เทคโนโลยีแฮชชิ่งและฐานข้อมูลโฟโต้ดีเอ็นเอในการตรวจพิสูจน์สื่อมัลติมีเดีย

Hashing Technology and PhotoDNA Database in Multimedia Forensics

อัศว์ณุต แสงทองดี[1] กชกร เพิงระนัย[2] และ ภิญโญ มีเปี่ยม[3]

[1]คณะคอมพิวเตอร์ วิศวกรรม และเทคโนโลยีดิจิทัล มหาวิทยาลัยทีซไซด์

[2,3]คณะนิติวิทยาศาสตร์ โรงเรียนนายร้อยตำรวจ

Usanut Sangtongdee[1], Kochchakorn Poengranai[2] and Pinyo Meephiam[3]

[1]School of Computing, Engineering and Digital Technology, Teesside University

[2,3]Faculty of Forensic Science, Royal Police Cadet Academy

Received September 20, 2020 | Revised October 29, 2020 | Accepted November 24, 2020

บทคัดย่อ

สื่อมัลติมีเดียมีความสำคัญในการตรวจพิสูจน์การกระทำความผิดมากยิ่งขึ้น โดยเฉพาะอย่างยิ่งอาชญากรรมล่วงละเมิดต่อเด็กซึ่งมีการรับส่งรูปภาพและสื่อที่มีเนื้อหารุนแรงทางเพศเพิ่มขึ้นทุกวันผ่านผู้ให้บริการสื่อสังคมออนไลน์ ทั้งนี้ เทคโนโลยีแฮชมีบทบาทสำคัญในการสืบค้นและคัดกรองไฟล์ดิจิทัล โดยค่าแฮชแบบไบนารีเป็นผลการคำนวณฟังก์ชันแบบเดิมที่ตรวจสอบความมั่นคงและความครบถ้วนสมบูรณ์ของไฟล์ ส่วนค่าแฮชรูปภาพแบบทนทานเป็นรูปแบบการคำนวณผลลัพธ์ทางคณิตศาสตร์แบบใหม่ซึ่งกำลังได้รับความนิยมในปัจจุบันในตรวจวัดความคล้ายของรูปภาพ โดยตำรวจ หน่วยงานบังคับใช้กฎหมาย องค์กรไม่แสวงหากำไร ผู้ให้บริการอินเทอร์เน็ต และสื่อสังคมออนไลน์ได้ร่วมกันพัฒนาเทคโนโลยีในการตรวจจับและคัดกรองเนื้อหาที่มีความรุนแรงทางเพศต่อเด็กและเยาวชน โดยฐานข้อมูลแฮชที่จัดบัญชีรายการค่าแฮชทั้งแบบทั่วไปและแบบโฟโต้ดีเอ็นเอถูกแบ่งปันไปยังหน่วยงานตำรวจเพื่อใช้สำหรับการตรวจวิเคราะห์ไฟล์สื่อลามกอนาจารที่ปรากฏในพยานหลักฐานอิเล็กทรอนิกส์ อย่างไรก็ตาม การเลือกใช้เครื่องมือตรวจพิสูจน์หลักฐานดิจิทัลควรพิจารณาให้รองรับกับการสืบค้นด้วยเทคโนโลยีแฮชที่ครอบคลุมทั้งแบบไบนารีและแบบทนทาน

คำสำคัญ: โฟโต้ดีเอ็นเอ, การทำซ้ำเสมือน, การแฮชรูปภาพแบบทนทาน, ฟังก์ชันแฮชชิ่งแบบไบนารี

Abstract

Multimedia files are currently vital evidence for cybercrime investigation, particularly online child abuse crimes which have been increasing daily with sexual abuse imagery being transmitted through social media platforms. Hashing technology plays an important role in performing the search and screening of digital files. Binary hashes result from the calculation of traditional functions that verify whether a file is equipped with confidentiality and integrity. Robust image hashing is another function in hashing technology that utilises a new form of mathematical calculation. This function becomes more popular in practice due to its ability to effectively detect image similarity. Police, law enforcement agencies, non-profit organisations, internet service providers, and social media providers have jointly established collaborations through database technology to detect and monitor unusual content related to child sexual abuse material. A hash database that accounts for both binary and PhotoDNA hash values is shared with various police departments to analyse sexual abuse material appeared in electronic evidence. However, in order to select digital forensic tools, it is essential for them to be comprehensively compatible with both binary and robust hash queries.

Keywords: PhotoDNA, Visual Copy, Robust Image Hashing, Binary Hash Function

บทนำ

ด้วยเทคโนโลยีสารสนเทศในปัจจุบันทำให้การรับส่งข้อมูลรูปภาพและวิดีโอเกิดขึ้นตลอดเวลา โดยการอัปโหลดสื่อมัลติมีเดียสามารถเกิดได้ทั้งจากโทรศัพท์มือถือ คอมพิวเตอร์ส่วนตัว และอุปกรณ์อิเล็กทรอนิกส์อื่น ๆ ทั้งนี้ ศูนย์อาชญากรรมทางดิจิทัลของไมโครซอฟท์ได้เคยรายงานไว้เมื่อปี พ.ศ. 2561 ว่าทุกหนึ่งนาทีจะมีไฟล์สื่อลามกอนาจารกระจัดกระจายอยู่ในระบบอินเทอร์เน็ตทั่วโลกไม่น้อยกว่าแปดร้อยภาพ (Microsoft, 2018) และมีไม่น้อยกว่าหนึ่งพันแปดร้อยล้านไฟล์ที่กระจายบนอินเทอร์เน็ตไปทั่วโลก (Nadeem et al., 2019) แต่เมื่อคาดการณ์ถึงปริมาณสื่อมัลติมีเดียเหล่านั้นรวมแล้ว องค์กรตำรวจสากลเปิดเผยว่าปริมาณสื่อลามกเกี่ยวกับเด็กที่ปรากฏในฐานข้อมูลนานาชาติมีไม่น้อยกว่าหนึ่งล้านไฟล์ โดยภายใต้จำนวนนี้มีจำนวนเพียงน้อยกว่าครึ่งหนึ่งที่ระบุตัวตนเหยื่อได้แล้ว (INTERPOL & ECPAT, 2018) โดยทั้งหมดนี้ได้ถูกตรวจสอบไปกว่าครึ่งหนึ่งแล้วว่าเป็นสื่อลามกที่มีเด็กปรากฏอยู่ในภาพ

กฎหมายคุ้มครองเด็กในหลายประเทศจะเอาผิดกับกรณีรูปภาพและวิดีโอเหล่านั้นปรากฏโดยชัดแจ้งว่า มีเด็กหรือเยาวชนร่วมอยู่ด้วยและมีการล่วงละเมิดที่เกิดขึ้นจริงเท่านั้น คนร้ายจำนวนมากจึงใช้วิธีปกปิดการกระทำผิดด้วยการบดบังรูปภาพ การวาดภาพแทนการบันทึกผ่านกล้อง การตัดต่อ

เปลี่ยนแปลงบุคคลในภาพ เป็นต้น อย่างไรก็ตาม บางประเทศมีกรอบกฎหมายที่เข้มงวด เช่น สหราชอาณาจักรและสหรัฐอเมริกาที่บัญญัติไว้ชัดเจนว่าหากภาพลามกนั้นไม่ใช่ไฟล์ต้นฉบับ แต่มีการลอกเลียนหรือปรับแต่ง เมื่อเจ้าหน้าที่รัฐสามารถพิสูจน์ให้สิ้นสงสัยได้ว่าเกี่ยวข้องกับเหยื่อที่เป็นเด็ก กรณีเหล่านี้ก็ถือว่าเป็นการกระทำผิดกฎหมาย

กระบวนการตรวจพิสูจน์รูปภาพและวิดีโอที่จัดในรูปแบบดิจิทัลมีการพัฒนามากขึ้นกว่าแต่ก่อน เนื่องจากเทคโนโลยีปัญญาประดิษฐ์ทำให้การตรวจสอบการเผยแพร่สื่อลามกที่เป็นแบบดิจิทัลถูกตรวจจับโดยเจ้าหน้าที่ตำรวจได้สะดวกยิ่งขึ้น (NetClean Technologies, 2019) ในอดีตการสืบค้นหาไฟล์มัลติมีเดียในการตรวจพิสูจน์หลักฐานทางคดีจะใช้เทคนิคการสืบค้นค่าแฮช ซึ่งค่านี้เกิดจากการผลคำนวณทางคณิตศาสตร์ที่จัดทำบัญชีไฟล์ทุกประเภทในก้อนสำเนาพยานหลักฐาน จากนั้นผู้ตรวจจะนำรายการค่าแฮชที่ได้รับจากผู้จับกุมมาดำเนินการสืบค้นและเปรียบเทียบจนกว่าจะพบเจอไฟล์ที่มีค่าแฮชตรงกันอย่างถูกต้อง แต่เมื่อศักยภาพเครื่องคอมพิวเตอร์สำหรับตรวจพิสูจน์มีมากขึ้น ประกอบกับบริษัทผู้พัฒนาเครื่องมือตรวจพิสูจน์หลักฐานได้ประยุกต์ใช้เทคโนโลยีการสืบค้นรูปภาพโดยตรง จึงเรียกเทคโนโลยีนี้ว่า การคำนวณค่าแฮชแบบทนทาน (Robust Hashing) การพัฒนาเช่นนี้ทำให้ซอฟต์แวร์การตรวจมัลติมีเดียสามารถตรวจเปรียบเทียบความคล้ายของตัวเนื้อหาที่ปรากฏในรูปภาพแทนการค้นค่าแฮชของตัวไฟล์แบบเดิมได้ ซึ่งเทคโนโลยีนี้ได้ถูกพัฒนาและนำมาใช้อย่างเป็นทางการในชื่อว่าโฟโต้ดีเอ็นเอ (PhotoDNA)

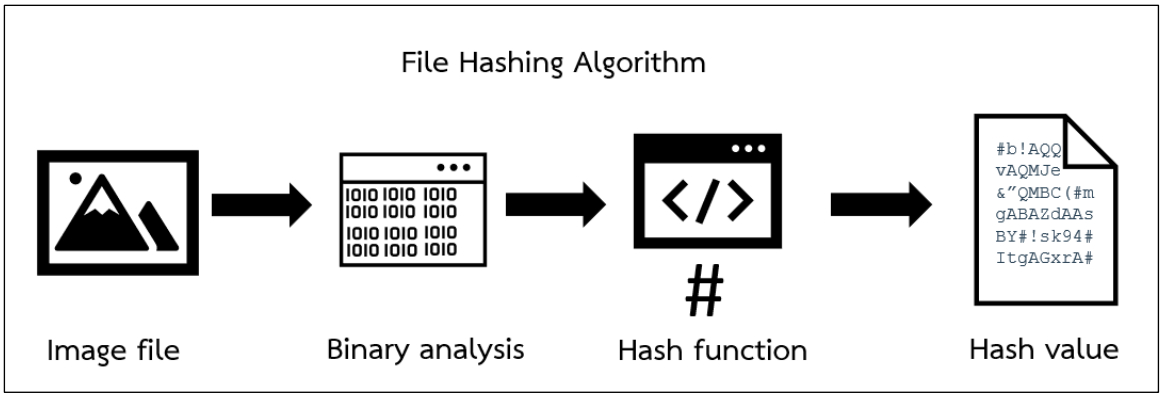
อย่างไรก็ตาม แม้ผู้ตรวจจะสามารถสืบค้นเนื้อหาที่ปรากฏในรูปภาพได้แล้ว แต่ข้อจำกัดการใช้เทคโนโลยีเดียวยังคงมีอยู่ ด้วยสาเหตุด้านประสิทธิภาพการตรวจจับใบหน้าและวัตถุในภาพที่เหมือนกัน หากมีการบดบังหรือความละเอียดของรูปภาพต้นฉบับแล้ว ย่อมทำให้การสืบค้นด้วยค่าแฮชแบบทนทานไม่สามารถคาดหวังผลลัพธ์ที่ต้องการได้ ในทางตรงกันข้าม หากมีการประยุกต์ใช้ควบคู่กับการสืบค้นแฮชชิ่งแบบไบนารีพร้อมกับการวิเคราะห์เชื่อมโยงกับร่องรอยดิจิทัลด้านอื่น ด้วยการบูรณาการเช่นนี้ย่อมทำให้การตรวจพิสูจน์เป็นไปอย่างมีประสิทธิภาพมากขึ้น ดังนั้น บทความนี้จึงต้องการนำเสนอเทคโนโลยีการสืบค้นมัลติมีเดียในงานตรวจพิสูจน์หลักฐานดิจิทัล และการวิเคราะห์เปรียบเทียบรวมถึงการนำเสนอแนวทางตรวจวิเคราะห์ที่เหมาะสมกับสภาพแวดล้อมการก่อเหตุอาชญากรรมในสังคมยุคดิจิทัล

บทความนี้จะไม่ได้กล่าวลึกลงในรายละเอียดสำหรับการอธิบายการทำงานของฟังก์ชันแฮชและซอฟต์แวร์โฟโต้ดีเอ็นเอ การทำให้เข้าใจในภาพรวมและความสำคัญของใช้ค่าแฮชในการตรวจสอบจับคู่สื่อมัลติมีเดียไม่ว่าจะเป็นภาพและวิดีโอเป็นประเด็นหลักที่ถูกกล่าวถึง ในส่วนถัดไปของบทความช่วงท้ายจะกล่าวถึงแนวทางการประยุกต์ใช้ฐานข้อมูลค่าแฮชที่ผู้ตรวจหรือภาคเอกชน สามารถแบ่งปันระหว่างกันได้ผ่านตัวกลางที่เป็นองค์กรสาธารณะที่ทำหน้าที่ช่วยเหลือเด็ก ส่วนสุดท้ายเป็นบทสรุปและข้อเสนอแนะสำหรับเป็นข้อพิจารณาในการเลือกใช้ประเภทของค่าแฮชที่เหมาะสำหรับงานตรวจพิสูจน์หลักฐานดิจิทัล

ฟังก์ชันแฮช (Hash Function)

เทคโนโลยีแฮช (Hashing) ถูกนำมาใช้ในหลายกรณีสำหรับการทำงานของผู้บังคับใช้กฎหมาย ผู้ให้บริการสื่อสังคมออนไลน์ องค์กรช่วยเหลือเด็ก และภาคธุรกิจ ในการร่วมกันตรวจคัดกรองเนื้อหารุนแรงและล่อแหลมที่เกี่ยวกับการล่วงละเมิดต่อเด็ก โดยเทคโนโลยีนี้ถูกนับได้ว่าเป็นเครื่องมือสำคัญในการสืบสวนทางดิจิทัลที่มีประสิทธิภาพและน่าเชื่อถือมากในปัจจุบัน

การคำนวณค่าแฮชเป็นกระบวนการหนึ่งทางวิทยาการคอมพิวเตอร์ใช้สำหรับการตรวจสอบความถูกต้องและความครบถ้วนสมบูรณ์ของไฟล์ที่จัดเก็บในรูปแบบดิจิทัล เมื่อไฟล์แต่ละไฟล์ถูกประมวลค่าแฮชเพื่อจัดทำบัญชี (File Index) แล้วนั้น ค่าแฮชของรูปภาพที่ถูกระบุไว้แล้วสามารถนำไปเปรียบเทียบกับไฟล์อื่นโดยมีค่าแฮชที่ตรงกันเพื่อระบุผลการจับคู่



ภาพที่ 1 รูปแบบอัลกอริทึมการคำนวณค่าแฮชไฟล์รูปภาพ

ฟังก์ชันแฮชเข้ารหัส (Cryptography Hash Function) เป็นแนวคิดพื้นฐานในการดำเนินการกับข้อมูล โดยคุณสมบัติของการทำงานนี้คือการย่อยข้อมูลให้เล็กลง ซึ่งพิจารณาหน่วยขนาดเล็กของข้อมูลคือบิต กลุ่มบิตที่ถูกนำเข้ามาคำนวณมาจากคอมพิวเตอร์ได้พิจารณาข้อมูลไบนารีที่บรรจุในไฟล์ดิจิทัล เมื่อกลุ่มบิตของไฟล์ที่กำลังถูกคำนวณมีขนาดความยาวที่แปรเปลี่ยนได้ (Variable Length) นี้ความยาวนั้นจะถูกลดขนาดโดยบีบให้เล็กลงให้เป็นความยาวชนิดคงที่ (Fixed-Length) จนได้ค่ารหัสอักขระที่เป็นตัวเลขชุดหนึ่งที่มีค่าเฉพาะและเอกลักษณ์ในตัวเอง ซึ่งนั้นคือค่าแฮช (Hash Value) โดยหลักการทำงานทางเดียว (One Way Function) เป็นคุณสมบัติสำคัญหนึ่งของค่าแฮช เมื่อไฟล์ได้ทำฟังก์ชันแฮชแล้วจะไม่สามารถปฏิบัติการย้อนกลับเพื่อให้ได้ข้อมูลที่เป็นรูปแบบไฟล์เดิมออกมาได้

การคำนวณหาค่าแฮชแบบทั่วไปในการประมวลผลของเครื่องคอมพิวเตอร์ถูกจัดว่าเป็นการหาค่าแฮชแบบไบนารี (Binary Hash) ซึ่งค่าแฮชลักษณะนี้ถูกสร้างขึ้นโดยอัลกอริทึมทางคณิตศาสตร์ที่ทำการแปลงขนาดของข้อมูลให้เล็กลง ความยาวของข้อมูลนั้นพิจารณาจากการตรวจสอบไฟล์ดิจิทัลให้อยู่ในค่าของเลขฐานสิบหก หรือเลขฐานสอง (Binary) อย่างใดอย่างหนึ่ง จากนั้นทำการคำนวณข้อมูลต้นฉบับให้มีขนาดสั้นลง ซึ่งอาจมีการคำนวณเกิดขึ้นหลายรอบ ทั้งนี้ขึ้นอยู่กับขนาดของไฟล์ที่ถูกคำนวณ

เมื่อจำนวนรอบการคำนวณใกล้ถึงจุดสิ้นสุดแล้ว ค่าตัวเลขที่ได้ในแต่ละรอบจะถูกกำหนดให้มีความแตกต่างกัน เมื่อเสร็จการคำนวณแล้วผลลัพธ์ที่ได้เป็นค่าแฮชของไฟล์ นักวิชาการทางด้านนิติวิทยาดิจิทัลเรียกอีกอย่างว่าลายนิ้วมือดิจิทัล (Digital Fingerprint)

เมื่อพิจารณาถึงหลักการรักษาความสมบูรณ์ของค่าแฮชแล้วจะเกี่ยวข้องกับอัลกอริทึมที่ถูกเรียกใช้งานเอ็มดีไฟว์ฟ (MD5) ชาร์วัน (SHA-1) และ ชาร์ทูว (SHA-2) เป็นสามอัลกอริทึมที่นิยมในกระบวนการฟังก์ชันแฮช (Saetang & Boonkrong, 2017) การทำงานของอัลกอริทึมทั้งสามจะแบ่งข้อมูลออกเป็นบล็อก โดยแต่ละบล็อกมีขนาด 512 บิต แต่เมื่อประมวลผลออกมาแล้วค่าแฮชที่ได้มีขนาดต่างกัน กล่าวคือ MD5 มีค่าแฮชเท่ากับ 128 บิต ส่วน SHA-1 มีค่าแฮชขนาด 160 บิต ทั้งนี้ให้พิจารณาความรัดกุมที่ขนาดค่าแฮชที่เป็นบิต หากมีค่ามากแสดงให้ว่าจำนวนรหัสตัวเลขที่บรรจุอยู่ในนั้นมีจำนวนมากกว่า ข้อนี้ทำให้ฟังก์ชันแฮชชนิดที่มีขนาดผลลัพธ์ที่จำนวนบิตมากกว่า มีโอกาสเกิดการชนกัน (Collision) หรือค่าเหมือนกันได้น้อยกว่ามาก

จากหลักที่เชื่อกันบนพื้นฐานความรู้ทางวิทยาการคอมพิวเตอร์ที่บ่งบอกว่าไฟล์ดิจิทัลที่เกิดขึ้นในขณะประมวลผลนั้น มีลักษณะจำเพาะและหน้าที่การทำงานแตกต่างกันออกไป ไฟล์รูปภาพที่จัดเก็บแบบดิจิทัลก็มีลักษณะเดียวกัน การทำให้รูปภาพมีลายนิ้วมือหรืออัตลักษณ์ของตัวเองโดยการใช้เทคนิคคำนวณผลทางคณิตศาสตร์ ทำให้ผลการคำนวณสามารถกำหนดค่าแฮชที่เป็นเอกลักษณ์เฉพาะเจาะจงของแต่ละไฟล์ภาพได้ โดยผลลัพธ์ที่เป็นค่าแฮชนี้จะอยู่ในแบบรหัสตัวเลขที่ไม่ซ้ำกัน และเมื่อนำไปเปรียบเทียบกับไฟล์อื่นจะมีผลของค่าแฮชที่แตกต่างกันอย่างสิ้นเชิงอีกด้วย

ฟังก์ชันแฮชรูปภาพแบบทนทาน (Robust Image Hashing)

กระบวนการฟังก์ชันแฮชสำหรับรูปภาพนั้น จะพิจารณาองค์ประกอบหลักสองประเด็นให้คงอยู่ได้แก่ ความแข็งแกร่งทนทาน (Robustness) และการวิเคราะห์แยกแยะ (Discrimination) โดยสองปัจจัยนี้ทำให้ประสิทธิภาพของค่าแฮชที่เป็นเอกลักษณ์ของไฟล์ที่ได้นั้น สามารถรักษาความสมบูรณ์และทนทานต่อการดัดแปลงแก้ไขในภายหลัง หนึ่งในความท้าทายของการพัฒนาฟังก์ชันแฮชประเภทนี้คือการหมุนกลับ (Rotation) ของภาพ (Tang, Zhang, Li, & Zhang, 2016) กล่าวคือ หากไฟล์ภาพที่ต้องการตรวจสอบความคล้ายถูกจัดให้หมุนอยู่ในตำแหน่งมากกว่าศูนย์องศาซึ่งเป็นตำแหน่งของภาพต้นฉบับแล้วโอกาสที่จะเกิดความผิดพลาดในการจับคู่ความเหมือนกันของภาพนั้นย่อมเกิดขึ้นได้

การคำนวณค่าแฮชปกติหรือแบบไบนารีแฮชนั้น เป็นการกระทำกับขนาดของไฟล์เพื่อให้ได้ผลลัพธ์ที่เป็นรหัสตัวเลขที่มีความเฉพาะแตกต่างจากไฟล์อื่น แต่ในทางตรงกันข้าม ค่าแฮชโรบัสหรือแบบทนทานนี้เป็นการคำนวณแบบเถรตรงหรือตรงไปตรงมา กล่าวให้ง่ายขึ้นคือการคำนวณหาค่าแฮชที่ทำกับรูปภาพเป็นการคิดประมวลจากเนื้อหาที่ปรากฏในภาพ ทั้งนี้ไม่ว่าจะเป็นใบหน้า สิ่งของ หรือยานพาหนะ เป็นต้น หากเมื่อปรากฏในภาพจะถูกนำมาคิดคำนวณทั้งสิ้น อย่างไรก็ตาม วัตถุเหล่านี้ไม่ได้ถูกระบุชนิดหรือจดจำในแบบปัญญาประดิษฐ์แต่อย่างใด

อัลกอริทึมหรือกระบวนการคิดแบบตรรกะของคอมพิวเตอร์ในการคำนวณค่าแฮชของรูปภาพชนิดทนทานนี้ประกอบขึ้นด้วยหลากหลายส่วน ทั้งการกำหนดขนาดภาพ การจัดทำตารางแบ่งโซนพื้นที่บนภาพ หรือแม้กระทั่งการกำหนดแทนค่าสีของภาพด้วยตัวเลข ดังนั้น ความแตกต่างที่เห็นได้ชัดเจนเมื่อเทียบกับค่าแฮชทั่วไปจึงอยู่ที่ค่าแฮชไบนารี เกิดจากการคำนวณขนาดไฟล์ของรูปภาพ แต่ค่าแฮชโรบัสคำนวณบนพื้นฐานของเนื้อหาที่ปรากฏบนภาพ ส่วนสาเหตุที่เรียกว่าค่าแฮชแบบทนทานนั้น เนื่องจากเมื่อผลการคำนวณในแต่อัลกอริทึมเสร็จแล้ว ขั้นตอนสุดท้ายต้องนำผลลัพธ์ที่ได้จากขั้นตอนวิธีย่อยทางโปรแกรมมารวมและคำนวณผลลัพธ์อันสุดท้าย จึงจะได้ค่าแฮชที่มีคุณลักษณะความแข็งแกร่งโดยทนทานต่อการแก้ไขเปลี่ยนแปลงได้เป็นอย่างดี ทั้งนี้ กระบวนการประมวลผลฟังก์ชันแฮชแบบนี้สามารถทำได้ดีกับไฟล์วิดีโอเช่นเดียวกับรูปภาพ (Singh & Farid, 2019) ด้วยเหตุดังกล่าวนี้ จึงทำให้ค่าแฮชรูปภาพแบบทนทาน (Robust Image Hash) สามารถรักษาความน่าเชื่อถือและความสมบูรณ์ของไฟล์ได้เป็นอย่างดี

ความแตกต่างที่สำคัญอีกประการหนึ่งหากไฟล์รูปภาพที่เนื้อหาปรากฏเหมือนกันทุกประการแต่มีการจัดเก็บไฟล์เป็นคนละประเภท ยกตัวอย่างเช่น รูปภาพแมวตัวเดียวกันจำนวนสองภาพแต่ไฟล์แรกเป็นชนิดเจเพ็ก (JPEG) และไฟล์ที่สองเป็นชนิดพีเอ็นจี (PNG) จากกรณีนี้ การคำนวณค่าไบนารีแฮชจะให้ผลลัพธ์เป็นรหัสตัวเลขที่แตกต่างกันอย่างสิ้นเชิง ในทางตรงกันข้ามหากคำนวณค่าแฮชแบบโรบัสจะได้ผลลัพธ์ที่เป็นรหัสตัวเลขตรงกัน ทั้งนี้ อย่างที่ทราบกันดีว่าการคำนวณค่าแฮชอย่างหลังเป็นการพิจารณาเนื้อหาที่ปรากฏบนภาพ

Kafae02.jpg	File name	cat-test1_in_PNG.png
.jpg	File extension	.png
image/jpeg	Mime type	image/png
357.31 KiB (365882 b)	File size	1.98 MiB (2079628 b)
Hash values	**Hashing type**	**Hash values**
857507F3DA46ED44D4B4B3241927C825AC933453	SHA-1	CB13A3386B46D4EFB973FA34C94FA35DB38F90C2
4A1DE3778F42455A1A4F93B2E2E72559	MD5	CC63E9836C5BFF58B0838910730059F4
OwUjZDAQSrNrTEt5EWFTzFwrdtYjb2CONAdTKy4NhjyfO1JNSXrSRXRHvEgTV1FlfQYsOTUdOzFsIz1AjCp4MBqPYDAKKzovjgRlByRfNC2gKwxnfhsgcB21HncyTihbWRYaJw5tDDE8Lg41aRwCrz2PDa6qNSgaGik0ER8xRg0MP1kDGSZXLEwVoBpSlykg	PhotoDNA	PAUjZDAQSrNrTEt5EWFTy1wrdtYjb1+ONAdTLC4NhjyfO1JNSXrSRXRHvEgTV1FlfQYsOTUdPDFsIz1AjCt4MBqPYDAKKzovjgRlByRfNC2gKwxnfxsgcBy1HncyTihbWRUaJw5tDDE8Lg41aRwCrz2PDa6qNSgaGik0ER8xRg0MP1oDGSZXLEwVnxpSlykg
Matched	Visual copies	Matched

หมายเหตุ: ขนาดของค่าแฮชแต่ละประเภทมีจำนวนตัวอักขระที่ต่างกันโดย PhotoDNA เท่ากับ 192 ตัว, SHA-1 เท่ากับ 40 ตัว, และ MD5 เท่ากับ 32 ตัว

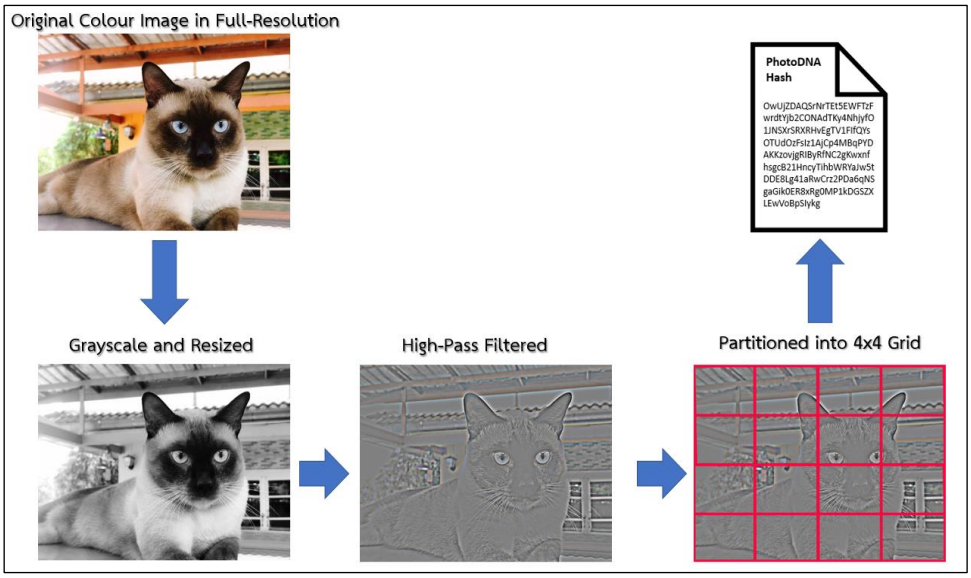
ภาพที่ 2 ผลการคำนวณค่าแฮชจากการใช้อัลกอริทึมต่าง ๆ

ความเหมือนกันที่มีของเทคนิคคำนวณแฮชทั้งสองนี้คือ การปฏิเสธกระบวนการวิศวกรรมย้อนกลับ (Reengineer Construction) กล่าวคือ ค่าแฮชที่ได้ทั้งสองแบบไม่สามารถคิดหรือประดิษฐ์ย้อนกลับไม่ว่าด้วยกระบวนการใดให้กลับกลายเป็นรูปภาพหรือสื่อมัลติมีเดียได้ ทั้งนี้ส่วนหนึ่งมาจากการคำนวณแฮชเป็นการพิจารณาจากโครงสร้างของไฟล์และเนื้อหาที่ปรากฏจากนั้นแปลงค่าเป็นตัวเลขทางคณิตศาสตร์ เช่นเดียวกับลายนิ้วมือและดีเอ็นเอของมนุษย์เมื่อมีการคำนวณหาค่าความเป็นเอกลักษณ์เสร็จแล้ว กระบวนทางกายภาพเหล่านี้ก็ไม่สามารถกระทำย้อนกลับให้ได้ตัวลายนิ้วมือหรือตัวเนื้อเยื่อที่ได้จากการเก็บวัตถุพยานทางกายภาพหรือชีววิทยา

โฟโต้ดีเอ็นเอ (PhotoDNA)

โฟโต้ดีเอ็นเอเป็นหนึ่งในเทคโนโลยีแฮชที่นิยมใช้กันอย่างแพร่หลาย โดยต้นกำเนิดจากการใช้แบบเจาะจง การประมวลผลค่าแฮชเช่นนี้แรกเริ่มถูกนำมาใช้สำหรับการตรวจสอบสื่อลามกอนาจารเด็กที่ถูกล่วงละเมิดทางเพศและเผยแพร่ผ่านเครือข่ายออนไลน์ ซึ่งไมโครซอฟท์เป็นผู้เริ่มพัฒนาร่วมกับศาสตราจารย์แฮนี่ ฟาริด (Hany Farid) จากมหาวิทยาลัยดาร์ตมัธ (Dartmouth College) ตั้งแต่ปี พ.ศ. 2552 โดยในระยะตั้งต้นนั้น โครงการนี้ถือเป็นกิจกรรมหนึ่งของไมโครซอฟท์ในการช่วยเหลือสังคม เมื่อระบบได้รับการพัฒนาแบบสมบูรณ์เรียบร้อยจึงถูกบริจาคให้กับองค์กรที่ช่วยเหลือเด็ก เช่น เน็กเม็ก (NCMEC) โปรเจควิค (Project VIC) เป็นต้น นอกจากการสนับสนุนให้กับองค์กรเพื่อการกุศลที่ไม่แสวงหากำไรแล้ว ไมโครซอฟท์ยังอนุญาตให้ผู้ให้บริการอินเทอร์เน็ตหลากหลายแพลตฟอร์มและบริษัทซอฟต์แวร์สามารถนำไปพัฒนาต่อยอด โดยขอให้คงวัตถุประสงค์หลักคือการสร้างซอฟต์แวร์เพื่อตรวจหาสื่อลามกอนาจารเด็ก โดยแพลตฟอร์มออนไลน์ที่กล่าวถึง ได้แก่ วันไดรฟ์ (OneDrive) กูเกิลจีเมล (Gmail) ทวิตเตอร์ (Twitter) เฟซบุ๊ก (Facebook) และอะโดบีซิสเต็มส์ (Adobe Systems) ส่วนองค์กรธุรกิจที่สำคัญในการนำโฟโต้ดีเอ็นเอไปพัฒนาต่อยอด ได้แก่ เน็ตคลีน (NetClean)

ภาพรวมของโฟโต้ดีเอ็นเอเป็นการสร้างค่าแฮชของสื่อมัลติมีเดียแต่ละไฟล์ โดยกำหนดให้มีหน้าที่คล้ายกันกับลายนิ้วมือเพื่อใช้สำหรับการค้นหาและการเปรียบเทียบความคล้ายคลึง การระบุสำเนาของไฟล์ที่คล้ายคลึงกันโดยไม่ต้องใช้ผู้ใช้ซึ่งเป็นมนุษย์เป็นผู้ยืนยันอีกต่อไป ผลการคำนวณค่าแฮชที่ได้นี้สามารถรักษาคุณสมบัติที่คล้ายคลึงกันของไฟล์ต้นฉบับกับไฟล์สำเนาที่มีการเปลี่ยนแปลงสีเพียงเล็กน้อย การแปลงภาพขาวดำ หรือแม้กระทั่งการปรับขนาดไฟล์ โดยการประมวลผลของโฟโต้ดีเอ็นเอนั้นจำเป็นต้องใช้งานกับเครื่องคอมพิวเตอร์ที่มีประสิทธิภาพสูง ด้วยเหตุนี้นักพัฒนาจึงได้ปรับปรุงการใช้งานให้ง่ายขึ้นแต่คงประสิทธิภาพไว้ด้วยการนำเทคโนโลยีนี้ไปให้บริการผ่านระบบประมวลผลแบบคลาวด์ (Cloud Service) ภายใต้ชื่อ PhotoDNA Cloud Service



ภาพที่ 3 กระบวนการจัดทำโฟโต้ดีเอ็นเอด้วยสามขั้นตอน
ที่มา: (Farid, 2018)

ขั้นตอนการทำงานของโฟโต้ดีเอ็นเอเริ่มจากเปลี่ยนรูปภาพเป็นโทนสีเทา (Greyscale) ต่อมาคือการปรับความคมชัดของภาพด้วยเครื่องมือไฮพาสฟิลเตอร์ (High-Pass Filter) ที่ช่วยกรองพื้นที่หรือโซนของภาพที่ไม่สำคัญให้ลดทอนลงและไม่ทำให้บดบังจุดเด่นของภาพ วิธีนี้ทำให้ภาพโทนสีเทามีความเด่นชัดจากการเน้นขอบ จุด และเฉดของสีให้แตกต่างกันมากขึ้น อันจะทำให้กระบวนการประมวลผลภาพมีความแม่นยำในการจำแนกยิ่งขึ้น จากนั้นทำการสร้างตารางหรือกริด (Grid) ให้เกิดกรอบสี่เหลี่ยมขนาดเล็ก โดยกริดเหล่านั้นถูกปรับขนาดและกำหนดค่าตัวเลขให้กับสี่เหลี่ยมแต่ละชิ้นค่าตัวเลขนั้นได้จากการแทนค่าสีแต่ละจุดหรือพิกเซล (Pixel) บนรูปถ่ายให้ถูกอ่านแบบดิจิทัลได้ ค่าเหล่านี้ที่กำหนดเข้าไปให้กรอบสี่เหลี่ยมถูกกำหนดให้เป็นข้อมูลเอกลักษณ์ (Signature) ของรูปภาพ ข้อมูลเอกลักษณ์นี้ถูกเรียกอีกอย่างว่า ดีเอ็นเอรูปถ่าย เมื่อต้องการตรวจสอบการจับคู่ผู้ให้บริการจึงดำเนินการตรวจสอบสี่เหลี่ยมเล็ก ๆ ในแต่ละไฟล์ภาพนั้นทีละชิ้นเพื่อจับคู่กับไฟล์ภาพอื่นที่กระจัดกระจายอยู่ (ภาพที่ 3) บนเครือข่ายระบบคอมพิวเตอร์ของผู้ให้บริการ การสืบค้นรูปจะมองหาคุณลักษณะความคล้ายกัน (Similarity Feature) โดยอาศัยค่าของดีเอ็นเอที่เกิดจากตารางกริดบนรูปถ่ายต้นฉบับ

เมื่อกล่าวถึงการวัดความคล้ายของรูปภาพ (Similarity Measure) จะสอดคล้องกับเทคนิควิธีการค้นคืนรูปภาพ (Content-Based Image Retrieval) ที่อาศัยการวิเคราะห์ของตัวเลขจากคุณลักษณะที่สกัด (Extraction) ออกมาจากรูปภาพทั้งสองรูป การวัดเช่นนี้พิจารณาจากคุณลักษณะของภาพด้านสี รูปทรง และลวดลาย ซึ่งล้วนแล้วแต่เป็นคุณลักษณะเด่นที่ปรากฏออกมาให้เห็นในแต่ละรูป ค่าความคล้ายคลึงที่ได้อยู่ในรูปแบบตัวเลขที่แสดงให้เห็นถึงความสัมพันธ์ของวัตถุต่อวัตถุที่ปรากฏในภาพเมื่อกำหนดให้ค่าเวกเตอร์ของวัตถุที่ปรากฏในภาพทั้ง x และ y มีความสัมพันธ์กันและผ่านการคำนวณด้วยเทคนิคที่เหมาะสม (Chinpanthana, 2017) ทั้งนี้ ภาพดิจิทัลหนึ่งภาพนั้นสามารถกำหนดฟังก์ชันให้กับ

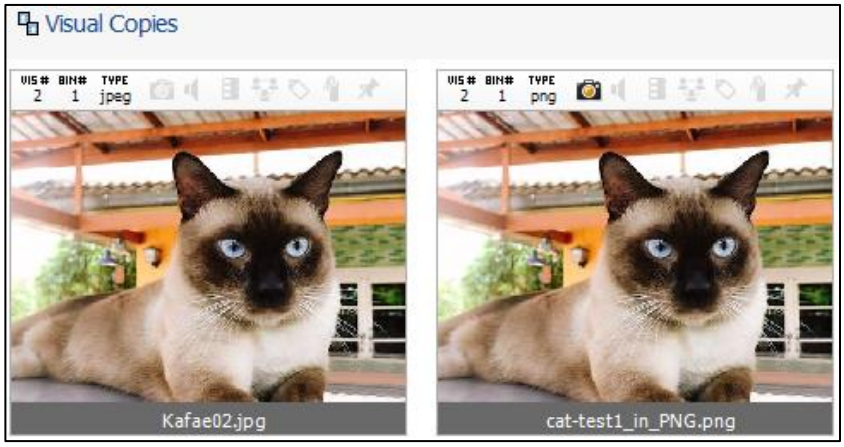
คอมพิวเตอร์ในการประมวลผลเป็นแบบสองมิติ โดยที่ค่า x และ y คือ พิกัดของภาพ ส่วน f เป็นรูปแบบฟังก์ชันที่แล้วแต่ในการประมวลผลจะพิจารณา ยกตัวอย่างเช่น ค่าความเข้มแสงของภาพ เป็นต้น

ระยะต่อมาจนถึงปัจจุบันการพัฒนาโฟโต้ดีเอ็นเอได้รับการต่อยอดประสิทธิภาพของซอฟต์แวร์ครั้งนี้ โดยเพิ่มฟังก์ชันการทำงานที่สำคัญเข้าไปคือการตรวจวิเคราะห์ไฟล์มัลติมีเดียที่เป็นวิดีโอ (ภาพเคลื่อนไหว) และเสียง (INTERPOL & ECPAT, 2018) นอกเหนือจากภาพนิ่งที่เป็นกลไกสำคัญของซอฟต์แวร์รุ่นแรก ด้วยการปรับปรุงซอฟต์แวร์ครั้งใหญ่ ส่งผลให้บริษัทผู้ให้บริการสื่อสังคมออนไลน์ทั้งเฟซบุ๊ก ทวิตเตอร์ กูเกิล และไมโครซอฟท์ ประกาศแผนการที่จะใช้โฟโต้ดีเอ็นเอเป็นเครื่องมือหลักในการตรวจจับเนื้อหาที่มีความรุนแรงที่เกิดขึ้นบนเครือข่ายการให้บริการของตน ดังนั้น การตรวจสอบเนื้อหาของกลุ่มก่อการร้ายจึงถูกกำหนดเป็นกรอบการตรวจสอบหลักที่เพิ่มมาจากสื่อลามกอนาจาร

เทคโนโลยีสืบค้นมัลติมีเดียสำหรับงานตรวจวิเคราะห์ดิจิทัล

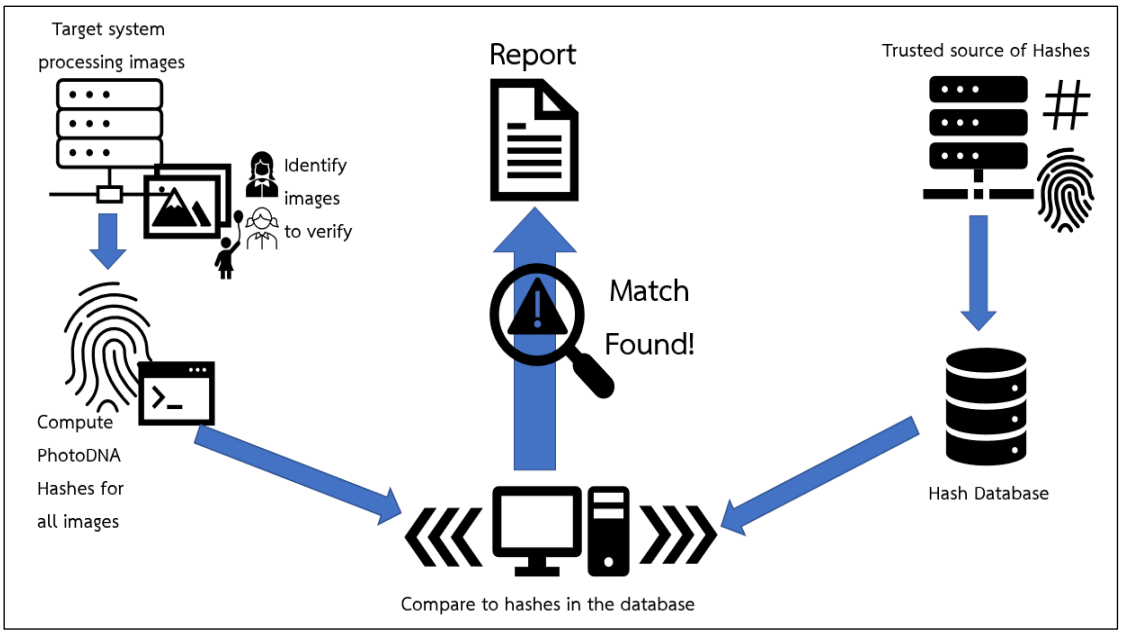
ศูนย์เพื่อเด็กหายและถูกแสวงหาประโยชน์แห่งชาติของสหรัฐอเมริกา (National Center for Missing and Exploited Children) หรือเรียกแบบย่อว่า เน็กเม็ก (NCMEC) มีหน้าที่จัดทำบัญชีรายการค่าแฮชที่ระบุเจาะจงไฟล์สื่อลามกอนาจารต่อเด็ก บัญชีค่าแฮช (Hash List) มีการปรับปรุงความถูกต้องและความทันสมัยอย่างสม่ำเสมอ โดยองค์กรที่ทำงานเพื่อเด็กและผู้ให้บริการอินเทอร์เน็ต (Internet Service Provider) สามารถเข้าร่วมกับเน็กเม็กเพื่อดำเนินการปรับปรุงข้อมูลค่าแฮชเหล่านี้ นอกจากนั้นเน็กเม็กยังเปิดให้องค์กรเหล่านั้นนำไปใช้ในกระบวนการตรวจสอบเนื้อหาความรุนแรงที่มีการเผยแพร่ในแพลตฟอร์มการบริการของตนเองแต่ละค่าย

โปรเจ็ควิค (Project VIC) เป็นโครงการที่เกิดจากความร่วมมือของหน่วยงานรัฐ องค์กรไม่แสวงหากำไร และองค์กรภาคเอกชนที่ให้บริการเข้าถึงระบบอินเทอร์เน็ตที่ช่วยกันตรวจตราเนื้อหาความรุนแรงต่อเด็กที่อาจถูกเผยแพร่บนเครือข่ายออนไลน์ โครงการนี้ยังถือว่าเป็นผู้ริเริ่มในการจัดเก็บบัญชีค่าแฮชรูปภาพและวิดีโอจำนวนมหาศาลไว้ เพื่อแบ่งปันให้หน่วยงานบังคับใช้กฎหมายที่ลงทะเบียนเข้าร่วมจากทั่วโลก เน็กเม็กเป็นองค์กรหลักที่ผลักดันและสนับสนุนโครงการนี้ให้เกิดผลเป็นรูปธรรม



ภาพที่ 4 ตัวอย่างผลการสืบค้นความเหมือนของภาพ (Visual Copy) ด้วยซอฟต์แวร์ Griffeye Analyse DI

การใช้ฐานข้อมูลแฮชสื่อมัลติมีเดียจำนวนหลายร้อยล้านรายการที่ระบุเนื้อหาการล่วงละเมิดทางเพศต่อเด็กจะถูกเก็บและระบุความน่าเชื่อถือให้กับเจ้าหน้าที่ตำรวจเมื่อนำไปใช้ว่ามาจากแหล่งข้อมูลของโปรเจ็ควิค การดำเนินการลักษณะนี้ช่วยให้เจ้าหน้าที่วิเคราะห์สามารถแยกแยะรูปภาพที่เคยถูกระบุมาก่อนให้เป็น “ไฟล์ที่รู้จัก” (Known Files) ส่วนผลลัพธ์ที่ได้จากการตรวจจับคู่กับฐานข้อมูลโฟโต้ดีเอ็นเอถูกจัดว่าเป็น “ภาพที่รู้จัก” (Known Images) (NetClean Technologies, 2019) จากนั้นจะทำการคัดกรองหรือฟิลเตอร์ (Filter) เอารูปภาพหรือไฟล์ที่ถูกระบุแล้วออกไป ขั้นตอนนี้ทำให้ปริมาณไฟล์ที่ต้องสงสัยถูกจำกัดวงให้แคบลง ผู้ตรวจจะสามารถคัดแยกไฟล์ต้องสงสัย รวมถึงสื่อลามกที่ไม่เคยถูกระบุมาก่อน แล้วมีการคำนวณค่าแฮชทั้งแบบไบนารีดั้งเดิมและแบบโรบัสอิมเมจไว้อยู่แล้ว อย่างไรก็ตาม การคำนวณค่าแฮชของซอฟต์แวร์โฟโต้ดีเอ็นเอนั้นไม่ได้ถูกติดตั้งในเครื่องมือการตรวจพิสูจน์ไว้โดยทั่วไป เครื่องมือตรวจพิสูจน์แบบโอเพนซอร์ซ (Open Source) พบว่า มีข้อจำกัดในการประมวลผลค่าแฮชของไฟล์ในสำเนาหลักฐานดิจิทัลที่อาจไม่รองรับอัลกอริทึมแบบ SHA-1 หรือดีกว่า (Tunyasevee & Witchuwanich, 2019) ดังนั้น ผู้ตรวจพิสูจน์จำเป็นต้องศึกษาเครื่องมือหรือซอฟต์แวร์ที่รองรับการติดตั้งโปรแกรมเสริมที่เปิดให้ใช้โฟโต้ดีเอ็นเอก่อนทุกครั้ง

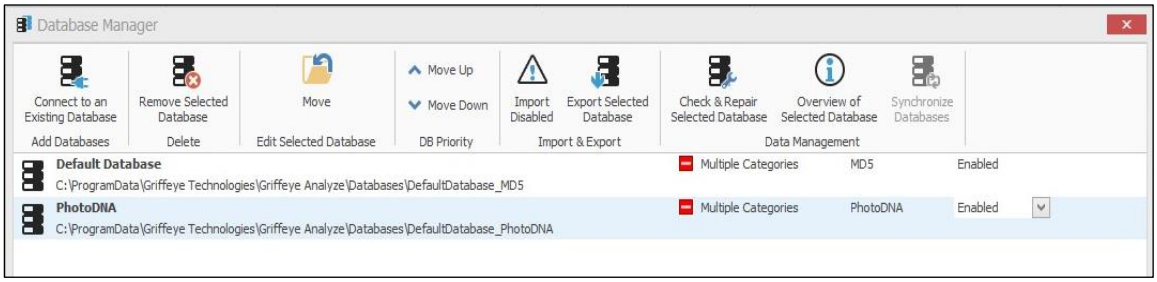


ภาพที่ 5 PhotoDNA Workflow
ที่มา: (Microsoft, 2018)

เมื่อไฟล์สื่อลามกเดิมที่เคยระบุถูกจับคู่ว่าพบเจอแล้ว (Matched) ไฟล์ต้องสงสัยใหม่ที่มีเนื้อหาเกี่ยวกับการล่วงละเมิดทางเพศต่อเด็กจะถูกกำหนดและส่งย้อนกลับไปยังฐานข้อมูลของโปรเจ็ควิคอีกครั้ง คราวนี้เมื่อทำการเริ่มต้นสืบสวนกับคดีใหม่หรือพยานหลักฐานอิเล็กทรอนิกส์รายการใหม่ ผู้ตรวจก็จะได้รับข้อมูลค่าแฮชที่มีการอัปเดตล่าสุดจากโปรเจ็ควิคไปด้วย การดำเนินการของโครงการนี้ส่งผลให้

เจ้าหน้าที่ตำรวจดำเนินการตรวจคัดแยกเหยื่อผู้เสียหายเป็นไปอย่างราบรื่น ซึ่งถือว่าเป็นการสนับสนุนที่มีประโยชน์อย่างยิ่ง หากผู้ตรวจต้องคัดกรองไฟล์สื่อลามกด้วยตนเองจะทำให้เกิดความล่าช้าและความผิดพลาดเกิดขึ้นได้ตลอดเวลา

ด้วยจุดเด่นของการประมวลค่าแฮชแบบหลายขั้นตอนของโฟโต้ดีเอ็นเอทำให้ได้รับการต่อยอดนำไปพัฒนาต่อยอดในเชิงคอมพิวเตอร์วิทัศน์อย่างกว้างขวาง หนึ่งในคุณลักษณะที่เพิ่มขึ้นมาในภายหลังคือการตรวจวัดระดับความลามก (Nudity Level) บริษัทไมโครซอฟท์ได้ใช้จุดเด่นของอัลกอริทึมแบบทนทาน (Robust) ตรวจวิเคราะห์รูปภาพ โดยให้ศึกษาจุดและเฉดสีที่ปรากฏในภาพว่ามีความคล้ายคลึงกับผิวหนังของมนุษย์มากเพียงใด (Jusas, Birvinskas, & Gahramanov, 2017) ยิ่งไปกว่านั้น ตัวซอฟต์แวร์ยังถูกเพิ่มประสิทธิภาพให้วิเคราะห์ภาพผิวหนัง ให้มีความน่าจะเป็นว่าเป็นบริเวณพื้นที่ส่วนใดของร่างกาย หากเป็นภาพถ่ายบริเวณหน้าอกหรือจุดซ่อนเร้นในร่างกาย การคำนวณระดับความลามกนี้จะมีแสดงค่าบ่งบอกในระดับที่สูงขึ้นกว่าปกติ อย่างไรก็ตาม คุณลักษณะนี้ยังต้องได้รับการพัฒนาอย่างต่อเนื่อง ทั้งนี้ผลการตรวจจับระดับความลามกมีโอกาสเกิดความผิดพลาดได้ง่ายหากเฉดสีในภาพไม่ได้แสดงตำแหน่งอวัยวะที่ต้องการระบุที่เกี่ยวกับทางเพศไว้อย่างชัดเจน



ภาพที่ 6 หน้าต่างการจัดการฐานข้อมูลค่าแฮชในซอฟต์แวร์ Griffeye Analyse DI

นอกจากหน่วยงานบังคับใช้กฎหมายที่มีการนำฐานข้อมูลค่าแฮชภาพสื่อลามกไปใช้อย่างเป็นรูปธรรมแล้ว ผู้ให้บริการอินเทอร์เน็ตและเว็บไซต์สื่อสังคมออนไลน์ได้มีการประยุกต์ฐานข้อมูลโฟโต้ดีเอ็นเอเข้าไปใช้งานร่วมกันกับโปรแกรมการทำงานหลักของแพลตฟอร์มตนเอง ซึ่งบัญชีรายการชุดค่าแฮชเหล่านั้นช่วยให้ผู้ให้บริการภาคเอกชนสามารถดำเนินการตรวจสอบเนื้อหาที่มีการเผยแพร่โดยลูกค้าหรือผู้ใช้บริการเครือข่ายของตน ทั้งนี้เป็นการรักษาความเชื่อมั่นและความปลอดภัยที่มีต่อเนื้อหาความรุนแรงที่ไม่เหมาะสมไม่ให้มีการเผยแพร่และส่งต่อออกไปในวงกว้าง โดยกระบวนการสืบค้นภายในเครือข่ายของผู้ให้บริการ ทำให้การระบุตัวตนผู้ใช้ที่ตกเป็นเหยื่อได้รับการตอบสนองและถูกปิดกั้นจากผู้ใช้รายอื่น การตรวจสอบไล่เรียงย้อนกลับไปยังผู้ใช้ที่เป็นต้นตอของการเผยแพร่ไฟล์ผิดกฎหมายก็สามารถดำเนินการได้อย่างมีประสิทธิภาพ อาศัยการสืบค้นเฉกเช่นเดียวกับที่เจ้าหน้าที่บังคับใช้กฎหมายดำเนินการกับผู้กระทำผิด ดังนี้ การดำเนินการแบบองค์รวมอย่างนี้ทำให้การประสานงานระหว่างตำรวจและสื่อสังคมออนไลน์เป็นไปอย่างราบรื่นในการกวาดล้างผู้ที่อัปโหลดและส่งต่อภาพสื่อลามกอนาจารเด็ก

บทวิเคราะห์ บทวิจารณ์ และข้อเสนอแนะ

บทวิเคราะห์

หลักการของคอมพิวเตอร์วิทัศน์ (Computer Vision) อยู่บนพื้นฐานของการทำความเข้าใจตรรกะการมองเห็นหรือเข้าใจบริบทของภาพถ่ายที่เกิดขึ้นในเครื่องจักรให้ใกล้เคียงกับรูปแบบระบบการมองเห็นของมนุษย์ ซึ่งเทคโนโลยีการสืบค้นภาพแบบย้อนกลับ (Reverse Image Search) เป็นเทคนิคหนึ่งในการอาศัยการจดจำผ่านการประมวลผลภาพถ่ายของคอมพิวเตอร์เพื่อสืบค้นภาพที่มีบริบท (Context) ที่ปรากฏในภาพต้นฉบับ แล้วแสดงผลการค้นหาที่ปรากฏภาพอื่นที่เหมือนหรือใกล้เคียง แต่การสืบค้นแบบย้อนกลับทำงานแตกต่างจากโฟโต้ดีเอ็นเอ โดยความแตกต่างอยู่ตรงที่ไม่ได้มีการแบ่งแยกภาพต้นฉบับออกเป็นตาราง แล้วทำการคำนวณความเป็นอัตลักษณ์หรือค่าแฮชของกรอบสี่เหลี่ยมหรือกริดเล็ก ๆ เหล่านั้นเสียก่อน

หากจะกล่าวให้เข้าใจโดยง่ายถึงเทคโนโลยีสืบค้นแบบโรบัสอิมเมจแฮชชิ่งว่าคืออะไร โฟโต้ดีเอ็นเอคือการอธิบายด้วยการยกตัวอย่างที่ชัดเจน ซอฟต์แวร์ถูกนำไปใช้บนหลักการทำงานสำคัญคือการระบุเจาะจงสำเนาของสื่อมัลติมีเดียที่เคยถูกระบุมาก่อนหน้า ทั้งนี้กระบวนการนี้เป็นการทำงานด้วยเครื่องจักรที่ไม่ต้องพึ่งพาการสังเกตหรือปฏิสัมพันธ์ของมนุษย์เพื่อยืนยันความถูกต้องอีกครั้ง

การทำงานของฟังก์ชันแฮชที่มีต่อข้อมูลไบนารีของไฟล์ดิจิทัลนั้นสามารถทำงานได้ดีกับการสืบค้นไฟล์ต้องสงสัยหรือที่เป็นเป้าหมายโดยคาดหวังผลในทันทีหรือแบบเรียลไทม์ (Realtime) ต่างกันกับการสืบค้นด้วยเทคนิคโฟโต้ดีเอ็นเอที่ต้องอาศัยฐานข้อมูลจากภายนอก จากนั้นจึงนำมาวิเคราะห์และเปรียบเทียบไฟล์แต่ละรายการที่ปรากฏอยู่ในก้อนสำเนาหลักฐานดิจิทัล (Forensic Image) ซึ่งต้องใช้เวลานานกว่ามาก

กรณีไฟล์ภาพที่เป็นหลักฐานได้ผ่านฟังก์ชันแฮชแล้วไม่สามารถย้อนกลับกระบวนการไปเป็นรูปภาพได้อีกครั้ง เหตุดังกล่าวถือว่าเป็นผลดีในการรักษาความสมบูรณ์ของไฟล์ให้คงเดิมโดยไม่เป็นกระบวนการสร้างความปนเปื้อนต่อพยานหลักฐานดิจิทัล อีกสิ่งหนึ่งที่สำคัญคือผลของค่าแฮชแต่ละชนิด ซึ่งถือว่าเป็นกระบวนการตรวจสอบความถูกต้องบนความแตกต่างของไฟล์มัลติมีเดียและผลคำนวณแฮชหรือแบบครอสแวลิเดชัน (Cross Validation) การดำเนินการเช่นนี้สามารถสร้างความน่าเชื่อต่อการรักษาความสมบูรณ์ของหลักฐานดิจิทัลที่ต้องถูกนำไปประกอบสำนวนการสอบสวนและการพิจารณาในชั้นศาลต่อไป นอกจากนั้น การตรวจเปรียบเทียบด้วยเครื่องคอมพิวเตอร์ย่อมก่อให้เกิดความคลาดเคลื่อนน้อยกว่าการตรวจวัดด้วยสายตาของมนุษย์ โดยเฉพาะเมื่อมีปริมาณไฟล์ภาพต้องสงสัยจำนวนมหาศาลในแต่ละคดี

บทวิจารณ์

แม้จุดเด่นของการสืบค้นสื่อมัลติเดียด้วยโฟโต้ดีเอ็นเอคือการค้นหาความคล้ายคลึงด้วยการพิจารณาเนื้อหาองค์ประกอบบนภาพซึ่งใกล้เคียงกับระบบการมองเห็นของมนุษย์ ซอฟต์แวร์นี้ยังมีข้อจำกัดในด้านการจดจำและระบุตัวบุคคลหรือสิ่งของในรูปภาพได้ กล่าวคือ การสืบค้นจะตั้งอยู่บนค่าแฮช

ที่ได้จากไฟล์ภาพต้นฉบับเท่านั้น และตัวซอฟต์แวร์ไม่ได้ถูกคิดค้นให้เรียนรู้ด้วยตนเองแบบปัญญาประดิษฐ์ เนื่องด้วยเป้าหมายแรกเริ่มของผู้พัฒนาคือ การตรวจเปรียบเทียบและจับคู่ตามความเหมือนระหว่างสองไฟล์ที่ถูกระบุก่อนหน้าและภายหลัง นอกเหนือไปจากนี้ ผลการคำนวณค่าแฮชแม้จะเป็นแบบพิเศษกล่าวคือ ได้ค่าแฮชแบบโรบัสอิมเมจที่อ้างถึงบริบทหรือเนื้อหาที่ปรากฏบนภาพ แต่หากจะดำเนินไล่เรียงกระบวนการแบบย้อนรอยไปยังจุดแรกเริ่มนั้น ซอฟต์แวร์นี้ไม่สามารถทำวิศวกรรมแบบย้อนกลับได้ หากอธิบายให้เข้าได้ง่ายยิ่งขึ้นคือ คอมพิวเตอร์ไม่อาจนำค่าแฮชที่ได้จากโฟโต้ดีเอ็นเอนำกลับมาสร้างให้เป็นรูปภาพหรือไฟล์มัลติมีเดียได้ ทั้งนี้เกิดจากโครงสร้างและองค์ประกอบของไฟล์ที่มีความแตกต่างกันอย่างสิ้นเชิง ค่าแฮชของซอฟต์แวร์นี้เป็นเพียงผลการคำนวณทางคณิตศาสตร์ที่ได้รหัสตัวเลขออกมาเท่านั้น ดังนั้น แม้จะนำค่าตัวเลขเหล่านั้นมารวมกันไม่ว่าในรูปแบบใดก็ตามก็ย่อมไม่สามารถประกอบร่างให้เป็นรูปภาพได้

ข้อเสนอแนะ

การสืบสวนเพื่อรวบรวมหลักฐานดิจิทัลในความผิดเกี่ยวกับการล่วงละเมิดทางเพศต่อเด็กที่เกิดขึ้นออนไลน์นั้น จำเป็นต้องได้รับความร่วมมือจากหน่วยงานหลายภาคส่วน ซึ่งการแบ่งปันข้อมูลถือเป็นองค์ประกอบสำคัญของความสำเร็จ สื่อลามกอนาจารที่ถูกเผยแพร่ผ่านเครือข่ายสื่อสังคมออนไลน์ไม่อาจตรวจจับด้วยวิธีการทำงานแบบดั้งเดิมที่ใช้ผู้เชี่ยวชาญที่เป็นมนุษย์มาทำดำเนินการด้วยตนเอง เทคโนโลยีปัญญาประดิษฐ์จึงได้เข้ามามีส่วนช่วยสนับสนุนการทำงานของเจ้าหน้าที่ตำรวจ เช่น การวิเคราะห์ใบหน้า โฟโต้ดีเอ็นเอ เป็นต้น

คณะทำงานปราบปรามการล่วงละเมิดทางเพศต่อเด็กทางอินเทอร์เน็ตของประเทศไทย (Thailand Internet Crimes Against Children) หรือเรียกอย่างย่อว่า ไทแคค (TICAC) เป็นชุดปฏิบัติการของสำนักงานตำรวจแห่งชาติที่ทำหน้าที่ป้องกันและปราบปรามปัญหาการเผยแพร่สื่อลามกอนาจารเด็กที่เกิดขึ้นในประเทศไทย ดังนั้น การจัดเก็บฐานข้อมูลค่าแฮชที่ได้จากการสืบสวนและวิเคราะห์หลักฐานดิจิทัลของไทแคคจึงเป็นสิ่งจำเป็นต่อการต่อสู้อาชญากรรมดิจิทัลต่อเด็กที่จะเกิดขึ้นในอนาคต การจัดทำฐานข้อมูลค่าแฮช ทั้งในแบบไบนารีและแบบโฟโต้ดีเอ็นเอเป็นสิ่งที่ต้องเร่งดำเนินการในระดับนโยบายส่วนการพัฒนาระบบฐานข้อมูลที่มารองรับการแบ่งปันบัญชีค่าแฮชรูปภาพและสื่อมัลติมีเดียต้องได้รับการคำนึงถึงความปลอดภัยของข้อมูล ตลอดจนสถาปัตยกรรมที่ถูกนำมาใช้พัฒนาต้องได้มาตรฐานสากลระดับเดียวกับที่องค์กรตำรวจสากลและศูนย์เพื่อเด็กหายและถูกแสวงหาประโยชน์แห่งชาติของสหรัฐอเมริกา ได้ดำเนินการกับฐานข้อมูลที่ตนเองดูแลควบคุมการเข้าถึงจากหน่วยงานที่เกี่ยวข้อง การเปิดโอกาสให้ภาคธุรกิจเข้ามามีส่วนร่วมเป็นสิ่งที่พึงกระทำ ซึ่งบริษัทที่มีเจตนารมณ์อย่างแรงกล้าในการช่วยเหลือเด็กและเยาวชนจากภัยทางเพศที่คุกคามผ่านเครือข่ายออนไลน์ส่วนใหญ่นั้นจะมีความยินดีในการพัฒนาต่อยอดระบบฐานข้อมูลค่าแฮชสื่อลามกอนาจารเด็กของประเทศไทยต่อไปอย่างแน่นอน

บทสรุป

การกรองและปิดกั้นข้อมูลที่มีเนื้อหารุนแรงต่อการล่วงละเมิดต่อเด็กบนเครือข่ายออนไลน์เป็นเรื่องที่ท้าทายหน่วยงานรัฐและองค์กรเอกชนผู้ให้บริการอินเทอร์เน็ต โดยผลกระทบจากการคัดกรองสื่อมัลติมีเดียที่มีการรับส่งข้อมูลผ่านบริการสื่อสังคมออนไลน์อย่างเคร่งครัดอาจส่งผลโดยตรงต่อผู้ใช้บริการ ดังนั้น ฐานข้อมูลค่าแฮชรูปภาพและสื่อมัลติมีเดียอื่นที่บ่งบอกการกระทำความรุนแรงต่อเด็กเป็นสิ่งสำคัญ บริษัทไมโครซอฟท์ร่วมกับนักวิจัยของวิทยาลัยดาร์ตมัธ ประเทศสหรัฐอเมริกาจึงร่วมกันคิดค้นอัลกอริทึมหรือขั้นตอนวิธีทางโปรแกรมคอมพิวเตอร์ในการคำนวณค่าแฮชของไฟล์รูปภาพ

ค่าแฮชนี้เป็นแบบทนทานหรือโรบัสที่มีการคำนวณผลลัพธ์จากการพิจารณาเนื้อหาที่ปรากฏอยู่บนภาพ รูปแบบนี้แตกต่างอย่างสิ้นเชิงกับการคำนวณค่าแฮชแบบดั้งเดิมกล่าวคือค่าไบนารีแฮชผลลัพธ์ที่ได้จากการค่าโรบัสแฮชทำให้การสืบค้นภาพในกระบวนการสืบสวนและแสวงหาหลักฐานดิจิทัลเป็นไปอย่างราบรื่นมากยิ่งขึ้น องค์กรภาคเอกชนสามารถนำซอฟต์แวร์โฟโต้ดีเอ็นเอที่เปิดให้บริการผ่านระบบคลาวด์คอมพิวติ้ง (Cloud Computing) ไปต่อยอดพัฒนาให้สอดรับการผลิตภัณฑ์และบริการของตนเองได้อย่างอิสระ อย่างไรก็ตาม ผู้คิดค้นโปรแกรมคาดหวังให้ฐานข้อมูลโฟโต้ดีเอ็นเอเป็นทางเลือกที่สำคัญในการตรวจวิเคราะห์ข้อมูลสื่อลามกอนาจารเด็ก

ทั้งนี้ ข้อจำกัดจากการใช้เทคนิคเดียวยังเป็นเรื่องที่หลีกเลี่ยงไม่ได้ ในกรณีนี้เช่นเดียวกันการตรวจเปรียบเทียบความคล้ายคลึงรูปภาพด้วยเทคโนโลยีที่พัฒนาให้กับโฟโต้ดีเอ็นเอไม่สามารถระบุความแม่นยำได้กับไฟล์ภาพที่เสียหายหรือมีปิดบังอำพรางเนื้อหาภาพที่เป็นส่วนใหญ่ได้ ดังนั้นในการพิจารณาเลือกใช้เทคนิคสำหรับสืบค้นไฟล์มัลติมีเดียนั้น ผู้ตรวจจำเป็นต้องพิจารณาเลือกใช้โดยคำนึงถึงความต้องการและประสิทธิภาพการทำงานของเครื่องคอมพิวเตอร์ที่ใช้ประมวลผลการ ตรวจพิสูจน์หลักฐานนั้นด้วย

เอกสารอ้างอิง

Chinpanthana, N. (2017). A Study of Low-Level Feature Extraction Techniques Used for Content-Based Image Retrieval System. **Christian University Journal**, 23(1), 130-139. (In Thai).

Farid, H. (2018). Reining in Online Abuses. **Technology & Innovation**, 19(3), 593-599. doi:10.21300 /19.3.2018.593.

INTERPOL & ECPAT. (2018). **Towards a Global Indicator on Unidentified Victims in Child Sexual Exploitation Material: Summary Report**. Bangkok, Thailand: ECPAT International.

_____. (2018). **Towards a Global Indicator on Unidentified Victims in Child Sexual Exploitation Material: Technical Report.** Bangkok, Thailand: ECPAT International.

Jusas, V., Birvinskas, D., and Gahramanov, E. (2017). Methods and Tools of Digital Triage in Forensic Context: Survey and Future Directions. **Symmetry**, 9(4), 1-20. doi: 10.3390/sym9040049.

Microsoft. (2018). **Microsoft Digital Crimes Unit**. Paper Presented at Roundtable on Best Practice for Enhancing Citizens' Digital Literacy, Taipei, Taiwan. Retrieved July 1, 2020. from http://mddb.apec.org/Documents/2018/TEL/TEL58-LSG IR/18_tel58_lsg_ir_005.pdf.

Nadeem, M. S., Franqueira, V. N., Xiaojun Zhai, Wei Ren, Wang, L., Choo, K.-K. R., and Xhafa, F. (2019). Privacy Verification of PhotoDNA Based on Machine Learning. **Security and Privacy for Big Data, Cloud Computing and Applications**. pp. 263-280. doi:/10.1049/ pbpc028e_ch12.

NetClean Technologies. (2019). **NetClean Report 2019: A Report about Child Sexual Abuse Crime.** Retrieved July 1, 2020. from https://www.netclean.com/wp-content/ uploads/sites/2/2017/06/Netclean_ report_20 19_spread.pdf.

Saetang, W. and Boonkrong, S. (2017). Effectiveness Analysis and Hash Function. **Journal of Food Health & Bioenvironmental Science**, 10(2), 81-94.

Singh, P., and Farid, H. (2019). **Robust Homomorphic Image Hashing**. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 11-18. Retrieved July 1, 2020. from https://openaccess.thecvf.com/content_ CVPRW_2019/papers/Media%20Forensics/Singh_Robust_Homomorphic_Image_Hash ing_CVPRW_ 2019_paper.pdf.

Tang, Z., Zhang, X., Li, X., and Zhang, S. (2016). Robust Image Hashing with Ring Partition and Invariant Vector Distance. **IEEE Transactions on Information Forensics and Security**, 11(1), 200–214. doi:10.1109/TIFS.2015.2485163.

Tunyasevee, W., and Witchuwanich, W. (2019). A Comparative Evaluation of Open Source and Commercial Tools for Digital Forensics**. Journal of Criminology and Forensic Science,** 5(2), 42-57. (In Thai).

ผู้เขียน

คำนำหน้า ชื่อ-สกุล	**Usanut Sangtongdee**
หน่วยงาน/สังกัด	School of Computing, Engineering and Digital Technology, Teesside University
ที่อยู่หน่วยงาน/สังกัด	Campus Heart, Southfield Rd, Middlesbrough, UK TS1 3BX
Email:	U.Sangtongdee@tees.ac.uk
คำนำหน้า ชื่อ-สกุล	**ร้อยตำรวจเอกหญิง กชกร เพิงระนัย**
หน่วยงาน/สังกัด	คณะนิติวิทยาศาสตร์ โรงเรียนนายร้อยตำรวจ
ที่อยู่หน่วยงาน/สังกัด	เลขที่ 90 หมู่ 7 ตำบลสามพราน อำเภอสามพราน จังหวัดนครปฐม 73110
Email:	Kochchakorn@rpca.ac.th
คำนำหน้า ชื่อ-สกุล	**ร้อยตำรวจเอก ภิญโญ มีเปี่ยม**
หน่วยงาน/สังกัด	คณะนิติวิทยาศาสตร์ โรงเรียนนายร้อยตำรวจ
ที่อยู่หน่วยงาน/สังกัด	เลขที่ 90 หมู่ 7 ตำบลสามพราน อำเภอสามพราน จังหวัดนครปฐม 73110
Email:	Pinyo@rpca.ac.th