



Utrecht University

DEPARTMENT OF INFORMATION AND COMPUTING SCIENCES

MASTER THESIS

Mobile Augmented Reality Interaction Using Gestures via Pen Tracking

Author:
Jerry VAN ANGEREN

Supervisors:
Dr. W.O. HÜRST
Dr. ir. R.W. POPPE

ICA-3496570

April 10, 2015

Acknowledgements

First I would like to thank my supervisors Wolfgang Hürst and Ronald Poppe. Wolfgang, thank you for providing additional materials as well as inspiring and encouraging me throughout the entire project. Ronald, thank you for the feedback on improving the image processing performance and helping with the statistical analysis. I also like to thank my employer for supplying me with the tablet stands used during the user study. Next, I like to thank everyone who participated in the pilot or user study and provided useful feedback as well as suggestions to improve the created technique. Finally, I would like to thank my family and friends for supporting and encouraging me till the very end.

Contents

1	Introduction	3
2	Scientific paper	4
A	Mobile augmented reality game using gesture-based interactions	14
A.1	Used libraries	14
A.2	Performance	15
A.3	Revised implementation	16
B	Pilot study	19
B.1	Motivation	19
B.2	Experimental setup	20
B.3	Procedure	20
B.4	Measures	20
B.5	Results and Analysis	21
B.5.1	Transformation and drawing gestures	21
B.5.2	Questionnaire and Informal interview	22
B.5.3	Discussion	23
C	User study - Additional results	24
C.1	Performance	24
C.2	Questionnaire	25
D	Conclusions and future work	29

Chapter 1

Introduction

In this thesis, we evaluate a novel interaction technique created for mobile augmented reality. We look at a free-hand gesture-based technique to create new virtual objects or manipulate existing ones. To create gestures a pen is used. The camera of a device captures a frame to locate the pen within. A drawn gesture has to be evaluated to determine whether it corresponds to any of the implemented ones. When a correctly drawn gesture is recognized, the proper corresponding action will be executed. A possible implementation of our interaction technique is shown in a proof-of-concept game.

The main goal of this thesis is as follows:

Evaluate the feasibility and usability of free-hand gestures for mobile augmented reality interaction.

For the first part of this goal, evaluating the feasibility, we have to verify whether the interaction technique can be created and in which manner. To address the second part of our goal, evaluating the usability, we have to look at possible ways of usage and whether users can learn to interact with the application using our free-hand gesture-based technique. Since the interaction between user and application take place in mid-air, we evaluate if users perform better when they are provided with visualizations of their drawing paths within the application. Further, we evaluate to which extent it is enjoyable to use our proposed technique. Overall, we will look at the following sub-goals:

- Create a mobile augmented reality system to track a pen and recognize certain gestures.
- Develop an application that allows us to verify the feasibility and usefulness of this interaction technique.
- Verify and evaluate the proposed application and interaction design via user studies.

Chapter 2 presents our scientific paper which contains the most important details and findings. Additional details and results are presented in the appendices. Appendix A contains additional details regarding the developed proof-of-concept game and changes we made based on given feedback from the user studies. The setup and results of a pilot study are presented in Appendix B, followed by additional results from the user study in Appendix C. Finally, Appendix D presents recommendations for future work and concludes this report.

Chapter 2

Scientific paper

In this chapter, we present our scientific paper. It outlines the most important results and findings from this thesis research. In the paper, we describe the development of a proof-of-concept game that we used to implement our free-hand gesture-based interaction technique. We discuss related approach from previous work and evaluate our technique in a user study. This work serves as a basis for a revised version which has been submitted to the International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2015).

Free-hand Drawn Gesture-based Augmented Reality Interaction For Games

Jerry van Angeren

Utrecht University

Information and Computing Sciences

j.vanangeren1@students.uu.nl

Abstract—In this work, we present a novel tracking-based gesture interaction technique using a tablet’s camera to interact in an augmented reality setting. We have created a proof-of-concept game in which transformation or drawing gestures are used to manipulate existing virtual objects or create new ones respectively. We evaluate the feasibility and usability of our technique in a user study with 25 participants. Results show that users are able to perform the implemented gestures after having some training and enjoy using our interaction technique. The performance and subjective data proved the potential of our interaction technique when used for augmented reality games.

I. INTRODUCTION

Usage of augmented reality (AR) present many opportunities for entertainment. The major influence on the game experience is based on the used interaction technique. When a mobile device such as a tablet or smartphone is used to interact with AR, either the touch screen or the built in camera is used to track objects [1]. Touch screen interactions are faster and more accurate compared with tracking-based interaction. However, in case of an AR board game which combines both virtual and real board pieces, the interactions between them are in different locations which leads to different focus. For example, a real board piece is translated by picking it up and placing it somewhere else while the touch screen has to be used for virtual objects.

Using a tracking-based interaction approach could possibly solve this issue. A hand or pen can be tracked behind the device on which it is visible. Moving around on the game board to interact with both real board pieces and virtual ones, instead of using the touch screen improves the immersion. However, there are some limitations when a tracking-based approach is used. Currently, image processing is required which is computationally heavy as well as less robust and less accurate compared with usage of the touch screen. These limitations could result in a lower potential and possibly have a negative effect on the game experience.

In this paper, we investigate a tracking-based approach to interact with virtual objects inside an augmented reality game on a mobile device such as a tablet or smart phone. We focus on free-hand drawn gestures which are created using a common pen. Our goal is to evaluate whether users can handle this interaction technique and whether they enjoy the game experience using a created proof-of-concept game. In this game, gestures are used to create or manipulate objects and are detected from tracked locations, see Fig. 1. When a gesture is detected, the corresponding action occurs.

The remainder of this paper continues with an overview of related work in Section II. Section III describes our created proof-of-concept game. We present the setup of our user study in Section IV, followed by the results in Section V. Finally, in Section VI we summarize our major findings, future work and conclude our work.



Fig. 1: The created augmented reality game with virtual objects and gesture-based interaction using a colored pen.

II. RELATED WORK

We discuss several tracking approaches used for mobile interaction. While not all approaches are created specifically for augmented reality, we focus on their potential usage in this field.

Most tracking approaches make use of markers, which are easily detected within a captured image. The markers are used to obtain the position, orientation and scale of present virtual objects and to display them correctly [2]. Moving a marker can lead to a change in the object’s orientation or position. Another possibility is to attach a marker to a physical object. The Magic Paddle, for example, is a piece of cardboard with a marker [3]. Moving the paddle around allows the user to place or move virtual objects inside the created environment or delete existing ones from it. Using a marker is typically robust, but it requires the presence of one or more markers. Moreover, allowing more complex interactions such as scaling require additional input modes.

Instead of using a physical object, the hand can also be used to interact with virtual object as shown in the Tether

prototype [4]. Their proposed system relies on reflective markers placed on the user’s hand and head as well as on the device. Since these markers are tracked inside a motion capture lab, interactions can be performed both in front of the device and behind it using finger movements. The markers on the head are tracked relative to the device to enhance the displayed realism using correct perspective. Since their prototype requires all markers, the possible usage is limited and thus also its potential. A different system, FingARtips detects markers attached to the fingertips to interact with AR [5]. However, since it also requires the tracking of markers, the same limitations are noticeable in this system.

While the use of markers typically leads to robust tracking, marker-less approaches are more appealing. Lee et al. [6] present an approach in which the hand is tracked using the skin-color. Movement of the hand allows interactions with present virtual objects. Due to using the skin-color and the required processing, the system became slow and also led to inaccuracies which reduce the potential of this approach. A different system combines the touch screen with the device’s internal sensors [7]. Based on the motions and orientation of the device, a 3D motion vector is created and applied on the virtual object to modify it accordingly.

Tracking approaches suggest high potential when the accuracy and speed are good. Another relevant factor is ease of use and comfort while using the interaction approach, especially when it is also used for the creation of objects. Free drawing approaches require tracking of a finger or object over time. Teddy is a system in which different contours can be drawn to create new objects or manipulate existing ones [8]. However, their method seems highly depending on the used algorithm and does not allow an untrained user to easily create new objects or manipulate present ones. Since drawing proved to be difficult, especially in mid-air, a virtual grid was introduced [9]. While the drawing performance of straight lines improved, limitations are introduced since round shapes are not supported.

Instead of drawing shapes directly, it is also possible to interact with virtual objects using gestures and poses. Vision-Wand shows such an approach in which gestures and postures are made using a plastic rod [10]. Based on the detected gesture or posture, the user is able to move, scale or rotate virtual objects. Additional interactions are achieved by making use of a pie-menu in which the user can select the desired action.

An overview of possible techniques to recognize user’s gestures is discussed in a survey which mainly focusses on hand, arm, body, face and head gestures [11]. Better recognition results are obtained if the gesture recognition allows rotation, scale and position invariance. While gestures are commonly used, performing them in mid-air can introduce discomfort to the user, based on the duration and location where a certain gesture has to be performed [12]. For this reason, gestures need to be intuitive as well as detected fast and robust.

Providing feedback on the performed gesture could help the user. However, usually there is a lag between performing and recognizing gestures. To overcome this issue of lag, a system called Ripples [13] converts this lag into a design element while using the touch screen. A contact visualization frameworks provides feedback to the user related to possible

interactions and whether interactions are performed successfully. The system showed promising results with less errors and faster interactions. Based on this, we have included a trace visualization to provide feedback from the interactions to the user.

Based on these previous works, we came up with a free-hand gesture drawing approach in which the user draws 2D gestures instead of contours. We have developed a game in which objects can be manipulated and new ones can be created using different gestures. Since the drawing of objects seems intuitive using a pen, we make use of pen tracking in our interaction technique. Since a marker-less approach is more appealing, we track the tip of a colored pen which is held by the user.

III. GESTURE-BASED INTERACTION USED FOR CREATING AND MODIFYING OBJECTS

In this section, we present our created augmented reality game in which gestures are performed using pen tracking. Since pen tracking can be used for a variety of applications or games, the created game should be seen as a proof-of-concept game to evaluate the interaction approach.



Fig. 2: Screenshot of the created game environment.

A. Pen Tracking

To achieve a robust pen tracking, the color of the used pen needs to be different from the background. While the specific color does not matter, we have used a pen with a pink tip in the remainder of our study.

Performance is important since the pen tracking is performed on a mobile device, in our case a tablet. The device’s camera is used to capture a frame and depending on the resolution, we resize it. When the resolution is larger than 640×480 , the frame is resized to a size equal or smaller than this size by decreasing the resolution with the same factor for both width and height. After resizing, the frame is converted to the HSV color space. On the hue channel, which encodes the color, we apply a threshold using predefined minimal and maximal values. A threshold using minimal values is applied on the saturation and value channels to make sure that bright colors are found. In the created binary image, only the tip of the pen is white and all other parts are black.

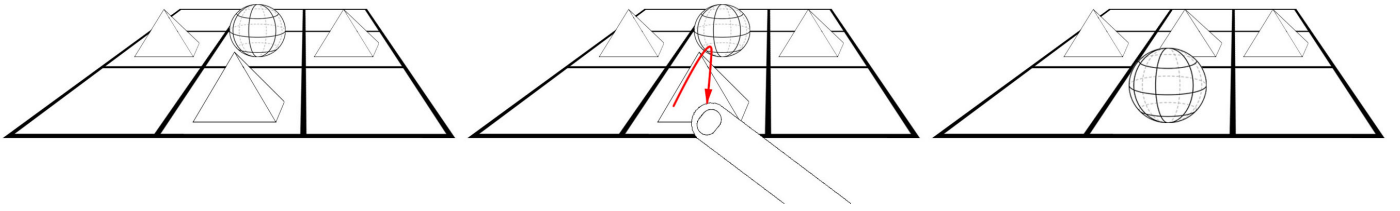


Fig. 3: Transformation gesture (center) which swaps two objects to create a set of three or more objects with the same shape.

To remove potential noise, we apply an erosion on the binary image. Finally, we search the largest contour and find the center using central moments. The found location is the center of the pen tip which will be used as input for the gesture drawing.

When a location is found in the previous frame, we limit the tracking to a rectangle around the last found location. This leads to a reduction of the required processing time and increases the frame rate with an average of 5 fps. Without a previous location, 12-50 milliseconds is required for the different processing steps, depending on the used device. Combined with all other processing steps, frame rates between 10 and 25 fps are found. Since a previous location is usually available, a frame rate of 15-30 fps is observed.

B. Augmented Reality Game

Playing with any kind of board game requires interaction with game pieces. In traditional board games, board pieces could be placed on the board or taken away from it. Virtual board pieces or characters, in augmented or virtual board games, may also be placed on the scene (creation), removed from it (deletion), or modified, for example by moving them or by transforming them in a way which may be impossible in the real world such as scaling. There are various ways to realize such creation or modification actions in virtual or augmented reality. In this research, we are investigating whether gestures are suitable for these basic interactions and in which way. We look at “drawing gestures” to create objects and “transformation gestures” to translate them.

To evaluate our interaction approach, we developed a game in which the interactions rely on pen tracking. The board game consists of a grid of tiles (3×3) shown on top of a detected marker, see Fig. 2. We use three basic shapes in the form of a sphere, a pyramid or a cube. Besides these shapes, a tile can also be empty. All tiles, including the empty ones, receive a randomly assigned texture; red, blue or orange. The goal of the game is to form a “matching group” in which three or more objects have the same shape, either in the same row or column. When such a group is found, the objects are removed from the board and new ones are added. Points are received for these matching groups and double points are given when all objects have the same color. At the beginning, random shapes are placed with randomly assigned colors.

When there are no more possible moves, the game ends. In this case, performing a drawing gesture or swapping two tiles does not lead to the formation of a matching group. The user can reset the board with new random objects to start a new game. This also leads to a reset of the obtained points,

only the high score is saved which is update when a higher score is obtained.

C. Implementation

The implementation makes use of several open-source libraries. Vuforia [14] is used to capture a frame and when it contains a highly textured marker, its position and orientation relative to the camera are determined. Based on this, a view matrix is created. The captured frame is passed to an image processing class which is used for the pen tracking and makes use of OpenCV [15] to detect the tip. Detected gestures are used to create new objects or manipulate existing ones. Finally, the board with the present objects is drawn based on the created view matrix using OpenGL ES 1.1 [16].

D. Gesture Interaction

To have a clear distinction between the manipulation and creation of objects, we use two different modes. In “switching mode”, two adjacent tiles can be switched while in “drawing mode” a new shape can be added on an empty tile. By default, the game starts in switching mode.

The found location of the pen tip is converted to coordinates of the grid first and tested against the locations of the tiles. Doing this, allows us to determine whether a tile should be selected. A selection occurs if the pen is held above a tile for at least 5 frames (approximately 0.15 – 0.25 seconds). When a tile is selected, a blue border is shown around it as shown in Fig. 2.

Swapping of two tile is done by first selecting one tile, followed by performing a transformation gesture. During this gesture, the user moves the pen from the selected tile to an adjacent tile in horizontal or vertical direction followed by moving it back towards the first tile. A move back towards the start location has to be performed to confirm the selection and prevent false interactions. When the second tile is not adjacent, the first one gets deselected. A schematic view of the transformation gesture is presented in Fig. 3. If two tiles are swapped and a matching group is created, the shapes are removed and the board is update. Otherwise, the swapped objects return back towards their starting locations. It is not allowed to swap empty tiles, only tiles which contain shapes can be moved.

To place a new object on the board, the user first selects an empty tile followed by a drawing gesture. The creation of a sphere, pyramid or cube is achieved by creating a circular, triangular or square gesture. An overview of all possible drawing gestures is shown in Fig. 4. Users are allowed to add

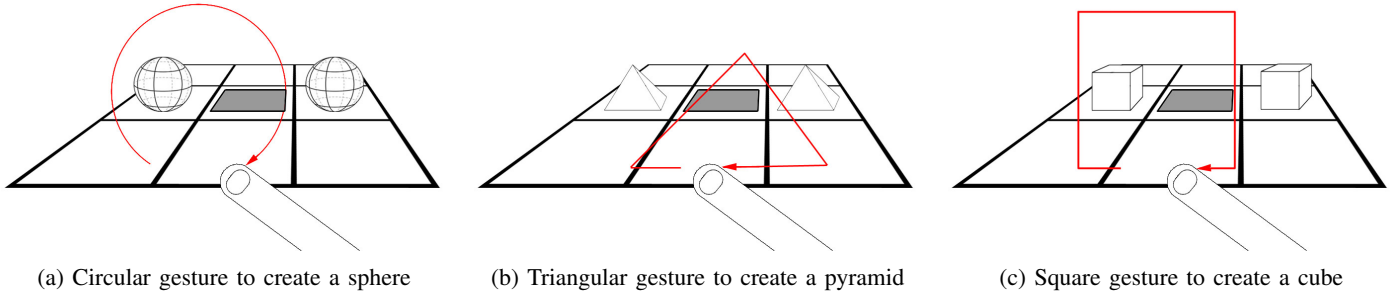


Fig. 4: Drawing gestures which can be used to create new virtual objects after an empty tile is selected and the application switched towards a drawing state.

any of the possible shapes since they could be used during a successive action.

The pen trace is created from tracked pen tip locations and is used to recognize the gestures. We limit the maximal length of the trace to 70 frames which is approximately 2 – 5 seconds. A longer trace allows slower movement but increases the change of false interactions and blocks parts of the board. On the other hand, a limited amount of points require fast drawing and thus skill.

If the trace crosses itself, the created closed shape is used for the gesture recognition. Based on the stored templates, a perfect shape for each gesture, a nearest neighbor matching occurs. To allow scale and position invariance, the drawn shape is normalized and matched with the dimensions of the stored templates. If the closest template is below a predefined threshold, a gesture is recognized, otherwise, the game returns to the switching mode.

When a gesture is recognized, the corresponding shape is placed on the selected empty tile in the color which was assigned to it. A test is performed to see if a matching group is formed and the board needs to be updated with new shapes or empty tiles. In case of no matching group, the game returns to the switching mode and the created object can be used to create a matching group.

IV. USER STUDY SETUP

To evaluate our proposed interaction technique, we set up a controlled user study. The goal is to evaluate both the feasibility and enjoyment. Objective data is captured to analyze the performance of the used gestures. Subjective data is used to analyze the user’s opinion and game experience.

The user study is split in two parts, both consisting of two tasks. To evaluate the influence of a visual trace, in one part visual feedback is shown while the other does not display the trace. In the first task of each part, a predefined set of transformation and drawing gestures had to be performed with a fixed amount of attempts. From captured data, we evaluate the feasibility and gesture recognition performance. During the second task, the user played the game for a couple of minutes followed by filling in a questionnaire to rate the game experience.

A. Experimental setup

Since our interaction technique uses gestures, the user needs to be able to perform these properly. Due to heavily image processing, usually a lag between movement and being visual on screen is noticed (approximately 17 – 32 milliseconds). However, since we require the processing some lag is unavoidable while it could influence the interaction behavior. We hypothesize that visual feedback on the tracked locations and thus the created gesture, could help the user’s performance. If for example the user want to draw a square gesture but the visualization shows a different shape, the user can alter their movement. We chose a trace visualization in which the tracked pen location is bright blue and previous detections fade to white with a decreasing width, see Fig. 2.

Since our experiment is split in two parts, both with and without visible trace, trace is a within-subject variable. We further hypothesized that a learning effect could occur by being more experienced due to playing for a longer period of time. On the other hand, we hypothesize that a shown visualization during the first part could improve the overall performance of the users. To evaluate this hypothesis, half of the users started with a visual trace while the other half started without it.

We also investigate the influence of the used device. This is achieved by using an Asus MeMOPad Smart tablet (10.1 inch display, 1280x800 pixels resolution, 5MP camera), a HTC Nexus 9 (8.9 inch display, 2048x1536 pixels resolution, 8MP camera) and an Asus Nexus 7 (2013) (7 inch display, 1920x1200 pixels resolution, 5MP camera). Participants were able to choose the used device. However, the Nexus 7 was only used when three participants were present due to the smaller size and longer processing time.

B. Participants

A total of 25 users (20 male, 5 female), at ages from 21 to 27 years (average 23.16 years) participated in our user study. Only 10 users had experience with AR to some extend. The majority of the users (23) were right-handed and none of them had any relevant knowledge about the subject or focus of this study when they started.

The sessions were conducted with up to three participants at the same time for practical reasons. When multiple users

were present at the same time, the different tasks were synchronized. Since the users had to play the game for a couple of minutes, we hoped for a competition between the participants and thus higher scores. Beside the competition, participants can discuss with each other which might result in more relevant feedback.

C. Procedure

To have equal conditions for all participants, the experiments took place on the same location with controlled external conditions. During the experiment, users are sitting comfortably on a chair in front of a table on which a highly textured marker is placed. In front of the marker, a tablet is placed on a stand which captures the marker and uses it to display the virtual objects properly. Gestures are created by moving a pen with a pink tip behind the device where it is captured by the device's camera.

An introduction talk explained the game and the used interaction techniques to the participant first. Then, all participants were assigned whether they started with or without having a visual trace. When multiple users were present, all of them had the same starting condition.

To make sure all users have the same preconditions, the participants first had to perform a tutorial. They were asked to perform the transformation gesture two times and all drawing gestures (circle, triangle and square) one time. For the drawing gestures, a maximum of five attempts could be used to create the desired shape. During the tutorial, no data was captured.

After that, the user had to perform the transformation gesture six times during which objective data is captured. To ensure that similar data is captured for all participants, we made use of a predefined set of objects. Using this set, only one successful transformation leads to a matching group. The next step focused on the creation of new objects using drawing gestures. Each shape had to be performed two times, leading to a total of nine drawing gestures. In this case, only one possible drawing gesture would form a matching group and no successful transformation gestures were possible. When the wrong shape was detected, the user had to perform the drawing gesture again. Again, a maximal of five attempts was allowed, otherwise the drawing gesture failed and the experiment continued.

Next, the users were asked to play a game with random objects for five minutes in which the goal was to reach a high score. However, the score was not displayed. This was done to prevent communication with other participants about it. Also, users could pay attention to the score which might influence the interaction and possibly the given ratings. When there was no possible move, the user could reset the board to start with new random objects. After these five minutes, the first part of the questionnaire was filled in.

After filling in the first part of the questionnaire, the users performed the same tasks using the other trace condition, either with or without trace. Again, the user performed the transformation and drawing gestures followed by playing the game for five minutes. Next, the remainder of the questionnaire is filled in. Finally, an informal interview was held to ask for a motivation about the given ratings and focus on the positives and negatives of the game and interaction technique.

D. Measures

To evaluate whether users can handle our interaction technique, we computed the percentage of correctly performed transformation and drawing gestures. These are obtained from the stored data since we knew which gesture had to be performed. Besides the percentage of correct performed gestures, the required frames for the gesture were stored as well. When less points were necessary, it is possible to conclude that a user became more experienced. In addition to this, a neutral observer took notes about special observations, interaction problems, etc.

Besides the objective data, subjective data is captured from a questionnaire. After each condition, they were asked to rate some statements regarding the game experience and level of control for both gestures. Since the drawing gestures consists of three different shapes, we also asked them about the level of control related to the individual shapes. After finishing the second trace condition and second part of the questionnaire, we asked them to compare the two trace conditions. Statements regarding the game experience and level of control were included in which they had to rate whether the experience became worse or improved. We also asked them if any discomfort was experienced during the experiment.

Finally, notes are taken from given feedback during the informal interviews. We are mainly interested in the user's opinion regarding positive aspects and possible improvement.

V. RESULTS AND DISCUSSION

The experiment took around 30–45 minutes per group, depending on the present participants. While the users had to choose the used device, the Nexus 7 is only used when three users were present at the same time. The reason for this was to benefit from the larger display and the faster processing time of the other devices. In total, the Nexus 7 was used 5 times, the Nexus 9, 7 times and the Asus MeMOPad Smart 13 times.

In this section, we discuss the captured objective and subjective data. To evaluate whether significant differences are found, we have used a paired t-test or repeated measures ANOVA with a confidence level $p = 0.05$. Since the user study consist of two parts, the differences between these two parts (part one and two) are evaluated to see whether a learning effect is noticeable. We also look at the presence of significant differences between with and without having a trace and the order of having and not having the visual trace. The performance of the transformation gestures are obtained by looking at made gestures which did not lead to a successful matching group combined with the correct performed ones. During the drawing gestures, we match the recognized shape with the desired shape since only one proper shape could be created. Also, a maximum of five attempts was allowed, otherwise the created shape failed.

A. Gesture performance: transformation gestures

Every participant finished all transformation gestures successfully. However, if we look at the performance measures in the first column of Table I, the percentage of correct transformation gestures is on average 67.07% which is rather low. This is especially caused by the low performance in the

TABLE I: Average percentage of correctly performed gestures over all participants, tasks and trials

	Translation gestures	Drawing gestures
Part 1	60.08	67.12
Part 2	74.06	76.53
Average	67.07	71.57
With trace	66.40	72.12
Without trace	67.74	71.01

first part of the user study in which the participants are using the interaction technique for the first time. Also, some users performed the transformation gesture on two objects which did not lead to a matching group. A wrong gesture could also be detected when the user moved the pen due to gripping. When we compare performance of the first with the second part of the user study, an improvement of 13.98% is shown.

Looking at the results with and without trace from the left column of Table I, the performance differences are small with a difference of 1.34%. We hypothesized that the presence of a visual trace might influence the performance and especially leads to faster learning when it is shown during the first part. Table II first column, shows the results of changing the order of having and not having a visual trace. To investigate whether a learning effect as well as an influence of the trace is noticeable, we perform a repeated measures ANOVA with part being repeated variable and trace order (trace in first or second part) as between-subject variable. Dependent variable is the average percentage of correct transformation gestures. A significant effect from the part is shown $F(1, 23) = 23.544, p < 0.001$ while no significant effect was found caused by the order or presence of a trace. While there is a significant improvement within the used order and thus a learning effect, the results between the two orders are rather similar with a difference of 3.67%. Overall, participants perform 2.9 – 4.4% better when they start with a visible trace instead of without. However, the main reason of a better performance seems to be caused by the presence of a learning effect.

TABLE II: Average percentage of correctly performed gestures when the order of with and without trace is altered. The top part starts without trace and is performed by 12 users while the bottom part starts with trace and is performed by 13 users.

	Translation gestures	Drawing gestures
Without trace	58.58	65.09
With trace	71.75	75.79
Average	65.16	70.15
	Translation gestures	Drawing gestures
With trace	61.46	69.03
Without trace	76.19	77.23
Average	68.83	72.90

B. Gesture performance: drawing gestures

The drawing gestures were not completed successfully by all participants. In total, one person failed the triangle and two failed the square drawing gesture. They all started without having a trace and successfully performed the gestures while

having a visual trace. Table I, right column, shows the average results from the drawing gestures and Table II, right column, shows the average percentage of correct gestures in which the order differs.

Again, we performed a repeated measures ANOVA with part and trace order as independent variables. Dependent variable is the average percentage of correct performed drawing gestures. Again, a significant improvement between the two parts is noticed $F(1, 23) = 4.867, p < 0.05$. An increase in performance of 9.41% is shown and thus the presence of a learning effect. If we look at the results with and without having a trace, no significant effect is shown since the difference is only 1.11%. The influence of the order did not proved to be significant either. Similar to the transformation gestures, the performance are rather similar with a better performance of 2.75% when starting with a visual trace. Also a better performance for the individual parts is shown when the user starts with having a trace which turned out to be 1.4 – 3.9% higher.

TABLE III: Average percentage of correctly performed drawing gestures over all participants, tasks and trials

	Circle	Triangle	Square
Part 1	89.29	69.01	52.17
Part 2	83.33	89.29	62.50
Average	86.21	77.95	56.98
With trace	92.59	75.76	56.82
Without trace	80.65	80.33	57.14

If we look at the individual shapes, see Table III, the best performance is shown for the circle and the worst performance while drawing a square. A performance improvement for the triangle and square is observed between the first and second part. This is especially noticeable for the triangle which improves with 20.28% during the second part. On the other hand, a decrease in performance of 5.96% is shown for the circular gesture. This is caused by the users who started with a visual trace, followed by not having the visualization. During the second part, over all 13 participants, a total of seven additional attempts were required compared to only one during the first part.

Similar results are shown when we look at the performance with and without having a trace. The circle proved to be 11.94% better while having a visual trace. Both other shapes showed similar results with differences of 4.57% or 0.32% for respectively the triangle and square.

To analyze the effect of the learning effect and trace, we performed a repeated measures ANOVA for each shape. Again, part and trace are the independent variables and the percentage correctly performed gestures was the dependent variable. None of the results proved to be significantly different. Looking at the largest difference, from the triangle, a marginal improvement is shown between the first and second part $F(1, 23) = 3.762, p = 0.065$. We notice better performance without trace for the triangle and square while the circle benefits from the trace. When we look at the difference in order, the trend of no significant differences is shown again which is similar with the results from Table II.

C. Different devices

Since we have used three different devices, the required frames to create the drawing gestures show a difference. The Nexus 7 and Asus MeMOPad Smart show rather similar results with averages of 145 and 169 frames. An average of 308 frames is required while using the Nexus 9, which is a lot higher. This can be explained by the fact of having better hardware and thus faster image processing. When the users moves the pen around, the Nexus 9 processes more captured frames which leads to the higher value.

TABLE IV: Average percentage of correctly performed gestures per device, over all participants, tasks and trials

	Transformation gestures	Drawing gestures
Nexus 7	73.84	70.24
Nexus 9	52.11	65.35
Asus MeMOPad Smart	69.47	77.11

On the other hand, the Nexus 9 showed the worst performance as presented in Table IV. This is especially caused by two participants who performed 12 and 15 incorrect translations and required more attempts during the drawing gestures as well. The best performances for the transformation gestures are shown when the Nexus 7 is used while the Asus MeMOPad Smart shows the best results for the drawing gestures. Overall, the performance of the used devices does not prove to be too different.

D. Questionnaire

The results from the questionnaire are used to evaluate the user experience during the performance of the gestures. We specifically focus on the game experience and level of control while having or not having a visual trace.

TABLE V: Average game experience over all participants

	With trace	Without trace
Overall	4.52	4.76
Transformation gestures	4.88	5.16
Drawing gestures	3.96	3.72

The average ratings on a scale between 1 and 7, are shown in Table V. Overall, users rated the game experience with an average of 4.64. If we look at the game experience with and without having a trace, a small difference of 0.24 is noticeable which is not significant $t(24) = 0.672, p = 0.508$. Similar results are shown for the transformation gesture and drawing gestures, with differences of respectively 0.28 and 0.24. However, the transformation gestures is slightly preferred without trace while the drawing gestures prefers a visual trace. Looking at the transformation and drawing gestures, again no significant effect from the trace is noticed. The users prefer making a translation gesture over a drawing gesture which is shown by a paired t-test on the averages for transformation and drawing gestures over both trace conditions $t(24) = 5.788, p < 0.001$.

If we look at the level of control during the different gestures this trend continues, see Table VI. When we combine the level of control during the transformation gesture with the

performance values, the observations of a limited influence from the trace are shown. A higher level of control is observed while performing the transformation gestures compared with the drawing gestures. The transformation gesture is rated significantly higher, as shown by a paired t-test $t(24) = 7.462, p < 0.001$. Based on these findings, we can conclude that the drawing gestures are found more difficult. Looking at the individual shapes, again no significant differences related to the trace are noticed. Differences between the different shapes are noticeable and show the best level of control for the circle and worst for the square. These values show similar results from the performance values in Table III since a lower level of control is likely to cause more mistakes and thus a worse performance. Having a higher level of control is likely to lead to a better performance due to less mistakes.

TABLE VI: Average level of control over all participants

	With trace	Without trace
Transformation gestures	5.48	5.28
Drawing gestures	3.56	3.24
Circle	4.76	4.92
Triangle	4.04	4.28
Square	3.28	2.92

E. Discomfort and informal interview

From the 25 participants, 8 experienced some discomfort while performing the user study. Most of them mentioned discomfort related to their arm, became tired or felt heavy. These results were noticed because they did not place their arm on the table but instead kept it unsupported in mid-air. The placement of the camera also caused these results, mostly with right-handed users. Especially the Nexus 9 and Nexus 7 showed this discomfort due to having the camera on the top left. This causes the user to have some difficulty with interactions on the left side when the device is placed in the tablet stand.

During the informal interview, most participants said that they enjoyed the proposed interaction technique. On the other hand, remarks about the trace being distracting were given frequently. This is mainly caused by the fact of having a long trace which fades slow. Due to having a long trace, a large part of the board is blocked and thus distracting while performing the transformation gestures. During the performance of the drawing gestures, users tend to watch the trace and thus move slower instead of focusing on the creation of the gesture with faster movements. Overall the majority of the users preferred the implementation without trace over having a visual trace.

False recognitions lead to frustrations for some participants. To overcome this issue, users suggested to add an undo button or even a pop-up dialog in which the recognized shaped needs to be confirmed. Also, they suggested to display a cursor on the last recognized location instead of having a visual trace. This allows the user to see the detected point while not having the blocked board or being distracted.

F. Discussion

If we combine the performance, results of the questionnaire and the feedback from the informal interview, the majority of

the users seemed to have enjoyed our interaction technique. Users are able to perform the gestures properly after having some training and improve over time.

Differences between the transformation and drawing gestures are found. Overall, the drawing gestures are rather hard to make, especially the performance of square gestures. This causes some frustration when the wrong shape is recognized. The level of control is the lowest for the square gesture and it might be preferred to include gestures which are easy to perform.

Our findings did not show a significant influence from a visualization in contrast with the presented results of the Ripples system [13]. However, users benefit from learning our proposed interaction technique when they first use it with a visualization. Feedback proved that the found location of the tip is helpful, especially since we use the center of the tip. On the other hand, the shown trace blocks parts of the board, which they found distracting.

While no significant influence from the trace is shown, differences in the obtained scores are found. On average, a score of 400 is obtained with a visual trace compared to 595 without trace. However, since the game is played using random objects and resetting the board also resets the score, it is hard to make a clear conclusion from these values.

The game experience and level of control per device are not evaluated due to the limited data for some devices. We expect that a limited lag between performance of a gesture and being shown on the display is preferred. This is achieved by having better hardware and we expect an increase in game experience and probably also increased level of control. However, future work will have to prove whether this is noticeable.

VI. CONCLUSION

Our research investigates a novel gesture-based augmented reality interaction approach. Using pen tracking, the user can create gestures to interact with virtual objects. To evaluate our interaction approach, we implemented a proof-of-concept game in which existing objects can be moved or new ones can be created using respectively transformation and drawing gestures. A user study is performed to evaluate both the performance of the gesture recognition and subjective ratings of game experience and level of control.

Participants were able to correctly perform the implemented gestures after having some training with an average performance of 70% correct. Since it is a new technique, users need to get used to it before they are able to properly perform the possible gestures. A significant learning effect is found which increased the recognition of the transformation and drawing gestures with respectively 13.98% and 9.41%. The increase in performance is shown within 30–45 minutes and from this, we expect a further increase when the interaction technique is used for a longer period of time. Looking at the individual drawing gestures, the circle, triangle and square are recognized with average accuracies of 86.21%, 77.95% and 56.98% respectively. From this, we can conclude that gestures with easy shapes lead to a better performance. Using a trace visualization did not show a significant effect on the performance.

The subjective data showed that the overall game experience was rated with a 4.64 on a scale between 1 and 7. Overall, they seemed to have enjoyed the interaction in which some frustration regarding false recognition is observed. Performing a translation gesture is preferred over the drawing gestures and resulted in a higher rating for both enjoyment and level of control. Results from the individual shapes show similar results as the performance, with the highest rating for the circular and lowest for the square gesture. Again, the used gestures are important to achieve higher enjoyment and better handling.

To aid the user and provide feedback, we have used a situation in which a trace visualization is shown and a different one without it. The results from the performance and subjective data did not show any significant influence caused by the displayed visualization. However, given feedback proved that most of the user did not like the trace since it blocks part of the view and is distracting. On the other hand, some kind of feedback is appreciated to see the tracked point and the current state of the application.

We are motivated by the positive reactions from the participants of the experiment. The analysis of the performance and given feedback helps us to improve the recognition accuracy as well as the game experience. Shown feedback can help users to improve the drawn gestures while it should not distract them during the gesture drawing. Since we tracked the center of the tip, feedback helps them to see the tracked location. On the other hand, improving the pen tracking by locating the extremity of the tip instead of the center will most likely improve the game experience and level of control. Since our proposed interaction technique is created with combining real game pieces and virtual ones in mind, a novel game should show the real advantages. Future work will have to show whether these improvements will improve the handling and enjoyment of this interaction technique.

REFERENCES

- [1] W. Hürst and C. van Wezel, "Multimodal interaction concepts for mobile augmented reality applications," in *Proceedings of the International Conference on Advances in Multimedia Modeling (MMM)*, 2011, pp. 157–167.
- [2] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top AR environment," in *Proceedings of the International Symposium on Augmented Reality (ISAR)*, 2000, pp. 111–119.
- [3] T. Kawashima, K. Imamoto, H. Kato, K. Tachibana, and M. Billinghurst, "Magic Paddle: A tangible augmented reality interface for object manipulation," in *Proceedings of the International Symposium on Mixed Reality (ISMR)*, 2001, pp. 194–195.
- [4] D. Lakatos, M. Blackshaw, A. Olwal, Z. Barryte, K. Perlin, and H. Ishii, "T(ether): Spatially-aware handhelds, gestures and proprioception for multi-user 3D modeling and animation," in *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*, 2014, pp. 90–93.
- [5] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn, "Fingartips: Gesture based direct manipulation in augmented reality," in *Proceedings of the International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia (GRAPHITE)*, 2004, pp. 212–221.
- [6] M. Lee, R. Green, and M. Billinghurst, "3d natural hand interaction for ar applications," in *Proceedings of the International Conference Image and Vision Computing New Zealand (IVCNZ)*, 2008, pp. 1–6.
- [7] T. Ha and W. Woo, "ARwand: Phone-based 3d object manipulation in augmented reality environment," in *Proceedings of the International Symposium on Ubiquitous Virtual Reality (ISUVR)*, 2011, pp. 44–47.

- [8] T. Igarashi, S. Matsuoka, and H. Tanaka, "Teddy: A sketching interface for 3D freeform design," in *Proceedings of the Conference on Computer graphics and Interactive Techniques (SIGGRAPH)*, 2007, pp. 409–416.
- [9] W. Hürst and J. Dekker, "Tracking-based interaction for object creation in mobile augmented reality," in *Proceedings of the ACM International Conference on Multimedia (MM)*, 2013, pp. 93–102.
- [10] X. Cao and R. Balakrishnan, "VisionWand: Interaction techniques for large displays using a passive wand tracked in 3D," in *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. ACM, 2003, pp. 173–182.
- [11] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, 2007.
- [12] D. Ahlström, K. Hasan, and P. Irani, "Are you comfortable doing that? Acceptance studies of around-device gestures in and for public settings," in *Proceedings of the International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI)*, 2014, pp. 193–202.
- [13] D. Wigdor, S. Williams, M. Cronin, R. Levy, K. White, M. Mazeev, and H. Benko, "Ripples: Utilizing per-contact visualizations to improve user interaction with touch displays," in *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, 2009, pp. 3–12.
- [14] "Vuforia developer portal," <https://developer.vuforia.com>.
- [15] G. Bradski, "The OpenCV library," *Doctor Dobbs Journal*, vol. 25, no. 11, pp. 120–126, 2000.
- [16] A. Munshi and J. Leech, "OpenGL ES common/common-lite profile specification, version 1.1.12," Khronos Group, Tech. Rep., 2008.

Appendix A

Mobile augmented reality game using gesture-based interactions

This appendix contains additional details regarding the developed proof-of-concept game. We elaborate the used libraries, some adjustments to improve the performance as well as improvements based on given feedback from the user study.

A.1 Used libraries

In addition to the details presented in the scientific paper (Section III-C, page 7), an overview of the used libraries is shown in Figure A.1. We have combined Vuforia, OpenCV and OpenGL ES with our own created code which takes care of updating the board and score, gesture recognition and additional required functionalities. The gray box represents our own created code.

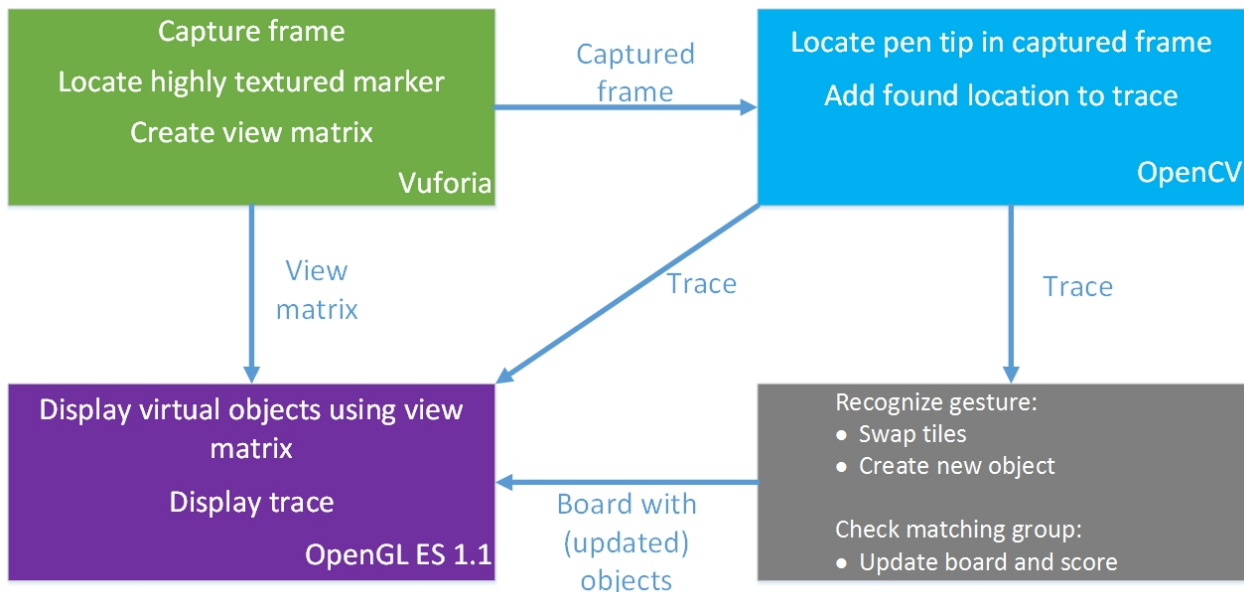


Figure A.1: Overview of the implementation and the used libraries.

For the augmented reality, we chose to use Vuforia version 3.0.9, the AR SDK from Qualcomm. This SDK is written in C/C++ and allows robust tracking of a highly textured marker even when it is partially occluded. Vuforia captures the camera frame first. In case of a present marker, the correct view matrix is computed using a default function from Vuforia.

To track the pen, we have to perform different image processing tasks. First, we convert the captured frame to the HSV (Hue, Saturation and Value) color space. The frame is captured in the RGB (Red, Green and Blue) color space which is influenced by lighting conditions. To overcome this, we use HSV which is illumination invariant. Next, we perform a threshold using predefined minimal and maximal HSV values corresponding to the pink cap. This operation results in a binary image in which pixels belonging to the pink cap are white and all others are black. The benefit of this operation is increasing the performance since following operations only involve one color channel instead of three. Next, to remove potential noise we perform an erosion. We then look for contours in the obtained binary image and search the largest which belongs to the pink cap. Finally, we find the center of the largest contour using central moments. This location is the tracked location and is used as input for the gesture drawing. We have performed these operations with OpenCV which contains many optimized functions to improve the performance. OpenCV is an open source computer vision library which is commonly used for implementations including image processing tasks in real-time. For our implementation, we have used the OpenCV4Android version 2.4.9.

Finally, to display everything, we have used OpenGL ES 1.1. The present objects are shown at the correct locations using the created view matrix from Vuforia. On top of the objects, the trace visualization is shown from the tracked locations.

A.2 Performance

Since our implementation is done on a mobile device, in our case a tablet, performance is an important factor. While we heavily rely on image processing tasks, the processing time needs to be as limited as possible to allow real-time interactions. Based on our (old) initial implementation we concluded that the performance had to be increased. A frame rate of 5 – 10 fps was found which does not allow real-time interactions. We therefore improved our implementation, leading to the frame rates shown in the scientific paper (15 – 30 fps).

An overview of the used image processing steps together with their processing time in milliseconds is presented in Table A.1. Looking at the individual steps, the major improvement is found by changing the noise reduction. In our initial implementation, an opening was used which is an image processing operation consisting of an erosion followed by a dilation. The improved version only makes use of an erosion and thus decreases the used time by 11.43 – 15.47msec for respectively the Asus MeMOPad Smart and Nexus 7.

In addition, we decreased the processing time of the Asus MeMOPad Smart by resizing the captured frame. The frame decreases from 1280×720 towards 640×360 pixels. While resizing requires on average 10.99msec, an improvement of 44.59msec is found in the remaining steps. A total improvement of 59% (33.60msec) is achieved for the Asus MeMOPad Smart. The Nexus 7 already captures

frames at 640×480 pixels, therefore we do not have to resize it. An increase in processing time is noticed for the contour operations due to having more objects in the captured frame. Overall, the performance for the Nexus 7 is improved by 33% (14.05msec).

Table A.1: Required processing time in msec per device. A performance increase is achieved by performing some steps different or by resizing the captured frame.

	Nexus 7 (old)	Nexus 7 (improved)	MeMOPad Smart (old)	MeMOPad Smart (improved)
Transform to HSV	9.93	8.98	22.06	5.15
Apply threshold	8.29	8.04	18.25	3.71
Noise reduction	23.10	7.63	12.96	1.53
Locate region of interest	0.01	0.01	0.01	0.01
Find and sort contours	1.43	3.99	3.46	1.84
Moments	0.04	0.09	0.10	0.07
Locate pen tip	0.03	0.05	0.08	0.03
Resize captured frame	-	-	-	10.99
Total	42.83	28.78	56.94	23.34

A.3 Revised implementation

Based on feedback from the user study, we have revised our implementation. In addition, some minor adjustments are included to make the game look more appealing. The revised implementation serves as a demo to illustrate the potential and usefulness of our interaction approach.

When the game starts, the user selects which hand will hold the pen. Based on the pressed button (left or right hand), the tracking algorithm is altered to find the tip either on the right or left side of the frame. Next, a tutorial is presented to show the different gestures which includes animations to clarify the used ones. If a user is familiar with the gestures, the tutorial can be skipped to play the game immediately. When the tutorial is finished, an options menu is shown in which the user can select the displayed visualization, a visible cursor or a trace respectively. We have included different levels to make the game more challenging. The first levels only allow transformation gestures while later ones will include both transformation and drawing gestures. Figure A.2 shows a flowchart from our developed game. During the user study no scores were displayed to prevent possible distraction. In our revised implementation, the obtained score is displayed to the user together with the progress of the current level which is shown by a circular bar. The current level is displayed as well.

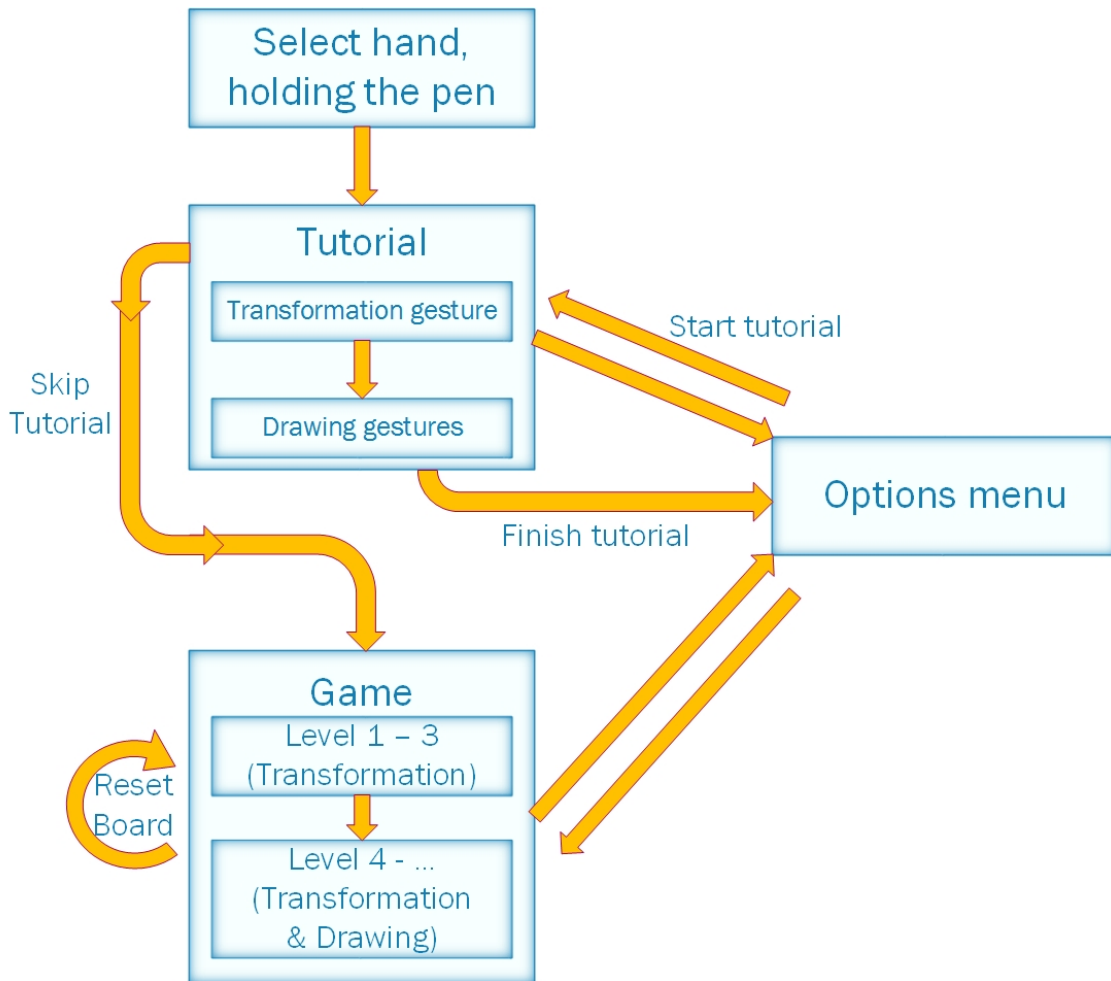


Figure A.2: Flowchart of created game.

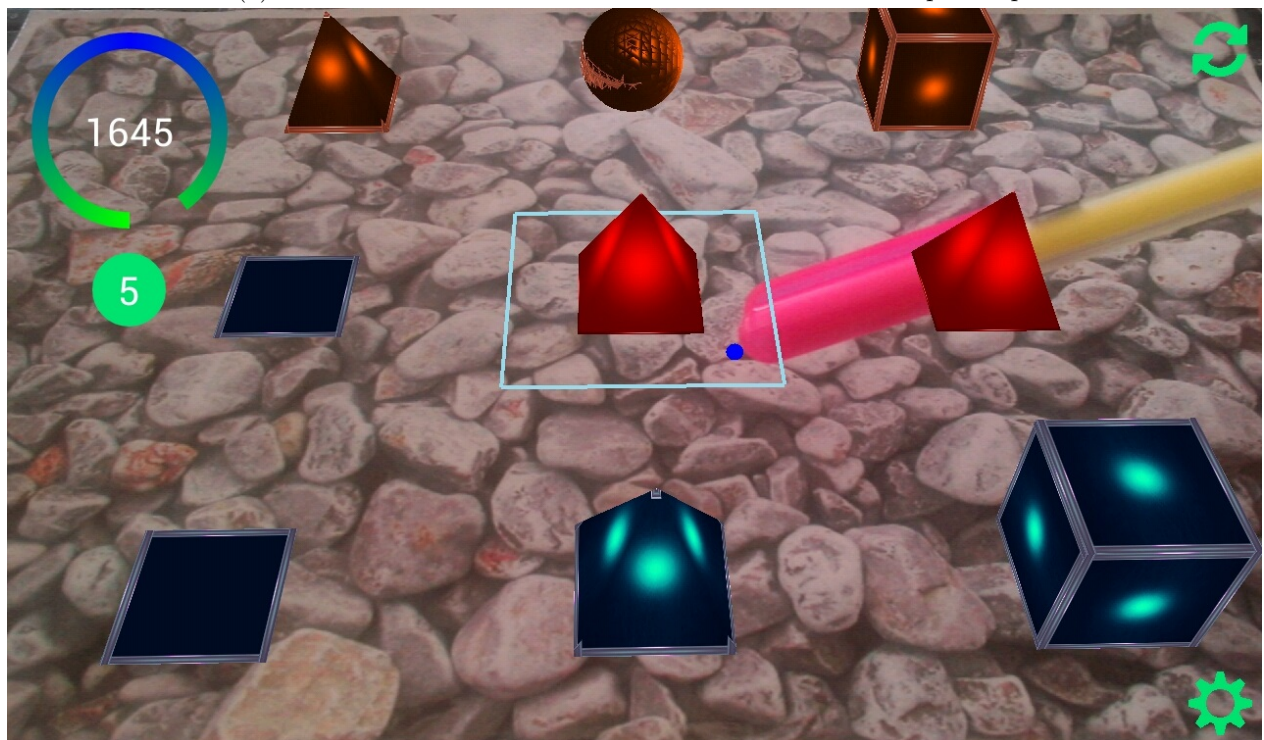
Besides these changes, the following aspects are changed based on given feedback from the user study:

- Track tip of pen instead of center of colored cap
- “Mixed mode”:
 - *Transformation gesture*: Show only a cursor instead of the trace visualization
 - *Drawing gesture*: Show a trace during tutorial and only a cursor while playing the levels

A comparison of the old and new visualization is shown in Figure A.3. The old version is shown in Figure A.3a, in which the center of the pink cap is tracked and a trace is shown which fades over time. Figure A.3b presents the revised version in which the tip of the pen is tracked and only a “cursor” is shown instead of the trace. Having a trace visualization instead of a cursor is possible by changing settings in the options menu. However, the trace will start at the tip of the pen instead of at the pink cap’s center.



(a) Old visualization: A trace which starts in the center of pink cap.



(b) New visualization: Only a cursor at the tip of the used pen.

Figure A.3: Different pen tracking and visualization styles.

Appendix B

Pilot study

This appendix contains details from the performed pilot study. We first look at the motivation of this study followed by the setup which covers the procedure, used devices and taken measures. Next, we look at the results and analysis of the captured data and given feedback.

B.1 Motivation

To evaluate whether users are able to use our proposed interaction technique, we performed a small informal pilot study. The main goal was to test whether users were able to interact with the application under different conditions and whether they enjoyed the game experience. Moreover, we wanted to evaluate the potential and possible issues or limitations of our technique since it was only tested by us. Since we performed a small feasibility study, 5 users (ages 23 - 27) participated. Most of them were students from the computer science program or had interests in this area.



Figure B.1: Overview of the setup during the pilot study.

B.2 Experimental setup

During the experiments, users were sitting comfortably in front of a table on which a highly textured marker was placed. An overview of the experimental setup is shown in figure B.1. A text marker with a pink cap was tracked by the device’s camera and used to create the different gestures. To evaluate the influence of visual feedback on the user’s interaction behavior, a trace consisting of maximal 40 visible points was displayed. Since the device might have an influence as well, we have used an Asus MeMOPad Smart (10.1 inch display, 1280×800 pixels resolution, 5MP camera) and an Asus Nexus 7 (2013) (7 inch display, 1920×1200 pixels resolution, 5MP camera). We also evaluate the influence of holding the device or placing it on a tablet stand.

B.3 Procedure

An introduction talk explained the game and the used interaction techniques to the participant first. Throughout the experiment, a set of predefined objects was used. This assures similar conditions for all participants in which only one possible transformation or drawing gesture leads to forming a group of three similar shaped objects. Performing a gesture on different objects was possible but did not lead to a matching group and thus not being able to proceed with the next task.

The experiments were split into two steps to focus on the transformation and drawing gestures. In the first step, users could only perform a transformation gesture to swap two items and in total nine actions had to be performed. The second step focused on the creation of objects using drawing gestures. Each shape (circle, triangle and square) had to be performed three times, also leading to a total of nine actions. In this case, only one possible drawing gesture could be used to form a matching group and no successful transformation gestures were possible.

To test our interaction method under different conditions, we want to evaluate the influence of a visible trace visualization, using a stand and using a smaller device. The following conditions were performed on the Asus MeMOPad Smart. First, the tablet was placed on a stand and a visible trace visualization was shown followed by a condition in which the tablet again placed on the stand but without having a trace visualization. Combining the interaction behavior with the user’s remarks and opinion allows us to evaluate whether an influence is noticeable from using a trace visualization.

Next, the trace was shown again but the user was not allowed to use the tablet stand. This condition is used to evaluate the influence of a tablet stand combined with possibly experienced discomfort. Finally, to test the influence of a smaller device, the user had to perform the two steps on the Nexus 7. Since the weight of the Nexus 7 is less than the Asus MeMOPad Smart, the users were not allowed to use the stand and thus had to hold the device. Also, a visible trace was shown again.

B.4 Measures

Notes were taken related to the recognized shapes, interaction technique, interaction problems, etc. Since our goal was to have an informal feasibility evaluation, further objective data was not captured during the pilot study. Subjective data was obtained by using a small questionnaire in which we

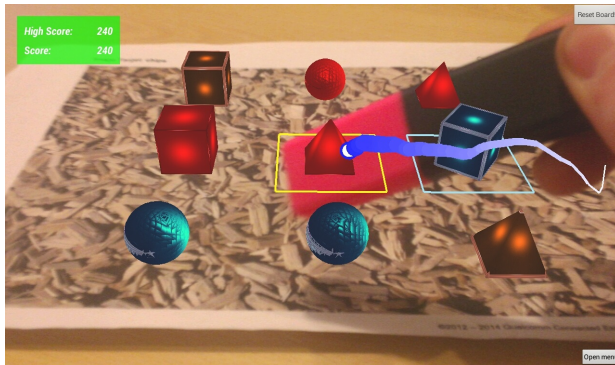
focus on the ease of use, level of control and usability. Finally, an informal interview was held with each of the participants to gather feedback about our interaction technique. Moreover, they were asked for positive and negative aspects to improve the implementation.

B.5 Results and Analysis

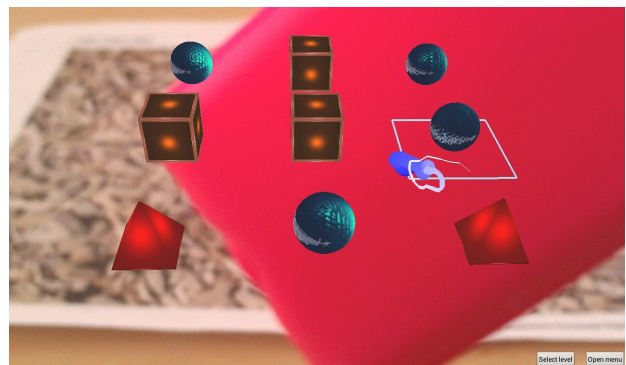
In this section, we discuss the results and analysis based on the feedback and notes taken during the experiments. We also discuss the major findings from the informal interviews with the participants.

B.5.1 Transformation and drawing gestures

All participants succeeded the transformation gesture tasks which some found more difficult than others. Since none of the participants ever used a similar interaction technique, an often heard comment was “I have to get used to it”. A comparison of two different interaction styles is shown in Figure B.2. The technique from Figure B.2a allows more precise interactions. Small pen movements lead to minor changes of the shown trace. Having the marker closer towards the camera results in a bigger visible area, see Figure B.2b. In this case, small movements of the pen lead to large changes of the visual trace. This also changes the interaction behavior. Moreover, the paper marker becomes occluded. If the text marker is held even closer to the camera, it cannot be found any more which leads to the disappearance of the virtual objects. Most users got used to the interaction method by increasing the distance from the camera while others continued to use the small distance.



(a) Text marker at a distance from camera.



(b) Text marker close to camera.

Figure B.2: Different ways of holding the text marker in front of the camera.

The drawing gestures were found challenging at first, especially when a square had to be created. A circle was the easiest shape and a triangle was also rather easy. Remarks like “nearly impossible” and “frustrating” were mentioned related to the drawing of a square in the first attempt. Since the shape needs to be drawn with only 40 points, one participant called it a vicious circle since the last points already disappeared without having a closed shape. This results in drawing the same part over and over again without having a successful shape which is recognized. However, after some attempts, they did all succeed to draw the correct shapes. Some users got confused when an incorrect recognition occurred. This was especially noticeable when the created shape looked rather perfect but a square was recognized as a circle due to having round corners. These false recognition also caused the experienced frustration.

Not having the visual trace “scared” some users at first but when they did use it for a couple of seconds, their opinion changed and they were able to interact properly. They had to get used to not having a trace visualization and test on which point the text marker was being tracked. For the transformation gestures, no real changes in moving behavior were noticeable and they all succeeded every task. A change in behavior and drawing abilities was shown during the creation of new objects, using the drawing gestures. Overall, the shapes were drawn better which caused better recognitions to occur. Some created much larger shapes while others moved way faster compared to the situation with a visual trace. Since there was no trace, the users did not see what they drew so the issue with false recognitions was not noticeable that much.

When the users were not allowed to use a stand, they first had to find a way to hold the tablet. Often heard comments included “It is hard to keep the tablet stable” and “It is shaking very much”. Every user was able to perform the transformation gesture tasks correctly and no real changes in moving behavior was noticeable. The drawing gestures did not show much differences either. However, this could be explained by the fact of being more experience. At this point, the participants have used the interaction technique for a longer period of time. As seen before, some confusion was noticed since the trace was visible during this condition. A smaller device (Nexus 7) led to similar results as shown while holding the Asus MeMOPad Smart. However, it was easier to hold the tablet for a longer period of time due to the relatively light device. Again no real changes in moving and gesture creating behavior was being observed.

B.5.2 Questionnaire and Informal interview

After the experiment, we asked the participants to fill in a questionnaire and performed an informal interview with them to ask their opinion about the created interaction technique. Basic information about the user’s age and experience with AR was taken from the questionnaire. Only two users had some experience with AR while one did not know what AR was. In contrast with some of the remarks during the experiment, the opinions were very optimistic and enthusiastic since it was a new and fun way to play a game. Because of this, remarks like “a new experience” and “cool” were mentioned. One person mentioned: “Nice to see what I have created”, after drawing a certain gesture and added: “Touchscreen interactions would not lead to the same feeling”.

The questionnaire and given feedback showed that the performance of a transformation gestures was found rather easy. This is noticed from comments like “Easy to understand without having a lot of experience”. Drawing gestures were found more difficult and especially the square gesture. However, everyone agreed that they did feel more in control after using it for a longer period of time and found the gesture interaction approach a suitable way to interact with AR. The drawing gestures benefit from not having a trace visualization which resulted in better recognitions and less attempts to create the desired shape.

Since our current implementation requires a paper marker, most participants did not see many people using our interaction technique in normal day life. After explaining them a possible application in which, for example, a newspaper contains a unique marker every day, most of them did see more potential. Related to the drawing gestures, the most often heard negative remark was “The trace fades too fast”. During the user study a maximal amount of 40 points was used and since some users were moving rather slow, this is limited. Increasing the number of traced points allows

slower movements and the drawing gestures will benefit from it. However, while it is an easy fix to have more points, we did not change the used points to have equal conditions for every participant. Also remarks about latency, between moving the pen and being visible on screen, were mentioned by multiple users. This is an issue which depends on the used device and required processing tasks. One person suggested using a device with a better camera and processor to have more fluent motions which could improve the usability.

B.5.3 Discussion

The study showed that all users were able to handle our interaction technique in a proper way after having some experience. Transformation gestures are rather easy and drawing gestures depend highly on the desired shape. Looking at the drawing gestures, squares are rather difficult to draw due to the limited amount of points which can be used. On the other hand, circles and triangles are rather easy to draw. Based on these findings, we decided to include more points during the user study. However, the maximal points has to be limited to prevent accidental crossings of the trace which cause incorrect recognitions.

During the pilot study, everyone started with a visual trace, followed by not having the visualization. The influence of a trace is noticeable and not having a visible trace leads to easier interactions and better recognitions. However, the used order might have an influence on the interaction behavior since users are more experienced during the second trace condition. Changing the order might result in different observations. Because of this, the user study included the different orders to evaluate whether an influence is noticeable.

Not using a stand or using a smaller device does not lead to much differences in interaction behavior. On the other hand, since the user study was rather short, an influence of the “novelty factor” has to be considered with the optimistic remarks. Conducting a user study in which users play the game over a longer period of time might result in less optimistic results. For this reason, the user study consisted of playing a game for a couple of minutes to evaluate this factor. Moreover, we did not want to focus on the discomfort caused by holding the device, therefore the tablet was placed on a stand during the user study.

We can also conclude that not all pens or text markers are suitable for this interaction method. Since we have used a big text marker, a large part of the paper marker might be blocked. Especially when the user holds the marker at a close distance from the camera. Based on these findings, a smaller pen was used during the user study while still having a pink cap which is tracked. Overall, the setup of the pilot study was good and therefore used again during the user study.

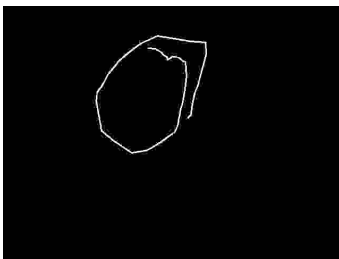
Appendix C

User study - Additional results

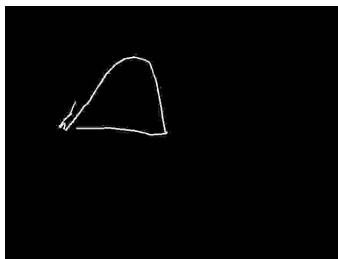
The most important results are presented in the scientific paper (pages 9 – 12). This appendix contains additional results related to the performance and questionnaire. We specifically focus on the differences from changing the order, starting with or without trace respectively.

C.1 Performance

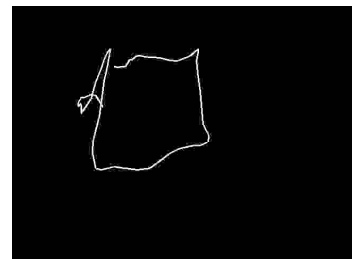
Based on the logged data, we were able to reconstruct drawn gestures from the participants. Typical results from the drawing gestures are shown in Figure C.1. Since we used a board with predefined objects, we were able to compare the detected shape with the desired one. This was used to calculate the percentage of correct performed gestures. Figures C.1a, C.1b and C.1c show correct recognitions of the desired shapes. On the other hand, recognized shapes which did not match the desired ones are shown in Figures C.1d, C.1e and C.1f.



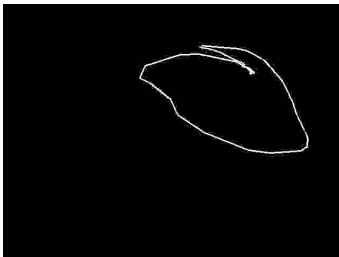
(a) Correct circular gesture.



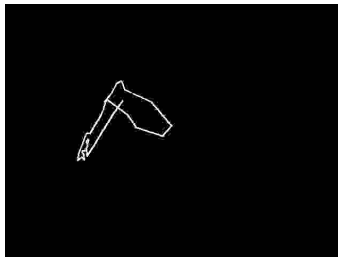
(b) Correct triangular gesture.



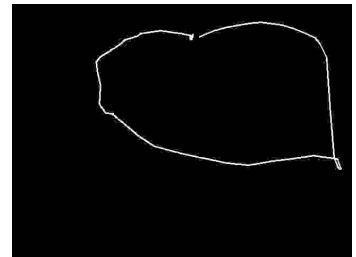
(c) Correct square gesture.



(d) Incorrect circular gesture.



(e) Incorrect triangular gesture.



(f) Incorrect square gesture.

Figure C.1: Typical drawing gestures seen during the user study.

The presented repeated measures ANOVA tests are performed by one of the supervisors. However, if we perform a paired t-test on the data, similar results are shown. Looking at the performances between the first and second part for the transformation gestures a significant difference is shown by the repeated measures ANOVA ($F(1, 23) = 23.544, p < 0.001$). When we perform a paired t-test on these values, ($t(24) = 4.965, p < 0.001$) also a significant improvement is shown. Similar results are shown when we look at the results with and without trace which did not show a significant difference ($t(24) = 0.334, p = 0.741$).

These observations continue when we look at the drawing gestures. A significant difference was shown between the first and second part with a repeated measure ANOVA ($F(1, 23) = 4.867, p < 0.05$) and also from a paired t-test ($t(24) = 2.239, p < 0.05$). Again, no significant difference is shown between the two trace conditions, with and without respectively, ($t(24) = 0.198, p = 0.844$).

Table C.1: Average percentage of correctly performed gestures when the order of with and without trace is altered. The top part starts without trace and is performed by 12 users while the bottom part starts with trace and is performed by 13 users.

	Circle	Triangle	Square
Without trace	82.76	69.70	50.00
With trace	88.89	85.71	60.00
Average	85.71	77.05	54.76
	Circle	Triangle	Square
With trace	96.30	68.42	54.17
Without trace	78.79	92.86	65.00
Average	86.67	78.79	59.09

Looking at the differences of starting with or without trace, the average results for the individual shapes do not show much differences, see Table C.1. Overall, starting with visual trace leads to a better performance of 0.96 – 4.33%. Despite a difference is shown within each shape, none of them is significant. If we perform a paired t-test on the largest difference, the triangle starting with trace, only a marginal learning effect is found ($t(12) = 1.954, p = 0.074$).

C.2 Questionnaire

Starting without instead of with trace results in a higher overall game experience as well as for the implemented gestures, see Figure C.2. Looking at the results when the first part is performed without trace (shown in orange), only a higher enjoyment is noticed while performing the drawing gestures. On the other hand, starting with trace (shown in green), a slightly higher enjoyment is observed when no visual trace is shown during the second part. The largest differences are found from the drawing gestures in which improvements of 1.08 and 0.54 are shown for starting without and with trace respectively. Overall, the transformation gesture is preferred over the drawing gesture

and especially without having a trace visualization. This can be explained by the fact of being easier to perform as mentioned by the majority of the participants.

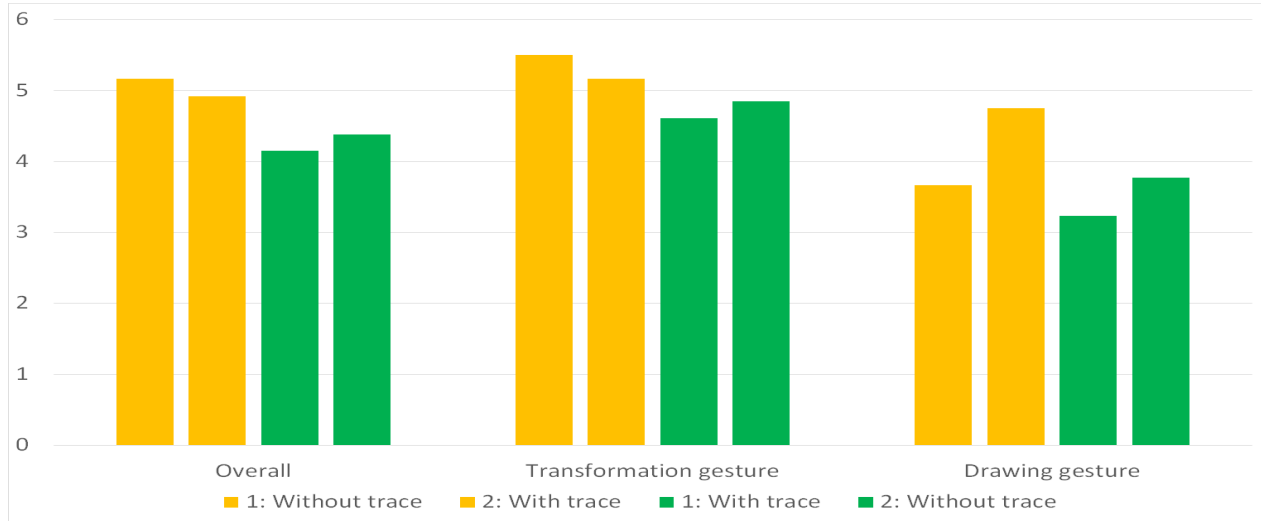


Figure C.2: Average game experience when the order of with and without trace is altered. Values over all participants who performed the given order, starting without or with trace respectively.

Looking at the experienced level of control, not many differences are shown for the transformation gestures, see Figure C.3. On the other hand, regardless of the used order, the drawing gestures showed improvements in the second part with values of 1.34 and 0.61 for starting without and with trace respectively. If we combine these values with the game experience, higher enjoyment and better level of control is noticed when users are more experienced during the second part. However, a higher level of control is shown when the user starts without trace. A difference of 0.31 is shown during the first part and 1.04 for the second.

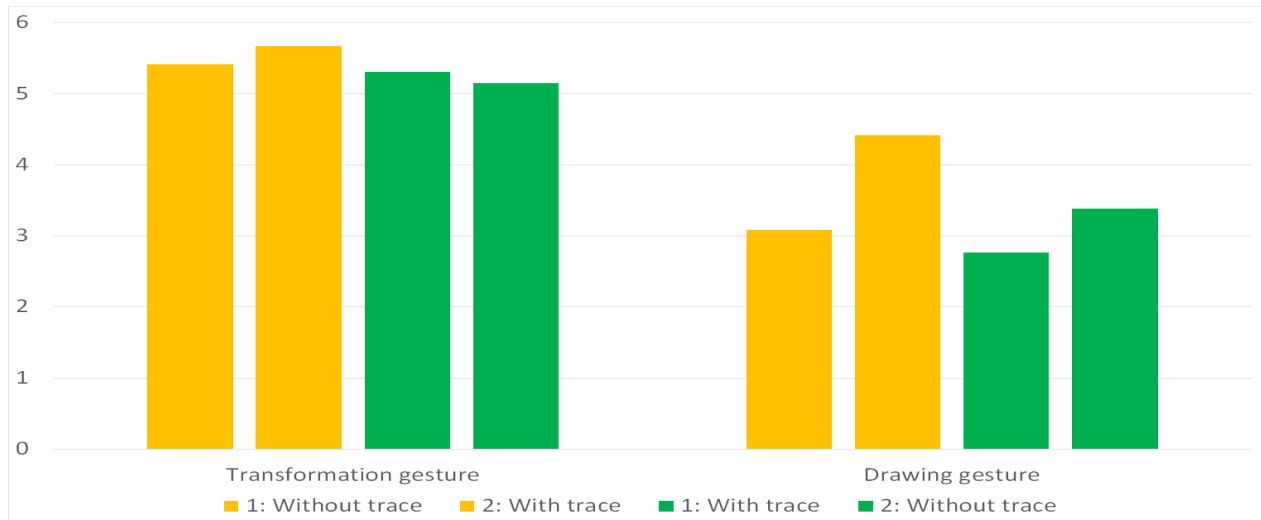


Figure C.3: Average level of control when the order of with and without trace is altered. Values over all participants who performed the given order, starting without or with trace respectively.

We also asked the participants to rate the experienced level of control for each shape and to rate whether the created shape looked like the desired one (perfectness), see Figure C.4. Overall, the

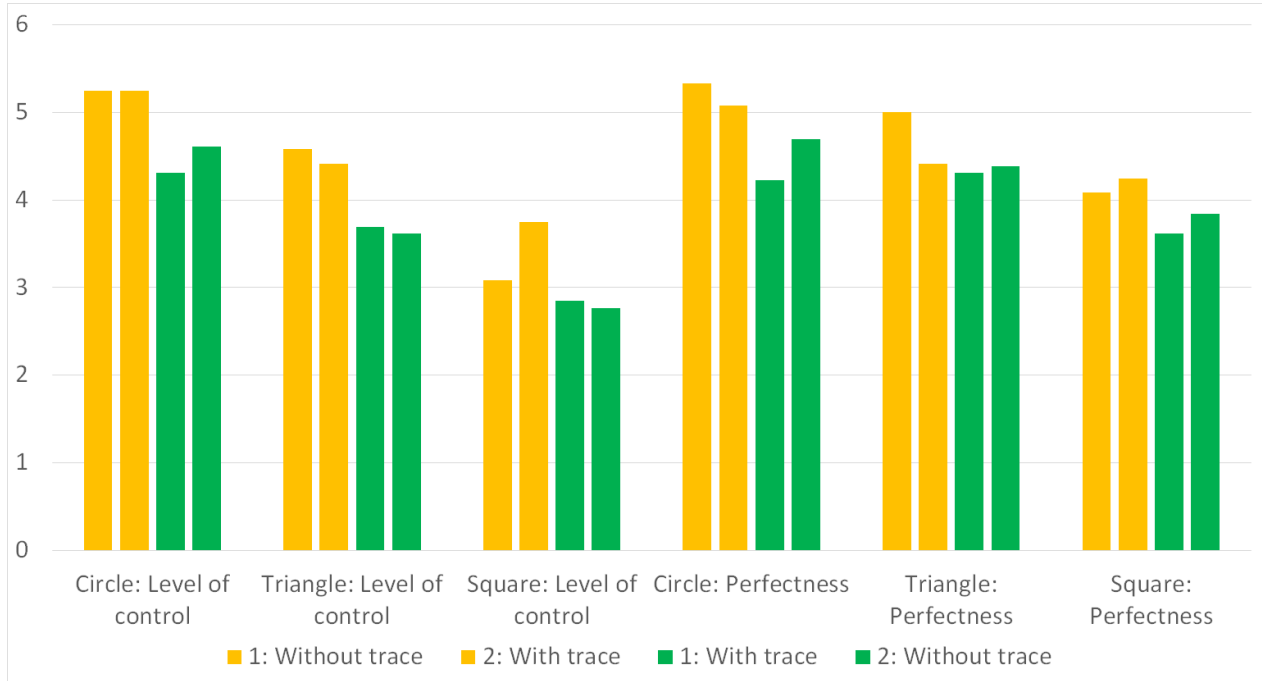


Figure C.4: Average level of control and shape perfectness for all individual drawing gesture shapes, when the order of with and without trace is altered. Values over all participants who performed the given order, starting without or with trace respectively.

ratings proved to be higher when a participant started without having a trace which continues the trends from the game experience and level of control. Limited differences are found from the level of control for each shape. However, a slight preference for a visible trace is noticed. Looking at the subjective perfectness of each shape, a slight preference without trace is noticeable for the circle and triangle. No clear preference was shown for the square gesture but a slightly better shape is created during the second part in which the user is more experienced.

Finally, we asked the participants to compare both trace conditions and rate whether different factors of the interaction technique became worse or better, see Figure C.5. Regardless of the order, an improvement is found during the second part for most of the aspects. Especially those who started without trace rated the second part higher compared with the first. The largest improvement is found for the experienced level of control during the drawing gestures. Based on the values, we can conclude that these users slightly prefer an implementation with visual trace. This confirms the results of the game experience and level of control for the individual shapes. However, these results are in contrast with the feedback from the informal interview which showed a preference of not having a visual trace. Since there is no clear preference for with or without trace, the ratings could be influenced due to being more experienced which leads to easier interactions. Looking at the participants who started with trace, the level of control during the drawing gesture and especially the square gesture became slightly worse. Half of them rated an improvement during the second part while the other half rated a decrease. Therefore, no clear explanation is found for this observation.

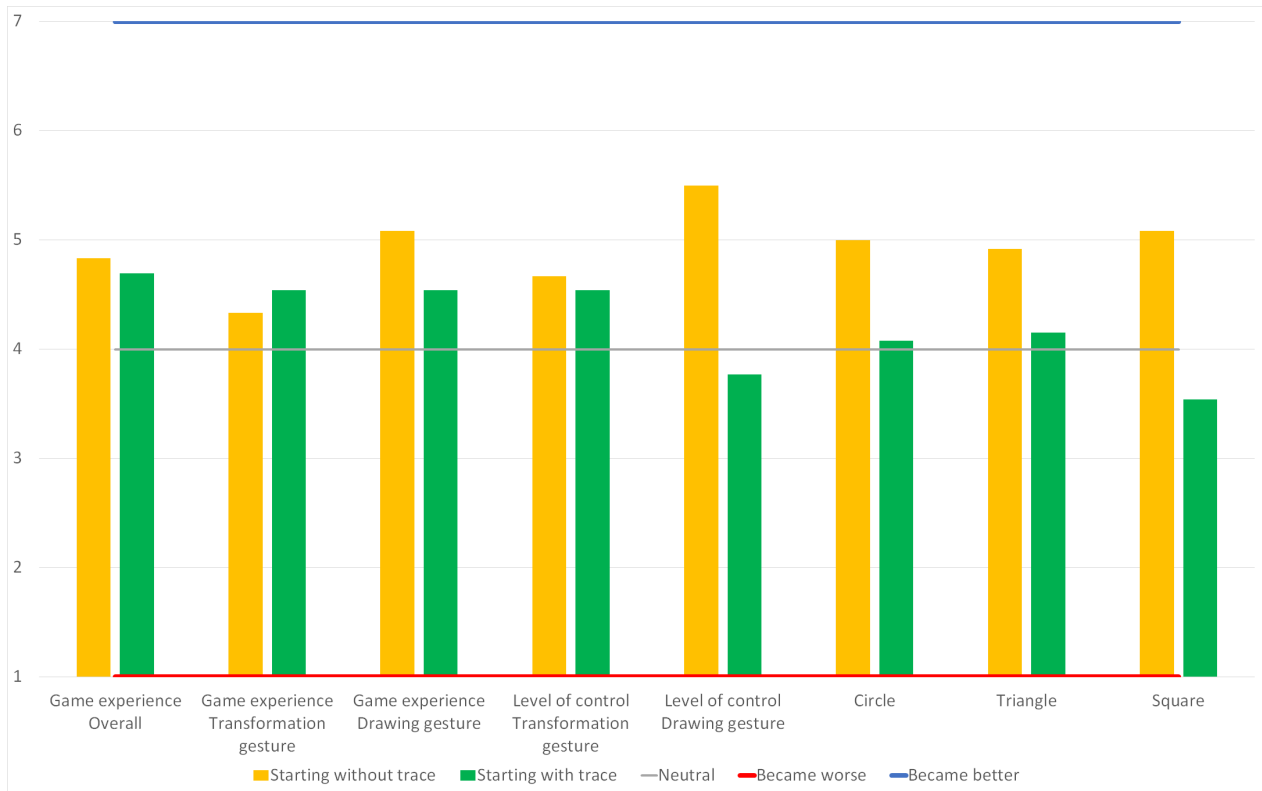


Figure C.5: Comparison between with and without visual trace. Values over all participants who performed the given order, starting without or with trace respectively.

Appendix D

Conclusions and future work

The main goal of this research was to evaluate the feasibility and usability of free-hand gestures for mobile augmented reality interaction. We therefore looked at multiple sub-goals as stated in the introduction (Chapter 1). The usability is evaluated by looking at possible ways of usage, ability to learn the technique and enjoyment while using it. Since interactions occur in mid-air, we also looked at the influence of provided visualizations.

To evaluate the feasibility and usefulness of the interaction approach we have developed a proof-of-concept game. In this game, existing objects can be moved and new ones can be created using respectively transformation and drawing gestures. The goal is to create groups of three or more objects with the same shape and reach a high score based on earned points. We have performed a pilot study which formed the basis of a formal user study to evaluate the feasibility and usability of our interaction approach. During both studies, the user had to perform different tasks using our developed game.

The pilot study proved the feasibility of our free-hand gesture-based interaction technique, showed its potential and motivated us to continue with this approach. Results showed that all participants enjoyed performing the different gestures. However, optimistic feedback might be influenced by the fact of being a novel technique combined with performing it for a short period of time. Since we gathered a limited amount of data by using only five participants, we were motivated to perform a formal user study to objectively evaluate the performance and gain more insights from more participants. From this user study, we have concluded that users are able to perform the implemented gestures with an average performance of 70% correct. Moreover, the performance improves when someone gets used to the technique. Based on this, we can conclude that users are able to handle our proposed interaction technique properly after having some training. Visual feedback did not prove to have a significant influence on the interaction behavior. However, some feedback is appreciated to see the tracked location and the state of the application.

While the results suggest high potential, some limitations have to be considered. Based on these, we recommend to make some changes for future work. First, the used libraries require a paper marker which is captured by the device's camera. When no marker is found or only a small part of it, no objects are displayed. The possible usage in normal day life is limited since the user requires to bring such a marker in order to play the game. A marker-less implementation suggests more potential. However, the implementation requires changes which may be computationally heavy. Next,

since the gestures are performed behind the device, a third of the users experienced some discomfort while playing the game, mostly related to their arm. Since the approach is created to combine real board pieces with virtual ones, this is an issue which is hard to overcome. Implementing a game which combines our approach with real board pieces could show the real benefit of this technique and possibly overcome some of the discomfort. Especially a turn-based multiplayer game allows the user to rest their arm when it is a different player's turn. Moreover, additional functionalities could be achieved by adding different gestures. These could for example be used to scale or rotate a virtual object. Another issue is the fact of having wrong recognitions during the creation of new objects. The used gesture recognition technique requires improvements or a different approach to obtain better results and limit the experienced frustration due to wrong recognitions. Currently, the circle and triangle are recognized correctly in most cases while the square gesture is recognized incorrect often. A possible solution could be having a different gesture matching approach instead of the used nearest neighbor matching. Using different gestures might also be a solution in case of shapes which do not resemble each other. Future work will have to prove whether these changes will improve the handling and enjoyment of this approach.