iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: II Month of publication: February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58103

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Keyboard and Mouse Free Music Controller

Anmol Sinha[1], Rahul Kumar Gupta[2], Varun Saxena[3]
[1,3]School of Engineering, Jawaharlal Nehru University, Delhi, India
[2]Electronics and Communication Engineering Department, JSS Academy of Technical Education, Noida, India

Abstract: Human-machine interface (HMI) is a crucial area of research as gestures have the potential to efficiently control and interact with computers. Many applications for hand detection have been created as a result of the pervasive use of built-in cameras in computers, smartphones, and tablets. For the majority of users, however, many of these are not useful. A straightforward concept for a keyboard- and mouse-free music controller is presented in this research. Using MATLAB code that integrates skin detection, area labelling, erosion, dilation, and motion differentiation, a music player controller is developed using the real-time frame tracking feature of a camera. Three hand detection algorithms are created and assessed for maximum performance and accuracy. Real-time hand detection for operating the music player is provided by the algorithm, which is created with efficiency and speed in mind.
Keywords: HMI, segmentation techniques, keyboard and mouse-free music controller,

I. INTRODUCTION

Gesture recognition has become a crucial component in various computer vision applications. Researchers have focused on developing specific applications, such as tracking humans, virtual mouse, controlling electronic devices, TV remote control, sign language recognition, and music control. Some notable works related to these tasks have been discussed in the following text. A vision-based remote control system for electronic appliances is suggested in one study. In order to pinpoint movement regions connected to waving hands, it uses skin colour detection. After the request is granted, the camera enlarges the area of the local hand, and finger tips are identified using a corner identification method based on the k-cosine function. The gesture state is determined by the number of open fingers, and the associated control command is activated when a series of gesture state transitions in the state transition queue occur. The robust and accurate performance of this method is independent of the operating environment because it obtains stable information over multiple frames. The suggested approach can make it easier to create low-cost, high-performance universal remote control systems [1]. In a different study, color-based pixel categorization is used to split skin in pixels at a time. The multilayer perceptron (MLP) classifier and the Bayesian classifier with the histogram approach are found to have classification rates that are greater than other examined classifiers. Because of the low dimension of the feature vector and the possibility of gathering a sizable training set, the Bayesian classifier with the histogram technique is appropriate for the problem of classifying pixels based on skin tone. Comparatively speaking to the MLP and other classifiers, the Bayesian classifier needs more memory. The study demonstrates that the selection of colour space has little bearing on the colour representation of pixel-wise skin segmentation. If only chrominance channels are employed, segmentation performance suffers, and there are large performance differences across the various chrominance options. Better segmentation outcomes are obtained with fine colour quantization (bigger histogram size). However, if a sizable and meaningful training dataset is employed, colour pdf estimation can be performed with histogram widths as little as 64 bits per channel [2-4]. Real-time hand gesture detection for music player control is the subject of another investigation. The motion and contour feature detection algorithm outperforms the other two hand detection methods when they are compared. To discriminate between these three hand orientations, three criteria for determination are used. In order to prevent lengthy lagging when the user moves their hand in front of the webcam, a quick and responsive "state machine" is employed. The average delay is 2.4 seconds. Lagging time and system robustness are trade-offs because the robust feature necessitates a thorough, occasionally redundant inspection method that invariably causes lagging time. The authors offer a MATLAB-programmed music player that is completely functional and has three method implementations that may be tested [5-7]. One concept under consideration uses fixed, user-defined movements to operate electronic devices like laptops and TVs. LabVIEW is the programme used to carry out the control. The suggested model is sensitive to light intensity and uneven backdrops, and it takes LabVIEW longer to detect the gesture. Future implementations of the same could also make use of dynamic gestures [8].

A different approach is created to reject accidental and irregular hand gestures (like children's illogical movements) and to provide visual feedback on the gestures captured. The writers created a collection of motions that are unique from one another but simple for the system to recognize.

This set comprises four distinct invariant moments that enable extremely precise and immediate categorization. Because there were so few hand motions, the control mechanism was 100% accurate. In TV mode, the "Volume" gesture can be mapped to the "Speed" function of a ceiling fan thanks to a special key mapping feature produced by the software interface. The authors want to use an Infrared camera in the future to deal with dim lighting. This system is currently ready to be implemented on dedicated hardware such as a digital TV set-top box [9]. The Covariance Method is suggested in this research study as a way to recognise hand motions. To conduct the study, a library of images of diverse static hand motions that are a subset of American Sign Language will be created (ASL). The Eigen values of the Eigen vectors are computed after the images have undergone preprocessing to remove noise. An image is converted into a feature vector (Eigen image) using a pattern recognition algorithm, and this feature vector is then compared to a trained collection of eight gestures. The suggested approach was successful in locating the right match [10-15]. The goal of this manuscript is to make hand gesture recognition possible for media player control. Setting up the media player and webcam starts the application. The media player starts playing music when the user holds their hand in front of the webcam for two seconds and then removes it. The music will stop if the user lifts their hand for longer than two seconds or if it is not recognized. To achieve this, the images are captured using a webcam and processed using image subtraction technique to extract the main object. The image is then converted into binary using threshold segmentation. Various region properties such as Eccentricity, Area, Orientation, and Centroid are used to label the detected hand as a separate entity in case of multiple object detection. One of the main challenges in hand gesture recognition is detecting the hand itself, which is done by identifying the skin color using the YCbCr color space.

II. PROPOSED METHODOLOGY

One of the main challenges in hand gesture recognition is detecting the hand itself, which is done by identifying the skin color using the YCbCr color space. The detected hand is then represented using the aforementioned parameters.

Overall, this manuscript provides a solution for hand gesture recognition that can be used to control a media player, enabling easy and efficient operation of the device.

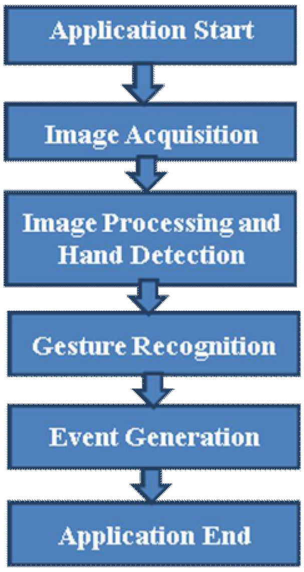


Figure 1: Flow chart

The system starts by switching on and acquiring images using a fixed webcam. The video is recorded continuously, and the frame rate is controlled by MATLAB's Image Acquisition Toolbox. The video is broken down into frames, which are further processed to extract the desired data. Hand detection is achieved using background subtraction, where the present image is subtracted from the previous image to isolate the moving object. Skin color detection is done using the YCbCr color space, and the detection is based on the filled area of the image once it is converted into binary. Hand gesture recognition is used to track the desired hand on the entire image, and the process continues for a specified number of frames. Once the gesture recognition is completed, an event is generated using the Windows Media Player, where specific gesture motions are assigned to specific activities.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

III. HAND GESTURE RECOGNITION

A. Skin and Shape Detection

The image of hand was separated from the background in the webcam frame by the authors using a skin pixel detector. They employed erosion and dilation to reduce noise and area labelling to identify the largest detected area, thinking that the hand would be the object with the most skin colour in the webcam frame. To determine the shape of the hand, contour extraction is computed on the obtained image [16]. It should be noted that skin tone and lighting conditions have an impact on how accurate this algorithm is.

B. Skin and Feature Detection

The second technique filters the collected image using a skin pixel detector before applying erosion and dilation to reduce noise. The fingertips for the thumb are then found using a feature extraction filter; these fingertips are represented by peaks along the contour. To recognise a hand, our system finds 30 fingertip points, five for each finger [7]. However, the stretching of the hand and avoiding noise brought on by face detection have an impact on the accuracy of this method.

C. Motion and Contour Feature Detection

As an enhancement over the two prior algorithms, which have some conditional restrictions, this approach has been devised. The image that was captured is subjected to the skin detector, erosion, and dilation noise reduction. The contour is then calculated by computing differentiation between two successively processed pictures. Based on the 3D depth of field, it is assumed that because the hand is closest to the webcam, its movement can be more easily recognised than that of other objects. The contour of the object with the greatest range of motion, in this case the hand, is produced by differentiating the two following photos. The second algorithm's technique is used to identify the hand features. This method is more reliable and accurate when compared to the first two algorithms.

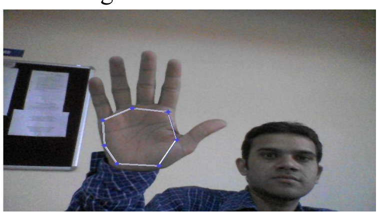
D. Detection Algorithm

The implementation specifics of the three hand gesture recognition algorithms are covered in this section. We will first present these stages since skin detection and erosion and dilation processing are prerequisites for all three algorithms and the result has been explained further. The distinct processing steps for each method are then independently described [17].

IV. RESULTS AND DISCUSSION

A. Skin Detection

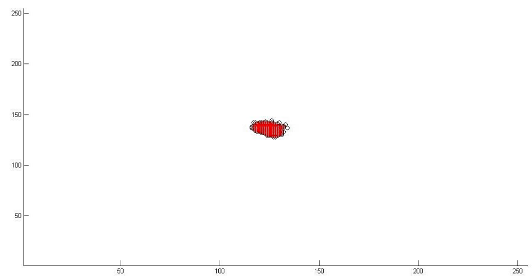
The In order to identify potential hand pixels, it is necessary to separate them from non-skin pixels. This can be achieved by analyzing the color space of the image. Human skin color typically falls within a certain range in a designated color space, and this can be used to detect skin pixels. Based on the representations of the YCbCr and HSV colour spaces, two distinct detectors have been created. The luminance-dependent Y value is first subtracted from the YCbCr representation. Following that, the Cb and Cr components are mapped into a mask using training image analysis. Based on whether or not they fall within the region designated as skin, skin pixels can then be divided which is shown in Figure 2.



(a)

(b)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com



(c)

Figure 2 a) Label Skin Region in CbCr color space b) CbCr color Space Projection c) Convex Mask

The V component is eliminated in the HSV representation, and the H and S components are projected onto a two-axes plain. A set of static thresholds is established after training with 10 to 20 photos. White pixels denote skin areas, whereas black pixels denote non-skin areas, in a binary mask produced by any skin segmentation technique. It should be noted that some areas of the face and light-colored background items may be detected as hand pixels, and additional processing may be required to filter them out. After analyzing the distribution of colors projected in the YCbCr and HSV color spaces, it has been determined that the skin color cluster is better defined in the Cb-Cr axis. Therefore, the Cb-Cr color projection has been chosen for subsequent algorithms to obtain the skin area. By using the Cb-Cr projection, the skin pixels can be more accurately separated from non-skin pixels. This is because the skin color cluster is more tightly grouped together in this color space, which makes it easier to identify and isolate skin pixels.

It is important to choose the appropriate color space representation for skin segmentation based on the specific requirements of the application. In this case, the Cb-Cr projection has been found to be the most effective for accurately identifying and segmenting skin pixels.

B. Noise Reduction by Erosion and Dilation

Several minor noisy regions may also appear after identifying probable hand skin patches as white pixels. We can utilise "disc" erosion to reduce the noise, though, if the area corresponding to the hand is the greatest. The little noisy patches, which under certain lighting conditions can resemble skin colour, are successfully removed by this method. Although the erosion process has the potential to reduce the size of the genuine hand region, this effect can be mitigated by utilising dilation to amplify the largest detected region. This improves the intended hand detection. After finishing this technique, we can use white pixels to outline the hand shape contour on the image.

The next step is to determine whether what we have obtained so far is a "hand." This requires a feature recognition model.

C. Convex Contour and area Property Recognize

We employ an approach that involves locating a convex contour that encloses the hand-shaped region in order to recognise a hand shape in an image. The detected hand-shaped region is then subtracted from this convex contour to produce roughly six distinct sections. The spaces between adjacent fingers are represented by four of these regions, while the remaining two are represented by the spaces between the convex contour and the hand's shape.

We use a centrifugal calculation to remove these two eccentric inter-spaces, leaving only the inter-spaces between the four fingers, in order to differentiate a hand shape from other shapes in the image. In order to assess whether the observed form is a hand, we set a threshold count of three or four. Specifically, if there are three or four separate regions left after applying the centrifugal calculation, we assume that the detected shape is a hand which is shown in figure 3 and 4.

Figure 3. Convex hull Subtract the hand region to get regions between fingers

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

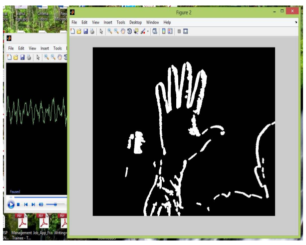
Figure 4. Hand regions between fingers

D. Differentiate Motion and Feature Recognize (Motion and Contour Feature Detection Algorithm)

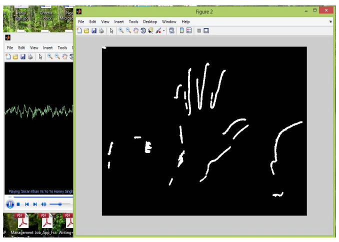
In our third algorithm, we utilize the webcam to capture successive frames and determine the difference between them. This approach is based on the assumption that the hand, being closest to the webcam, will be captured with even the slightest movement. By detecting the hand's contour based on the difference between two consecutive frames, we can filter out certain-length vertical edges using edge and vertical direction detection. For a hand to be considered "perfect," it should display 10 vertical straight edges [8]. To detect a hand, we set a threshold count of six to ten vertical straight edges. If there are between six to ten such edges detected, we assume that a hand is present shown in figure 5.

Figure 5. Hand Detected Motion Based

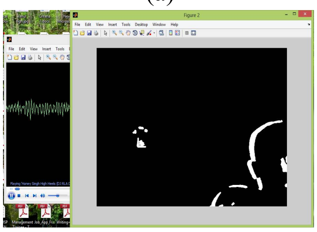
The results of gesture based music control has been shown in Figure 6. In the Figure 6 present all the controls of music player such as pause, play, next, and previous etc.



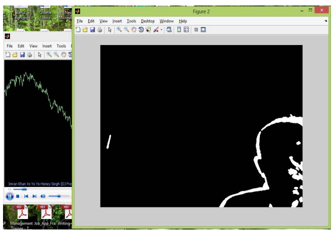
(a)



(b)

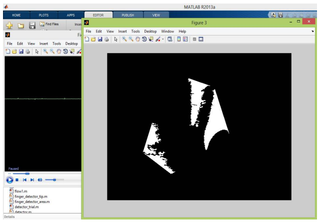


(c)

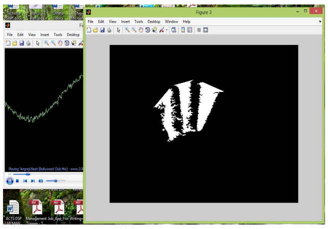


(d)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com



(e)



(f)

Figure 6 Music Control Using Hand Gesture a Pause b Play c Next d Previous e Hand detection using contour and area based: Pause f Play

V. CONCLUSION

To achieve optimal performance and precision, three hand detection algorithms were designed and evaluated individually. The primary objective of these algorithms was to enable the detection, recognition, and interpretation of hand gestures for triggering specific events, such as controlling a media player. The real-time hand detection algorithm was optimized for efficiency and rapid response to ensure seamless control of the music player using hand gestures. The hand gestures were captured in 3D using multiple cameras, enabling precise detection of gestures in real-time. However, detecting gestures represented by partially occluded hands proved to be more challenging. To control various aspects of the media player, such as music volume, playback control (previous, next, play, pause), and video zoom, a method for detecting such gestures would need to be developed. This would require careful consideration of various factors, including the degree of occlusion and the complexity of the gesture. Nonetheless, the development of such a method would enhance the overall user experience, enabling seamless control of the media player using hand gestures.

A. Conflicts of Interest

There is no conflict of interest between authors.

B. Funding Statement

NA

C. Acknowledgments

I present my gratitude to my supervisor and organisation for continuous support throughout the work

REFERENCES

[1] Daeho Lee and Youngtae Park, “Vision-Based Remote Control System by Motion Detection and Open Finger Counting”, IEEE Transactions on Consumer Electronics, Vol. 55, No. 4, NOVEMBER 2009.
[2] Son Lam Phung, Member, IEEE, Abdesselam Bouzerdoum, Sr. Member, IEEE, and Douglas Chai, Sr. Member, IEEE, “Skin Segmentation Using Color Pixel Classification: Analysis and Comparison”, IEEE transactions on Pattern Analysis and Machine Intelligence, VOL. 27, NO. 1, JANUARY 2005.
[3] Junhao Jiang, Junji Ma, Yiye Jin, “Computer Music Controller Based on Hand Gestures Recognition Through Web-cam” 2011.
[4] Gilles Bailly , Dong-Bach Vo , Eric Lecolinet ,Yves Guiard, “Gesture-Aware Remote Controls:Guidelines and Interaction Techniques” 2008.
[5] Prashan Premaratne, Q. Nguyen, “Consumer electronics control system based on hand gesture moment invariants”, University of Wollongong ,Research Online, IET Computer Vision, 1(1), 2007, 35-41.
[6] Ginu Thomas, Rahul Vivek Purohit, “Hand Gesture Recognition via Covariance Method”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-5, June 2013.
[7] Rafiqul Zaman Khan and Noor Adnan Ibraheem, “Hand Gesture Recognition: A Literature Review” International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012.
[8] Jianchao Yang, Student Member, IEEE, John Wright, Member, IEEE, Thomas S. Huang, Fellow, IEEE, and Yi Ma, Senior Member, IEEE, “Image Super-Resolution Via Sparse Representation” IEEE Transactions on Image Processing, VOL. 19, NO. 11, NOVEMBER 2010 2861.
[9] Vladimir Vezhnevets, Vassili Sazonov, Alla Andreeva, Graphics and Media Laboratory, Faculty of Computational Mathematics and Cybernetics,Moscow State University, Moscow, Russia, “A Survey on Pixel-Based Skin Color Detection Techniques”
[10] D.Devasena, P.Lakshana, A.Poovizhiarasi, D.Velvizhi, “Controlling Of Electronic Equipment Using Gesture Recognition”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-3, Issue-2, December 2013.
[11] Prof Kamal K Vyas Director (SIET) SGI Sikar, Amita Pareek Tech-Advisor, eYuG, Dr Sandhya Vyas HOD (S & H), BBVP Pilani , “Gesture Recognition and Control Part 1 - Basics, Literature Review & Different Techniques”, International Journal on Recent and Innovation Trends in Computing and Communication ISSN 2321 – 8169 Volume: 1 Issue: 7 575 – 581.

[12] Prateem Chakraborty, Prashant Sarawgi, Ankit Mehrotra, Gaurav Agarwal, Ratika Pradhan, “Hand Gesture Recognition-A Comparative Study”, Proceedings of the International Multi Conference of Engineers and Computer Scientists 2008 Vol I IMECS 2008, 19-21 March, 2008, Hong Kong.
[13] Rafael C. Gonzalez & Richard, E. Woods, “Digital Image Processing Using MATLAB”, Pearson Prentice Hall (December 26, 2003)
[14] Anil Kr. Jain, “Fundamental of Digital Image Processing”, Prentice Hall; 1 edition (October 3, 1988)
[15] Erdem Yoruk, Ender Konukoglu, Bulent Sankur, and Jerome Darbon, “Shape-Based Hand Recognition”, IEEE Transactions of Image Processing, VOL. 15, NO. 7, July 2006
[16] R. Lockton and A. Fitzgibbon, “Real-time gesture recognition using deterministic boosting”, In Proc. of British Machine Vision Conference, 2002
[17] P. Viola and M. Jones, “Robust Real-time Object Detection”, Int. Journal of Compuer Vision, 2002.
[18] M. J. Jones and J. M. Rehg, “Statistical Color Models with Application to Skin Detection”, Int. Journal of Computer Vision, 46(1):81-96, Jan 2002.
[19] D. Chai and K N. Ngan, “Face Segmentation Using Skin Color Map inVideophone Applications,” IEEE Trans. Circuits and Systems for Video Technology, vol. 9, no. 4, pp. 551-564, 1999.
[20] K. Sobottka and I. Pitas, “A Novel Method for Automatic Face Segmentation, Facial Feature Extraction and Tracking,” Signal Processing: Image Comm., vol. 12, no. 3, pp. 263-281, 1998.
[21] C. Garcia and G. Tziritas, “Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis,” IEEE Trans. Multimedia, vol. 1, no. 3, pp. 264-277, 1999.