

ชื่อเรื่อง เทคนิคการลิปซิงค์อัจฉริยะเพื่อความสะดวกในการทำสื่อวิดีโอ
โดย นางสาวณัฐมล กองไธสง

บทคัดย่อ

การผลิตวิดีโอั้นหลายครั้งที่ต้องการการประสานเสียงและภาพที่มีความแม่นยำ ซึ่งทำได้ยากหากภาพหรือเสียงมีคุณภาพที่ไม่ดีพอ การลิปซิงค์ (Lip Synchronization) ได้ถูกนำมาใช้ในการแก้ไขวิดีโอ ซึ่งเป็นการซิงค์เสียงหรือภาพใหม่กับส่วนประกอบภาพเดิม อย่างไรก็ตาม การลิปซิงค์แบบธรรมดาอาจต้องใช้เวลาานาน เพื่อที่จะให้ภาพและเสียงตรงกัน บางครั้งการบันทึกวิดีโอใหม่ทั้งหมดไม่สามารถทำได้ เนื่องด้วยข้อจำกัดทางการปฏิบัติต่างๆ เช่น บางกรณีอาจจะไม่สามารถสร้างสภาพแวดล้อมเดิมได้ นอกจากนี้ยังมีบางสถานการณ์ที่ส่วนประกอบทางภาพของวิดีโอต้องถูกบันทึกใหม่ แต่เสียงที่มีนั้นคืออยู่แล้วและไม่จำเป็นต้องบันทึกใหม่ เป็นต้น

จากปัญหาดังกล่าว ดังนั้นผู้จัดทำจึงเสนออัลกอริทึมที่ใช้ในการประสานภาพปากและเสียงอัตโนมัติ โดยใช้เทคนิค Dynamic-Time Warping (DTW) ซึ่งเทคนิคใหม่นี้จะถูกใช้ในการค้นหาฟังก์ชันการจับคู่เวลาระหว่างสัญญาณเสียงเดิมและสัญญาณเสียงใหม่ เฟรมวิดีโอจะถูกจัดเรียงอัตโนมัติเพื่อให้ตรงกับเสียงที่ต้องการแม้ว่าวิดีโอใหม่จะมีเสียงที่ไม่ตรงกับตำแหน่งปากของวิดีโอตั้งต้น

ระบบดังกล่าวได้พัฒนาโดยใช้ภาษา Python และได้ทดสอบกับสถานการณ์ต่างๆ เช่น พื้นหลังที่มีเสียงรบกวนและเสียงคุณภาพต่ำ ผลการทดสอบได้แสดงให้เห็นถึงผลลัพธ์ที่ดีในการแก้ไขวิดีโอได้สะดวกขึ้น แม้จะยังมีข้อจำกัดบางอย่างของระบบนี้อยู่ ซึ่งได้อภิปรายในรายงานนี้

Thesis Title: Intelligent lip synchronization for convenient video production
By Miss Nadtamon Kongthaisong

ABSTRACT

Video production often requires precise synchronization between the visual and audio components, which can be challenging when both components are not available in good quality at the same time. Lip synchronization, commonly used in scenarios such as entertainment videos or video fixing, involves synchronizing a new audio recording to the original visual component.

However, traditional lip synchronization methods can be time-consuming, as they require manual matching of speech with the original audio to align lip movements with the speech audio. In some cases, re-recording the entire video may not be feasible due to practical constraints, such as the inability to recreate the original scenario. Additionally, there are situations where the visual component of a video needs to be re-recorded but the accompanying audio is already available and in good qualities. On the other hand, there are situations where the visual component can be used but the audio needs re-recording.

To address these challenges, we propose an automatic lip synchronization algorithm that utilizes Dynamic-Time Warping (DTW). DTW is used to find time-sequence mappings between the original and re-recorded audio signals, and the video frames are automatically rearranged to match the desired audio. The system is implemented in Python and tested with various scenarios, including noisy backgrounds and low audio quality. The results demonstrate promising outcomes in making video editing more convenient, although certain limitations are discussed in this report.