Extracting Cardiovascular-Induced Chest Vibrations from Ordinary Chest Videos: A Comparative Study

M. Rahman and A. Taebi

Biomedical Engineering Program, Mississippi State University, Mississippi State, Mississippi, USA
{mmr510, a.taebi}@msstate.edu

***Abstract*— Seismocardiography (SCG) has attracted significant interest for monitoring cardiac health and diagnosing cardiovascular conditions. While traditional SCG methods rely on uncomfortable chest-mounted accelerometers, recent research explores non-contact approaches, including analyzing video recordings of the chest. In this study, three computer vision-based methods including Lucas-Kanade optical flow, template tracking, and Gunnar-Farneback optical flow were evaluated for extracting SCG signals from ordinary camera-recorded chest videos. The study focused on right-to-left and head-to-foot SCG signals obtained from 13 healthy subjects during breath-hold at the end of exhalation and inhalation. Comparative analysis was performed by calculating the mean squared error (MSE) and root MSE (RMSE) between the vision-based SCG signals and the gold-standard accelerometer signals. Visual and quantitative analyses showed that the Lucas-Kanade and template tracking methods estimated vision-based SCG signals closely resembling the accelerometer data, particularly in the head-to-foot direction. The Lucas-Kanade method had MSE values ranging from 0.14 to 0.93, RMSE values from 0.38 to 0.96, average correlation values of 0.82±0.09. The template tracking method showed MSE values between 0.12 to 0.94, RMSE values from 0.35 to 0.97, and average correlation values of 0.83±0.10. In comparison, the Farneback method had higher MSE values ranging from 0.20 to 1.07, RMSE values from 0.44 to 1.03, and average correlation values of 0.76±0.11. These results suggest the effectiveness of Lucas-Kanade and template tracking methods for non-contact SCG signal extraction from chest video data.**

Keywords*— *Seismocardiography, heart vibrations, contactless cardiovascular monitoring, vision-based SCG.

I. Introduction

Cardiovascular diseases are a significant global health concern, being the leading cause of mortality with approximately 17.9 million deaths annually, and their economic burden is substantial, with costs projected to reach 47 trillion USD by 2030 [1–3]. Despite advancements in prevention, diagnosis, and treatment, the impact of these conditions on healthcare systems, productivity, and societal costs remains significant. In that regard, early detection and management of cardiovascular diseases are crucial not only for improving patient outcomes but also for alleviating the strain on healthcare resources [1].

Seismocardiography (SCG) emerges as a promising noninvasive technique for capturing cardiac vibrations from the chest surface, offering valuable information on cardiac activity through the detection of both low-frequency infrasonic vibrations and higher-frequency components related to cardiac events [4–8]. Unlike electrocardiography, SCG does not rely on the electrical activity of the heart, making it a unique and complementary method for assessing cardiac mechanical health. Traditionally, SCG involves attaching accelerometers directly to the chest, which can be inconvenient and costly, particularly when an array of accelerometers is used to measure SCG signals from multiple chest locations. This has prompted researchers to explore alternative strategies including non-contact SCG approaches utilizing infrared speckle vibrometry [9–11] and millimeter and microwave radar [12–15], which offer the potential to capture SCG signals without physical contact.

Advancements in computer vision have further expanded the possibilities for non-contact SCG signal extraction. Recently, we proposed three innovative computer vision-based methods for SCG extraction from chest videos: Lucas-Kanade optical flow-based SCG extraction [16], template tracking-based SCG extraction [17], and Gunnar-Farneback optical flow-based SCG extraction [18]. These methods are developed to provide accurate and reliable SCG signal extraction pipelines from chest videos captured by ordinary cameras, potentially revolutionizing the field of cardiac diagnostics by providing a more accessible and comfortable alternative to traditional methods. The objective of this paper is to compare the accuracy of these three methods in extracting SCG signals from ordinary videos, thereby contributing to the ongoing development of non-invasive and efficient diagnostic tools for cardiovascular health. By advancing non-contact SCG techniques through computer vision, we aim to address the limitations of traditional methods and offer new pathways for early diagnosis and continuous monitoring of cardiovascular diseases through more convenient and accessible methods, ultimately contributing to better patient outcomes and reduced healthcare burdens.

II. Materials and Methods

A. Study Population

This research study was conducted under a protocol reviewed and approved by the Institutional Review Board at Mississippi State University. Data were collected from 14 human subjects (including 4 females), with an average age of 23.50 ± 5.16 years and a body mass index of 23.93 ± 4.07 kg/m^2. Prior to participation, all subjects were informed about the study's goals and procedures and signed an informed consent form. Additionally, all participants confirmed they had no history of cardiovascular disease. Due to incomplete data recording from one of the female subjects, the study analysis was based on the data from the remaining 13 subjects.

B. Data Acquisition Protocol

We attached a tri-axial accelerometer (356A32, PCB Piezotronics, Depew, NY) to the left sternal border near the fourth costal notch while the subjects were in a supine position. The accelerometer was used to record gold-standard SCG signals at a sampling frequency of 5000 Hz. A sticker patterned with a QR code was attached to the top face of the accelerometer to create a high-contrast artificial region and facilitate the object tracking as shown in Figure 1. Chest videos were recorded using an iPhone (13 Pro, Apple Inc., Cupertino, CA) at 60 fps with a resolution of 3840×2160 pixels. To prevent phone vibrations during the recordings, we used a phone holder and a Bluetooth control to remotely start and stop the video recordings. Synchronization between the accelerometer and video data was achieved using a microphone connected to both the smartphone and the data acquisition system recording the accelerometer output (416, iWorx Systems, Inc., Dover, NH). At the beginning and end of each recording, the microphone was tapped, and these timestamps were utilized to synchronize the videos and the SCG data obtained from the accelerometer. We recorded data in two breath-holding sessions, each lasting about 15 seconds: one at the end of exhalation (BHEE) and the other at the end of inhalation (BHEI).

C. Vision-based SCG Signal Extraction

We used three computer vision-based methods to extract SCG signals from the region of the sticker attached to the accelerometer in the video. All image processing and computer vision techniques were implemented using Python 3.10, while signal processing was performed using MATLAB R2022a. To extract the SCG signals from the sticker, we first identified the sticker's location in the first video frame. For this purpose, we utilized the same deep learning model used in [18], where a YOLOv7 [19] object detection method was trained on a custom QR code dataset.

After locating the sticker, we extracted the right-to-left and head-to-foot components of the SCG signals from the chest videos using computer vision techniques, including Lucas-Kanade optical flow [16], template tracking [17], and Gunnar-Farneback optical flow [18].

1) Lucas–Kanade Optical Flow: We employed the Lucas-Kanade optical flow method [20, 21] to track the sticker across video frames and extract the SCG signal in both right-to-left and head-to-foot directions, as described in [16]. This dense optical flow technique estimates motion vectors by assuming that the flow remains approximately constant within a small local neighborhood of a pixel. It analyzes differences between consecutive video frames, solving the basic optical flow equations using the least squares criterion. The goal is to find the optimal alignment of pixels, representing the motion vector at each pixel.

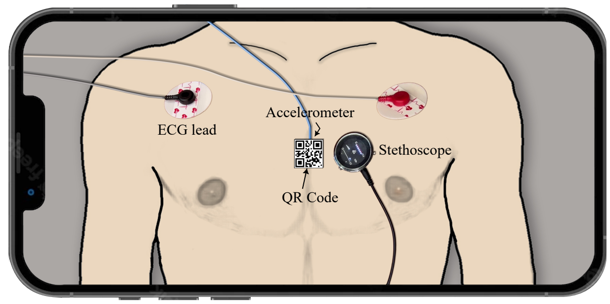


Figure 1. Data acquisition and sensor placement setup. Chest videos were recorded using an ordinary camera phone. Simultaneously, an accelerometer was used to record SCG signals. A sticker patterned with a QR code was attached to the top face of the accelerometer to facilitate the extraction of SCG signals from the chest videos.

2) Template Tracking: Template tracking [22] is a technique used to follow an object's movement across video frames. The process begins by selecting a template from the initial frame. Subsequently, this template is compared to image regions in the subsequent frames. The objective is to find the best match with the template, indicating the new location of the object. This is accomplished by sliding the template over the image and identifying the location with the highest similarity between the template and the overlapped region of the new video frame. To extract the SCG signal, we adopted the workflow outlined in [17] from the video using template tracking. Initially, template tracking was employed to follow the identified sticker across successive video frames, after which a sub-pixel refinement process was applied to enhance displacement accuracy at the sub-pixel level.

3) Gunnar-Farneback Optical Flow: This method [23] is a dense motion estimation technique that calculates the motion for every pixel between two frames. It constructs an image pyramid, where each successive level has a lower resolution than the preceding one. At each level, the algorithm uses polynomial expansion to approximate the pixel neighborhoods. By minimizing the sum of squared differences between the predicted and actual pixel intensities, it iteratively determines the best displacement for each pixel. In our implementation, once the sticker's location was identified in the first frame, we tracked its motion across subsequent frames using the Gunnar-Farneback optical flow algorithm to extract the SCG signal, following the method in [18]. For the Gunnar-Farneback algorithm, we set the pyramid scale factor to 0.5, using five pyramid levels. The averaging window size was set to 20, with 10 iterations applied at each pyramid level. We used a pixel neighborhood size of five and applied Gaussian smoothing with a standard deviation of 1.1. This process produced an optical flow map with a displacement vector (dx, dy) for each pixel. To calculate the overall displacement for each region of interest (RoI), we took the median displacement of all pixels within the RoI and obtained the second derivative of the displacement as the acceleration (SCG) signal in the right-to-left and head-to-foot directions.

D. Signal Denoising

After extracting vision-based SCG signals from the QR code sticker motions using the three aforementioned algorithms in both right-to-left and head-to-foot directions, we applied a band-pass filter with cutoff frequencies of 1 and 30 Hz to both the vision-based SCG signals and the accelerometer signal to remove low-frequency respiration vibrations and high-frequency SCG components above 30 Hz. Next, we resampled the vision-based SCG signals from 60 Hz to 5000 Hz to match the sampling frequency of the gold standard accelerometer signal. During the resampling process, we used MATLAB's resample function, which employs an FIR antialiasing low-pass filter and accounts for the signal delay caused by this filter. Finally, to enhance signal quality for subsequent analysis and interpretation, the signals were centered around zero by subtracting their mean values and normalized by dividing by their standard deviations.

III. Results and Discussion

A. Accuracy of the Vision-Based SCG

We compared the vision-based SCG signals with the gold-standard signals recorded by the accelerometer. Figure 2 shows the SCG signals captured in both the right-to-left and head-to-foot directions for subject 8 (S08) during BHEE. Visual inspection reveals a close resemblance between the signals generated by all three vision-based methods and the gold standard accelerometer signal. To quantify the similarity between the signals, we calculated the mean squared error (MSE) and the root MSE (RMSE) between the vision-based SCG signals and the accelerometer signal (Figure 3). Results show that the vision-based SCG signals in the head-to-foot direction generally have lower MSE and RMSE values compared to those in the right-to-left direction. This indicates better performance of the vision-based methods in capturing the head-to-foot vibrations than the right-to-left SCG components.

Among the three methods, Lucas-Kanade and template tracking achieved comparable performance, with consistently lower MSE and RMSE values compared to the Farneback method. For both the BHEI and BHEE tasks combined, the Lucas-Kanade method's MSE ranged from 0.52 to 1.46 in the right-to-left direction and 0.14 to 0.93 in the head-to-foot direction (Figure 3). The template tracking method exhibited similar results, with MSE values between 0.58 and 1.47 for the right-to-left SCG and 0.12 to 0.94 for the head-to-foot SCG. The Farneback method had higher dissimilarity to the accelerometer signal, with MSE ranging from 0.67 to 1.56 and 0.20 to 1.07 for the right-to-left and head-to-foot SCG signals, respectively.

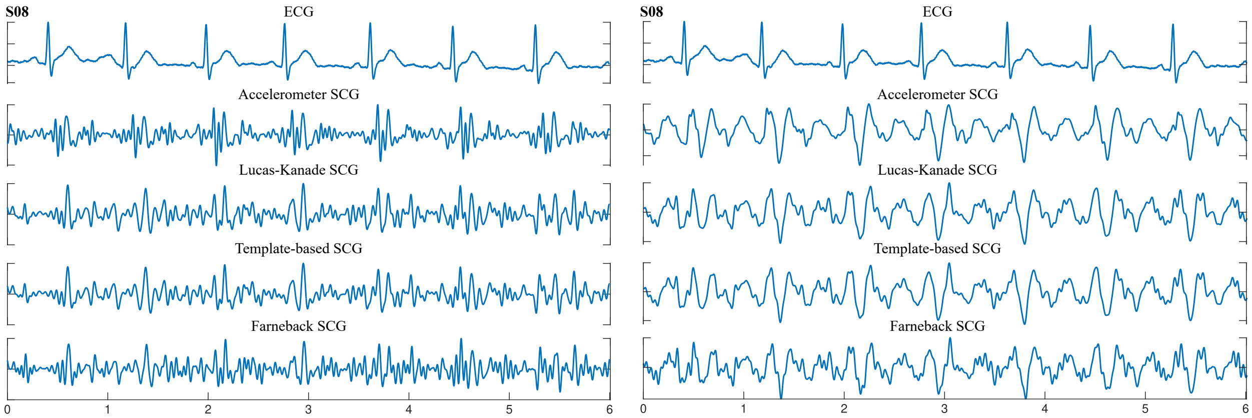
The RMSE values followed a similar trend. The Lucas-Kanade method achieved a range of 0.72 to 1.21 (right-to-left) and 0.38 to 0.96 (head-to-foot), and the template tracking method showed RMSE values between 0.76 to 1.21 (right-to-left) and 0.35 to 0.96 (head-to-foot). The Farneback method again yielded higher RMSE values, ranging from 0.82 to 1.25 (right-to-left) and 0.44 to 1.03 (head-to-foot).

In addition to MSE and RMSE, we also calculated the correlation of each method with the gold standard accelerometer signal, with the correlation range set between -1 to 1. Among 13 subjects, the Lucas-Kanade method had an average correlation of 0.50 ± 0.13 in the right-to-left direction and 0.82 ± 0.09 in the head-to-foot direction. The template tracking method had an average correlation of 0.50 ± 0.12 in the right-to-left direction and 0.83 ± 0.10 in the head-to-foot direction. The Farneback method had an average correlation of 0.44 ± 0.13 in the right-to-left direction and 0.76 ± 0.11 in the head-to-foot direction. When comparing the correlation, the Lucas-Kanade and template tracking methods performed slightly better than the Farneback method. Across all methods, the highest correlation was found in the head-to-foot direction.

This quantitative analysis corroborates the visual inspection observations presented in Figure 2, reinforcing the superior performance of the Lucas-Kanade and template tracking methods in capturing the vision-based SCG signal, particularly in the head-to-foot direction.

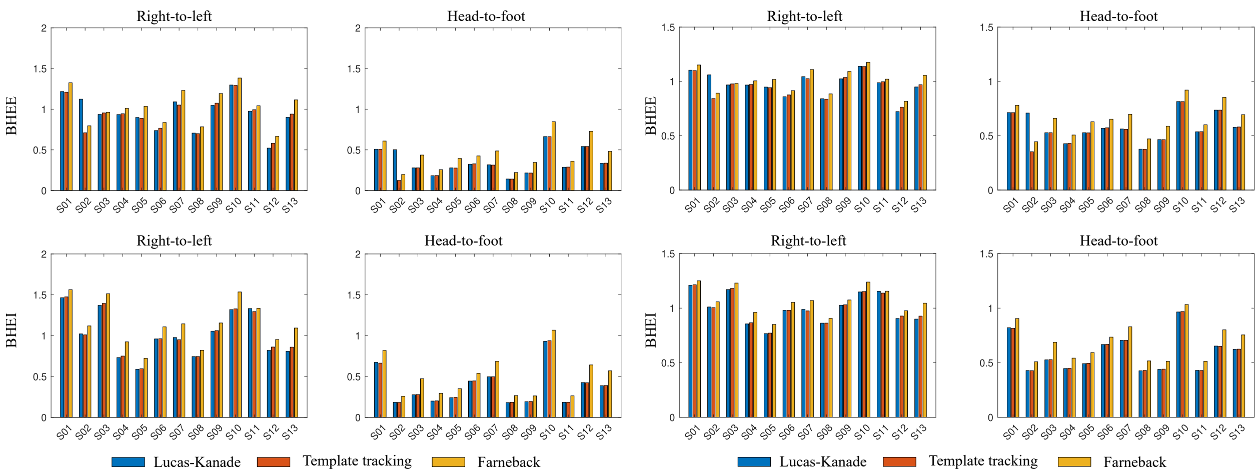
B. Implications of Findings

The lower MSE and RMSE values, along with higher correlation values in the head-to-foot direction suggest that vertical vibrations are more accurately captured by vision-based methods than horizontal vibrations. This might be due to the nature of SCG signals where head-to-foot vibrations are more pronounced and easier to detect. This finding is crucial as it provides insight into the directional sensitivity of our vision-based SCG methods, which can guide the design of future SCG monitoring systems. The consistently lower error met-



(a) SCG signal in right-to-left direction (b) SCG signal in head-to-foot direction

Figure 2. A 6-second sample of the vision-based SCG signals in the right-to-left (a) and head-to-foot (b) directions recorded during breath-hold at the end of exhalation from subject S08. The gold-standard SCG signal recorded by the accelerometer and the subject's electrocardiogram are shown for reference.



(a) MSE between vision-based and accelerometer signals (b) RMSE between vision-based and accelerometer signals

Figure 3. Comparison of the vision-based and gold-standard SCG signals during both breath-holding tasks. The MSE and RMSE values are plotted for the right-to-left and head-to-foot SCG components.

rics for Lucas-Kanade and template tracking methods highlight their robustness and reliability in non-contact SCG measurement. The Farneback method, while still viable, showed higher error values, indicating it may be less suitable for precise SCG capture in the context of this study.

C. Potential Applications and Future Work

These findings have significant implications for the development of non-invasive, vision-based cardiac monitoring systems. The ability to accurately capture SCG signals without direct contact offers numerous advantages, including increased patient comfort, reduced risk of infection in particular groups of patients, and ease of deployment in various settings, such as home monitoring and telemedicine.

Future research could explore the integration of these vision-based methods with advanced machine learning algorithms to further improve the accuracy and reliability of SCG signal extraction. Additionally, expanding the study to include a more diverse population and varying clinical conditions can help generalize the findings and validate the efficacy of these methods across different scenarios. Further studies could also investigate the optimization of camera settings, lighting conditions, and environmental factors to enhance the robustness of vision-based SCG monitoring systems. Exploring the combination of vision-based methods with other non-contact sensing technologies, such as radar and lidar, could provide a multi-modal approach to cardiac monitoring, offering even greater accuracy and reliability.

IV. Conclusion

This study compared the performance of three computer vision methods based on Lucas-Kanade optical flow, template tracking, and Gunnar-Farneback optical flow for extracting SCG signals from chest videos. Results underscored the efficacy of these methods in SCG signal extraction and highlighted the potential for these methods to enhance non-invasive cardiac diagnostics. The results indicated that the Lucas-Kanade optical flow and template tracking methods outperformed the Farneback method. These methods exhibited superior accuracy, particularly in capturing SCG signals in the head-to-foot direction. As non-contact alternatives to traditional accelerometer-based measurements, these vision-based techniques can hold promise for advancing SCG research, with the potential to revolutionize cardiac monitoring and diagnostics.

Acknowledgements

This work was supported by the National Science Foundation (Grant No. 2340020) and a SMART Business Act grant (No. 2024-04) through the Mississippi Institutes of Higher Learning. A.T. has a financial interest in the vision-based SCG technology that could potentially benefit from this project. This relationship has been disclosed to the appropriate entities at MSU and a management plan has been implemented in accordance with the university's policies and procedures regarding Financial Conflict of Interest in Sponsored Activities.

References

[1] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, A. Alonso, A. Z. Beaton, M. S. Bittencourt, A. K. Boehme, A. E. Buxton, A. P. Carson, Y. Commodore-Mensah *et al.*, "Heart disease and stroke statistics—2022 update: a report from the american heart association," *Circulation*, vol. 145, no. 8, pp. e153–e639, 2022.

[2] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, N. C. Barengo, A. Z. Beaton, and et al., "Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the gbd 2019 study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982–3021, 2020.

[3] W. H. Organization, "Cardiovascular diseases (CVDs)," https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 2021, accessed: 2024-06-30.

[4] A. Taebi, B. E. Solar, A. J. Bomar, R. H. Sandler, and H. A. Mansy, "Recent advances in seismocardiography," *Vibration*, vol. 2, no. 1, pp. 64–86, 2019.

[5] M. M. Rahman and A. Taebi, "Reconstruction of 3-axis seismocardiogram from right-to-left and head-to-foot components using a long short-term memory network," *2023 IEEE 19th International Conference on Body Sensor Networks (BSN)*. IEEE, 2023, pp. 1–4.

[6] J. Cook, M. Umar, F. Khalili, and A. Taebi, "Body acoustics for the non-invasive diagnosis of medical conditions," *Bioengineering*, vol. 9, no. 4, p. 149, 2022.

[7] Y. Huang, D. Huang, J. Huang, G. Wang, L. Pan, H. Lu, M. He, and W. Wang, "Camera-based blood pressure monitoring based on multi-site and multi-wavelength pulse transit time features," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[8] E. J. Wang, J. Zhu, M. Jain, T.-J. Lee, E. Saba, L. Nachman, and S. N. Patel, "Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera," *Proceedings of the 2018 CHI conference on human factors in computing Systems*, 2018, pp. 1–9.

[9] S. Que, W. Verkruijsse, M. van Gastel, and S. Stuijk, "Contactless heartbeat measurement using speckle vibrometry," *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 4604–4610.

[10] L. Liu, D. Yu, H. Lu, C. Shan, and W. Wang, "Camera-based seismocardiogram for heart rate variability monitoring," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[11] C. Wang, F. Jin, S. Trivedi, M. Goncharovsky, I. Wang, G. P. Zientara, J. Khurgin, and F.-S. Choa, "From head to toe: Non-contact laser heart rate and respiration rate detection under obstructive conditions," *CLEO: Applications and Technology*. Optica Publishing Group, 2024, pp. ATu3B–5.

[12] Z. Xia, M. M. H. Shandhi, O. T. Inan, and Y. Zhang, "Non-contact sensing of seismocardiogram signals using microwave doppler radar," *IEEE Sensors Journal*, vol. 18, no. 14, pp. 5956–5964, 2018.

[13] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," *Proceedings of the 26th annual international conference on mobile computing and networking*, 2020, pp. 1–14.

[14] K. Gupta, M. Srinivas, J. Soumya, O. J. Pandey, and L. R. Cenkeramaddi, "Automatic contact-less monitoring of breathing rate and heart rate utilizing the fusion of mmwave radar and camera steering system," *IEEE Sensors Journal*, vol. 22, no. 22, pp. 22 179–22 191, 2022.

[15] C. Xu, H. Li, Z. Li, H. Zhang, A. S. Rathore, X. Chen, K. Wang, M.-c. Huang, and W. Xu, "Cardiacwave: A mmwave-based scheme of non-contact and high-definition heart activity computing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–26, 2021.

[16] M. M. Rahman, J. Cook, and A. Taebi, "Non-contact heart vibration measurement using computer vision-based seismocardiography," *Scientific Reports*, vol. 13, no. 1, p. 11787, 2023.

[17] M. M. Rahman, B. Kakavand, W. V. Wurm, W. L. Holman, M. R. Movahed, and A. Taebi, "From video to vital signs: A new method for contactless multichannel seismocardiography," *npj Cardiovascular Health*, 2024.

[18] M. M. Rahman and A. Taebi, "Contactless seismocardiography via gunnar-farneback optical flow," *IEEE International Conference on Body Sensor Networks (BSN)*. IEEE, 2024.

[19] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464–7475.

[20] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI'81. Morgan Kaufmann Publishers Inc., 1981, p. 674–679.

[21] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International journal of computer vision*, vol. 56, no. 3, pp. 221–255, 2004.

[22] R. Brunelli, *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[23] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*. Springer, 2003, pp. 363–370.