DOI: 10.61782/fa.2025.0158

FORUM ACUSTICUM EURONOISE 2025

AUDIO-VIDEO ACQUISITION SYSTEM FOR DYNAMIC ROAD PAVEMENT MONITORING

Federica Petrongolo[1] **Laura Fiorella**[1] **Gonzalo de León**[2]
Domenico Profumo[3] **Francesco Artuso**[4*] **Alessandro Monticelli**[4]
Francesco D'Alessandro[3] **Elena Ascari**[1] **Gaetano Licitra**[1] **Luca Fredianelli**[1]
[1] Institute for Chemical-Physical Processes of the Italian Research Council (CNR-IPCF), Pisa, Italy
[2] University of Pisa, Department of Earth Sciences, Via Santa Maria, 53, 56126 Pisa, Italy
[3] iPOOL S.r.l., Via Antonio Cocchi, 3, 56121 Pisa, Italy
[4] University of Pisa, Physics Department, Largo Bruno Pontecorvo, 3, 56127 Pisa, Italy

ABSTRACT

Pavement quality plays a crucial role in ride comfort, security, and vehicle pollution emissions, including tyre rolling noise. In addition, tempestive evaluation of the pavement conditions and localization of distresses can reduce the overall environmental impact of road traffic transport. These measures enhance road maintenance planning, extend pavement lifetime and reduce cost related effects. In order to have a deeper understanding of the road condition, a multimodal-system can be useful to extract diverse features across different domains. However, achieving synchronous connectivity between various devices, both online and offline, can be challenging. In this work, a new dynamic multimodal-system is presented. The system is based on the integration of a Tyre Cavity Microphone and a video recording device, enabling a road surface characterization both in terms of visual and acoustical inspection. Key challenges, such as alignment and spatial positioning, are tackled. Special attention is given to the synchronization between the multiple measuring devices. Additionally, the intrinsic constraints of each sensor are described. The main goal of the ongoing project is to create a system that can localize and assess pavements' state during normal traffic conditions.

**Corresponding author*: *francesco.artuso@phd.unipi.it.*

Keywords: *tyre cavity noise, road damage detection, image analysis, multimodal-system, road traffic noise managment*

1. INTRODUCTION

The condition of road surfaces directly or indirectly affects various aspects of human well-being, including driving comfort, safety and noise pollution [1–11]. Thus, a programmed monitoring is of paramount importance for the effective maintenance of roads and for the assurance of optimal driving conditions [12]. Common monitoring techniques [13] are often characterized by their intrusiveness. This is exemplified by the installation of sensors within the pavement, necessitating a destructive approach and a frequently challenging implementation [14–16]. Conversely, non-intrusive techniques employ instruments and sensors on a mobile laboratory, including image analysis [17–19], Ground Penetrating Radar [20], or acoustic methods for tyre-road noise measurement [21]. Traditional methods are therefore costly, time-consuming and can disrupt traffic flow. Consequently, recent research has focused on developing efficient monitoring techniques exploiting machine learning and neural networks, which can mainly be grouped into Tyre Cavity Noise (TCN) analysis [22, 23] and image analysis [24].

TCN method is defined as the acoustic signals captured by a microphone placed within the tyre cavity of a mobile lab, which have the capacity to indicate interactions between the rolling tyre and road anomalies. The data are acquired during normal driving conditions without disrupting traffic, and are subsequently analysed. The analysis involves the extraction of features derived from temporal and spectral characteristics, such as peak amplitude, kurtosis, spectral centroid, and power spectral density. These features are processed using clustering techniques or by employing a neural network to classify the pavement [23, 25]. The TCN analysis method is subject to certain limitations. For instance, the information is referred only to the surface trodden by the wheel and the audio signal from two distinct damage sites may be similar or different. External factors (e.g. tyre hardness, air temperature) may also influence the TCN. A factor that must be considered when using the TCN method is the induced resonance phenomenon that arises when toroidal viscoelastic material comes into contact with a rigid surface [26]. Moreover, the position of the microphone significantly impacts the measurement of the system's response, as the acoustic pressure inside the tyre is not equal in all points and the microphone follows the rotation of the tyre [27–30].

The method based on the analysis of video or image data is undergoing rapid development within this field. It facilitates the automatic and unattended detection of defects in pavements through the use of artificial intelligence. Such systems have demonstrated an ability to precisely localize potholes, cracks, and severities [24,31–33]. However, the limits of this method lie in the fact that the information extracted from the video camera only yields geometrical information coming from an illuminated scene, therefore the atmospheric conditions could influence the obtained results. Furthermore, the perspective of the camera can result in significant variations in the appearance of road distresses [34].

With a view to evaluating road deterioration for the reduction of noise exposure, the present work is intended to combine both audio and video approaches for effective road condition monitoring. This goal is achieved by integrating image-based object detection systems with TCN measurements, thus circumventing the limitations inherent in each individual method. The paper discusses the main challenges encountered in obtaining a multimodal system, mainly due to connection, communication, data synchronisation or misalignment [35, 36].

2. MATERIALS

The system for the monitoring of road pavement condition consists of a standard vehicle equipped with various sensors. The sensors employed in the present work for damages detection comprise both an audio and a video device. The audio is recorded through a Tyre Cavity Microphone (TCM), which is an analog high dynamic range micro-electro-mechanical system (MEMS) microphone mounted on a printed circuit board (PCB). The PCB is placed inside the tyre cavity, fixed to the rim and it is connected to a microcontroller, the wires connecting them pass through a valve. The microcontroller serves for data acquisition, it is placed within a specific case mounted outside, on the wheel centre cup, as illustrated in Fig. 1. The video is captured by an action camera mounted at the rear of the vehicle via an appropriate case and adhesive mount. The camera is placed in such a manner as to record the road behind the vehicle. The system is further comprised of two components: a GPS module, which is mounted on the vehicle's roof, and a processing unit, which is located within the vehicle. The vehicle equipped with the sensors appears as shown in Fig. 2.

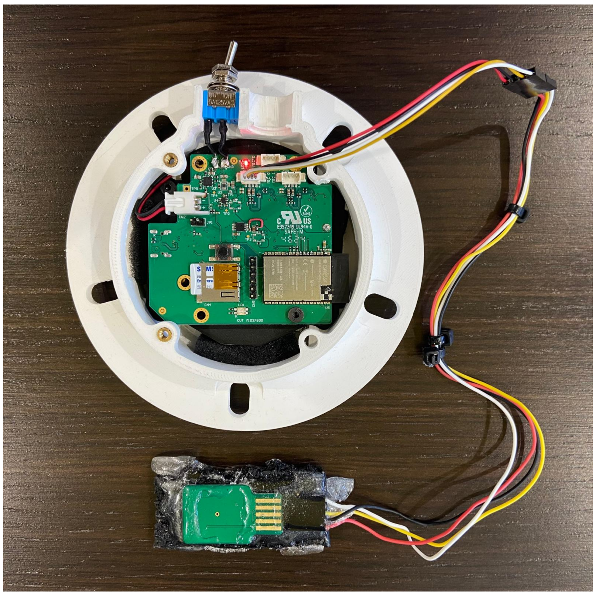

Figure 1. The Tyre Cavity Microphone connected to the microcontroller.

3. CHALLENGES IN A MULTIMODAL SYSTEM

The challenges experienced are attributable to both hardware and software origins.

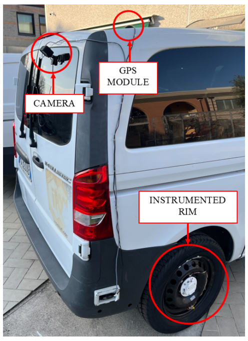

Figure 2. The equipped vehicle.

3.1 Hardware

The primary challenge in implementing a system comprising multiple sensors is to ensure the establishment of an effective connection and communication infrastructure. The processing unit occupies a central role in this system, as it is responsible for communicating with the devices, receiving the acquired data, and processing it. In this regard, the initial priority is to ensure optimal connectivity for the effective functioning of the system. The TCM is connected to the microcontroller that provides a WiFi connection with the computer on board. Similarly, the camera is connected to the processing unit in WiFi thanks to its internal network. In order to manage two wireless connections simultaneously, the computer is equipped with an additional WiFi adapter operating at 300 Mbps. Finally, the GPS module is connected via USB to the processing unit. The connection system is shown in Fig. 3.

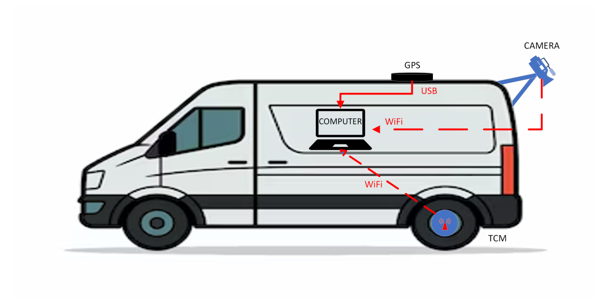

Figure 3. The connection system.

The second hardware obstacle concerns the necessity of maintaining a sufficient acquisition frequency, which varies between the different sensors. Consequently, the number of acquisitions is not uniform across all sensors, thereby not allowing immediate matching. A bottleneck is the low GPS frequency at 1 Hz. An example to illustrate the disparity in data sampling is to consider that the vehicle travelling at 50 km/h covers 14 metres in 1 second, but GPS would only provide 1 position for the entire segment. Spatial accuracy can be improved with a linear fitting of latitude and longitude separately, thereby providing an approximation of the travelled path, as illustrated in Fig. 4. This methodology enables the acquisition of the coordinates, whether real or resampled, of the initial and final points of a fixed distance. The method is quite effective, but when combined with the uncertainty associated with the GPS position, errors can occur in the case of a very curved path (such as the upper roundabout in the figure).

Figure 4. An example of linear fitting of GPS data.

The action camera, during the acquisition, records the video at 120 Frames Per Second (FPS) and, in the meanwhile, collects metadata, including timestamps and GPS coordinates. However, the frequency of the metadata acquisition is limited to 18 Hz, meaning that location information is not available for every frame. To address this lack of data, a linear interpolation method is used, enabling georeferencing of every frame of the video.

Finally, the TCM allows to set the frequency of acquisition and it is on the order of magnitude of the kHz, thus ensuring the sampling of the signal of interest.

3.2 Software

The use of multiple devices also necessitates the management of different types of data, each of which poses its own unique challenge. The video results are too expensive in terms of both memory and duration to be analysed in their entirety; therefore, a solution may be found in the extraction of frames. However, this approach does not guarantee a solution to the issue. The video is recorded with a resolution of 1080p and a frame rate of 120 FPS to ensure optimal resolution during movement; at a speed of 30 km/h, the analysis of 120 frames covers a distance of only 8 meters. In addition, there is significant overlap between adjacent frames, leading to repeated analysis of the same segment of the road. Ideally, the frames would be selected to provide full coverage of the road pavement without overlap. This issue has been addressed by computing the vehicle speed from the GPS coordinates, the time passed, and the visual coverage of the camera frame. This approach reduces the overlap to a few centimeters, avoiding the analysis of the same portion of road several times, and significantly reduces the number of frames. Assuming a visual coverage of 5 meters and a speed of 30 km/h, a total of two frames is sufficient to cover the entire distance in one second: the number of frames is decreased by a factor of 60. It is advantageous to select only the significant video frames, as this reduces both the number of images and the required memory occupancy. Furthermore, reducing the number of collected frames also accelerates the extraction process (see Tab. 1). In Fig. 5 it is shown how frames extraction based on the speed and the distance covered by the framing lead to a small overlap between subsequent frames. With a complete extraction, there are fifty intermediate frames between the two of Fig. 5, with a small shift between each other (see Fig. 6). This results to be a redundant information.

Figure 5. Two subsequent frames extracted basing on the speed and the visual coverage.

Finally, it is imperative to ensure the synchronised operation of all the sensors. The occurrence of asynchrony has consequences such as the misalignment of recordings and the impossibility of an effective correlation. This compromises the reliability of the system and invalidates the use of a multimodal system to yield a more accurate analysis.

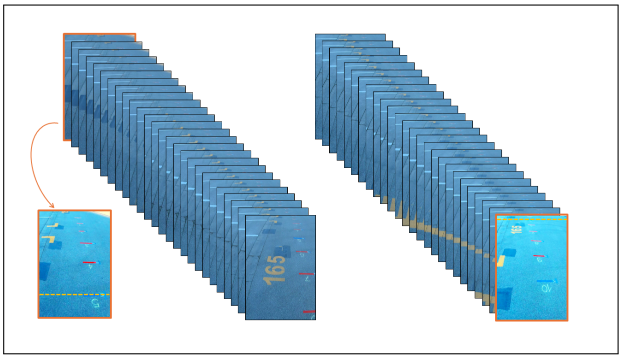

Figure 6. Oversampling of video frames.

Table 1. Comparison between extracting all the frames and extracting them basing on speed for a video of 5 seconds.

	All frames	Extracted frames
Number of frames	600	10
Size [GB]	1.83	0.294
Time [s]	178.13	3.46

In order to establish a synchronous system, it is necessary to acquire data concurrently. This can be facilitated by utilising a computer that is integrated into the system and connected to three sensors. The computer serves as the primary controller, starting the recording of all sensors simultaneously. Post-acquisition, the recordings must be correlated, and this is achieved on a space basis, whereby a fixed distance is defined. During the measurement, the GPS provides the coordinates of the initial and final points every time the predefined distance is traversed. The use of an external GPS module is preferable as it provides spatial data in real-time, whereas the internal GPS of the camera saves coordinates in the video metadata, which can only be accessed once the recording is complete. Therefore, the external GPS module enables the pre-processing of data while the measurement is ongoing. Subsequent to determining the range of space, the corresponding time interval is extracted and used to correlate audio and video recordings within the current space segment. The audio signal is segmented into intervals based on the computed time ranges. The frames extracted from the camera video are organised into folders, with each folder corresponding to a specific time range. It is worth to notice that the time intervals have to be computed differently for audio and video. The position points of beginning and end of the fixed distance range are valid for the audio signal, indeed the TCM is placed in line with the GPS module, while the camera framing is shifted. The time interval used for frames separation must therefore be calculated on shifted points. The shift of the initial point is the sum of the visual coverage of the camera framing and the distance between the GPS module and the beginning of the framing, while the shift of the final point is the distance between the GPS module and the beginning of the framing (see Fig. 7).

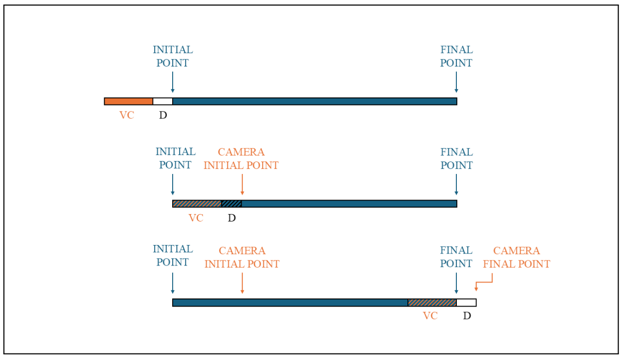

Figure 7. Space interval for video frames: VC is the visual coverage of the camera framing, D is the distance between the GPS module and the beginning of the framing.

At this phase, a correspondence exists between each audio segment and a folder of frames, referring to the same portion of space, making them available for analysis, as illustrated in Fig. 8.

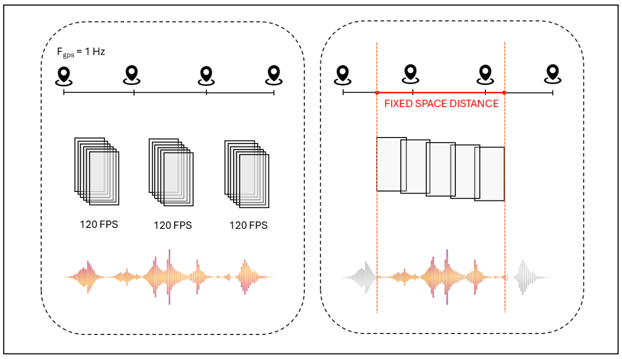

Figure 8. Pre/Post processing of data.

4. CONCLUSION

The multimodal monitoring system developed for automated conditions and damage detection of pavements integrates audio, video and GPS sensors. This work was aimed to improve the system's reliability and spatial accuracy by ensuring correlation between sensor outputs. This goal was addressed through an accurate management of both hardware and software components, such as ensuring robust connectivity, managing varying sensor acquisition frequencies and optimising data synchronisation. Linear interpolation and fitting methods have been employed to successfully resolve discrepancies arising from varying data acquisition rates and frequencies, thereby enhancing the precision of georeferenced data. The synchronisation of audio and video data was based on spatial intervals instead of time intervals and was achieved by ensuring consistent alignment despite differences in sensor placement and temporal recording offset. The segmentation of audio recordings and the organisation of video frames spatially is expected to simplify analysis, reduce redundancy, and optimise computational resources. In the following of the project, comprehensive field tests will be conducted to validate the effectiveness and accuracy in real-world conditions of the system. Subsequent steps will include analysing system performance under various road and weather conditions, improving sensor integration and developing advanced algorithms for automated damage detection and classification.

5. ACKNOWLEDGMENTS

This research is co-funded by GIOVANISI Tuscany regional program FSE+ 2021-2027, as part of the Project PRIOMAPS (PRIOritizzazione degli interventi di MAnutenzione di Pavimentazioni Stradali per la riduzione dell'esposizione al rumore).
This research is part of the NOMADIA research project (Novel On-road Maintenance Assessment with Deep-learning-based Intelligent Algorithms) funded by the Bando per progetti di Ricerca Industriale e Sviluppo Sperimentale BANDO RI&SS ARTES 4.0 N.5 - 2023, related

to the Piano Nazionale di Ripresa e Resilienza (PNRR) for Italy.

6. REFERENCES

[1] Q. Li, F. Qiao, and L. Y. and, "Impacts of pavement types on in-vehicle noise and human health," *Journal of the Air & Waste Management Association*, vol. 66, no. 1, pp. 87–96, 2016.

[2] E. Ascari, M. Cerchiai, L. Fredianelli, D. Melluso, and G. Licitra, "Tuning user-defined pavements in cnossos-eu towards reliable estimates of road noise exposure," *Transportation Research Part D: Transport and Environment*, vol. 130, p. 104195, 2024.

[3] E. Ascari, M. Cerchiai, L. Fredianelli, D. Melluso, F. Rampino, and G. Licitra, "Decision trees and labeling of low noise pavements as support for noise action plans," *Environmental Pollution*, vol. 337, p. 122487, 2023.

[4] G. Licitra, M. Bernardini, R. Moreno, F. Bianco, and L. Fredianelli, "Cnossos-eu coefficients for electric vehicle noise emission," *Applied Acoustics*, vol. 211, p. 109511, 2023.

[5] L. Pellecuer, G. J. Assaf, and M. St-Jacques, "Influence of pavement condition on environmental costs," *Journal of Transportation Engineering*, vol. 140, no. 10, p. 04014050, 2014.

[6] R. Moreno, F. Bianco, S. Carpita, A. Monticelli, L. Fredianelli, and G. Licitra, "Adjusted controlled pass-by (cpb) method for urban road traffic noise assessment," *Sustainability*, vol. 15, no. 6, 2023.

[7] H. B. Zhang and H. Tan, "Research for noise reduction performance of rubberized asphalt overlay," in *Advances in Civil Engineering and Transportation IV*, pp. 1335–1338, 2015.

[8] T. Tsubota, T. Yoshii, H. Shirayanagi, and S. Kurauchi, "Effect of pavement conditions on accident risk in rural expressways," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3613–3618, 2018.

[9] R. Mkwata and E. E. M. Chong, "Effect of pavement surface conditions on road traffic accident - a review," *E3S Web Conf.*, vol. 347, p. 01017, 2022.

[10] E. F. Freitas, F. F. Martins, A. Oliveira, I. R. Segundo, and H. Torres, "Traffic noise and pavement distresses: Modelling and assessment of input parameters influence through data mining techniques," *Applied Acoustics*, vol. 138, pp. 147–155, 2018.

[11] E. Freitas, H. Torres, C. Vuye, and P. Pereira, "Effect of road pavement defects on tyre-road noise," in *INTER-NOISE and NOISE-CON Congress and Conference Proc.*, pp. 5534–5544, 2019.

[12] S. Kephalopoulos, M. Paviotti, and F. A. Ledee, "Common noise assessment methods in europe (cnossos-eu)," 2012.

[13] Y. Hou, Q. Li, C. Zhang, G. Lu, Z. Ye, Y. Chen, L. Wang, and D. Cao, "The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis," *Engineering*, vol. 7, no. 6, pp. 845–856, 2021.

[14] R. S. Rollings and D. W. Pittman, "Field instrumentation and performance monitoring of rigid pavements," *Journal of Transportation Engineering*, vol. 118, no. 3, pp. 361–370, 1992.

[15] D. Timm and A. Priest, "Dynamic pavement response data collection and processing at the ncat test track," 2004.

[16] Y. Chen, M. Q. Feng, and C.-A. Tan, "Bridge structural condition assessment based on vibration and traffic monitoring," *Journal of Engineering Mechanics*, vol. 135, no. 8, pp. 747–758, 2009.

[17] Z. Othman, A. Abdullah, F. Kasmin, and S. S. S. Ahmad, "Road crack detection using adaptive multi resolution thresholding techniques," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, p. 1874, 2019.

[18] P. Li, B. Zhou, and F. Yan, "A method to improve the accuracy of pavement crack identification by combining a semantic segmentation and edge detection model," *Applied Sciences*, vol. 12, p. 4714, 2022.

[19] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3708–3712, 2016.

[20] R. M. Morey, N. C. H. R. Program., and N. R. C. U. T. R. Board., *Ground penetrating radar for evaluating subsurface conditions for transportation facilities*. Washington, D.C: National Academy Press, 1998.

[21] S.-K. Lee, J. Yoo, C.-H. Lee, K. An, Y.-S. Yoon, J. Lee, G.-H. Yeom, and S.-U. Hwang, "Road type classification using deep learning for tire-pavement interaction noise data in autonomous driving vehicle," *Applied Acoustics*, vol. 212, p. 109597, 2023.

[22] J. Masino, J. Pinay, M. Reischl, and F. Gauterin, "Road surface prediction from acoustical measurements in the tire cavity using support vector machine," *Applied Acoustics*, vol. 125, pp. 41–48, 2017.

[23] G. Schiaffino, L. G. D. Pizzo, S. Silvestri, F. Bianco, G. Licitra, and F. G. Praticò, "Machine learning techniques applied to road health status recognition through tyre cavity noise analysis," *Journal of Physics: Conference Series*, vol. 2162, no. 1, p. 012011, 2022.

[24] N.-D. Hoang and Q.-L. Nguyen, "A novel method for asphalt pavement crack classification based on image processing and machine learning," *Engineering with Computers*, vol. 35, p. 487–498, 2019.

[25] S. Kanka, F. Artuso, F. Bianco, S. Carpita, A. Moro, F. Fidecaro, and G. Licitra, "Estimation of the pavement condition index via tyre cavity noise measurements and ai signal processing," in *Proc. of the 10th Convention of the European Acoustics Association. FA2023*, pp. 933–939, 2022.

[26] Z. Mohamed, "Structural-acoustic analysis of tire cavity system," 2020.

[27] H. Yamauchi and Y. Akiyoshi, "Theoretical analysis of tire acoustic cavity noise and proposal of improvement technique," *JSAE Review*, vol. 23, no. 1, pp. 89–94, 2002.

[28] Z. Feng, P. Gu, Y. Chen, and Z. Li, "Modeling and experimental investigation of tire cavity noise generation mechanisms for a rolling tire," *SAE International Journal of Passenger Cars - Mechanical Systems*, vol. 2, pp. 1414–1423, 2009.

[29] X. Hu, X. Liu, X. Wan, Y. Shan, and J. Yi, "Experimental analysis of sound field in the tire cavity arising from the acoustic cavity resonance," *Applied Acoustics*, vol. 161, p. 107172, 2020.

[30] X. Hu, X. Liu, Y. Shan, and T. He, "Research on split and evolution of acoustic cavity resonance frequency of rotating loaded tire," *Journal of Vibration and Control*, vol. 29, no. 1-2, pp. 466–478, 2023.

[31] J. Guan, X. Yang, L. Ding, X. Cheng, V. Lee, and C. Jin, "Automated pixel-level pavement distress detection based on stereo vision and deep learning," *Automation in Construction*, vol. 129, p. 103788, 2021.

[32] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, 2018.

[33] X. Chen, X. Zhang, J. Li, M. Ren, and B. Zhou, "A new method for automated monitoring of road pavement aging conditions based on recurrent neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24510–24523, 2022.

[34] Z. Zhang, F. Liu, Y. Huang, and Y. Hou, "Detection and statistics system of pavement distresses based on street view videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 10, pp. 15106–15115, 2024.

[35] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video object localization and tracking," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 22–31, 2001.

[36] M. Pucher, D. Schabus, P. Schallauer, Y. Lypetskyy, F. Graf, H. Rainer, M. Stadtschnitzer, S. Sternig, J. Birchbauer, W. Schneider, and B. Schalko, "Multimodal highway monitoring for robust incident detection," in *13th International IEEE Conference on Intelligent Transportation Systems*, pp. 837–842, 2010.