

Automated Opencv-Based Presentation Controller

G Ramesh^{1*}, Karanam Madhavi¹, P. Jaya Prakash Reddy¹, Abhiram Pedamallu¹, Raj Kumar²

¹Département of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad India.

²Uttaranchal School of Computing Sciences, Uttaranchal University, Dehradun, 248007, India

Abstract. Présentations have proved to be an effective way of communication in today's world. Effective présentations are essential as they express one's idea or analysis in professional meetings. A keyboard or a remote-controlled physical device is generally used to operate the présentations. The display can be performed using hand gestures by utilizing the features provided by Computer Vision (CV). The slides in the présentation can be navigated using specific hand gestures for each operation, like navigating forward or backward in slides. This model also allows the user to virtually draw on the screen in real-time by using their hands. So this model helps eliminate the use of a physical device and is cost-effective. This automation results in a software product that works as an efficient controller. Using this, the user can operate the slides efficiently without any physical contact. This automated model is cost-effective, easy to use, and doesn't require any physical device. This can also see tremendous market potential as présentations have become necessary in every field today. Many features and gestures can be added to further development of this model.

1. Introduction

This era of the world is a very innovative and competitive place as there are many new inventions and discoveries at a fast and incredible rate. The nature of humans is to invent tools to help ease things and their surroundings. So, in this corporate and business-driven world, an effective way of communication in companies is in the form of présentations. Computer Vision (CV) is the field of Artificial Intelligence (AI) that enables computers and systems to detect and recognize images, text, videos, and more. Humans have the gift of analyzing any situation and reacting to it based on past experiences. If AI helps the computer to think, CV enables it to see and understand. This project aims to present an effective and

* Corresponding author: ramesh680@gmail.com

reliable software system that can be used to accurately control a presentation using simple hand gestures.

Our contributions to this paper are as follows:

- A literature review is made to identify the main cause of this project.
- The working and ways to understand it.
- Summarize and conclude the daily uses of the controller.

2. Literature Review

This section reviews various use cases of Computer Vision that helped build this project. Hu, et al. [1] proposed a methodology for local relation networks for image recognition in computer vision applications. This paper introduces a novel image feature extractor known as the "local relation layer." This innovative layer is designed to enhance the aggregation of information from local pixel pairs within an image. The unique aspect of this approach lies in its ability to dynamically compute aggregation weights by considering the compositional relationship between these pixel pairs.

By incorporating a relational perspective, the local relation layer is capable of combining visual elements in a more intelligent and effective manner. This, in turn, results in the creation of higher-level entities with improved efficiency, thereby contributing to enhanced semantic inference capabilities.

In essence, the local relation layer offers a fresh approach to feature extraction in images, leveraging its adaptive aggregation weights and compositional understanding of local pixel pairs. This approach holds promise for advancing the field of image analysis and contributing to more efficient and accurate semantic inference tasks.

Tejaswini Priyanka, et al. [2] have been proposed Convolution Neural Network (CNN) Algorithm to effectively capture and identify the user's facial features from a real-time webcam feed. This enables the model to then analyze these facial cues such as the movements of lips and eyes to accurately detect the user's emotions.

Rautaray [3] explored real-time hand gesture recognition. The core focus of this research initiative revolves around the practical implementation of an application that harnesses the capabilities of computer vision algorithms and advanced gesture recognition techniques. The ultimate objective of this endeavor is to create an accessible and affordable interface device, which facilitates interaction with virtual objects through the intuitive use of hand gestures. This innovation aims to democratize the ability to engage with virtual environments by providing a cost-effective solution that recognizes and interprets gestures, thereby enabling users to seamlessly manipulate and navigate within these digital realms. The prototype architecture of the application comprises a central computational module that applies the cam shift technique for tracking hands and their motions.

Huang, et al. [4] studied a model-based hand gesture recognition system. This paper presents a comprehensive model-based approach for recognizing hand gestures, encompassing three distinct phases: feature extraction, training, and recognition. During the feature extraction stage, a synergistic approach is adopted, combining both spatial (edge-based) and temporal (motion-based) information from each frame to derive feature images. Subsequently, the training phase leverages two distinct methodologies: principal component analysis (PCA) is employed to capture spatial shape variations, while hidden Markov models (HMM) are utilized to encapsulate temporal shape variations. This combined framework allows for a robust and nuanced understanding of hand gestures, effectively enabling the system to learn and differentiate various gestures for subsequent recognition tasks.

Zhao, et al. [5] researched exploring self-attention for image recognition. Recent work has shown that self-attention can be a fundamental building block for image recognition models. We examine variations of self-attention and assess their effectiveness for image recognition. We consider two forms of self-attention. One is pairwise self-attention, which generalizes standard dot-product attention and is fundamentally a set operator. The other is patchwise self-attention, which is strictly more powerful than convolution.

In [6] the authors developed an innovative sequential model aimed at identifying blurry or duplicated images in a systematic manner. This model automatically organizes these identified images into their designated folders, which are generated by the model itself. Users have the option to review these folders at their convenience, or they can opt to remove them if they prefer. Additionally, the model generates a video compilation showcasing the original, high-quality images, excluding any duplicates. This video provides users with a rapid and effortless way to assess and browse through their images. In [7] the authors explored a novel morphological filtering algorithm has been introduced to address the challenge of low contrast in Scanning Electron Microscopic (SEM) images of composite materials. The utilization of the discrete wavelet transform, while common, has been observed to lead to a decline in the quality and robustness of the watermarked image. This study adopts the lifting wavelet transform technique as a substitute for the discrete wavelet transform presented in [8].

Oudah, et al. [9] reviewed hand gesture recognition based on computer vision. Hand gestures constitute a vital mode of nonverbal communication, holding applicability across diverse domains like facilitating communication among the hearing-impaired, steering robot operations, human-computer interaction (HCI), automating home systems, and advancing medical applications. In the realm of research, a plethora of techniques have been explored for the analysis of hand gestures, encompassing avenues such as instrumented sensor technology and computer vision methodologies. These approaches have paved the way for the categorization of hand signs into various dimensions, including posture and gesture distinctions, as well as a dichotomy between dynamic and static gestures, or even a synthesis of both.

This paper undertakes a comprehensive exploration of the body of literature concerning hand gesture techniques. Its objective is to illuminate the strengths and constraints inherent to these techniques within varying contextual settings. By delving into this analysis, the paper effectively unveils the nuanced landscape of hand gesture research, shedding light on the merits and limitations of different methodologies in diverse scenarios.

2.1. Methodology

This section provides the methodology to build the project of creating a Presentation Controller. Figure 1 shows the conceptual framework that is followed to complete this project.

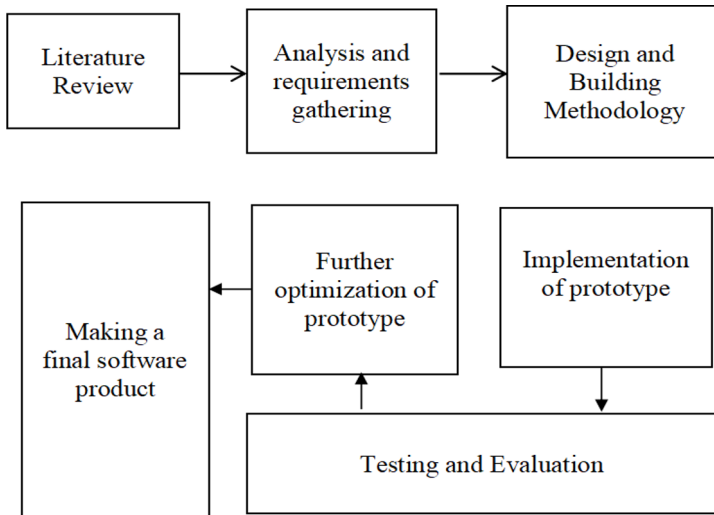


Fig 1: Conceptual framework

With the help of a literature review, valuable insights are gathered to analyze the problem. We followed the conceptual framework as shown in Figure-1 to complete this project. Analysis and requirements gathering helped us recognize the needs of the project. The working methodology of the project is discussed further.

2.2. Setting up the webcam and required OpenCV modules

The first step in making this project is setting up the software environment and a programming language to write the code in. We used Python programming language and the PyCharm platform. The necessary slides for the presentation are gathered. The built-in modules like cv zone are imported, which will help write the code for hand gestures. The code is written based on the following algorithm.

Algorithm-1: Setting up camera and listing presentation images

1. Begin
2. Importing HandDetector from cvzone.HandTrackingModule, cvzone, numpy
3. Declare parameters width, height, folderpath
4. Print the images for presentation
5. Set the width and height
6. Connect the webcam
7. End

Algorithm-1

As presented in Algorithm-1, preparing and printing images is done along with importing valuable libraries. To get started, the webcam is set up by adjusting the width and height parameters. We then proceed to the second algorithm, which continues the preceding one.

2.3. Implementation of gestures and virtual pointer

Algorithm-2: Implementation of gestures and drawing

1. Begin
2. Detecting hands by setting max hand as 1 and detection confidence at 80%
3. Creating gesture threshold line
4. Preparing hand gesture controls
5. Setting gestures to be detected only above threshold line
6. Creating a circular pointer to draw
7. Preparing an undo command
8. End

Algorithm-2

As presented in Algorithm-2, first, the code is written to detect human hands. We have set the confidence of seeing a pointer to 80% and the maximum number of hands to be taken as input to 1. This proposed project will stand out because of the accuracy of the product. By creating a threshold line at approximately 40 percent below the top of the screen, one can move their hands freely while presenting. This is important because speakers generally move their hands while speaking to a group. There are hand gesture controls by using which the presenter can control the presentation. By pointing our thumb finger, we can navigate to the previous slide and navigate to the next slide by pointing a little finger.

2.4. Working procedure of the proposed system

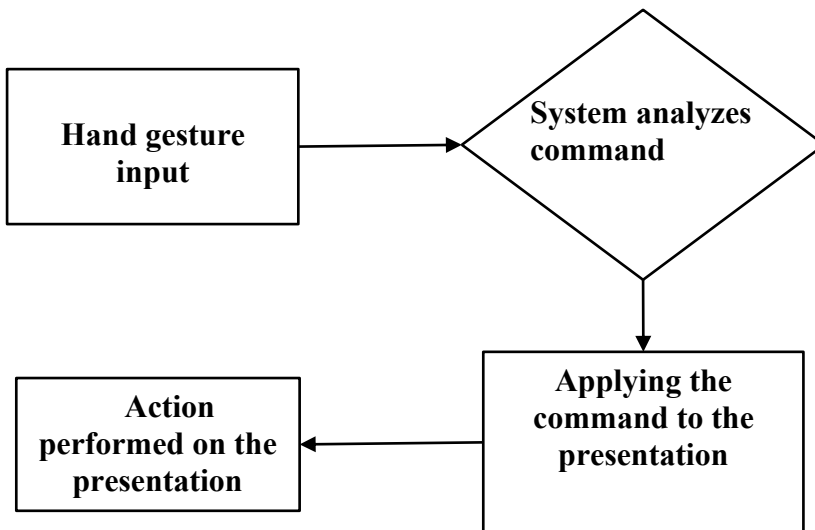


Fig 2: Working of the system

The working of the system is depicted in the figure 2. Also, in this project, we made virtual drawings and undo features using hand gestures. The presenter can access the pointer using

the index and middle fingers and also virtually draw using the index finger. However, to undo the previous command, one can use the first three fingers of their hand to point. In this way, the project is built by effectively detecting and implementing all the discussed features. This project makes it easier and simpler to give presentations. Hence, this results in a cost-effective, easy-to-use, reliable, and accurate presentation control system.

In our study, we evaluated the performance of a presentation controller using computer vision techniques focusing on hand gesture detection. The system was tested using a dataset of 50 presentations, each with an average length of 15 minutes. The system could accurately detect and track the presenter's hand gestures, such as moving forward and backward through slides, with an average accuracy of 80%.

The results of our study demonstrate that hand gesture detection is a viable method for controlling presentations using computer vision techniques. Our study's high accuracy of hand gesture detection indicates that the system can accurately respond to the presenter's commands.

However, it is also notable that, in this study, we didn't implement facial expression detection, which is also a promising area for presentation control. Facial expressions can provide additional cues about the presenter's intent and can be used to improve the system's overall performance. We plan on adding facial expression detection in our future work, which is expected to enhance the system's capability further.

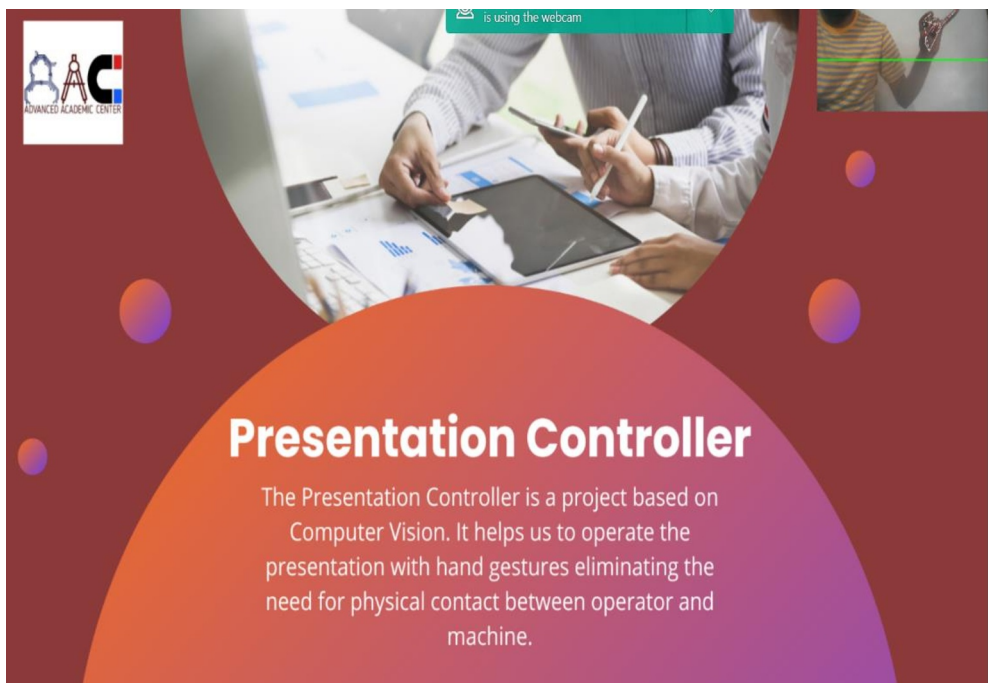


Fig 3: Sample presentation

As shown in figure-3, one can give a specific gesture as input which can be seen in the top right corner. In this figure, the input is the "move to the right" gesture shown using a little finger. The following slide in the presentation is displayed using this gesture. The sample presentation is taken, for example.



Fig 4: Sample presentation

As shown in figure-4, a virtual pointer (in red) can be inserted, and one can draw (in red) on the screen virtually by moving their index finger and the pointer using both index and middle fingers. However, it is to be noted that the changes done using the pointer will disappear when moving to the next or previous slide. There are other gestures like moving to the left using the thumb and undoing the previous command using three fingers. So, the features can be extended to a large extent in further paper development. Hence, this study finds many applications and use cases in the market.

Additionally, the testing was conducted on a specific dataset, and it may not generalize to other cases like different persons, lighting, background, etc. Further experimentation and testing in different scenarios would improve the system's robustness and make it more suitable for real-world applications.

Hence, our study shows hand gesture detection is a promising method for controlling presentations using computer vision techniques. Adding facial expression detection is expected to enhance the system's performance even more. Our study highlights the potential of computer vision techniques for improving the experience of giving presentations and paving the way for future research and development in this area.

3. Conclusion

With the help of the proposed system, one can easily navigate through a presentation using simple hand gestures without any help from a physical device like a remote. This proposed model is cost-effective, reliable, and easy to use. This software system implementation requires simple hardware like a good-quality camera or a webcam. The software can be further improved, and many more features can be added, such as facial recognition (person-based recognition to use the product). However, with the help of the current project, one can move left and right through slides in a presentation, create a pointer and virtually draw on the

screen using hand movements and undo the previous commands using simple hand gestures. This project has much potential to be improvised further and add many hand gestures. This system can also find significant market potential by which this idea can be commercialized by building a mobile application. The proposed system stands out as an innovative method by which a software product can improve the operation of a presentation.

References

1. Hu, H., Zhang, Z., Xie, Z., & Lin, S. *Local Relation Networks for Image Recognition*. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). (2019). doi:10.1109/iccv.2019.00356
2. V. Tejaswini Priyanka, Y. Reshma Reddy, D. Vajja, G. Ramesh and S. Gomathy. A Novel Emotion based Music Recommendation System using CNN. 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 592-596, (2023). doi: 10.1109/ICICCS56967.2023.10142330.
3. Rautaray, S. S, International Journal of UbiComp, 3(1), 21–31, (2012). doi:10.5121/iju.2012.3103.
4. Huang, C.-L., & Jeng, S, Machine Vision and Applications, 12(5), 243–258, (2001). doi:10.1007/s001380050144
5. Zhao, H., Jia, J., & Koltun, V, Exploring Self-Attention for Image Recognition. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (2020). doi:10.1109/cvpr42600.2020.01009.
6. Ramesh, G., Anugu, A., Madhavi, K., Surekha, P, Automated Identification and Classification of Blur Images, Duplicate Images Using Open CV. In: Luhach, A.K., Jat, D.S., Bin Ghazali, K.H., Gao, XZ., Lingras, P. (eds) Advanced Informatics for Computing Research. ICAICR 2020. Communications in Computer and Information Science, 1393. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-16-3660-8_52\(2020\)](https://doi.org/10.1007/978-981-16-3660-8_52(2020)).
7. Somasekar, J Ramesh, G, IJEMS, 29(6) [December 2022], NIScPR-CSIR, India, (2022).
8. Kumar, S.K., Reddy, P.D.K., Ramesh, G., Maddumala, V.R, Traitement du Signal, 36(3), 233-237, (2019). <https://doi.org/10.18280/ts.360305>.
9. Oudah, M., Al-Naji, A., & Chahl, J, Journal of Imaging, 6(8), 73, (2020). doi:10.3390/jimaging6080073.