NÖHÜ Müh. Bilim. Derg. / NOHU J. Eng. Sci., 16 (2026), 1580491

Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences

Araştırma makalesi / Research article

www.dergipark.org.tr/tr/pub/ngumuh / www.dergipark.org.tr/en/pub/ngumuh

Implementing hand recognition solutions to control intelligent lighting in office

Ofiste akıllı aydınlatmayı kontrol etmek için el tanıma çözümlerinin uygulanması

Abdulvahit Karail[1*], Veysel Böcekçi[2], İsmail Kıyak[3]

[1] *Institute of Pure and Applied Sciences, Marmara University, İstanbul, Türkiye*
[2,3] *Faculty of Technology, Electrical and Electronics Engineering, Marmara University, İstanbul, Türkiye*

Abstract

Solutions obtained through multi-system approaches, ranging from simple models to detailed frameworks and from internet-enabled tools to integrated electronic structures, reflect advances in communication technologies. The primary objective of the research is to integrate the MediaPipe framework into office environments. An approach to more efficiently utilizing office-like environments has been presented using Mediapipe, an AI-based tool. The fundamental premise of this approach is that factors such as ambient light intensity and distance from the camera play a critical role in traditional image processing methods, making control processes difficult and, in some cases, impossible. Office environments are being transformed into smart systems through MediaPipe-based methods, enabling light and sound levels to be controlled with hand gestures. The proposed approach aims to increase the functionality of offices and support user productivity, while also offering a multidisciplinary perspective by integrating AI-based recognition solutions with lighting systems and image processing algorithms.

Keywords: Artificial intelligence, Image processing, Mediapipe, Lighting control, Smart lighting

Öz

İletişim teknolojilerinde çözümler; basit modeller, detaylı çerçeveler, internet tabanlı araçlar ve bütünleşik elektronik yapılar gibi çoklu sistemler aracılığıyla elde edilir. Bu araştırmanın temel amacı Mediapipe çerçevesini ofislere entegre edip kullanmaktır. Yapay zeka tabanlı bir araç olan Mediapipe kullanılarak ofis benzeri ortamların daha verimli şekilde kullanılmasına yönelik bir yaklaşım sunulmuştur. Bu yaklaşımın temel çıkış noktası, klasik görüntü işleme yöntemlerinde ortamın ışık şiddeti ve kameraya olan mesafe gibi faktörlerin kritik rol oynaması, kontrol süreçlerini zorlaştırması ve bazı durumlarda uygulanabilirliği imkansız hale getirmesidir. Bu çalışma, MediaPipe'ı kullanarak ofislerimizi akıllı bir sisteme dönüştürmekte, aydınlatma ve ses seviyesini el hareketleriyle ayarlama imkanı sunmaktadır. Bu yenilikçi yaklaşım hem kullanıcı verimliliğini arttırdı hem de ofis hayatını daha kullanışlı hale getirmiştir. Bu araştırma, tanıma çözümleri ve yapay zeka ile geliştirilmiş aydınlatma sistemiyle görüntü işleme algoritmalarını entegre etmektedir.

Anahtar kelimeler: Yapay zeka, Görüntü işleme, Mediapipe, Aydınlatma kontrolü, Akıllı aydınlatma

1 Introduction

Ongoing advancements in motion recognition technologies and hand gesture prediction have emerged as a significant area of research and application in both academia and industry due to their potential to redefine human-computer interaction. These technologies offer a more intuitive and efficient form of interaction between users and technological systems, creating a direct and meaningful communication bridge [1].

Gesture recognition technologies are being developed in the field of motion recognition systems, particularly to detect human gestures and generate appropriate responses to them. By leveraging diverse technologies, these recognition systems identify and construe physical actions, such as hand or body gestures, exchanging them into commands for interaction with digital interfaces. To get hold of the required information, a range of strategies has been deployed in the refinement of gesture recognition systems [2,3].

Hand-based control systems are advanced using sensor-based or vision-based approaches and are generally divided into two main categories. As in the example developed by Yaseen and colleagues [4] using MediaPipe, vision-based systems can recognize even complex hand movements with high accuracy.

The system developed by Saha and colleagues [5], which is based on combining accelerometer and gyroscope data, stands out as a prominent example in the field of sensor-based motion recognition. This approach enables the detection and classification of hand movements with high accuracy.

Gesture recognition systems, which leverage the power of deep learning and machine learning algorithms that emerged with the rise of artificial intelligence, have made computer-human interaction more natural and intuitive, constituting new areas of utilize from virtual reality to smart homes [6,7]. To enable the classification of gestures by learning complex patterns in visual data, advanced

* Sorumlu yazar / Corresponding author, e-posta / e-mail: karail925@gmail.com (A. Karail)
Geliş / Received: 06.11.2024 Kabul / Accepted: 23.10.2025 Yayımlanma / Published: 17.02.2026
doi: 10.28948/ngumuh.1580491

algorithms such as artificial neural networks are utilized in gesture recognition applications. These systems perform shape analysis by extracting object boundaries using the Canny edge detection method and enhance detection accuracy by identifying hand and skin regions.

Hand gesture recognition technology makes human-computer interaction more intuitive by creating a natural interface for users, thanks to computers' ability to detect and classify human hand movements. Historically, hand gesture recognition systems were constricted to fundamental camera inputs and classical image analysis; however, the incorporation of advanced techniques, count in deep learning architectures and depth-sensing modules, has markedly enhanced their recognition aptitudes. Used in many fields such as virtual reality, augmented reality and drone control, these systems are able to recognize complex hand gestures with high accuracy thanks to artificial neural networks.

However, factors such as lighting conditions, noise and different skin tones still pose some challenges.

Gesture detection and recognition are achieved through processing visual data from traditional cameras, which face difficulties due to varying lighting conditions and occlusions [8].

The tracking of hand movements is won through a multi-modal touch that integrates reflected infrared light with cameras to enhance system accuracy [9].

Depth sensors in devices like the Microsoft Kinect capture 3D information, which enhances gesture recognition by providing spatial context [10].

Hand gesture recognition systems have been used in various fields [3,11]. Technology, which finds application in a wide range of fields such as automotive, smart homes, education, retail, art and design, and accessibility, stands out with its gesture recognition feature.

Thanks to deep learning and computer vision techniques, it has become capable of performing complex tasks such as hand gesture tracking and sign language interpretation. In this way, an important step has been taken in facilitating the communication and daily lives of people with disabilities [12].

It is used in medicine to observe real-time user movements and to improve the clarity of visual representations of medical data [13].

In a study, its application in 3D games was explored [14].

Another study investigated the use of real-time hand movements to control robots, aiming to simplify robot-human interaction [15].

A separate study explored techniques for mobile device interaction that do not require sensors to be attached to the human body [16].

An auxiliary research study assessed the efficacy of gesture-based interaction within virtual reality environments and investigated the challenges inherent in gesture recognition [17]. Among gesture recognition techniques, hand gesture recognition stands out for its extensive application areas and ease of use. It is one of the most widely used methods.

A video-based hand gesture detection and tracking system developed by Fang and his colleagues [18] demonstrated the potential of this technology, while also revealing the need to address the issue of background clutter. Hand tracking and recognition were performed using methods such as optical flow and color marking. By applying a series of algorithms, six different hand movements could be accurately identified.

Innovative lighting systems stand out thanks to their adaptive lighting features, integrating sensing modules with smart control algorithms into an artificial intelligence framework. This design not only meets user needs but also reduces energy consumption by enabling the effective use of natural light.

Contribution utilities such as boost light quality, circadian rhythm regulation [19], raised energy efficiency [20], and accelerated plant enlargement [21], these systems indicate crucial prospective for deployment across industrial research, architectural design, and human physiology applications.

The establishment of sustainable working environments and the improvement of occupational health depend largely on providing stable and adequate lighting levels.

This research, which focuses on the use of MediaPipe in the management of indoor lighting environments, systematically analyzes and evaluates lighting quality by integrating artificial intelligence systems.

As in many other fields today, artificial intelligence enables significant transformation in lighting systems. By analyzing movements and activities in the office environment, it determines the most appropriate lighting level for each area and ensures that lights in unused areas are automatically dimmed or turned off completely. The lighting level and color temperature can be adjusted according to ambient conditions (time of day, weather, and user preferences).

This increases visual comfort and productivity. AI makes lighting systems smarter, more efficient and more personalized, positively impacting both the user experience and the environment.

With the help of artificial intelligence, it was aimed to perform long-term measurements, which are difficult to determine with classical methods, in a shorter time frame and with minimum error. Utilizing a camera alongside an integrated system simplified the lighting control process and proficient, avoiding problems related to light levels in conventional image processing, eliminating the need to facilitate a hand-held control button, which removes the need for interfacing with the lectern.

Consequently, the goal is to manage the lighting in an office with a solution that is both effective and visually appealing. This research developed computer programs and models using MediaPipe to recognize hand poses and control LED lighting.

In this project, MediaPipe was utilized to control the lighting system, while an AI-based system was employed to assess the indoor lighting quality. The proliferation of smart lighting systems has been driven by advancements in technology and artificial intelligence, leading to the widespread adoption of smart LED lamps, smart offices.

Artificial intelligence has enabled the rapid and cost-effective adoption of intelligent lighting systems, revolutionizing traditional lighting control and promoting energy-efficient, user-centered lighting. This integration of camera technology and the installed system enhanced the comfort and effectiveness of lighting control.

This study explores the use of the MediaPipe library in Python programming for office lighting system. This study successfully implemented image processing techniques to enhance smart lighting applications, demonstrating their effectiveness in various lighting scenarios.

Intelligent lighting systems, powered by artificial intelligence, have revolutionized traditional lighting control by offering enhanced convenience, speed, and cost-effectiveness.

Through its innovative hand gesture control interface, this research has developed a versatile solution that can benefit a wide range of users. From enhancing accessibility for the elderly and disabled to optimizing office productivity, this study's contributions are significant and far-reaching.

2 Processes and system

A machine learning pipeline powers this hand recognition solution, utilizing a model that processes the palm-detected region to accurately pinpoint hand landmarks.

By minimizing adjustments for rotation, translation, and scaling, accurate palm image cropping enables the network to allocate resources more effectively, leading to improved landmark localization accuracy.

In an office environment, a presenter can adjust light color, brightness, and sound levels through the use of hand gesture controls. Figure 1 shows the system setup diagram.

Figure 1. System setup diagram

2.1 Models

MediaPipe is an open-source platform that supports a wide range of computer vision applications such as video analysis and audio processing. It is widely used for tasks like face recognition, object tracking, and hand gesture analysis; its modular architecture allows for easy integration of different artificial intelligence models, providing application flexibility. It offers high performance optimized for different processors such as GPUs and CPUs, while its cross-platform compatibility enables models to be efficiently deployed across various platforms, from research environments to commercial applications.

The MediaPipe platform facilitates a flexible architecture that supports a wide range of hardware, from personal computers to high-performance servers and embedded systems, enabling developers to explore different application strategies and produce effective solutions in various fields. One of MediaPipe's core components is its hand model, which tracks 21 joints per hand to predict hand position in real time with high accuracy.

The MediaPipe Hands model, which represents significant progress for augmented and virtual reality applications, reliably converts user actions into commands thanks to its ability to track hand movements with high accuracy and precision, according to the work of Zhang and colleagues [22].

Patel and his colleagues [23] have developed a smart system for injured patients to recover faster. This system constantly monitors the patient's movements and instantly reports whether they are doing the exercises correctly. In this way, patients undergo a more effective rehabilitation process in line with the guidance of their physiotherapists.

MediaPipe consists of small, customizable modules called Calculators. Each calculator performs a specific step of the hand tracking process [24].

MediaPipe is a powerful tool that uses machine learning and computer vision technologies to deliver fast and accurate solutions in many different application domains. It creates accelerated inferences on GPU or CPU platforms [25].

The palm is positioned using the object detection algorithm and this data is used to identify the fingers by the hand gesture model. It uses machine learning algorithms to accurately detect and track the position and orientation of a person's hands using a camera. MediaPipe Hands can actually identify 21 different 3D landmarks in a single image. These processes include palm recognition model and hand sign recognition model [26].

MediaPipe analyzes hand gestures in real time through a multi-stage process that starts by detecting palms on a video or image. It then creates a detailed map of the hand by detecting 21 specific points on these palms and classifies different hand gestures by tracking changes in the positions of these points.

2.1.1 Hand landmark model

The hand landmark model detects landmarks in hand images using deep learning techniques. These landmarks are used to determine the 3D pose of the hand and the angles of its joints. These landmarks help to make sense of the position and shape of the hand.

The hand shape recognition method based on deformable part models proposed by Zhu et al. [27] has paved the way for advanced systems that can track the complex movements and positions of the hand more accurately and in real time.

Pavlakos and colleagues' work [28], which makes a significant contribution to 3D hand position estimation and enables the reconstruction of complex hand movements from image data, has enabled a more spatially accurate understanding of hand movements through the application of deep learning techniques.

By developing a deep learning-based architecture that integrates the local feature extraction capabilities of CNNs with the temporal sequence learning abilities of RNNs, Lai and Yanushkevich [29] have made significant progress in the field of dynamic hand gesture recognition with this approach.

The method used by León et al. [30] merges depth sensors with CNNs for hand tracking in different lighting environments and hand configurations.

Xu and his team [31] acquired more faultless and trustworthy outputs by combining the power of RGB and depth data in 3D hand sign detection.

Biswas et al. [32] developed a real-time hand tracking system on the MediaPipe framework by combining the ability of CNNs to extract local features from images with the power of spatio-temporal features to capture motion information.

In this study, which is an important step in the field of hand signal recognition, different aspects of hand movements were analyzed by generating three different outputs. It accurately estimated the 3D coordinates of 21 joint points within the detected hand regions. The model generates a coherent internal representation by transforming hand poses into an abstract mathematical structure. In this way, it can accurately predict the joints in partially visible hands by filling in missing information and in self-closing hands. Figure 2 illustrates the three outputs generated by the model.

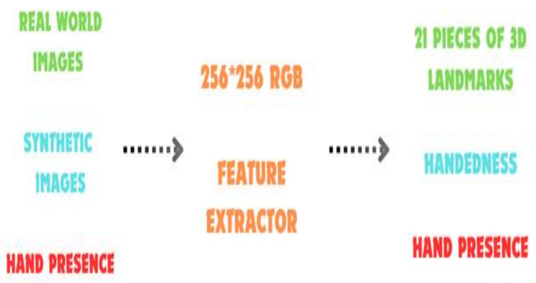

Figure 2. The design of a hand landmark model

1. Each of the 21 hand landmarks is defined by its x, y, and depth coordinates relative to the others.
2. A hand flag indicates the visibility of a hand in the input image, representing its likelihood of being present.
3. The system incorporates a binary classification to distinguish between left and right hands.

The hand landmarks depicted in Figure 3, obtained from MediaPipe's official documentation, were used to define the commands [33].

The specified architecture is used to predict the 21 landmarks, with their coordinates learned from real-world images and datasets [34].

To resolve tracking issues, an additional output model has been implemented [35].

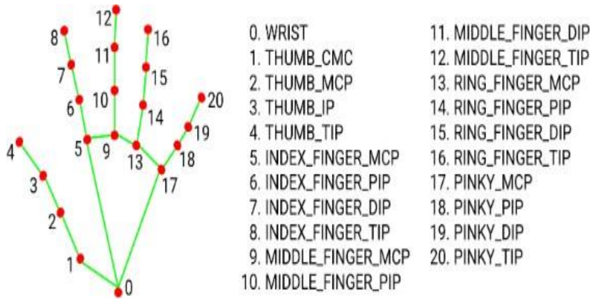

Figure 3. MediaPipe hand landmarks

The primary goal of this output is to determine the probability of a well-aligned hand being located within the designated image region. This information can be used to address tracking challenges and improve the overall accuracy of the hand tracking system.

MediaPipe's hand landmark feature enables the detection of key points on hands within images. By utilizing this functionality, salient hand points can be identified, and visual effects can be applied to them. The model, which determines the position and boundaries of hands detected in both image and world coordinates, has the ability to work on both continuous video streams and static images. This hand detection process, which can recognize multiple hands simultaneously, including left and right hands, can be performed in real time or video-based using instant detection techniques.

2.2 Benefits of hand gesture lighting system control

Hand gesture-controlled lighting technologies developed in recent years have significantly improved system usability as well as increasing user interaction. Hand gesture-controlled lighting technologies offer notable advantages, particularly in terms of user comfort and operational efficiency.

Gesture-based control methods, which replace traditional keys and buttons, eliminate the need for physical contact in lighting systems, providing hygienic advantages and offering superior interaction comfort for users.

The fact that hand gestures are a natural form of communication, combined with customizable structures that can adapt to different lighting needs and preferences, facilitates users' quick and effective adoption of the systems. By eliminating the need for buttons or switches, it makes control more accessible and practical for individuals with physical disabilities.

These systems provide selective lighting based on space usage characteristics and light intensity requirements, increasing energy efficiency and preventing unnecessary consumption, thanks to their integration with ambient light sensors, artificial lighting automatically adjusts according to existing light levels.

Motion-based lighting modulation which enables precise adjustments in dimly lit environments, maintains its applicability especially in dark nighttime settings while eliminating the need for physical buttons and switches, thereby lending the space a modern and aesthetically pleasing design approach.

As a method that enriches the user experience and increases the functionality of systems, the use of hand gestures as input in lighting systems stands out; it also contributes to optimizing power consumption, enhancing operational safety, and ensuring integration compatible with visual design, becoming a fundamental component of smart home applications.

2.3 *User acceptance of gesture lighting systems*

With the development of smart home technologies, motion-based lighting control has become an increasingly common and preferred feature in recent years. While these systems have the potential to enhance the user experience, they also bring some challenges. How quickly and easily users accept this technology depends on a variety of factors.

Factors affecting user acceptance are given below:

- Ease of use
- Technological literacy
- Aesthetic appearance
- Reliability

It is important how easy the system is to learn and whether it allows the user to move naturally. How quickly commands are recognized and responded to directly affect user satisfaction. As soon as the commands are given, they are instantly recognized and the lighting and sound system is controlled. Users are very satisfied with this control.

The user's aptitude for technology affects the speed at which they learn new systems. Being open to new technologies accelerates the acceptance process. The harmony of the system with the interior space positively affects the user experience. The system works in harmony with the office environment.

Since it is important to provide a simple and understandable interface rather than complex settings, a simple and understandable interface was used. A key feature of the system is its user-centric design, which provides the flexibility for individuals to modify system behavior based on their personal preferences.

The implementation of gesture-based lighting control technology has the potential to meaningfully enhance user experience. Its accomplished adoption is contingent upon addressing user necessities and expectations. Therefore, the recommendations and considerations outlined previously are crucial for increasing user acceptance of this technology.

3 Control system details

3.1 *Model construction*

To properly train the palm detector, one of three important datasets must be selected. These datasets are: the Google internal dataset, the Synthetic dataset, and the Wild dataset. In this study, agreed to utilize the Wild dataset because the inherent assortment and accurate localization of the dataset are critical for robust performance across a wide range of computing environments. The option to train the palm detector on bounding boxes is favorable because it maintains a precise and compact representation of the spatial and dimensional properties of the palm. Furthermore, Feature Pyramid Network (FPN) benefits as an efficacious tool to boost the situational awareness of a model, which is vital for location-based tasks.

Since FPN advance the contextual consciousness of the system by combining multi-scale features, FPN facilitates the production of more accurate and reliable results [22].

Habibie et al.'s [36] study used various real-world datasets, such as the Wild dataset, and this work highlights the importance of using real-world data to address limitations inherent in pose estimation models. By rising the number of training sessions in diverse environments, the model can generalize better and make more accurate predictions.

The crucial role of the Google dataset in the development of real-time pose estimation models was highlighted in the research conducted by Toshev and Szegedy [37]. Their findings show that the Google dataset significantly cultivates both the accuracy and computational efficiency of models required for practical, real-world applications.

The work of Dhulipala et al. [38] was a significant milestone in the field of hand gesture recognition systems because it highlighted the usefulness of synthetic data in hand gesture recognition system.

The basic stages of an object detection and hand pointing system are described below:

Object detection consists of three basic operations: Determining bounding boxes around objects of interest, refining these boxes through resampling, and using a robust classifier to accurately categorize objects. An example of the use of the bounding box in these is that when the palm is placed inside the bounding box, a hand pointing model can precisely identify 21 key points on the hand using regression.

The model's primary function is to create a detailed reference point map for hand kinematics by identifying 21 critical points on a hand in real time. This process also includes tasks such as detecting the presence of a hand, determining whether it is right-handed or left-handed, and classifying its orientation. Simon and colleagues [34] further improve the accuracy of this critical point detection by utilizing various image processing techniques.

MediaPipe concludes the first stage of the process by converting raw image data into a format that can be processed by the model, thereby enabling the achievement of results with higher accuracy and efficiency in subsequent stages.

MediaPipe's trained model, which provides a powerful foundation for numerous applications that analyze hand and finger movements and stands out from others with its wide range of applications, also forms its second stage with feature extraction that enables complex information in images to be represented in a meaningful way.

MediaPipe, which has the ability to track by extracting the reference coordinates of hand and finger movements, is a powerful tool that can be used in many different fields. The model enables the spatially accurate positioning of hand and finger movements.

The 3D points manufactured by the model are symbolized by x, y, z coordinates, as given in the Equation (1).

$$p_i = (x_i, y_i, z_i) \tag{1}$$

p_i: landmark point
x_i, y_i, z_i: coordinates

MediaPipe's touch, bottomed on mathematical and geometric calculations, enables a deeper understanding of hand movements and gestures and provided accurate results.

The angle between two fingers is calculated using trigonometric functions such as the dot product or the law of cosines, as seen Equation (2).

$$\cos Q = \frac{(p_2 - p_1) * (p_3 - p_1)}{\|p_2 - p_1\| \ \|p_3 - p_1\|} \tag{2}$$

p_1, p_2, p_3: The coordinates of three different landmark points.

The Euclidean formula provides detailed information about the position of the fingers, their size and their relationship to each other, as seen Equation (3).

$$d = (\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{3}$$

The system is controlled by three main hand gestures. Each action provides access to the main menus, and once in the main menu, access to the submenus is possible.

By combining mathematical calculations, such as Euclidean distance and angle computations, with algorithmic approaches, Mediapipe effectively analyze hand movements and positions in real-time, providing accurate and reliable results. By integrating a hand recognition system, this research demonstrates a practical application of human-computer interaction in the field of smart office technology, specifically for lighting control.

There are 3 main menus:

1. Led color control
2. Led brightness control
3. Sound system control

The hand recognition solution enabled basic control of both the lighting and sound systems, allowing users to turn them on and off, increased and decreased illumination level, sound system. Furthermore, by adjusting the position of the thumb, users could seamlessly transition between the colors on the color scale. A 270-degree finger movement facilitated the smooth change of LED color. A simple up-and-down movement of the thumb and index finger was used to effectively control the LED brightness and sound system, offering a user-friendly and intuitive interface.

The first operation is to control the lamp which involves turning the lamp on and off. The first menu was used to change the color temperature, which allowed the LED to appear in a range of colors. Changes to the lighting level, including increasing and decreasing, controlled through the second menu. The control for turning the computer's sound system on/off and adjusting the volume increase/decrease was done in menu 3. On the screen, there are three buttons that are numbered. These numbered buttons allow for the selection of the desired menu. The first button is used to select the LED color, second button is used to control the LED brightness level, and third button is used to control the sound system.

3.2 Led control system

The study incorporated hand recognition technology to control the smart lighting system, enabling users to interact with the system through hand gestures. The core of this method is to develop a control strategy that operates the lighting system, allowing it to be turned on and off, with dimming adjustments from 0% to 100% to increase brightness and from 100% to 0% to decrease it. Furthermore, color transition was achieved by translating the hand positions to the corresponding colors on the color scale, which was performed with the gesture of the thumb and index finger defined for the system. Hand gestures were also used to control the sound system. Within this framework, the commands obtained in this part have been turned into digital data. The control strategy for the smart lighting system includes the following steps, respectively:

1. Hand recognition with MediaPipe
2. Transformation of commands into digital information
3. Digital analogue conversion
4. Analog information inputs
5. Lighting system

3.3 Results

The project was divided into four phases: building a lighting dataset utilizing data from the MediaPipe hand detection system, the study includes the definition of specific codes for each movement and the development of software to control the lighting system accordingly. Following the creation of lighting control datasets for an office, MediaPipe was used for hand gesture recognition, gestures were defined, and the program used these gestures for lighting control.

Gesture lighting control adds a modern touch to living spaces by personalizing the user experience. This technology is designed with the needs and expectations of users at the center, increasing both comfort and energy efficiency.

Artificial intelligence-supported lighting systems in offices offer a customized environment according to the needs of employees. By analyzing movements, these systems create a pleasant workspace that does not tire the eyes and optimize energy consumption. Thus, both the health and happiness of employees are supported and an environmentally friendly approach is adopted.

The lighting control is based on the comparison between points 0 and 4 on the hand and used hand landmarks, determined whether the Led is on or off according to their position.

The following test parameters were used to obtain the results:

1. Camera detection distance
2. The minimum amount of ambient light level for the camera to detect
3. Tests performed during day and night
4. Different color temperatures

A common standard is the International Commission on Illumination (CIE) guidelines, which recommend a minimum illuminance level of 300 lux for general office areas.

During the tests, the camera was measured at different distances and different illuminance level as seen in Table 1, Table 2 and Table 3 evaluated with different illuminance levels and environments with different color temperatures. Each test included 100 trials, resulting in the following data. According to the test results for the office environment, the following system operation success rates were obtained.

Table 1. Camera detection distance test results

Brightness Level of the Office	Camera Detection Distance	Menu 1	Menu 2	Menu 3
300 lux	1.5 m	% 98	% 97	% 98
400 lux	1.5 m	% 97	% 95	% 95
500 lux	1.5 m	% 97	% 96	% 96

Table 2. Camera detection distance test results

Brightness Level of the Office	Camera Detection Distance	Menu 1	Menu 2	Menu 3
300 lux	2.5 m	% 98	% 96	% 95
400 lux	2.5 m	% 96	% 95	% 95
500 lux	2.5 m	% 94	% 94	% 95

Table 3. Camera detection distance test results

Brightness Level of the Office	Camera Detection Distance	Menu 1	Menu 2	Menu 3
300 lux	3.5 m	% 96	% 94	% 92
400 lux	3.5 m	% 95	% 94	% 94
500 lux	3.5 m	% 92	% 93	% 94

In this study, different illumination values were tested based on a fixed detection distance, as well as detection distances based on a fixed illumination value. These results demonstrate the high success rate of the model and further demonstrate its robustness.

Movement for camera detection distance, menu 1 (color change control), the success rate for 300 lux illumination level was 97.34 percent on average. Camera detection distance, movement for menu 1 (color change control), for controlling the lighting system based on hand gestures, for 400 lux illumination level was 96 percent on average. Movement for camera detection distance, menu 1 (color change control), for controlling the lighting system based on hand gestures, for 500 lux illumination level was 94.34 percent on average.

For the 2nd menu, the success rate for the 300 lux illumination level was 95.67 percent on average. Success rate in menu 2 for gesture-based lighting system control, 94.67 percent on average for 400 lux lighting level. Success rate in menu 2 for gesture-based lighting system control, 93.34 percent on average for 500 lux lighting level.

For the 3rd menu (sound system control), the success rate for 300 lux illumination level was 95 percent on average. For the 3rd menu (sound system control), the success rate for 400 lux illumination level was 94.67 percent on average. For the 3rd menu (sound system control), the success rate for 500 lux illumination level was 95 percent on average.

Table 4 shows the test results for low illumination levels for the office.

Table 4. Office illuminance level test results

Minimum office brightness level	Menu 1	Menu 2	Menu 3
1 lux	% 0	% 0	% 0
1.5 lux	% 0	% 0	% 0
2 lux	% 81	% 79	% 76
3 lux	% 84	% 83	% 80

As shown in Table 4, luxmeter measurements were taken to determine the minimum illumination level so that the system could be detected and controlled by the camera. Values of 1 lux, 1.5 lux, 2 lux, and 3 lux were used to determine the minimum illumination level. The lower limit of the illumination level required for system operation was determined to be 2 lux.

Table 5 shows the office minimum illuminance test results.

Table 5. Office minimum illuminance test results

Minimum office brightness level	Menu 1	Menu 2	Menu 3
2 lux	% 81	% 79	% 76

As shown in Table 5, the minimum illuminance level was measured with a luxmeter so that the system could be detected and controlled with the camera. The lower limit of the illumination level required for system operation is 2 lux.

As can be seen in Table 6, the maximum distance for camera detection was measured as 3.8 meters.

Table 6. Camera detection maximum distance test results

Brightness Level of the Office	Camera Detection Max Distance	Menu 1	Menu 2	Menu 3
300 lux	3.8 m	% 95	% 85	% 86
400 lux	3.8 m	% 94	% 84	% 84
500 lux	3.8 m	% 95	% 84	% 83

4 Conclusion

Hand gesture recognition has transitioned from using simple feature-based methods to employing complex deep learning techniques. They are commonly utilized in user interfaces, gaming, healthcare, and several other sectors. Although there have been considerable advancements, issues like accuracy, real-time processing, and personalization continue to pose challenges. Research moving forward will likely focus on integrating cutting-edge technologies and increasing robustness and efficiency.

Thanks to the method we proposed, a new perspective has been gained in the control of environments such as conference halls, and at the same time, a higher efficiency has been achieved compared to classical image processing methods.

MediaPipe has emerged as the framework for computer vision and machine learning tasks. Mediapipe's knack to seamlessly operate across various platforms, deliver accurate real-time outputs, and deliver an intuitive user experience has solidified its position in the industry. The aim of this study is to highlights its widespread utilize in fields such as hand gesture recognition, pose estimation, face detection, and object tracking. The creation of more accurate and well-rounded hand gesture recognition systems is a direct result of major progress in the fields of image processing and machine learning.

This study demonstrates that advanced computational tools such as artificial intelligence and MediaPipe can be effectively applied in office environments and that lighting and sound systems can be controlled with high efficiency. Users can easily manage the office's lighting and sound systems using defined hand gestures; the primary goal of the study is to provide an intuitive and user-friendly experience using these gestures. In this context, an interaction model has been developed by assigning a specific command, such as adjusting lighting or sound levels, to each unique hand gesture. Leveraging artificial intelligence models and MediaPipe's effective hand position prediction capabilities, the study offers an innovative and practical solution for office lighting management by establishing a direct and reliable relationship between users' hand gestures and lighting commands.

The results acquired in the study were verified with an experimental study with 100 iterations to verify the reliability of the system and the effectiveness of the control commands.

Conflict of Interest

The authors declare that there is no conflict of interest.

Similarity rate (iThenticate): 8%

References

[1] H. Mokhtar and M. Pramod, Brightness factor matching for gesture recognition system using scaled normalization. International Journal of Computer Science and Information Technology. 3, 2011. https://doi.org/10.5121/ijcsit.2011.3203

[2] W. Simei, L. Marcus, K. Susumu and I. Akira. A rotation invariant approach on static-gesture recognition using boundary histograms and neural networks, 2003.

[3] J. J. LaViola, A survey of hand posture and gesture recognition techniques and technology, 1999.

[4] Yaseen, O.-J. Kwon; J. Kim, S. Jamil, J. Lee and F. Ullah, Next-Gen dynamic hand gesture recognition: MediaPipe, Inception-v3 and LSTM-Based enhanced deep learning model. Electronics, 13, 3233, 2024. https://doi.org/10.3390/electronics13163233

[5] U. Saha, S. Saha, M. T. Kabir, S. A. Fattah and M. Saquib, Decoding human activities: Analyzing wearable accelerometer and gyroscope data for activity recognition. IEEE Sensors Letters, vol. 8, no. 8, pp. 1-4, Art no. 7003904, 2024. doi: 10.1109/LSENS.2024.3423340.

[6] R. Z. Khan, N. A. Ibraheem. Survey on gesture recognition for hand image postures, International Journal of Computer and Information Science, Vol. 5(3), 2012. https://doi.org/10.5539/cis.v5n3p110

[7] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture, Elsevier, Computer Vision and Image Understanding, Vol. 81, pp. 231–268, 2001. https://doi.org/10.1006/cviu.2000.0897

[8] C. Stauffer and W.E.L.Grimson, Learning patterns of activity using real-time tracking., 747-757., IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 747 - 757, 2000. https://doi.org/10.1109/34.868677

[9] M. Z. Islam, L. Yu, H. Abuella, J. F. O'Hara, C. Crick and S. Ekin, Hand gesture recognition through reflected infrared light wave signals, 2023 10th International Conference on Electrical and Electronics Engineering (ICEEE), Istanbul, Turkiye, pp. 1-5, 2023. doi: 10.1109/ICEEE59925.2023.00008.

[10] M. Tang, Recognizing hand gestures with microsoft's kinect. Palo Alto: Department of Electrical Engineering of Stanford University: [sn] 23, 2011.

[11] S. Mitra, and T. Acharya. Gesture recognition: A survey. IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 37 (3), pp. 311- 324, 2007. https://doi.org/10.1109/TSMCC.2007.893280.

[12] A. Silvia, and N. L. Husni, Hand contour recognition in language signs codes using shape based hand gestures methods. 1st International Conference on Computer Science and Engineering, Sriwijaya University, Palembang, Indonesia, September 2014. https://www.semanticscholar.org/paper/Hand_Contour_Recognition_in_Language_Signs_Codes_Silvia_Husni/4d3bd0bc9615482b808e5763824ef4b65717484f?utm_source=direct_link

[13] J. Wachs, H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith and J. Handler, A real-time hand gesture interface for medical visualization applications. Applications of Soft Computing. Advances in Intelligent and Soft Computing, 36. Springer, Berlin, Heidelberg, 2006. https://doi.org/10.1007/978-3-540-36266-1_15

[14] H. Kaur and J. Rani. A review: Study of various techniques of hand gesture recognition. In 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), pp. 1–5, 2016. https://doi.org/10.1109/ICPEICES.2016.7853514

[15] Q. Gao, Y. Chen, Z. Ju, and Y. Liang. Dynamic hand gesture recognition based on 3d hand pose estimation for human-robot interaction. IEEE Sensors Journal, PP:1–1, 2021. https://doi.org/10.1109/JSEN.2021.3059685

[16] A. Shimada, T. Yamashita, and R. Taniguchi. Hand gesture based tv control system towards both user-amp; machine-friendly gesture applications. In The 19th Korea-Japan JointWorkshop on Frontiers of Computer Vision, pp. 121–126, 2013. https://doi.org/10.1109/FCV.2013.6485473

[17] Y. Li, J. Huang, F. Tian, H. Wang, and G. Dai. Gesture interaction in virtual reality. Virtual Reality & Intelligent Hardware, 1(1), 84–112, 2019. https://doi.org/10.3724/SP.J.2096-5796.2018.0006

[18] Y. Fang, K. Wang, J. Cheng and H. Lu, A real-time hand gesture recognition method. In 2007 IEEE International Conference on Multimedia and IEEE, Expo pp. 995-998, 2007. https://doi.org/10.1109/ICME.2007.4284820

[19] J.H. Oh, S.J. Yang and Y.R. Do. Healthy, natural, efficient and tunable lighting: four-package white leds for optimizing the circadian effect, color quality and vision performance, Light Sci. Appl. 3 (2) e141, 2014. https://doi.org/10.1038/lsa.2014.22

[20] R.F. Karlicek. Smart lighting-beyond simple illumination. IEEE Photonics Society Summer Topical Meeting Series, IEEE 147–148, 2012. https://doi.org/10.1109/PHOSST.2012.6280791

[21] G.D. Massa, H.-H. Kim, R.M. Wheeler and C.A. Mitchell. Plant productivity in response to led lighting. HortScience 43 (7), 1951–1956, 2008. https://doi.org/10.21273/HORTSCI.43.7.1951

[22] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C. Chang and M. Grundmann, MediaPipe hands: On-device real-time hand tracking, 2020. https://doi.org/10.48550/arXiv.2006.10214

[23] S. Patel, H. Park, P. Bonato, L. Chan and M. Rodgers A review of wearable sensors and systems with application in rehabilitation. J NeuroEngineering Rehabil 9, 21, 2012. https://doi.org/10.1186/1743-0003-9-21

[24] C. Lugaresi, J. Tang, H. Nash, C. Mc-Clanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. G. Yong, J. Lee, Wan-TehChang, W. Hua, M. Georg, and M. Grundmann. Mediapipe: A framework for building perception pipelines. Volume abs/1906.08172, 2019. https://doi.org/10.48550/arXiv.1906.08172

[25] I. Grishchenko and V. Bazarevsky, Mediapipe holistic—simultaneous face, hand and pose prediction, on device. Google Research, Retrieved June 15 (2020): 2021. https://research.google/blog/mediapipe-holistic-simultaneous-face-hand-and-pose-prediction-on-device/

[26] Hand landmarks detection guide for Python. https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker/python ,Accessed 2025

[27] M. Zhu, C. Zhang, J. Wang, L. Sun and M. Fu, Robust hand gesture recognition using a deformable dual-stream fusion network based on CNN-TCN for FMCW radar. Sensors, 23, 8570, 2023. https://doi.org/10.3390/s23208570

[28] G. Pavlakos, D. Shan, I. Radosavovic, A. Kanazawa, D. Fouhey and J. Malik, Reconstructing hands in 3d with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9826-9836, 2024. https://doi.org/10.48550/arXiv.2312.05251

[29] K. Lai and S. N. Yanushkevich, CNN+RNN depth and skeleton based dynamic hand gesture recognition, 24th International Conference on Pattern Recognition (ICPR), pp. 3451-3456, Beijing, China, 2018. https://doi.org/10.1109/ICPR.2018.8545718.

[30] D. G. León, J. Gröli, S. R. Yeduri, D. Rossier, R. Mosqueron and O. J. Pandey, Video hand gestures recognition using depth camera and lightweight CNN, IEEE Sensors Journal, 22(14), 14610-14619, 2022. https://doi.org/10.1109/JSEN.2022.3181518

[31] C. Xu, J. Zhou, W. Cai, Y. Jiang, Y. Li and Y. Liu, Robust 3D hand detection from a single RGB-D image in unconstrained environments. Sensors, 20, 6360, 2020, https://doi.org/10.3390/s20216360

[32] S. Biswas, A. Nandy, A. K. Naskar and R. Saw, MediaPipe with LSTM Architecture For Real-Time Hand Gesture Recognition. In: Kaur, H., Jakhetiya, V., Goyal, P., Khanna, P., Raman, B., Kumar, S. (eds) Computer Vision and Image Processing. CVIP 2023. Communications in Computer and Information Science, vol 2010. Springer, Cham, 2024, https://doi.org/10.1007/978-3-031-58174-8_36

[33] Hand landmarks detection guide. https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker , Accessed 13 March 2003.

[34] T. Simon, H. Joo, I. A. Matthews, and Y. Sheikh. Hand keypoint detection in single images using multiview bootstrapping. CoRR, abs/1704.07809, 2017. https://doi.org/10.48550/arXiv.1704.07809

[35] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann. Real-time facial surface geometry from monocular video on mobile gpus. CoRR, abs/1907.06724, 2019. https://doi.org/10.48550/arXiv.1907.06724

[36] I. Habibie, W. Xu, D. Mehta, G. Pons-Moll and C. Theobalt, In the wild human pose estimation using explicit 2D features and intermediate 3D representations, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10897-10906, Long Beach, CA, USA, 2019. https://doi.org/10.1109/CVPR.2019.01116

[37] A. Toshev and C. Szegedy, DeepPose: Human pose estimation via deep neural networks, 2014 IEEE

Conference on Computer Vision and Pattern Recognition, pp. 1653-1660, Columbus, OH, USA, 2014. https://doi.org/10.1109/CVPR.2014.214.

[38] P. V. Dhulipala, S. Oncken, S. Claypool and S. Kalafatis, Comparison of training for hand gesture recognition for synthetic and real datasets. Natural Language Processing, Information Retrieval and AI Trends 2025, https://doi.org/10.5121/csit.2024.150214