

การพัฒนาระบบสร้างวิดีโอที่ได้จากปัญญาประดิษฐ์ด้วยเสียงจริงของผู้ใช้

โดย นายพงศภัค บุญประชม

นายสิรภพ เกยวิจิตร

บทคัดย่อ

วิดีโอที่สร้างจากปัญญาประดิษฐ์มักฟังเหมือนเสียงสังเคราะห์ซึ่งขาดเอกลักษณ์ของมนุษย์ โครงการนี้จึงนำเสนอการพัฒนากรอบแนวคิดและระบบต้นแบบสำหรับการนำเสียงจริงของผู้ใช้มาประยุกต์ใช้ร่วมกับวิดีโอ AI โดยอาศัยเทคนิคการประมวลผลสัญญาณระบบถูกพัฒนาขึ้นโดยใช้ Mel-Frequency Cepstral Coefficients (MFCC) สำหรับการสกัดคุณลักษณะ และใช้อัลกอริทึม Dynamic Time Warping (DTW) เพื่อซิงโครไนซ์เสียงจริงเข้ากับเฟรมวิดีโอ พร้อมทั้งพัฒนาส่วนติดต่อผู้ใช้งานผ่านเว็บ (Web-based Interface) เพื่ออำนวยความสะดวกในการทดสอบ การประเมินผลประสิทธิภาพแบ่งออกเป็น 2 ระยะ ได้แก่ การทดสอบนำร่องและการทดสอบขยายผล ผลการทดสอบนำร่องแสดงให้เห็นว่าระบบสามารถซิงโครไนซ์เสียงและปากได้ในระดับที่น่าพอใจ อย่างไรก็ตาม เมื่อขยายขอบเขตการทดสอบไปยังกลุ่มผู้พูดที่หลากหลายและรูปแบบประโยคที่ซับซ้อนขึ้นพบว่าความแม่นยำในการซิงโครไนซ์ลดลง การวิเคราะห์ผลชี้ให้เห็นว่า การประยุกต์ใช้อัลกอริทึม DTW ในรูปแบบพื้นฐานที่ดำเนินการในโครงการนี้ ยังขาดเงื่อนไขบังคับ (Constraints) ที่เพียงพอต่อการรองรับความแปรปรวนของจังหวะการพูดที่ซับซ้อน การศึกษานี้จึงสรุปเป็นข้อเสนอแนะว่า การพัฒนาในอนาคตควรพิจารณาปรับปรุงอัลกอริทึมให้มีความยืดหยุ่นยิ่งขึ้น หรือสำรวจเทคนิคการประมวลผลสัญญาณขั้นสูงรูปแบบอื่น เพื่อยกระดับความแม่นยำในการซิงโครไนซ์ให้รองรับความแปรปรวนของผู้พูดได้อย่างมีประสิทธิภาพ

Development of an AI-Generated Video System Using User's Real Voice

By Mr. Phongsaphat Boonprachom

Mr. Siraphop Ngoeivijit

ABSTRACT

Current AI-generated videos often rely on synthetic voices, lacking the individuality and emotional nuance of real human speech. This project presents the development of a framework for integrating a user's authentic voice into AI-generated video content using signal processing techniques. The proposed system employs a pipeline consisting of Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction and Dynamic Time Warping (DTW) for temporal synchronization between the authentic voice and the synthetic video frames. An end-to-end workflow was constructed, accessible via a web-based user interface. Experimental evaluations were conducted in two phases: a pilot study with controlled samples and an extended study with various speakers. The pilot results demonstrated satisfactory synchronization performance. However, in the extended evaluation involving diverse speaker