Computer Vision Based Virtual Musical Instruments

Abhinandan Adhikari
UG Scholar
Computer Science & Engineering
Shri Shankaracharya Institute of Professional Management and Technology Raipur
Chhattisgarh, India
abhinandan.adhikari@ssipmt.com

T. Siddhartha Rao
UG Scholar
Computer Science & Engineering
Shri Shankaracharya Institute of Professional Management and Technology Raipur
Chhattisgarh, India
t.rao@ssipmt.com

Kaveri Kar
Assistant Professor
Computer Science & Engineering
Shri Shankaracharya Institute of Professional Management and Technology Raipur
Chhattisgarh, India
k.kar@ssipmt.com

Tushar Dhruw
UG Scholar
Computer Science & Engineering
Shri Shankaracharya Institute of Professional Management and Technology Raipur
Chhattisgarh, India
tushar.dhruw@ssipmt.com

Article Info
Page Number: 9600 - 9608
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received: *15 September 2022*
Revised: *25 October 2022*
Accepted: *14 November 2022*
Publication: *21 December 2022*

Abstract: **—** Most prospective musicians will ultimately need to make an investment in a musical instrument kit in order to pursue their art. This study intends to speed up musicians' exposure to the playing of musical instruments without the associated expenditures and to enable musicians to practise, at least sporadically, without a full musical instrument kit. Consequently, a larger audience may now enjoy the pleasure of playing musical instruments. The option we investigate is the creation of a virtual musical instrument set prototype that would just need users to have a laptop with a camera and readily available markers to simulate the tips of drum sticks or any blue/red color object for playing drum kit and knee motion, such as colorful papers for playing other musical instruments. This was done using OpenCV, a Python-based program that uses the concept of color-based blob recognition to find the markers.

Keywords— Virtual Musical Instrument, Image Processing, Object tracking, OpenCV, Computer Vision.

1. Introduction

In the world of today, there has been a lot of technological advancement. Technology of today is integrated with a method known as Artificial Intelligence. This project also draws on a minor amount of AI. In this project, camera-based colored object movement gesture detection is used to play the musical instruments virtually with just holding colored object in front of computer's webcam. It may be accessed more readily by using colored object detection for rapid camera access and an intuitive user interface. By eliminating the need for a real musical instrument, this method saves time and money.

The most widely used percussion instruments in the music industry nowadays are either drums or pianos. Learning to play the piano or the drums is a common first step for beginners who want to become percussion musicians. Contrary to other instruments like the guitar or keyboard, a standard drum set is typically expensive, occupies a large amount of space, and is difficult to carry. The objective of this study is to develop a system that will allow aspiring musicians to use computer vision to play and practise musical instruments like pianos and drums in the absence of a physical instrument. The goal is to convert a video of a person playing virtual drums or piano into an audio synthesis of suitable drum/piano samples in real-time, while taking into account realistic movements with an actual drum set or genuine piano/keyboard. The project aims to develop an implementation that would make it simpler for people without the resources or equipment to practise and study the real drums and real pianos. It does this by utilising the built-in web camera on a laptop.

This study doesn't need any additional tools or technology, unlike other studies that have looked into virtual musical instrument performance. Traditionally, the instrument is played by using extra devices. Use of the system merely necessitates a web-connected camera or a laptop with a built-in camera. The only need for an effective result is to employ a high resolution camera to recognise the colored object held by the person correctly playing the virtual drum. The system will still function without a high-quality camera but there will be a lot of errors. In this study, a high-resolution camera might also refer to a high definition (HD) camera. By doing away with human intervention and the reliance on physical equipment to perform the musical instruments, the suggested method will also prevent the spread of COVID-19.

We are developing such application which is combination of AI & Web. Implement such code where camera can recognize (Red/Blue) colored object movement & responds according to it. After completing this project user will be able to play the musical instruments virtually by just holding a colored object. The AI-powered Virtual Musical Instrument is a musical instrument set up that does not require a physical or real musical instruments. The application detects colored object movements with the help of computer's webcam, then tracks and depending upon the movements of colored object it detects the coordinate of musical instrument and plays the respective musical instrument's sound.

2. Literature Review

A. *Virtual Drum*

Several portable drum system implementations included sensors that could be mounted to the sticks to detect the direction and velocity of the drum sticks. The gathered information is fed into a drum system, which plays the proper sound sample for each distinct drum component. These sensors have the advantage of being more responsive to velocity changes and requiring less setup than utilizing a camera. Cost constraints, however, may prevent the general population from using these sensors. Freedrum[1], which has this as its commercial application and charges about $235 for a full kit, is one example. Both the feet and the drumsticks have sensors as part of their implementation.

There are also works that use computer vision to build virtual drums. These systems have the benefit of making cameras easily available. However, real-time implementation of the methods could be challenging and complex. Without the use of drumsticks, Bering [2] and Famador's work locates and tracks the position of the hands in the video. However, they do not use their feet to move, which is necessary to identify the bass drum and the hi-hat control. Orange markers are placed to the ends of the drumsticks in a different experiment conducted by Brown et al. [3] in order to identify and monitor the movement of the drumsticks. Once more, their work is unrelated to mobility.

In Rojo's [4] work, coloured markers and a dynamics calculation for volume control of the synthetic drum sounds were also utilised. The most widely used and commercially successful computer vision implementation makes use of a high-speed camera and reflecting balls mounted to the ends of drum sticks. The Aerodrums [5], which cost $199, are the name of this item.

B. Virtual Piano

Over the past few decades, there have been numerous research projects aimed at creating virtual and augmented musical instruments. In order to improve user psycho-pleasure, boost instrument accessibility, and offer performance assistance, virtual reality and/or augmented reality techniques are used. Rogers et al.(2014)[6], Dirkse (2009)[7], Lin and Liu (2006)[8], Chow et al. (2013) [9], and A virtual piano created by Broersen and Nijholt in 2009 [10] that serves educational purposes and supports multi-agent playback was built. It is, however, difficult to play because it requires a genuine synthesiser or a mouse or keyboard as an input method. Other applications used gestural controllers that resemble instruments, including video cameras [Modler and Myatt](2008)[11], motion capture systems [Nymoen et al.](2011)[12], multi-touch devices [Ren et al.](2012)[13], data gloves [Mitchell et al.] (2012)[14], and more recently depth sensors. With the help of machine learning algorithms and a 3D depth sensor, Digito [Gillian and Paradiso](2012)[15] recognises hand gestures to play notes. The note is played by making a tapping motion with the right hand's tip of the index finger.

The user interface of Digito differs too much from actual piano playing with various fingers. Leap Motion Controller has been used in some applications to build virtual pianos that employ 3D finger placement to detect tapping [Heavers], but in these applications, users' hands are not permitted to interact with other objects, which is awkward and unnatural for a pianist. Leap Motion as a tracking device and algorithm for playing the piano was thoroughly examined by Han and Gold(2014) [16], who found that even though Leap Motion provides accurate tracking for free hand postures, it can be challenging for players to accurately judge the position and height of the virtual keyboard without practise.

3. Methodology

The study's workflow is shown in Figure 1. The system first takes a frame that was read from the camera and decodes it. The image is subsequently processed by being converted to the YCbCr color model. A collection of color spaces known as YCbCr is utilised in digital photography and video systems. The image is then given a Gaussian filter. The range of skin color values for YCbCr is used to produce a mask from the converted frame, and it is then used to detect all the contours for the areas whose colors fall inside this range. The computer recognises the colored objects and draws a rectangle around each one. As a result, it is simpler to follow the contours because variables can be created from the bounding rectangles' coordinates. If a specific corner of the bounding rectangle has coordinates that are within the boundaries of the drum pads. The simulation continues until the user clicks the 'Q' key on the keyboard to signal its conclusion.

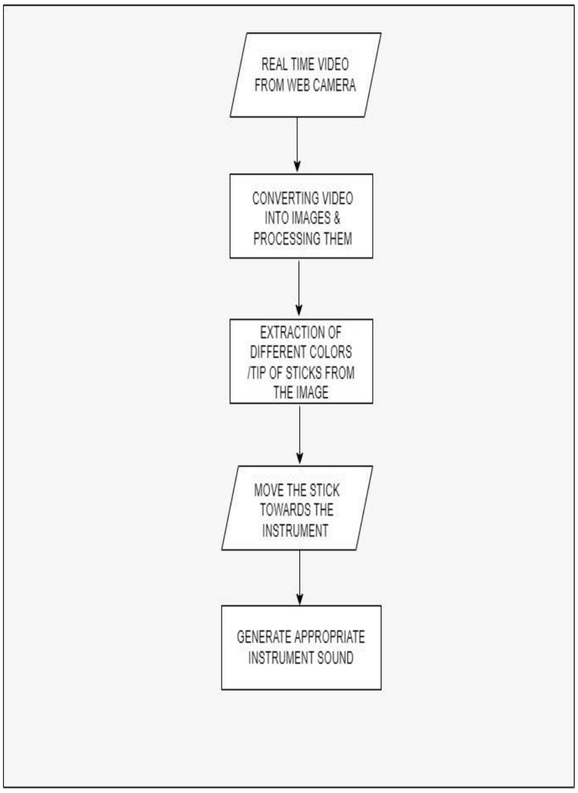

Fig. 1. Workflow Diagram

The system analyses the video stream to track the movements of the drumstick or colored object held by the person while playing the musical instruments and produces the corresponding sounds of an instrument in real time.

We are interested in the spatial positioning of drumstick movement. The types of instruments are determined by their spatial locations; for example, a hit-hat is placed farthest to the right and a snare drum is placed close by but slightly lower.

There are numerous methods, and characteristics like edge, texture, and color may be used to pinpoint the object. After extensive empirical testing, we have chosen to make color the primary feature. It is common knowledge that color is affected by an object's luminous and reflecting qualities. Prior to further processing, the color coding from a camera, which is by default in RGB, is converted into HSV. The drumsticks' tips are given a small trick. The colors picked—a dark blue or dark red in this experimental setting which are different from the surroundings.

The prototyped system uses blob detection based on color to find the keypoints. As a result, this approach presupposes that the red and blue color spectrums have different color ranges.

The background of the camera view shouldn't include any identically colored objects that are larger than the size of the thing being held by the person as seen by the camera.

In order to identify each keypoint for each frame that a camera records, thresholding is done. Each batch of extracted pixels is given a dilatation treatment to make them appear more blob-like. For each threshold, the largest blob that contains the desired keypoint is identified, and its centre is computed to identify the keypoint's location.

C. Frame Reading and Processing

In the experiment, many colour models, including HSV, were used. HSV is one of the two most often used methods for expressing a point in an RGB colour model. S represents for saturation, V for value, and H stands for hue. YCbCr was discovered to be the system's most efficient material in this investigation. A Gaussian filter was then applied to the picture to minimise noise and make image segmentation simpler, as shown in equation 1.

$$G_{2\,D}(x, y, \sigma) = \frac{1}{2\pi r^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

In this investigation, we employed a fixed threshold. The thresholding technique is shown next, where ϴ is the threshold and a[m,n] is the currently being processed frame. Values of ϴ are frequently used to determine skin tone. The value ϴ is to identify skin in a picture were reported by Basilio, et al. [17]. The cv2.inRange() function of OpenCV was utilised to put this into practise.

If a[m,n] ≥ ϴ a[m,n] = object = 1
Else a[m,n] = background = 0

Thresholding Algorithm

D. Contour Detection and Tracking

Detecting the colored objects and making a rectangle around it is contour detection. To put this into practise, we extracted the contours from the thresholded image using the cv2.FindContours() function. After that, the machine chooses which contours to track.

E. Sound Generation

Our virtual drum set has the following 12 drum parts: Ride, Ride Bell, Hithat close, Crash, Snare, Snare Rim, Tom Hi, Tom Mid, Tom Low, Kick, Hit Hat Open, Hit Hat. The arrangement of these components is shown in Figure 4. Once a drum pad hit has been detected, the generated drum sound is based on the computed position of the hit given the pre-defined bounding boxes.

One thread is generated for each drum set component to conduct the sound generating procedure for that particular component. The causes are (i) many components may be struck simultaneously, and each one's sound may be independently produced; (ii) when the same component is struck repeatedly, the effect of the previous sound should be replaced by the current one. Please be aware that there can be intricate interactions between successive impacts in the real world. However, the aforementioned presumptions result in a realistic-sounding rendered sound for the purposes of this air drum modelling prototype.

Two main procedures may be used to summarise the system's implementation:

1. In order to extract the appropriate colour pixels in a given frame, the recognition of the object carried by the person uses color-based object detection. Following the color-based object recognition method, the localisation of the item is determined from each frame using an algorithm.
2. The module that creates sounds uses spatial information. Sounds that are appropriate are produced using this information.

4. Results And Discussion

Our Virtual Musical Instruments was tested by using a red or blue colored object. Because we wanted our system to be dependable and reasonably priced to operate on things that people would easily access, we decided to use any red or blue colored objects as sticks. Other things including bottle caps and drinking straws were also recognised as legitimate markers due to the system's use of color detection.

The application was also tested to work on other laptop models; using the same application calibrations, it performed equally on both HP and Lenovo laptops.

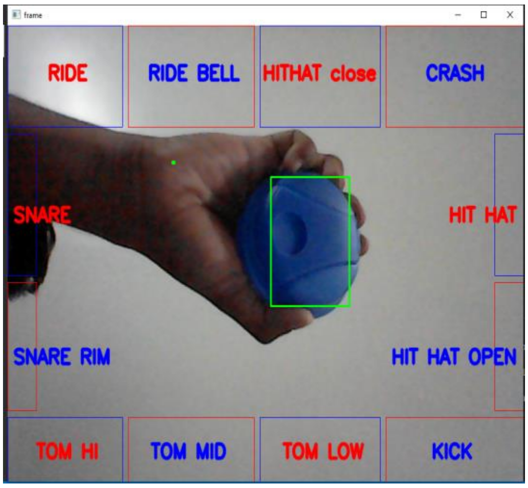

Fig. 2. Virtual Drum Kit Set

The figure 2 shows the Virtual Drum Kit Set during run time. It displays the mirror image of the user. The contours found in the image are outlined in green color.

Fig. 3. Mask Layer

The figure 3 depicts the mask layer. It creates the image mask of the blue colored object that is show in figure 2. Masking is a popular method for obtaining the Region of Interest (ROI). Using the draw function and bitwise operation in OpenCV, any masking shape may be created.

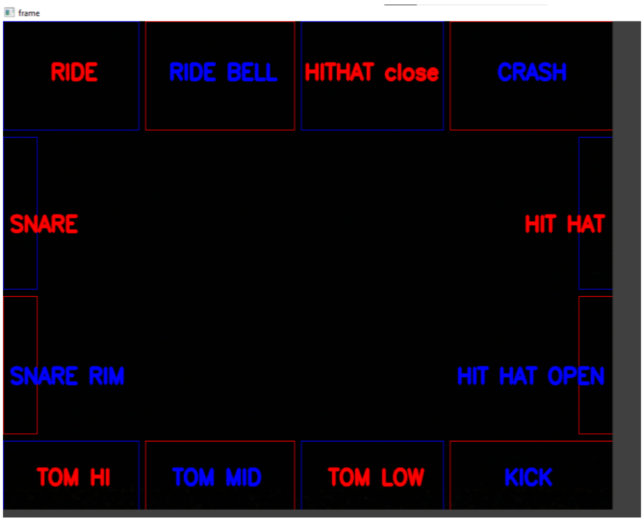

Fig. 4. Drum Area

The figure 4 depicts the drum area. Here we have 12 instruments of drum kit set: Ride, Ride Bell, Hithat close, Crash, Snare, Snare Rim, Tom Hi, Tom Mid, Tom Low, Kick, Hit Hat Open, Hit Hat.

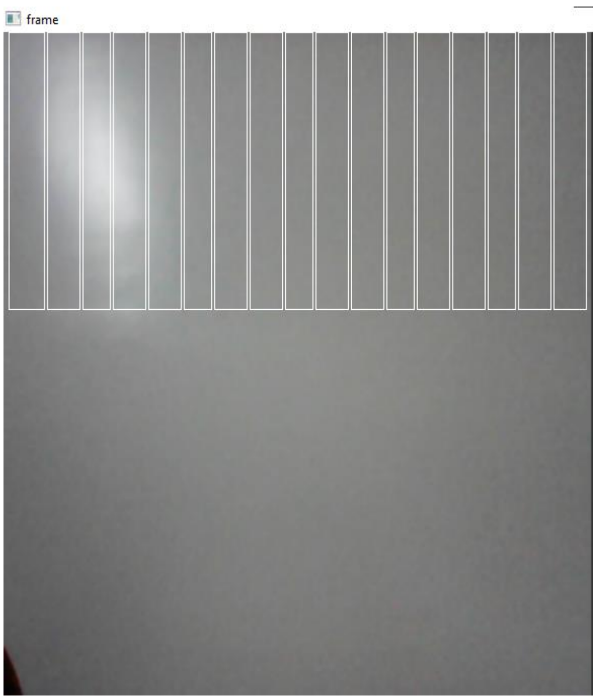

Fig. 5. Virtual Piano

The figure 5 depicts the virtual piano. It contains the keys of original piano.

5. Conclusion

According to our experiments, we can conclude that we have succeeded in creating a prototype system for a virtual drum and piano that is inexpensive to use and accessible to beginning performers. Although color-based detection for real-time detection is oversimplified, it is practical for the objective of obtaining the highest number of hits per minute in real-time due to its speed. With the help of this application, the user will be able to play the musical instruments virtually holding the Colored object.

The cost of a musical instrument kit is an expense that most prospective musicians will have to make in order to pursue their passion. This project attempts to speed musicians' introduction to the instruments experience while minimizing costs, as well as allowing musicians to practise, at least casually, without a full musical instrument set virtually. As a result, a larger audience can enjoy the pleasure of music. The instruments that our project will include are Drum kit, Keyboards (Piano) etc.

6. Future Scope

As computer technology continues to develop, people are going for smaller and smaller electronic devices. And increasingly we are recognizing the human to computer interaction and vision based gestures or object based gestures.

In the future, we'll be able to recognise hand and finger motions so that users can digitally play instruments by waving their hands or fingers around. The most efficient and expressive method of human communication, hand gestures constitute a universally recognised language.

References

[1] "Freedrum: The Drumkit That Fits in Your Pocket," 2017. [Online].

[2] S. R. F. Bering and S. M. W. Famador, "Virtual Drum Simulator Using Computer Vision," pp. 370–375, 2017.

[3] B. Brown, D. Quenneville, and M. Shashoua, "CCS453 Computer Vision Spring 2016 Final Project: Air-drumming through Video Object-detection", 2016.

[4] F.Rojo, "Air Drums Using OpenCV and Python," 2012. [Online].

[5] "Aerodrums," 2019. [Online]. Available: https://aerodrums.com/ .

[6] ROGERS, K., ROÂ¨ HLIG, A., WEING, M., GUGENHEIMER, J., KOÂ¨ NINGS, B., KLEPSCH, M., SCHAUB, F., RUKZIO, E., SEUFERT, T., AND WEBER, M. 2014. Piano: Faster piano learning with interactive projection. In Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, ACM, 149158.

[7] DIRKSE, S. 2009. A survey of the development of sight-reading skills in instructional piano methods for average-age beginners and a sample primer-level sight-reading curriculum. University of South Carolina.

[8] LIN, C.-C., AND LIU, D. S.-M. 2006. An intelligent virtual piano tutor. In Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications, ACM, 353 356.

[9] CHOW, J., FENG, H., AMOR, R., AND WUÂ¨ NSCHE, B. C. 2013

[10] BROERSEN, A., AND NIJHOLT, A. 2002. Developing a virtual piano playing environment. In IEEE International conference on Advanced Learning Technologies (ICALT 2002), 278282.

[11] MODLER, P., AND MYATT, T. 2008. Video based recognition of hand gestures by neural networks for the control of sound and music. In Proceedings of the InternationalConference on New Interfaces for Musical Expression (NIME), Cite seer, 57. NYMOEN, K., SKOGSTAD, S. A. V. D., AND JENSENIUS, A. R. 2011. Sound saber-a motion capture instrument. OIKONOMIDIS, I., KYRIAZIS, N., ARGYROS, A., ETAL. 2011.

[12] NYMOEN, K., SKOGSTAD, S. A. V. D., AND JENSENIUS, A. R. 2011. Sound saber-a motion capture instrument. OIKONOMIDIS, I., KYRIAZIS, N., ARGYROS, A., ET AL. 2011.

[13] REN, Z., MEHRA, R., COPOSKY, J., AND LIN, M. 2012. Designing virtual instruments with touch-enabled interface. In CHI12 Extended Abstracts on Human Factors in Computing Systems, ACM, 433436.

[14] GIRSHICK, R., SHOTTON, J., KOHLI, P., CRIMINISI, A., AND FITZGIBBON, A. 2011. Efficient regression of general-activity human poses from depth images. In IEEE International Confer-MITCHELL, T. J., MADGWICK, S., AND HEAP, I. 2012. Musical interaction with hand posture and orientation: A toolbox of gestural control mechanisms.

[15] Kumar, S. A. S., Naveen, R., Dhabliya, D., Shankar, B. M., & Rajesh, B. N. (2020). Electronic currency note sterilizer machine. Materials Today: Proceedings, 37(Part 2), 1442-1444. doi:10.1016/j.matpr.2020.07.064

[16] GILLIAN, N., AND PARADISO, J. A. 2012. Digito: A fine-grain gestural controlled virtual musical instrument. In Proc. NIME, vol. 2012.

[17] Venu, S., Kotti, J., Pankajam, A., Dhabliya, D., Rao, G. N., Bansal, R., . . . Sammy, F. (2022). Secure big data processing in multihoming networks with AI-enabled IoT. Wireless Communications and Mobile Computing, 2022 doi:10.1155/2022/3893875

[18] HAN, J., AND GOLD, N. 2014. Lessons learned in exploring the leap motion tm sensor for gesture-based instrument design. In Proceedings of the International Conference on New Interfaces for Musical Expression, 371374.

[19] Basilio, J. A. M., Torres, G. A., Pérez, G. S., Medina, L. K. T., & Meana, H. M. P. (2011). Explicit image detection using YCbCr space color model as skin detection. Applications of Mathematics and Computer Engineering, 123-128.