ORIGINAL ARTICLE

How to shoot yourself right with a smartphone?

Alena Tesařová[1] · Adam Herout[1] · Daniel Bambušek[1] · Vojtěch Juřík[2]

Abstract
This article presents a new use case of using handheld augmented reality to set up a smartphone camera for taking a self-photograph needed for evaluating the user's sports exercise. The problem addressed is that of settings up the rear smartphone camera so that users can take photographs of themselves performing sports poses without the need for assistance from another person. The proposed solution uses an augmented reality avatar with the same body properties as the user, which is placed into the scene so that the user can set up the camera to best capture the avatar. We also propose a straightforward mode of interaction for placing the avatar into the scene and setting up the camera for the desired purpose. We conducted a user study to explore the usability and user experience of the proposed solution. The results showed that the augmented reality visual aids, especially the avatar, enhanced the effectiveness of taking the self-photograph and that users perceived augmented reality elements favorably.

1 Introduction

An increasingly popular trend is using fitness and personal health applications to help users monitor their health and fitness goals and aid motivation in the comfort of their homes (Woźniak et al. 2020; Garbett et al. 2021; Rothkrantz 2021). Working out individually and independently has many advantages, such as time savings and gym fees. Moreover, many people are embarrassed or do not feel comfortable exercising in the presence of other people. However, people exercising at home, without a coach or trainer, often do not know if they are doing the exercise correctly. Therefore, they may turn to a remote trainer to help them maintain motivation and achieve the skills required to train effectively. In this case, it can be helpful to use a smartphone camera to take self-photographs of themselves while exercising to monitor whether the exercise is being done correctly, and based on feedback from a remote trainer, make the necessary corrections. However, the issue with this tool is how to set up the smartphone camera quickly, so the subject's whole body is optimally visible while exercising.

In this situation, the athlete is working out at home, and the space for exercise and camera placement is limited. They can be expected to own a smartphone with a rear-facing camera, as 84% of the world's population owns a smartphone (Turner 2022). They will be focusing on individual movements and want to do them correctly. With the aid of a smartphone camera, it is necessary for the athlete to see their posture *ex post*, after the workout, whatever the exercise, with maximum resolution and the highest quality. At the same time, the setting up of the camera must be as painless and intuitive as possible.

In order to achieve this, a basic approach is to place the smartphone on the ground, use a timer and capture a short video clip, adjust the camera setting, and repeat the whole process multiple times until the photograph meets the requirements. Interviews with people who do this show that people tend to use the yoga mat as a 'marker', whereby the yoga mat is lined up at the bottom of the camera image.

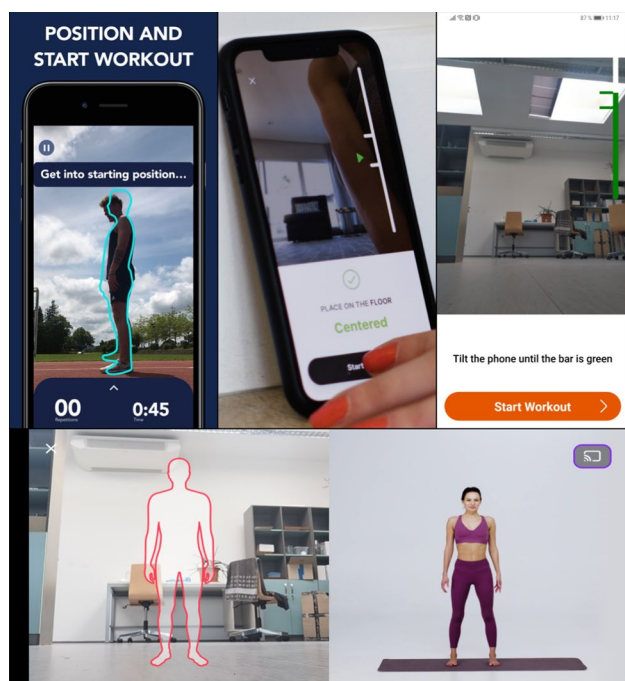✉ Alena Tesařová
  atesarova@fit.vut.cz

  Adam Herout
  herout@fit.vut.cz

  Daniel Bambušek
  bambusekd@fit.vut.cz

  Vojtěch Juřík
  jurik.vojtech@mail.muni.cz

[1] Faculty of Information Technology, Brno University of Technology, Božetěchova 2, 612 00 Brno, Czech Republic

[2] Department of Psychology, Faculty of Arts, Masaryk University, Arne Nováka 1/1, 602 00 Brno, Czech Republic

Fig. 1 Conventional methods for setting up a front-facing camera used in mobile fitness applications (person outline and mobile phone calibration), left to right: *VAY Fitness Coach*, *Onyx Home Workout*, *InfiGro: AI Fitness Personal Training Assistant*, *Zenia Yoga & Flexibility*

It serves as the initial setting, and then the iterative process continues and tends to involve fewer repetitions.

As this process is tedious and leads to many failed attempts, existing exercise applications try to help users with their user interface (UI) (see Fig. 1). The most common approach is to require the use of the front camera and display a static posture shadow (*Zenia* (2022), *VAY fitness* (2022)). This shadow is statically mounted in the center, at the bottom part of the screen. The second option is to measure the angle between the phone and the ground. This technique is used by the *Onyx Home Workout* (2022) app and *InfiGro* (2022). These methods assume that the user can view the display as they stand in front of the camera. It can be a problem as the smartphone may be difficult to see, and the user may lose focus while exercising. Also, the rear smartphone cameras typically offer better resolution, field of view, framerate, focus, and other advantages, and the light from the active display may be disturbing, especially in group settings.

In this study, we, therefore, define our problem as setting up the rear-facing smartphone camera so that the user can take a self-photograph in a specific pose. We propose a new use of augmented reality (AR) to help set up the camera. Specifically, we suggest using an AR avatar representing the user during the camera setup. To represent the user properly, the avatar should have the same body properties as the user.

This method is compared with two other variants. Section 3 contains a detailed description of the proposed augmented element together with the implementation details and overall design of the experimental app. Section 4 describes the user study and discusses the obtained results.

2 Related work

There is an increasing number of fitness and health apps on the market, and as of the last quarter of 2021, there were estimated to be over 65,000 (Ceci 2022). There have been several reviews of existing mobile fitness apps (Tavares et al. 2020; Neupane et al. 2021). Tavares et al. (2020) reviewed 36 related mobile apps from Google Play, updated between 2017 and 2020, none of which used a smartphone camera to capture the user. Capturing the user is crucial in artificial intelligence (AI) apps. Garbett et al. (2021) concentrated mainly on artificial intelligence for computer vision in fitness instructor apps. In this study, participants used three apps over eight days and compared their experiences. The study revealed one problem related to spatial limitations: the participants needed much space and struggled to set the camera up so the phone could track them. Spatial limitations and AI requirements result in the user prioritizing the camera's viewpoint rather than their own well-being.

Augmented reality and virtual reality (VR) for visual feedback and scene prototyping are not new. They have been used in mobile apps as motivation tools, in entertainment, and in marketing. One of the most well-known use cases was developed by *IKEA*. With the aid of AR visual feedback, it is possible to place furniture in into space in the correct scale and view through a smart device what the furniture would look like (Ozturkcan 2021). This opportunity to allow the customer to see a virtual replica of an actual product attracted many other retailers by offering a more direct and engaging experience (Swilley 2016; Cehovin and Ruban 2017).

A study by Alturki and Gay (2019) conducted a systematic review of AR and VR in mobile apps. They studied their use in tourism, transportation, and education. They concluded that AR and VR technology is likely to have a positive effect on the field of fitness, which encouraged them to develop a new fitness app using elements of AR.

Uchiyama and Saito (2007) created an AR support system for pool games that can simulate ball behavior through the LCD of a camera, which illustrates another practical use case for AR visual aid. AR supporting elements can also help children in education. For example, AR-Maze is a programming tool based on AR that helps children learn loop logic, parameters, and other techniques in an intuitive way (Jin et al. 2018; Kim and Shim 2022). AR elements are used in
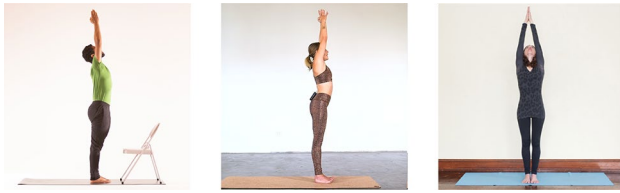
Fig. 2 The goal is for the user to take a self-photograph when practicing yoga. The body should be upright and centered, and the space below feet and over fingers should be small but not zero

sports such as basketball, rock climbing, air hockey, table tennis, golf, skiing, etc. (Soltani and Morice 2020).

Recently, frameworks such as Google ARCore[1] and Apple ARKit[2] have enabled the fast and easy development of AR applications for smartphones and tablets. Both deliver mandatory functionalities for AR using their closed-source implementation of visual-inertial simultaneous localization and mapping (Liu et al. 2018; Taketomi et al. 2017; Terashima and Hasegawa 2017), and both have their strengths and weaknesses (Nowacki and Woda 2019). However, they are usable for simple, small-scale environments (Feigl et al. 2020) and noncomplicated use cases only, as hologram drifting can often rise to 31 cm in challenging scenarios (Scargill et al. 2021).

3 Proposed augmented reality technique for shooting oneself

Figure 2 illustrates the objectives for taking a self-photograph of an exercise. The user needs to shoot themselves without ever departing from the frame in any direction, but at the same time, they want to make the best use of the image's resolution, avoiding unnecessary margins. We focus on square images because yoga (and many other sports) includes both 'vertical' and 'horizontal' poses. Nevertheless, the proposed method can easily be used for non-square images without limitations.

We informally and qualitatively interviewed athletes who had already taken pictures and videos of themselves while exercising. They typically used a built-in smartphone camera app, and to set up the camera properly, they recorded short videos of themselves walking to the spot and standing in the position, then played the video back, looking at themselves, assuming a pose similar to that in Fig. 2. Having obtained this feedback for the camera setup, they adjusted the camera and repeated the action. An incorrect setup would lead to

useless camera sessions (typically with parts of their bodies out of view) and user frustration.

The interviewed athletes tended to take advantage of the yoga mat and set up the camera reasonably well by aligning the frame with the mat at the bottom of the view and with empirically located margins on the left and right sides. In cases where the camera was set at different heights (on the floor, on a tripod at waist level, at shoulder level), the camera still needed several iterations of adjustment in the vertical direction.

This common practice of the athletes has been replicated in a very rudimentary AR solution proposed as a baseline reference, i.e., indicating the desired position of the yoga mat in the camera stream by rendering a fixed (though parameterizable) trapezoid shape (see Fig. 5 middle image).

3.1 AR avatar and the mode of interaction

We propose using AR to solve the problem of setting up the camera to capture sports poses. The user/athlete will have an AR avatar of the same height and other body proportions as a placeholder for prototyping the shot. The avatar can be relatively static or more dynamic, which means making motions that well represent the possible movements to be captured. The avatar is expected to provide direct and intuitive visual feedback toward the following criteria:

1. Where it is placed, i.e., the location within the space.
2. The extent of the avatar's limbs and their relation to the boundaries of the image, most importantly to the top and bottom margins, but also to the left and right ones.
3. The posture of the avatar should be upright.
4. The posture should be centered in the image.
5. No interaction with furniture and other obstacles, i.e., verifying that the posture is not occluded and does not collide with the environment.

The avatar must be parameterized by its height, corresponding to the user's height. The current experiments use this one parameter. Optionally, the avatar can be parameterized with more parameters, such as ratios of lengths of limbs, torso, head, and approximate weight.

We propose the following manner of interaction (Fig. 5 shows sample screenshots). The avatar is always present in the (augmented) camera stream, i.e., it does not 'appear and disappear'. The interaction is, at any given moment, in either of the following two modes:

- *Placement*—The avatar is always prominent in the camera field of view. It is horizontally centered with the camera stream with its feet in the bottom part of the field of view, placed on the ground as detected by AR technology. By moving (by translation and rotation) the hand-

held device, the user places the avatar in the room at their chosen location.

- *Interaction*—Once the avatar is placed in a suitable location, it is fixed in world coordinates and the camera is moved to correctly capture the avatar in the camera stream, meeting the criteria enumerated above.

The only user control of the interaction is by switching between these modes. In this experimental application, the switching is performed by touching any place in the camera stream view, thereby avoiding the need for a dedicated UI element. The two modes are visually distinguished. In the *placement* mode, the avatar is grayed and semitransparent to indicate that the body is at a tentative location and that the UI's primary goal has not yet been resolved. Once the avatar is fixed in its location, the *interaction* mode is indicated by a full-color and opaque avatar.

The task of setting up the camera is completed by the athlete taking an actual self-photograph to verify and possibly correct the setup. This step is equivalent to taking a self-photograph in the alternative approaches (i.e., the use of just the camera without any visual aid or a visual trapezoid indicating the suggested location of the yoga mat). The idea behind the use of the avatar is that it should keep to a minimum the number of iterations of taking a self-photograph/reviewing it/adjusting the camera.

3.2 Implementation details

The prototype smartphone application was implemented in the Unity 3D engine,[3] and the AR part is based on the AR foundation library.[4] We chose the AR foundation because it encapsulates both the *ARCore* and the *ARKit* frameworks for mobile AR and therefore allows for multi-platform development (Linowes 2021). Moreover, it is actively maintained and continuously developed and is reliable enough for simple use cases and small-scale environments (Feigl et al. 2020; Scargill et al. 2021).

The application workflow is as follows:

1. Device tracking and the detection of horizontal planes are established by the AR foundation.
2. By moving the camera in a physical environment, planes are continually detected and visualized to the user by the white dotted polygon.
3. In each frame, a ray is cast from the center of the camera into the scene to determine the placement of the avatar.
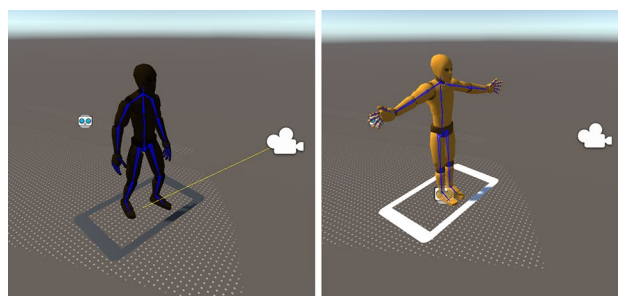
Fig. 3 Avatar placement procedure. Left: the avatar is placed at the intersection of the ray emitted from the camera and the plane detected by the AR foundation (visualized by the white dotted polygon). Right: when the avatar is grounded, it is colored, and a local anchor is created to improve the avatar's tracking
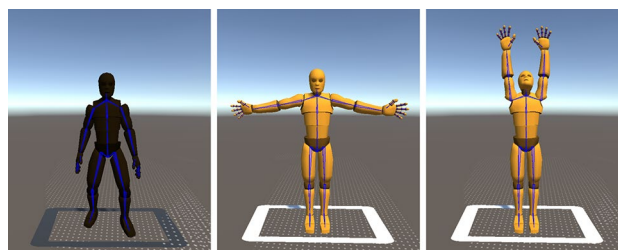
Fig. 4 Avatar animation states. Left: the idle state, which is active during the *placement* mode. Middle and right: the stretching state, which is active during the *interaction* mode. The avatar is smoothly transitioning between these two stretching poses
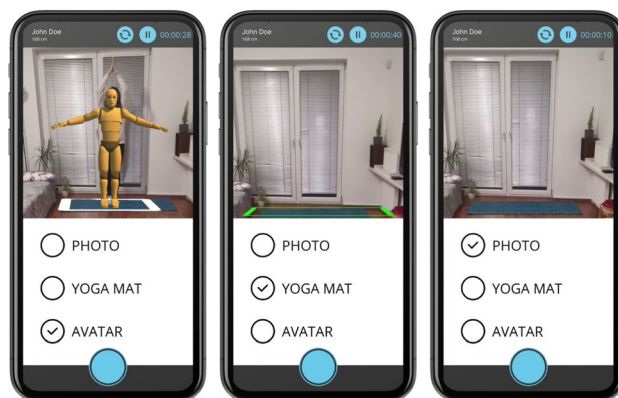
Fig. 5 The experimental application supports the three methods of taking photographs. Left: AR avatar, middle: yoga mat, and right: no visual aid. The app also tracks the time it takes the participants to perform the given task

4. If a ray hits the ground plane detected by the AR foundation, the avatar is displayed at the intersection of the ray and the plane. The avatar is grayed out, rotated toward the camera, and is in an idle state of animation (see Fig. 3).

5. By casting a ray in each frame, the avatar appears to be moving on top of the ground plane as the camera moves.

6. When the user touches anywhere on the screen, raycasting is paused, the avatar is grounded, colored, and transitions to the stretching states (see Figs. 3 and 4). A local anchor is created at the avatar's position to improve its tracking.

7. When the user touches anywhere on the screen, the anchor is deleted, and the procedure goes back to step 3.2.

The avatar model and the idle animation are taken from the *Basic Motions FREE* asset, available publicly at the Unity Asset Store.[5] The model is fully rigged. The rest of the animations were created directly in Unity. When in the *placement* mode, the avatar's animator plays the idle animation in a loop. When in the *interaction* mode, the animator plays the stretching animation in a circle (see Fig. 4). This animation smoothly transitions between the skeleton's T-pose (which lasts approximately 4 s) and Y-pose (which lasts about 7.5 s). When in the Y-pose, the avatar flutters its fingers to create more life-like credibility. The complete animation, including transitions between poses, is 15.07 s long.

3.3 Experimental app design

To evaluate the proposed AR solution of shooting oneself, we designed an experimental application with three different modes of visual aid: the AR avatar, yoga mat screen overlay, and no visual aid. These can be seen in Fig. 5. The camera view is intentionally square because we wanted to make sure that when the user has set the camera so that they are fully visible in the vertical dimension, they will also be visible in various horizontal positions. Unfortunately, the AR foundation does not currently provide access to controlling the viewport ratio of the camera. It automatically stretches the camera's image to fill the device's whole screen. We achieved the squared view by overlaying a part of the view with UI elements. This is not an ideal solution for a production app, as the edges of the camera view are unnecessarily cut off. Nevertheless, it is sufficient for an experimental evaluation.

The main element of the UI is a squared camera view, under which three radio buttons are located for selecting one of the three visual aid modes—Photo (A), Yoga mat (B), and Avatar (C). The first mode does not contain any visual aid. Mode B displays a green isosceles trapezoid as a screen overlay. This trapezoid represents the desired position for placing a yoga mat. It is assumed that by fitting an actual
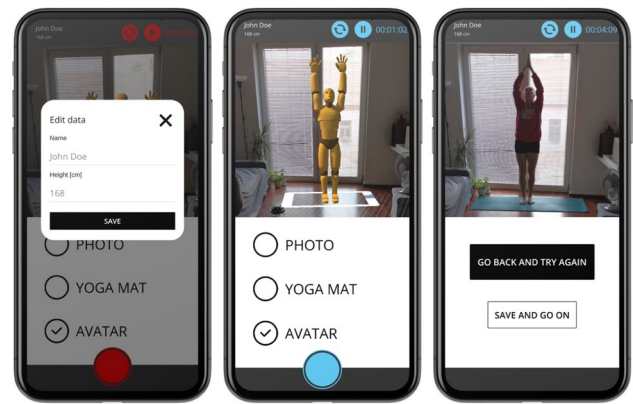
Fig. 6 Experimental app UI. Left: a modal window for changing the participant's name and height. Middle: standard application view displaying the AR avatar. Right: confirmation window that allows the user to either repeat the shot or save the photograph as the final attempt

yoga mat into the overlaid trapezoid, the user would be sufficiently visible during the exercise. Mode C is the avatar mode that displays the AR avatar and complies with the placement and interaction procedures described in Sect. 3.1.

Several additional UI elements serve solely for gathering the necessary participant information and evaluating the experiment. These are inputs for the participant's name, height, and buttons for controlling the stopwatch. The height parameter scales the avatar to match the user's approximate size. The lengths of the subject's limbs are not taken into account at the moment. Lastly, there is a big blue button for capturing the photograph located at the bottom of the screen. Once pressed, a confirmation screen is displayed with the image taken, where the user can either repeat the shoot or save the photograph as the final attempt (see Fig. 6). The photo capture is implemented as a screenshot where the visual aid elements are hidden. This screenshot is cropped to the size and position of the camera view.

The application saves all captured photographs to the device's persistent data storage along with metadata built into the photograph's filename. These are the participant's name and height, the visual aid mode, the time it took to take the picture from the beginning of the task (in seconds), a complete timestamp of the photo capture, and an indicator of the final attempt.

4 User study and results

The goal of the presented user study was to evaluate the proposed AR-based method alongside the two baseline variants of the method commonly used by athletes (Fig. 5).

5 https://assetstore.unity.com/packages/3d/animations/basic-motions-free-154271.

Fig. 7 Visualization of the testing process. There were 56 participants in total, and the experimental procedure was identical for each, apart from the order of variants. The first method tested was pseudorandomly selected, and the remaining variants were subsequently presented in the given order: A—no visual aid, B—yoga mat, C—AR avatar
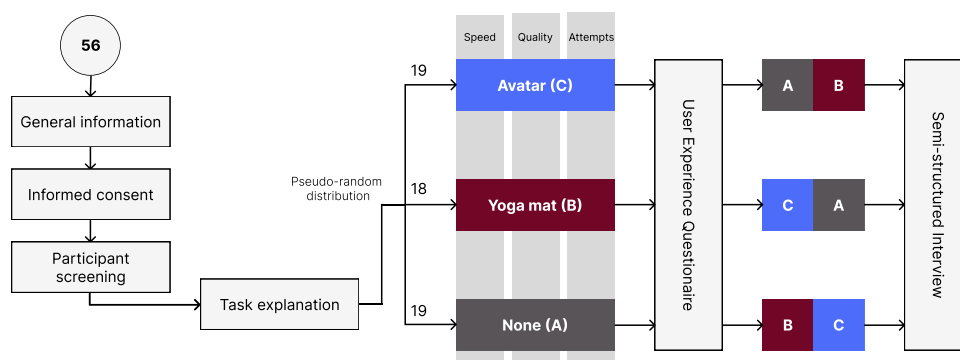
4.1 Methodology of the experiments

The user testing followed a controlled experimental approach. All the experimental runs were conducted in the same university classroom with the same experimental setup and conditions (e.g., light conditions and room setting) for all participants. We used a between-subjects experimental design (Passer 2021) because the study aimed to explore the first user encounter with the visual aids in question, and it was necessary to prevent a confounding issue (specifically the learning effect caused by practice in taking the photographs and the anticipated "attractiveness" of the added visual aids, including the virtual avatar). The AR avatar was a qualitatively distinctive functional element, which was expected to attract significant attention per se, and with a repeated measures design, the presence of AR elements would have biased the data. Since counterbalancing would not have prevented users from feeling that the AR elements were something that the study had initially set out to explore, we primarily present between-subjects comparisons as the main outcome of the study.

Objective performance was captured in the first trial, where it was possible to make the between-subjects comparisons separately for three different groups of users without confounding. Subsequently, user comfort when using a specific variant was captured with the use of the User Experience Questionnaire, Version 8 (Schrepp 2019), where participants were asked to evaluate 22 individual aspects (ease, speed, excitement, motivation, etc.) of the process of setting up the camera and creating self-photographs.

The assessment of subjective user preferences between the variants was supplemented by within-subjects measurements (Passer 2021) of the remaining variants, where counterbalancing was applied. In this regard, after the first key measurement in the assigned group (A/B/C) was taken, each participant complemented the experiment by also trying the remaining two variants in a predefined order (A–B–C; B–C–A; C–A–B), in order to obtain the relative user experience from all the available variants. The specific preferences and opinions or comments on all the experienced variants were recorded within a final semi-structured interview and then further analyzed.

4.1.1 The experimental procedure

The experimental procedure was maintained identically for all participants except for the order of presented variants; the general scheme of the procedure is depicted in Fig. 7.

Participants were invited to take part in the experiment via social media, e-mails, and personal contacts. After arriving in the lab, the participants were welcomed by the experimenter and seated. Subsequently, the general purpose of the experiment was explained, and participants were given relevant information regarding the procedure so they could read and sign the informed consent. They were informed that their participation was voluntary and that they could leave the experiment whenever they wanted without any negative consequences. Participants then completed the questionnaire with basic demographic information (age, gender, and field of study) and were pseudorandomly assigned to the experimental conditions (A/B/C). After this, the participants were instructed to use the given device with the specific functionality (i.e., AR elements/no visual aids) to produce self-photographs of a certain quality. The quality parameters were precisely specified (Sect. 4.1.2) and described to participants, and they were shown reference photographs. Although participants were told that they could take as many attempts as necessary to produce the final photograph, the quality of the outcome was crucial, we urged them to be time-efficient, i.e., to produce a photograph of the requested quality in the shortest possible time.

In the next step, participants were given brief instruction on how to use the device (an ordinary smartphone with a camera with the interface described in Sect. 3.3 and a tripod holder). After this training, participants were asked to take satisfactory photographs of themselves, which fulfilled the required parameters. After a number of attempts, participants announced that they had produced the final photograph, and the experimenter saved the output. The participant then evaluated the photograph-taking process in
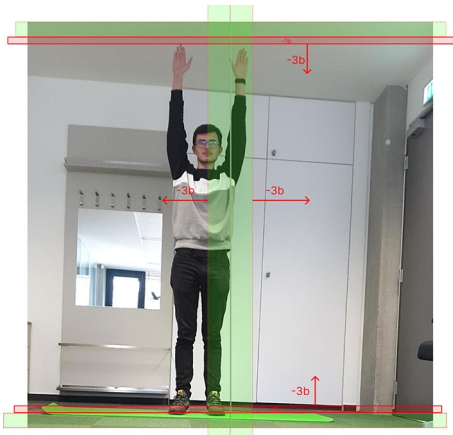
Fig. 8 Example of a photograph evaluation. The maximum number of points was 11. This photograph would be penalized by 3 points because the space between the fingertips and the top of the screen is too broad and by an additional 3 points because it is not precisely centered. Overall, this photograph would be awarded 5 points

the questionnaire. After completing the questionnaire, the participant underwent two identical trials with the other two variants of visual aids. This was followed by the questioning of participants in the framework of a semi-structured interview, where the experimenter asked them to choose their most and least preferred variant, their reasons for their evaluation, any other comments on the experiment as a whole, and their ideas for carrying out the given task in new ways.

4.1.2 Objective measures of usability

The objective measures of usability, such as speed and quality of the produced photographs, were captured and assessed. Specifically, the time requirement, the number of attempts it took to produce the requested picture, and the quality of the produced photographs produced (on a scale of 0–11 points) were measured for all participants for all three variants. The time requirement and the number of attempts were recorded automatically by the application. The quality of the photographs was later evaluated on the basis of an evaluation scheme developed for this purpose. The scheme was based on the following three rules, the violation of which was penalized by a deduction of points from the maximum score (11 points per photograph) (Fig. 8):

1. The photograph should be centered.
2. The photograph should contain the participant's whole body (parts of hands and legs must not be out of the picture),
3. The space of the photograph should be used effectively (no significant spaces above or below the participant's body).

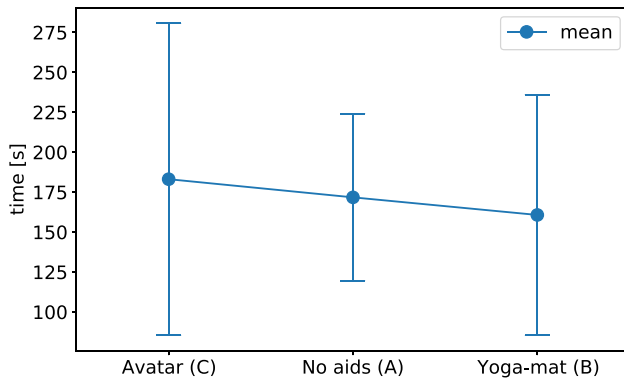### 4.1.3 User experience questionnaire

Participants' subjective evaluations of the visual aids and further comments on the process of taking the photographs were also recorded. Since the goal of the user study was to assess the subjective user experience of participants using various visual aids, user experiences were explored in detail by means of the User Experience Questionnaire, Version 8 (Schrepp 2019), which contained 22 separate categories measured on a Likert scale (1–7) (Jebb et al. 2021) where the low values represented low levels of the parameter. The list of categories is presented below:

1. Annoying/enjoyable
2. Incomprehensible/understandable
3. Dull/creative
4. Difficult to learn/easy to learn
5. Inferior/valuable
6. Boring/exciting
7. Not interesting/interesting
8. Slow/fast
9. Conventional/inventive
10. Obstructive/supportive
11. Bad/good
12. Complicated/easy
13. Unlikable/pleasing
14. Usual/leading edge
15. Unpleasant/pleasant
16. Demotivating/motivating
17. Inefficient/efficient
18. Confusing/clear
19. Impractical/practical
20. Unattractive/attractive
21. Unfriendly/friendly
22. Conservative/innovative

The evaluation procedure culminated in a semi-structured interview, where participants were interviewed about their overall experience with all variants and were asked to comment on their preferences and on the possible direction of the future development of the app.

4.1.4 Participants

All participants in the study were students of local universities, typically from the fields of information technology and psychology. The total number of participants in the study was 56 (46 males, 10 females), aged between 19 and 34 years (*mean* = 22.6 years; SD = 3.84). The participants were randomly assigned to the three experimental conditions in approximately the same number and ratio of individuals for each gender and field of study in each condition.

Fig. 9 Participants' response times for each condition

Fig. 10 Quality of the final photographs for each condition

Fig. 11 Number of attempts needed to produce a satisfactory photograph for each condition
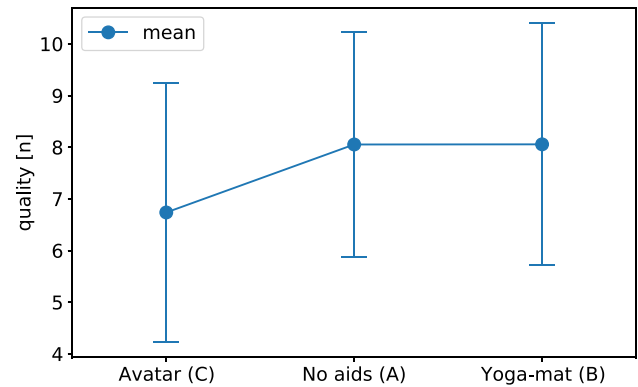
4.1.5 Data analysis

Since Levene's test of homogeneity (Ruscio and Roche 2012) in combination with a visual inspection of the normality distribution of all analyzed data, indicated violations of F-test assumptions, the Kruskal–Wallis H-test (Bewick et al. 2004) was used as a nonparametric alternative for one-way analysis of variance in the three different conditions (A, B, C) as a between-subject factor. In the analysis, the omega-squared (ω^2) effect size was reported due to the small sample size used (Okada 2013). The data were analyzed with the use of JASP (version 0.16.1) and Python (Pandas library version 1.1.4).
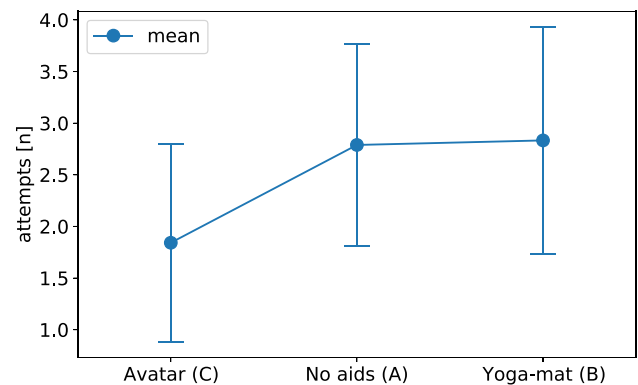
4.2 Results—between-subjects experiments

This part analyzed the first experience with the photograph-taking process with a specific variant (A/B/C). Since the analysis was always done on a single mode of interaction, the results can be compared between subjects. This contrasts with Sect. 4.3, where the participant has already been 'exposed' to all the modes of interaction, which allow for a comparison between them within one subject but invalidate the quantitative comparison between subjects.

4.2.1 Usability: time

The Kruskal–Wallis test showed that there were no differences in the time required when producing photographs with the different modes A/B/C ($H(2) = 0.520$; $p = .771$) with no effect ($\omega^2 = .00$). Participants in the first ($M = 171.737$, SD = 52.23), second ($M = 160.722$, SD = 75.23), and third condition ($M = 183.105$, SD = 97.67), did not differ in the time required to complete the given task (see Fig. 9).

4.2.2 Usability: quality of the captured photos

Regarding the quality of produced photographs, the Kruskal–Wallis test identified no differences between the experimental conditions ($H(2) = 2.974$; $p = .226$) with a small effect ($\omega^2 = .034$). As seen in Fig. 10, participants in the second condition using the augmented virtual avatar (variant C) scored slightly worse ($M = 6.737$, SD = 2.5) than two other groups, namely variant B with the virtual yoga mat frame ($M = 8.056$, SD = 2.34) and variant A with no additional visual aids ($M = 8.053$, SD = 2.17).

4.2.3 Usability: number of attempts of shooting oneself

The Kruskal–Wallis test demonstrated that, as for the number of attempts needed to complete the task, participants using the virtual avatar (variant C) needed significantly fewer ($M = 1.842$, SD = 0.96) to produce the final outcome of satisfactory quality ($H(2) = 10.822$; $p = .004$) with a large effect ($\omega^2 = .146$), when compared to the control variant A ($M = 2.789$, SD = 0.976) and the yoga mat variant B ($M = 2.833$, SD = 1.10). As seen in Fig. 11, the number of
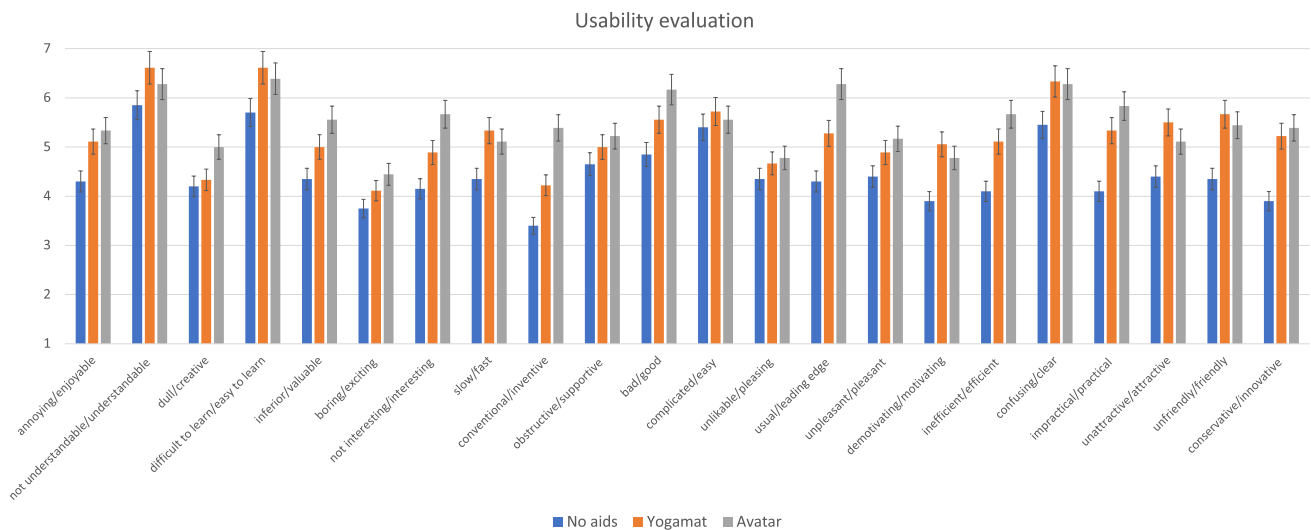
Fig. 12 User experience questionnaire evaluation. Blue: no visual (A), orange: Yoga mat (B), gray: Avatar (C). In 22 different categories exploring user experience, participants consistently evaluated AR variants (especially the avatar variant) more favorably (colour figure online)

attempts taken to achieve the required quality of the picture was similar for variants A and B but was significantly lower for variant C. Dunn's *post hoc* comparisons (Dunn 1964) identified a significant difference between the variants A (no aids) and variant C (avatar) in favor of the avatar ($z = 2.820$; p(Holm) < .01); and between variant B (yoga mat) and C (avatar) in favor of the avatar ($z = 2.865$; p(Holm) < .01).

4.2.4 Subjective user experience: first experience

The measures of objective performance (usability) were complemented by detailed subjective evaluations obtained by means of the User Experience Questionnaire (Schrepp 2019), Version 8, which was adapted for our purposes (4 of the 26 individual categories measured by the UEQ were removed before use since they did not apply to our construct. They were: *unpredictable/predictable, secure/not secure, met expectations/did not meet expectations, and organized/cluttered*). As seen in Fig. 12, in the 22 different categories exploring user experience, the participants consistently evaluated the AR favorably (especially the avatar variant). This clearly testifies to the subjectively perceived feeling of user comfort when working with an application with advanced visual aids such as AR elements.

4.3 Results—within-subject measurements and interviews

#### 4.3.1 Subjective user experience: relative comparison of the approaches

After participants had completed the process of taking photographs in all three modes, they were questioned about their

overall experience with each variant. In this relative evaluation, the majority of subjectively reported preferences were for the virtual avatar variant, which was in accordance with the fewer attempts participants needed to reach a satisfactory outcome and was confirmed by the verbal comments expressed in the semi-structured interview. Here, more than two-thirds of participants reported that the virtual avatar was the best option for taking a self-photograph (see Fig. 13).

4.3.2 Semi-structured interviews

The experiment ended with the semi-structured interview, where various aspects of the users' experiences of the variants were discussed. Participants reported their subjective preferences for the variants and the specific functions they found helpful in the process of taking the photographs. Based on the transcription of the interviews, we conducted a thematic analysis (Smith 2015) and identified several topics of interest specified in Table 1.
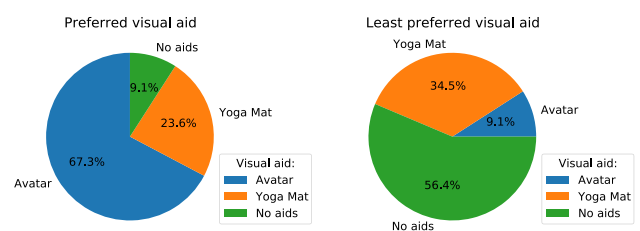
Fig. 13 Overall preferences for the variants. The chart on the left shows that the avatar variant was most preferred by participants. The right chart indicated that the variant with no aids was the least preferred

Table 1 Semi-quantification of the thematic analysis: subthemes in the transcription of the interview with participants

Major theme		Subtheme	No. of partic.	% of total p.
What was the biggest problem with the whole process?	1	Problems with the visual aids (avatar, yoga mat)	11	20.4
	2	Perfectionate details	6	11.1
	3	Tripod setup	17	31.5
	4	Problem with the estimating distance	14	25.9
What helped you to take the photograph the fastest way?	1	Visual aids in app (avatar, yoga mat)	17	31.5
	2	Furniture in the room	9	16.7
	3	Ceiling and the floor	8	14.8
	4	Previous experience with tripod and taking photographs	4	7.4
What should the ideal app have that will allow you to take a self-photograph on the first try and in the shortest time that fulfills all the conditions?	1	Add voice and visual feedback for error indication	17	31.5
	2	Add cameras to see result from different angles	2	3.7
	3	Automatic cropping, zooming, and centering final photo	12	22.2
	4	Automatic avatar placing on the mat	3	5.5
	5	Use remote control for taking the photograph (using a smart watch)	3	5.5

The participants were first asked about the most significant problem during the photograph-taking process. In addition to reported difficulties using the tripod, some participants complained about the visual aids, namely the yoga mat and avatar. One participant said: "*The avatar moves his arms around, and you have to wait until he stops moving, which is annoying.*" Another commented: "*I struggled to place the avatar on the yoga mat.*" There were also comments about the yoga mat frame. For example, one participant said: "*The frame of the mat did not fit with reality.*"

Secondly, we inquired about elements that had helped the participants capture the photograph. In this case, most participants mentioned the visual aids in the app (avatar, yoga mat). One participant commented: "*With the avatar, I was sure of how I would look exactly.*" Another said: "*It was best with the avatar when estimating distance so that there was not much space above the hands.*"

The final question was not related to the application, it was aimed at collecting suggestions for future research and getting interesting ideas from the participants. We asked them to imagine an ideal application for taking a self-photograph that met all the requirements on the first try. Over 31% of participants suggested adding verbal or visual feedback for error indication so that instead of going to the camera physically, they would simply move in the proposed direction by listening to the app's commands. Another interesting idea was to add automatic image cropping, zooming, and centering. This solution proposes that if the athlete wants the camera to track them while exercising, they just place the phone on the ground, and the app manages all the cropping and zooming autonomously.

In the qualitative part of the questionnaire, the respondents also frequently mentioned the use of the features of the room. It was proposed that the situation would be much more difficult outdoors or in a large room with a high ceiling, and we would like to explore these contexts. Moreover, the participants often "struggled" with the tripod since the smartphone was placed close to the ground, and manipulating was not very easy. We are currently looking for better tripods or other more convenient tools for holding the smartphone and setting it up. Another problem that the participants mentioned was related to the avatar, insofar as the animation was long and slow, and they had to wait for the avatar to raise his hands to the highest position. We are going to explore more suitable animation sequences that would not lead to such frustration.

5 Discussion

The data for all variants showed a similar distribution in relation to the length of time it took participants to take the self-photographs; therefore, no significant trend was identified regarding time. Moreover, the quality of the final photographs was also found to be similar for all variants. However, the number of attempts it took to produce a satisfactory picture differed significantly ($p = .004$), as the AR avatar significantly reduced the number of attempts required to produce a good self-photograph. The subjective self-reported preferences supported the use of advanced AR elements, primarily the virtual avatar, since people evaluated the photographs-taking process with the aid of the virtual avatar more favorably (from the detailed analysis of the User Experience Questionnaire) than with the aid of the other variants. Moreover, they also described it as subjectively the most preferred variant.

On the other hand, the least preferred variant was the one with no visual aids. In the semi-structured interviews, participants consistently reported the AR elements to be promoting, motivating, and effective, and that they would prefer using an interface with AR for such a photograph-taking task. The observed contradiction regarding the fact that the most helpful elements for the participants in taking a photograph (AR visual aids) were also frequently criticized was probably due to the fact that the AR visual aids are a significant aspect of the app; therefore, the participants felt that they should comment on them. In the frame of questions asked within the semi-structured interview, participants tended to comment primarily on the AR elements with regard to their functionality and user comfort. Participants also made several suggestions for future AR interface development, such as voice and visual feedback and automatic cropping, zooming, and centering, which may be valuable for future research.

At first glance, the results of this study are surprising since we initially expected the elements of augmented reality to influence objective performance in the self-shooting process, especially in terms of the time required and the quality of the pictures. However, participants were shown an example of the ideal outcome of the self-shooting process, and therefore participants tended to generate a final photograph of adequate quality, and thus the overall quality of the photograph produced was generally similar. On the other hand, it was surprising that the time required to complete the task did not significantly differ; there was no difference between the variants with augmented elements and the variant with no visual aids at all, which from our perspective was initially disappointing.

However, the number of attempts needed to produce a satisfactory photograph was partially based on participants' subjective feelings since it was their decision when to stop trying, i.e., once they were subjectively satisfied. The data showed that the number of attempts in the virtual avatar condition was significantly lower ($p = .004$) than in the other variants. This means that the participants systematically reached satisfactory results earlier (in terms of the number of attempts) using the virtual avatar visual aid. However, the time per se remained the same for all variants since the procedure with the virtual avatar involved more steps compared to the other variants. It is logical to assume that what makes a process tedious and unpleasant is not the time itself but the necessity to repeat the same task over and over again. Therefore, the aim of reducing the actual *time* it takes to complete a task may not be the most pertinent goal.

The detailed subjective evaluation of user experiences was generally more favorable toward variants with AR visual aids. After a participant's first experience with the photographs-taking process with one of the three variants, those using the AR-based interfaces evaluated these more

positively on almost all the subscales of the UEQ. Moreover, taking into account their relative experience of all the variants, the majority of users (almost 70%) reported the avatar variant as preferable.

5.1 Limitations of the study and future research

There are several limitations of this study that should be taken into consideration in the future research. Firstly, as participants mentioned in the semi-structured interviews, the room format helped people reference their spatial position and achieve the outcome. They referenced the proportions of windows, doors, and furniture to aid the setting of the camera. This limitation was partially addressed by the between-subjects experimental design, where conditions were identical for all groups. However, in the future research, it would be beneficial to use a larger room, such as a gym or a hall, or even an outdoor location, where proximity cues would not apply when taking the photographs.

Furthermore, in an attempt to simplify the experimental process, we decided to use a mini tripod, which participants sometimes perceived as bothersome to set up. Manipulation with the smartphone camera was difficult because it was close to the floor. Subsequent research could be improved by using a 'proper' tripod or, at least, experimenting with three different sizes: small (a mini tripod as used in our study), medium (waist height), and high (shoulder height). The different heights of the camera might bring new requirements for the adjustment process and could possibly reveal new aspects of the three methods.

Despite the effects observed within this experimental design, conclusions from research conducted on a sample of university students from specific fields cannot be directly generalized to the general population since there may be a difference in performance due to different technical skills or expertise. This is also considered a limitation of this study.

The comments of the participants highlighted several potentially fruitful ideas for the future development of the tool, especially voice and visual feedback and automatic cropping.

6 Conclusions

We proposed a new use case for augmented reality: setting up a camera for "shooting oneself" for evaluating sports training. The proposed user interface was very simple, with no additional user interface elements, only switching between two modes by tapping anywhere on the camera stream.

We anticipated that this user interface would reduce the time required for setting up the camera. However, this assumption was not confirmed by the experiment. There

was no significant difference between the three variants (the avatar variant was the slowest, but the difference was not significant). Nevertheless, the avatar method required the least number of attempts and was perceived as the least demanding and the most pleasant method.

Participants in the study indicated that they used different proximity cues, which were relatively abundant in the experimental room. This might have simplified the setup process considerably, especially for the baseline variant without visual aids. We intend to conduct a similar study outdoors or in a room much larger and without similar visual cues. It is interesting to explore what visual cues are effective when taking a self-photograph ("shooting oneself") and how augmented reality can make it even more effective.

Data availability Data sharing is not applicable to this article, as no datasets were generated or analyzed during the study.

Declarations

Conflict of interest The authors declare they have no conflict of interest.

References

Alturki R, Gay V (2019) Augmented and virtual reality in mobile fitness applications: a survey. Springer, Cham, pp 67–75

Bewick V, Cheek L, Ball J (2004) Statistics review 10: further nonparametric methods. Crit Care 8(196):2857. https://doi.org/10.1186/cc2857

Ceci L (2022) Number of mhealth apps available in the google play store from 1st quarter 2015 to 1st quarter 2022. https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/

Cehovin F, Ruban B (2017) The impact of augmented reality applications on consumer search and evaluation behavior. Master's thesis, Copenhagen business school, Denmark, Denmark. https://research-api.cbs.dk/ws/portalfiles/portal/58424530/federica_cehovin_og_bernice_ruban.pdf

Dunn OJ (1964) Multiple comparisons using rank sums. Technometrics 6(3):241–252. https://doi.org/10.1080/00401706.1964.10490181

Feigl T, Porada A, Steiner S, Löffler C, Mutschler C, Philippsen M (2020) Localization limitations of arcore, arkit, and hololens in dynamic large-scale industry environments. In: VISIGRAPP (1: GRAPP), pp 307–318

Garbett A, Degutyte Z, Hodge J, Astell A (2021) Towards understanding people's experiences of ai computer vision fitness instructor apps. In: Designing interactive systems conference 2021, DIS'21, New York, NY, USA. Association for Computing Machinery, pp 1619–1637

Infivolve, Inc. (2022) infigro: Ai fitness personal training assistant on Google Play. Accessed 20 Apr 2022

Jebb AT, Ng V, Tay L (2021) A review of key Likert scale development advances: 1995–2019. Front Psychol. https://doi.org/10.3389/fpsyg.2021.637547

Jin Q, Wang D, Deng X, Zheng N, Chiu S (2018) Ar-maze: a tangible programming tool for children based on ar technology. In: Proceedings of the 17th ACM conference on interaction design and children, IDC'18, New York, NY, USA. Association for Computing Machinery, pp 611–616

Kim J, Shim J (2022) Development of an ar-based ai education app for non-majors. IEEE Access 10:14149–14156. https://doi.org/10.1109/ACCESS.2022.3145355

Linowes J (2021) Augmented reality with unity AR foundation: a practical guide to cross-platform AR development with Unity 2020 and later versions. Packt Publishing, Birmingham

Liu H, Chen M, Zhang G, Bao H, Bao Y (2018) Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1974–1982

Neupane A, Hansen D, Fails JA, Sharma A (2021) The role of steps and game elements in gamified fitness tracker apps: a systematic review. Multimodal Technol Interact. https://doi.org/10.3390/mti5020005

Nowacki P, Woda M (2019) Capabilities of arcore and arkit platforms for ar/vr applications. In: International conference on dependability and complex systems. Springer, pp 358–370

Okada K (2013) Is omega squared less biased? A comparison of three major effect size indices in one-way anova. Behaviormetrika 40:129–147. https://doi.org/10.1016/j.compedu.2020.103923

Onyx Inc. (2022) Onyx: Home workout on App Store. https://apps.apple.com/us/app/onyx-home-workout/id1440639203. Accessed 20 Apr 2022

Ozturkcan S (2021) Service innovation: using augmented reality in the ikea place app. J Inf Technol Teach Cases 11(1):8–13. https://doi.org/10.1177/2043886920947110

Passer M (2021) Research methods: concepts and connections second edition, 3rd edn. Worth Publishers, New York

Rothkrantz L (2021) Personalized digital fitness coach. Association for Computing Machinery, New York, pp 123–127

Ruscio J, Roche B (2012) Variance heterogeneity in published psychological research. Methodology 8(1):1–11. https://doi.org/10.1027/1614-2241/a000034

Scargill T, Chen J, Gorlatova M (2021) Here to stay: measuring hologram stability in markerless smartphone augmented reality. arXiv preprint arXiv:2109.14757

Schrepp M (2019) User experience questionnaire handbook, p 15. https://doi.org/10.13140/RG.2.1.2815.0245

Smith J (2015) Qualitative psychology: a practical guide to research methods. SAGE Publications, New York

Soltani P, Morice AH (2020) Augmented reality tools for sports education and training. Comput Educ 155:103923. https://doi.org/10.1016/j.compedu.2020.103923

Swilley E (2016) Moving virtual retail into reality: examining metaverse and augmented reality in the online shopping

experience. In: Campbell C, Ma JJ (eds) Looking forward, looking back: drawing on the past to shape the future of marketing. Springer, Cham, pp 675–677

Taketomi T, Uchiyama H, Ikeda S (2017) Visual slam algorithms: a survey from 2010 to 2016. IPSJ Trans Comput Vis Appl 9(1):1–11

Tavares B, Pires IM, Gonçalo M, Garcia N, Zdravevski E, Lameski P, Trajkovik V, Jevremovic A (2020) Mobile applications for training plan using android devices: a systematic review and a taxonomy proposal. Information. https://doi.org/10.3390/info11070343

Terashima T, Hasegawa O (2017) A visual-slam for first person vision and mobile robots. In: 2017 fifteenth IAPR international conference on machine vision applications (MVA). IEEE, pp 73–76

Turner A (2022) June 2022 mobile user statistics: discover the number of phones in the world & smartphone penetration by country or region. https://www.bankmycell.com/blog/how-many-phones-are-in-the-world

Uchiyama H, Saito H (2007) Ar display of visual aids for supporting pool games by online markerless tracking. In: 17th international conference on artificial reality and telexistence (ICAT 2007), pp 172–179

Vay, Inc. (2022) Vay: The #1 human motion analysis. Accessed 20 Apr 2022

Woźniak PW, Kucharski PP, de Graaf MM, Niess J (2020) Exploring understandable algorithms to suggest fitness tracker goals that foster commitment. Association for Computing Machinery, New York

Zenia, Inc. (2022) Cover art zenia: Interactive fitness app on Google Play. Accessed 20 Apr 2022