A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014

Arindam Dey 1*, Mark Billinghurst 1, Robert W. Lindeman 2 and J. Edward Swan II 3

1 Empathic Computing Laboratory, University of South Australia, Mawson Lakes, SA, Australia, 2 Human Interface Technology Lab New Zealand (HIT Lab NZ), University of Canterbury, Christchurch, New Zealand, 3 Mississippi State University, Starkville, MS, United States

Augmented Reality (AR) interfaces have been studied extensively over the last few decades, with a growing number of user-based experiments. In this paper, we systematically review 10 years of the most influential AR user studies, from 2005 to 2014. A total of 291 papers with 369 individual user studies have been reviewed and classified based on their application areas. The primary contribution of the review is to present the broad landscape of user-based AR research, and to provide a high-level view of how that landscape has changed. We summarize the high-level contributions from each category of papers, and present examples of the most influential user studies. We also identify areas where there have been few user studies, and opportunities for future research. Among other things, we find that there is a growing trend toward handheld AR user studies, and that most studies are conducted in laboratory settings and do not involve pilot testing. This research will be useful for AR researchers who want to follow best practices in designing their own AR user studies.

Keywords: augmented reality, systematic review, user studies, usability, experimentation, classifications

1. INTRODUCTION

Augmented Reality (AR) is a technology field that involves the seamless overlay of computer generated virtual images on the real world, in such a way that the virtual content is aligned with real world objects, and can be viewed and interacted with in real time (Azuma, 1997). AR research and development has made rapid progress in the last few decades, moving from research laboratories to widespread availability on consumer devices. Since the early beginnings in the 1960’s, more advanced and portable hardware has become available, and registration accuracy, graphics quality, and device size have been largely addressed to a satisfactory level, which has led to a rapid growth in the adoption of AR technology. AR is now being used in a wide range of application domains, including Education (Furió et al., 2013; Fonseca et al., 2014a; Ibáñez et al., 2014), Engineering (Henderson and Feiner, 2009; Henderson S. J. and Feiner, 2011; Irizarry et al., 2013), and Entertainment (Dow et al., 2007; Haugstvedt and Krogstie, 2012; Vazquez-Alvarez et al., 2012). However, to be widely accepted by end users, AR usability and user experience issues still need to be improved.

To help the AR community improve usability, this paper provides an overview of 10 years of AR user studies, from 2005 to 2014. Our work builds on the previous reviews of AR usability research shown in Table 1. These years were chosen because they cover an important gap in other reviews, and also are far enough from the present to enable the impact of the papers to be measured. Our goals are to provide a broad overview of user-based AR research, to help researchers find example papers that contain related studies, to help identify areas where there have been few user studies...
TABLE 1 | Summary of earlier surveys of AR usability studies.

| Publication                        | Venues considered                      | Coverage years | Total reviewed publications |
|------------------------------------|----------------------------------------|----------------|-----------------------------|
| Swan and Gabbard, 2005              | IEEE ISMAR, ISWC, IEEE VR, and Presence | 1992–2004       | 21                          |
| Dünser et al., 2008                | All venues in IEEE Xplore, ACM Digital Library, and Springer Link | 1992–2007       | 166                         |
| Bai and Blackwell, 2012            | IEEE ISMAR                             | 2001–2010       | 71                          |
| This survey [2017]                 | All venues indexed in Scopus           | 2005–2014       | 291                         |

conducted, and to highlight exemplary user studies that embody best practices. We therefore hope the scholarship in this paper leads to new research contributions by providing outstanding examples of AR user studies that can help current AR researchers.

1.1. Previous User Study Survey Papers

Expanding on the studies shown in Table 1, Swan and Gabbard (2005) conducted the first comprehensive survey of AR user studies. They reviewed 1,104 AR papers published in four important venues between 1992 and 2004; among these papers they found only 21 that reported formal user studies. They classified these user study papers into three categories: (1) low-level perceptual and cognitive issues such as depth perception, (2) interaction techniques such as virtual object manipulation, and (3) collaborative tasks. The next comprehensive survey was by Dünser et al. (2008), who used a list of search queries across several common bibliographic databases, and found 165 AR-related publications reporting user studies. In addition to classifying the papers into the same categories as Swan and Gabbard (2005), they additionally classified the papers based on user study methods such as objective, subjective, qualitative, and informal. In another literature survey, Bai and Blackwell (2012) reviewed 71 AR papers reporting user studies, but they only considered papers published in the International Symposium on Mixed and Augmented Reality (ISMAR) between 2001 and 2010. They also followed the classification of Swan and Gabbard (2005), but additionally identified a new category of studies that investigated user experience (UX) issues. Their review thoroughly reported the evaluation goals, performance measures, UX factors investigated, and measurement instruments used. Additionally, they also reviewed the demographics of the studies’ participants. However there has been no comprehensive study since 2010, and none of these earlier studies used an impact measure to determine the significance of the papers reviewed.

1.1.1. Survey Papers of AR Subsets

Some researchers have also published review papers focused on more specific classes of user studies. For example, Kruijff et al. (2010) reviewed AR papers focusing on the perceptual pipeline, and identified challenges that arise from the environment, capturing, augmentation, display technologies, and user. Similarly, Livingston et al. (2013) published a review of user studies in the AR X-ray vision domain. As such, their review deeply analyzed perceptual studies in a niche AR application area. Finally, Rankohi and Waugh (2013) reviewed AR studies in the construction industry, although their review additionally considers papers without user studies. In addition to these papers, many other AR papers have included literature reviews which may include a few related user studies such as Wang et al. (2013), Carmigniani et al. (2011), and Papagiannakis et al. (2008).

1.2. Novelty and Contribution

These reviews are valued by the research community, as shown by the number of times they have been cited (e.g., 166 Google Scholar citations for Dünser et al., 2008). However, due to a number of factors there is a need for a more recent review. Firstly, while early research in AR was primarily based on head-mounted displays (HMDs), in the last few years there has been a rapid increase in the use of handheld AR devices, and more advanced hardware and sensors have become available. These new wearable and mobile devices have created new research directions, which have likely impacted the categories and methods used in AR user studies. In addition, in recent years the AR field has expanded, resulting in a dramatic increase in the number of published AR papers, and papers with user studies in them. Therefore, there is a need for a new categorization of current AR user research, as well as the opportunity to consider new classification measures such as paper impact, as reviewing all published papers has become less plausible. Finally, AR papers are now appearing in a wider range of research venues, so it is important to have a survey that covers many different journals and conferences.

1.2.1. New Contributions Over Existing Surveys

Compared to these earlier reviews, there are a number of important differences with the current survey, including:

- we have considered a larger number of publications from a wide range of sources
- our review covers more recent years than earlier surveys
- we have used paper impact to help filter the papers reviewed
- we consider a wider range of classification categories
- we also review issues experienced by the users.

1.2.2. New Aims of This Survey

To capture the latest trends in usability research in AR, we have conducted a thorough, systematic literature review of 10 years of AR papers published between 2005 and 2014 that contain a user study. We classified these papers based on their application areas, methodologies used, and type of display examined. Our aims are to:

1. identify the primary application areas for user research in AR
2. describe the methodologies and environments that are commonly used
3. propose future research opportunities and guidelines for making AR more user friendly.

The rest of the paper is organized as follows: section 2 details the method we followed to select the papers to review, and how we conducted the reviews. Section 3 then provides a high-level overview of the papers and studies, and introduces the classifications. The following sections report on each of the classifications in more detail, highlighting one of the more impactful user studies from each classification type. Section 5 concludes by summarizing the review and identifying opportunities for future research. Finally, in the appendix we have included a list of all papers reviewed in each of the categories with detailed information.

2. METHODOLOGY

We followed a systematic review process divided into two phases: the search process and the review process.

2.1. Search Process

One of our goals was to make this review as inclusive as practically possible. We therefore considered all papers published in conferences and journals between 2005 and 2014, which include the term “Augmented Reality,” and involve user studies. We searched the Scopus bibliographic database, using the same search terms that were used by Dünser et al. (2008) (Table 2). This initial search resulted in a total of 1,147 unique papers. We then scanned each one to identify whether or not it actually reported on AR research; excluding papers not related to AR reduced the number to 1,063. We next removed any paper that did not actually report on a user study, which reduced our pool to 604 papers. We then examined these 604 papers, and kept only those papers that provided all of the following information: (i) participant demographics (number, age, and gender), (ii) design of the user study, and (iii) the experimental task. Only 396 papers satisfied all three of these criteria. Finally, unlike previous surveys of AR usability studies, we next considered how much impact each paper had, to ensure that we were reviewing papers that others had cited. For each paper we used Google Scholar to find the total citations to date, and calculated its Average Citation Count (ACC)

\[ ACC = \frac{\text{total lifetime citations}}{\text{lifetime (years)}} \]  

For example, if a paper was published in 2010 (a 5 year lifetime until 2014) and had a total of 10 citations in Google Scholar in April 2015, its ACC would be 10/5 = 2.0. Based on this formula, we included all papers that had an ACC of at least 1.5, showing that they had at least a moderate impact in the field. This resulted in a final set of 291 papers that we reviewed in detail. We deliberately excluded papers more recent than 2015 because most of these hadn’t gather significant citations yet.

2.2. Reviewing Process

In order to review this many papers, we randomly divided them among the authors for individual review. However, we first performed a norming process, where all of the authors first reviewed the same five randomly selected papers. We then met to discuss our reviews, and reached a consensus about what review data would be captured. We determined that our reviews would focus on the following attributes:

- application areas and keywords
- experimental design (within-subjects, between-subjects, or mixed-factorial)
- type of data collected (qualitative or quantitative)
- participant demographics (age, gender, number, etc.)
- experimental tasks and environments
- type of experiment (pilot, formal, field, heuristic, or case study)
- senses augmented (visual, haptic, olfactory, etc.)
- type of display used (handheld, head-mounted display, desktop, etc.).

In order to systematically enter this information for each paper, we developed a Google Form. During the reviews we also flagged certain papers for additional discussion. Overall, this reviewing phase encompassed approximately 2 months. During this time, we regularly met and discussed the flagged papers; we also clarified any concerns and generally strove to maintain consistency. At the end of the review process we had identified the small number of papers where the classification was unclear, so we held a final meeting to arrive at a consensus view.

2.3. Limitations and Validity Concerns

Although we strove to be systematic and thorough as we selected and reviewed these 291 papers, we can identify several limitations and validity concerns with our methods. The first involves using the Scopus bibliographic database. Although using such a database has the advantage of covering a wide range of publication venues and topics, and although it did cover all of the venues where the authors are used to seeing AR research, it remains possible that Scopus missed publication venues and papers that should have been included. Second, although the search terms we used seem intuitive (Table 2), there may have been papers that did not use “Augmented Reality” as a keyword when describing an AR experience. For example, some papers may have used the term “Mixed Reality,” or “Artificial Reality.”

| TABLE 2 | Search terms used in the Scopus database. |
| "Augmented reality" AND "user evaluation(s)" |
| "Augmented reality" AND "user study/-ies" |
| "Augmented reality" AND "feedback" |
| "Augmented reality" AND "experiment(s)" |
| "Augmented reality" AND "pilot study" |
| "Augmented reality" AND participant AND study |
| "Augmented reality" AND participant AND experiment |
| "Augmented reality" AND subject AND study |
| "Augmented reality" AND subject AND experiment |

We searched in Title, Abstract, and Keywords fields.
Finally, although using the ACC as a selection factor narrowed the initial 604 papers to 291, it is possible that the ACC excluded papers that should have been included. In particular, because citations are accumulated over time, it is quite likely that we missed some papers from the last several years of our 10-year review period that may soon prove influential.

3. HIGH-LEVEL OVERVIEW OF REVIEWED PAPERS

Overall, the 291 papers report a total of 369 studies. Table 3 gives summary statistics for the papers, and Table 4 gives summary statistics for the studies. These tables contain bar graphs that visually depict the magnitude of the numbers; each color indicates the number of columns are spanned by the bars. For example, in Table 3 the columns Paper, Mean ACC, and Mean Author Count are summarized individually, and the longest bar in each column is scaled according to the largest number in that column. However, Publications spans two columns, and the largest value is 59, and so all of the other bars for Publications are scaled according to 59.

Figure 1 further summarizes the 291 papers through four graphs, all of which indicate changes over the 10 year period between 2005 and 2014. Figure 1A shows the fraction of the total number of AR papers that report user studies, Figure 1B analyzes the kind of display used, Figure 1C categorizes the experiments into application areas, and Figure 1D categorizes the papers according to the kind of experiment that was conducted.

3.1. Fraction of User Studies Over Time

Figure 1A shows the total number of AR papers published between 2005 and 2014, categorized by papers with and without a user study. As the graph shows, the number of AR papers published in 2014 is five times that published in 2005. However,
the proportion of user study papers among all AR papers has remained low, less than 10% of all publication for each year.

### 3.2. Study Design

As shown in Table 4, most of the papers (213, or 73%) used a within-subjects design, 43 papers (15%) used a between-subjects design, and 12 papers (4%) used a mixed-factorial design. However, there were 23 papers (8%) which used different study designs than the ones mentioned above, such as Baudisch et al. (2013), Benko et al. (2014), and Olsson et al. (2009).

### 3.3. Study Type

We found that it was relatively rare for researchers to report on conducting pilot studies before their main study. Only 55 papers (19%) reported conducting at least one pilot study in their experimentation process and just 25 of them reported the pilot studies with adequate details such as study design, participants, and results. This shows that the importance of pilot studies is not well recognized. The majority of the papers (221, or 76%) conducted the experiments in controlled laboratory environments, while only 44 papers (15%) conducted the experiments in a natural environment or as a field study (Figure 1D). This shows a lack of experimentation in real world conditions. Most of the experiments were formal user studies, and there were almost no heuristic studies, which may indicate that the heuristics of AR applications are not fully developed and there exists a need for heuristics and standardization.

### 3.4. Data Type

In terms of data collection, a total of 139 papers (48%) collected both quantitative and qualitative data, 78 (27%) papers only qualitative, and 74 (25%) only quantitative. For the experimental task, we found that the most popular task involved performance (178, or 61%), followed by filling out questionnaires (146, or 50%), perceptual tasks (53, or 18%), interviews (41, or 14%) and collaborative tasks (21, or 7%). In terms of dependent measures, subjective ratings were the most popular with 167 papers (57%), followed by error/accuracy measures (130, or 45%), and task completion time (123, or 42%). We defined task as any activity that was carried out by the participants to provide data—both quantitative and/or qualitative—about the experimental system(s). Note that many experiments used more than one experimental task or dependent measure, so the percentages sum to more than 100%. Finally, the bulk of the user studies were conducted in an indoor environment (246, or 83%), not outdoors (43, or 15%), or a combination of both settings (6, or 2%).

---

**Table 3: Summary of the 291 reviewed papers.**

| Application Area          | Paper | Mean ACC | Mean Count | Publication | Display* | Data Collected |
|---------------------------|-------|----------|------------|-------------|----------|----------------|
| Collaboration             | 12    | 6.4      | 3.58       | 1           | HMD      | 1 5 6          |
| Education                 | 42    | 5.5      | 3.24       | 23          | HHD      | 7 19 16        |
| Entertainment and Gaming  | 14    | 4.2      | 4.71       | 2           | Other    | 2 5 7          |
| Industrial                | 30    | 5.7      | 3.87       | 14          | Other    | 1 4 22         |
| Interaction               | 67    | 5.1      | 3.64       | 8           | HMD      | 11 42          |
| Medical                   | 43    | 4.2      | 6.02       | 35          | HHD      | 14 42          |
| Navigation and Driving    | 24    | 4.2      | 4.58       | 10          | Other    | 1 2 15         |
| Perception                | 51    | 4.9      | 3.75       | 16          | HMD      | 1 2 17         |
| Tourism and Exploration   | 8     | 6.3      | 3.63       | 4           | Other    | 1 7            |
| **Overall**               | 291   | 5.2      | 4.1        | 113         | 178      | 102 100 115    |

*HMD = Head Mounted Display, HHD = Hand-Held Display

**Table 4: Summary of the 369 user studies reported by the 291 reviewed papers.**

| Application Area          | Study Type | Study Design | Study Location | Median Participants |
|---------------------------|------------|--------------|----------------|---------------------|
| Collaboration             | Lab        | Field        | Heuristic      | Pilot              | 14             | 1 14 1         | 12             |
| Education                 | 28         | 14           | 13             |                   | 31 16           | 7 49 2         | 28             |
| Entertainment and Gaming  | 13         | 5            |                |                   | 14 2            | 2 13 4         | 17             |
| Industrial                | 29         | 6            | 1              |                   | 28 6            | 2 32 4         | 15             |
| Interaction               | 21         | 5            | 1              |                   | 20 2 6          | 5 25 7         | 14             |
| Medical                   | 39         | 7            | 4              |                   | 36 11           | 7 54           | 13             |
| Navigation and Driving    | 23         | 5            |                |                   | 21 3            | 1 21 7         | 18             |
| Perception                | 60         | 3            | 1              |                   | 52 12           | 1 52 17        | 16             |
| Tourism and Exploration   | 3          | 6            |                |                   | 6               | 3 9            | 28             |
| **Overall**               | 278        | 54           | 6              | 31                 | 272 52 16       | 29 319 44 6    | 16             |
3.5. Senses
As expected, an overwhelming majority of papers (281, or 96%) augmented the visual sense. Haptic and Auditory senses were augmented in 27 (9%) and 21 (7%) papers respectively. Only six papers (2%) reported augmenting only the auditory sense and five (2%) papers reported augmenting only the haptic sense. This shows that there is an opportunity for conducting more user studies exploring non-visual senses.

3.6. Participants
The demographics of the participants showed that most of the studies were run with young participants, mostly university students. A total of 182 papers (62%) used participants with an approximate mean age of less than 30 years. A total of 227 papers (78%) reported involving female participants in their experiments, but the ratio of female participants to male participants was low (43% of total participants in those 227 papers). When all 291 papers are considered only 36% of participants were female. Many papers (117, or 40%) did not explicitly mention the source of participant recruitment. From those that did, most (102, or 35%) sourced their participants from universities, whereas only 36 papers (12%) mentioned sourcing participants from the general public. This shows that many AR user studies use young male university students as their subjects, rather than a more representative cross section of the population.

3.7. Displays
We also recorded the displays used in these experiments (Table 3). Most of the papers used either HMDs (102 papers, or 35%) or handhelds (100 papers, or 34%), including six papers that used both. Since 2009, the number of papers using HMDs started to decrease while the number of papers using handheld displays increased (Figure 1B). For example, between 2010 and 2014 (204 papers in our review), 50 papers used HMDs and 79 used handhelds, including one paper that used both, and since 2011 papers using handheld displays consistently outnumbered papers using HMDs. This trend—that handheld mobile AR has recently become the primary display for AR user studies—is of course driven by the ubiquity of smartphones.

3.8. Categorization
We categorized the papers into nine different application areas (Tables 3, 4): (i) Perception (51 papers, or 18%), (ii) Medical (43, or 15%), (iii) Education (42, or 14%), (iv) Entertainment and Gaming (14, or 5%), (v) Industrial (30, or 10%), (vi) Navigation and Driving (24, or 9%), (vii) Tourism and Exploration (8, or 2%), (viii) Collaboration (12, or 4%), and (ix) Interaction (67, or 23%). Figure 1C shows the change over time in number of AR papers with user studies in these categories. The Perception and Interaction categories are rather general areas of AR research, and contain work that reports on more low-level experiments, possibly across multiple application areas. Our analysis shows that there are fewer AR user studies published in Collaboration, Tourism and Exploration, and Entertainment and Gaming, identifying future application areas for user studies. There is also a noticeable increase in the number of user studies in educational applications over time. The drop in number of papers in 2014 is due to the selection criteria of papers having at least 1.5 average citations per year, as these papers were too recent to be cited often. Interestingly, although there were relatively few of them, papers in Collaboration, Tourism and Exploration categories received noticeably higher ACC scores than other categories.

3.9. Average Authors
As shown in Table 3, most categories had a similar average number of authors for each paper, ranging between 3.24 (Education) and 3.87 (Industrial). However papers in the Medical domain had the highest average number of authors (6.02), which indicates the multidisciplinary nature of this research area. In contrast to all other categories, most of the papers in the Medical category were published in journals, compared to the common AR publications venues, which are mostly conferences. Entertainment and Gaming (4.71), and Navigation and Driving (4.58) also had considerably higher numbers of authors per paper on average.

3.10. Individual Studies
While a total of 369 studies were reported in these 291 papers (Table 4), the majority of the papers (231, or 80%) reported only one user study. Forty-seven (16.2%), nine (3.1%), two (<1%), and one (<1%) papers reported two, three, four, and five studies respectively, including pilot studies. In terms of the number of participants used (median) in each study, Tourism and Exploration, and Education were the highest among all categories with an average of 28 participants per study. Other categories used between 12 and 18 participants per study, while the overall median stands at 16 participants. Based on this insight, it can be claimed that 12 to 18 participants per study is a typical range in the AR community. Out of the 369 studies 31 (8.4%) were pilot studies, six (1.6%) heuristic evaluation, 54 (14.6%) field studies, and rest of the 278 (75.3%) were formal controlled user studies. Most of the studies (272, or 73.7%) were designed as within-subjects, 52 (14.1%) between-subjects, and 16 (4.3%) as mixed-factors (Table 4).

In the following section we review user studies in each of the nine application areas separately. We provide a commentary on each category and also discuss a representative paper with the highest ACCs in each application area, so that readers can understand typical user studies from that domain. We present tables summarizing all of the papers from these areas at the end of the paper.

4. APPLICATION AREAS
4.1. Collaboration
A total of 15 studies were reported in 12 papers in the Collaboration application area. The majority of the studies investigated some form of remote collaboration (Table 5), although Henrysson et al. (2005a) presented a face-to-face collaborative AR game. Interestingly, out of the 15 studies, eight reported using handheld displays, seven used HMDs, and six used some form of desktop display. This makes sense as...
TABLE 5 | Summary of user studies in Collaboration application area.

| References         | Topic                          | Data type | Displays used | Dependent variables                                    | Study type | Participants (female) |
|--------------------|--------------------------------|-----------|---------------|--------------------------------------------------------|------------|-----------------------|
| Almeida et al., 2012 | AR based video meetings       | S         | DT            | Rating                                                 | Formal     | 10 (0)                |
| Chastine et al., 2007 | Collaboration                | S         | HMD           | Interview answers                                     | Formal     | 16 (4)                |
| Chen et al., 2013 | Remote collaboration          | O + S     | HH            | Time, Subjective feedback                             | Field      | 16 (7)                |
| Gauglitz et al., 2012 | Remote collaboration with an expert | O + S    | HH, DT        | Error/Accuracy, Rating, Completed task count          | Formal     | 48 (21)               |
| Gauglitz et al., 2014a | Annotations in remote Collaboration | S         | HH, DT, DT touchscreen | User preference                                      | Field      | 11 (5)                |
| Gauglitz et al., 2014b | Remote collaboration           | O + S     | HH, DT        | Time, Error/Accuracy, Rating                          | Formal     | 60 (29)               |
| Grasset et al., 2005 | Collaboration                 | O + S     | HMD           | Time, Error/Accuracy, Rating                          | Formal     | 14 (2)                |
| Henrysson et al., 2005a | Games, Interaction, Tangible Interfaces | O         | HH            | Rating                                                 | Formal     | 12 (0)                |
| Kasahara and Rekimoto, 2014 | Remote Collaboration          | O + S     | HMD           | Time, Rating, Body movement                           | Formal     | 10 (0)                |
| Poelman et al., 2012 | Remote Collaboration, Crime Scene Investigation | S         | HMD           | Observation and discussion                            | Field      | 5 (0)                 |
| Sodhi et al., 2013 | Remote Collaboration          | S         | HH            | Rating                                                 | Formal     | 8 (1)                 |
| Wang and Dunston, 2011 | Collaboration                 | O + S     | HMD           | Time, NASA TLX                                        | Formal     | 16 (4)                |

S, Subjective; O, Objective; DT, Desktop; HH, handheld. Participant numbers are absolute values, and where more than one study was reported in the paper we used average counts.

collaborative interfaces often require at least one collaborator to be stationary and desktop displays can be beneficial in such setups. One noticeable feature was the low number of studies performed in the wild or in natural settings (field studies). Only three out of 15 studies were performed in natural settings and there were no pilot studies reported, which is an area for potential improvement. While 14 out of 15 studies were designed to be within-subjects, only 12 participants were recruited per study. On average, roughly one-third of the participants were females in all studies considered together. All studies were performed in indoor locations except for (Gauglitz et al., 2014b), which was performed in outdoors. While a majority of the studies (8) collected both objective (quantitative) and subjective (qualitative) data, five studies were based on only subjective data, and two studies were based on only objective data, both of which were reported in one paper (Henrysson et al., 2005a). Besides subjective feedback or ratings, task completion time and error/accuracy were other prominent dependent variables used. Only one study used NASA TLX (Wang and Dunston, 2011).

4.1.1. Representative Paper
As an example of the type of collaborative AR experiments conducted, we discuss the paper of Henrysson et al. (2005a) in more detail. They developed an AR-based face-to-face collaboration tool using a mobile phone and reported on two user studies. This paper received an ACC of 22.9, which is the highest in this category of papers. In the first study, six pairs of participants played a table-top tennis game in three conditions—face to face AR, face to face non-AR, and non-face to face collaboration. In the second experiment, the authors added (and varied) audio and haptic feedback to the games and only evaluated face to face AR. The same six pairs were recruited for this study as well. Authors collected both quantitative and qualitative (survey and interview) data, although they focused more on the latter. They asked questions regarding the usability of system and asked participants to rank the conditions. They explored several usability issues and provided design guidelines for developing face to face collaborative AR applications using handheld displays. For example, designing applications that have a focus on a single shared work space.

4.1.2. Discussion
The work done in this category is mostly directed toward remote collaboration. With the advent of modern head mounted devices such the Microsoft HoloLens, new types of collaborations can be created, including opportunities for enhanced face to face collaboration. Work needs to be done toward making AR-based remote collaboration akin to the real world with not only shared understanding of the task but also shared understanding of the other collaborators emotional and physiological states. New gesture-based and gaze-based interactions and collaboration across multiple platforms (e.g., between AR and virtual reality users) are novel future research directions in this area.

4.2. Education
Fifty-five studies were reported in 42 papers in the Education application area (Table 6). As expected, all studies reported some kind of teaching and learning applications, with a few niche areas, such as music training, educational games, and teaching body movements. Out of 55 studies, 24 used handheld displays, 8 used HMDs, 16 used some form of desktop displays, and 11 used spatial or large-scale displays. One study had augmented only sound feedback and used a head-mounted speaker (Hatala and Wakkary, 2005). Again, a trend of using handheld displays is prominent in this application area as well. Among all the
### TABLE 6 | Summary of user studies in Education application area.

| References | Topic | Data type | Displays used | Dependent measures | Study type | Participants (female) |
|------------|-------|-----------|---------------|---------------------|------------|-----------------------|
| Anderson and Bischof, 2014 | Training, Learning | O + S | DT | IMI Score, Isolation, Muscle control | Formal | 12 (6) |
| Arvanitis et al., 2009 | Learning | O + S | HMD | Rating, Physiological measures, sense of wellbeing | Formal | 5 (2) |
| Asai et al., 2005 | AR Instructions/Annotations | S | HMD, HH | Rating | Formal | 22 (15) |
| Cai et al., 2013 | Education | O + S | S/LS | Rating, Exam questions correct | Formal | 50 (30) |
| Cai et al., 2014 | Learning and Teaching | O + S | DT | Error/Accuracy, Rating | Formal | 29 (13) |
| Chang et al., 2013 | Training | O + S | S/LS | Error/Accuracy, Rating | Formal | 3 (1) |
| Chiang et al., 2014 | Education | O | HH | Learning outcomes | Field | 57 (NA) |
| Cocciolo and Rabina, 2013 | Learning, Tourism | S | HH | Interview response | Field | 34 (74) |
| Dünser et al., 2012b | Education | O | HH | Error/Accuracy | Pilot | 10 (10) |
| Fonseca et al., 2014b | Education | O + S | HH | Rating, Learning test scores | Formal | 48 (18) |
| Fonseca et al., 2014a | Education | S | HH, DT | Rating | Field | 57 (29) |
| Freitas and Campos, 2008 | Educational AR Game for 2nd Graders (7-8 years old) | O + S | S/LS, DT | Error/Accuracy | Qualitative Observation | Field | 54 (32) |
| Furió et al., 2013 | Education | S | Unspecified | Rating, Multi-choice question responses | Pilot | 117 (74) |
| Gama et al., 2012 | AR for movement training Museum guide | S | DT | Head-mounted speakers | Rating | Pilot | 10 (NA) |
| Hallata and Wakkary, 2005 | AR for movement training Museum guide | O + S | DT | Head-mounted speakers | Rating, Subjective response | Field | 8 (4) |
| Hou and Wang, 2013 | Training | O | DT | Time, Error/Accuracy | Formal | 28 (14) |
| Hsiao et al., 2012 | Education | O + S | S/LS, DT | Error/Accuracy, Rating | Formal | 884 (489) |
| Hunter et al., 2010 | Educational AR | O | HH | Rating | Field | 9 (NA) |
| Ibañez et al., 2014 | Education | O + S | HH | Rating, knowledge learnt - pre and post test | Formal | 60 (15) |
| Iwata et al., 2011 | Self-learning, Gaming | S | DT | Rating | Formal | 18 (1) |
| Juan et al., 2011b | AR handheld gaming, Educational gaming | S | HH | Rating | Formal | 38 (14) |
| Juan et al., 2011a | AR educational game | S | DT | Rating, knowledge of animals | Formal | 31 (14) |
| Kurt, 2010 | Learning | O | HH | Rating, field notes, observations | Field | 55 (NA) |
| Li et al., 2011 | AR for education | O + S | HH | Error/Accuracy, Rating | Formal | 36 (20) |
| Liarokapis, 2005 | Music Education | S | HMZ, DT | Ease of use/usability | Pilot | 9 (NA) |
| Lin et al., 2013 | Education | O | HH | Correct answers in physics tests | Formal | 40 (25) |
| Luckin and Fraser, 2011 | Learning and Teaching | O + S | DT | Rating, Items remembered, Engagement | Field | 304 (NA) |
| Martin-Gutierrez, 2011 | AR use in the classroom Interactive learning systems | S | DT | Rating | Formal | 47 (NA) |
| Oh and Byun, 2012 | Interactive learning systems | S | HH | Rating | Pilot | 15 (6) |
| Salvador-Herranz et al., 2013 | Education | S | S/LS | Rating | Pilot | 21 (9) |
| Santos et al., 2013 | AR X-Ray techniques for education | S | HH | Rating | Pilot | 27.3 (15.6) |
| Schwerdtfeger and Klinker, 2008 | AR enabled instructions | O + S | HMD | Time, Error/Accuracy, Subjective feedback | Formal | 24 (10) |
| Shatte et al., 2014 | Library management Education | O | HH | Time | Formal | 35 (NA) |
| Sommerauer and Müller, 2014 | Education | O | HH | Error/Accuracy, Rating | Field | 101 (39) |

(Continued)
TABLE 6 | Continued

| References               | Topic                | Data type | Displays used | Dependent measures                  | Study type | Participants (female) |
|--------------------------|----------------------|-----------|---------------|-------------------------------------|------------|-----------------------|
| Sumadio and Rambli, 2010 | Education            | O + S     | DT            | Rating                              | Formal     | 33 (20)               |
| Szymczak et al., 2012    | Multi-sensory AR     | S         | HH            | Rating                              | Field      | 17 (10.5)             |
| Toyama et al., 2013      | Reading Assistance   | O         | HMD           | Error/Accuracy                      | Pilot      | 12 (NA)               |
| Weing et al., 2013       | Music Education      | S         | S/LS          | Interview questions                 | Pilot      | 4 (0)                 |
| Wojciechowski and Cellary, 2013 | Education          | S         | DT            | Rating                              | Formal     | 42 (NA)               |
| Yamabe and Nakajima, 2013 | Training             | S         | S/LS, DT      | Rating                              | Field      | 10 (1.5)              |
| Zhang et al., 2014       | Teaching             | O + S     | HH            | Error/Accuracy, flow experience      | Field      | 147 (64)              |

S, Subjective; O, Objective; DT, Desktop; HH, handheld. Participant numbers are absolute values and where more than one studies were reported in the paper we used average counts.

Studies reported, 13 were pilot studies, 14 field studies, and 28 controlled lab-based experiments. Thirty-one studies were designed as within-subjects studies, and 16 as between-subjects. Six studies had only one condition tested. The median number of participants was 28, jointly highest among all application areas. Almost 43% of participants were females. Forty-nine studies were performed in indoor locations, four in outdoor locations, and two studies were performed in both locations. Twenty-five studies collected only subjective data, 10 objective data, and 20 studies collected both types of data. While subjective rating was the primary dependent measure used in most of the studies, some specific measures were also noticed, such as pre- and post-test scores, number of items remembered, and engagement. From the keywords used in the papers, it appears that learning was the most common keyword and interactivity, users, and environments also received noticeable importance from the authors.

4.2.1. Representative Paper

The paper from Fonseca et al. (2014a) received the highest ACC (22) in the Education application area of AR. They developed a mobile phone-based AR teaching tool for 3D model visualization and architectural projects for classroom learning. They recruited a total of 57 students (29 females) in this study and collected qualitative data through questionnaires and quantitative data through pre- and post-tests. This data was collected over several months of instruction. The primary dependent variable was the academic performance improvement of the students. Authors used five-point Likert-scale questions as the primary instrument. They reported that using the AR tool in the classroom was correlated with increased motivation and academic achievement. This type of longitudinal study is not common in the AR literature, but is helpful in measuring the actual real-world impact of any application or intervention.

4.2.2. Discussion

The papers in this category covered a diverse range of education and training application areas. There are some papers used AR to teach physically or cognitively impaired patients, while a couple more promoted physical activity. This set of papers focused on both objective and subjective outcomes. For example, Anderson and Bischof (2014) reported a system called ARM trainer to train amputees in the use of myoelectric prostheses that provided an improved user experience over the current standard of care. In a similar work, Gama et al. (2012) presented a pilot study for upper body motor movements where users were taught to move body parts in accordance to the instructions of an expert such as physiotherapist and showed that AR-based system was preferred by the participants. Their system can be applied to teach other kinds of upper body movements beyond just rehabilitation purposes. In another paper, Chang et al. (2013) reported a study where AR helped cognitively impaired people to gain vocational job skills and the gained skills were maintained even after the intervention. Hsiao et al. (2012) and Hsiao (2010) presented a couple of studies where physical activity was included in the learning experience to promote “learning while exercising”. There are few other papers that gamified the AR learning content and they primarily focused on subjective data. Iwata et al. (2011) presented ARGo an AR version of the GO game to investigate and promote self-learning. Juan et al. (2011b) developed ARGreenet game to create awareness for recycling. Three papers investigated education content themed around tourism and mainly focused on subjective opinion. For example, Hatala and Wakkary (2005) created a museum guide educating users about the objects in the museum and Szymczak et al. (2012) created multi-sensory application for teaching about the historic sites in a city. There were several other papers that proposed and evaluated different pedagogical approaches using AR including two papers that specifically designed for teaching music such as Liarokapis (2005) and Weing et al. (2013). Overall these papers show that in the education space a variety of evaluation methods can be used, focusing both on educational outcomes and application usability. Integrating methods of intelligent tutoring systems (Anderson et al., 1985) with AR could provide effective tools for education. Another interesting area to explore further is making these educational interfaces adaptive to the users cognitive load.

4.3. Entertainment and Gaming

We reviewed a total of 14 papers in the Entertainment and Gaming area with 18 studies were reported in these papers (Table 7). A majority of the papers reported a gaming application while fewer papers reported about other forms of entertainment
TABLE 7 | Summary of user studies in Entertainment and Gaming application area.

| References                  | Topic                   | Data type | Displays used | Dependent measures                                      | Study type       | Participants (female) |
|-----------------------------|-------------------------|-----------|---------------|----------------------------------------------------------|------------------|-----------------------|
| Baudisch et al., 2013       | Gaming                  | S         | Audio interface | Rating                                                   | Formal           | 30 (7)                |
| Dow et al., 2007            | Entertainment           | O + S     | HMD           | Time, Observations, subject interviews                   | Formal           | 12 (6)                |
| Grubert et al., 2012        | Mobile gaming           | O + S     | HH            | Time, Rating, Fatigue                                    | Formal           | 16 (6)                |
| Haugstvedt and Krogstie, 2012 | Mobile AR for           | S         | HH            | Rating                                                   | Field            | 121 (60.5)            |
| Henze and Boll, 2010a       | Mobile music listening  | S         | HH            | Rating                                                   | Formal           | 15 (3.5)              |
| Kern et al., 2006           | Gaming                  | S         | DT            | Subjective opinion                                       | Formal           | 3 (2)                 |
| Mulloni et al., 2008        | Handheld AR Gaming      | S         | HH            | Rating                                                   | Field            | 12 (6)                |
| Oda and Feiner, 2009        | Handheld AR gaming      | O + S     | HH            | Rating, Distance                                         | Formal           | 18 (3)                |
| Schinke et al., 2010        | AR tourism information systems, outdoor AR | O         | HH            | Error/Accuracy                                           | Formal           | 26 (13)               |
| Vazquez-Alvarez et al., 2012 | Tourism, Navigation    | O + S     | Headphone     | Time, Rating, Distance covered, Walking speed, Times stopped | Field            | 8 (2)                 |
| Wither et al., 2010         | AR story-telling        | O + S     | HH            | Rating, Subjective feedback                              | Formal           | 16 (9)                |
| Xu et al., 2008             | AR Gaming, Collaboration| O + S     | AR GamePad (Gizmondo) | Rating                                               | Formal           | 18 (5)                |
| Xu et al., 2011             | Handheld AR for social tabletop games | O        | No displays used | Coding of recorded video                                 | Field            | 9 (NA)                |
| Zhou et al., 2007           | Gaming, Audio AR        | O + S     | HMD           | Time, Error/Accuracy, Rating                             | Formal           | 40 (13)               |

S, Subjective; O, Objective; DT, Desktop; HH, handheld. Participant numbers are absolute values and where more than one studies were reported in the paper we used average counts.

applications. Out of the 18 studies, nine were carried out using handheld displays and four studies used HMDs. One of the reported studies, interestingly, did not use any display (Xu et al., 2011). Again, the increasing use of handheld displays is expected as this kind of display provides greater mobility than HMDs. Five studies were conducted as field studies and the rest of the 13 studies were controlled lab-based experiments. Fourteen studies were designed as within-subjects and two were between-subjects. The median number of participants in these studies was 17. Roughly 41.5% of participants were females. Thirteen studies were performed in indoor areas, four were in outdoor locations, and one study was conducted in both locations. Eight studies collected only subjective data, another eight collected both subjective and objective data, and the remaining two collected only objective data. Subjective preference was the primary measure of interest. However, task completion time was also another important measure. In this area, error/accuracy was not found to be a measure in the studies used. In terms of the keywords used by the authors, besides games, mobile and handheld were other prominent keywords. These results highlight the utility of handheld displays for AR Entertainment and Gaming studies.

4.3.1. Representative Paper

Dow et al. (2007) presented a qualitative user study exploring the impact of immersive technologies on presence and engagement, using interactive drama, where players had to converse with characters and manipulate objects in the scene. This paper received the highest ACC (9.5) in this category of papers. They compared two versions of desktop 3D based interfaces with an immersive AR based interface in a lab-based environment. Participants communicated in the desktop versions using keyboards and voice. The AR version used a video see-through HMD. They recruited 12 participants (six females) in the within-subjects study, each of whom had to experience interactive dramas. This paper is unusual because user data was collected mostly from open-ended interviews and observation of participant behaviors, and not task performance or subjective questions. They reported that immersive AR caused an increased level of user Presence, however, higher presence did not always lead to more engagement.

4.3.2. Discussion

It is clear that advances in mobile connectivity, CPU and GPU processing capabilities, wearable form factors, tracking robustness, and accessibility to commercial-grade game creation tools is leading to more interest in AR for entertainment. There is significant evidence from both AR and VR research of the power of immersion to provide a deeper sense of presence, leading to new opportunities for enjoyment in Mixed Reality (a continuum encompassing both AR and VR Milgram et al., 1995) spaces. Natural user interaction will be key to sustaining the use of AR in entertainment, as users will shy away from long term use of technologies that induce fatigue. In this sense, wearable AR will probably be more attractive for entertainment AR applications. In these types of entertainment applications, new types of evaluation measures will need to be used, as shown by the work of Dow et al. (2007).
4.4. Industrial

There was a total of 30 papers reviewed that focused on Industrial applications, and together they reported 36 user studies. A majority of the studies reported maintenance and manufacturing/assembly related tasks (Table 8). Eleven studies used handheld displays, 21 used HMDs, four used spatial or large screen displays, and two used desktop displays. The prevalence of HMDs was expected as most of the applications in this area require use of both hands at times, and as such HMDs are more suitable as displays. Twenty-nine studies were executed in a formal lab-based environment and only six studies were executed in their natural setups. We believe performing more industrial AR studies in the natural environment will lead to more usable results, as controlled environments may not expose the users

| References | Topic | Data type | Displays used | Dependent measures | Study type | Participants (female) |
|------------|-------|-----------|---------------|--------------------|------------|-----------------------|
| Allen et al., 2011 | AR for architectural planning | O + S | HMD | Rating | Formal | 18 (7) |
| Bruno et al., 2010 | Industrial prototyping | O + S | HMD | Error/Accuracy | Formal | 30 (NA) |
| Bunnun et al., 2013 | Modeling | O + S | HH | Time, Error/Accuracy, Rating | Formal | 10 (5) |
| Fiorentino et al., 2013 | None | O + S | HMD | Time, Error/Accuracy, Rating | Formal | 14 (3) |
| Fiorentino et al., 2014 | Maintenance | O + S | S/LS | Time, Error/Accuracy, Rating | Formal | 14 (3) |
| Gavish et al., 2013 | Industrial maintenance and assembly | O + S | HH | Time, Error/Accuracy, Rating, unsolved errors | Formal | 40 (1) |
| Hakkarainen et al., 2008 | Object assembly | S | HH | Rating | Pilot | 8 (NA) |
| Hartl et al., 2013 | Document Verification | O + S | HH | Time, Error/Accuracy, Rating, NASA TLX, AttrakDiff UX survey | Formal | 17 (1) |
| Henderson and Feiner, 2009 | AR Maintenance | O + S | HMD, DT | Time, Rating | Field | 6 (0) |
| Henderson and Feiner, 2010 | Industrial AR | O + S | HMD | Time, Error/Accuracy, Rating | Formal | 15 (4) |
| Henderson S. and Feiner, 2011 | Maintenance and repair, Defense | O + S | HMD | Time, Error/Accuracy, Rating, Head movement, Supporting task focus | Formal | 6 (0) |
| Henderson S. J. and Feiner, 2011 | AR for Industrial Tasks, Assembly Tasks | O + S | HMD | Time, Error/Accuracy, Rating | Formal | 11.3 (2.3) |
| Irizarry et al., 2013 | Construction | O + S | HH | Time, Rating | Formal | 30 (9) |
| Lau et al., 2012 | Tangible UI, 3D modeling | O + S | HMD | Time, Error/Accuracy, Rating | Formal | 10 (4) |
| Magnusson et al., 2010 | Pointing in space | O | HH | Time, Error/Accuracy, Rating | Formal | 6 (3) |
| Markov-Vetter et al., 2012 | Flight | O + S | HMD | Time, Error/Accuracy, Rating | Formal | 6 (1) |
| Marner et al., 2013 | None | O + S | S/LS | Time, Error/Accuracy, Rating | Formal | 24 (6) |
| Olsson et al., 2009 | Design | S | HH | Rating | Formal | 23 (10) |
| Petersen and Stricker, 2009 | Industrial assembly | O + S | S/LS | Time | Formal | 15 (10) |
| Rauhala et al., 2006 | Humidity data visualization | O + S | HH | Error/Accuracy | Formal | 10 (3) |
| Reif and Günther, 2009 | Storage facility management | O + S | HMD | Time, Error/Accuracy, Rating | Formal | 16 (3) |
| Rosenthal et al., 2010 | AR for guiding manual tasks | O | S/LS | Time, Error/Accuracy | Formal | 30 (17) |
| Schall et al., 2013a | Surveying | S | HH | Rating | Field | 16 (4) |
| Schoenfelder and Schmalstieg, 2008 | Industrial building acceptance | O + S | HMD, HH | Error/Accuracy, Navigational activity | Formal | 36 (9) |
| Schwerdtfeger et al., 2009 | Stock Picking | O + S | HMD | Time, Error/Accuracy, Rating, NASA TLX | Formal | 13.5 (3.5) |
| Schwerdtfeger et al., 2011 | Order picking in a warehouse | O + S | HMD | Time, Error/Accuracy, Observation | Field | 22.3 (10) |
| Tumler et al., 2008 | Industrial AR, Order Picking Task | O | HMD | Rating, heart rate | Formal | 12 (0) |
| Vignais et al., 2013 | Ergonomics | O + S | HMD | Time, Rating, Articulation score | Formal | 12 (0) |
| Yeh et al., 2012 | Construction | O | Projected AR | Time, Error/Accuracy | Formal | 34 (7) |
| Yuan et al., 2008 | Assembly, manufacturing | O + S | HMD, DT | Time, Error/Accuracy | Formal | 14 (4) |

S, Subjective; O, Objective; DT, Desktop; HH, handheld. Participant numbers are absolute values and where more than one studies were reported in the paper we used average counts.
to the issues that they face in real-world setups. Twenty-eight studies were designed as within-subjects and six as between-subjects. One study was designed to collect exploratory feedback from a focus group (Olsson et al., 2009). The median number of participants used in these studies was 15 and roughly 23% of them were females. Thirty-two studies were performed in indoor locations and four in outdoor locations. Five studies were based on only subjective data, four on only objective data, and rest of the 27 collected both kinds of data. Use of NASA TLX was very common in this application area, which was expected given the nature of the tasks. Time and error/accuracy were other commonly used measurements along with subjective feedback. The keywords used by the authors to describe their papers highlight a strong interest in interaction, interfaces, and users. Guidance and maintenance are other prominent keywords that authors used.

4.4.1. Representative Paper
As an example of the papers written in this area, Henderson S. and Feiner (2011) published a work exploring AR documentation for maintenance and repair tasks in a military vehicle, which received the highest ACC (26.25) in the Industrial area. They used a video see-through HMD to implement the study application. In the within-subjects study, the authors recruited six male participants who were professional military mechanics and they performed the tasks in the field settings. They had to perform 18 different maintenance tasks using three conditions—AR, LCD, and HUD. Several quantitative and qualitative (questionnaire) data were collected. As dependent variables they used task completion time, task localization time, head movement, and errors. The AR condition resulted in faster locating tasks and fewer head-movements. Qualitatively, AR was also reported to be more intuitive and satisfying. This paper provides an outstanding example of how to collect both qualitative and quantitative measures in an industrial setting, and so get a better indication of the user experience.

4.4.2. Discussion
Majority of the work in this category focused on maintenance and assembly tasks, whereas a few investigated architecture and planning tasks. Another prominent line of work in this category is military applications. Some work also cover surveying and item selection (stock picking). It will be interesting to investigate non-verbal communication cues in collaborative industrial applications where people form multiple cultural background can easily work together. As most of the industrial tasks require specific training and working in a particular environment, we assert that there needs to be more studies that recruit participants from the real users and perform studies in the field when possible.

4.5. Interaction
There were 71 papers in the Interaction design area and 83 user studies reported in these papers (see Table 9). Interaction is a very general area in AR, and the topics covered by these papers were diverse. Forty studies used handheld displays, 33 used HMDs, eight used desktop displays, 12 used spatial or large-screen displays, and 10 studies used a combination of multiple display types. Seventy-one studies were conducted in a lab-based environment, five studies were field studies, and six were pilot studies. Jones et al. (2013) were the only authors to conduct a heuristic evaluation. The median number of participants used in these studies was 14, and approximately 32% of participants were females. Seventy-five studies were performed in indoor locations, seven in outdoor locations, and one study used both locations. Sixteen studies collected only subjective data, 14 collected only objective data, and 53 studies collected both types of data. Task completion time and error/accuracy were the most commonly used dependent variables. A few studies used the NASA TLX workload survey (Robertson et al., 2007; Henze and Boll, 2010b) and most of the studies used different forms of subjective ratings, such as ranking conditions and rating on a Likert scale. The keywords used by authors identify that the papers in general were focused on interaction, interface, user, mobile, and display devices.

4.5.1. Representative Paper
Boring et al. (2010) presented a user study for remote manipulation of content on distant displays using their system, which was named Touch Projector and was implemented on an iPhone 3G. This paper received the highest ACC (31) in the Interaction category of papers. They implemented multiple interaction methods on this application, e.g., manual zoom, automatic zoom, and freezing. The user study involved 12 volunteers (four females) and was designed as a within-subjects study. In the experiment, participants selected targets and dragged targets between displays using the different conditions. Both quantitative and qualitative data (informal feedback) were collected. The main dependent variables were task completion time, failed trials, and docking offset. They reported that participants achieved highest performance with automatic zooming and temporary image freezing. This is a typical study in the AR domain based within a controlled laboratory environment. As usual in interaction studies, a significant amount of the study was focused on user performance with different input conditions, and this paper shows the benefit of capturing different types of performance measures, not just task completion time.

4.5.2. Discussion
User interaction is a cross-cutting focus of research, and as such, does not fall neatly within an application category, but deeply influences user experience in all categories. The balance of expressiveness and efficiency is a core concept in general human-computer interaction, but is of even greater importance in AR interaction, because of the desire to interact while on the go, the danger of increased fatigue, and the need to interact seamlessly with both real and virtual content. Both qualitative and quantitative evaluations will continue to be important in assessing usability in AR applications, and we encourage researchers to continue with this approach. It is also important to capture as many different performance measures as possible from the interaction user study to fully understand how a user interacts with the system.
### TABLE 9 | Summary of user studies in Interaction application area.

| References        | Topic                              | Displays used | Dependent measures                                      | Study type | Participants (female) |
|-------------------|------------------------------------|---------------|--------------------------------------------------------|------------|-----------------------|
| Ajanki et al., 2011 | AR Optical See-Through Calibration | O + S HMD, HH | Error/Accuracy, Rating                                  | Formal     | 7.5 (1.5)             |
| Axholt et al., 2011 | Phone-based AR interaction methods | O + S HH      | Time, Error/Accuracy, Rating                           | Formal     | 11 (1)                |
| Bai et al., 2012   | Handheld AR                         | O + S HMD     | Time, Rating                                           | Formal     | 32 (16)               |
| Bai et al., 2013a  | Basic AR interaction methods        | O + S HMD     | Time, Error/Accuracy, Rating                           | Formal     | 5 (0)                 |
| Barošević et al., 2012 | Object selection               | O + S HMD     | Time, Error/Accuracy, Rating                           | Formal     | 12 (2)                |
| Benko and Feiner, 2007 | NA                                   | O + S S/LS    | Time, Error/Accuracy, Rating                           | Formal     | 11 (5)                |
| Boring et al., 2010 | Mobile AR                           | O + S HH      | Time, Error/Accuracy, Docking offset, subjective feedback | Formal     | 12 (4)                |
| Boring et al., 2011 | AR control of large displays      | S HH          | Discussions                                           | Field      | 15 (5)                |
| Choi and Kim, 2013 | Spatial AR                          | O HH, S/LS    | Time, Number of clicks                                 | Formal     | 13 (4)                |
| Chun and Höllerer, 2013 | Handheld AR                      | O + S HH      | Time, Error/Accuracy, Rating                           | Field      | 30 (23)               |
| Datcu and Lukosch, 2013 | Basic AR interaction             | O + S HMD     | Time, Error/Accuracy, Rating                           | Field      | 25 (9)                |
| Denning et al., 2014 | None                                | S HMD         | Interview analysis                                     | Field      | 31 (13)               |
| Derker et al., 2009 | None                               | O + S HMD     | Time, Error/Accuracy, Rating                           | Field      | 22 (11)               |
| Fröhlich et al., 2006 | Spatial information appliances | S HH          | Rating, discussion                                     | Field      | 12 (5)                |
| Grassel et al., 2012 | None                                | S HH          | User preference                                        | Pilot      | 7 (4)                 |
| Gupta et al., 2012 | NA                                 | O + S DT      | Time, Error/Accuracy, Interviews                        | Formal     | 16 (8)                |
| Hürst and Van Wezel, 2013 | Gesture-based interaction | O + S HH      | Time, Rating, Interview                                | Formal     | 21 (5)                |
| Ha and Woo, 2010   | Object manipulation for tangible UIs | O + S DT      | Time, Rating                                           | Formal     | 20 (5)                |
| Hennings and Feiner, 2008 | AR Affordances for user interaction. | O + S HMD     | Time, Error/Accuracy, Rating                           | Formal     | 15 (4)                |
| Henningsson et al., 2006b | Handheld AR Interaction          | O + S HH      | Time, Rating                                           | Formal     | 9 (2)                 |
| Henningsson et al., 2007 | Mobile AR                           | O + S HH      | Time, Rating                                           | Formal     | 12.5 (1.5)            |
| Henze and Boll, 2010b | Object manipulation              | O + S S/LS    | Time, NASA TLX                                         | Formal     | 12 (4)                |
| Hoang and Thomas, 2010 | Selection of objects around the user | O + S HMD    | Time, Error/Accuracy, Rating                           | Field      | 16 (2)                |
| Jo et al., 2011    | None                               | S S/LS        | Rating, Simulator sickness                             | Formal     | 12.5 (2)              |
| Kerr et al., 2011  | Outdoor wearable AR                | S HMD         | Rating                                                | Pilot      | 8 (NA)                |
| Ko et al., 2013    | Handheld AR                        | S HH          | Rating                                                | Field      | 20 (10)               |
| Kron and Schmidt, 2005 | Telepresence                       | S HMD         | Rating                                                | Field      | 20 (0)                |
| Langlotz et al., 2013 | Spatialized audio in AR          | S HH          | Rating                                                | Pilot      | 30 (8)                |
| Lee and Billingham, 2008 | Multimodal interaction technique   | O + S HH, DT  | Freq. of speech and gesture commands                   | Formal     | 12 (2)                |
| Lee et al., 2009   | None                               | O HMD         | Number of collisions with virtual wire                 | Formal     | 14 (5)                |
| Lee et al., 2010   | NA                                 | O HMD         | No. of collisions in path tracking task                | Formal     | 13.5 (4)              |
| Lee and Billingham, 2011 | Handheld outlining of AR objects     | O HH          | Time, Error/Accuracy                                   | Formal     | 8 (3)                 |
| Lee et al., 2013b  | Spatial Interaction                | S DT          | Rating                                                | Formal     | 10 (2)                |
| Lee et al., 2013c  | Multimodal (speech-gesture) interaction | O + S DT  | Time, Error/Accuracy, Rating                           | Formal     | 25 (3)                |
| Lehnen et al., 2012 | Interaction in Mobile AR            | O + S HH      | Time, perceived mental workload                       | Formal     | 17 (7)                |
| Leitinger et al., 2013 | None                               | O Optical-ST DT | Time                                             | Formal     | 10 (4)                |
| Looser et al., 2007 | Tabletop AR; Object Selection       | O + S HMD     | Time, Error/Accuracy, Rating                           | Formal     | 16 (1)                |
| Lv, 2013          | Mobile AR                          | S HH          | Rating                                                | Formal     | 15 (6)                |
| Maer et al., 2011  | None                               | O HMD         | Error/Accuracy                                        | Formal     | 24 (NA)               |
| Mossel et al., 2013b | None                                | O + S HH      | Time, Rating, No. of steps to do task                  | Formal     | 28 (12)               |
| Mossel et al., 2013a | 3D Interaction in AR                | O + S HH      | Time, Error/Accuracy, Rating                           | Formal     | 28 (12)               |
| Muloni et al., 2013 | AR tracking initialization/calibration | O + S HH      | Error/Accuracy, Rating                                 | Formal     | 7 (2)                 |
| Ofek et al., 2013  | None                               | O + S S/LS    | Time, Error/Accuracy, Number of word detection        | Formal     | 48 (26)               |

(Continued)
4.6. Medicine
One of the most promising areas for applying AR is in medical sciences. However, most of the medical-related AR papers were published in medical journals rather than the most common AR publication venues. As we considered all venues in our review, we were able to identify 43 medical papers reporting AR studies and they in total reported 54 user studies. The specific topics were diverse, including laparoscopic surgery, rehabilitation and recovery, phobia treatment, and other medical training. This application area was dominated by desktop displays (34 studies), while 16 studies used HMDs, and handheld displays were used in only one study. This is very much expected, as often in medical setups, a clear view is needed along with free hands without adding any physical load. As expected, all studies were performed in indoor locations. Thirty-six studies were within-subjects and 11 were between-subjects. The median number of participants was 13, and approximately only 14.2% of participants were females, which is considerably lower than the gender-ratio in the profession of medicine. Twenty-two studies collected only objective data, 19 collected only subjective data, and 13 studies collected both types of data. Besides time and accuracy, various domain-specific surveys and other instruments were used in these studies as shown in Table 10.

The keywords used by authors suggest that AR-based research was primarily used in training and simulation. Laparoscopy, rehabilitation, and phobia were topics of primary interest. One difference between the keywords used in medical science vs. other AR fields is the omission of the word user, which indicates that the interfaces designed for medical AR were primarily focused on achieving higher precision and not on user experience. This is understandable as the users are highly trained professionals who need to learn to use new complex interfaces. The precision of the interface is of utmost importance, as poor performance can be life threatening.

4.6.1. Representative Paper
Archip et al. (2007) reported on a study that used AR visualization for image-guided neurosurgery, which received the highest ACC (15.6) in this category of papers. Researchers recruited 11 patients (six females) with brain tumors who underwent surgery. Quantitative data about alignment accuracy was collected as a dependent variable. They found that using AR produced a significant improvement in alignment accuracy compared to the non-AR system already in use. An interesting aspect of the paper was that it focused purely on one user performance measure, alignment accuracy, and there was no qualitative data captured from users about how they felt about the system. This appears to be typical for many medical related AR papers.
### TABLE 10 | Summary of user studies in Medical application areas.

| References                  | Topic                      | Data type | Displays used | Dependent measures                                                                 | Study type | Participants (female) |
|-----------------------------|----------------------------|-----------|---------------|-------------------------------------------------------------------------------------|------------|-----------------------|
| Akinbiyi et al., 2006       | Surgery                    | O         | TV screen    | Time, Error/Accuracy, Number of broken structure, Applied force                     | Formal     | 9 (3)                 |
| Albrecht et al., 2013       | Medical Education          | O + S     | HH           | Rating, Number of correct answers to exam                                           | Formal     | 10 (4)                |
| Anderson et al., 2013       | Movement training          | O         | S/LS         | Error/Accuracy                                                                     | Formal     | 8 (2)                 |
| Archip et al., 2007         | Image-guided surgery       | O         | DT           | Error/Accuracy                                                                      | Field      | 8 (1)                 |
| Bai et al., 2013b           | Autism                     | O + S     | S/LS         | Time, Rating, video analysis of type of play                                         | Formal     | 12 (2)                |
| Bichemie et al., 2007       | Image-Guided Surgery       | O + S     | DT           | Time, Error/Accuracy, discussion, qualitative                                         | Formal     | 12 (2)                |
| Botden et al., 2007         | Surgical Simulation        | O + S     | DT           | Rating, surgical effectiveness measures                                             | Formal     | 90 (NA)               |
| Botden et al., 2008         | Laparoscopic surgery       | S         | DT           | Rating                                                                               | Formal     | 55 (6)                |
| Botella et al., 2005        | Psychology/Phobia          | S         | HMD          | Rating, user behavior                                                               | Formal     | 1 (1)                 |
| Botella et al., 2010        | Phobia therapy             | O + S     | HMD          | Rating                                                                               | Formal     | 6 (6)                 |
| Botella et al., 2011        | Phobia treatment           | O         | DT           | Rating, Anxiety Disorders Interview Schedule, Behavioral avoidance test, Fear of spiders questionnaire, Spider phobia beliefs questionnaire, Subjective units of discomfort scale | Field      | 1 (1)                 |
| Bretón-López et al., 2010  | Phobia                     | S         | HMD          | Rating                                                                               | Formal     | 6 (6)                 |
| Brinkman et al., 2012       | Laparoscopic surgical training | O     | DT           | Time                                                                                 | Formal     | 36 (NA)               |
| Chintamani et al., 2010     | Teleoperation              | O         | DT           | Time, Error/Accuracy, Path Distance: Deviation From Path: Distance From Receptacle: | Formal     | 13.5 (2.5)            |
| Dixon et al., 2011          | Image-guided surgical planning | O + S   | Laparoscope  | Error/Accuracy                                                                      | Formal     | 12 (NA)               |
| Espay et al., 2010          | Rehabilitation/training, Gait assistance for Parkinson's disease | O + S   | HMD          | Rating, gait performance                                                           | Field      | 13 (7)                |
| Fichtinger G. et al., 2005  | Medical                    | O         | DT           | Error/Accuracy                                                                      | Pilot      | NA (NA)               |
| Fichtinger G. D. et al., 2005 | Medical                     | O         | DT           | Error/Accuracy                                                                      | Pilot      | NA (NA)               |
| Grasso et al., 2013         | Medicine                   | O         | DT           | Time, Number of scans, Dose                                                          | Field      | 3 (NA)                |
| Horeman et al., 2012        | Laparoscopic Training      | O         | DT           | Time, Force applied                                                                 | Pilot      | 12 (NA)               |
| Horeman et al., 2014        | Surgical Training          | O         | DT           | Time, Error/Accuracy, Path length, motion volume                                    | Formal     | 25 (18)               |
| Jeon et al., 2012           | Medical Training           | O + S     | DT           | Time, Similarity score                                                              | Formal     | 12 (2)                |
| Juan and Prez, 2010         | Phobia Treatment           | S         | HMD          | Rating, SUS                                                                         | Formal     | 20 (4)                |
| Juan and Joele, 2011        | Phobia                     | S         | HMD          | Rating                                                                               | Formal     | 24 (6)                |
| King et al., 2010           | Medicine                   | O + S     | DT           | Fugl-Meyer Assessment, Wolf Motor Function Test, DASH questionnaire                 | Formal     | 4 (NA)                |
| Letblanc et al., 2010       | Medical Training           | O + S     | DT           | Rating                                                                               | Formal     | 34 (NA)               |
| Lee et al., 2013d           | Medical procedure training | O + S     | DT           | Error/Accuracy, Rating                                                              | Formal     | 40 (NA)               |
| Luo et al., 2005b           | Medical AR                 | O         | HMD          | Grasping force                                                                       | Pilot      | 1 (0)                 |
| Luo et al., 2005a           | Stroke rehabilitation      | O         | HMD          | Clinical measures related to hand grasping performance                               | Field      | 3 (0)                 |
| Markovic et al., 2014       | Artificial Limbs           | O         | HMD          | Time, Error/Accuracy, Rating                                                         | Formal     | 13 (NA)               |
| Nicolau et al., 2005        | Medicine                   | O         | DT           | Time, Error/Accuracy                                                                | Formal     | 2 (NA)                |
| Nilsson and Johansson, 2007 | Cognitive System Engineering | S         | HMD          | Rating                                                                               | Field      | 12 (NA)               |
| Regenbrecht et al., 2011    | Stroke recovery and rehabilitation | O + S   | DT           | Time, Error/Accuracy                                                                | Formal     | 64 (10)               |
| Regenbrecht et al., 2012    | Medical rehabilitation     | S         | DT           | Rating, Discussion                                                                  | Formal     | 36.2 (5.7)            |
| Regenbrecht et al., 2014    | AR for rehabilitation      | S         | DT           | Rating, Interview                                                                   | Formal     | 44 (8)                |
| Ritter et al., 2007         | Laparoscopic surgery       | O         | DT           | path length, smoothness                                                             | Formal     | 60 (NA)               |
| Teber et al., 2009          | Laparoscopic Surgery       | O         | HMD          | Time, Error/Accuracy                                                                | Field      | 1 (NA)                |
| Thomas et al., 2010         | Anatomical Education       | S         | DT           | Rating                                                                               | Formal     | 34 (21)               |

(Continued)
**4.6.2. Discussion**

AR medical applications are typically designed for highly trained medical practitioners, which are a specialist set of users compared to other types of user studies. The overwhelming focus is on improving user performance in medical tasks, and so most of the user studies are heavily performance focused. However, there is an opportunity to include more qualitative measures in medical AR studies, especially those that relate to user estimation of their physical and cognitive workload, such as the NASA TLX survey. In many cases medical AR interfaces are aiming to improve user performance in medical tasks compared to traditional medical systems. This means that comparative evaluations will need to be carried out and previous experience with the existing systems will need to be taken into account.

**4.7. Navigation and Driving**

A total of 24 papers reported 28 user studies in the Navigation and Driving application areas (see Table 11). A majority of the studies reported applications for car driving. However, there were also pedestrian navigation applications for both indoors and outdoors. Fifteen studies used handheld displays, five used HMDs, and two used heads-up displays (HUDs). Spatial or large-screen displays were used in four studies. Twenty-three of the studies were performed in controlled setups and the remaining five were executed in the field. Twenty-two studies were designed as within-subjects, three as between-subjects, and the remaining three were mixed-factors studies. Approximately 38% of participants were females in these studies, where the median number of participants used was 18. Seven studies were performed in an outdoor environment and the rest in indoor locations. This indicates an opportunity to design and test hybrid AR navigation applications that can be used in both indoor and outdoor locations. Seven studies collected only objective data, 18 studies collected a combination of both objective and subjective data, whereas only three studies were based only on subjective data. Task completion time and error/accuracy were the most commonly used dependent variables. Other domain specific variables used were headway variation (deviation from intended path), targets found, number of steps, etc.

Analysis of author-specified keywords suggests that mobile received a strong importance, which is also evident by the profuse use of handheld displays in these studies, since these applications are about mobility. Acceptance was one of the noticeable keywords, which indicates that the studies intended to investigate whether or not a navigation interface is acceptable by the users, given the fact that, in many cases, a navigational tool can affect the safety of the user.

**4.7.1. Representative Paper**

Morrison et al. (2009) published a paper reporting on a field study that compared a mobile augmented reality map (MapLens) and a 2D map in a between-subjects field study, which received the highest ACC (16.3) in this application area of our review. MapLens was implemented on a Nokia N95 mobile phone and use AR to show virtual points of interest overlaid on a real map. The experimental task was to play a location-based treasure hunt type game outdoors using either MapLens or a 2D map. Researchers collected both quantitative and qualitative (photos, videos, field notes, and questionnaires) data. A total of 37 participants (20 female) took part in the study. The authors found that the AR map created more collaborations between players, and argued that AR maps are more useful as a collaboration tool. This work is important, because it provides an outstanding example of an AR Field study evaluation, which is not very common in the AR domain. User testing in the field can uncover several usability issues that normal lab-based testing cannot identify, particularly in the Navigation application area. For example, Morrison et al. (2009) were able to identify the challenges for a person of using a handheld AR device while trying to maintain awareness of the world around themselves.

**4.7.2. Discussion**

Navigation is an area where AR technology could provide significant benefit, due to the ability to overlay virtual cues on the real world. This will be increasingly important as AR displays become more common in cars (e.g., windscreen heads up displays) and consumers begin to wear head mounted displays outdoors. Most navigation studies have related to vehicle driving, and so there is a significant opportunity for pedestrian navigation studies. However human movement is more complex and erratic.
than driving, so these types of studies will be more challenging. Navigation studies will need to take into consideration the user's spatial ability, how to convey depth cues, and methods for spatial information display. The current user studies show how important it is to conduct navigation studies outdoors in a realistic testing environment, and the need to capture a variety of qualitative and quantitative data.

### 4.8. Perception

Similar to Interaction, Perception is another general field of study within AR, and appears in 51 papers in our review. There were a total of 71 studies reported in these papers. The primary focus was on visual perception (see Table 12) such as perception of depth/distance, color, and text. A few studies also reported perception of touch (haptic feedback). AR X-ray vision was also a common interface reported in this area. Perception of egocentric distance received significant attention, while exocentric distance was studied less. Also, near-to-medium-field distance estimation was studied more than far-field distances. A comprehensive review of depth perception studies in AR can be found in Dey and Sandor (2014), which also reports similar facts about AR perceptual studies as found in this review.

Twenty-one studies used handheld displays, 34 studies used HMDs, and 9 studies used desktop displays. The Phantom haptic display was used by two studies where haptic feedback was studied. Sixty studies were performed as controlled lab-based experiments, and only three studies were performed in the field. Seven studies were pilot studies and there was one heuristic study (Veas et al., 2012). Fifty-three studies were within-subjects, 12 between-subjects, and six mixed-factors. Overall, the median number of participants used in these studies was 16, and 27.3% of participants were females. Fifty-two studies were performed in indoor locations, only 17 studies were executed outdoors, and two studies used both locations. This indicates that indoor visual perception is well studied whereas more work needed to be done in outdoor settings.

### Table 11

| References                     | Topic                              | Data type | Displays used | Dependent measures                                               | Study type | Participants (female) |
|-------------------------------|------------------------------------|-----------|---------------|------------------------------------------------------------------|------------|-----------------------|
| Arning et al., 2012           | AR navigation systems              | O + S     | HH, S/LS      | Rating, Navigation performance metrics                           | Formal     | 24 (19)               |
| Avery et al., 2008            | Outdoor AR navigation              | O + S     | HMD           | Time, Error/Accuracy, Rating                                      | Formal     | 34 (9)                |
| Choi et al., 2011             | Outdoor AR (Mobile)                | O + S     | HH            | Time, Rating, Clicks on screen, number targets found              | Formal     | 12 (1)                |
| Dünnser et al., 2012a         | Outdoor navigation                | O + S     | HH            | Time, Rating, Discussion; video coding                           | Formal     | 22 (11)              |
| Fröhlich et al., 2011         | In-car navigation using AR         | O + S     | "in-car" screen | Rating, See measures.                                               | Formal     | 31 (11)              |
| Gee et al., 2011              | Cooperative AR, Automatic map building | O + S   | HH            | Time, Error/Accuracy                                               | Formal     | 12 (2)                |
| Goldiez et al., 2007          | Navigation                        | O         | HMD           | Time, Percentage of maze covered                                  | Formal     | 120 (60)              |
| Ha et al., 2012               | Path editing using tangible user interfaces | O + S | HMD   | Time, Error/Accuracy, Rating                                      | Formal     | 16.5 (1)              |
| Heller et al., 2014           | Navigation                        | O         | HH            | Path, orientation                                                 | Formal     | 16.5 (2.5)           |
| Kjeldskov et al., 2013        | Mobile urban map-based info.       | O + S     | HH            | Rating, Interviews                                                | Field      | 58 (6)                |
| Möller et al., 2012           | Indoor navigation                 | S         | DT            | Rating                                                            | Formal     | 81 (39)               |
| Möller et al., 2014           | Navigation                        | O + S     | HH            | Time, Error/Accuracy, Rating                                      | Formal     | 12 (1)                |
| Morrison et al., 2009         | AR Map, Augmenting a paper map     | S         | HH            | Rating, Primarily an observational study                          | Field      | 37 (20)               |
| Moussa et al., 2012           | AR for driving analysis           | O + S     | HMD           | Error/Accuracy, steps                                             | Formal     | 44 (18)               |
| Mulloni et al., 2011a         | Indoor navigation                 | O         | HH            | Time, Error/Accuracy, steps                                       | Formal     | 10 (5)                |
| Mulloni et al., 2011b         | Navigation                        | O + S     | HH            | Time, Rating, where AR was used                                   | Field      | 9 (NA)                |
| Ng-Thow-Hing et al., 2013     | Automotive Augmented Reality      | O         | HUD on car windshield | Error/Accuracy, steps                                             | Formal     | 16 (8)                |
| Rohs et al., 2007             | Mobile maps on handheld display    | O + S     | HH            | Time, Error/Accuracy, Rating                                      | Formal     | 18 (10)               |
| Rohs et al., 2009a             | Map navigation                    | O         | HH            | Time, Error/Accuracy                                              | Formal     | 17 (12)               |
| Rusch et al., 2013            | Driving                            | O         | S/LS           | Time, Error/Accuracy                                              | Formal     | 27 (14)               |
| Schall et al., 2013b          | Driving                            | O         | Projection    | Time, Error/Accuracy, Response Rate, Time to collision, Headway variation | Formal | 20 (7) |
| Tönnis et al., 2005           | Driving                            | O + S     | S/LS           | Time, Error/Accuracy, Rating                                      | Formal     | 12 (2)                |
| Tönnis and Kinker, 2007       | Driving                            | O + S     | S/LS           | Time, Error/Accuracy, Rating                                      | Formal     | 24 (10)               |
| Tangmanee and Teeravarunyou, 2012 | Vehicle Navigation               | O + S     | S/LS           | Time, Subjective questions, number of eye fixations              | Formal     | 5 (2)                |

S, Subjective; O, Objective; DT, Desktop; HH, handheld. Participant numbers are absolute values and where more than one studies were reported in the paper we used average counts.
### TABLE 12 | Summary of user studies in Perception application area.

| References                     | Topic                     | Data type | Displays used | Dependent measures                                      | Study type  | Participants (female) |
|--------------------------------|---------------------------|-----------|---------------|----------------------------------------------------------|-------------|-----------------------|
| Blum et al., 2010              | General AR                | S         | HMD, DT       | Rating                                                   | Formal      | 18 (4)                |
| Dey et al., 2010               | X-ray vision              | O + S     | HH            | Time, Error/Accuracy, Rating, NASA TLX                   | Formal      | 20 (2)                |
| Dey et al., 2012               | X-ray vision              | O + S     | HH            | Error/Accuracy                                           | Formal      | 20 (NA)               |
| Gabbard et al., 2005           | Outdoor AR                | O         | HMD           | Time, Error/Accuracy                                     | Formal      | 18 (6)                |
| Gabbard et al., 2006           | Perception                | O         | HMD           | Time, Error/Accuracy                                     | Formal      | 18 (6)                |
| Gabbard et al., 2007           | Text legibility           | O         | HMD           | Time, Error/Accuracy                                     | Formal      | 24 (12)               |
| Gabbard and Swan II, 2008      | Outdoor AR                | O         | HMD           | Time, Error/Accuracy                                     | Formal      | 24 (12)               |
| Gandy et al., 2010             | AR Testbed Design         | O + S     | HMD           | Error/Accuracy, Physiological measures                   | Formal      | 20 (6)                |
| Grechkin et al., 2010          | Distance estimation       | O         | Ambient display | Distance walked                                          | Formal      | 53.5 (23.5)           |
| Gustafsson and Gyllenswärd, 2005 | None                      | S         | HH            | Interview questions                                      | Pilot       | 15 (4)                |
| Hincapié-Ramos et al., 2014    | None                      | S         | HH            | Interview questions                                      | Pilot       | 8 (2)                 |
| Iwai et al., 2013              | Spatial AR                | O + S     | S/LS          | Time, Error/Accuracy                                     | Formal      | 10 (1)                |
| Jankowski et al., 2010         | Text readability          | O + S     | DT            | Time, Error/Accuracy, Rating                             | Formal      | 20 (4)                |
| Jeon and Choi, 2011            | Haptic rendering of stiffness | S     | Phantom       | Psychophysical PSE                                       | Formal      | 12 (4)                |
| Jeon and Harders, 2012         | Haptic AR                 | S         | HMD, Phantom  | Rating                                                   | Pilot       | 6 (2)                 |
| Jones et al., 2008             | None                      | O         | HMD           | Error/Accuracy                                           | Formal      | NA (NA)               |
| Jones et al., 2011             | None                      | O         | HMD           | Error/Accuracy                                           | Formal      | 21.75 (NA)            |
| Kellner et al., 2012           | None                      | O         | HMD           | Time, Error/Accuracy                                     | Formal      | 14.5 (6)              |
| Kerber et al., 2013            | None                      | O         | HH            | Error/Accuracy                                           | Formal      | 12 (2)                |
| Kim, 2013                      | Context in handheld AR    | S         | HH, HH projectors | Rating                                                   | Field       | 20 (10)               |
| Knörlein et al., 2009          | None                      | O         | HMD           | Correct selection of strongest force                     | Formal      | 14 (7)                |
| Lee et al., 2012               | AR haptic perception      | O + S     | DT            | Rating, Perceived location                               | Formal      | 14 (5)                |
| Lee et al., 2013a              | NA                        | O + S     | HMD           | Time, Rating                                             | Formal      | 48 (28)               |
| Lindeman et al., 2007          | None                      | O         | AudioBone bone-conducting headset | Error/Accuracy, Frequency                            | Formal      | 24 (2)                |
| Liu et al., 2010               | Displays                  | O + S     | HMD           | Error, Rating                                            | Formal      | 10 (2)                |
| Liu et al., 2012               | Handheld AR               | O         | HH            | Time, Error/Accuracy                                     | Formal      | 16 (4)                |
| Livingston et al., 2005        | NA                        | O + S     | HMD           | Error/Accuracy                                           | Formal      | 8 (NA)                |
| Livingston, 2007               | Visual acuity in AR displays | O      | HMD, DT       | Time, Error/Accuracy                                     | Formal      | 5 (1)                 |
| Livingston and Ai, 2008        | Tracking error            | O + S     | HMD           | Time, Error/Accuracy, Rating                             | Formal      | 11 (1)                |
| Livingston et al., 2009c       | Basic visual perception   | O         | HMD           | Time, Error/Accuracy                                     | Formal      | 20 (5.5)              |
| Livingston et al., 2009b       | Basic perception in AR    | O         | HMD           | Time, Error/Accuracy                                     | Formal      | 11 (2)                |
| Livingston et al., 2009a       | Object depth perception   | S         | HMD           | Time, Error/Accuracy, Rating                             | Formal      | 12 (4)                |
| Livingston et al., 2011        | Military situation awareness | O      | HMD           | Error/Accuracy                                           | Formal      | 14 (3)                |
| Lu et al., 2012                | Visual search             | O         | DT            | Time, Error/Accuracy                                     | Formal      | 20.5 (7)              |
| Mercier-Ganady et al., 2014    | None                      | O + S     | S/LS          | Rating, BCI output                                       | Formal      | 12 (NA)               |
| Olsson et al., 2012            | Mobile AR                 | S         | None          | Rating                                                   | Formal      | 262 (133)             |
| Peterson et al., 2009          | None                      | O + S     | Projection HUD | Time, Error/Accuracy, Rating                           | Formal      | 16 (NA)               |
| Puchar et al., 2014            | None                      | O + S     | HH            | Time, Error/Accuracy, Subject preference                 | Formal      | 15 (4)                |
| Salamin et al., 2006           | Unspecified               | S         | HMD           | Rating, Able to perform tasks                            | Pilot       | 6 (0)                 |
| Sandor et al., 2010            | X-Ray Vision              | O + S     | HH            | Time, Rating                                             | Formal      | 21.5 (1)              |
| Singh et al., 2010             | NA                        | O         | HMD           | Error/Accuracy, Distance to object                       | Formal      | 18 (7)                |
| Singh et al., 2012             | Depth Perception          | O         | HMD           | Error/Accuracy                                           | Formal      | 40 (NA)               |
| Suzuki et al., 2013            | None                      | O + S     | HMD           | Rating, Cardio-visual and tactile-visual feedback module | Formal      | 21 (11)               |

(Continued)
is needed to investigate outdoor visual perception. Outdoor locations present additional challenges for visualizations such as brightness, screen-glare, and tracking (when mobile). This is an area to focus on as a research community. Thirty-two studies were based on only objective data, 14 used only subjective data, and 25 studies collected both kinds of data. Time and error/accuracy were most commonly used dependent measures along with subjective feedback.

Keywords used by authors indicate an emphasis on depth and visual perception, which is expected, as most of the AR interfaces augment the visual sense. Other prominent keywords were X-ray and see-through, which are the areas that have received a significant amount of attention from the community over the last decade.

4.8.1. Representative Paper

A recent paper by Suzuki et al. (2013), reporting on the interaction of exteroceptive and interoceptive signals in virtual cardiac rubber hand perception, received the highest ACC (13.5) in this category of papers. The authors reported on a lab-based within-subjects user study using 21 participants (11 female) who wore a head-mounted display and experienced a tactile feedback simulating cardiac sensation. Both quantitative and qualitative (survey) data were collected. The main dependent variables were proprioceptive drift and virtual hand ownership. Authors reported that ownership of the virtual hand was significantly higher when tactile sensation was presented synchronously with the heart-beat of the participant than when provided asynchronously. This shows the benefit of combing perceptual cues to improve the user experience.

4.8.2. Discussion

A key focus of AR is trying to create a perceptual illusion that the AR content is seamlessly part of the user’s real world. In order to measure how well this is occurring it is important to conduct perceptual user studies. Most studies to date have focused on visual perception, but there is a significant opportunity to conduct studies on non-visual cues, such as audio and haptic perception. One of the challenges of such studies is being able to measure the users perception of an AR cue, and also their confidence in how well they can perceive the cue. For example, asking users to estimate the distance on an AR object from them, and how sure they are about that estimation. New experimental methods may need to be developed to do this well.

4.9. Tourism and Exploration

Tourism is one of the relatively less explored areas of AR user studies, represented by only eight papers in our review (Table 13). A total of nine studies were reported, and the primary focus of the papers was on museum-based applications (five papers). Three studies used handheld displays, three used large-screen or spatial displays, and the rest head mounted displays. Six studies were conducted in the field, in the environment where the applications were meant to be used, and only three studies were performed in lab-based controlled environments. Six studies were designed to be within-subjects. This area of studies used a markedly higher number of participants compared to other areas, with the median number of participants being 28, with approximately 38% of them female. All studies were performed in indoor locations. While we are aware of studies in this area that have been performed in outdoor locations, these did not meet the inclusion criteria of our review. Seven studies were based completely on subjective data and two others used both subjective and objective data. As the nature of the interfaces were primarily personal experiences, the over reliance on subjective data is understandable. An analysis of keywords in the papers found that the focus was on museums. User was the most prominent keyword among all, which is very much expected for an interface technology such as AR.

4.9.1. Representative Paper

The highest ACC (19) in this application area was received by an article published by Olsson et al. (2013) about the expectations of user experience of mobile augmented reality (MAR) services in a shopping context. Authors used semi-structured interviews as their research methodology and conducted 16 interview sessions with 28 participants (16 female) in two different shopping centers. Hence, their collected data was purely qualitative. The interviews were conducted individually, in pairs, and in groups. The authors reported on: (1) the characteristics of the expected user experience and, (2) central user requirements related to MAR in a shopping context. Users expected the MAR systems to be playful, inspiring, lively, collective, and surprising, along with providing context-aware and awareness-increasing services.
This type of exploratory study is not common in the AR domain. However, it is a good example of how qualitative data can be used to identify user expectations and conceptualize user-centered AR applications. It is also an interesting study because people were asked what they expected of a mobile AR service, without actually seeing or trying the service out.

4.9.2. Discussion
One of the big advantages of studies done in this area is the relatively large sample sizes, as well as the common use of “in the wild” studies, that assess users outside of controlled environments. For these reasons, we see this application area as useful for exploring applied user interface designs, using real end-users in real environments. We also think that this category will continue to be attractive for applications that use handheld devices, as opposed to head-worn AR devices, since these are so common, and get out of the way of the content when someone wants to enjoy the physically beautiful/important works.

5. CONCLUSION
5.1. Overall Summary
In this paper, we reported on 10 years of user studies published in AR papers. We reviewed papers from a wide range of journals and conferences as indexed by Scopus, which included 291 papers and 369 individual studies. Overall, on average, the number of user study papers among all AR papers published was less than 10% over the 10-year period we reviewed. Our exploration shows that although there has been an increase in the number of studies, the relative percentage appears the same. In addition, since 2011 there has been a shift toward more studies using handheld displays. Most studies were formal user studies, with little field testing and even fewer heuristic evaluations. Over the years there was an increase in AR user studies of educational applications, but there were few collaborative user studies. The use of pilot studies was also less than expected. The most popular data collection method involved filling out questionnaires, which led to subjective ratings being the most widely used dependent measure.

5.2. Findings and Suggestions
This analysis suggests opportunities for increased user studies in collaboration, more use of field studies, and a wider range of evaluation methods. We also find that participant populations are dominated by mostly young, educated, male participants, which suggests the field could benefit by incorporating a more diverse selection of participants. On a similar note, except for the Education and Tourism application categories, the median number of participants used in AR studies was between 12 and 18, which appears to be low compared to other fields of human-subject research. We have also noticed that within-subjects designs are dominant in AR, and these require fewer participants to achieve adequate statistical power. This is in contrast to general research in Psychology, where between-subject designs dominate.

Although formal, lab-based experiments dominated overall, the Education and Tourism application areas had higher ratios of field studies to formal lab-based studies, which required more participants. Researchers working in other application areas of AR could take inspiration from Education and Tourism papers and seek to perform more studies in real-world usage scenarios.

Similarly, because the social and environmental impact of outdoor locations differ from indoor locations, results obtained from indoor studies cannot be directly generalized to outdoor environments. Therefore, more user studies conducted outdoors are needed, especially ethnographic observational studies that report on how people naturally use AR applications. Finally, out of our initial 615 papers, 219 papers (35%) did not report either participant demographics, study design, or experimental task, and so could not be included in our survey. Any user study without these details is hard to replicate, and the results cannot be accurately generalized. This suggests a general need to improve the reporting quality of user studies, and education of researchers in the field on how to conduct good AR user studies.

5.3. Final Thoughts and Future Plans
For this survey, our goal has been to provide a comprehensive account of the AR user studies performed over the last decade.
We hope that researchers and practitioners in a particular application area can use the respective summaries when planning their own research agendas. In the future, we plan to explore each individual application area in more depth, and create more detailed and focused reviews. We would also like to create a publicly-accessible, open database containing AR user study papers, where new papers can be added and accessed to inform and plan future research.

REFERENCES

Ajani, A., Billinghurst, M., Gamper, H., Järvenpää, T., Kandemir, M., Kaski, S., et al. (2011). An augmented reality interface to contextual information. *Virt. Real. 15*, 161–173. doi: 10.1007/s10055-010-0183-5

Akinbiyi, T., Reiley, C. E., Saha, S., Burschka, D., Haser, C. J., Yuh, D. D., et al. (2006). "Dynamic augmented reality for sensory substitution in robot-assisted surgical systems," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 567–570.

Albrecht, U.-V., Folta-Schoofs, K., Behrends, M., and Von Jan, U. (2013). Effects of mobile augmented reality learning compared to textbook learning on medical students: randomized controlled pilot study. *J. Med. Int. Res. 15*, doi: 10.2196/jmir.2497

Allen, M., Regenbrecht, H., and Abbott, M. (2011). “Smart-phone augmented reality for public participation in urban planning,” in *Proceedings of the 23rd Australian Computer-Human Interaction Conference, OzCHI 2011*, 11–20.

Almeida, I., Oikawa, M., Carres, J., Miyazaki, K., Kato, H., and Billinghurst, M. (2012). “AR-based video-mediated communication: a social presence enhancing experience,” in *Proceedings - 2012 14th Symposium on Virtual and Augmented Reality, SVR 2012*, 125–130.

Alvarez-Santos, V., Iglesias, R., Pardo, X., Regueiro, C., and Canedo-Rodriguez, A. (2014). Gesture-based interaction with voice feedback for a tour-guide robot. *J. Vis. Commun. Image Represent. 25*, 499–509. doi: 10.1016/j.jvci.2013.03.017

Anderson, F., and Bischof, W. F. (2014). Augmented reality improves myoelectric prostheses training. *Int. J. Disabil. Hum. Dev. 13*, 349–354. doi: 10.1515/ijdh-2014-0327

Anderson, J. R., Boyle, C. F., and Reiser, B. J. (1985). Intelligent tutoring systems. *Science 228*, 456–462.

Anderson, F., Grossman, T., Matejka, J., and Fitzmaurice, G. (2013). "YouMove: enhancing movement training with an augmented reality mirror," in *UIST 2013 - Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 311–320.

Archip, N., Clatz, O., Whalen, S., Kacher, D., Fedorov, A., Kot, A., et al. (2007). Non-rigid alignment of pre-operative MRI, SNeMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. *Neuroimage 35*, 509–514. doi: 10.1016/j.neuroimage.2007.04.007

Arun, K., Zollei, M., Li, M., and Kobelt, L. (2012). ”Insights into user experiences and acceptance of mobile indoor navigation devices,” in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, MUM 2012*, 312–320.

Aryanitis, T., Petrou, A., Knight, J., Savas, S., Grifka, S., Gargalakos, M., et al. (2009). Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities. *Pers. Ubiquit. Comput. 13*, 243–250. doi: 10.1007/s00779-007-0187-7

Asai, K., Kobayashi, H., and Kondo, T. (2005). “Augmented instructions - A fusion of augmented reality and printed learning materials,” in *Proceedings - 5th IEEE International Conference on Advanced Learning Technologies, ICALT 2005*, Vol. 2005, 213–215.

Asai, K., Sugimoto, Y., and Billinghurst, M. (2010). “Exhibition of lunar surface navigation system facilitating collaboration between children and parents in Science Museum,” in *Proceedings - VRCAI 2010, ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Application to Industry*, 119–124.

Avery, B., Thomas, B. H., and Piekarski, W. (2008). “User evaluation of see-through vision for mobile outdoor augmented reality,” in *Proceedings - 7th IEEE International Symposium on Mixed and Augmented Reality 2008*, ISMAR 2008, 69–72.

AUTHOR CONTRIBUTIONS

All authors contributed significantly to the whole review process and the manuscript. AD initiated the process with Scopus database search, initial data collection, and analysis. AD, MB, RL, and JS all reviewed and collected data for an equal number of papers. All authors contributed almost equally to writing the paper, where AD and MB took the lead.

Axholt, M., Cooper, M., Skoglund, M., Ellis, S., O’Connell, S., and Ynnerman, A. (2011). “Parameter estimation variance of the single point active alignment method in optical see-through head mounted display calibration,” in *Proceedings - IEEE Virtual Reality*, 27–34.

Azuma, R. T. (1997). A survey of augmented reality. *Presence 6*, 355–385.

Bai, Z., and Blackwell, A. F. (2012). Analytic review of usability evaluation in ISMAR. *Interact. Comput. 24*, 450–460. doi: 10.1016/j.intcom.2012.07.004

Bai, H., Lee, G. A., and Billinghurst, M. (2012). “Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces,” in *ACM International Conference Proceeding Series*, 126–131.

Bai, H., Gao, L., El-Sana, J. B. J., and Billinghurst, M. (2013a). “Markerless 3D gesture-based interaction for handheld augmented reality interfaces,” in *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, SA 2013.

Bai, Z., Blackwell, A. F., and Coulouris, G. (2013b). “Through the looking glass: pretend play for children with autism,” in *2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013*, 49–58.

Bai, H., Lee, G. A., and Billinghurst, M. (2014). “Using 3D hand gestures and touch input for wearable AR interaction,” in *Conference on Human Factors in Computing Systems - Proceedings*, 1321–1326.

Baldauf, M., Lasinger, K., and Fröhlich, P. (2012). “Private public screens - Detached multi-user interaction with large displays through mobile augmented reality,” in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, MUM 2012*.

Baricivèc, D., Lee, C., Turk, M., Höllerer, T., and Bowman, D. (2012). “A handheld AR magic lens with user-perspective rendering,” in *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality 2012*, Science and Technology Papers, 197–206.

Baudisch, P., Pohl, H., Reinicke, S., Wittmers, E., Lühne, P., Knust, M., et al. (2013). “Imaginary reality gaming: Ball games without a ball,” in *ISMAR 2013 - Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (St. Andrews, UK)*, 405–410.

Benko, H., and Feiner, S. (2007). “Balloon selection: a multi-finger technique for accurate low-fatigue 3D selection,” in *IEEE Symposium on 3D User Interfaces 2007 - Proceedings*, 3DUI 2007, 79–86.

Benko, H., Wilson, A., and Zannier, F. (2014). “Dyadic projected spatial augmented reality,” in *UIST 2014 - Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (Honolulu, HI)*, 645–656.

Bichlmair, C., Heininger, S., Rustace, M., and Navah, N. (2007). “Laparoscopic virtual mirror for understanding vessel structure: evaluation study by twelve surgeons,” in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR*.

Blum, T., Wieczorek, M., Aichert, A., Tibrewal, R., and Navah, N. (2010). “The effect of out-of-focus blur on visual discomfort when using stereo displays,” in *9th IEEE International Symposium on Mixed and Augmented Reality 2010: Science and Technology, ISMAR 2010 - Proceedings*, 13–17.

Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. (2010). “Touch projector: Mobile interaction through video,” in *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 4, (Atlanta, GA), 2287–2296.

Boring, S., Gehring, S., Wiethoff, A., Blöckner, M., Schöning, J., and Butz, A. (2011). “Multi-user interaction on media facades through live video on mobile devices,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2721–2724.

Botden, S., Buzinsk, S., Schijven, M., and Jakimowicz, J. (2007). Augmented versus virtual reality laparoscopic simulation: what is the difference? A comparison of the ProMIS augmented reality laparoscopic simulator versus...
LapSim virtual reality laparoscopic simulator. *World J. Surg.* 31, 764–772. doi: 10.1007/s00268-006-0724-y

Botden, S., Buzinsk, S., Schijven, M., and Jakimowicz, J. (2008). ProMIS augmented reality training of laparoscopic procedures face validity. *Surg. Endosc.* Other Intervent. Tech. 26, 2172–2178. doi: 10.1007/s00464-012-1849-9

Bruno, F., Cosco, F., Anglica, A., and Muzzuppapa, M. (2010). “Mixed prototyping for products usability evaluation,” in *Proceedings of the ASME Design Engineering Technical Conference*, Vol. 3, 1381–1390.

Bunnn, P., Subramanian, S., and Mayol-Cuevas, W. W. (2013). *In situ* interactive image-based model building for Augmented Reality from a handheld device. *Virt. Real.* 17, 137–146. doi: 10.1007/s11055-011-0206-x

Cai, S., Chang, F.-K., and Wang, X. (2013). Using the augmented reality 3D technique for a convex imaging experiment in a physics course. *Int. J. Eng. Educ.* 29, 856–865.

Cai, S., Wang, X., and Chang, F.-K. (2014). A case study of Augmented Reality simulation system application in a chemistry course. *Comput. Hum. Behav.* 31, 31–40. doi: 10.1016/j.chb.2014.04.018

Cavigliani, J., Forbuth, B., Anisetti, M., Ceravolo, P., Damiani, E., and Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multim. Tools Appl.* 51, 341–377. doi: 10.1007/s11042-010-0660-6

Chan, Y.-J., Kang, Y.-S., and Huang, P.-C. (2013). An augmented reality (AR)-based vocational task prompting system for people with cognitive impairments. *Res. Dev. Disabil.* 34, 3049–3056. doi: 10.1016/j.ridd.2013.06.026

Chastine, J., Nagel, K., Zhu, Y., and Yearsovich, L. (2007). “Understanding the design space of referencing in collaborative augmented reality environments,” in *Proceedings - Graphics Interface*, 207–214.

Chen, S., Chen, M., Kunz, A., Yantaç, A., Bergmark, M., Sundin, A., et al. (2013). “SiMarbeta: mobile sketch-gesture-video remote support for car drivers,” in *ACM International Conference Proceeding Series*, 69–76.

Chiang, T., Yang, S., and Hwang, G.-J. (2013). Students' online interactive patterns in augmented reality-based inquiry activities. *Comput. Educ.* 78, 97–108. doi: 10.1016/j.compedu.2014.05.006

Chintamani, K., Cao, A., Ellis, R., and Pandya, A. (2010). Improved telemanipulator navigation under display-control misalignments using augmented reality cues. *IEEE Trans. Syst. Man Cybern. A Syst. Humans* 40, 29–39. doi: 10.1109/TSMCA.2009.2030166

Choi, J., and Kim, G. J. (2013). Usability of one-handed interaction methods for handheld projection-based augmented reality. *Pers. Ubiquit. Comput.* 17, 399–409. doi: 10.1007/s11077-011-0502-1

Choi, J., Jang, B., and Kim, G. J. (2011). Organizing and presenting geospatial tags in location-based augmented reality. *Pers. Ubiquit. Comput.* 15, 641–647. doi: 10.1007/s11077-010-0343-3

Chun, W. H., and Höllerer, T. (2013). “Real-time hand interaction for augmented reality on mobile phones,” in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 307–314.

Coccioni, A., and Rabina, D. (2013). Does place affect user engagement and understanding?: mobile learner perceptions on the streets of New York. *J. Document.* 69, 98–120. doi: 10.1108/0022041131293342

Datcu, D., and Lukosch, S. (2013). “Free-hands interaction in augmented reality,” in *SUI 2013 - Proceedings of the ACM Symposium on Spatial User Interaction*, 33–40.

Denning, T., Dehlavi, Z., and Kohn, T. (2014). “In situ with bystanders of augmented reality glasses: perspectives on recording and privacy-mediating technologies,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2377–2386.

Dey, A., and Sander, C. (2014). Lessons learned: evaluating visualizations for occluded objects in handheld augmented reality. *Int. J. Hum. Comput. Stud.* 72, 704–716. doi: 10.1016/j.ijhcs.2014.04.001

Dey, A., Cunningham, A., and Sander, C. (2010). “Evaluating depth perception of photorealistic Mixed Reality visualizations for occluded objects in outdoor environments,” in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, 211–218.

Dey, A., Jarvis, G., Sander, C., and Reitmayr, G. (2012). “Tablet versus phone: depth perception in handheld augmented reality,” in *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality*, 2012, *Science and Technology Papers*, 187–196.

Dierker, A., Mertes, C., Hermann, T., Hanheide, M., and Sagerer, G. (2009). “Mediated attention with multimodal augmented reality,” in *ICMI-MLMI’09 - Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interfaces*, 245–252.

Dixon, B., Daly, M., Chan, H., Vescan, A., Witterick, I., and Irish, J. (2011). Augmented image guidance improves skull base navigation and reduces task workload in trainees: a preclinical trial. *Laryngoscope* 121, 2060–2064. doi: 10.1002/lary.22153

Dow, S., Mehta, M., Harin, E., MacIntyre, B., and Mateas, M. (2007). “Presence and engagement in an interactive drama,” in *Conference on Human Factors in Computing Systems - Proceedings (San Jose, CA)*, 1475–1484.

Dünser, A., Grasset, R., and Billingshurst, M. (2008). *A Survey of Evaluation Techniques Used in Augmented Reality Studies*. Technical Report.

Dünser, A., Billingshurst, M., Wen, J., Lehtinen, V., and Nurminen, A. (2012a). Exploring the use of handheld AR for outdoor navigation. *Comput. Graph.* 36, 1084–1095. doi: 10.1016/j.cag.2012.10.001

Dünser, A., Walker, L., Horner, H., and Bentall, D. (2012b). “Creating interactive physics education books with augmented reality,” in *Proceedings of the 24th Australian Computer-Human Interaction Conference*, 107–114.

Espay, A., Baram, Y., Dwivedi, A., Shukla, R., Gartner, M., Gaines, L., et al. (2010). At-home training with closed-loop augmented-reality cueing device for improving gait in patients with Parkinson disease. *J. Rehabil. Res. Dev.* 47, 573–582. doi: 10.1682/JRRD.2009.10.0165

Fichtinger, G., Deuguet, A., Masamune, K., Balogh, E., Fischer, G., Mathieu, H., et al. (2005). Image overlay guidance for needle insertion in CT scanner. *IEEE Trans. Biomed. Eng.* 52, 1415–1424. doi: 10.1109/TBME.2005.851493

Fichtinger, G. D., Deuguet, A., Fischer, G., Iordachita, I., Balogh, E. B., Masamune, K., et al. (2005). Image overlay for CT-guided needle insertions. *Comput. Aided Surg.* 10, 241–255. doi: 10.3109/10929080500230486

Fiorentino, M., Debardens, S., Uva, A. E., and Monno, G. (2013). Augmented reality text style readability with see-through head-mounted displays in industrial context. *Presence* 22, 171–190. doi: 10.1162/PRES_a_00146

Fiorentino, M., Uva, A. E., Gattullo, M., Debardens, S., and Monno, G. (2014). Augmented reality on large screen for interactive maintenance instructions. *Comput. Indust.* 65, 270–278. doi: 10.1016/j.compind.2013.11.004

Fonseca, D., Redondo, E., and Villagra, S. (2014b). “Mixed-methods research: a new approach to evaluating the motivation and satisfaction of university students using advanced visual technologies,” in *Universal Access in the Information Society*.

Fonseca, D., Marti, N., Redondo, E., Navarro, L., and Sánchez, A. (2014a). Relationship between student profile, tool use, participation, and academic performance with the use of Augmented Reality technology for visualized architecture models. *Comput. Hum. Behav.* 31, 434–445. doi: 10.1016/j.chb.2013.03.006

Freitas, R., and Campos, P. (2008). “SMART: a system of augmented reality for teaching 2nd grade students,” in *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, BCS HCI 2008*, Vol. 2, 27–30.

Fröhlich, P., Simon, R., Baillie, L., and Anegg, H. (2006). “Comparing conceptual designs for mobile access to geo-spatial information,” in *ACM International Conference Proceeding Series*, Vol. 159, 109–112.
Lee, M., and Billinghurst, M. (2008). “A Wizard of Oz Study for an AR Multimodal Interface,” in ICMIF’08: Proceedings of the 10th International Conference on Multimodal Interfaces, 249–256.

Lee, C., Bonebrake, S., Höllerer, T., and Bowman, D. (2009). “A replication study testing the validity of AR simulation in VR for controlled experiments,” in Science and Technology Proceedings - IEEE 2009 International Symposium on Mixed and Augmented Reality, ISMAR 2009, 203–204.

Lee, C., Bonebrake, S., Höllerer, T., and Bowman, D. (2010). “The role of latency in the validity of AR simulation,” in Proceedings - IEEE Virtual Reality, 11–18.

Lee, J., Kim, Y., and Kim, G. J. (2012). “Funneling and salutation effects for tactile interaction with virtual objectvirtual,” in Conference on Human Factors in Computing Systems - Proceedings, 3141–3148.

Lee, C., Rincon, G., Meyer, G., Höllerer, T., and Bowman, D. (2013a). The effects of visual realism on search tasks in mixed reality simulation. IEEE Trans. Visual. Comput. Graph. 19, 547–556. doi: 10.1109/TVCG.2013.41

Lee, J., Olwal, A., Ishii, H., and Boulanger, C. (2013b). “SpaceTop: integrating 2D and spatial 3D interactions in a see-through desktop environment,” in Conference on Human Factors in Computing Systems - Proceedings, 189–192.

Leithinger, D., Pollmer, S., Olwal, A., Luescher, S., Hogge, A., Lee, J., et al. (2013). “Sublimate: state-changing virtual and physical rendering to augment interaction with shape displays,” in Conference on Human Factors in Computing Systems - Proceedings, 1441–1450.

Long, M. A., Dey, A., Sandor, C., and Thomas, B. H. (2013). “Objective measures for the effectiveness of augmented reality displays,” in Proceedings - ISMAR 2013, 39–48.

Markovic, M., Dosen, S., Cipriani, C., Popovic, D., and Farina, D. (2014). “Pointing for non-visual orientation and navigation,” in NordiCHI 2010: Extending Boundaries - Proceedings of the 6th Nordic Conference on Human-Computer Interaction, 735–738.

Maier, P., Dey, A., Waechter, C. A. L., Sandor, C., Tšnnis, M., and Klinker, C. (2012). “Sublimate: state-changing virtual and physical rendering to augment interaction with shape displays,” in Conference on Human Factors in Computing Systems - Proceedings, 735–738.

Livingston, M., A., Ai, Z., Kline, T., Waldinger, H., and Kamper, D. (2005a). “Integration of augmented reality and assistive devices for post-stroke hand opening rehabilitation,” in Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, Vol. 7, 6855–6858.

Lucin, R., and Fraser, D. (2011). Limitless or pointless? An evaluation of augmented reality technology in the school and home. Int. J. Technol. Enhanced Learn. 3, 510–524. doi: 10.1504/IJTEL.2011.042102

Lu, J., Kline, T., Fischer, H., Stubblefield, K., Kenyon, R., and Kamper, D. (2005b). “An augmented reality training environment for post-stroke finger extension rehabilitation,” in Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics, Vol. 2005, 329–332.

LV, Z. (2013). “Wearable smartphone: wearable hybrid framework for hand and foot gesture interaction on smartphone,” in Proceedings of the IEEE International Conference on Computer Vision, 436–443.

Magguson, C., Molina, M., Rassmus-Grön, K., and Szmyczak, D. (2010). "Pointing for non-visual orientation and navigation," in NordiCHI 2010: Extending Boundaries - Proceedings of the 6th Nordic Conference on Human-Computer Interaction, 735–738.

Maier, P., Dey, A., Waechter, C. A. L., Sandor, C., Tšnnis, M., and Klinker, C. (2012). “Sublimate: state-changing virtual and physical rendering to augment interaction with shape displays,” in Conference on Human Factors in Computing Systems - Proceedings, 735–738.

Markovic-Vetter, D., Moll, E., and Staadt, O. (2012). “Evaluation of 3D selection tasks in parabolic flight conditions: pointing task in augmented reality user interfaces,” in Proceedings - VRCAI 2012: 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, 287–293.

Markovic, M., Dosen, S., Cipriani, C., Popovic, D., and Farina, D. (2014). Stereovision and augmented reality for closed-loop control of grasping in hand prostheses. J. Neural Eng. 11, 046001. doi: 10.1088/1741-2560/11/4/046001

Marner, M. R., Irfitt, A., and Thomas, B. H. (2013). “Improving procedural task performance with Augmented Reality annotations,” in 2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013, 39–48.

Marín-Gutierrez, J., et al. (2011). “Generic user manual for maintenance of mountain bike brakes based on augmented reality,” in Proceedings of the 28th International Symposium on Automation and Robotics in Construction, ISARC 2011, 1401–1404.
Mercier-Ganady, J., Lotte, F., Loub-Escande, E., Marchal, M., and Lecuyer, A. (2014). "The Mind-Mirror: see your brain in action in your head using EEG and augmented reality," in Proceedings - IEEE Virtual Reality, 33–38.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). "Augmented reality: a class of displays on the reality-virtuality continuum," in Telemanipulator and Telepresence Technologies, Vol. 2351 (International Society for Optics and Photonics), 282–293.

Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., et al. (2009). “Like bees around the hive: a comparative study of a mobile augmented reality map,” in Conference on Human Factors in Computing Systems - Proceedings (Boston, MA), 1889–1898.

Mossel, A., Venditti, B., and Kaufmann, H. (2013a). “3Touch and homero-s: intuitive manipulation techniques for one-handed handheld augmented reality,” in ACM International Conference Proceeding Series.

Mossel, A., Venditti, B., and Kaufmann, H. (2013b). “Drillsample: precise selection in dense handheld augmented reality environments,” in ACM International Conference Proceeding Series.

Moussa, G., Radwan, E., and Hussain, K. (2012). Augmented reality vehicle system: left-turn maneuver study. Transport. Res. C Emerging Technol. 21, 1–16. doi: 10.1016/j.trc.2011.08.005

Mulloni, A., Wagner, D., and Schmalstieg, D. (2008). “Mobility and social interaction as core gameplay elements in multi-player augmented reality,” in Proceedings - 3rd International Conference on Digital Interactive Media in Entertainment and Arts, DIMEA 2008, 472–478.

Mulloni, A., Seichter, H., and Schmalstieg, D. (2011a). “Handheld augmented reality indoor navigation with activity-based instructions,” in Mobile HCI 2011 - 13th International Conference on Human-Computer Interaction with Mobile Devices and Services, 211–220.

Mulloni, A., Seichter, H., and Schmalstieg, D. (2011b). “User experiences with augmented reality aided navigation on phones,” in 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, 229–230.

Mulloni, A., Ramachandran, M., Reitmayr, G., Wagner, D., Grasset, R., and Diaz, S. (2013). “User friendly SLAM initialization,” in 2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013, 153–162.

Møller, A., Kranz, M., Huitl, R., Diewald, S., and Roatler, L. (2012). “A mobile indoor navigation system interface adapted to vision-based localization,” in Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, MUM 2012.

Møller, A., Kranz, M., Diewald, S., Roatler, L., Huitl, R., Stockinger, T., et al. (2014). “Experimental evaluation of user interfaces for visual indoor navigation,” in Conference on Human Factors in Computing Systems - Proceedings, 3607–3616.

Ng-Thow-Hing, V., Bark, K., Beckwith, L., Tran, C., Bhandari, R., and Sridhar, S. (2012). “Experimental evaluation of mobile augmented reality scenarios,” in Journal of Ambient Intell. Smart Environ. 4, 29–47. doi: 10.3233/JAI-2011-0127

Olsson, T., Lagerström, E., and Väänänen-Vainio-Mattila, K. (2013). “User expectations of mobile augmented reality services: an initial user study,” in VTT Sym. (Viltion Teknillinen Tutkimuskeskus) (Helsinki), 177–184.

Olsson, T., Kärkkäinen, T., Lagerström, E., and Väänänen-Vainio-Mattila, K. (2013). User evaluation of mobile augmented reality services: a user study in the context of shopping centres. Pers. Ubiquit. Comput. 17, 287–304. doi: 10.1007/s00779-011-0494-x

Papagiannakis, G., Singh, G., and Magenbek-Thalmann, N. (2008). A survey of mobile and wireless technologies for augmented reality systems. Comput. Animat. Virt. Worlds 19, 3–22. doi: 10.1002/cav.191

Pescarin, S., Pagano, A., Wallergård, M., Hupperetz, W., and Ray, C. (2012). “Archeovirtual 2011: an evaluation approach to virtual museums,” in Proceedings of the 2012 18th International Conference on Virtual Systems and Multimedia, VSMM 2012: Virtual Systems in the Information Society, 25–32.

Petersen, N., and Stricker, D. (2009). “Continuous natural user interface: reducing the gap between real and digital world,” in Science and Technology Proceedings - IEEE 2009 International Symposium on Mixed and Augmented Reality, ISMAR 2009, 23–26.

Petersen, S., Axholt, M., Cooper, M., and Ellis, S. (2009). “Visual clutter management in augmented reality: effects of three label separation methods on spatial judgments,” in 3DUI - IEEE Symposium on 3D User Interfaces 2009 - Proceedings, 111–118.

Poelman, R., Akman, O., Lukosch, S., and Jonker, P. (2012). "As if being there: mediated reality for crime scene investigation," in Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, 1267–1276.

Porter, S. R., Marner, M. R., Smith, R. T., Zucco, J. E., and Thomas, B. H. (2010). “Validating spatial augmented reality for interactive rapid prototyping,” in 9th IEEE International Symposium on Mixed and Augmented Reality 2010: Science and Technology, ISMAR 2010 - Proceedings, 265–266.

Pucihar, K., Coulton, P., and Alexander, J. (2014). “The use of surrounding visual context in handheld AR-device vs. user perspective rendering,” in Conference on Human Factors in Computing Systems - Proceedings, 197–206.

Pusch, A., Martin, O., and Coquillart, S. (2008). "HEMP - Hand-displacement-based Pseudo-haptics: a study of a force field application," in 3DUI - IEEE Symposium on 3D User Interfaces 2008, 59–66.

Pusch, A., Martin, O., and Coquillart, S. (2009). "HEMP-hand-displacement-based pseudo-haptics: a study of a force field application and a behavioural analysis," in J. Hum. Comput. Stud. 67, 256–268. doi: 10.1017/jhcs.2008.09.015

Rankohi, S., and Waugh, L. (2013). Review and analysis of augmented reality literature for construction industry. Visual. Eng. 1, 1–18. doi: 10.1186/2213-7459-1-9

Raulaha, M., Gunnarsson, A.-S., Henrysson, A., and Ynnerman, A. (2006). "A navigation system: an initial user study," in Conference on Human Factors in Computing Systems - Proceedings, 159, 145–148.

Regenbrecht, H., McGregor, G., Ott, C., Hoermann, S., Schubert, T., Hale, L., et al. (2011). "Out of reach! - A novel AR interface approach for motor rehabilitation," in 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, 219–228.

Regenbrecht, H., Hoermann, S., McGregor, G., Dixon, B., Franz, E., Ott, C., et al. (2012). Visual manipulations for motor rehabilitation. Comput. Graph. (Pergamon) 36, 819–834.

Regenbrecht, H., Hoermann, S., Ott, C., Muller, L., and Franz, E. (2014). "Manipulating the experience of reality for rehabilitation applications," in Proc. IEEE 102, 170–184. doi: 10.1109/PROC.2013.2294178

Reif, R., and Güntherner, W. A. (2009). Pick-by-vision: augmented reality supported order picking. Vis. Comput. 25, 461–467. doi: 10.1007/s00371-009-0348-y

Ritter, E., Kindelan, T., Michael, C., Pimentel, E., and Bowyer, M. (2007). Concurrent validity of augmented reality metrics applied to the fundamentals of laparoscopic surgery (FLS). Surg. Endosc. Other Intervent. Techniq. 21, 1441–1445. doi: 10.1007/s00464-007-9261-5
