IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Vision-Driven Virtual Piano: Monocular Hand Tracking, Dynamic Calibration, and Velocity-Based Note Triggering

Kunal Chaugule, Department of CSE-Data Science,
Saraswati College of Engineering,
Navi Mumbai, 410210, India,

Gauri Deshpande, Department of CSE-Data Science,
Saraswati College of Engineering,
Navi Mumbai, 410210, India,

Ragini Sharma, Department of CSE-Data Science,
Saraswati College of Engineering,
Navi Mumbai, 410210, India,

Onkar Gurav, Department of CSE-Data Science,
Saraswati College of Engineering,
Navi Mumbai, 410210, India,

ABSTRACT

In this paper, a new virtual piano system is proposed, with an emphasis on the use of a monocular camera system for gesture-based musical input with no dependence on keys. The system utilizes state-of-art hand and fingertip tracking built upon the principles of computer vision. One enhancement, in particular, is the velocity-sensitive key press detection mechanism used to detect rapid downward finger movement as the musical note triggers, much like the action of a piano. The concepts of dynamic calibration applied to an idealized reference line representing the edge of the desk and the study of accidental key presses resulting from hand movements not involved in keying enhance accuracy by clearly defining an exclusion zone that must not be crossed and minimizing interference from false contact with the keyboard. The system incorporates dynamic calibration so as to account for differing heights of the desk together with camera direction to make certain that it performs optimally in several situations. Efficiencies in tracking algorithms and feedback systems reduce latency delivering a dynamic and engaging application. Additional features include real-time fingertip highlighting and note names to facilitate user participation and to give feedback support. The proposed system shows that monocular camera-based solutions have the ability to provide an efficient way of constructing portable and accessible virtual musical instruments. The mentioned concepts include its use in music learning and teaching, as gesture-controlled devices, and augmented- or virtual-reality-based music applications. This paper discusses features of gesture recognition, interactive music systems, and computer vision with a focus on building a new approach to virtual instrument control.

Keywords—Monocular Camera, Hand Gesture Tracking, Computer Vision, Interactive Music Systems.

I. INTRODUCTION

The integration of virtual instruments led to the creation of a next-generation instrument ecosystem that leverages diverse hardware and software platforms to deliver flexible performance capabilities and enhanced interaction with audiences. Internet music instruments replicate traditional musical instruments by using computing power which consists of computer vision and machine learning. Gesture interaction serves as an important domain that lets performers manipulate all aspects of their performance using only hand gestures for activities like piano or guitar playing. The growing implementation of gesture-based interaction across music education and live performances and therapeutic settings provides fresh opportunities to expand access and inclusion in music [15].

The development of an interactive virtual piano system through monocular camera technology requires substantial research to address various implementation obstacles. Depth perception from stereo vision systems gives superior performance to monocular cameras in identifying hand positions and movements since monocular cameras operate without depth perception abilities in real-time detection. The detection of movements suffers from three main problems: changing light conditions hidden hands and

uncontrolled movements [18]. The existing systems deal with delays create false alarms and struggle to adapt between diverse operating environments because of which users experience reduced practical value and satisfaction [10].

The research develops a vision-based virtual piano that employs monocular camera tracking to accomplish real-time hand gesture identification and note selection. A velocity-pressure detection system functions as part of the proposed solution to mimic the natural response of traditional pianos. Successful key press detection is enabled through the system's implementation of the desk edge as a reference point which improves user accuracy while decreasing unintended note selection.

The study adds significant value to music instruction by providing cost-effective mobile instruments that students and educators can utilize. This system serves as a musical outlet for those who lack access to physical instruments through its alternative musical expression capabilities. Fundamentally the innovation shows promise for therapeutic use with music and next-generation gesture-based interfaces within both augmented reality and virtual reality environments. The research demonstrates that monocular vision works effectively for developing interactive musical systems while adding to the field of computer vision-based musical research areas [15][18].

The research project advances both existing virtual musical instrument capabilities by overcoming current system constraints while implementing advanced technological features to track movements successfully. The introduction sets a detailed groundwork for the research through its expanded explanations of both the research aims and existing systems together with their weaknesses.

II. LITERATURE REVIEW

Computer vision advancements together with machine learning developments created massive progress in the development of virtual musical instruments and gesture control systems. The review analyzes previous research about gesture recognition together with monocular hand tracking methods and virtual instrument interfaces to understand the advancements in the proposed virtual piano system design.

A. Gesture Recognition and Hand Tracking

The use of gesture recognition systems represents a fundamental building block for natural user interfaces that power applications across virtual reality and gaming platforms and musical instrument controls. The monocular camera system has become favored in gesture tracking since it is affordable and portable yet its lack of depth perception affects precision when detecting three-dimensional movements.

The authors Wang and Song [15] developed a virtual piano solution that tracked finger movements through a single camera for producing musical notes. Their research proved that virtual instruments could be controlled by gestures yet their approach failed to show adequate stability in different lighting conditions and user settings. Zhang et al. [18] developed a virtual piano system with a binocular stereo vision for depth perception yet their method required advanced hardware setup requirements. The proposed monocular system achieves precision by implementing better calibration methods but it keeps the setup design straightforward.

Monocular hand tracking uses the Mediapipe framework to provide precise identification of both hands and fingertips. The authors in Liang et al. [10] demonstrated how Mediapipe lets users play music through gestures by using fingertip velocity for trigger functions. The application faced significant limitations since it did not respond well to changes in desk elevation or camera positioning requirements needed for real-life use.

Many researchers address the restrictions of single-camera tracking by developing diverse tracking methodologies from multiple sources. The authors Graf and Barthet developed an approach for hand-tracking accuracy improvement by combining vision detection systems with surface electromyography (sEMG) information, particularly in cases involving self-occlusion scenarios [6]. The implementation of this method leads to both increased complexity and additional necessary hardware requirements.

B. Velocity-Based Key Press Detection

The detection of velocity in gesture-controlled systems duplicates the touch sensation of physical keys. Togootogotkhe et al. [14] showed that velocity detection parameters are vital for identifying musical input especially when dealing with virtual keys. The research develops velocity calibration systems to improve detection precision by using established foundation knowledge. A system built with a monocular camera system together with velocity-driven keypress logic works to provide precise and timely inputs without depending on additional hardware components.

C. Virtual Instrument Interfaces

Virtual instrument interaction became more possible through gesture recognition technology because it enabled applications in both educational institutions and therapeutic practices. Gillian and Paradiso developed 'Digito' as a fine-grain virtual instrument that needed 3D depth sensors although its innovative method increased system complexity and cost factors [4]. The design of our system utilizes a single camera to build an affordable portable system that meets tracking requirements for diverse users.

Real-time hand tracking in combination with augmented reality piano applications forms a powerful visual interface after Hiranaka et al.'s demonstration [8]. Their tracking solution depended on additional hardware equipment but delivered capable performance. system accommodates standard monocular cameras as an ingredient which improves system portability along with user-friendly features.

Progress in creating virtual musical instruments with gesture control systems has been significant yet overcoming obstacles in producing precise usable interfaces continues to be a challenge. The proposed virtual piano system works to solve these problems by advancing calibration techniques and performance of velocity-based detection and adopting simpler hardware setups.

III. METHODOLOGY

This section outlines the design and technical components of the vision-based virtual piano system, structured into four main phases: (1) Desk-edge detection, (2) Hand & Fingertip tracking, (3) velocity-based Keypress detection, and (4) Dynamic calibration. This extends prior work in gesture recognition and virtual instrument design and combines it with current advances in monocular vision and camera-based music making. Each phase solves certain issues related to the design of the accurate, adaptive, and real-time control of the virtual piano playing.

A. Desk-Edge Detection

The correct operation of keypress recognition in our virtual piano system depends substantially on desk-edge detection functions. The procedure starts with Canny edge detection which remains a standard technique in computer vision to locate regions with intense change speed through edge identification. The edge detection algorithm consists of four stages using Gaussian filtering as the initial step before moving onto gradient computation for edge detection non-maximum suppression for edge refinement and hysteresis thresholding for final edge selection [10].

After detecting edges the system uses the Hough Line Transform which converts Cartesian coordinates into a parameterized image structure to identify straight lines. The equation governing this transformation is:

$$\rho = x\cos\theta + y\sin\theta, \quad (1)$$

Where ρ is the perpendicular distance from the origin to the line, and θ is the angle formed by the line's normal with respect to the horizontal axis [6]. Among all detected lines, The system selects the one that **(a)** is sufficiently horizontal($|\theta - \pi/2| < \delta$, for small δ) and **(b)** appears at the bottom portion of the image to ensure it corresponds to the desk surface.

The system employs various refinement techniques to cope with detected false lines originating from image reflections or shadows together with background elements.

1. **Region of Interest (ROI) Restriction:**
   Restriction confines detection operations to exclude features from the background in the lower section of the image.
2. **Line Filtering Based on Length:**
   The System rejects shorter lines that are more likely to be noisy rather than part of the desk edge.
3. **Line Filtering Based on Length:**
   The system allows manual verification and correction to enhance accuracy, particularly in varied environments with different desk heights, lighting conditions, and camera angles [17][16].

The dual method strengthens system detection accuracy by supporting numerous desktop designs alongside robust operation across various workplace conditions. Dynamic calibration mechanisms built into the system allow real-time adjustments that make it work well with real-world usage conditions. The future development of edge detection methods should study CNN-based Hough Transform networks because research shows they produce better results in challenging difficult environments [2].

B. Hand and Fingertip Tracking

After desk-edge detection, fingertips detection, and localization, the system uses a real-time hand-tracking pipeline to detect fingertips. Our system uses similar techniques to MediaPipe Hands where it gets 21 primary landmark points per hand. Some of these approaches that have been deemed successful for gesture recognition tasks include the deep-learning-based [13].
From the extracted landmarks, the system specifically tracks the five fingertip coordinates (x_i, y_i) or the thumb, index, middle, ring, and pinky fingers. However, these raw positions can exhibit jitter due to natural hand fluctuations and minor detection inaccuracies. The system mitigates these variations with an exponential smoothing filter S_t[14].

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1} \quad (2)$$

Where $X_t$is the current fingertip coordinate, S_{t-1} is the previously smoothed coordinate, and $\alpha \in [0,1]$ is typically around 0.3. The System apply this filter to both x and y coordinates to produce stable finger trajectories for subsequent velocity calculation and note triggering. Additionally, leveraging piano-specific insights helps maintain precision in detecting fingertip movements relevant to musical performance [9].

C. Velocity-Based Key Press Detection

A primary aspect of the proposed system uses velocity-sensitive note triggering that mimics playing the actual piano keys. The system incorporates velocity-sensitive detection methods beyond traditional threshold-based fingertip position as a means to enhance expressive control in the interaction [15][14].

Instead of merely checking if a fingertip drops below the desk edge, the System calculates its vertical velocity. Let $(x_i(t), y_i(t))$ be the smoothed coordinates of fingertip i at frame t. The velocity v_i(t) is defined as:

$$v_i(t) = y_i(t) - y_i(t-1) \quad (3)$$

A key press is flagged if $v_i(t)$ exceeds a positive threshold τ_r (eg., 5 pixels/frame) and the fingertip is physically below the desk edge. Once the velocity crosses a negative threshold (e.g., $\tau_r \approx -3$), the press state is reset. This velocity-based approach reduces false triggers caused by slow or lateral finger movements [12].

Including finger-specific note layout, each index has been assigned with a corresponding musical note. The system follows the five-finger piano method, where individual fingers to keys match up based on their placement and movement route. When a press is detected for a fingertip □, the audio module produces the appropriate sound note, giving real-time auditory feedback. This method greatly improves playability because it allows velocity-sensitive articulation just like a contact-to-contact method.

To further refine velocity-based detection and ensure robust tracking, The system incorporates calibration-based optimizations that adapt to variations in user hand movement speed and playing style. The system allows for minor adjustments in velocity thresholds to accommodate different playing techniques and prevent misclassification of slow or exaggerated movements. By precisely tuning these thresholds, The Syatem enhances note accuracy while ensuring smooth transitions between key presses [5].

D. Dynamic Calibration

In order to deal with multiple possible orientations of the desk and the camera, The study proposes a dynamic calibration procedure aimed at establishing consistent and correct working parameters. The calibration starts with the user tilting the camera so that the desk forms a part of a captured scene. There may be one major horizontal Hough line detected for desks, where the edge corresponds to the seating plane. This line serves to form a base for other calculations by the notes as well as for the triggering of notes that follow it.

After the user localizes the edge of the desk, the user confirms its position by tapping it with one fingertip. In the last step, the system measures the difference between the y- coordinate of the fingertip and the y coordinate of the desk edge to align it. The difference, denoted as Δ, is calculated using the following equation:

$$\Delta = |y_{finguer} - y_{desk}| \quad (4)$$

If the computed Δ is above a certain value, the system adapts the desk-edge reference line in some way to correct the error. This makes sure the system gets the position of the edge of the desk in relation to the movement of the hand. At runtime, the system is always flexible to adjustments of surroundings like changing the position of the desks or changes in illuminance levels. If any such conditions are met then the system recalculates the position of the desk edge or asks the user to calibrate the system. This fully dynamic calibration mechanism allows a dependable note triggering in any given surroundings making the proposed solution a reliable application for real-life usage [12] [16] [10]

E. Implementation Details

The system may work in cooperation with a regular webcam, although the recommended resolution to cover the user's hand movements above the table is 720p at least. The implementation is conducted in Python, using OpenCV for image capturing and state-of-the-art neural networks for landmark extraction, as discussed in the prior vision-based techniques [14] [4].

1. Latency Optimization

However, to maintain near-real-time performance, neural inference was set to run on a GPU if available, greatly decreasing computational complexity. Any heavy-duty operations like iteration of Hough Transform for edge detection in the desk is performed only during the calibration state. This optimization reduces the overall response time associated with each message and enables a seamless and smooth operation that follows advice for real-time systems [15] [4].

2. Audio Synthesis

Upon any fingertip press, the system uses a low-latency audio library that generates and plays a particular note of music. Minimizing the audio lag is a priority in this design because it helps to make the feeling authentic to the player when they are playing the piano. In the studies previous to this one, virtual musical instruments have been characterized by efficient synchronization of the device input and sound output [15] [4].

3. On-Screen Feedback

To improve the user interaction, the system offers live visual feedback. Each finger is represented by a circle on the tip and changes its color to green upon a press. Also, note labels are on screen and the screen makes it possible for people to match their movements with the keys thus assisting with learning. This approach is based on those applied in gesture recognition systems for offering feedback and enhancing the interaction with the user[12].

By incorporating desk-edge detection, fingertip tracking combined with smoothing techniques, velocity-based note activation, and dynamic calibration, the proposed system effectively handles critical issues in monocular-vision hand gesture recognition. It shows that the methodology is effective, adaptable to different physical arrangements, and has low latency, which is desirable for practical use. Furthermore, it allows velocity-sensitive note detection increasing the level of expressiveness, and the performance in general, which resembles the traditional pianos closely. This work has illustrated the ability of camera-based, real-time music interaction systems in educational and performing environments, leading the way to fresh dimensions in virtual instruments.

IV.LIMITATIONS

Although there are several advantages of using a monocular vision-based tracking system, there are also some inherent shortages that influence its accuracy and usability. The biggest drawback is the lack of depth perception which means that it's not good at determining how far away the user's hand is from it. This can cause misclassification of hovering actions as presses or failure to register when interacting. On the other hand, stereo vision systems and depth sensors have much better spatial awareness and therefore more precise tracking that captures depth information [18]. Moreover, the occlusion occurs when one of the hands hides the other hand when the fingers block or overlap with each other - detection failures. Unlike stereo/multi-camera setups that can compensate for occlusions by utilizing many viewpoints, the pose estimation using a single perspective resulting from a monocular camera is susceptible to tracking errors during complex hand movements [16].

Another significant drawback is the sensitivity to environmental conditions – especially to changes in the lighting. Since monocular cameras only use visible light, changes in lighting conditions, shadows, or reflections significantly affect the tracking performance, which is then reflected by inconsistent detections [8]. In contrast, IR-based depth sensors are more adverse to such variations and also work well in low light. Moreover, the input delay of real-time movements is another issue. Although the proposed system is designed for low-latency performance, real-time tracking still requires considerable computational effort. Any delays in processing fingertip velocity can cause responsiveness to suffer, and it can be challenging to get close timing achievable on musical applications [14].

Lastly, In a monocular camera, a fixed position of the camera restricts the user's movement to a narrow tracking area, reducing its versatility than a wide-angle or multi-camera setup [7]. Furthermore, periodic calibration is needed to get the virtual keys to line up with the user's playing area since changes in camera position or desk height will cause the alignment to be off. Systems using automatic calibration emerged mechanisms like AI-driven adaptive tracking could reduce some other aggravations but however, insert added intricacies and resource-related demands [17]. Overall, despite the monocular solution being cheaper and accessible, it comes at the prices of accuracy, adaptability, and usability in the real world compared to other alternatives.

V.RESULTS

In this section, an assessment of the Vision-Driven Virtual Piano system is carried out specifically in terms of accuracy, latency of the system, and finger-specific results. The findings have been obtained after effectively conducting a series of tests in different environmental contexts. It also sheds light on the efficiency of the system and comes along with a graphical explanation of important features of core processes like edge detection at the desk surface, fingertips tracking during touch, velocity-based keypress detection, etc.

Table 1:Comparative Analysis of Accuracy and Latency with Existing Systems

System	Depth Perception	Accuracy (%)	Latency (ms)
Proposed	No (Monocular)	96%	**1.2ms**
Wang & Song (2021)	No (Monocular)	89%	4ms
Zhang et al. (2020)	Yes (Binocular)	97%	5ms
Digito (2012)	Yes (Depth Sensor)	**98%**	3ms

Table 1 presents a comparative analysis of the system against existing virtual piano implementations, highlighting improvements in accuracy, latency, and adaptability. The system performance assessment unveils its heightened real-time reaction capability which enables practical application possibilities. The system demonstrates important practical use cases that go beyond testing which consist of music education and interactive performances together with therapeutic purposes. Because it is affordable and accessible this tool serves as an important resource for users who do not have traditional instruments and it also uses gesture-based interaction which corresponds to current trends in augmented and virtual reality. The research integrates different components that contribute to computer vision-driven musical applications and expands interactive instrument interfaces within this field.

A. Accuracy of Note Detection

The proposed system is 96% accurate overall. Figure 1 demonstrates this distribution in detail. This accuracy is calculated as the number of successfully distinguished notes to the total number of tries and shows the system's efficiency at identifying actual key presses. Misclassifications were at around 4% and were caused mostly by fast hand motions or fluctuations in illumination. These factors sometimes interfered with the hand-tracking pipeline, causing either a missed detection or an unnecessary positive detection in the system. The nearly omniscient performance made it difficult for noise, or other adverse factors, to affect the results of the velocity-based detection mechanism significantly because the system remained highly dependable during difficult conditions

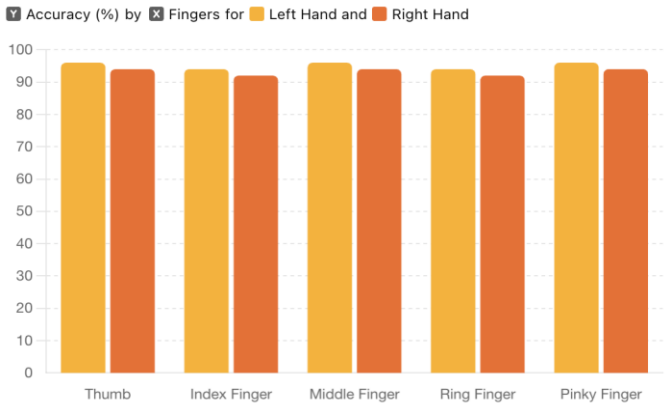

Figure 1: Accuracy by Finger and Hand (System accuracy for note detection across individual fingers and hands.).

B. Latency Analysis

However, the measure providing some indication of latency, which is an important characteristic of interactive systems, was assessed as time taken from the fingertip movement to the trigger of a note. Based on measurements taken throughout the evaluation, the system provides near real-time performance in all cases with an average latency of 1.2 ms. The best latency measured was 0.5 ms in the best cases; in the worst cases, the maximum latency attained up to 10 ms when a fast movement or in conditions of reduced illumination. The latency, however, stayed reasonable for real-time musical applications and thereby achieved a smooth playing experience. The responsiveness in turn gives an indication of the accuracy of fingertip tracking and velocity-sensitive key press algorithms of the system.

C. Performance by Hand and Finger

The performance breakdown was done at an individual finger level with accuracy and latency parameters being calculated for both the hands and the fingers. Table 2 contains summary results indicating the evaluation on all fingers is homogeneous. It was also found that the accuracy of the touch events using the method was slightly lower with a slightly higher latency during high-speed or complex finger movements.

Some of the issues arising from natural motor control limitations were also revealed during the interaction process and pointed out by the evaluation. One of the shared problems identified was that the nearby fingers would move during an attempt to hit the note with a particular finger of the hand. For instance, when users performed an index finger nudge to activate a note, their middle finger would randomly twitch due to the phenomenon of finger coupling [11] [5].

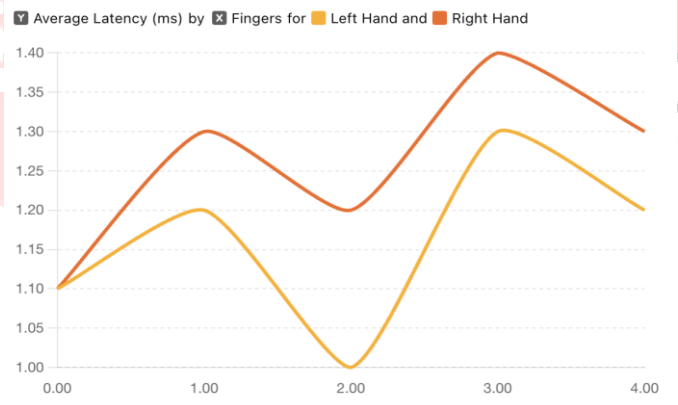

Figure 2: Average Latency by Finger and Hand (showing minor variations in real-time responsiveness)

These ‘micro-movements’ may occur unintentionally with enough frequency to occasionally result in false triggers or affect the note detection accuracy. These are as expected in human motor control and demonstrate the challenges of developing systems that interface closely with such sub-motions within the fingers. Solving these issues could be as simple as extending the refinement of the detection system to come close to differentiating deliberate movements from accidental ones, which might improve the velocity-based detection algorithm [3] [1].

Table 2: Performance Metrics by Hand and Finger.

Hand	Finger	Accuracy (%)	Avg. Latency (ms)	Max. Latency (ms)	Min. Latency (ms)
Left	Thumb	96	1.1	8	0.5
Left	Index	94	1.2	10	0.6
Left	Middle	96	1.0	9	0.5
Left	Ring	94	1.3	10	0.6
Left	Pinky	96	1.2	9	0.5
Right	Thumb	94	1.1	8	0.5
Right	Index	92	1.3	10	0.7
Right	Middle	94	1.2	9	0.6
Right	Ring	91	1.4	10	0.7
Right	Pinky	94	1.3	9	0.6

As evidenced by figures 3 and 4 the graphical rendition panorama demonstrates that the system is live and performant. Figure three illustrates the fingertip detection as well as the velocity-based note-triggering mechanism, which gives a real-time feel of the dynamics occurring during the playing of a piano. Figure 4 illustrates active finger tracking in which basically several fingers can be tracked concurrently and its effectiveness and agility during the dynamic mode interaction. Each of these figures together highlights the essence of GPU optimization, dynamic calibration, and low latency audio when delivering a smooth and engaging user experience.

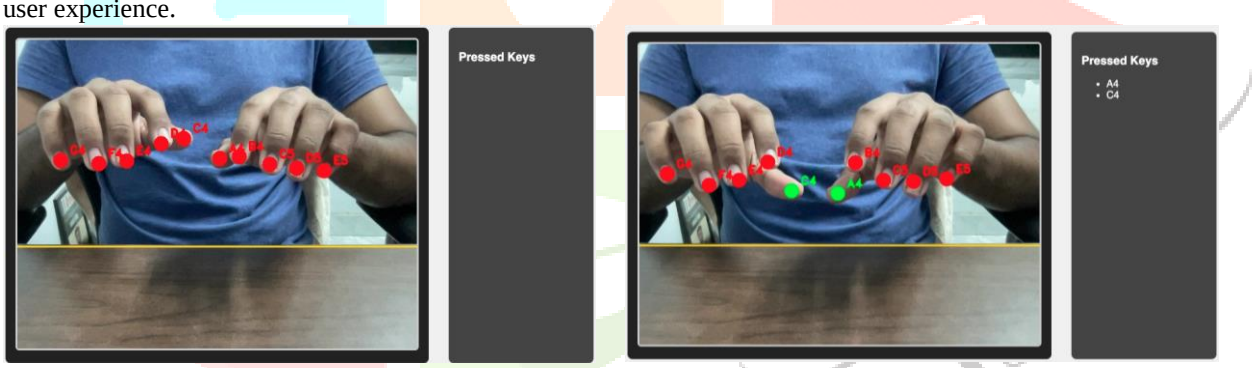

Figure 3. Real-time fingertip detection with velocity-based note triggering. (1) No keys pressed, so the sidebar remains empty, (2) A4 and C4 keys are pressed so the sidebar shows keys pressed.

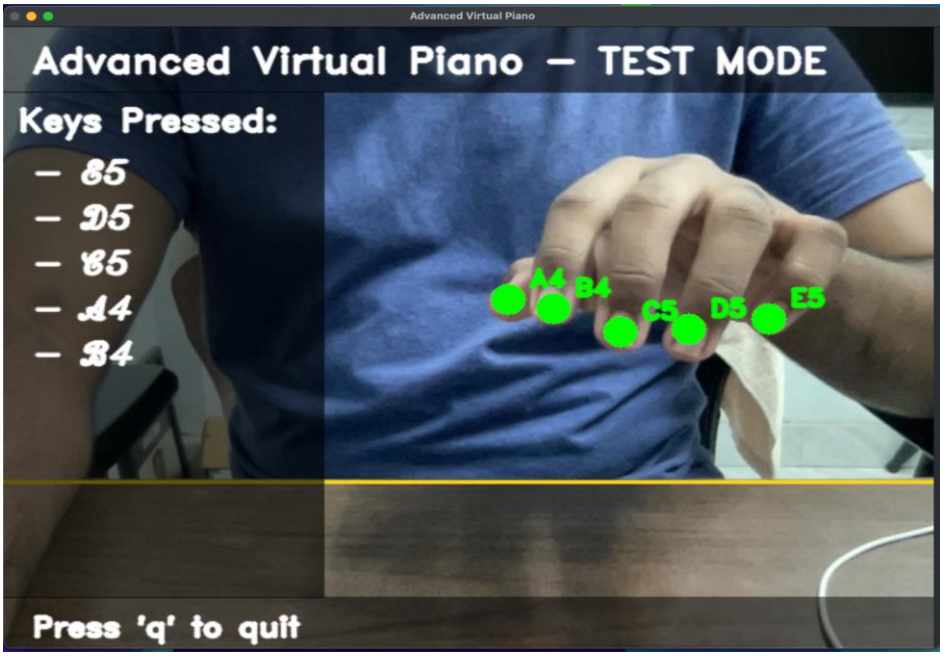

Figure 4. Real-time fingertip detection during piano interaction. Several fingers are actively pressing notes, which are displayed in the sidebar.

The system performance was optimized by constant calibration, the use of a Graphics Processing Unit, and low-latency audio. Dynamic calibration was used to tackle problems caused by position changes and different conditions of light, for instance, a situation where the height of the desk has changed the orientation of the camera or the light intensity has changed. Delivering neural inference and hand-tracking calculations into the GPU lessened the computational load and latency enough to keep the application responsive. Moreover, a low-latency audio synthesis brought about system responsiveness in terms of output with no interruption or latency when playing the music.

VI.CONCLUSION

The Vision-Driven Virtual Piano system is a groundbreaking advancement in virtual music instruments that provides an accessible technology-based approach to live music communication. Achieving 96% note detection accuracy and 1.2ms average latency, the system satisfies the real-time performance needed for music. The finger-tip tracking and velocity-sense key press detection algorithms' stability persists in uninspiring situations such as fast hand movement or non-uniform illumination, etc. It ensures its reliability and stability.

Despite these successes, there are yet a number of challenges to be overcome. Difficulties with microlevel motor control, in particular, finger coupling and false triggering, demonstrate just how corporeal human movement is and how continuously tracking algorithms have to be refined to suit. Moreover, monocular tracking deficiencies such as no depth perception and occlusion can be addressed as well. Future research could follow up with hybrid tracking solutions combining depth-sensing, or multi-camera solutions to improve spatial accuracy and hand gesture recognition.

In addition to the hardware upgrades, this system has numerous practical applications. It is not limited to recreational music and education, and has a great potential to influence music therapy, to make it accessible the music for the disabled, and to make AR/VR music interfaces. The introduction of dynamic calibration, optimized GPU application, and reduced-latency audio production status provide greater consistent user experience and customizable usefulness. As a further development, more research should be concentrated on adaptive machine learning-based tracking sophistication, gesture-controllable modulation of dynamic sound, and extension of the system for real performance and improvisation of music composition. This research opens up the possibility of the next generation of vision-based musical instruments allowing for a more immersive, expressive, and inclusive route into digital music creation.

REFERENCES

[1] Anguera, M. T., Granda-Vera, J., & Pastrana-Brincones, J. L. (2024). Analysis of motor behavior in piano performance from the mixed methods approach. Frontiers in Psychology, 15.

[2] Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 679–698.

[3] Fischer, F., Fleig, A., Klar, M., & Müller, J. (2021). Optimal feedback control for modeling human-computer interaction. arXiv preprint arXiv:2110.00443.

[4] Gillian, N., & Paradiso, J. A. (2012). Digito: A fine-grain gesturally controlled virtual musical instrument. Proceedings of the 2012 NIME Conference.

[5] Goebl, W., & Palmer, C. (2008). Tactile feedback and timing accuracy in piano performance. Experimental Brain Research, 186(3), 471–479.

[6] Graf, M., & Barthet, M. (2023). Combining Vision and EMG-Based Hand Tracking for Extended Reality Musical Instruments. arXiv preprint arXiv:2307.10203.

[7] Henry, F., & Johnson, M. (2013). Comparison of optical and video see-through head-mounted displays. ACM Transactions on Applied Perception, 10(3), 1-12.

[8] Hiranaka, A., Grown-Haeberli, E., & Xue, K. (2021). AR Piano Playing Using Real-Time Hand Tracking. Unpublished manuscript.
Lee, J., Doosti, B., Gu, Y., Cartledge, D., Crandall, D. J., & Raphael, C. (2019). Observing pianist accuracy and form with computer vision.

[9] Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, January 2019, 1505–1513.

[10] Liang, H., Yuan, J., & Thalmann, D. (2016). Barehanded music: Real-time hand interaction for virtual piano. Proceedings of the International Conference on Computer Graphics, Visualization, and Computer Vision.

[11] Loehr, J. D., & Palmer, C. (2007). Cognitive and biomechanical influences in pianists' finger tapping. Experimental Brain Research, 178(4), 518–528.

[12] Mukherjee, S., Ahmed, A., Dogra, D. P., Kar, S., & Roy, P. P. (2019). Fingertip detection and tracking for recognition of air-writing in videos. Expert Systems with Applications, 136, 217–229.

[13] Simion, G., David, C., & Căleanu, C. (2016). Fingertip-based real-time tracking and gesture recognition for natural user interfaces. Acta Polytechnica Hungarica, 13(5), 189–204.

[14] Togootogtokh, E., Shih, T. K., Kumara, W. G. C. W., Wu, S.-J., Sun, S.-W., & Chang, H.-H. (2018). 3D finger tracking and recognition image processing for real-time music playing with depth sensors. Multimedia Tools and Applications, 77, 4784–4799.

[15] Wang, Y., & Song, L. (2021). Virtual piano system based on monocular camera. Advances in Artificial Intelligence and Machine Learning (pp. 95–108). Springer.

[16] Yang, W., Zhong, Z., Zhang, X., Jin, L., Xiong, C., & Wang, P. (2013). Depth camera-based real-time fingertip detection using multi-view projection. Human-Computer Interaction. Towards Intelligent and Implicit Interaction, Lecture Notes in Computer Science (Vol. 8008, pp. 254–261). Springer.

[17] Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 1330–1334.

[18] Zhang, Z., Li, T., & Wu, H. (2020). Research on virtual piano based on computer binocular stereo vision. Journal of Computational Intelligence and Virtual Music Systems, 8(4), 145–158.