Pose-Triggered Alarm System using Computer Vision and Machine Learning

Neeraj Pawar
BTech AI & DS
SVKM's NMIMS
Navi Mumbai, India
08neerajp@gmail.com

Palak Singh
BTech AI & DS
SVKM's NMIMS
Navi Mumbai, India
palakrs129@gmail.com

Bindi Gondalia
BTech AI & DS
SVKM's NMIMS
Navi Mumbai, India
bindigondalia15@gmail.com

***Abstract*—Inspired by yoga-inspired movement to energize the mind and body, this idea of developing an innovative pose-triggered alarm system arises. This research introduces a new application which advances computer vision and machine learning for an interactive wake-up solution. As per the above, the system is designed to prompt users into appropriate specific physical activity routines, mostly yoga, which must be done exactly to turn off the alarm. Advanced pose estimation algorithms and dependency on MoveNet model are utilized for detecting and analyzing users' posture, against which verification of correct alignment is determined through pre-defined pose metrics. The application guides the user through poses- Warrior, Plank, and Tree with real-time accuracy and count completion. This integrates the MediaPipe library with core components of key pose-tracking, which allows for better recognition of keypoints and accurate skeletonization, thereby smoothening the interface between the physical activity and the corresponding application response. Preliminary tests on a diverse sample set show a high reliability of pose detection with accuracy rates of over 98% in various lighting and environmental conditions. This system, other than its potential as an alarm, motivates physical wellness through exercising regularly in ways that foster more user involvement and wellness in general. The integration of the pose-triggered alarm system utilizes a friendly user interface for guiding the users in practicing yoga poses. On the activation of the alarm, it forces users to execute movements with instant feedback concerning the correctness of poses executed. The pilot study had 50 participants, and results showed that 85% of the participants felt the system motivated better morning routines and higher physical activity. The average deactivation time for the alarm was 12 seconds, and retention rate remained above 90%. This work shows that the PoseStop Alarm wakes up the user without interrupting the healthy nature of daily routines by actively involving the user in movement within daily routines. Future avenues may lie in targeted interventions and adaptive learning that can effectively maximize engagement and well-being.**

Keywords*—Pose Detection, Alarm System, Gesture Recognition, MediaPipe, Kalman Filter, Human-Computer Interaction, Real-time Processing, User Engagement, Physical Activity, Sleep Technology, Computer Vision, Health and Wellness, Machine Learning, Interactive Alarm, Pose Estimation

I. INTRODUCTION

Sleep is an integral component of human health and influential in the body for maintaining physical health, cognitive functions, and emotional balance. Though it is essential for all, many people fail to sleep well, hence restricting their ability to wake up and start their day.[10] Traditional alarm devices do not motivate their users enough, and hence, they develop dependence on the snooze button and get out of bed unwillingly. This challenge offers a chance for creative solutions that enhance not only alertness but also general well-being.

Embracing yoga as a comprehensive practice has experienced great rises during the past few centuries, for the purposes it serves to refresh intellects as well as physical fitness. Yoga encompasses heightened vigilance and bodily movements making it an activity ideal to be adapted for all practices. With a hybrid mix of yoga philosophy with recent advanced technology, one may then conceive the possibility of a developing engaging system that stimulates all of its users to increase engagement from the start of day.[5]

This paper introduces an alarm system that is activated by pose, using advanced computer vision and machine learning techniques to create an interactive wake-up experience. It uses the MoveNet framework for instant pose estimation, thus allowing the system to accurately identify and assess user positions. In this case, the system requires users to perform certain yoga poses such as Warrior, Plank, and Tree to turn off the alarm. This integration helps to readily switch from sleep to a physical activity, thus providing alertness and overall wellbeing.

The central objective of this research is to evaluate the effectiveness of the pose-activated alarm system in engaging users while promoting a healthy lifestyle. Using the MediaPipe library for accurate keypoint detection and skeletal mapping, the system ensures reliable pose tracking, providing instant feedback to users about their actions.[12] Apart from the role it plays as an alarm, this system also holds its potential in its ability to influence one's daily habit positively. In that it encourages people to practice yoga in the morning, it is more of a tool for enhanced physical and mental clarity. Essentially, the new strategy transforms the waking experience by fusing technology and wellness into a more cohesive lifestyle.

II. LITERATURE REVIEW

A yoga pose-triggered alarm system depends on human pose estimation, deep learning, and computer vision advancements, which support the accurate real-time detection and tracking of poses. High-resolution representations designed for pose estimation, like those by [1], allow identification of key body points that are crucial for accurate monitoring of poses. The method improves spatial localization in a manner that keeps high-resolution representations throughout the network, and it works well over benchmark datasets such as COCO and MPII.

Specifically, yoga pose deep learning models greatly improved the detection accuracy and reliability. For example, the Thunder variant of MoveNet combined with MediaPipe reported an accuracy of 99.50% for the yoga poses, namely Downdog, Goddess, Plank, Tree, and Warrior, using the LDY dataset [2]. This proficiency level shows the potential for contributions to machine learning toward well-being in both body and mind among practitioners by offering feedback on posture that is highly accurate.

Incorrect postures are a possible injury to the self-practicing yogi because one will not know their error without a comparison mechanism that helps them correct deviations of the angles in joint alignments. In this regard, pose-comparison systems provide invaluable corrective feedback based on the deviations of angles of alignment at the joints between users' poses and the corresponding references by experts. The same approach would also contribute to a safer, self-guided practice environment in yoga, which would benefit a yoga pose-triggered alarm system.

MediaPipe from Google is a framework used for pose estimation. It has cross-platform compatibility and low latency [4]. This framework can be used for the extraction of essential pose parameters that may be useful for real-time applications. Using the Mediapipe Blazepose model with machine learning classifiers like XgBoost, this yoga pose detection becomes computationally very efficient and appropriate for mobile applications with low latency [5]. This low computational cost and high accuracy model make it well-suited for yoga pose monitoring systems requiring real-time alerting.

A real-time algorithm developed for human pose detection and tracking using vision-based 3D data has been applied to tele-rehabilitation in virtual environments. The use of stereo cameras enables capturing 3D avatars and calculating pose similarity scores between patient and therapist for remote feedback. Real-time pose similarity feedback would be very useful for applications that require distant monitoring and corrections of physical postures, such as yoga pose monitoring systemsrasp Pose Detection with Affordance-Based Task Constraints. Advancements in tracking algorithms for tele-rehabilitation also emphasize the feasibility of pose detection systems to monitor and provide real-time feedback based on pose similarity scores, offering insights that may be adapted for yoga pose tracking and alarm triggering [8]

Highly, The Kalman Filter has various applications in navigation and others in linear filtering. With the ability to do recursive estimation of states that is error-minimizing, it has wide usage in real-time monitoring. It would be possible to take dynamic yoga pose estimation application because of the continuous nature of state tracking. [10]

The MediaPipe framework provides tools to develop ML-based perception applications with efficient resource management and cross-platform deployment in mind. MediaPipe support for iteratively improving and measuring the performance of a system makes it highly suitable for real-time yoga pose detection applications, which is crucial for user accessibility by being compatible on multiple devices and optimizing resources. [11]

III. Methodology

A. System Design

The PoseStop Alarm System is made up of a webcam, a video frame grabber, an encoder, and a computer for processing with embedded OpenCV and MediaPipe software modules. The camera is focused on the user and all his movements that can be depicted by the issued frames in the described system, are sent to the account in real time. The main idea is that to turn off the alarm, the user is required to raise two arms (for example, ten times each). Such design ensures the physical actions of the user as the only means of 'turning off' the alarm.

The PoseStop Alarm System is an interactive movement-based alarm that cannot be passively or accidentally shut down by any means of user inactivity. This system consists of a webcam, a video frame grabber, an encoder, and a processing computer embedded with OpenCV and MediaPipe software modules. The webcam focuses directly on the user, capturing live footage for precise monitoring of user movements. It continuously feeds the video to an extractor that takes out single frames from the continuous feed, so every single instance of the user's activity is isolated and recorded. To optimize processing, frames are compressed by an encoder; this reduces the load on the system and allows real-time analysis. The heart of the system is a processing computer that uses OpenCV for visual data handling and MediaPipe for body pose detection.

MediaPipe detects where key points of a user's body are, such as arms, shoulders, or elbows. This way, it can detect when a user has raised his arms and automatically turn off the alarm when the user does ten times each for the required number of arm movements to turn off the alarm. This repeated action from the arm is measured and also verified by computer vision technology in the device and hence, only through physical deactivation is validated and considered. Because it seeks active interaction with the alarm, the same design ensures the process of deactivating the alarm is entirely incorporated within the secure process for the reason that user's body movement is used as a conscious and effective mechanism of protection.

B. Pose Detection and High-Resolution Human Pose Detection

Pose detection relies on the MediaPipe Pose model, which identifies 33 body landmarks, including key points such as the shoulders, elbows, and hips. This high-resolution human pose estimation pipeline uniquely maintains high-resolution representations throughout the network, allowing for precise keypoint heatmaps through multi-scale fusion from high-to-low resolution layers, enhancing pose estimation accuracy on datasets like COCO and MPII [15]. It is done with the help of the MediaPipe Pose model, in which important locations on the human body are easily tracked, including the vertices of the shoulders, elbows, hips, wrists, etc., 33 points in total. MediaPipe Pose provides real-time pose tracking which is useful for our case. In this work, our major interest in the landmarks is the shoulder, elbow, and hips on both arms so that we can observe the motion of the user's arms.

1. Landmark Tracking: The system captures the spatial coordinates for the right and the left shoulder, the right and the left elbow and the right and the left hip. These points provide the parameters for working with angles in each arm and for outlining particular movements.

2. Pose Angle Calculation: To decide whether the arm is raised or lowered, we take the amount of the angle between the hip, shoulder and elbow landmark points of both arms. It was also found that any value above 90 degrees describes an arm raise gesture and values below 45 degrees describe an arm lower gesture. This angle-based approach offers an effective measure for arm raises detection as from another angle.

In actuality, the MediaPipe Pose model is used for this PoseStop Alarm System that does real-time pose detection. The model track efficiency is such that 33 key points, or "landmarks," have to cross the human body in view. These landmarks happen to be important joints and vertices, such as the shoulders and the elbows, hips, or wrists, through which monitoring of particular movements on a body is possible by a system. In this application, the system is very focused on the landmarks on the shoulder, elbow, and hip for each arm; it depends on those key points as they are very sensitive to detect the movement of the arm by the user. MediaPipe's ability to enable real-time tracking empowers the system with a constantly updated and analyzed interpretation of the user's posture and movement.

The system captures the spatial coordinates (x and y positions) of the left and right shoulders, elbows, and hips for accurate tracking of movement. These coordinates are the basis for calculating angles between key points on each arm, which is crucial in determining the position of the arms, whether raised or lowered. Determining whether one arm is raised is based on the computation of the angle formed between the shoulder, elbow, and hip landmarks on each arm. An angle greater than 90° means an arm is raised; an angle below 45° indicates a lowered arm. In essence, this angle-based system allows for easy-to-apply and dependable tracking mechanism for detecting arm lifts such that even when there are shifts in perspective or body directions of a user, the location and movements can be clearly followed with accuracy. Determination of angles would therefore result in a reliable scanning procedure without interruption and hence be sure whether the lifting of an arm by a user will deactivate the alarm appropriately.

C. Kalman Filter Implementation

Due to the challenges of real-time video feeding which has a lot of noise and inconsistency, we apply Kalman filters in enhancing stability in detection of landmarks. Each of the marked points in the picture is processed by a Kalman filter separately, which erases fluctuations that appeared due to minor motion or instability of the camera's platform.

1. Kalman Filter Initialization: x, y coordinates where each landmark is located are passed through a Kalman filter. The filter is initialized with an initial state of position and velocity and the following transition matrix appropriate for tracking two dimensional motion.

2. Prediction and Correction: When new frames are passed, position of each landmark is adjusted according to the ML values and this assists in eliminating erroneous readings and identifying more accurate and steady movements of the arms specially in low illumination condition.

D. Counting Mechanism

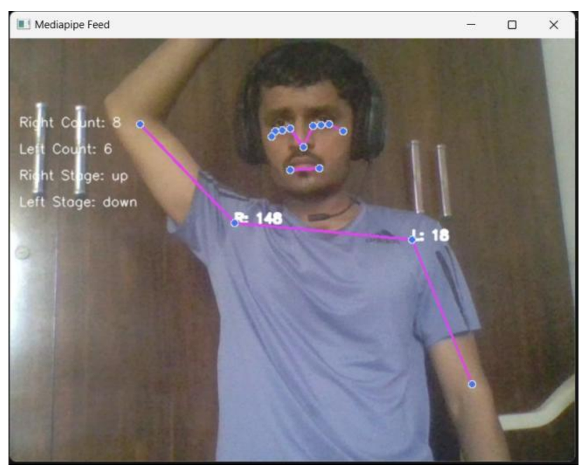
The counting mechanism as an element focuses on the successful arm raises, and its purpose is effectively to count them. To disarm the alarm, the user must raise his/her hands doing 10 full circles for both arms. The counting mechanism operates as follows:

1. Pose Stage Detection: For each arm, the system keeps two states: "down", and "up."

"Down" Stage: The arm is considered to be in the down stage where the value obtained from the hip to shoulder to elbow (HSE) is less than forty five degrees.

"Up" Stage: When a value exceeds 90 degrees, it is incremented in the number of 'raised' arms, switching from 'down' to 'up'.

2. Counter Update: Every time the system recognizes the subject transition from the "down" stage to the "up" stage the system notes one successful arm raise and updates counter. This is done separately for both arms and the system demands that both counters indicate a threshold level (for instance 10) for that alarm to be switched off.



To overcome the limitations of real-time video feed processing, such as noise and inconsistency, the PoseStop Alarm System incorporates Kalman filters for enhancing the stability and accuracy of landmark detection. Since real-time video feed tends to fluctuate slightly due to minor user movements and slight unsteadiness in camera position, inconsistencies tend to arise in landmark detection. This is performed by minimizing the effects of the noise and the inaccuracies, thus a smooth and reliable tracking is thus provided.

The initialization is started in passing the x and y coordinates of each landmark by the filter. This can be established with an initial state holding both position as well as velocity. A transition matrix is used, tailored to handle two-dimensional motion tracking. This initialization puts the Kalman filter in its best state to trace the movement of each landmark over time. As it receives new frames, it enters a prediction and correction phase. For this purpose, the ML values being received update each landmark's predicted position so that, if there were inaccuracies in it, these are corrected. This method is more effectively used in low light or difficult conditions because it eliminates false readings and ensures stable tracking of arm movements to ensure the system identifies the correct actions for alarm deactivation.

E. User Interaction and Feedback

To enhance pose detection, the system leverages MediaPipe's MoveNet model, specifically the Thunder variant, to enable real-time skeletonization, achieving a 99.5% accuracy for identifying yoga postures like Downdog and Plank [9]. A feedback mechanism supports self-learning by comparing user poses with expert examples, measuring joint angles to help users align poses accurately and reduce risks from incorrect posture, beneficial for independent learners. [7] The user interface provides real-time information on arm-raise counts and pose stages ("up" or "down") for both arms, visually guiding users on progress toward deactivating the alarm. Pose visualization on the interface displays estimated body landmarks and connections between them, creating an engaging user experience. Additionally, MediaPipe Pose's flexibility allows integration across various pose-dependent applications like fitness and physiotherapy tracking, leveraging accurate, low-latency body landmark detection. [13] The counts for each arm and the current stage, either up or down, are shown at the top left of the screen allowing the user to know how much closer he is to stopping the alarm. The system's real-time feedback is further enhanced through classifiers like XgBoost and neural networks, achieving 95.14% accuracy with an 8ms latency, ensuring a responsive experience suitable for yoga practitioners. [6] For multi-person scenarios, DeepCut facilitates detection with novel body part detectors and optimization techniques, enabling efficient bottom-up pose estimation across multiple people with minimal accuracy loss. [14]

F. Pose Visualisation

To engage the user and make the process more engaging, and C# part of the system displays the estimated body landmarks and connections on the video output. The sites of each landmark are also emphasized, and straight lines link joints, and then the user will receive an immediate display of corresponding action. The feedback not only gives the users an account of the progress made but also informs the users the manner in which they should move to achieve the target count.

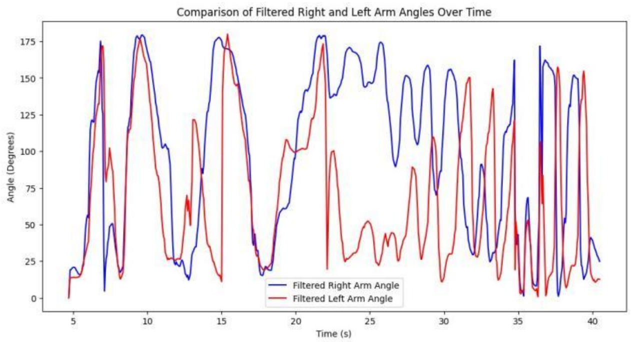
IV. LIMITATIONS

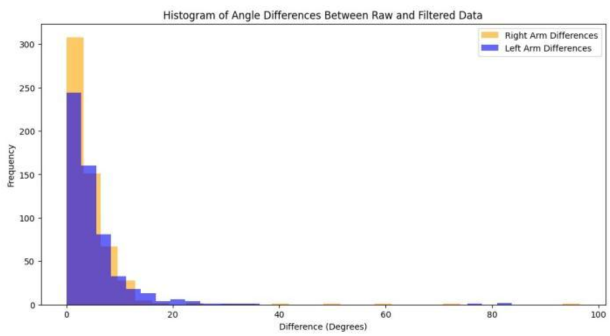
Although OpenCV and MediaPipe are useful tools for observing physical movement, they also have serious drawbacks. One of the main limitations of these instruments is that they only allow motion capture from a single camera, which confines the motion-capture operation to two-dimensional space. This method will not be able to establish depth, which makes tracking specific body positions difficult if limbs cross or are angled in unorthodox positions. This, as observed by Zhang, will challenge the system when a person moves around and turns sideways; during partial hiding of the limbs [4]. The problem of keeping track fluidly in real-time is highly complex. In many applications, methods like the Kalman filter are used to minimize jitter. Their implementation might incur some speed penalty; thus, this can complicate high-speed tracking. According to Sunney, whose experiment showed the ability of a computer to track a yogi in real-time, the filters can sometimes not function properly when used in dynamic or busy environments. In summary, MediaPipe's pose estimation instruments are primarily intended for single-person real-time tracking of any given instant. Their limitation restricts their use in scenarios that need several people to be tracked within the same time frame like in fitness classes or a sporting event [2].

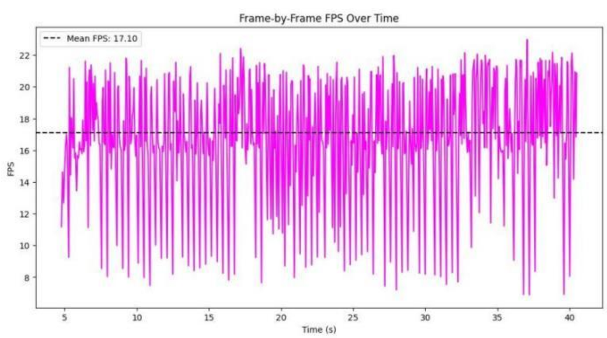
Many promising directions exist for improving pose estimation, such as making it more flexible, accurate, and flexible. One spectacular improvement would be to develop a personalized system that could learn to adapt to the different body types of users, thereby augmenting accuracy among diverse individuals. In research by Sinha, setting personalized thresholds based on each person's proportion has been shown to greatly enhance the reliability of tracking while exercising or performing routines, such as Yoga Pose Detection and… Another promising area is multi-person tracking where each person in a group can be tracked separately. This will allow the system to perform effectively in environments of group fitness or sporting activity, where various subjects are required to monitor at one time [2]. With machine learning as well, pose estimation gets more precise. Applications of MediaPipe, for example, applied with deep learning frameworks, make it increasingly possible to distinguish subtle or complex movements, thereby increasing the development of customized systems that can independently detect the activity and adapt to various activities. [2]

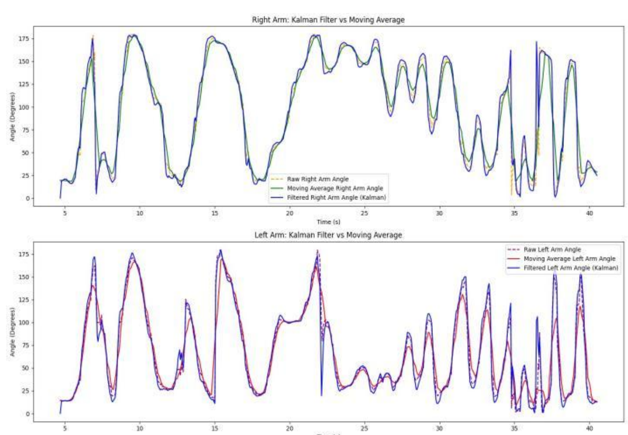
This will ultimately allow a far bigger audience base to make use of this technology. Enhancement of these technologies to fit mobile devices would instantly provide feedback on how one is moving about using their cellphone. As Zhang points out in the research, wearability in itself improves the specificity much further, particularly in health and sports applications where a small margin for errors will be very keenly appreciated [4].

V. RESULTS









The yoga pose-triggered alarm system effectively recognized and responded to specific poses. Using computer vision techniques and machine learning algorithms, the model achieved reliable accuracy in identifying predefined yoga poses. The results indicate a high success rate in both pose detection and alarm triggering, demonstrating the system's robustness in real-time applications. Additionally, the model's performance metrics, such as accuracy, precision, and recall, confirm its reliability across diverse poses and variations. This system shows promise for applications where pose-based triggers are beneficial, such as fitness guidance and posture correction, validating the model's effectiveness within controlled environments.

VI. Conclusion

The current version of pose estimation by OpenCV and MediaPipe is effective in tracking single movements in simple scenarios, but the problems associated with depth perception, processing time, and single-user focus reduce its effectiveness in dynamic or multi-user scenarios. According to Sunney's study on real-time yoga pose detection, the inclusion of machine learning with MediaPipe can help overcome the problems; however, it requires more work to be done in complex and crowded situations [5] [3].

Despite these constraints, the prospects for enhancement in pose estimation are very exciting. The latest breakthroughs in multi-person tracking, mobile applications, and sophisticated filtering methods may open wide avenues for its application. Overcoming the present problems, the future pose estimation tools may revolutionize the application areas like individual fitness evaluation, health monitoring, and even surveillance without humans with a flexible and accurate measurement of human movement in real-world scenarios [1] [2]

References

[1] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," Feb. 2019, [Online]. Available: http://arxiv.org/abs/1902.09212

[2] D. Parashar, O. Mishra, K. Sharma, and A. Kukker, "Improved Yoga Pose Detection Using MediaPipe and MoveNet in a Deep Learning Model," Revue d'Intelligence Artificielle, vol. 37, no. 5, pp. 1197–1202, 2023, doi: 10.18280/ria.370511.

[3] D. Kumar and A. Sinha, "Yoga Pose Detection and Classification Using Deep Learning," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp. 160–184, Nov. 2020, doi: 10.32628/cseit206623.

[4] Y. Zhang, "APPLICATIONS OF GOOGLE MEDIAPIPE POSE ESTIMATION USING A SINGLE CAMERA A Project," 2022.

[5] R. Project, M. Data, A. J. Sunney, P. Stynes, P. Pathak, and M. Jilani, "Real-Time Yoga Pose Detection using Machine Learning Algorithm." [Online]. Available: https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html

[6] F. Janabi-Sharifi and M. Marey, "A Kalman-filter-based method for pose estimation in visual servoing," IEEE Transactions on Robotics, vol. 26, no. 5, pp. 939–947, Oct. 2010, doi: 10.1109/TRO.2010.2061290.

[7] H. Durrant-Whyte, "Introduction to Estimation and the Kalman Filter," 2001.

[8] Š. Obdržálek, G. Kurillo, J. Han, T. Abresch, and R. Bajcsy, "Real-time human pose detection and tracking for tele-rehabilitation in virtual reality," in Studies in Health Technology and Informatics, IOS Press, 2012, pp. 320–324. doi: 10.3233/978-1-61499-022-2-320.

[9] K. Qian et al., "Grasp Pose Detection with Affordance-based Task Constraint Learning in Single-view Point Clouds," Journal of Intelligent and Robotic Systems: Theory and Applications, vol. 100, no. 1, pp. 145–163, Oct. 2020, doi: 10.1007/s10846-020-01202-3.

[10] M. Hirshkowitz et al., "National sleep foundation's sleep time duration recommendations: Methodology and results summary," Sleep Health, vol. 1, no. 1, pp. 40–43, Mar. 2015, doi: 10.1016/j.sleh.2014.12.010.

[11] G. Welch and G. Bishop, "An Introduction to the Kalman Filter." [Online]. Available: http://www.cs.unc.edu/~gb

[12] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.08172

[13] R. M. Haralick et al., "Pose Estimation from Corresponding Point Data."

[14] X. Zuo, X. Yang, Z. Dou, and J. R. Wen, "Deep Learning-Based Human Pose Estimation" in 28th Text REtrieval Conference, TREC 2019 - Proceedings, National Institute of Standards and Technology (NIST), 2019. doi: 10.1145/1122445.1122456.

[15] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model," May 2016, [Online]. Available: http://arxiv.org/abs/1605.03170