The Literature Survey on Virtual Piano

Yashwanth G†, Saifulla Khan††, T V Rahul Reddy†††, Sukanya H A Varsha N†††††

†UG Students, Department of Information Science And Technology Vidya Vikas Institute of Engineering & Technolgy Karnataka, India.

Summary
This paper presents an efficient data-driven approach to trace fingertip and detect finger tapping for virtual piano using an RGB-D camera. We collect 7200 depth images covering the foremost common finger articulation for enjoying piano, and train a random regression forest using depth context features of randomly sampled pixels in training images. within the online tracking stage, we firstly segment the hand from the plane in touch by fusing the knowledge from both colour 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 supported the estimated fingertip motion. Finally, we build a kinematic chain and recover the articulation parameters for every finger. In contrast to the prevailing hand tracking algorithms that always require hands are within the air and can't interact with physical objects, our method is meant 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 then tap fingers on the desk, like playing a true piano. Preliminary results show that our method can recognize most of the beginner's piano-playing gestures in real-time for soothing rhythms.
Keywords:
Fingerend Tracking, Finger Beating Detection, Virtual Piano, RGB-D picture, Computer- Human Interaction.

1. Introduction

Recent years have witnessed rapid progress of hand pose tracking and hand motion analysis using consumer depth sensors. State-of the-art techniques [Tagliasacchi et al. 2015][Sun et al. 2015] are ready 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 within the air and cannot interact with physical objects. Such a requirement diminishes their utility for virtual instrument applications thanks to two reasons: First, users can quickly get tired when hands aren't supported by some object . Second, mid-air interactions don't 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 place their hands on a desk, move them sideways then tap fingers on the desk, like playing a true piano. There are two major technical challenges during this application. First, the system must track the positions of fingertips and detect their status, i.e., whether a finger is tapping or not. thanks to frequent interaction between fingers and desk, the prevailing hand tracking algorithms often fail. Second, piano-playing gestures are usually fast and sophisticated , involving highly flexible hand articulation and causing severe hand self-occlusion.

To tackle these challenges, we propose a virtual-piano tailored method to trace 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 foremost common finger articulation for enjoying 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 random forest. Then we use the trajectories of those 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.

Supported the locations of fingertips and therefore the finger tapping status, the system can hereby determine which piano key's pressed and play the corresponding sound. Preliminary results show that our method can recognize the essential 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 ahead of the user. We render the hand and therefore the refore the piano supported the coordinates of the desk surface and the detected hand pose from the RGB-D images.

2. Related work

Hand pose tracking and evaluation may be a fundamental hassle in laptop portraits and vision, and is central for several 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

Manuscript received April 5, 2025
Manuscript revised April 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.4.6

employing 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 doesn't obstruct consumer movements, hereby is specifically effective to VR applications. This introduce short evaluations related work accessible pose monitoring, finger motion popularity and digital instrument.

A. Hand Pose Tracking

Algorithms for imaginative and prescient-primarily based hand pose monitoring are often widely labeled as generative version-fitting methods and discriminative strategies. Each magnificence of algorithms have their very own merits and disadvantages . The version-becoming strategies [Melax et al. 2013][Tagliasacchi et al. 2015] reconstruct hand poses with the help of becoming a 3-d articulated hand model to intensity photographs. These methods work properly in controlled environments, but, they typically require calibration and their outcomes are sensitive to initialization. The discriminative techniques require an annotated dataset to look at a regressor offline, after which use it to expect the hand pose on-line. Such techniques are strong to initialization, however their accuracy closely depends on the dimensions of the training dataset. Therefore, the dataset must be fairly massive to hide the viable hand and finger articulations for a specific 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 not interacting with other items. the aim is that, hand motion itself is of excessive ranges-of-freedom and as a consequence involves many training information to symbolize such flexibility. Thus, if the hand is interacting with unknown gadgets, there might be greater unpredictable complexity, e.G., hand self-occlusion and occlusion between item and hand, and therefore the big look versions of every the hand and interacting objects.

These problems avoid researches during this vicinity. There are some preceding paintings which will help hand interacting with gadgets, however they either assume that the geometrical information of the item is understood in order that item and hand can supplement every other to reinforce pose estimation [Oikonomidis et al. 2011] through the physical constraints among them or confine the possible hand articulations to be inside alittle set of templates [Rogez et al. 2014]. In [Oikonomidis et al. 2011] the sort and specific length of the interacting item are assumed to be known earlier. The poses of the thing and hand are jointly solved during a generative version-fitting framework the utilization of a multi-digicam placing to reduce prediction ambiguity, which maximizes the fashions' compatibility to the image inputs and minimizes the intersection among hand and objects to seek out their pose.

In [Rogez et al. 2014] the hand is permitted to possess interaction with different items, like bottles, desk surfaces, etc., and therefore the hand pose is predicted during 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 alittle set of pre-described templates. this is often faraway from our want to play the piano within the proposed software, wherein we would like to music the accurate articulated fingertip positions and wrist positions, in order that the device can encounter whether or not a finger tap is finished with the help of the performer or not.

Among the discriminative strategies, random regression woodland and its versions have validated powerful to capture hand pose thorough photos [Xu and Cheng 2013; Tang et al. 2013; Liang et al. 2015; Sun et al. 2015]. In [Xu and Cheng 2013], it's wont to regress for hand joint angles directly. With a pre-skilled woodland, a group of voting pixels forged their votes for each joint angle, which could be fused into numerous candidate hand poses. An greater model-matching degree is needed to get the foremost pose. In [Tang et al. 2013] a transductive regression forest is proposed to alleviate the discrepancy among synthesis and actual-global facts to reinforce 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, so as that infeasible handpostures are often averted. In [Sun et al. 2015], a hierarchical regression scheme is made upon the regression forest for hand pose estimation, wherein the idea joints of hand skeleton are predicted first and other joints are expected subsequently based totally on the basis joints, which proves to enhance prediction accuracy in large part. While those techniques paintings simplest for in-air arms, we advocate to form use of the regression forest for hand in interplay with planar gadgets.

B. Finger Action Recognition

To extract discriminative capabilities and find effective going to know fashions are the two key troubles in every pattern reputation hassle. Actions are spatio-temporal patterns, which needs complete features accumulating data from time domain also as area domain to define the matter . It's usually recognized that absolute 3-d joints positions are useful in detecting moves of physical body . Multi-digicam movement capture (MoCap) structures[Campbell and Bobick 1995] has been broadly wont to reap correct 3D joint function of physical 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 could also 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 therefore 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 as compared to the closing guidelines. Yi et al. [2015] detected the falling fringe of Y coordinate as a faucet, and altered the technique in [Palshikar et al. 2009] to return across the peak value of modifications in Y coordinate. However, their strategies only consider the tapping action together instantaneous motion rather than 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 Instuments

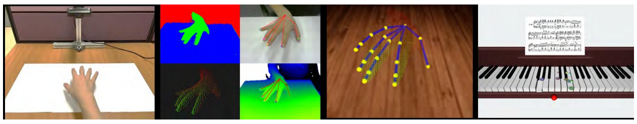

Figure 1: Virtual musical instruments

As shown in Figure 1, There are many research efforts to develop virtual and augmented musical instruments within the past decades. computer game and/or augmented reality techniques are utilized to extend instrument accessibility, improve user's psycho-pleasure and supply 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 beneficial for educational purpose. However, it uses a true synthesizer or mouse/keyboard as data input device , making it non-intuitive to play. Other applications involve instrument-like gestural controllers, like 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] may be a gesturally controlled virtual instrument which utilizes 3D depth sensor to acknowledge hand gesture with machine learning algorithms and triggers the note to be played by employing a "tap" style gesture with the tip of the index of the proper hand. However, the user experience of Digito is just 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 aren't allowed to interact with any object, which is unnatural and uncomfortable for pianist . Han and Gold [2014] conducted an in depth examination on Leap Motion because the tracking device and algorithm for enjoying piano, which shows that although Leap Motion provide accurate tracking for free of charge hand postures, when there's no interaction of hand with any object, it's difficult for player to work out the position and height of the virtual keyboard without prior practices. Our approach allows users to place their bare hand on a planar object and tap thereon , like playing on a true piano.

3. Overview

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

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

* Tapping Detection converts five fingertip locations into global frame of reference , computes the peak relative to the reference plane and therefore the relative positions of every Pair-wise fingertip, and eventually generates tapping event supported the spatial-temporal features retrieved from motion trajectory data.

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

4. Hand Segmentation

To form certain high exceptional of hand pose estimation, the hand area desires to be segmented correctly from the background within the depth photo. To locate the palms, we feature call at step with-pixel skin shade detection [Hammer and Beyerer 2013; Li and Kitani 2013]. However, the detected pores and skin masks isn't always reliable and history pixels are often misclassified into the hand vicinity. to enhance the outcomes, we advocate to first in shape a plane to the desk floor using the RANSAC algorithm [Fischler and Bolles 1981] within the depth photo after which differentiate the points that don't fit the plane because the hand region. However, because the hand occupies a huge a part of the foreground intensity image, it introduces many outliers for 3-d aircraft fitting. This big selection of outliers can have an impact on the RANSAC algorithm because it will need tons greater iterations to seek out the good set of things that in shape the plane. to deal with this trouble, we discover the hand vicinity inside the depth image with the pores and skin coloration detection outcomes, then use RANSAC to suit a plane with the ultimate factors.

Supported the detected plane, the hand is often better segmented in intensity photos. additionally, we use the traditional vector targeted on the desk because the beginning of the coordinate to assemble the digital piano for interplay in figure 2.

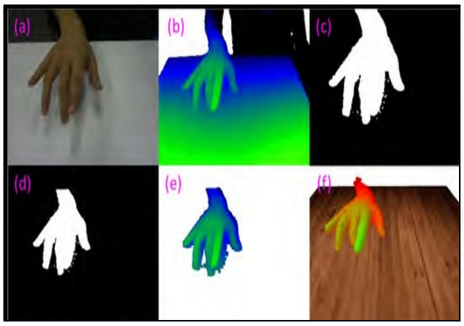

Figure 2: Segmentation of Hand. (a) and (b): RGB-D images input; (c) skin masking in RGB image; (d) skin masking mapped to depth of image; (e) segmentation of hand without 3D plane fitting; (f) fitted 3D plane and final segmentation of hand.

.

5. Hand Pose tracking

Once the hand is segmented, we will 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 variety of split nodes and leaf nodes. Each split node contains one split function learnt from the training data to branch to the kid node supported 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

6. 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 Depth Senser 325 sensor on top of the desk and ahead of the user. The system was tested on a PC with an Intel i7 3.3GHz CPU and 16GB RAM. it's worth noting that the time cost to process one frame is merely 20ms, which is efficient enough for real-time tracking. After training for 20 minutes, a user with little musical playground can play an easy adagio melody with our virtual piano. See the accompanying video.

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

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

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 those images is 320 X 240. the themes can either put their fork over or on the desk. The hand poses collected cover the foremost frequent gestures for enjoying piano within the view of depth camera, and therefore the poses are music score independent. In each of the image we manually annotate the 3D positions of the seven joints of the hand. during this experiment we set the amount of trees within the forest to be 3. During training, we randomly sample 150 pixels from each training image and generated 6000 candidate split functions to find out the tree structure. The tree stops growing if its depth exceeds 20 or the node sample is a smaller amount than 50. During testing, variety of 500 voting pixels are randomly sampled from the segmented hand region to predict the hand joint positions.

To gather the training data set for tapping, we manually label several sequences of RGB images, including over 100 taps for every 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 opposite frames are labeled as non-tapping frames. These annotated data are then wont to train the SVM classifier for tap detection.

Performance. We perform 4-fold cross validation on this dataset to guage the performance of the proposed method. The prediction performance of a joint is evaluated in terms of the share of its predictions that are within a distance of DT centimeters from the bottom truth within the test images. This metric is averaged for all the seven joints to get the general evaluation. to raised understand the performance of the tactic , we present the results for various DT in order that the distribution of the predictions over different intervals of DT are often observed, as shown in Figure 7. the typical error between the bottom truth hand joint positions and therefore the predicted positions is 1:3cm. Figure 8 shows the hand pose prediction results on some sample frames within the dataset. we will see that the proposed method can accurately recover the positions of the hand joints, when fingers are within the air and on the reference plane. In contrast, the commercial products, like Leap Motion, Intel RealSense and SoftKinetic, aren't ready to detect the hand joints for those cases. to check 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 results 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 ready to accurately track (multiple) hands once they are within the air, however, they fail when hands are interacting with physical objects. In contrast, our algorithm is specifically
designed for hand interaction with planar objects, hereby has better performance and accuracy within 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 isn't quite robust to hand-shape variations, e.g., the prediction accuracy drops when the form and/or size of player's hand are significantly different from those within the training dataset. (2) Thumb under may be a common gesture, where the thumb is brought under the hand so as to pass
the 3rd or 4th finger for enjoying the size . thanks to severe occlusion, the depth sensor isn't ready to capture the thumb and our tracking algorithm cannot detect it either. (3) Our current implementation isn't efficient and accurate enough to detect the tapping event during a fast tempo. (4) Our method supports two-hand tracking. However, thanks to the limited viewing volume of the DepthSenser325 sensor, users can only play with one hand for about 2 octaves.

7. Conclusion & Future Work

This paper provided a virtual piano application that allows 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 this hand tracking algorithms, our method is meant 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 could also be similarly incorporated with head-established display, which incorporates Oculus Rift, to supply with consumer an immersive visible and audial environment which can also additionally support remote gaining knowledge of and gamification in musical tool gaining knowledge of. during a broader experience, our work affords a pipeline to unravel hand integration with planar objects and a general solution to such sort of application, which pulls the community's attention to the effort of cutting-edge mid-air hand tracking strategies.
within the future, we'll 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 couple of graphical machine studying version could be implemented to expect occluded finger role and tapping moment with the assistance of domain information. We can also increase a hand normalization set of rules in order that players whose palms are notably specific from those of the education dataset can use our gadget. Moreover, we'll behavior a proper person have a glance at to assess the efficacy of the proposed machine.

Acknowledgment

This assignment became partially funded with the help of Singapore MOE2013-T2-2-011, MOE RG40/12, MOE RG23/15, the Economic Development Board and therefore the National Research Foundation of Singapore, the NSFC Grants (61322206, 61521002) and therefore 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 Interface 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 International 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.

[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.