รายงานวิจัยฉบับสมบูรณ์

โครงการ

การสร้างตัวแทนคุณลักษณะขั้นสูงที่มีประสิทธิภาพโดยการใช้ข้อมูลที่ไม่มีการกำหนดเป้าหมาย

โดย นาย เอกชัย ไพศาลกิตติสกุล

พฤษภาคม 2558

สัญญาเลขที่ TRG5680074

รายงานวิจัยฉบับสมบูรณ์

โครงการ

การสร้างตัวแทนคุณลักษณะขั้นสูงที่มีประสิทธิภาพโดยการใช้ข้อมูลที่ไม่มีการกำหนดเป้าหมาย

นาย เอกชัย ไพศาลกิตติสกุล
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเกษตรศาสตร์

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
และมหาวิทยาลัยเกษตรศาสตร์

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) และมหาวิทยาลัยเกษตรศาสตร์ที่ให้การสนับสนุนทุนวิจัยในโครงงานนี้ นอกจากนี้ขอขอบคุณ รศ.ดร.สมหญิง ไทยนิมิต นักวิจัยที่ปรึกษาที่ได้ช่วยให้ข้อเสนอแนะและข้อคิดเห็น ตลอดจนสละเวลาอ่านงานวิจัยนี้

รศ.ดร.เอกชัย ไพศาลกิตติสกุล

พฤษภาคม 2558

สารบัญ

	หน้า
บทคัดย่อภาษาอังกฤษ	1
บทคัดย่อภาษาไทย	2
บทนำ	3
วัตถุประสงค์	7
วิธีทดลอง	7
ผลการทดลอง	11
สรุปและวิจารณ์ผลการทดลอง	15
ข้อเสนอแนะสำหรับงานวิจัยในอนาคต	15
เอกสารอ้างอิง	16
ภาคผนวก	
A. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.	19
B. ผลงานวิจัยที่ได้ส่งเพื่อรับการตีพิมพ์ในวารสารวิชาการนานาชาติ	20

Abstract

Project Code : TRG5680074

Project Title : An Effective formalism of higher-level feature representations using unlabeled data

Investigator : Assoc.Prof.Dr Ekachai Phaisangittisagul
Department of Electrical Engineering, Faculty of Engineering, Kasetsart University

E-mail Address : fengecp@ku.ac.th

Project Period : 2 years (June 2013 – June 2015)

Abstract:

High-level feature representation plays a crucial role in transforming raw input data (low-level) into a new informative representation for learning algorithms to improve the performance on supervised learning problems in computer vision tasks. In particular, dictionary learning for sparse coding has been widely used to generate high-level feature representation. In sparse coding, an input data can be represented as a sparse linear combination of a set of training overcomplete dictionary. However, one problem in traditional sparse coding is quite slow to find the corresponding coding coefficients due to an ℓ_0 / ℓ_1-norm optimization. A process was proposed to create not only a discriminative sparse coding but also an effective method to compute the coding coefficients with low computational effort. More specifically, a linear model of sparse coding prediction was introduced to estimate the coding coefficients by simply computing the matrix-vector product. Subsequently, the predicted coding coefficients were used as a high-level feature representation to train a classifier. The experimental results demonstrated that the proposed method achieved promising classification results on well-known benchmark image databases and also outperformed in terms of computation time on the test data.

Keywords : dictionary learning, high-level feature representation, K-SVD, object classification, sparse coding, supervised learning

บทคัดย่อ

รหัสโครงการ : TRG5680074

ชื่อโครงการ : การสร้างตัวแทนคุณลักษณะขั้นสูงที่มีประสิทธิภาพโดยการใช้ข้อมูลที่ไม่มีการกำหนดเป้าหมาย

ชื่อนักวิจัย : รศ.ดร.เอกชัย ไพศาลกิตติสกุล
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

ที่อยู่อีเมล์ : fengecp@ku.ac.th

ระยะเวลาโครงการ : 2 ปี (มิถุนายน 2556 – มิถุนายน 2558)

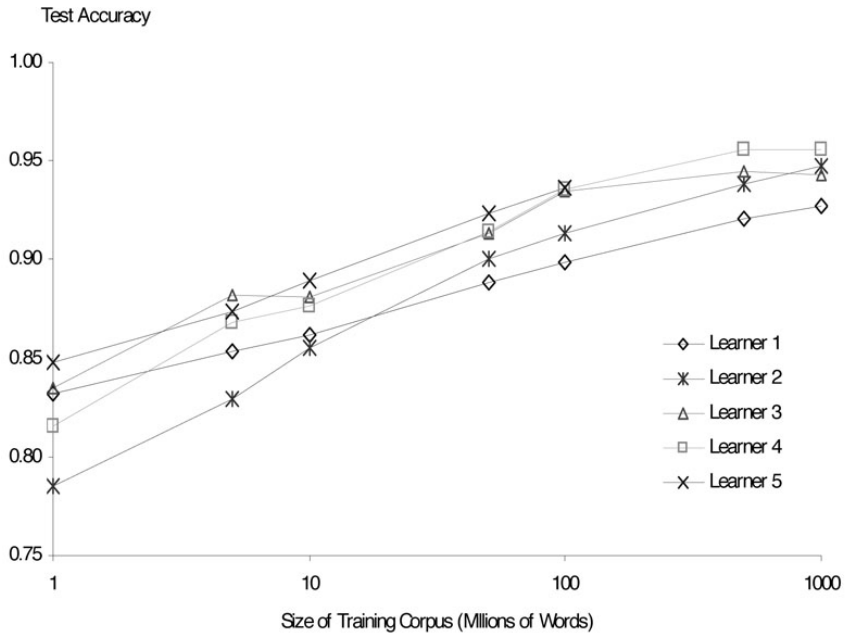
บทคัดย่อ:

ตัวแทนคุณลักษณะขั้นสูงมีบทบาทสำคัญในการแปลงข้อมูลอินพุทเดิมให้กลายเป็นคุณลักษณะใหม่ที่เป็นประโยชน์ต่อขั้นตอนวิธีการเรียนรู้เพื่อใช้ในการปรับปรุงสมรรถนะในปัญหาการเรียนรู้แบบมีผู้ฝึกสอนในงานคอมพิวเตอร์วิทัศน์ด้านต่างๆ โดยเฉพาะอย่างยิ่งการเรียนรู้ดิกชันนารีสำหรับสำหรับสปาซโค๊ดดิ้งซึ่งเป็นวิธีการที่นิยมใช้กันอย่างแพร่หลายในการสร้างตัวแทนคุณลักษณะขั้นสูง โดยสปาซโค๊ดดิ้งเป็นวิธีการแสดงข้อมูลในรูปของผลรวมเชิงเส้นของโอเวอร์คอมพลีทดิกชันนารีโดยบังคับให้จำนวนเบสิสที่ใช้มีจำนวนน้อย แต่อย่างไรก็ตามการคำนวณหาค่าสัมประสิทธิ์ประกอบในผลรวมเชิงเส้นค่อนข้างใช้เวลาในการคำนวณเนื่องจากส่วน l_0 / l_1 ในการบังคับเงื่อนไขของความเป็นสปาซ ซึ่งในส่วนงานวิจัยนี้ได้ทำการนำเสนอ ขั้นตอนวิธีในการสร้างคุณลักษณะขั้นสูงสปาซโค๊ดดิ้งซึ่งนอกจากจะมีประสิทธิภาพในการใช้จำแนกรูปแบบแล้ว ยังสามารถทำการคำนวณค่าสัมประสิทธิ์ได้จากผลคูณระหว่างเมตริกซ์กับเวคเตอร์ ทำให้ช่วยลดเวลาในการคำนวณลงอีกด้วย ทั้งนี้ค่าสัมประสิทธิ์สปาซโค๊ดดิ้งที่ได้จะถูกนำไปใช้ในการสร้างโมเดลสำหรับจำแนกรูปแบต่อไป ซึ่งขั้นตอนวิธีที่นำเสนอได้ทำการทดสอบเปรียบเทียบผลกับวิธีการอื่นๆ ที่ได้เผยแพร่ในวารสารวิชาการโดยใช้ฐานข้อมูลมาตรฐาน ซึ่งจากผลการทดลองที่ได้ปรากฏว่าขั้นตอนวิธีที่ได้นำเสนอนอกจากมีประสิทธิภาพในการจำแนกที่ดีแล้ว ยังใช่เวลาในการคำนวณน้อยกว่าขั้นตอนวิธีการอื่นๆ อีกด้วย

คำสำคัญ : การเรียนรู้ดิกชันนารี, ตัวแทนคุณลักษณะขั้นสูง, K-SVD, การจำแนกวัตถุ, สปาซโค๊ดดิ่ง, การเรียนรู้แบบมีผู้ฝึกสอน

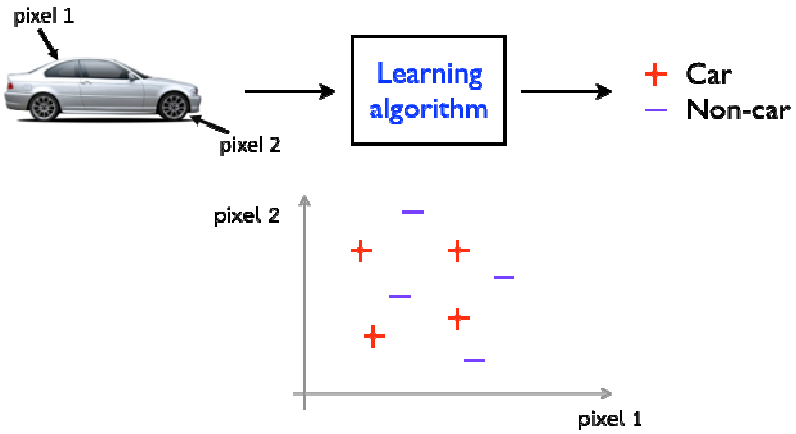
บทนำ

การเรียนรู้ของเครื่อง (machine learning) มีบทบาทที่สำคัญมากในสาขาปัญญาประดิษฐ์ (artificial intelligence) สำหรับการประยุกต์ในการจำแนกรูปแบบ (pattern classification) การประมาณค่าของฟังก์ชั่น (regression) และการจัดกลุ่ม (clustering) โดยการเรียนรู้ของเครื่องจะมุ่งไปที่การพัฒนาขั้นตอนวิธีเพื่อให้คอมพิวเตอร์มีความสามารถในการเรียนรู้และสามารถปรับตัวต่อการเปลี่ยนแปลงของสิ่งแวดล้อมโดยอาศัยการวิเคราะห์ข้อมูลที่ได้รับจากอุปกรณ์ตรวจวัดต่างๆ (sensors) ซึ่งการปรับตัวนี้จะกระทำผ่านการเปลี่ยนแปลงค่าพารามิเตอร์ของโมเดลที่สร้างขึ้น การเรียนรู้จากข้อมูลที่ใช้ในการฝึกสอน (training data) โดยส่วนใหญ่จะประกอบไปด้วยข้อมูลส่วนที่เป็นอินพุทและเอาท์พุทเป้าหมาย เราเรียกการเรียนรู้ลักษณะดังกล่าวนี้ว่า การเรียนรู้แบบมีผู้ฝึกสอน (supervised learning) ทั้งนี้โดยมีการตั้งสมมุติฐานว่าลักษณะของข้อมูลที่ใช้ในการฝึกสอนและข้อมูลที่จะนำมาใช้ในการทดสอบ (testing data) จะมีคุณลักษณะใกล้เคียงหรือคล้ายคลึงกัน จึงจำทำให้โมเดลที่สร้างขึ้นทำงานได้อย่างมีประสิทธิภาพ อย่างไรก็ตามปัญหาที่พบในการใช้งานจริงคือบางครั้งสมมุติฐาน ดังกล่าวไม่เป็นจริง ดังนั้นเมื่อคุณลักษณะของข้อมูลที่ใช้มีการเปลี่ยนแปลงไปจากเดิม ทำให้ต้องมีการพัฒนาโมเดลขึ้นใหม่สำหรับรองรับข้อมูลดังกล่าว ยกตัวอย่างเช่น ปัญหาการระบุตำแหน่งของผู้ใช้ WiFi [1] เพื่อใช้สำหรับตรวจจับหาตำแหน่งของผู้ใช้โดยการอาศัยข้อมูลการใช้ WiFi ก่อนหน้านี้ ซึ่งในทางปฏิบัติต้องใช้ความพยายามอย่างมากในการเก็บรวบรวมข้อมูลของสัญญาณของผู้ใช้ที่ตำแหน่งต่างๆ โดยเฉพาะอย่างยิ่งในบริเวณสภาพแวดล้อมที่กว้างสำหรับใช้เป็นข้อมูลในการพัฒนาโมเดลขึ้น นอกจากนี้เนื่องจากสัญญาณ WiFi เป็นฟังก์ชันเวลา ดังนั้นการพยากรณ์ตำแหน่งผู้ใช้สัญญาณ ณ เวลาหนึ่งให้ได้ค่าที่แม่นยำจึงเป็นเรื่องท้าทาย



รูปที่ 1. แสดงเปอร์เซ็นต์ความถูกต้องจากโมเดลการเรียนรู้แบบต่างๆ โดยเปรียบเทียบกับจำนวนข้อมูลที่ใช้ในการฝึกสอน

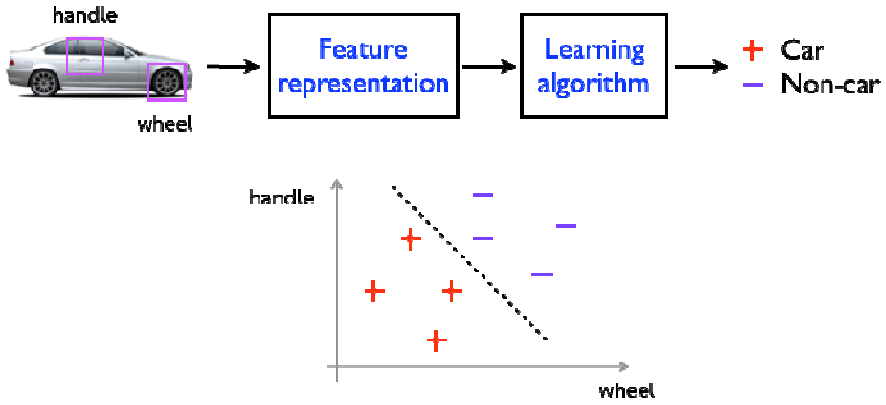
อีกปัญหาหนึ่งของการเรียนรู้แบบมีผู้ฝึกสอนก็คือเมื่อจำนวนข้อมูลที่ใช้ในการเรียนรู้หรือฝึกสอนมีจำนวนน้อย จากการศึกษาของ M. Banko และ E. Brill [2] พบว่าขั้นตอนวิธีที่ดีที่สุดไม่ได้หมายความว่าจะให้ผลลัพธ์ดีที่สุด แต่ขั้นตอนวิธีที่ได้จากการเรียนรู้ข้อมูลที่มีจำนวนมากกว่าอาจให้ผลลัพธ์ที่ดีกว่าดังแสดงในรูปที่ 1. จากภาพจะเห็นว่าแต่ละโมเดลที่พัฒนาจะให้ค่าความถูกต้องเพิ่มสูงขึ้นเมื่อจำนวนข้อมูลที่ใช้เพิ่มขึ้น ดังนั้นโมเดลที่พัฒนาขึ้นจากขั้นตอนวิธีใดๆ สามารถให้ประสิทธิภาพที่ดีกว่าโมเดลอื่นๆ ถ้าหากมีจำนวนข้อมูลที่ใช้ในการฝึกสอนมากกว่าได้ [3]



รูปที่ 2. ตัวอย่างโมเดลการจำแนกรถยนต์อาศัยการเรียนรู้แบบทั่วไป

สำหรับการประยุกต์การเรียนรู้ของเครื่องในส่วนการจำแนกวัตถุ เช่น การจำแนกรถยนต์จากข้อมูลอินพุทภาพดิจิตอล ขนาด 100x100 พิกเซล ซึ่งรูปภาพอินพุทนี้สามารถแสดงให้อยู่ในรูปของเวคเตอร์ขนาด $R^{10,000}$ จากนั้นอินพุทนี้จะถูกนำไปใช้ในการเรียนรู้ของโมเดลที่จะพัฒนาขึ้น โดยในกรณีนี้ค่าของสมาชิกแต่ละตัวในเวคเตอร์จะแทนค่าความเข้มแสงของแต่ละพิกเซล ในทางปฏิบัติการสร้างโมเดลสำหรับการจำแนกรถยนต์จากอินพุทเวคเตอร์ดังกล่าว จะต้องคำนึงถึงผลของการเปลี่ยนมุมกล้องจากอินพุทรูปภาพ สภาวะแสง การที่รถยนต์ถูกบดบังด้วยวัตถุอื่นๆ และลักษณะรูปร่างรถยนต์ที่เปลี่ยนแปลงไป ซึ่งปัจจัยต่างๆ เหล่านี้ต่างมีผลต่อประสิทธิภาพในการจำแนกวัถตุทั้งสิ้น ซึ่งปัญหาดังกล่าวมักจะเกิดขึ้นสำหรับการสร้างโมเดลโดยอาศัยการเรียนรู้แบบมีผู้ฝึกสอน ยกตัวอย่างเช่น ถ้าต้องการสร้างโมเดลสำหรับการจำแนกรถยนต์โดยใช้ค่าพิกเซลความเข้มแสงจำนวน 2 พิกเซล ดังรูปที่ 2 จะพบว่าโมเดลที่ได้ประสบปัญหาในการจำแนกเนื่องจากโมเดลนี้พิจารณา ข้อมูลอินพุทภาพเพียง 2 พิกเซล จากทั้งหมด 10,000 พิเซล แต่สำหรับในกรณีที่เรามีข้อมูลคุณลักษณะเด่นจากข้อมูลภาพรถยนต์ เช่น ตำแหน่งแสดงล้อของรถยนต์ หรือตำแหน่งที่เปิดประตูรถดังตัวอย่างแสดงในรูปที่ 3 มาใช้ในการสร้างโมเดลสำหรับการจำแนกจะทำให้ผลลัพธ์ที่ได้จากการจำแนกมีประสิทธิที่ดีและยังช่วยให้โมเดลที่ได้มีความเรียบง่ายอีกด้วย ซึ่งคุณลักษณะดังกล่าวเรียกว่า คุณลักษณะขั้นสูง (high-level feature) ซึ่งคุณลักษณะดังกล่าวถือเป็นคุณสมบัติที่จำเป็นสำหรับการสร้างโมเดลที่มีประสิทธิภาพ จากตัวอย่างที่ผ่านมาสะท้อนให้เห็นถึง

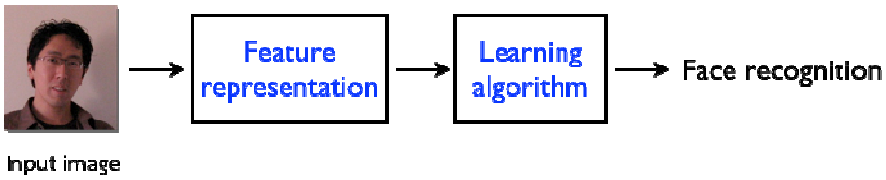
ความสำคัญของการเลือกหรือการสร้างคุณลักษณะของข้อมูลที่จะนำไปใช้ในการเรียนรู้ของโมเดลที่ต้องการพัฒนา อย่างไรก็ตาม ณ ปัจจุบันยังไม่มีวิธีการที่แน่นอนชัดเจนในการสร้างคุณลักษณะขั้นสูงดังกล่าว ซึ่งในทางปฏิบัติยังต้องอาศัยความเชี่ยวชาญเฉพาะด้านของมนุษย์อยู่ รวมถึงระยะเวลา และประสบการณ์ ตลอดจนความยากลำบากเมื่อมีฐานข้อมูลมีขนาดใหญ่ ดังนั้นจึงได้มีการวิจัยพัฒนาและศึกษาการ พัฒนาขั้นตอนวิธีในการสร้างคุณลักษณะขั้นสูงในการเรียนรู้แบบมีผู้ฝึกสอนโดยปราศจากความเชี่ยวชาญของมนุษย์ สำหรับการนำไปใช้ในการแก้ปัญหาต่างๆ



รูปที่ 3. ตังอย่างการโมเดลการเรียนรู้ที่อาศัยคุณลักษณะขั้นสูง

อย่างไรก็ตามในทางปฏิบัติ ถ้าจำนวนข้อมูลที่ใช้ในการฝึกสอนมีมากเพียงพอ มีความเป็นไปได้ที่เราจะสามารถพัฒนาโมเดลซึ่งอาศัยการเรียนรู้แบบมีผู้ฝึกสอนที่มีประสิทธิภาพขึ้นได้ แต่ในความเป็นจริงหลายๆ กรณีจำนวนข้อมูลที่ใช้ในการฝึกสอนมีอยู่อย่างจำกัดเนื่องมาจากหลากหลายสาเหตุ เช่น เสียค่าใช้จ่ายในการเก็บรวบรวมข้อมูลสูง ข้อมูลที่ใช้เป็นข้อมูลที่หาได้ยากและมีอยู่อย่างจำกัด ส่งผลให้การออกพัฒนาโมเดลที่ได้มีความสลับซับซ้อนสูง ทำให้การนำโมเดลไปใช้ในการ แก้ไขปัญหาจริงได้ผลลัพธ์ไม่ดีเท่าที่ควร (generalization error) เนื่องจากปัญหา overfitting ดังนั้นขั้นตอนวิธีการเรียนรู้ แบบมีผู้ฝึกสอนจึงให้ผลลัพธ์ไม่ดีเท่าที่ควรในกรณีที่จำนวนข้อมูลมีไม่มากเพียงพอต่อการเรียนรู้ของโมเดล

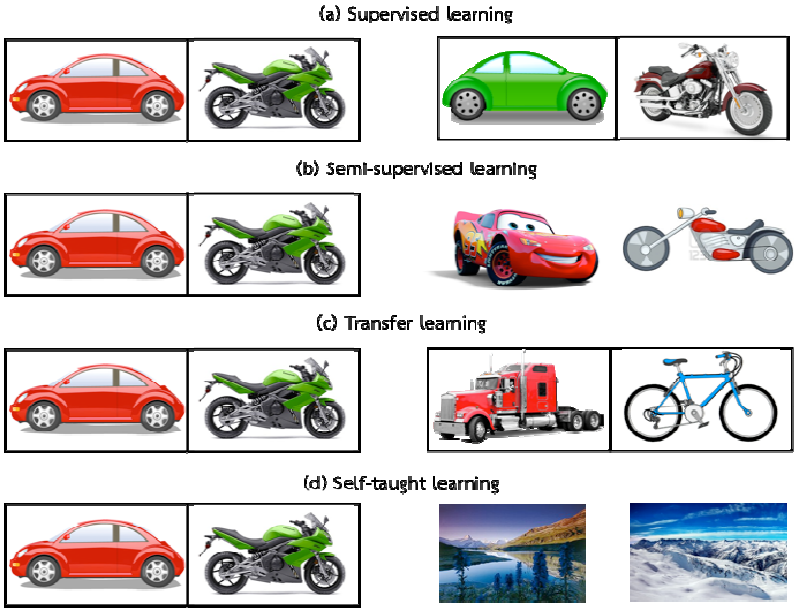
วิธีการหนึ่งที่นำมาใช้ในการแก้ไขปัญหาในกรณีที่มีจำนวนข้อมูลจำกัดก็คืออาศัยการสร้างคุณลักษณะใหม่ขึ้นมาแทนที่ข้อมูลอินพุท (x) ผ่านฟังก์ชั่นไม่เป็นเชิงเส้น (nonlinear function, $\varphi(x)$) จากนั้นจึงนำคุณลักษณะใหม่ที่ได้นำไปใช้ในการฝึกสอนโมเดลต่อไป ซึ่งกระบวนการดังกล่าวสามารถเขียนแสดงดังรูปที่ 4. โดยวิธีการเลือกฟังก์ชั่นไม่เชิงเส้นนั้นจะอาศัยการลองผิดลองถูก (trial and error) ซึ่งยังคงใช้ความพยายามจากมนุษย์เป็นหลัก



รูปที่ 4. แสดงกระบวนการสร้างโมเดลสำหรับรู้จำใบหน้าโดยอาศัยการสร้างตัวแทนคุณลักษณะขั้นสูง

ในความเป็นจริงสำหรับการสร้างตัวแทนคุณลักษณะขั้นสูงในการประยุกต์ใช้ในการจำแนกวัตถุ มนุษย์อาจใช้ระยะเวลาที่ยาวนานมากและต้องอาศัยผู้ที่มีความรู้ความชำนาญที่เกี่ยวข้องโดยตรง นอกจากนี้คุณลักษณะบางตัวอาจมีประโยชน์ต่อโมเดล แต่เมื่อนำคุณลักษณะที่หามาได้หลายๆ ตัวรวบรวมเข้าด้วยกันมาใช้สำหรับการฝึกสอนโมเดลอาจได้คุณลักษณะที่ไม่ดีอย่างที่คาดหวังไว้ก็ได้ ฉะนั้นจากที่กล่าวมาจะเห็นว่าการสร้างตัวแทนคุณลักษณะขั้นสูงนั้นเป็นสิ่งที่มีความจำเป็นอย่างยิ่งสำหรับการฝึกสอนโมเดลในการเรียนรู้ของเครื่อง แต่การได้มาซึ่งคุณลักษณะขั้นสูงที่เป็นประโยชน์ก็เป็นสิ่งที่ต้องใช้เวลาและต้องอาศัยความเชี่ยวชาญโดยเฉพาะ จากปัญหาดังกล่าวได้มีกลุ่มนักวิจัยจากหลายๆ สถาบันพยายามนำเสนอวิธีการสร้างคุณลักษณะขั้นสูงโดยอาศัยข้อมูลที่ไม่มีการระบุเป้าหมาย (unlabeled data) เพื่อช่วยในการเรียนรู้แบบมีผู้ฝึกสอนให้มีประสิทธิภาพเพิ่มขึ้น ซึ่งเรียกว่า semi-transfer learning (SSL) [4] โดยถูกพัฒนาต่อยอดจาก semi-supervised learning ซึ่งข้อดีของหลักการดังกล่าวคือช่วยแก้ไขปัญหาการพัฒนาโมเดลในกรณีที่มีจำนวนข้อมูลที่ใช้ในการฝึกสอนอย่างจำกัด โดยข้อมูลที่ไม่มีการระบุเป้าหมายที่นำมาใช้จะมีความสัมพันธ์หรือเกี่ยวข้องกับข้อมูลที่ใช้ในการฝึกสอนที่มีอยู่แต่ไม่จำเป็นต้องเหมือนกัน ซึ่งเป็นข้อแตกต่างกันที่สำคัญระหว่างการเรียนรู้แบบ semi-supervised และ transfer learning ซึ่งหลักการดังกล่าวพัฒนามาจากการเรียนรู้ของมนุษย์โดยอาศัยความรู้และประสบการณ์ที่เรียนรู้ในอดีตจากโดเมนหนึ่งนำไปประยุกต์ใช้ในการแก้ไขปัญหาในอีกโดเมนหนึ่ง ยกตัวอย่างเช่น การเรียนรู้จำลักษณะผลส้มสามารถนำไปช่วยในการรู้จำผลแอปเปิลได้ หรือในกรณีการจำแนกใบหน้ามนุษย์ เราสามารถทำการจัดเก็บฐานข้อมูลที่เกี่ยวกับรูปใบหน้ามนุษย์จากอินเทอร์เน็ต โดยที่รูปใบหน้าดังกล่าวไม่จำเป็นต้องเป็นรูปของบุคคลที่เก็บอยู่ในฐานข้อมูลการฝึกสอน แต่ข้อมูลดังกล่าวสามารถถ่ายโอนไปยังโมเดลสำหรับการรู้จำคุณลักษณะหรือองค์ประกอบของใบหน้ามนุษย์ได้ช่วยให้โมเดลสามารถตรวจจับหาบริเวณส่วนใบหน้าก่อนที่จะทำการระบุอัตลักษณ์ของบุคคลในภาพโดยที่โมเดลที่ใช้ไม่จำเป็นต้องมีความซับซ้อนเป็นพิเศษ อย่างไรก็ตามเมื่อไม่กี่ปีที่ผ่านมาได้มีกลุ่มนักวิจัยจากมหาวิทยาลัย Stanford University สหรัฐอเมริกาได้เสนอวิธีการประยุกต์ใช้ประโยชน์จากข้อมูลที่ไม่มีการระบุเป้าหมายมาช่วยในการดึงคุณลักษณะเด่นจากข้อมูลที่ใช้ในการฝึกสอนโมเดล มีชื่อเรียกว่า self-taught learning [5] โดยมีเป้าหมายเพื่อปรับปรุงประสิทธิภาพการเรียนรู้ของโมเดลให้ดีขึ้น โดยที่ข้อมูลที่ไม่มีการระบุเป้าหมายนี้ไม่จำเป็นต้องสัมพันธ์หรือเกี่ยงข้องกับข้อมูลที่ไม่มีการระบุเป้าหมายเหมือนอย่างกรณีของการเรียนรู้แบบ semi-supervised learning และ transfer learning แต่อย่างใด ทำให้การเรียนรู้โดยวิธีนี้สามารถนำไปประยุกต์ใช้ในการแก้ไขปัญหาได้หลากหลายและกว้างขว้างกว่า ทั้งนี้ self-taught learning จะอาศัยข้อมูลที่ไม่มีการ

กำหนดเป้าหมายมาทำการสร้างตัวดึงคุณลักษณะเด่น (feature extractor) สำหรับใช้ ในการดึงคุณลักษณะขั้นสูงจากข้อมูลที่ใช้ในการฝึกสอนเพื่อใช้ในการเรียนรู้ของโมเดลต่อไป โดยผลที่ได้นอกจากจะสามารถช่วยลดความยากลำบากในการดึงคุณลักษณะเด่นของมนุษย์แล้ว ยังช่วยทำให้ได้โมเดลที่มีประสิทธิภาพในการนำไปใช้ในการแก้ไขปัญหาการรู้จำวัตถุอื่นๆ โดยทั่วไป รูปที่ 5 แสดงการเปรียบเทียบหลักการเรียนรู้แบบต่างๆ ที่ กล่าวถึงโดยรูปที่ตีกรอบหมายถึงข้อมูลที่มีการระบุเป้าหมาย (labeled data) ส่วนรูปที่ไม่มีการตีกรอบคือข้อมูลที่ไม่มีการระบุเป้าหมาย



รูปที่ 5. แสดงการเปรียบเทียบการเรียนรู้แบบต่างๆ

วัตถุประสงค์

สำหรับในงานวิจัยนี้มีวัตถุประสงค์ดังต่อไปนี้

1. เพื่อพัฒนาขั้นตอนวิธีการสร้างคุณลักษณะขั้นสูงสำหรับการจำแนกรูปแบบที่มีประสิทธิภาพ
2. เพื่อพัฒนาวิธีการคำนวณหาคุณลักษณะขั้นสูงจากขั้นตอนวิธีที่นำเสนอที่มีความรวดเร็ว

วิธีทดลอง

วิธีการหนึ่งในการสร้างคุณลักษณะขั้นสูงซึ่งได้รับความนิยมอย่างมากเมื่อไม่นานมานี้คือ สปาซโค๊ดดิ้ง [6], [7] โดยข้อมูลจะถูกแสดงในรูปผลรวมเชิงเส้นของโอเวอร์คอมพลีทเบสิสเซตแบบสปาซ โดยมีขั้นตอนวิธีดังนี้คือ กำหนดให้มีชุดข้อมูลที่ใช้ในการเรียนรู้ จำนวน m ตัวอย่าง

$$X=\left[x^{(1)}, x^{(2)}, \ldots, x^{(m)}\right] \quad \text{and} \quad Y=\left[y^{(1)}, y^{(2)}, \ldots, y^{(m)}\right]$$

$$X_u=\left[x_u^{(1)}, x_u^{(2)}, \ldots, x_u^{(k)}\right]$$

$$x^{(i)}, x_u^{(i)} \in R^n, \quad \text{and} \quad y^{(i)} \in \{1, 2, \ldots, C\}$$

โดย $x^{(i)}$ คือข้อมูลอินพุทแบบมีการระบุเป้าหมาย

$x_u^{(i)}$ คือข้อมูลอินพุทแบบไม่มีการระบุเป้าหมาย

$y^{(i)}$ คือข้อมูลเป้าหมายที่สอดคล้องกับข้อมูลอินพุท $x^{(i)}$

สำหรับกระบวนการสร้างคุณลักษณะขั้นสูงตามหลักการของสปาซโค๊ดดิ้งสามารถแบ่งออกเป็น 2 ประเภทหลักๆ ดังนี้

1. การเรียนรู้การสร้างตัวแทนคุณลักษณะขั้นสูงโดยอาศัยข้อมูลแบบไม่มีการระบุเป้าหมาย ในการสร้างตัวแทนคุณลักษณะขั้นสูงจากข้อมูลที่ไม่มีการกำหนดเป้าหมายนั้น เพื่อเพิ่มประสิทธิภาพในการคำนวณและความยืดหยุ่นในการสร้างตัวแทน จึงได้นำหลักการของสปาซโค๊ดดิ้ง [7] มาประยุกต์ใช้ด้วย โดยในขั้นตอนวิธีนี้จะทำการเรียนรู้เพื่อหาชุดเบสิสหรือดิกชันนารี (D) ซึ่งใช้ในการสร้างตัวแทนแสดงข้อมูลอินพุทโดยมีเงื่อนไขกำหนดให้อินพุทใดๆ จะประกอบด้วยผลรวมเชิงเส้นของชุดเบสิสที่มีจำนวนน้อย ในทางปฏิบัติจำนวนของชุดเบสิสจะมีจำนวนมากกว่าจำนวนมิติของอินพุท (n) ซึ่งรู้จักในชื่อของโอเวอร์คอมพลีทเบสิสหรือดิกชันนารีโดยมีสมการที่จะต้องทำการหาคำตอบดังนี้

$$\langle D, A \rangle = \arg\min \sum \| X_u - DA \| \quad \textbf{s.t.} \quad \| a_i \|_0 \leq \gamma \qquad (1)$$

โดยที่ D คือชุดเบสิสหรือดิกชันนารี ส่วน $A = [a^{(1)}, a^{(2)}, \ldots, a^{(k)}]$ คือส่วนสัมประสิทธิ์ เรียกว่า coding coefficient ส่วน γ เรียกว่า sparsity constraint ซึ่งมีค่ามากกว่า 0 ทำหน้าที่ในการกำหนดน้ำหนักความสำคัญระหว่างเทอมที่ 1 และ 2 ในสมการที่ (1) โดยพารามิเตอร์ที่จะต้องทำการหาค่าที่เหมาะสมได้แก่ ชุดเบสิส D และสัมประสิทธิ์ของสปาซโค๊ดดิ้ง A โดยวัตถุประสงค์ของสมการที่ (1) และเงื่อนไขที่กำหนด ต้องการลดค่าความผิดพลาดที่เกิดจากตัวแทนที่สร้างขึ้นจากผลรวมเชิงเส้นของดิกชันนารีกับอินพุท $x^{(i)}$ และบังคับให้ $L_0 - Norm$ ของ $a^{(i)}$ มีลักษณะสปาซ (สมาชิกส่วนใหญ่มีค่าเท่ากับศูนย์) โดยค่าพารามิเตอร์สามารถทำการหาค่าที่เหมาะสมโดยอาศัยหลักการของการลดลงของค่าเกรเดียนท์ (gradient decent) โดยทำการสลับการปรับค่าพารามิ-เตอร์ ดังกล่าวทีละตัว ส่วนพารามิเตอร์ที่ไม่ได้ทำการปรับจะกำหนดให้มีค่าคงที่จนกระทั่งคำตอบลู่เข้าหาค่าที่เหมาะสม จากนั้นนำชุดดิกชันนารีที่ได้ไปทำการคำนวณหาค่าสัมประสิทธิ์ของ X โดยอาศัยดิกชันนารี D ที่ได้สร้างขึ้นจากข้อมูลแบบไม่มีผู้ฝึกสอนก่อนหน้านี้ โดยสามารถหาค่าได้จากปัญหาดังต่อไปนี้

$$\langle A \rangle = \arg\min \sum \| X - DA \| \quad \textbf{s.t.} \quad \| a_i \|_0 \leq \gamma \qquad (2)$$

2. การเรียนรู้การสร้างคุณลักษณะขั้นสูงโดยอาศัยข้อมูลแบบมีการระบุเป้าหมาย

การสร้างตัวแทนคุณลักษณะขั้นสูงโดยใช้ข้อมูลที่มีการระบุเป้าหมายมีลักษณะคล้ายกับวิธีก่อนหน้านี้ เพียงแต่ในวิธีการนี้การหาชุดเบสิสหรือดิกชันนารีให้ทำการคำนวณจากข้อมูลที่มีการระบุเป้าหมายโดยตรง [8]-[12] ซึ่งสามารถคำนวณหาได้จากฟังก์ชั่นต่อไปนี้

$$\langle D, A \rangle = \arg\min \sum \|X - DA\| \quad \textbf{s.t.} \quad \|a_i\|_0 \le \gamma \qquad (3)$$

ทั้งนี้ค่าสัมประสิทธิ์ที่ได้จากทั้งสองวิธีถือว่าเป็นคุณลักษณะขั้นสูง ซึ่งจะถูกนำไปใช้ในการสร้างโมเดลสำหรับการพยากรณ์ต่อไป ในทางปฏิบัติคุณลักษณะขั้นสูงที่ได้จากการใช้ข้อมูลที่มีการระบุเป้าหมายจะให้ผลลัพธ์ในการพยากรณ์ที่แม่นยำกว่า แต่วิธีการสร้างคุณลักษณะขั้นสูงที่อาศัยข้อมูลที่ไม่มีการระบุเป้าหมายจะเหมาะสมสำหรับกรณีที่ข้อมูลที่มีการระบุเป้าหมายมีจำนวนจำกัด

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อทำการพัฒนาขั้นตอนวิธีในการสร้างคุณลักษณะขั้นสูงแบบอัตโนมัติที่มีประสิทธิภาพสำหรับใช้ในการจำแนกรูปแบบและมีความรวดเร็วในการคำนวณ ซึ่งวิธีการที่นำเสนอตั้งชื่อเรียกว่า predictive sparse coding based on KSVD (SPC-KSVD) โดยมีขั้นตอนดังต่อไปนี้

<u>ขั้นตอนที่ 1</u>: การสร้างดิกชันนารีสำหรับการจำแนกจากข้อมูลที่มีการระบุเป้าหมาย โดยกำหนดให้

$X = [X_1, X_2, \ldots, X_C] \in R^{n \times m}$ โดย $X_i \in R^{n \times m_i}$ แทนข้อมูลอินพุทในกลุ่ม i^{th}-class จำนวน m_i ตัวอย่าง กำหนดให้ดิกชันนารี $D \in R^{n \times K}$ และค่าสัมประสิทธิ์ $A \in R^{K \times m}$ ทำการคำนวณหาค่าจากฟังก์ชั่นจุดประสงค์ต่อไปนี้

$$\langle D, T, W, A \rangle = \arg\min_{D,T,W,A} \|X - DA\|_2^2 + \lambda_1 \|M - TA\|_2^2 + \lambda_2 \|H - WA\|_2^2 \qquad (4)$$
$$s.t. \quad \|a_i\|_0 \le \gamma, \forall i$$

ซึ่ง $\lambda_1, \lambda_2,$ และ γ คือพารามิเตอร์ที่กำหนดความสำคัญของแต่ละเทอมและความสปาซของสัมประสิทธิ์ A ตามลำดับ

โดยเทอมที่ 1 ทำหน้าที่ควบคุมค่าความผิดพลาดที่เกิดจากการสร้างอินพุทใหม่กลับคืนจากผลคูณระหว่างดิกชันนารีและค่าสัมประสิทธิ์ (DA) ส่วนเทอมที่ 2 เป็นการบังคับให้สัมประสิทธิ์ที่มาจากอินพุทกลุ่มเดียวกันมีความสอดคล้องกัน โดย $M = [M_1, M_2, \ldots, M_m] \in R^{K \times m}$ และ M_i ถูกนิยามขึ้นจากค่าเฉลี่ยของดิกชันนารีย่อยที่มาจาก i^{th}-class เดียวกัน ยกตัวอย่างเช่น

สมมุติว่ามีข้อมูล $X = [x_1, x_2, x_3, x_4, x_5]$ และกลุ่ม class ของข้อมูลคือ $Y = [1^{st}, 3^{rd}, 2^{nd}, 3^{rd}, 1^{st}]$ ดังนั้นเมตริกซ์ M สามารถกำหนดได้ดังนี้

$$M = \begin{bmatrix} \tilde{d}_1 & 0 & 0 & 0 & \tilde{d}_1 \\ 0 & 0 & \tilde{d}_2 & 0 & 0 \\ 0 & \tilde{d}_3 & 0 & \tilde{d}_3 & 0 \end{bmatrix}$$

โดย $\tilde{d}_i$ เป็นค่าเฉลี่ยของดิกชันนารีซึ่งเป็นตัวแทนของ i^{th}-class สำหรับเทอมสุดท้ายในสมการที่ 4 ทำหน้าที่บังคับให้สัมประสิทธิ์สามารถแยกแยะระหว่างกลุ่มที่แตกต่างกัน ซึ่งจะทำให้การจำแนกข้อมูลเป็นไปอย่างมีประสิทธิภาพ การหาคำตอบจากฟังก์ชั่นจุดประสงค์ซึ่งแสดงในสมการที่ 4 สามารถใช้หลักการของขั้นตอนวิธี K-SVD [13]

<u>ขั้นตอนที่ 2</u>: การสร้างโมเดลเชิงเส้นสำหรับการพยากรณ์หาค่าสัมประสิทธิ์
เนื่องจากการหาค่าสัมประสิทธิ์ต้องใช้เวลาและทรัพยากรมากในการคำนวณหาค่าที่เหมาะสม เนื่องจากเงื่อนไขบังคับของความสปาซ ผู้วิจัยจึงคิดค้นวิธีการคำนวณที่มีความรวดเร็วและยังคงรักษาคุณลักษณะขั้นสูงสำหรับการจำแนกเดิมไว้ด้วย โดยวิธีการที่นำเสนอพัฒนามาจากปัญหาของการสร้างตัวแทนแบบสปาซในกรณีที่มีดิกชันนารีเรียบร้อยแล้ว ถ้ากำหนด ให้มีดิกชันนารี D ต้องการคำนวณหาค่าสัมประสิทธิ์ A ซึ่งสามารถหาค่าได้โดย

$$\begin{aligned} a_i &= \arg\min_a \|x_i - Da\|_2^2 \\ s.t.\ &\|a\|_0 \le \gamma \end{aligned} \qquad (5)$$

โดย $\|.\|_0$ คือ zero-norm ซึ่งจะมีค่าเท่ากับจำนวนสมาชิกที่มีค่าไม่เท่ากับศูนย์ ในที่นี้ได้ทำการนิยามปัญหาที่จะใช้ในการคำนวณหาค่าสัมประสิทธิ์อย่างมีประสิทธิภาพดังนี้

$$P = \arg\min_P \|A - PX\|_2^2 \qquad (6)$$

ซึ่ง P คือ projection matrix ที่ใช้สำหรับทำการคำนวณหาค่าสัมประสิทธิ์ A ในการหาคำตอบจากสมการที่ 6 สามารถหาค่าที่เหมาะสมจากฟังก์ชันต่อไปนี้

$$\begin{aligned} P^T &= \arg\min_{P^T} \left\|A^T - X^T P^T\right\|_2^2 \\ s.t.\ &\left\|p_i^T\right\|_1 \le \beta \end{aligned} \qquad (7)$$

ในการหาค่า P จากสมการที่ 7 สามารถประยุกต์ใช้ขั้นตอนวิธีของ Orthogonal Matching Pursuit (OMP) algorithm [14], [15]

ขั้นตอนที่ 3: การสร้างโมเดลสำหรับการจำแนกรูปแบบ

สำหรับโมเดลที่นำมาใช้ในการจำแนกรูปแบบมีชื่อเรียกว่า softmax regression [16] ซึ่งเป็นโมเดลแบบมีผู้ฝึกสอนที่ทำหน้าที่ประมาณค่า posterior probability ของแต่ละกลุ่ม class สำหรับข้อมูลอินพุท x_i ใดๆ ซึ่งค่าความน่าจะเป็นของแต่ละ class สามารถคำนวณได้ดังต่อไปนี้

$$h_\theta(x_i) = \begin{bmatrix} p(y_i = 1 \mid x_i; \theta_1) \\ p(y_i = 2 \mid x_i; \theta_2) \\ \text{M} \\ p(y_i = C \mid x_i; \theta_C) \end{bmatrix} = \frac{1}{\sum_{j=1}^{C} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ \text{M} \\ e^{\theta_C^T x_i} \end{bmatrix}$$

โดยที่ $\theta = [\theta_1, \theta_2, \ldots, \theta_c]$ คือเมตริกซ์ของพารามิเตอร์สำหรับโมเดล softmax regression ซึ่งคำนวณมาจากฟังก์ชั่นจุดประสงค์ต่อไปนี้โดยวิธีการลดลงของค่าเกรเดียน

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{C} 1\{y_i = j\} \log \frac{e^{\theta_l^T x_i}}{\sum_{l=1}^{C} e^{\theta_l^T x_i}}\right] + \frac{\lambda}{2}\sum_{i=1}^{C}\sum_{j=1}^{n}\theta_{ij}^2$$

ซึ่ง λ คือพารามิเตอร์สำหรับ regularization สำหรับป้องกันการเกิด overfitting ของโมเดล ในขณะที่ $1\{\cdot\}$ เรียกว่า indicator function ซึ่งจะให้ค่าเป็น 1 เมื่อ argument เป็นจริงและค่า 0 เมื่อ argument เป็นเท็จ ทั้งนี้การจะทำการพยากรณ์ว่าอินพุท x_i ใน class ใด จะทำการพิจารณาจากค่า $p(y = c \mid x_i)$ ที่มากที่สุดนั้นเอง

ผลการทดลอง

ในการทดลองเพื่อเปรียบเทียบประสิทธิภาพขั้นตอนวิธีการที่ได้นำเสนอกับวิธีการเรียนรู้ของดิกชันนารีสำหรับการจำแนกอื่นๆ ข้อมูลที่ใช้ทดสอบจะเป็นข้อมูลกลางที่ได้รับการยอมรับสำหรับการรู้จำรูปแบบได้แก่ Extended YaleB [17], AR face [18], Caltech101 [19] และ 15 scene [20] นอกจากนี้ค่าเวลาเฉลี่ยในการพยากรณ์ของวิธีการที่นำเสนอจะถูกทำการวัดผลบนเครื่องคอมพิวเตอร์โน๊ตบุ๊กซึ่งมีหน่วยประมวลผลกลางเป็น Intel core i5 ขนาด 1.8GHz โดยมีขนาดหน่วยความจำหลัก 8 GB ทั้งนี้ขั้นตอนวิธีที่ได้นำเสนอจะต้องมีการกำหนดค่าพารามิเตอร์ดังต่อไปนี้ $K, \lambda_1, \lambda_2, \gamma$ และ β โดยการเลือกค่าที่เหมาะสมจะใช้เทคนิค cross validation สำหรับขั้นตอนวิธีการอื่นๆ ที่นำมาใช้ในการเปรียบเทียบในการทดลองนี้ประกอบด้วย sparse representation-based classification (SRC), discriminative K-SVD (DKSVD), label-consistent K-SVD (LC-KSVD) และ Fisher discrimination dictionary learning (FDDL) จากผลการทดลองใน [11] ปรากฏ

ว่า วิธีการ LC-KSVD ให้ผลเวลาที่ใช้ในการคำนวณที่ดีที่สุด ดังนั้นในการทดลองนี้เวลาที่ใช้ในการคำนวณของวิธีการที่ได้นำเสนอจะถูกเปรียบเทียบกับผลที่ได้จากวิธีการ LC-KSVD เท่านั้น

- ข้อมูล Extended YaleB

ข้อมูลชุดนี้ประกอบด้วยภาพใบหน้าด้านหน้าจาก 38 บุคคล จำนวนทั้งสิ้น 2,414 รูป ภายใต้สภาวะการควบคุมระดับความเข้มแสงดังตัวอย่างแสดงในรูปที่ 1 โดยรูปของแต่ละบุคคลมีจำนวน 64 ภาพ ขนาด 192x168 พิกเซล จากนั้นทำการลดจำนวนมิติเหลือขนาด 504 โดยวิธี randomly generated matrix หรือเรียกว่า randomface พารามิเตอร์ที่เลือกใช้มีค่า $\lambda_1 = 4, \lambda_2 = 5, \gamma = 60$ และ $\beta = 60$ โดยขนาดของดิกชันนารีมีค่าเท่ากับ 950 สมาชิก โดยผลการจำแนกประเภทและเวลาที่ใช้ในการคำนวณเฉลี่ยต่อหนึ่งตัวอย่างสำหรับการทดสอบถูกแสดงในตารางที่ 1 ซึ่งความถูกต้องที่ดีที่สุดได้จากวิธีการ SRC แต่วิธีการที่นำเสนอเร็วกว่าวิธี LC-KSVD ประมาณเกือบ 4.5 เท่า

รูปที่ 1: แสดงตัวอย่างภาพใบหน้าจากฐานข้อมูล Extended YaleB [17]

ตารางที่ 1: แสดงผลการจำแนกและเวลาที่ใช้สำหรับฐานข้อมูล Extended YaleB

Method	Accuarcy(%)	Testing time (ms)
K-SVD	93.1	-
SRC	97.2	-
D-KSVD	94.1	-
FDDL	91.9	-
LC-KSVD	95.0	0.408
PSC-KSVD	94.5	0.09

- ข้อมูล AR face

ข้อมูลชุดนี้ประกอบด้วยภาพสีของใบหน้าด้านหน้าอีกเช่นกัน ของ 126 บุคคล จำนวนทั้งสิ้น 4,000 รูป โดยทิ้งช่วงห่างในการถ่ายภาพประมาณ 2 สัปดาห์ ทั้งนี้แต่ละบุคคลนอกจากจะมีการเปลี่ยนแปลงระดับความเข้มแสงแล้ว สีหน้าที่แสดงออกและการเพิ่มวัตถุปิดบังบางส่วนของใบหน้าจากแว่นตา ผ้าพันคออีกด้วยดังตัวอย่างแสดงในรูปที่ 2 ในการทดลองรูปภาพจำนวน 2,000 ภาพ จากสุภาพบุรุษจำนวน 50 คน และสุภาพสตรีจำนวน 50 คน ถูกสุ่มเลือกสำหรับใช้ในการทดลอง โดยรูปของแต่ละบุคคลมีขนาด 165x120 พิกเซล จากนั้นทำการลดจำนวนมิติ

เหลือขนาด 504 โดยใช้ randomface matrix พารามิเตอร์ที่เลือกใช้มีค่า $\lambda_1 = 3, \lambda_2 = 2, \gamma = 60$ และ $\beta = 100$ โดยขนาดของดิกชันนารีมีค่าเท่ากับ 500 สมาชิก โดยผลการจำแนกประเภทและเวลาที่ใช้ในการคำนวณเฉลี่ยต่อหนึ่งตัวอย่างสำหรับการทดสอบถูกแสดงในตารางที่ 2 ซึ่งวิธีการที่ได้นำเสนอให้ค่าความถูกต้องที่ดีที่สุดที่ 98% และเร็วกว่าวิธี LC-KSVD ประมาณเกือบ 6 เท่า

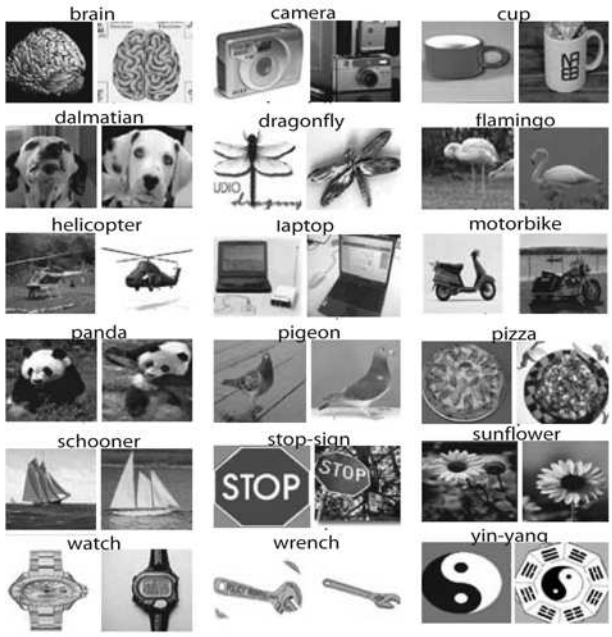
รูปที่ 2: แสดงตัวอย่างภาพใบหน้าจากฐานข้อมูล AR face [18]

ตารางที่ 2: แสดงผลการจำแนกและเวลาที่ใช้สำหรับฐานข้อมูล AR face

Method	Accuarcy(%)	Testing time (ms)
K-SVD	86.5	-
SRC	97.5	-
D-KSVD	88.8	-
FDDL	92.0	-
LC-KSVD	93.7	0.344
PSC-KSVD	98.0	0.055

- <u>ข้อมูล Caltech101</u>

ฐานข้อมูลสำหรับการจำแนกวัตถุซึ่งเป็นที่รู้จักกันอย่างดีได้แก่ Caltech101 ซึ่งประกอบด้วยภาพจำนวนทั้งสิ้น 9,144 รูป จาก 102 กลุ่ม โดยจำนวนข้อมูลในแต่ละกลุ่มอยู่ระหว่าง 31 ถึง 800 รูป และมีขนาด 300x200 พิกเซล ตัวอย่างภาพของข้อมูลนี้แสดงในรูปที่ 3 ในการดึงคุณลักษณะเด่นขั้นตอนวิธีการได้แก่ scale-invariant feature transform (SIFT) และ spatial pyramid matching (SPM) ถูกนำมาใช้ นอกจากนี้ principal component analysis (PCA) ถูกใช้ในการลดจำนวนมิติให้เหลือ 3,000 พารามิเตอร์ที่เลือกใช้ $\lambda_1 = 10, \lambda_2 = 5,\ \gamma = 60$ และ $\beta = 50$ โดยขนาดของดิกชันนารีมีค่าเท่ากับ 3,060 ผลการจำแนกประเภทและเวลาที่ใช้ในการคำนวณเฉลี่ยต่อหนึ่งตัวอย่างสำหรับการทดสอบถูกแสดงในตารางที่ 3 ซึ่งวิธีการที่ได้นำเสนอให้ค่าความถูกต้องที่ดีที่สุดที่ 73.9% และเร็วกว่าวิธี LC-KSVD ประมาณ 12 เท่า



รูปที่ 3: แสดงตัวอย่างภาพใบหน้าจากฐานข้อมูล Caltech101 [19]

ตารางที่ 3: แสดงผลการจำแนกและเวลาที่ใช้สำหรับฐานข้อมูล Caltech101

Method	Accuarcy(%)	Testing time (ms)
K-SVD	73.2	-
SRC	70.7	-
D-KSVD	73.0	-
LC-KSVD	73.6	2.392
PSC-KSVD	73.9	0.198

- <u>ข้อมูล 15 scene</u>

  ข้อมูลชุดนี้ประกอบด้วยภาพทิวทัศน์จำนวน 15 กลุ่ม โดยในแต่ละกลุ่มประกอบด้วยจำนวนตัวอย่างระหว่าง 200 ถึง 400 และมีขนาดเฉลี่ยเท่ากับ 300x250 พิกเซล รูปที่ 4 แสดงตัวอย่างภาพจากฐานข้อมูลนี้ พารามิเตอร์ที่ใช้มีค่า $\lambda_1 = 6, \lambda_2 = 2,\ \gamma = 80$ และ $\beta = 100$ โดยขนาดของดิกชันนารีมีค่าเท่ากับ 450 สมาชิก หรือ 30 สมาชิกต่อกลุ่ม ซึ่งวิธีการที่ได้นำเสนอให้ค่าความถูกต้องที่ดีที่สุดที่ 95.8% และเร็วกว่าวิธี LC-KSVD ประมาณ 8 เท่า

รูปที่ 4: แสดงตัวอย่างภาพใบหน้าจากฐานข้อมูล 15 scene [20]

ตารางที่ 4: แสดงผลการจำแนกและเวลาที่ใช้สำหรับฐานข้อมูล 15 scene

Method	Accuarcy(%)	Testing time (ms)
K-SVD	86.7	-
SRC	91.8	-
D-KSVD	89.1	-
LC-KSVD	94.2	0.396
PSC-KSVD	95.8	0.050

สรุปและวิจารณ์ผลการทดลอง

จากการทดลองจะเห็นว่าการเรียนรู้ของดิกชันนารีมีความสำคัญต่อการสร้างตัวแทนคุณลักษณะขั้นสูงแบบสปาซสำหรับการจำแนกรูปแบบ อย่างไรก็ตามการหาค่าสัมประสิทธิ์ต้องใช้เวลาในการคำนวณพอสมควรเนื่องจากเงื่อนไขบังคับความเป็นสปาซ ในงานวิจัยนี้ได้ทำการเสนอวิธีการสร้างคุณลักษณะขั้นสูงโดยที่ยังสามารถคำนวณหาค่าได้อย่างรวดเร็วโดยอาศัยผลคูณระหว่างเมตริกซ์กับเวคเตอร์ จากการทดลองกับข้อมูลมาตรฐานได้แก่ Extended YaleB, AR face, Caltech101, และ 15 scene พบว่าขั้นตอนวิธีการที่นำเสนอให้ผลการจำแนกรูปแบบมีความถูกต้องสูง นอกจากนี้ยังใช้เวลาในการคำนวณน้อยที่สุดเปรียบเทียบกับวิธีการอื่นๆ ดังนั้นวิธีการที่นำเสนอสามารถพิจารณานำไปประยุกต์ใช้ในการแก้ไขปัญหาแบบเวลาจริงได้

ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

ขั้นตอนวิธีที่ได้นำเสนอในงานวิจัยชิ้นนี้จะต้องมีการกำหนดค่าพารามิเตอร์ได้แก่ ขนาดของดิกชันนารี ค่า $\lambda_1, \lambda_2, \beta,$ และ γ ซึ่งในการทดลองใช้เทคนิค cross validation ในการค้นหาค่าที่เหมาะสม ดังนั้นการศึกษาวิธีการเลือกค่าที่เหมาะสมสามารถช่วยให้การออกแบบขั้นตอนวิธีมีความรวดเร็วมากยิ่งขึ้น

เอกสารอ้างอิง

[1] Pan, S. J., Yang, Q. “A survey on transfer learning”. IEEE Trans. on Knowledge and Data Engineering, Vol. 22, no. 10, 2010, pp. 1345-1359.

[2] Banko, M., Brill, E. “Mitigating the paucity-of-data problem: exploring the effect of training corpus size on classifier performance for natural language processing”, First International Conference on Human language technology research, 2001, pp. 1-5.

[3] Brants, T., Popat, A. C, Xu, P., Och, F. J., Dean, J. “Large language models in machine translation”, EMNLP-CoNLL. 2007.

[4] Zhu, X. “Semi-supervised learning literature survey”, Technical Report: 1530. Computer sciences, University of Wisconsin-Madison. 2005.

[5] Raina, R., Battle, A., Lee, H., Parker, B., Ng, A. Y. “Self-taught learning: Transfer learning from unlabeled data”, In Proceedings of the twenty-fourth International Conference of Machine Learning, 2007, pp. 759-766.

[6] Olshausen, B. A., Field, D. J. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, Nature, 381, pp. 607-609, 1996.

[7] K. Huang and S. Aviyente. “Sparse representation for signal classification”, Proc. Conf. Neural Information Processing Systems, pp. 609-616, 2006.

[8] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.31(2), pp. 210--227, 2009.

[9] Q. Zhang and B.X. Li, Discriminative K-SVD for dictionary learning in face recognition. Proc. IEEE conf. Computer Vision and Pattern Recognition, pp. 2691-2698, 2010.

[10] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, Locality-Constrained Linear Coding for Image Classification, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 3360 - 3367, 2010.

[11] Z. Jiang, Z. Lin, and L. Davis, Label consistent K-SVD: Learning a discriminative dictionary for recognition. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 35(11), pp. 2651-2664, 2013.

[12] M. Yang, L. Zhang, X.C. Feng, and D. Zhang, Fisher discrimination dictionary learning for sparse representation, Proc. IEEE Int'l Conf. Computer Vision, pp. 543 - 550, 2011.

[13] M. Aharon, M. Elad, and A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Processing, vol. 54(11), pp. 4311–4322, 2006.

[14] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, Conf. Rec. 27th Asilomar Conf. Signals, Systems, and Computers, vol. 1. pp. 40-44, 1993.

[15] J. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inf. Theory, vol. 50, pp. 2231-2242, 2004.

[16] A. Ng, J. Ngiam, C. Foo, Y. Mai, C. Suen, Softmax regression, http://ufldl.stanford.edu/wiki /index.php/ Softmax_Regression, 2013.

[17] A. Georghiades, P. Belhumeur, and D. Kriegman."From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23(6), pp. 643-660, 2001.

[18] A. Martinez, R. Benavente, The AR face database. CVC Technical Report (1998)

[191] L. Fei Fei, R. Fergus, and P. Perona, Learning Generative Visual Models from Few Training Samples: An Incremental Bayesian Approach Tested on 101 Object Categories, Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Generative Model Based Vision, pp. 746-751, 2004.

[20] S. Lazebnik, C. Schmid, and J. Ponce, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2169-2178, 2007.

ภาคผนวก

A. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

1. ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ
   - อยู่ระหว่างการพิจารณาตีพิมพ์ของทางวารสาร

2. การนำผลงานวิจัยไปใช้ประโยชน์
   - เชิงวิชาการ
     ได้มีการพัฒนาการเรียนการสอนด้านวิชา Principles of Machine Learning
   - สร้างนักวิจัยใหม่ในด้าน Machine learning โดยเน้นทางด้าน Sparse coding
     ประกอบด้วยนิสิตระดับปริญญาโทจำนวน 1 คน และ ระดับปริญญาตรีจำนวน 2 คน

3. อื่นๆ (เช่น ผลงานตีพิมพ์ในวารสารวิชาการในประเทศ การเสนอผลงานในที่ประชุมวิชาการ หนังสือ การจดสิทธิบัตร)

   การตีพิมพ์ผลงานวิจัยในการประชุมทางวิชาการระดับนานาชาติ
   - E. Phaisangittisagul, R. Chongprachawat, Post-processing of unsupervised dictionary learning in handwritten digit recognition*, in* International Symposium on Communications and Information Technologies (ISCIT 2014), Incheon, South Korea, September 24-26, 2014, pp.166-170.

B. ผลงานวิจัยที่ได้ส่งเพื่อการตีพิมพ์ในวารสารวิชาการนานาชาติ

Fast Predictive High-Level Feature Representation based on Discriminative Dictionary Learning

Ekachai Phaisangittisagul[a,*], Somying Thainimit[a]

[a] *Department of Electrical Engineering, Faculty of Engineering, Kasetsart University*
50 Phaholyothin Rd., Jatujak, Bangkok, 10900, Thailand

Abstract

High-level feature representation plays a crucial role in transforming raw input data (low-level) into a new informative representation for learning algorithms to improve the performance of supervised learning problems in computer vision tasks. In particular, dictionary learning for sparse coding has been widely used to generate high-level feature representation. In sparse coding, input data can be represented as a sparse linear combination of a trained overcomplete dictionary. However, one problem in traditional sparse coding is that it is quite slow to find the corresponding coding coefficients due to an ℓ_0/ℓ_1 optimization. A process was proposed to create not only a discriminative sparse coding but also an effective method to compute the coding coefficients with low computational effort. More specifically, a linear model of sparse coding prediction was introduced to estimate the coding coefficients by simply computing the matrix-vector product. Subsequently, the predicted coding coefficients were used as a high-level feature representation to train a classifier. The experimental results demonstrated that the proposed method achieved promising classification results on well-known benchmark image databases and also outperformed in terms of computation time on the test data.

Keywords: dictionary learning, high-level feature representation, K-SVD, object classification, sparse coding, supervised learning

1. Introduction

In supervised learning problems, a typical objective is to create a mapping model from input data to a target output. For example, a common task in object classification such as face classification, handwritten classification, or text classification is to build a learning model to map an input image represented by pixel intensity values to a predefined category of the object appeared in the image. Although, there are many effective supervised learning algorithms, learning input data (low-level features) directly is difficult to achieve high classification performance. This is due to the fact that learning a mapping from pixel intensity values to object class label is a complex nonlinear function to discover. In fact, the performance of supervised learning is highly dependent on the choice of data representation. As a result, many groups of researchers have attempted to propose methods that are able to capture latent and high-fidelity representation (high-level features) from the raw input data.

In signal processing, a signal or function $f(t)$ can often be described as a linear decomposition, $f(t) = \sum_{j=1}^{k} a_j \, \varphi_j$ where a_j are the coefficients and φ_j are the set of functions. A signal can be uniquely represented as a linear combination of those functions as long as it forms a basis set. This basis set is sometimes known in the computer vision community as a dictionary. In addition, the calculation of the coefficients can be done efficiently by an inner product between the input signal and the basis set. For example, a basis

*Corresponding author. Tel.: +66 2 797 0999 ext. 1506.
Email address: fengecp@ku.ac.th (Ekachai Phaisangittisagul)

set of a Fourier series consists of sines and cosines (or equivalently complex exponentials) functions at different harmonic frequencies. However, this predefined set of bases is less effective to model complex local structure of the images. Recently, an overcomplete basis set was proposed from an inspiration of a mechanism of human vision system. Olshausen and Field [35], [36] have revealed that the receptive fields can extract meaningful information from images based on sparse coding. Many studies have shown that sparse representation has been successfully applied to various applications of image restoration and compressed sensing [3],[4],[5],[9],[25],[46],[50],[54],[56]. Another essential motivation of the sparse coding is that it can be applied to unlabeled data for dictionary learning due to a limitation of labeled data. As a result, many researchers have focused on exploiting a sparse coding to learn a high-level feature representation [9],[23],[40],[44] in diverse tasks. Conceptually, sparse coding can be viewed as a way of constructing an approximation of an input data by a sparse linear combination of a set of overcomplete dictionary according to:

$$< D, A >= \arg\min_{D,A} ||X - DA||_2^2 + \lambda \|A\|_p \tag{1}$$

where X is a set of n-dimensional m input data, i.e., $X = [x_1, \ldots, x_m] \in \mathbb{R}^{n \times m}$. $D = [d_1, \ldots, d_K] \in \mathbb{R}^{n \times K} (K > n)$ is an overcomplete dictionary of K atoms and $A = [a_1, \ldots, a_m] \in \mathbb{R}^{K \times m}$ is a set of coding coefficients. λ is a regularization parameter. Finally, $\|A\|_p$ is an ℓ_p-norm regularization constraint on A to control the number of nonzero elements in its coding. In (1), the dictionary (D) and the corresponding coding coefficients (A) are the parameters to be optimized. Usually, this non-convex optimization problem is solved by alternatively iterating between the dictionary and coding coefficient updating. Note that all the dictionary atoms (d_i) should have unit norm to avoid the scenario in which the dictionary elements (atoms) have arbitrary large norm so that the coding coefficients are forced to have small value. In sparse representation, the key success counts on the choice of the dictionary. Traditionally, the dictionary can be obtained by either taking from off-the-shelf bases (e.g., wavelets) [15] or learning from the data. Although, off-the-shelf dictionaries might be universal to represent all types of data, these dictionaries are not well represented for specific applications (e.g., text or face classifications). To better capture the salient features of the data, various dictionary learning methods have been developed for particular tasks such as reconstruction and classification problems. The current prevailing approach to create a dictionary is based on learning from data which can be divided into two categories: unsupervised dictionary learning and supervised dictionary learning. An unsupervised dictionary learning does not used class information of the data to produce the dictionary. The learned dictionary is built with the goal of minimizing the residual error between reconstructed data and original data. Aharon *et al.* [1] proposed a K-SVD algorithm which can create a learned overcomplete dictionary from a set of unlabeled data. Such unsupervised dictionary learning led to improve results in image denoising [9],[31],[60], image compression [4], and image super-resolution [51]. However, without using label information from the data, the unsupervised dictionary learning is not powerful for classification tasks. The other category of dictionary learning is based on supervised learning in which the class label of the data is available to exploit in the dictionary learning. As a result, discrimination capability could be boosted by a dictionary resulting from a learning process or sparse coding, or both and thus better classification performance is achieved [31],[17],[29],[41],[57]. One of the main focuses of this study is to create a high-level feature representation for object classification with discriminative capability based on supervised dictionary learning strategy.

In general, an existing supervised dictionary learning can be categorized into three classes. In the first class, each dictionary atom is shared to all classes while the coding coefficients are exploited as a high-level feature to train a classifier [58]. Some approaches proposed to jointly learn a shared dictionary with a classifier while enforcing the coding coefficients to be discriminative [17]. For example, Mairal *et al.* [27] proposed an effective method to learn a shared dictionary for creating a discriminative model. Other strategies merged or chose dictionary atoms from an initial large dictionary using different criteria such as intraclass and interclass discrimination [15], mutual information of class distribution [24], and submodular dictionary learning [16]. The second class of supervised dictionary learning is designed to improve discrimination power among classes [6],[26],[28],[38],[41],[43],[52],[57],[59],[61] called class-specific dictionary learning in which each dictionary atom is learned for a single class only. In addition, the corresponding reconstruction error can be

used for class assignment. Yang *et al.* [55] applied the Fisher discrimination criteria to the objective function to encourage discriminative representation in the coding coefficients. Ramirez *et al.* [41] introduced class-specific dictionary to be independent by adding an incoherence promoting term. Wang *et al.* [48] presented a margin-based perspective to dictionary learning with improved classification power. The last class of supervised dictionary learning is to combine the shared dictionary with the class-specific dictionary named hybrid dictionary learning. Kong et al. [19] shown that the combination of class-specific dictionary and a common pattern pool led to more compact and more discriminative dictionary for classification. Zhou *et al.* [61] proposed a joint dictionary learning algorithm to leverage the correlation within a group of the visually similar object categories to enhance the discrimination of the learned dictionary.

In summary of sparse representation for classification, it consists of two phases: coding and classification. In coding, a sparse representation or coding coefficients (a_i) is determined from a dictionary (D) with some sparsity constraint. Then, the coding coefficients are used to build a classification model for class prediction. Most proposed dictionary learning algorithms have focused on designing synthesis dictionary to not only well represent the original data but also produce better classification results while adopting ℓ_o/ℓ_1-norm sparsity constraints. Traditionally, most of the existing sparse representation is based on an iterative learning process to solve for the solution leading to time-consuming in calculation on both training and testing phases, thus prohibiting real-time applications. Although, efficient sparse coding optimization algorithms have been proposed [22], solving sparse representation is still computational challenge, especially in large dataset. For that reason, the idea of using linear projection to predict the coding coefficients is very attractive to explore. To this end, a novel method to fast compute the coding coefficients which only involve matrix-vector multiplication was proposed in this study. The goal of the algorithm was to make the prediction of the coding coefficients as close as the optimal set of coefficients acquired from the objective function of sparse representation constraints. The main contribution of this study consists of the following aspects.

- A new sparse representation constraint to create a new high-level feature representation based on discriminative dictionary learning is introduced.
- An implementation of linear projection model to predict a set of coding coefficients is presented.

The rest of the paper is organized as follows. Section 2 describes the related works of sparse representation for supervised dictionary learning. An approach to build a new high-level feature representation and a model to predict a set of sparse coding coefficients are proposed in section 3. The detail of the classifier to assign the class label to the test data is introduced in section 4. In section 5, experimental results and discussion are provided to justify the effectiveness of the proposed method. Finally, the conclusion is drawn in section 6.

2. Supervised dictionary learning for classification

Most existing sparse representation approaches aim to sparsely represent the original data using the dictionary without considering the label information. However, recent research [17] showed that better classification performance can be obtained by combining the class label to the objective function. Wright *et al.* [50] proposed an algorithm called sparse representation-based classification (SRC) for robust face recognition. In SRC, a predefined dictionary acquired from the entire training data was used to determine the coding coefficients of the query data and then they were assigned to a given class based on minimum reconstruction error. The procedures of SRC can be summarized as follows:

Step 1. Determine the coding coefficient $\hat{a}$ of a test data x_{test} over dictionary $D = [D_1, D_2, \ldots, D_C]$ which is obtained from the entire set of training data, where $D_i = X_i$ is a subset of the training data from the i^{th} class.

$$\hat{a} = \arg\min_a \|x_{test} - D \cdot a\|_2^2 + \lambda \|a\|_1 \tag{2}$$

Step 2. Assign x_{test} to a class label c using the following criteria:

$$c = \arg\min_i \|x_{test} - D_i \cdot \hat{a}\|_2^2, \quad i = 1, \ldots, C \tag{3}$$

Based on the Fisher discrimination criterion, Yang *et al.* [55] introduced the Fisher discrimination dictionary learning approach so that a structure dictionary was able to determine coding coefficients useful for pattern recognition. In this method, a dictionary $D = [D_1, \ldots, D_C]$ was composed of a set of the class-specific subdictionary from each category (D_i). Similarly, the input dataset $X = [X_1, \ldots, X_C]$ was defined as the composition of the input data subset where X_i was the subset of data from the i^{th} class. The following was the formulation of the objective function for the dictionary learning:

$$< D, A >= \arg\min_{D,A} \mathcal{R}(D, A, X) + \lambda_1 \|A\|_1 + \lambda_2 f(A) \tag{4}$$

where $\mathcal{R}(D, A, X)$ and $f(A)$ are called the discriminative fidelity term and the discrimination coefficient term, respectively. The discriminative fidelity term was designed to reserve the good representation between the coding coefficients a_i and the subdictionary D_i but not with $D_j, i \neq j$. Consequently, the function of the discriminative fidelity term in (4) was defined by:

$$\begin{aligned} \mathcal{R}(D, A, X_i) &= \|X_i - DA_i\|_F^2 + \|X_i - D_i A_i^i\|_F^2 \\ &+ \sum_{i \neq j} \|D_j A_i^j\|_F^2 \end{aligned} \tag{5}$$

where $A_i = [A_i^1, \ldots, A_i^c, \ldots, A_i^C]$ is the coding coefficient to represent the subset data X_i and A_i^c is the coding coefficient of X_i associated with the subdictionary D_c. With regard to the discrimination coefficient term, the Fisher discrimination criterion [8] was employed to minimize the within-class scatter of A, denoted by $S_W(A)$, and to maximize the between-class scatter of A, denoted by $S_B(A)$. To this end, $f(A)$ can be denoted by:

$$f(A) = \mathbf{tr}(S_W(A)) - \mathbf{tr}(S_B(A)) + \eta \|A\|_F^2 \tag{6}$$

Since $f(A) = \mathbf{tr}(S_W(A)) - \mathbf{tr}(S_B(A))$ is non-convex and unstable, adding the $\|A\|_F^2$ term in (6) can help to solve the problem [55].

Although the SRC algorithm yielded an impressive performance on face recognition, using all training data as a dictionary was computationally expensive to solve for the coding coefficients with large datasets. Consequently, recent sparse representation research groups [29],[31],[17],[41],[57] introduced an additional discriminative constraint to the objective function as shown in (7) to achieve better classification performance.

$$\begin{aligned} < D, W, A > &= \arg\min_{D,W,A} \|X - DA\|_2^2 + \sum_{i=1}^{m} \mathcal{L}\{h_i, f(x_i : W)\} \\ &+ \lambda \|A\|_1 + \alpha \|W\|_1 \end{aligned} \tag{7}$$

where $\mathcal{L}$ is a classification loss penalty function. h_i is a class label of x_i and W is a model parameters. In this case, an approach to jointly optimize the dictionary learning and the classification predictor was proposed by Zhang *et al.* [58] and it was called the discriminative K-SVD (D-KSVD). Their method combined the classification error term into the original objective function of the spare coding. Not only was the class label incorporated with the reconstruction constraint but also a linear predictor was optimized during the dictionary learning process based on the K-SVD algorithm. The problem of dictionary learning with the combination between reconstruction and discriminative constraints can be formulated by [58]:

$$\begin{aligned} < D, W, A > &= \arg\min_{D,W,A} \|X - DA\|_2^2 + \lambda_1 \|H - WA\|_2^2 \\ &+ \lambda_2 \|A\|_1 + \lambda_3 \|W\|_1 \end{aligned} \tag{8}$$

where the term $\|H - WA\|_2^2$ represents the classification error. $H = [h_1, \ldots, h_m] \in \mathbb{R}^{C \times m}$ is the class label of the entire input data X, i.e., $h_i = [0, \ldots, 0, 1, 0, \ldots, 0]^t \in \mathbb{R}^C$ in which the nonzero position denotes the class label of data x_i. W represents the parameter of the classification model. λ_1, λ_2 and λ_3 are the parameters to control the contribution of the individual terms. The benefit of this approach is its ability to simultaneously learn the dictionary and the classification model. After obtaining D, A, and W, the classification of the query data (x_{test}) can be performed by $f(x_{test}; W) = W \cdot a_{test} \in \mathbb{R}^C$ and then assigned the class label to the largest value among all elements of $f(x_{test}; W)$.

In addition to using classification error in the objective function, Jiang *et al.* [17] proposed a method named the label consistent KSVD to learn a discriminative dictionary for sparse coding by introducing a label consistent constraint in the dictionary learning process. The new objective function was defined by:

$$\begin{aligned} < D, T, W, A > &= \arg \min_{D,T,W,A} \|X - DA\|_2^2 + \lambda_1 \|Q - TA\|_2^2 \\ &+ \lambda_2 \|H - WA\|_2^2 + \lambda_3 \|A\|_1 \end{aligned} \tag{9}$$

where H and W are the same notations as in the D-KSVD algorithm. An additional term $\|Q - TA\|_2^2$ called the discriminative sparse-code error was used to enforce consistency of the coding coefficients from the same class; in other words, it imposed coding coefficients from the same class to have similar coefficients. In this respect, $Q = [q_1, \ldots, q_m] \in \mathbb{R}^{K \times m}$ was used to control the label consistency and was defined as $q_i = [0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0]^T \in \mathbb{R}^K$ where K is the dictionary size. The nonzero elements in q_i indicated the share class label between the input data x_i and the corresponding subdictionary of the x_i. Although, the classification mechanism of the test data using Fisher discrimination dictionary learning, discriminative K-SVD and label consistent K-SVD was fast to predict a class label, the bottleneck of these methods is the calculation of the coding coefficients $a_{test} = \arg\min_a \|x_{test} - Da\|_2^2 + \lambda \|a\|_1$ of a query data.

3. Predictive high-level feature representation

The discriminative dictionary learning methods described previously aim to learn a sparse representation of the training data by integrating a class label in consideration with an ℓ_p-norm sparse regularization; therefore, making the calculation of sparse coding coefficients on the test data is inefficient for real-time applications due to nonlinear optimization. In this work, a new process is proposed to create not only discriminative coding coefficients but also an effective method to compute the coding coefficients with low computational effort. In particular, it would become more efficient if the coding coefficients can be simply computed by the matrix-vector product. Then, the predicted coding coefficients are subsequently used as a high-level feature representation to train a classifier for predicting a class label of the test data.

3.1. Discriminative dictionary learning

Define a set of n-dimensional training data $X = [X_1, \ldots, X_C] \in \mathbb{R}^{n \times m}$ from C classes, where $X_i \in \mathbb{R}^{n \times m_i}$ is a subset of training data in the i^{th} class with the number of data m_i samples. Note that $m = \sum_{i=1}^{C} m_i$. To obtain discriminative sparse coding coefficients $A \in \mathbb{R}^{K \times m}$ with the learned dictionary $D \in \mathbb{R}^{n \times K}$, the objective function for dictionary learning process is formulated by:

$$\begin{aligned} < D, T, W, A > &= \arg \min_{D,T,W,A} \|X - DA\|_2^2 + \lambda_1 \|M - TA\|_2^2 \\ &+ \lambda_2 \|H - WA\|_2^2 \\ \text{s.t.} \quad &\|a_i\|_0 \leq \gamma, \quad \forall\, i \end{aligned} \tag{10}$$

The objective function in (10) is modified from the label-consistent K-SVD algorithm [17]. The first term $\|X - DA\|_2^2$ represents a reconstruction error as for [17]. The second term $\|M - TA\|_2^2$ is used to control the consistency of the coding coefficients drawn from the same class where $M = [M_1, \ldots, M_m] \in \mathbb{R}^{K \times m}$ and M_i

is composed of the centroid of the class-specific dictionary based on the subset training data X_i. A well-known K-SVD algorithm is applied to each subset of the training data X_i to determine M_i by computing the mean ($\tilde{d}_i$) of the learned dictionary D_i where D_i is a learned dictionary of a subset training data of the i^{th} class. For example, suppose $X = [x_1, x_2, x_3, x_4, x_5]$ is a set of training data and the corresponding class label is $Y = [1^{st}, 3^{rd}, 4^{th}, 2^{nd}, 1^{st}]$, then the M can be created by:

$$M = \begin{bmatrix} \tilde{d}_1 & 0 & 0 & 0 & \tilde{d}_1 \\ 0 & 0 & 0 & \tilde{d}_2 & 0 \\ 0 & \tilde{d}_3 & 0 & 0 & 0 \\ 0 & 0 & \tilde{d}_4 & 0 & 0 \end{bmatrix}$$

In this case, the mean vector of each subset dictionary is used as a class dictionary representative to encourage coding consistency from the same class. The last term $\|H - WA\|_2^2$ is designed to promote the discriminative representation of the coding coefficients. W is a mapping function of the coding coefficients to class label. The matrix $H = [h_1, \ldots, h_m] \in \mathbb{R}^{C \times m}$ represents the corresponding class label of each training data. For example, if a training data x_i belongs to class 2^{nd}, h_i can be defined as a vector of $[0, 1, 0, \ldots, 0]^T$ in which the nonzero position points out the class label of x_i. To this end, the learned dictionary in this process can build the coding coefficients not only for sparse representation and very similar coding with the same class but also for discriminative high-level feature representation.

3.2. High-level feature prediction

As mentioned, a computational burden of sparse representation results from the ℓ_0/ℓ_1-norm sparsity regularization. It is very useful to investigate whether the coding coefficient can be computed without the cost of ℓ_0/ℓ_1-norm sparsity constraint. In particular, if the coding coefficients can be determined by linear projection rather than sparse coding optimization, computation of the coding coefficients will be more efficient. In fact, the method to predict the coding coefficients is not new. In [18], a nonlinear feedforward predictor with specific architecture and encoder was proposed to approximate the coding coefficients. Lee *et al.* [22] introduced an efficient sparse coding algorithm based on iteratively solving two convex optimization problems to increase the speed of coefficients' computation. There are two aspects that differ from our proposed method. First, our predicted coding coefficients are approximated by a simple linear projection function. Second, the proposed dictionary learning integrates class information to promote the coding consistency and discriminative representation for classification problems.

Based on the proposed objective function of the dictionary learning in (10), the results of the optimization problem consist of D, T, W, and A in which the coding coefficient A is subsequently used to train a classifier. Thus, the ultimate goal is to formulate the mapping function for fast approximation of the coding coefficients: $f(X; P) = PX \approx A$. Recall from the sparse representation problem, given a dictionary D, the coding coefficient of input x_i can be computed by:

$$\begin{aligned} a_i &= \arg\min_a \|x_i - Da\|_2^2 \\ &\quad s.t. \quad \|a\|_0 \leq \gamma, \quad or \\ a_i &= \arg\min_a \|x_i - Da\|_2^2 + \gamma\|a\|_1 \end{aligned} \tag{11}$$

where γ is a nonzero sparsity constraint. In this case, the P matrix can be solved by:

$$P = \arg\min_P \|A - PX\|_2^2 \tag{12}$$

To find the solution, the problem in (12) can be revised so that the $P = [p_1, \ldots, p_m] \in \mathbb{R}^{K \times m}$ matrix can be solved by standard pursuit algorithms [45] as follows.

$$P^T = \arg\min_{P^T} \|A^T - X^T P^T\|_2^2 \quad s.t. \quad \|p_i^T\|_1 \leq \beta \tag{13}$$

where β is also a nonzero sparsity constraint. To this end, the computation of the coding coefficients or a set of high-level features simply involves matrix-vector multiplication resulting in rapid calculation without the burden of ℓ_0 or ℓ_1-norm regularization.

3.3. Method of solving parameters

To simultaneously solve all parameters in the objective function shown in (10), the K-SVD algorithm can be used with a similar procedure introduced in the discriminative K-SVD [58] and the label-consistent K-SVD [17] algorithms. In the K-SVD algorithm, the sparse representation problem is originally formulated by:

$$\begin{aligned} < D, A > &= \arg\min_{D,A} ||X - DA||_F^2 \\ &\text{s.t.} \quad \|a_i\|_0 \leq \gamma \end{aligned} \tag{14}$$

Thus, the objective function must conform with (14) and a new formulation of the problem can be written as:

$$\begin{aligned} < \hat{D}, A > &= \arg\min_{\hat{D},A} ||\hat{X} - \hat{D}A||_2^2 \\ &\text{s.t.} \quad \|a_i\|_0 \leq \gamma \end{aligned} \tag{15}$$

where

$$\hat{X} = \begin{pmatrix} X \\ \sqrt{\lambda_1}M \\ \sqrt{\lambda_2}H \end{pmatrix} \quad \text{and} \quad \hat{D} = \begin{pmatrix} D \\ \sqrt{\lambda_1}T \\ \sqrt{\lambda_2}W \end{pmatrix} \tag{16}$$

Equation (15) is now exactly equivalent to the K-SVD formulation and the K-SVD algorithm can be used to solve for $\hat{D}$ and A. Note that the column vectors in $\hat{D}$ need to be normalized in order to prevent them becoming arbitrarily small or large. After determining A, the orthogonal matching pursuit algorithm (OMP) [7],[37],[45] is applied to (13) to solve for P. Upon completion of finding the projection matrix P, the coding coefficients A of all the training data can be determined by the matrix-vector product of P and X. Finally, these coding coefficients will be used to train the prediction model for classification. The proposed method can be summarized as follows.

Algorithm - Predictive sparse coding K-SVD
Input: $X, M, H, K, \lambda_1, \lambda_2, \gamma, \beta$
Output: A, P
- Step 1: Compute $D^{(0)}, T^{(0)}, W^{(0)}$:
  - $D^{(0)}$ is created based on combining class-specific dictionary from each training class data using standard K-SVD algorithm.
  - The initial coding coefficients $A^{(0)}$ is calculated by (11).
  - The initial values of $T^{(0)}$ and $W^{(0)}$ are computed by (17) and (18), respectively.
- Step 2: Formulate $\hat{X}$ and $\hat{D}$ according to (16).
- Step 3: Solve $\hat{D}$ from (15) by standard K-SVD algorithm to acquire D, T, and W.
- Step 4: Determine P from (13) using OMP algorithm.
- Step 5: Calculate an approximated set of coding coefficients (A) from the product of P and X. These features are considered a high-level feature representation and are used to train the classifier.

Note that the initializations of $T^{(0)}$ and $W^{(0)}$ are obtained from the ridge regression model [14] based on the least squared problem with an ℓ_2-norm regularization. In particular, the ridge regression can be used to solve $T^{(0)}$ for the following problem:

$$
\begin{aligned}
T &= \arg\min_{T} \|M - TA\|_2^2 + \lambda \|T\|_2^2 \\
\text{which yields} \quad T &= MA^T(AA^T + \lambda_T I)^{-1} \qquad (17)
\end{aligned}
$$

Similarly, $W^{(0)}$ can be determined by:

$$
W = HA^T(AA^T + \lambda_W I)^{-1} \qquad (18)
$$

where λ_T and λ_W are the regularization parameters to control the smoothness of the regression models.

4. Classification method

With the classification, a simple supervised learning model is employed based on discriminative learning algorithm called softmax regression [34], or equivalently a generalized logistic regression, to make the class prediction in this study. Conceptually, the softmax regression is designed to estimate the posterior probability of the class label on the given input data $p(y_i = c|x_i)$ where $c = 1, \ldots, C$ and the probabilities for each class can be computed by:

$$
h_\theta(x_i) = \begin{bmatrix} p(y_i = 1|x_i; \theta_1) \\ p(y_i = 2|x_i; \theta_2) \\ \vdots \\ p(y_i = C|x_i; \theta_C) \end{bmatrix} = \frac{1}{\sum_{j=1}^{C} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ \vdots \\ e^{\theta_C^T x_i} \end{bmatrix}
$$

where $\theta = [\theta_1, \ldots, \theta_C]^T$ is a set of parameters of the softmax model. These parameters can be obtained by optimizing the following cost function based on the gradient decent method.

$$
\begin{aligned}
J(\theta) &= -\frac{1}{m} \left[\sum_{i=1}^{m} \sum_{j=1}^{C} \mathbb{1}\{y_i = j\} \log \frac{e^{\theta_j^T x_i}}{\sum_{l=1}^{C} e^{\theta_l^T x_i}} \right] \\
&+ \frac{\lambda}{2} \sum_{i=1}^{C} \sum_{j=1}^{n} \theta_{ij}^2 \qquad (19)
\end{aligned}
$$

where λ is a regularization parameter to control model complexity and $\mathbb{1}\{\cdot\}$ is called an indicator function which gives the value 1 if $\mathbb{1}\{true\}$ and 0 otherwise. For classification of the test data, the softmax regression will predict to class c^{th} which yields the highest probability of $p(y = c|x_{test})$.

5. Experiments

In this section, the classification performance of the proposed method was evaluated using various visual classification datasets: Extended YaleB [13], AR face [33], Caltech101 [10], and 15 scene category [20]. These benchmark datasets have been widely used in previous work [17],[55],[58] to demonstrate the performance of dictionary learning. Furthermore, average processing time to predict a single test data was investigated to compare the methods. All experiments were performed on a laptop cpmputer with 1.8GHz Intel core i5 CPU and 8 GB in RAM.

In the proposed method, there are five parameters to specify: K, λ_1, λ_2, γ, and β that were determined by a cross validation technique on the training data. The classification results of the proposed approach, which named the predictive sparse coding K-SVD algorithm (PSC-KSVD), were analyzed to compare performance with the following methods: sparse representation-based classification (SRC), discriminative K-SVD

Figure 1: Sample images from Extended YaleB dataset [13]

Table 1: Classification results and computation time on the Extended YaleB dataset

Method	Accuracy (%)	Testing time (ms)
K-SVD [1]	93.1	-
SRC [50]	97.2	-
D-KSVD [58]	94.1	-
LLC [47]	90.7	-
FDDL [55]	91.9	-
LC-KSVD [17]	95.0	0.408
PSC-KSVD	94.5	0.09

(DKSVD), label-consistent K-SVD (LC-KSVD), locality constrained linear coding (LLC), and Fisher discrimination dictionary learning (FDDL). Based on the results of [17], LC-KSVD was the fastest in terms of the average time to predict a single test data compared to other approaches mentioned previously. Consequently, it was legitimate to only compare the computation time of the proposed method with that of the LC-KSVD algorithm.

5.1. Extended YaleB dataset

A set of face images extracted from extended YaleB database consists of 2,414 frontal face images of 38 persons under various laboratory-controlled illumination conditions as shown in Fig.1. There are 64 images of each person. Half of the images per category are randomly chosen as a training data and the rest are used for testing. The original images were normalized and cropped to 192×168 pixels. To reduce the dimension of the images, a randomly generated matrix [40] called randomface was used to transform image array $\mathbb{R}^{192\times168}$ to a vector of $\mathbb{R}^{504}$. The learned dictionary of our method contained $K = 950$ atoms or 25 atoms for each category. The parameters of our method were set to $\lambda_1 = 4$, $\lambda_2 = 5$, $\gamma = 60$, and $\beta = 100$. The experimental results are shown in Table 1. The best classification accuracy of 97.2% was achieved by the SRC. Although, our method obtained an accuracy of 94.5% which was slightly lower than the LC-KSVD method, the average time of predicting one test data was approximately 4.5 times faster than that of the LC-KSVD algorithm.

The effects of parameter selection of λ_1 and λ_2 on the classification accuracy is illustrated in Table 2. Each bar represents the classification accuracy of the change of λ_1 and λ_2 varied between 1 and 6 while the other parameters are fixed. The best accuracy was achieved at $\lambda_1 = 4$ and $\lambda_2 = 5$. In this case, the variance of the classification accuracy is 2×10^{-5} so the proposed method is robust to the adjustment of λ_1 and λ_2.

5.2. AR face dataset

Another challenging face image database is the AR face dataset [33] containing over 4,000 color images of 126 persons taken in sessions two weeks apart. The challenge of the AR face dataset involves different variations in illumination conditions, facial expressions, and facial disguises with sunglass and scarf occlusion, as illustrated in Fig.2. In the experiments, a subset of 2,600 images of 50 female and 50 male were employed and from these 20 images were randomly chosen for training and the remaining 6 images for testing. Similar

Table 2: Effect of parameters λ_1 and λ_2 on classification accuracy on the Extended YaleB database

		λ_1					
		1	2	3	4	5	6
λ_2	1	0.933	0.934	0.926	0.931	0.928	0.931
	2	0.935	0.939	0.937	0.935	0.937	0.932
	3	0.932	0.927	0.936	0.939	0.929	0.934
	4	0.939	0.927	0.933	0.936	0.937	0.931
	5	0.938	0.940	0.927	**0.945**	0.930	0.937
	6	0.932	0.933	0.938	0.930	0.940	0.922

Figure 2: Sample images from AR face dataset [33]

to the Extended YaleB dataset, the randomly generated matrix was applied to reduce the dimensionality from the original image size $\mathbb{R}^{165\times120}$ to a 540-dimensional vector. The setting of the model parameters for this dataset was $\lambda_1 = 3$, $\lambda_2 = 2$, $\gamma = 60$, and $\beta = 100$. The learned dictionary has $K = 500$ atoms with 5 atoms per person. The experimental results are reported in Table 3. The proposed method was superior to all of the other methods with a classification accuracy of 98%. Also, the computation time of our method was approximately 6 times faster than the result of LC-KSVD algorithm.

In addition, the impact of increasing dictionary size on the classification accuracy and the computation time for the test data were investigated and the results are shown in Fig.3. It can be seen that the accuracy performance was improved when using a large dictionary size, but this benefit was reduced with a dictionary size greater than 600 atoms. As expected, the computation time was proportional to the dictionary size; that is, the larger the dictionary size, the higher the computation time to predict the test data.

5.3. Caltech101 dataset

One of the well-known datasets in object recognition problem is Caltech101 [10] which contains 9,144 images from 102 classes (101 distinct classes and a background class). The number of images in each category varies from 31 to 800 and the size of each image is roughly 300×200 pixels. A sample of images

Table 3: Classification results and computation time on the AR face dataset

Method	Accuracy (%)	Testing time (ms)
K-SVD [1]	86.5	-
SRC [50]	97.5	-
D-KSVD [58]	88.8	-
LLC [47]	88.7	-
FDDL [55]	92.0	-
LC-KSVD [17]	93.7	0.344
PSC-KSVD	98.0	0.055

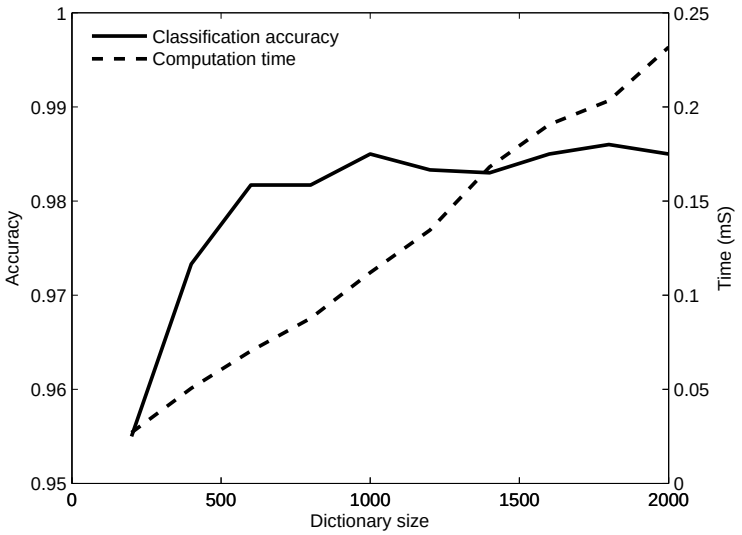


Figure 3: Effect of dictionary size on classification performance and computational speed of the AR face dataset

Table 4: Classification results and computation time on the Caltech101 dataset

Method	Accuracy (%)	Testing time (ms)
K-SVD [1]	73.2	-
SRC [50]	70.7	-
D-KSVD [58]	73.0	-
LC-KSVD [17]	73.6	2.392
PSC-KSVD	73.9	0.198

from this dataset are shown in Fig.4. According to the experimental setting in [17],[20], 30 images per category were randomly chosen for training and the rest were used for testing. Based on [20], scale-invariant feature transform (or SIFT) descriptors were computed for three different grid sizes $(1 \times 1, 2 \times 2, 4 \times 4)$ and then the spatial pyramid matching (SPM) algorithm was applied for feature extraction. Finally, the principal component analysis was adopted to reduce the dimension of the samples to 3,000. Table 4 presents the experimental results of classification accuracy and the average time of each testing run. The proposed method achieved the highest performance of 73.9% accuracy with parameter settings of $\lambda_1 = 10$, $\lambda_2 = 5$, $\gamma = 60$, and $\beta = 50$. A dictionary size of 3,060 atoms was chosen which was the same as in [17] corresponding to 30 atoms per category. Moreover, the improvement of computation time of the proposed method was approximately 12 times faster than that of the LC-KSVD method.

Table 5: Effect of parameters γ and β on classification performance of the Caltech101 dataset

		γ				
		20	40	60	80	100
β	50	0.738	0.738	**0.739**	0.738	0.738
	100	0.734	0.734	0.734	0..734	0.734
	150	0.735	0.735	0.735	0.735	0.735
	200	0.738	0.738	0.738	0.738	0.737

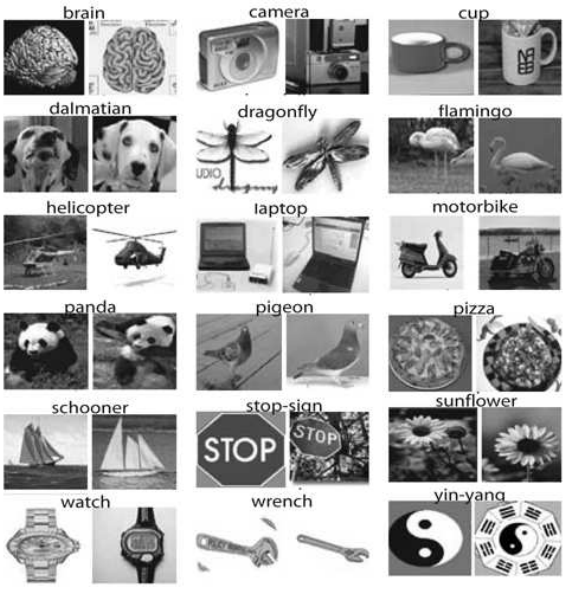


Figure 4: Sample images from Caltech101 dataset [10]

Table 6: Classification results and computation time on the 15 scene dataset

Method	Accuracy (%)	Testing time (ms)
K-SVD [1]	86.7	-
SRC [50]	91.8	-
D-KSVD [58]	89.1	-
LLC [47]	89.2	-
LC-KSVD [17]	94.2	0.396
PSC-KSVD	95.8	0.05

To investigate the effect of the selection of γ and β parameters on the classification accuracy, different values of γ and β were evaluated while other parameters remained fixed. Table 5 demonstrates the classification accuracy with changes in γ and β. The variance of the classification accuracy was extremely small (approximately 3.81×10^{-6}), which suggested the performance was not sensitive to the variation of γ and β parameters.

5.4. Fifteen scene dataset

Scene recognition is an interesting and challenging task. In the experiments, the scene dataset contained 15 natural scene categories collected by Fei *et al.* [11]. Each class consists of 200 to 400 images with the average size of 300×250 pixels. Similar to the other datasets used in this study, a set of training data was randomly chosen with 100 images per category and the rest was used for testing. Some sample images from this dataset are illustrated in Fig.5. The dictionary comprised 450 atoms with 30 atoms for each class. The parameters of the proposed method were set to $\lambda_1 = 6$, $\lambda_2 = 2$, $\gamma = 80$, and $\beta = 100$. The results of classification and computation time of different methods are summarized in Table 6 and Fig.6 displays a confusion matrix of the proposed method. The best result achieved by the proposed method was 95.8% accuracy and it was approximately 8 times faster computation time than the LC-KSVD method.

The previous results of the computation time shown in Fig.3 indicates that the computational speed on the test data was dependent on the dictionary size. To investigate other parameters that might affect the computational speed, a selection of γ and β parameters was evaluated and the results from using different

Figure 5: Sample images from 15 scene dataset [11]

	suburb	coast	forest	highway	insidecity	mountain	opencountry	street	tallbuilding	office	bedroom	industrial	kitchen	livingroom	store
suburb	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
coast	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
forest	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
highway	0.00	0.00	0.00	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
insidecity	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mountain	0.00	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
opencountry	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
street	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
tallbuilding	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.98	0.01	0.00	0.00	0.00	0.00	0.00
office	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.02
bedroom	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.97	0.00	0.00	0.00	0.02
industrial	0.00	0.00	0.01	0.01	0.00	0.02	0.01	0.00	0.00	0.02	0.01	0.95	0.03	0.01	0.02
kitchen	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.95	0.00	0.00
livingroom	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.98	0.01
store	0.01	0.01	0.01	0.02	0.01	0.03	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.01	0.91

Figure 6: Confusion matrix of 15 scenes data with dictionary size = 450

values of γ and β are depicted in Fig.7. As expected, the choice of the β parameter and not γ had a linear relationship with the computation time since β based on (13) is employed to control the sparsity of the P matrix used for predicting the coding coefficients ($A = PX$). Consequently, a small value of β leads to a reduced computation time cpmpared with a large value because the small value of β enforces most components in the P matrix to be zeros. In summary, the selection of parameters, λ_1, λ_2, γ, and K is designed to control the new high-level feature representation for classification while the computation time depends not only on the dictionary size but also on the β value.

6. Conclusion

The importance of supervised dictionary learning based on sparse representation has been promoted for its strong capability in classification. However, solving coding coefficients typically requires the use of ℓ_0/ℓ_1-norm regularization which leads to expensive computation. In this study, a novel method called predictive sparse coding K-SVD (PSC-KSVD) was proposed to efficiently compute the coding coefficients simply using

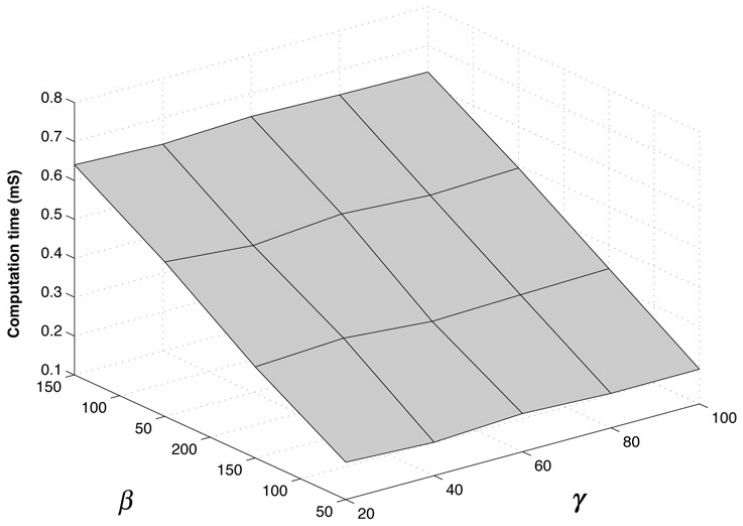


Figure 7: Effect of parameters γ and β on computational speed on 15 scene dataset

the matrix-vector product. In addition, the proposed approach is also able to create discriminative sparse coding coefficients based on class consistency and discriminative constraints. The solution of the proposed dictionary learning follows the procedure derived from the K-SVD algorithm while the OMP algorithm is used to determine the predictive matrix (P) for computing the coding coefficients.

Performance was tested on several visual datasets involving face, object, and scene recognitions. In addition, the effect of the parameter selections was analyzed. The experimental results indicated that the proposed method provides promising classification accuracy compared with other previous dictionary learning techniques for classification and it also clearly outperforms in terms of computation time for the test data. As a result, it can be considered as a suitable method to apply in real-time classification problems.

Acknowledgment

The authors would like to thank the Thailand Research Funds (TRF) under grant research no. TRG5680074 and Kasetsart University Research and Development Institute (KURDI) for supporting this research.

References

[1] M. Aharon, M. Elad, and A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Processing, vol. 54(11), pp. 4311-4322, 2006.

[2] F. Bergeaud and S. Mallat, Matching pursuit of images, In Proc. Int'l Conf. Image Processing, vol. 1, pp. 53-56, 1995.

[3] D. Bradley and J. Bagnell, Differential Sparse Coding, Proc. Conf. Neural Information Processing Systems, pp. 113-120, 2008.

[4] O. Bryt and M. Elad, Compression of facial images using the K-SVD algorithm. Journal of Visual Communication and Image Representation, vol. 19(4), pp. 270-282, 2008.

[5] E. Candes, Compressive sampling. Proc. Int. Congress of Mathematics, vol. 3, pp. 1433-1452, 2006.

[6] A. Castrodad and G. Sapiro, Sparse modeling of human actions from motion imagery, Int'l Journal of Computer Vision, vol. 100, pp. 1-15, 2012.

[7] S. Chen, S. A. Billings, and W. Luo, Orthogonal least squares methods and their application to non-linear system identification, Int. J. Contr., vol. 50(5), pp. 1873-1896, 1989.

[8] R. Duda, P. Hart, and D. Stork, Pattern classification (2nd ed.), Wiley-Interscience, 2000.

[9] M. Elad and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Processing, vol. 15(12), pp. 3736-3745, 2006.

[10] L. Fei Fei, R. Fergus, and P. Perona, Learning Generative Visual Models from Few Training Samples: An Incremental Bayesian Approach Tested on 101 Object Categories, Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Generative Model Based Vision, pp. 746-751, 2004.

[11] L. Fei-Fei and P. Perona, A Bayesian hierarchical model for learning natural scene categories, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 524-531, 2005.
[12] B. Fulkerson, A. Vedaldi, and S. Soatto, Localizing Objects with Smart Dictionaries, Proc. European Conf. Computer Vision, pp. 179-192, 2008.
[13] A. Georghiades, P. Belhumeur, and D. Kriegman, From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23(6), pp. 643-660, 2001.
[14] G. Golub, P. Hansen, and D. O'leary, Tikhonov Regularization and Total Least Squares, SIM J. Matrix Analysis Applications, vol. 21(1), pp. 185-194, 1999.
[15] K. Huang and S. Aviyente, Sparse representation for signal classification. Proc. Conf. Neural Information Processing Systems, pp. 609-616, 2006.
[16] Z. Jiang, G. Zhang, and L. Davis, Submodular Dictionary Learning for Sparse Coding, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 3418-3425, 2012.
[17] Z. Jiang, Z. Lin, and L. Davis, Label consistent K-SVD: Learning a discriminative dictionary for recognition. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 35(11), pp. 2651-2664, 2013.
[18] G. Karol and Y. LeCun, Learning Fast Approximations of Sparse Coding, Proc. Int'l Conf. Machine Learning, pp. 399-406, 2010.
[19] S. Kong and D.H., Wang, A dictionary learning approach for classification: Separating the particularity and the commonality, European conference on Computer Vision, pp. 186-199, 2012.
[20] S. Lazebnik, C. Schmid, and J. Ponce, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Cate- gories, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2169-2178, 2007.
[21] S. Lazebnik and M. Raginsky, Supervised Learning of Quantizer Codebooks by Information Loss Minimization, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31(7), pp. 1294- 1309, 2009.
[22] H. Lee, A. Battle, R. Raina, and A.Y. Ng, Efficient Sparse Coding Algorithms, Proc. Conf. Neural Information Processing Systems, pp. 801-808, 2006.
[23] Y. Li, A. Cichocki, and S. Amari, Analysis of sparse representation and blind source separation, Neural Computation, vol. 16(6), pp. 1193-1234, 2004.
[24] J. Liu and M. Shah, Learning Human Actions via Information Maximization, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
[25] J. Mairal, M. Elad, and G. Sapiro, Sparse representation for color image restoration. IEEE Trans. Image Processing, vol. 17(1), pp. 53-69, 2008.
[26] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, Discriminative Learned Dictionaries for Local Image Analysis, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
[27] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zissserman. Learning discriminative dictionaries for local image analysis, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
[28] J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce, Discriminative Sparse Image Models for Class-Specific Edge Detection and Image Interpretation, Proc. European Conf. Computer Vision, pp. 43-56, 2008.
[29] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, A. Supervised dictionary learning. Proc. Conf. Neural Information Processing Systems, pp. 1033-1040, 2009.
[30] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, Online Learning for Matrix Factorization and Sparse Coding, J. Machine Learning Research, vol. 11, pp. 19-60, 2010.
[31] J. Mairal, F. Bach, and J. Ponce, Task-Driven Dictionary Learning. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 34(4), pp. 791-804, 2012.
[32] S. Mallat and Z. Zhang, Matching Pursuits with Time-Frequency Dictionaries, IEEE Trans. Signal Processing, pp. 3397-3415, 1993.
[33] A. Martinez, R. Benavente, The AR face database. CVC Technical Report (1998)
[34] A. Ng, J. Ngiam, C. Foo, Y. Mai, C. Suen, Softmax regression, http://ufldl.stanford.edu/wiki/index.php/Softmax_Regression, 2013.
[35] B.A. Olshausen, and D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, vol. 381, pp. 607-609, 1996.
[36] B.A. Olshausen, and D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by v1? Vision Research, vol. 37(23), pp. 3311-3325, 1997.
[37] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, Conf. Rec. 27^{th} Asilomar Conf. Signals, Systems, and Computers, vol. 1. pp. 40-44, 1993.
[38] F. Perronnin, Universal and Adapted Vocabularies for Generic Visual Categorization, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30(7), pp. 1243-1256, 2008.
[39] Q. Qiu, Z. Jiang, and R. Chellappa, Sparse Dictionary-Based Representation and Recognition of Action Attributes, Proc. IEEE Int'l Conf. Computer Vision, pp. 707-714, 2011.
[40] R. Raina, A. Battle, H. Lee, B. Packer, A. Ng, Self-taught learning: Transfer learning from unlabeled data, Proc. Int'l Conf. Machine Learning, pp. 759-766, 2007.
[41] I. Ramirez, P. Sprechmann, and G. Sapiro, Classification and clustering via dictionary learning with structured incoherence and shared features. Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 3501-3508, 2010.
[42] A. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, Efficient Learning of Sparse Representations with an Energy-Based Model, Proc. Conf. Neural Information Processing Systems, pp. 1137-1144, 2007.
[43] R. Sivalingam, D. Boley, V. Morellas, and N. Papanikolopoulos, Positive Definite Dictionary Learning for Region Covari-

ances, Proc. IEEE Int'l Conf. Computer Vision, pp. 1013 - 1019, 2011.
[44] J. Starck, M. Elad, and D. Donoho, Image decomposition via the combination of sparse representation and a variational approach, IEEE Trans. on Image Processing, vol.14(10), pp. 1570-1582, 2005.
[45] J. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inf. Theory, vol. 50, pp. 2231-2242, 2004.
[46] A. Wagner, J. Wright, A. Ganesh, Z.H. Zhou, and Y. Ma, Towards a Practical Face Recognition System: Robust Registration and Illumination by Sparse Representation. Proc. IEEE conf. Computer Vision and Pattern Recognition, pp. 597-604, 2009.
[47] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, Locality-Constrained Linear Coding for Image Classification, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 3360 - 3367, 2010.
[48] Z. Wang, J. Yang, N. Nasrabadi, T. Huang, A max-margin perspective on sparse representation-based classification, Proc. IEEE Int'l Conf. Computer Vision, pp. 1217 - 1224, 2013.
[49] J. Winn, A. Criminisi, and T. Minka, Object Categorization by Learned Universal Visual Dictionary, Proc. IEEE Int'l Conf. Computer Vision, pp. 1800-1807, 2005.
[50] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.31(2), pp. 210–227, 2009.
[51] J. Yang, J. Wright, T. Huang, and Y. Ma, Image Super-resolution as Sparse Representation of Raw Patches, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
[52] L. Yang, R. Jin, R. Sukthankar, and F. Jurie, Unifying Discriminative Visual Codebook Genearation with Classifier Training for Object Category Recognition, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
[53] J. Yang, K. Yu, Y. Gong, and T. Huang, Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1794-1801, 2009.
[54] M. Yang and L. Zhang, Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary. European conference on Computer Vision, pp. 448-461, 2010.
[55] M. Yang, L. Zhang, X.C. Feng, and D. Zhang, Fisher discrimination dictionary learning for sparse representation, Proc. IEEE Int'l Conf. Computer Vision, pp. 543 - 550, 2011.
[56] M. Yang, L. Zhang, J. Yang, and D. Zhang, Robust sparse coding for face recognition. Proc. IEEE conf. Computer Vision and Pattern Recognition, pp. 625-632, 2011.
[57] M. Yang, L. Zhang, X.C. Feng, and D. Zhang, Sparse representation based Fisher discrimination dictionary learning for image classification, Inter. Journal on Computer vision, vol. 9(3), pp. 209-232, 2014.
[58] Q. Zhang and B.X. Li, Discriminative K-SVD for dictionary learning in face recognition. Proc. IEEE conf. Computer Vision and Pattern Recognition, pp. 2691-2698, 2010.
[59] W. Zhang, A. Surve, X. Fern, and T. Dietterich, Learning non-redundant codebooks for classifying complex objects, Proc. Int'l Conf. Machine Learning, pp. 1241-1248, 2009.
[60] M. Zhou, H. Chen, J. Paisley, L. Ren, L.B. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, Nonparametric Bayesian Dictionary Learning for Analysis of Noisy and Incomplete Images. IEEE Trans. Image Processing, vol. 21, no. 1, pp. 130-144, 2012.
[61] N. Zhou and J. Fan, Learning inter-related visual dictionary for object recognition, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 3490-3497, 2012.