Developing a Sound-Based Method to Synchronize Multiple Videos Recorded by Multiple Sound Sources

Davaanyam Jargal
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Rena Kato
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Tomoki Taniguchi
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Kosei Shibata
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Takahiro Koga
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Obada Al Aama
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: jargal.davaanyam711@mail.kyutech.jp, waga@brain.kyutech.ac.jp

Abstract

The surround environmental monitoring from a vehicle on the road is important for the driving behavior analysis as well as the monitoring of driver's operations. A light weight mobile camera is useful for the record of multiple directions from the vehicle simultaneously. However, the timing synchronization is an important issue need to be solved. In this purpose, we proposed the sound-based method to synchronize different videos recorded with environmental sounds. In this task, the extraction of common sound features and amplifying of the features are necessary to superimpose those sound profiles to find consisting time points. In the validation the effectiveness, we used recoded videos from the bus driven by an expert driver.

Keywords: Wavelet, Bandpass filter, Gaussian filter, Autocorrelation, Synchronize

1. Introduction

Synchronizing video from multiple cameras is essential for various applications, from video surveillance to scientific research, where time alignment can dramatically influence data analysis and interpretation. Many traditional techniques utilize hardware-based synchronization via either GPS or timestamping, which can be costly and vulnerable to environmental disruptions. Misaligned timing the analysis of events and makes correlation and integration of video and sensor data difficult (e.g. to analyze driver behavior and for object detection). While video capture has become ubiquitous through portable devices, synchronizing footage from multiple unlinked perspectives remains a challenge. In challenging multi-video synchronization, Wu et al. [1] proposed a deep learning framework which merges pose-matching with temporal encoding to align the video streams. Although their method uses visual information in an innovative way, it requires a lot of technical power to perform real-time calculations. Brassarote et al. [2] have shown the benefit of using non-decimated wavelet transforms (NDWT) for performing shift-invariant analysis, which is a method that is attractive for non-stationary signals. Wavelet-based approaches are computationally inexpensive and conserves important aspects of the signal [3]. Calibration of the sound signal reinforces the frequency features with sound peaks or sustainable components, which is useful for synchronization of the camera seen in the video and audio signal [4]. In Band-pass filters play an integral role in extracting audio signals within a precise frequency range if the target band is clear, while discarding irrelevant background noise falling outside the targeted band. These versatile filters see widespread use in signal processing applications to refine data quality, as evidenced in mechanical systems designed for vibration examination and ecological surveillance [5]. The renowned Gaussian smoothing filter ensures the removal of high-frequency

interference while safeguarding key characteristics of the acoustic signals. Recursively implemented Gaussian feature enhancement methods, as pioneered by Young and Van Vliet, offer computational efficiency and accuracy, allowing for rapid filtering without compromising functionality [6]. The objective of auto-correlation analysis is described as a mathematical way of pattern recognition. For instance, it helps to find a periodic signal possibly hidden in noise or a fundamental frequency which is contained in its harmonics [7] [8].

Testing the proposed approach on video recordings obtained using RGB cameras placed on a bus used by an expert driver. We demonstrated the effectiveness of our proposed method using sample videos attached with a moving vehicle in real traffic environments. With results showing how precise our timing synchronization capability is, this has the potential for much better video based environmental monitoring and analysis of that data. Accurate monitoring of the environment and synchronization of multiple streams of video has become a significant source of demand in some fields such as autonomous driving, surveillance, and monitoring of traffic situations.

2. Methodology

2.1. Processing sequence.

We used an RGB camera with a sound recorder inside. There is sound in the input video, with a high reliability and rich availability. Thus, audio signals can be used for the supportive information to analyze. We placed the cameras on the bus as shown in Fig. 1. Then, a lot of noise were contaminated in the raw audio signals. Cameras are manually started individually, therefore individual onset timings of sound signals started differently, which are not synchronized automatically. The goal of the present study is to process raw audio signals and use them to identify and synchronize the sound that is heard when a vehicle starts moving, recorded in video.

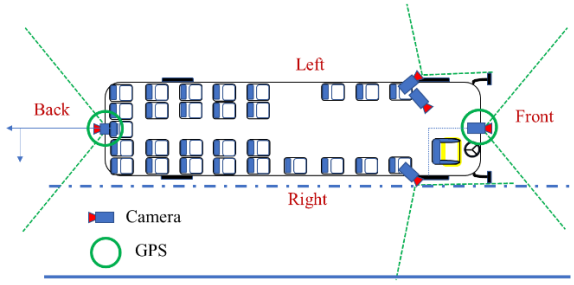

Fig. 1 Processing sequence.

First, we separated the video and audio contained in the data. Second, the relevant frequency components are identified, utilizing the magnitude scalogram obtained via the continuous wavelet transform (CWT). Then, the identified frequency band is retained and the rest is filtered out. Then Gaussian filtering was applied to smooth the signal and to reduce high-frequency noise even further. Finally, autocorrelation can detect and synchronize similar patterns in the signal. An outline of the processing sequence and the theories where applied is shown Fig. 2.

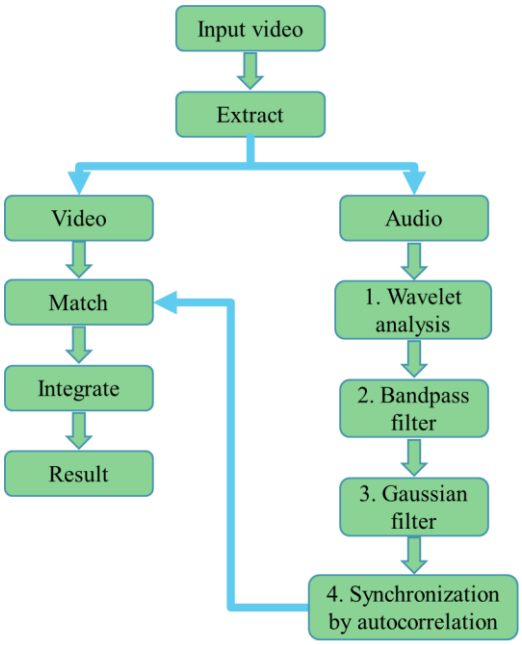

Fig. 2 Processing sequence.

There is a 12-minute (758 s) audio signal extracted from the video as in shown Fig. 3.

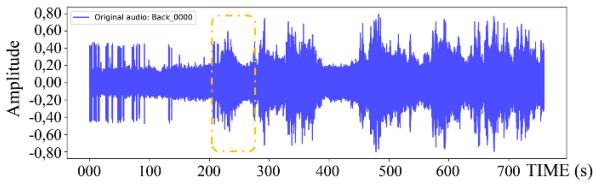

Fig. 3 Input audio signal

The bus starts moving about 3 minutes. At this point, a warning beep sounds, making it easier to synchronize the images of cameras that started working at different times. The sound of a bus starting to move, as shown in Fig. 4. However, no obvious pattern is observed in this raw signal.

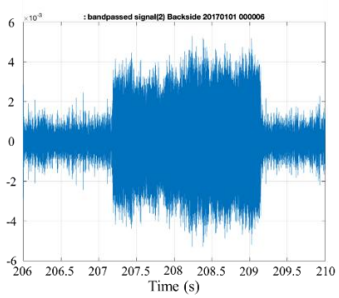

Fig. 4 The section containing the beep signal is cut off.

2.2. Continuous Wavelet Transform Method

Wavelet-based approaches are computationally inexpensive and conserves important aspects of the signal. Moreover, only complex wavelet can provide real and imaginary parts to obtain amplitude and phase information from time series. The Morse wavelet is expressed as follows in Eq. 1.

$$\psi_{P,\gamma}(\omega) = U(\omega) a_{\beta,\gamma} \omega^{\beta} \cdot e^{-\omega^{\gamma}} \quad (1)$$

Continuous Wavelet Transform (CWT) breaks down the signal into time-frequency components, which allows for identifying distinct feature differences of audio streams equation as Eq. 2. Morse wavelet has a good balance between the localization of time and frequency.

$$Wf(u,s) = \frac{1}{2} Wf_a \cdot (u,s) \quad (2)$$

2.3. Application of Bandpass Filters in Audio Signal Processing.

The audio signal has a complex structure, and at this stage, the amount of noise and unnecessary signals is reduced, preventing the ingress of low-frequency noise and unwanted high-frequency noise, allowing the next stage to work with an attractive sound quality. The band pass filter is defined as follows and is shown in Eq. 3 [4] [5].

$$h(t) = 2f_h sin(2f_h t) - 2f_l sin(2f_l t) \quad (3)$$

The real signal can be processed as shown in Eq. 4.

$$y(t) = x(t) \cdot h(t) \quad (4)$$

The expanded formula is shown in Eq. 5.

$$y(t) = x(t) \cdot 2f_h sin(2f_h t) - 2f_l sin(2f_l t) \quad (5)$$

2.4. Application of Gaussian Filters in Audio Signal Processing.

The bandpass-filtered signal may still contain residual noise, particularly at higher frequencies. Gaussian filtering smoothens the signal while retaining its essential features. The Gaussian filter kernel is defined as in shown Eq. 6 [5].

$$G(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(t-t_i)^2}{2\sigma^2}\right) \quad (6)$$

where $\boldsymbol{t}$ is the time from the origin in the horizontal axis, and σ is the standard deviation of the Gaussian distribution. The filtered signal is obtained by convolving the original signal with the Gaussian kernel defined as in shown Eq.7.

$$Z_{filt}(t_i) = \sum_{j-1}^{N} w_j \, z_j \exp\left(-\frac{(t-t_i)^2}{2\sigma^2}\right) \quad (7)$$

Gaussian smoothing effectively reduces noise while avoiding the sharp attenuation introduced by other types of filters.

2.5. Signal Synchronization Using Autocorrelation

Auto-correlation, otherwise called serial correlation in the discrete time setting, defines the connection between a signal and its lagged version in terms of time. Simply put, it measures the degree of association between two points of a time series on the basis of their distance in time. The objective of auto-correlation analysis is described as a mathematical way of pattern recognition. For instance, it helps to find a periodic signal possibly hidden in noise or a fundamental frequency which is contained in its harmonics. This method finds application in the processing of signals and time-domain signals among other applications. The autocorrelation function as in shown Eq. 8 [6]. $\boldsymbol{X(t)}$ is the value (or realization) produced by a given process at time t. Suppose that the process has mean $\boldsymbol{\mu_t}$ and variance $\boldsymbol{\sigma_t^2}$ at time $\boldsymbol{t}$, for each $\boldsymbol{t}$. Then the definition of the autocorrelation function between times $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$

$$R_{XX}(t_1,t_2) = E\left[X_{t_1}\overline{X_{t_2}}\right] \quad (8)$$

where $\boldsymbol{E}$ is the expected value operator and the bar represents complex conjugation. Subtracting the mean before multiplication yields the auto-covariance function between times $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ as shown in Eq. 9, Eq. 10.

$$R_{XX}(t_1,t_2) = E\left[\left(X_{t_1} - \mu_{t_1}\right)\overline{\left(X_{t_2} - \mu_{t_2}\right)}\right] \quad (9)$$

$$R_{XX}(t_1,t_2) = E\left[X_{t_1}\overline{X_{t_2}}\right] - \mu_{t_1}\overline{\mu_{t_2}} \quad (10)$$

3. Results and Discussion

3.1. Wavelet Analysis

In CWT analysis provides a detailed representation of the time-frequency distribution of an audio signal. The wavelet transform is illustrated in Fig. 5 as a scalogram. The magnitude scalogram represents the energy distributed over time and frequency, clearly in a semblance. Looking at the results of the first wavelet processing, our target sound does not look very good.

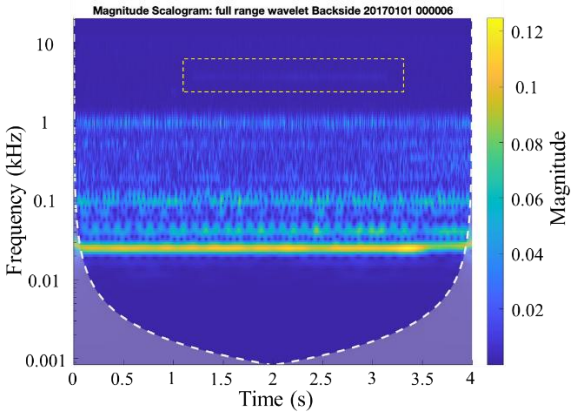

Fig. 5 Results using Wavelet analysis

3.2. Bandpass Filter

The band-pass filter effectively removes the low and high frequency components that interfere with further processing of the signal. This filter has been used in many experimental studies and has already been shown to give good results.

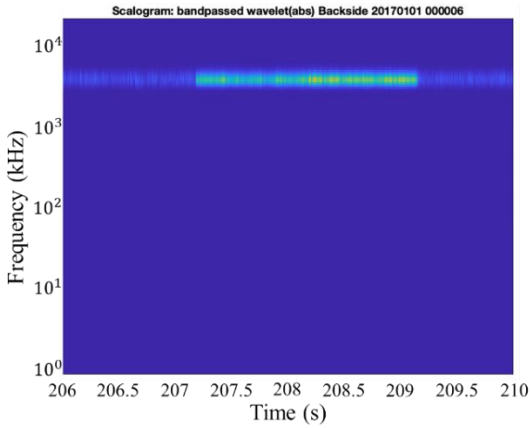

Fig. 6 Results using Bandpass filter

In the signal frequency components from 3.7k Hz to 4k Hz, those were judged to be relevant for the study and were retained by bandpass filtering as in shown Fig. 6. Unwanted low-frequency and high-frequency noise was effectively eliminated. The filtered signal retained most of the relevant components of the sound signal. But even now, this signal is still affected by noise.

3.3. Gaussian filter

The Gaussian filter effectively filters out noise without losing the characteristics of the signal, also it can also correct a signal that has been altered to some extent by the effects of noise. This employed a Gaussian filter for smoothing out any residual noise present in the bandpass-filtered signal. The standard deviation was adjusted thereby attaining smoothness with minimal distortion, shown in Fig. 7.

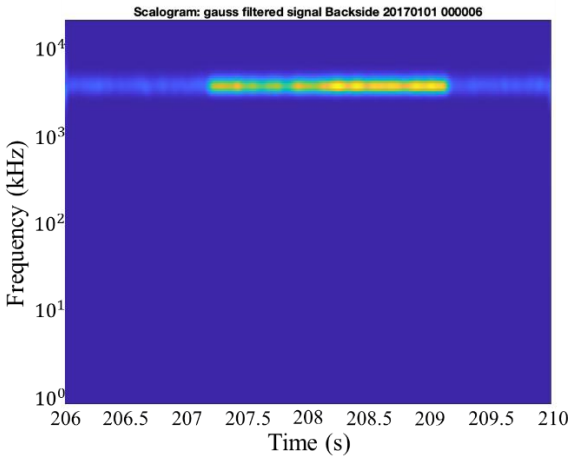

Fig. 7 Results using Gaussian filter

We have performed a wavelet transformation of the filtered signal. The signal was then reconstructed using the inverse wavelet.

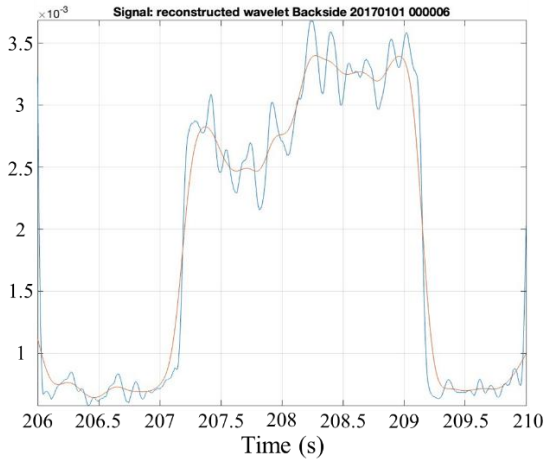

Fig. 8 Results using Reconstruct and Gaussian filter

This reconstructed signal was processed by Gaussian filtering, as shown in Fig. 8. Reconstruction and Gaussian filtering make the signal clearer and improve its characteristics.

3.4. Signal Synchronization Using Autocorrelation

We have to synchronize the signals. To do this, we use autocorrelation. Autocorrelation is useful to detect signal coherence. We used the autocorrelation function on this filtered data. Then we can determine the time of such synchronization. Finally, we have achieved our goal and the time synchronization as shown in Fig. 9.

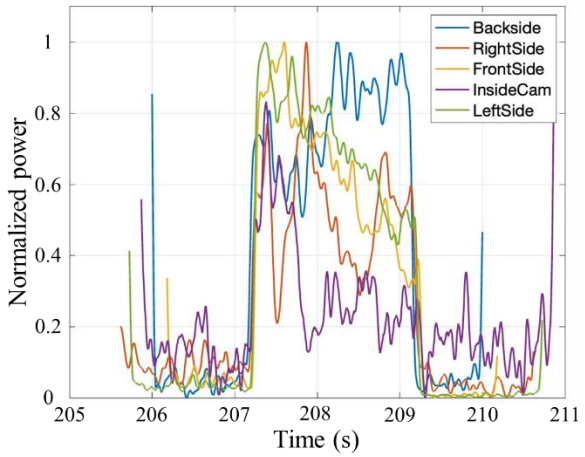

Fig. 9 The synchronized results

We had five cameras inside the bus. All of these cameras were started at different times. But each of these cameras recorded the sound inside the bus. So I was able to recognize the sound signal that the bus made when it started moving and use it for synchronization.

4. Conclusion

In this paper, a method for synchronizing multi-camera videos based on audio signals is proposed. This method uses wavelet transform, bandpass filtering, Gaussian noise reduction and autocorrelation function to achieve synchronization. Wavelet analysis enables frequency and amplitude analysis without losing the exact time scale. The important information of the frequency components is recorded in the magnitude scalogram. Bandpass filtering removes undesirable section of the signal, while Gaussian smoothing reduces noise. Wavelet transforming the Gaussian-filtered signal and reconstructing it further reveals the characteristics and shape of the signal. Finally, the use of autocorrelation enables very accurate synchronization.

In future work will be focused on real-time implementation as well as combining this method with previous multi-camera systems to improve operation in dynamic environments.

Acknowledgements

This work was supported in part by JSPS KAKENHI (JP17H06383, JP24K07387) and Kitakyushu Foundation for the Advancement of Industry, Science and Technology (FAIS). Authors also deeply appreciate expert drivers and safety operation managers in Nishitetsu group for offering a precious opportunity to examine professional procedures for the safety assessments in the real driving on the road.

References

1. Wu, Xinyi, et al. "Multi-video temporal synchronization by matching pose features of shared moving subjects." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019, pp. 0-0.
2. Brassarote, Gabriela de Oliveira Nascimento, E. M. Souza, and João Francisco Galera Monico. "Non-decimated wavelet transform for a shift-invariant analysis." TEMA (São Carlos) 19.1, 2018, pp. 93-110.
3. Singh, Balbir, and Hiroaki Wagatsuma. "Two-stage wavelet shrinkage and EEG-EOG signal contamination model to realize quantitative validations for the artifact removal from multiresource biosignals." Biomedical Signal Processing and Control 47, 2019, pp. 96-114.
4. Barik, B., Kalirasu, A., & Prathap Kumar, A. V. High efficient band pass filter design and analysis on impact of resonator on its performance. Journal of Shanghai Jiaotong University, 17(3), 2021, pp.180–190.
5. Shahruz, S. M. Design of mechanical band-pass filters for energy scavenging. Journal of Sound and Vibration, 292(4–5), 2006, pp. 987–998.
6. Young, I. T., & Van Vliet, L. J. Recursive implementation of the Gaussian filter. Signal Processing, 44(2), 1995, pp. 139–151.
7. Reynolds, K. M., and L. V. Madden. "Analysis of epidemics using spatio-temporal autocorrelation." Phytopathology 78.2 1988, pp. 240-246.
8. Gubner, John A. Probability and Random Processes for Electrical and Computer Engineers. Cambridge University Press. 2006

Authors Introduction

Mr. Davaanyam Jargal

He received his Master's degree from the School of Mechanical Engineering and Transportation, Mongolian University of Science and Technology, Mongolia in 2010. He is currently a Doctoral course student in Kyushu Institute of Technology, Japan

Dr. Obada Al Aama

He graduated from the Department of Communication and Electronics Engineering at Al-Baath University, Syria, in 2013. He obtained his M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2019 and 2024, respectively. He is currently a researcher in the Department of Human Intelligence Systems at the Kyushu Institute of Technology, Japan. His research interests focus on robotics and autonomous driving.

Takahiro Koga

He received his Master's degree in Engineering from the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology in Japan. He is currently a researcher in Kyushu Institute of Technology.

Tomoki Taniguchi

He received his Bachelor's degree in Computer Science and Systems Engineering from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan, in 2024. He is currently a master's student at Kyushu Institute of Technology, Japan.

Ms. Rena Kato

She received her associate's degree from the Department of Electrical and Electronic Systems Engineering, National Institute of Technology, Toyota College, Japan in 2024. She is currently enrolled in the Department of Systems Design and Informatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan.

Hakaru Tamukoh

He received a B.Eng. degree from Miyazaki University, Japan 2001. He received his M.Eng and Ph.D. from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of the 21st-century Center of Excellent Program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor at Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He was an associate professor from February 2013 to March 2021 and is currently a professor at the Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interests include hardware/software complex systems, digital hardware design, neural networks, soft computing, and computing service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI, and RSJ.

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. In 2009, he joined Kyushu Institute of Technology, where he is currently an Associate Professor of the Department of Human Intelli- gence Systems. His research interests include non- linear dynamics and robotics. He is a member of IEEE.