Gesture-based interaction via finger tracking for mobile augmented reality

Wolfgang Hürst · Casper van Wezel

Abstract The goal of this research is to explore new interaction metaphors for augmented reality on mobile phones, i.e. applications where users look at the live image of the device's video camera and 3D virtual objects enrich the scene that they see. Common interaction concepts for such applications are often limited to pure 2D pointing and clicking on the device's touch screen. Such an interaction with virtual objects is not only restrictive but also difficult, for example, due to the small form factor. In this article, we investigate the potential of finger tracking for gesture-based interaction. We present two experiments evaluating canonical operations such as translation, rotation, and scaling of virtual objects with respect to performance (time and accuracy) and engagement (subjective user feedback). Our results indicate a high entertainment value, but low accuracy if objects are manipulated in midair, suggesting great possibilities for leisure applications but limited usage for serious tasks.

1 Introduction and motivation

In mobile *augmented reality* (AR), users look at the live image of the video camera on their mobile phone and the scene that they see (i.e. the reality) is enriched by integrated tridimensional virtual objects (i.e. an augmented reality; cf. Fig. 1). It is assumed that this technique offers tremendous potential for applications in areas such as cultural heritage, entertainment, or tourism. Imagine, for example, a fantasy game with flying dragons, hidden doors, trolls that can change their appearance based on the environment, exploding walls, and objects that catch fire. Huynh et al. present a user study of a collaboratively played augmented realty board game using

W. Hürst (✉) · C. van Wezel
Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, De Uithof, 3584 CC Utrecht, The Netherlands
e-mail: huerst@uu.nl
URL: http://people.cs.uu.nl/huerst

Fig. 1 Mobile augmented reality enhances reality (represented by the live video stream on the mobile) with 3D virtual objects (superimposed at a position in the video that corresponds to a related spot in the real world) as illustrated here by a virtual flower growing in a real plant pot

handhelds that shows that such games are indeed "fun to play" and "enable a kind of social play experience unlike non-AR computer games" [21].

If virtual objects are just associated with a particular position in the real world but have no direct relation to physical objects therein, the accelerometer and compass integrated in modern phones are sufficient to create an AR environment. Using location information, for example via GPS, users can even move in these AR worlds. If virtual objects have some sort of connection to the real world other than just a global position, such as a virtual plant growing in a real plant pot, further information is needed, for example the concrete location of the pot within each video frame. This is usually achieved either by the markers (e.g. [38, 42]) or via natural feature tracking (e.g. [43–45]). For general information about augmented reality and related techniques, we refer to [8].

In order to explore the full potential of AR on mobile phones, users need to be able to create, access, modify, and annotated virtual objects and their relations to the real world by manipulations in 3D. However, current interaction with such AR environments is often limited to pure 2D pointing and clicking via the device's touch screen. Such interaction features several problems in relation to mobile AR. First, users are required to hold the device in the direction of a virtual object's position in the real world, thus forcing them to take a position that might not be optimal for interaction (cf. Fig. 2, left). Second, in general interface design, we can, for example, make icons big enough so they can easily be hit even with larger fingers. In AR, the size of the virtual objects on the screen is dictated by the real world, and thus they might be too small for easy control via touch screen. In addition, moving fingers over the screen covers large parts of the content (cf. Fig. 2, center). Third, interactions in 3D, such as putting a virtual flower into a real plant pot, have to be done by 2D operations on the device's screen (cf. Fig. 2, right). Finally, in a setting with real and virtual objects, such as an AR board game, users must constantly switch between operations on the board (to manipulate real game pieces) and interaction on the touch screen (to interact with virtual characters and objects), as illustrated in the left and center image in Fig. 3. It

Fig. 2 Potential problems with mobile augmented reality interaction via touch screen

Fig. 3 Direct interaction with real objects on a table (*left*), remote interaction via touch screen with a virtual object (*center*), and finger interaction with a virtual object that is not associated to the table or any other physical object (*right*) —notice that the virtual object (*red balloon*) is only visible on the touch screen and just added for illustration purposes

seems obvious that a smooth integration of both interaction tasks is more natural and should lead to better game play experience. Some of these intuitive arguments are confirmed by recent studies such as [35] that identified differences in task characteristics when comparing laboratory studies with building-selection tasks in an urban area.

To deal with these issues, we investigate the potential of gesture interaction based on finger tracking in front of a phone's camera. While this approach allows us to deal with the last problem by offering a smooth integration of interaction with both virtual and real objects, it also contains many potential issues, not just in terms of a robust tracking of fingers and hand gestures, but also considering the actual interaction experience. In this article, we focus on the second one. Using markers attached to the fingers, we evaluate canonical interactions such as translating, scaling, and rotating virtual objects in a mobile AR setting. Our goal is to verify the usability of this concept, identify typical usage scenarios as well as potential limitations, and investigate useful implementations. For work related to marker-less hand tracking in relation to mobile and augmented reality interaction, we refer to [3, 39, 41, 46–48].

In the remainder of this article, we start by describing the context of our research and discussing related work in Section 2. Section 3 presents a first user study evaluating finger-based interaction with virtual objects in midair [22]. We compare this concept to touch screen and another sensor-based approach in order to investigate the general feasibility of finger-based interaction. Based on a detailed analysis, we describe a second experiment in Section 4 investigating different finger-based interaction concepts for objects related to a physical object (e.g. a virtual object on a real board in an AR board game), such as illustrated in Fig. 3, left and center. Final conclusions and future work are presented in Section 5.

2 Context and related work

Most current commercial AR games on mobiles such as smart phones or the recently released Nintendo 3DS feature rather simple interaction with virtual objects such as selecting them or making them explode by clicking on a touch screen. However, finger gestures on a touch screen impose certain problems in such a setting, as discussed in Section 1. Alternative approaches for navigation in virtual reality environments on mobile phones include tilting [9, 10, 14]. This interaction concept is unsuitable in AR, because the orientation of the device is needed to create the augmented world. For the same reason, interactions based on motion gestures, such as the ones studied by [36, 37] are not applicable to this scenario.

One of the most common approaches for interaction in AR are tangible interfaces, where users take real objects that are recognized and tracked by the system to manipulate virtual objects [5, 16,

24, 40]. While this technique proved to work very well and intuitive in many scenarios, it suffers in others from forcing the user to utilize and control real objects even when interacting with pure virtual ones. Another approach that recently gained lots of attention due to research projects such as MIT's Sixth Sense [29, 30] and commercial tools such as Microsoft's Kinect [25] is interaction via finger or hand tracking. Related work demonstrated the benefits of using hand tracking for manipulation of virtual reality and 3D data, for example in Computer Aided Design [46].

On mobile phones, both industry [13] and academia [31, 41] have started to explore utilizing the user-facing camera for gesture-based control of traditional interfaces. Examples for tangible interaction in context with AR on mobile phones are [33, 34]. [1, 18] use a combination of tangible interaction, joy pad, and key based interaction. [19, 20] evaluate the usage of movement of the phone in combination with a traditional button interface. [23] gives an example for body part tracking (e.g. hands) to interact with an AR environment in a setting using external cameras and large screens. [11] features one of the first approaches to utilize finger tracking in AR using a camera and projector mounted on the ceiling. A comparable setup is used by [26] where interaction with an augmented desk environment was done by automatic finger tracking. [12] uses markers attached to the index finger to realize interaction in an AR board game. [28] presents a detailed study on the usability and usefulness of finger tracking for AR environments using head mounted displays.

Particularly the work by [28] demonstrates promising potential for the usage of finger tracking in AR interaction. However, it is unclear if these positive results still apply when mobile phones are used to create the augmented reality instead of head mounted displays. Examples using mobile handheld devices include [39] who creates AR objects in the palm of ones hand by automatic hand detection and tracking. [4] presents a system where free-hand drawings are used to create AR objects in a mechanical engineering e-learning application. Hagbi et al. use hand-drawn shapes as markers for AR objects [17]. However, both do not enable explicit object manipulation, which is the main focus of our research. Using finger tracking for gesture-based interaction, as proposed in our work, has also been investigated by [3, 6, 48]. [6] describes design options for such a concept and discusses related advantages. The authors claim that a related user test on biomechanics showed no significant differences in physical comfort between their approach and other ways of interaction, but no further information is given. The concept was also identified as being "useful", "innovative", and "fun" based on a video illustrating the approach and interviews with potential users. However, no actual implementation or prototype was tested. [3] and [48] both focus on realizing a marker-less hand tracking, but provide limited information on actual usability. [3] gives a detailed description of potential use cases, but no testing has been done.

The goal of the research presented in this article is to further investigate and explore the full potential of finger tracking for mobile AR interaction. Usability of this concept could be critical due to several limiting factors. First, there are restrictions imposed by the hardware. The camera's field of view (FOV) defines the region in which a hand or finger can be tracked. For example, the camera of the phone we used in our first experiment features a horizontal and vertical viewing angle of approximately 55 and 42°, respectively. The camera's resolution has an influence on how fine-grained the interaction can be. For example, a low resolution does not enable us to make a very detailed tracking, especially if the finger is very close to the camera. Second, there are restrictions imposed by human biomechanics. Every arm has a reachable workspace which can be defined as the volume within which all points can be reached by a chosen reference point on the wrist [27]. It is restricted, among other characteristics, by the human arm length, which, for example, in the Netherlands is on average about 0.5968 m for males and 0.5541 m for females. In addition, the phone has to be held at a certain distance from the eyes. For example, the Least Distance of Distinctive Vision (LDDV), which describes the

closest distance at which a person with normal vision can comfortably look at something, is commonly around 25 cm [2]. Considering all these issues, we get a rather limited area that can be used for interaction in a mobile AR scenario, as illustrated in Fig. 4. [48] observed that the size of area (1) in the image is around 15 to 25 cm.

3 Interaction with objects floating in midair
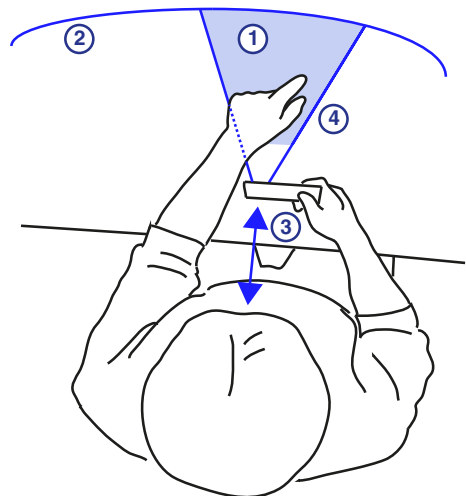
3.1 Interaction concepts and tasks

3.1.1 Goal and motivation

In order to evaluate the general feasibility of finger tracking in relation to mobile AR, we start by comparing it with standard interaction via touch screen and another interaction concept that depends on how the device is moved (utilizing accelerometer and compass sensors). Our overall aim is to target more complex operations with virtual objects than pure clicking. In the ideal case, a system should support all canonical object manipulations in 3D, i.e. selection, translation, scaling, and rotation. However, for this initial study and in order to be able to better compare it to touch screen based interaction (which per default only delivers two-dimensional data), we restrict ourselves here to the three *tasks* of *selecting virtual objects*, *selecting entries in context menus*, and *translation of 3D objects in 2D* (i.e. left/right and up/down). Objects are floating in midair, i.e. they have a given position in the real world, but no direct connection to objects therein (cf. Fig. 3, right).

3.1.2 Framework and setup

AR environment We implemented all three interaction concepts described below on a Motorola Droid/Milestone phone with Android OS version 2.1. Since we restricted ourselves to manipulations of virtual objects floating in midair in this study, no natural feature tracking was needed. The AR environment used in this experiment thus relied solely on sensor information from the integrated compass and accelerometer.

Fig. 4 Area that can be used for interaction (1), depending on the reachable workspace (2), the least distance of distinctive vision (3), and the camera's FOV (4)

3.1.3 Interaction concepts

Touch screen-based concept For the standard touch screen-based interaction, the three tasks have been implemented in the following way: selecting an object is achieved by simply clicking on it on the touch screen (cf. Fig. 5, left). This selection evokes a context menu that was implemented in a pie menu style [7], which has proven to be more effective for pen and finger based interaction (compared to list-like menus as commonly used for mouse or touchpad based interaction; cf. Fig. 5, right). One of these menu entries puts the object in "translation mode" in which a user can move an object around by clicking on it and dragging it across the screen. If the device is moved without dragging the object, it stays at its position with respect to the real world (represented by the live video stream). Leaving translation mode by clicking at a related icon fixes the object at its new final position in the real world.

In terms of usability, we expect this approach to be simple and intuitive because it conforms to standard touch screen based interaction. In case of menu selection, we can also expect it to be reliable and accurate because we are in full control of the interface design (e.g. we can make the menu entries large enough and placed as far apart of each other so they can be hit easily with your finger). We expect some accuracy problems though when selecting virtual object, especially if they are rather small, very close to each other, or overlapping because they are placed behind each other in the 3D world. In addition, we expect that users might feel uncomfortable when interacting with the touch screen while they have to hold the device upright and targeted towards a specific position in the real world (cf. Fig. 2, center). This might be particularly true in the translation task for situations where the object has to be moved to a position that is not shown in the original video image (e.g. if users want to place it somewhere behind them).

Device-based concept Our second interaction concept uses the position and orientation of the device (defined by the data delivered from the integrated accelerometer and compass) for interaction. In this case, a reticule is visualized in the center of the screen (cf. Fig. 6, left) and used for selection and translation. Holding it over an object for a certain amount of time (1.25 s in our implementation) selects the object and evokes the pie menu. In order to avoid accidental selection of random objects, a progress bar is shown above the object to illustrate the time till it is selected. Menu selection works in the same way by moving the reticule over one of the entries and holding it till the bar is filled (cf. Fig. 6, right). In translation mode, the object sticks to the reticule in the center of the screen while the device is moved around. It can be placed at a target position by clicking anywhere on the touch screen. This action also forces the system to leave translation mode and go back to normal interaction.
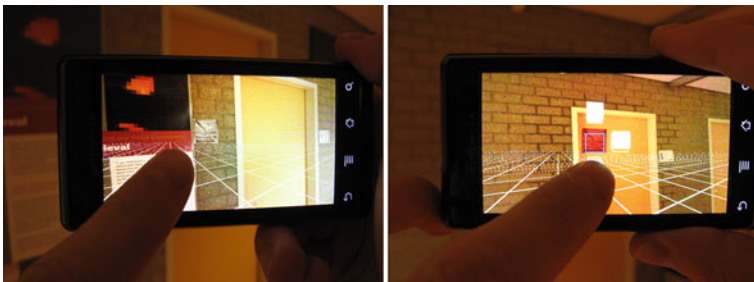

Fig. 5 Touch screen interaction: select an object (*left*) or an entry from a context menu that pops up around the object after selection (*right*)

Fig. 6 Device based interaction: selection by pointing a reticule (*left*) to the target till the related bar is filled (*right*)

Compared to touch screen interaction, we expect this "device based" interaction to take longer when selecting objects and menu entries because users can not directly select them but have to wait till the progress bar is filled. In terms of accuracy, it might allow for a more precise selection because the reticule could be pointed more accurately at a rather small target than your finger. However, holding the device at one position over a longer period of time (even if it's just 1.25 s) might prove to be critical especially when the device is hold straight up into the air. Translation with this approach seems intuitive and might be easier to handle because people just have to move the device (in contrast to the touch screen where they have to move the device and drag the object at the same time). However, placing the object at a target position by clicking on the touch screen might introduce some inaccuracy because we can expect that the device shakes a little when being touched while held up straight in the air. This is also true for the touch screen based interaction, but might be more critical here because for touch screen interaction the finger already rests on the screen. Hence, we just have to it release it whereas here we have to explicitly click the icon (i.e. perform a click and release action).

Finger-based concept Touch screen based interaction seems intuitive because it conforms to regular smart phone interaction with almost all common applications (including most current commercial mobile AR programs). However, it only allows to remotely controlling the tridimensional augmented reality via 2D input on the touch screen. If we track the users' index finger when their hand is moved in front of the device (i.e. when it appears in the live video on the screen), we can realize a finger based interaction where the tip of the finger can be used to directly interact with objects, i.e. select and manipulate them. In the ideal case, we can track the finger in all three dimensions and thus enable full manipulation of objects in 3D. However, since 3D tracking with a single camera is difficult and noisy (especially on a mobile phone with a moving camera and relatively low processing power) we restrict ourselves to 2D interactions in the study presented in this paper. In order to avoid influences of noisy input from the tracking algorithm, we also decided to use a robust marker based tracking approach where users attach a small sticker to the tip of their index finger (cf. Fig. 7). Object selection is done by "touching" an object (i.e. holding the finger at the position in the real world where the virtual object is displayed on the screen) until an associated progress bar is filled. Menu entries can be selected in a similar fashion. In translation mode, objects can be moved by "pushing" them. For example, an object is moved to the right by approaching it with the finger from the left side and pushing it rightwards. Clicking anywhere on the touch screen places the object at its final position and leaves translation mode.
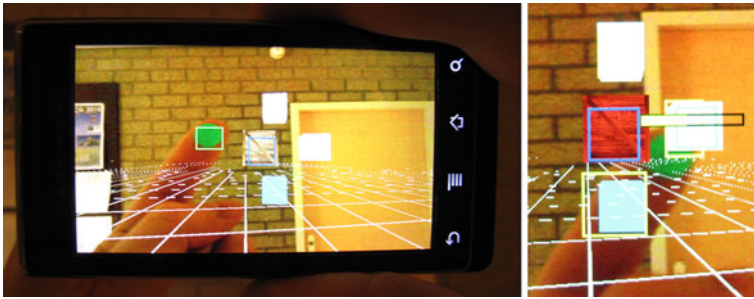
Fig. 7 Interaction by using a green marker for tracking that is attached to the tip o the finger

3.2 First evaluation: objects in midair

3.2.1 Setup and procedure

Participants In the previous section we discussed some intuitive assumptions about the usability and potential problems with the respective concepts. In Section 2, we also described potential limitations of finger-based interaction in mobile AR in general. For example, Fig. 8 illustrates the problem that occurs when a finger is too close to the camera. In order to verify these expected advantages and disadvantages we set up a user study with 18 participants (12 male and 6 female, 5 users at ages 15–20 years, 8 at ages 21–30, 1 at ages 31–40 and 41–50, and 3 at ages 51–52). For the finger tracking, users were free in where to place the marker on their finger tip. Only one user placed it on his nail. All others used it on the inner side of their finger as shown in Figs. 7 and 8. Eleven participants held the device in their right hand and used the left hand for interaction. For the other seven it was the other way around. No differences could be observed in the evaluation related to marker placement or hand usage.

Task A within-group study was used, i.e. each participant tested each interface (subsequently called *touch screen*, *device*, and *finger*) and task (subsequently called *object selection*, *menu selection*, and *translation*). Interfaces were presented in different order to the participants to exclude potential learning effects. For each user, tasks were done in the following order: object selection, then menu selection, then translation, because this would also be the natural order in a real usage case. For each task, there was one introduction test in which the interaction method was explained to the subject, one practice test in which the subject could try it out, and finally three "real" tests (four in case of the translation task) that were used in
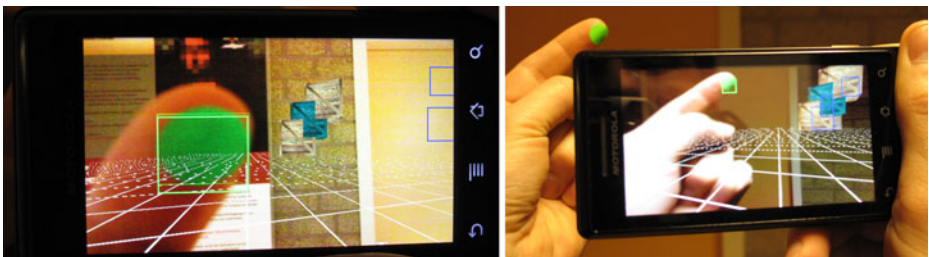

Fig. 8 Holding the finger too close to the camera makes it impossible to select small objects (*left*). Moving your hand away from the camera decreases the size of the finger in the image but can result in an uncomfortable position because you have to stretch out your arm (*right*)

the evaluation. Subjects were aware that the first two tests were not part of the actual experiment and were told to perform the other tests as fast and accurate as possible. The three tests used in the object selection task can be classified as easy (objects were placed far away from each other), medium (objects were closer together), and hard (objects overlapped; cf. Fig. 9). In the menu selection task, the menu contained three entries and users had to select one of the entries on top, at the bottom, and to the right in each of the three tests (cf. Figs. 5, 6, and 7, right). In the translation task, subjects had to move an object to an indicated target position (cf. Fig. 10). The view in one image covered a range of 72.5°. In two of the four tests, the target position was placed within the same window as the object that had to be moved (at an angle of 35° between target and initial object position to the left and right, respectively). In the other two tests, the target was outside of the initial view but users were told in which direction they had to look to find it. It was placed at an angel of 130° between target and object, to the left in one case, and to the right in the other one. The order of tests was randomized for each participant to avoid any order-related influences on the results.

Evaluation For the evaluation, we logged the time it took to complete the task, success or failure, and all input data (i.e. the sensor data delivered from the accelerometer, compass, and the marker tracker). Since entertainment and leisure applications play an important role in mobile computing, we were not only interested in pure accuracy and performance, but also in issues such as fun, engagement, and individual preference. Hence, users had to fill out a related questionnaire and were interviewed and asked about their feedback and further comments at the end of the evaluation.

3.2.2 Results and discussion

Time Figure 9 shows the time it took the subjects to complete the tests averaged over all users for each task (Fig. 11, top left) and for individual tests within one task (Fig. 11, top right and bottom). The averages in Fig. 11, top left show that touch screen is the fastest approach for both selection tasks. This observation conforms to our expectations mentioned in the previous section. For device and finger, selection times are longer but still seem reasonable.

Menu selection task Looking at the individual tests used in each of these tasks, times in the menu selection task seem to be independent of the position of the selected menu entry (cf. Fig. 11, bottom left). Almost all tests were performed correctly: only three mistakes happened over all with the device approach and one with the finger approach.

Object selection task In case of the object selection task (cf. Fig. 11, top right), touch screen interaction again performed fastest and there were no differences for the different levels of difficulty of the tasks. However, there was one mistake among all easy tests and in five of the

Fig. 9 Object selection task: test case (single object, *left*), easy (multiple non-overlapping objects, *center*), and hard (multiple overlapping objects, *right*)

Fig. 10 Translation task: move object to the target (*white square on the right*)
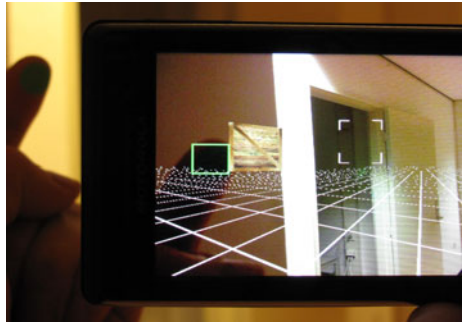
hard tests the wrong object was selected thus confirming our assumption that interaction via touch screen will be critical in terms of accuracy for small or close objects. Finger interaction worked more accurate with only two mistakes in the hard test. However, this came at the price of a large increase of selection time. Looking into the data we realized that this was mostly due to subjects holding the finger relatively close to the camera resulting in a large marker that made it difficult to select an individual object that was partly overlapped by others. Once the users moved their hand further away from the camera, selection worked well as indicated by the low amounts of errors. For the device approach, there was a relatively large number of errors for the hard test, but looking into the data we realized that this was only due to a mistake that we made in the setup of the experiment: in all six cases, the reticule was already placed over an object when the test started and the subjects did not move the device away from it fast enough to avoid accidental selection. If we eliminate these users from the test set, the time illustrated for the device approach in the hard test illustrated in Fig. 11, top right increases from 10,618 msec to 14,741 msec which is still in about the same range as the time used for the tests with easy and medium levels of difficulty. Since all tests in which the initialization problem did not happen have been solved correctly, we can conclude that being forced to point the device to a particular position over a longer period of time did not result in accuracy problems as we suspected.

Fig. 11 Time it took to solve the tasks (in milliseconds, averaged over all users)

Translation task In the translation task, we see an expected increase in time in case of the finger and touch screen approach if the target position is outside of the original viewing window (cf. Fig. 11, bottom right). In contrast to this, there are only small increases in the time it took to solve the tasks when the device approach is used. In order to verify the quality of the results for the translation task, we calculated the difference between the center of the target and the actual placement of the object by the participants in the tests. Figure 12 illustrates these differences averaged over all users. It can be seen that the device approach was not only the fastest but also very accurate, especially in the more difficult cases with an angel of 130° between targets and initial object position. Finger based interaction on the other hand seems very inaccurate. A closer look into the data reveals that the reason for the high difference between target position and actual object placement are actually many situations in which the participants accidentally hit the touch screen and thus placed the object at a random position before reaching the actual target. This illustrates a basic problem with finger-based interaction, i.e. that users have to concentrate on two issues: moving the device and "pushing" the object with their finger at the same time. In case of device-based interaction on the other hand, users could comfortably take the device in both hands while moving around thus resulting in no erroneous input and a high accuracy as well as fast solution time. A closer look into the data also revealed that the lower performance in case of the touch based interaction are mostly due to one outlier who did extremely bad (most likely also due to an accidental input) but otherwise they are at about the same level of accuracy as the device approach (but of course at the price of a worse performance time; cf. Fig. 11).

Fun and engagement Based on this formal analysis of the data, we can conclude that touch based interaction seems appropriate for selection tasks. For translation tasks, the device based approach seems to be more suitable, whereas finger based interaction seems less useful in general. However, especially in entertainment and leisure applications, speed and accuracy are not the only relevant issues, but fun and engagement can be equally important. In fact, for gaming, mastering an inaccurate interface might actually be the whole purpose of the game (think of balancing a marble through a maze by tilting your phone appropriately—an interaction task that can be fun but is by no means easy to handle). Hence, we asked the participants to rank the interaction concepts based on performance, fun, and both. Results are summarized in Table 1. Rankings for performance clearly reflect the results illustrated in Figs. 11 and 12 with touch based interaction ranked highest by 11 participants, and device ranked first by six. Only one user listed finger based interaction as top choice for performance. However, when it comes to fun and engagement, the vast majority of subjects ranked it as their top choice, whereas device and touch were ranked first only four and one time, respectively. Consequently, rankings are more diverse when users were asked to consider both issues. No clear "winner" can be identified here, and in fact, many users commented that
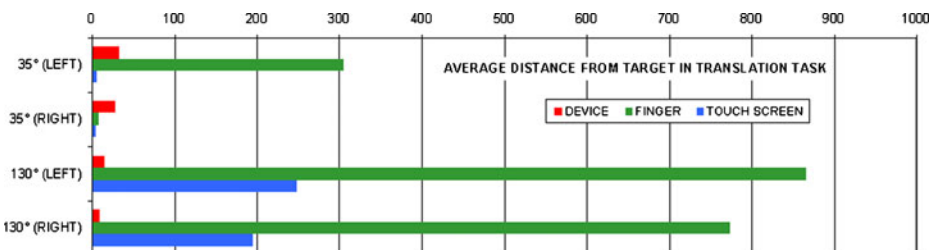

Fig. 12 Correctness of translation task (squared Manhattan distance in screen units)

Table 1 Times how often an interface was ranked first, second, and third with respect to performance versus fun and engagement versus both (*T* touch screen, *D* device, *F* finger)

	Performance			Fun			Perf. & fun		
	T	D	F	T	D	F	T	D	F
Ranked 1st	11	6	1	1	4	13	8	3	7
Ranked 2nd	3	6	9	6	10	2	6	8	4
Ranked 3rd	4	6	8	11	4	3	4	7	7

it depends on the task: touch and device based interaction are more appropriate for applications requiring accurate placement and control, whereas finger based interaction was considered as very suitable for gaming applications. Typical comments about the finger-based approach characterized it as "fun" and "the coolest" of all three. However, its handling was also criticized as summarized by one user who described it as "challenging but fun". When asked about the usefulness of the three approaches for the three tasks of the evaluation, rankings clearly reflect the objective data discussed above. Table 2 shows that all participants ranked touch based interaction first for both selection tasks, but the vast majority voted for the device approach in case of translation. It is somehow surprising though that despite its low performance, finger based interaction was ranked second by eight, seven, and nine users, respectively, in each of the three tasks—another indication that people enjoyed and liked the approach although it is much harder to handle than the other two.

3.2.3 Summary of major findings

Touch screen based interaction achieved the best results in the shortest time in both selection tasks and was rated highest by users in terms of performance. Device based interaction achieved the best results in the shortest time in the translation task and was highly ranked by the users for this purpose. Finger tracking based interaction seems to offer great potential in terms of fun and engagement, because users responded very enthusiastically about it considering these aspects. However, both qualitative as well as objective data suggest a very low performance when trying to perform canonical operations such as selection and translation of virtual objects floating in mid air. This makes it rather questionable if this concept is suitable for interaction, especially for serious applications. Major problems identified include that holding up the device over longer periods of time can be tiresome and especially moving objects over long distances is difficult to control. However, these characteristics are typical for interaction in midair but might not appear if

Table 2 Times how often an interface was ranked first, second, and third with respect to the individual tasks (*T* touch screen, *D* device, *F* finger)

	Object select.			Menu select.			Translation		
	T	D	F	T	D	F	T	D	F
Ranked 1st	18	–	–	18	–	–	3	15	–
Ranked 2nd	–	10	8	–	11	7	6	3	9
Ranked 3rd	–	8	10	–	7	11	9	–	9

objects have a clear relation to the environment, such as virtual play figures in AR board game, where people can rest their arm on the table and the interaction range is naturally restricted (cf. Fig. 3 left/center vs. right). Another major issue was low tracking performance because people held their fingers too close to the camera. Again, we expect a different experience if the objects are not floating in midair but are directly related to physical objects in the real environment. In addition, we identified that pushing the object from different sides—although originally considered as an intuitive interaction—turned out to be awkward and hard to handle, for example when pushing an object from the right using the left hand. This might be overcome by investigating different versions for gestures that are more natural and easier to control.

4 Interaction with objects related to the real world

4.1 Interaction concepts and tasks

4.1.1 Goal and motivation

In order to further investigate the potential of finger-based interaction and to verify if our above arguments about the manipulation of objects that have a connection to physical ones are true, we set up this second experiment. Because we cannot expect a better performance of finger interaction compared to touch screen and device based due to our results from the first study, our major focus is on game play experience. Consequently, we do our evaluation in a board game scenario. Our major goal with this study is to identify if and what kind of interactions are possible, natural, and useful for achieving an engaging game play. Hence, we set up an experiment comparing different realizations of the canonical interactions *translation*, *scaling*, and *rotation* of two kinds of virtual objects—ones related to the actual game board and ones floating slightly above it (cf. Fig. 3, left and center).

4.1.2 Framework and setup

AR environment To create the AR environment for this second experiment, we used the Qualcomm Augmented Reality (QCAR) SDK. This SDK provides a robust and fast framework for natural feature tracking. Because of the natural feature tracking, we could use a game board that was created in the style of the common board game Ludo (cf. Fig. 13) and no artificial markers were needed. Our implementation was done on an HTC Desire HD smartphone running Android 2.2.3 and featuring a 4.3-inch screen with a resolution of 480×800 pixels and an 8-megapixel camera.

Marker-based finger tracking For the marker-based finger tracking we extended our color tracker from the first experiment to track two markers—a green and red one on the user's thumb and index finger, respectively (cf. Fig. 14). Interaction was implemented by creating a bounding box around the trackers and the virtual objects on the board. The degree to which those bounding boxes overlapped or touched each other resulted in different interactions (cf. below). Notice that with this approach, fingers can again only be tracked in 2D (left/right and up/down). However, interaction in 3D (far/near) can be done by moving the fingers along with the device (cf. Fig. 15). Using more than two markers significantly reduces tracking performance because of occlusion of fingers. In addition, we did some informal testing that confirmed that using more than two fingers becomes less natural and harder to handle.
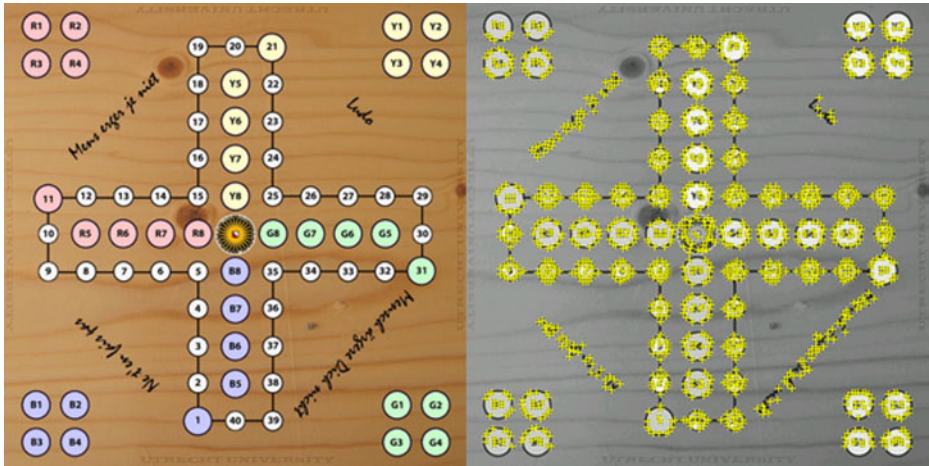
Fig. 13 Game board used in the evaluations and extracted natural features for tracking and creating the AR environment

4.1.3 Interaction concepts

Translation When people are moving objects on a real board game, we usually observe two kinds of interactions: they are either pushed over the board or grabbed with two fingers, moved in midair, and placed at a target position. For our comparative study we implemented both versions using our marker tracking as illustrated in Fig. 16. For pushing, an object is moved whenever a finger touches it from a particular side. Hence, only the index finger is used for this interaction. Grabbing on the other hand uses both fingers. Once both markers come close enough to the bounding box, the virtual object is "grabbed". It can now be moved around in midair and dropped anywhere on the board by moving the fingers away from each other.

Scaling Scaling is an interaction that can only be done with virtual objects. Hence, there is no natural equivalent. However, gesture-based scaling is quite common on touch screens using a pinch-to-zoom gesture: moving two fingers away from or closer to each other results in zooming out and in, respectively. While this seems to be a natural gesture for scaling of virtual objects in AR as well, there is one major problem: in contrast to a touch screen, we are not able to automatically determine when an object is "touched", i.e. when a scaling gesture appears and when not. To deal with this issue, we implemented and evaluated two versions. In the first one,

Fig. 14 A green marker (*thumb*) and a red one (*index finger*) are used to track fingers and manipulate virtual objects on the board game
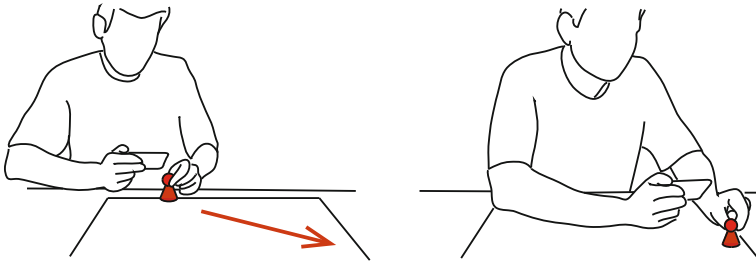
Fig. 15 By moving the device with the hand, users can manipulate virtual objects in 3D even if the marker tracking of the thumb and index finger is only done in 2D—notice that the virtual object (*red pawn*) is only visible on the touch screen and just added for illustration purposes

we distinguish between two gestures: touching an object with two fingers close to each other and then increasing the distance between them results in an enlargement. Grabbing the object from two different sides and decreasing the distance between the two fingers scales it down (cf. Fig. 17). In both cases, the object is released by moving the fingers in the opposite direction. With this, scaling down and enlarging are two separate interactions. Hence, we implemented a second version where increasing and decreasing the distance between the two marked fingers continuously changes the size of an object. To fix it at a particular size, users have to tap anywhere on the touch screen. While this interaction seems much more flexible and natural for the actual scaling, it is unclear if users will feel comfortable touching the screen with one finger of the hand that currently holds it (cf. Fig. 18).

Rotation Rotating an object is usually done by grabbing it with two fingers and turning it around. This can be implemented by first grabbing the object (similarly to the two-finger-based interaction for translation), and then rotating it in dependence to the change of the angle at the center of the line between the two markers (cf. Fig. 19). However, in our informal pre-testing we got the impression that such a gesture is hard to perform when holding your phone in the other hand. Hence, we implemented a second version where users select an object by touching it. Once selected, the index finger can be used to rotate the object by a circular movement. Tabbing anywhere on the touch screen is used to stop rotation and fix the object at a certain orientation (similarly to the approach for scaling illustrated in Fig. 18).

Visualization and lack of haptic feedback One of the major problems we face in virtual and augmented reality applications where users can interact directly with virtual objects by "touching" them with their real hands is the lack of haptic feedback. In order to deal with this issue, we implemented a visualization that clearly indicates the interpretation of a particular interaction and the related status of the object (e.g. selected or not). In addition, we increased a tolerance distance between the bounding boxes of the markers and object in all interaction concepts
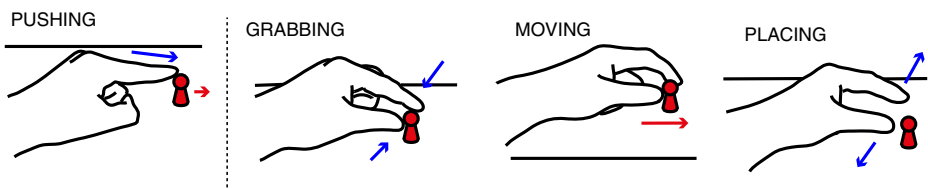

Fig. 16 Translation of virtual objects by pushing them with one finger (*left*) or grabbing, lifting, moving, and placing them with two fingers (*right*)—notice that the red pawn represents a virtual object that was only added here for illustration purposes. In the actual implementation it is only visible on the phone's touch screen
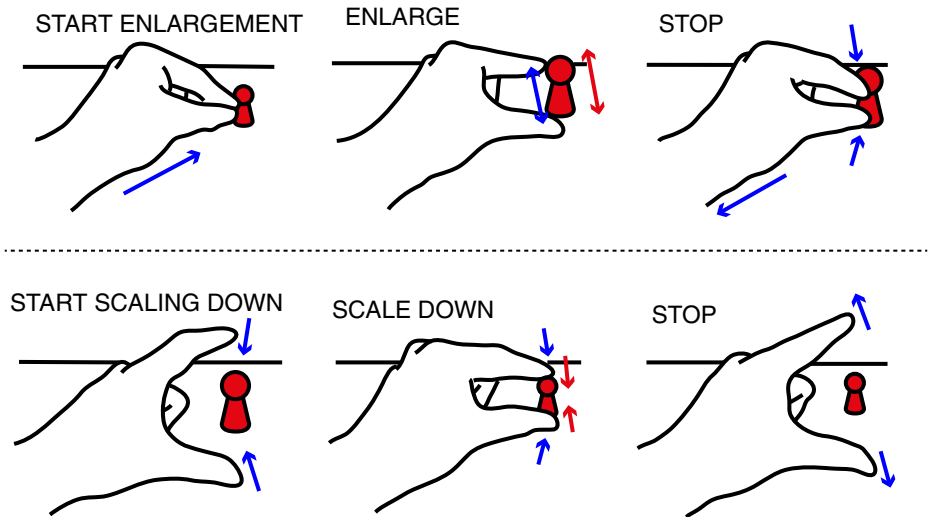
Fig. 17 Scaling of virtual objects by using different gestures—the virtual object (*red pawn*) is only visible on the touch screen and just added for illustration purposes

discussed before. Hence, users do not have to "touch" the object directly but it already becomes selected and reacts to gestures in a certain area around it—indicated by a transparent feedback on the board as illustrated in the snapshots in Figs. 20, 21, 22 and 23.

4.2 Second evaluation: objects with real world relation

4.2.1 Setup and procedure

Participants In the previous section we introduced two different interaction types for each of the three tasks translation, scaling, and rotation. In order to evaluate their usability, we set up an experiment with 24 participants (20 male, 4 female) at ages from 18 to 25 years (average 21.34 years). For this experiment, we have purposely chosen this age range because it represents one of the major target groups for entertainment-related applications on mobile phones. Studies with other target groups are part of our future work.

Tasks and levels For each interaction type, we specified some objectives—subsequently called levels—which users had to solve (cf. Figs. 9, 10, 11 and 12). For translation,
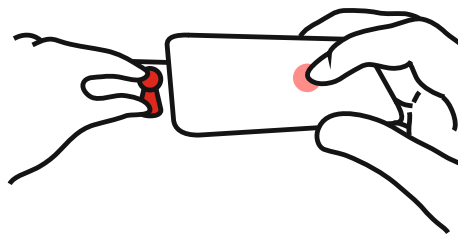

Fig. 18 Ending scaling mode by clicking anywhere on the touch screen—notice that the light red dot on the screen was only added to illustrate a tap on the screen but does not actually appear in the implementation. The red pawn illustrates a virtual object that is displayed at this position on the screen but does not exist in reality
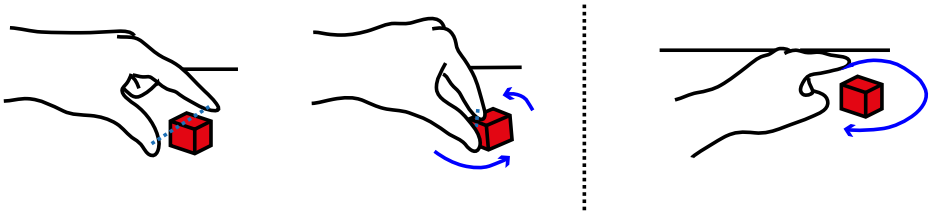
Fig. 19 Rotation of virtual objects by grabbing them with two fingers and turning them (*left*) or by selecting them and making related circular movements with one finger—notice that the red cube presents a virtual object that was only added here for illustration purposes; in the actual implementation it is only visible on the phone's touch screen

participants had to move a pawn figure from a particular place on the board to a labeled position over a small, medium, and large distance. For scaling, they had to shrink or enlarge a smaller and larger version of the pawn to its original size that was indicated on the board. Both levels were done twice—once in midair and once on the board. For rotation, they had to do a long clockwise rotation and a short counter clockwise rotation on the board, as well as a long clockwise rotation and a short clockwise rotation in midair. Translation was only tested on the board because we already evaluated similar midair interactions in the first experiment (cf. Section 3). Each level was preceded by a practice task, resulting in an overall number of 32 tasks per experiment (4 translation, 2*3 scaling, 3*2 rotation, times 2 because each interaction type was tested in two versions).

Evaluation The experiment was set up as a within subject approach, i.e. every subject evaluated each setting. Hence, the interaction design served as independent variable. Dependent variables for measurement were time to complete one level and the accuracy at which a specific level was solved. In addition, we logged all sensor and tracking data for further analysis. Because improving the game play experience is one of the major aims of our approach, we also gathered subjective user feedback via a questionnaire and a concluding interview and discussion with each participant. During the tests, a neutral observer made notes about special observations, interaction problems, etc. Hence, the overall test started with a short introduction, followed by the levels for each task (including a practice level for each task), and concluded with a questionnaire and subsequent interview. To avoid the influence of learning effects, the order of levels within one interaction type as well as the order of midair and on board interactions have been counter balanced over all participants.
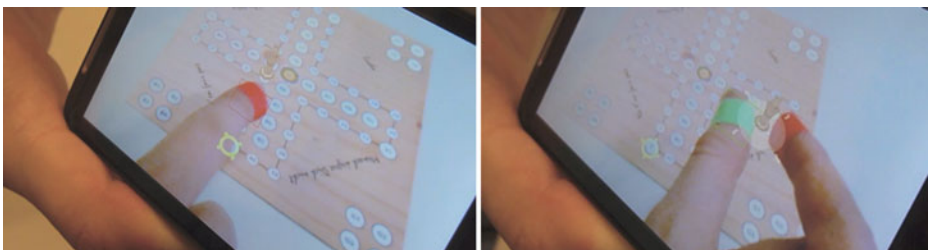

Fig. 20 Translation with one finger (*left*) and two fingers (*right*) as used in the evaluation
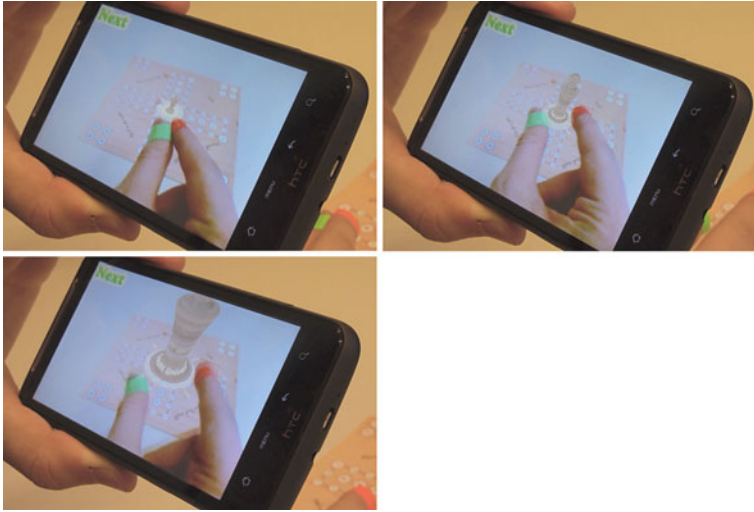
Fig. 21 Scaling via gestures as used in the evaluation

4.2.2 Results

Time and accuracy The major aim of our proposed approach is to improve the interaction experience and making it more natural by smoothly integrating interactions with real and virtual objects. How people experience an interaction depends on various factors including subjective issues such as ease of use, engagement, and personal enjoyment. These qualitative factors can be influenced by more objective issues, such as performance in achieving a task, which in turn can be measured in different ways. Figures 24 and 25 illustrate the average time it took users to solve a task and the achieved accuracy, respectively. Average times are indicated in milliseconds. Since issues such as a "perfectly accurate" placement of objects in an actual game play situation are prune to subjective interpretation, accuracy is measured in two categories: a perfect solution (i.e. perfect position, size, and angle for translation, scaling, and rotation tasks, respectively) and a solution that would be considered as good enough in an actual game play situation. The latter was measured by using experimentally determined thresholds (i.e. distances, sizes, and angles at which a translation, scaling, and

Fig. 22 Rotation with one finger as used in the evaluation—notice that the blue dice was used as a reference and the task was to rotate the red one into a similar orientation
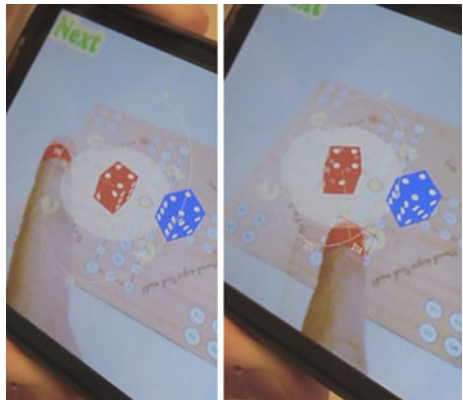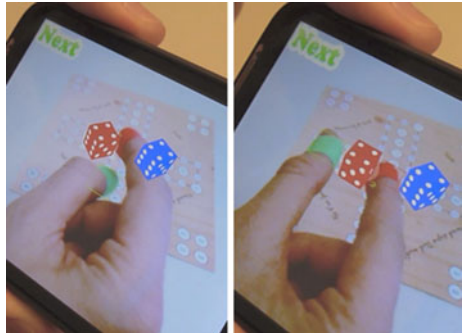
Fig. 23 Rotation with two fingers as used in the evaluation—notice that rotating at larger angles is not possible without releasing and re-grabbing the dice again

rotation operation would be considered as done successfully by most participants). In the diagrams, blue indicates translation levels. Scaling and rotation levels are shown in green and red, respectively. For scaling and rotation, darker colors indicate interactions done on the board, whereas lighter colors represent interactions in midair. The lower part of each bar illustrated the amount of "perfect" solutions.

Translation task For translation (blue bars), the first three entries in both figures represent the interaction with two fingers. The latter three show the results for pushing an object with one finger. Although the bar charts show clear differences in time for different levels, a statistically significance could not be proven. The same holds for the accuracy measures. In general, both operations—one- and two-finger-based translation—worked equally well. The comparable performance indicators are reflected in the subjective feedback from the users, as well. In the final questionnaire, translation using two fingers was rated slightly higher than one-finger based translation with an average rating of 7.27 vs. 7.02 on a scale from 0 (worst) to 10 (best). However, considering the participants' comments in the following interview, there seems to be a slight preference towards using two fingers, as it was often noted as being more natural and intuitive. Although this was not reflected in worse performance values, some uses commented that pushing the object with one finger makes the exact placement a little harder because you have to push it from different sides if you overshoot the target.
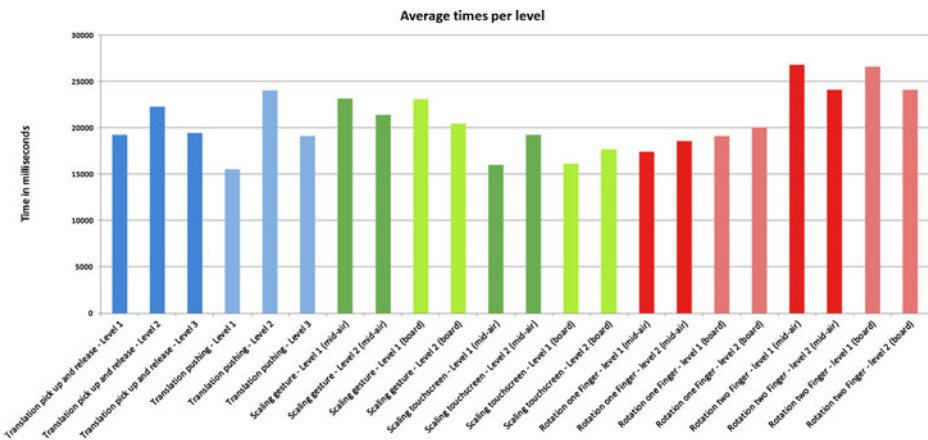

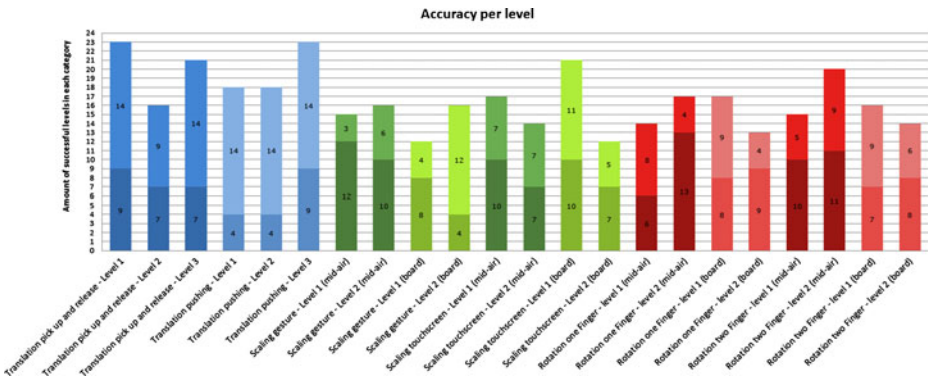**Fig. 24** Average time to solve the different levels (in milliseconds)

Fig. 25 Accuracy per level (measured in number of correct solutions)

Scaling task The first four green bars in the figures represent the gesture-based scaling results, the latter ones the interaction via touch screens. Here, both interaction concepts show a significant difference in the average time to solve the tasks for both midair and on board interaction (*t*-test, $p<0.05$). Within one interaction type, average times for midair and on board operations are almost the same and not statistically significant difference could be proven. However, considering accuracy, there is a significant difference between midair and on board tasks within each of the two interaction concepts (*t*-test, $p<0.05$). Especially for shrinking an object on the board, touch screen interactions seem more useful achieving a correct result in 87.5% of the cases versus a success rate of only 66.7% for gesture-based interaction. The slight differences in favor of the touch screen approach are also reflected in the users' rating with touch screen interaction scoring an average of 7.15 and 7.17 for on board and midair operations, respectively, and gesture-based interaction with an average rating of 6.65 and 6.67 (again on a scale from 0 to 10). However, as we anticipated (cf. Fig. 18), several users complained during the interviews that it was awkward doing the gestures with one hand and clicking on the touch screen while holding the device with the other hand and that it put an additional cognitive load on them. The slightly lower rating for the gesture-based scaling was mostly due to the impression that it was harder to control, and that users had to remember certain interaction patterns.

Rotation task In case of the rotation task, the first four red bars in the diagrams show the results for interaction with one finger, the latter ones the two finger interaction. A statistically significant difference was observed for the average solution time between these two approaches (*t*-test, $p<0.05$), whereas the midair and on board interaction in each of the two concepts only showed slight differences. Also, no statistically significance could be proven for the accuracy at which the different levels have been solved, despite a higher performance for one of the levels in midair interaction using two fingers. Considering both the rating in the questionnaire and the comments by the users, rotation using two fingers was not considered very useful. Whereas rotation with one finger achieved the highest marks of all interaction concepts—7.52 for on board and 7.40 for midair interaction—using two fingers was rated extremely low—5.58 for on board and 5.33 for midair interaction. The major reason for this low rating and the related negative comments by the users is that rotation at larger angles forces people to release and grab the object again due to the limited rotation range of the human wrist. Consequently, the participants preferred the one-finger-based interaction although this is the less natural one of the two. Nevertheless, they described it as intuitive and easy to understand and use.

Discussion and interview In order to start an open discussion and informal interview, users were asked three open questions at the end related to their overall impression, engagement, and the potential of finger-based gestures for usage in an actual AR board game, respectively. Overall, the feedback was very positive with only two participants mentioning that they would in general prefer touch screen based interaction. One of them was generally not convinced of the practicability of this kind of interaction whereas the other still described it as "usable and original". On the other hand, five participants explicitly noted that they would consider it as annoying to hold the device in such a way and operate the touch screen at the same time. Other negative remarks were mainly related to the rotation with two fingers due to the problem already described above. Remarks about the other interaction tasks and concepts were very positive, especially in relation to game play experience with words such as "cool" and "fun" being frequently used by the participants to describe their overall impression. Three participants also made positive comments about the visual feedback noting that it helped them to solve the tasks. In general, we were surprised on how quickly all participants were able to handle the interaction especially considering their lack of experience with this kind of input. Nine users noted that the concepts need some getting used to, but all agreed that only little practice is needed before they can be used in a beneficial way.

Real and virtual world integration Many people explicitly noted that the integration of real and virtual world was done very nicely as reflected by one comment noting that "the mix between AR and the natural operation feels convincing and emotionally good." In order to further explore how well the interactions really integrated into natural game play interactions on regular board games—one of the major goals of our research—we analyzed the tracking data to gain more insight into how users actually moved their hand and fingers while holding the phone. Figure 26 illustrates the average size of the markers during the scaling and rotation operations (measured in amount of recognized pixels). Green bars indicate the amount of recognized pixels for the green marker on the user's thumb, red bars the ones for the red marker around the index finger. These values are an indication on how far away the users held their hand from the camera with lower values relating to further distances. A statistically significant difference (*t*-test, $p < 0.05$) can be observed between operations in midair that are illustrated by darker colors and the ones done on the board. This verifies that for on board operations people actually moved their hand to the position on the board where the virtual object was located, although this was not necessary because hand gestures were only tracked in two dimensions. Hence, using finger-based gestures in such a setting seems to overcome one of the major problems identified in our first study, i.e. that people have a tendency to hold their fingers too close to the camera, thus resulting in a lower tracking
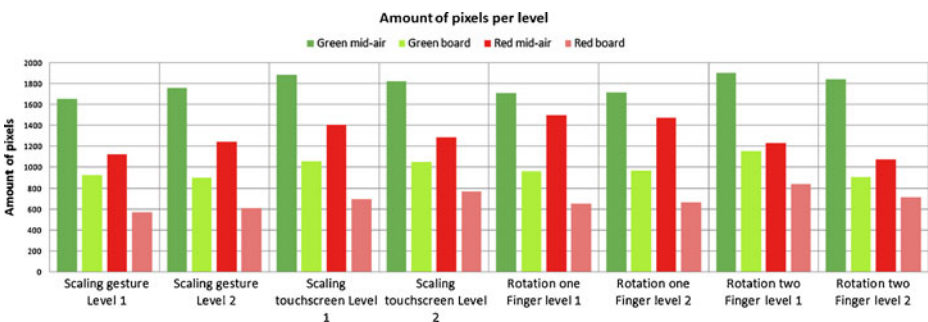

Fig. 26 Average amounts of pixel for each marker that are visible per level

performance and noisier input. It also confirms the subjective comments from the users that explicitly noted the naturalness and seamless integration of virtual and real world interaction.

5 Conclusion and future work

In our first experiment, where virtual objects were floating in midair, gesture-based interaction on mobile phones using finger tracking achieved a high level of user engagement and positive feedback in terms of fun and entertainment. However, the concept failed performance-wise thus making it unfeasible in concrete applications. Our second experiment proved that in a different setting, i.e. augmented reality board games, we can not only maintain the high level of engagement and entertainment achieved previously, but we can also achieve a performance that makes this kind of interaction useful in actual game play. Most importantly, both quantitative and qualitative data confirmed that our interaction concept nicely integrates virtual and real world interactions in a natural and intuitive way. However, when looking at the concrete implementations of the concept for different tasks, it becomes clear that trying to simulate natural interactions as good as possible does not always result in the best interaction experience. Whereas the more natural approach of using two fingers for translation was generally preferred, in both other cases the less natural one was rated higher. For scaling, users had a slight preference towards using the touch screen to stop the operation instead of a different set of finger gestures. For rotation, most users disliked the operation with two fingers and preferred the less natural, but more flexible and easier to perform one-finger-based interaction.

Reviewing these results with respect to the motivating example used at the very beginning of this article – i.e. a fantasy game with virtual characters such as flying dragons, hidden doors, etc. —our evaluation shows that interaction based on finger tracking can indeed be a valuable alternative for controlling such virtual characters. However, certain restrictions apply. For example, virtual objects should not float freely in the air but have some relation to the physical environment—such as virtual characters in a board game that are usually on the board or floating just a bit above them. Observations from our first experiment suggest that other approaches might outperform finger-based interaction considering quality measures such as accuracy and time to solve a task. However, the positive user feedback in both tests suggests additional value in areas where fun and an engaging experience is more important than pure performance values. Potential areas are not necessarily restricted to gaming, but include, for example, tourist attractions or e-learning applications in museums. Instead of just using mobile AR to passively provide people with additional information—as often done by today's applications—users would be able to actively explore these virtual enhancements of the real world.

Our work provides a first step in this direction, but further research is needed to achieve this vision. One of the most important issues is how to switch between different types of operations. In order to verify the general feasibility of the concept, we restricted ourselves to canonical interactions so far. How to best combine them, for example to enable users to seamlessly rotate and scale a virtual object, is still an open question. We did some initial tests where the system automatically distinguishes between different gestures for various tasks. First results are promising but also indicate that users might get easily confused and face a cognitive overload. Alternative approaches, such as back-of-device-interaction as proposed by [3] could provide a good alternative for task selection. When integrating several gestures into a more complex application, it might also be suitable to not purely rely on finger tracking, but complement it with traditional interaction styles, such as touch screen interaction. Hence, further studies are needed to identify what interaction mode is most suitable for a particular task and in what particular context. In relation to this, it is also important to investigate what

kinds of gestures are natural and intuitive for certain tasks, similarly to [15] and [47] who evaluated this in the context of non-mobile devices. These studies should also consider social aspects. For example, [6] reports that some users mentioned in their interviews that such an interaction might be perceived as "strange" or "weird". A recent study also showed "that location and audience had a significant impact on a user's willingness to perform gestures" [32]. Our experiments did not consider these issues because they have been done in a lab setting. Nevertheless, we got the impression that some people also felt "socially" more comfortable in the second test—where they could sit on a table—compared to the first one—where they were standing and had to make gestures in midair, thus also indicating the relevance of the user's global context. Marker-less hand and finger tracking was purposely excluded from this work because of its focus on interaction concepts. However, implementing a reliable real-time maker-less tracking on a mobile phone and evaluating the influence of the most likely noisier input on the interaction experience is an important part of our current and future work.

References

1. Andel M, Petrovski A, Henrysson A, Ollila M (2006) Interactive collaborative scene assembly using AR on mobile phones. In: Pan Z et al. (Eds.): ICAT 2006, LNCS 4282, Springer-Verlag Berlin Heidelberg, 1008–1017
2. Atchison DA, Smith G (2000) Optics of the human eye, Vol. 1, Butterworth-Heinemann
3. Baldauf M, Zambanini S, Fröhlich P, Reichl P (2011) Markerless visual fingertip detection for natural mobile device interaction. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (MobileHCI '11). ACM, New York, NY, USA, 539–544
4. Bergig O, Hagbi N, El-Sana J, Billinghurst M (2009) In-place 3D sketching for authoring and augmenting mechanical systems. In Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '09). IEEE Computer Society, Washington, DC, USA, 87–94
5. Billinghurst M, Kato H, Myojin S (2009) Advanced interaction techniques for augmented reality applications. In *Proceedings of the 3rd International Conference on Virtual and Mixed Reality: Held as Part of HCI International 2009* (VMR '09), Randall Shumaker (Ed.). Springer-Verlag, Berlin, Heidelberg, 13–22
6. Caballero ML, Chang T-R, Menéndez M, Occhialini V (2010) Behand: augmented virtuality gestural interaction for mobile phones. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (MobileHCI '10). ACM, New York, NY, USA, 451–454
7. Callahan J, Hopkins D, Weiser M, Shneiderman B (1988) An empirical comparison of pie vs. linear menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*
8. Carmigniani J, Furht B, Anisetti M, Ceravolo P, Damiani E, Ivkovic M (2011) Augmented reality technologies, systems and applications. Multimed Tools Appl 51(1):341–377
9. Chehimi F, Coulton P (2008) Motion controlled mobile 3D multiplayer gaming. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology* (ACE '08)
10. Chehimi F, Coulton P, Edwards R (2008) 3D motion control of connected augmented virtuality on mobile phones. In *Proceedings of the 2008 International Symposium on Ubiquitous Virtual Reality* (ISUVR '08). IEEE Computer Society, Washington, DC, USA, 67–70
11. Crowley JL, Bérard F, Coutaz J (1995) Finger tracking as an input device for augmented reality. In *Proceedings of the International Workshop on Face and Gesture Recognition*, Zurich, Switzerland, 195–200
12. Dorfmüller-Ulhaas K, Schmalstieg D (2001) Finger tracking for interaction in augmented environments. In *Proceedings of IEEE and ACM International Symposium on Augmented Reality, 2001*, 55–64
13. EyeSight's Touch Free Interface Technology Software: http://www.eyesight-tech.com/technology/
14. Gilbertson P, Coulton P, Chehimi F, Vajk T (2008) Using "tilt" as an interface to control "no-button" 3-D mobile games. *Comput. Entertain.* 6, 3, Article 38 (November 2008)

15. Grandhi SA, Joue G, Mittelberg I (2011) Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (CHI '11). ACM, New York, NY, USA, 821–824

16. Ha T, Woo W (2010) An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment. In *Proceedings of the 2010 IEEE Symposium on 3D User Interfaces* (3DUI '10). IEEE Computer Society, Washington, DC, USA, 91–98

17. Hagbi N, Bergig O, El-Sana J, Billinghurst M (2009) Shape recognition and pose estimation for mobile augmented reality. In Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '09). IEEE Computer Society, Washington, DC, USA, 65–71

18. Henrysson A, Billinghurst M, Ollila M (2005) Virtual object manipulation using a mobile phone. In *Proceedings of the 2005 international conference on Augmented tele-existence* (ICAT '05). ACM, New York, NY, USA, 164–171

19. Henrysson A, Billinghurst M (2007) Using a mobile phone for 6 DOF mesh editing. In *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI* (CHINZ '07). ACM, New York, NY, USA, 9–16

20. Henrysson A, Ollila M, Billinghurst M (2005) Mobile phone based AR scene assembly. In *Proceedings of the 4th international conference on Mobile and ubiquitous multimedia* (MUM '05). ACM, New York, NY, USA, 95–102

21. Huynh DT, Raveendran K, Xu Y, Spreen K, MacIntyre B (2009) Art of defense: a collaborative handheld augmented reality board game. In Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games (Sandbox '09), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 135–142

22. Hürst W, van Wezel C (2011) Multimodal interaction concepts for mobile augmented reality applications. In *Proceedings of the 17th international conference on Advances in multimedia modeling—Volume Part II* (MMM'11), Kuo-Tien Lee, Jun-Wei Hsieh, Wen-Hsiang Tsai, Hong-Yuan Mark Liao, and Tsuhan Chen (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 157–167

23. Jung J, Cho K, Yang HS (2009) Real-time robust body part tracking for augmented reality interface. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry* (VRCAI '09), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 203–207

24. Kato H, Billinghurst M, Poupyrev I, Imamoto K, Tachibana K (2000) Virtual object manipulation on a table-top AR environment. *International Symposium on Augmented Reality*, IEEE and ACM International Symposium on Augmented Reality (ISAR'00), 2000, p. 111

25. Kinect - Xbox.com, http://www.xbox.com/kinect

26. Koike H, Sato Y, Kobayashi Y (2001) Integrating paper and digital information on EnhancedDesk: a method for realtime finger tracking on an augmented desk system. ACM Trans Comput-Hum Interact 8(4):307–322

27. Kumar A, Waldron K (1981) The workspace of a mechanical manipulator. Trans ASME J Mech Des 103:665–672

28. Lee M, Green R, Billinghurst M (2008) 3D natural hand interaction for AR applications. Lincoln, New Zealand: 23rd International Conference Image and Vision Computing New Zealand, 26–28 Nov 2008

29. Mistry P, Maes P (2009) SixthSense: a wearable gestural interface. In ACM SIGGRAPH ASIA 2009 Sketches (SIGGRAPH ASIA '09). ACM, New York, NY, USA, Article 11, 1 pages

30. Mistry P, Maes P, Chang L (2009) WUW—wear Ur world: a wearable gestural interface. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* (CHI EA '09). ACM, New York, NY, USA, 4111–4116

31. Niikura T, Hirobe Y, Cassinelli A, Watanabe Y, Komuro T, Ishikawa M (2010) In-air typing interface for mobile devices with vibration feedback. In *ACM SIGGRAPH 2010 Emerging Technologies* (SIGGRAPH '10). ACM, New York, NY, USA, Article 15, 1 pages

32. Rico J, Brewster S (2010) Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the 28th international conference on Human factors in computing systems* (CHI '10). ACM, New York, NY, USA, 887–896

33. Rohs M (2004) Real-world interaction with camera-phones. In *Proceedings of 2nd International Symposium on Ubiquitous Computing Systems* (UCS 2004), Springer, 74–89

34. Rohs M (2005) Visual code widgets for marker-based interaction. In *Proceedings of the Fifth International Workshop on Smart Appliances and Wearable Computing—Volume 05* (ICDCSW '05), Vol. 5. IEEE Computer Society, Washington, DC, USA, 506–513

35. Rohs M, Oulasvirta A, Suomalainen T (2011) Interaction with magic lenses: real-world validation of a Fitts' Law model. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (CHI '11). ACM, New York, NY, USA, 2725–2728

36. Ruiz J, Li Y. DoubleFlip: a motion gesture delimiter for interaction. In *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology* (UIST '10). ACM, New York, NY, USA, 449–450

37. Ruiz J, Li Y (2011) DoubleFlip: a motion gesture delimiter for mobile interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (CHI '11). ACM, New York, NY, USA, 2717–2720
38. Schmalstieg D, Wagner D (2007) Experiences with handheld augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (ISMAR '07). IEEE Computer Society, Washington, DC, USA, 1–13
39. Seo B-K, Choi J, Han J-H, Park H, Park J-I (2008) One-handed interaction with augmented virtual objects on mobile devices. In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (VRCAI '08). ACM, New York, NY, USA, Article 8, 6 pages
40. Shaer O, Hornecker E (2010) Tangible user interfaces: past, present, and future directions. Found Trends Hum-Comput Interact 3(1&2):1–137
41. Terajima K, Komuro T, Ishikawa M (2009) Fast finger tracking system for in-air typing interface. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems* (CHI EA '09). ACM, New York, NY, USA, 3739–3744
42. Wagner D, Langlotz T, Schmalstieg D (2008) Robust and unobtrusive marker tracking on mobile phones. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (ISMAR '08). IEEE Computer Society, Washington, DC, USA, 121–124
43. Wagner D, Schmalstieg D, Bischof H (2009) Multiple target detection and tracking with guaranteed framerates on mobile phones. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality* (ISMAR '09). IEEE Computer Society, Washington, DC, USA, 57–64
44. Wagner D, Schmalstieg D (2009) Making augmented reality practical on mobile phones, part 1. IEEE Comput Graph Appl 29(3):12–15
45. Wagner D, Schmalstieg D (2009) Making augmented reality practical on mobile phones, part 2. IEEE Comput Graph Appl 29(4):6–9
46. Wang R, Paris S, Popović J (2011) 6D hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (UIST '11). ACM, New York, NY, USA, 549–558
47. Wright M, Lin C-J, O'Neill E, Cosker D, Johnson P (2011) 3D gesture recognition: an evaluation of user and system performance. In *Proceedings of the 9th international conference on Pervasive computing* (Pervasive'11), Kent Lyons, Jeffrey Hightower, and Elaine M. Huang (Eds.). Springer-Verlag, Berlin, Heidelberg, 294–313
48. Yousefi S, Kondori FA, Li H (2011) 3D gestural interaction for stereoscopic visualization on mobile devices. In *Proceedings of the 14th international conference on Computer analysis of images and patterns—Volume Part II* (CAIP'11), Ainhoa Berciano, Daniel Diaz-Pernil, Helena Molina-Abril, Pedro Real, and Walter Kropatsch (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 555–562

Wolfgang Hürst is an assistant professor at the Department of Information and Computing Sciences and a lecturer in the bachelor program "Game technology" and the master program "Game and Media Technology" at Utrecht University, The Netherlands. His research interests include multimedia systems and technologies, human-computer interaction, mobile computing, information retrieval, and computer-supported teaching and learning. Hürst has a PhD in computer science from the University of Freiburg, Germany, and a master in computer science from the University of Karlsruhe (TH), Germany. From January 1996 till March 1997 he was a visiting researcher at the Language Technologies Institute at Carnegie Mellon University in Pittsburgh, PA, USA. From March 2005 till October 2007 he worked as a teaching and research associate at the Faculty for Applied Sciences at the University of Freiburg, Germany.

Casper van Wezel holds a Master degree in Game and Media Technology from Utrecht University, the Netherlands. His research and interests during this Master programme included multimedia systems and technology, image processing, mobile computing and human-computer interaction. Casper is currently an entrepreneur, starting his own software company. In addition he is also working as a software engineer at a telecom company.