Abstract 1(2022-10-01)
Visual Dubbing Pipeline using Two-Pass Identity Transfer
Speaker : Dhyey Devendrakumar Patel

Dubbing is more popular in filmmaking to gain good quality of dialogue. In this seminar, initially speaker talks about the traditional and visual dubbing. he further discussed that traditional dubbing needs to such that it should matches with original actor. In the video example, speaker showed example of how dubbing took place where original video was spoken on French and dubbed into the English language, but the dubbing is fully synchronized with the dubbing language. Such kind of things reduce the user experience. Secondly speaker talked about the advantage of visual dubbing by comparing with traditional dubbing. He discussed that visual dubbing could correct the lip motion to match the audio also it makes easier the translator task. He further discussed about the automatic dubbing pipeline for industrial application, but it has some challenges like small dataset especially for TV advertisement, mouth expression should be same and maintain quality of original video. He further discussed and classified he visual dubbing into the parts first is audio based and expression bases. Next speaker discusses about how dubbing takes place. For example, actor A is speaking language B and dubber D dubs in language C so it should be converted to actor a speaks language C, but it consists of 6 important parameters like pose, identity, expression of mouth and face, background, and resolution (high or low).

Speaker discussed about the smoothing and skin color conversion which helps that dubbing looks more realistic. He described about how transfer network will help for visual dubbing which has set and corridor. It has shared encoder and GAN module and two dubber one for actor and one for dubber

In final stage, speaker has separately all the remaining parameter where composite of expression takes place. Now speaker described about over smoothness of the video which he needed to overcome for maintain the quality of the video. He did with the help of spatio-temporal stabilization which helped him to maintain the temporary smoothness of the video. Speaker discussed about the facial landmark to create the binary mouth mask. He had edited the face mask of original actor video and composting the expression of actor frame and synthesised expression which led to get the result.

At last, speaker had compared different method with his method like FOMM, Wav2Lip and Ours.

Where he described the head posture how position of head is different in all the videos. Also he shared the quantitative metric to measure lip sync quality (lower is better) and visual quality (results indicates lower is better). Lastly, speaker invited 28 participants to evaluate his results where he showed traditional and visual dubbing and asked three questions to participants. 86% people found that lips are synchronized with the audio, 61 to 80 says visual quality is good. Some

Limitation are double chin issue, inconsistent nasolabial fold, and improper intensity of expression.

Speaker discussed that in future enhancement he will be going to solve the double chin issue and additional control channel for expression intensity.