Study on Assessing User Experience of Augmented Reality Applications

Lei Wang[1](✉) and Meiyu Lv[1,2]

[1] School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing 100876, China
wl_bupt@163.com

[2] Beijing Key Laboratory of Network and Network Culture, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract. With the development of the augmented reality technology and the popularisation of smartphones, the application of augmented reality based on mobile devices demonstrates an optimistic development prospect. The current development of mobile augmented reality is mainly technology-oriented, mostly emphasises on technology advancement as a basis of measure while placing insufficient emphasis upon user experience. User-centric design is increasingly important in the design of mobile applications. As it is crucial to quantify and evaluate user experiences of AR application to gain insight in pivotal areas for future development, which this research proposes that the application of the Delphi-AHP method is capable of identifying those areas via five first-level indicators and 20 second-level indicators. This method is tested and verified with six model display applications, which discovered the most important first-level indicators to affect user experience is a system's functionality and its display.

Keywords: Augmented reality · User experience · Evaluation method

1 Introduction

Human-computer interaction is changing drastically alongside the novel innovations and roles in computers and mobile devices. "Augmented reality is a cutting-edge technology that allows the user to observe real-world objects supplemented with computer-generated objects" [1]. By transforming visual interactions away from 2D to 3D, from point touch to gesture, human-machine interaction is converted to a fusion of human-machine-environment interaction, where the change of display and interaction bring new challenges to user experience (UX) design.

Ames et al. [2] define user experience as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service". According to the same definition, UX includes the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments that occur before, during and after use of a system. The term "user experience" is used along with "usability" with various degrees of relativity. We consider UX to be a larger entity,

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2020, LNCS 12190, pp. 208–222, 2020.
https://doi.org/10.1007/978-3-030-49695-1_14

encompassing usability and including both pragmatic and hedonic aspects of a system as UX is considered a significant factor for interactive products and services [3].

The gradual maturity of AR technology brought prosperity in the AR application market, particularly in the relatively matured smartphone industry where a large variety of applications based on AR interaction emerge constantly. As the market saturates, UX has become the main factor of competitiveness for interactive products and services [4]. The improvements and measures of user experience of a product is a continues process of iterative optimisation, where the necessity to have a clear assessment and cognition of the current user experience is essential. Therefore, the study on how to evaluate the user experience is critical at gaining the first-hand advantage within an increasingly competitive market environment.

Baidu AI Interaction Design Institute divides mobile AR application into three categories according to the distance between user and virtual model, and the differences between interactive contents in AR scenes: 1) Human-based AR, the user interacts with the face and human body; 2) Space-based AR, the user interacts with scenes and objects in the space in which they are located, such as placing objects such as furniture, models, people/animals on the ground and interacting with them; 3) Environment-based AR, the user can experience and interact with large scenes in the physical world [5]. Due to the diversity of applications and its potentials, space-based AR is the most widely used application. Therefore, this paper will use space-based AR as the research object as it has the capability to merge and interact with virtual displays with real-world environments.

The rest of this paper is as follows: Sect. 2 introduces the concept of the mobile augmented reality and an overview of the relevant studies. Section 3 constructs the user experience evaluation system of AR applications via the Delphi-AHP methodology. Section 4 applies the system to evaluate the six model display applications, which justifies the validity of the system. Section 5 concludes the research and suggests future research directions.

2 Study Basis

2.1 Mobile Augmented Reality

AR is a technology to superimpose the virtual information with reality while allowing interactions between the user and virtual information and to enhance people's perception of the real world through supplementing reality with additional information and data. In 1970, Professor Ivan Sutherland of the Massachusetts Institute of Technology first came up with the term "Augmented Reality." In 1997, Ronald Azuma [1] of the University of Northern Carolina proposed a clearer definition of the concept of AR, which is a more widely known definition that suggests that AR should have the following three characteristics:

1) Combination of virtuality and reality: Virtual object and real-world display and the same visual space.
2) Real-time interaction: The user can naturally interact with virtual objects and their real surroundings in real-time.

3) 3D registration: Calculate the location of the camera so that virtual objects are properly placed in the real-world perception.

Mobile AR is experienced via a smartphone or a handheld device [1]. In mobile AR, the "reality" is the camera view and the "virtual" is the superimposed enhanced graphics. For mobile AR in this article, we have added one more component [6]:

4) Complete accessibility via mobile devices (e.g. cellular phones, tablets, etc.).: The user can experience AR anywhere and interact in any environment.

With the advancement in relevant technologies, mobile AR has become the most prosperous field in AR application. AR not only delivers an enhanced visual sense but also the profound influence of cognitive understanding. Previous cognitive understanding of surroundings is achieved through a combination of traditional ways, such as dictionaries, maps, instructions or field trips. However, this information can be presented directly and naturally via the mobile AR application, overwhelmingly simplifies the procedures and reduces the opportunity cost for users.

2.2 Relevant Study Status

The study of mobile AR mainly focuses on the development of technology, such as a variety of output devices and interactive displayers, as well as the process of identifying and tracking tangible objects [4]. Although the general framework to assess user experience in other fields are robust, but the theoretical framework and research knowledge related to user experience within modern fields such as mobile AR are highly insufficient [6, 7].

Only less than 10% of AR literature discusses topics upon user-centric designs and the methodologies to assess user experience in AR application. Currently, literature mainly focuses on three aspects: Assessing the usability of augmented reality with human factors considered; User experience in augmented reality, concentrating on certain specific applications; Augmented reality design and evaluation methods.

Kalahti [9] proposed a heuristic algorithm to assess the usability of augmented reality applications. Based on the literature review, we developed a preliminary version of the heuristic algorithm and was evaluated by four experts. As a result, six evaluation criteria were formed: 1) Interaction methods and controls, 2) Display of virtual objects, 3) Relationship between virtual objects and the real-world setting, 4) Information related to virtual objects, 5) Ease of operation, and 6) Physical user comfort.

Li et al. [10] set up the evaluation criteria of VR/AR teaching experience according to three aspects: User interface, Usability and Interactive design. The indices of user interface include: Information display, Menu setting, Geographical location, Progress access, Operation, Input, Visual field, Customisation; the indices of usability include: Repetition strategy, Expectation, Entry strategy, Recognition ability, Consistency; the interactive design indices include: Initial experience, Operating environment, Goal, Task, Control, Feedback, Progress and plot.

Pribeanu et al. [11] proposed a multidimensional model to measure the perceived quality of AR applications. The model is specified by a second-order factor (perceived quality) and three dimensions: Ergonomic quality, Study quality, and Hedonic quality.

The purpose of this model is to introduce the previous research to a coherent framework of the AR-based educational platforms and to provide guidance for researchers and practitioners.

Dunser et al. [8] combined the particularity and the operating background of AR with the user-centric design principle and proposed that the design of augmented reality should consider the following eight aspects: Affordance, Reducing cognitive overhead, Low physical effort, Learnability, User satisfaction, Flexibility of use, Responsiveness and feedback, Tolerance to faults and unintended mistakes.

The current study on the assessment of AR mainly reflects the user-centric design and focuses on evaluating usability, with little emphasis on user experience and emotions [12], with no design evaluation system for mobile AR applications. Considering the different displaying methods and functions of AR across different applications, these applications cannot be evaluated by one system. Therefore, this paper focuses on the study of the model display applications.

3 Construction of User Experience Evaluation System

3.1 Determination of Evaluation Indexes

Many experts and scholars have paid attention to the importance of user experience in augmented reality and have made corresponding research on the application of AR in different fields. In this paper, through literature analysis on AR, study of AR industry specifications, user experience framework, collection of a large number of evaluation indices from existing research results in related fields, and in combination with real experience feedbacks from AR users, a set of indicators has been preliminarily determined, including 5 first-level indicators and 20 second-level indicators (as shown in Table 1).

Table 1. Preliminarily determined indicators set

First-level indicators	Display	Interaction	Performance	Function	Aesthetic experience
Second-level indicators	• Physical rationality • Intuitive visualization • Model quality • Virtual-real fusion	• Consistency • Feedback • Sense of control • Flexibility • Fault tolerance and error proof • Easy to learn • Novelty of interaction • Fluency • Clear operation • Status visibility • Humanized help • Guidance	• Stability • Response speed • Matching accuracy • Recognition ability	• Satisfaction of user's requirements • Efficiency improvement • Model richness • Cognitive improvement	• Immersion • Entertainment • Enjoyment • Novelty • Attractiveness • Gratification

On this basis, the expert consultation table is drafted, and the Delphi method is used to conduct two rounds of consultation for the selected discipline experts. Nine experts who are engaged in human-computer interaction, user experience, AR related work and research with rich theoretical knowledge are invited to consult on the above index set through Delphi method. Those experts first filled in the form of "judgment basis" and "familiarity level" to identify their authority in this field. Then they evaluated those indicators according to five levels: very important, important, average, unimportant and very unimportant, and give comments. The index is modified per expert's comments to further improve the index system.

The results of the first-round consultation are shown in Table 2, with multiple coefficients of variation >0.2, showing that experts differ greatly in their evaluation of the same index. According to the data analysis results of the first-round expert consultation and experts' comments, the following modification has been made: nine secondary indices (Intuitive visualization, Consistency, Novelty of interaction, Fluency, Efficiency improvement, Model richness, Entertainment, Novelty, Attractiveness) are deleted; two indexes (Sense of control and Flexibility, Feedback and Status visibility) are combined; the description of two indexes are modified (with "Stability" changed to "System stability", and "Recognition ability" to "Recognition rate"); and two indexes (Stability of synthetic environment, Detail presentation) are added.

Table 2. Delphi results of the first round

	Mean value	Standard deviation	Coefficient of variation	Median
C1-1	4.333	0.816	0.188	5
C1-2	4.556	1.257	0.276	5
C1-3	4.222	0.916	0.217	4
C1-4	4.222	0.786	0.186	4
C2-1	4.556	0.685	0.150	5
C2-2	4.889	0.314	0.064	5
C2-3	4.667	0.667	0.143	5
C2-4	4.444	0.685	0.154	5
C2-5	4.444	0.831	0.187	5
C2-6	4.667	0.667	0.143	5
C2-7	3.333	1.155	0.346	3
C2-8	4.556	0.685	0.150	5
C2-9	4.444	0.831	0.187	5
C2-10	4.889	0.314	0.064	5
C2-11	4.556	0.497	0.109	5
C2-12	4.333	0.816	0.188	5
C3-1	4.889	0.314	0.064	5
C3-2	4.778	0.416	0.087	5
C3-3	4.333	0.943	0.218	5
C3-4	4.111	1.197	0.291	5

(*continued*)

Table 2. (*continued*)

	Mean value	Standard deviation	Coefficient of variation	Median
C4-1	4.556	0.956	0.210	5
C4-2	4.222	0.786	0.186	4
C4-3	3.222	1.227	0.381	3
C4-4	3.556	1.343	0.378	4
C5-1	4.778	0.416	0.087	5
C5-2	4.556	0.685	0.150	5
C5-3	4.444	0.831	0.187	5
C5-4	4.111	1.100	0.268	5
C5-5	4.444	0.956	0.215	5
C5-6	4.111	0.994	0.242	4

Table 3. Concordance coefficient of expert consultation opinions

	First round	Second round
Number of indexes	30	20
Concordance coefficient	0.401	0.629
Chi-Square	102.004	32.096
P	0.000	0.000

While the results of the first-round consultation are communicated back to the experts, the second round consultation is carried out. In the second round, the average number of secondary indicators is more than 4 points, and the full score ratio is more than 0.5. The expert concordance coefficient increased from 0.4 in the first round to 0.63 (as shown in Table 3), indicating that in the second round of expert consultation, the experts' opinions are more unified, and the degree of concordance is higher. Through the non-parametric test, the results show that there is a significant difference between the two rounds of expert consultation results. Therefore, experts have a high degree of concordance on the structure and indicators of the evaluation system, and the results are effective. Thus, the index system structure is determined, as shown in Fig. 1.

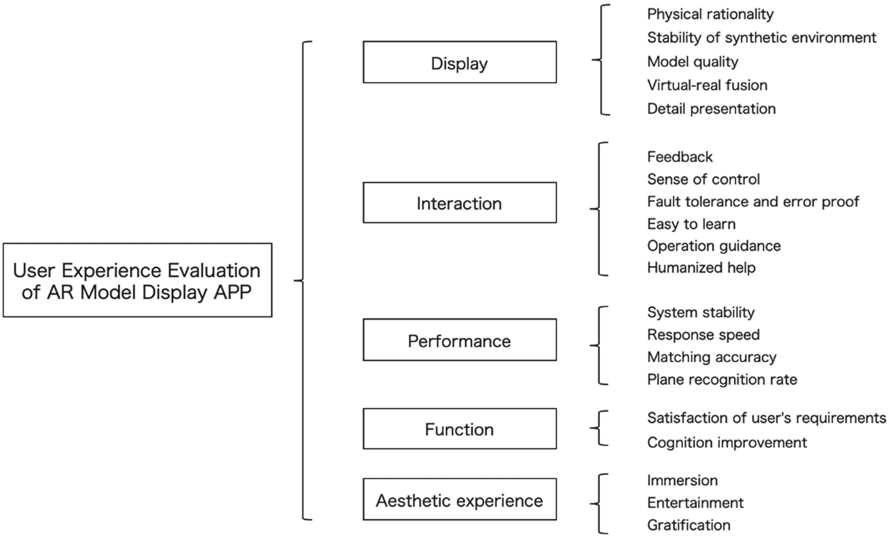

Fig. 1. Index system structure

3.2 Determination of Index Weight

The weight of the index is a representation of the importance of the index. The weight is calculated by comparing the importance of each index and then attaching the weight to it. Analytic Hierarchy Process (AHP) is a statistical system method proposed in the 1970s by Thomas L. Saaty, an American operational research scientist. It combines qualitative method with a quantitative method. AHP relies on the judgment of experts to obtain the priority scale [13], which uses the absolute judgment scale to compare. The absolute scale defines the relative importance of one indicator over another to a given objective (as shown in Table 4). Seven experts in the fields of augmented reality and human-computer interaction were invited to participate in the evaluation of the importance of those indicators.

Table 4. AHP scoring scale and meaning

Relative importance	Definition	Meaning
1	Equally important	Two goals are equally important
3	Slightly more important	one goal is slightly more important than the other
5	Quite more important	one goal is quite important than another
7	Obviously more important	one goal is obviously more important than another
9	Absolutely more important	one goal is absolutely more important than another
2, 4, 6, 8	Intermediate value of two adjacent judgments	Used during compromise

In this paper, the weight of the five first-level indicators of the display, interaction, performance, function and aesthetic experience is taken as an example of AHP calculation method. The specific steps are as follows:

① According to the scoring scale table of AHP, the comparison matrix A is formed by pairwise comparison of relevant elements.

$$A = \begin{bmatrix} 1 & 3 & 5 & 1 & 2 \\ 1/3 & 1 & 2 & 1/3 & 1/2 \\ 1/5 & 1/2 & 1 & 1/5 & 1/3 \\ 1 & 3 & 5 & 1 & 2 \\ 1/2 & 2 & 3 & 1/2 & 1 \end{bmatrix}$$

② Calculate the maximum eigenvalue and weight vector.

Set A = (aij) be n-order matrix, normalise each column of A to get B = (bij), and use formula (1) to calculate.

$$\bar{a}_{\mathrm{ij}} = \mathrm{a_{ij}} / \sum\nolimits_{k=1}^{n} a_{kj}, \mathrm{i}, \mathrm{j} = 1, 2, \ldots, \mathrm{n}. \tag{1}$$

$$B = \begin{bmatrix} 0.330 & 0.316 & 0.313 & 0.330 & 0.343 \\ 0.110 & 0.105 & 0.125 & 0.110 & 0.086 \\ 0.066 & 0.053 & 0.063 & 0.066 & 0.057 \\ 0.330 & 0.316 & 0.313 & 0.330 & 0.343 \\ 0.165 & 0.211 & 0.188 & 0.165 & 0.171 \end{bmatrix}$$

$$\tilde{w}_i = \sum\nolimits_{j=1}^{n} \bar{\mathrm{a}}_{\mathrm{ij}}, \mathrm{i} = 1, 2, \ldots., \mathrm{n}. \tag{2}$$

Use formula (2) to obtain the sum of lines:

$$C = \begin{bmatrix} 1.630 \\ 0.536 \\ 0.304 \\ 1.630 \\ 0.899 \end{bmatrix}$$

$$w_i = \tilde{w}_\iota / \mathrm{n}. \tag{3}$$

Use formula (3) to normalize C to obtain:

$$\begin{bmatrix} 0.326 \\ 0.107 \\ 0.061 \\ 0.326 \\ 0.180 \end{bmatrix}$$

Calculate the maximum eigenvalue:

$$\lambda_{max} = \frac{1}{n}\sum\nolimits_{n=1}^{n} \frac{(AW)_i}{W_i} = \left(\frac{1.630}{0.326} + \frac{0.536}{0.107} + \frac{0.304}{0.061} + \frac{1.630}{0.326} + \frac{0.899}{0.180}\right)/5 = 5.051$$

③ Calculate consistency index (CI): CI $= \frac{\lambda_{max}-n}{n-1} = \frac{0.051}{4} = 0.004$

④ By comparing the random consistency index table, RI = 1.12 is obtained by n = 5.

⑤ Calculate CR: CR $= \frac{CI}{RI} = \frac{0.004}{1.12} = 0.0034 < 0.1$

Seven experts conducted a consistency analysis on the comparison results of the first-level indicators. The CR values were all less than 0.1, so the consistency test was all passed. Seven experts take the weighted average value of each indicator's weight (as shown in Table 5) to obtain the comprehensive weight of the first level indicators. In view of the above method, data analysis is made on each influencing factors of the secondary indicators. After the weight of each index is obtained, the total weight is then calculated. The specific method is to multiply the judgment matrix of each index by the weight of its upper layer. The results are shown in Table 6.

Table 5. Weight results of first-level indicators

First level weight	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Weight
Display	0.3263	0.3103	0.5883	0.3600	0.4691	0.4030	0.3263	0.3103
Interaction	0.3263	0.2414	0.1491	0.2800	0.2010	0.2444	0.3263	0.2414
Performance	0.1799	0.1724	0.1491	0.2000	0.2010	0.1367	0.1799	0.1724
Function	0.1068	0.1724	0.0585	0.1200	0.0862	0.1367	0.1068	0.1724
Aesthetic experience	0.0607	0.1034	0.0549	0.0400	0.0427	0.0791	0.0607	0.1034
CR	0.3263	0.3103	0.5883	0.3600	0.4691	0.4030	0.3263	0.3103

It can be concluded from the index weight result that the "function" (0.3156) is the most important one in the first level index, and the "practicability of the product" is the aspect that users consider more. Besides, users are prone to use products that have practical functions, and the practicability will also affect whether users to use the product again. Even if the experience of augmented reality is good, users will still consider whether it is necessary to apply AR technology on a certain product and whether it can meet users' needs or improve their efficiency and cognition.

The "display" effect of the model (0.2443) is also an important dimension in the evaluation. The location, size, light and shadow effect, degree of refinement, stability, floating sensation and degree of integration with the environment all affect users' realistic sense of the model.

The experience immersion is also the aspect that users pay much attention to. The introduction of sound and the sound effect of feedback make the interaction more

Table 6. Index weight and meaning

First level index	Weight	Two level indexes	Explanations	Weight	Total weight
Display	0.2443	Physical rationality	The position where the model is placed. The shape, size, shadow, action and feedback of the model conform to the physical reality	0.2471	0.0604
		Stability of synthetic environment	The synthetic environment will not change physically with the change of time and space	0.2491	0.0609
		Model quality	The model style, level of sophistication, and the expected style of application are consistent	0.2965	0.0724
		Virtual-real fusion	The virtual scene and the real scene are well combined, and the light effects, shadows, and occlusion are properly handled	0.0787	0.0192
		Detail presentation	Provides a way to view model details	0.1286	0.0314
Interaction	0.1545	Feedback	Timely and obvious feedback, consistent with cognition	0.1959	0.0303
		Sense of control	Flexible and controllable operation, interactive operation is efficient and easy to control	0.2695	0.0416
		Fault tolerance and error proof	The bearing capacity of the wrong operation, e.g. the function of deletion and reset is provided in case where model is misplaced or cannot be recovered	0.0827	0.0128
		Easy to learn	Interactive operation conforms to cognition, and the interactive mode is easy to learn and remember	0.2049	0.0317
		Operation guidance	Provide guidance on operation in time	0.1925	0.0298
		Humanized help	Provide timely and visible help center, and provide operable improvement suggestions in case of abnormality	0.0546	0.0084
Performance	0.1359	System stability	The system can complete the task smoothly and stably	0.4056	0.0551
		Response speed	Including environment identification, loading speed of model, and adopting appropriate strategies to reduce users' perception of loading time	0.1171	0.0159
		Matching accuracy	The accuracy of position matching between virtual object and real space	0.2585	0.0351
		Plane recognition rate	Horizontal Plane recognition success rate	0.2188	0.0297
Function	0.3156	Satisfaction of user's requirements	The extent to which the function meets the user's needs	0.6513	0.2056
		Cognitive improvement	Enhance users' cognition after use	0.3487	0.1100
Aesthetic experience	0.1498	Immersion	To focus on the task in the current target situation	0.4087	0.0612
		Entertainment	Interesting and entertaining in terms of user experience	0.2724	0.0408
		Gratification	Satisfaction with senses and information	0.3189	0.0478

interesting and enhance the user's immersion. The stronger the sense of reality, the better the user's immersive experience.

4 Experimental Design

4.1 Methodology

Taking the existing AR model display products on the market as an example, user experience evaluation is conducted with the above evaluation index system to verify the rationality of model construction and the effectiveness of the practical application. According to the AR ranking, two categories of six AR applications in the top, middle and bottom of the ranking are selected as the evaluation objects, which are "education" apps (Assemblr, JigSpace, Homebrew Club) and "furniture" apps (Housecraft, IKEA Place, Chuangzaojia).

The experiment is divided into two parts: product experience and rating scale. First, the main test personnel introduced the experiment, showed the task scene and task content to the test subjects before experiencing each app. After understanding the task, the test subjects completed the task in order according to the task table. The experiment used an in-group design and in order to avoid the influence of learning effect on scores, the sequence of experimental products was processed according to Latin square design. After experiencing all the apps, the test subjects rate the six apps according to the indicators constructed above. To ensure that users understand the meaning of indicators, a specific explanation is made to each indicator. The score table uses a Likert 5-point scale to obtain users' experimental experience.

Experiment task: the experience of AR model display products mainly includes the following aspects: initial startup, scene building, model interaction and abnormal conditions. Five experiment tasks are designed to correspond to these four aspects (Table 7).

Table 7. Experimental tasks

Task number	Task content	Focus of test
1	Start the application, browse the app homepage and operation guide, and select a model	Initial startup (application startup, content guidance, model loading, technical preparation)
2	Identify the surrounding environment and place the model	Scene Building (environment identification and model placement)
3	Interact with the model, including displacement, scaling, rotation, selection, deletion	Model interaction
4	Greatly enlarge, reduce and move the model, and then restore it to its original position and state as much as possible	Abnormal status (model lost, model out of screen)
5	Get up and walk around the house, scan around with the camera and focus on the model again	Stability of synthetic environment

Subjects: after screening and invitation confirmation, a total of 14 users participated in the experiment, including 7 males and 7 females. The age is between 20 and 30. In order to ensure consistency of AR understanding among subjects, users who have not used AR-related products in their previous experience are selected as test subjects.

Device: the experimental device is an iPhone XR with the system of IOS 12.3.1. The experiment is conducted indoors under a Wi-Fi network.

4.2 Result

By scoring six AR products, six groups of data of display, interaction, performance, function, and aesthetic experience are obtained by synthetic weight. According to the results of variance analysis of each group of data (Table 8), the two-tailed significance of display, interaction, performance, function, and aesthetic experience are 0.000, 0.000, 0.013, 0.003, and 0.000 respectively, all less than 0.05, indicating that the five main evaluation indexes in the evaluation system can effectively distinguish the experience of different AR products, and further indicating that the evaluation indexes analysed in this paper have certain reference value and practical significance.

Table 8. Results of variance analysis

	Score						F	P
	Assemblr	JigSpace	Homebrew club	IKEA Place	Housecraft	Chuangzaojia		
Display	3.64	4.28	3.03	4.40	4.40	3.62	5.722	0.000**
Interaction	3.14	3.81	1.86	3.11	3.49	3.04	5.700	0.000**
Performance	3.95	4.13	3.10	4.04	4.27	3.43	3.327	0.013*
Function	3.21	4.21	2.58	3.87	4.29	3.17	4.310	0.003**
Aesthetic experience	2.91	3.89	1.73	3.66	4.06	3.20	3.327	0.000**

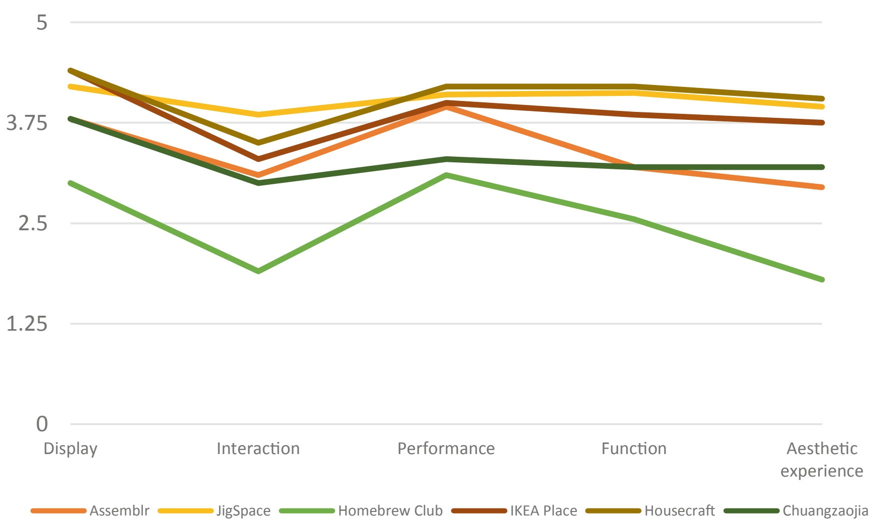

Fig. 2. Scoring results of six apps

In terms of display, Housecraft, IKEA Place, JigSpace perform well. Housecraft handles the textures and details of the models very well, and the models are solid, allowing for home furniture stacking. The effect of collision is also designed, and the sense of reality is better (Fig. 2).

In terms of interaction, there is a large gap between scores. JigSpace performed best because the interaction was cognitive, natural and controlled. In the case of model abnormality, it provides users with reset buttons. Bounds are set for zooming. However, the existing problems are that when the screen is out or the model is lost, there is no guide to retrieve them. Moreover, no operation guide or tutorials are provided during the process of scanning the plane and users are confused about what to do. Although Assembler gives a graphical representation of the operation in the interface, many users feel that the description is ambiguous, and the interaction cannot be successfully completed after attempts made in accordance to the guidelines. The lowest score was given to the Homebrew Club as interaction is slow, unstable and inconsistent with the cognition. Users expressed that they could not control it well after multiple and stressful attempts. The model can be added infinitely but cannot be deleted in the scene as the operational icons are too ambiguous.

In terms of performance, Housecraft scored the best, with faster speed in the process of identifying the environment and more accurate plane recognition. It provides operational guidance for identifying the plane, and gives feedback on the identification progress, reducing the user's perception of the loading time. In the experience process, it rarely shows an abnormal state and the test subject can successfully complete the given task, indicating a highly satisfying system stability.

In terms of functions, Housecraft scored well in furniture applications. It displays the projection of the furniture on the ground while placing the furniture. It well reflected the width, height and floor area of the furniture and helped users make purchase decisions. With strong practicability, it is well in line with the scene of the user's home design. Among educational applications, JigSpace allows users to understand how different things work via step-by-step procedures. Combined with animation and explanation, the complex structure is clear at a glance, where many users express that the function is very practical, enhancing their understanding of the application.

In terms of the overall experience, Housecraft performed the best as many users were amazed and fascinated during the experience. The sense of fun or immersion was good and many users expressed that they would like to use it again when they purchase furniture in the future. The Homebrew Club received the lowest score in terms of the overall experience, as it was not smooth in operation, unable to complete tasks at a satisfactory level, poor functionality, and had no practical value or value-adding for users.

5 Conclusions

This paper constructs the user experience evaluation system for the AR model display applications. Firstly, it collects the evaluation indices through analysing previous literature, initially establishing the hierarchical evaluation system. Through exercising the

Delphi method, 5 first-level indicators (Display, Interaction, Performance, Function and Aesthetic experience) and 20 second-level indicators are obtained. The index weight is then quantified by the analytic hierarchy process. After which, the system and its indicators are robustly tested through applying it to six mobile AR products in the market, verifying the validity of the main evaluation indices. The results suggest that function (0.32) and display (0.24), two second-level indices, have a greater influence over the user experience. Firstly, designers of relevant applications should prioritise its practicality, whether the use of AR technology is necessary, and whether the use of such technology can efficiently and effectively resolve challenges or result in cognitive improvement. Functional practicality affects the user's willingness to use and future possibility to reuse the application. Secondly, the display of the model is also an important dimension of evaluation. The location of the model, the size of the model, the light and shadow effects, the quality of the model, the stability of display, and the degree of virtual-real fusion with the environment will affect the sense of the reality of the model. By constructing a user experience evaluation system for the AR model display applications, the design elements and weights that affect the AR user experience are defined, providing informative insights to application designs of related applications. In addition, it can help the designer to adjust the design and evaluate each design's impact upon user experience. According to the conclusion of this paper, the goal of the augmented reality system design needs to be clear and precise. By considering the influence of various factors on the user experience, the user experience level of augmented reality applications will be comprehensively promoted, and the development of mobile augmented reality products is further initiated.

References

1. Azuma, R.T.: A survey of augmented reality. Teleoperators Virtual Environ. **6**(4), 355–385 (1997)
2. Ames, S., Wolffsohn, J., Mcbrien, N.: The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. Optom. Vis. Sci. **82**(3), 168–176 (2005)
3. Olsson, T., Saari, T., Lucero, A., Arrasvuori, J., Vaananen, K.: Reflections on experience-driven design: a case study on designing for playful experiences. In: DPPI, pp. 164–174 (2013)
4. Irshad, S., Rambli, D.: User experience of mobile augmented reality: a review of studies, pp. 125–130. IEEE (2014)
5. Baidu Al Interaction Design Lab: A guide to Mobile terminal space level AR interaction design. https://mp.weixin.qq.com/s/2vZBuCxxPz_yJxu38PO45g. Accessed 12 Sept 2019
6. Keating, G., Guest, D., Konertz, A., Padovani, N., Villa, A.: Design, User Experience, and Usability, Theory, Methods, Tools and Practice, pp. 135–141. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21708-1
7. Olsson, T., Kärkkäinen, T., Lagerstam, E., Ventä-Olkkonen, L.: User evaluation of mobile augmented reality scenarios. J. Ambient Intell. Smart Environ. **4**(1), 29–47 (2012)
8. Dünser, A., Grasset, R., Billinghurst, M.: A survey of evaluation techniques used in augmented reality studies, pp. 123–145. ACM (2008)

9. Kalalahti, J.: Developing usability evaluation heuristics for augmented reality applications. Lutpub.lut.fi. (2015). https://lutpub.lut.fi/handle/10024/103081
10. Li, X.P., Zhao, F.N., Zhang, S.G.: Research on the design and application of VR/AR teaching experience. In: China Educational Technology, p. 10 (2018)
11. Pribeanu, C., Balog, A., Iordache, D.: Measuring the perceived quality of an AR-based learning application: a multidimensional model. Interact. Learn. Environ. **25**(4), 482–495 (2016)
12. Zhou, G.Z.: Research on the construction and application of mobile AR user experience model. In: PLA Information Engineering University (2013)
13. Saaty, T.L.: Decision making with the analytic hierarchy process. Serv. Sci. **1**(1), 83–98 (2008)