Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

The Literature Survey on Virtual Piano

Yashwanth G[1]
UG Students[1],
Department of Information Science and Technology
Vidya Vikas Institute of Engineering & Technolgy
Karnataka, India

Saifulla Khan[2]
UG Students[2],
Department of Information Science and Technology
Vidya Vikas Institute of Engineering & Technolgy
Karnataka, India

T V Rahul Reddy[3]
UG Students[5],
Department of Information Science and Technology
Vidya Vikas Institute of Engineering & Technolgy
Karnataka, India

Sukanya H A[4]
UG Students[4],
Department of Information Science and Technology
Vidya Vikas Institute of Engineering & Technolgy
Karnataka, India

Varsha N[5]
Faculty[5],
Department of Information
Science and Technology
Vidya Vikas Institute of Engineering & Technolgy
Karnataka, India

***Abstract*—This paper presents an efficient data-driven approach to track fingertip and detect finger tapping for virtual piano using an RGB-D camera. We collect 7200 depth images covering the most common finger articulation for playing piano, and train a random regression forest using depth context features of randomly sampled pixels in training images. In the online tracking stage, we firstly segment the hand from the plane in contact by fusing the information from both color and depth images. Then we use the trained random forest to estimate the 3D position of fingertips and wrist in each frame, and predict finger tapping based on the estimated fingertip motion.**

Finally, we build a kinematic chain and recover the articulation parameters for each finger. In contrast to the existing hand tracking algorithms that often require hands are in the air and cannot interact with physical objects, our method is designed for hand interaction with planar objects, which is desired for the virtual piano application. Using our prototype system, users can put their hands on a desk, move them sideways and then tap fingers on the desk, like playing a real piano. Preliminary results show that our method can recognize most of the beginner's piano-playing gestures in real-time for soothing rhythms.

***Keywords*— *Fingertip Tracking, Finger Tapping Detection, Virtual Piano, RGB-D images, Human-Computer Interaction*.**

I. INTRODUCTION

Recent years have witnessed rapid progress of hand pose tracking and hand motion analysis using consumer depth sensors. State-ofthe-art techniques [Tagliasacchi et al. 2015][Sun et al. 2015] are able to accurately track hand motion and handle intricate geometric configurations with complex contact patterns among fingers in real-time.

However, most of them require that hands are in the airand cannot interact with physical objects. Such a requirement diminishes their utility for virtual instrument applications due to two reasons: First, users can quickly get tired when hands are not supported by some physical object. Second, mid-air interactions do not provide user any feedback, hence users may feel difficult to position their fingers and map them to the keys or strings of virtual instrument.

This paper aims at developing a virtual piano application, which allows users to put their hands on a desk, move them sideways and then tap fingers on the desk, like playing a real piano. There are two major technical challenges in this application. First, the system must track the positions of fingertips and detect their status, i.e., whether a finger is tapping or not. Due to frequent interaction between fingers and desk, the existing hand tracking algorithms often fail. Second, piano-playing gestures are usually fast and complex, involving highly flexible hand articulation and causing severe hand self-occlusion.

To tackle these challenges, we propose a virtual-piano tailored method to track fingertip and detect finger tapping using an RGB-D camera in real-time. We first collect a training dataset with 7200 RGB-D images, covering the most common finger articulation for playing piano. After manually labeling the positions of seven hand joints including five fingertips, thumb MCP joint and wrist center, we train a random regression forest to predict them using depth context features of spatial-voting pixels randomly sampled over the training images. During online testing, we first predict the positions of the hand joints from raw RGB-D images with the trained

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

random forest. Then we use the trajectories of these joints to detect and locate finger tapping using support vector machine (SVM) classification. The virtual piano is registered onto the desk surface using pre-detected normal vector and centroid of the desk surface.

Based on the locations of fingertips and the finger tapping status, the system can hereby determine which piano key is pressed and play the corresponding sound. Preliminary results show that our method can recognize the basic piano-playing gestures in real-time for soothing rhythms. Figure 1 illustrates our virtual piano application with a DepthSenser 325 sensor on top of the desk and in front of the user. We render the hand and the piano based on the coordinates of the desk surface and the detected hand pose from the RGB-D images.

II. RELATED WORK

Hand pose tracking and evaluation is a fundamental hassle in laptop portraits and vision, and is central for many human-computer interfaces. Early gesture reputation programs resort to the usage of facts gloves or uniquely colored gloves/markers on palms or hands [Aristidou and Lasenby 2010]. In latest years there has been a developing interest in non-invasive setup using a unmarried commodity RGB-D sensor, consisting of Microsoft Kinect, Intel Real Sense, or reason-designed hardware, e.g., the Leap Motion Controller. Such unmarried-dig cam acquisition does not obstruct consumer movements, hereby is specifically effective to VR applications. This phase in short evaluations related work accessible pose monitoring, finger motion popularity and digital musical instrument.

A. Hand Pose Tracking

Algorithms for imaginative and prescient-primarily based hand pose monitoring can be widely labeled as generative version-fitting methods and discriminative strategies. Each magnificence of algorithms have their very own merits and drawbacks. The version-becoming strategies [Melax et al. 2013][Tagliasacchi et al. 2015] reconstruct hand poses with the aid of becoming a 3-d articulated hand model to intensity photographs. These methods work properly in controlled environments, but, they usually require calibration and their outcomes are sensitive to initialization. The discriminative techniques require an annotated dataset to examine a regressor offline, after which use it to are expecting the hand pose on-line. Such techniques are strong to initialization, however their accuracy closely depends on the size of the training dataset. Therefore, the dataset must be fairly massive to cover the viable hand and finger articulations for a selected software.The latest strategies (e.G.,[Tang et al. 2013; Sun et al. 2015; Xu and Cheng 2013]) require that the hand is inside the air and no longer interacting with other items. The purpose is that, hand motion itself is of excessive ranges-of-freedom and as a consequence calls for lots of training information to symbolize such flexibility. Thus, if the hand is interacting with unknown gadgets, there could be greater unpredictable complexity, e.G., hand self-occlusion and occlusion between item and hand, and the big look versions of each the hand and interacting objects.

These problems avoid researches in this vicinity. There are some preceding paintings that may help hand interacting with gadgets, however they either assume that the geometrical information of the item is known so that item and hand can supplement every other to enhance pose estimation [Oikonomidis et al. 2011] through the physical constraints among them or confine the possible hand articulations to be inside a small set of templates [Rogez et al. 2014]. In [Oikonomidis et al. 2011] the type and specific length of the interacting item are assumed to be known earlier. The poses of the object and hand are jointly solved in a generative version-fitting framework the use of a multi-digicam placing to lessen prediction ambiguity, which maximizes the fashions' compatibility to the picture inputs and minimizes the intersection among hand and objects to find their pose.

In [Rogez et al. 2014] the hand is authorized to have interaction with different items, such as bottles, desk surfaces, etc., and the hand pose is expected in a discriminative way via education a multi-class cascade classifier on a dataset that covers many interacting examples between hand and objects. However, in their hand pose estimation framework, the hand posture is best assumed to belong to a small set of pre-described templates. This is far from our want to play the piano in the proposed software, wherein we want to music the accurate articulated fingertip positions and wrist positions, so that the device can come across whether or not a finger tap is finished with the aid of the performer or now not.

Among the discriminative strategies, random regression woodland and its versions have validated powerful to capture hand pose in depth photos [Xu and Cheng 2013; Tang et al. 2013; Liang et al. 2015; Sun et al. 2015]. In [Xu and Cheng 2013], it is used to regress for hand joint angles directly. With a pre-skilled woodland, a set of vote casting pixels forged their votes for every joint angle, which might be fused into numerous candidate hand poses. An greater model-matching degree is wanted to discover the foremost pose. In [Tang et al. 2013] a transductive regression wooded area is proposed to alleviate the discrepancy among synthesis and actual-global facts to enhance prediction accuracy. In [Liang et al. 2015] a multi-modal prediction fusion set of rules is proposed to utilize hand motion constraints to remedy the ambiguous pose predictions from random regression woodland, in order that infeasible handpostures can be averted. In [Sun et al. 2015], a hierarchical regression scheme is constructed upon the regression forest for hand pose estimation, wherein the basis joints of hand skeleton are predicted first and other joints are expected subsequently based totally on the root joints, which proves to improve prediction accuracy in large part. While those techniques paintings simplest for in-air arms, we advocate to make use of the regression wooded area for hand in interplay with planar gadgets.

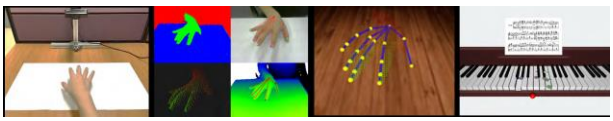
B. Finger Action Recognition

To extract discriminative capabilities and find effective getting to know fashions are the 2 key troubles in every pattern reputation hassle. Actions are spatio-temporal patterns, which requires complete features accumulating data from time domain as well as area domain to define the problem. It's usually recognized that absolute 3-d joints

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

positions are useful in detecting moves of human body. Multi-digicam movement capture (MoCap) structures[Campbell and Bobick 1995] has been broadly used to reap correct 3D joint function of human body. Similar techniques include facts glove (http://www.5dt.Com), which offers accurate tracking and haptic comments. Action recognition based totally on three-D joints positions retrieved from such devices has been properly-studied. There had been many exceptional temporal fashions in detecting human frame movements. Lv and Nevatia [2006] used Hidden Markov Model over pre-defined relative positions received from the 3-d joints. Han et al. [2010] used conditional random field over three-D joint positions. For movements with complicated articulated structure, motions of individual joints are sometimes correlated. Relative positions among joints may be extra discriminative features than absolutely the role of individual joint [Zhu et al. 2008]. However, these strategies tracks human frame motion involving many intermediate joints, and the motions are generally easily observant and has larger variations in among than finger motions. Besides, for tapping gesture, y coordinate, particularly transferring direction of tapping finger, embeds extra records in comparison to the closing guidelines. Yi et al. [2015] detected the falling edge of Y coordinate as a faucet, and changed the technique in [Palshikar et al. 2009] to come across the height value of modifications in Y coordinate. However, their strategies only consider the tapping action as one instantaneous motion instead of separate moves: up and down. Our technique makes use of the relative role of pair-smart fingertip and character joints motion trends as functions for tapping detection, and considers each up and down instructions.

C. *Virtual Musical Instruments*



There are many research efforts to develop virtual and augmented musical instruments in the past decades. Virtual reality and/or augmented reality techniques are utilized to increase instrument accessibility, improve user's psycho-pleasure and provide performance guidance [Rogers et al. 2014] [Dirkse 2009] [Chow et al. 2013] [Lin and Liu 2006]. Broersen and Nijholt [2002] developed a virtual piano, which allows multi-agents to play and is useful for educational purpose. However, it uses a real synthesizer or mouse/keyboard as input device, making it non-intuitive to play. Other applications involve instrument-like gestural controllers, such as video camera [Modler and Myatt 2008] [Yeh et al. 2010], motion capture [Nymoen et al. 2011], multi-touch device [Ren et al. 2012], data glove [Mitchell et al. 2012], and more recently depth sensors. Digito [Gillian and Paradiso 2012] is a gesturally controlled virtual musical instrument which utilizes 3D depth sensor to recognize hand gesture with machine learning algorithms and triggers the note to be played by using a "tap" style gesture with the tip of the index finger of the right hand. However, the user experience of Digito is too much different from real playing piano with different fingers. Some applications are developed using Leap Motion Controller to construct virtual piano using 3D positioning of fingers to detect the tapping [Heavers], but in these applications user's hands are not allowed to interact with any object, which is unnatural and uncomfortable for piano player. Han and Gold [2014] conducted a detailed examination on Leap Motion as the tracking device and algorithm for playing piano, which shows that although Leap Motion provide accurate tracking for free hand postures, when there is no interaction of hand with any object, it's difficult for player to determine the position and height of the virtual keyboard without prior practices. Our approach allows users to put their bare hand on a planar object and tap on it, like playing on a real piano

III. OVERVIEW

We develop a virtual piano application enabling fingertip tracking and tapping gesture detection, which can let users play a virtual piano keyboard on any plane as a force feedback. We especially design the application for starter-level piano players, who start playing with slow and simple practice songs. In such cases, the fingertip motions can be accurately tracked, and tapping can be identified relatively robustly based on hand joint trajectories only. Our application is developed with DepthSenser 325 as the RGB-D sensor, which consists of three components:

* Fingertip tracking takes RGB-D images as inputs, extracts the hand from the reference plane, and computes the positions of hand joints;

* Tapping Detection converts five fingertip locations into global coordinate system, computes the height relative to the reference plane and the relative positions of each
Pair-wise fingertip, and finally generates tapping event based on the spatial-temporal features retrieved from motion trajectory data.

* Rendering and Feedback takes the tapping event as input, triggers virtual piano key event and sound system, and finally provides a visual and sound feedback to the user.

IV. HAND SEGMENTATION

To make certain high exceptional of hand pose estimation, the hand area desires to be segmented correctly from the background in the depth photo. To locate the palms, we carry out in step with-pixel skin shade detection [Hammer and Beyerer 2013; Li and Kitani 2013]. However, the detected pores and skin masks is not always reliable and history pixels can be misclassified into the hand vicinity. To improve the outcomes, we advocate to first in shape a plane to the desk floor using the RANSAC algorithm [Fischler and Bolles 1981] in the depth photo after which differentiate the points that do not fit the plane as the hand region. However, as the hand occupies a massive part of the foreground intensity image, it introduces many outliers for 3-d aircraft fitting. This big range of outliers can have an effect on the RANSAC algorithm as it will need tons greater iterations to find the great set of factors that in shape the plane. To address this trouble, we find the hand vicinity inside the depth image with the pores and skin coloration detection outcomes, then use RANSAC to fit a plane with the final factors.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Based on the detected plane, the hand can be better segmented in intensity photos. In addition, we use the normal vector targeted on the desk as the beginning of the coordinate to assemble the digital piano for interplay.

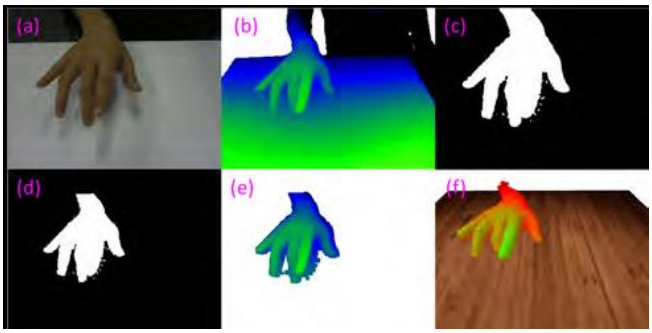


Figure: Hand segmentation. (a) and (b): input RGB-D images; (c) skin mask in RGB image; (d) skin mask mapped to depth image; (e) hand segmentation without 3D plane fitting; (f) fitted 3D plane and final hand segmentation.

V. HAND POSE TRACKING

Once the hand is segmented, we can then use the random regression forest [Girshick et al. 2011] to predict the 3D positions of the seven joints of the hand. The regression forest is an ensemble of several random regression trees, each of which consists of a number of split nodes and leaf nodes. Each split node contains one split function learnt from the training data to branch to the child node based on the feature values of the descriptor of an input pixel i. Each leaf node contains the distributions over the 3D relative offsets to the joint positions, which are collected from the training samples.

VI. EXPERIMENTS & DISCUSSIONS

Experiment Setup. We implemented the fingertip tracking and tapping detection algorithm in C++/OpenCV and rendered the virtual piano using OpenGL. We adopted a DepthSenser 325 sensor on top of the desk and in front of the user. The system was tested on a PC with an Intel i7 3.3GHz CPU and 16GB RAM. It is worth noting that the time cost to process one frame is only 20ms, which is efficient enough for real-time tracking. After training for 20 minutes, a user with little musical playground can play a simple adagio melody with our virtual piano. See the accompanying video.

Table 1: Precision and recall for individual tapping down class.

	Precision	Recall
Thumb	88.99%	97.00%
Index	100.00%	95.00%
Middle	96.77%	90.00%
Ring	85.34%	99.00%
Pinky	100.00%	87.00%

Training. To validate the effectiveness of the proposed hand pose tracking algorithm, we collect a dataset of real-world hand images consisting of around 7.2k depth images of two subjects performing various finger tapping postures to play the virtual piano. The resolution of these images is 320 X 240. The subjects can either put their hand over or on the desk. The hand poses collected cover the most frequent gestures for playing piano in the view of depth camera, and the poses are music score independent. In each of the image we manually annotate the 3D positions of the seven joints of
the hand. In this experiment we set the number of trees in the forest to be 3. During training, we randomly sample 150 pixels from each training image and generated 6000 candidate split functions to learn the tree structure. The tree stops growing if its depth exceeds 20 or the node sample is less than 50. During testing, a number of 500 voting pixels are randomly sampled from the segmented hand region to predict the hand joint positions.

To collect the training data set for tapping, we manually label several sequences of RGB images, including over 100 taps for each finger. We label the tapping down moment frame and its previous 3 frames as TD frames, and label the tapping up moment frame and its following 3 frames as TU frames. The other frames are labeled as non-tapping frames. These annotated data are then used to train the SVM classifier for tap detection.

Performance. We perform 4-fold cross validation on this dataset to evaluate the performance of the proposed method. The prediction performance of a joint is evaluated in terms of the percentage of its predictions that are within a distance of DT centimeters from the ground truth in the test images. This metric is averaged for all the seven joints to obtain the overall evaluation. To better understand the performance of the method, we present the results for different DT so that the distribution of the predictions over different intervals of DT can be observed, as shown in Figure 7. The average error between the ground truth hand joint positions and the predicted positions is 1:3cm. Figure 8 shows the hand pose prediction results on some sample frames in the dataset. We can see that the proposed method can accurately recover the positions of the hand joints, when fingers are in the air and on the reference plane. In contrast, the commercial products, such as Leap Motion, Intel RealSense and SoftKinetic, are not able to detect the hand joints for those cases. To test tapping detection algorithm, we ask 2 users to perform 100 taps totally on each finger with around 1 second time difference in between. We consider a tapping down and tapping up event classified successfully if the finger which performed the action is correctly identified within 0.3 second. The result of tapping down detection is shown in Table 1.

Comparison. We compare our method with two state-of-the-art techniques, a model-based algorithm [Tagliasacchi et al. 2015] and Leap Motion Controller – the leading commercial product for hand tracking. These methods are able to accurately track (multiple) hands when they are in the air, however, they fail when hands are interacting with physical objects. In contrast, our algorithm is specifically

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

designed for hand interaction with planar objects, hereby has better performance and accuracy in the virtual piano application.

Limitations. Although our method can track most of the beginner's piano-playing gestures for soothing rhythms in realtime, our virtual piano has several limitations compared with playing real piano. (1) The proposed tracking algorithm is not quite robust to hand-shape variations, e.g., the prediction accuracy drops when the shape and/or size of player's hand are significantly different from the ones in the training dataset. (2) Thumb under is a common gesture, where the thumb is brought under the hand in order to pass
the 3rd or 4th finger for playing the scale. Due to severe occlusion, the depth sensor is not able to capture the thumb and our tracking algorithm cannot detect it either. (3) Our current implementation is not efficient and accurate enough to detect the tapping event in a fast tempo. (4) Our method supports two-hand tracking. However, due to the limited viewing volume of the DepthSenser325 sensor, users can only play with a single hand for about 2 octaves.

VII. CONCLUSION & FUTURE WORK

This paper provided a virtual piano application that permits customers to play with naked palms on or near a planar surface. Taking the RGB-D pics as input, our approach makes use of an offline trained random regression forest to music the fingertips and locate the finger tapping. Compared with the present hand tracking algorithms, our method is designed for hand interplay with planar gadgets. Preliminary consequences display that our method can apprehend most of the amateur's piano-gambling gestures for soothing rhythms in real-time.

The machine may be similarly incorporated with head-established display, which includes Oculus Rift, to provide with consumer an immersive visible and audial environment which may also in addition support remote gaining knowledge of and gamification in musical tool gaining knowledge of. In a broader experience, our work affords a pipeline to solve hand integration with planar objects and a general solution to such form of application, which draws the community's attention to the hassle of cutting-edge mid-air hand tracking strategies.

In the future, we will extend the gesture database for intermediate and advanced gamers and enhance the accuracy of our monitoring set of rules for allegro rhythms. To locate self-occluded gestures, a few graphical machine studying version might be implemented to expect occluded finger role and tapping moment with the help of domain information. We may also increase a hand normalization set of rules so that players whose palms are notably specific from the ones of the education dataset can use our gadget. Moreover, we will behavior a formal person have a look at to assess the efficacy of the proposed machine.

ACKNOWLEDGMENT

This assignment turned into in part funded with the aid of Singapore MOE2013-T2-2-011, MOE RG40/12, MOE RG23/15, the Economic Development Board and the National Research Foundation of Singapore, the NSFC Grants (61322206, 61521002) and the Foundation of TNList.

REFERENCES

[1] ARISTIDOU, A., AND LASENBY, J. 2010. Motion capture with constrained inverse kinematics for real-time hand tracking. In International Symposium on Communications, Control and Signal Processing, IEEE, 1–5.

[2] BROERSEN, A., AND NIJHOLT, A. 2002. Developing a virtual piano playing environment. In IEEE International conference on Advanced Learning Technologies (ICALT 2002), 278–282.

[3] CAMPBELL, L. W., AND BOBICK, A. E. 1995. Recognition of human body motion using phase space constraints. In Computer Vision, 1995. Proceedings., Fifth International Conference on, IEEE, 624–630.

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

[5] Music education using augmented reality with a head mounted display. In Proceedings of the Fourteenth Australasian User In- terface Conference-Volume 139, Australian Computer Society, Inc., 73–79.

[6] COMANICIU, D., AND MEER, P. 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. PAMI 24, 5, 603–619.

[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] FISCHLER, M. A., AND BOLLES, R. C. 1981. Random sample

[9] consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24, 6, 381–395.

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

[11] 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.

[12] 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 International Conference on New Interfaces for Musical Expression (NIME), Cite seer, 5–7. 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] Full do tracking of a hand interacting with an object by modeling occlusions and physical constraints. In IEEE International Conference on Computer Vision, IEEE, 2088–2095.

[14] PALSHIKAR, G., ET AL. 2009. Simple algorithms for peak detection in time-series. In Proc. 1st Int. Conf. Advanced Data Analysis, Business Analytics and Intelligence.

[15] REN, Z., MEHRA, R., COPOSKY, J., AND LIN, M. 2012. Designing virtual instruments with touch-enabled interface. In CHI'12 Extended Abstracts on Human Factors in Computing Systems, ACM, 433–436.

[16] ROGERS, K., RO¨ HLIG, A., WEING, M., GUGENHEIMER, J., KO¨ 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, 149–158.

[17] HAMMER, J. H., AND BEYERER, J. 2013. Robust hand tracking in real-time using a single head-mounted rgb camera. In Inter- national Conference on Human-Computer Interaction, Springer, 252–261.

[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, 371–374.

[19] HAN, L., WU, X., LIANG, W., HOU, G., AND JIA, Y. 2010. Discriminative human action recognition in the learned hierarchical manifold space. Image and Vision Computing 28, 5, 836–849.

[20] HEAVERS, M. Vimeo video: Leap motion air piano. https://vimeo. com/67143314.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

[21] LI, C., AND KITANI, K. M. 2013. Pixel-level hand detection in ego-centric videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 3570–3577.

[22] LIANG, H., YUAN, J., AND THALMANN, D. 2015. Resolving ambiguous hand pose predictions by exploiting part correlations. IEEE Trans. Circuits and Systems for Video Technology 25, 7, 1125–1139.

[23] 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.

[24] LV, F., AND NEVATIA, R. 2006. Recognition and segmentation of 3-d human action using hmm and multi-class ad boost. In Computer Vision–ECCV 2006. Springer, 359–372.

[25] MELAX, S., KESELMAN, L., AND ORSTEN, S. 2013. Dynamics based 3d skeletal hand tracking. In Proceedings of Graphics Interface 2013, 63–70.

[26] M. M., AND RAMANAN, D. 2014. 3d hand pose detection in egocentric rgb-d images. In ECCV Workshop on Consumer Depth Cameras for Computer Vision, Springer, 356–371.

[27] SUN, X., WEI, Y., LIANG, S., TANG, X., AND SUN, J. 2015. Cascaded hand pose regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 824–832.

[28] TAGLIASACCHI, A., SCHRO¨ DER, M., TKACH, A., BOUAZIZ, S., BOTSCH, M., AND PAULY, M. 2015. Robust articulated-icp for real-time hand tracking. Computer Graphics Forum 34, 5, 101–114.TANG, D., YU, T.-H., AND KIM, T.-K. 2013. Real-time articulated hand pose estimation using semi-supervised transductive regression forests. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 3224–3231.

[29] XU, C., AND CHENG, L. 2013. Efficient hand pose estimation from a single depth image. In IEEE International Conference on Computer Vision, IEEE, 3456–3462.

[30] YEH, C.-H., TSENG, W.-Y., BAI, J.-C., YEH, R.-N., WANG,S.-C., AND SUNG, P.-Y. 2010. Virtual piano design via single- view video based on MultiFinder actions recognition. In 2010 3rd International Conference on Human-Centric Computing, 1–5. YI, X., YU, C., ZHANG, M., GAO, S., SUN, K., AND SHI, Y. 2015. Ask: Enabling ten-finger freehand typing in air based on 3d hand tracking data. In Annual ACM Symposium on User In- terrace Software and Technology.

[31] ZHU, L. L., CHEN, Y., LU, Y., LIN, C., AND YUILLE, A. 2008.Max margin and/or graph learning for parsing the human body. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 1–8.