TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

A Low-Jitter Hand Tracking System for Improving Typing Efficiency in Virtual Reality Workspace

Tianshu Xu
Division of Advanced Science and Technology,
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
xutianshu@jaist.ac.jp

Wen Gu
Center for Innovative Distance Education and Research,
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
wgu@jaist.ac.jp

Koichi Ota
Center for Innovative Distance Education and Research,
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
ota@jaist.ac.jp

Shinobu Hasegawa
Center for Innovative Distance Education and Research,
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
hasegawa@jaist.ac.jp

***Abstract*— Virtual reality technology has the potential to revolutionize immersive experiences in various applications, including office settings. However, efficient text entry in VR remains a significant challenge. This study addresses this challenge by proposing a machine learning-based solution, the 2S-LSTM typing method, to enhance text entry performance in VR. The 2S-LSTM leverages the back of the hand image. It employs a two-stream Long Short-Term Memory (LSTM) network, combined with a Kalman Filter (KF), to improve hand position tracking accuracy and reduce jitter. The results from questionnaire-based evaluations and typing data analysis demonstrate the superiority of the 2S-LSTM solution over existing solutions like Oculus Quest 2 and Leap Motion in terms of typing efficiency, fatigue reduction, accurate hand position replication, and positive user experience. These findings contribute to the advancement of text entry in VR environments and pave the way for immersive work experiences in the office and beyond.**

Keywords—virtual reality, typing efficiency, immersive work experience.

I. INTRODUCTION

Virtual Reality (VR) holds immense potential for transforming immersive experiences into practical tools for various applications, including office settings. It allows users to interact with and visualize data without the limitations of physical screens, creating a more immersive and engaging workspace. However, one significant challenge that hinders the realization of this vision is the lack of robust text entry capabilities in VR [1-3]. When users wear a Head Mounted Display (HMD), they face difficulties in entering long text as they are unable to see their hands and the keyboard. Existing solutions, such as wearable devices or specialized controllers, have been developed to address this issue but often reduce text input efficiency [1-4]. Moreover, another hurdle to the efficient text input in VR is the presence of jitter, which refers to image rendering problems that cause virtual hands to shake, leading to inconsistencies between hand actions and virtual responses [5]. Therefore, text entry in VR remains a significant challenge.

The objective of this study is to tackle the problem of text entry in VR, particularly in the context of immersive office experiences. Existing literature categorizes text entry support in VR into two main areas: traditional hardware solutions and machine learning approaches. Traditional hardware solutions, such as wearable devices or specialized controllers, while providing some level of text input support, often suffer from inconvenience and additional costs [1-4]. On the other hand, the machine learning approach has gained attention in recent years and shows promise for VR text entry. However, they may not match the performance of traditional hardware solutions [6-8].

With this in mind, our research focuses on the machine learning approach to address the limitations and additional costs associated with traditional hardware solutions. Existing machine learning solutions primarily track users' actions [6-8]. However, typing behavior in VR presents unique challenges, such as the blocking of typing fingers by the palm, which makes it difficult to obtain a complete palm contour and track hand positions accurately. To overcome these challenges, our proposed research investigates the utilization of the back of the hand image. By leveraging information from the back of the hand image, even when the fingers are blocked, we aim to maintain precise hand position tracking. To achieve this, we propose a method that divides the back of the hand image into hand images and motion images. We introduce a two-stream Long Short-Term Memory (LSTM) network to process these images separately, and we apply a Kalman Filter (KF) to reduce jitter and enhance the accuracy of hand position tracking.

II. RELATED WORK

A. Typing in VR

HMD coupled with a keyboard is a basis for a full Virtual Reality Workspace (VRWS) in which users can enjoy a motion-independent robust and immersive virtual office environment[1]. However, one barrier is no robust text entry. Entering long text will become difficult because user wearing the HMD cannot see their hands and keyboard. In order to see the hands and keyboard while typing in VR, hands and keyboards should be recognized and shown in virtual reality. Microsoft focuses specifically on investigating the method for virtually representing a user's hands in VR in several different ways[4].

Heretofore, many text input support solutions are considered to maintain user's typing efficiency in VRWS. At present, there are mainly two approaches to solve the typing inefficiency problems: the traditional hardware area and the machine learning area. Among all other solutions, the machine learning solution is considered as the most potential support solution [6-8]. Hwang et al. propose a method to estimate 3D human pose from a monocular fisheye camera mounted on a VR headset [6]. Erwin et al. introduce a system to recognize 3D hand poses from a wrist-worn camera via a deep neural network [7]. Jang et al. present a metaphoric gesture interface for manipulating virtual objects with an egocentric viewpoint [8].

B. Hand Tracking

Hand tracking is a technology that enables the detection and tracking of the position, depth, speed, and orientation of a user's hands using various methods such as headset cameras [9], LiDAR arrays [10], or external sensor stations [11]. This tracking data is analyzed and processed to create a virtual, real-time representation of the user's hands and their movements within the virtual world. This representation is subsequently transmitted to the respective application or video game being used, allowing users to interact naturally with the virtual environment using their hands.

Unfortunately, LiDAR arrays, or external sensor stations, this kind of wearable hand tracking solutions often hinder typing efficiency due to the requirement of wearing extra devices. Deep learning solutions offer cost advantages as they only require the cameras embedded in the HMD, eliminating the need for additional hardware [6-8]. This also means that the Deep learning solution has less impact on typing efficiency because it does not need to wear extra devices.

However, typing as a task presents unique challenges. When users wear an HMD and type in a VR environment, the fingers are often obstructed by the back of the hand. It makes the HMD's cameras difficult to capture the complete view of the typing hands. As a result, the accurate tracking of typing hands positions becomes challenging. A study has been conducted to estimate finger positions during typing by utilizing subtle variations on the back of the hand, using a wrist-mounted camera [7]. Inspired by their work, our approach also focuses on visual features on the back of the hand, extending it to support richer, full typing hands position estimation. Our approach builds upon the insights from their research, focusing on the visual features on the back of the hand, and extending it into a robust and practical VR typing support system.

C. Jitter

In VR systems, jitter refers to small fluctuations in the signal and is a significant factor that can adversely affect motor performance and user experience. Despite continuous technological advancements, effectively reducing or eliminating jitter remains a challenge, especially in tracking systems that are integrated into various HMDs. The impact of jitter on VR systems has been extensively studied by various researchers. Teather et al. [12] conducted an analysis and found that even a small amount of spatial jitter (0.3 mm) in the input device could noticeably decrease user performance. Moreover, it has been observed that larger jitter levels have a more pronounced negative effect on user performance, especially when dealing with smaller targets [13]. Batmaz et al. [14] also support this finding, noting that user performance declines significantly in terms of time, error rate, and throughput as the jitter level increases. Additionally, Moaaz et al. [15] conducted experiments where they artificially introduced 0.5°, 1°, and 1.5° jitter to the VR system, leading to a substantial increase in the user's error rate with each the increment in jitter level.

III. Methodology and Evaluation

In order to achieve our goal of developing a powerful and low-jitter VR assisted typing system, we carried out the following steps:

A. Data Collection

We required a dataset of "obscured typing hands" to train the network model since the existing hand databases mainly consisted of complete hand images. Due to the scarcity of such data, we conducted our own data collection process.

A total of 11 students from JAIST participated in the data collection phase, including 4 females, aged between 25 and 31. The participants were instructed to use a wearable camera while typing on a computer. The camera device, a 4K high-definition camera worn on the ear, was used to capture images of the "obscured typing hands, as shown in Fig. 1. Each participant engaged in a one-hour typing session, resulting in a total of 21,900 images collected.

Subsequently, following the steps outlined in related research [7], we employed OpenCV to apply image processing techniques for data augmentation. Specifically, we adjusted the hand color and brightness of these images to create variations. As a result, we obtained a dataset consisting of 438,000 images, approximately 20 times larger than the original dataset. To ensure unbiased evaluation, we randomly divided the dataset into training and testing sets, with 80% of the images allocated for training and the remaining 20% for testing purposes.

B. Network Architecture

The characteristics that affect the typing efficiency in VRWS are investigated. In this step, the unique characteristics of typing behavior, such as finger blocked by palm or typing finger small, will be considered thoroughly. By using or overcoming these characteristics, expected to get better tracking results. The following steps are conducted to develop a low-jitter hand tracking system to improve typing efficiency in VRWS.

1) In this research, the training sequence of length τ is 10. For each τ, we use the hand position labels $y_{1:\tau}$ and two input streams: original image $I_{1:\tau}$ and motion history image (MHI)

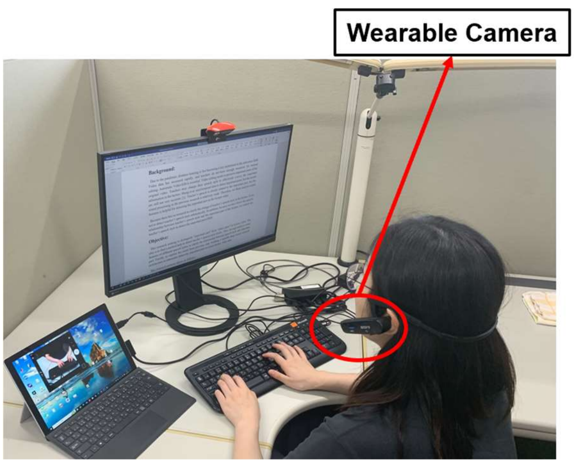

Fig. 1. In this scene, participants are wearing cameras while typing.

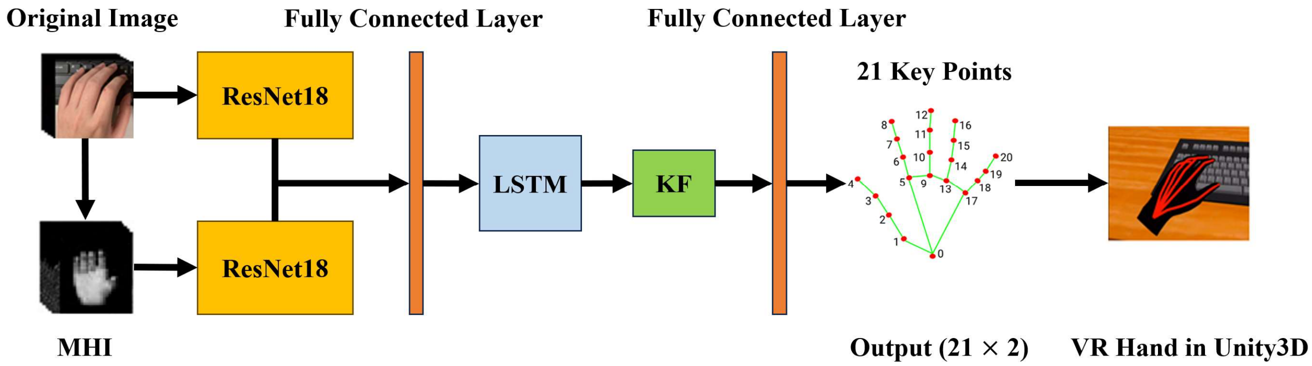

Fig. 2. 2S-LSTM network overview.

$X_{1:\tau}$. We referred the design of BlazePalm [16], each hand position label includes 42 key points (21 key points in one hand). The key points show in Fig. 2.

2) Original image and MHI are separately processed through a ResNet18 network to extract visual features. Subsequently, a fully connected layer is used to combine two visual features into a unified visual feature ϕ. Following this, the visual feature sequence $\phi_{1:\tau}$ is fed into an LSTM layer to extract temporal feature sequence $\psi_{1:\tau}$.

3) To reduce jitter and improve the quality of the output, a Kalman Filter (KF) is applied [7-8]. The KF serves to stabilize the sequence of features extracted by the network, enhancing the accuracy and robustness of hand position estimation, especially in the presence of occlusions and complex backgrounds. Then, the output is passed through another fully connected layer. This step serves to map the temporal feature to the estimated position of the typing hands$\tilde{y}_{1:\tau}$.

4) To visualize $\tilde{y}_{1:\tau}$, we implemented a hand simulator using Unity3D. This simulator can map $\tilde{y}$ to a both hand model consisting of 42 key points. By associating these key points with $\tilde{y}$, we are able to dynamically reproduce and simulate the movements and positions of typing hands in real time.

To this end, we propose a 2-stream LSTM-KF network, and the architecture, as shown in Fig. 2.

C. Ablation Study

Starting from the 2S-LSTM network to VGG16 and ResNet18, eliminate each of the proposed structural elements from the proposed architecture one by one by comparing the two-stream LSTM-KF network to the ordinary LSTM network. The architectures are: 2S ResNet18+LSTM+KF (RGB, MHI), 2S ResNet18+LST (RGB, MHI), ResNet18 + LSTM (RGB), ResNet18 (RGB), VGG16 (RGB). 2S stands for 2-stream, and "Ours" means 2S ResNet18+LSTM+KF (RGB, MHI). In this research, we used our dataset to do this test. We calculated the "accuracy of hand positions" for each model. For the specific calculation of the "accuracy of hand positions," considering the size of the letter key on the keyboard is approximately 1.5cm × 1.5cm, we assume that it is easy to make typing errors when the distance between the fingertip and the center of the target key on the keyboard exceeds half the length of a key (0.75cm). Therefore, we set the threshold to 0.75cm. When the distance between the key point and the ground truth is less than the threshold, that key point is valid. The "accuracy of hand positions" is calculated as the total number of valid key points divided by the total number of key points, as shown in TABLE I. Based on the results, we can conclude that the 2S-LSTM-KF performs better.

TABLE I. RESULTS OF ABLATION STUDY

Architecture	Accuracy of hand positions
VGG16 (RGB)	0.46
ResNet18 (RGB)	0.48
ResNet18 + LSTM (RGB)	0.53
2-steam ResNet18 + LSTM (RGB, MHI)	0.79
Ours	**0.86**

IV. EXPERIMENT

By conducting a comparative experiment between the developed assistance solution (2S-LSTM) and two existing solutions (oculus quest 2, leap motion), it aims to validate the effectiveness of the proposed method in improving typing efficiency. This experiment has obtained approval from the JAIST Life Sciences Committee.

A. Participant

We recruited 24 participants but 23 were right-handed and 1 left-handed (16 males and 8 females, average age M=26, standard deviation SD=4) with normal or corrected-to-normal vision. 7 participants had prior VR experience. We balanced the 6 participant groups by gender and experimental order. All participants had a certain level of English proficiency and could not have enough touch-typing skills.

B. Equipment

We experimented on a desktop PC with an NVIDIA GeForce GTX 1080 Ti graphics card. We used an HTC VIVE Pro Eye headset to apply the 2S-LSTM network. We used Oculus Quest 2 and Leap Motion as the baseline. The VR environment and other VR models used in the experiment were created by Unity3D. Several USB cameras were used to record experimental data from the participants.

C. Equipmental Conditions

1) Regular Typing (Normal): Participants first performed typing tasks without wearing the HMD for 30mins. This condition served as a baseline to measure participants' regular typing ability.

2) HMD Typing: Participants wore the HMD and performed typing tasks using three different typing assistance

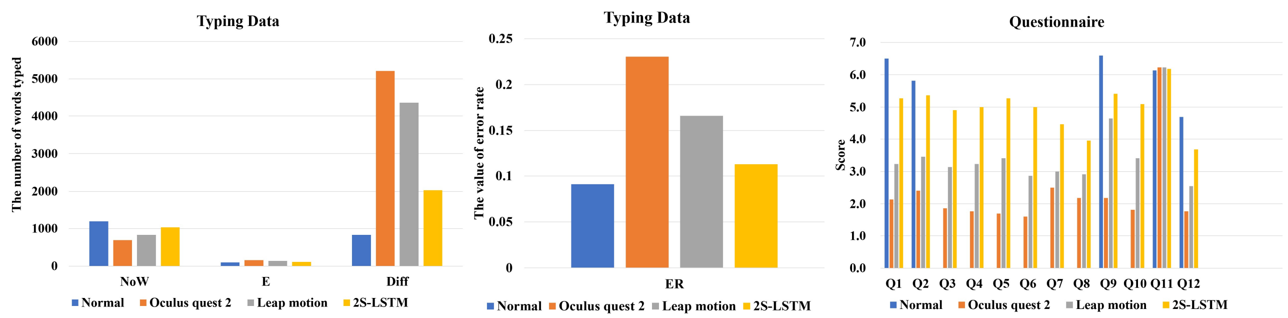

Fig. 3. Average Results for Each Condition.

solutions: Oculus Quest 2, Leap Motion, and the developed 2S-LSTM solution. Participants performed each task for 30 mins. The order of the solutions was counterbalanced among participants to minimize any order effects.

D. Experiment Procedure

1) Pre-Experiment Session: Participants were provided a brief training session to familiarize themselves with the HMD and the typing assistance solutions. This session ensured that participants understood the task requirements and could comfortably perform typing tasks.

2) Typing Tasks: Participants were presented with a set of short sentences sourced from CNN news articles to type. The sentences were standardized across participants to ensure consistency. Participants were instructed to type as accurately and quickly as possible while maintaining a comfortable typing pace.

3) Break and comfort: Participants were allowed to take breaks at any time during the experiment to ensure their comfort and prevent symptoms such as "VR sickness."

4) Typing hands position: The experimental setup involved recording the typing actions of the participants using a combination of a USB camera and a virtual camera within the real environment and VR environment. These cameras captured the real hand position and the virtual hand positions when the participants pressed the keys on the keyboard. By combining these recordings, a dataset was created for each typing session. In cases where the hand tracking accuracy was high and the impact of jitter was minimal, it was expected that the typing postures of the real and virtual hands would closely resemble each other. By comparing the typing postures of the real and virtual hands, the level of fidelity and jitter in replicating the hand movements in the virtual environment could be evaluated.

5) Data Collection: During the typing tasks, the following data need to be collected:

a) Total number of words (NoW) entered (including errors) in normal, Oculus Quest 2, Leap Motion, and 2S-LSTM conditions.

b) Number of errors (E) in normal, Oculus Quest 2, Leap Motion, and 2S-LSTM conditions.

c) Error rate (ER) in normal, Oculus Quest 2, Leap Motion, and 2S-LSTM conditions.

d) Difference (Diff.) of hand positions in HMD typing conditions. Note that the difference between real and virtual hand positions was quantified at 21 * 2 key points of hand and the difference is summed for 100 inputs.

E. Questionnaire

After each typing task, participants were asked to complete a questionnaire. Each question of the questionnaire consisted of a 7-point Likert scale, ranging from 1 (negative) to 7 (positive). The questionnaire is shown in Table II. The questionnaire was provided after each of the four experimental conditions: Normal, Oculus Quest 2, Leap Motion, and 2S-LSTM. The questions which marked as "only for VR typing" were not asked in the normal condition.

TABLE II. QUESTIONNAIRE

	Question
1.	Did you perform at your normal typing efficiency during this typing session?
2.	How fatigued did you feel during the typing session?
3.	To what extent did the virtual hands replicate real hand position during this typing session? (Only for VR typing)
4.	Would you be willing to replace traditional typing with this typing scheme? (Only for VR typing)
5.	Please evaluate the level of jitter in this VR typing system. (Only for VR typing)
6.	How much did Jitter have a negative impact on you in last task? (Only for VR typing)
7.	How much dizziness did you experience during the typing task? (Only for VR typing)
8.	How comfortable did you find in last typing task? (Only for VR typing)
9.	Were there any times during the typing when you just wanted to give up?
10.	Would you like to use this typing system again in the future? (Only for VR typing)
11.	Please evaluate your level of focus during the typing process.
12.	Did your typing fluent in last task?

V. RESULT

In order to evaluate the impact of the factors on user performance, we conducted statistical tests using SPSS software. First, we conduct tests to examine the normality and homogeneity of variance of all the collected data.

A. Typing Data

The average results of typing data are shown in Fig. 3. We conducted tests for normality and homogeneity of variances. Since the sample size for all the collected data is less than 50, the Shapiro-Wilk (S-W) test was used for the normality test. The results show that E and ER for all conditions followed the normal distribution (P-values of E: 0.421, 0.137, 0.188, 0.484 respectively; P-values of ER: 0.082, 0.138, 0.338, 0.344 respectively). However, the tests for homogeneity of variances indicated that E (P = 0.011) and ER (P = 0.000**) did not meet the assumption of equal variances.

Moreover, none of the conditions exhibited normal distributions for Now and Diff. values (P values of NoW: 0.001, 0.012, 0.011, 0.001 respectively; P values of Diff.: 0.001, 0.013, 0.011 respectively). Therefore, non-parametric tests were employed to analyze the total number of words typed, number of errors, error rates, and Diff. values. Since there were more than two conditions, the Kruskal-Wallis test was used to examine the differences among conditions. The results indicated significant differences among the conditions for the NoW, E, ER, and Diff. (P values are all less than 0.05).

We conducted multiple comparisons using the Mann-Whitney U test with Bonferroni's adjustment. For NoW, the comparison of 2S-LSTM and Leap Motion is no significant difference (P = 0.357). For E, the comparison of 2S-LSTM and Leap Motion also no significant difference (P = 0.313). For other comparisons, the p-values are all less than 0.05. In summary, the number of NoW is Normal > 2S-LSTM = Leap Motion > Oculus, the number of E is Oculus > Leap Motion = 2S-LSTM > Normal, and the number of Diff. is Oculus > Leap Motion > 2S-LSTM, respectively.

B. Questionnaire

The sample sizes for all questions are less than 50, the Shapiro-Wilk (S-W) test was used. However, the data for all these questions did not exhibit normal distribution characteristics (P values are all less than 0.05). Therefore, non-parametric tests were employed. Since there are more than two experimental conditions, the Kruskal-Wallis test statistic was used for analysis. The results showed that there was no significant difference among the different conditions for Question 11 (H= 0.446, p = 0.93). For other questions, the different typing conditions demonstrated significant differences (all p values are less than 0.05). The average score of each question in different conditions is shown in Fig. 3.

We also performed multiple comparisons using the Mann-Whitney U test with Bonferroni's adjustment. For Question 2, the comparison of 2S-LSTM and Normal is no significant difference (P = 0.514). For Question 7, Question 8, and Question 12, the comparison of Leap Motion and Oculus is no significant difference (P = 0.445, P = 0.102, p = 0.101). For Question 9, the comparison of 2S-LSTM and Leap Motion is no significant difference (P = 0.054). The results were (Normal >) 2S-LSTM > Leap Motion > Oculus for most questions.

VI. Discussion

A. Typing Data

Notably, it is evident from our statistical analysis that the 2S-LSTM outperformed the Oculus Quest 2 and Leap Motion. These findings highlight the importance of considering the specific typing scheme when evaluating typing efficiency, error rates, and Diff. values. The Mann-Whitney U test with Bonferroni's adjustment was conducted to obtain these results.

From the results of the Mann-Whitney U test with Bonferroni's adjustment, we can conclude that there is no significant difference between 2S-LSTM and Leap Motion in the number of inputs and errors quantity per unit time. Our method utilizes a regular RGB camera on HMD, while Leap Motion employs a depth camera. Therefore, achieving similar results to Leap Motion by using a regular device is still considered a positive outcome. Additionally, there is a significant difference between 2S-LSTM and the other methods in Diff.. This result indicates that using the original image and MHI, combined with the implementation of KF to reduce jitter, indeed leads to a reduce the Diff.. Considering the deployment cost and the other results obtained in this research, we have reasons to believe that our approach is superior to the Leap Motion and Oculus solutions.

B. Questionnaire

The questionnaire focused on various aspects such as typing efficiency, fatigue, replication of hand position, the willingness to replace traditional typing, evaluation of jitter, negative impact of jitter, dizziness, comfort, willingness to continue using the system, focus level, and typing fluency. The statistical analysis involved non-parametric tests due to the data not exhibiting normal distribution characteristics. The questionnaire results indicated no significant difference among the different conditions for Question 11, which evaluated the level of focus during the typing process. This suggests that the different typing conditions, including the use of 2S-LSTM, Oculus Quest 2, and Leap Motion, did not significantly affect the participants' focus level.

Notably, it can be observed that the 2S-LSTM condition generally outperformed the Oculus Quest 2 and Leap Motion conditions in terms of typing efficiency, fatigue, replication of hand position, willingness to replace traditional typing, evaluation of jitter, negative impact of jitter, dizziness, comfort, and typing fluency. These findings suggest that the 2S-LSTM typing solution showed promising results in various aspects compared to the existing solutions of Oculus Quest 2 and Leap Motion. The 2S-LSTM condition exhibited higher typing efficiency, lower fatigue levels, better replication of hand position, and a more positive user experience.

From the results of the Mann-Whitney U test with Bonferroni's adjustment, for question 2: "How fatigued did you feel during the typing session?", there was no significant difference between 2S-LSTM and Normal. This result shows 2S-LSTM performs excellently in the VR typing task, and users do not experience additional fatigue from VR. For Question 7: "How much dizziness did you experience during the typing task?", Question 8: "How comfortable did you find the last typing task?", and Question 12: "Did your typing feel fluent in the last task?", there was no significant difference between Leap Motion and Oculus. However, our approach showed significant differences in these questions compared to Leap Motion and Oculus, and the questionnaire results are more positive. This outcome suggests that, compared to the existing VR systems, our approach hard to make user fill dizziness in VR typing tasks and also superior in comfort and typing fluency. Furthermore, for Question 9: "Were there any times during the typing which you just wanted to give up?", there has no significant difference between our approach and Leap Motion. Although our approach performed better in reducing dizziness, improving comfort, and better typing fluency, users still want to give up while using our approach to type. We suggest that there might be some hidden flaws in our approach that lead to user dissatisfaction. Therefore, further discussions and investigations regarding this issue are essential for future improvements.

There still have some limitations in this research. Firstly, the sample size for the questionnaire was limited to a specific number of participants. Expanding the sample size and including a more diverse group of participants could enhance the generalizability of the findings. Additionally, the study focused on specific typing tasks and conditions, and further investigation is needed to evaluate the solution's performance in different contexts and for various user profiles. In conclusion, the results of the questionnaire highlighted significant differences among the different typing conditions, with the 2S-LSTM solution demonstrating superior performance compared to the Oculus Quest 2 and Leap Motion solutions. These findings support the effectiveness of the developed solution in improving typing efficiency, reducing fatigue, and providing a more comfortable and satisfactory typing experience.

VII. Conclusion

This study addresses the challenge of text entry in virtual reality (VR) environments, specifically in immersive office experiences. By leveraging machine learning approaches, the proposed 2S-LSTM typing solution, utilizing the back of the hand image, demonstrates superior performance compared to existing solutions such as Oculus Quest 2 and Leap Motion. The 2S-LSTM solution significantly improves typing efficiency, reduces fatigue, accurately replicates hand position, and provides a more positive user experience. These findings underscore the potential of the developed solution in enhancing typing performance and user satisfaction in VR environments.

Furthermore, the outcomes of this study are expected to contribute significantly to the fields of distance learning and telecommuting, as addressing the challenges of text entry in VR can facilitate the development and widespread adoption of VR technology in various applications. Future research and development efforts can focus on refining the solution and exploring its potential applications in practical settings. Additionally, expanding the sample size, incorporating additional typing metrics, and further investigating factors influencing typing performance in VR environments can provide valuable insights for the development and refinement of VR typing systems.

Reference

[1] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson, "Text entry in immersive head-mounted display-based virtual reality using standard keyboards," In IEEE Virtual Reality (VR) 2018, page to appear. IEEE, 2018.

[2] D. A. Bowman, C. J. Rhoton, and M. S. Pinho, "Text input techniques for immersive virtual environments: An empirical comparison," In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 46, pp.2154–2158, 2002.

[3] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, "3D User Interfaces: Theory and Practice," Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.

[4] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson, "Effects of hand representations for typing in virtual reality," In IEEE Virtual Reality (VR) 2018, page to appear. IEEE, 2018.

[5] J. P. Stauffert, F. Niebling, M. E. Latoschik, "Effects of Latency Jitter on Simulator Sickness in a Search Task," In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp.121-127. IEEE Computer Society, 2018.

[6] D. Hwang, K. Aso, and H. Koike, "MonoEye: Monocular Fisheye Camera-based 3D Human Pose Estimation," In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp.988–989. IEEE VR, 2019.

[7] W. Erwin, Y. Ye, H. S. Yeo, A. Quigley, and H. Koike, M. Kitani, "Back-Hand-Pose: 3D Hand Pose Estimation for a Wrist-Worn Camera via Dorsum Deformation Network," Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, pp.1147–1160. Association for Computing Machinery, 2020.

[8] Y. Jang, I. Jeon, T. Kim, and W. Woo, "Metaphoric Hand Gestures for Orientation-Aware VR Object Manipulation with an Egocentric Viewpoint," IEEE Transactions on Human-Machine Systems 47, 1, pp.113–127, IEEE, 2017.

[9] B. Gavin, "Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges," Virtual Reality and Human Behaviour., Volume 2 - 2021.

[10] Heng Wang, Bin Wang , Bingbing Liu , Xiaoli Meng , Guanghong Yang, "Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle," Robotics and Autonomous Systems, Volume 88, pp.71-78, February, 2017.

[11] Lizy Abraham ,Andrea Urru, Niccolò Normani, "Hand Tracking and Gesture Recognition Using Lensless Smart Sensors," Special Issue Emerging Algorithms and Applications in Vision Sensors System based on Artificial Intelligence, sensors 2018.

[12] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie, "Effects of tracking technology, latency, and spatial jitter on object movement," In 3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on, pp.43–50. IEEE, 2009.

[13] A. Pavlovych and W. Stuerzlinger, "The tradeoff between spatial jitter and latency in pointing tasks," in proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems, pp.187–196. ACM, 2009.

[14] A. U. Batmaz, M. R. Seraji, J. Kneifel, and W. Stuerzlinger, "No jitter please: Effects of rotational and positional jitter on 3D mid air interaction," In Future Technologies Conference, vol. AISC 1289. Springer, Nov 2020.

[15] M. H. Mughrabi, A. K. Mutasim, W. Stuerzlinger, A. U. Batmaz, "My Eyes Hurt: Effects of Jitter in 3D Gaze Tracking," 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp.12-16, March, 2022.

[16] https://github.com/vidursatija/BlazePalm