UDC 004.5:004.93, 004.42, 004.92
DOI https://doi.org/10.32782/IT/2024-3-4

Serhii ZELINSKYI
PhD Student at the Department of Computer Engineering, Faculty of Radiophysics, Electronics and Computer Systems, Taras Shevchenko National University of Kyiv, 60, Volodymyrska Str., Kyiv, Ukraine, 01033
ORCID: *0009-0003-6271-2601*

Yuriy BOYKO
Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Computer Engineering, Faculty of Radiophysics, Electronics, and Computer Systems, Taras Shevchenko National University of Kyiv, 60, Volodymyrska Str., Kyiv, Ukraine, 01033
ORCID: *0000-0003-1417-7424*
Scopus Author ID: *24722552300*

To cite this article: Zelinskyi, S., Boyko, Yu. (2024). Doslidzhennia vzaiemodii na osnovi pohliadu ta zhestiv u veb-seredovyshchi: porivniannia z vzaiemodiieiu kompiuternoiu mysheiu dlia manipuliuvannia obiektamy [Exploring gaze-gesture interaction on the web: a comparison with mouse input for object manipulation]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 33–42, doi: https://doi.org/10.32782/IT/2024-3-4

EXPLORING GAZE-GESTURE INTERACTION ON THE WEB: A COMPARISON WITH MOUSE INPUT FOR OBJECT MANIPULATION

The purpose *of this study is to implement and evaluate a web-based gaze-gesture interaction method for object manipulation using a standard web camera. This method combines gaze tracking for object selection with hand gestures for natural manipulation tasks like rotating, scaling, and dragging. Unlike other implementations of such interaction that require specialized hardware, this method uses widely available technology, making advanced interaction techniques more accessible.*

The scientific novelty *lies in developing a gaze-gesture interaction system that operates entirely on a web platform using standard hardware, removing the need for expensive, specialized equipment and enabling broader adoption.*

The methodology *involved creating a web-based system using computer vision algorithms for real-time gaze tracking and gesture recognition. A user study was conducted where participants completed object manipulation tasks using both the gaze-gesture input and traditional mouse input, with task completion times recorded and analyzed.*

Conclusion. *The study shows that gaze-gesture interaction is particularly effective for tasks requiring simultaneous actions, such as rotating and scaling objects, outperforming mouse input in these scenarios. While mouse input remains more efficient for simpler tasks, gaze-gesture interaction offers strong potential for enhancing complex task interactions on web platforms, contributing to the development of more accessible and intuitive input methods.*

Key words: *gaze-gesture interaction, mouse interaction, web-based interaction, object manipulation, human-computer interaction, gaze tracking, accessible technology, hand gesture recognition.*

Сергій ЗЕЛІНСЬКИЙ
аспірант кафедри комп’ютерної інженерії, факультет радіофізики, електроніки та комп’ютерних систем, Київський національний університет імені Тараса Шевченка, вул. Володимирська, 60, м. Київ, Україна, 01033
ORCID: *0009-0003-6271-2601*

Юрій БОЙКО
кандидат фізико-математичних наук, доцент, завідувач кафедри комп’ютерної інженерії факультету радіофізики, електроніки та комп’ютерних, Київський національний університет імені Тараса Шевченка, вул. Володимирська, 60, м. Київ, Україна, 01033
ORCID: *0000-0003-1417-7424*
Scopus Author ID: *24722552300*

Бібліографічний опис статті: Зелінський, С., Бойко, Ю. (2024). Дослідження взаємодії на основі погляду та жестів у веб-середовищі: порівняння з взаємодією комп'ютерною мишею для маніпулювання об'єктами. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 33–42, doi: https://doi.org/10.32782/IT/2024-3-4

ДОСЛІДЖЕННЯ ВЗАЄМОДІЇ НА ОСНОВІ ПОГЛЯДУ ТА ЖЕСТІВ У ВЕБ-СЕРЕДОВИЩІ: ПОРІВНЯННЯ З ВЗАЄМОДІЄЮ КОМП'ЮТЕРНОЮ МИШЕЮ ДЛЯ МАНІПУЛЮВАННЯ ОБ'ЄКТАМИ

***Метою** цього дослідження є реалізація та оцінка веб-орієнтованого методу взаємодії на основі погляду та жестів для маніпулювання об'єктами, використовуючи стандартну веб-камеру. Цей метод поєднує відстеження погляду для вибору об'єктів з жестами рук для природних маніпуляцій, таких як обертання, масштабування та перетягування. На відміну від інших реалізацій цієї взаємодії, що вимагають спеціалізованого обладнання, цей метод використовує широко доступні технології, роблячи передові техніки взаємодії більш доступними.*

***Наукова новизна** полягає в розробці системи взаємодії на основі погляду та жестів, яка працює повністю на веб-платформі з використанням стандартного обладнання, що усуває потребу в дорогому спеціалізованому обладнанні та дозволяє ширше впровадження.*

***Методологія** включає створення веб-системи з використанням алгоритмів комп'ютерного зору для відстеження погляду в реальному часі та розпізнавання жестів. Було проведено дослідження за участю користувачів, які виконували завдання з маніпулювання об'єктами, використовуючи як метод введення на основі погляду та жестів, так і традиційне введення комп'ютерною мишею, з подальшим записом та аналізом часу виконання завдань.*

***Висновок.** Дослідження показує, що взаємодія на основі погляду та жестів є особливо ефективною для завдань, що вимагають одночасних дій, таких як обертання та масштабування об'єктів, перевершуючи введення комп'ютерною мишею у таких сценаріях. Хоча введення комп'ютерною мишею залишається більш ефективним для простіших завдань, взаємодія на основі погляду та жестів має великий потенціал для підвищення ефективності користувачів при виконанні складних завдань на веб-платформах, сприяючи розвитку більш доступних та інтуїтивно зрозумілих методів введення.*

***Ключові слова:** взаємодія на основі погляду та жестів, взаємодія комп'ютерною мишею, веб-орієнтована взаємодія, маніпулювання об'єктами, взаємодія людина-комп'ютер, відстеження погляду, доступні технології, розпізнавання жестів рук.*

Introduction. The interaction between users and computer systems has greatly evolved, expanding from traditional input devices like the mouse and keyboard to more natural methods, including eye-tracking, hand gestures, facial expressions, or their combinations. These advancements are designed to create more engaging and efficient user experiences, especially in complex digital environments.

Motivation. Pointing to graphical elements is one of the fundamental tasks in human-computer interaction (HCI) (Argelaguet & Andujar, 2013). Eye-tracking stands out compared to traditional mouse input due to its speed (Ware & Mikaelian, 1986). This allows users to quickly select objects simply by looking at them. On the other hand, hand gestures allow for more natural and intuitive manipulation of objects, enabling users to rotate, scale, and move items as they would in the physical world. The combination of gaze tracking and hand gestures has gained considerable attention, especially in virtual reality (VR) (Pfeuffer et al., 2017), where it enhances user experience by creating seamless and immersive interactions. However, the adoption of this interaction method outside of VR, particularly on the web, has been limited due to hardware requirements.

Objective. The primary objective of this study is to implement and evaluate a gaze and hand gesture-based interaction (gaze-gesture interaction) technique on the web using only a standard computer web camera. By removing the need for specialized hardware, the study aims to make this advanced interaction method more accessible and practical for a broader range of users.

Approach. To achieve this objective, the study involves the development of a web-based platform that leverages a standard web camera for gaze tracking and gesture recognition. The effectiveness of the gaze-gesture interaction method is evaluated by comparing it to the traditional mouse input method in scenarios where users are required to manipulate objects, such as rotating, scaling, and dragging. The comparison is based on task completion time.

Related Work. In recent years, new types of input methods have appeared in the field of human-computer interaction. Traditional devices like the mouse and keyboard have been supplemented by more advanced techniques, including gaze tracking, hand gestures, and facial expressions. For instance, (Rozado et al., 2017) describe the FaceSwitch, which supports motor-impaired users to interact with a computer hands-free by using gaze pointing for target selection and facial gestures for target-specific action commands. (Wachs et al., 2011) examines various applications of vision-based hand-gesture interfaces across different fields, such as medical systems, assistive technologies, entertainment, crisis management, disaster relief, and human-robot interaction.

Gaze and Hand Gesture Interaction

Gaze tracking is recognized for its speed, allowing users to quickly select and focus on objects simply by looking at them. This makes gaze interaction faster than other input modalities (Ware & Mikaelian, 1986), particularly in scenarios where quick selections are necessary. Hand gestures complement gaze tracking by providing a more natural way to manipulate objects, enabling actions like rotating, scaling, and dragging with intuitive hand movements. (Slambekova et al., 2012) presented a framework for enabling the use of both gaze and hand gestures for interaction within a 3D virtual world.

A set of interaction techniques combining gaze and free-space hand gestures has been presented in (Chatterjee et al., 2015). Results showed that the combination of gaze and gesture can outperform systems using gaze or gesture alone. Another study showed the gaze-assisted techniques to outperform hands-only input and gives insight into trade-offs in combining gaze with direct or indirect, and spatial or semantic freehand gestures (Lystbæk et al., 2022).

(Ryu et al., 2019) proposed a spatial interaction technique called gaze-grasp pose interaction (GG Interaction) that can be used in 3D virtual spaces for object manipulation.

Another research introduced a novel virtual mouse system that enables users to control an on-screen pointer using hand and eye gestures, providing a contactless input method (Reddy et al., 2023).

The combination of gaze tracking and hand gestures has been studied in the context of virtual reality (VR). In VR environments, this interaction method allows users to look at objects and manipulate them with their hands, creating a highly immersive experience (Pfeuffer et al., 2017).

Despite the advantages of the interaction method based on the combination of gaze and hand gestures, this approach has its own challenges. In (Pfeuffer, 2024), the author discusses design principles and issues, focusing on interfaces that use gaze and pinch interaction.

Limitations of Current Implementations. Most implementations from the presented research rely on sophisticated equipment like eye-trackers, cameras with infrared (IR) sensors, or devices such as Microsoft Kinect. Some studies have even developed custom-made eye-tracking devices to enable this interaction method (Hales, 2013). The need for specialized hardware creates a barrier to the broad adoption of such an interaction method, particularly on more accessible platforms like the web.

Web-Based Interaction and Accessibility. With the increasing popularity of web applications, there is a growing interest in implementing advanced interaction techniques using widely available hardware. Recent advancements in computer vision and machine learning have made it possible to track gaze and recognize gestures directly in the browser using standard web cameras. This, in turn, allows us to bring gaze and gesture-based interaction methods to the web. However, the effectiveness and practicality of these web-based implementations in real-world scenarios remain underexplored.

Summary. Existing research highlights the advantages of combining gaze tracking and hand gestures, particularly in VR environments, where they enhance user experience and interaction efficiency. However, the reliance on specialized hardware limits the broader adoption of this method, especially on the web. This study aims to address this gap by implementing and evaluating a web-based gaze-gesture interaction method that utilizes a standard web camera. Through this research, we want to make this advanced interaction technique more accessible and practical for everyday use.

Methods

System Overview

The developed system integrates eye-tracking and hand gesture recognition technologies to enable intuitive interaction with on-screen objects. Fig. 1 illustrates the system architecture diagram. The system is composed of four primary modules:

1. **Eye-Tracking Module.** Responsible for detecting the user's gaze and identifying the on-screen object they are focusing on.
2. **Hand Detection Module.** Detects and tracks the user's hands, providing real-time positional data of key hand landmarks.
3. **Gesture Recognition Module.** Interprets the hand landmarks to recognize specific gestures, such as pinch gestures, and sends this information to the Object Manipulation module.
4. **Object Manipulation Module.** Receives data from the Eye-Tracking and Gesture Recognition modules and determines the appropriate transformation (e.g., selection, dragging, scaling,

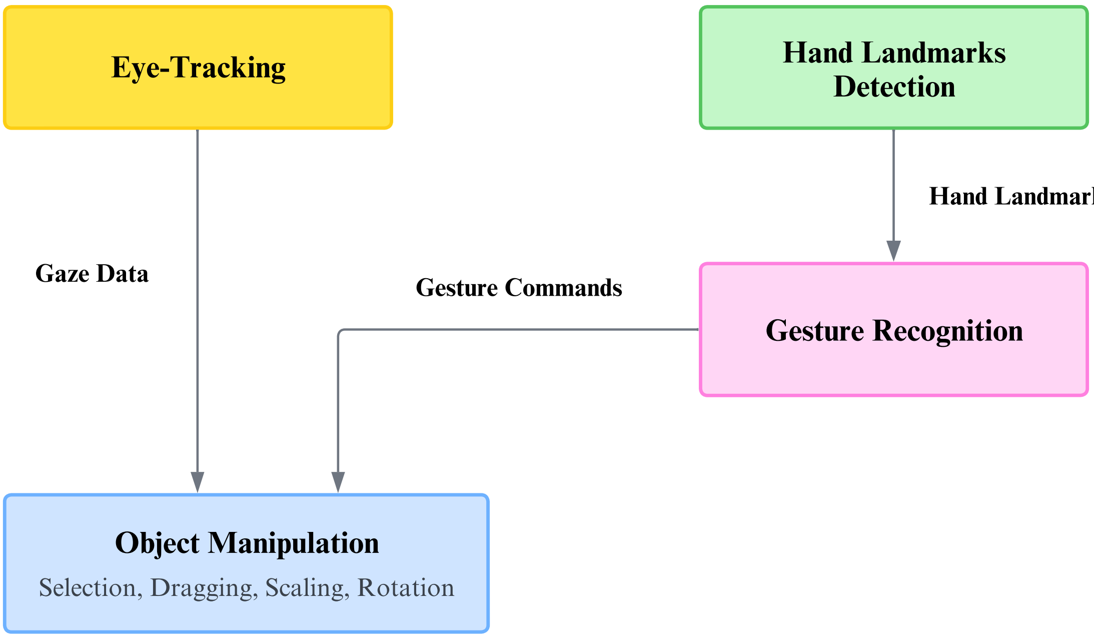

Fig. 1. System architecture overview

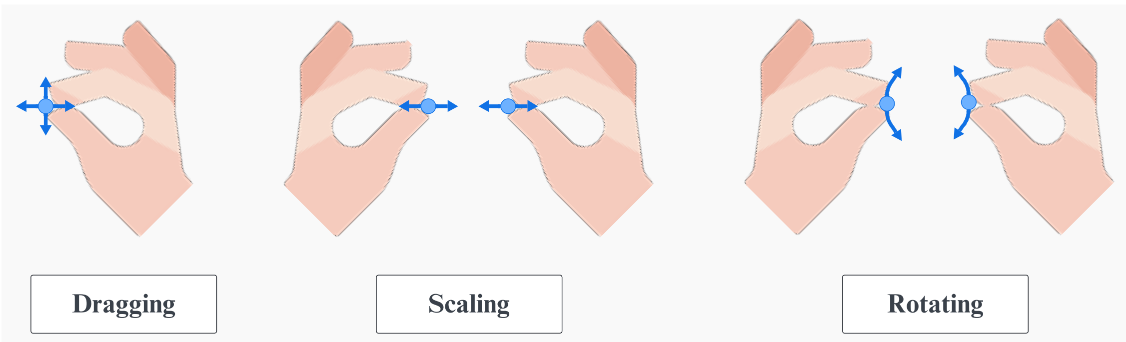

Fig. 2. Illustration of the hand gestures recognized by the system for dragging, scaling, and rotating objects

rotation) to be applied to the selected object based on the user's input.

Eye-Tracking Module. The Eye-Tracking Module, which was implemented using WebGazer.js (Papoutsaki et al., 2016), continuously monitors the user's gaze direction to determine the point of regard on the screen. This information is used to identify which object the user intends to interact with. The gaze data is crucial for initiating the selection process, allowing for hands-free interaction.

Hand Detection Module. The Hand Detection Module utilizes the MediaPipe Hand Landmarker (Google, n.d.) to accurately identify and track the positions of key landmarks on the user's hands. This module is responsible for providing the positional data necessary for gesture recognition. The detected hand landmarks are continuously updated to reflect the user's hand movements in real time.

Gesture Recognition Module. The Gesture Recognition Module processes the hand landmark data provided by the Hand Detection Module to identify specific gestures. For instance, the system recognizes pinch gestures by calculating the Euclidean distance between the thumb and index fingertip landmarks. When this distance falls below a specified threshold, the gesture is classified as a pinch. The recognized gestures are then passed on to the Object Manipulation Module, which uses the input to perform actions such as dragging, scaling, and rotating the selected object.

Fig. 2 illustrates gestures the system can detect and respond to. They include:

- **Dragging.** A single-hand pinch gesture is used to drag objects across the screen.
- **Scaling.** A two-hand pinch gesture, where the distance between the hands increases or decreases, is used to scale the object.
- **Rotating.** A two-hand pinch gesture, where the hands rotate relative to each other, is used to rotate the object.

Object Manipulation Module. The Object Manipulation Module is responsible for applying the appropriate transformation to the selected object based on the combined input from the Eye-Tracking and Gesture Recognition modules. This module determines whether to select, drag, scale, or rotate the object depending on the user's gaze position and the recognized gestures. For instance, if the system detects a gaze fixated on an object and a pinch gesture, it triggers the selection and dragging of that object. If pinch gestures are detected on both hands, the system initiates scaling and/or rotation actions depending on the movement direction.

Interaction Workflow

The interaction process in the developed system is designed to provide an intuitive and seamless user experience by integrating gaze-based object selection with hand gesture-driven manipulation. The interaction flowchart is displayed in Fig. 3. The interaction workflow is as follows:

1. **Gaze at Object.** The user initiates the interaction by gazing at the desired object on the screen. The eye-tracking module detects the gaze point and selects the object for potential manipulation.

2. **Hand Gesture Detection.** The system continuously monitors the user's hands to detect any gestures. The hand detection module identifies the positions of key hand landmarks and sends this data to the gesture recognition module.

3. **Gesture Recognition.** Once a gesture is detected, the system interprets the specific type of gesture being performed. The recognition module identifies single-hand pinch gestures for dragging and two-hand pinch gestures for scaling and rotating.

4. **Object Manipulation.** Based on the recognized gestures and gaze data, the system performs the corresponding action – whether it be dragging, scaling, rotating, or a combination of scaling and rotating – on the selected object.

The system supports simultaneous scaling and rotation of the selected object. When a two-hand pinch gesture is detected, the system analyzes both the distance between the hands (for scaling) and the rotational movement (for rotation). If both actions are detected simultaneously, the system applies both transformations concurrently to the selected object.

System Interface Overview

The interface used during the evaluation sessions is illustrated in Fig. 4. This screenshot captures the key elements of the system that were critical for conducting the experiments:

1. **Video Feed and Hand Landmark Overlay.** The interface features a live video feed with hand landmarks overlayed to provide real-time feedback on recognized gestures.

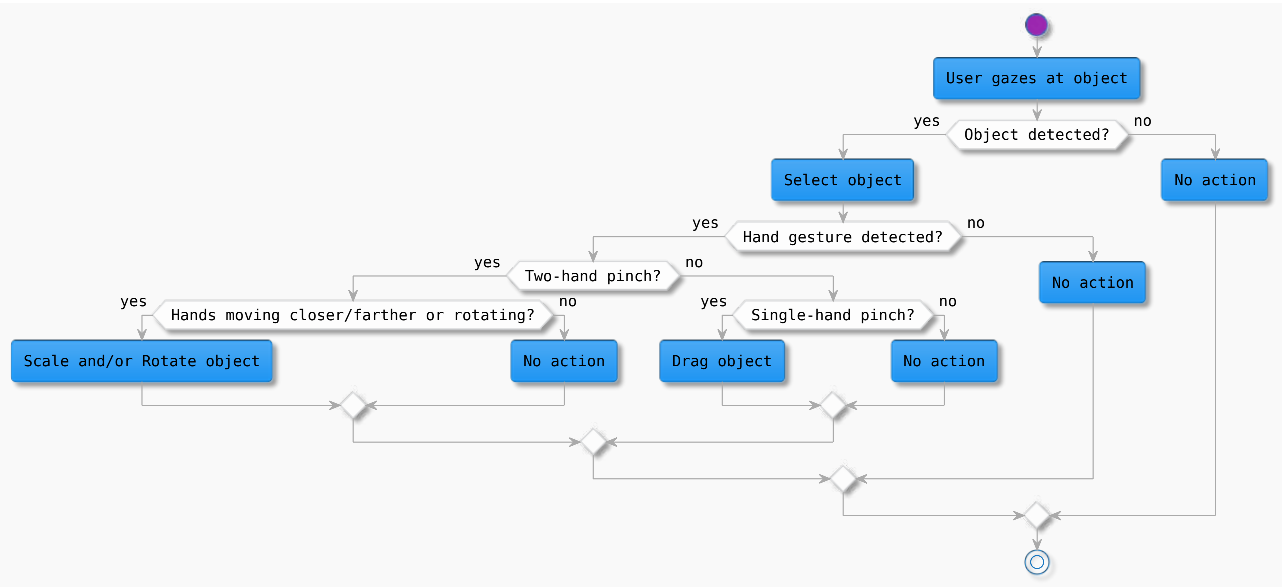

Fig. 3. Interaction workflow

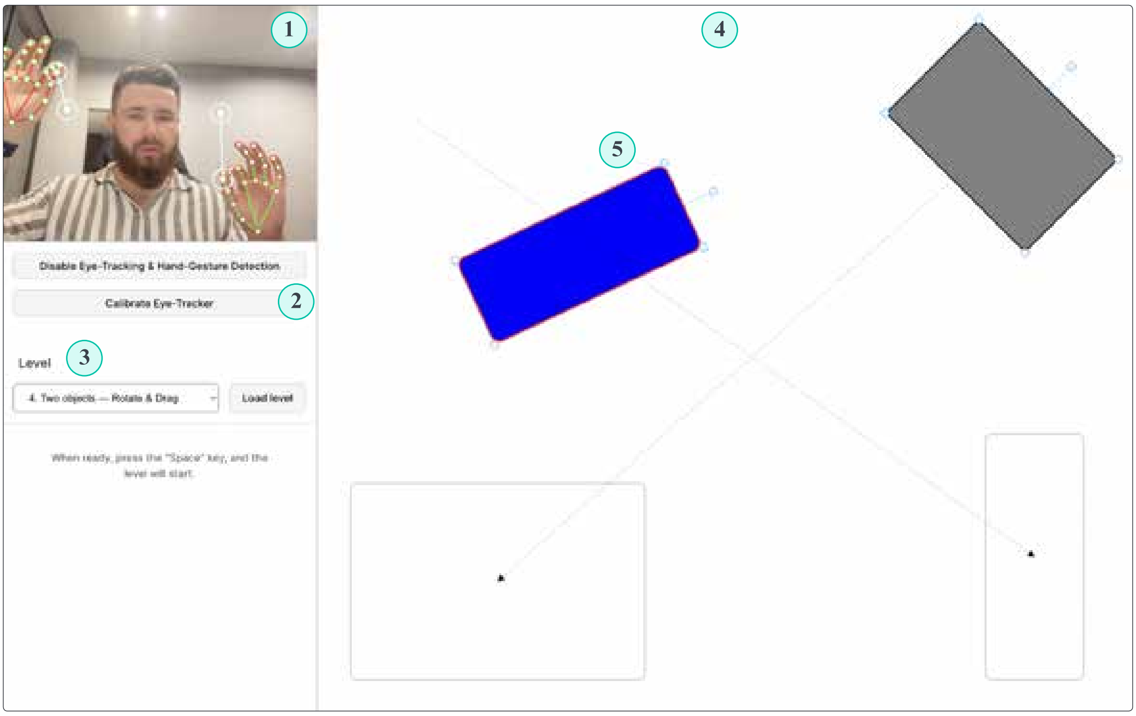

Fig. 4. Screenshot of the developed system interface – manipulating selected object

2. **Eye-Tracker Calibration Button.** A calibration tool is provided to ensure the accuracy of the eye-tracking system before each session starts.
3. **Level Picker.** A tool for selecting different levels, varying in difficulty.
4. **Object Manipulation Stage.** This area displays the objects participants need to manipulate and position within their respective designated areas.
5. **Selected Object.** The object the current transformation is applied to.

The tasks for participants were structured into 5 different levels, with increasing complexity from Level 1 to Level 5. Each level can be completed using either mouse input or gaze-gesture Input.

To complete a level, users must position objects in the designated areas (the system displays a designated area for each object using hint arrows). After completing a level, the system displays the time taken by a user to complete it. The complexity of the levels is based on the number of objects and transformations users need to apply to the objects in order to complete the level:

- **Level 1.** Drag one object to the designated area.
- **Level 2.** Drag two objects to their respective designated areas.
- **Level 3.** Scale and drag two objects to the designated areas.
- **Level 4.** Rotate and drag two objects to the designated areas.
- **Level 5.** Rotate, scale, and drag two objects to their designated areas.

Fig. 5 contains an overview of all evaluation levels.

The user interface of the completed level is illustrated in Fig. 6. The following key elements, except those already mentioned, are numerically highlighted:

1. **Time Display.** After the current level is completed, the system displays the time taken by the user to complete it.
2. **Gaze Indicator.** Displays the user's gaze point.
3. **Matched Object.** The object is successfully positioned within the designated area.

Evaluation and Results

Evaluation Approach

To evaluate the effectiveness of the gaze-gesture interaction method in comparison to the traditional mouse input method, a user study was conducted involving 10 participants. Each participant completed 5 levels, designed to progressively increase in complexity, using both interaction methods. Each participant completed each level 5 times (5 trials) for each method.

The levels were structured as follows:

- **Level 1.** Simple object dragging: participants had to **drag one object** to a designated area.
- **Level 2.** Dragging multiple objects: participants had to **drag two objects** to their respective designated areas.

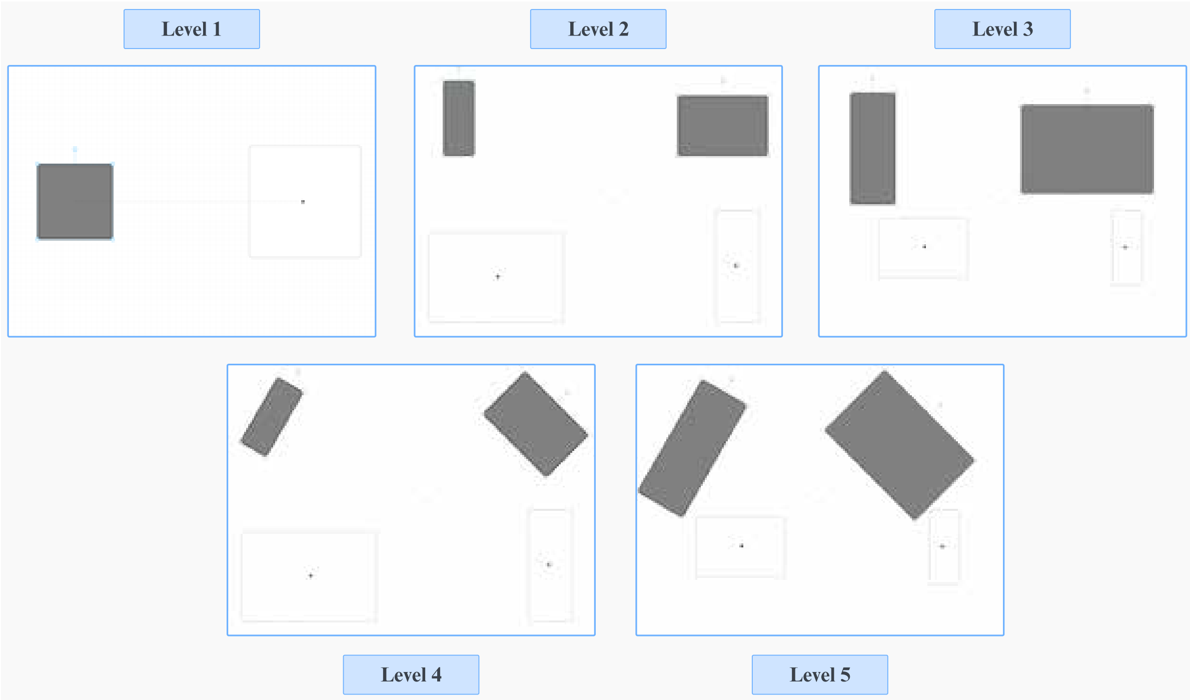

Fig. 5. Overview of all evaluation levels

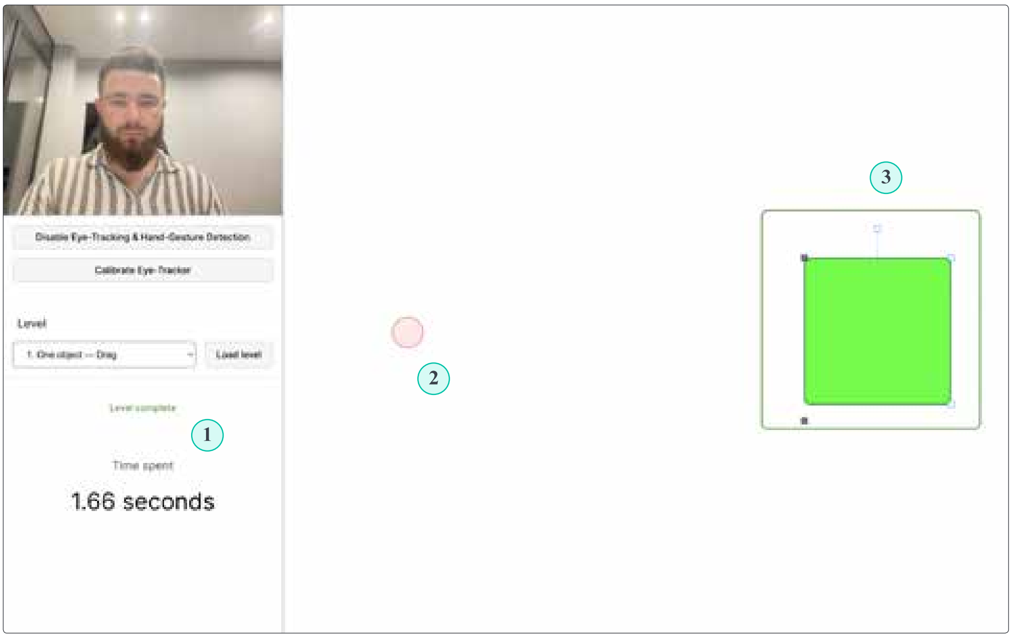

Fig. 6. Screenshot of the developed system interface – completed level

- **Level 3.** Scaling and dragging objects: participants had to **scale and drag two objects** to their respective designated areas.
- **Level 4.** Rotating and dragging objects: participants had to **rotate and drag two objects** to their respective designated areas.
- **Level 5.** A combination of rotating, scaling, and dragging objects: participants had to **rotate, scale, and drag two objects** to their respective designated areas.

The primary metric used for evaluation was the time taken to complete each level, recorded in seconds.

Results. The average task completion time for each level, aggregated across all participants, is shown in Fig. 7. The chart compares the performance of the mouse input method with the gaze-gesture interaction method.

As shown in the chart, the mouse input method generally resulted in faster task completion times

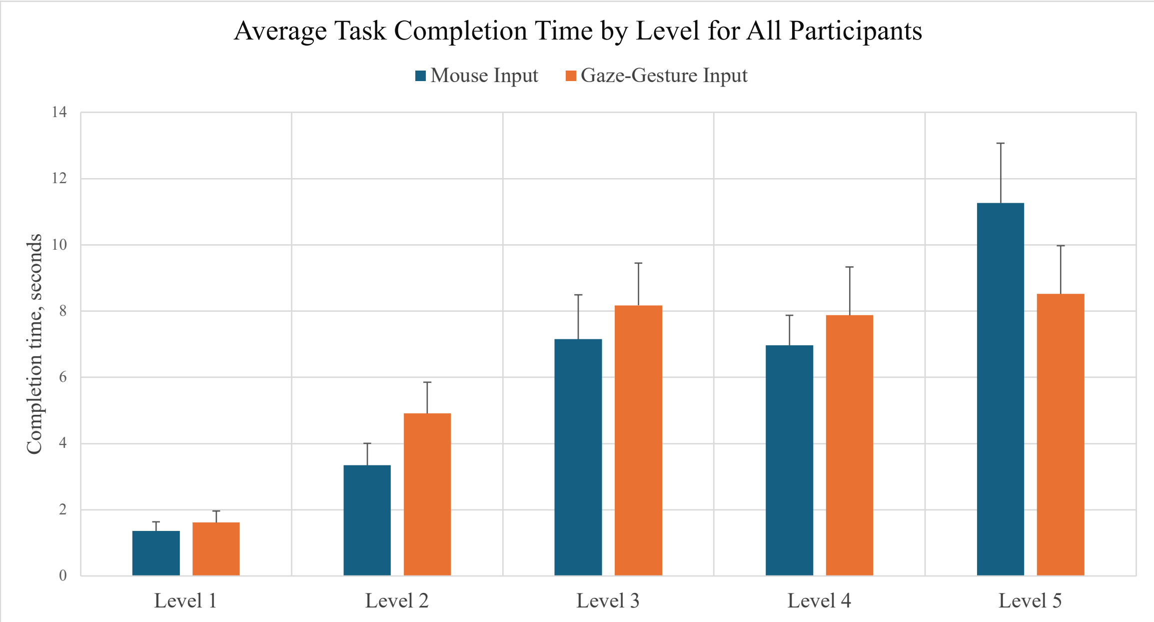

Fig. 7. Average task completion time by level across all participants

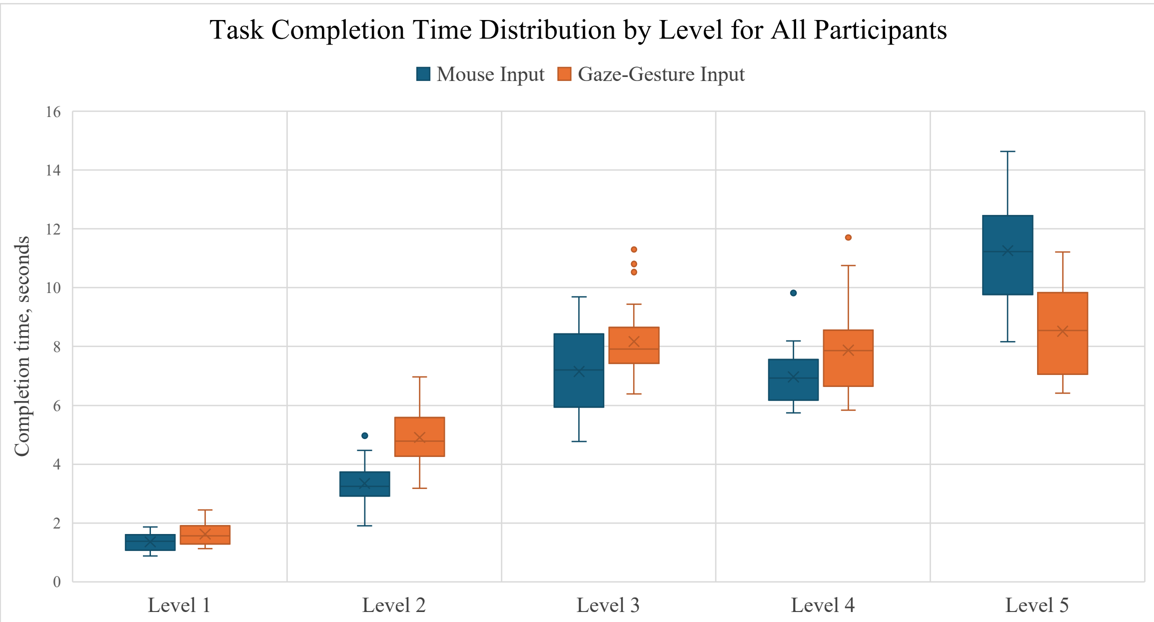

Fig. 8. Task completion time distributed by level for all participants

across most levels, particularly for simpler tasks. However, in Level 5, where participants were required to rotate, scale, and drag objects, the gaze-gesture interaction method outperformed the mouse interaction method. This suggests that the gaze-gesture interaction method is particularly well-suited for complex tasks that involve multiple simultaneous actions.

The box plots of the task completion times for each level are shown in Fig. 8. It illustrates variability in participants' performance for both interaction methods. The mouse input method generally has lower variability, particularly in the earlier levels. In contrast, the gaze-gesture interaction method showed higher variability, which is expected given the novelty of the interaction technique. However, in Level 5, the gaze-gesture interaction method not only matched but exceeded the mouse method in efficiency, as indicated by the lower median completion time. This supports the observation that participants were able to leverage the simultaneous nature of gaze and hand gestures to perform complex actions more efficiently than when using a mouse, which required separate, sequential actions.

Discussion. While the mouse interaction method remains more efficient for simpler tasks, its sequential nature limits its effectiveness in complex scenarios. This study highlights the potential of the gaze-gesture interaction method, particularly in complex tasks that involve simultaneous actions like rotating, scaling, and dragging objects. Implemented and evaluated on a web-based platform using a standard web camera, this method requires no additional hardware, making it accessible and practical for a wide range of users. Additionally, with the gaze-gesture interaction method, it is possible to simultaneously manipulate multiple objects – users can capture and control two objects independently with each hand, which is not possible with traditional mouse input.

Potential Applications

Given its web-based implementation and reliance on standard technology, the gaze-gesture interaction method has several potential applications:

- **Web-Based Interactive Applications.** This method could be used in web applications that require complex object manipulations, such as online design tools, interactive maps, interactive educational platforms, and virtual reality environments accessed through a browser.
- **Remote Collaboration Tools.** It could enhance web-based collaboration platforms by allowing users to interact more naturally with shared content, such as rotating and scaling 3D models in real time.
- **Accessibility Tools.** The gaze-gesture method could provide an alternative input mechanism for users with limited mobility, offering more intuitive control over web interfaces with no need for traditional input devices.

Future Directions

Future research could explore hybrid approaches that combine the gaze-gesture interaction method with traditional input devices, as this may be useful in specific use cases. Additionally, further investigation into a wider range of applications for the gaze-gesture interaction method could help assess and expand its utility across various digital environments.

Conclusions. This study evaluated the gaze-gesture interaction method, implemented and tested on a web-based platform using a standard web camera with no additional hardware requirements. The results demonstrate that while the traditional mouse input method is more efficient for simpler tasks, the gaze-gesture method shows better results in scenarios that require simultaneous actions, like rotating, scaling, and dragging objects.

The ability to perform multiple actions concurrently is a significant advantage of the gaze-gesture method, making it particularly suitable for complex, multitasking environments. The implementation on the web using a standard camera demonstrates the accessibility and practicality of this approach, eliminating the need for specialized hardware.

Future work should focus on exploring hybrid approaches that combine the gaze-gesture interaction method with traditional input devices, as this may offer benefits in certain scenarios. Additionally, further research could investigate broader applications of the gaze-gesture interaction method across various digital environments to assess its potential to enhance interaction efficiency and user experience.

BIBLIOGRAPHY:

1. Argelaguet F., Andujar C. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 2013. 37(3), 121–136. https://doi.org/10.1016/j.cag.2012.12.003

2. Chatterjee I., Xiao R., Harrison C. Gaze+gesture: Expressive, precise and targeted free-space interactions. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 2015. 131–138. https://doi.org/10.1145/2818346.2820752

3. Google. (n.d.). MediaPipe hand landmarker. URL: fhttps://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker June 26, 2024

4. Hales J. Interacting with objects in the environment by gaze and hand gestures. 2013. URL:https://api.semanticscholar.org/CorpusID:52206471

5. Lystbæk M. N., Rosenberg P., Pfeuffer K., Grønbæk J. E., Gellersen H. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, *6*(ETRA), 2022. 1–18. https://doi.org/10.1145/3530886

6. Papoutsaki A., Sangkloy P., Laskey J., Daskalova N., Huang J., Hays J. WebGazer: Scalable webcam eye tracking using user interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 2016. 3839–3845.

7. Pfeuffer K. Design principles & issues for gaze and pinch interaction. *ArXiv*, abs/2401.10948. 2024. URL: https://api.semanticscholar.org/CorpusID:267069101

8. Pfeuffer K., Mayer B., Mardanbegi D., Gellersen H. Gaze + pinch interaction in virtual reality. *Proceedings of the 5th Symposium on Spatial User Interaction*, 2017. 99–108. https://doi.org/10.1145/3131277.3132180

9. Reddy N. L., Murugeswari R., Imran Md., Subhash N., Reddy N. V. K., Adarsh N. B. Virtual mouse using hand and eye gestures. *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, 2023. 1–5. https://doi.org/10.1109/ICDSAAI59313.2023.10452550

10. Rozado D., Niu J., Lochner M. Fast human-computer interaction by combining gaze pointing and face gestures. *ACM Transactions on Accessible Computing*, 2017. 10(3), 1–18. https://doi.org/10.1145/3075301

11. Ryu K., Lee J.-J., Park J.-M. GG interaction: A gaze–grasp pose interaction for 3D virtual object selection. *Journal on Multimodal User Interfaces*, 2019. 13(4), 383–393. https://doi.org/10.1007/s12193-019-00305-y

12. Slambekova D., Bailey R., Geigel J. Gaze and gesture based object manipulation in virtual worlds. *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology*, 2012. 203–204. https://doi.org/10.1145/2407336.2407380

13. Wachs J. P., Kölsch M., Stern H., Edan Y. Vision-based hand-gesture applications. *Communications of the ACM*, 2011. *54*(2), 60–71. https://doi.org/10.1145/1897816.1897838

14. Ware C., Mikaelian H. H. An evaluation of an eye tracker as a device for computer input2. *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, 1986. 183–188. https://doi.org/10.1145/29933.275627

REFERENCES:

1. Argelaguet, F., & Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, *37*(3), 121–136. https://doi.org/10.1016/j.cag.2012.12.003

2. Chatterjee, I., Xiao, R., & Harrison, C. (2015). Gaze+gesture: Expressive, precise and targeted free-space interactions. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 131–138. https://doi.org/10.1145/2818346.2820752

3. Google. (n.d.). MediaPipe hand landmarker. Retrieved June 26, 2024, from https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

4. Hales, J. (2013). Interacting with objects in the environment by gaze and hand gestures. https://api.semanticscholar.org/CorpusID:52206471

5. Lystbæk, M. N., Rosenberg, P., Pfeuffer, K., Grønbæk, J. E., & Gellersen, H. (2022). Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, *6*(ETRA), 1–18. https://doi.org/10.1145/3530886

6. Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 3839–3845.

7. Pfeuffer, K. (2024). Design principles & issues for gaze and pinch interaction. *ArXiv*, abs/2401.10948. https://api.semanticscholar.org/CorpusID:267069101

8. Pfeuffer, K., Mayer, B., Mardanbegi, D., & Gellersen, H. (2017). Gaze + pinch interaction in virtual reality. *Proceedings of the 5th Symposium on Spatial User Interaction*, 99–108. https://doi.org/10.1145/3131277.3132180

9. Reddy, N. L., Murugeswari, R., Imran, Md., Subhash, N., Reddy, N. V. K., & Adarsh, N. B. (2023). Virtual mouse using hand and eye gestures. *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, 1–5. https://doi.org/10.1109/ICDSAAI59313.2023.10452550

10. Rozado, D., Niu, J., & Lochner, M. (2017). Fast human-computer interaction by combining gaze pointing and face gestures. *ACM Transactions on Accessible Computing*, *10*(3), 1–18. https://doi.org/10.1145/3075301

11. Ryu, K., Lee, J.-J., & Park, J.-M. (2019). GG interaction: A gaze–grasp pose interaction for 3D virtual object selection. *Journal on Multimodal User Interfaces*, *13*(4), 383–393. https://doi.org/10.1007/s12193-019-00305-y

12. Slambekova, D., Bailey, R., & Geigel, J. (2012). Gaze and gesture based object manipulation in virtual worlds. *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology*, 203–204. https://doi.org/10.1145/2407336.2407380

13. Wachs, J. P., Kölsch, M., Stern, H., & Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM*, *54*(2), 60–71. https://doi.org/10.1145/1897816.1897838

14. Ware, C., & Mikaelian, H. H. (1986). An evaluation of an eye tracker as a device for computer input2. *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, 183–188. https://doi.org/10.1145/29933.275627