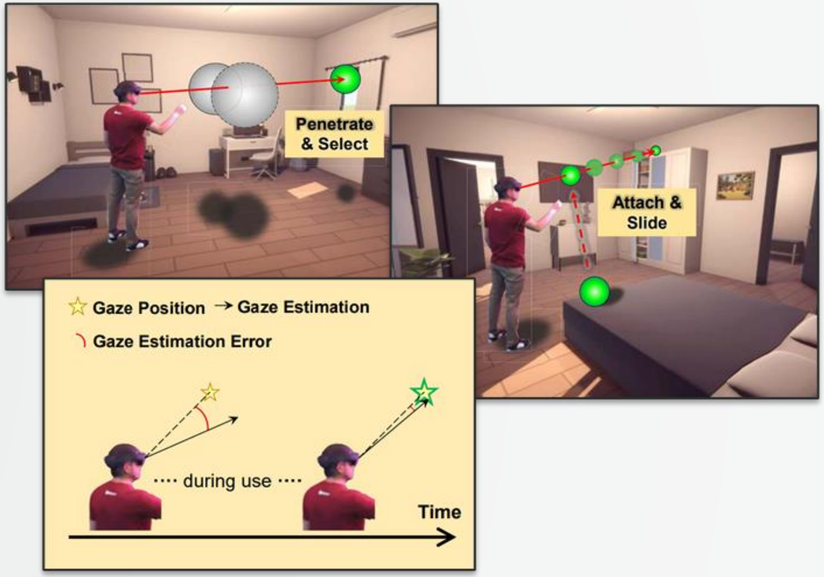
Exploring 3D Interaction with Gaze Guidance in Augmented Reality

- 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR), pp. 22-32;

- Authors: Yiwei Bao; Jiaxi Wang; Zhimin Wang; Feng Lu;

- HCI, April 15th 2024
- Bruno Gomes, 103320
- Alexandre Martins, 103552
- LECI



Why did we choose this paper?

- We believe that VR and AR are imporant techonologies that will become very present in our future lives;
- We were curious to see how our gaze would be integrated with AR environments;
- We wanted to get a new insight into how AR technology works;
- To be aware of the difficulties that the users face when using AR devices;
- Discover more innovative ways to manipulate entities while using AR;

Introduction

- VR/AR technologies:
  - Merge digital and physical worlds;
  - Increase user immersion compared to traditional 2D screens;
- Current interaction methods:
  - VR devices: Handheld controllers (eg., HTC Vive, Oculus Quest);
  - AR devices: Prefer hand gestures (eg, Hololens);
- Gazed-based Interactions:
  - Advantages: Fast and intuitive;
  - Limitations: Calibration is cumbersome and Midas Touch problem.
- Research objectives:
  - Object selection with 3D occlusions;
  - Depth dimension object translation;
  - Reducing explicit gaze calibration

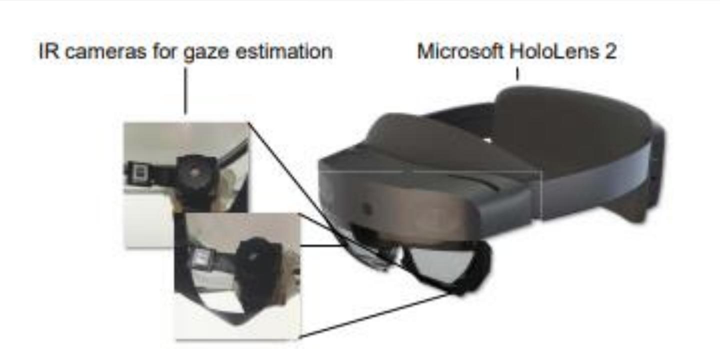
Study 01

Study 1 - Overview

- In Study 1, we evaluated and compared three hand-eye coordination techniques.
- The aim was to investigate whether gaze could assist in 3D object selection and translation.
- Two tasks were designed for each of the three interaction techniques.
- In the Heavy Occlusion task, participants needed to select target objects within multiple occlusions and move them to four target positions.
- In the Varies Depth task, participants were required to move non-occluded objects to different target positions with a maximum depth distance of 6.5 meters.

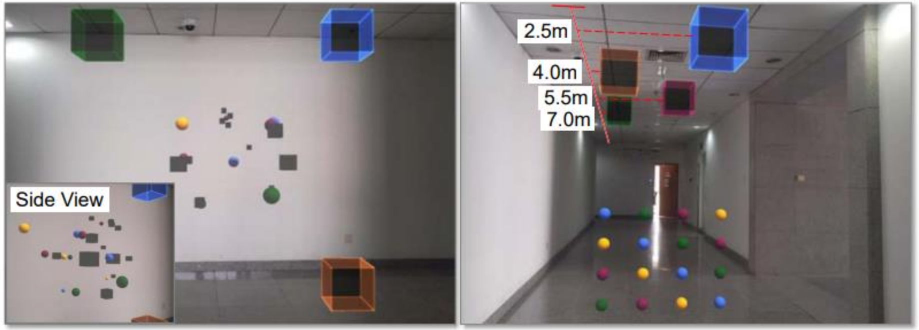
Study 1 - Participants and Devices

- Participants:
  - 15 university students recruited (14 males and 1 female).
  - Age range: 22 to 29 years old (mean=24.2).
  - All participants are familiar with computers and digital games.
- Devices:
  - Microsoft HoloLens 2 used for experiments.
  - Installed two infrared cameras with 30 FPS and 320 X 240 resolution for gaze estimation.
  - Software implemented in C# with Unity3D.



Study 1 – Task Design (1)

- Task Design:
  - Two search and move tasks were designed for Study 1: the Heavy Occlusion (HO) task and the Varies Depth (VD) task.
  - Goal of both tasks: Find spheres with four different colors and move them to corresponding color-coded target areas.
  - Sphere radius varies from 0.06 to 0.2 meters.

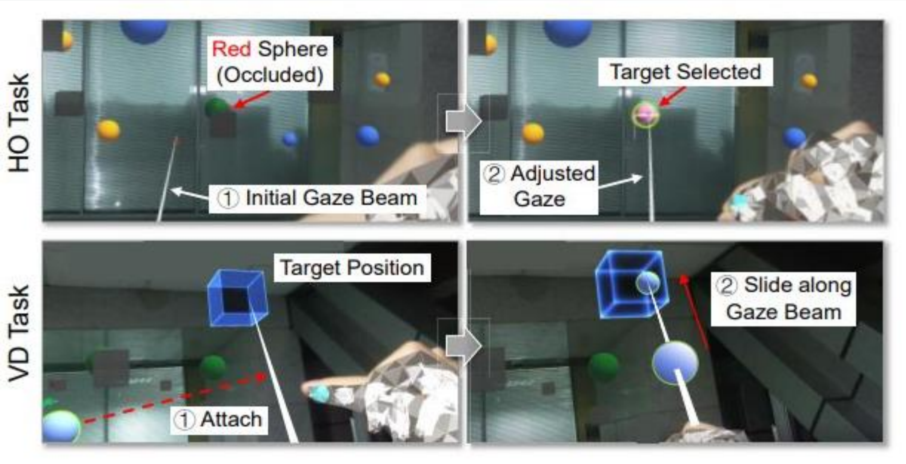


(a) Heavy Occlusion (HO) Task (b) Varies Depth (VD) Task

Study 1 – Task Design (2)

- Heavy Occlusion (HO) task:
  - Consists of 12 target spheres, with 3 spheres for each color.
  - Involves 12 gray cubes distributed as interference.
  - Only 2 target spheres are fully visible, while 6 are partially occluded and 4 are completely occluded.
  - Objects placed 5m away from participants.
  - Four target positions located on the corners of a 3m X 3m rectangle, 3m away.

- Varies Depth (VD) task:
  - Involves moving objects away in the depth axis for different distances.
  - Includes 16 target spheres without interference cubes.
  - Target positions located 1.5m high and at distances of 2.5m, 4m, 5.5m, and 7m in the depth axis.



Study 1 – Evaluation Metrics
1) Objective Measures

- Measurements for Selection in 3D:
  - Total Selection Time: Total completion time minus total translation time. Represents time spent observing and selecting.
  - Total Selection Count: Total number of selections made during the HO task. More selections imply more redundant work.
  - Invalid Selection Ratio: Proportion of interference cube selection count to Total Selection Count. Lower ratios indicate more effective selection in 3D spaces with occlusion.
- Measurements for Translation in 3D:
  - Hand Translation Distance: Total distance of hand translations, accumulated across frames. Longer distance suggests less effective translation.
  - Average Translation Count: Selection count of target objects divided by the number of target objects. Represents the number of translations required to move an object successfully to the target position.

Study 1 – Evaluation Metrics
2) Subjective Measures

- NASA-TLX [14]:
  - Utilizes a 21-point Likert scale to measure various aspects including mental demand, physical demand, temporal demand, effort, performance, and frustration level of participants.
- Borg CR10 [17]:
  - Employs a 10-point scale to measure the level of arm fatigue. It uses verbal anchors and numbers to map the magnitude of exertion to a scalar invariant scale.
- Subjective Ranking:
  - Assesses participants' preferences across all techniques.
- Open Questions:
  - Participants respond to open-ended questions about general evaluation, intuitiveness, frustration level, suggestions for improvement, and comparison to former techniques.

Study 1 – Experiment Procedure(1)

- Pre-Study Questionnaire:
  - Participants complete a questionnaire to assess their familiarity with gaze-related interaction techniques, hand-related interaction techniques, and AR systems.
- Introduction to Techniques:
  - Participants receive a brief introduction to all three interaction techniques.
- Warm-Up Trial:
  - Participants engage in a warm-up trial with random objects to become familiar with the techniques. They can interact freely with objects until they master the techniques.
- Formal Experiments:
  - Participants complete the HO task and the VD task using each of the three techniques in a predetermined order. The order is counterbalanced using the Latin Square approach.

Study 1 – Experiment Procedure(2)

- Calibration:
  - Participants undergo a standard 9-point calibration at the beginning of each technique to ensure accurate gaze estimation.

- Questionnaire and Break:
  - After completing each technique, participants fill out a questionnaire to collect subjective measures. They then take a 5-minute break to relax their arms before starting the next technique.

- Duration: The entire experiment lasts approximately 68 minutes.

Study 1 – Results
1) Objective Measures(1)

- Total Selection Time:
  - RH spent significantly more time than others in both tasks (HO: $p = 0.023$, $p = 0.008$; VD: $p = 0.002$, $p = 0.002$), indicating more efficient selection in GP and GB across occlusions.

- Total Selection Count:
  - RH took significantly more selections in the HO task ($p < 0.001$), attributed to needing to move occluded objects, unlike GP and GB.

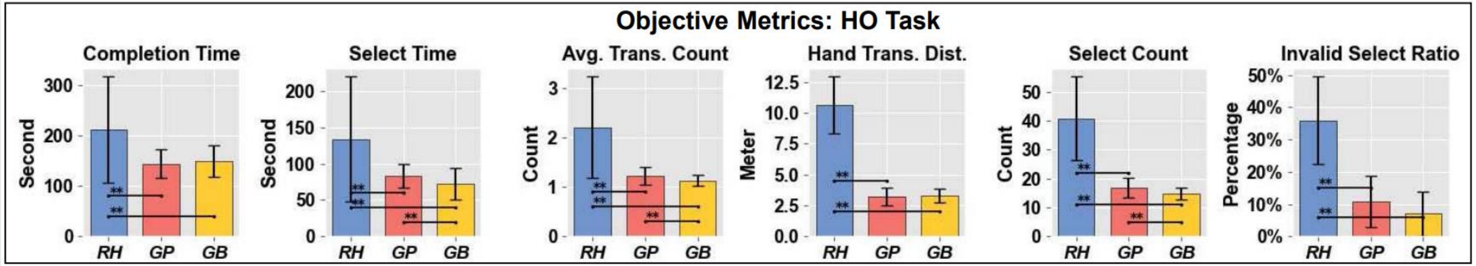
Study 1 – Results
1) Objective Measures(2)

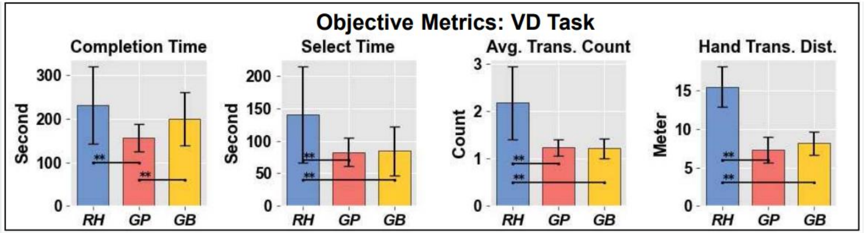
- Hand Translation Distance:
  - RH had significantly longer translation distance in both tasks (HO: $p < 0.001$, $p < 0.001$; VD: $p < 0.001$, $p < 0.001$), supporting GP and GB's advantage.

- Overall Completion Time:
  - RH took significantly more time in the HO task ($p = 0.01$, $p = 0.021$), indicating GP and GB's superiority in occlusion settings.

Study 1 – Results
1) Objective Measures(3)

- Conclusion:
  - GP and GB perform better in efficiency and effectiveness compared to RH, especially in occlusion tasks. However, GB may lag in translation demanding settings due to its multi-step design.





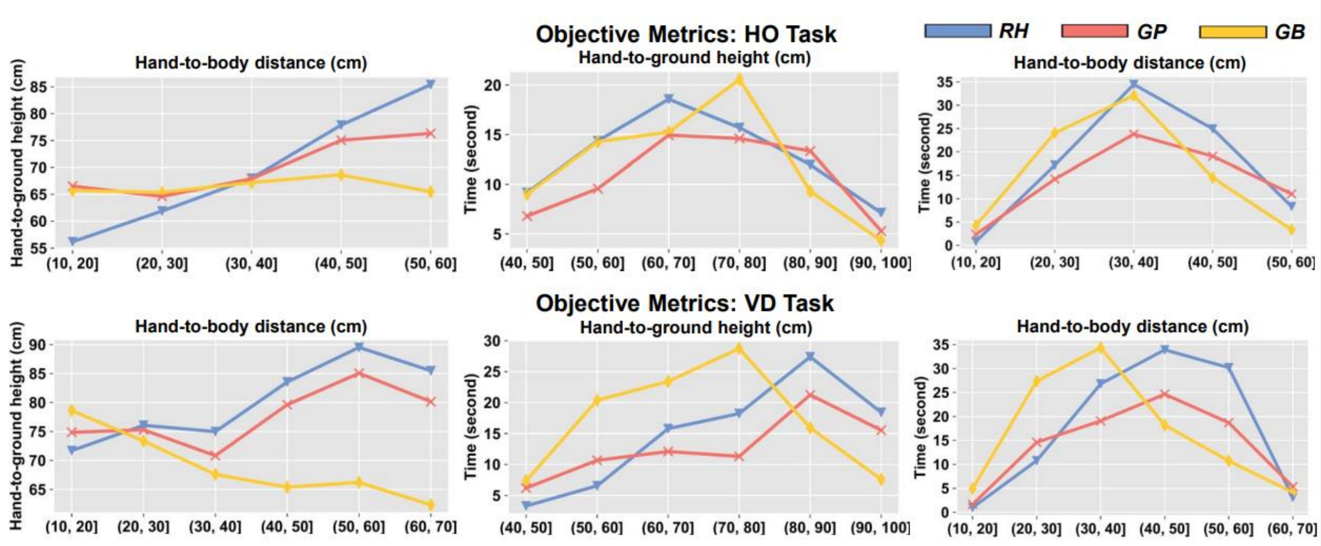
Study 1 – Results
2) Subjective Measures(1)

- **Physical Demand:**
  - Gaze Beam Guided Interaction showed significantly lower physical demand compared to other techniques (all $p < 0.001$), with GP also having lower demand than RH ($p = 0.0058$).
- **Effort:**
  - Users exerted less effort with all techniques (RH-GP, $p = 0.009$; RH-GB, $p = 0.002$; GP-GB, $p = 0.005$), showing a decreasing trend across the board.
- **Frustration:**
  - Similar to effort, frustration levels decreased for all techniques (RH-GP, $p = 0.004$; RH-GB, $p = 0.001$; GP-GB, $p = 0.01$).
- **Arm Fatigue:**
  - Significant differences were observed among techniques ($p < 0.001$), with GB causing the least arm fatigue.
- **User Preference:**
  - Participants preferred GB over the other two techniques (RH-GB, $p < 0.001$; GP-GB, $p = 0.005$) and GP over RH ($p < 0.001$).

Study 1 – Results

2) Subjective Measures(2)

- **Conclusion:**
  - Gaze Beam Guided Interaction significantly reduces physical demand, effort, and frustration levels compared to other techniques. It also minimizes arm fatigue, making it the preferred choice among participants.



Study 1 – Discussion

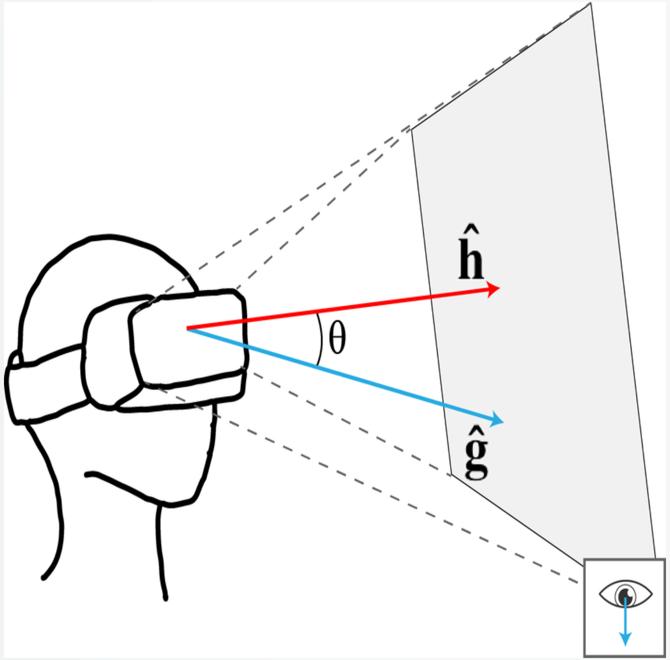
- **Participant Feedback:**
  - Remote Hand faced criticism for inaccuracy and arm fatigue, while GP and GB received positive remarks for accuracy, with GP noted for arm fatigue in distant translations and GB for its multi-step design.
- **Objective Arm Fatigue Measures:**
  - GB showed lower arm fatigue, supported by hand height and duration metrics, indicating its advantage in distant translations and simplified depth adjustment.
- **Key Findings:**
  - Gaze Position Guided Interaction and Gaze Beam Guided Interaction improved object selection efficiency, especially in occluded scenarios.
  - Gaze Beam Guided Interaction notably reduced arm fatigue during 3D translations, particularly in depth adjustments.

Study 2 - Overview

- **Objective**:
  - Evaluate the proposed Implicit Online Gaze Calibration;
- **Task**:
  - Participants complete a search and move task similar to the Heavy Occlusion task from study 1;
  - Interaction method: Gaze Beam Guided Interaction;
- **Data Collection**:
  - Record all Implicit Calibration Samples during the task.
- **Analysis:**
  - Compare the accuracy of online calibrated gaze estimation with standard methods after task completion;

Study 2 - Procedure

- **Task objective**:
  - Find colored spheres and move them to target positions;
- **Formal Task**:
  - Starts with a standard 9 point calibration;
  - 16 globes are randomly placed if front of the participant;
  - task is completed after moving all globes to target positions;
  - Implicit Calibration samples are recorded during the task;
- **Calibration and Testing:**
  - Personal coefficients derived from Implicit Calibration Samples;
  - Post-task : Participants do the 9 point calibration again;
  - Compare angular error between estimated gaze ray and ground truth gaze ray.



Study 2 - Results

- **Gaze Estimation Error:**
  - Results from participant P12 were excluded due to an unusually high error (16°);
  - **Error Comparision**:
    - Implicit Online Gaze Calibration (IOGC): **1.87°**
    - 9 points calibration: **2.57°**
    - Average Coefficients: **2.55°**

- **Participant Performance:**
    - IOGC performed better than both 9 pts and Avg. methods;
    - Glasses and eye appearance causes estimation errors around 1° to 6°;
    - IOGC achieved the best performance for 8 out of 11 participants;

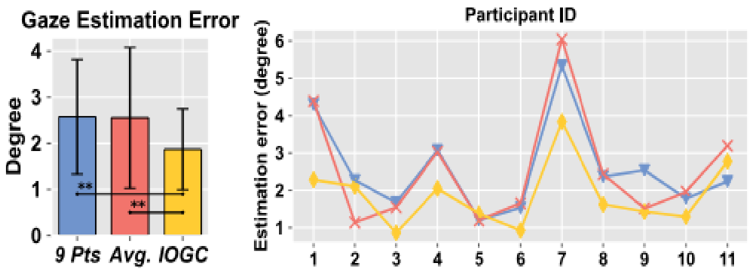


Figure 10: Comparison on average gaze estimation results of the three methods (left). Detailed results of 11 valid participants are shown on the right.

Conclusion

- **Study 1 findings**:
  - Gaze, combined with hand interaction, enhances selection efficiency even with multiple occlusions.
  - Reduces arm fatigue during depth translation tasks.

- **Implicit Online Gaze Calibration**:
  - Introduced a method that eliminates the need for explicit calibration;
  - Calibration is made by interacting with the environment;

- **Study 2 findings**:
  - Implicit calibration is better that two baseline calibration techniques;

Bibliography

- https://www.researchgate.net/profile/Rawan_Alghofaili/publication/332747757/figure/fig5/AS:756186420899846@1557300300966/Computing-the-gaze-angle.ppm
- https://www.robotlab.com/hubfs/VR-AR.jpeg
- https://cdn-dynmedia-1.microsoft.com/is/image/microsoftcorp/gldn-Hololens2-ImmHero:VP2-859x540
- https://www.cnet.com/a/img/resize/636e3a23602386fe52d7fc5bbb08f31568f9c315/hub/2021/06/19/65e980ba-f9c0-4eea-9b1b-00464d773488/htc-vive-pro-2.jpg?auto=webp&fit=crop&height=1200&width=1200
- https://ieeexplore.ieee.org/document/10108465