Overview of Audio & Video Synchronization

A. Thenmozhi[1] Dr. P. Kannan[2]
[1]PG Scholar [2]Professor & Head of Department
[1,2]Department of Electronics & Communication Engineering
[1,2]Anna University, Chennai, India

Abstract— The audio and video synchronization is extremely necessary. The synchronization loss between image and sound continues to disturb observers and irritate telecasters. The demand is to assure synchronization without adjusting content at the same time as still remaining price low. The objective of the synchronization is to line up both the audio and video signals that are processed individually. The multimedia means the combination of various forms of information such as text, speech, music, images, graphics and video. For many audio-visual applications, the integration and synchronization of audio and video signals is essential. This paper describes the recent techniques in exploiting the audio-visual synchronization that is very significant in multimedia communication. The applications include lip synchronization, joint audio-video coding, and person verification.

Key words: Synchronization, Integration, Multimedia, Lip Synchronization, Audio-Video Coding

I. INTRODUCTION

Audio-to-video synchronization (also known as audio-video sync, audio/video sync, lip sync, or by the lack of it: lip sync error, lip flap) refers to the relative timing of the audio (sound) and the video (image) parts during creation, post production, transmission, reception and play-back processing. When the sound and video have a timing related cause and effect, AV-sync can be an issue in television, videoconferencing or film. An essential aspect of any multimedia presentation is the way time-dependent and time-independent media objects, which make up the presentation, are arranged in time. In this text the term synchronization is used to refer on a general level to the specification and enforcement of temporal relations between media objects.

The transmission of multimedia over packet networks and the use of independent systems for audio and video, means that listeners often experience a time lag between a remote user's audible words, and the associated lip movements. The more quality demanding applications of multimedia conferencing, such as remote language teaching, require both lip synchronization and good quality audio in order to be effective. Lip synchronization can be provided by artificially delaying either the audio or the video output, so that words and sounds are matched in time. With the development of multimedia and network technology, various multimedia services such as video on demand, video conference, distance learning is in great demand. In these multimedia applications, video and audio are stored, transmitted and presented. During the presentation time, the temporal relationship between audio and video have to be preserved in order to offer the best perceptual quality.

II. PROPOSED METHODOLOGY

The proposed framework is automatically measuring and maintaining the perfect synchronization between audio and video using audio-visual signatures. Figure 1 shows the proposed framework for audio and video synchronization based on audio-visual signatures. The framework based on signatures extracted from audio and video streams for automatically measuring and maintaining synchronization between the two streams. The audio signature is based on projections of a coarse representation of the spectrogram onto random vectors. The video signature is based on projections of a coarse representation of the difference image between two consecutive frames onto random vectors. The time alignment present at the signature generator between the two streams is recorded by combining audio and video signatures into a combined synchronization signature. At the detector after video and audio streams go through different processing operations, we extract the signatures again. The signatures extracted before and after processing from the audio and the video are compared independently using a Hamming distance based correlator to estimate the relative misalignment introduced due to processing in each of the streams. Then, the estimated relative misalignment between the audio and video streams is used to preserve the same alignment between the streams that was present before processing.

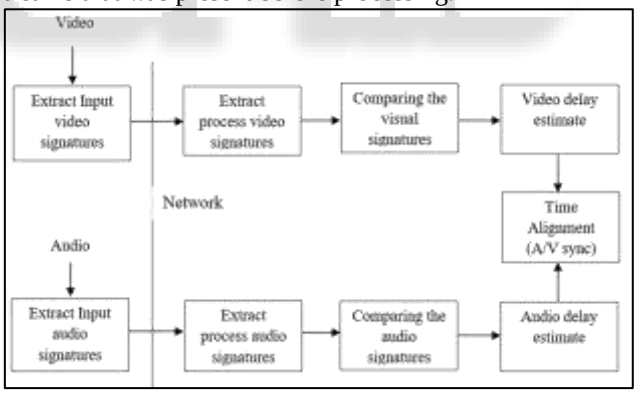

Fig. 1: Audio & Video Sync using Audio-Visual Signatures

III. RESULTS & DISCUSSIONS

The test content used for the performance assessment of the system consisted of 5 seconds A/V clips.

The input video is divided into chunks for generating the audio signature. Fig. 3.1, shows the audio signature generation using spectrograph. The input video is divided into frames for generating the video signature. Fig. 3.2, shows the video signature generation.

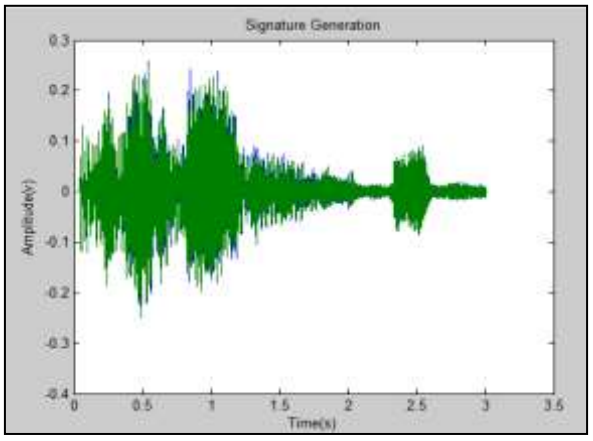

Fig. 3.1: The Audio Signature Extraction

Fig. 3.2: The Video Signature Generation

Fig. 3.3, shows the relative time alignment between the audio and video stream. It decodes the corresponding video frame that is given as input with proper time alignment between the input and processed video frames. Fig. 3.4, shows the decoded input video frame.

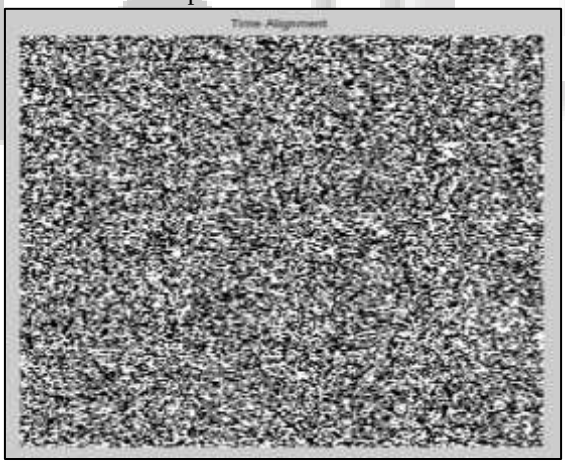

Fig. 3.5: The Audio-Video Stream Time Alignment

Fig. 3.4: The Decoded Video Frame

Fig. 3.5: The A/V Synchronization

Fig.10, shows the audio and video synchronization using signature. The A/V sync using signatures provides perfect synchronization between the corresponding audio and video streams.

IV. CONCLUSION

Thus the audio and video synchronization using signature methodology was implemented and their performances were analyzed sufficiently and appropriately. The proposed system would automatically estimate and preserve the perfect synchronization between the audio and video streams and it would maintain the perceptual quality of audio and video. It was shown the guarantee and quite simple process suitable for the real world multimedia application and offline applications.

REFERENCES

[1] Alka Jindal, Sucharu Aggarwal, "Comprehensive overview of various lip synchronization techniques" IEEE International transaction on Biometrics and Security technologies, 2008.
[2] Anitha Sheela.k, Balakrishna Gudla, Srinivasa Rao Chalamala, Yegnanarayana.B, "Improved lip contour extraction for visual speech recognition" IEEE International transaction on Consumer Electronics,pp.459-462, 2015.
[3] N. J. Bryan, G. J. Mysore and P. Smaragdis, "Clustering and synchronizing multicamera video via landmark cross-correlation," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2012, pp. 2389–2392.
[4] Claus Bauer, Kent Terry, Regunathan Radhakrishnan, "Audio and video signature for synchronization" IEEE International conference on Multimedia and Exposition Community (ICME), pp.1549-1552, 2008.
[5] N. Dave, N. M. Patel. "Phoneme and Viseme based Approach for Lip Synchronization.", International Journal of Signal Processing, Image Processing and Pattern Recognition, pp. 385-394, 2014.
[6] Dragan Sekulovski, Hans Weda, Mauro Barbieri and Prarthana Shrestha, "Synchronization of Multiple Camera Videos Using Audio-Visual Features," in IEEE Transactions On Multimedia, Vol. 12, No. 1, January 2010.
[7] Fumei Liu, Wenliang, Zeliang Zhang, "Review of the visual feature extraction research" IEEE 5th International Conference on software Engineering and Service Science, pp.449-452, 2014.

[8] Josef Chalaupka, Nguyen Thein Chuong, "Visual feature extraction for isolated word visual only speech recognition of Vietnamese" IEEE 36th International conference on Telecommunication and signal processing (TSP), pp.459-463, 2013.
[9] K. Kumar, V. Libal, E. Marcheret, J. Navratil, G.Potamianos and G. Ramaswamy, "Audio-Visual speech synchronization detection using a bimodal linear prediction model". in Computer Vision and Pattern Recognition Workshops, 2009, p. 54.
[10] Laszlo Boszormenyi, Mario Guggenberger, Mathias Lux, "Audio Align-synchronization of A/V streams based on audio data" IEEE International journal on Multimedia, pp.382-383, 2012.

