Sign language comprehension and vocalization: A Game-Changer for Auditory-challenged and Mute Communication

[1]Anandhakumar Dharmalingam,[2]Dr.K.V.Krishna Kishore,[3]Yakkala Hemanth
[2] Department of Computer Science and Engineering,VFSTR(Deemed to be University), Vadlamudi,Guntur
[1,3]Department of Information Technology,VFSTR(Deemed to be University), Vadlamudi,Guntur
[1]anandhakumardharmalingam@gmail.com
[2]kishorekvk_1@yahoo.com
[3]yakkalahemanth1412@gmail.com

ABSTRACT

Sign language plays a crucial role in facilitating communication between individuals who are unable to speak and those who can. Mute individuals face significant challenges in conveying their message to the hearing population, particularly in times of emergency when swift communication is essential. Since most individuals are not trained in sign language, it becomes particularly difficult to comprehend their messages. To overcome this problem, sign language can be converted into spoken language to enable effective communication. There are two primary methods for detecting hand gestures, namely vision-based and non-vision-based techniques, which are used to convert the detected information into voice. While the former employs a camera to detect gestures, the latter is used in this project. It is worth noting that many mute individuals are also deaf, making it necessary to convert spoken language into sign language for effective communication.

Keywords:

Hand gesture recognition, Voice conversion, Deaf and dumb, Sign language, Vision-based technique, on-vision-based technique Communication, Emergency communication.

I. Introduction

Nowadays, there is constant discussion about new technologies that enhance our daily lives and make things simpler. Technology has completely transformed how we live, and it continues to evolve at a rapid pace. Various fields of technology, such as artificial intelligence and smartphones, are heavily researched, leading to new discoveries that simplify life for everyone. However, there has been limited research on technologies specifically designed for individuals who are deaf and dumb, compared to other industries.

Communication gap is one of the major challenges faced by deaf and dumb individuals when interacting with regular people. This communication barrier often makes them feel uncomfortable and marginalized in society, as the struggle to convey their emotions effectively. To bridge this gap, the Hand Gesture Recognition and Voice Conversion (HGRVC) system was developed to detect and track hand gestures of deaf and dumb individuals using a web camera. The system employs pre-processing techniques to resize the captured images, and aims to convert hand gestures into text for communication purposes. The goal of the project is to create a database of hand gesture images, which can be matched in real-time to convert the gestures into text.

The paper is divided into five sections. The first section provides an overview of the existing research on the topic. The second section discusses the hand gesture recognition approach employed in the system. The third section focuses on the operation of the system, detailing the image detection process. The fourth section covers the outcomes of the system,

including the generated text-based output that helps bridge the communication gap between deaf-mute individuals and others. Finally, the last section concludes the paper by discussing the system's accuracy and significance in addressing the communication challenges faced by deaf and dumb individuals.

II. LITERATURE SURVEY

In the literature review, we have examined various projects related to the topic and studied their system behavior. Shweta S. Shinde, Rajesh M. Autee, and Vitthal K. Bhosale [1] proposed a method that uses angle and peak calculation techniques in MATLAB to extract hand gesture features, which are then converted into speech using MATLAB's built-in commands. R.K., Valliammai.V., and Padmavathi.S. [2] proposed a system for Indian hand sign language that does not require any additional external hardware and is implemented using MATLAB. They capture runtime live images, apply image processing using the HIS model, and extract features using the distance transform method, achieving satisfactory results for most hand gestures.

Anchal Sood and Anju Mishra [3] proposed a sign recognition system based on the Harris method for feature extraction, where the features are extracted and stored in an Nx2 matrix after image pre-processing. The system has some limitations, such as errors caused by the range value for skin segmentation when the backdrop color varies from very light brown to relatively dark brown, but the results are efficient.

Prashant G. Ahire, Kshitija B. Tilekar, Tejaswini A. Jawake, and Pramod B. Warale [4] designed a hand gesture recognition system using MATLAB, where real-time video input is processed through image processing stages and a correlation-based approach is used for mapping. The audio output is generated using the Google TTS API, and the proposed system yields efficient outcomes.

Mrs. Neela Harish and Dr. S. Poonguzhali [5] proposed a hardware-based approach using a data glove with sensory components like flex sensors, an accelerometer, and a PIC microcontroller for input and output handling, and achieved pleasant and efficient outcomes.

Sonal Kumari and Suman K. Mitra [6] designed a system based on hand action recognition using the background subtraction approach for image processing and employed the Direct Fourier transform (DTE) algorithm for feature extraction in MATLAB.

III. RELATE WORK

Voice conversion techniques are used to convert sign language gestures into spoken language to enable communication between deaf and dumb individuals and the hearing population. There are two main approaches to voice conversion: rule-based and data-driven.
Rule-based techniques involve defining a set of rules that map sign language gestures to spoken language. These rules are typically defined by experts in sign language and linguistics. The advantage of rule-based techniques is that they are transparent and easy to understand. However, they are often limited by the complexity of the rules, which can make it difficult to handle the wide variety of sign language gestures.

Data-driven techniques, on the other hand, use machine learning algorithms to learn the mapping between sign language gestures and spoken language from a large dataset of

sign language videos and corresponding spoken language transcripts. The advantage of data-driven techniques is that they can handle the complexity and variability of sign language gestures more effectively than rule-based techniques. However, they are often less transparent and more difficult to understand than rule-based techniques.

Deep learning techniques, such as neural networks, are commonly used in data-driven voice conversion systems. The most common approach is to use a sequence-to-sequence model, which maps a sequence of sign language gestures to a sequence of spoken language words. Using a supervised learning method, the model is trained on a sizable dataset of sign language movies and related spoken language transcripts.
In a study by Lee et al. (2018), a deep learning-based voice conversion system was developed for Korean Sign Language (KSL) using a sequence-to-sequence model. The system achieved a word recognition rate of 85.6%, demonstrating the effectiveness of deep learning techniques in voice conversion.

Challenges and Future Research Directions Despite the recent advancements in hand gesture recognition and voice conversion technologies, there are still several challenges that need to be addressed. One of the main challenges is the development of robust and accurate hand gesture recognition systems that can handle the variability and complexity of sign language gestures. This requires the use of advanced machine learning techniques, such as deep learning, and the development of large annotated datasets of sign language gestures.
Another challenge is the development of natural and accurate voice conversion systems that can handle the wide variety of sign language gestures and produce natural-sounding speech. This calls for the creation of substantial datasets of sign language films and related spoken language transcripts, as well as the utilization of cutting-edge machine learning techniques like neural networks.

In addition, there is a need for the development of user-friendly and accessible hand gesture recognition and voice conversion systems that can be used by deaf and dumb individuals in various settings, including emergency situations. This requires the development of low-cost and portable hardware devices, as well as the integration of these devices with existing communication technologies, such as smartphones and tablets.

Conclusion Voice conversion and hand gesture recognition technologies have the potential to improve communication between deaf and dumb individuals and the hearing population. These technologies have the potential to facilitate effective communication in various settings, including emergency situations. The development of robust and accurate hand gesture recognition systems and natural and accurate voice conversion systems is critical for the success of these technologies. The use of advanced machine learning techniques, such as deep learning, and the development of large datasets of sign language videos and corresponding spoken language transcripts are essential for the development of these systems. The integration of these technologies with existing communication technologies and the development of user-friendly and accessible hardware devices are also important for their widespread adoption.

IV. Proposal work

In the proposed system, hand motions or gestures of individuals who are deaf and dumb will be detected and converted into human hearing voice signals. Deep learning techniques, specifically convolutional neural networks (CNNs) and recurrent neural networks

(RNNs), will be used to train the system on hand gesture images, which can then be used to predict gestures from a webcam.

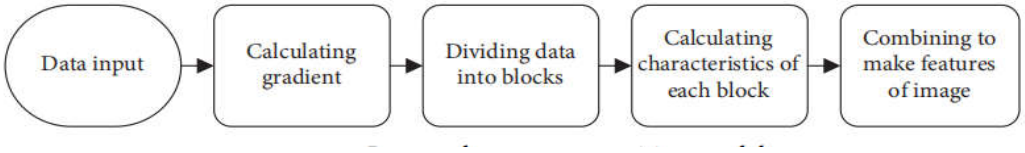
The system will consist of two main components: a hand gesture recognition system and a voice conversion system. The hand gesture recognition system will use CNNs and RNNs to detect and recognize hand gestures used in sign language, and convert them into text. The voice conversion system will then convert the text into spoken language, allowing the hearing population to understand the message.

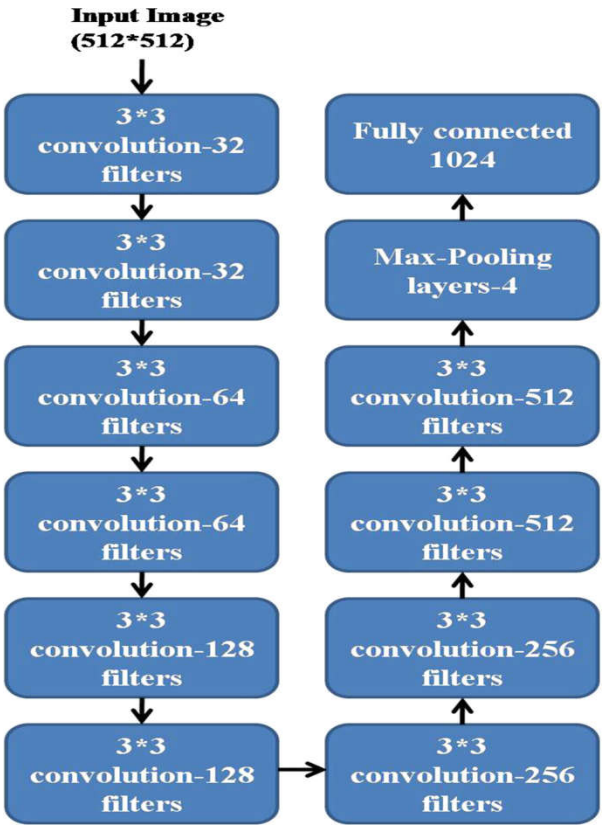
The hand gesture recognition system will be trained on a large dataset of sign language videos with corresponding text annotations. It will be designed to detect and recognize a wide range of sign language gestures, including those used in emergency situations.

The voice conversion system will be developed using a sequence-to-sequence model, which maps text to spoken language. It will be trained on a large dataset of sign language videos with corresponding spoken language transcripts, in order to produce natural-sounding speech for effective communication.

The proposed system will be tested on a dataset of sign language videos with corresponding spoken language transcripts, and will be evaluated based on its accuracy in detecting and recognizing hand gestures, as well as its ability to generate natural-sounding speech.

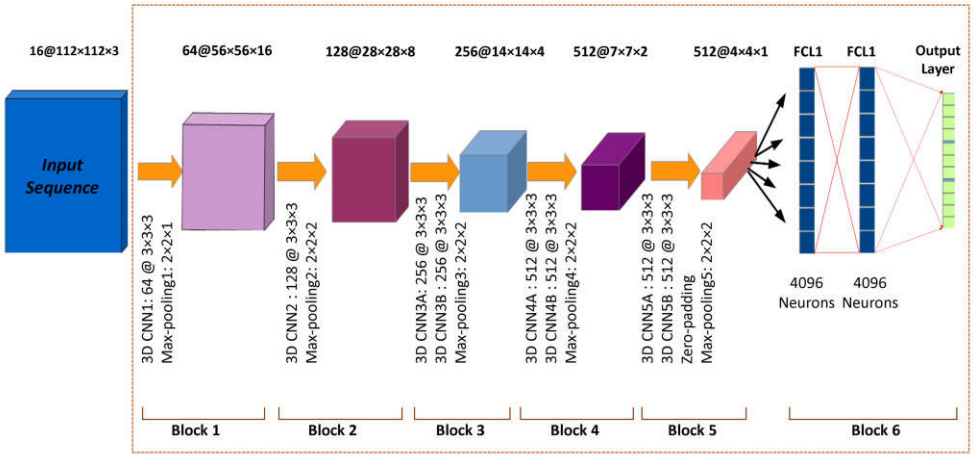
Architecture of Proposal System:



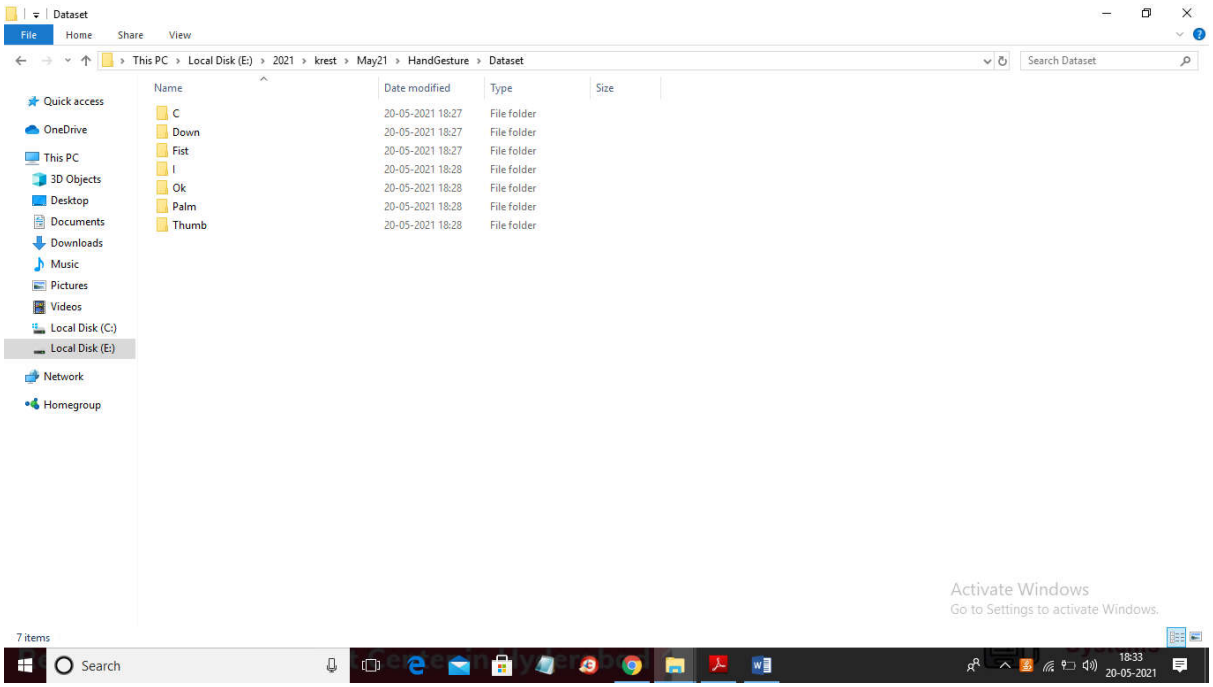


V. RESULT ANALYSIS

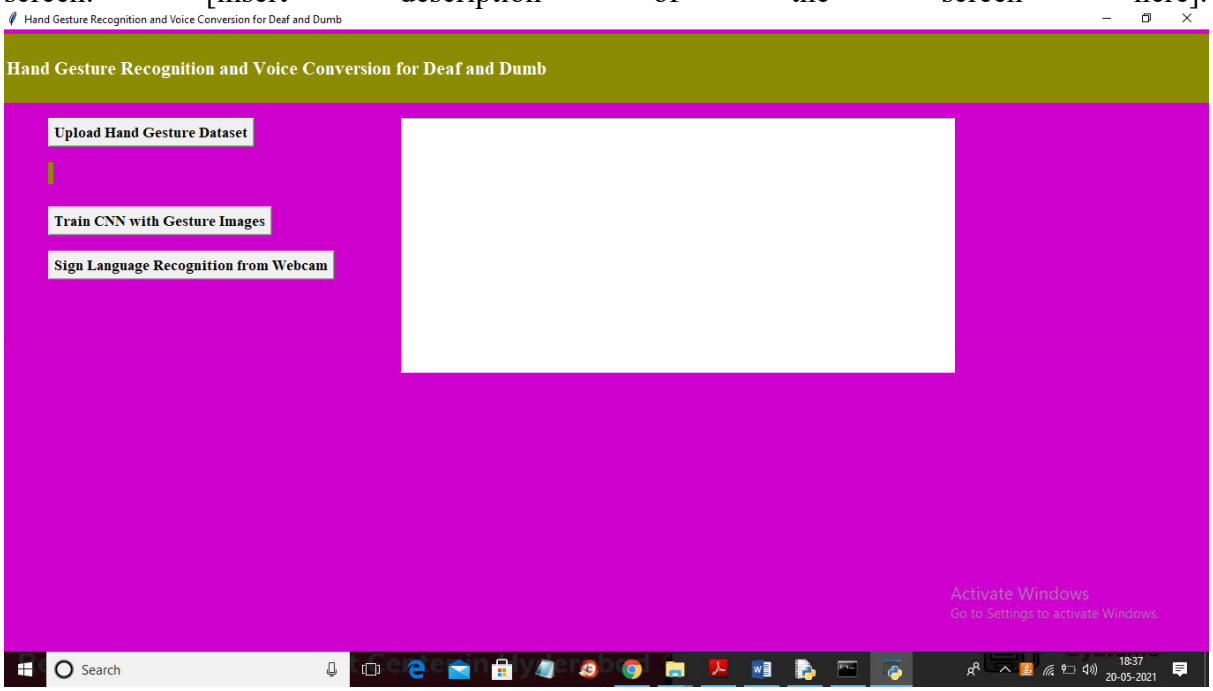
In this paper, the author proposes building a machine learning model that can predict hand gestures from a webcam and convert them into voice signals to facilitate understanding by the general population of what deaf and dumb individuals are expressing. The initial attempt to use SVM algorithm in Python for hand gesture recognition was found to be inaccurate. Therefore, the author opted to use deep learning techniques, specifically Convolutional Neural Networks (CNNs), to train hand gesture images and predict gestures from a webcam.



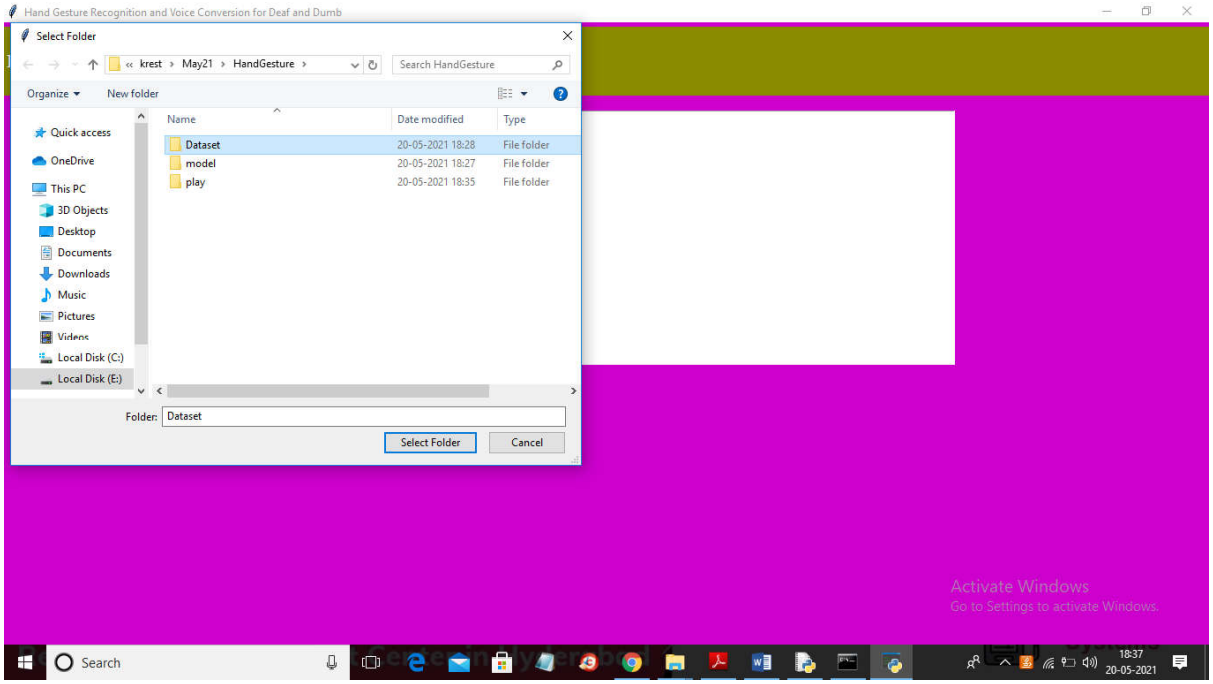
The dataset used for training the CNN model includes: [insert dataset details here].



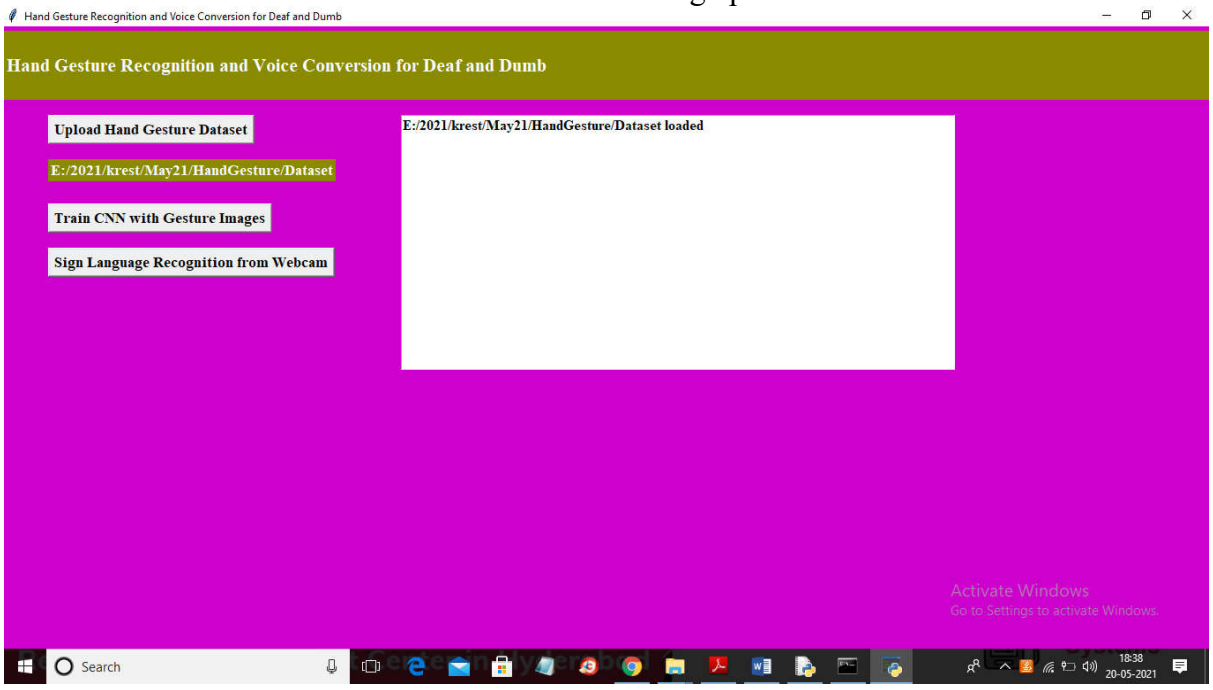
In the dataset used for training the CNN model, there are seven different types of hand gesture images, including thumbs up, thumbs down, C, I, OK, etc. The accuracy of the computer program in predicting these gestures is estimated to be around 80% (8 out of 10) when the gestures are properly shown in front of the webcam. To run the project, simply double-click on the 'run.bat' file, which will launch the program and display the following screen: [insert description of the screen here].



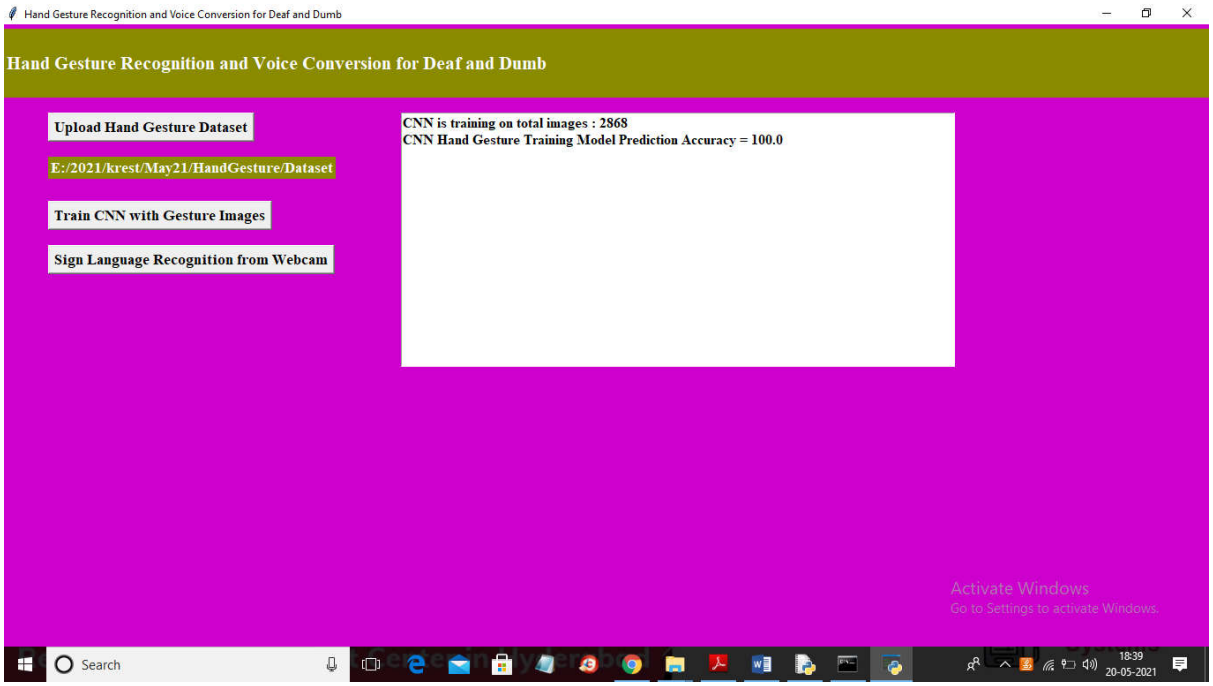
In above screen click on ‘Upload Hand Gesture Dataset’ button to upload dataset



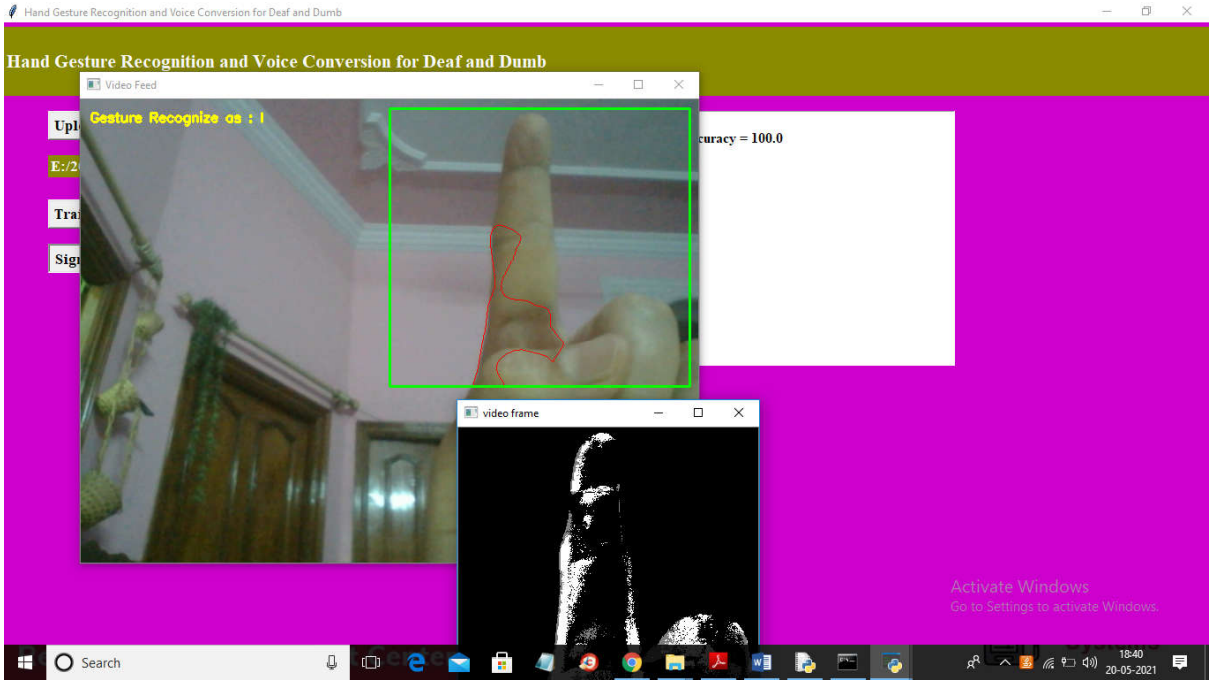
Selecting the "Dataset" folder in the upper screen, uploading it, and then clicking "Select Folder" to load the dataset will bring up the screen below.



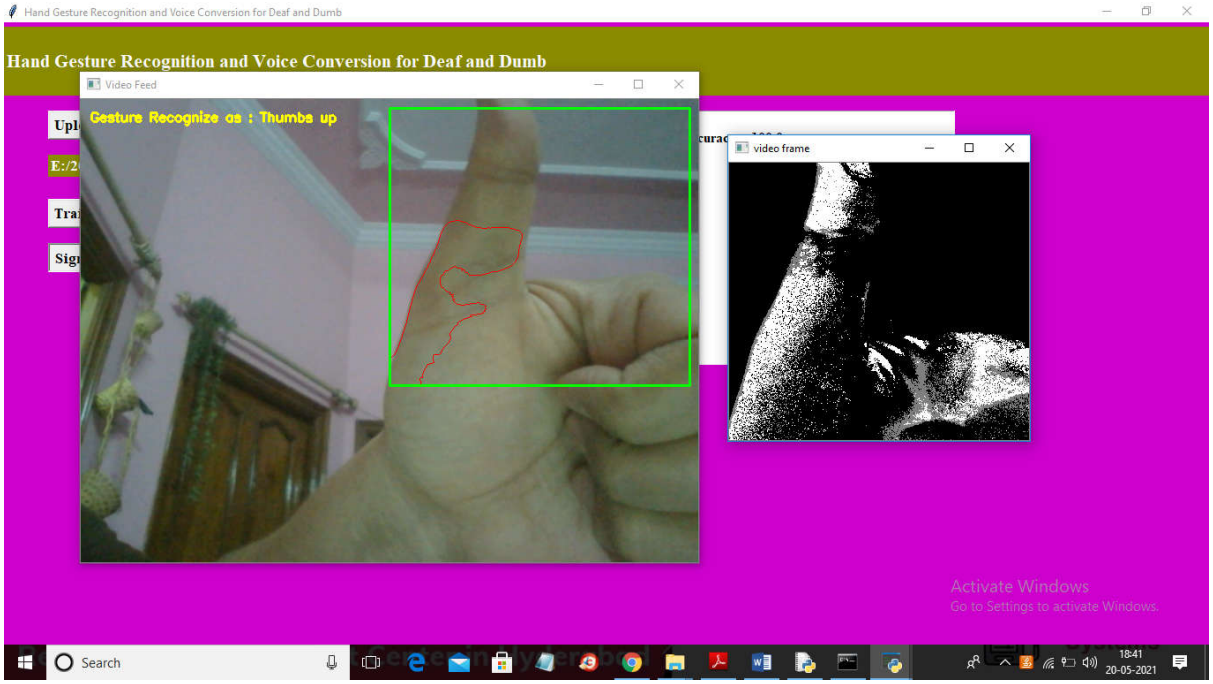
In above screen dataset loaded and now click on ‘Train CNN Gesture Images’ button to train Model

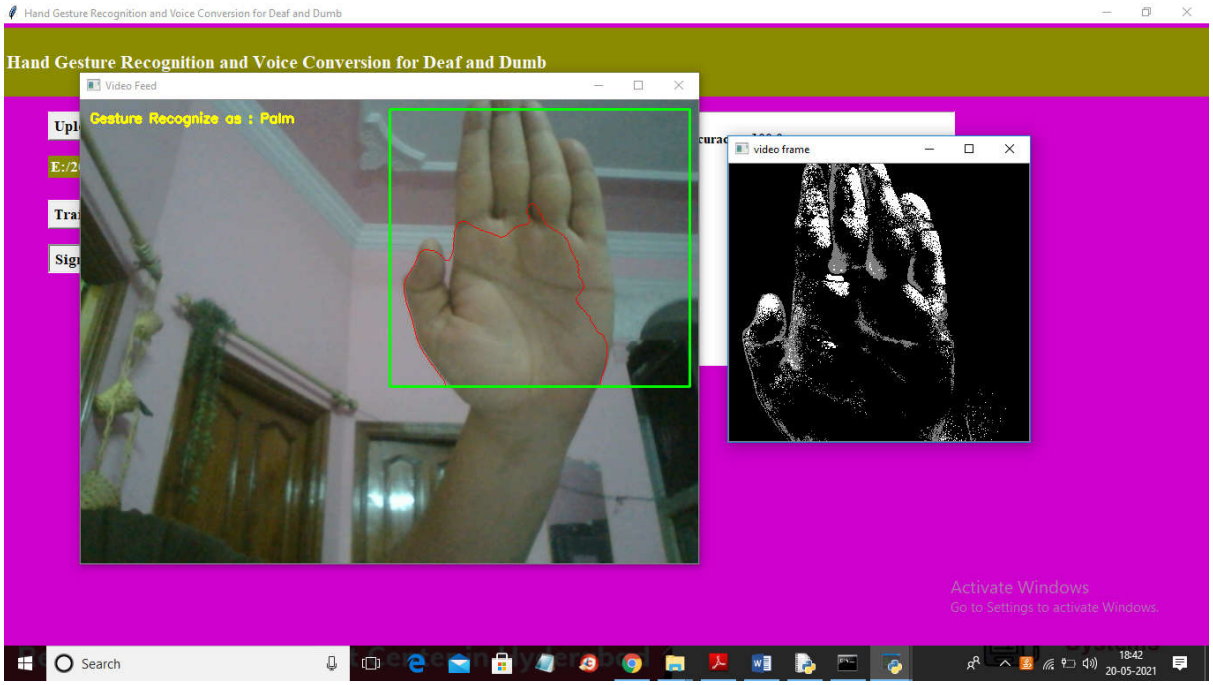


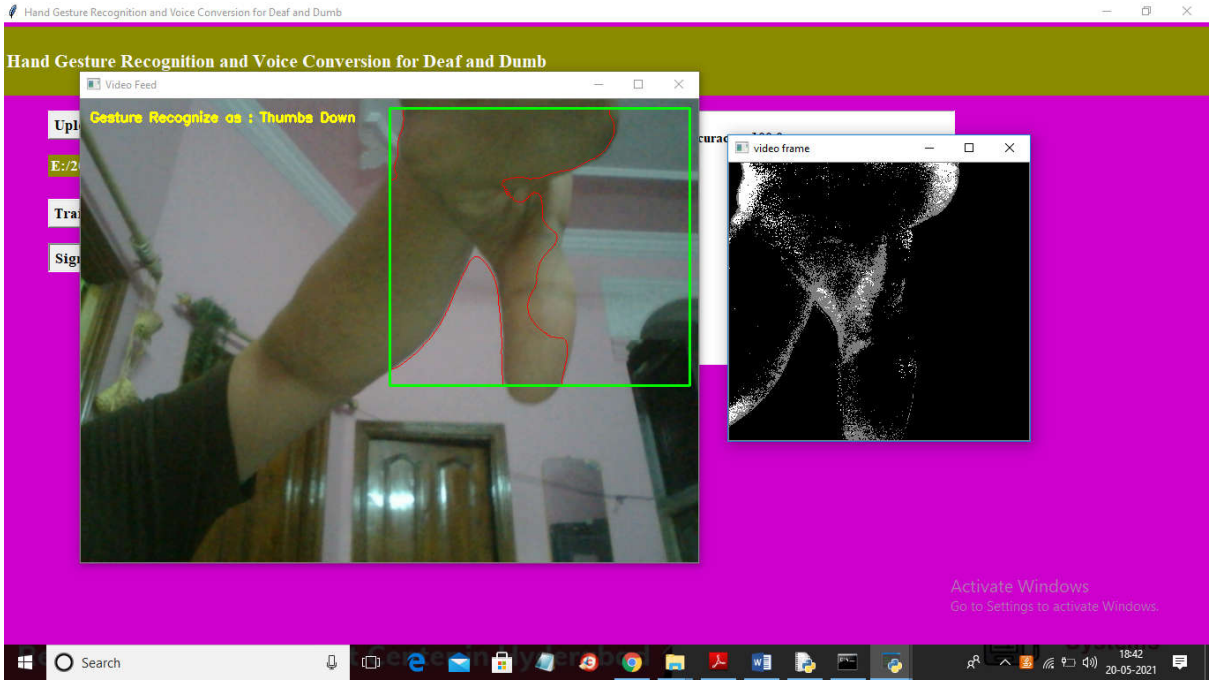
Using the 'Sign Language Recognition from Webcam' button, go to the below-screen after the CNN model has been built in the screen above.

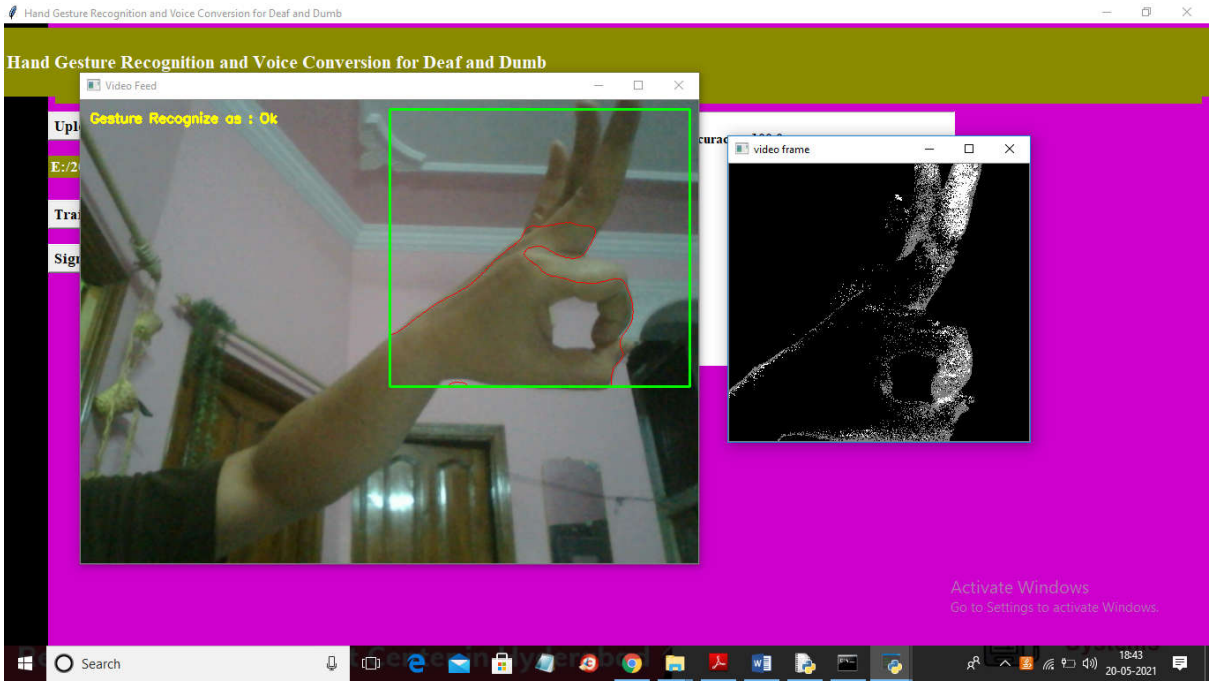


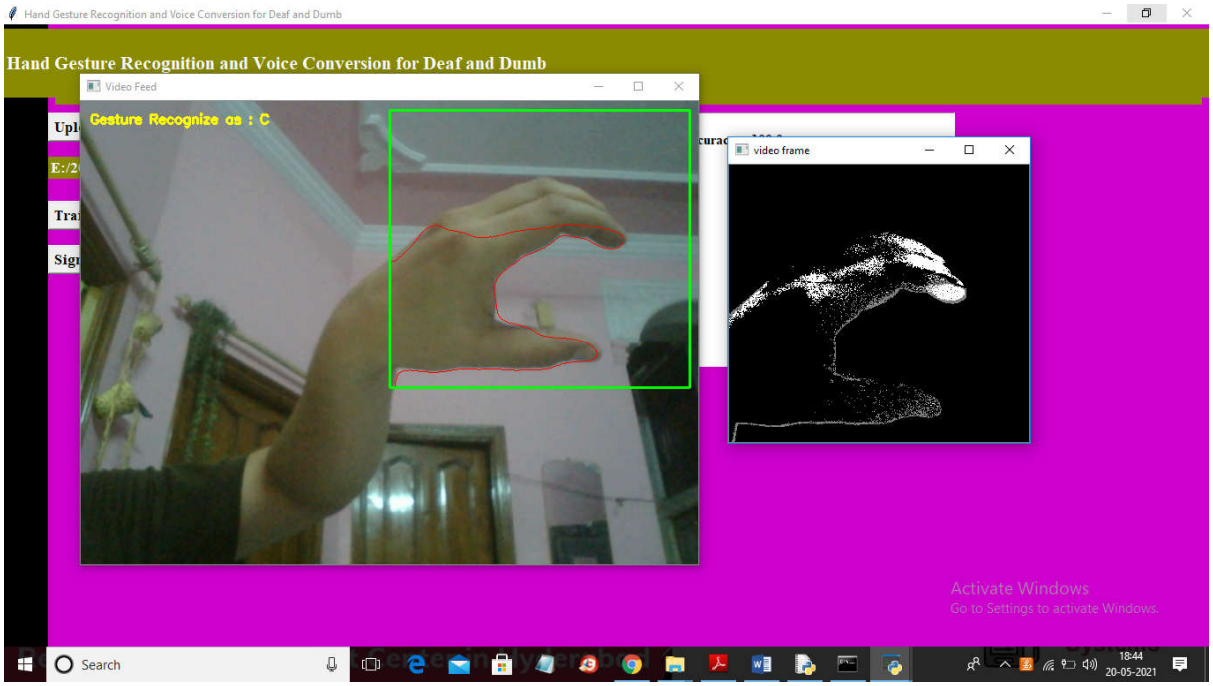
In above screen just show ur hand gesture in green colour rectangle box then application will recognize and then play it as voice


Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Video Feed
Gesture Recognize as : Thumbs up
video frame
Activate Windows
Go to Settings to activate Windows.
Search
18:41
20-05-2021


Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Video Feed
Gesture Recognize as : Palm
video frame
Activate Windows
Go to Settings to activate Windows.
Search
18:42
20-05-2021


Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Video Feed
Gesture Recognize as : Thumbs Down
video frame
Activate Windows
Go to Settings to activate Windows.
Search
18:42
20-05-2021


Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Hand Gesture Recognition and Voice Conversion for Deaf and Dumb
Video Feed
Gesture Recognize as : Ok
video frame
Activate Windows
Go to Settings to activate Windows.
Search
18:43
20-05-2021



In this work, it is important to accurately perform hand gestures as shown on the screen. Initially, there may be some incorrect predictions due to hand positioning or other factors. However, as the user adjusts their hand gesture to align with the expected gesture pattern, the accuracy of the predictions is likely to improve.

The project follows a series of modules for each prediction, including,

1)Image extraction from webcam: The system captures an image from the webcam in real-time, which serves as the input for hand gesture recognition.

2)Image pre-processing: The captured image is converted into binary or grey format, and background removal techniques are applied to isolate the hand region from the background noise or clutter.

3)Feature extraction: Relevant features, such as angles, peaks, or other characteristic parameters, are extracted from the pre-processed image to represent the hand gesture.

4)Recognition and audio output: The system uses a recognition algorithm or a trained machine learning model to match the extracted features with predefined gesture patterns. Once the gesture is recognized, the system generates an audio output using text-to-speech (TTS) technology or other audio playback methods to provide feedback or instructions to the user.

It is important to note that the accuracy of the predictions may improve as the user corrects their hand gesture to align with the expected pattern. Once the correct gesture is recognized, the system will play the corresponding audio output, providing feedback or instructions to the user based on the recognized gesture.

CONCLUSION

A system has been developed for recognizing hand gestures and converting voice for individuals who are deaf and mute. The process utilizes image processing, accepting an

image as input and producing text and speech as output. The system has demonstrated up to 89% accuracy and has proven to be effective in the majority of test cases.

REFERENCES

[1] Shinde, Shweta S., Rajesh M. Autee, and Vitthal K. Bhosale. "Real time two way communication approach for hearing impaired and dumb person based on image processing." Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on. IEEE, 2016.
[2] Shangeetha, R. K., V. Valliammai, and S. Padmavathi. "Computer vision based approach for Indian Sign Language character recognition." Machine Vision and Image Processing (MVIP), 2012 International Conference on. IEEE, 2012.
[3] Sood, Anchal, and Anju Mishra. "AAWAAZ: A communication system for deaf and dumb." Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), 2016 5th International Conference on. IEEE, 2016.
[4] Ahire, Prashant G., et al. "Two Way Communicator between Deaf and Dumb People and Normal People." Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on. IEEE, 2015.
[5] Ms R. Vinitha and Ms A. Theerthana. "Design And Development Of Hand GestureRecognition System For Speech Impaired People."
[6] Kumari, Sonal, and Suman K. Mitra. "Human action recognition using DFT." Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on. IEEE, 2011.

[7] A. Argyros and M. Lourakis, “Vision-based interpretation of hand gestures for remote control of a computer mouse,” in *Proc. Workshop Com*

put. Human Interact., 2006, pp. 40–51.

[8] C. Wang and K. Wang, *Hand Gesture Recognition Using Adaboost With SIFT for Human Robot Interaction*, vol. 370. Berlin, Germany: Springer

Verlag, 2008.

[9] A. Barczak and F. Dadgostar, “Real-time hand tracking using a set of co-operative classifiifiers based on Haar-like features,” *Res. Lett. Inf. Math. Sci.*, vol. 7, pp. 29–42, 2005.

[10] Q. Chen, N. Georganas, and E. Petriu, “Real-time vision-based hand gesture recognition using Haar-like features,” in *Proc. IEEE IMTC*, 2007, pp. 1–6.

[11] P. Viola and M. Jones, “Robust real-time object detection,” *Int. J. Comput.*

Vis., vol. 2, no. 57, pp. 137–154, 2004.

[12] S. Wagner, B. Alefs, and C. Picus, “Framework for a portable gesture interface,” in *Proc. Int. Conf. Autom. Face Gesture Recog.*, 2006, pp. 275–280.

[13] M. Kolsch and M. Turk, “Analysis of rotational robustness of hand detection with a Viola-Jones detector,” in *Proc. 17th ICPR*, 2004, pp. 107–110.

[14] Y. Ke and R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2004, pp. II-506–II-513.

[15] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, “SURF: Speeded up robust

features,” *Comput. Vis. Image Understand. (CVIU)*, vol. 110, no. 3, pp. 346–359, 2008.

[16] W. Zhao, Y. Jiang, and C. Ngo, “Keyframe retrieval by keypoints: Can

point-to-point matching help?” in *Proc. 5th Int. Conf. Image Video Retrieval*, 2006, pp. 72–81.

[17] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2006, pp. 2169–2178.