตุได้ดียิ่งขึ้น

จากการทบทวนวรรณกรรมผู้วิจัยพบว่างานวิจัยส่วนใหญ่มุ่งเน้นในการวิเคราะห์วีดีทัศน์ให้ถูกต้อง โดยใช้คุณลักษณะระดับต่ำของวีดีทัศน์ และการใช้ทิศทางการเคลื่อนที่ของวัตถุมาประกอบเพื่อเพิ่มประสิทธิภาพการตรวจจับอุบัติเหตุให้ถูกต้องมากยิ่งขึ้น แต่อย่างไรก็ตามงานวิจัยที่มีอยู่ยังมีข้อจำกัดบางประการ เช่น การใช้ สี รูปทรง หรือเส้นขอบยังมีความน่าเชื่อถือน้อย เนื่องจากค่าเหล่านี้จะเปลี่ยนแปลงไปตามคุณลักษณะของวีดีทัศน์ เช่น ขนาดภาพ มุมกล้อง การหมุนของภาพ เป็นต้น

งานวิจัยนี้จึงพยายามจะลดข้อจำกัดดังกล่าวให้น้อยลงโดยเสนอเทคนิคใหม่ซึ่งอธิบายในหัวข้อถัดไป



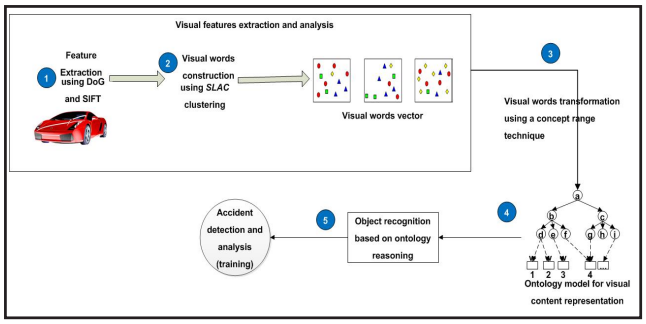
ภาพที่ 2 กรอบแนวคิดของระบบวิเคราะห์วีดีทัศน์เพื่อตรวจจับการเกิดอุบัติเหตุบนถนน

3. กรอบแนวคิด

ในบทความนี้ผู้วิจัยจะนำเสนอเทคนิคการตรวจสอบอุบัติเหตุบนท้องถนนโดยโครงสร้างการทำงานของระบบที่จะสร้างขึ้นประกอบด้วย 2 ส่วนหลักแสดงภาพที่ 2

1) การตรวจสอบวัตถุหรือยานพาหนะในคลิปวีดีทัศน์ (object or vehicle detection) คือส่วนที่ระบบทำการรู้จำเพื่อตรวจสอบวัตถุที่เป็นรถยนต์ในวีดีทัศน์ โดยดูจากคุณลักษณะต่างๆ ของภาพ กระบวนการตรวจสอบวัตถุในคลิปวีดีทัศน์แสดงดังภาพที่ 3 โดยงานวิจัยนี้จะไม่ใช้คุณลักษณะระดับต่ำทั่วไป (low level features) เช่น สี ขอบ รูปทรง ซึ่งเรียกว่าคุณลักษณะแบบโกลบอล (global) แต่จะใช้คุณลักษณะระดับต่ำที่เรียกว่า SIFT (Scale Invariant Feature Transform-SIFT) [12] ระบบจะถูกทำการสอน (train) โดยใช้คลิปวีดีทัศน์อุบัติเหตุประมาณ 50 คลิป เพื่อให้ระบบสามารถเรียนรู้ลักษณะต่างๆ ของรถยนต์ให้มากที่สุดและทำการสร้างตัวแบบ (model) ที่เป็นมาตรฐานขึ้นเพื่อนำไปใช้ภายหลัง โดยในระบบต้นแบบนี้จะตรวจสอบเฉพาะรถยนต์ตั้งแต่ 4 ล้อขึ้นไป ซึ่งระบบยังไม่สามารถตรวจสอบกับรถมอเตอร์ไซต์และรถจักรยานได้ในขณะนี้

2) การตรวจสอบการเกิดอุบัติเหตุจากการเคลื่อนไหวของวัตถุ (Motion-based accident detection) ในส่วนนี้ระบบจะทำการตรวจสอบว่ามีรถคันใดไม่วิ่งไปตามทิศทางที่ตัวแบบได้คำนวณไว้ ซึ่งระบบจะคาดเดาว่ารถคันนั้นอาจจะเกิดอุบัติเหตุขึ้นได้ เช่น การเปลี่ยนทิศทางแบบกระทันหันหรือมีความเร็วลดลงอย่างเฉียบพลัน อาจจะเป็นสัญญาณว่าเกิดอุบัติเหตุขึ้นในวีดีทัศน์ดังกล่าว



ภาพที่ 3 ขั้นตอนการตรวจสอบวัตถุในคลิปวีดีทัศน์

3.1 การตรวจสอบยานพาหนะโดยใช้ซิฟท์

ในปี 2004 นักวิจัยท่านหนึ่งชื่อ Lowe ได้เสนอคุณลักษณะของภาพประเภทหนึ่งซึ่งไม่เปลี่ยนแปลงไปตามขนาดของภาพ มุมกล้องความสว่างหรือการหมุนของภาพคุณลักษณะนี้เรียกว่าคุณลักษณซิฟท์ (SIFT) ซึ่งเป็นคุณลักษณะแบบโลคอล โดยจะหาจุดสำคัญต่างๆ ในภาพ (keypoints) ซึ่งจะแสดงในรูปแบบของเว็กเตอร์ 128 ค่า (128 dimensional) และถูกใช้เป็นประโยชน์ในการแยกประเภทของภาพ (Image classification) จากผลการเปรียบเทียบของนักวิจัยชื่อ Mikolajczyk [13] ทำการทดสอบการแยกประเภทของภาพโดยใช้คุณลักษณะของภาพหลายๆ คุณลักษณะและผลการทดสอบสรุปว่าซิฟท์ให้ประสิทธิภาพสูงสุดในการแยกประเภทของภาพ ดังนั้นผู้วิจัยจึงแนวคิดที่จะนำคุณลักษณะซิฟท์มาใช้ในการตรวจสอบหาวัตถุที่เป็นรถยนต์หรือยานพาหนะในวีดีทัศน์ในการวิจัยนี้เพื่อเพิ่มประสิทธิภาพการตรวจสอบฉากการเกิดอุบัติเหตุและทำให้ระบบทนต่อความสว่าง ความละเอียด และมุมของวีดีทัศน์ที่เปลี่ยนไปได้

3.2 การวิเคราะห์และการดึงคุณลักษณะซิฟท์

ในขั้นตอนนี้ประกอบด้วย 2 กระบวนการย่อยเพื่อประมวลผลและวิเคราะห์คุณลักษณะแบบโลคอล (local feature) ของรูปรถยนต์ขั้นตอนแรกคือการดึงเอาคุณลักษณะของภาพออกมาและเก็บในรูปแบบที่จะนำไปใช้ต่อไปเพื่อแปลความหมายได้ง่ายขั้นตอนนี้จะเรียกว่า “การวิเคราะห์ภาพ (image processing and analysis)” ขั้นตอนที่สองคือแปลความหมายของคุณลักษณะเหล่านั้นให้เป็นคอนเซฟท์ระดับสูง (high level conceptualization) นั่นคือแปลข้อมูลรูปภาพที่ได้ว่าเป็นรถยนต์หรือไม่ที่ปรากฏในวีดีทัศน์

ในงานวิจัยนี้จะใช้เทคนิค bag-of-visual-word (BVW) สำหรับการรู้จำวัตถุต่างๆ เพื่อช่วยในการจำแนกภาพ ประโยชน์หลักของวิธีการนี้คือแก้ข้อจำกัดที่พบในคุณลักษณะแบบโกลบอลเช่นขนาดรูป ความสว่างในรูปการอิ่มตัวของสีหรือมุมกล้องที่ต่างกันทำให้การรู้จำมีประสิทธิภาพมากยิ่งขึ้นอย่างไรก็ตามวิธีการ BVW แบบเดิมมีข้อจำกัดอยู่หลายประการตัวอย่างเช่นไม่มีเก็บรักษาความหมายที่อยู่ในภาพเอาไว้ดังนั้นในงานวิจัยนี้จะนำเสนอวิธีการเพื่อลดข้อจำกัดดังกล่าวเพื่อเพิ่มประสิทธิภาพการค้นคืนภาพให้สูงขึ้นซึ่งขั้นตอนการสร้าง BVW นี้จะประกอบด้วย 3 ขั้นตอนหลักๆ ดังต่อไปนี้

1) การสกัดคุณลักษณะระดับต่ำของภาพ (Low-level feature extraction)

ขั้นตอนนี้เริ่มจากการใช้อัลกอริทึม Difference of Gaussian (DOG) [12] ประมวลผลภาพเพื่อทำการดึงเอาจุดสำคัญ (Keypoint) ของภาพออกมากและจุดสำคัญบนภาพเหล่านี้ข้อมูลเหล่านี้มีลักษณะเป็นเว็กเตอร์ของตัวเลขที่เป็นตัวอธิบายภาพและหลังจากนั้นข้อมูลเหล่านี้จะถูกแปลงเป็นข้อมูลที่เรียกว่า SIFT descriptor ดังนั้นผลลัพธ์สุดท้ายของขั้นตอนนี้คือภาพแต่ละภาพจะถูกแทนด้วยข้อมูลเว็กเตอร์ที่มีมิติ (dimension) ที่เท่ากันทุกภาพ

2) การสร้างวิชวลเวิร์ด (Visual word construction)

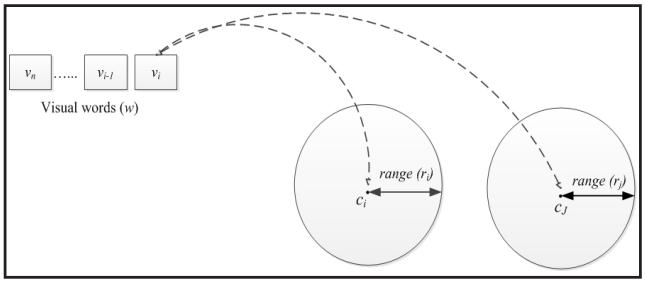
คุณลักษณะของซิฟท์ที่เหมือนกันจะถูกจัดกลุ่มไว้ในกลุ่มเดียวกันโดยใช้เทคนิคการจัดกลุ่ม (clustering algorithm) เช่น k-mean แต่ข้อเสียของ k-mean คือไม่คำนึงถึงตำแหน่งของจุดสำคัญต่างๆ บนภาพหรือที่เรียกว่า “Spatial location” ซึ่งทำให้สูญเสียการเชื่อมโยงกันระหว่างความหมายของภาพและคุณลักษณะของภาพดังกล่าวดังนั้นในงานวิจัยนี้จึงนำเสนอการใช้เทคนิคที่ชื่อว่า Semantic Local Adaptive Clustering (SLAC) [14] แทน k-mean โดยเทคนิคดังกล่าวจะเก็บคุณลักษณะที่สัมพันธ์กันมาอยู่ในกลุ่มเดียวกันแต่ละกลุ่มนี้จะเรียกว่าวิชวลเวิร์ด (visual word) และวิชวลเวิร์ดเหล่านี้จะถูกนำไปสร้าง BVW ที่มีโครงสร้างในรูปแบบเว็กเตอร์ต่อไป การนำคุณลักษณะของรูปที่สัมพันธ์กันมาไว้ในกลุ่มเดียวกันทำให้คุณภาพของวิชวลเวิร์ดที่ได้ดีขึ้นและใช้เป็นตัวแทนของภาพได้ดียิ่งขึ้นนั่นเอง

3) การแปลงโครงสร้างของวิชวลเวิร์ด (Visual word transformation)

นักวิจัยหลายท่านได้พยายามลดความกำกวมของวิชวลเวิร์ดโดยการแปลงโครงสร้างของวิชวลเวิร์ดจากโครงสร้างแบบเว็กเตอร์เป็นโครงสร้างแบบลำดับชั้น [15]–[17]

โดยอัลกอริทึมในการจัดกลุ่มข้อมูลตัวอย่างเช่น Hierarchical Spatial Markov model [17] Agglomerative clustering [18] และวิธีการ Hierarchical Latent Dirichlet Allocation algorithm หรือวิธีการ hLDA [19] อย่างไรก็ตาม วิธีการเหล่านี้ยังมีข้อจำกัดบางประการซึ่งมีผลต่อการแปลความหมายของรูปภาพซึ่งสรุปได้ดังต่อไปนี้ประการแรกโครงสร้างลำดับชั้นที่ได้จากอัลกอริทึมเหล่านี้ส่วนใหญ่มีโครงสร้างเป็นต้นไม้ไบนารี่ (binary tree) เท่านั้นแต่ในความเป็นจริงแล้วข้อมูลที่อยู่ในรูปภาพอาจจะมีความสัมพันธ์กันมากกว่าโครงสร้างแบบไบนารี่ก็ได้ซึ่งทำให้มีผลเสียต่อการแทนข้อมูลในภาพไม่ถูกต้องและประการที่สองไม่มีการสืบทอดคุณสมบัติระหว่างโหนดแม่และโหนดลูกในลักษณะ Multiple-inheritance คือการที่โหนดลูกสามารถสืบทอดคุณสมบัติมาจากโหนดแม่หลายๆ โหนดดังนั้นการใช้โครงสร้างแบบไบนารี่จึงไม่สามารถใช้แทนข้อมูลในวีดีทัศน์ได้อย่างมีประสิทธิภาพเท่าที่ควรในงานวิจัยนี้จึงนำเสนอแนวคิดใหม่ในการใช้แปลงโครงสร้างของวิชวลเวิร์ดซึ่งจะอธิบายได้ดังต่อไปนี้

ในขั้นแรกของการสอนระบบ (training) ภาพรถยนต์ที่สนใจในวีดีทัศน์จะถูกแยกออกจากพื้นหลังโดยใช้คนทำ (manual) ทั้งนี้เพื่อให้ได้ข้อมูลของวัตถุที่เราต้องการจริงๆ และกำจัดข้อมูลรบกวน (noise) ออกเพื่อจุดประสงค์ในการสอนระบบให้รู้จำข้อมูลภาพรถยนต์แบบต่างๆ ผลที่ได้ของการแยกวัตถุที่สนใจออกมานี้ทำให้เราได้ข้อมูลชุดหนึ่งของรถยนต์แบบต่างๆ และข้อมูลแต่ละชุดนั้นมีความเกี่ยวข้องกันเชิงความหมาย (semantically relevant) เนื่องจากเป็นข้อมูลภาพของวัตถุเดียวกันนั่นเองหลักจากนั้นคุณลักษณะของรถยนต์แต่ละประเภท (C) แต่ละชุดจะถูกนำไปสร้าง BVW (ϖ) ซึ่ง ($\varpi_i \in C_i$) เมื่อเราได้ BVW ของรถยนต์แต่ละประเภทมาแล้วจะนำมาคำนวณค่าที่เรียกว่า "concept range" (ภาพที่ 4) ซึ่งเป็นเทคนิคที่ใช้ในการนำข้อมูลเข้าออนโทโลยีหรือโครงสร้างลำดับชั้นสำหรับข้อมูลภาพนั่นเอง



ภาพที่ 4 *แสดงการวิธีการรู้จำวัตถุ (รถยนต์) ของระบบโดยใช้ Concept range*

กำหนดให้รัศมี (r_i) ของคอนเซฟท์ i หมายถึงระยะทางระหว่างจุดศูนย์กลางของคอนเซฟท์ (c) และจุดศูนย์กลาง (v) ของวิชวลเวิร์ด (ϖ) ที่อยู่ไกลที่สุดของวัตถุนั้นๆ ซึ่งสามารถเขียนเป็นสมการได้ดังสมการที่ 1

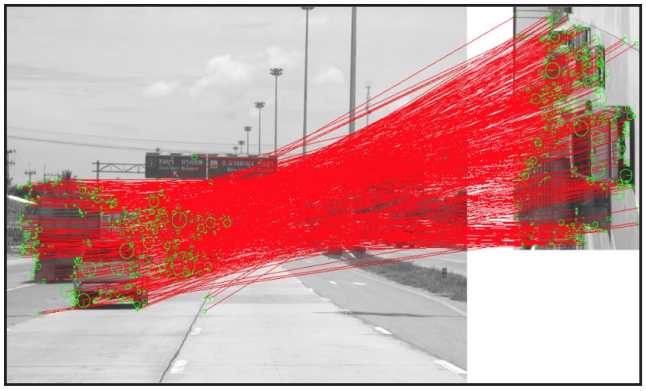
$$r_i = max|v\text{-}c|,\ v \in \varpi \quad (1)$$

ข้อดีของวิธีการนี้คือจะสามารถช่วยลดความกำกวมของวิชวลเวิร์ดได้และช่วยให้การจำแนกรูปภาพในวีดีทัศน์ให้มีประสิทธิภาพมากยิ่งขึ้นนอกจากนี้วิธีการนี้ยังแก้ปัญหาของต้นไม้ไบนารีได้ด้วยคือสามารถสืบทอดคุณลักษณะมาจากโหนดแม่ได้มากกว่าหนึ่งโหนด (multiple parents) เนื่องจากค่าของวิชวลเวิร์ดสามารถอยู่ในช่วงของรัศมีของคอนเซฟท์ได้มากกว่าหนึ่งคอนเซฟท์นั่นเองและในขณะเดียวกันวิชวลเวิร์ดบางอันก็จะถูกทิ้งไปหากค่าจุดศูนย์กลางของวิชวลเวิร์ดนั่นไม่อยู่ในรัศมีของคอนเซฟท์ใดๆ เลยซึ่งแนวคิดนี้สอดรับกับความจริงที่ว่าคำหนึ่งคำ (หรือวิชวลเวิร์ด) สามารถมีได้หลายความหมาย (polysemy) ด้วยเช่นกันตัวอย่างเช่นวิชวลเวิร์ดรถกระบะ อาจจะอยู่ในคอนเซฟท์ของรถเก๋งได้เช่นกันเนื่องจากมีลักษณะคล้ายๆ กัน เช่น ล้อรถ หรือไฟเลี้ยว เป็นต้นจากภาพที่ 4 จะเห็นว่า ผู้วิจัยทำการทดลองโดยตัดภาพเดียวกันให้มีขนาดต่างกันและทำการหมุนภาพ 90 องศาด้วยความสามารถของคุณลักษณะซิฟท์ถึงแม้ว่าภาพมีขนาดและการหมุนของภาพที่ต่างกัน (ภาพ 5-ก) แต่ระบบสามารถเชื่อมโยงจุดสำคัญ (keypoints) ของรูปภาพทั้ง 2 รูปภาพได้อย่างแม่นยำแสดงได้ดังภาพ 5-ข แนวคิดในการนำออนโทโลยีมาใช้ในการรู้จำภาพผู้เขียนได้นำเสนอไว้ในบทความ [20] ซึ่งผู้อ่านสามารถอ่านรายละเอียดได้จากบทความดังกล่าว

3.3 การรู้จำภาพรถยนต์

การรู้จำภาพที่เหมือนกันนั้นจะใช้หลักการของฮิสโตแกรมความถี่ของวิชวลเวิร์ดเป็นตัวแทนของวัตถุต่างๆ โดยถ้าความถี่ของวิชวลเวิร์ด $f(v_i)$ มีค่ามากกว่าค่าที่ได้จากการทดลอง (threshold=200) ซึ่งได้จากค่าเฉลี่ยที่ทดลองซ้ำๆ หลายๆ ครั้งระบบจะตีความว่ามีวัตถุดังกล่าวอยู่ในกลุ่มภาพนั้นๆ แต่เนื่องจากภาพแต่ละภาพมีขนาดไม่เท่ากันดังนั้นการพิจารณาเฉพาะความถี่อย่างเดียวอาจจะทำให้เกิดความผิดพลาดในการตีความดังนั้นผู้วิจัยจึงทำการนอร์มอไรซ์ (normalization) โดยนำความถี่ของวิชวลเวิร์ดทั้งหมดของทุกภาพในระบบมาหารแสดงสูตรการคำนวณได้ดังสมการที่ 2

(ก)



(ข)

ภาพที่ 5 *แสดงผลการสกัดคุณลักษณะซิฟท์ออกมาจากรูปภาพ 2 ภาพที่มีมุมกล้องและขนาดของภาพที่ต่างกัน*

$$n_i = \frac{f(v_i)}{\sum_{j=0}^{N} f(v_j)} \quad (2)$$

หลังจากการเพิ่มประสิทธิภาพของการรู้จำรถยนต์โดยใช้คุณลักษณะซิฟท์แล้ว สิ่งสำคัญอีกประการหนึ่งคือ การตรวจสอบการชนการของวัตถุ ในบทความนี้เสนอการตรวจสอบการชนกันของรถยนต์จากการเคลื่อนที่ของวัตถุซึ่งจะอธิบายในหัวข้อถัดไป

3.4 การตรวจจับการเคลื่อนที่ของรถยนต์ในวีดีทัศน์

สำหรับการตรวจจับการเคลื่อนที่ของรถยนต์จะทำการตรวจสอบความแตกต่างของพิกัดของวัตถุรถยนต์ (x, y) จากเฟรม (I) ที่ต่อเนื่องกัน $(I_k(x, y), I_{k-1}(x, y))$ ในวีดีทัศน์ โดยภาพจะถูกแปลงเป็นภาพสีเทา (gray-scale image) ดังนั้นตำแหน่งการเคลื่อนที่ของรถยนต์สามารถคำนวณได้จากสมการที่ 3

$$L(x, y) = |I_k(x, y) - I_{k-1}(x, y))| \geq 0 \quad (3)$$

ถ้า $L(x, y)$ มีค่ามากกว่า 0 แสดงว่ารถยนต์มีการเคลื่อนที่

3.5 การวิเคราะห์การเกิดการชนกันของวัตถุในวีดีทัศน์

ในหัวข้อนี้เป็นขั้นตอนการตรวจสอบว่าเกิดการชนกันของรถยนต์หรือไม่ โดยใช้สมมติฐานที่ว่าเมื่อรถเกิดอุบัติเหตุจะมีการเปลี่ยนแปลงของความเร็วอย่างฉับพลัน (acceleration) มีการเปลี่ยนขนาดของภาพรถยนต์ (size) และทิศทางของการเคลื่อนที่ของรถยนต์ (direction) ซึ่งข้อมูลเหล่านี้จะถูกนำมาใช้ในการวิเคราะห์การเกิดอุบัติเหตุในวีดีทัศน์

สมมติฐานที่ 1 รถยนต์จะมีการเปลี่ยนแปลงความเร็วอย่างฉับพลันเมื่อเกิดอุบัติเหตุขึ้น

เมื่อรถยนต์เกิดอุบัติเหตุ การเปลี่ยนแปลงความเร็ว (S) ของรถยนต์จะมีค่าสูง เช่น มีการเพิ่มความเร็วหรือลดความเร็วอย่างฉับพลัน สำหรับเทคนิคในการตรวจจับการเปลี่ยนแปลงความเร็วนี้เราสามารถใช้กระบวนการตรวจจับการเคลื่อนที่ของรถยนต์มาช่วย ในการคำนวณหาความเร็วรถยนต์จะคำนวณจากเฟรมที่รถยนต์เข้ามาในเฟรม (F_0) จนถึงเฟรมที่รถยนต์ออกไปจากเฟรม (F_n) และอัตราความเร็วของแต่ละเฟรม (frame rate, FR) ความเร็วของรถยนต์สามารถคำนวณได้ดังสมการที่ 4-6 คือการนำความเร็วของรถ ณ เวลาที่เข้ามาในกล้องลบกับความเร็วของรถ ณ สุดท้ายก่อนที่รถจะจอดนิ่ง

$$\Delta t = (F_n - F_0) \times FR \quad (4)$$

$$speed = dist \mid \Delta t \quad (5)$$

$$\Delta t = speed_{k-1} - speed_k \quad (6)$$

Δt คือเวลาทั้งหมดที่รถยนต์ใช้ในการเคลื่อนที่จากเฟรมเริ่มต้น (F_0) ไปยังเฟรมสุดท้าย (F_n) และ $dist$ คือ ระยะทางจริงที่รถยนต์เคลื่อนที่จากเฟรม (F_0) ไปยังเฟรม (F_n) ระบบจะคิดว่ามีอุบัติเหตุเกิดขึ้นถ้า Δs เพิ่มขึ้นแล้วหยุดอย่างฉับพลันในช่วงเวลา Δt

สมมติฐานที่ 2 เมื่อเกิดอุบัติเหตุขึ้นจะส่งผลให้การเปลี่ยนแปลงขนาดของภาพรถยนต์จะมีค่าสูง (high variation of the car size)

โดยปกติภาพของรถยนต์จะมีการเปลี่ยนแปลง เช่น เมื่อรถวิ่งห่างออกไปจากกล้อง ขนาดของรถยนต์จะเล็กลงหรือจะมีขนาดใหญ่ขึ้นเมื่อรถยนต์วิ่งเข้าหากล้อง แต่การเปลี่ยนแปลงขนาด (variation rate) ของรถยนต์จะมีค่าน้อย

แต่ในทางตรงกันข้ามเมื่อรถยนต์เกิดอุบัติเหตุการเปลี่ยนแปลงขนาดของรถยนต์จะเกิดขึ้นอย่างฉับพลันและส่งผลให้ความแตกต่างของขนาดรถยนต์จะมีค่าสูงกว่าปกติในช่วงเวลาอันสั้นซึ่งการเปลี่ยนแปลงขนาดของภาพรถยนต์นี้สามารถใช้เป็นปัจจัยในการวิเคราะห์การเกิดอุบัติเหตุได้ สมการในการคำนวณการเปลี่ยนแปลงขนาดของรถยนต์แสดงดังสมการที่ 7

$$\Delta a = A_t - A_{t-1} \quad (7)$$

เมื่อ A คือขนาดรถยนต์ที่คำนวณได้ ณ เวลา t

สมมติฐานที่ 3 เมื่อเกิดอุบัติเหตุจะส่งผลให้เกิดการเปลี่ยนแปลงทิศทางของรถยนต์อย่างเฉียบพลันในวีดีทัศน์ ทิศทาง (Direction) การวิ่งของรถยนต์สามารถตรวจสอบได้จากเฟรมของวีดีทัศน์ (frame) ที่ต่อเนื่องกัน แต่หากเกิดอุบัติเหตุเกิดขึ้น ทิศทางการวิ่งของรถยนต์จะเปลี่ยนแปลงอย่างฉับพลันโดยตรวจสอบจากมุมบนระนาบเว็กเตอร์ (vector) ระหว่างเฟรมของวีดีทัศน์ กำหนดให้ V_n เป็นเว็กเตอร์การเคลื่อนที่ของรถยนต์บนระนาบที่ควรจะเป็น (โดยการคำนวณของระบบ) และ V_i เป็นเว็กเตอร์การเคลื่อนที่ของรถยนต์ที่ตรวจสอบได้จริงจากภาพวีดีทัศน์ของเฟรมที่ต่อจากเฟรมของ V_n และ θ เป็นมุมที่อยู่บนระนาบของเว็กเตอร์ V_n และ V_i โดยคำนวณได้จากสมการที่ 8 [21]

$$\Delta d = \cos\theta = \frac{V_n - V_i}{|V_n||V_i|} \quad (8)$$

ถ้าค่า $\cos\theta$ มีค่าสูงกว่าค่าที่ได้จากการทดลอง (threshold = 4) แสดงว่ารถยนต์คันดังกล่าวมีการเคลื่อนที่ที่ไม่ปกติและมีความน่าจะเป็นของการเกิดอุบัติเหตุได้

เมื่อทำการตรวจสอบปัจจัยที่ใช้ในการวิเคราะห์การเกิดอุบัติเหตุในแต่ละวีดีทัศน์ครบทั้งหมด ระบบจะนำค่าที่ได้ทั้ง 3 ปัจจัยมาคำนวณ หากผลรวมของค่าที่ได้จากทั้ง 3 ปัจจัย (สมการ 9) มีค่ามากกว่าค่าที่ได้จากการทดลอง (threshold = 150) ระบบจะสรุปว่าเกิดอุบัติเหตุเกิดขึ้นในวีดีทัศน์นั้นๆ

$$\Delta s + \Delta a + \Delta d \geq threshold \quad (9)$$

ดังนั้นสรุปอัลกอริทึมในการตรวจจับอุบัติเหตุในวีดีทัศน์ได้ดังอัลกอริทึมที่ 1

อัลกอริทึม 1 การตรวจจับอุบัติเหตุในวีดีทัศน์
1: ตรวจจับภาพรถยนต์ในวีดีทัศน์ (vehicle object detection) 2: ตรวจจับการเคลื่อนที่ของรถยนต์ (moving vehicle tracking) 3: ตรวจสอบปัจจัยการเกิดอุบัติเหตุในวีดีทัศน์โดยคำนวณค่าและ 4: ประเมินการเกิดอุบัติเหตุจากผลรวมของปัจจัยต่างๆ (สมการที่ 9) และทำการสรุปว่าเกิดอุบัติเหตุในวีดีทัศน์นั้นๆ หรือไม่

4. การทดลองและอภิปรายผล

สำหรับต้นแบบที่ถูกพัฒนาขึ้นนี้จะจำกัดการทดลองอยู่ที่ภาพในวีดีทัศน์ที่รวมรวมจากเว็บไซต์ต่างๆ และยูทูบ (Youtube.com) และสำหรับในการวัดประสิทธิภาพการตรวจจับการเกิดอุบัติเหตุในวีดีทัศน์ที่โดยใช้อัลกอริทึมที่นำเสนอนั้นสามารถประเมินได้จากสมการที่ 10 เรียกว่าอัตราความถูกต้องในการตรวจสอบข้อมูล (correct detection rate, δ)

$$\delta = \frac{\text{no.of correct detected accident video } (\alpha)}{\text{no.of detected accident video } (\beta)} \quad (10)$$

โดยที่ α หมายถึงจำนวนวีดีทัศน์ที่ระบบสามารถตรวจสอบว่าเกิดอุบัติเหตุขึ้นได้อย่างถูกต้อง และ β คือจำนวนวีดีทัศน์ทั้งหมดที่ระบบระบุว่ามีอุบัติเหตุเกิดขึ้นทั้งหมด ซึ่งค่านี้จะใช้เป็นตัวบอกถึงประสิทธิภาพความถูกต้องของการตรวจสอบการเกิดอุบัติเหตุในวีดีทัศน์ นอกจากนี้ยังมีค่าซึ่งบอกถึงความสามารถในการค้นหาวีดีทัศน์ที่เกิดเหตุการณ์อุบัติเหตุทั้งหมดถูกค้นพบมากน้อยเพียงใด โดยแสดงดังสมการที่ 11

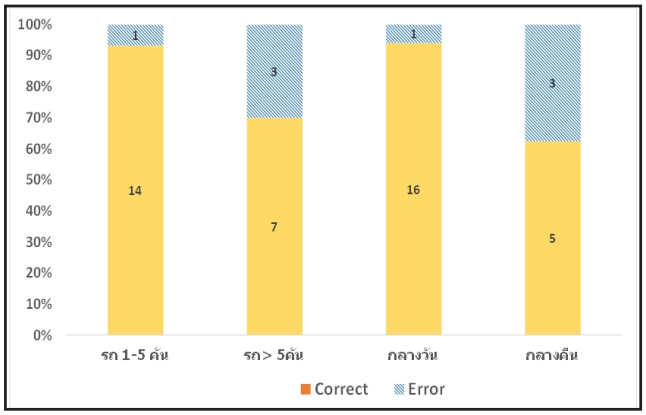
$$\varepsilon = \frac{\text{no.of correct detected accidentvideos } (\alpha)}{\text{no.of all accidentvideos in repository } (\phi)} \quad (11)$$

การทดลองนี้ทำการรวบรวมวีดีทัศน์ทั้งหมดจำนวน 50 วีดีทัศน์ ซึ่งแบ่งเป็นวีดีทัศน์ที่ไม่เกิดอุบัติเหตุ (No) ต่อวีดีทัศน์ที่เกิดอุบัติเหตุ (Yes) 25:25 ซึ่งจากการทดลองได้ผลลัพธ์การทดลองดังตารางที่ 1

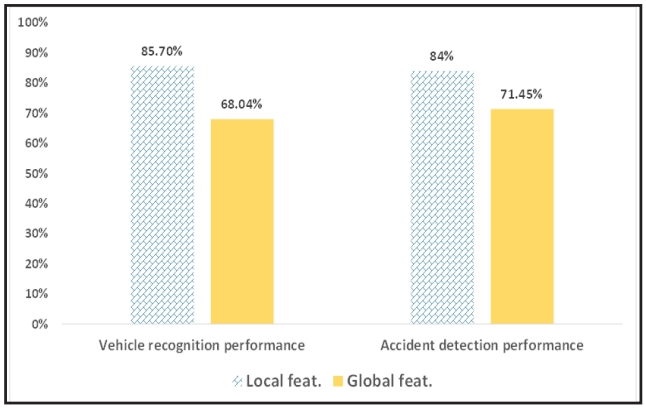
ตารางที่ 1 *ตารางค่าความสับสน (Confusion matrix)*

	Yes	No
Yes	21	8
No	4	29

จากตารางที่ 1 สามารถคำนวณค่าตามสมการที่ 10 และ 11 ได้ดังนี้ δ = 21/29 = 72.41% และ ε = 21/25 = 84.00% จะเห็นว่าระบบมีค่าเฉลี่ยความถูกต้องในการตรวจจับเหตุการณ์อุบัติเหตุสูงถึง 72.41% และสามารถค้นพบวีดีทัศน์ที่เกิดเหตุการณ์อุบัติเหตุถึง 84% จากวีดีทัศน์ที่มีเหตุการณ์อุบัติเหตุทั้งหมดในฐานข้อมูล จากการศึกษารายละเอียดพบว่าวีดีทัศน์ที่ระบบตรวจสอบผิดพลาดเกิดจากสาเหตุหลักคือ 1) ในวีดีทัศน์มีจำนวนรถยนต์มากเกินไป (มากกว่า 5 คันขึ้นไป) เมื่อจำนวนรถยนต์ในฉากนั้นๆ มากเกินไปทำให้การตรวจจับการเคลื่อนที่ของรถยนต์มีการผิดพลาดสูง เป็นผลทำให้ไม่สามารถตรวจสอบได้ว่ารถยนต์เหล่านั้นมีทิศทางการวิ่งที่ผิดปกติหรือไม่ และ 2) การตรวจจับภาพรถยนต์ในบางวีดีทัศน์ยังมีประสิทธิภาพไม่สูงมากนักสำหรับภาพวีดีทัศน์ในเวลากลางคืน ทำให้แสงมีผลต่อการหาจุดสำคัญของ (keypoints) ของซิฟท์ เมื่อหาจุดสำคัญได้ไม่ถูกต้อง มีผลทำให้ข้อมูลใน BVW ของรถยนต์มีค่าผิดเพี้ยนไป จึงทำให้การตรวจจับวัตถุที่เป็นรถยนต์มีความผิดพลาดสูงและทำให้ตรวจสอบการเกิดอุบัติเหตุไม่ได้หากเราพิจารณาแบ่งกลุ่มวีดีทัศน์อุบัติเหตุทั้งหมดออกเป็นกลุ่มตามปัญหาที่พบคือ 1) วีดีทัศน์ที่มีจำนวนรถยนต์ตั้งแต่ 1-5 คัน 2) วีดีทัศน์ที่มีจำนวนรถยนต์มากกว่า 5 คัน 3) วีดีทัศน์ในเวลากลางวัน และ 4) วีดีทัศน์ในเวลากลางคืน ซึ่งเมื่อพิจารณาแยกกลุ่มพบข้อมูลแสดงดังจากภาพที่ 6 ซึ่งจะเห็นว่าถ้าวีดีทัศน์มีจำนวนยานพาหนะมากกว่า 5 คันระบบจะตรวจสอบผิดพลาดมากกว่าปกติ (ผิดถึง 30%) เช่นเดียวกับวีดีทัศน์ที่เป็นภาพกลางคืน ระบบจะทำงานผิดพลาด 3 วีดีทัศน์จากทั้งหมด 8 วีดีทัศน์(37.5%)



ภาพที่ 6 *แสดงสัดส่วนการตรวจสอบการเกิดอุบัติเหตุแบ่งตามกลุ่มของวีดีทัศน์*



ภาพที่ 7 *เปรียบเทียบประสิทธิภาพการตรวจจับอุบัติเหตุในวีดีทัศน์โดยใช้คุณลักษณะโลคอลและโกลบอลของภาพในวีดีทัศน์*

นอกจากนี้เมื่อทำการทดลองถึงประสิทธิภาพของคุณลักษณะโลคอลของภาพในวีดีทัศน์ โดยทำการเปลี่ยนจากการใช้คุณลักษณะโลคอลเป็นคุณลักษณะโกลบอล เช่น สี รูปทรงและเส้นขอบ ผลการทดลองแสดงให้เห็นว่าการใช้คุณลักษณะแบบโลคอลช่วยเพิ่มประสิทธิภาพการรู้จำยานพาหนะเพิ่มขึ้น (ภาพที่ 7) 17.66% (85.70%-68.04%) ซึ่งส่งผลให้การตรวจจับอุบัติเหตุในวีดีทัศน์มีประสิทธิภาพเพิ่มขึ้น 12.55% (84%-71.45%) เหตุผลสำคัญเนื่องจากคุณลักษณะแบบโลคอลสามารถรู้จำวัตถุได้ ทั้งที่มีสี ขนาด รูปทรง และมุมกล้องที่เปลี่ยนไปได้ดีกว่าคุณลักษณะแบบโกลบอลนั่นเอง จึงทำให้การตรวจจับอุบัติเหตุในวีดีทัศน์สูงขึ้นตามไปด้วย

5. บทสรุป

งานวิจัยนี้ได้พัฒนาเทคนิคในการตรวจจับการเกิดอุบัติเหตุในภาพวีดีทัศน์ โดยใช้ปัจจัย 3 ประการในการวิเคราะห์การเกิดอุบัติเหตุบนถนน ได้แก่ ความเร็ว ขนาดของรถยนต์ และทิศทางการเคลื่อนที่ของรถยนต์ร่วมกับคุณลักษณะของภาพแบบโลคอล ผลการทดลองชี้ให้เห็นว่าปัจจัยทั้ง 3 สามารถช่วยให้การตรวจสอบการเกิดอุบัติเหตุมีความถูกต้อง 72.41% จุดเด่นของงานวิจัยนี้คือความน่าเสถียร (robust) ต่อการเปลี่ยนแปลงของมุมกล้องในลักษณะต่างๆ เช่น ขนาดและการหมุนของมุมกล้องที่เปลี่ยนแปลงไป ซึ่งจะไม่มีผลต่อประสิทธิภาพการทำงานของระบบ นอกจากนี้งานวิจัยนี้ยังนำเสนอการประยุกต์ใช้ออนโทโลยีมาร่วมกับคุณลักษณะของรูปภาพอีกด้วย ซึ่งทำให้การรู้จำรูปภาพ

รถยนต์มีถูกต้องมากยิ่งขึ้น

อย่างไรก็ตามระบบที่พัฒนาขึ้นนี้เป็นเพียงระบบต้นแบบเท่านั้นซึ่งยังมีข้อจำกัดอยู่หลายประการ เช่น หากจำนวนรถยนต์ในวีดีทัศน์มีมากเกินไป รวมทั้งหากภาพในวีดีทัศน์เกิดขึ้นในเวลากลางคืนหรือมีแสงน้อย จะเป็นสาเหตุที่ทำให้ระบบจะมีการทำงานที่ผิดพลาดสูงในงานวิจัยของ Lee และคณะ [22] ได้นำเสนอเทคนิคที่ใช้ในการตรวจจับอุบัติเหตุในช่วงเวลากลางคืนได้อย่างมีประสิทธิภาพและสามารถใช้เป็นแนวทางมาประยุกต์ใช้กับงานวิจัยที่นำเสนอในบทความนี้ได้ นอกจากนี้การสกัดเอาคุณลักษณะแบบโลคอล (SIFT) มีความซับซ้อนและใช้เวลามากในการประมวลผล ดังนั้นระบบที่นำเสนอจึงยังไม่เหมาะสมกับการตรวจสอบอุบัติเหตุแบบ real-time ข้อจำกัดอีกประการหนึ่งของระบบคือการทดลองเปรียบเทียบกับวิธีการที่นิยมใช้กันในบทความอื่นๆ เช่น HMM รวมถึงการทดสอบด้วยจำนวนวีดีทัศน์ที่มากขึ้นเพื่อให้ได้ผลการทดสอบที่น่าเชื่อถือมากกว่าเดิมซึ่งเป็นข้อจำกัดที่ผู้วิจัยจะต้องทำการพัฒนาต่อไปในอนาคต

6. เอกสารอ้างอิง

[1] Y. Zou, G. Shi, H. Shi, and Y. Wang. “Image Sequences Based Traffic Incident Detection for Signaled Intersections Using HMM.” *In Proceedings of the 9th International Conference on Hybrid Intelligent Systems*, Vol. 1, pp. 257 –261, 2009.

[2] H. Ikeda, Y. Kaneko, T. Matsuo, and K. Tsuji. “Abnormal Incident Detection System Employing Image Processing Technology.” *In Proceedings of the IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems*, pp. 748 –752, 1999.

[3] C.-P. Lin, J.-C. Tai, and K.-T. Song, “Traffic monitoring based on Real-time Image Tracking.” *In Proceedings of the IEEE International Conference on Robotics and Automation*, Vol. 2, pp. 2091 – 2096, 2003.

[4] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi. “Traffic Monitoring and Accident Detection at Intersections.” *IEEE Transaction of Intelligent Transportation Systems*, Vol. 1, No. 2, pp. 108 –118, 2000.

[5] O. Akoz and M. E. Karsligil. “Video-based traffic accident Analysis at Intersections using Partial Vehicle Trajectories.” *In Proceedings of the IEEE 18th Signal Processing and Communications Applications Conference*, pp. 499 –502, 2010.

[6] C. Prabha, R. Sunitha, and R. Anitha. “Automatic vehicle accident detection and messaging system Using GSM and GPS modem.” *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE)*, Vol. 3, No. 7, pp. 10723–10727, 2014.

[7] L.-W. Tsai, J.-W. Hsieh, and K.-C. Fan. “Vehicle Detection using Normalized Color and Edge Map.” *In Proceedings of the IEEE International Conference on Image Processing (ICIP 20005)*, Vol. 2, pp. 598–601, 2005.

[8] A. Bensrhair, M. Bertozzi, A. Broggi, A. Fascioli, S. Mousset, and G. Toulminet. “Stereo vision-based feature extraction for vehicle detection.” *In Proceedings of the IEEE Intelligent Vehicle Symposium*, Vol. 2, pp. 465–470, 2002.

[9] K. Huh, J. Park, J. Hwang, and D. Hong. “A stereo vision-based obstacle detection system in vehicles.” *Optics Lasers Engineering*, Vol. 46, No. 2, pp. 168–178, 2008.

[10] C. Tomasi and T. Kanade. “Detection and Tracking of Point Features.” *International Journal of Computer Vision, Technical report*, 1991.

[11] W. Kruger, W. Enkelmann, and S. Rossle. “Real-time estimation and tracking of optical flow vectors for obstacle detection.” *In Proceedings of the Intelligent Vehicles Symposium*, pp. 304–309, 1995.

[12] D. G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints.” *International Journal of Computer Vision (IJCV)*, Vol. 60, No. 2, pp. 91–110, 2004.

[13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. “A Comparison of Affine Region Detectors.” *International Journal of Computer Vision (IJCV)*, Vol. 65,

No. 1, pp. 43–72, 2005.

[14] L. AlSumait and C. Domeniconi. "Text Clustering with Local Semantic Kernels." *Survey of Text Mining II: Clustering, Classification, and Retrieval,* M. W. Berry and M. Castellanos, Eds. London, United Kingdom: Springer-Verlag London Limited, pp. 87–105, 2008.

[15] Y.-G. Jiang and C.-W. Ngo. "Visual Word Proximity and Linguistics for Semantic Video Indexing and Near-Duplicate Retrieval." *Computer Vision and Image Understanding*, Vol. 113, No. 3, pp. 405–414, 2009.

[16] J. Sivic, B. C. Russell, A. Zisserman, W. T. Freeman, and A. A. Efros. "Unsupervised Discovery of Visual Object Class Hierarchies." *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[17] L. Wang, Z. Lu, and H. H. Ip. "Image Categorization Based on a Hierarchical Spatial Markov Model." *In Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns*, pp. 766–773, 2009.

[18] B. Walter, K. Bala, M. Kulkarni, and K. Pingali. "Fast Agglomerative Clustering for Rendering." *In Proceedings of the IEEE Symposium on Interactive Ray Tracing*, pp. 81–86, 2008.

[19] D. M. Blei, A. Y. Ng, and M. I. Jordan. "Latent Dirichlet Allocation." *Journal of Machine Learning Research (JMLR)*, Vol. 3, pp. 993–1022, 2003.

[20] S. Poslad and K. Kesorn. "A Multi-Modal Incompleteness Ontology model (MMIO) to enhance information fusion for image retrieval." *Information Fusion*, Vol. 20, pp. 225–241, 2014.

[21] Y.-K. Ki and D.-Y. Lee. "A Traffic Accident Recording and Reporting Model at Intersections." *IEEE Transaction on Intelligent Transportation Systems*, Vol. 8, No. 2, pp. 188–194, 2007.

[22] I. J. Lee. "An accident detection system on highway through CCTV with calogero-moser system." *In Proceedings of the 18th Asia-Pacific Conference on Communications*, pp. 522–525, 2012.