FlowAR: How Different Augmented Reality Visualizations of Online Fitness Videos Support Flow for At-Home Yoga Exercises

Hye-Young Jo*
Industrial Design, KAIST
Daejeon, Republic of Korea
hyeyoungjo@kaist.ac.kr

Laurenz Seidel
Hasso-Plattner-Institute,
University of Potsdam
Potsdam, Germany
laurenz.seidel@student.hpi.uni-potsdam.de

Michel Pahud
Microsoft Research
Redmond, WA, USA
mpahud@microsoft.com

Mike Sinclair
Microsoft Research
Redmond, WA, USA
sinclair@microsoft.com

Andrea Bianchi
Industrial Design &
School of Computing, KAIST
Daejeon, Republic of Korea
andrea@kaist.ac.kr

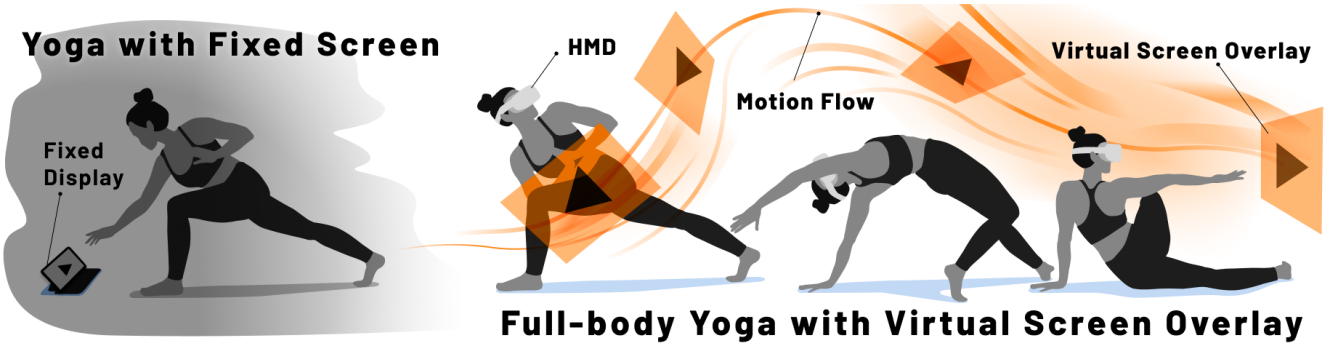

Figure 1: Many home yoga exercise videos require keeping the screen in view, disrupting the ability to perform poses and the overall *motion flow* (left). *FlowAR* enables viewing *full-body and fluid motion* yoga exercises via static or dynamic augmented reality video overlays around the yogi, which are visible using a head-mounted display and are superimposed onto the view of the surrounding space (right).

ABSTRACT

Online fitness video tutorials are an increasingly popular way to stay fit at home without a personal trainer. However, to keep the screen playing the video in view, users typically disrupt their balance and break the motion flow — two main pillars for the correct execution of yoga poses. While past research partially addressed this problem, these approaches supported only a limited view of the instructor and simple movements. To enable the fluid execution of complex full-body yoga exercises, we propose *FlowAR*, an augmented reality system for home workouts that shows training video tutorials as always-present virtual static and dynamic overlays around the user. We tested different overlay layouts in a study with 16 participants, using motion capture equipment for baseline performance. Then, we iterated the prototype and tested it in a furnished lab simulating home settings with 12 users. Our results highlight the advantages of different visualizations and the system's general applicability.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; *Visualization application domains.*

KEYWORDS

augmented reality, fitness video, home workouts, yoga

ACM Reference Format:
Hye-Young Jo, Laurenz Seidel, Michel Pahud, Mike Sinclair, and Andrea Bianchi. 2023. *FlowAR*: How Different Augmented Reality Visualizations of Online Fitness Videos Support Flow for At-Home Yoga Exercises. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3544548.3580897

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580897

1 INTRODUCTION

Fitness videos are a convenient, popular and affordable way to practice physical exercises at home [34, 61]. Examples include videos with fitness exercises, yoga, calisthenics training, and dance tutorials — many of which are offered on online streaming services, such as YouTube [48]. However, video tutorials require users to split their attention between the screen (e.g., a TV, computer, or mobile device), and their body movements. Such distractions interrupt the users' *motion flow* and hinder the learning and the effectiveness of the workout [74].

Motion flow interruptions due to continuous screen monitoring while performing an exercise [16] are particularly significant for **yoga**, one of the most popular training activities (with 300 million practitioners worldwide [70]). Yoga is characterized by poses with large movements that engage the whole body by standing, sitting, and lying supine or prone [69], and emphasizes the fluid transition between poses to build balance and strengthen the muscles [4]. It becomes, therefore, difficult for yoga practitioners (i.e. yogis) to follow the reference instructions on a screen and to perform the poses correctly without disrupting the exercise flow.

To mitigate this problem in yoga and other fitness activities, researchers employed displays that physically *move* along with the user or simulate the instructor's movements. For example, Hoang et al. [24] proposed to instrument the user with a Head-Mounted Display (HMD) to offer a First Person View of a Tai Chi 3D avatar instructor in Virtual Reality. Nakamura et al. [46] used instead a physical projector mounted on a moving robot to provide the affordance of how the dance instructor moves in the video [46]. However, these approaches require recording the instructor's movements via motion capture systems, and therefore they cannot directly leverage the large number of fitness videos already existing online. Furthermore, the limited viewing window offered by a moving target (i.e., a moving projected screen [35] or an avatar in First Person View [78]) limit the applicability of these methods to simple and small motions that are not compatible with the variety of complex full-body poses of typical yoga exercises [67].

In this paper, we propose *FlowAR*, a system that supports yoga workouts with a variety of complex full-body movements (including rotations, flexions, and extensions at various levels [38]) without interrupting the exercise's natural motion flow. Our system leverages the large number of videos already existing online and displays them around the user as an augmented reality overlay rendered via a wearable HMD. We test different placement layouts and question *how different visualization methods of the screen showing the yoga instructor impact the quality of the yoga exercise*. To answer that, we conducted a user study in lab settings with 16 participants using a motion capture system to determine the baseline performance. We performed an in-depth analysis through a heuristic expert evaluation, quantitative motion performance analysis, and qualitative evaluation, which resulted in identifying the optimal visualization method of the instructional video. Based on these results, we modified *FlowAR* for deployment outside of a motion capture studio, using instead a 3D pose estimation algorithm that works with commonly available videos. We then tested the feasibility of the system in a furnished lab simulating realistic home settings via a second user study with 12 participants of various yoga expertise.

In summary, our work provides the following three contributions:

(1) We introduce *FlowAR*, a system that supports at-home training with commonly available yoga videos via a series of virtual screen layouts displayed around the user as an augmented reality overlay.
(2) Using motion capture data obtained in a user study, we validated the feasibility and effectiveness of our system and answered the question of which screen layout visualization is best suited for yoga training.
(3) We integrated into the system a state-of-the-art 3D pose estimation that uses conventional videos supporting testing of *FlowAR* outside of a motion capture studio. We then show the real-world applicability and performance of the system through a user study in a furnished lab, such as in one's home, and with yogis of various levels of experience.

2 RELATED WORK

Our work takes inspiration from previous motion training research in sports and within the HCI community. In this survey of related work, we highlight the importance of motion flow in fitness and yoga and review how it was supported and studied using technology-based motion training tools with various modalities, input user interfaces, and different types of visualizations.

2.1 Importance of uninterrupted flow for physical exercises

Motion training videos, like yoga or aerobic exercises, typically show the instructor's movements which a practitioner simultaneously replicates in their home — a cost-effective and flexible alternative to offline classes [48]. However, unlike personal training with an instructor who can provide immediate feedback or slow the pace when needed, at-home training in front of a screen (e.g., mobile phone, TV, etc...) can cause undesired interruptions of the motion flow during physical activity. For example, a yogi exercising on a new pose sequence might lose sight of the video content, causing a break from immersion [11] and hindering the formation of muscle memory [19].

Sport and HCI literature show that when this state of *flow* [30] is interrupted, both the focus and quality of the exercise deteriorate. In dance, for example, looking at the screen introduces delays and mismatches with the music [46]. In Tai Chi, the limited screen estate makes it difficult to learn large body movements that span over several meters [28]. Furthermore, while the head direction is considered an important standard movement in Tai Chi, the practitioner cannot freely move the head while maintaining a line of sight with the screen [20]. Similarly, it was observed that in golf, the correct posture *collapses* when the player moves the head to peek at the video on the screen [26].

In yoga, maintaining a constant motion flow is essential for blood circulation, injury prevention, and meditation [29, 53]. Yoga practice is inherently fluid as yogis move from one pose to another, and each body motion creates a supporting base for the next pose or movement [62]. However, it is not easy to maintain motion flow when practicing at-home yoga with videos because viewers are typically standing far away from a screen to perform full-body

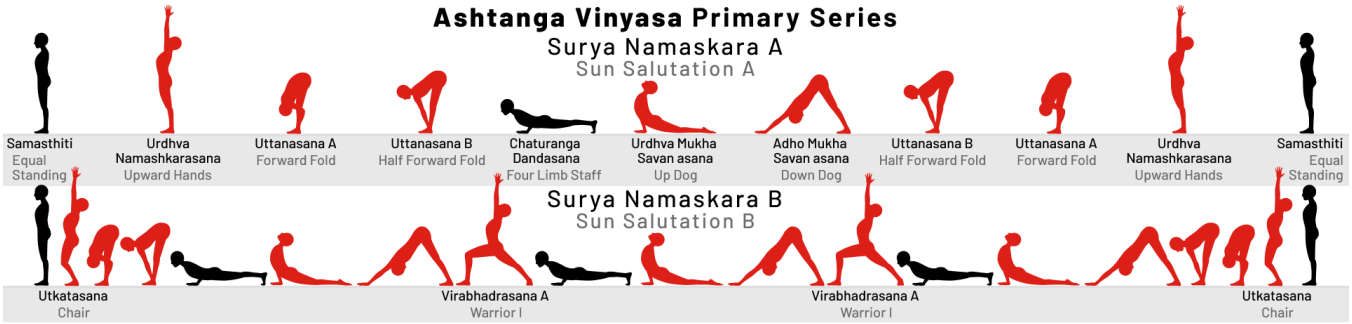

Figure 2: *Ashtanga Vinyasa Yoga* Primary Series's Sun Salutations A and B. Poses where the yogi face away from the front direction are highlighted in red.

workouts and have to interrupt their motions to peek at the video reference on screen [16]. For example, Figure 2 shows the standard primary series of poses for Ashtanga yoga [50] (a classification of classical yoga), which include the Sun salutation A and B sequences [31] — two among the most common exercises in beginner yoga classes. The image shows, highlighted in red, that 22 poses out of 30 require the yogi to face away from the front direction, where the screen is usually placed (a similar argument could be made for any fixed location of the screen). Previous works on yoga training tools exist [8, 40, 65, 66, 75], but rather than focusing on motion training with videos, they focus on providing feedback for limited postures. The rest of this related work surveys systems and techniques that support motion flow in yoga and other sports.

2.2 Visualizing trainers via movable displays or different modalities

A major disruption when using a video tutorial for training is the inherently fixed location of the display with the video. Addressing this problem, researchers proposed interfaces based on movable [46] or projected displays [35, 60]. For example, Nakamura et al. [46] developed a dance training system using a robot-mounted display that moves like the instructor in the video. Kosmalla et al. [35] presented a system that supports indoor climbing using a projector to display the instructor's video directly next to the user. However, movable physical or projected displays require considerable setup and expense yet still allow for a limited range of user motions. To mitigate the inconveniences caused by looking at a physical display fixed in space, researchers in the medical field experimented with always-following displays using Head-Mounted Displays [76]. Research demonstrated that surgeons wearing HMDs were better able to remain attentive and move their arms comfortably [41], resulting in shorter operation times [76]. However, these only tested delicate, unidirectional surgical movements, comparing two visualization techniques. Unlike them, we test large, multidirectional yoga motions using various visualization layouts.

Another problem with video training exercises is the inherent limitations of 2D visualizations, which cannot represent depth [54] and are susceptible to occlusions (e.g., when one part of the body obstructs the view of another part). To address this problem, researchers have attempted to use different visualization techniques and modalities. For example, Xia et al. [72] used audio feedback, generated from the original video, to guide the magnitude and direction of the user's motion. However, this technique is limited to the motion of a single joint at a time, making full-body yoga movements not possible.

2.3 Playback control via user input

Another common disruption of the exercise flow is caused when the user attempts to navigate the video or control the video playback, using click-and-drag or touch commands. To solve this problem, researchers [7, 11, 18, 19] experimented with intangible user interfaces that do not require the users to manipulate controllers to pause, play or navigate a video. Clarke et al. [11] developed a speed-adaptive system for Tai Chi that compares the instructor's poses, extracted from a video using computer vision techniques, and the real-time practitioner's pose, estimated via a Microsoft Kinect. When the poses do not match, the system adjusts the video's playback speed to synchronize with the user's movements. Chang et al. [7] developed a content-based voice navigation system that automatically extracts keywords from the video narration and allows users to play, skip or rewind specific parts of the video via voice commands.

These user interfaces reduce the disruption of the motion flow when the user is attempting to directly control the video playback, but these works suffer the same limitations as traditional at-home exercises. In fact, for these works, the video is still displayed on a static screen, and the ability to control the video and playback speed does not necessarily translate into better motion flow.

2.4 3D and immersive visualization training environments

Several researchers [1, 6, 26, 42] proposed to represent the trainer's motion via animated 3D characters instead of using simple videos and then project them on multiple screens or 3D immersive caves [23, 25, 37]. To achieve high fidelity of the motion, these animations are typically generated using the 3D motion data that can be acquired using a motion capture system (e.g., OptiTrack, Vicon) with post-processing (e.g, data cleaning, rigging, rendering). The advantage of these systems is that they can show the trainer from any desired or custom viewing angle.

To fully exploit the 3D content, past research [9, 17, 74] also adopted HMDs to visualize the virtual trainer in stereoscopic view.

These works explored the effect of using different viewing perspectives (first-person view, superimposition, third-person view, group training) of a 3D virtual trainer [71, 78] in various domains, such as Tai Chi [9, 24], ski [71], and simple guidance of mid-air movements [12]. Also in the domain of yoga, several immersive commercial applications use HMDs combined with 3D virtual trainers in front and surrounding layouts [32, 63, 68].

Similar to the above works, the system presented in this paper is also based on HMD visualizations. However, differently from any prior work, our system uses plain and commonly available fitness videos rather than 3D content that has to be generated ad-hoc (e.g., via motion tracking or animation tools). By testing visualizations of the screen with different layouts, we also contribute to the scientific literature on motion-guided training.

3 THE *FLOWAR* SYSTEM

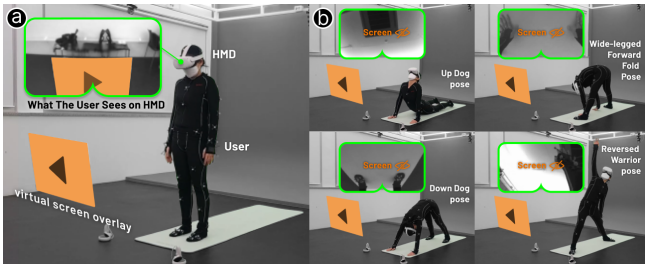

Figure 3: ⓐ A yogi seeing a virtual screen overlay through the HMD and ⓑ examples of yoga poses in which the yogi would not be able to see the screen locked in the front location. The green highlighted parts show what the user would see on the HMD. The real world is in greyscale (the device's default setting), and the video overlay is displayed in colors.

We present *FlowAR*, an augmented reality training system that uses commonly available videos to support the practice of multi-directional full-body yoga poses and their transitions in a seamless flow. Users wear the Oculus Quest HMD, generally used for virtual reality experiences but also has augmented reality capabilities, and see the surrounding environment via a real-time video stream captured by the HMD's built-in cameras and an augmented Computer-Generated (CG) video screen overlay (Figure 3.ⓐ). This technique involving a see-through video with a CG overlay is described as one of the types of augmented reality in the *Reality-Virtuality Continuum* [44]. We chose this technique to allow users to see their limbs and possible obstacles (e.g., people or furniture) and prevent collisions, considering that users might exercise in their living rooms or bedrooms. We chose the Oculus Quest as HMD, despite this device being more popular for Virtual Reality rather than Augmented Reality applications, because of its large field of view [51], its light weight (503 grams, which is 63 grams lighter than the popular Hololens 2 [1]), and its low cost (under 400 USD, as of September 2022).

As shown in Figure 3.ⓑ, the yogi cannot see the screen in a fixed location when performing various dynamic yoga poses with different body directions. To solve this problem, *FlowAR* renders a virtual video screen overlay around the yogi, using various layout configurations. These are designed to remain in sight despite the users' head movements. By combining a virtual overlay of the video with a live-feed camera stream from the HMD, the system allows users not only to see the instructor but also to track their limbs and movements, as well as the surrounding environment.

Below here, we explain the proposed screen layouts and their working principles, followed by a description of the prototype implementation.

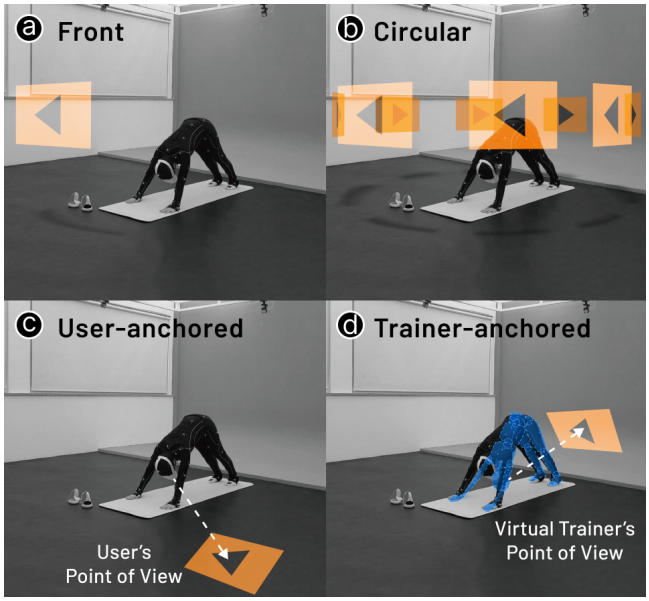

Figure 4: A yogi practicing the *Down Dog* pose using four augmented screen overlays: ⓐ Front (baseline) ⓑ Circular ⓒ User-anchored ⓓ Trainer-anchored. The virtual trainer (blue) is invisible to the user and shown here only for clarity.

3.1 Augmented Screens Layouts

Figure 4 shows a user practicing the *Down Dog* pose. As visible in ⓐ, a single screen fixed in space (physical or displayed as an augmented overlay) does not provide enough coverage and ends up outside of the user's field of view. We propose three alternative layout: **Circular**, **User-anchored**, and **Trainer-anchored**. The first is a *static* layout (i.e., the screens do not move in respect of the user), while the latter two are *dynamic* layouts (i.e., the screens move with or in respect to the user).

Circular: the Circular layout (ⓑ) consists of eight evenly spread screens surrounding the user in a circle, each screen displaying the same video. This panoramic view of the instructor is inspired by previous works [9, 20, 32], which used copies of a 3D virtual coach surrounding a user immersed in Virtual Reality, to support immediate access to the reference movements. In the above example of the *Down Dog* pose, despite the head not facing forward, the yogi is still able to see the trainer on the rear screen.

User-anchored: Differently from the *Circular* static layout, which places the virtual screens in the world independently of

[1] https://www.microsoft.com/en-us/hololens/hardware

the user's position or movement, the *User-anchored* layout shows a screen that follows the user's head (i.e., a local coordinate system). In practice, the screen is placed at a constant distance and orientation from the user's head, resembling an always-visible Head-Up Display(HUD). The motivation for this choice is that the HUD overlay is known to support dual tasks such as monitoring visual information while engaging in physical activities that require situational awareness, such as reading while driving [22], or watching educational videos while walking [49]. HUD was also employed for yoga training with 3D characters in Virtual Reality following the user's viewpoint [68]. We borrow this technique to show an AR screen overlay with the trainer's video, thus ensuring that the user can always see the reference motion in any posture. In the *Down Dog* pose example, the virtual screen remains in front of the yogi, appearing on the floor (ⓒ).

Trainer-anchored: like the *User-anchored* layout, also the *Trainer-anchored* layout is dynamic, but in this case, the screen moves following the trainer's viewpoint (the anchor point is the trainer's head). In other words, the screen moves following the correct reference motion, hinting to the user the distance from the screen and the direction they should face to correctly perform the exercise (ⓓ). The concept of gently *nudging* users to alter their posture is known in ergonomics literature [57], but it is exploited for the first time here in the domain of fitness videos. In the *Down Dog* pose example, the yogi can maintain a good view of the screen as long as she/he moves following the trainer.

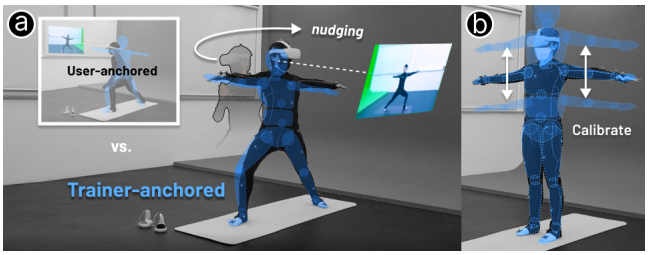

Figure 5: Implementation of *Trainer-anchored* layout: ⓐ the *Trainer-anchored* layout is nudging a user to alter posture while the *User-anchored* layout is following the user. ⓑ A virtual trainer is calibrated to the user's height.

From the technical point of view, the *Trainer-anchored* layout is more challenging than the others, because it requires the coordinates of the 3D motion performed by the trainer to compute the screen position/orientation. This 3D motion data can be obtained in various ways, such as 3D motion capture systems [6, 9] and 3D pose estimation [11, 42]. We will show the feasibility of both methods in the studies presented in this paper. Furthermore, it is not sufficient to directly map the trainer's 3D motion coordinates to the screen, as the differences in height between the trainer and user could result in incorrect postures (e.g., if the trainer is taller than the user, looking forward for the trainer might translate to looking upward for the user). Therefore a user-dependent calibration and coordinates re-targeting process are necessary (see *Implementation* section for details).

3.2 Implementation

FlowAR relies on the augmented reality capabilities of the Oculus Quest 2 HMD [2] to render the screen's layout overlay. Using its 4 built-in front cameras, the Oculus Quest 2 HMD provides a live see-through camera stream (passthrough API [3]) that shows the surrounding environment in greyscale. This is augmented by a graphical overlay with the screen layout displaying the prerecorded fitness videos. We used C# and Unity 3D to compute the screen placement in the space, and the rendering was achieved at 70 fps. The code and video were preloaded on the HMD allowing its untethered usage.

Following the ergonomic recommendation in [57], using Unity, we set the initial distance of all the screens from the user to 1.3 m and rotated them 6° backward. To select the screen size used in all our layouts, we informally tested various sizes in the range of 11-inch (e.g., table PC) to 80-inch displays, finally choosing the 32-inch screen size because it provides enough real estate to cover a wide range of body-movements without completely obstructing the view of the surrounding environment.

For the implementation of the *Trainer-anchored* layout, we first collected 3D motion data either directly from an expert trainer via motion capture (*Study 1*) or using a regular video fed in a 3D pose estimation algorithm (*Study 2*). For 3D motion capture data, we used the *Motive Body* software by OptiTrack to extract the joints' position and orientation from the point cloud and manually cleaned the data (e.g., labeling missing, wrong, or swapped markers). This data was exported to Unity and mapped to the 18 joints of an invisible 3D skeleton representing the virtual trainer: head, neck, spine (3 joints), pelvis, shoulders (2), elbows (2), wrists (2), hips (2), knees (2), and ankles (2).

Because the final computed height placement of the screens in the *Trainer-anchored* layout depends on the trainer's height and relative joints length (e.g., legs vs. torso), the virtual joints of the invisible 3D character representing the trainer had to be translated and scaled to match those of the actual user. This re-targeting process requires a one-time calibration, in which users' body parts and height were manually measured and input into the system (the process could be automated as in [56], but it is beyond the goal of this paper). The result is that the reference movement of the trainer matches the body dimensions of the user. Finally, the virtual screen placement and orientation are computed to perpendicularly face the head of the virtual trainer (1.3 m away, 6° tilted backward). The small motion jitter of the screen is attenuated using a rolling average filter of 0.5 seconds applied to translation and orientation.

4 STUDY 1: BASELINE PERFORMANCE

We conducted a user study to gauge how a fitness video, visualized using different augmented screen layouts, affects the user's correct execution of a yoga exercise (measured as posture accuracy and motion flow). We therefore tested the three proposed screens layout —**Circular**, **User-anchored**, **Trainer-anchored** — vs. a baseline condition (**Front**), which consists of a single static screen placed in front of the user, resembling a TV monitor. For comparison purposes, the *Front* layout was also presented as an augmented

[2]https://store.facebook.com/kr/quest/products/quest-2
[3]https://developer.oculus.com/blog/mixed-reality-with-passthrough

overlay visible with the HMD and with the same parameters of the other conditions (32-inch, 1.3 m away, 6° tilted).

We recruited a professional yoga instructor to help us design a balanced yoga sequence of various movements and poses (Figure 6.ⓑ), as well as to evaluate the quality of the user's exercises via a heuristic evaluation. The instructor has more than ten years of yoga experience and was certified for KPJAYI (K. Pattabhi Jois Ashtanga Yoga Institute), Sharath Yoga Center, in 2010. We also employed a full-body motion tracking system to record both the trainer's and all users' motion data. The trainer's data served both as baseline performance and to compute the screen position and orientation for the *Trainer-anchored* layout.

4.1 Participants

We recruited sixteen participants (8 females, 8 males) aged 20-31 (M: 26.5, SD: 3.1) via a posting on our institution's announcement web portal. Nine participants reported to workout from home using YouTube videos at least once a week (M: 1.6, SD: 1.1). Five out of sixteen participants had less than three months of yoga experience, four through offline classes in yoga studios and one through YouTube videos. Eleven participants reported that they had previously experienced virtual reality, and five of them had also experienced augmented reality.

4.2 Motion tracking and yoga sequence

After wearing a 3D motion tracking suit with 50 passive retro-reflective markers, the instructor performed the reference yoga exercise. We simultaneously recorded the 3D motion data using 8 OptiTrack infrared *Prime* cameras at 120 fps, and a 1080p (30 fps) audio-video footage showing a perspective view of the instructor, using a mobile phone (Samsung Galaxy Note 8). The exercise sequence was determined to target beginner yogis, to fit a 4.7 m x 3.5 m x 2.8 m height capture region, and to account for wearing an HMD, like in Figure 3.ⓐ. The yoga instructor was free to choose the poses of the sequence, but we asked to avoid poses that would require the yogi to place the head on the ground, so to prevent HMD from bumping into ground. The resulting recorded sequence is an 8-minute long exercise, composed of a sitting warm-up session (4 minutes) and a standing main workout session (4 minutes).

The warm-up session serves as a distraction task to reduce learning effects carried over across sessions. Each workout session consists of several poses, made of motion transitions among different body alignments (postures). In other words, a pose is *not* a *static* posture, but rather a *fluid and slow transition* among postures, as explained in [59]. For example, Pose A (Extended Side Angle Pose) starts from a lunging posture and, through a fluid motion transition, terminates with an extended arm. All the poses of our session are visible in Figure 6. Some poses (A, B, C) were repeated for both the left and right side of the body to complement the exercise (e.g., A pose is complemented by A'). Finally, the video is accompanied by the trainer's voice narration (recorded during the trainer's reference exercise) which describes the poses performed.

4.3 Experiment design

The experiment followed a within-subjects design with four conditions, clustered in two order groups: **static** layouts (*Circular* and *Front*) and **dynamic** layouts (*User-anchored* and *Trainer-anchored*). During the study, each participant tested all four screen layout conditions following a Latin-square design balanced for the presentation group (static vs. dynamic), resulting in eight possible sequence orders. We chose to cluster the four conditions in two groups (static vs. dynamic layouts) presented in balanced alternating order to minimize the learning effects and possible confusion carried over apparently similar layout conditions (e.g., from *User-anchored* to *Trainer-anchored*). Overall, each participant experienced a preparation phase, four exercise sessions, and a final interview (Figure 6.ⓐ).

During the preparation phase, the participants completed a demographic survey and watched twice the instructor's video sequence on a tablet PC to familiarize themselves with the pose sequence. Afterward, the participants wore the motion capture suit and the HMD and completed a calibration session. The participants then performed the four workout sessions (warm-up and main) with a questionnaire at the end, followed by a rest between sessions. After the four workout sessions, we conducted an interview to gather the participants' feedback. The experiment took about 2 hours to complete, and participants were compensated with 30 USD in local currency.

4.4 Data collection and analysis

Besides demographics and interviews, we collected 3D motion data for each participant. We performed both qualitative and quantitative analyses of this data, such as a holistic evaluation of each key pose by the expert instructor, and a quantitative performance analysis of timing and posture errors (similar to [24]). Following previous work [10], we also collected participants' feedback on the level of perceived competence and value/usefulness of the different screen layouts. All interviews were audio-recorded, transcribed, translated, and analyzed using open and axial coding methods [14]. Here are more details about each of the evaluation methods we used:

4.4.1 Heuristic expert evaluation. We performed a heuristic evaluation with the expert yoga instructor who composed the sequence using a questionnaire followed by a semi-structured interview. We modified the questionnaire together with the instructor by selecting a subset of the existing metrics presented in prior work [47] according to their relevance to our exercise. With this questionnaire, the expert evaluated participants' yoga flow competence for each of the 7 key poses on a 5-point Likert scale, considering the following four factors: *form*, *ease*, *ability to follow instructions*, and *holding duration*. These metrics were finally averaged in a single score.

For the evaluation, the instructor received a video for each participants with a 3D character animation of the motion performed by each participant for all layout configurations (e.g., see Appendix Figure 14.ⓐ). The 3D character mapped to the participant's movements (shown in yellow) were then overlaid with a semi-transparent character representing the trainer's motion reference (in blue). We provided the video both in its entirety and with the 10-20 seconds windows around the target poses. We used a 3D character animation (with the captured motion data) instead of a plain video of the participants to limit the scope of the evaluation only to the motion flow and the pose correctness, reducing visual bias due to other

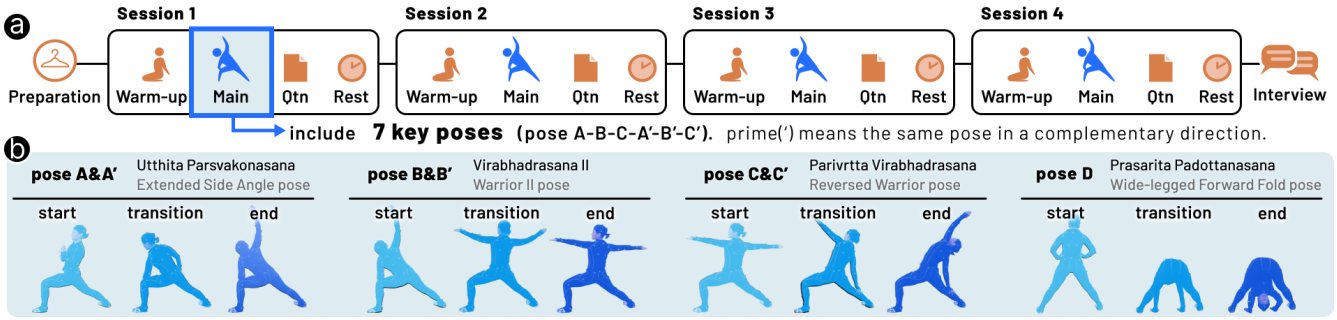

Figure 6: Study 1 procedure: ⓐ overall procedure with four sessions, and the ⓑ seven key motions(A-B-C-A'-B'-C'-D) for the main workout sequence.

factors (e.g., physical look, breathing rate, facial expressions, providing anonymity, etc...). Finally, we also interviewed the instructor to gather more detailed feedback about the given ratings.

4.4.2 Motion performance analysis. To quantify the participants' performance, we measured their temporal and spatial deviation from the trainer's reference movements using the motion data obtained from the optical tracking system. The Python scrips to compute the temporal and spatial deviations are open source and available from this repository [4].

Temporal deviations occur when the participant performs a movement slower or faster than the reference. Like previous work in the domain of movement processing [64], we performed a Dynamic Time Warping (DTW) [3] to map the participant's motion onto the trainer's reference motion. Specifically, because we are dealing with compound movements, we performed a multi-dimensional DTW-D [58] on the z-normalized three-dimensional position of tracked joints. The result from the DTW-D is a frame-by-frame mapping between the trainer's and participants' motions, which we then used to calculate the deviation of the participant's motion from the reference. To compute the cumulative timing error, we finally converted the absolute frame difference into seconds and returned its mean value.

Spatial deviations occur when the participant's movement does not match the expert's movement (i.e., wrong posture or body alignment). Similar to OneBody [24], we report the Mean Per Joint Angle Error [27] which is the aggregated angular difference of selected joints when comparing the user data with the reference. This metric is advantageous for our study setup because it is invariant with the user's absolute position in space. In our analysis, we report the mean angular error of shoulders, elbows, hips, and knee joints. To eliminate errors due to temporal deviations, we again use the frame matching previously computed using the DTW-D.

4.4.3 User feedback from questionnaires and interview. To collect the participants' confidence with the performed exercise, like in [36], we used the Intrinsic Motivation Inventory questionnaire (IMI [52]) with *Perceived Competence* as subscale. We also added the *Value/Usefulness* subscale to record their impression on the utility of each visualization layout used for yoga. Participants answered each question in the IMI subscales using a 7-Likert scale (see Appendix.Table 2). We also collected the participant's perceived motion sickness on a 4-point Likert scale using the Simulator Sickness Questionnaire (SSQ [33], see Appendix.Table 1) to verify if any of the screen layout caused disorientation during the exercise.

In the final semi-structured interview, we asked participants to elaborate on their ratings, and further explain their impressions or preferences toward each overlay visualization.

5 RESULTS OF STUDY 1

The study results are organized into three parts: (1) expert evaluation, (2) motion performance analysis, and (3) user feedback.

The effect of different overlay conditions on the expert ratings for each pose, performance inaccuracy (temporal/spatial deviations), and questionnaire results were all tested using a Friedman test ($\alpha = 0.05$), and pairwise post-hoc analysis was performed using Wilcoxon signed-rank test with Bonferroni correction ($p = 0.0083$).

5.1 Expert evaluation

Expert scores for individual key poses and their averages are shown in Figure 7 with the corresponding Friedman test results at the bottom. Overall, the mean scores show a statistical difference for screen layout condition ($\chi^2(3) = 9.162, p = 0.027$), but pairwise comparisons analysis with Wilcoxon signed-rank test was not significant. Nonetheless, static layouts received, on average lower ratings ($M : 2.760, SD : 0.028$) than dynamic layouts ($M : 2.915, SD : 0.007$).

We also performed an analysis for individual poses (Figure 7). Our statistical analysis with Friedman tests did not reveal differences amongst screen conditions for specific poses. Nonetheless, Figure 7 shows that, in all poses, either of the dynamic layouts was always rated the highest (pose C' was a tie). Specifically, the scores for the *User-anchored* layout were the highest during the first half of the exercise (poses A-C — $M : 2.880, SD : 0.250$), while the *Trainer-anchored* layout led to the highest rankings for the second half of the exercise (poses A'-C', D — $M : 3.110, SD : 0.208$), suggesting some kind of learning effect.

Finally, upon suggestion from the expert, through video analysis, we counted the number of times the head direction abruptly changed from the gazing point narrated in the video (e.g., caused by distractions or screen monitoring). Raw results for these look-away

[4] https://github.com/hyeyoungjo/dtw-for-two-movements

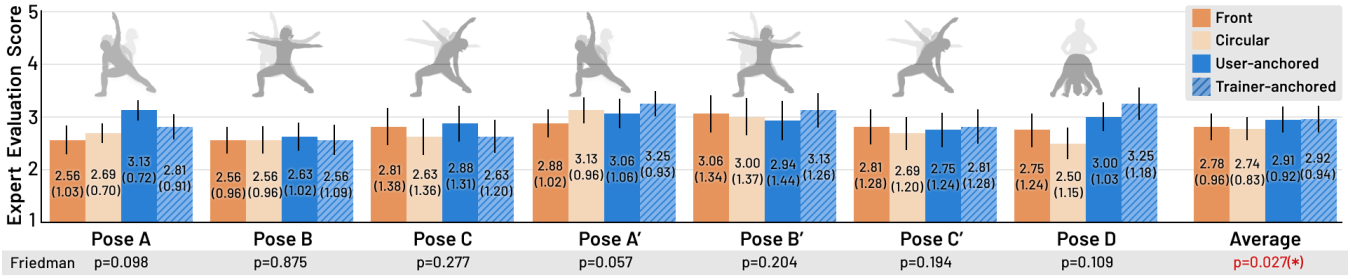

Figure 7: Expert evaluation results show the expert ratings per pose and averaged across participants. Mean scores (and standard deviation) are displayed on the bars, while Friedman test results are at the bottom.

instances are shown in Figure 8. A Friedman test reveals differences across screen layouts ($\chi^2(3) = 43.268, p<0.001$), and pairwise comparisons revealed that, unsurprisingly, both *User-anchored* and *Trainer-anchored* had fewer screen look-aways than any of the static conditions.

We finally collected several observations from the expert during the interview. The main important findings that emerged are the followings:

Screen look-aways damage the posture and break the motion flow: the expert instructor highly emphasized the importance of maintaining the correct alignment of the head and of the body, stressing that *"yoga is a training that develops the ability to recognize one's body in space"* and that *gaze* direction leads to a natural motion flow. In the words of the trainer:

> *"The gaze is tightly related with the flow of poses and breath [...] If the head direction goes against the motion flow, it's not the right posture even if all the rest of the body follows the instructions."*

According to the instructor, jerky head movements such as looking at a screen to follow the instruction caused harmful effects such as *"stiff necks"*, *"shrunk shoulders"*, *"lack of focus"*, and *"direction confusion"*. These observations were more common for the static layouts.

> *"It seemed that they [the users of static layouts] could not concentrate. Their heads were constantly moving back and forth [...] confusing the left foot with the right [...] Some of them even did the left side of the exercise twice [poses A-C instead of A'-C']."* (The trainer)

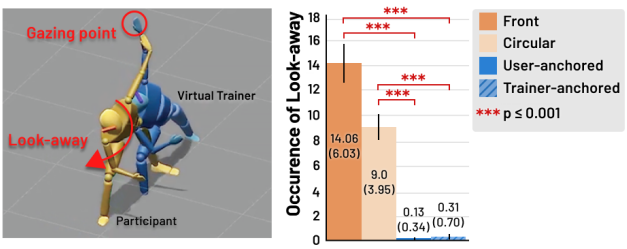

Figure 8: Look-aways. An example of look-away (left) and a graph with the occurrence mean count of look-aways (right).

Dynamic layouts support different types of users: Although mean scores were similar for the *User-anchored* ($M: 2.910, SD: 0.920$) and the *Trainer-anchored* layout ($M: 2.920, SD: 0.940$), the expert noted that the participants' motion flow afforded by these layouts was different. Specifically, the expert qualified the motion flow during the *User-anchored* condition as *"good"* and *"natural-looking"* across all participants, but for the *Trainer-anchored* condition the motions were described as ranging between *"great"* and *"unstable"*, depending on the participant. For example, despite the overall high scores of poses A'($M: 3.250, SD: 0.931$) and C'($M: 2.813, SD: 1.276$), some participants performed poorly, showing unbalanced postures when stretching backward due to lack of core muscles. In other words, while the *User-anchored* layout supported a stable motion flow across participants, performance using the *Trainer-anchored* layout seemed to bond with the participant's pre-existing physical capabilities

> *"Their [Trainer-anchored layout users] posture was unstable and unnatural. It looked like they could not put any strength to tighten their abs. So, the overall posture looked not in a good balance."* (The trainer)

5.2 Motion Performance Analysis

Figure 9 shows the mean timing and spatial errors (angle deviations from reference) for all participants in individual poses and on average. A Friedman test was conducted for both types of errors and test scores are indicated at the bottom of Figure 9.

Looking at the cumulative mean timing errors, we report statistically significant differences for screen layouts ($\chi^2(3) = 8.025, p = 0.045$) but no pairwise statistical differences. Overall, dynamic layouts led to smaller mean timing errors ($M: 1.095, SD: 0.078$) than the static layouts ($M: 1.420, SD: 0.085$), with timing errors being the smallest in the *User-anchored* condition ($M: 1.042, SD: 0.438$), closely followed by the *Trainer-anchored* ($M: 1.148, SD: 0.634$) condition — though no statistical differences were found. As for the correctness of postures, no statistical difference was found among the mean joint angle errors ($\chi^2(3) = 4.350, p = 0.226$) but only for individual poses: B ($\chi^2(3) = 7.950, p = 0.047$), A' ($\chi^2(3) = 12.675, p = 0.005$), C' ($\chi^2(3) = 9.300, p = 0.026$), and D ($\chi^2(3) = 8.175, p = 0.043$). Also, joint angle errors found in the dynamic layouts were lower than static ones ($M: 16.750, SD: 0.212$ vs. $M: 18.185, SD: 0.686$).

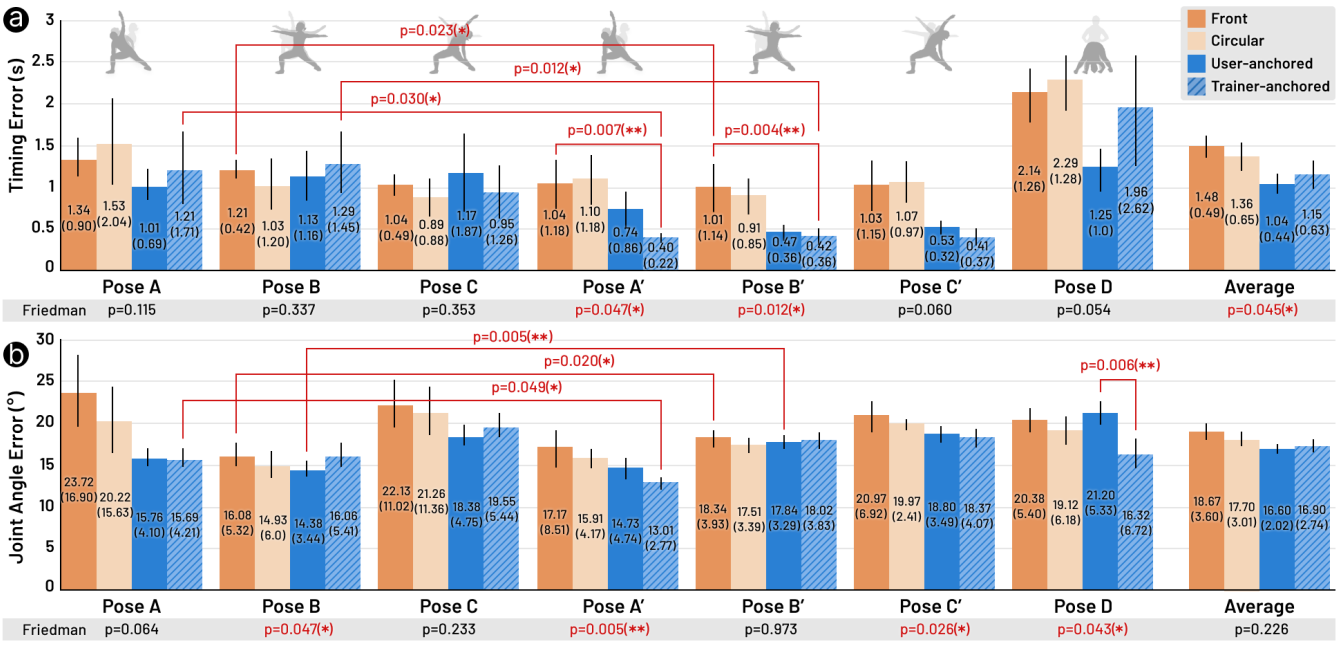

Figure 9: Performance analysis result shows the cumulative ⓐ temporal inaccuracies (timing errors) and ⓑ spatial inaccuracies (joint angle error) for each pose and average. Mean scores (and standard deviation) are displayed on the bars, and Friedman test results are at the bottom.

5.3 User Feedback

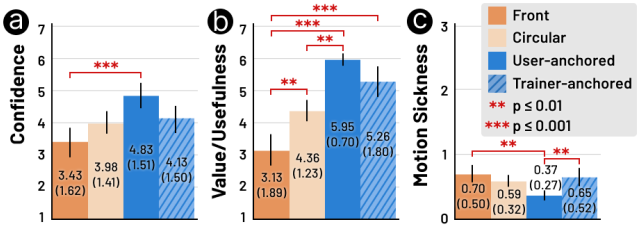

Figure 10: Questionnaire results about the level of ⓐ confidence, ⓑ value/usefulness, and ⓒ motion sickness. Bars represent mean scores for all users.

Figure 10 shows the summary of the participants' responses to the questionnaires. Users gave significantly different scores on perceived competence (i.e., confidence) ($\chi^2(3) = 12.898, p = 0.005$), value/usefulness ($\chi^2(3) = 29.013$, p < 0.001), and motion sickness level ($\chi^2(3) = 11.776, p = 0.008$) for different conditions. Overall, participants reported to feel more confidence and perceived more value for the dynamic layouts, rather than for the static ones (confidence level: $M : 4.480, SD : 0.495$ vs $M : 3.705, SD : 0.389$; perceived value: $M : 5.605, SD : 0.488$ vs. $M : 3.745, SD : 0.870$). As to motion sickness, all conditions scored slightly less than 1, meaning that the participants felt able to work out in all conditions without experiencing any severe motion sickness. It is worth noting that the *User-anchored* layout caused significantly less motion sickness than the baseline *Front* condition ($Z = -2.878, p = 0.004$).

The interviews with participants corroborate many of the results presented above. All sixteen preferred dynamic layouts over static layouts — 10 of which selected the *User-anchored* layout as their favorite (*"[The User-anchored layout] was the most comfortable because it was always in front of me wherever I looked [...] it gave me more energy to spare. So, I was able better observe the details of the movements."* — P11). The *Trainer-anchored* layout received mixed reviews. Three participants (P6, P8, P11) strongly disliked it because they felt it was too challenging (*"I am not flexible as the instructor, so I could not see the screen in some poses, even if I tried hard ."* — P6). Other participants instead appreciated that they could better notice their own mistakes with this layout. As a result, 13 of 16 participants mentioned this self-correcting mechanism positively in the interviews: *"I felt like I was learning properly. When I had to look up to see the screen, I realized the teacher was straightening the back. So, I was able to correct my posture."* — P4).

More generally and regardless of the preference for screen layouts, the participants agreed that the usage of HMDs rather than fixed displays helped them stay focused on the exercise, supporting their flow, corroborating prior work about the usage of HMDs [76]. For example, P5 commented that: *"My concentration level was much higher with the HMD. I felt like I was alone even though I could see the experimenter in the background. It was fun because it just felt different."* On the other hand, the participants also voiced discomfort when using HMDs for certain poses where the head was upside down or very close to other body parts (e.g., *Down Dog* and *Wide-legged Forward Fold*). For example, P3 commented that the *"HMD felt heavy when I lowered my head. It does not hold in place [because of the strap], so the video becomes opaque. It is a similar discomfort*

of when I work out wearing glasses." We indeed observed that when in these poses, participants often reached their HMDs with their hands to readjust their position.

5.4 Summary of results

The results from the expert evaluation, motion performance analysis, and user feedback combined lead to three main findings:

(1) In general and across all poses, dynamic layouts (*User-* and *Trainer-anchored*) achieved better scores from the expert and led to fewer timing and posture errors.
(2) All participants as well as the expert preferred the dynamic layouts over the static ones. Specifically, most participants generally preferred the *User-anchored* layout because it felt easier. However, the *Trainer-anchored* layout displayed some potential of nudging users with more physical training toward better exercise practice (e.g., closer to the instructor's reference posture).
(3) Head movements (like screen monitoring) were the main factor for the lower performance and ratings of the static layouts. Dynamic layouts are less affected by sudden head movements.

6 *FLOWAR* WITH POSE ESTIMATION

The paper's original motivation was to support at-home fitness exercises using commonly available online tutorial videos (i.e., from YouTube). However, up to this point, we relied on a motion capture system to generate the instructor's 3D motion data — data that we used to both enable the *Trainer-anchored* screen layout and to perform the user-trainer comparative performance analysis of *Study 1*. In this section, we describe how we modified *FlowAR* to enable the acquisition of 3D motion data directly from the video tutorial using 3D pose estimation.

Google's *BlazePose* [5] is a state-of-the-art computer-vision algorithm capable of extracting 2D/3D features (33 landmarks such as joints and facial features) from a 2D video, and it is typically used for fitness or sports applications. We selected BlazePose for its easy accessibility (open-source), speed, and proven accuracy in yoga poses. Prior work demonstrated a PCK@0.2 (Percent of Correct Points with 20% tolerance) of 84.5 for a 2D yoga dataset [2], and an MPJPE (Mean Per Joint Positional Error) of 121 mm of for a 3D yoga dataset [15] — numbers that show that, for the domain of yoga fitness, BlazePose outperforms other state-of-the-art pose estimation algorithms [2, 15]. Although pose estimation can be performed in real-time, we pre-computed it offline to achieve the highest accuracy. We used a PC equipped with an AMD Ryzen 9 5900X CPU at 3.70 GHz, 64GB of RAM, and an NVIDIA GeForce RTX 3080 (10 GB) graphic card. The average time of data generation was measured as 0.1 seconds per frame. It took, for example, 30 minutes to extract the 2D/3D motion data from a 10-minute long video (30 fps).

We used a Python script [6] to precompute and export the landmarks from a fitness video using the bespoke 3D and 2D pose estimations. Then, we input the extracted video instructor's 2D and 3D landmark coordinates with the user's height information into the Unity engine, where an automated script reconstructed the instructor's character in 3D space and calibrated its dimensions with the provided data. Specifically, we used the 3D coordinates to calculate the joints' locations and their relative movements, while we used the 2D coordinates to determine the initial vertical position and estimate the horizontal translations of the whole target skeleton in the world coordinates. For calibration, we also adjusted the virtual instructor's joint length accounting for the user's measurement. To place a video screen in front of the virtual trainer's head in the *Trainer-anchored* layout, we used the vector between the nose and the rest of the facial feature points to infer the correct screen orientation and position. Finally, as before, we applied a rolling average filter of 0.5 seconds for a smooth transition.

7 STUDY 2: APPLICABILITY IN HOME-LIKE SETTINGS

The first user study aimed to assess the overall feasibility of *FlowAR* and to determine which screen visualizations performed best. We concluded that participants preferred the dynamic screen layouts and that these led to better performance. However, because the first study was conducted in an empty, spacious lab using motion capture data to generate the *Trainer-anchored* layout, it did not offer insights into the system's real-world applicability at home, where there are possible distractions caused by the surrounding environment and no motion capture system to support the *Trainer-anchored* layout. This second study further explores the differences between the two dynamic screen layouts (*User-* and *Trainer- anchored*) but in a furnished and smaller lab simulating the home environment. In this study, tracking was achieved using 3D pose estimation in place of a motion capture system which demonstrates the feasibility of using these visualizations on commodity hardware at home. We also considered the participant's experience with yoga to test whether it affects layout preference and to obtain rich qualitative feedback.

7.1 Experiment room and material

In a real-world use case, users would be able to train from their homes using online fitness videos augmented by the *FlowAR* overlays, and without the need for pre-captured or edited motion data. Thus, we designed a user study that uses a YouTube video[7] as input, from which the motion data was extracted via pose estimation. The video was chosen to showcase various narrated yoga poses that do not require placing the forehead on the ground and with the instructor always well visible on screen. Furthermore, we were interested in providing a surrounding environment that resembles a real home, and seeing whether the background and objects around the user interfered with the overlays. Thus we prepared a 4.0 m x 3.7 m x 2.5 m space with various furniture and a camera facing the user. We chose a lab-study setup with a room-like environment rather than the users' homes to *partially control* the environment across participants and thus be able to pin changes of performance to the *FlowAR* system rather than the uncontrolled surrounding environment. As shown in Figure 11.ⓐ, participants wore their workout clothes instead of motion capture suits and practiced yoga on a matte surrounded by household appliances and furniture.

[5] https://google.github.io/mediapipe/solutions/pose
[6] https://github.com/hyeyoungjo/video-pose-estimation
[7] https://youtu.be/Wkmarh2Ps_o

7.2 Experiment design

This study follows a within-subject design with two layout conditions (*User-anchored* vs. *Trainer-anchored*), presented in a fully balanced order. We recruited 12 participants (6 females, 6 males) aged 22-31 (M:25.4, SD:3.0), 6 of which had no yoga experience. For participants with yoga experience, that varied between 2 to 24 months of offline attendance at a yoga studio. Like before, we collected motion data for analysis, but this time using video recordings of participants and subsequently extracted their motion data with the BlazePose algorithm. As before, we collected user preferences via questionnaires and interviews, but also task workload ratings using a NASA-TLX questionnaire [21].

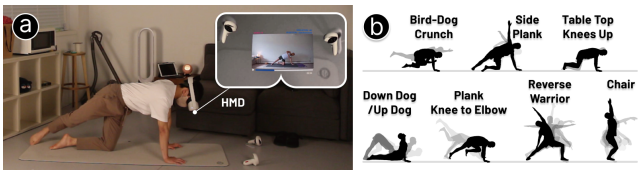

Figure 11: FlowAR study 2: ⓐ setup and ⓑ seven key motions from YouTube video.

The study procedure closely resembled that of our previous study. As before, after an initial preparation phase (i.e., demographics, watching the training video, wearing HMD, calibration), the participants completed two 10-minute long workout sessions (one per condition) with rest in between. To prevent the HMD from sliding in bending poses, we substitute the Oculus Quest 2's default elastic strap with the rigid one (Elite Strap[8]). Also, We carefully ensured that the order of presentations was balanced for the *yoga expertise* variable. For the workout, we clipped the YouTube video to 10-minutes. In the video, there is no warm-up session, and the exercise consists of 7 key-poses (see Figure 11. ⓑ for details). After the workout, we conducted a 30 minutes post-hoc interview on the general impression of two screen layouts and the usability of *FlowAR* in home settings. The experiment lasted 1.5 hour, and participants were compensated with 15 USD in local currency.

7.3 Results

7.3.1 Performance analysis. The performance analysis with motion data followed the same procedure as that described in the first study. A Wilcoxon signed-rank test ($\alpha = 0.05$) was used to compare variables between conditions (*User-* and *Trainer-anchored* layouts). Figure 12 shows an overview of all mean timing and spatial errors per pose, split between the participants with or without prior yoga experience (*Experienced* vs. *Inexperienced*). The line graphs above the bar charts additionally show a timeline of the participants' motion deviation from the reference.

A close analysis of errors shows that, in both conditions and for all poses, angle error on average amounted 10° and that timing errors never exceed 2 seconds. Regardless of yoga experience, we report no statistical difference for spatial errors (Inexperienced: $Z = -0.314, p = 0.753$, Experienced: $Z = -0.736, p = 0.462$) or time errors (Inexperienced: $Z = -1.572, p = 0.116$, Experienced: $Z = -1.051, p = 0.293$). Interestingly, mean timing errors with the *User-anchored* layout are, for inexperienced participants, slightly higher than those in the *Trainer-anchored* layout ($M : 0.600, SD : 0.159$ vs. $M : 0.482, SD : 0.114$). However, this trend is the opposite for experienced users ($M : 0.450, SD : 0.115$ vs. $M : 0.525, SD : 0.185$). Finally, a further pairwise Wilcoxon signed rank test for each pose shows that inexperienced participants training with specific poses significantly made smaller timing errors in the *Trainer-anchored* layout (*Side Plank*: $Z = -2.207, p = 0.027$, and *Down/Up Dog*: $Z = -2.023, p = 0.043$).

7.3.2 User feedback. Figure 13 shows the mean ratings from the questionnaire, split by level of yoga experience. Confidence and value/usefulness scores were uniformly high, showing no statistical differences. Some participants additionally remarked that they preferred this type of system compared to their home workout experience: *"Usually, I have to put too much effort to work out at home. So I tend to stick with a couple of familiar videos to work out without moving the display. I think I would try new exercises with this system."* (P3).

Despite the apparent similar high rankings, we also note that inexperienced yoga participants gave higher scores to the *Trainer-anchored* layout (confidence: $M : 5.000, SD : 1.265$ vs. $M : 4.167, SD : 1.472$, value/usefulness: $M : 6.500, SD : 0.837$ vs. $M : 5.500, SD : 0.837$) because it made it easier for them to understand the instructor's intentions. For example, P10 commented that *"Whenever I looked at the screen, my body moved naturally to the correct position. So, I didn't have to think about left or right."* This could indicate that inexperienced participants prefer an active-guiding system.

On the other hand, 4 of 6 experienced participants stated that they prefer the *User-anchored* layout because it supports a more flexible on-demand learning and observation of details (*"I like that I can always see the instructor's postures [...] In Plank Knee To Elbow, I saw the instructor flexing his feet, which he forgot to mention in the video narration."* — P12). Interestingly, while inexperienced participants generally positively rated the guidance offered by *Trainer-anchored* layout, experienced participants reported that the *Trainer-anchored* layout was both physically and mentally more taxing for them and that it negatively affected their performance resulting in a significantly higher workload ($Z = -2.023, p = 0.043$). When asked why that was the case, we received two types of comments. Some participants disliked being nudged toward specific postures (*"I felt chased"* — P6), while others noticed tracking imperfections.

We also collected more general feedback about the overall usability of the system and on potential distractions caused by the surrounding environment. As in the first study, participants reported about the overall heaviness of the HMD (*"It's too heavy. I wish I could practice with smart glasses."* — P9), but the newly employed strap was effective in stabilizing the HMD, without requiring further adjustments from the users. As to possible distractions caused by the surrounding furniture, all participants stated that the home-like settings did not distract them but rather allowed them to relax and focus on their workouts (*"It was comfortable. It felt like I was working out alone at my home. I didn't notice the household items [...] it was nice to see the instructor's posture more clearly."* — P6). Finally, several participants reported rotational inaccuracies in the

[8]https://www.meta.com/kr/en/quest/accessories/quest-2-elite-strap/

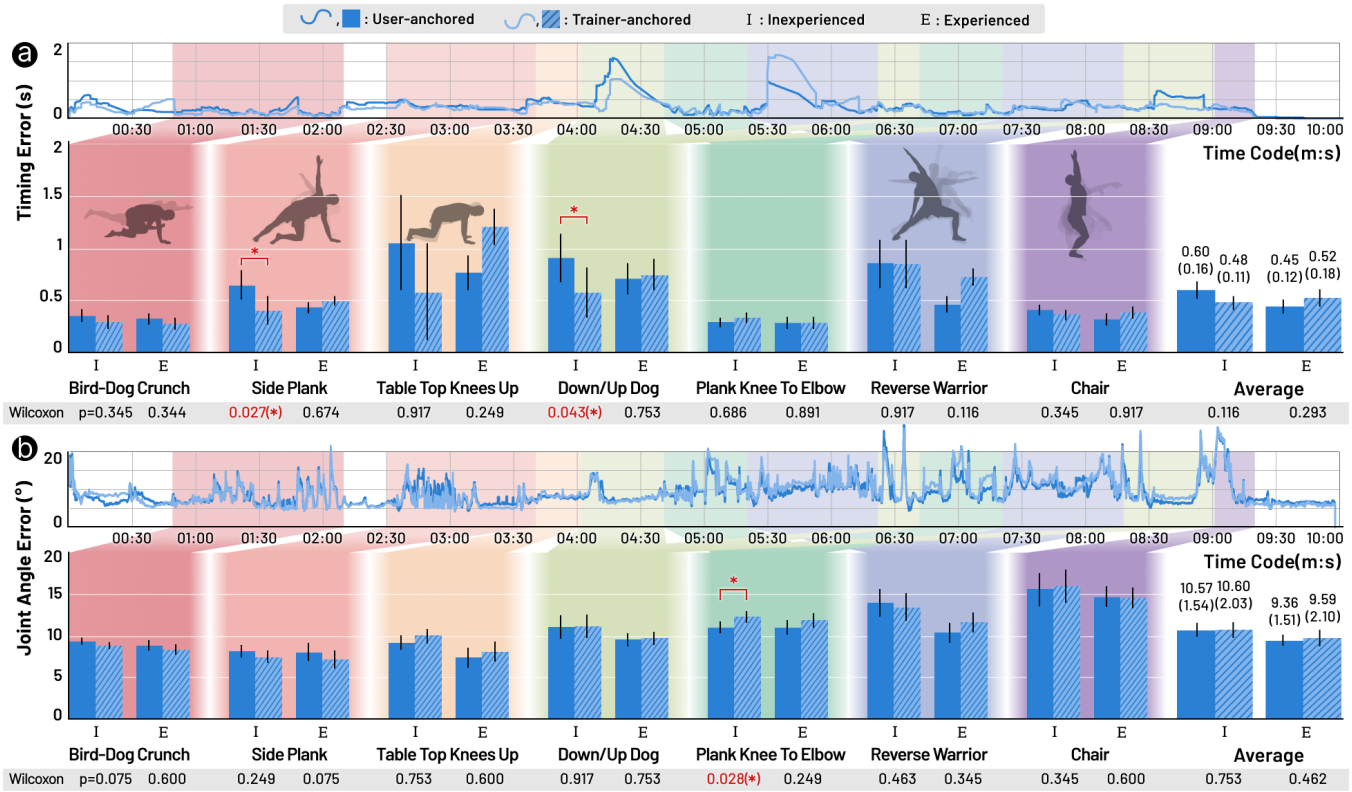

Figure 12: Performance analysis result: ⓐ temporal inaccuracy (timing Error) ⓑ spatial inaccuracy (joint angle error). Mean scores (and standard deviation) are displayed on the bars. Wilcoxon signed-rank test results are written at the bottom.

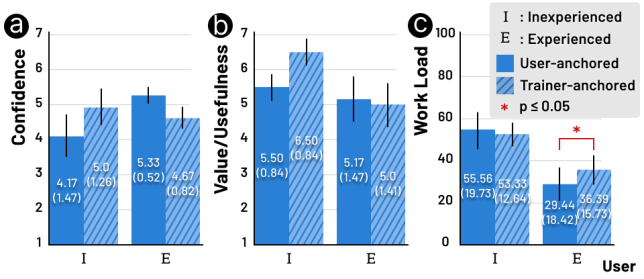

Figure 13: Questionnaire results about the level of ⓐ confidence, ⓑ value/usefulness, and ⓒ work load. Mean ranking scores for the user questionnaires.

screen placement during the *Trainer-anchored* layout. In fact, the overlay is placed using the instructor's head orientation as a proxy for the direction of the gaze, which is not necessarily the same, but that the pose estimation algorithm cannot determine. Despite this inaccuracy, however, the users commented that this problem *"... did not interfere with my performance"* (P9), but also would have appreciated more "concrete feedback on my movement" (P2).

7.4 Summary of results

Three main results emerged from this study:

(1) *FlowAR* can be feasibly deployed in home settings without being affected by the surrounding environment, and participants overall scored it high for perceived value.
(2) Both the performance analysis and the users' scores seem to indicate that inexperienced yoga participants preferred the *Trainer-anchored* layout because it helped them learn, while experienced participants preferred the *User-anchored* layout because it imposed fewer restrictions on their movements and felt more accurate.
(3) From the interview feedback, it emerged that one of the main usability limitations of *FlowAR* is related to the bulkiness and weight of the HMD.

8 DISCUSSION

Combining the results from our studies, we highlight a few emerging themes.

New knowledge about which visualization is most suitable to keep the motion flow

The main contribution of this work is to show that *FlowAR* is effective in keeping users in their motion flow while performing yoga exercises at home following a training video. This is important because maintaining the immersion in a flow is essential in yoga for supporting meditation, blood circulation, and prevention of physical injuries [29, 53]. *FlowAR* enables motion flow by displaying

the reference video using augmented screen overlay visualizations, mitigating the interruptions caused by peeking at the trainer's reference motion displayed on a static screen.

Our analysis of quantitative data and user preferences supports that dynamic screen layouts translate into better motion flow, fewer interruptions due to undesired head movements, and higher user confidence/satisfaction than static visualizations. This corroborates prior work with similar research question [9], but also adds quantitative and expert heuristic metrics to further shed light on the effect of even more screen visualizations and the root cause of motion flow interruptions. For example, the *Circular* layout was already proposed as an alternative to a fixed screen layout (i.e., our *Front* condition) for training in martial art via a Virtual Reality system [9]. For that system as well, no difference emerged between the two static conditions, but no clear explanation was offered about what could be the cause. Through our in-depth quantitative performance analysis with an expert yoga instructor, we identified that the immobility of the screens of both the *Front* and *Circular* layouts cause the *look-aways* that ultimately distract the users and disrupt the motion flow.

Our studies also reveal that there is no optimal dynamic visualization and that both the *User-anchored* and the *Trainer-anchored* layouts have trade-offs. Specifically, it appears from the interviews that experienced users tend to favor layouts that support their existing workflows, while inexperienced users prefer more guidance. Interestingly, this duality of needs for users with different expertise has already been highlighted for other domains (e.g., arts [79]). Our paper presents evidence that this generalizable knowledge also applies to the specific yoga domain. Therefore, we see an opportunity for a system capable of adaptively and dynamically changing the guidance based on the user's expertise (which might also change over time). For example, we could use direct projections on the user's body, like in [66], to provide explicit feedback for beginners. In contrast, we could use a third-person view video streaming like in [20] or a small avatar in place of a flat-screen overlay [73] for users with more experienced proprioception. Similarly, we could use adaptive speed control for slowing down the video playback speed when the user can't keep the pace of the instructor, like in [11], or providing more subtle or more complex guidance via auditory feedback [72] or motion path visualizations, like in [71].

Simplicity may support deployability

The main difference between the proposed *FlowAR* system and previous work is that *FlowAR* does not require a professional-level 3D motion capture system [6, 9, 46] nor the need of authoring 3D animated fitness tutorials [71, 78]. Instead, *FlowAR* leverages ready-available online fitness videos and simple camera-based tracking.

While prior works showed creative solutions for sports training that made good usage of 3D tracking [6, 9], professional-level 3D motion capture systems are expensive, require calibration before each usage, and are unfeasible to install in the users' homes. On the contrary, *FlowAR* uses a much more affordable tracking device that can be easily deployed in the users' homes. We demonstrated this claim in our second study, showing that *FlowAR*'s tracking is robust enough for deployment in a home-like furnished environment.

Furthermore, prior work also relied on ad-hoc 3D content for the training sessions. Although through this process, the creators of the training material have absolute control over every aspect of the tutorial, generating 3D animated avatars is typically long, expensive, and burdensome [43]. Furthermore, despite best efforts, it is unlikely that 3D content creators will ever match the sheer number of existing online training videos (as an example, a survey [77] conducted in 2020 by the Culture&Trends team of YouTube reported that there are already 2,000 channels about yoga). On the contrary, *FlowAR* extracts motion information from existing training videos via a lightweight pose estimation algorithm without requiring additional manual labor.

The simplicity of the tracking and content generation are two unique aspects of *FlowAR* that differentiate it from prior work and would ultimately support a large audience of yogis and sports practitioners who workout at home using training videos.We, however, acknowledge that there are examples of sport training systems that do not require professional-level 3D motion capture or ad-hoc 3D content [11], but these still rely on a fixed display for visualizing the instructional videos and constrain the user's motion. *FlowAR*, in contrast, supports multidirectional workout movements without limiting the user's sight to a fixed location resulting, as we have demonstrated in this paper, in less *look-aways* and improved motion flow.

Opportunities beyond Augmented Reality and Yoga

The greatest technical limitation and source of user discomfort of *FlowAR* is the bulkiness of the Oculus Quest 2 HMD. Still, at the same time, this device offers the opportunity to go beyond augmented reality overlays rendered on top of a live see-through video and to immerse users in a completely virtual environment. The augmented reality offered by the device-embedded cameras allows users to monitor their bodies and movements. With Virtual Reality (VR), this would be possible, but, as shown in previous work [71, 78], tracking and 3D reconstruction of the users' body as a digital avatar are also necessary. For example, *FlowAR* could be used with already existing VR yoga applications where the instructors are represented by 3D avatars [32, 63, 68] or 360 videos [39], but, differently from these work, it could leverage on the same real-time video-based tracking approach [15](BlazePose 3D) to avoid the need for authoring time and preparation of the instructional videos. Overall, an advantage of choosing VR over Augmented Reality is that it enables users to see their own body from an external perspective as a 3D avatar [25], as well as to customize or tailor their avatars [36].

Finally, by constructing completely digital environments and allowing people from different places to join the same virtual space remotely, VR would also naturally lead to the development of group training courses with multiple people practicing together, and also for sports and fitness exercises different from yoga. As shown in previous work, these could include sports where the body posture during the training is important, such as pilates, weight training [25], golf [26], dance [42], and martial art [23]. We believe that *FlowAR*'s screen layouts and the findings presented in this paper would remain relevant regardless of the choice of implementation — Augmented or Virtual Reality — or the choice of training fitness videos.

9 LIMITATIONS AND FUTURE WORK

Despite the positives, *FlowAR* also presents opportunities for improvement. *FlowAR* heavily relies on the HMD for rendering the screen overlays. Heavy and bulky HMDs can cause fatigue for long sessions or impede some yoga exercises, for example, where the head contacts the floor. Smaller and lighter visors, such as future smart glasses or contact lenses, might help address these problems. Other technical limitations of *FlowAR* are related to the accuracy of the 3D pose estimation from the video. The accuracy varies depending on the quality of the input video (e.g., camera angle, lighting conditions), occlusions caused by movements, or the direction in which the trainer is facing. We note, however, that these limitations are inherent to the BlazePose algorithm, and we direct readers to prior work that addresses those [45]. Finally, prior work about yoga training made a point about the importance of supporting breathing instruction [5] and symmetry indicators [13], which we did not account for in our implementation. Future work will need to address these points and more (e.g., pauses and other distractions) for real-world implementation, perhaps using visual or haptic feedback (as in [55, 73]), or even adaptive tracking systems of the user's behavior [11].

Our studies also have some limitations that should be considered when interpreting our results. The relatively small sample size (12-16) and homogeneity of the participants (recruited within the same institution) should be considered when generalizing the results. Furthermore, the study was designed to understand a limited number of visualizations/conditions with a single motion training sequence (constructed by the expert yoga trainer we recruited or sampled from a YouTube video). Future work should attempt to test our results with a richer set of sequences and yoga poses, as well as for other types of fitness exercises beyond yoga.

10 CONCLUSIONS

In conclusion, our paper presented *FlowAR*, a novel augmented reality system for home training using commonly available online yoga videos. The videos are displayed in an HMD as virtual overlays rendered on top of a live camera feed, allowing the user to see the training instructions as well as their motion and the surrounding environment. We experimented with different visualization layouts, clustered into two groups (static and dynamic), and used a motion capture system to detect the trainer's motion. We also added a pose estimation algorithm that allows generating the trainer's reference motion from a commonly available online video and tested the system's applicability and robustness to noise in a furnished lab.

Through a multi-stage analysis (heuristic evaluation from an expert, spatial and temporal performance analysis with motion data, and user feedback), we learned that displaying the visual information on a static overlay screen interrupts the motion flow of the exercise and that dynamic screen layouts are superior to static layouts for both user's performance and satisfaction. We also learned that the two dynamic screen layouts (*User-anchored* and *Trainer-anchored*) are equally efficient, but that the users' preferences might depend on individual experience. We conclude that the level of prior expertise of the users should be considered when determining how much guidance the system should provide. Future avenues of research include addressing the problems related to the bulkiness of the HMD, providing real-time feedback to the users during training, and applying the visualization overlays to the training of other indoor sports.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C1012233).

REFERENCES

[1] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 311–320.

[2] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. 2020. Blazepose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204* (2020).

[3] Richard Bellman and Robert Kalaba. 1959. On adaptive control processes. *IRE Transactions on Automatic Control* 4, 2 (1959), 1–9.

[4] Christina Brown. 2017. *The Modern Yoga Bible*. Hachette UK.

[5] Richard P Brown and Patricia L Gerbarg. 2009. Yoga breathing, meditation, and longevity. *Annals of the New York Academy of Sciences* 1172, 1 (2009), 54–62.

[6] Jacky CP Chan, Howard Leung, Jeff KT Tang, and Taku Komura. 2010. A virtual reality dance training system using motion capture technology. *IEEE transactions on learning technologies* 4, 2 (2010), 187–195.

[7] Minsuk Chang, Mina Huh, and Juho Kim. 2021. Rubyslippers: Supporting content-based voice navigation for how-to videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[8] Hua-Tsung Chen, Yu-Zhen He, Chien-Li Chou, Suh-Yin Lee, Bao-Shuh P Lin, and Jen-Yu Yu. 2013. Computer-assisted self-training system for sports exercise using kinects. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 1–4.

[9] Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins, and Randy Pausch. 2003. Training for physical tasks in virtual environments: Tai Chi. In *IEEE Virtual Reality, 2003. Proceedings.* IEEE, 87–94.

[10] Rachel B Clancy, Matthew P Herring, and Mark J Campbell. 2017. Motivation measures in sport: A critical review and bibliometric analysis. *Frontiers in psychology* 8 (2017), 348.

[11] Christopher Clarke, Doga Cavdir, Patrick Chiu, Laurent Denoue, and Don Kimber. 2020. Reactive Video: Adaptive Video Playback Based on User Motion for Supporting Physical Activity. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 196–208.

[12] Maximilian Dürr, Rebecca Weber, Ulrike Pfeil, and Harald Reiterer. 2020. EGuide: Investigating different visual appearances and guidance techniques for egocentric guidance visualizations. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 311–322.

[13] Alicia Garcia-Falgueras. 2016. An introduction to proprioception concept in Pilates and yoga. *British Journal of Medicine and Medical Research* 15, 3 (2016).

[14] Graham R Gibbs. 2007. Thematic coding and categorizing. *Analyzing qualitative data* 703 (2007), 38–56.

[15] Ivan Grishchenko, Valentin Bazarevsky, Andrei Zanfir, Eduard Gabriel Bazavan, Mihai Zanfir, Richard Yee, Karthik Raveendran, Matsvei Zhdanovich, Matthias Grundmann, and Cristian Sminchisescu. 2022. BlazePose GHUM Holistic: Real-time 3D Human Landmarks and Pose Estimation. *arXiv preprint arXiv:2206.11678* (2022).

[16] Jiajing Guo and Susan R Fussell. 2022. " It's Great to Exercise Together on Zoom!": Understanding the Practices and Challenges of Live Stream Group Fitness Classes. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–28.

[17] Koaburo Hachimura, Hiromu Kato, and Hideyuki Tamura. 2004. A prototype dance training support system with motion capture and mixed reality technologies. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, 217–222.

[18] Perttu Hämäläinen. 2004. Interactive video mirrors for sports training. In *Proceedings of the third Nordic conference on Human-computer interaction*. 199–202.

[19] Natsuki Hamanishi and Jun Rekimoto. 2020. PoseAsQuery: Full-Body Interface for Repeated Observation of a Person in a Video with Ambiguous Pose Indexes and Performed Poses. In *Proceedings of the Augmented Humans International Conference*. 1–11.

[20] Ping-Hsuan Han, Yang-Sheng Chen, Yilun Zhong, Han-Lei Wang, and Yi-Ping Hung. 2017. My Tai-Chi coaches: an augmented-learning tool for practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference*. 1–4.

[21] S. G. Hart. 1986. NASA Task Load Index (TLX). Volume 1.0; Paper and Pencil Package.

[22] Renate Häuslschmid, Benjamin Fritzsche, and Andreas Butz. 2018. Can a helmet-mounted display make motorcycling safer?. In *23rd International Conference on Intelligent User Interfaces*. 467–476.

[23] Tianyu He, Xiaoming Chen, Zhibo Chen, Ye Li, Sen Liu, Junhui Hou, and Ying He. 2017. Immersive and collaborative Taichi motion learning in various VR environments. In *2017 IEEE Virtual Reality (VR)*. IEEE, 307–308.

[24] Thuong N Hoang, Martin Reinoso, Frank Vetere, and Egemen Tanin. 2016. Onebody: remote posture guidance system using first person view in virtual environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. 1–10.

[25] Felix Hülsmann, Cornelia Frank, Irene Senna, Marc O Ernst, Thomas Schack, and Mario Botsch. 2019. Superimposed skilled performance in a virtual mirror improves motor performance and cognitive representation of a full body motor action. *Frontiers in Robotics and AI* 6 (2019), 43.

[26] Atsuki Ikeda, Dong-Hyun Hwang, and Hideki Koike. 2019. A real-time projection system for golf training using virtual shadow. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1527–1528.

[27] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. 2013. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence* 36, 7 (2013), 1325–1339.

[28] Takahiro Iwaanaguchi, Mikio Shinya, Satoshi Nakajima, and Michio Shiraishi. 2015. Cyber tai chi-cg-based video materials for tai chi chuan self-study. In *2015 International Conference on Cyberworlds (CW)*. IEEE, 365–368.

[29] Bellur Krishnamukar Sundara Iyengar. 1979. *Light on yoga: the definitive guide to yoga practice*. Schocken Books.

[30] Susan A Jackson and Mihaly Csikszentmihalyi. 1999. *Flow in sports*. Human Kinetics.

[31] Josée L Jarry, Felicia M Chang, and Loreana La Civita. 2017. Ashtanga yoga for psychological well-being: initial effectiveness study. *Mindfulness* 8, 5 (2017), 1269–1279.

[32] Jijia. 2021. Home YogaVR. virtual reality application. Retrieved September 15, 2022 from https://www.oculus.com/experiences/quest/4828252637184827/

[33] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

[34] Minseong Kim. 2022. How can I Be as attractive as a Fitness YouTuber in the era of COVID-19? The impact of digital attributes on flow experience, satisfaction, and behavioral intention. *Journal of Retailing and Consumer Services* 64 (2022), 102778.

[35] Felix Kosmalla, Florian Daiber, Frederik Wiehr, and Antonio Krüger. 2017. Climbvis: Investigating in-situ visualizations for understanding climbing movements by demonstration. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 270–279.

[36] Jordan Koulouris, Zoe Jeffery, James Best, Eamonn O'neill, and Christof Lutteroth. 2020. Me vs. Super (wo) man: Effects of Customization and Identification in a VR Exergame. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–17.

[37] Matthew Kyan, Guoyu Sun, Haiyan Li, Ling Zhong, Paisarn Muneesawang, Nan Dong, Bruce Elder, and Ling Guan. 2015. An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 2 (2015), 1–37.

[38] Ray Long. 2009. *The key muscles of yoga*. Bandha Yoga Publications LLC.

[39] Immerse Labs Ltd. 2022. Yogistream. virtual reality application. Retrieved September 15, 2022 from https://www.oculus.com/experiences/quest/4832913850073489/

[40] Zhiqiang Luo, Weiting Yang, Zhong Qiang Ding, Lili Liu, I-Ming Chen, Song Huat Yeo, Keck Voon Ling, and Henry Been-Lirn Duh. 2011. "Left Arm Up!" Interactive Yoga Training in Virtual Environment. In *2011 IEEE virtual reality conference*. IEEE, 261–262.

[41] SK Maithel, L Villegas, N Stylopoulos, S Dawson, and DB Jones. 2005. Simulated laparoscopy using a head-mounted display vs traditional video monitor: an assessment of performance and muscle fatigue. *Surgical Endoscopy And Other Interventional Techniques* 19, 3 (2005), 406–411.

[42] Zoe Marquardt, João Beira, Natalia Em, Isabel Paiva, and Sebastian Kox. 2012. Super Mirror: a kinect interface for ballet dancers. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 1619–1624.

[43] Alberto Menache. 2000. *Understanding motion capture for computer animation and video games*. Morgan kaufmann.

[44] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, Vol. 2351. Spie, 282–292.

[45] Sarah Mroz, Natalie Baddour, Connor McGuirk, Pascale Juneau, Albert Tu, Kevin Cheung, and Edward Lemaire. 2021. Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose. In *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. IEEE, 1–4.

[46] Akio Nakamura, Sou Tabata, Tomoya Ueda, Shinichiro Kiyofuji, and Yoshinori Kuno. 2005. Dance training system with active vibro-devices and a mobile image display. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3075–3080.

[47] Francesca M Nicosia, Nadra E Lisha, Margaret A Chesney, Leslee L Subak, Traci M Plaut, and Alison Huang. 2020. Strategies for evaluating self-efficacy and observed success in the practice of yoga postures for therapeutic indications: methods from a yoga intervention for urinary incontinence among middle-aged and older women. *BMC complementary medicine and therapies* 20, 1 (2020), 1–13.

[48] Kate Parker, Riaz Uddin, Nicola D Ridgers, Helen Brown, Jenny Veitch, Jo Salmon, Anna Timperio, Shannon Sahlqvist, Samuel Cassar, Kim Toffoletti, et al. 2021. The use of digital platforms for adults' and adolescents' physical activity during the COVID-19 pandemic (our life at home): Survey study. *Journal of medical Internet research* 23, 2 (2021), e23389.

[49] Ashwin Ram and Shengdong Zhao. 2021. LSVP: Towards Effective On-the-go Video Learning Using Optical Head-Mounted Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.

[50] David Riley. 2004. Hatha yoga and the treatment of illness. *Alternative therapies in health and medicine* 10, 2 (2004), 20–25.

[51] Jannick P Rolland, Richard L Holloway, and Henry Fuchs. 1995. Comparison of optical and video see-through, head-mounted displays. In *Telemanipulator and Telepresence Technologies*, Vol. 2351. SPIE, 293–307.

[52] Richard M Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology* 43, 3 (1982), 450.

[53] Paul Salmon, Elizabeth Lush, Megan Jablonski, and Sandra E Sephton. 2009. Yoga and mindfulness: Clinical aspects of an ancient mind/body practice. *Cognitive and behavioral practice* 16, 1 (2009), 59–72.

[54] Krishanu Sarker, Mohamed Masoud, Saeid Belkasim, and Shihao Ji. 2018. Towards robust human activity recognition from rgb video stream with limited labeled data. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 145–151.

[55] Christian Schönauer, Kenichiro Fukushi, Alex Olwal, Hannes Kaufmann, and Ramesh Raskar. 2012. Multimodal motion guidance: techniques for adaptive and dynamic feedback. In *Proceedings of the 14th ACM international conference on Multimodal interaction*. 133–140.

[56] Ari Shapiro, Andrew Feng, Ruizhe Wang, Hao Li, Mark Bolas, Gerard Medioni, and Evan Suma. 2014. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds* 25, 3-4 (2014), 201–211.

[57] Joon Gi Shin, Doheon Kim, Chaehan So, and Daniel Saakes. 2020. Body Follows Eye: Unobtrusive Posture Manipulation Through a Dynamic Content Position in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[58] Mohammad Shokoohi-Yekta, Bing Hu, Hongxia Jin, Jun Wang, and Eamonn Keogh. 2017. Generalizing DTW to the multi-dimensional case requires an adaptive approach. *Data mining and knowledge discovery* 31, 1 (2017), 1–31.

[59] C Alexander Simpkins and Annellen M Simpkins. 2012. *Yoga Basics: The Basic Poses and Routines you Need to be Healthy and Relaxed*. Tuttle Publishing.

[60] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 179–188.

[61] Karina Sokolova and Charles Perez. 2021. You follow fitness influencers on YouTube. But do you actually exercise? How parasocial relationships, and watching fitness influencers, relate to intentions to exercise. *Journal of Retailing and Consumer Services* 58 (2021), 102276.

[62] Mark Stephens. 2012. *Yoga sequencing: Designing transformative yoga classes*. North Atlantic Books.

[63] Soaring Roc Studio. 2021. Rhythm Yoga. virtual reality application. Retrieved September 15, 2022 from https://www.viveport.com/apps/70755b3a-4397-46a8-a33b-5abb9f832f32/

[64] Adam Switonski, Henryk Josinski, and Konrad Wojciechowski. 2019. Dynamic time warping in classification and selection of motion capture data. *Multidimensional Systems and Signal Processing* 30, 3 (2019), 1437–1468.

[65] Martijn Ten Bhömer and Hanxiao Du. 2018. Designing Personalized Movement-based Representations to Support Yoga. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*. 283–287.

[66] Laia Turmo Vidal, Elena Márquez Segura, Christopher Boyer, and Annika Waern. 2019. Enlightened Yoga: Designing an Augmented Class with Wearable Lights to Support Instruction. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 1017–1031.

[67] Manisha Verma, Sudhakar Kumawat, Yuta Nakashima, and Shanmuganathan Raman. 2020. Yoga-82: a new dataset for fine-grained classification of human poses. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 1038–1039.

[68] VirtuaLiron. 2019. VirtuaLiron. virtual reality application. Retrieved September 15, 2022 from https://www.viveport.com/cdb9dec4-1d00-40f0-958f-8e2075b2b27f

[69] Elizabeth Whissell, Lin Wang, Pan Li, Jing Xian Li, and Zhen Wei. 2021. Biomechanical Characteristics on the Lower Extremity of Three Typical Yoga Manoeuvres. *Applied Bionics and Biomechanics* 2021 (2021).

[70] Jason Wise. 2022. *How many people do yoga in 2022? (yoga statistics).* Retrieved Sep 15, 2022 from https://earthweb.com/how-many-people-do-yoga/

[71] Erwin Wu, Florian Perteneder, Hideki Koike, and Takayuki Nozawa. 2019. How to vizski: Visualizing captured skier motion in a vr ski training simulator. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. 1–9.

[72] Chengshuo Xia, Xinrui Fang, Riku Arakawa, and Yuta Sugiura. 2022. VoLearn: A Cross-Modal Operable Motion-Learning System Combined with Virtual Avatar and Auditory Feedback. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–26.

[73] Shuo Yan, Gangyi Ding, Zheng Guan, Ningxiao Sun, Hongsong Li, and Longfei Zhang. 2015. Outsideme: Augmenting dancer's external self-image by using a mixed reality system. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 965–970.

[74] Ungyeon Yang and Gerard Jounghyun Kim. 2002. Implementation and Evaluation of "Just Follow Me": An Immersive, VR-Based, Motion-Training System. *Presence* 11, 3 (2002), 304–323. https://doi.org/10.1162/105474602317473240

[75] Yinghua Yang. 2018. YogAR. augmented reality application. Retrieved September 15, 2022 from https://www.microsoft.com/en-us/p/yogar/9p47cc593fh9?activetab=pivot:overviewtab

[76] Jang W Yoon, Robert E Chen, Esther J Kim, Oluwaseun O Akinduro, Panagiotis Kerezoudis, Phillip K Han, Phong Si, William D Freeman, Roberto J Diaz, Ricardo J Komotar, et al. 2018. Augmented reality for the surgeon: systematic review. *The International Journal of Medical Robotics and Computer Assisted Surgery* 14, 4 (2018), e1914.

[77] YouTube. 2021. *Community Spotlight: Yoga.* Retrieved Sep 14, 2022 from https://www.youtube.com/trends/articles/stay-home-workout-at-home/

[78] Xingyao Yu, Katrin Angerbauer, Peter Mohr, Denis Kalkofen, and Michael Sedlmair. 2020. Perspective matters: Design implications for motion guidance in mixed reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 577–587.

[79] Amit Zoran, Roy Shilkrot, Suranga Nanyakkara, and Joseph Paradiso. 2014. The hybrid artisans: A case study in smart tools. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 3 (2014), 1–29.

A APPENDIX

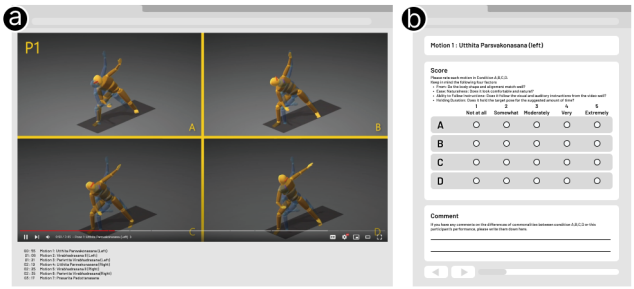

Figure 14: Expert Evaluation Material: ⓐ video and ⓑ Google survey. The video shows each participant's motion in yellow 3D character animation overlapped with the blue semi-transparent virtual trainer for each screen layout condition(A-D). The order of screen layouts is randomized.

Table 1: 16 symptoms in Simulator Sickness Questionnaire (SSQ) [33]. Users rate each symptom on a 4-point scale (0: none, 1: slight, 2: moderate, 3: severe) for each symptom.

No.	Symptoms	No.	Symptoms
1	General discomfort	9	Difficulty concentrating
2	Fatigue	10	Fullness of head
3	Headache	11	Blurred vision
4	Eyestrain	12	Dizzy (eyes open)
5	Difficulty focusing	13	Dizzy (eyes closed)
6	Increased Salivation	14	Vertigo
7	Sweating	15	Stomach awareness
8	Nausea	16	Burping

Table 2: Items in Intrinsic Motivation Inventory (IMI) [52]. Users rate each item on a 7-Likert scale (1: not at all true, 4: somewhat true, 7: very true) for each item.

Subscales of IMI	No.	Items
Perceived Competence	1	I think I am pretty good at this activity.
	2	I think I did pretty well at this activity, compared to other students.
	3	After working at this activity for a while, I felt pretty competent.
	4	I am satisfied with my performance at this task.
	5	I was pretty skilled at this activity.
	6	This was an activity that I couldn't do very well. (Reverse score)
Value / Usefulness	1	I believe this *screen layout* could be of some value to me.
	2	I think that doing this *screen layout* is useful for *learning yoga movement.*
	3	I think this *screen layout* is important in *physical training.*
	4	I would be willing to do this again because it has some value to me.
	5	I think doing this *screen layout* help me to *practice yoga.*
	6	I believe doing this *screen layout* could be beneficial to me.
	7	I think this is an important *screen layout.*